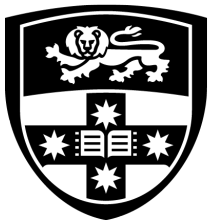
A Deep Learning Framework for Real-Time Cancer Targeting in Radiation Therapy



THE UNIVERSITY OF
SYDNEY

Adam Mylonas

Faculty of Medicine and Health
The University of Sydney

A thesis submitted to fulfil requirements for
the degree of *Doctor of Philosophy*

May 2025

Statement of Originality

This is to certify that to the best of my knowledge, the content of this thesis is my own work. This thesis has not been submitted for any degree or other purposes.

I certify that the intellectual content of this thesis is the product of my own work and that all the assistance received in preparing this thesis and sources have been acknowledged.

26 May, 2025

Adam Mylonas

Abstract

Growing evidence highlights the detrimental effects of underdosing tumours and overdosing organs at risk during high dose radiation therapy, emphasising the need for precision in treatment delivery. Tumour motion, due to normal physiological processes, is one factor that compromises accuracy. To address this, real-time motion management technologies have been developed to continuously monitor the tumour position. These technologies, though effective, are either prohibitively expensive or require the implantation of fiducial markers. These barriers were highlighted in a recent international survey which revealed that 71% of responding centres wish to implement real-time motion management for another treatment site but are limited by resources and capacity.

This thesis presents the first large-scale proof-of-principle for x-ray-based markerless segmentation in globally available radiation therapy systems. It provides an important step towards making real-time motion management treatments accessible for all patients, eliminating the need for expensive, dedicated equipment or fiducial marker implantation. The proposed markerless approach relies solely on x-ray images acquired during treatment which in principle, covers the majority of linear accelerators, thus overcoming the resource and capacity barriers to real-time motion management.

The first two studies of this thesis investigate a deep learning framework for prostate (Chapter 3) and pancreas head (Chapter 4) markerless segmentation. The deep learning framework was found to be highly accurate across a large cohort of patients. The third study investigates the use of residual contrast agents from liver cancer chemotherapy as a surrogate for real-time motion monitoring (Chapter 5). The approach was found to successfully track the contrast agents for a cohort of patients with liver cancer. Chapter 6 describes a software application developed as part of this thesis for ground truth data labelling. The thesis concludes by summarising the key findings and exploring future research directions for deep learning-based markerless tumour tracking.

Acknowledgements

First, I want to express my gratitude to my supervisors Doan Trang Nguyen, Paul Keall, Jeremy Booth, and Andre Kyme.

Trang, thank you for your support and guidance over the past seven years since I joined Image X as a second-year undergraduate student, knowing little about medical physics. Your passion and dedication to your research inspired me to pursue a PhD in medical physics. Your mentorship has been invaluable from my early days as a summer research student to now as a PhD candidate. I am especially grateful that, despite your demanding role as CEO at SeeTreat, you have continued to make time to help me navigate the complexities of research and encouraged me through each challenge.

Paul, it has been an honour to work alongside a world-class academic and to learn from the very best. Thank you for the wisdom and insight you've shared over the years. I appreciate all of the thought-provoking questions and positivity you have brought to my research, which have driven me to become the best researcher and communicator I could be. It has truly been a privilege to be part of the incredible Image X Institute, which you have shaped into the inspiring team it is today.

Jeremy, I am grateful for the time you've taken out of your busy clinical schedule to support my research. Your guidance, along with the clinical context you've provided, has been instrumental in shaping my work. Thank you for welcoming me into the medical physics team over the years. It has been a privilege to collaborate with your team to solve these challenging problems and ensure that our work will ultimately benefit patients.

Andre, although you joined as a supervisor partway through my PhD, you have been a valuable part of my journey from the beginning. Your support enabled me to enter the field of medical physics and discover my passion for research, and I am grateful for the fresh insights you have brought to my work. I am also thankful for the opportunity to work alongside you in your teaching team and witness your passion for teaching.

Thank you to everyone at Image X for making my time there so enjoyable. Having spent a significant part of my life at Image X, I couldn't have chosen a better place to begin my journey in research. While there are too many people to name, I am grateful to everyone who has reviewed my papers, abstracts, and presentations. In particular, I would like to thank **Marco Mueller** and **Mark Gardner** for their invaluable contributions to our deep learning markerless tracking research. A special shout-out to the Image X Run Club—I've really enjoyed our Tuesday runs, which have helped me rediscover my passion for running.

I would like to thank my collaborators at the Royal North Shore Hospital, especially **Ryan Brown** and **Maegan Stewart**, for your time and effort in collecting the data used in this thesis, as well as for your invaluable clinical guidance. I am also grateful to the teams at the Princess Alexandra Hospital, Calvary Mater Newcastle Hospital, and Olivia Newton-John Cancer Wellness & Research Centre for your contributions to the ROCK-RT clinical trial. I appreciate your contributions to this research, despite your busy clinical schedules.

Finally, my deepest gratitude goes to my wife, **Caroline**, my parents, **Peter** and **Rose**, and my friends, for your steadfast support throughout my PhD and beyond. Caroline, thank you for selflessly supporting me throughout my PhD journey. Words are insufficient to express my love and gratitude to you.

Funding Acknowledgements

The research reported in this thesis was supported by the Australian Government Research Training Program Fee Offset, the University of Sydney Postgraduate Award, and the Cancer Institute NSW Translational Program Grant Supplementary Scholarship awarded to the PhD candidate.

Publications

The work in this thesis includes results from previously published work described below:

Chapter 2 was published as "**Mylonas, A.**, Booth, J., & Nguyen, D. T. A review of artificial intelligence applications for motion tracking in radiotherapy. *Journal of Medical Imaging and Radiation Oncology*, **65**, 596-611 (2021)." I completed the literature review and prepared the manuscript with guidance from the co-authors.

Chapter 3 is under review as "**Mylonas, A.**, Li, Z., Mueller, M., Booth, J., Brown, R., Gardner, M., Kneebone, A., Eade, E., Keall, P., & Nguyen, D. T. Patient-specific prostate segmentation in kilovoltage images for radiation therapy intrafraction monitoring via deep learning. *Communications Medicine* (submitted)." I developed the deep learning implementation and designed the experiments, and I prepared the manuscript with guidance from the co-authors.

Appendix A includes a publication on which I am featured as a co-author, published as "Plant, N., **Mylonas, A.**, Sengupta, C., Nguyen, D. T., Silvester, S., Pryor, D., Greer, P., Lee, Y. Y. D., Ramachandran, P., Seshadri, V., Trada, Y., Khor, R., Wang, T., Hardcastle, N., & Keall, P. Radio-opaque contrast agents for liver cancer targeting with KIM during radiation therapy (ROCK-RT): An observational feasibility study. *Radiation Oncology*, **19**, 139 (2024)." I assisted in developing the protocol and manuscript, and I generated the results presented in the publication.

Appendix B includes a publication on which I am featured as a co-author, published as "Gardner, M., Bouchta, Y.B., **Mylonas, A.**, Mueller, M., Cheng, C., Chlap, P., Finnegan, R., Sykes, J., Keall, P.J. & Nguyen, D.T., Realistic CT data augmentation for accurate deep-learning based segmentation of head and neck tumors in kV images acquired during radiation therapy. *Medical Physics*, **50**, 4206-4219 (2023)." The deep learning framework developed in this thesis was used in this publication for head and neck cancer. I assisted with the initial implementation of the deep learning model.

Appendix C includes a publication on which I am featured as a co-author, published as "Madden, L., Ahmed, A., Stewart, M., Chrystall, D., **Mylonas, A.**, Brown, R., Nguyen, D.T., Keall, P. & Booth, J., CBCT-DRRs superior to CT-DRRs for target-tracking applications for pancreatic SBRT. *Biomedical Physics & Engineering Express*, **10**, 035039 (2024)." I was involved in testing several algorithms within the deep learning framework.

The following works were published but not included in this thesis:

Chrystall, D., **Mylonas, A.**, Hewson, E., Martin, J., Keall, P., Booth, J., & Nguyen, D. T. Deep learning enables MV-based real-time image guided radiation therapy for prostate cancer patients. *Physics in Medicine & Biology*, **68**, 095016 (2023)

Ahmed, A.M., Gargett, M., Madden, L., **Mylonas, A.**, Chrystall, D., Brown, R., Briggs, A., Nguyen, T., Keall, P., Kneebone, A. & Hruby, G. Evaluation of deep learning based implanted fiducial markers tracking in pancreatic cancer patients. *Biomedical Physics & Engineering Express*, **9**, 035008 (2023)

Authorship Declaration

In addition to the statements above, in cases where I am not the corresponding author of a published item, permission to include the published material has been granted by the corresponding author.

26 May, 2025

Adam Mylonas

As supervisor for the candidature upon which this thesis is based, I can confirm that the authorship attribution statements above are correct.

26 May, 2025

Paul Keall

Contents

1

Introduction

Cancer is one of the most significant global health challenges of modern society. Responsible for nearly one in six deaths worldwide in 2022, cancer is projected to result in 35 million new cases annually by 2050.[1] The increasing incidence and poor treatment outcomes highlight the need for improved cancer therapies. This thesis will focus on cancers of the prostate, liver, and pancreas. Prostate cancer was the fourth most common cancer and eighth leading cause of cancer-related deaths in 2022, with over 1.4 million new cases and almost 400,000 deaths.[1] Although less common, liver and pancreatic cancer are the third and sixth leading causes of cancer-related deaths, respectively, due to their higher mortality rates.[1]

Radiation therapy is an effective cancer treatment for curative and palliative purposes, and can be used in isolation or as part of a multi-modal treatment approach. It is estimated that one in every two cancer patients could benefit from radiation therapy.[2] However, normal physiological motion during treatment, known as intrafraction motion, poses a challenge for radiation therapy and will be addressed in this thesis. Motion can compromise tumour control in radiation therapy, potentially reducing treatment effectiveness and simultaneously increasing radiation toxicity to healthy tissue.

1.1 Motion During Radiation Therapy

During treatment, the tumour and surrounding anatomy are dynamically moving as a consequence of normal physiological processes. Different treatment sites are affected to varying

degrees of intrafraction motion by factors such as breathing, cardiovascular activity, digestion, bladder filling, and musculoskeletal movement. Intrafraction motion of the prostate is typically less than 3 mm but can exceed 10 mm.[3–5] There is an increased likelihood of prostate displacement with elapsed time.[3,4] Intrafraction motion of the liver and pancreas are considerably larger, up to 40 mm, on account of diaphragm motion from respiration.[6,7] While intrafraction motion may seem negligible, it can have a substantial impact on tumour control and side effects.[8–10]

For effective treatments, the spatial accuracy of the radiation beam is crucial to ensure that cancerous cells are killed, while sparing healthy tissue. Figure 1.1 illustrates intrafraction motion of the prostate caused by bowel gas. Without motion management, the prostate may not receive the planned dose and the healthy tissue of the rectum may receive excessive dose. The importance of real-time motion management has led to technological advancements including the adaption of conventional linear accelerators (linacs) and the development of specialised treatment machines, such as magnetic resonance imaging (MRI) and robotic linacs.

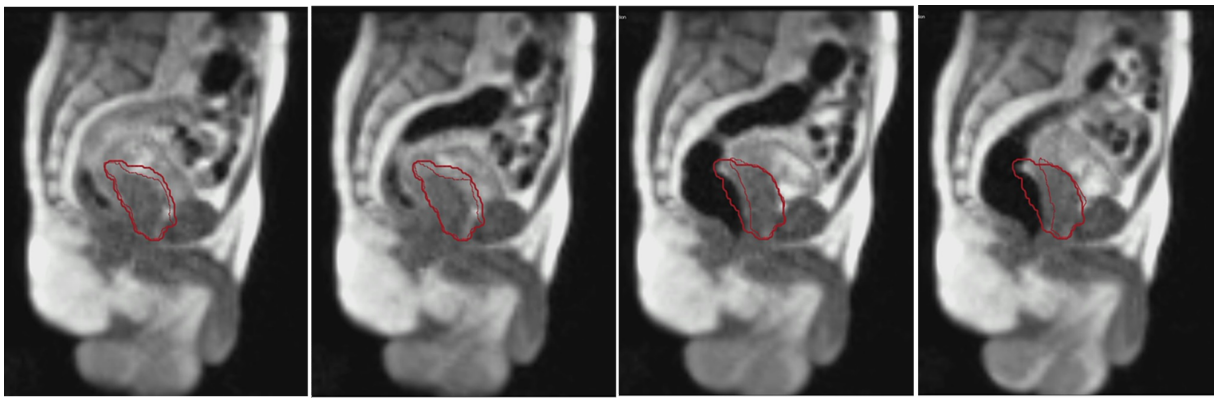


Fig. 1.1 | Intrafraction motion of the prostate. Sequential MRI images demonstrating intrafraction motion of the prostate due to rectal distention. The prostate position (thin red line) is compared against the planned treatment position (thick red line). Reprinted from Hegde et al.[11]

In radiation therapy, there are three main volumes which are crucial in treatment planning: the Gross Tumour Volume (GTV), Clinical Target Volume (CTV), and Planning Target Volume (PTV). The GTV represents the tumour that can seen, palpated, or imaged, while the CTV encompasses the GTV and accounts for microscopic disease. The PTV contains the CTV with additional margins to account for setup uncertainties and patient motion. Ideally, the CTV to PTV margin can be reduced significantly if target motion is managed, leading to more dose conformity to the target and less radiation dose to healthy tissue.

The clinical motivator of real-time motion management is to ensure dose coverage of the target. Figure 1.2 illustrates the effect of motion on dose distributions, highlighting the deterioration in dose coverage with no motion management. However, with motion management there is closer alignment between the planned and delivered dose. Lovelock et al.[12] found that the dose in 10% of prostate patients failed to meet clinical PTV coverage requirements in the absence of motion management. For liver radiation therapy without motion management, Nankali et al.[13] observed that the dose coverage deteriorated with the magnitude of the geometric error. Karava et al.[8] showed a significant improvement in tumour coverage and sparing of organs at risk (OARs) for pancreatic cancer when using tumour tracking.

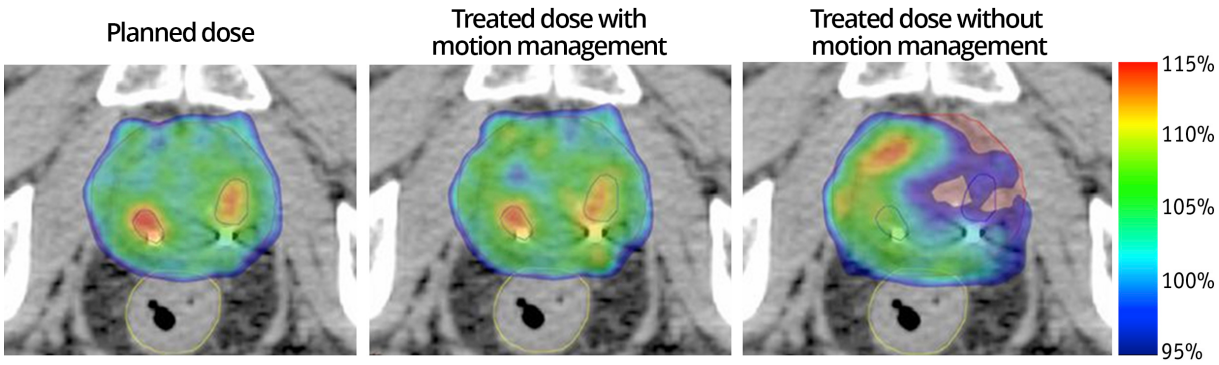


Fig. 1.2 | The effect of motion on dose distributions. The planned dose, treated dose with motion management, and treated dose without motion management for the prostate (red). Adapted from Colvill et al.[14]

An advantage of improved target coverage through motion management is the ability to reduce treatment margins. A lung cancer clinical trial found that motion tracking enabled a reduction of the PTV by 26.4% while maintaining coverage to the GTV.[15,16] Figure 1.3 shows the smaller margin that can be achieved when using motion tracking. A pancreatic cancer study found that MRI-guidance enabled a PTV margin reduction from 3–5 mm to 2–3 mm.[17] Similarly, motion guidance demonstrated the ability to reduce prostate cancer PTV margins to 2 mm.[9,18]

There are two main advantages of reducing treatment margins. First, a reduction in treatment margins enables dose escalation, which has shown improved treatment outcomes. For lung cancer, high dose treatments demonstrated superior local control of the primary disease.[19] For liver cancer, the elimination of motion allowed for dose escalation resulting in increased tumour control probability.[10] Due to the increased precision required for dose escalation, a white paper by the American Society for Radiation Oncology emphasises the necessity of image guidance and motion management for dose escalated treatments.[20]

The other benefit of reduced treatment margins is a reduction of radiation toxicity. Improved targeting of the cancer means reduced radiation dose to the surrounding healthy OARs, thereby minimising radiation toxicity. The MIRAGE clinical trial investigated the reduction of PTV margins from 4 mm using computed tomography (CT) guidance to 2 mm using MRI-guided radiation therapy.[9] MRI guidance reduced $\geq$ grade 2 genitourinary toxicities by 19.0% and gastrointestinal toxicities by 10.5%.[9] Moreover, there were no grade 3 genitourinary toxicities, and no gastrointestinal toxicities $\geq$ 2 in the MRI arm.[9]

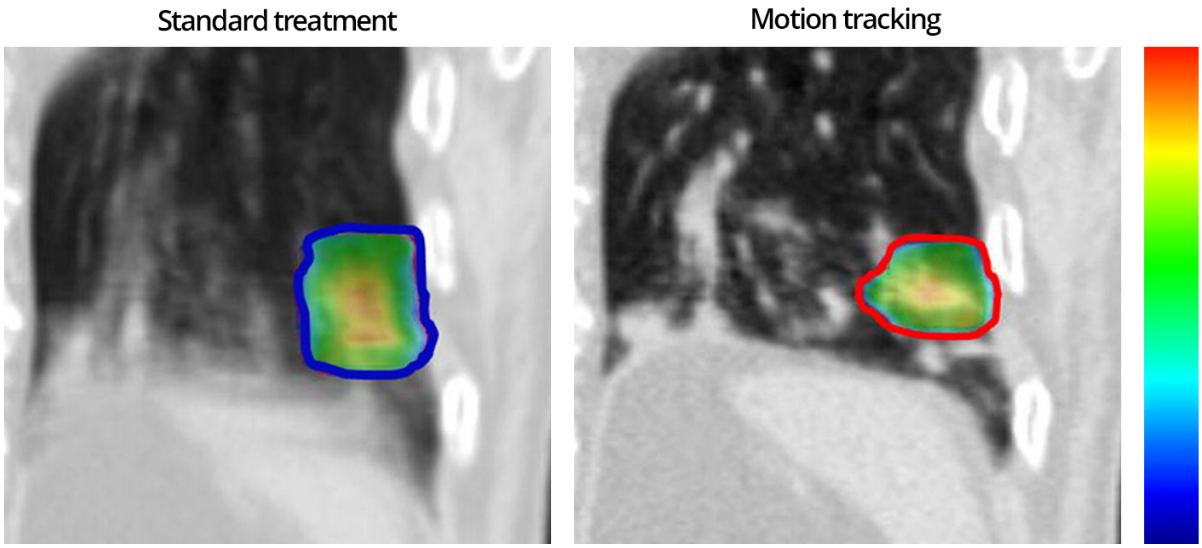


Fig. 1.3 | Example of reduced margins with motion tracking. Motion tracking allows for the reduction of margins and a more targeted approach. Adapted from Booth et al.[16]

1.2 Motivation and Rationale of this Thesis

For standard linacs, the low radiographic contrast of treatment sites such as the prostate, pancreas, and liver necessitates the implantation of fiducial markers (Fig. 1.4). Fiducial markers act as a surrogate of the tumour and can be used for real-time motion monitoring. However, the invasive implantation procedure can delay treatments, increases costs, and introduces additional risks to the patient. Treatment delays result from both the implantation procedure itself and a waiting period for the markers to stabilise prior to treatment planning.[21] However, the position of the markers may not remain stable and can migrate over the course of treatment.[22,23] Additionally, CT image artefacts from metal markers can interfere with treatment planning.[24] Finally, there are risks associated with marker implantation that vary by treatment site, including infection, haematuria, pneumothorax, bleeding, and discomfort.[25–27]

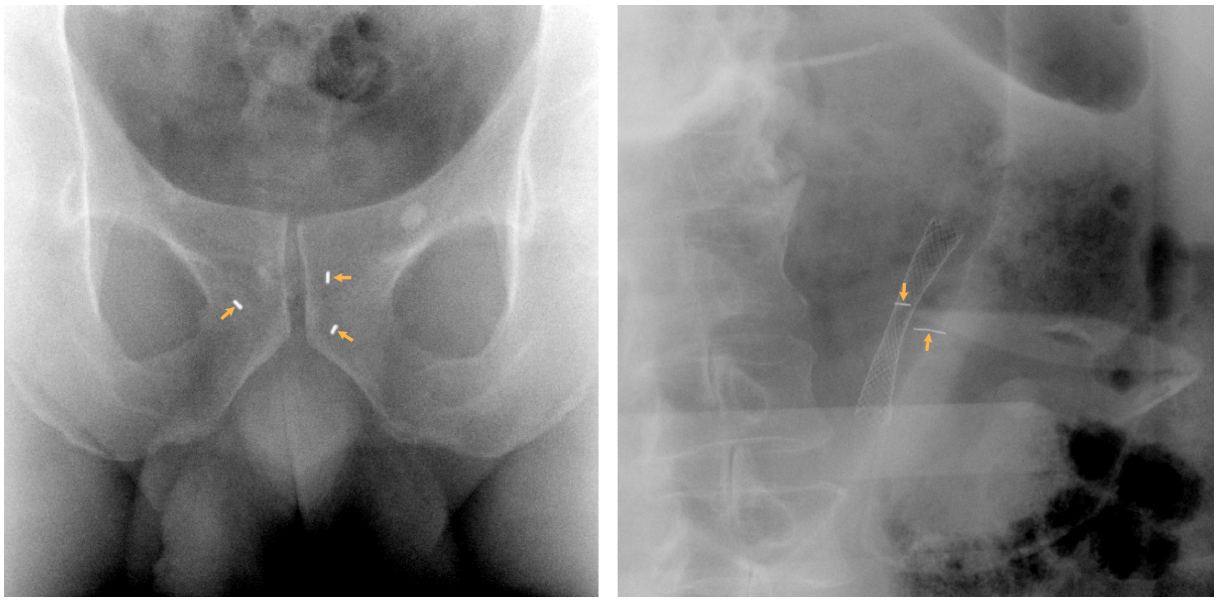


Fig. 1.4 | Implanted fiducial markers. Examples of fiducial markers implanted in the prostate (left) and pancreas (right). The target is not visible in these intrafraction kilovoltage images, requiring the implantation of fiducial markers.

Specialised technologies, such as the MRI-linac, can track the tumour and surrounding organs without fiducial markers. However, MRI-linacs are an expensive treatment option for patients and are not widely available compared to standard linacs.[28] An international survey found that 71% of radiation therapy centres wish to expand real-time motion management to additional treatment sites but are limited by resources and capacity.[29] These survey results highlight the demand for inexpensive motion management solutions that can be integrated into existing systems and workflows. Ideally, real-time motion monitoring should not require any additional procedures or hardware. A markerless-based approach implemented on standard linacs would make real-time motion monitoring accessible to all patients without additional hardware.

1.3 Thesis Aims

The overarching objective of this thesis is to develop the first real-time markerless tumour tracking framework for prostate, pancreatic, and liver radiation therapy. It was hypothesised that deep learning can be used to accurately track the target organ without fiducial markers on a standard linac. The specific aims of this thesis are:

(i) To conduct the first large-scale proof of principle for x-ray-based markerless prostate segmentation, suitable for clinical implementation on standard linacs.

(ii) To explore the potential of x-ray-based markerless pancreas head segmentation, focusing on accuracy and identifying technical challenges.

(iii) To investigate the feasibility of using residual contrast agents from chemotherapy for real-time markerless liver cancer tracking.

1.4 Thesis Structure

Chapter 2 provides an overview of radiation therapy and real-time motion management on a standard linac. The second part of Chapter 2 presents a literature review of artificial intelligence (AI) applications for real-time motion management, as of March 2021. Chapter 3 presents the first study of this thesis, detailing the deep learning framework and its application to prostate tracking, using data from the SPARK (NCT02397317[30]) and OPTIMAL (NCT03386045[31]) clinical trials. This approach is extended to pancreatic cancer in Chapter 4 by tracking the pancreas head using data from the Span-C clinical trial (NCT03505229[32]). Chapter 5 investigates the use of residual contrast agents from chemotherapy to track the liver using data from the ROCK-RT clinical trial (NCT05169177[33]). Chapter 6 describes a software application designed for data visualisation and labelling ground truth data for deep learning models. Finally, Chapter 7 summarises the key findings and discusses future directions of research for deep learning-based markerless tumour tracking.

1.5 References

1. Bray, F. et al. Global cancer statistics 2022: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA: a cancer journal for clinicians* **74**, 229–263 (2024).

2. Barton, M. B. et al. Estimating the demand for radiotherapy from the evidence: a review of changes from 2003 to 2012. *Radiotherapy and oncology* **112**, 140–144 (2014).

3. Langen, K. M. et al. Observations on real-time prostate gland motion using electromagnetic tracking. *International Journal of Radiation Oncology* Biology* Physics* **71**, 1084–1090 (2008).

4. Su, Z., Zhang, L., Murphy, M. & Williamson, J. Analysis of prostate patient setup and tracking data: potential intervention strategies. *International Journal of Radiation Oncology* Biology* Physics* **81**, 880–887 (2011).

5. Litzenberg, D. W. et al. Influence of intrafraction motion on margins for prostate radiotherapy. *International Journal of Radiation Oncology* Biology* Physics* **65**, 548–553 (2006).

6. Minn, A. Y. et al. Pancreatic tumor motion on a single planning 4D-CT does not correlate with intrafraction tumor motion during treatment. *American journal of clinical oncology* **32**, 364–368 (2009).

7. Worm, E. S., Høyer, M., Fledelius, W. & Poulsen, P. R. Three-dimensional, time-resolved, intrafraction motion monitoring throughout stereotactic liver radiation therapy on a conventional linear accelerator. *International Journal of Radiation Oncology* Biology* Physics* **86**, 190–197 (2013).

8. Karava, K. et al. Potential dosimetric benefits of adaptive tumor tracking over the internal target volume concept for stereotactic body radiation therapy of pancreatic cancer. *Radiation Oncology* **12**, 1–9 (2017).

9. Kishan, A. U. et al. Magnetic resonance imaging–guided vs computed tomography–guided stereotactic body radiotherapy for prostate cancer: the MIRAGE randomized clinical trial. *JAMA oncology* **9**, 365–373 (2023).

10. Gargett, M., Haddad, C., Kneebone, A., Booth, J. & Hardcastle, N. Clinical impact of removing respiratory motion during liver SABR. *Radiation Oncology* **14**, 1–9 (2019).

11. Hegde, J. V. et al. Magnetic resonance imaging guidance mitigates the effects of intrafraction prostate motion during stereotactic body radiotherapy for prostate cancer. *Cureus* **10** (2018).

12. Lovelock, D. M., Messineo, A. P., Cox, B. W., Kollmeier, M. A. & Zelefsky, M. J. Continuous monitoring and intrafraction target position correction during treatment improves target coverage for patients undergoing SBRT prostate therapy. *International Journal of Radiation Oncology* Biology* Physics* **91**, 588–594 (2015).

13. Nankali, S. et al. Geometric and dosimetric comparison of four intrafraction motion adaptation strategies for stereotactic liver radiotherapy. *Physics in Medicine & Biology* **63**, 145010 (2018).

14. Colvill, E. et al. Multileaf collimator tracking improves dose delivery for prostate cancer radiation therapy: results of the first clinical trial. *International Journal of Radiation Oncology* Biology* Physics* **92**, 1141–1147 (2015).

15. Booth, J. et al. MLC tracking for lung SABR is feasible, efficient and delivers high-precision target dose and lower normal tissue dose. *Radiotherapy and Oncology* **155**, 131–137 (2021).

16. Booth, J. T. et al. The first patient treatment of electromagnetic-guided real time adaptive radiotherapy using MLC tracking for lung SABR. *Radiotherapy and Oncology* **121**, 19–25 (2016).

17. Fasala, A. et al. Intrafraction motion and impact of margin reduction for MR-Linac online adaptive radiotherapy for pancreatic cancer treatments. *Journal of Medical Radiation Sciences* (2024).

18. Goddard, L., Jeong, K., Tang, J., Garg, M. & Tomé, W. A. Reducing PTV margins for prostate SBRT with motion compensation and gating techniques. *Journal of Applied Clinical Medical Physics* **24**, e13861 (2023).

19. Ball, D. et al. Stereotactic ablative radiotherapy versus standard radiotherapy in stage 1 non-small-cell lung cancer (TROG 09.02 CHISEL): a phase 3, open-label, randomised controlled trial. *The Lancet Oncology* **20**, 494–503 (2019).

20. Solberg, T. D. et al. Quality and safety considerations in stereotactic radiosurgery and stereotactic body radiation therapy: Executive summary. *Practical radiation oncology* **2**, 2–9 (2012).

21. Bertholet, J. et al. Real-time intrafraction motion monitoring in external beam radiotherapy. *Physics in medicine & biology* **64**, 15TR01 (2019).

22. Bhagat, N. et al. Complications associated with the percutaneous insertion of fiducial markers in the thorax. *Cardiovascular and interventional radiology* **33**, 1186–1191 (2010).

23. Kitamura, K. et al. Registration accuracy and possible migration of internal fiducial gold marker implanted in prostate and liver treated with real-time tumor-tracking radiation therapy (RTRT). *Radiotherapy and oncology* **62**, 275–281 (2002).

24. Osman, S. O. et al. Fiducial markers visibility and artefacts in prostate cancer radiotherapy multi-modality imaging. *Radiation Oncology* **14**, 1–13 (2019).

25. Kothary, N. et al. Safety and efficacy of percutaneous fiducial marker implantation for image-guided radiation therapy. *Journal of vascular and interventional radiology* **20**, 235–239 (2009).

26. Loh, J. et al. Infections after fiducial marker implantation for prostate radiotherapy: are we underestimating the risks? *Radiation Oncology* **10**, 1–5 (2015).

27. Igdem, S. et al. Implantation of fiducial markers for image guidance in prostate radiotherapy: patient-reported toxicity. *The British Journal of Radiology* **82**, 941–945 (2009).

28. van Herk, M., McWilliam, A., Dubec, M., Faivre-Finn, C. & Choudhury, A. Magnetic resonance imaging–guided radiation therapy: a short strengths, weaknesses, opportunities, and threats analysis (2018).

29. Anastasi, G. et al. Patterns of practice for adaptive and real-time radiation therapy (POP-ART RT) part I: Intra-fraction breathing motion management. *Radiotherapy and Oncology* **153**, 79–87 (2020).

30. Keall, P. et al. Stereotactic prostate adaptive radiotherapy utilising kilovoltage intrafraction monitoring: the TROG 15.01 SPARK trial. *BMC cancer* **17**, 1–7 (2017).

31. Royal North Shore Hospital. Optimal Prostate Study. `https://clinicaltrials.gov/study/NCT03386045` (2023).

32. Royal North Shore Hospital. Span-C-SBRT for Pancreatic Cancer (Span-C). `https://clinicaltrials.gov/study/NCT03505229` (2023).

33. Plant, N. et al. Radio-opaque contrast agents for liver cancer targeting with KIM during radiation therapy (ROCK-RT): an observational feasibility study. *Radiation Oncology* **19**, 139 (2024).

2

Background and Literature Review

This chapter is divided into two parts. Part 1 reviews the fundamental principles of external beam radiation therapy and its role in cancer treatment. We introduce the anatomy of a standard linac commonly used in most clinics and the typical workflow involved in treatment planning and delivery. Given the focus of this thesis on motion management, we present an overview of motion management techniques on a standard linac, highlighting current challenges.

Part 2 critically surveys the prior work on AI applications for marker-based and markerless motion management in radiation therapy. The literature review covers publications up to March 2021 and was published as "**Mylonas, A.**, Booth, J., & Nguyen, D. T. A review of artificial intelligence applications for motion tracking in radiotherapy. *Journal of Medical Imaging and Radiation Oncology*, **65**, 596-611 (2021)."[1]

2.1 External Beam Radiation Therapy

External beam radiation therapy (EBRT) is the most common form of radiation therapy,[2] and it is estimated that half of all patients with cancer could benefit from it.[3] EBRT aims to control tumour growth by using a linac to deliver ionising radiation to a localised area to damage and kill cancer cells. As the ionising radiation passes through tissues, it interacts with cells, depositing energy that causes direct and indirect damage to the deoxyribonucleic acid (DNA). Damaging the DNA blocks the ability of cells to divide and proliferate further.[4]

The challenge of radiation therapy is to maximise the radiation dose to cancer cells while minimising the radiation dose to healthy cells. As the radiation must pass through healthy tissue to reach the cancer, it is not possible to completely spare the healthy tissue. However, normal cells can typically repair damage more efficiently than cancer cells due cancer cells having defective DNA repair mechanisms that more readily result in cell death.[2,5]

Most patients receiving EBRT are treated on a standard clinical linac, as shown in Figure 2.1. Clinical linacs consist of a gantry rotating around the patient on the treatment couch. The treatment head on the gantry generates megavoltage (MV) radiation, which can be captured by the electric portable imaging device (EPID) to obtain images from the treatment beam. Within the treatment head are jaws and a multileaf collimator (MLC) which are used to collimate the treatment beam. Modern clinical linacs are fitted with an onboard kilovoltage (kV) imaging system perpendicular to the treatment beam. The kV system consists of a source and detector which is used for patient positioning and verification on the day of treatment.

Once a patient has been prescribed radiation therapy by their radiation oncologist, they will receive a planning CT in the same position as for treatment, including any necessary immobilisation devices. Clinicians will contour the tumour volume, treatment volumes, and surrounding OARs. The treatment plan will then be created based on the prescription. This process involves optimising the dose distribution using inverse planning[6] with parameters that include the fractionation rate, gantry rotation, dose rate, number of fields, and field shaping using the MLC.

Once the plan has been approved, the patient will receive the treatment over multiple daily sessions, known as fractions. Fractionation helps to maximise tumour damage while minimising healthy tissue damage by providing time for healthy cells to repair DNA damage.[7] Additionally, fractionation allows tumour cells to cycle into a more radio-sensitive phase of the

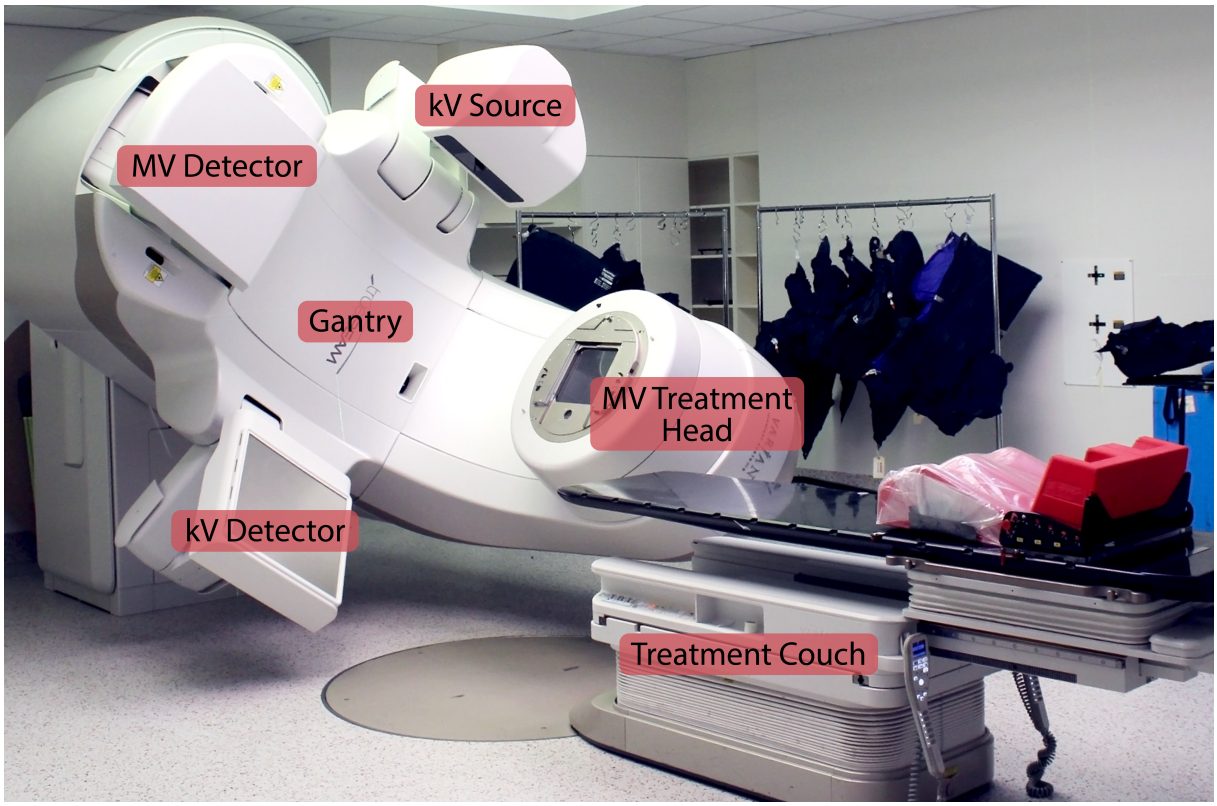


Fig. 2.1 | A standard clinical linear accelerator. The treatment head, which generates megavoltage (MV) radiation, is mounted on a gantry that rotates around the patient on the treatment couch. The gantry includes an onboard kilovoltage (kV) imaging system, and an electric portable imaging device (EPID) that detects the MV treatment beam which passes through the patient.

cell cycle[8] and re-oxygenate, increasing their radio-sensitivity.[7] For each fraction, the patient will be positioned on the treatment couch using a combination of lasers, external markers such as tattoos, surface imaging, and the onboard kV imager. Once the patient has been correctly positioned, the radiation is delivered as per the plan.

EBRT has evolved over the decades to improve tumour targeting while minimising the dose to healthy tissues. The introduction of 3D conformal radiation therapy (3D-CRT) improved dose distributions by conforming the beam to the target geometry using the MLC.[9] Intensity-modulated radiation therapy (IMRT) was a major advancement that modulated beam intensities to reduce the radiation dose to OARs, while increasing the radiation dose to the target.[10] Volumetric modulated arc therapy (VMAT) built on IMRT by delivering radiation in continuous arcs.[11] These advancements allowed for the development of stereotactic body radiation therapy (SBRT), also known as stereotactic ablative radiotherapy (SABR), delivering higher doses in fewer fractions to improve tumour control.[12]

2.2 Motion Management in Radiation Therapy

Motion during treatment is a major challenge for radiation therapy. Motion can lead to underdosing of the tumour, reducing tumour control, and overdosing of the healthy tissue, increasing radiation toxicity. To address this, motion management technologies have been developed such as immobilisation devices, internal and external monitoring, and real-time adaptive radiation therapy. One simple approach to account for motion uncertainties is to increase treatment margins.[13] However, increasing the margins may not fully account for all motion and also results in a higher radiation dose to healthy tissue.

Patient immobilisation devices have been developed to ensure the patient is in a reproducible position and has restricted mobility during treatment.[14] Supports moulded to the patient's body can be used to ensure reproducible positioning and stability. In the case of head and neck cancer, thermoplastic masks are moulded over a patient's face to reduce head motion. Abdominal compression is another form of immobilisation used to restrict respiratory motion for abdominal cancers.[15] While these techniques are used to reduce motion, movement can still occur necessitating daily or real-time image guidance.[16]

External motion monitoring involves tracking the external surface of the patient during treatment as a surrogate for the (internal) target motion and is typically used for lung and abdominal radiation therapy. Surface-guided radiation therapy uses cameras to create a 3D surface of the patient and track any changes.[17] Alternatively, cameras can track an external marker block on the patient, known as the real-time position management (RPM) block. While these methods can monitor motion in real-time, surface motion may not be directly correlated with internal motion. Implanted fiducial markers and stents were found to be better predictors for target motion than the RPM block.[18]

The ideal solution to motion management is to directly monitor the target pose using internal monitoring of the anatomy. This can be achieved using equipment on a modern linac such as the EPID or onboard kV system.[19] MV imaging uses the treatment beam and EPID to capture images in the beam's-eye-view. However, MV imaging has poor soft tissue contrast and is only possible when the treatment beam is on.[19] Alternatively, the kV imaging system has better imaging contrast compared to MV images, and can monitor the target continuously and independently of the MV beam state.[19] For these approaches, fiducial markers are required for anatomical sites with poor soft tissue contrast. The markers have numerous disadvantages, as

discussed in Chapter 1. The tumour and surrounding organs can be tracked without markers using an MRI-linac, although it is an expensive treatment option and not as widely available as standard linacs.[20]

One limitation of using the onboard kV system for motion management is the introduction of additional imaging dose to the patient. Frequent or continuous kV imaging contributes to cumulative radiation exposure, which should be carefully managed. Crocker et al. reported that the total imaging dose for kV prostate monitoring was between 0.1 and 0.5% of the treatment dose.[21] Strategies to minimise the dose include optimising exposure settings, reducing image acquisition frequency, collimating the kV field to the target region, and using predictive tracking models to decrease the number of required images. These approaches aim to maintain accurate target localisation while reducing the additional imaging dose associated with kV motion management.

With the development of real-time motion management, techniques have been developed for real-time adaption. Gating is one technique where the treatment beam is paused when the target moves beyond a preset motion threshold.[22] More sophisticated approaches involve moving the patient or adapting the treatment beam in real-time. Couch tracking, which has only been implemented experimentally, uses the robotic treatment couch to move the patient and compensate for motion.[23] Finally, the treatment beam can be adapted to follow the target using the MLC leaves.[24] MLC tracking could be extended to use real-time dose-based optimisation based on the target motion.[25]

The demand for accurate motion management has driven the development of real-time image guidance and adaptive technologies. However, these solutions typically rely on fiducial markers or are prohibitively expensive. AI has shown great potential for motion management in radiation therapy and will be the focus of this thesis. The remainder of this chapter presents a literature review of AI applications for motion management in radiation therapy which is current as of March 2021, when the review was submitted.[1]

2.3 References

1. Mylonas, A., Booth, J. & Nguyen, D. T. A review of artificial intelligence applications for motion tracking in radiotherapy. *Journal of Medical Imaging and Radiation Oncology* **65**, 596–611 (2021).

2. Baskar, R., Lee, K. A., Yeo, R. & Yeoh, K.-W. Cancer and radiation therapy: current advances and future directions. *International journal of medical sciences* **9**, 193 (2012).

3. Barton, M. B. et al. Estimating the demand for radiotherapy from the evidence: a review of changes from 2003 to 2012. *Radiotherapy and oncology* **112**, 140–144 (2014).

4. Jackson, S. P. & Bartek, J. The DNA-damage response in human biology and disease. *Nature* **461**, 1071–1078 (2009).

5. Begg, A. C., Stewart, F. A. & Vens, C. Strategies to improve radiotherapy with targeted drugs. *Nature Reviews Cancer* **11**, 239–253 (2011).

6. Webb, S. The physical basis of IMRT and inverse planning. *The British journal of radiology* **76**, 678–689 (2003).

7. Thoms, J. & Bristow, R. G. DNA repair targeting and radiotherapy: a focus on the therapeutic ratio. In *Seminars in radiation oncology*, vol. 20, 217–222 (Elsevier, 2010).

8. Short, S., Woodcock, M., Marples, B. & Joiner, M. Effects of cell cycle phase on low-dose hyper-radiosensitivity. *International journal of radiation biology* **79**, 99–105 (2003).

9. Vijayakumar, S. & Chen, G. Implementation of three dimensional conformal radiation therapy: prospects, opportunities, and challenges. *International journal of radiation oncology, biology, physics* **33**, 979–983 (1995).

10. Bortfeld, T. IMRT: a review and preview. *Physics in Medicine & Biology* **51**, R363 (2006).

11. Otto, K. Volumetric modulated arc therapy: IMRT in a single gantry arc. *Medical physics* **35**, 310–317 (2008).

12. Chang, B. K. & Timmerman, R. D. Stereotactic body radiation therapy: a comprehensive review. *American journal of clinical oncology* **30**, 637–644 (2007).

13. Burnet, N. G., Thomas, S. J., Burton, K. E. & Jefferies, S. J. Defining the tumour and target volumes for radiotherapy. *Cancer Imaging* **4**, 153 (2004).

14. Saw, C. B., Yakoob, R., Enke, C. A., Lau, T. P. & Ayyangar, K. M. Immobilization devices for intensity-modulated radiation therapy (IMRT). *Medical Dosimetry* **26**, 71–77 (2001).

15. Giraud, P. & Houle, A. Respiratory gating for radiotherapy: main technical aspects and clinical benefits. *International Scholarly Research Notices* **2013**, 519602 (2013).

16. Dawson, L. A. & Sharpe, M. B. Image-guided radiotherapy: rationale, benefits, and limitations. *The lancet oncology* **7**, 848–858 (2006).

17. Freislederer, P. et al. Recent advances in surface guided radiation therapy. *Radiation Oncology* **15**, 1–11 (2020).

18. Huguet, F. et al. Modeling pancreatic tumor motion using 4-dimensional computed tomography and surrogate markers. *International Journal of Radiation Oncology* Biology* Physics* **91**, 579–587 (2015).

19. Bertholet, J. et al. Real-time intrafraction motion monitoring in external beam radiotherapy. *Physics in medicine & biology* **64**, 15TR01 (2019).

20. van Herk, M., McWilliam, A., Dubec, M., Faivre-Finn, C. & Choudhury, A. Magnetic resonance imaging–guided radiation therapy: a short strengths, weaknesses, opportunities, and threats analysis (2018).

21. Crocker, J. K., Ng, J. A., Keall, P. J. & Booth, J. T. Measurement of patient imaging dose for real-time kilovoltage x-ray intrafraction tumour position monitoring in prostate patients. *Physics in Medicine & Biology* **57**, 2969 (2012).

22. Shirato, H. et al. Physical aspects of a real-time tumor-tracking system for gated radiotherapy. *International Journal of Radiation Oncology* Biology* Physics* **48**, 1187–1195 (2000).

23. D D'Souza, W., Naqvi, S. A. & Cedric, X. Y. Real-time intra-fraction-motion tracking using the treatment couch: a feasibility study. *Physics in Medicine & Biology* **50**, 4021 (2005).

24. Keall, P. J. et al. The first clinical implementation of electromagnetic transponder-guided MLC tracking. *Medical physics* **41**, 020702 (2014).

25. Mejnertsen, L., Hewson, E., Nguyen, D. T., Booth, J. & Keall, P. Dose-based optimisation for multi-leaf collimator tracking during radiation therapy. *Physics in Medicine & Biology* **66**, 065027 (2021).

Journal of Medical Imaging and Radiation Oncology **65** (2021) 596–611

RADIATION ONCOLOGY—REVIEW ARTICLE

Journal of Medical Imaging and Radiation Oncology

A review of artificial intelligence applications for motion tracking in radiotherapy

Adam Mylonas,[1,2] Jeremy Booth[3,4] and Doan Trang Nguyen[1,2,3]

1 ACRF Image X Institute, Faculty of Medicine and Health, The University of Sydney, Sydney, New South Wales, Australia
2 School of Biomedical Engineering, University of Technology Sydney, Sydney, New South Wales, Australia
3 Northern Sydney Cancer Centre, Royal North Shore Hospital, St Leonards, New South Wales, Australia
4 Institute of Medical Physics, School of Physics, The University of Sydney, Sydney, New South Wales, Australia

A Mylonas B.Eng(Biomedical)(Hons I), B.Med.Sci; **J Booth** B.MedPhys(Hons I), PhD; **D Trang Nguyen** B.Eng(Hons 1A), PhD.

Correspondence
Mr Adam Mylonas, ACRF Image X Institute, Faculty of Medicine and Health, The University of Sydney, Sydney, NSW, Australia.
Email: adam.mylonas@sydney.edu.au

Conflict of interest: The authors A Mylonas and D T Nguyen are inventors on two patent applications related to AI in motion tracking.

Submitted 5 March 2021; accepted 29 June 2021.

doi:10.1111/1754-9485.13285

Abstract

During radiotherapy, the organs and tumour move as a result of the dynamic nature of the body; this is known as intrafraction motion. Intrafraction motion can result in tumour underdose and healthy tissue overdose, thereby reducing the effectiveness of the treatment while increasing toxicity to the patients. There is a growing appreciation of intrafraction target motion management by the radiation oncology community. Real-time image-guided radiation therapy (IGRT) can track the target and account for the motion, improving the radiation dose to the tumour and reducing the dose to healthy tissue. Recently, artificial intelligence (AI)-based approaches have been applied to motion management and have shown great potential. In this review, four main categories of motion management using AI are summarised: marker-based tracking, markerless tracking, full anatomy monitoring and motion prediction. Marker-based and markerless tracking approaches focus on tracking the individual target throughout the treatment. Full anatomy algorithms monitor for intrafraction changes in the full anatomy within the field of view. Motion prediction algorithms can be used to account for the latencies due to the time for the system to localise, process and act.

Key words: artificial intelligence; deep learning; machine learning; motion tracking; radiation oncology.

Introduction

Since the early days, radiotherapy has been driven by new technological developments. Image-guided radiation therapy (IGRT) improves the treatment accuracy and enables the success of highly conformal treatment delivery techniques such as intensity-modulated arc radiation therapy and volumetric modulated arc therapy. With hypofractionated and ultrahypofractionated treatments becoming more mainstream, particularly with stereotactic body radiation therapy (SBRT) becoming the standard of care for non-small cell lung cancer and prostate cancer, ASTRO recommends not only daily IGRT but also real-time IGRT as the target positioning method for patient treatments.[1,2] This is due to a growing appreciation of intrafraction target motion management by the radiation oncology community.

Although the prostate was assumed to move negligibly, greater understanding of prostate motion led to changes in intrafraction motion management. This was a result of the introduction of kilovoltage (kV) and ultrasound imaging during treatment in the early 2000s driven by the highly sculpted dynamic MLC deliveries. The range and distribution of intrafraction prostate motion observed in patients has been reported in several studies.[3–5] An analysis of prostate intrafraction translational motion in 427 patients found that displacement larger than 2 mm occurred in 66% of fractions, while 28% of fractions had exceeded 3 mm displacement.[6] Most of the prostate motion is reported to be in the anterior–posterior and superior–inferior directions.[4] In some cases, intrafraction prostate motion can be up to 15 mm, which typically exceeds the planning target volume (PTV) expansion for radiotherapy treatment of the prostate. In addition to translational motion, the prostate motion also includes rotation. Intrafraction rotation motion is more commonly seen in the left–right axis, with an average of 2.5° (±2.3°) reported.[5] However, in extreme cases, the

rotational motion can reach beyond 10°.[7,8] The physiological origins of prostate intrafraction motion are frequently bladder and rectal volume changes due to bladder filling, rectal gas passing or rectal filling. These same processes also result in intrafraction motion and deformation seen during other pelvic cancer treatments, such as cervical cancer. Using cine MRI images, motion of the cervical clinical target volume (CTV) was reported to be as high as 10.6 mm.[9]

Respiration-induced motion is faster and has greater variability than the random motion of the prostate and has been shown to affect treatment accuracy. The motion of the target within the thorax and abdomen is thus needed to be monitored in real-time with a high frequency. For thoracic and upper abdominal targets, intrafraction motion (translation, rotation and deformation) of several centimetres has been observed in the liver,[10,11] lung[12,13] and pancreas.[14] For targets located close to the heart, cardiac motion is also a concern for motion management for two reasons. Firstly, the target will exhibit respiratory and cardiac induced motion. Secondly, the heart and the associated large vessels are typically dose-limiting organs due to high radiation sensitivity. Cardiac motion of the heart and its substructures was reported to be between 7 mm and 15 mm in a MRI study.[15] However, the cardiac component of the motion was found to be within 4–5 mm magnitude.[16] Although inclusion of an internal target volume during planning to account for motion is a widely used method, it has been shown to be less optimal than monitoring and accounting for motion in real-time both lung and liver SBRT treatments.[17,18]

The evidence for the dosimetric impact of intrafraction motion is mounting. A study on liver cancer treatment with radiation therapy showed that the dose to the liver could be reduced by as much as 20% due to liver motion.[19] Another study on prostate cancer patients showed that prostate motion could result in 19% tumour underdose.[20] For lung cancer patients, when motion is compensated for, not only will the tumour receive the full dosage, but the surrounding organs benefit with a reported reduction in mean lung dose by 17% and in heart dose by 56%.[18]

A number of different intrafraction real-time guidance methods have been used during radiotherapy treatments. The first real-time tracking radiotherapy system, pioneered by Shirato et al.,[21] uses four orthogonal room-mounted X-ray systems to track the prostate position based on segmented positions of implanted markers.[21–26] Other commercial systems including CyberKnife[27] (Accuray, Sunnyvale, CA, USA) and ExacTrac[28] (BrainLab, Munich, Germany) use two room-mounted X-ray imagers synchronously to monitor the target motion during treatment. Due to gantry occlusion, an image pair can only be taken at cardinal angles. This limitation is not seen in the four X-ray system developed by Shirato et al.[21]

Non-radiographic methods for target motion monitoring have also been developed in the past two decades. Calypso (Varian, Palo Alto, CA, USA)[4] and RayPilot (Micropos, Gothenburg, Sweden)[29] utilise implanted markers, which also act as electromagnetic transponders, transmitting positional signals to an external receiver. Emerging real-time guidance technologies include ultrasonography[30] and integrated magnetic resonance imaging (MRI)-radiation therapy systems.[31,32] While these technologies are becoming more common, this review will focus primarily on kV IGRT.

Up to date, we have witnessed artificial intelligence (AI) lending its power to the ideal goal of monitoring the target during radiotherapy using a standard linear accelerator. Machine learning (ML) and deep learning (DL) are key AI technologies that are being increasingly used in radiotherapy applications. These technologies excel at extracting features from training data and making predictions on new unseen data. Most of the published research to date that use these technologies centre around localising the target locations on 2D kV images such that the 3D position of the target can be reconstructed. The target in this context could be a surrogate such as arbitrary-shaped fiducial markers,[33] the gross tumour volume,[34] bony anatomy surrounding the target[35] or the soft tissue target with very low radiological contrast.[36,37]

AI-related motion tracking publications have been predominately produced over the past two decades. In recent years, there has been a surge of publications related to this field (Fig. 1). Almost half of all articles published over the past two decades have been published in just the past three years (Fig. 1) and now focus primarily on DL. Therefore, it is important to evaluate the shift in AI-related motion tracking research from ML to DL over the past two decades. Furthermore, a review of this field will summarise initial progress made with AI techniques and provide insight for further research. In this review, we aim to examine the state of the art of AI techniques that have been applied to motion monitoring, starting from marker-based approaches to full anatomy reconstruction. Marker-based approaches are typically used for prostate real-time IGRT since the markers act as a high contrast surrogate for the target. However, the development of markerless tracking technologies is beneficial to the patient and the healthcare system by obviating the need for implanted markers. Full anatomy algorithms can monitor for intrafraction changes in the full anatomy within the field of view. Another important technology is motion prediction algorithms which can be used to account for the latencies due to the time for the system to localise, to process and to act. AI algorithms have helped to advance these key technologies for motion tracking in radiotherapy.

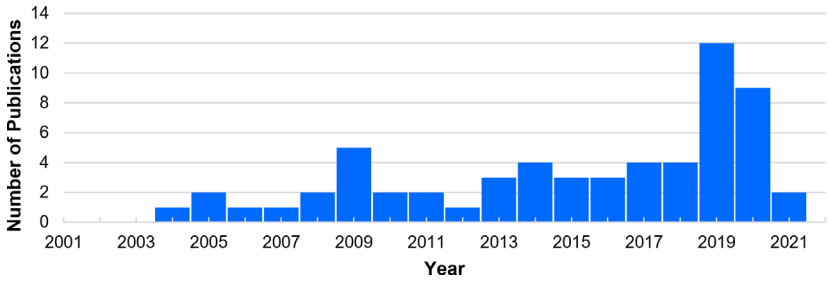


Figure 1. Number of publications over the past two decades for AI in intrafraction motion tracking

AI technologies

The research discipline of AI aims to develop intelligent systems so advanced that they can solve problems to the same ability or even better than humans. ML is a branch of AI in which systems can automatically learn from data and improve from experience without being programmed to do so. There are a large variety of statistical algorithms used in ML such as artificial neural networks (ANN), support vector machines (SVM), decision trees, decision forests, logistic regression and Bayesian classifiers.[38] DL is a subset of ML which solves problems using neural networks which have a structure similar to the human neural system. The name DL comes from the structure (or architecture) of the networks. These networks have multiple layers in which different operations are applied to train the network, unlike traditional ML algorithms which only had a small number of layers.

Convolutional neural networks (CNN) are one of the main DL approaches with a multiple layer architecture. Traditional CNN architectures are typically used for image classification problems (Fig. 2a). Traditional CNNs are designed to receive an input image of a specified size and output a single-label classification. CNNs can be adapted for image segmentation using a common approach known as regions with CNN (R-CNN). R-CNNs first generate region proposals that potentially contain the object of interest and then performs classification on the regions using a CNN. The traditional CNN architecture can be modified to perform detection and semantic segmentation (Fig. 2b). Semantic segmentation is the technique of partitioning an image into labelled segments by classifying each pixel. A common architecture used for semantic segmentation is the encoder–decoder network (Fig. 2a), which allows for an arbitrarily shaped input and output of the same size.

Method

The authors conducted a topical literature review using the online databases PubMed and Web of Science from February to May 2021. The search criteria initially restricted the articles to the field of radiotherapy using: (radiation therapy OR radiotherapy). These search results were refined towards AI using: (machine learning OR deep learning OR neural networks OR artificial intelligence). Finally, the articles were further refined to tracking applications using: (fiducial markers OR markerless), or refined to motion prediction applications using: (motion AND prediction). An additional search was completed to review applications of AI to motion management in MRI-guided radiotherapy. The search criteria initially restricted the articles to the field of MRI-guided radiotherapy using: (mr OR mri) AND linac. These search results were refined towards AI using the aforementioned AI search terms. In addition, the reference lists of selected articles were hand searched for other relevant articles. Articles returned in the search results that were not related to intrafraction motion tracking were excluded from the review.

Marker-based approaches

Image-guided radiation therapy for some types of cancer typically relies on markers to determine the target location. For motion tracking in kV images, markers are implanted into the tumour or surrounding tissue to act as a surrogate for the target due to a lack of contrast in the intrafraction images (Fig. 3). A range of AI approaches for marker tracking have been demonstrated to have high accuracy (Table 1).

Many methods for segmenting markers use traditional approaches such as template matching.[39–43] Template matching approaches require prior information of the marker. Typically for regular shaped markers (Fig. 1a; e.g. cylindrical gold fiducials), the templates are constructed using known marker properties.[40,41] Other types of markers such as coiled markers (Fig. 1b; e.g. Visicoil™) may deform into arbitrary shapes upon implantation resulting in unknown marker properties. In order to develop a template, an additional learning period is required to determine the marker properties.[39,42,43] Cone-beam CT (CBCT) projections are commonly used as learning data for the template generation.[41] This learning period has the drawback of subjecting the patient to additional radiation due to the extra imaging

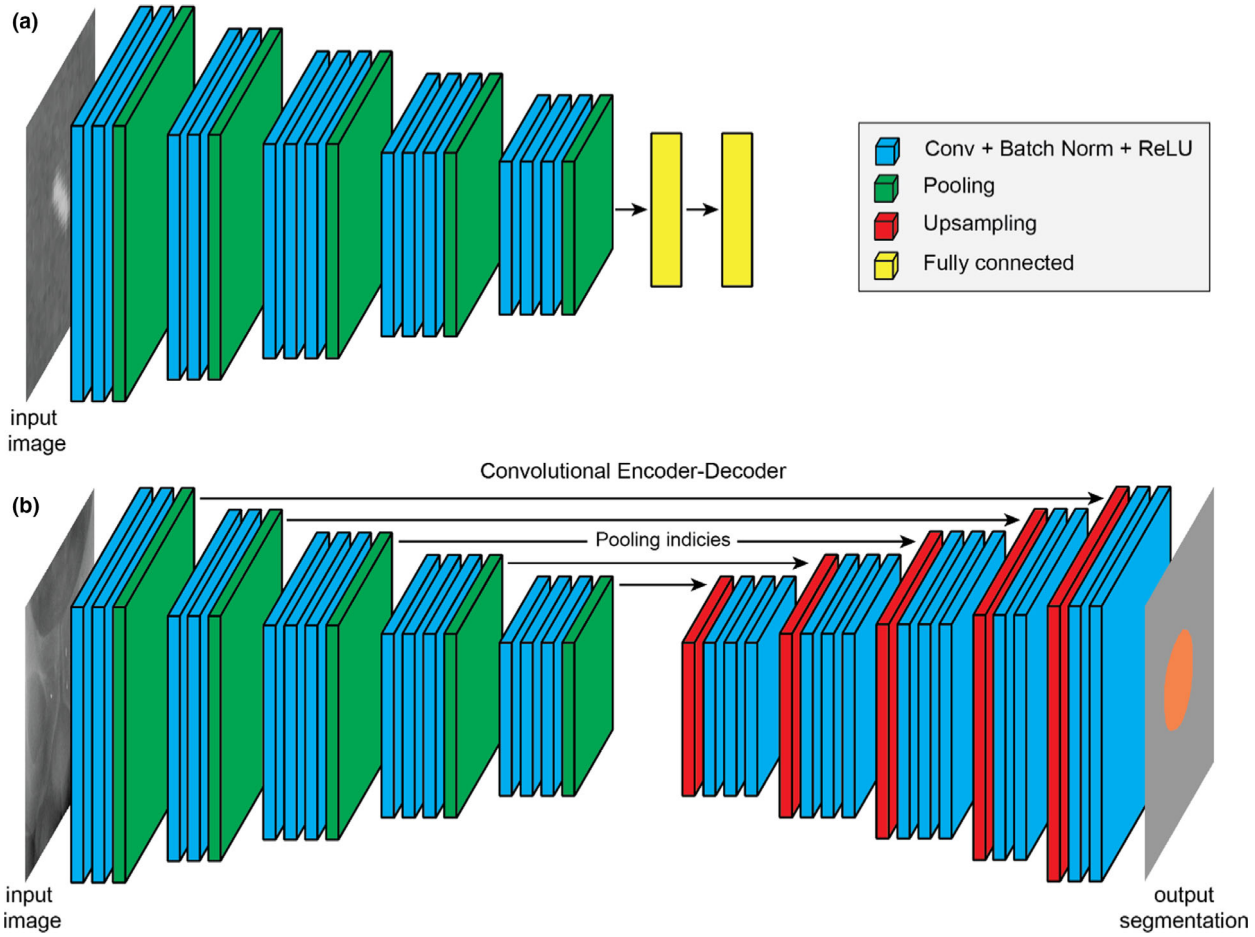


Figure 2. The basic architecture of (A) a CNN for image classification and (B) an encoder–decoder network for semantic segmentation. The network layers include convolutions (Conv), batch normalisation (Batch Norm), rectified linear unit (ReLU), pooling, upsampling and fully connected layers

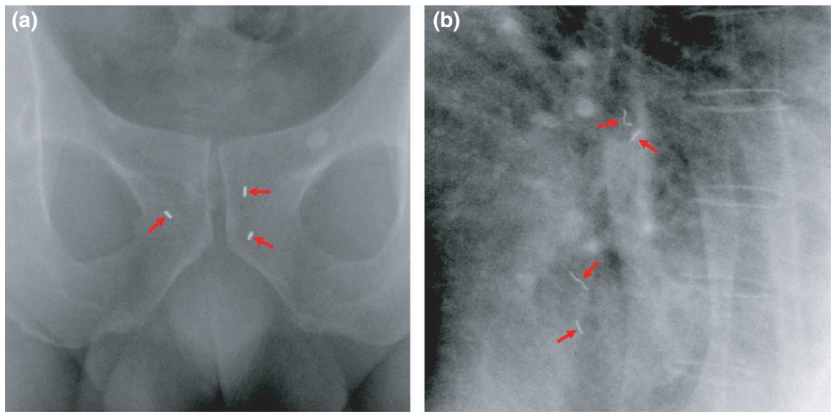


Figure 3. Example kV projections of (a) regular shaped gold fiducial markers implanted in the prostate and (b) arbitrarily shaped coiled markers implanted in the lung

dose. DL methods have been successfully used for the automatic detection of markers in X-ray and MRI images to overcome these issues.[33,44,45]

Tracking fiducial markers using the megavoltage electronic portal imaging device (MV-EPID) is ideal as it eliminates the need for additional imaging equipment and

Table 1. Summary of the methods to track implanted markers

Study	Imaging modality	Marker type	Method	Error	Other metrics
Lin et al. 2012 [45]	MV	Regular	ML	RMSE error: 0.5 mm	-
Mylonas et al. 2019 [33]	kV	Regular	CNN	Mean (±SD) error: −0.8 ± 1.3 pixels and 0.3 ± 1.2 pixels in the x and y directions, respectively	Sensitivity (recall): 99.4% Specificity: 99.3%
Mylonas et al. 2019 [33]	kV	Arbitrary	CNN	Mean (±SD) error: −0.9 ± 2.8 pixels and 0.2 ± 2.2 pixels in the x and y directions, respectively	Sensitivity (recall): 98.5% Specificity: 99.6%
Liang et al. 2020 [44]	X-ray	Regular	U-Net	Mean (±SD) error: 0.3 ± 0.5 pixels	Sensitivity (recall): 95.6% Precision: 98.6%

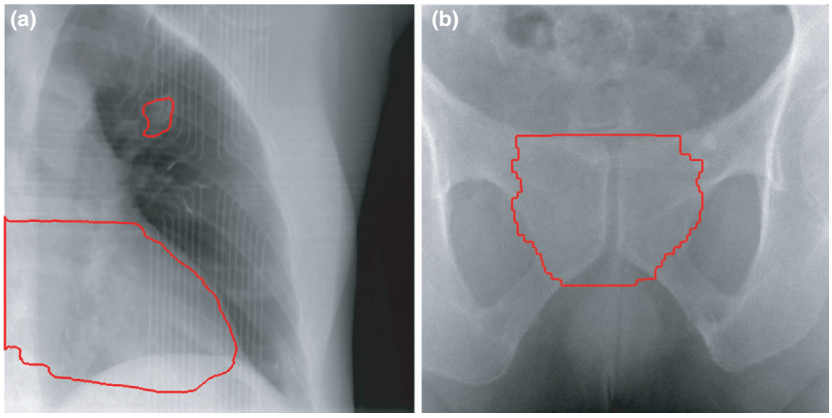


Figure 4. Example segmentations of the (a) heart and lung tumour and (b) the prostate

provides target information in the view of the treatment beam. However, accurate tracking is challenging for VMAT treatments due to low contrast and MLC leaves occluding markers. Lin et al.[45] detected the markers using discriminant analysis for the first frame and mean-shift feature space analysis for subsequent frames. To reduce the computation time, the authors exploited the fact that the markers moved slightly between each frame. An average root-mean-square error (RMSE) of 0.5 mm was achieved using this approach.

Mylonas et al.[33] implemented a framework that uses a CNN to automatically detect regular and arbitrarily shaped markers in intrafraction fluoroscopic images. The CNN had three convolutional layers with two pooling layers and one fully connected layer. The CNN was used to classify whether sub-images of the kV projections contained a marker or background. High classification performances were achieved for both types of markers. For the regular shaped markers implanted in the prostate, a sensitivity and specificity of 99.4% and 99.3%, respectively, were achieved. These results improve upon a marker detection rate of 94.8% for a template matching approach reported by Fledelius et al.[41] For the arbitrarily shaped markers implanted in the lung, a sensitivity and specificity of 98.5% and 99.6%, respectively, were achieved. This improves upon the traditional method developed by Bertholet et al.,[39] which reported a sensitivity and specificity of 67% and 94%, respectively. While the CNN approach improves upon the current standard, it should be noted that the systems were tested on different data sets. The CNN was incorporated into a tracking system to segment multiple markers for intrafraction monitoring. The computation time was adequate for real-time applications with a time of 36 ms on a CPU and 9 ms on a GPU. Sub-millimetre accuracy was achieved for both regular and arbitrarily shaped marker tracking.

A study by Liang et al.[44] presented an automated framework that was used to evaluate intrafraction motion in X-ray images acquired by CyberKnife. The framework includes a fully convolutional network (FCN)-based fiducial marker detection module. The authors utilised the U-Net architecture to perform semantic segmentation of the markers in full-size X-ray images. The U-Net architecture is an implementation of the encoder–decoder network architecture shown in Fig. 2b. The network has a contracting path to obtain the feature information and a symmetrical expanding path which

Table 2. Summary of the markerless-based approaches for fluoroscopic-guided radiotherapy

Study	Imaging modality	Method	Site	Results
Tong et al. 2009 [50]	X-ray	ML	Lung	Mean error (95% confidence level): 2.1 (4.6) pixels (pixel size is about 0.5 mm)
Terunuma et al. 2018 [51]	Synthetic X-ray	Encoder-decoder CNN	Lung	Mean error: approximately 1 mm
Edmunds et al. 2019 [52]	X-ray	Mask R-CNN	Diaphragm	Mean error: 4.4 mm
Hirai et al. 2019 [53]	X-ray	DNN	Liver and lung	Mean (±SD) accuracy: 1.6 ± 0.7 mm
Mori et al. 2019 [54]	X-ray	ML	Chest	Accuracy: <0.5 mm
Zhao et al. 2019 [37]	Synthetic X-ray	Faster R-CNN	Pancreas	MAD: less than 2.6 mm across all directions
Zhao et al. 2019 [36]	Synthetic X-ray	Faster R-CNN	Prostate	MAD (±SD): 1.6 ± 0.4 mm, 1.6 ± 0.4 mm, and 1.7 ± 0.4 mm in anterior-posterior, lateral, and oblique directions, respectively
Roggen et al. 2020 [35]	X-ray	Faster R-CNN	Vertebrae	Sensitivity of 1.5 mm with a median below 0.5 mm for single vertebrae
Sakata et al. 2020 [55]	X-ray	ML	Lung	Mean (±SD) error, Euclidean distance: 1.0 ± 0.3 mm
Takahashi et al. 2020 [56]	X-ray	FCN	Lung	Mean (±SD) error: mean (±SD): 0.5 ± 0.2 mm and 0.3 ± 0.2 mm for 2 cm and 3 cm tumours, respectively

Table 3. Summary of the markerless-based approaches for MRI-guided radiotherapy

Study	Imaging modality	Method	Site	Results
Cervino et al. 2011 [59]	cine MRI	ML	Lung	Mean tracking error of 1.5 mm
Yun et al. 2015 [60,61]	cine MRI	PCNN	Lung	Centroid tracking accuracy of 1.0–1.4 mm
Bourque et al 2017 [62]	cine MRI	ML	Lung	RMSE: 1.3 ± 0.5 mm
Fast et al. 2017 [63]	cine MRI	PCNN	Lung	Median Dice similarity index: ≥0.91
Friedrich et al. 2021 [64]	cine MRI	U-net	Liver	Average Dice coefficient ≥83% Mean contour distance ≤0.5 pixel

increases the resolution of the encoder feature maps. The method achieved a precision and recall of 98.6% and 95.6%, respectively, and the mean (±SD) centroid error was 0.3 ± 0.5 pixels. Their proposed method was used to retrospectively analyse the intrafraction motion of markers implanted in the liver. This method could be further developed for use in real-time tumour tracking systems. The computation time of this technique would need to be examined before being implemented in a real-time tracking system.

Markerless-based approaches

The development of markerless IGRT would be beneficial to patients and the healthcare system by eliminating the need for the surgical implantation of markers. A markerless approach would avoid extra appointments, costs, discomfort and potential complications associated with marker implantation. Percutaneous marker implantation can have complications such as haemorrhaging.[46,47] Furthermore, some patients are not candidates for markers due to contraindications. This prevents some patients from being treated with real-time IGRT techniques. Marker migration is another potential issue that can result in tracking errors[48] and markers produce metal artefacts in CT images that may result in treatment-planning errors.[49] Therefore, the ability to track tumours without markers would be beneficial to cancer patients and enable wide access to real-time IGRT. It is difficult to develop highly accurate markerless tracking methods for organs such as the prostate or pancreas due to the low contrast in intrafraction images (Fig. 4). However, AI approaches have shown great potential for markerless tracking (Tables 2 and 3).

Several ML approaches have been employed for markerless lung tumour tracking. A novel method developed by Lin et al.[50] uses principal component analysis (PCA) to determine the motion patterns of the surrogate anatomical features. The tumour position is predicted based on the motion patterns through regression using ML approaches. The tracking system was tested on the fluoroscopic images for ten lung cancer patients. Compared to other methods, ANN regression was shown to perform the best with a mean tracking error of 2.1 pixels and mean maximum error (95% confidence level) of 4.6 pixels where one pixel is approximately 0.5 mm.

Mori et al.[54,57] and Sakata et al.[55] used a ML algorithm to produce a tumour likelihood map using fluoroscopic images. A binary classifier optimised by the extremely randomised tree method was trained to classify images as positive (including tumour) and negative (not including tumour). The classifier was trained using digital reconstructed radiographs (DRR) generated from the 4DCT. During the treatment, sliding window classification was used to generate the tumour likelihood map. The tumour position was predicted using the tumour

likelihood map and prior probability tumour positions. The system achieved an accuracy of <0.5 mm when tested on a moving chest phantom.[54] When tested on images from eight patients, the mean (±SD) tumour position error was 1.0 ± 0.3 mm in 3D space based on paired X-ray images.[55] Furthermore, the ML algorithm was shown to be accurate when trained with DRRs generated from a simulated 4DCT.[57]

Zhao et al.[36] implemented a DL approach for markerless tracking for prostate and pancreatic[37] cancer. The faster R-CNN model used in these studies consisted of two networks that share information: a region-proposed network (RPN) and a R-CNN. The authors used a RPN to generate proposals for the R-CNN to reduce the computation time and enable real-time target detection. The feature extractor was the VGG16 ConvNet. A model was trained for three different directions: anteroposterior, left–right and an oblique. The network input can be either digitally reconstructed radiography (DRR) or X-ray projection, and the output is a bounding box of the PTV. The models were evaluated on simulated kV projection images. For the pancreatic cancer model, the median absolute deviation (MAD) between the predicted and actual position was less than 2.6 mm for all cases and directions.[37] For the prostate cancer model, the MAD (±SD) between the predicted and actual positions were 1.6 ± 0.4 mm, 1.6 ± 0.4 mm and 1.7 ± 0.4 mm in anterior–posterior, lateral and oblique directions, respectively.[36] The prostate cancer model was additionally evaluated on kV images of one prostate cancer patient. The predicted prostate position was consistent with the position of the markers for the two kV images shown.[36] These results demonstrated the feasibility of markerless prostate and pancreas tracking for IGRT. For clinical implementation, it would be beneficial to expand this method for tracking through all angles of a radiotherapy treatment arc. For both methods, the model produced the prediction in less than 200 ms which makes this method feasible for real-time monitoring.

A DL method developed by Hirai et al.[53] successfully estimated lung and liver tumour positions for markerless tracking. A deep neural network (DNN) was trained to produce a target probability map (TPM) that estimated the tumour position. The DNN was based on an encoder and decoder network architecture (Fig. 2b). For training, DRR images were cropped to produce sub-images of the target and surrounding anatomy. The network takes the pattern of sub-images to produce the TPM. During tracking, the fluoroscopic images are cropped at the same positions used during the training process on the DRR images and the target positions were calculated based on the TPMs. The framework was implemented on both lung and liver cases. The accuracy was calculated using the Euclidian distance in the 3D space between the calculated and reference tumour position. An overall mean accuracy (±SD) of 1.6 ± 0.7 mm was achieved for all patients. The framework was more accurate on liver cases achieving a mean accuracy (±SD) of 1.4 ± 0.8 mm in comparison to 1.9 ± 0.7 mm for lung cases. This method was shown to be feasible for real-time since the computation time was less than 40 ms.

A study by Roggen et al.[35] evaluated a DL model to detect vertebra motion in 2D X-ray projection images. The authors used the Mask R-CNN framework based on ResNet101 to perform image segmentation. First, the framework detects several regions of interest and then a Faster R-CNN is used to process each region to determine the actual presence of the objects of interest. The framework was trained and tested using the CBCT projections acquired using an on-board kV imaging system. The ground truth was the forward projection of the 3D contour of the vertebrae. The network takes a 2D kV projection image as input and outputs a segmentation mask and bounding box of each vertebrae. This approach was validated on images from one patient and a patient-like full-body phantom. When validated on the patient images, positional changes in the range of 1.5 mm could be detected, with a median below 0.5 mm. Sub-millimetre accuracy was achieved when the detection of several vertebrae was combined. Additionally, this method can detect spinal rotations larger than 1° based on the detection of all vertebrae in the image. Further development is still required to achieve the same level of accuracy and provided additional advantages compared to current systems such as the Cyberknife system. For spinal tracking, the Cyberknife system has an accuracy of <0.2 mm for translational and <0.3° for rotational directions.[58]

A R-CNN was also used by Edmunds et al.[52] to segment the diaphragm for lung tumour tracking with a mean diaphragm apex error of 4.4 mm. Terunuma et al.[51] used a encoder–decoder CNN architecture for lung tumour segmentation in DRRs. When tested on DRRs, the tracking error was approximately 1 mm with a processing time of 25 ms/frame for contouring and tracking. A DL approach for segmenting lung tumours was also developed by Takahashi et al.[56] Their approach uses a FCN trained on patient-specific DRRs. The system was tested in a phantom study with tumours of 1, 2 and 3 cm in size. Sub-millimetre accuracy was achieved with a mean (±SD) error of 0.5 ± 0.2 mm and 0.3 ± 0.2 mm for 2 cm and 3 cm tumours, respectively. However, the mean error increased to over 1 mm for a smaller 1 cm tumour. The tracking system had an approximate processing time of 33 ms/frame.

Currently, MR images can only be acquired in 2D modality during treatment delivery.[65] Therefore, the AI approaches for motion management during MRI-guided radiotherapy summarised in Table 3 are designed for 2D images.[59–64] A study by Cervino et al. evaluated two approaches for tracking lung tumours in cine MRI images.[59] Cervino et al. compared an ANN model and a template matching approach combined with surrogate

tracking using the diaphragm. It was observed that the template matching approach detected the tumour location more accurately than the ANN. Yun et al.[60] and Fast et al.[63] used a pulse-coupled neural network (PCNN) to track lung tumours in cine MRI images. Yun et al.[60] reported a centroid tracking accuracy of 1.0–1.4 mm. However, the PCNN was outperformed by multi-template matching and deformable image registration in the study by Fast et al.[63] A recent study compared B-spline deformable image registration and U-net for segmentation of liver tumours in cine MRI images.[64] The U-net was found to outperform B-spline and achieved a mean contour distance of <0.5 pixels. A ML approach that incorporated motion prediction for tracking lung tumours was proposed by Bourque et al.[62] A RMSE of 1.3 ± 0.5 mm and 2.0 ± 0.8 mm with and without prediction, respectively, shows the benefit of including motion prediction into the tracking algorithm.

Algorithms for monitoring changes in the full anatomy

Another approach for motion tracking is to monitor for shifts in the full anatomy or features of the surrounding anatomy. It can be useful to monitor for shifts in the full anatomy during the treatment to ensure that the target volume and surrounding organs at risk remain in position. If there are significant shifts, gating can be performed and the patient can be repositioned. There are a variety of AI solutions for gating based on the full anatomy.[34,66,67]

An approach used in several studies is to use a binary classifier for online treatment target verification by determining when to turn the beam ON or OFF.[34,67,68] Tang et al.[67] proposed using an ANN binary classifier to classify the lung tumour as inside or outside of the beam aperture in cine EPID images. The ANN was trained on DRRs with simulated tumour shifts to represent cine EPID images. The MLC contour was superimposed on the DRRs at different locations to simulate different tumour positions. The CTV defined by the physician was projected onto the DRRs. The percentage of the CTV in the beam aperture was calculated for each image. A percentage threshold was set to divide the images into the two classes: inside or outside of the beam aperture. The ANN was validated retrospectively on five lung patients. The ANN had a classification accuracy of 98.0%, recall of 97.6% and precision of 99.7%. The proposed method could be implemented for real-time gating during radiotherapy when the ANN determines the tumour has moved outside of the beam aperture.

Similarly, Lin et al.[34] developed a solution for gated radiotherapy using a binary classifier and evaluated the performance of several types of ML algorithms. The authors used a combination of dimensionality reduction technique and ML algorithms. The binary classifier was developed to classify images as beam ON or beam OFF based on the gating window size set during treatment planning. The method was validated on fluoroscopic images of nine lung cancer patients. The ANN combined with PCA was determined to be the best candidate for real-time gated radiotherapy. The ANN with PCA method had a mean (±SD) classification accuracy, recall and target coverage of $96.3 \pm 1.6\%$, $89.9 \pm 7.6\%$ and $97.8 \pm 1.7\%$ respectively. Cui et al.[68] investigated the use of a SVM for the classification problem. However, the ANN approach was shown to be more accurate than SVM in the study completed by Lin et al.[34]

An approach for volumetric reconstruction using DL developed Shen et al.[66] could be used for image-guided radiotherapy. An encoder–decoder model was trained to generate 3D volumetric CTs from 2D X-ray projections. The model can be trained on the planning CT to make the approach patient-specific. The training data set of DRR-CT pairs was generated by producing DRR images based on the CT. The authors investigated using 1, 2, 5 and 10 projection views as input to obtain the CT. It was found that the underlying semantic representations of the 3D scene were similar when different 2D projection views are utilised. The model was validated on both abdominal and lung CTs. The mean absolute error between the predicted and ground truth images for single-view reconstruction was 0.018 and 0.025 for abdominal and lung CTs, respectively. This single-view imaging approach is a potential solution to image-guided interventional procedures. Another study proposed a method for producing a volumetric image using a single X-ray projection using PCA.[69]

Similarly for MRI-guided radiotherapy, DL methods have the potential to reduce the computation time for image reconstruction from 2D cine MRI images. Real-time reconstruction would allow for volumetric segmentation of the target and treatment adaption. Terpstra et al. compared four conventional and DL approaches for image reconstruction and to estimate the motion of abdominal cancer.[70] The authors used a combination of methods to achieve a RMSE up to 1 mm, with a processing time of 60 ms.

Ultrasound-guided radiotherapy can be challenging due to the major changes in the image content due to motion of the anatomy and the position of the ultrasound probe. Zhu et al.[71] used a pretrained CNN for image registration of 3D ultrasound prostate images. The registration errors using the CNN were smaller than 5 mm in 81% of the cases which was an improvement over a manual approach that had errors smaller than 5 mm in only 61% of the cases.

Prediction of tumour motion

There will be a latency in any system from the time to localise, to process and to act. For any action to be considered real-time, then the system must predict (estimate) localisation at a time in the future equal to the

system latency. Common system latencies for say an MLC tracking system are 100-400 ms. Prediction algorithms are not new, and many have been applied to the motion of targets in radiotherapy using statistical approaches. More recently, DL approaches have been applied to this problem.

Several studies have used ML algorithms for predicting tumour motion based on past motion,[72–89] including in MRI-guided radiotherapy[90–92] and ultrasound-guided radiotherapy.[93] A comparison study of ML algorithms was made by Sharp et al. and showed that most ML algorithms have a lower localisation error compared to no prediction.[94] Murphy and Dieterich[95] evaluated a linear adaptive filter and an adaptive nonlinear neural networks (NN) for the prediction of highly irregular breathing patterns. The ANN was found to be more robust as it successfully predicted all respiratory patterns, while the linear filter failed for several patterns. For the successful cases, the ANN achieved comparable or better accuracy compared to the linear filter. As the latency was increased, the ANN become significantly more accurate than the linear filter. This indicates the ANN is superior at predicting the future respiratory signal pattern. Ruan et al.[96] showed that a local regression method can outperform ANN for predicting motion 1 s in advance. Based on initial studies, Goodband et al.[76] and Murphy et al.[82] proceed to optimise the performance of NNs to further improve the prediction accuracy. These studies show that different configurations of parameters should be tested to optimise the ANN model. A study by Kraus et al.[97] compared the performance of respiratory motion predictors based on NN, linear regression (LR), kernel density estimation (KDE) and support vector regression (SVR). The authors developed patient-independent models. The predictors were validated on various sampling rates and latencies using 3D lung tumour motion traces and the errors were averaged over all sampling rates and latencies. The prediction errors were normalised to the errors using no prediction to give 0.44, 0.46, 0.49 and 0.55 for NN, SVR, LR and KDE, respectively. This shows that all the predictors approximately halved the prediction errors compared to using no prediction. There were only small differences between the predictors with the ANN approach being the best predictor.

The majority of the aforementioned studies focus on predicting respiratory motion for lung cancer treatments. Dick et al.[98,99] used an ANN to predict the motion of abdominal tumours based on the lung–diaphragm border. The studies demonstrated that it is feasible to predict the location of liver tumours based on the lung–diaphragm border. This method shows the potential for markerless tracking of abdominal tumours by tracking a surrogate organ.

More recent approaches use DL to predict the motion trajectories of tumours. These methods commonly use CNNs, long short-term memory (LSTM) or fuzzy DL.[100–105] Lee et al.[100] showed that a CNN model performs better than a RNN. Lin et al.[101] developed a LSTM-based generalised model to predict the patient's respiratory signal. The neural network was trained and tested using real-time position management (RPM) data. When tested on unseen respiratory data with a 500 ms prediction window, the average relative mean absolute error was 0.112 and the root-mean-square error was 0.139. Wang et al.[105] compared the use of LSTM networks and SVR to predict external respiratory motion and internal liver motion. It was found that the LSTM network performed better than SVR. These results demonstrate the potential of LSTM networks for respiratory signal prediction. Romaguera et al.[103] incorporated convolutional LSTMs into an recurrent encoder–decoder model. The model proposed by Romaguera et al. finds the deformation between input images and then predicts the future motion. This method has shown the potential to predict motion on MRI, ultrasound and CT data sets.

In the study of Park et al.,[102] a fuzzy DL model was used to predict intra- and inter-fractional variation of lung tumours. The fuzzy DL model combines fuzzy logic and an ANN component. The ANN component gives the model the ability to self-learn features and set network parameters during training. The fuzzy logic component improves the reasoning ability for uncertainty. This feature is useful for predicting inconsistent inter-fractional variation. Using 192.3 ms prediction, the fuzzy method had an RMSE of 0.5 $\pm$ 0.8 mm, representing an improvement of 30.0% over a CNN and hybrid motion estimation based on extended Kalman filter. The average computation time for the model was 1.5 ms. Torshabi et al.[104] also successfully used an adaptive neuro-fuzzy model to predict the tumour motion based of external surrogate. Therefore, these studies show the potential for combining fuzzy logic and NNs to achieve real-time prediction.

A unique method developed by Kim et al.[106] involved using a CNN to predict motion of the patient in advance based on their facial expressions. A camera was installed in the treatment room to monitor the patient's facial expressions. The system had an accuracy of 85.6% for detecting uncomfortable expressions which can be used to provide warning about potential sudden movements. A limitation of this approach is that only movements associated with an uncomfortable expression will be detected.

Discussion

In the past 5 years, the world has undergone a dramatic change with the advances of AI. Self-driving vehicles are becoming a reality and AI defeated a world champion at the last traditional board game that had been unconquered by computers.[107,108] Medicine, too, has started to feel the impact of AI. In particular, AI has been used for a range of medical imaging applications including organ segmentation and localisation, lesion detection and segmentation, image registration and predicting clinical outcomes.[109] DL approaches are now commonly

used for medical image analysis due to their success in recent years.[109,110] There is a gap for intrafraction motion management that had the potential to be filled by AI-based approaches. In radiotherapy, intrafraction motion is still without a satisfactory commercial solution on a standard linear accelerator. The requirement of accuracy and precision in real time for a variety of targets is often a barrier for in-house developed solutions to become mainstream: it is far easier to ensure a solution work 90% of the time in a controlled clinical trial single-centre environment compared to a commercial solution where it needs to be accurate 99.5% for every centre and every patient. As discussed in the introduction, the clinical benefit of real-time intrafraction motion monitoring is clear with mounting evidence. In the past 5 years, we have witnessed the growing application of AI for target segmentation on kV images to address the gap in intrafraction monitoring solutions. With AI, more robust solutions for implanted marker tracking was developed with claimed sensitivity up to 99.4%.[33] More excitingly, the power of AI was harnessed for markerless tracking of the target without markers despite the low soft tissue contrast on kV images, typically acquired on a standard linear accelerator. Our literature search also shows innovations of AI methods to monitor for changes in the full anatomy. Motion of a target with some periodicity can also be predicted ahead of time such that the frequency of measurements can be reduced with no decrease in tumour detection accuracy.

The use of AI for marker-based tracking approaches has been demonstrated to be highly accurate through several studies. All of the marker-based tracking approaches discussed achieved sub-millimetre accuracy.[33,44,45] Mylonas et al.[33] and Liang et al.[44] use different DL approaches due to the differences in the image acquisition methods. The tracking window approach proposed by Mylonas et al.[33] is effective for continuous fluoroscopic imaging where the markers stay within the tracking window in subsequent image acquisitions. However, this approach is not suitable for the discrete X-ray image acquisition method used by CyberKnife. To overcome this, Liang et al.[44] utilised the U-Net architecture to perform semantic segmentation of the markers in full-size X-ray images. The method developed by Mylonas et al.[33] was successfully applied to both regular and arbitrarily shaped markers. The performance of these systems is approaching the upper limit for accuracy compared to the ground truth. Due to the previously discussed disadvantages of marker-based tracking, future work would more likely be focused on implementing markerless-based approaches.

To date, several studies have demonstrated the potential to eliminate the use of markers altogether through markerless tracking.[35–37,50–55,57,59–64] The implementation of markerless tracking is considered the ideal technique for the patient and the healthcare system due to its non-invasiveness. Markerless-based approaches for radiotherapy using AI have been developed for the pancreas, prostate, liver, lungs and vertebrae thus far. The ML approaches were only applied to lung tumour tracking,[50,54,55,59,62] while DL approaches were used for other sites. This may indicate that ML is outperformed by DL at sites that have a lower contrast. Several studies applied an R-CNN approach for tracking the target.[35–37,52] The R-CNN approach helps to reduce the computation time by producing region proposals. This is ideal for real-time applications since it is important to have minimal latency. The majority of markerless tracking approaches in cine MRI images focused on lung tumour. Other low-contrast sites such as the pancreas and prostate stand to benefit from MRI-guided radiotherapy due to improved tissue contrast. The markerless approaches do not achieve the same level of accuracy in comparison to marker-based approaches. Therefore, there is still the opportunity to further improve the performance of these systems.

Another approach for IGRT is the monitoring of changes based on the full surrounding anatomy. The majority of these methods do not produce a geometric measurement that can allow for real-time motion compensation. These approaches are ideal for performing gating during the treatment. Several studies used binary classifiers to determine whether the target is within view for gating purposes.[34,67,68] Novel single-view reconstruction methods show potential for use during IGRT [66,69]. Single projections acquired during the radiotherapy fraction could be used for 3D reconstruction during the treatment. The reconstructed volumes could be compared to the pre-treatment CBCT to determine whether intrafraction shifts have occurred. However, the typical time for the 3D reconstruction method developed by Shen et al. is around 0.5 s.[66] Therefore, further refinement is required to improve the efficiency of this method to allow for real-time implementation. For ultrasound-guided radiotherapy, a CNN can perform image registration of 3D ultrasound prostate images for motion management.[71] Proton therapy could benefit from the implementation of the aforementioned algorithms. Clinical results of proton therapy to date have not met the original high expectations. A commonly cited reason for this discrepancy is the lack of effective image guidance for proton therapy.[111–113] One prospective randomised trial for lung cancer estimated a threefold reduction in lung toxicity with proton therapy compared with X-ray therapy from 15% to 5%. The actual results in practice were the opposite, with 11% lung toxicity (radiation pneumonitis grade ≥3) in the proton therapy arm and 7% in the X-ray arm.[114] To this end, using AI to reconstruct real-time 3D volumes could be useful in both target localisation and 3D range uncertainty quantification during proton therapy treatments of internal targets.

For motion adaptation, a latency typically exists from the time the motion was observed and when the radiotherapy system can adapt it. The latency of the motion management system needs to be minimised to enable

real-time adaption. Real-time is defined by the AAPM Task Group 264 as a system latency below 500 ms.[115] To achieve real-time motion adaption, there is a prolific body of work that looks at motion prediction using ML and DL methods. AI-based predictors have demonstrated high accuracy in predicting the motion of tumours in the lung and abdomen. While there were mixed results in earlier studies comparing various ML algorithms, ANNs were close to or were the best performing methods. DL approaches have further improved the accuracy of motion prediction more recent studies.

Finally, the use of AI in motion tracking is in its early days. At the time of this review, the number of AI papers for intrafraction motion monitoring is modest compared with other areas of radiotherapy such as automatic contouring. Yet, the change was dramatic and hinted at unprecedented potential. Solutions such as markerless tracking of the liver, pancreas or prostate, which were often cited as the reason for MRI-linac use are no longer far-fetched ideas for standard linear accelerators. However, most of the featured studies are initial tests conducted with often off-the-shelf algorithms, adopted from other areas of medicine and computer vision and small data sets. Furthermore, there is a range of potential pitfalls. Similar to other areas of medicine where quality assurance is key to ensure patient safety, there is currently a lack of standardisation in data sets and reporting metrics. Many studies use DRRs due to a lack of large, annotated data sets. The development of publicly available data sets and the standardisation of clinically relevant evaluation metrics would enable the growth and benchmarking of AI solutions. A major concern is the robustness of algorithms since the performance of the algorithms may be correlated to the particular data set used for training.[116] For example, if an organisation upgraded the hardware or software in their imaging system, the imaging data may change in some form. This data set shift could potentially affect the performance of the AI algorithm that was developed using data prior to the change. Large publicly available data sets have the potential to solve this and improve the robustness of the algorithms, reducing the effects of data set shifts. Ethical and legal concerns are another key issue that may prevent clinical implementation. Data privacy is one challenge that organisations will have to deal with in order to create large, annotated data sets for the development of AI solutions. Patients may have concerns about who has access to their data, whether their data are being used for subsequent projects or being commercialised, and the risk of privacy breaches.[117] There is also the potential for bias to occur if the models are trained and then deployed on different patient demographics, leading to specific patient populations receiving improved treatments over others.[116]

While there are several concerns associated with AI that make clinical implementation difficult, ML and DL are becoming increasingly explored for motion tracking due to the apparent strengths of these approaches. With the surge of AI publications in other research fields including medical imaging, there are many well-established algorithms.[109,110] Therefore, it is relatively easy to choose and implement an already established algorithm depending on the specific application. Another advantage is that many of the discussed AI approaches have low latency during inference which meet requirements outlined by the AAPM Task Group 264.[115] As discussed above, the crucial next steps to translate AI solutions for intrafraction motion monitoring are (i) the development of large annotated database and (ii) standardisation of the reporting metrics. This will help to ensure that the AI algorithms are robust to different systems and populations, achieving similar impact of the ImageNet data set on the development of deep learning.[118] Furthermore, the development of consensus regulations and guidelines regarding the use of AI would be beneficial and help enable clinical implementation. At the time of writing, the authors believe that we are at the inflection point for real-time intrafraction motion monitoring and the use of AI for motion tracking will continue to develop and increase.

Conclusion

The growth of AI in the field of medicine and image processing has led to its implementation in motion monitoring during radiotherapy. AI approaches have shown great success in tracking markers with improved accuracy compared to traditional methods. Furthermore, AI methods have demonstrated the potential to eliminate the use of markers altogether through markerless tracking. The algorithms for monitoring intrafraction changes in the full anatomy demonstrate the potential for monitoring the full anatomy in the field of view. There is a latency associated with these AI approaches and other traditional methods due to the time for computation and motion compensation. However, AI has also been demonstrated to solve this problem by accurately predicting future motion. With the continued improvement of AI techniques, we can expect to see further refinement of monitoring the target during radiotherapy using a standard linear accelerator.

Acknowledgements

We thank the three anonymous reviewers whose comments helped improve this manuscript. We would also like to thank Dr Mark Gardner and Marco Mueller for reviewing this manuscript and providing helpful suggestions.

Funding

D T Nguyen is funded by an Early Career Research Fellowship from the Australian National Health and Medical

Research Council (NHMRC) and the Cancer Institute of New South Wales: (i) Cancer Institute NSW Early Career Fellowship 2018-ECF007; (ii) NHMRC Council Early Career Fellowship GNT1138807.

DATA AVAILABILITY STATEMENT

Data sharing is not applicable to this article as no new data were created or analyzed in this study.

References

1. Videtic GMM, Donington J, Giuliani M *et al*. Stereotactic body radiation therapy for early-stage non-small cell lung cancer: Executive Summary of an ASTRO Evidence-Based Guideline. *Pract Radiat Oncol* 2017; **7**: 295–301.
2. Morgan SC, Hoffman K, Loblaw A *et al*. ASTRO/ASCO/AUA guideline on hypofractionation for localized prostate cancer. *Practical Radiation Oncology* 2018; **8**: 354–60.
3. Aubry J-F, Beaulieu L, Girouard L-M *et al*. Measurements of intrafraction motion and interfraction and intrafraction rotation of prostate by three-dimensional analysis of daily portal imaging with radiopaque markers. *Int J Radiat Oncol Biol Phys* 2004; **60**: 30–9.
4. Kupelian P, Willoughby T, Mahadevan A *et al*. Multi-institutional clinical experience with the Calypso System in localization and continuous, real-time monitoring of the prostate gland during external radiotherapy. *Int J Radiat Oncol Biol Phys* 2007; **67**: 1088–98.
5. Deutschmann H, Kametriser G, Steininger P *et al*. First clinical release of an online, adaptive, aperture-based image-guided radiotherapy strategy in intensity-modulated radiotherapy to correct for inter- and intrafractional rotations of the prostate. *Int J Radiat Oncol Biol Phys* 2012; **83**: 1624–32.
6. Kotte ANTJ, Hofman P, Lagendijk JJW, van Vulpen M, van der Heide UA. Intrafraction motion of the prostate during external-beam radiation therapy: analysis of 427 patients with implanted fiducial markers. *Int J Radiat Oncol Biol Phys* 2007; **69**: 419–25.
7. Li JS, Jin L, Pollack A *et al*. Gains from real-time tracking of prostate motion during external beam radiation therapy. *Int J Radiat Oncol Biol Phys* 2009; **75**: 1613–20.
8. Amro H, Hamstra DA, Mcshan DL *et al*. The dosimetric impact of prostate rotations during electromagnetically guided external beam radiation therapy. *Int J Radiat Oncol Biol Phys* 2013; **85**: 230–6.
9. Kerkhof EM, van der Put RW, Raaymakers BW, van der Heide UA, Jürgenliemk-Schulz IM, Lagendijk JJW. Intrafraction motion in patients with cervical cancer: The benefit of soft tissue registration using MRI. *Radiother Oncol* 2009; **93**: 115–21.
10. Bertholet J, Worm ES, Fledelius W, Høyer M, Poulsen PR. Time-resolved intrafraction target translations and rotations during stereotactic liver radiation therapy: implications for marker-based localization accuracy. *Int J Radiat Oncol Biol Phys* 2016; **95**: 802–9.
11. Worm ES, Bertholet J, Høyer M *et al*. Fiducial marker guided stereotactic liver radiotherapy: is a time delay between marker implantation and planning CT needed? *Radiother Oncol* 2016; **121**: 75–8.
12. Huang C-Y, Tehrani JN, Ng JA, Booth JT, Keall PJ. Six degrees-of-freedom prostate and lung tumor motion measurements using kilovoltage intrafraction monitoring. *Int J Radiat Oncol Biol Phys* 2015; **91**: 368–75.
13. Steiner E, Shieh C-C, Caillet V *et al*. Both four-dimensional computed tomography and four-dimensional cone beam computed tomography under-predict lung target motion during radiotherapy. *Radiother Oncol* 2019; **135**: 65–73.
14. Jones BL, Schefter T, Miften M. Adaptive motion mapping in pancreatic SBRT patients using Fourier transforms. *Radiother Oncol* 2015; **115**: 217–22.
15. Guzhva L, Flampouri S, Mendenhall NP, Morris CG, Hoppe BS. Intrafractional displacement of cardiac substructures among patients with mediastinal lymphoma or lung cancer. *Adv Radiation Oncology* 2019; **4**: 500–6.
16. Lydiard S, Pontré B, Lowe BS, Ball H, Sasso G, Keall P. Cardiac radioablation for atrial fibrillation: target motion characterization and treatment delivery considerations. *Med Phys* 2021; **48**: 931–41.
17. Booth JT, Caillet V, Hardcastle N *et al*. The first patient treatment of electromagnetic-guided real time adaptive radiotherapy using MLC tracking for lung SABR. *Radiother Oncol* 2016; **121**: 19–25.
18. Caillet V, Keall PJ, Colvill E *et al*. MLC tracking for lung SABR reduces planning target volumes and dose to organs at risk. *Radiother Oncol* 2017; **124**: 18–24.
19. Poulsen PR, Worm ES, Hansen R, Larsen LP, Grau C, Høyer M. Respiratory gating based on internal electromagnetic motion monitoring during stereotactic liver radiation therapy: First Results. *Acta Oncol* 2016; **54**: 1445–52.
20. Colvill E, Poulsen PR, Booth JT, O'Brien RT, Ng JA, Keall PJ. DMLC tracking and gating can improve dose coverage for prostate VMAT. *Med Phys* 2014; **41**.
21. Shirato H, Shimizu S, Kunieda T *et al*. Physical aspects of a real-time tumor-tracking system for gated radiotherapy. *Int J Radiat Oncol Biol Phys* 2000; **48**: 1187–95.
22. King CR, Brooks JD, Gill H, Pawlicki T, Cotrutz C, Presti JC. Stereotactic body radiotherapy for localized prostate cancer: Interim results of a prospective phase II clinical trial. *Int J Radiat Oncol Biol Phys* 2009; **73**: 1043–8.
23. Kitamura K, Shirato H, Shimizu S *et al*. Registration accuracy and possible migration of internal fiducial gold marker implanted in prostate and liver treated with real-time tumor-tracking radiation therapy (RTRT). *Int J Radiat Oncol Biol Phys* 2002; **62**: 275–81.
24. Sazawa A, Shinohara N, Harabayashi T, Abe T, Shirato H, Nonomura K. Alternative approach in the treatment

of adrenal metastasis with a real-time tracking radiotherapy in patients with hormone refractory prostate cancer. *Int J Urol* 2009; **16**: 410–2.
25. Shimizu S, Shirato H, Kitamura K *et al*. Use of an implanted marker and real-time tracking of the marker for the positioning of prostate and bladder cancers. *Int J Radiat Oncol Biol Phys* 2000; **48**: 1591–7.
26. Shirato H, Harada T, Harabayashi T *et al*. Feasibility of insertion/implantation of 2.0-mm-diameter gold internal fiducial markers for precise setup and real-time tumor tracking in radiotherapy. *Int J Radiat Oncol Biol Phys* 2003; **56**: 240–7.
27. Ozhasoglu C, Saw CB, Chen H *et al*. Synchrony – Cyberknife Respiratory Compensation Technology. *Med Dosim* 2008; **33**: 117–23.
28. Jin J-Y, Yin F-F, Tenn SE, Medin PM, Solberg TD. Use of the BrainLAB ExacTrac X-Ray 6D system in image-guided radiotherapy. *Med Dosim* 2008; **33**: 124–34.
29. Castellanos E, Ericsson MH, Sorcini B, Green U, Nilsson S, Lennernäs B. RayPilot – electromagnetic real-time positioning in radiotherapy of prostate cancer – initial clinical results. *Radiother Oncol* 2012; **103**(Suppl 1): S433.
30. Ballhausen H, Li M, Hegemann N-S, Ganswindt U, Belka C. Intra-fraction motion of the prostate is a random walk. *Phys Med Biol* 2015; **60**: 549–63.
31. Fallone BG, Murray B, Rathee S *et al*. First MR images obtained during megavoltage photon irradiation from a prototype integrated linac-MR system. *Med Phys* 2009; **36**: 2084–8.
32. Raaymakers BW, Lagendijk JJW, Overweg J *et al*. Integrating a 1.5 T MRI scanner with a 6 MV accelerator: proof of concept. *Phys Med Biol* 2009; **54**: N229–37.
33. Mylonas A, Keall PJ, Booth JT *et al*. A deep learning framework for automatic detection of arbitrarily shaped fiducial markers in intrafraction fluoroscopic images. *Med Phys* 2019; **46**: 2286–97.
34. Lin T, Li R, Tang X, Dy JG, Jiang SB. Markerless gating for lung cancer radiotherapy based on machine learning techniques. *Phys Med Biol* 2009; **54**: 1555.
35. Roggen T, Bobic M, Givehchi N, Scheib SG. Deep Learning model for markerless tracking in spinal SBRT. *Physica Med* 2020; **74**: 66–73.
36. Zhao W, Han B, Yang Y *et al*. Incorporating imaging information from deep neural network layers into image guided radiation therapy (IGRT). *Radiother Oncol* 2019; **140**: 167–74.
37. Zhao W, Shen L, Han B *et al*. Markerless pancreatic tumor target localization enabled by deep learning. International Journal of Radiation Oncology* Biology*. *Physics* 2019; **105**: 432–9.
38. Hastie T, Tibshirani R, Friedman J. *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. Springer Science & Business Media, New York, 2009.
39. Bertholet J, Wan H, Toftegaard J *et al*. Fully automatic segmentation of arbitrarily shaped fiducial markers in cone-beam CT projections. *Phys Med Biol* 2017; **62**: 1327.
40. Fledelius W, Worm E, Elstrøm UV *et al*. Robust automatic segmentation of multiple implanted cylindrical gold fiducial markers in cone-beam CT projections. *Med Phys* 2011; **38**: 6351–61.
41. Fledelius W, Worm E, Høyer M, Grau C, Poulsen P. Real-time segmentation of multiple implanted cylindrical liver markers in kilovoltage and megavoltage x-ray images. *Phys Med Biol* 2014; **59**: 2787.
42. Poulsen PR, Fledelius W, Keall PJ *et al*. A method for robust segmentation of arbitrarily shaped radiopaque structures in cone-beam CT projections. *Med Phys* 2011; **38**: 2151–6.
43. Regmi R, Lovelock DM, Hunt M *et al*. Automatic tracking of arbitrarily shaped implanted markers in kilovoltage projection images: a feasibility study. *Med Phys* 2014; **41**: 71906.
44. Liang Z, Zhou Q, Yang J *et al*. Artificial intelligence-based framework in evaluating intrafraction motion for liver cancer robotic stereotactic body radiation therapy with fiducial tracking. *Med Phys* 2020; **47**: 5482–9.
45. Lin WY, Lin SF, Yang SC, Liou SC, Nath R, Liu W. Real-time automatic fiducial marker tracking in low contrast cine-MV images. *Med Phys* 2013; **40**: 11715.
46. Kothary N, Heit JJ, Louie JD *et al*. Safety and efficacy of percutaneous fiducial marker implantation for image-guided radiation therapy. *J Vasc Interv Radiol* 2009; **20**: 235–9.
47. Park WG, Yan BM, Schellenberg D *et al*. EUS-guided gold fiducial insertion for image-guided radiation therapy of pancreatic cancer: 50 successful cases without fluoroscopy. *Gastrointest Endosc* 2010; **71**: 513–8.
48. Bhagat N, Fidelman N, Durack JC *et al*. Complications associated with the percutaneous insertion of fiducial markers in the thorax. *Cardiovasc Intervent Radiol* 2010; **33**: 1186–91.
49. Lim YK, Kwak J, Kim DW *et al*. Microscopic gold particle-based fiducial markers for proton therapy of prostate cancer. *Int J Radiat Oncol Biol Phys* 2009; **74**: 1609–16.
50. Lin T, Cervino LI, Tang X, Vasconcelos N, Jiang SB. Fluoroscopic tumor tracking for image-guided lung cancer radiotherapy. *Phys Med Biol* 2009; **54**: 981.
51. Terunuma T, Tokui A, Sakae T. Novel real-time tumor-contouring method using deep learning to prevent mistracking in X-ray fluoroscopy. *Radiol Phys Technol* 2018; **11**: 43–53.
52. Edmunds D, Sharp G, Winey B. Automatic diaphragm segmentation for real-time lung tumor tracking on cone-beam CT projections: a convolutional neural network approach. *Biomed Phys Eng Express* 2019; **5**: 35005.
53. Hirai R, Sakata Y, Tanizawa A, Mori S. Real-time tumor tracking using fluoroscopic imaging with deep neural network analysis. *Physica Med* 2019; **59**: 22–9.
54. Mori S, Sakata Y, Hirai R *et al*. Commissioning of a fluoroscopic-based real-time markerless tumor tracking system in a superconducting rotating gantry for carbon-ion pencil beam scanning treatment. *Med Phys* 2019; **46**: 1561–74.

55. Sakata Y, Hirai R, Kobuna K, Tanizawa A, Mori S. A machine learning-based real-time tumor tracking system for fluoroscopic gating of lung radiotherapy. *Phys Med Biol* 2020; **65**: 85014.
56. Takahashi W, Oshikawa S, Mori S. Real-time markerless tumour tracking with patient-specific deep learning using a personalised data generation strategy: proof of concept by phantom study. *Br J Radiol* 2020; **93**: 20190420.
57. Mori S, Hirai R, Sakata Y. Simulated four-dimensional CT for markerless tumor tracking using a deep learning network with multi-task learning. *Physica Med* 2020; **80**: 151–8.
58. Fürweger C, Drexler C, Kufeld M, Muacevic A, Wowra B, Schlaefer A. Patient motion and targeting accuracy in robotic spinal radiosurgery: 260 single-fraction fiducial-free cases. *Int J Radiat Oncol Biol Physics* 2010; **78**: 937–45.
59. Cervino LI, Du J, Jiang SB. MRI-guided tumor tracking in lung cancer radiotherapy. *Phys Med Biol* 2011; **56**: 3773.
60. Yun J, Yip E, Gabos Z, Wachowicz K, Rathee S, Fallone B. Neural-network based autocontouring algorithm for intrafractional lung-tumor tracking using Linac-MR. *Med Phys* 2015; **42**: 2296–310.
61. Yun J, Yip E, Gabos Z, Wachowicz K, Rathee S, Fallone B. Improved lung tumor autocontouring algorithm for intrafractional tumor tracking using 0.5 T linac-MR. *Biomed Phys Eng Express* 2016; **2**: 67004.
62. Bourque AE, Carrier J-F, Filion É, Bedwani S. A particle filter motion prediction algorithm based on an autoregressive model for real-time MRI-guided radiotherapy of lung cancer. *Biomed Phys Eng Express* 2017; **3**: 35001.
63. Fast MF, Eiben B, Menten MJ *et al*. Tumour auto-contouring on 2d cine MRI for locally advanced lung cancer: a comparative study. *Radiother Oncol* 2017; **125**: 485–91.
64. Friedrich F, Hörner-Rieber J, Renkamp CK *et al*. Stability of conventional and machine learning-based tumor auto-segmentation techniques using undersampled dynamic radial bSSFP acquisitions on a 0.35 T hybrid MR-linac system. *Med Phys* 2021; **48**: 587–96.
65. Cusumano D, Boldrini L, Dhont J *et al*. Artificial Intelligence in magnetic Resonance guided Radiotherapy: Medical and physical considerations on state of art and future perspectives. *Physica Med* 2021; **85**: 175–91.
66. Shen L, Zhao W, Xing L. Patient-specific reconstruction of volumetric computed tomography images from a single projection view via deep learning. *Nat Biomed Eng* 2019; **3**: 880–8.
67. Tang X, Lin T, Jiang S. A feasibility study of treatment verification using EPID cine images for hypofractionated lung radiotherapy. *Phys Med Biol* 2009; **54**: S1.
68. Cui Y, Dy JG, Alexander B, Jiang SB. Fluoroscopic gating without implanted fiducial markers for lung cancer radiotherapy based on support vector machines. *Phys Med Biol* 2008; **53**: N315.
69. Xu Y, Yan H, Ouyang L *et al*. A method for volumetric imaging in radiotherapy using single x-ray projection. *Med Phys* 2015; **42**: 2498–509.
70. Terpstra ML, Maspero M, d'Agata F *et al*. Deep learning-based image reconstruction and motion estimation from undersampled radial k-space for real-time MRI-guided radiotherapy. *Phys Med Biol* 2020; **65**: 155015.
71. Zhu N, Najafi M, Han B, Hancock S, Hristov D. Feasibility of image registration for ultrasound-guided prostate radiotherapy based on similarity measurement by a convolutional neural network. *Technol Cancer Res Treat* 2019; **18**: 1533033818821964.
72. Balasubramanian A, Shamsuddin R, Prabhakaran B, Sawant A. Predictive modeling of respiratory tumor motion for real-time prediction of baseline shifts. *Phys Med Biol* 2017; **62**: 1791.
73. Bukhari W, Hong S. Real-time prediction and gating of respiratory motion using an extended Kalman filter and Gaussian process regression. *Phys Med Biol* 2014; **60**: 233.
74. Bukovsky I, Homma N, Ichiji K *et al*. A fast neural network approach to predict lung tumor motion during respiration for radiation therapy applications. *Biomed Res Int* 2015; **2015**: 1–13.
75. Dürichen R, Wissel T, Ernst F, Schweikard A, editors. Respiratory motion compensation with relevance vector machines. International Conference on Medical Image Computing and Computer-Assisted Intervention; 2013: Springer.
76. Goodband JH, Haas OC, Mills J. A comparison of neural network approaches for on-line prediction in IGRT. *Med Phys* 2008; **35**: 1113–22.
77. Hong S, Bukhari W. Real-time prediction of respiratory motion using a cascade structure of an extended Kalman filter and support vector regression. *Phys Med Biol* 2014; **59**: 3555.
78. Isaksson M, Jalden J, Murphy MJ. On using an adaptive neural network to predict lung tumor motion during respiration for radiotherapy applications. *Med Phys* 2005; **32**: 3801–9.
79. Jiang K, Fujii F, Shiinoki T. Prediction of lung tumor motion using nonlinear autoregressive model with exogenous input. *Phys Med Biol* 2019; **64**: 21NT02
80. Jurkovic I-A, Stathakis S, Papanikolaou N, Mavroidis P. Prediction of lung tumor motion extent through artificial neural network (ANN) using tumor size and location data. *Biomed Phys Eng Express* 2016; **2**: 25012.
81. Mafi M, Moghadam SM. Real-time prediction of tumor motion using a dynamic neural network. *Med Biol Eng Comput* 2020; **58**: 529–39.
82. Murphy MJ, Pokhrel D. Optimization of an adaptive neural network to predict breathing. *Med Phys* 2009; **36**: 40–7.
83. Riaz N, Shanker P, Wiersma R *et al*. Predicting respiratory tumor motion with multi-dimensional

adaptive filters and support vector regression. *Phys Med Biol* 2009; **54**: 5735.
84. Ruan D, Keall P. Online prediction of respiratory motion: multidimensional processing with low-dimensional feature learning. *Phys Med Biol* 2010; **55**: 3011.
85. Sun W, Jiang M, Ren L, Dang J, You T, Yin F. Respiratory signal prediction based on adaptive boosting and multi-layer perceptron neural network. *Phys Med Biol* 2017; **62**: 6822.
86. Sun W, Wei Q, Ren L, Dang J, Yin F-F. Adaptive respiratory signal prediction using dual multi-layer perceptron neural networks. *Phys Med Biol* 2020; **65**: 185005.
87. Teo TP, Ahmed SB, Kawalec P *et al*. Feasibility of predicting tumor motion using online data acquired during treatment and a generalized neural network optimized with offline patient tumor trajectories. *Med Phys* 2018; **45**: 830–45.
88. Torshabi AE, Pella A, Riboldi M, Baroni G. Targeting accuracy in real-time tumor tracking via external surrogates: a comparative study. *Technol Cancer Res Treat* 2010; **9**: 551–61.
89. Yan H, Yin F-F, Zhu G-P, Ajlouni M, Kim JH. Adaptive prediction of internal target motion using external marker motion: a technical study. *Phys Med Biol* 2005; **51**: 31.
90. Dhont J, Vandemeulebroucke J, Cusumano D *et al*. Multi-object tracking in MRI-guided radiotherapy using the tracking-learning-detection framework. *Radiother Oncol* 2019; **138**: 25–9.
91. Mirzapour SA, Mazur T, Sharp G, Salari E. Intra-fraction motion prediction in MRI-guided radiation therapy using Markov processes. *Phys Med Biol* 2019; **64**: 195006.
92. Yun J, Mackenzie M, Rathee S, Robinson D, Fallone B. An artificial neural network (ANN)-based lung-tumor motion predictor for intrafractional MR tumor tracking. *Med Phys* 2012; **39**: 4423–33.
93. Prall M, Kaderka R, Saito N *et al*. Ion beam tracking using ultrasound motion detection. *Med Phys* 2014; **41**: 41708.
94. Sharp GC, Jiang SB, Shimizu S, Shirato H. Prediction of respiratory tumour motion for real-time image-guided radiotherapy. *Phys Med Biol* 2004; **49**: 425.
95. Murphy MJ, Dieterich S. Comparative performance of linear and nonlinear neural networks to predict irregular breathing. *Phys Med Biol* 2006; **51**: 5903.
96. Ruan D, Fessler JA, Balter J. Real-time prediction of respiratory motion based on local regression methods. *Phys Med Biol* 2007; **52**: 7137.
97. Krauss A, Nill S, Oelfke U. The comparative performance of four respiratory motion predictors for real-time tumour tracking. *Phys Med Biol* 2011; **56**: 5303.
98. Dick D, Wu X, Hatoum GF, Zhao W. A fiducial-less tracking method for radiation therapy of liver tumors by diaphragm disparity analysis part 2: validation study by using clinical data. *J Radiat Oncol* 2018; **7**: 345–56.
99. Dick D, Wu X, Hatoum GF, Zhao W. A fiducial-less tracking method for radiation therapy of liver tumors by diaphragm disparity analysis part 1: simulation study using machine learning through artificial neural network. *J Radiat Oncol* 2018; **7**: 275–84.
100. Lee SJ, Motai Y, Weiss E, Sun SS. Customized prediction of respiratory motion with clustering from multiple patient interaction. *ACM Transact Intelligent Syst Technol* 2013; **4**: 1–17.
101. Lin H, Shi C, Wang B, Chan MF, Tang X, Ji W. Towards real-time respiratory motion prediction based on long short-term memory neural networks. *Phys Med Biol* 2019; **64**: 85010.
102. Park S, Lee SJ, Weiss E, Motai Y. Intra-and inter-fractional variation prediction of lung tumors using fuzzy deep learning. *IEEE J Transl Eng Health Med* 2016; **4**: 1–12.
103. Romaguera LV, Plantefève R, Romero FP, Hébert F, Carrier J-F, Kadoury S. Prediction of in-plane organ deformation during free-breathing radiotherapy via discriminative spatial transformer networks. *Med Image Anal* 2020; **64**: 101754.
104. Torshabi AE. Investigation of the robustness of adaptive neuro-fuzzy inference system for tracking moving tumors in external radiotherapy. *Australas Phys Eng Sci Med* 2014; **37**: 771–8.
105. Wang G, Li Z, Li G *et al*. Real-time liver tracking algorithm based on LSTM and SVR networks for use in surface-guided radiation therapy. *Radiat Oncol* 2021; **16**: 1–12.
106. Kim KH, Park K, Kim H *et al*. Facial expression monitoring system for predicting patient's sudden movement during radiotherapy using deep learning. *J Appl Clin Med Phys* 2020; **21**: 191–9.
107. Ma Y, Wang Z, Yang H, Yang L. Artificial intelligence applications in the development of autonomous vehicles: a survey. *IEEE/CAA J Automatica Sinica* 2020; **7**: 315–29.
108. Silver D, Schrittwieser J, Simonyan K *et al*. Mastering the game of go without human knowledge. *Nature* 2017; **550**: 354–9.
109. Savadjiev P, Chong J, Dohan A *et al*. Demystification of AI-driven medical image interpretation: past, present and future. *Eur Radiol* 2019; **29**: 1616–24.
110. Ker J, Wang L, Rao J, Lim T. Deep learning applications in medical image analysis. *Ieee Access.* 2017; **6**: 9375–89.
111. Oborn BM, Dowdell S, Metcalfe PE, Crozier S, Mohan R, Keall PJ. Future of medical physics: real-time MRI-guided proton therapy. *Med Phys* 2017; **44**: e77–90.
112. Liao Z, Mohan R. Future of Protons Depends on Precision. Journal of clinical oncology: official journal of the American Society of. *Clin Oncol (R Coll Radiol)* 2018; **36**: 2002.
113. Beddok A, Vela A, Calugaru V *et al*. Proton therapy for head and neck squamous cell carcinomas: a review of the physical and clinical challenges. *Radiother Oncol* 2020; **147**: 30–9.

114. Liao Z, Lee JJ, Komaki R *et al*. Bayesian adaptive randomization trial of passive scattering proton therapy and intensity-modulated photon radiotherapy for locally advanced non–small-cell lung cancer. *J Clin Oncol* 2018; **36**: 1813.
115. Keall PJ, Sawant A, Berbeco RI *et al*. AAPM Task Group 264: The safe clinical implementation of MLC tracking in radiotherapy. *Med Phys* 2021; **48**: e44-64.
116. Subbaswamy A, Saria S. From development to deployment: dataset shift, causality, and shift-stable models in health AI. *Biostatistics* 2020; **21**: 345–52.
117. Ahmad OF, Stoyanov D, Lovat LB. Barriers and pitfalls for artificial intelligence in gastroenterology: ethical and regulatory issues. *Tech Innov Gastrointest Endosc* 2020; **22**: 80–4.
118. Deng J, Dong W, Socher R, Li L-J, Li K, Fei-Fei L, editors. Imagenet: A large-scale hierarchical image database. 2009 IEEE conference on computer vision and pattern recognition; 2009: Ieee.

3

Patient-Specific Prostate Segmentation in Kilovoltage Images for Intrafraction Monitoring via Deep Learning

This chapter introduces the deep learning framework utilised throughout this thesis for real-time, markerless tracking on a standard linac. A conditional Generative Adversarial Network was applied to markerless prostate segmentation and was validated on multi-centre data from two clinical trials. Aim 1 is addressed by conducting the first large-scale proof of principle for x-ray-based markerless prostate segmentation on a standard linac. The framework demonstrated accuracy within several millimetres and no correlation between the observed motion and error, indicating its ability to detect cases of large motion. This chapter is under review in *Communication Medicine* as "**Mylonas, A.**, Li, Z., Mueller, M., Booth, J., Brown, R., Gardner, M., Kneebone, A., Eade, E., Keall, P., & Nguyen, D. T. Patient-specific prostate segmentation in kilovoltage images for radiation therapy intrafraction monitoring via deep learning."

3.1 Abstract

During radiation therapy, the natural movement of organs can lead to cancer underdose and healthy tissue overdose, compromising treatment efficacy. Real-time image-guided adaptive radiation therapy (IGART) can track the tumour and account for the motion. Typically, fiducial markers are implanted as a surrogate for the tumour position due to the low radiographic contrast of soft tissues in kilovoltage (kV) images. A segmentation approach that does not require markers would eliminate the costs, time delays, and risks inherent to marker implantation. In this study, we take an important step towards clinical implementation through the first large scale proof-of-principle of x-ray-based prostate markerless segmentation for globally available cancer therapy systems. We demonstrate the feasibility of a deep learning approach with kV images across an entire radiation therapy delivery arc from 30 prostate cancer patient treatments. Markerless segmentation via deep learning will allow for IGART on conventional cancer therapy systems without requiring additional hardware, leading to more efficient radiation therapy for more patients.

3.2 Introduction

Image guidance is vital for interventional procedures such as radiation therapy, needle biopsy and surgery. During radiation treatments, the tumour and surrounding anatomy are dynamically moving, in accordance with normal physiological processes. This makes image guidance a necessity throughout the duration of treatment to monitor tumour motion and, therefore, ensure adequate dose coverage of the tumour. Motion monitoring is essential for high dose treatments, such as stereotactic body radiation therapy (SBRT), where high radiation dose per treatment fraction is prescribed with small geometric margins, ultimately demanding high precision.[1,2] For prostate cancer, the effect of motion can result in up to 19% less radiation dose delivered to the prostate in one fraction compared to the prescribed dose per fraction.[3] Hewson et al.[1] found that SBRT prostate cancer patients treated without real-time motion adaptation resulted in underdosing to the clinical treatment volume (CTV) of up to -5.6% and overdosing to the rectum and bladder of up to 1.2% and 8.5% respectively. With mounting evidence on the detrimental effects of underdosing tumours and overdosing organs at risk during treatment, the American Society for Radiation Oncology recommended imaging during treatment

to continuously monitor the tumour motion for high dose radiation treatments.[4]

Real-time image-guided adaptive radiation therapy (IGART) can be used to estimate the tumour location during radiation therapy to improve dose coverage and reduce the radiation dose to healthy tissue.[5] IGART can be performed by acquiring kilovoltage (kV) images during treatment using the on-board imager that is routinely installed on conventional radiation therapy treatment systems. A robust segmentation method can be used to accurately determine the tumour position. For conventional therapy systems, real-time motion monitoring methods typically track implanted fiducial markers as surrogates to the tumour, especially for organs and tumours with low radiographic contrast, such as the prostate.[6–11] Fiducial markers and the implantation procedure adds time delays, additional costs, and risks. The treatment delays are a result of surgery wait time and the time for the markers to stabilise.[12] Some risks associated with the implantation of markers include infection, haematuria, rectal bleeding, and patient discomfort from the surgery.[13,14] Furthermore, marker migration can result in tracking errors,[15] and the metal artefacts produced from markers in computed tomography (CT) images may result in treatment image matching errors.[16] Currently, patients who are not candidates for marker implantation due to contraindications cannot receive real-time IGART. A worldwide survey relating to respiratory motion management found that 71% of radiation therapy centres wish to implement targeted radiation therapy but are limited by resources and capacity.[17] Ideally, real-time motion monitoring should not require any additional procedures or hardware. A markerless-based approach using a conventional therapy system would help make real-time IGART accessible to all patients without additional hardware.

In this study, we perform markerless prostate segmentation in kV images with a patient-specific model by leveraging deep learning. During treatment, the treatment beam rotates around the patient delivering radiation from multiple angles. It is important to track the prostate continuously to account for motion that may result in suboptimal tumour control and increased toxicities. Although random motion within the planning treatment volume (PTV) is accounted for in the prescribed dose, large motion can have significant dosimetric impact. For prostate radiation therapy, the MIRAGE trial[18] demonstrated reduced toxicities when using magnetic resonance imaging (MRI) guidance while the SPARK trial[19] achieved dose improvements using x-ray guidance. In the SPARK trial, the prostate CTV received a dose 5% less than the planned dose in 11% of treatments without real-time tracking, compared to 0% with real-time tracking.[19] In a Memorial Sloan Kettering trial, the prostate CTV received

a dose 5% less than the planned dose in 6% of patients without real-time tracking, compared to 0% with real-time tracking.[20] Furthermore, motion management is critical for the coverage of smaller targets or organs at risk (such as focal boosting[21] or urethral sparing), which are more susceptible to motion.

Recently, artificial intelligence (AI) for fluoroscopic and MRI guided radiation therapy has shown great potential.[22] Markerless-based approaches have been developed for lung,[23–25] diaphragm,[26] liver,[24] pancreas,[27,28] prostate,[29] and head and neck.[30] We take an important step towards clinical implementation through the first large scale proof-of-principle of x-ray-based markerless tracking for globally available cancer therapy systems (Fig. 3.1). Our method uses a deep learning model trained on synthetic two-dimensional (2D) images derived from the three-dimensional (3D) planning CT and prostate contour. Our study presents the full offline method development that can be incorporated into the treatment workflow, as shown in Fig. 3.2. Here we use a patient-specific conditional Generative Adversarial Network (cGAN) to segment the prostate in 2D kV images. A patient-specific model is advantageous as it requires less data than training a generalised model. Furthermore, it allows the model to learn features most relevant to the specific patient under treatment and can be applicable to patients imaged using different imaging systems. This patient-specific approach can eliminate potential biases that may be present in generalised models.[31] Our approach leverages the patient's own imaging and planning data that are available prior to the commencement of their treatment. The cGAN model was evaluated on two datasets using imaging data with and without markers from four different clinical sites in Australia. The datasets each have different uncertainties related to the ground truth creation. As such, we assessed the model's performance on both datasets to gauge its effectiveness considering these varying uncertainties. Our results indicate that the prostate can be segmented without markers in x-ray images with a high degree of accuracy.

3.3 Results

Fig. 3.1 provides a detailed overview of our deep learning study. For our study, two separate datasets were used to evaluate the performance of the cGAN segmentation: the masked dataset (Fig. 3.1a) and the markerless dataset (Fig. 3.1b). Ground truth annotation by clinicians in each kV image was not feasible due to the low soft tissue contrast of the images. There-

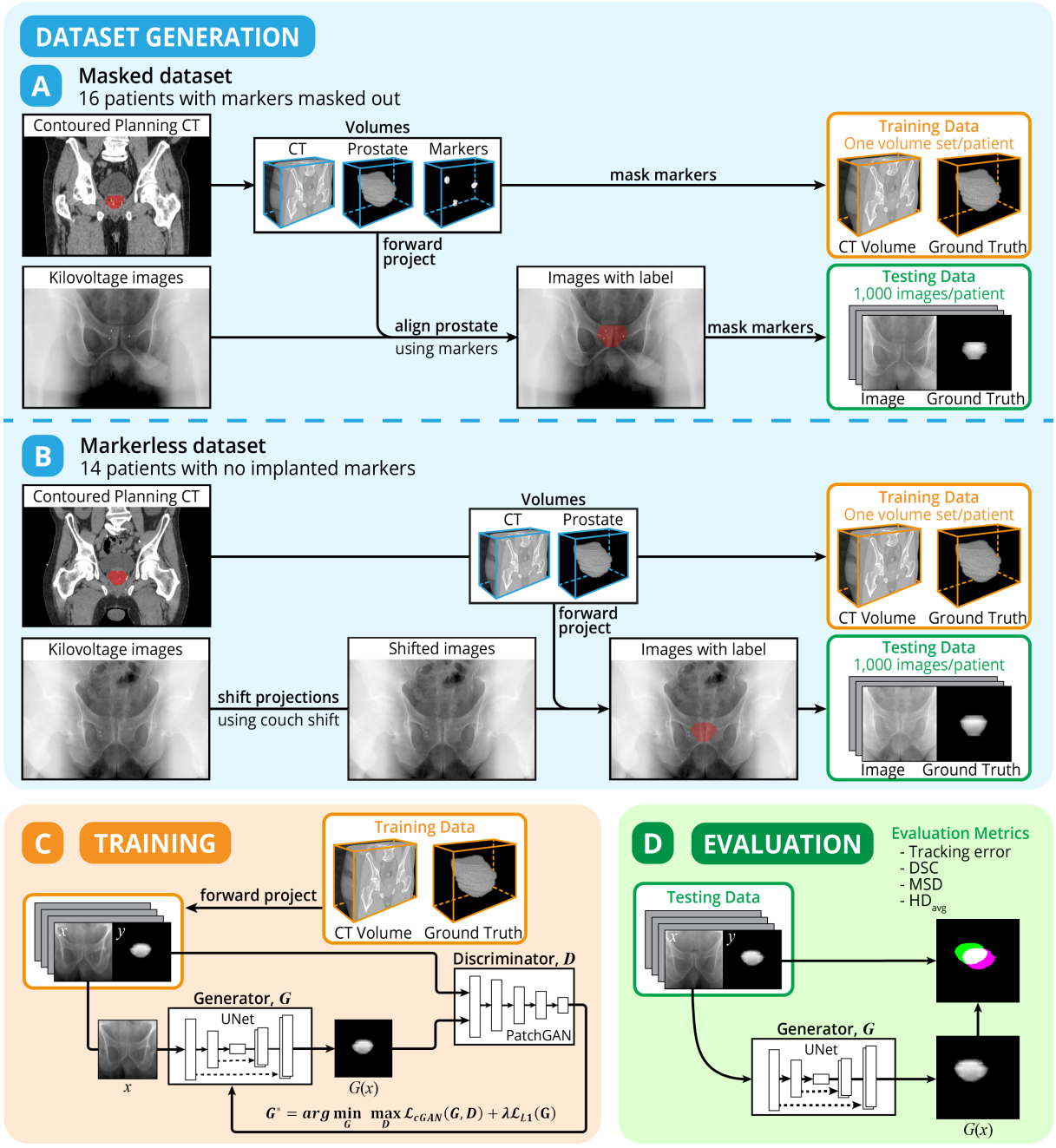


Fig. 3.1 | Study Overview. **(a)** The masked dataset was generated from imaging data of prostate cancer patients with implanted fiducial markers. The markers were used to align the ground truth contour and were masked out to avoid biasing the model. **(b)** The markerless dataset was generated from imaging data of prostate cancer patients with no implanted fiducial markers. The kilovoltage images were shifted based on the soft tissue registration for each fraction. **(c)** The data were used to train a conditional Generative Adversarial Network (cGAN) for each patient consisting of a UNet generator network, *G*, and a PatchGAN discriminator network, *D*. **(d)** The cGAN model was evaluated using the testing data and the performance was quantified using the centroid tracking error, Dice similarity coefficient (DSC), mean surface distance (MSD), and average Hausdorff distance (HD_{avg}).

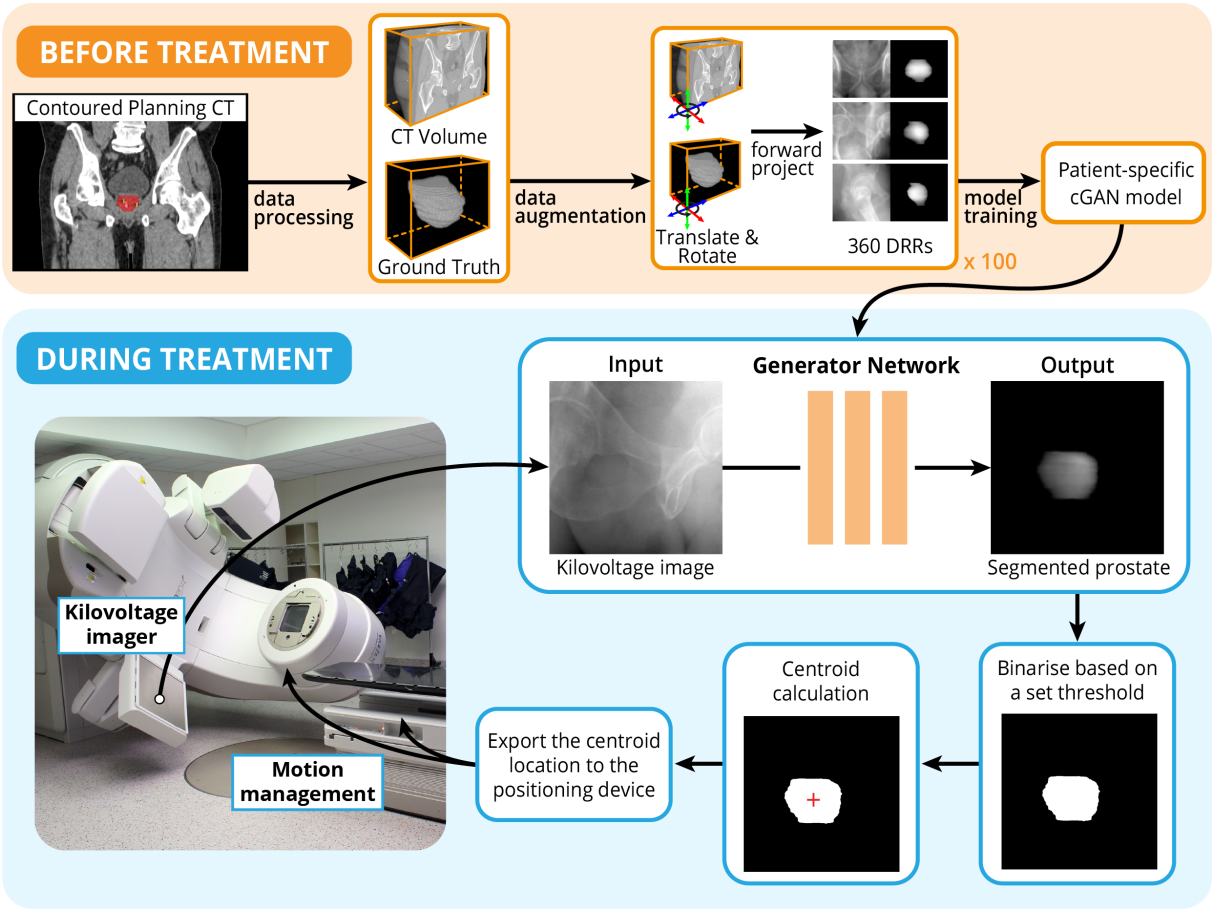


Fig. 3.2 | Simulated real-time clinical study of the deep learning method. The workflow for automatic prostate tracking is comprised of two key components: before treatment and during treatment. A patient-specific conditional Generative Adversarial Network (cGAN) is trained prior to the patient's treatment using 36,000 synthetic 2D images derived from the 3D planning CT and prostate contour data. The cGAN generator network is used during the treatment to segment the prostate. The location of the segmented prostate can be used for motion management.

fore, the ground truth was generated using a different approach for each dataset. The masked dataset was generated from imaging data of 16 prostate cancer patients with implanted fiducial markers. The markers were used to annotate the real-time location of the prostate in the kV images. During testing of the model, the real-time prediction of the prostate location could be compared with the ground-truth prostate location, using the implanted markers as surrogates. The fiducial markers were removed in the training and testing data to avoid biasing the deep learning model. Marker removal was achieved by smoothly interpolating inwards from the pixel values surrounding the marker with Poisson noise subsequently applied to the kV images. Manual visual inspection of each kV image was performed to ensure that the markers were masked and no longer visible.

The markerless dataset included imaging data of 14 prostate cancer patients with no im-

planted fiducial markers. A rigid shift of the kV images was applied based on an expert's registration performed between the planning CT and reconstructed cone beam computed tomography (CBCT). Therefore, the ground truth in the markerless dataset was defined by the average location of the prostate rather than the real-time location. The markerless dataset does not provide a gold standard for intrafraction motion as the ground truth is based on the average location. The observed motion may be a result of the prostate not being at the average location rather than the detection of intrafraction motion.

For this approach to be clinically feasible, the patient-specific model must be trained using data available prior to the patient's first treatment. In the conventional clinical workflow, a patient will receive a planning CT several days or weeks prior to the first treatment which is used by clinicians to contour the relevant volumes and plan the treatment. Therefore, we can use this available data to train the model with sufficient time prior to the patient's first treatment. The inputs to the model were the planning CT and prostate contour 3D volumes (Fig. 3.1c). The volumes were forward projected to produce 36,000 synthetic 2D images simulating kV images from different projections angles. The images produced from the planning CT were each paired with the respective image produced from the prostate contour. To segment the prostate in the 2D kV images, we used a cGAN with a UNet for the generator architecture and a PatchGAN for the discriminator architecture. Details about the network architecture are included in Methods and Extended Data Fig. 3.1. The mean (± standard deviation) training time including data generation and augmentation was 9.7 ± 0.7 hours with a maximum time of 10.7 hours when trained on a desktop computer with an Intel® Xeon® Gold 6248R processor (3.0 GHz) with 256 GB RAM and a NVIDIA® RTX A6000 Graphics Processing Unit (GPU).

We evaluated the models using kV images from two fractions of each patient's treatment (Fig. 3.1d). The kV images were acquired during patient setup to reconstruct the CBCT. The centroid errors of the cGAN segmentations for all patients are represented in Fig. 3.3 and Extended Data Tables 3.1 and 3.2. The centroid errors are presented in the anterior-posterior/lateral (AP/LAT) and superior-inferior (SI) directions with a pixel size of 0.26 mm. The AP and LAT directions are combined due to the rotational geometry of the treatment beam. For the masked dataset, the mean (± standard deviation) error across all patients was 0.7 ± 1.9 mm and -0.2 ± 1.9 mm in the AP/LAT and SI directions, respectively (Extended Data Table 3.1). The mean absolute deviation (MAD) was 1.4 mm and 1.5 mm in the AP/LAT and SI directions, respectively. The 5^{th} and 95^{th} percentiles were -2.4 and 4.0 mm in the AP/LAT

direction, and -3.4 and 3.5 mm in the SI direction (Extended Data Table 3.1). Similar performance was observed for the markerless dataset. The mean error for all patients was 0.1 ± 1.8 mm and -0.6 ± 1.9 mm in the AP/LAT and SI directions, respectively (Extended Data Table 3.2). The MAD was 1.4 mm and 1.6 mm in the AP/LAT and SI directions, respectively. The 5th and 95th percentiles were -2.8 and 3.1 mm in the AP/LAT direction, and -4.0 and 2.6 mm in the SI direction (Extended Data Table 3.2).

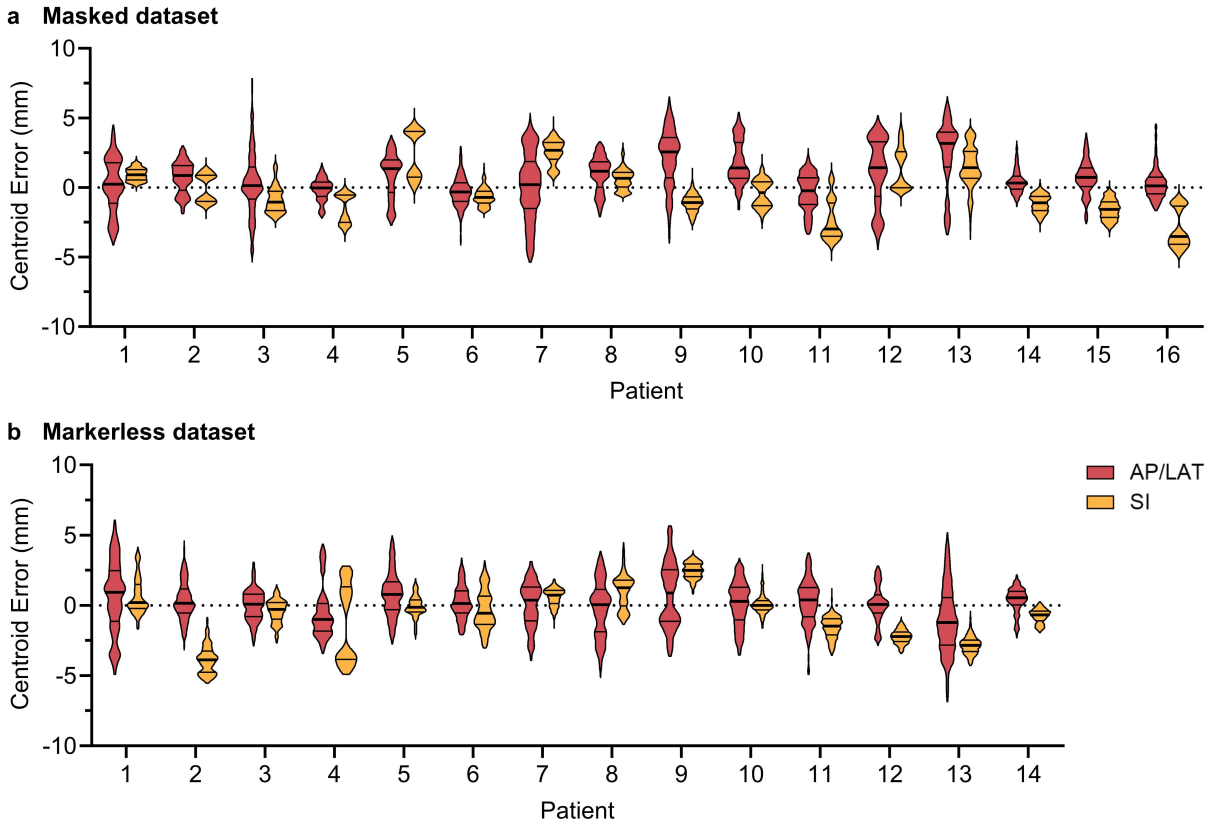


Fig. 3.3 | Violin plots of the centroid error results. The errors in the centroid location of the cGAN segmentation versus the ground truth for each patient in the anterior-posterior/lateral (AP/LAT) directions (red) and superior-inferior (SI) directions (yellow) for the **(a)** masked dataset and **(b)** markerless dataset. The bars represent the median (thick) and quartiles (thin).

The correlation between the magnitude of motion and the absolute errors for all patients in the masked dataset is shown in Fig. 3.4. The error was found to be independent of the observed motion with low Pearson correlation coefficients of 0.113 and -0.073 in the AP/LAT and SI directions, respectively. Fig. 3.4 shows a Bland-Altman plot of the observed motion versus the centroid error. The mean is low in both directions suggesting minimal bias between the measurements. The difference between the measurements does not tend to get larger as the average increases.

The performance of the cGAN in terms of the Dice similarity coefficient (DSC), mean sur-

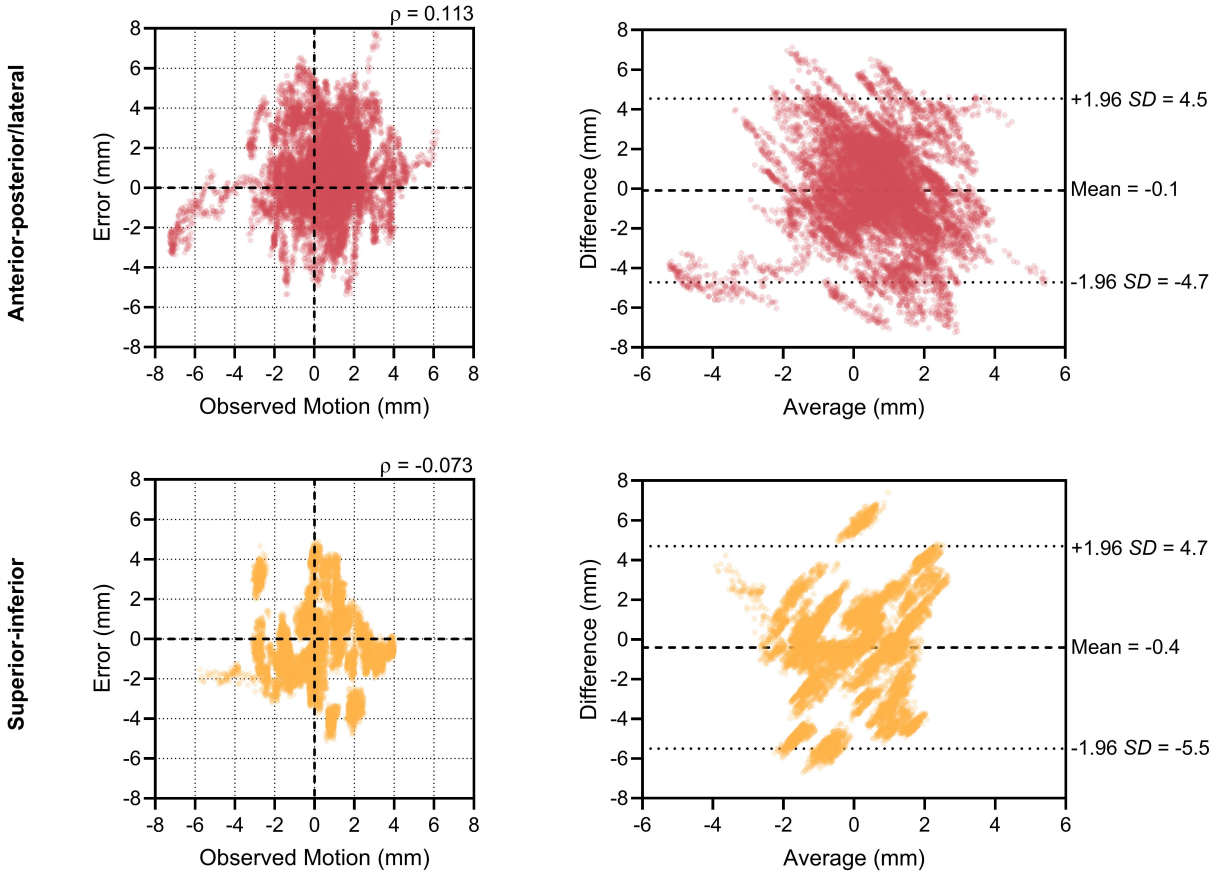


Fig. 3.4 | Comparisons between the observed motion and centroid error for the masked dataset. Scatter plots (left) comparing the observed motion and the error for each direction. The Pearson correlation coefficient is displayed on each scatter plot. Bland-Altman plots (right) comparing the observed motion and the error for each direction.

face distance (MSD), and average Hausdorff distance (HD_{avg}) is presented in Fig. 3.5 and Extended Data Tables 3.3 and 3.4. For the masked dataset, the predicted and ground truth segmentations have high agreement, with a mean (± standard deviation) DSC across all patients of 0.91 ± 0.04. The 5^{th} and 95^{th} percentiles of the DSC were 0.84 and 0.95. The mean MSD and HD_{avg} were 1.73 ± 0.67 mm and 1.83 ± 0.71 mm respectively (Extended Data Table 3.3). Similar performance was observed for the markerless dataset with a mean DSC for all patients of 0.91 ± 0.04 and 5^{th} and 95^{th} percentiles of 0.84 and 0.95. The mean MSD and HD_{avg} were 1.77 ± 0.70 mm and 1.86 ± 0.72 mm respectively (Extended Data Table 3.4).

The mean centroid error and DSC results at each imager angle are presented in Fig. 3.6. For the masked dataset, the mean error in the SI direction was consistent across all angles while the error in the AP/LAT direction has a positive offset for the first 180 degrees (Fig. 3.6a). For the masked dataset, the mean error in the AP/LAT and SI direction was consistent across all angles (Fig. 3.6b). The DSC results show consistently worse performance for each dataset at

the post-oblique angles (Fig. 3.6a,b). At these angles, the x-ray path length is towards the maximum as it beams through the pelvis. The difference in the model performance in the range of 270–330° for the masked dataset and 290–90° for the markerless dataset is a result of the variable number of samples at each angle. The kV images were obtained through a sub-arc acquisition method. Consequently, there are fewer samples available for angles within the aforementioned ranges compared to other segments of the arc.

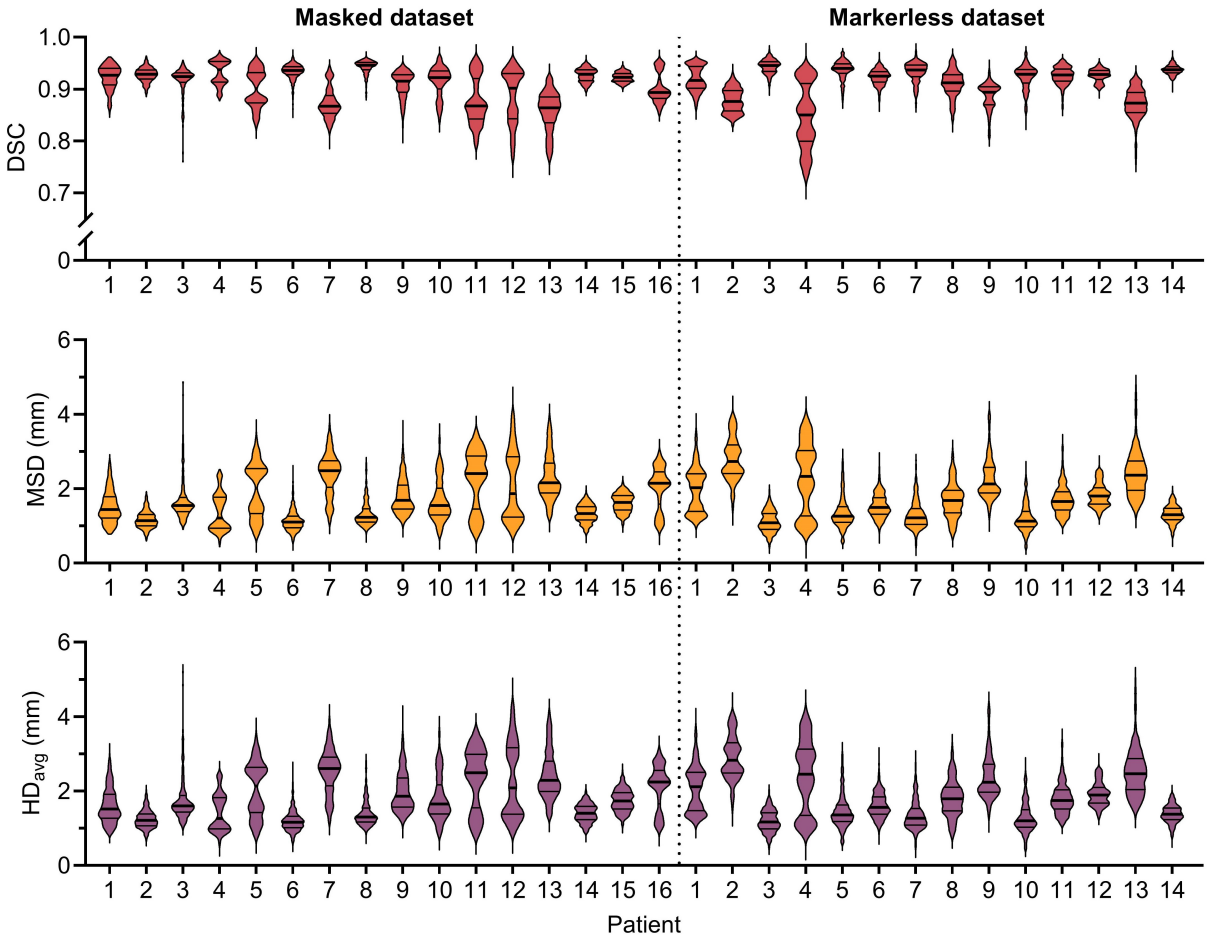


Fig. 3.5 | Violin plots of the geometric assessment metrics. The DSC (red), MSD (yellow), and HD_{avg} (purple) between the cGAN segmentation and ground truth for each patient. The bars represent the median (thick) and quartiles (thin). DSC, Dice similarity coefficient; MSD, mean surface distance; HD_{avg}, average Hausdorff distance.

The time taken for the trained network to generate the segmentation was approximately 10 milliseconds per image using the desktop computer described earlier. An example of the cGAN and ground truth segmentations at different image angles of four patients are shown in Fig. 3.7. This figure demonstrates the best and worst performing patient in each dataset based on the mean DSC (Fig. 3.7). It can be observed that for all patients, there is strong agreement in the shape of the segmentation. While there is strong agreement in shape, there is an offset

in the centroid positions for the worse performing segmentations.

The results presented in Fig. 3.8 show example tracking results for the average performing patient in each dataset based on the mean DSC. One fraction each with 500 kV images are shown. The tracking system was able to maintain tracking throughout both fractions. The error in the SI direction is typically lower compared to the AP/LAT direction (Fig 3.8c). The predicted and ground truth segmentations agree well with each other with a DSC around 0.9 across both patients (Fig. 3.8d).

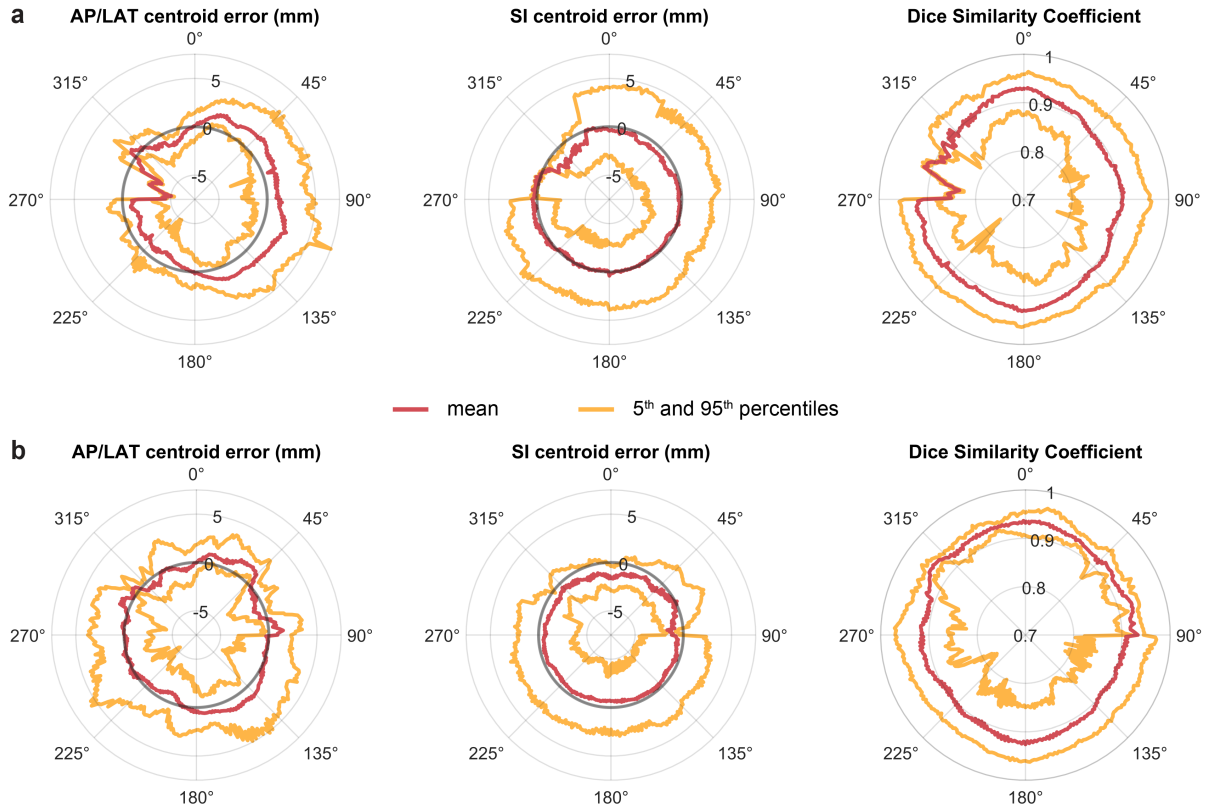


Fig. 3.6 | Centroid error and Dice similarity coefficient (DSC) results based on imager angle. The mean (red) with the 5th and 95th percentiles (yellow) of the centroid error and DSC across the entire radiation therapy treatment arc for the **(a)** masked dataset and **(b)** markerless dataset. AP/LAT, anterior-posterior/lateral; SI, superior-inferior.

3.4 Discussion

Our results demonstrate that a patient-specific deep learning model can be used to detect cases of high motion using real-time markerless prostate segmentation. We accomplished this using data acquired from a conventional radiation therapy system at four different cancer treatment centres. The tracking accuracy was 0.7 ± 1.9 mm and -0.2 ± 1.9 mm in the AP/LAT and SI direc-

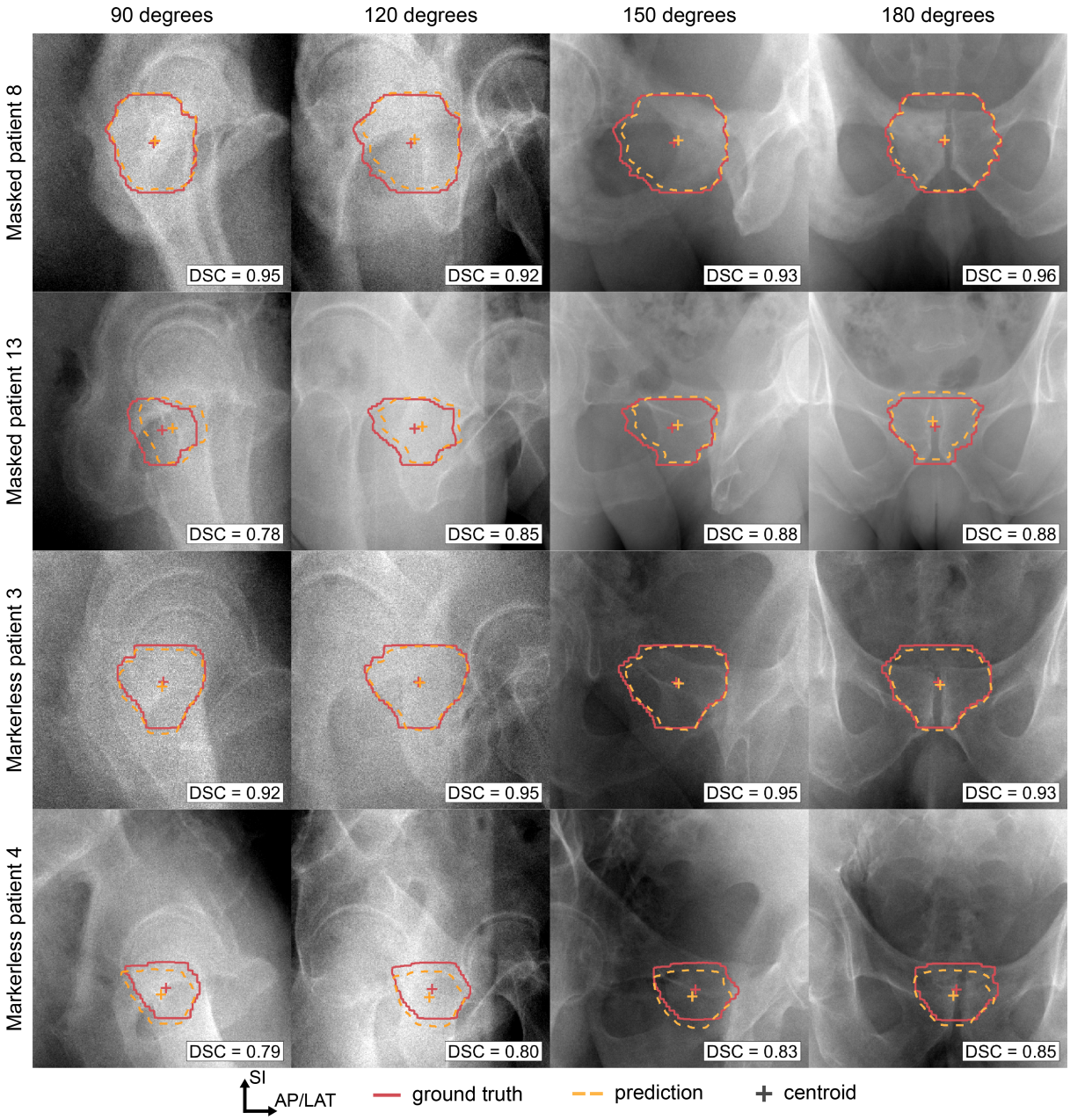


Fig. 3.7 | Example ground truth and prediction segmentations for four patients. Comparison of the ground truth (red) versus conditional Generative Adversarial Network segmentation (yellow) at imager angles of 90, 120, 150, and 180 degrees. The best and worst performing patient in each dataset based on the mean Dice similarity coefficient (DSC) are shown. The DSC is reported for each image. AP/LAT, anterior-posterior/lateral; SI, superior-inferior.

tions for the masked dataset. A similar performance was observed for the markerless dataset with 0.1 ± 1.8 mm and -0.6 ± 1.9 mm in the AP/LAT and SI directions, respectively. The masked dataset results indicate that the method maintains consistent accuracy across varying magnitudes of motion, as the errors are uncorrelated with the observed motion (Pearson correlation coefficients: 0.113 and -0.073 for AP/LAT and SI directions, respectively). The Bland-Altman analysis further supports this, showing minimal bias between the measurements and no trend

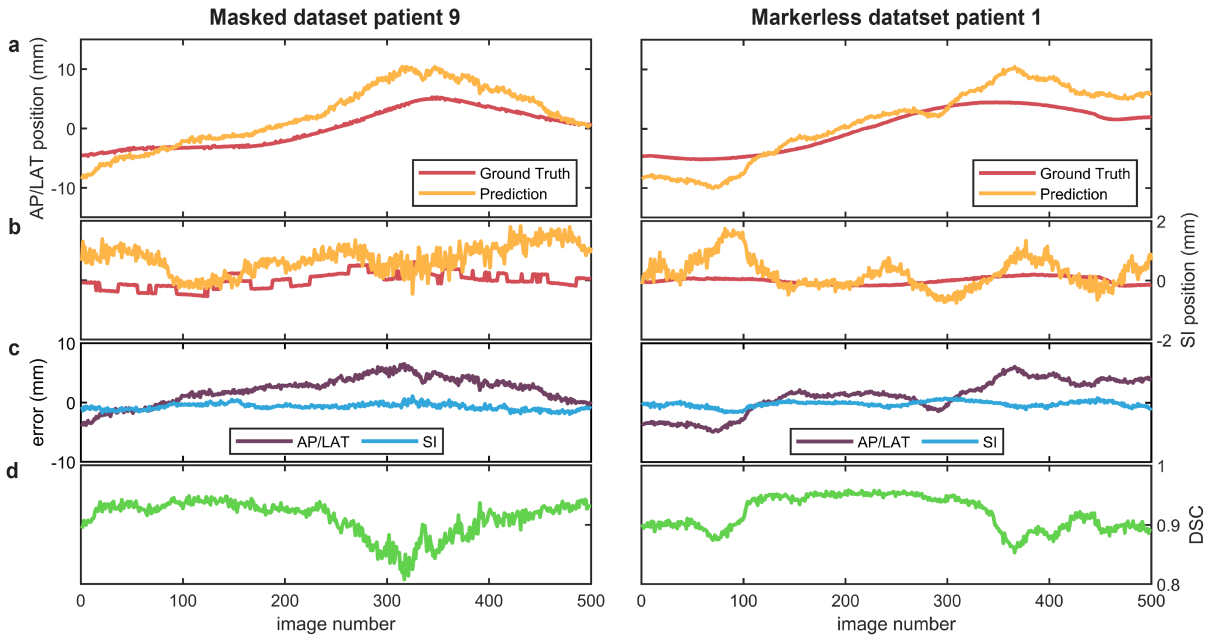


Fig. 3.8 | Tracking and Dice similarity coefficient (DSC) results for two patients. The prostate centroid position in the **(a)** anterior-posterior/lateral (AP/LAT) and **(b)** superior-inferior (SI) directions of the conditional Generative Adversarial Network (cGAN) segmentation in comparison to the ground truth. **(c)** The centroid error between the cGAN segmentation and the ground truth. **(d)** DSC between the cGAN segmentation and ground truth.

of larger errors with increasing averages.

From a study of 17 patients, Langen et al. reported the 3D prostate displacement was >3 mm and >5 mm approximately 14% and 3% of the time, respectively.[32] These prostate motion measurements are similar to those reported by Su et al. indicating 17% and 5% of time for the corresponding displacements.[33] Both studies highlighted the increased likelihood of displacement of the prostate with elapsed time.[32,33] Therefore, real-time motion monitoring is essential for treatment regimens where the treatment time is increased, such as with hypofractionated SBRT. For radiation therapy treatments, a PTV margin is applied around the prostate to account for treatment setup errors and intrafraction motion. For prostate radiation therapy, a PTV of 10 mm is used when using skin marks or bony anatomy for set up and 5–8 mm when using soft tissue registration or implanted markers.[34] These margins were further reduced for motion management treatments. In the SPARK trial, which used x-ray guidance, margins of 3 mm posteriorly and 5 mm in other directions were used.[19] In the MIRAGE trial, which used MRI guidance, 2 mm isotropic margins were used.[18] Our model successfully tracks within the acceptable margins for standard and x-ray guidance treatments.

Our results are insensitive for motion in the plane perpendicular to the detector plane as

model estimates the position in the 2D kV image frame of reference. For clinical use, there will need to be implementation of an algorithm to infer the 3D target coordinates from the 2D kV images. For marker-based tumour tracking, a method for 3D target estimation using the marker positions in 2D images has been clinically implemented.[35] Other successful approaches include using a 3D Gaussian PDF,[36] Bayesian inference,[37,38] or a Kalman filter.[39] These estimation methods could be adapted using the segmentation boundary or centroid for our approach. While our accuracy is reported in 2D, making direct comparisons with other studies difficult, we can expect that the model would be useful for detecting high motion cases. Given the high mean DSC achieved on both datasets (masked: 0.91 ± 0.04; markerless: 0.91 ± 0.04), gating could be performed when a defined percentage of prostate moves outside of a set treatment boundary.

Current x-ray-based methods for tracking the prostate during radiation therapy rely on implanted fiducial markers.[6–11] Further refinement of the method is required to reach the sub-millimetre accuracy offered by marker-based methods, which remain the clinical benchmark. There is minimal literature on tracking the prostate without markers. Zhao et al.[29] developed a deep learning model to identify the rectangular bounding box containing the prostate in simulated images for ten patients at three specific angles: anterior-posterior, left-lateral and oblique.[29] Zhao et al. achieved a MAD of 1.58–1.67 mm across the three angles when tested on simulated images. Our large multi-centre study demonstrated comparable performance on treatment-acquired kV images with a MAD of 1.4–1.6 mm across all angles and datasets. However, a larger error distribution was observed in our study. The larger error distribution may be attributed to the more challenging evaluation on treatment-acquired kV images with scatter presence, compared to evaluation on simulated images at defined angles. Zhao et al. reported an increase in the MAD to 2.29 mm when evaluating the models on kV images for a subset of patients demonstrating the increased difficulty.

Deep learning approaches have been developed for other sites including lung,[23–25] diaphragm,[26] liver,[24] pancreas,[27,28] prostate,[29] and head and neck.[30] Our method performs with a similar accuracy when compared to the markerless approaches for other anatomical sites. However, it should be noted that markerless tracking of the prostate and pancreas is more difficult than the other sites due to the lower soft tissue contrast. MRI linear accelerators (linacs) are an alternative for IGART due to improved soft tissue contrast compared to kV images.[12] However, MRI-linacs are an expensive treatment option for patients and are not widely

available compared to standard linacs.[40] Deep learning approaches for markerless MRI-guided radiation therapy have typically focussed only on lung[41,42] and liver[43] tracking.

Our model has several features that make it an ideal candidate for clinical implementation. First, the model takes 9.7 ± 0.7 hours on average to train including data generation and augmentation, making it feasible for patient-specific training in between the patient's planning session and the first treatment. Multiple models can be trained simultaneously, and the training time could be further optimised through computer enhancements. Second, the inference time of the model was 10 ms on average per image. This makes the model suitable for real-time applications as the AAPM Task Group 264 defines real-time as a system latency below 500 ms.[44] A single model approach for tracking at all angles across the entire treatment arc is less computationally intensive and more clinically relevant than using several models for discrete angles. Third, the model produces a segmentation of the prostate which when combined with 3D target estimation can be beneficial for other applications such as real-time dose optimisation.[45] Finally, the model is patient-specific, allowing it to learn features relevant to the specific patient and imaging system. The robustness of health AI algorithms is a major concern.[46] Often the performance of an algorithm can be correlated to the particular data used for training. However, this is not a concern for our study as the patient specific model has been tested across four different cancer treatment centres, achieving a similar performance across all patients. As the model is agnostic to the prostate characteristics, it could be easily adapted to include other treatment targets such as the pelvic lymph nodes or organs at risk.

However, there are risks associated with tumour tracking algorithms. If the algorithm produces inaccurate results, it could potentially disrupt a treatment that is proceeding as expected. One strategy to address this would be to incorporate a confidence metric so that low confidence results do not disrupt the treatment. Additionally, the beam delivery could be paused only when the tumour motion exceeds a pre-set tolerance for a pre-defined time to negate random errors. Other risks can include failure to detect clinically significant motion. One potential solution would be to incorporate redundancy through multiple tracking algorithms. The additional algorithms tracking the surrounding organs at risk may assist with detecting all significant motion.

Our study is not without limitations. The first limitation are the uncertainties related to the ground truth of the prostate in the kV images. Due to the low soft tissue contrast in the kV images, it is not possible to manually contour or label the prostate. Therefore, other solu-

tions were required to generate the ground truth. We developed two methods of producing the ground truth. The aim was that results together, from both methods, would reduce the uncertainties related to the results. The masked dataset was generated from imaging data of prostate cancer patients with implanted fiducial markers. The markers were used to annotate the real-time location of the prostate in the kV images and were then masked out. The method for localisation of the markers has shown sub-millimetre accuracy.[10] However, fiducial markers are subject to 1–2 mm surrogacy errors[47–49] and may therefore limit the accuracy of the ground truth prostate segmentation.

The markerless dataset was generated from imaging data with no implanted fiducial markers. Since it is not possible to annotate the real-time location, the kV images were rigidly shifted based on image registration performed between the planning CT and treatment CBCT giving an average location. The uncertainties related to soft-tissue registration between the CBCT and planning CT are typically between 1–3 mm.[50,51] The results from these two ground truths combined provide increased certainty of the model performance for markerless tracking. The two approaches rely on the planning CT prostate contour which can have uncertainties relating to intraobserver and interobserver variability. Additionally, utilising the planning CT contour for training does not account for potential deformation of the prostate between planning and treatment. Prostate deformation has been reported to have a mean and standard deviation of 0.6 mm and 1.7 mm, respectively.[52] Furthermore, it was found that anterior-posterior displacement of the prostate centre of gravity is highly correlated with deformation in the middle-anterior and posterior segments.[52] Therefore, the prostate deformation and resulting shift in the centre of gravity may affect the model performance.

The augmentation applied to the training dataset only accounted for the motion of the entire patient, excluding anatomical motion. In previous work, we used realistic deformations to train a deep learning model for head and neck tumour segmentation, demonstrating that this approach enhanced robustness to patient motion.[30] However, head and neck tumours exhibit different motion and deformation characteristics compared to the prostate. Initial experiments using a similar augmentation approach to shift the prostate and deform surrounding soft tissue did not improve performance and significantly increased training time due deformation computations. Therefore, we focused on optimising performance and training time by using rigid shifts alone.

The second limitation is that the model was evaluated using kV images acquired during

patient setup rather than intrafraction kV images. For the masked dataset, the intrafraction kV images were collimated. As a result, the surrounding anatomy is not visible making the images unsuitable for segmentation. Additionally, for the markerless dataset we prioritised the quality of our ground truth and hence used kV images from patient setup for this study. To generate the ground truth for the markerless dataset, image registration was required between the treatment 3D CBCT and planning CT to determine the average location. Therefore, we could not generate the ground truth for the intrafraction kV images in the markerless dataset. While kV images from patient setup provide superior quality to intrafraction kV images, state-of-the-art clinical systems provide solutions to minimise the effect of megavoltage (MV) scatter from the treatment beam and provide an improved kV image quality. One such solution is trigged imaging that is incorporated into Varian systems.[53] Triggered imaging improves kV image quality by placing the treatment beam on hold prior to acquisition of each triggered image in order to eliminate the effect of MV scatter. The simple solution of frame averaging has been previously used to reduce noise in the images.[54] As our model is trained on a case-by-case basis, it can benefit from future research that improves kV image quality.

3.5 Outlook

Our study is the first large scale proof-of-principle of x-ray-based markerless tracking for globally available cancer therapy systems, providing an important step towards clinical implementation. The results demonstrate the potential of our method to be expanded to other soft tissue organs such as the pancreas, liver, and kidneys. Our approach only requires x-ray images during treatment which in principle, covers all linear accelerators from Elekta, Varian, and other manufacturers. Our markerless method will help make IGART treatments accessible for all patients, eliminating the time delays, costs and risk associated with marker implantation. Future work will look at the development of a single software solution with the model for clinical implementation. The software solution can then be experimentally tested on a conventional radiation therapy system using a phantom and prospectively with a quality assurance procedure.[55]

3.6 Methods

Masked dataset. The masked dataset was generated using imaging data of patients with implanted fiducial markers which were masked-out for training and analysis (Fig. 3.1a). The dataset was constructed using the imaging data of 16 prostate cancer patients undergoing radiation therapy in the TROG 15.01 SPARK trial.[56] The patients that were used in this study were treated on the TrueBeam (Varian Medical Systems, Palo Alto) Linac across three different sites (Calvary Mater Newcastle, Australia; Westmead Hospital, Australia; Peter MacCallum Cancer Centre, Australia). We collected the planning CT, physician contours, and kV images from two fractions of each patient associated with this cohort. The kV images were acquired during patient setup to reconstruct the CBCT. We utilised 500 kV images from each fraction, giving a total of 16,000 kV images. Each patient had three cylindrical gold fiducial markers implanted in their prostate. The training data were the 3D planning CT and 3D prostate contour. The test data were generated using the kV images from two fractions. The prostate contour volume was projected as a 2D digitally reconstructed radiograph (DRR) for each kV image angle. The DRRs were generated using the Reconstruction Toolkit[57] and the Insight Toolkit.[58] We employed the traditional DRR algorithm as derived by Madden et al.[59] The ground truth was generated by aligning the prostate only DRR with the kV images based on the implanted fiducial markers. The marker positions were labelled using a previously developed deep learning framework[10] with manual verification. Following alignment, the fiducial markers were masked out in all volumes, DRRs, and kV images to avoid biasing the model. The masking algorithm uses the *regionfill* function in MATLAB (The MathWorks, Inc.) to smoothly interpolate inwards from the pixel values surrounding the marker by calculating the discrete Laplacian and solving the Dirichlet boundary value problem. Poisson noise was applied to the kV images following interpolation.

Markerless dataset. The markerless dataset was generated using imaging data of patients with no implanted fiducial markers (Fig. 3.1b). The dataset was constructed using the imaging data of 14 prostate cancer patients undergoing radiation therapy in the OPTIMAL trial (NCT03386045).[60] This trial treated patients on the TrueBeam STx (Varian Medical Systems, Palo Alto) Linac at Royal North Shore Hospital, Australia. We collected the planning CT, physician contours, and kV images from two fractions associated with this cohort. As with the masked dataset, the kV images were acquired during patient setup to reconstruct the CBCT.

We utilised 500 kV images from each fraction, giving a total of 14,000 kV images. The training data were generated in the same way as the masked dataset. Since these patients did not have implanted fiducial markers, the ground truth was generated using shifts based on image registration performed between the planning CT and treatment CBCT. The 2D kV images were shifted based on the 3D couch shift using equation:

$$\hat{\mathbf{P}}\left(\hat{\mathbf{M}}(t) \mid \theta\right) = \begin{pmatrix} \hat{\mathbf{u}} \\ \hat{\mathbf{v}} \end{pmatrix} = \frac{SID}{SAD - (\hat{\mathbf{x}} \cdot \cos\theta + \hat{\mathbf{z}} \cdot \sin\theta)} \begin{pmatrix} \hat{\mathbf{x}} \cdot \sin\theta - \hat{\mathbf{z}} \cdot \cos\theta \\ \hat{\mathbf{y}} \end{pmatrix} \tag{3.1}$$

where SID is the linac source-isocentre distance, SAD is the linac source-aperture distance, $(\hat{x}, \hat{y}, \hat{z})$ are the 3D shifts, and $(\hat{u}, \hat{v})$ are the resulting 2D shifts.

After shifting the kV images, prostate only DRRs were generated for each kV image angle. Therefore, the ground truth in the markerless dataset is defined by the average location of the prostate instead of the real-time location.

cGAN framework. The tracking system uses a cGAN model for segmentation of the prostate. The training of the model involves adversarial learning between the generator network, *G*, and the discriminator network, *D*. The cGAN model is trained to replicate a prostate segmentation given a pelvis kV image as input. The generator, *G*, takes the input kV image, *x*, and creates a segmentation image *G(x)*. The discriminator, *D*, classifies whether the paired image, *xy*, comes from the training set or the generator network as shown in Fig. 3.1c. The cGAN was initialised with a normal distribution and trained to minimise the loss function:

$$G^* = arg\min_G \max_D \mathcal{L}_{\text{cGAN}}(G, D) + \lambda \mathcal{L}_{L1}(G) \tag{3.2}$$

where λ is a constant (set to 100 for this implementation) and:

$$\mathcal{L}_{\text{cGAN}}(G, D) = \mathbb{E}_{(x,y)}\left[\log D(x, y)\right] + \mathbb{E}_x[\log(1 - D(x, G(x)))] \tag{3.3}$$

$$\mathcal{L}_{\text{L1}} = \mathbb{E}_{(x,y)} \left\| y - G(x) \right\|_1 \tag{3.4}$$

The cGAN implementation used for this study was based on the Pix2pix model.[61] A UNet was used for the generator architecture and a PatchGAN for the discriminator architecture. The detailed network architectures are shown in Extended Data Fig. 3.1. A patient-specific model was trained for each patient using 36,000 DRRs per patient. The DRRs were produced using

two volumetric images: the planning CT and prostate contour volume. The volumes were forward projected to produce one DRR each degree over 360 degrees, generating 360 images. Prior to training, data augmentation for each patient was performed 100 times by randomly shifting the CT geometry up to 10 mm and rotating up to 10 degrees and then computing a new set of DRRs to replicate possible treatment setup error and anatomical motion.[62] In total 36,000 DRRs were created for each patient with a size of 512 × 512 pixels. The images produced from the planning CT were each paired with the respective image produced from the prostate contour. Each model was trained for 20 epochs with a batch size of four and a learning rate of 0.0002 using the Adam optimiser. The models were trained on a desktop computer with an Intel® Xeon® Gold 6248R processor (3.0 GHz) with 256 GB RAM and a NVIDIA® RTX A6000 GPU.

Clinical application. The cGAN model can be incorporated into the treatment workflow for intrafraction monitoring of the prostate in kV images (Fig. 3.2). For clinical implementation, the central component of the workflow is the deep learning model for prostate segmentation. The generator network of the cGAN produces a prediction image (prostate segmentation) based on a kV input image. The prediction image is binarised using a set threshold value to give the final segmentation, and the centroid of the segmentation is calculated. If multiple unconnected regions are detected, the centroid of the most significant region is calculated. The calculated centroid location can be exported to the positioning systems to allow for real-time motion adjustments during the treatment.

Analysis. The models were tested on the unseen kV images to evaluate the accuracy of the prostate segmentation and the tracking system. The models were tested using the kV images from two fractions of each patient, giving 1,000 test images per patient (500 per fraction). For the masked dataset, ground truth generation involved manual verification of 48,000 marker positions to ensure high-quality data. Due to the time-consuming nature, we prioritised evaluating variability between patients and limited the analysis to two fractions per patient to include more patients. For the markerless dataset, only two fractions of data were available for each patient. The cGAN segmentation was binarised based on a 0.1 threshold and was compared to the ground truth segmentation for the analysis. The generator's ability to produce accurate prostate segmentations was evaluated for each patient model. The performance was quantified by calculating the DSC, MSD, and HD_{avg}. If multiple unconnected regions were present in the cGAN segmentation, the metrics were calculated using the largest region. Fi-

nally, the generator's ability to be used in an automated tracking system was evaluated by comparing the centroid of the segmentations. If multiple unconnected regions were present in the cGAN segmentation, the centroid was calculated using the largest region. The tracking system error was defined as the cGAN segmentation centroid minus the ground truth segmentation centroid. The errors were calculated in the AP/LAT and SI directions. The errors were reported at the patient coordinate system using the source-isocentre distance/source-detector distance ratio (=1.5) as the correction factor giving a pixel size of 0.26 mm. The overall errors were quantified by calculating the mean error, MAD, and the 5^{th} and 95^{th} percentiles of the errors. The correlation between the observed motion and centroid error for the masked dataset was evaluated using the Pearson correlation coefficient and Bland-Altman analysis.

3.7 References

1. Hewson, E. A. et al. Is multileaf collimator tracking or gating a better intrafraction motion adaptation strategy? An analysis of the TROG 15.01 stereotactic prostate ablative radiotherapy with KIM (SPARK) trial. *Radiotherapy and Oncology* **151**, 234–241 (2020).

2. Caillet, V., Booth, J. T. & Keall, P. IGRT and motion management during lung SBRT delivery. *Physica Medica* **44**, 113–122 (2017).

3. Colvill, E. et al. DMLC tracking and gating can improve dose coverage for prostate VMAT. *Medical physics* **41**, 091705 (2014).

4. Solberg, T. D. et al. Quality and safety considerations in stereotactic radiosurgery and stereotactic body radiation therapy: Executive summary. *Practical radiation oncology* **2**, 2–9 (2012).

5. Keall, P. J. et al. The first clinical implementation of real-time image-guided adaptive radiotherapy using a standard linear accelerator. *Radiotherapy and Oncology* **127**, 6–11 (2018).

6. Fledelius, W., Worm, E., Høyer, M., Grau, C. & Poulsen, P. R. Real-time segmentation of multiple implanted cylindrical liver markers in kilovoltage and megavoltage x-ray images. *Physics in Medicine & Biology* **59**, 2787 (2014).

7. Bertholet, J. et al. Fully automatic segmentation of arbitrarily shaped fiducial markers in cone-beam CT projections. *Physics in Medicine & Biology* **62**, 1327 (2017).

8. Poulsen, P. R. et al. A method for robust segmentation of arbitrarily shaped radiopaque structures in cone-beam CT projections. *Medical physics* **38**, 2151–2156 (2011).

9. Regmi, R. et al. Automatic tracking of arbitrarily shaped implanted markers in kilovoltage projection images: a feasibility study. *Medical physics* **41**, 071906 (2014).

10. Mylonas, A. et al. A deep learning framework for automatic detection of arbitrarily shaped fiducial markers in intrafraction fluoroscopic images. *Medical physics* **46**, 2286–2297 (2019).

11. Campbell, W. G., Miften, M. & Jones, B. L. Automated target tracking in kilovoltage images using dynamic templates of fiducial marker clusters. *Medical physics* **44**, 364–374 (2017).

12. Bertholet, J. et al. Real-time intrafraction motion monitoring in external beam radiotherapy. *Physics in medicine & biology* **64**, 15TR01 (2019).

13. Loh, J. et al. Infections after fiducial marker implantation for prostate radiotherapy: are we underestimating the risks? *Radiation Oncology* **10**, 1–5 (2015).

14. Igdem, S. et al. Implantation of fiducial markers for image guidance in prostate radiotherapy: patient-reported toxicity. *The British Journal of Radiology* **82**, 941–945 (2009).

15. Bhagat, N. et al. Complications associated with the percutaneous insertion of fiducial markers in the thorax. *Cardiovascular and interventional radiology* **33**, 1186–1191 (2010).

16. Lim, Y. K. et al. Microscopic gold particle-based fiducial markers for proton therapy of prostate cancer. *International Journal of Radiation Oncology* Biology* Physics* **74**, 1609–1616 (2009).

17. Anastasi, G. et al. Patterns of practice for adaptive and real-time radiation therapy (POP-ART RT) part I: Intra-fraction breathing motion management. *Radiotherapy and Oncology* **153**, 79–87 (2020).

18. Kishan, A. U. et al. Magnetic resonance imaging–guided vs computed tomography–guided stereotactic body radiotherapy for prostate cancer: the MIRAGE randomized clinical trial. *JAMA oncology* **9**, 365–373 (2023).

19. Keall, P. et al. Real-time image guided ablative prostate cancer radiation therapy: results from the TROG 15.01 SPARK trial. *International Journal of Radiation Oncology* Biology* Physics* **107**, 530–538 (2020).

20. Lovelock, D. M., Messineo, A. P., Cox, B. W., Kollmeier, M. A. & Zelefsky, M. J. Continuous monitoring and intrafraction target position correction during treatment improves target coverage for patients undergoing SBRT prostate therapy. *International Journal of Radiation Oncology* Biology* Physics* **91**, 588–594 (2015).

21. Kerkmeijer, L. G. et al. Focal boost to the intraprostatic tumor in external beam radiotherapy for patients with localized prostate cancer: results from the FLAME randomized phase III trial. *Journal of Clinical Oncology* **39**, 787–796 (2021).

22. Mylonas, A., Booth, J. & Nguyen, D. T. A review of artificial intelligence applications for motion tracking in radiotherapy. *Journal of Medical Imaging and Radiation Oncology* **65**, 596–611 (2021).

23. Terunuma, T., Tokui, A. & Sakae, T. Novel real-time tumor-contouring method using deep learning to prevent mistracking in X-ray fluoroscopy. *Radiological physics and technology* **11**, 43–53 (2018).

24. Hirai, R., Sakata, Y., Tanizawa, A. & Mori, S. Real-time tumor tracking using fluoroscopic imaging with deep neural network analysis. *Physica Medica* **59**, 22–29 (2019).

25. Takahashi, W., Oshikawa, S. & Mori, S. Real-time markerless tumour tracking with patient-specific deep learning using a personalised data generation strategy: proof of concept by phantom study. *The British journal of radiology* **93**, 20190420 (2020).

26. Edmunds, D., Sharp, G. & Winey, B. Automatic diaphragm segmentation for real-time lung tumor tracking on cone-beam CT projections: a convolutional neural network approach. *Biomedical physics & engineering express* **5**, 035005 (2019).

27. Zhou, D., Nakamura, M., Mukumoto, N., Yoshimura, M. & Mizowaki, T. Development of a deep learning-based patient-specific target contour prediction model for markerless tumor positioning. *Medical physics* **49**, 1382–1390 (2022).

28. Zhao, W. et al. Markerless pancreatic tumor target localization enabled by deep learning. *International Journal of Radiation Oncology* Biology* Physics* **105**, 432–439 (2019).

29. Zhao, W. et al. Incorporating imaging information from deep neural network layers into image guided radiation therapy (IGRT). *Radiotherapy and Oncology* **140**, 167–174 (2019).

30. Gardner, M. et al. Realistic CT data augmentation for accurate deep-learning based segmentation of head and neck tumors in kV images acquired during radiation therapy. *Medical Physics* **50**, 4206–4219 (2023).

31. Zhang, A., Xing, L., Zou, J. & Wu, J. C. Shifting machine learning for healthcare from development to deployment and from models to data. *Nature Biomedical Engineering* **6**, 1330–1345 (2022).

32. Langen, K. M. et al. Observations on real-time prostate gland motion using electromagnetic tracking. *International Journal of Radiation Oncology* Biology* Physics* **71**, 1084–1090 (2008).

33. Su, Z., Zhang, L., Murphy, M. & Williamson, J. Analysis of prostate patient setup and tracking data: potential intervention strategies. *International Journal of Radiation Oncology* Biology* Physics* **81**, 880–887 (2011).

34. Yartsev, S. & Bauman, G. Target margins in radiotherapy of prostate cancer. *The British journal of radiology* **89**, 20160312 (2016).

35. Nguyen, D. T. et al. The first clinical implementation of a real-time six degree of freedom target tracking system during radiation therapy based on Kilovoltage Intrafraction Monitoring (KIM). *Radiotherapy and Oncology* **123**, 37–42 (2017).

36. Poulsen, P. R., Cho, B. & Keall, P. J. A method to estimate mean position, motion magnitude, motion correlation, and trajectory of a tumor from cone-beam CT projections for image-guided radiotherapy. *International Journal of Radiation Oncology* Biology* Physics* **72**, 1587–1596 (2008).

37. Shieh, C.-C. et al. A Bayesian approach for three-dimensional markerless tumor tracking using kV imaging during lung radiotherapy. *Physics in Medicine & Biology* **62**, 3065 (2017).

38. Li, R., Fahimian, B. P. & Xing, L. A Bayesian approach to real-time 3D tumor localization via monoscopic x-ray imaging during treatment delivery. *Medical physics* **38**, 4205–4214 (2011).

39. Nguyen, D. T. et al. A real-time IGRT method using a Kalman filter framework to extract 3D positions from 2D projections. *Physics in Medicine & Biology* **66**, 214001 (2021).

40. van Herk, M., McWilliam, A., Dubec, M., Faivre-Finn, C. & Choudhury, A. Magnetic resonance imaging–guided radiation therapy: a short strengths, weaknesses, opportunities, and threats analysis (2018).

41. Yun, J. et al. Neural-network based autocontouring algorithm for intrafractional lung-tumor tracking using Linac-MR. *Medical physics* **42**, 2296–2310 (2015).

42. Fast, M. F. et al. Tumour auto-contouring on 2d cine MRI for locally advanced lung cancer: A comparative study. *Radiotherapy and Oncology* **125**, 485–491 (2017).

43. Friedrich, F. et al. Stability of conventional and machine learning-based tumor auto-segmentation techniques using undersampled dynamic radial bSSFP acquisitions on a 0.35 T hybrid MR-linac system. *Medical Physics* **48**, 587–596 (2021).

44. Keall, P. J. et al. AAPM Task Group 264: The safe clinical implementation of MLC tracking in radiotherapy. *Medical physics* **48**, e44–e64 (2021).

45. Mejnertsen, L., Hewson, E., Nguyen, D. T., Booth, J. & Keall, P. Dose-based optimisation for multi-leaf collimator tracking during radiation therapy. *Physics in Medicine & Biology* **66**, 065027 (2021).

46. Subbaswamy, A. & Saria, S. From development to deployment: dataset shift, causality, and shift-stable models in health AI. *Biostatistics* **21**, 345–352 (2020).

47. Nichol, A. M. et al. A magnetic resonance imaging study of prostate deformation relative to implanted gold fiducial markers. *International Journal of Radiation Oncology* Biology* Physics* **67**, 48–56 (2007).

48. Delouya, G., Carrier, J.-F., Béliveau-Nadeau, D., Donath, D. & Taussky, D. Migration of intraprostatic fiducial markers and its influence on the matching quality in external beam radiation therapy for prostate cancer. *Radiotherapy and Oncology* **96**, 43–47 (2010).

49. Poggi, M. M., Gant, D. A., Sewchand, W. & Warlick, W. B. Marker seed migration in prostate localization. *International Journal of Radiation Oncology* Biology* Physics* **56**, 1248–1251 (2003).

50. Létourneau, D. et al. Assessment of residual error for online cone-beam CT-guided treatment of prostate cancer patients. *International Journal of Radiation Oncology* Biology* Physics* **62**, 1239–1246 (2005).

51. Deegan, T. et al. Assessment of cone beam CT registration for prostate radiation therapy: Fiducial marker and soft tissue methods. *Journal of medical imaging and radiation oncology* **59**, 91–98 (2015).

52. Nakazawa, T. et al. Analysis of prostate deformation during a course of radiation therapy for prostate cancer. *PLoS One* **10**, e0131822 (2015).

53. Kaur, G., Lehmann, J., Greer, P. & Simpson, J. Assessment of the accuracy of truebeam intrafraction motion review (IMR) system for prostate treatment guidance. *Australasian Physical & Engineering Sciences in Medicine* **42**, 585–598 (2019).

54. Ng, J. A. et al. Kilovoltage intrafraction monitoring for prostate intensity modulated arc therapy: first clinical results. *International Journal of Radiation Oncology* Biology* Physics* **84**, e655–e661 (2012).

55. Mueller, M. et al. The first prospective implementation of markerless lung target tracking in an experimental quality assurance procedure on a standard linear accelerator. *Physics in Medicine & Biology* **65**, 025008 (2020).

56. Keall, P. et al. Stereotactic prostate adaptive radiotherapy utilising kilovoltage intrafraction monitoring: the TROG 15.01 SPARK trial. *BMC cancer* **17**, 1–7 (2017).

57. Rit, S. et al. The Reconstruction Toolkit (RTK), an open-source cone-beam CT reconstruction toolkit based on the Insight Toolkit (ITK). In *Journal of Physics: Conference Series*, vol. 489, 012079 (IOP Publishing, 2014).

58. McCormick, M., Liu, X., Jomier, J., Marion, C. & Ibanez, L. ITK: enabling reproducible research and open science. *Frontiers in neuroinformatics* **8**, 13 (2014).

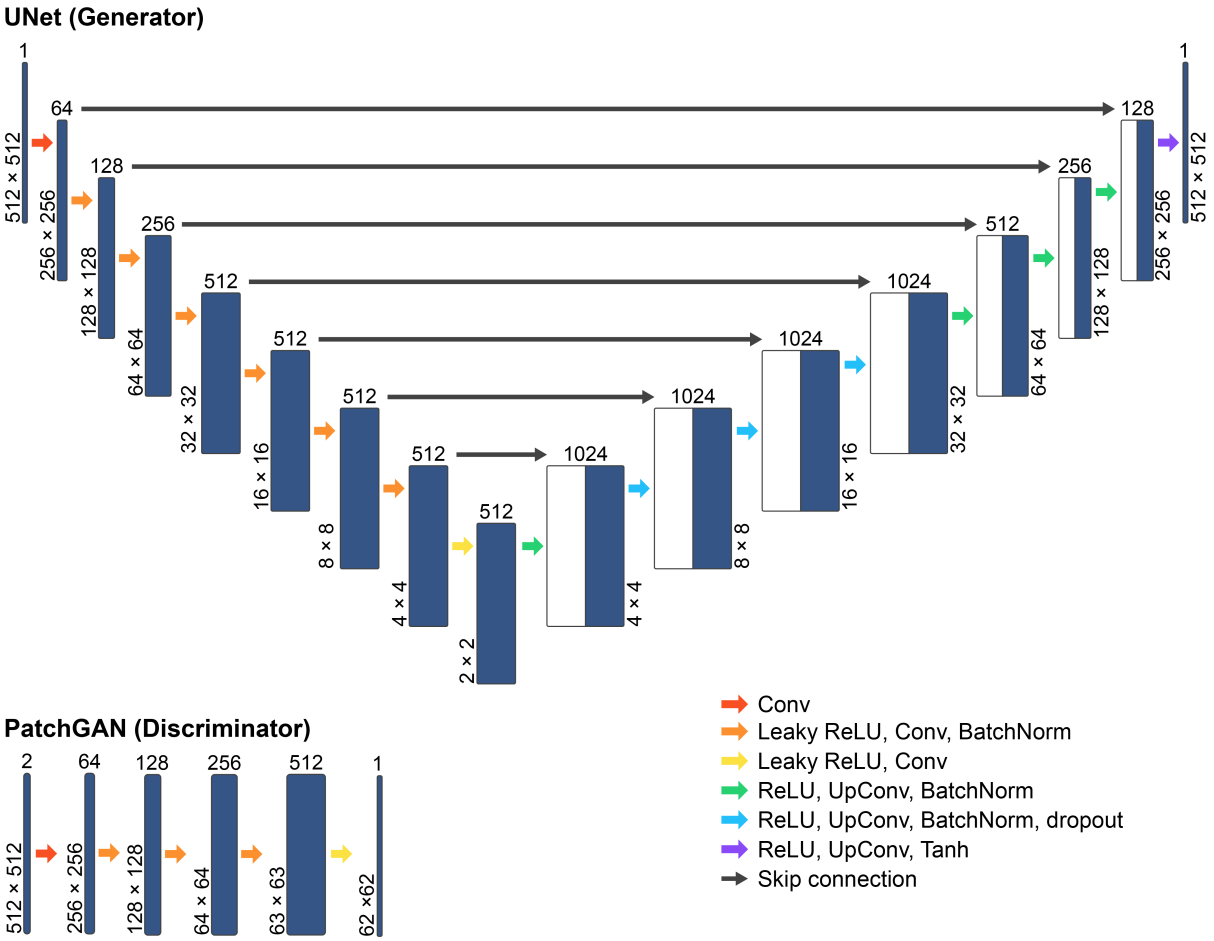
59. Madden, L. et al. CBCT-DRRs superior to CT-DRRs for target-tracking applications for pancreatic SBRT. *Biomedical Physics & Engineering Express* **10**, 035039 (2024).

60. Royal North Shore Hospital. Optimal Prostate Study. `https://clinicaltrials.gov/study/NCT03386045` (2023).

61. Isola, P., Zhu, J.-Y., Zhou, T. & Efros, A. A. Image-to-image translation with conditional adversarial networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, 1125–1134 (2017).

62. Byrne, T. E. A review of prostate motion with considerations for the treatment of prostate cancer. *Medical Dosimetry* **30**, 155–161 (2005).

3.8 Extended Data



Extended Data Fig. 3.1 | UNet (Generator) and PatchGAN (Discriminator) network architectures. The UNet takes a kilovoltage image as input and produces the prostate segmentation. The PatchGAN takes 70 × 70 patches of the input image to determine if the segmentations are real or fake. Conv, convolution; ReLU, rectified linear unit; BatchNorm, batch normalisation; UpConv, upconvolution.

Extended Data Table 3.1 | Individual patient centroid errors for the masked dataset.

	AP/LAT Centroid Error (mm)			SI Centroid Error (mm)		
Patient	**Mean**	**5th percentile**	**95th percentile**	**Mean**	**5th percentile**	**95th percentile**
1	0.2 ± 1.9	-3.2	3.1	0.9 ± 0.5	0.3	1.7
2	0.7 ± 1.1	-1.1	2.4	0.0 ± 1.0	-1.3	1.3
3	0.3 ± 2.3	-3.4	4.6	-0.9 ± 0.9	-2.1	1.2
4	-0.1 ± 0.8	-1.7	1.1	-1.5 ± 1.0	-2.9	-0.2
5	0.9 ± 1.6	-2.1	3.0	2.4 ± 1.7	0.1	4.4
6	-0.3 ± 1.1	-2.0	1.5	-0.7 ± 0.6	-1.6	0.5
7	0.1 ± 2.4	-4.2	3.7	2.5 ± 0.9	1.0	3.8
8	1.0 ± 1.2	-1.3	2.8	0.6 ± 0.7	-0.5	1.8
9	2.1 ± 2.1	-1.5	5.1	-1.1 ± 0.6	-2.3	-0.1
10	1.8 ± 1.6	-0.5	4.4	-0.4 ± 1.0	-1.8	1.0
11	-0.3 ± 1.3	-2.8	1.5	-2.4 ± 1.4	-4.0	0.5
12	1.1 ± 2.4	-3.0	4.1	1.4 ± 1.5	-0.3	3.9
13	2.4 ± 2.4	-2.5	5.3	1.5 ± 1.5	-1.2	3.8
14	0.4 ± 0.8	-0.7	2.1	-1.2 ± 0.6	-2.2	-0.2
15	0.7 ± 1.2	-1.3	2.8	-1.6 ± 0.7	-2.7	-0.4
16	0.3 ± 1.1	-1.2	2.5	-3.0 ± 1.4	-4.6	-0.8
Overall	0.7 ± 1.9	-2.4	4.0	-0.2 ± 1.9	-3.4	3.5

The mean, 5th percentiles, and 95th percentiles of the errors in the centroid location of the cGAN segmentation versus the ground truth for each patient in the masked dataset. The errors are calculated in the anterior-posterior/lateral (AP/LAT) directions and superior-inferior (SI) directions. Data shown as mean ± standard deviations.

Extended Data Table 3.2 | Individual patient centroid errors for the markerless dataset.

	AP/LAT Centroid Error (mm)			SI Centroid Error (mm)		
Patient	**Mean**	**5th percentile**	**95th percentile**	**Mean**	**5th percentile**	**95th percentile**
1	0.6 ± 2.5	-3.7	4.6	0.7 ± 1.3	-0.8	3.4
2	0.3 ± 1.4	-1.8	2.7	-3.9 ± 1.0	-5.2	-2.0
3	0.0 ± 1.1	-1.8	1.9	-0.4 ± 0.8	-1.8	0.9
4	-0.5 ± 1.8	-2.6	3.5	-1.3 ± 2.7	-4.6	2.5
5	0.8 ± 1.6	-1.8	3.7	-0.1 ± 0.8	-1.5	1.3
6	0.2 ± 1.2	-1.6	2.3	-0.3 ± 1.4	-2.4	2.1
7	0.1 ± 1.6	-2.9	2.3	0.6 ± 0.6	-0.5	1.4
8	-0.2 ± 2.0	-3.4	2.8	1.0 ± 1.2	-0.9	3.0
9	0.8 ± 2.3	-2.7	4.5	2.5 ± 0.6	1.5	3.4
10	0.1 ± 1.6	-2.6	2.6	0.1 ± 0.7	-0.9	1.5
11	0.2 ± 1.5	-2.2	2.9	-1.6 ± 0.8	-3.0	-0.5
12	0.1 ± 1.3	-2.3	2.3	-2.2 ± 0.5	-3.1	-1.4
13	-1.0 ± 2.3	-4.3	3.2	-2.8 ± 0.6	-3.9	-1.8
14	0.4 ± 0.9	-1.6	1.6	-0.8 ± 0.5	-1.7	0.0
Overall	0.1 ± 1.8	-2.8	3.1	-0.6 ± 1.9	-4.0	2.6

The mean, 5th percentiles, and 95th percentiles of the errors in the centroid location of the cGAN segmentation versus the ground truth for each patient in the markerless dataset. The errors are calculated in the anterior-posterior/lateral (AP/LAT) directions and superior-inferior (SI) directions. Data shown as mean ± standard deviations.

Extended Data Table 3.3 | Individual patient geometric assessment metrics for the masked dataset.

Patient	DSC	MSD (mm)	HD_{avg} (mm)
1	0.92 ± 0.02	1.52 ± 0.44	1.62 ± 0.47
2	0.93 ± 0.01	1.17 ± 0.23	1.24 ± 0.24
3	0.91 ± 0.03	1.71 ± 0.61	1.79 ± 0.66
4	0.93 ± 0.02	1.37 ± 0.49	1.42 ± 0.49
5	0.90 ± 0.04	1.96 ± 0.71	2.05 ± 0.72
6	0.93 ± 0.02	1.14 ± 0.29	1.20 ± 0.30
7	0.87 ± 0.03	2.40 ± 0.56	2.53 ± 0.58
8	0.94 ± 0.01	1.32 ± 0.34	1.40 ± 0.35
9	0.91 ± 0.03	1.80 ± 0.47	1.99 ± 0.55
10	0.91 ± 0.03	1.69 ± 0.56	1.81 ± 0.61
11	0.88 ± 0.05	2.24 ± 0.79	2.33 ± 0.80
12	0.88 ± 0.05	2.07 ± 0.92	2.27 ± 0.99
13	0.86 ± 0.04	2.31 ± 0.61	2.43 ± 0.64
14	0.93 ± 0.01	1.33 ± 0.24	1.41 ± 0.24
15	0.92 ± 0.01	1.62 ± 0.25	1.74 ± 0.29
16	0.90 ± 0.03	2.03 ± 0.57	2.12 ± 0.59
Overall	0.91 ± 0.04	1.73 ± 0.67	1.83 ± 0.71

The mean ± standard deviations of the Dice similarity coefficient (DSC), mean surface distance (MSD), and average Hausdorff distance (HD_{avg}) for each patient in the masked dataset.

Extended Data Table 3.4 | Individual patient geometric assessment metrics for the markerless dataset.

Patient	DSC	MSD (mm)	HD_{avg} (mm)
1	0.92 ± 0.02	1.97 ± 0.59	2.07 ± 0.62
2	0.88 ± 0.02	2.80 ± 0.56	2.90 ± 0.58
3	0.94 ± 0.01	1.12 ± 0.29	1.19 ± 0.29
4	0.85 ± 0.06	2.29 ± 0.94	2.39 ± 0.96
5	0.94 ± 0.02	1.36 ± 0.43	1.46 ± 0.46
6	0.92 ± 0.02	1.54 ± 0.32	1.61 ± 0.34
7	0.93 ± 0.02	1.30 ± 0.38	1.35 ± 0.40
8	0.91 ± 0.03	1.69 ± 0.47	1.81 ± 0.50
9	0.89 ± 0.03	2.27 ± 0.55	2.39 ± 0.60
10	0.92 ± 0.02	1.20 ± 0.37	1.28 ± 0.41
11	0.93 ± 0.02	1.71 ± 0.40	1.81 ± 0.42
12	0.93 ± 0.01	1.83 ± 0.31	1.92 ± 0.31
13	0.87 ± 0.03	2.41 ± 0.61	2.52 ± 0.64
14	0.94 ± 0.01	1.31 ± 0.24	1.38 ± 0.24
Overall	0.91 ± 0.04	1.77 ± 0.70	1.86 ± 0.72

The mean ± standard deviations of the Dice similarity coefficient (DSC), mean surface distance (MSD), and average Hausdorff distance (HD_{avg}) for each patient in the markerless dataset.

4

Pancreas Head Segmentation Using Deep Learning for Real-Time Targeting During Radiation Therapy

This chapter extends the deep learning framework introduced in Chapter 3 to pancreatic cancer. Aim 2 is addressed by implementing markerless pancreas head segmentation and identifying the technical challenges of larger motion and deformation compared to the prostate. The pancreas head was selected for this study as it is the most common site of pancreatic cancer. The larger motion of the pancreas head necessitated enhancements to the deep learning framework through deformable data augmentation, achieving a tracking accuracy within several millimetres. This chapter has been prepared for submission to *Medical Physics* as "**Mylonas, A.**, Nguyen, D. T., Stewart, M., Booth, J., Kyme, A., Gardner, M., Mueller, M., & Keall, P. Deep learning segmentation of the pancreas head for real-time targeting during radiation therapy."

4.1 Abstract

Dose escalation has demonstrated promising results for pancreatic cancer radiation therapy. However, the movement of the tumour and surrounding organs can compromise treatment efficacy and increase radiation toxicity. Real-time image-guided adaptive radiation therapy can overcome this challenge by accounting for the motion in real-time to improve dose coverage. The poor soft tissue contrast of the pancreas in kilovoltage images necessitates the implantation of fiducial markers to act as a surrogate. However, the invasive marker implantation procedure carries risks and not all patients are eligible. Therefore, we present a method for pancreas head segmentation in x-ray images for real-time motion management during radiation therapy. We achieved a centroid accuracy of -0.2 ± 3.9 mm and 0.6 ± 3.4 mm in the anterior-posterior/lateral and superior-inferior directions, respectively. The deep learning method can be implemented on a standard linear accelerator eliminating the need for costly, real-time systems and fiducial markers.

4.2 Introduction

Pancreatic cancer has one of the poorest prognoses among cancers with a 5-year survival rate of 10%.[1] Although it was the twelfth most common cancer globally, it ranks as the sixth leading cause of cancer mortality.[2] In 2022, there were an estimated 511,000 new cases and 467,000 deaths globally.[2] Countries with a higher Human Development Index have rates four times greater than those in lower index countries.[2] While incidence rates are the highest in Europe, North America, and Australia/New Zealand, mortality rates are relatively stable.[2,3] The poor treatment outcomes highlight the need for improved cancer therapies such as radiation therapy.

Dose escalation in radiation therapy has shown promising results for improved survival and tumour control for pancreatic cancer.[4–6] A clinical trial investigating hypofractionated ablative radiation therapy demonstrated a 2-year survival rate of 38%.[5] These findings are reinforced by another study, which reported a 2-year survival rate of 36% for dose escalation versus 19% for conventional radiation therapy.[6] However, dose escalated treatments require greater certainty in the radiation hitting the target. This remains a key challenge due to the large motion of the pancreas resulting from respiration. The greatest motion of the pancreas

is found in the superior-inferior direction with an average motion of 15 mm.[7,8]

Motion management strategies for pancreatic radiation therapy include abdominal compression, breath hold, free breathing gating, external guidance, and internal guidance.[9] Methods such as abdominal compression, breath hold, and free breathing gating vary in effectiveness on a per-patient basis.[10,11] External monitoring methods include surface monitoring and the real-time position management (RPM) external marker block. While these methods can monitor motion in real-time, surface motion may not be directly correlated with internal motion. Huguet et al. found that implanted fiducial markers and stents are better predictors of target motion than the RPM block.[12] Internal monitoring can be achieved using the onboard kilovoltage (kV) or megavoltage (MV) imager on standard linear accelerators (linacs) to directly measure the motion.[13]

The low soft tissue contrast of the pancreas necessitates fiducial markers for existing kV real-time motion management systems.[13,14] However, the implantation of fiducial markers is invasive and associated with complications including patient discomfort, infection, bleeding, and reactions to anaesthesia.[15,16] Furthermore, marker migration can result in tracking errors,[17] and CT metal artefacts produced from markers may result in treatment image matching errors.[18] Magnetic resonance imaging (MRI) guidance can be used to visualise the target and surrounding anatomy. However, MRI-linacs are an expensive treatment option for patients and are not widely available compared to standard linacs.[19] This limitation is highlighted in an international survey that found that 71% of radiation therapy centres wish to implement targeted radiation therapy but are limited by resources and capacity.[20]

One solution to overcome the resources and capacity barrier is to use the existing equipment on modern clinical linacs without fiducial markers. A markerless-based would be a step towards making real-time motion management accessible to all patients without additional hardware. The demand for markerless solutions has led to the use of artificial intelligence (AI) for fluoroscopic- and MRI-guided radiation therapy.[21] Markerless-based approaches have been developed for lung,[22–24] diaphragm,[25] liver,[23] pancreas,[26,27] prostate,[28] and head and neck.[29] Therefore, the goal of this work was to implement a deep learning-based pancreas head segmentation framework for real-time pancreatic cancer targeting during radiation therapy. This study focuses on the pancreas head as it is the most common site of pancreatic cancer.[30] Furthermore, the length of the pancreas means the entire organ is not always within the field of view.

4.3 Methods and Materials

4.3.1 Data

The analysis was conducted using imaging data from the Stereotactic Body Radiotherapy (SBRT) for High Risk Localised Pancreatic Cancer (Span-C; NCT03505229[31]) clinical trial. The Span-C clinical trial is investigating the safety, efficacy, and feasibility of SBRT for pancreatic cancer, with a target accrual of 40 patients. Patients with pancreatic cancer who received neoadjuvant chemotherapy and had successful implantation of fiducial markers are eligible. Patients were treated at Royal North Shore Hospital, Australia, and to date, data has been received for 24 patients.

The planning CT, structure set, and kV images from two fractions of each patient associated with this cohort were collected (Fig. 4.1a). The kV images were acquired during patient setup to reconstruct the cone-beam CT (CBCT) under breath-hold conditions. Patients imaged with half-fan geometry were excluded from this study, as the target area was not fully captured within the field of view. Consequently, the first 17 patients in the trial were excluded, while the remaining patients, imaged using full-fan geometry, were included. The training datasets were the 3D planning CT and pancreas head contour from the structure set (Fig. 4.1a).

The test datasets were generated using 1,000 kV images from two fractions (500 images per fraction), except for Patient 1 who had 499 images in one fraction. The ground truth segmentation was created by forward projecting the pancreas head contour to create a 2D digitally reconstructed radiograph (DRR) at each kV image angle. The DRRs were generated using the Reconstruction Toolkit[32] and the Insight Toolkit[33] employing the DRR algorithm derived by Madden et al.[34] The ground truth segmentation was aligned in the kV images based on the implanted fiducial markers (Figure 4.1a). The markers were manually labelled using the open-source Contour Alignment Tool v1.6.3 (`www.github.com/Image-X-Institute/contour-alignment-tool`). Following alignment, the fiducial markers were masked out in all images to avoid biasing the model. The masking algorithm uses the *regionfill* function in MATLAB (The MathWorks, Inc.) to smoothly interpolate inwards from the pixel values surrounding the marker by calculating the discrete Laplacian and solving the Dirichlet boundary value problem. Poisson noise was applied to the kV images following interpolation.

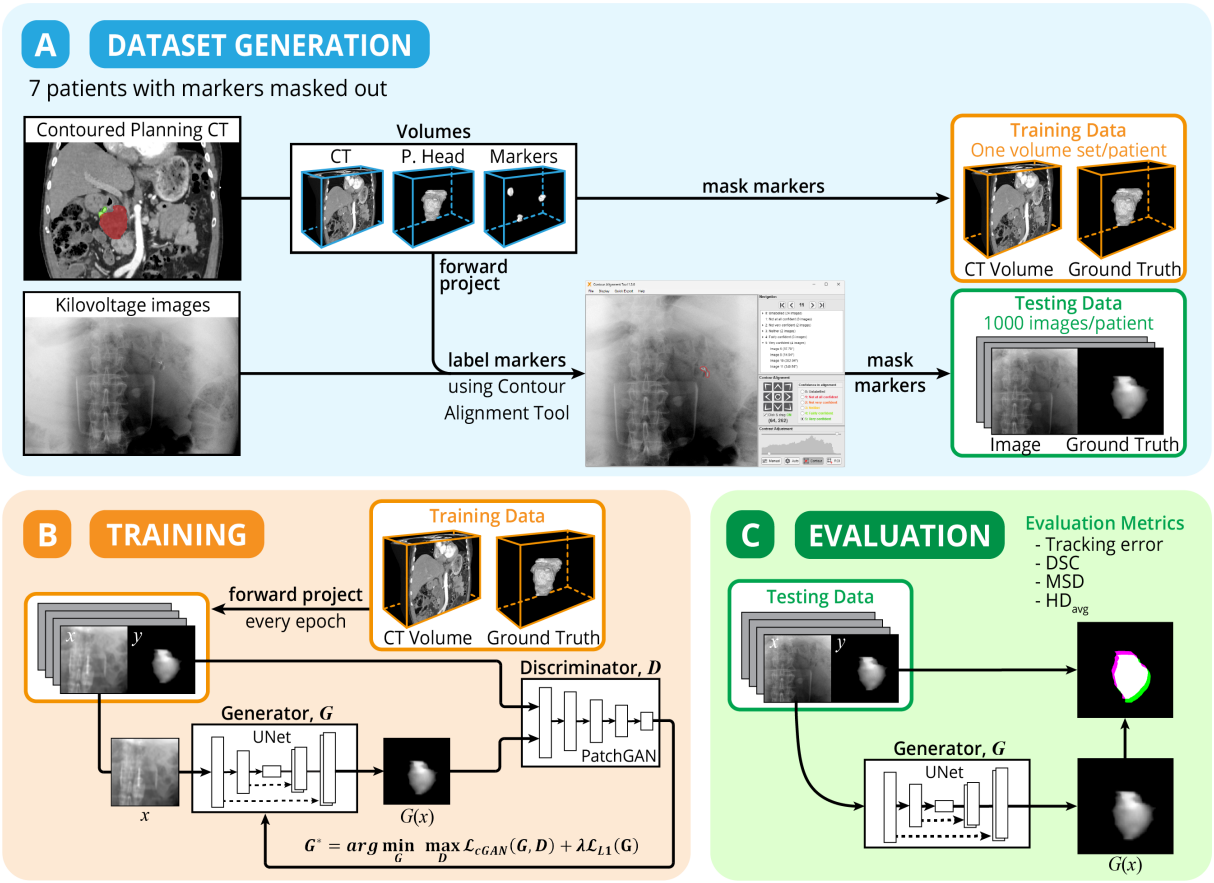


Fig. 4.1 | **Study overview. (a)** The datasets were generated from imaging data of seven patients with pancreatic cancer and implanted fiducial markers. The markers were used to align the ground truth contour and were masked out. The marker positions were manually labelled using the Contour Alignment Tool v1.6.3. **(b)** A conditional Generative Adversarial Network (cGAN) was trained for each patient consisting of a UNet generator network, *G*, and a PatchGAN discriminator network, *D*. **(c)** The cGAN model performance was quantified using the centroid tracking error, Dice similarity coefficient (DSC), mean surface distance (MSD), and average Hausdorff distance (HD_{avg}).

4.3.2 Deep Learning Model

A patient-specific conditional Generative Adversarial Network (cGAN) model was used to segment the pancreas head in 2D kV images based on the Pix2pix model.[35] The cGAN model consists of a UNet generator and a PatchGAN discriminator. The generator, *G*, takes the input kV image, *x*, and creates a segmentation image, *G(x)*, of the pancreas head. The discriminator, *D*, classifies whether the paired image, *xy*, comes from the training set or the generator network as shown in Figure 4.1b.

The framework leverages the patient's pretreatment imaging and planning data, which are available prior to the commencement of their treatment. A patient-specific model was trained for each patient using DRRs produced from the 3D planning CT and pancreas head contour.

For each patient, three models were trained over 120 degree intervals (1–120, 121–240, and 241–360). Initial testing demonstrated poor performance when a single network for the full treatment arc. During training, the 3D volumes were forward projected to produce one 512 × 512 DRR each degree over 120 degrees, generating 120 images.

Data augmentation was performed by applying random rigid shifts (up to 5 mm) and rotations (up to 10 deg) of the CT geometry in all directions. Additionally, deformable augmentation was performed by applying a random shift of the pancreas head contour (up to 10 mm) in all directions. Deformation strength was modelled using a Gaussian function according to:

$$S(\mathrm{x}) = \exp\left(-\frac{||\mathrm{x}-\mathrm{c}||^2}{2\sigma^2}\right) \tag{4.1}$$

where c is the centroid of the pancreas head and σ controls the decay rate of the deformation (set to 20 in this implementation).

Data augmentation was applied ten times prior to each epoch, giving a total training set of 1,200 images for each epoch. The cGAN model was trained for 20 epochs with a batch size of four and a learning rate of 0.0002 using the Adam optimiser. The models were trained on a desktop computer with an Intel® Xeon® Gold 6248R processor (3.0 GHz) with 256 GB RAM and a NVIDIA® RTX A6000 GPU. The cGAN was initialised with a normal distribution and trained to minimise the loss function:

$$G^* = arg\min_G \max_D \mathcal{L}_{\text{cGAN}}(G,D) + \lambda \mathcal{L}_{L1}(G) \tag{4.2}$$

where λ is a constant (set to 100 for this implementation) and:

$$\mathcal{L}_{\text{cGAN}}(G,D) = \mathbb{E}_{(x,y)}\left[\log D(x,y)\right] + \mathbb{E}_x[\log(1 - D(x,G(x)))] \tag{4.3}$$

$$\mathcal{L}_{\text{L1}} = \mathbb{E}_{(x,y)}\left\|y - G(x)\right\|_1 \tag{4.4}$$

4.3.3 Analysis

The models were tested using the kV images acquired as part of radiation therapy to evaluate the accuracy of the pancreas head segmentation and the tracking system (Fig. 4.1c). Each patient-specific model was tested on 500 kV images from two fractions of each patient, giving 1000 test images per patient. To refine the segmentation, template matching was performed

using normalised 2D cross-correlation with the known DRR at each angle.[36] The performance of the generator was quantified by calculating the Dice similarity coefficient (DSC), mean surface distance (MSD), and average Hausdorff distance (HD_{avg}).

In real-time tracking applications, the centroid of the segmentation is crucial and frequently utilised. The segmentation centroids were used to evaluate the generator's performance in an automated tracking system. The tracking system error was defined as the cGAN segmentation centroid minus the ground truth segmentation centroid. The ground truth for this study was defined as the pancreas head DRR contour, shifted by the vector difference between the mean position of the implanted markers in the kV image and the mean position of the markers in the corresponding DRR. The marker positions were manually labelled in each kV image.

The centroid errors were calculated in the anterior-posterior/lateral (AP/LAT) and superior-inferior (SI) directions, and were reported at the patient coordinate system using the source-isocentre distance/source-detector distance ratio as the correction factor. The centroid errors were quantified by calculating the mean error, mean absolute deviation (MAD), and the 5^{th} and 95^{th} percentiles of the errors. Segmentations that were not within 5 cm of the expected plan location were classed as an unsuccessful detection and were excluded from the analysis.

4.4 Results

4.4.1 Tracking Accuracy

The pancreas head was detected in 97% of images for Patients 1 and 5 and 100% of images for all remaining patients (Fig. 4.2a). The cGAN output for these unsuccessful cases typically had a large amount of noise with no discernible segmentation. The centroid errors of the cGAN segmentations are represented in Figure 4.2b. The mean (± standard deviation) centroid error was -0.2 ± 3.9 mm and 0.6 ± 3.4 mm in the AP/LAT and SI directions, respectively (Fig. 4.2b). The MAD was 3.1 mm in the AP/LAT direction and 2.7 mm in the SI direction. Additionally, the 5^{th} and 95^{th} percentiles were -5.9 and 7.2 mm in the AP/LAT direction, and -4.7 and 7.0 mm in the SI direction. There is a similar distribution of errors in the AP/LAT direction with some outliers for patient 6. For some patients, there was an observable offset in the SI direction.

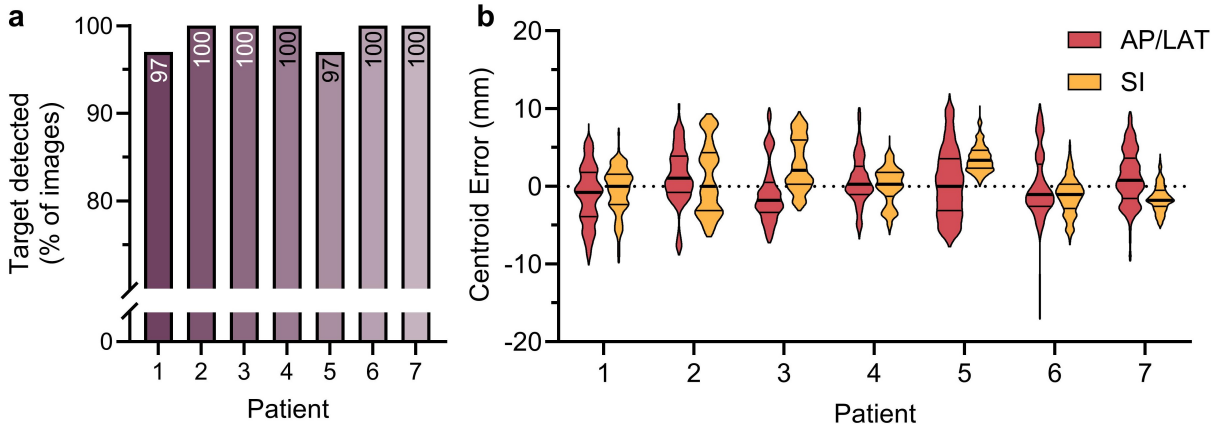


Fig. 4.2 | **Pancreas head detection and centroid error results. (a)** The percentage of images where the conditional Generative Adversarial Network detected the pancreas head. **(b)** Violin plots of the centroid error results in the anterior-posterior/lateral (AP/LAT) directions (red) and superior-inferior (SI) directions (yellow). The bars represent the median (thick) and quartiles (thin).

4.4.2 Segmentation Performance

The segmentation accuracy was assessed using DSC, MSD and HD_{avg} (Fig. 4.3). The agreement between the ground truth and cGAN predictions was characterised by a mean (± standard deviation) DSC across all patients of 0.85 ± 0.09. The 5^{th} and 95^{th} percentiles of the DSC were 0.69 and 0.96. The mean (± standard deviation) MSD and HD_{avg} were 2.6 ± 1.5 mm and 2.7 ± 1.5 mm respectively. The 5^{th} and 95^{th} percentiles were 0.7 and 5.3 mm for MSD, and 0.7 and 5.3 mm for HD_{avg}. As observed for the centroid errors, there were some outliers in all metrics for Patient 6.

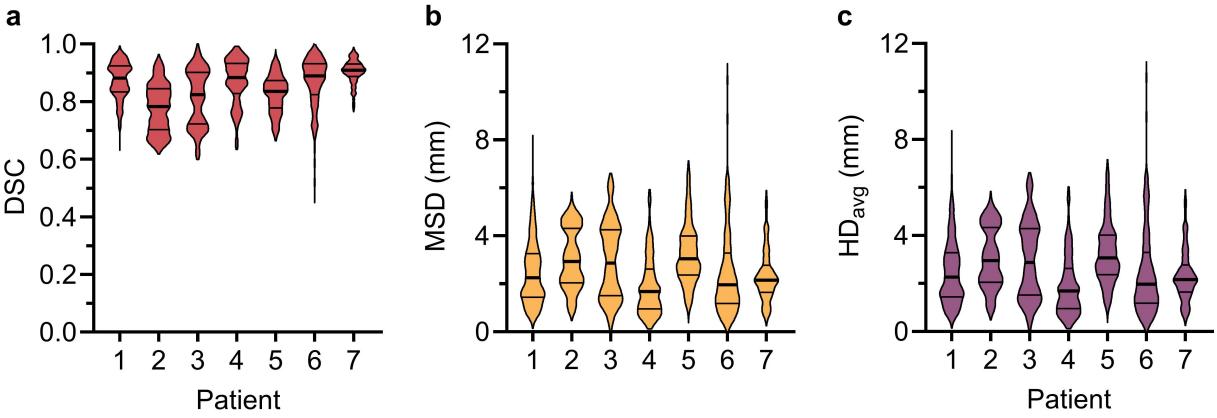


Fig. 4.3 | **Violin plots of the geometric assessment metrics.** The **(a)** DSC (red), **(b)** MSD (yellow), and **(c)** HD_{avg} (purple) between the conditional Generative Adversarial Network segmentation and ground truth for each patient. The bars represent the median (thick) and quartiles (thin). DSC, Dice similarity coefficient; MSD, mean surface distance; HD_{avg}, average Hausdorff distance.

The mean centroid errors and DSC results versus the imager angles are presented in Figure 4.4. The mean centroid errors were consistent across all angles with the exception at 240 degrees, which is the transition angle between two of the cGAN models. The shift was significant in the AP/LAT direction, while it was less noticeable in the SI direction. Similarly, there was a small decrease in the DSC at 240 degrees. However, there was no noticeable difference at the other two transition angles of 0 and 120 degrees. It should be noted that the kV images were obtained through a sub-arc acquisition method. Consequently, the angular range of 250 to 290 has approximately double the number of samples compared to other angles.

The mean training time inclusive of data generation and augmentation was 55 minutes per network, giving a total time of approximately 165 minutes to train all three networks per patient. The time taken for the trained network to generate the segmentation was approximately 10 milliseconds per image using the desktop computer described earlier. Examples of the cGAN and ground truth segmentations at different imager angles are shown in Figure 4.5. This figure demonstrates the results for the average performing patients based on the mean DSC.

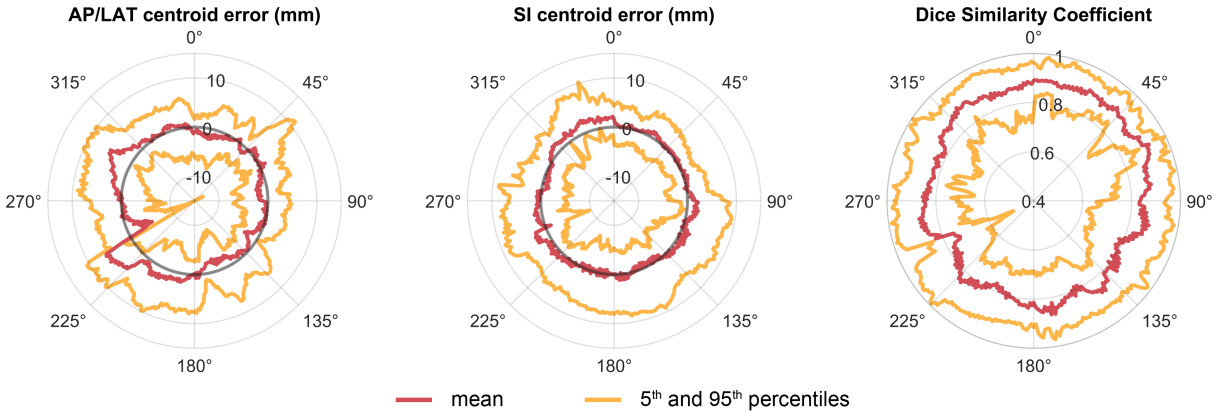


Fig. 4.4 | Centroid error and Dice similarity coefficient (DSC) results based on imager angle. The mean (red) with the 5th and 95th percentiles (yellow) of the centroid error and DSC across the entire radiation therapy treatment arc. AP/LAT, anterior-posterior/lateral; SI, superior-inferior.

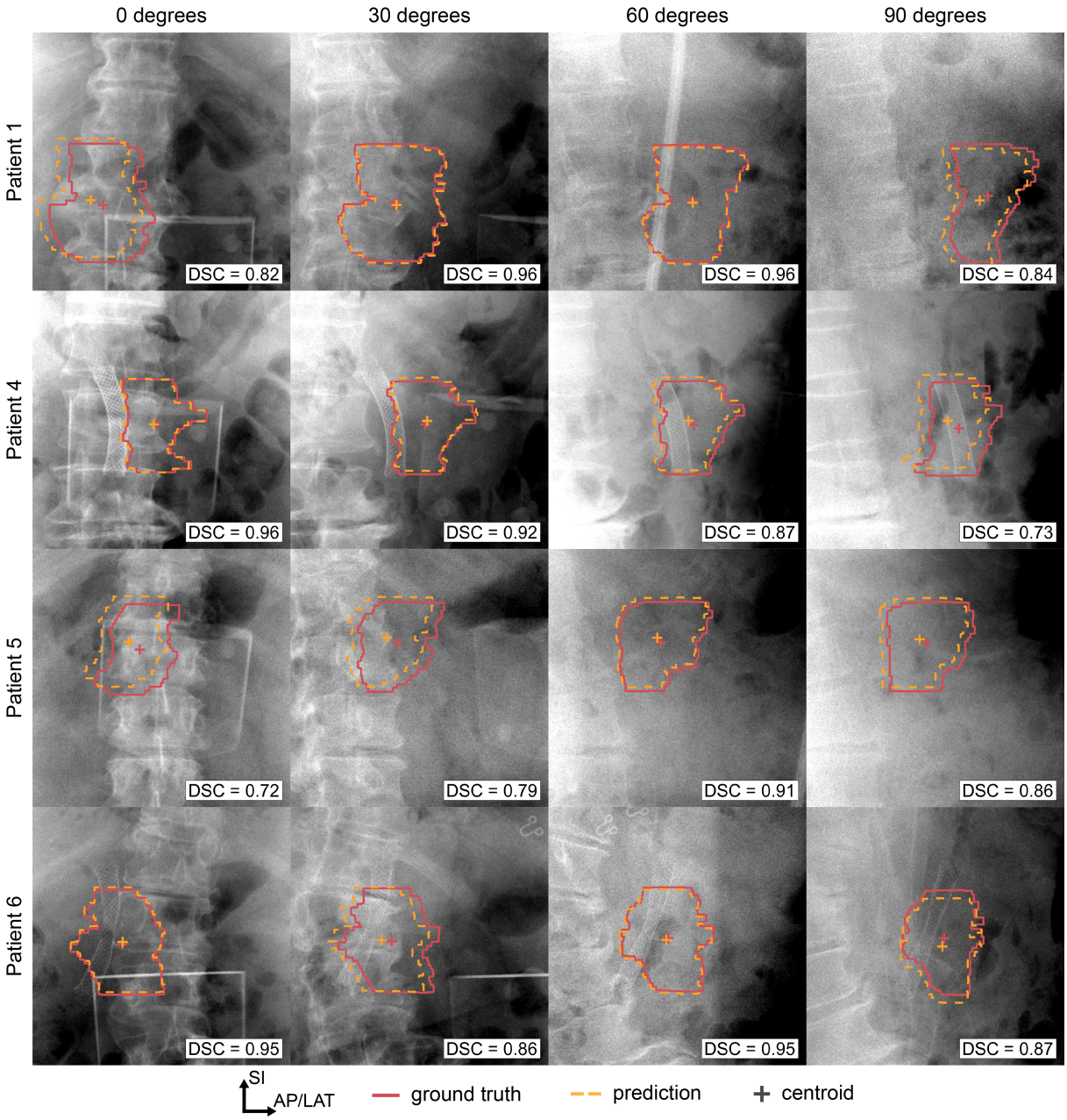


Fig. 4.5 | Example prediction and ground truth segmentations for each patient. Comparison of the ground truth (red) versus conditional Generative Adversarial Network segmentation (yellow) at imager angles of approximately 0, 30, 60, and 90 degrees. The Dice similarity coefficient (DSC) is reported for each image. AP/LAT, anterior-posterior/lateral; SI, superior-inferior.

4.5 Discussion

A framework for pancreas head segmentation for real-time motion management in pancreatic cancer radiation therapy was developed. This was achieved using a patient-specific deep learning model trained on planning data and implemented using data from a conventional radiation therapy system. The pancreas head was chosen for this study for two reasons: first, the

pancreas head is the most common site of pancreatic cancer[30]; and second, due to the length of the pancreas, the entire organ is not always within the field of view. The tracking accuracy was -0.2 ± 3.9 mm and 0.6 ± 3.4 mm in the AP/LAT and SI directions, respectively. The motion of the pancreas is heavily influenced by breathing motion resulting in an average motion of 15 mm in the SI direction.[7,8] Pancreas motion is reduced to within 5 mm with breath-hold but can reach 10 mm for some patients.[37,38]. The accuracy of our deep learning approach is within the expected motion amplitudes for the pancreas during breath hold. However, further improvements are needed to reach the localisation accuracy achievable with implanted fiducial markers, which remain the clinical gold standard for motion monitoring.

Zhao et al.[27] developed a deep learning model to identify the rectangular bounding box containing the planning treatment volume for patients with pancreatic cancer. The model was evaluated on simulated images of two patients for three specific angles: anterior-posterior, left-right, and oblique. Zhao et al. achieved MADs of less than 2.6 mm across all cases and in all directions compared to our MADs of 3.1 mm in the AP/LAT direction and 2.7 mm in the SI direction. The larger errors in our study may be attributed to the more challenging evaluation on treatment-acquired kV images with scatter presence, compared to evaluation on simulated images at defined angles. Additionally, our model has been validated on a larger patient cohort, which may include a wider range of challenging scenarios. Zhou et al.[26] developed a patient-specific, deep learning model for markerless pancreatic tumour contouring, achieving a mean DSC of 0.98. Similarly to Zhao et al. the system was trained and tested on DRR images and the performance on treatment-acquired kV images may vary.

There are several advantages to our approach. First, the model takes less than 3 hours to train all three networks per patient, including data generation and augmentation. This training time frame is feasible for training between the simulation session and first treatment fraction. Furthermore, the training could be optimised by training each network concurrently. Second, the model inference time was 10 ms on average. The AAPM Task Group 264 defines real-time as a system latency below 500 ms[39]. While target segmentation is one component of real-time adaptive treatments, the model's current inference time makes it suitable for real-time applications. Third, as the model is trained on a case-by-case basis, future improvements to kV imaging are likely to improve the model's performance. Fourth, our model produces a segmentation in comparison to the bounding box approach by Zhao et al.[27] The segmentation could be used for other applications such as real-time dose optimisation.[40]

There are some limitations with our pancreas head segmentation method. First, the model was evaluated using kV images acquired during patient setup rather than intrafraction kV images. The kV images acquired during patient setup are of higher quality than intrafraction images. We required images where the fiducial markers were visible for manual labelling of the ground truth. However, this approach could be used in conjunction with triggered imaging[41] or frame averaging[42] to improve kV image quality.

Second, there are uncertainties related to the ground truth of the pancreas head in the kV images. It is not possible to manually contour the pancreas head in kV images due to the poor soft tissue contrast. Therefore, implanted fiducial markers were used to annotate the real-time location of the pancreas and were subsequently masked out. However, the fiducial markers may move over the course of treatment due to tissue deformation[43] or migration[44]. In our study, some fiducial markers were external to the pancreas head. These factors may reduce the accuracy of the ground truth generation method.

Finally, for our framework, the ground truth is based on the planning CT contour which can have uncertainties relating to intraobserver and interobserver variability and does not account for pancreas deformation. Liu et al.[45] quantified interfractional pancreatic head deformation by the maximum overlap ratio between the volume in the planning CT and daily CBCT. The average maximum overlap ratio was 80.2% for the pancreas head, highlighting the large interfractional deformation.[45] While the model may detect deformation in the kV images, the ground truth does not account for the deformation, potentially impacting the model evaluation.

4.6 Conclusions

A framework for markerless pancreas head segmentation during radiation therapy was developed and tested on patient kV images. Our method requires only a brief training period using the patient's planning data and relies on kV image acquisition during treatment. Therefore, the framework has the potential to be implemented as a software only solution on existing, standard linacs. Centroid accuracy within several millimetres was achieved with a mean DSC of 0.85. Future work will look at further evaluation in scenarios with MV scatter and experimental studies on a physical phantom using patient motion traces.

4.7 References

1. Mizrahi, J. D., Surana, R., Valle, J. W. & Shroff, R. T. Pancreatic cancer. *The Lancet* **395**, 2008–2020 (2020).

2. Bray, F. et al. Global cancer statistics 2022: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA: a cancer journal for clinicians* **74**, 229–263 (2024).

3. Carioli, G. et al. European cancer mortality predictions for the year 2021 with focus on pancreatic and female lung cancer. *Annals of Oncology* **32**, 478–487 (2021).

4. Petrelli, F. et al. Stereotactic body radiation therapy for locally advanced pancreatic cancer: a systematic review and pooled analysis of 19 trials. *International Journal of Radiation Oncology* Biology* Physics* **97**, 313–322 (2017).

5. Reyngold, M. et al. Association of ablative radiation therapy with survival among patients with inoperable pancreatic cancer. *JAMA oncology* **7**, 735–738 (2021).

6. Krishnan, S. et al. Focal radiation therapy dose escalation improves overall survival in locally advanced pancreatic cancer patients receiving induction chemotherapy and consolidative chemoradiation. *International Journal of Radiation Oncology* Biology* Physics* **94**, 755–765 (2016).

7. Heerkens, H. D. et al. MRI-based tumor motion characterization and gating schemes for radiation therapy of pancreatic cancer. *Radiotherapy and Oncology* **111**, 252–257 (2014).

8. Gwynne, S. et al. Respiratory movement of upper abdominal organs and its effect on radiotherapy planning in pancreatic cancer1. *Clinical Oncology* **21**, 713–719 (2009).

9. Reyngold, M., Parikh, P. & Crane, C. H. Ablative radiation therapy for locally advanced pancreatic cancer: techniques and results. *Radiation Oncology* **14**, 1–8 (2019).

10. Van Gelder, R. et al. Experience with an abdominal compression band for radiotherapy of upper abdominal tumours. *Journal of medical radiation sciences* **65**, 48–54 (2018).

11. Keall, P. J. et al. The management of respiratory motion in radiation oncology report of AAPM Task Group 76 a. *Medical physics* **33**, 3874–3900 (2006).

12. Huguet, F. et al. Modeling pancreatic tumor motion using 4-dimensional computed tomography and surrogate markers. *International Journal of Radiation Oncology* Biology* Physics* **91**, 579–587 (2015).

13. Keall, P. et al. Real-time image guided ablative prostate cancer radiation therapy: results from the TROG 15.01 SPARK trial. *International Journal of Radiation Oncology* Biology* Physics* **107**, 530–538 (2020).

14. Vinogradskiy, Y., Goodman, K. A., Schefter, T., Miften, M. & Jones, B. L. The clinical and dosimetric impact of real-time target tracking in pancreatic SBRT. *International Journal of Radiation Oncology* Biology* Physics* **103**, 268–275 (2019).

15. Dutta, D. et al. Prospective evaluation of fiducial marker placement quality and toxicity in liver CyberKnife stereotactic body radiotherapy. *Radiation Oncology Journal* **38**, 253 (2020).

16. Park, S. H. et al. Efficacy and safety of ultrasound-guided implantation of fiducial markers in the liver for stereotactic body radiation therapy. *PLoS One* **12**, e0179676 (2017).

17. Bhagat, N. et al. Complications associated with the percutaneous insertion of fiducial markers in the thorax. *Cardiovascular and interventional radiology* **33**, 1186–1191 (2010).

18. Nemoto, H. et al. Evaluation of computed tomography metal artifact and CyberKnife fiducial recognition for novel size fiducial markers. *Journal of Applied Clinical Medical Physics* **24**, e14142 (2023).

19. van Herk, M., McWilliam, A., Dubec, M., Faivre-Finn, C. & Choudhury, A. Magnetic resonance imaging–guided radiation therapy: a short strengths, weaknesses, opportunities, and threats analysis (2018).

20. Anastasi, G. et al. Patterns of practice for adaptive and real-time radiation therapy (POP-ART RT) part I: Intra-fraction breathing motion management. *Radiotherapy and Oncology* **153**, 79–87 (2020).

21. Mylonas, A., Booth, J. & Nguyen, D. T. A review of artificial intelligence applications for motion tracking in radiotherapy. *Journal of Medical Imaging and Radiation Oncology* **65**, 596–611 (2021).

22. Terunuma, T., Tokui, A. & Sakae, T. Novel real-time tumor-contouring method using deep learning to prevent mistracking in X-ray fluoroscopy. *Radiological physics and technology* **11**, 43–53 (2018).

23. Hirai, R., Sakata, Y., Tanizawa, A. & Mori, S. Real-time tumor tracking using fluoroscopic imaging with deep neural network analysis. *Physica Medica* **59**, 22–29 (2019).

24. Takahashi, W., Oshikawa, S. & Mori, S. Real-time markerless tumour tracking with patient-specific deep learning using a personalised data generation strategy: proof of concept by phantom study. *The British journal of radiology* **93**, 20190420 (2020).

25. Edmunds, D., Sharp, G. & Winey, B. Automatic diaphragm segmentation for real-time lung tumor tracking on cone-beam CT projections: a convolutional neural network approach. *Biomedical physics & engineering express* **5**, 035005 (2019).

26. Zhou, D., Nakamura, M., Mukumoto, N., Yoshimura, M. & Mizowaki, T. Development of a deep learning-based patient-specific target contour prediction model for markerless tumor positioning. *Medical physics* **49**, 1382–1390 (2022).

27. Zhao, W. et al. Markerless pancreatic tumor target localization enabled by deep learning. *International Journal of Radiation Oncology* Biology* Physics* **105**, 432–439 (2019).

28. Zhao, W. et al. Incorporating imaging information from deep neural network layers into image guided radiation therapy (IGRT). *Radiotherapy and Oncology* **140**, 167–174 (2019).

29. Gardner, M. et al. Realistic CT data augmentation for accurate deep-learning based segmentation of head and neck tumors in kV images acquired during radiation therapy. *Medical Physics* **50**, 4206–4219 (2023).

30. Modolell, I., Guarner, L. & Malagelada, J. Vagaries of clinical presentation of pancreatic and biliary tract cancer. *Annals of oncology* **10**, S82–S84 (1999).

31. Royal North Shore Hospital. Span-C-SBRT for Pancreatic Cancer (Span-C). `https://clinicaltrials.gov/study/NCT03505229` (2023).

32. Rit, S. et al. The Reconstruction Toolkit (RTK), an open-source cone-beam CT reconstruction toolkit based on the Insight Toolkit (ITK). In *Journal of Physics: Conference Series*, vol. 489, 012079 (IOP Publishing, 2014).

33. McCormick, M., Liu, X., Jomier, J., Marion, C. & Ibanez, L. ITK: enabling reproducible research and open science. *Frontiers in neuroinformatics* **8**, 13 (2014).

34. Madden, L. et al. CBCT-DRRs superior to CT-DRRs for target-tracking applications for pancreatic SBRT. *Biomedical Physics & Engineering Express* **10**, 035039 (2024).

35. Isola, P., Zhu, J.-Y., Zhou, T. & Efros, A. A. Image-to-image translation with conditional adversarial networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, 1125–1134 (2017).

36. Lewis, J. P. et al. Fast template matching. In *Vision interface*, vol. 95, 15–19 (Quebec City, QC, Canada, 1995).

37. Lens, E., van der Horst, A., Versteijne, E., Bel, A. & van Tienhoven, G. Considerable pancreatic tumor motion during breath-holding. *Acta Oncologica* **55**, 1360–1368 (2016).

38. Zeng, C. et al. Intrafraction tumor motion during deep inspiration breath hold pancreatic cancer treatment. *Journal of applied clinical medical physics* **20**, 37–43 (2019).

39. Keall, P. J. et al. AAPM Task Group 264: The safe clinical implementation of MLC tracking in radiotherapy. *Medical physics* **48**, e44–e64 (2021).

40. Mejnertsen, L., Hewson, E., Nguyen, D. T., Booth, J. & Keall, P. Dose-based optimisation for multi-leaf collimator tracking during radiation therapy. *Physics in Medicine & Biology* **66**, 065027 (2021).

41. Kaur, G., Lehmann, J., Greer, P. & Simpson, J. Assessment of the accuracy of truebeam intrafraction motion review (IMR) system for prostate treatment guidance. *Australasian Physical & Engineering Sciences in Medicine* **42**, 585–598 (2019).

42. Ng, J. A. et al. Kilovoltage intrafraction monitoring for prostate intensity modulated arc therapy: first clinical results. *International Journal of Radiation Oncology* Biology* Physics* **84**, e655–e661 (2012).

43. Van der Horst, A. et al. Interfractional position variation of pancreatic tumors quantified using intratumoral fiducial markers and daily cone beam computed tomography. *International Journal of Radiation Oncology* Biology* Physics* **87**, 202–208 (2013).

44. Sanders, M. K. et al. EUS-guided fiducial placement for stereotactic body radiotherapy in locally advanced and recurrent pancreatic cancer. *Gastrointestinal endoscopy* **71**, 1178–1184 (2010).

45. Liu, F., Erickson, B., Peng, C. & Li, X. A. Characterization and management of interfractional anatomic changes for pancreatic cancer radiotherapy. *International Journal of Radiation Oncology* Biology* Physics* **83**, e423–e429 (2012).

5

Deep Learning-Based Contrast Agent Segmentation for Liver Cancer Targeting During Radiation Therapy

Chemotherapy is often delivered via transarterial chemoemobilisation for patients with liver cancer, leaving residual radiopaque contrast agents. The contrast agents remain visible in treatment imaging for patients who subsequently receive radiation therapy. Therefore, this chapter addresses Aim 3 by identifying a cohort of patients where real-time markerless tracking can be successfully achieved using the contrast agents. This chapter has been prepared for submission to the *International Journal of Radiation Oncology - Biology - Physics* as "**Mylonas, A.**, Nguyen, D.T., Seshadri, V., Ramachandran, R., Lye, J., Khor, R., Dipuglia, A., Greer, P., Kyme, A., Booth, J., Gardner, M., Sengupta, C., Mueller, M., & Keall, P. Deep learning-based contrast agent segmentation for liver cancer targeting during radiation therapy."

The clinical trial protocol and initial results discussed in this chapter are presented in Appendix A and have been published as "Plant, N., **Mylonas, A.**, Sengupta, C., Nguyen, D. T., Silvester, S., Pryor, D., Greer, P., Lee, Y. Y. D., Ramachandran, P., Seshadri, V., Trada, Y., Khor, R., Wang, T., Hardcastle, N., & Keall, P. Radio-opaque contrast agents for liver cancer targeting with KIM during radiation therapy (ROCK-RT): An observational feasibility study. *Radiation Oncology*, **19**, 139 (2024)."

5.1 Abstract

Liver cancer is associated with increasing incidence and poor treatment outcomes in many countries. The motion of the tumour and surrounding organs limits the effectiveness of radiation therapy in this context. To overcome this challenge, real-time image-guided adaptive radiation therapy can improve dose coverage by accounting for the motion in real-time. However, implanted fiducial markers are typically required for motion tracking due to poor soft tissue contrast in kilovoltage images. Patients with hepatocellular carcinoma, the most common form of primary liver cancer, often receive chemotherapy via transcatheter arterial chemoembolisation. The radiopaque chemoembolisation agents can remain visible in imaging for months afterwards making it a potential surrogate to aid tracking. Therefore, we present the first method for using the chemoembolisation agents as a method for real-time guidance during radiation therapy. The mean tracking accuracy was -1.0 ± 4.7 mm and 1.4 ± 3.1 mm in the AP/LAT and SI directions, respectively, with a mean Dice similarity coefficient of 0.78 ± 0.19. Our deep learning approach has the potential to increase targeting accuracy without the need for an additional invasive procedure to implant fiducial markers.

5.2 Introduction

Liver cancer is the sixth most common cancer globally and the third leading cause of cancer related deaths, with an estimated 865,000 new cases and 757,000 deaths in 2022.[1] While incidence rates have been decreasing in high-risk countries, there has been an increase or stabilisation in most countries across Europe, North America, Australia/New Zealand, and South America.[1] The increasing incidence and poor outcomes highlight the need for improved liver cancer therapy. Radiation therapy is a widely used liver cancer treatment that is recommended for one in two patients with cancer.[2]

One of the challenges of radiation therapy is the significant movement of tumours and surrounding anatomy that can occur during treatment, particularly for tumours in the liver. Respiration is the main contributor to liver motion, which predominates in the superior-inferior and anterior-posterior directions. While liver stereotactic ablative body radiotherapy (SABR) is associated with improved control rates for primary liver tumours and metastases,[3] the higher SABR radiation dose to the normal liver tissue carries a significant risk of radia-

tion induced liver damage.[4] In this context, motion management is crucial for improving dose coverage and reducing the radiation dose to healthy tissue.[5,6]

The most commonly used respiratory motion management methods for thoracic and abdominal cancers include internal target volume (ITV), breath hold, abdominal compression, and free breathing gating.[7] Common to these methods is the lack of internal anatomical imaging that could enable more accurate targeting and smaller margins.[7,8] In the case of breath hold, the technique may not be well correlated with the internal motion,[9,10] and does not account for non-breathing associated motion of liver tumours. Furthermore, it may not be suitable for all patients. Eccles et al.[11] demonstrated that only 62% of patients were suitable for exhale breath hold with the active breathing coordinator, while Hardcastle et al.[12] found that 68% of patients achieved reproducible voluntary exhale breath hold.

Real-time image-guided adaptive radiation therapy (IGART) can enable the visualisation of internal anatomy during treatment, addressing the limitations of other motion management methods. IGART can be performed by acquiring kilovoltage (kV) images during treatment using the on-board imager that is on modern radiation therapy treatment systems. The desire for real-time tumour tracking has led to the use of implanted markers as surrogates of the tumour, especially for organs and tumours with low radiographic contrast such as the liver.[5] However, the invasive marker implantation procedure adds time delays, additional costs, and risks. Complications associated with marker implantation include patient discomfort, infection, bleeding, and reactions to anaesthesia.[13,14] Some patients with cirrhosis, underlying portal hypertension and associated thrombocytopaenia may pose a higher bleeding risk. Furthermore, marker migration can result in tracking errors,[15] and the metal artefacts produced from markers in computed tomography (CT) images may result in treatment image matching errors.[16]

An international survey found that 71% of radiation therapy centres wish to implement targeted radiation therapy but are limited by resources and capacity.[17] An ideal solution, and the premise underpinning this research, is to achieve the real-time tracking benefits of implanted markers without the need for the invasive implantation procedure. A markerless-based approach using a conventional therapy system would help make real-time IGART accessible to all patients without additional hardware. The concept of this study is to leverage the contrast agents typical of hepatocellular carcinoma (HCC) treatment protocols instead of implanted fiducials, and the benefits of deep learning for image-based tracking, to enable robust mark-

erless targeted radiation therapy on conventional treatment hardware.

HCC accounts for 75%–85% of primary liver cancer cases.[1] For HCC patients, chemotherapy is often delivered via transcatheter arterial chemoemobilisation (TACE). The procedure involves injecting chemotherapy drugs into the artery supplying the tumour combined with embolic particles to restrict blood supply and keep the drug from washing away, and a radiopaque contrast agent. The contrast agent allows direct visualisation of the treated tumour at the time of TACE and remains visible on imaging for months afterwards (Figure 5.1). The most common radiopaque agent used in this setting is ethiodised oil, also known as Lipiodol. For patients undergoing chemoradiation treatment, the retained radiopaque contrast agent can be used as a surrogate tumour marker to aid target verification before and during radiation therapy.[18–20] Therefore, the contrast agent could be used as a non-invasive method for real-time tumour tracking.

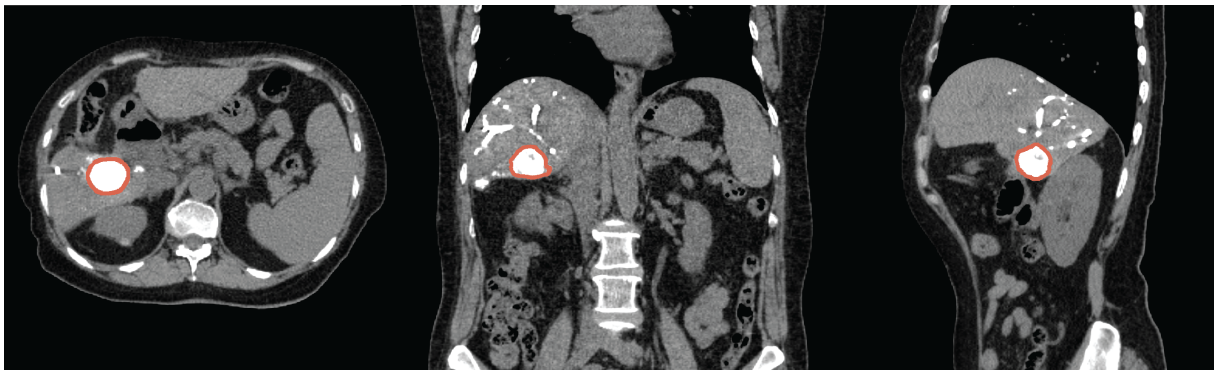


Fig. 5.1 | Lipiodol mass in the liver. Axial, coronal, and sagittal views (left to right) showing a lipiodol contrast agent mass (red) in the liver. The current study aims to use this contrast for real-time image guidance during liver radiation therapy.

Artificial intelligence has shown great potential for fluoroscopic and magnetic resonance imaging guided radiation therapy.[21] Markerless approaches have been developed for lung[22–24], diaphragm[25], liver[23], pancreas[26,27], prostate[28], and head and neck[29]. We hypothesise that deep learning can be used to accurately segment the contrast agents. Therefore, the goal of this work was to develop and investigate deep learning-based contrast agent segmentation for real-time liver cancer targeting during radiation therapy.

5.3 Methods and Materials

5.3.1 Data

The analysis was conducted using imaging data from the ethics-approved Radio-opaque Contrast Agents for Liver Cancer Targeting With KIM During Radiation Therapy (ROCK-RT; NCT05169177) study.[30] ROCK-RT is an observational study to investigate whether radiopaque contrast agents can be detected by tumour-tracking software. Patients with liver cancer who received SABR after TACE are eligible, with a target accrual of 50 patients.[30] To date, 15 patients have been recruited with data received for seven patients. One patient has been excluded from this study as the contrast agent was not fully contained within the kV imager field of view.

The patients were treated across two different sites in Australia (Princess Alexandra Hospital, Australia; and Olivia Newton John Cancer Wellness & Research Centre, Australia). We collected the planning CT, structure set, and kV images from two fractions of each patient associated with this cohort (Fig. 5.2a). The kV images were acquired during patient setup to reconstruct the CBCT. The training datasets consisted of the 3D planning CT and contrast agent contour.

Test datasets were generated using 50 kV images per fraction, resulting in 100 test images per patient, evenly spaced across the imaging arc. The ground truth was manually labelled in each kV image by performing a rigid shift of the 2D contour using the Contour Alignment Tool v1.6.3 (`www.github.com/Image-X-Institute/contour-alignment-tool`). The Contour Alignment Tool generates the ground truth by forward projecting the 3D contrast agent contour to give a 2D digitally reconstructed radiograph (DRR) at each kV image angle. DRRs were generated using the Reconstruction Toolkit[31] and the Insight Toolkit[32] using the DRR algorithm as derived by Madden et al.[33]. The confidence in the ground truth labelling was scored on a 5-point Likert scale in the Contour Alignment Tool with the following levels: *not at all confident*, *not very confident*, *neither*, *fairly confident*, and *very confident*.

5.3.2 Deep Learning Model

The central component of the system was a conditional Generative Adversarial Network (cGAN) model for segmentation of the contrast agent in 2D kV images. The cGAN implementation used for this study was based on the Pix2pix model.[34] The model uses two convolutional neu-

ral networks called the generator and discriminator. The generator, G, takes the input kV image, x, and creates a segmentation image, $G(x)$, of the contrast agent. The discriminator, D, classifies whether the paired image, xy, comes from the training set or the generator network as shown in Figure 5.2b. Through adversarial learning, the generator learns to create realistic outputs while the discriminator learns to distinguish them from real data, driving both networks to improve iteratively. A UNet was used for the generator architecture and a PatchGAN for the discriminator architecture.

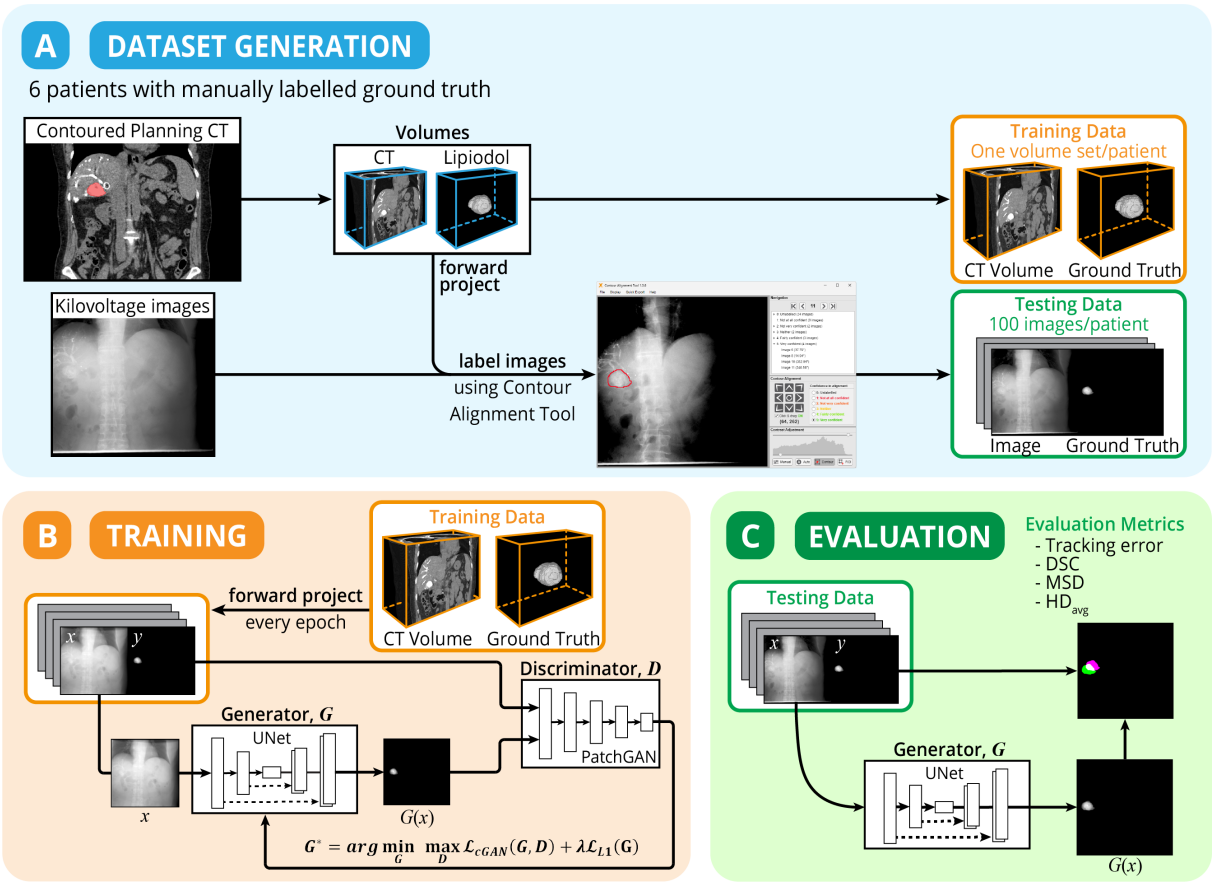


Fig. 5.2 | **Study overview. (a)** The datasets were generated from imaging data of six patients with liver cancer with retained chemoembolisation agents. The ground truth was manually labelled using the Contour Alignment Tool v1.6.3. **(b)** A conditional Generative Adversarial Network (cGAN) was trained for each patient consisting of a UNet generator network, G, and a PatchGAN discriminator network, D. **(c)** The cGAN model performance was quantified using the centroid tracking error, Dice similarity coefficient (DSC), mean surface distance (MSD), and average Hausdorff distance (HD_{avg}).

The model is trained using the patient's pretreatment imaging and planning data, which are available prior to the commencement of their treatment. A patient-specific model was trained for each patient using DRRs produced from the 3D planning CT and contrast agent contour. For each patient, three models were trained over 120 degree intervals (1–120, 121–

240, and 241–360). The sub-interval approach was adopted because initial testing revealed poor performance when using a single model for the full-range. Specifically, for angles separated by 180 degrees, the model may be unable to distinguish between opposite sides resulting in segmentation errors where the target was mirrored. The volumes were forward-projected to produce one 512 × 512 DRR each degree over 120 degrees, generating 120 images. Subsequently, noise was applied to each image by sampling from a Poisson distribution, where the mean of the distribution (λ) for each element is determined by the corresponding value in the image.

During training, data augmentation was performed by applying random rigid shifts (up to 5 mm) and rotations (up to 10 deg) of the CT geometry in all directions. Additionally, deformable augmentation was performed by applying a random shift of the contrast agent mass (up to 10 mm) in all directions. Deformation strength was modelled using a Gaussian function according to:

$$S(\mathrm{x}) = \exp\left(-\frac{||\mathrm{x}-\mathrm{c}||^2}{2\sigma^2}\right) \tag{5.1}$$

where c is the centroid of the contrast volume and σ, which controls the decay rate of the deformation, was set to 20.

Data augmentation was applied ten times prior to each epoch, giving a total training set of 1,200 images. Each model was trained for 20 epochs with a batch size of four and a learning rate of 0.0002 using the Adam optimiser. The models were trained on a desktop computer with an Intel® Xeon® Gold 6248R processor (3.0 GHz) with 256 GB RAM and a NVIDIA® RTX A6000 GPU. The cGAN was initialised with a normal distribution and trained to minimise the loss function:

$$G^* = arg\min_G \max_D \mathcal{L}_{\text{cGAN}}(G,D) + \lambda\mathcal{L}_{L1}(G) \tag{5.2}$$

where λ is a constant (set to 100 for this implementation) and:

$$\mathcal{L}_{\text{cGAN}}(G,D) = \mathbb{E}_{(x,y)}\left[\log D(x,y)\right] + \mathbb{E}_x[\log(1-D(x,G(x)))] \tag{5.3}$$

$$\mathcal{L}_{\text{L1}} = \mathbb{E}_{(x,y)}\left\|y - G(x)\right\|_1 \tag{5.4}$$

5.3.3 Analysis

The contrast agent may consist in a single mass or multiple smaller masses. The volume and mean intensity of each mass was analysed. Since some masses may be too small or diluted to track, a subset was selected for tracking in cases with multiple masses. The masses were filtered using criteria based on each patient's specific data: only contrast volumes with a volume greater than the patient's mean contrast volume and intensity above the patient's mean contrast Hounsfield units (HU) were retained. These thresholds were based on observed tracking performance, ensuring that only the most reliably trackable volumes were included for each patient.

The models were tested using the kV images acquired as part of radiation therapy to evaluate the accuracy of the contrast agent segmentation and the tracking system (Fig. 5.2c). Each patient-specific model was tested on kV images from two fractions of each patient, giving 100 test images per patient (50 per fraction). To refine the network-derived segmentation, template matching was performed using normalised 2D cross-correlation,[35] with the known contrast agent DRR at each angle used as the template. The performance of the generator was quantified by calculating the Dice similarity coefficient (DSC), mean surface distance (MSD), and average Hausdorff distance (HD_{avg}).

The tracking system error was quantified as the distance between the cGAN and ground truth segmentation centroids. The errors were calculated in the anterior-posterior/lateral (AP/LAT) and superior-inferior (SI) directions, and were reported in the patient coordinate system using the source-isocentre distance/source-detector distance ratio as the correction factor. The centroid errors were quantified by calculating the mean error, mean absolute deviation (MAD), and the 5^{th} and 95^{th} percentiles of the errors. Segmentations that were not within 5 cm of the expected plan location were classed as an unsuccessful detection and were excluded from the analysis.

5.4 Results

5.4.1 Contrast Agent Analysis

Variable distributions of contrast agent masses were observed across the six patients as shown in Figure 5.3. Three patients had a single mass (Patients 3, 5, and 6) and three had a discon-

tinuous mass (Patients 1, 2, and 4) with up to 11 discrete regions (Fig. 5.3c). An example of a discontinuous mass is shown in Figure 5.3d. For Patients 1 and 2, there was a clear distinction between a grouping of larger and smaller masses (Fig. 5.3a). However, Patient 4 was characterised by nine very small ($<$0.36 cm^3) volumes. The mean intensity of the volumes ranged from 90 to 795 HU (Fig. 5.3b). The filtering criteria was applied to reduce the number of small and dilute contrast agent masses, resulting in one to three suitable masses to track per patient (Fig. 5.3c and Extended Data Tables 5.1–5.6).

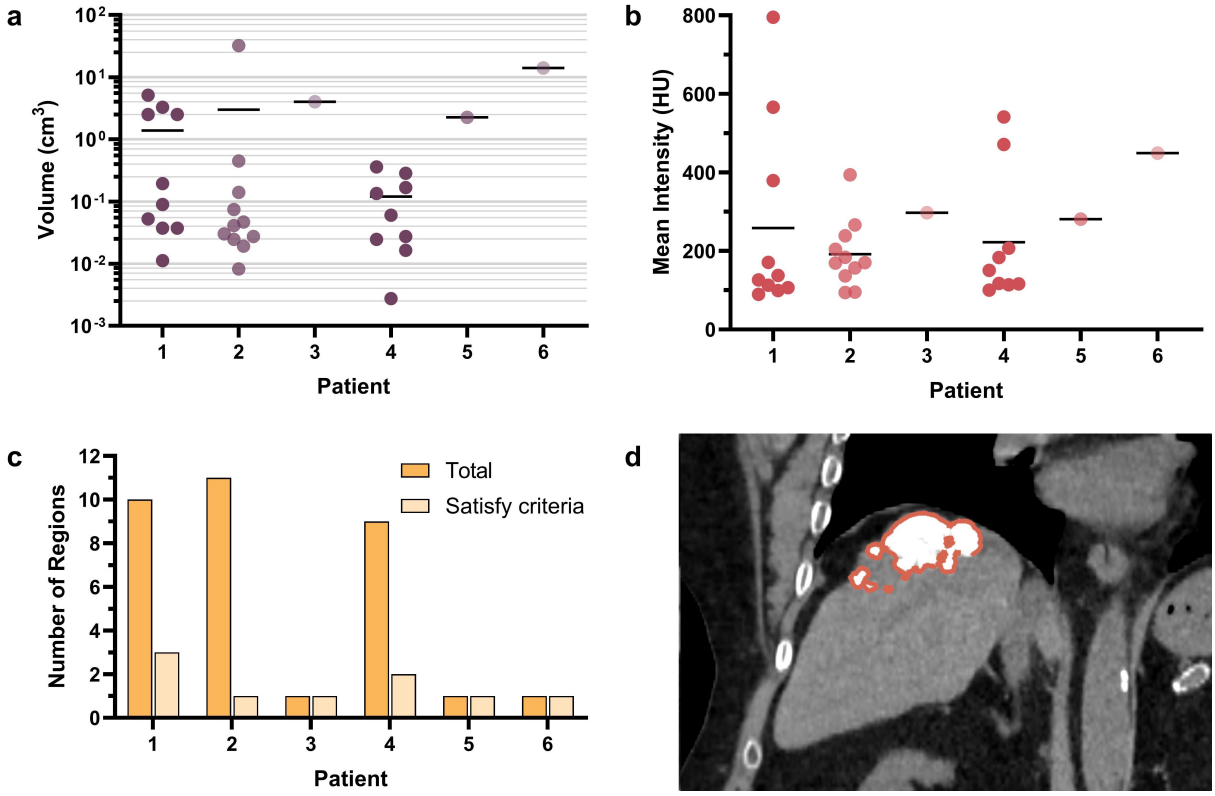


Fig. 5.3 | Analysis of contrast agent masses. (a) Volume and **(b)** mean intensity of each contrast agent mass, with horizontal bars indicating the mean. **(c)** Total number of volumes before and after applying the filter criteria. **(d)** Example of a discontinuous volume from Patient 2.

The confidence in the ground truth labels is shown in Figure 5.4. Labels for Patients 3 and 5 had the lowest overall confidence with none receiving a score $>$3 (*neither*). Labels for Patient 6 had the highest confidence with all labels receiving a score of 5 (*very confident*). Overall, the majority of labels (55%) across all patients were scored with a confidence ≥ 4. There were no observable trends between the mass volume or intensity and the confidence of the label.

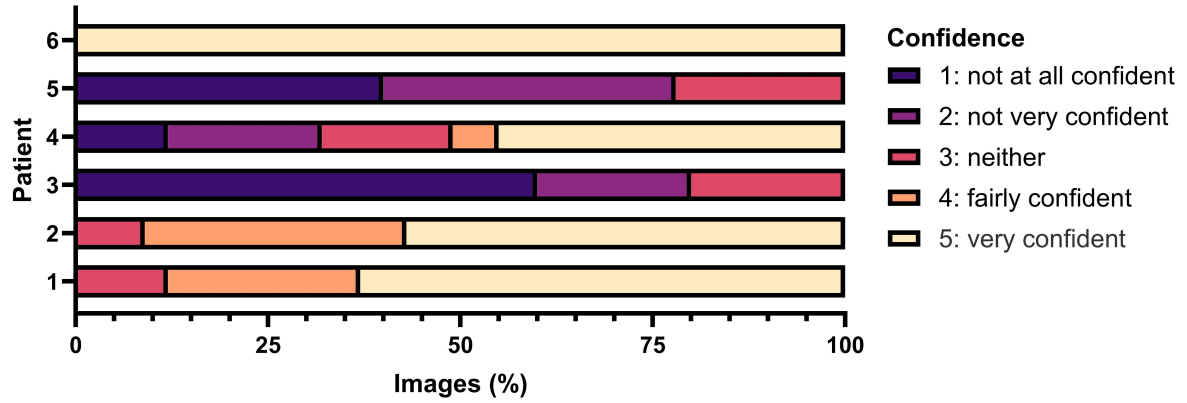


Fig. 5.4 | Confidence of ground truth labels. The distribution of label confidences on a 5-point Likert scale from *not at all confident* to *very confident*.

5.4.2 Tracking Accuracy

In some images, the cGAN model was unable to detect any contrast agent. Segmentations that were not within 5 cm of the expected plan location were excluded from the analysis (Fig. 5.5a). The contrast agent was detected in 100% of images for four patients (2, 3, 5 and 6). Whereas there was a 93% and 0% detection rate for Patients 1 and 4, respectively. Therefore, Patient 4 was excluded from the remaining analysis as the contrast was not detected in any kV images. The largest tracked mass for Patient 4 was 0.29 cm^3 which was substantially smaller than the next largest tracked mass for any other patient (2.26 cm^3).

The centroid errors of the cGAN segmentations for all patients are represented in Figure 5.5b. The mean (± standard deviation) centroid error was -1.0 ± 4.7 mm and 1.4 ± 3.1 mm in the AP/LAT and SI directions, respectively (Fig. 5.5b). The MAD was 3.3 mm in the AP/LAT direction and 1.9 mm in the SI direction. Additionally, the 5^{th} and 95^{th} percentiles were -10.7 and 5.2 mm in the AP/LAT direction, and -2.1 and 4.7 mm in the SI direction. The violin plots demonstrate a non-normal nature of the distributions. Notably, some outliers from Patient 1 contributed to larger errors, increasing the overall variance.

5.4.3 Segmentation Performance

The performance of the cGAN in terms of the DSC, MSD, and HD_{avg} is presented in Figure 5.6. The agreement between the ground truth and cGAN predictions was characterised by a mean (± standard deviation) DSC across all patients of 0.78 ± 0.19. The 5^{th} and 95^{th} percentiles of the DSC were 0.32 and 0.96. The mean MSD and HD_{avg} were 2.29 ± 2.07 mm and 2.25 ± 1.76

mm respectively. The 5th and 95th percentiles were 0.46 and 5.41 mm for MSD, and 0.46 and 5.57 mm for HD_{avg}.

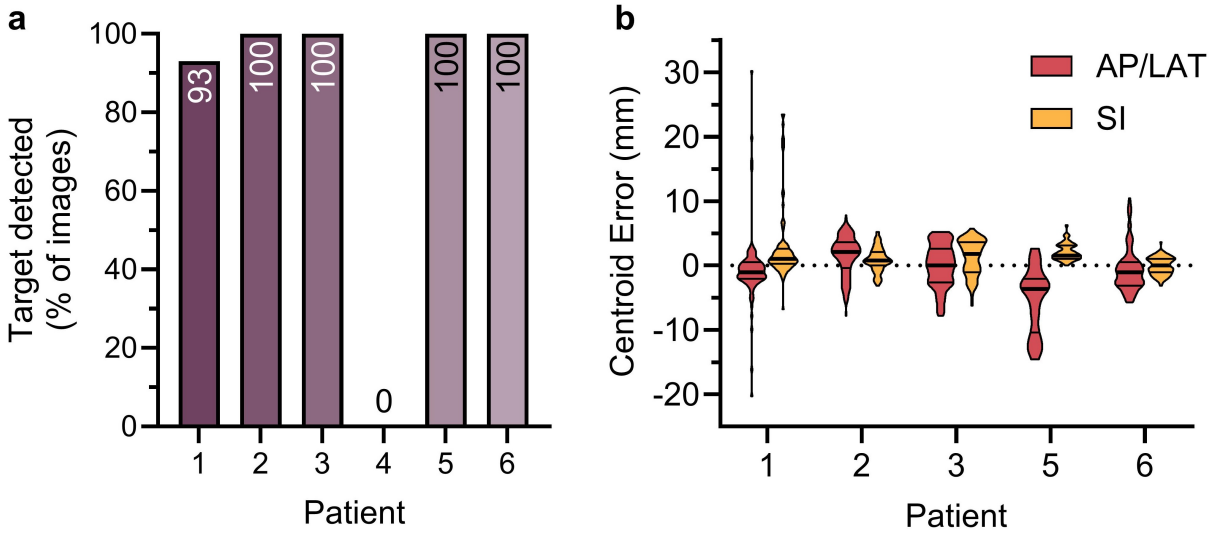


Fig. 5.5 | Mass detection and centroid error results. (a) The percentage of images where the conditional Generative Adversarial Network detected the contrast agent mass. **(b)** Violin plots of the centroid error results in the anterior-posterior/lateral (AP/LAT) directions (red) and superior-inferior (SI) directions (yellow). The bars represent the median (thick) and quartiles (thin).

As observed with the centroid error, large outliers were present in the segmentation metrics for Patient 1. Additionally, Patient 5 showed a noticeable decrease in DSC performance, which can be attributed to the sensitivity of DSC in small volumes. Examples of the cGAN and ground truth segmentations at different imager angles for all patients are shown in Figure 5.7. At these angles it can be observed that there is good agreement in the segmentations with a DSC greater than 0.8 for most images.

5.5 Discussion

We present a segmentation system which uses chemoemobilisation agents for real-time motion tracking in liver cancer radiation therapy. This was achieved using a patient-specific deep learning model trained on data from a conventional radiation therapy system at two different treatment centres. The tracking accuracy was -1.0 ± 4.7 mm and 1.4 ± 3.1 mm in the AP/LAT and SI directions, respectively with a MAD of 3.3 mm in the AP/LAT direction and 1.9 mm in the SI direction. Intrafraction motion of the liver is considerably large due to the contribution of breathing motion.[36] From a study of 11 patients, Worm et al. reported mean 3D

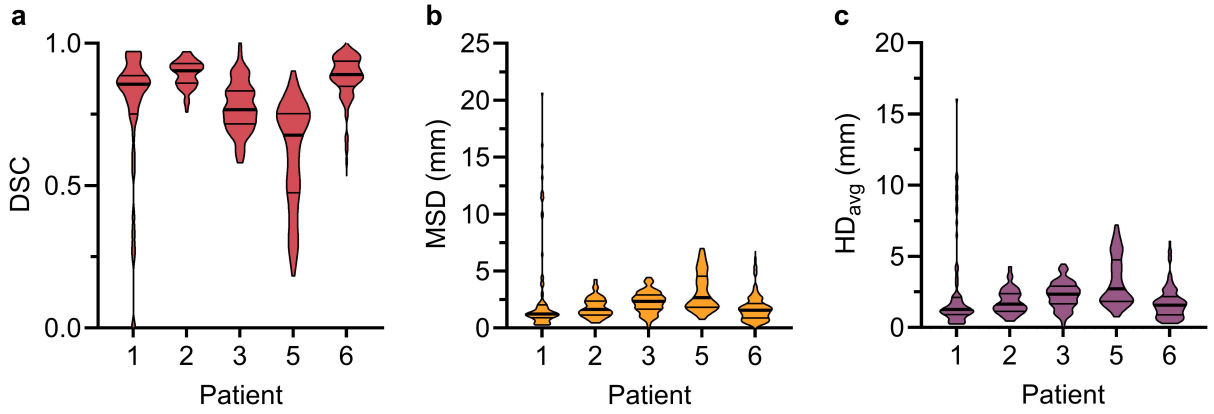


Fig. 5.6 | Violin plots of the geometric assessment metrics. The **(a)** DSC (red), **(b)** MSD (yellow), and **(c)** HD_{avg} (purple) between the conditional Generative Adversarial Network segmentation and ground truth for each patient. The bars represent the median (thick) and quartiles (thin). DSC, Dice similarity coefficient; MSD, mean surface distance; HD_{avg}, average Hausdorff distance.

intrafraction motion of 17.6 mm during liver SABR with a maximum of 39.5 mm.[37] There was a weak increase in mean intrafraction displacement with elapsed time.[37] In another study, Kitamura et al. reported average tumour motion of 4 ± 4 mm, 9 ± 5 mm, and 5 ± 3 mm in the left-right, SI, and AP directions.[38] The deep learning approach achieved an accuracy within several millimetres which is less than the lower range of expected motion amplitudes for the liver.

To the best of our knowledge, this is the first study to use residual contrast agent from TACE for real-time image guidance during radiation therapy. Yue et al. previously investigated the use of lipiodol for CBCT image guidance.[18] It was found that through using lipiodol the planning treatment volume margins could be reduced from 5 to 2.5 mm in the mediolateral direction, 8 to 4.0 mm in the SI direction, and 8 to 2.9 mm in AP direction. Margin reduction is important for high dose treatments such as SABR as the higher radiation dose carries a significant risk of radiation-induced liver damage.[4] Through motion management, the dose coverage is improved to the tumour and there is lower dose to the surrounding healthy tissue.[5,39]

For this study, the ground truth was manually labelled in all kV images using the Contour Alignment Tool. Overall, the majority of images could be labelled with some confidence. However, for two patients there was low confidence in the labels. There were no observable trends between the lipiodol volume or intensity and the confidence in the label. Patient anatomy can play a role in the image quality and the visibility of the contrast agent, and may have contributed to the confidence scoring. The patterns of retained lipiodol can be classified into four

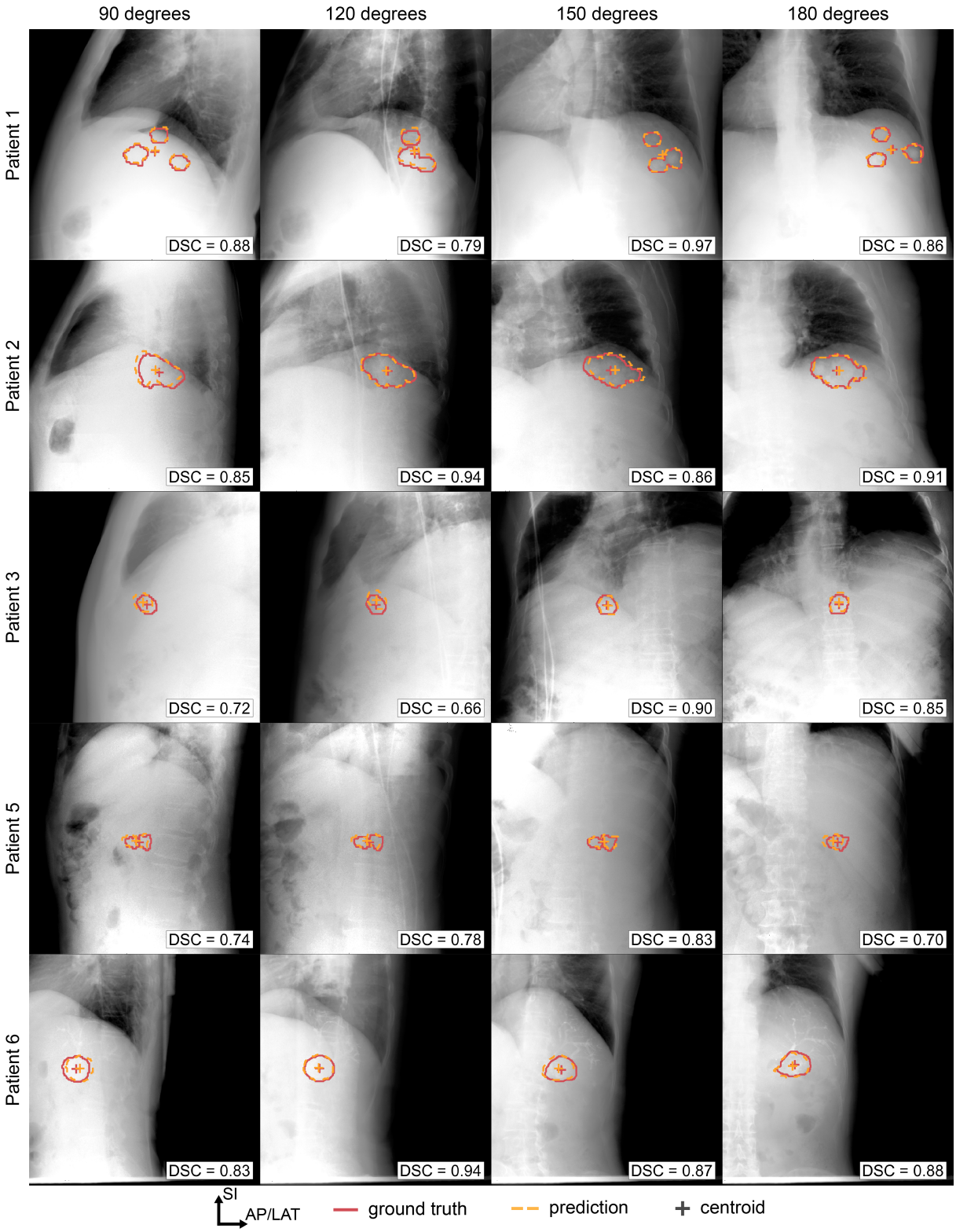


Fig. 5.7 | Example ground truth and prediction segmentations for each patient. Comparison of the ground truth (red) versus conditional Generative Adversarial Network segmentation (yellow) at imager angles of approximately 90, 120, 150, and 180 degrees. The Dice similarity coefficient (DSC) is reported for each image. AP/LAT, anterior-posterior/lateral; SI, superior-inferior.

categories: homogenous accumulation (type 1), partial defect (type 2), sporadic accumulation (type 3), and punctuate or absent accumulation (type 4).[18] Types 1 and 2 are charcterised by a single continuous accumulation, whereas types 3 and 4 consist of discontinuous masses. Yue et al. reported that only types 1 and 2 are clinically acceptable to use as a surrogate for CBCT guidance.[18]

Our study is not without limitations. One factor was the inability to track smaller and/or multiple masses. The deep learning model trained on Patient 4 failed to detect the two smaller masses, which had volumes (mean intensities) of 0.29 cm^3 (541 HU) and 0.17 cm^3 (471 HU). The next largest volume that was tracked across the other five patients was 2.26 cm^3. Patient 3 was the next worse in performance with several outlier results. The volumes for these two patients would fall into categories 3 or 4, which are not clinically acceptable for CBCT guidance. This limitation likely extends to real-time kV guidance as well. Our results show that for the remaining patients there is a higher level of accuracy. Based on the current results, this approach may need to be limited to retention categories 1 and 2 and to volumes larger than 1 cm^3. The exclusion criteria will be refined as further analysis is conducted on additional patients.

The second limitation was that the model was evaluated on kV images during patient setup. The kV images acquired during patient setup are of higher quality than intrafraction images with MV scatter. As the ground truth in this study was manually labelled, we require images of higher quality where the target was visible. However, this approach could be applied in an environment where MV scatter is managed through triggered imaging[40] and frame averaging[41]. In scenarios with MV scatter, further evaluation would be required before deployment.

5.6 Conclusions

An algorithm for liver contrast agent segmentation during radiation therapy treatment was developed and evaluated on patient kV images. The preliminary results demonstrate that real-time deep learning-based segmentation within several millimetres is achievable for some types of retention patterns. This solution could enable a non-invasive method for real-time tumour tracking without the costs, time and side effects for marker implantation. The method will undergo continuous refinement and evaluation on additional patients as ROCK-RT recruitment progresses.

5.7 References

1. Bray, F. et al. Global cancer statistics 2022: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA: a cancer journal for clinicians* **74**, 229–263 (2024).

2. Barton, M. B. et al. Estimating the demand for radiotherapy from the evidence: a review of changes from 2003 to 2012. *Radiotherapy and oncology* **112**, 140–144 (2014).

3. Ohri, N. et al. Local control after stereotactic body radiation therapy for liver tumors. *International Journal of Radiation Oncology* Biology* Physics* **110**, 188–195 (2021).

4. Miften, M. et al. Radiation dose-volume effects for liver SBRT. *International Journal of Radiation Oncology* Biology* Physics* **110**, 196–205 (2021).

5. Sengupta, C. et al. The first clinical implementation of real-time 6 degree-of-freedom image-guided radiotherapy for liver SABR patients. *Radiotherapy and Oncology* **190**, 110031 (2024).

6. Poulsen, P. R. et al. Respiratory gating based on internal electromagnetic motion monitoring during stereotactic liver radiation therapy: first results. *Acta Oncologica* **54**, 1445–1452 (2015).

7. Ball, H. J. et al. Results from the AAPM Task Group 324 respiratory motion management in radiation oncology survey. *Journal of applied clinical medical physics* **23**, e13810 (2022).

8. Distefano, G., Garikipati, S., Grimes, H. & Hatton, M. Current status of stereotactic ablative body radiotherapy in the UK: six years of progress. *BJR open* **1** (2019).

9. Hoisak, J. D., Sixel, K. E., Tirona, R., Cheung, P. C. & Pignol, J.-P. Correlation of lung tumor motion with external surrogate indicators of respiration. *International Journal of Radiation Oncology* Biology* Physics* **60**, 1298–1306 (2004).

10. Dawson, L. A. et al. The reproducibility of organ position using active breathing control (ABC) during liver radiotherapy. *International Journal of Radiation Oncology* Biology* Physics* **51**, 1410–1421 (2001).

11. Eccles, C., Brock, K. K., Bissonnette, J.-P., Hawkins, M. & Dawson, L. A. Reproducibility of liver position using active breathing coordinator for liver cancer radiotherapy. *International Journal of Radiation Oncology* Biology* Physics* **64**, 751–759 (2006).

12. Hardcastle, N. et al. Selection of motion management in liver stereotactic body radiotherapy and its impact on treatment time. *Physics and Imaging in Radiation Oncology* **25**, 100407 (2023).

13. Dutta, D. et al. Prospective evaluation of fiducial marker placement quality and toxicity in liver CyberKnife stereotactic body radiotherapy. *Radiation Oncology Journal* **38**, 253 (2020).

14. Park, S. H. et al. Efficacy and safety of ultrasound-guided implantation of fiducial markers in the liver for stereotactic body radiation therapy. *PLoS One* **12**, e0179676 (2017).

15. Bhagat, N. et al. Complications associated with the percutaneous insertion of fiducial markers in the thorax. *Cardiovascular and interventional radiology* **33**, 1186–1191 (2010).

16. Nemoto, H. et al. Evaluation of computed tomography metal artifact and CyberKnife fiducial recognition for novel size fiducial markers. *Journal of Applied Clinical Medical Physics* **24**, e14142 (2023).

17. Anastasi, G. et al. Patterns of practice for adaptive and real-time radiation therapy (POP-ART RT) part I: Intra-fraction breathing motion management. *Radiotherapy and Oncology* **153**, 79–87 (2020).

18. Yue, J. et al. Lipiodol: a potential direct surrogate for cone-beam computed tomography image guidance in radiotherapy of liver tumor. *International Journal of Radiation Oncology* Biology* Physics* **82**, 834–841 (2012).

19. Hawkins, M. A. et al. Assessment of residual error in liver position using kV cone-beam computed tomography for liver cancer high-precision radiation therapy. *International Journal of Radiation Oncology* Biology* Physics* **66**, 610–619 (2006).

20. Nakagawa, K. et al. Contrast medium-assisted stereotactic image-guided radiotherapy using kilovoltage cone-beam computed tomography. *Japanese Journal of Radiology* **26**, 570 (2008).

21. Mylonas, A., Booth, J. & Nguyen, D. T. A review of artificial intelligence applications for motion tracking in radiotherapy. *Journal of Medical Imaging and Radiation Oncology* **65**, 596–611 (2021).

22. Terunuma, T., Tokui, A. & Sakae, T. Novel real-time tumor-contouring method using deep learning to prevent mistracking in X-ray fluoroscopy. *Radiological physics and technology* **11**, 43–53 (2018).

23. Hirai, R., Sakata, Y., Tanizawa, A. & Mori, S. Real-time tumor tracking using fluoroscopic imaging with deep neural network analysis. *Physica Medica* **59**, 22–29 (2019).

24. Takahashi, W., Oshikawa, S. & Mori, S. Real-time markerless tumour tracking with patient-specific deep learning using a personalised data generation strategy: proof of concept by phantom study. *The British journal of radiology* **93**, 20190420 (2020).

25. Edmunds, D., Sharp, G. & Winey, B. Automatic diaphragm segmentation for real-time lung tumor tracking on cone-beam CT projections: a convolutional neural network approach. *Biomedical physics & engineering express* **5**, 035005 (2019).

26. Zhou, D., Nakamura, M., Mukumoto, N., Yoshimura, M. & Mizowaki, T. Development of a deep learning-based patient-specific target contour prediction model for markerless tumor positioning. *Medical physics* **49**, 1382–1390 (2022).

27. Zhao, W. et al. Markerless pancreatic tumor target localization enabled by deep learning. *International Journal of Radiation Oncology* Biology* Physics* **105**, 432–439 (2019).

28. Zhao, W. et al. Incorporating imaging information from deep neural network layers into image guided radiation therapy (IGRT). *Radiotherapy and Oncology* **140**, 167–174 (2019).

29. Gardner, M. et al. Realistic CT data augmentation for accurate deep-learning based segmentation of head and neck tumors in kV images acquired during radiation therapy. *Medical Physics* **50**, 4206–4219 (2023).

30. Plant, N. et al. Radio-opaque contrast agents for liver cancer targeting with KIM during radiation therapy (ROCK-RT): an observational feasibility study. *Radiation Oncology* **19**, 139 (2024).

31. Rit, S. et al. The Reconstruction Toolkit (RTK), an open-source cone-beam CT reconstruction toolkit based on the Insight Toolkit (ITK). In *Journal of Physics: Conference Series*, vol. 489, 012079 (IOP Publishing, 2014).

32. McCormick, M., Liu, X., Jomier, J., Marion, C. & Ibanez, L. ITK: enabling reproducible research and open science. *Frontiers in neuroinformatics* **8**, 13 (2014).

33. Madden, L. et al. CBCT-DRRs superior to CT-DRRs for target-tracking applications for pancreatic SBRT. *Biomedical Physics & Engineering Express* **10**, 035039 (2024).

34. Isola, P., Zhu, J.-Y., Zhou, T. & Efros, A. A. Image-to-image translation with conditional adversarial networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, 1125–1134 (2017).

35. Lewis, J. P. et al. Fast template matching. In *Vision interface*, vol. 95, 15–19 (Quebec City, QC, Canada, 1995).

36. Shirato, H., Seppenwoolde, Y., Kitamura, K., Onimura, R. & Shimizu, S. Intrafractional tumor motion: lung and liver. In *Seminars in radiation oncology*, vol. 14, 10–18 (Elsevier, 2004).

37. Worm, E. S., Høyer, M., Fledelius, W. & Poulsen, P. R. Three-dimensional, time-resolved, intrafraction motion monitoring throughout stereotactic liver radiation therapy on a conventional linear accelerator. *International Journal of Radiation Oncology* Biology* Physics* **86**, 190–197 (2013).

38. Kitamura, K. et al. Tumor location, cirrhosis, and surgical history contribute to tumor movement in the liver, as measured during stereotactic irradiation using a real-time tumor-tracking radiotherapy system. *International Journal of Radiation Oncology* Biology* Physics* **56**, 221–228 (2003).

39. Gargett, M., Haddad, C., Kneebone, A., Booth, J. & Hardcastle, N. Clinical impact of removing respiratory motion during liver SABR. *Radiation Oncology* **14**, 1–9 (2019).

40. Kaur, G., Lehmann, J., Greer, P. & Simpson, J. Assessment of the accuracy of truebeam intrafraction motion review (IMR) system for prostate treatment guidance. *Australasian Physical & Engineering Sciences in Medicine* **42**, 585–598 (2019).

41. Ng, J. A. et al. Kilovoltage intrafraction monitoring for prostate intensity modulated arc therapy: first clinical results. *International Journal of Radiation Oncology* Biology* Physics* **84**, e655–e661 (2012).

5.8 Extended Data

The following data presents the contrast agent mass measurements for each patient. The masses were filtered using criteria based on each patient's specific data: only contrast volumes with a volume greater than the patient's mean volume and intensity above the patient's mean intensity were retained. Individual mass measurements exceeding the mean values are highlighted in bold. Additionally, masses that met both criteria and were used for tracking purposes are highlighted in blue.

Extended Data Table 5.1 | Patient 1 contrast agent mass measurements.

Contrast volume	Volume (cm^3)	Mean Intensity (HU)
1	**3.29**	**795**
2	**2.51**	**379**
3	**2.51**	**566**
4	0.01	126
5	0.05	106
6	0.04	90
7	0.04	137
8	0.09	113
9	**5.12**	171
10	0.19	99
Mean	1.39	258

Extended Data Table 5.2 | Patient 2 contrast agent mass measurements.

Contrast volume	Volume (cm^3)	Mean Intensity (HU)
1	0.07	136
2	0.03	94
3	**32.07**	**394**
4	0.02	95
5	0.14	**266**
6	0.45	**239**
7	0.02	157
8	0.05	184
9	0.04	170
10	0.03	169
11	0.01	**204**
Mean	2.99	192

Extended Data Table 5.3 | Patient 3 contrast agent mass measurements.

Contrast volume	Volume (cm^3)	Mean Intensity (HU)
1	**4.03**	**297**
Mean	4.03	297

Extended Data Table 5.4 | Patient 4 contrast agent mass measurements.

Contrast volume	Volume (cm^3)	Mean Intensity (HU)
1	**0.17**	**471**
2	**0.29**	**541**
3	0.02	116
4	**0.13**	207
5	**0.36**	183
6	0.03	117
7	0.06	151
8	0.00	114
9	0.02	100
Mean	0.12	222

Extended Data Table 5.5 | Patient 5 contrast agent mass measurements.

Contrast volume	Volume (cm^3)	Mean Intensity (HU)
1	**2.26**	**281**
Mean	2.26	281

Extended Data Table 5.6 | Patient 6 contrast agent mass measurements.

Contrast volume	Volume (cm^3)	Mean Intensity (HU)
1	**14.02**	**449**
Mean	14.02	449

6

Contour Alignment Tool: An Open-Source Application for Visualising and Aligning 3D Contours in 2D Intrafraction Images

AI models for markerless segmentation require extensively labelled datasets for training and evaluation. These datasets can be difficult and time consuming to generate. Therefore, the objective of this chapter is to develop an application that allows users to quickly and easily label intrafraction images, leveraging the radiation therapy planning data. The developed Contour Alignment Tool (`www.github.com/Image-X-Institute/contour-alignment-tool`) is open-source and was used to label the test datasets in Chapters 4 and 5. This chapter has been prepared for submission to *SoftwareX* as "**Mylonas, A.**, Nguyen, D.T., Shieh, C.C., Seshadri, V., Ramachandran, P., & Keall, P. Contour Alignment Tool: An open-source application for visualising and aligning 3D contours in 2D intrafraction images."

6.1 Abstract

Real-time image-guided radiation therapy can be used to track the tumour during radiation therapy to improve dose coverage. Recently, artificial intelligence (AI) has shown great potential for x-ray and magnetic resonance imaging guided radiation therapy. However, an obstacle to AI implementation in radiation therapy is the lack of large, expertly labelled datasets which are time consuming to create. The contouring of 2D intrafraction images can be optimised by leveraging the 3D treatment planning data. This chapter presents the Contour Alignment Tool (`www.github.com/Image-X-Institute/contour-alignment-tool`), an open-source application that allows users to quickly and easily match contours projected from 3D radiation therapy treatment planning data to intrafraction images acquired during treatment. The contours can be moved rigidly to the correct position and exported as binary masks facilitating the creation of labelled data for machine learning projects. The tool is being used for two large, ethics-approved multi-centre clinical trials (ROCK-RT, NCT05169177; LEARN, NCT05184790).

6.2 Introduction

During radiation therapy, the tumour and surrounding anatomy are dynamically moving which can lead to underdosing tumours and overdosing the healthy tissue. The mounting evidence of the detrimental effects of underdosing tumours and overdosing organs at risk led the American Society for Radiation Oncology to recommend continuous imaging for tumour motion monitoring during high dose radiation treatments.[1] Real-time image guided radiation therapy (IGRT) monitors and accounts for the tumour motion, ensuring adequate dose coverage.[2] Current motion management solutions typically rely on implanted fiducial markers or expensive hardware such as magnetic resonance imaging (MRI) linear accelerators (linacs) to visualise the tumour.[3] However, a worldwide survey found that 71% of radiation therapy centres wish to implement targeted radiation therapy for respiratory motion management but are limited by resources and capacity.[4] Therefore, a solution that can be implemented on a conventional linac, without the need for implanting fiducial markers, would be beneficial for cancer centres and patients.

Artificial intelligence (AI) has shown great potential for real-time x-ray IGRT without fiducial markers.[5] Markerless-based approaches have been developed for lung,[6–8] diaphragm,[9]

liver,[7] pancreas,[10,11] prostate,[12] and head and neck.[13] However, an obstacle to AI implementation in the clinic is the lack of large, expertly labelled datasets. For the described task, images acquired during treatment are required to be labelled with the contour of the target and in some cases the surrounding organs. The task of data labelling is time consuming due to the large quantity of images required to train and evaluate AI models. The goal of this work was to develop an open-source software solution that allows users to efficiently contour intrafraction images. The Contour Alignment Tool can be used to generate large, expertly labelled datasets that are required to develop AI-based real-time IGRT solutions implemented on conventional linacs.

6.3 Methods and Materials

The Contour Alignment Tool consists of a main graphical user interface (GUI) that allows the user to load the relevant data and then contour the images. The Contour Alignment Tool has been implemented using MATLAB App Designer (The MathWorks Inc., Natick, Massachusetts, USA) and has been tested in the R2021a release. MATLAB App Designer is an integrated environment for designing a GUI with the supporting code. The Contour Alignment Tool code requires the Image Processing Toolbox, Statistics and Machine Learning Toolbox, and Parallel Computing Toolbox for development and compilation. Additionally, the tool can be run as a standalone application with MATLAB Runtime. MATLAB Compiler is required to build the standalone application. The standalone application, code, and resources are publicly available on GitHub (`www.github.com/Image-X-Institute/contour-alignment-tool`).

6.3.1 Treatment Data Import

The Contour Alignment Tool imports DICOM planning data using the DICOM functions in the MATLAB Image Processing Toolbox and intrafraction images using in-house developed code. After launching the application, the user can select the treatment plans and intrafraction images to be imported using the GUI (Fig. 6.1). The required planning data includes the CT, plan, and structure set DICOM files. These files can be imported individually. Alternatively, if all the DICOM files are contained in one folder, the user can select the folder to automatically import all files at once. Once the structure set file has been imported, the structures will be listed in a drop-down menu for user selection.

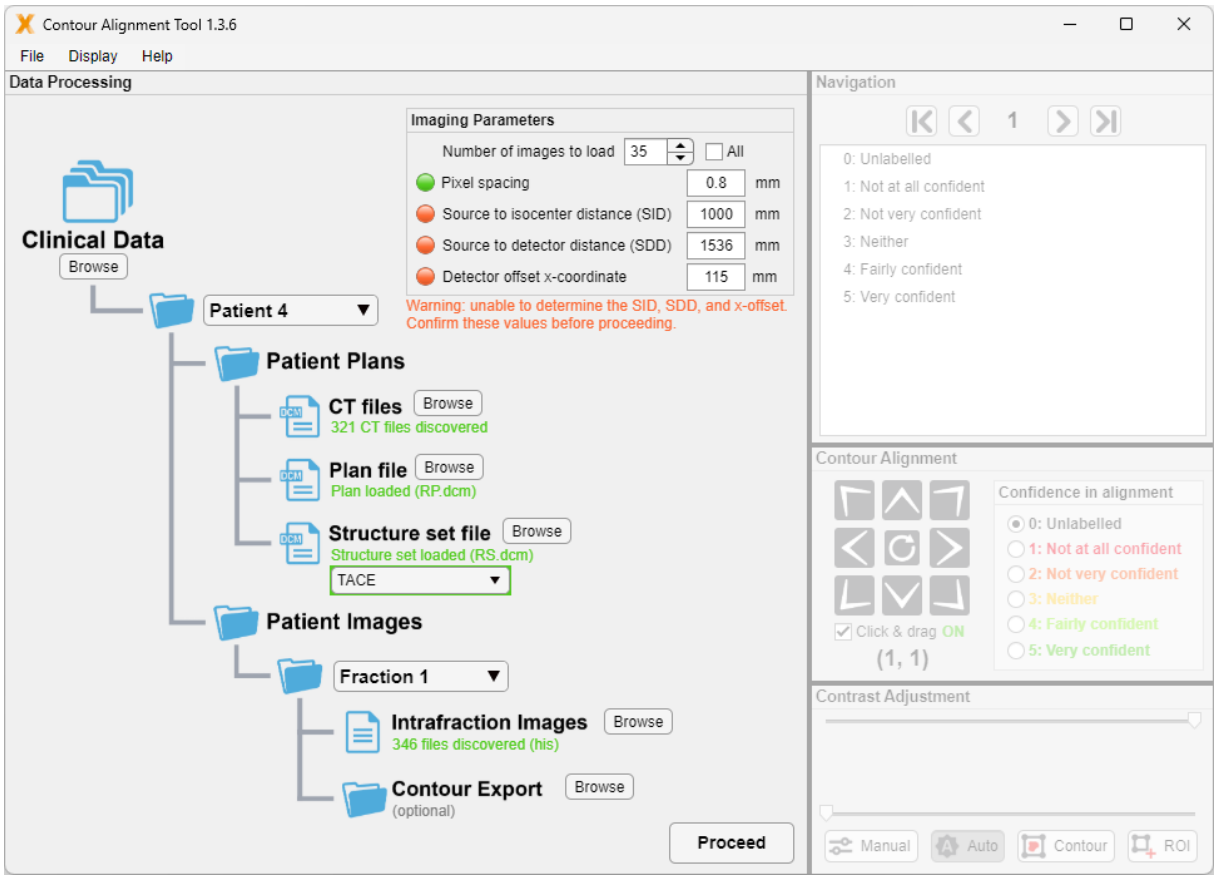


Fig. 6.1 | Treatment data import user interface. Graphical user interface for importing patient planning data and intrafraction images.

The Contour Alignment Tool can import intrafraction images of the specified file formats, including TIFF, XIM, HND, HNC, and HIS from Varian (Palo Alto, CA, USA) and Elekta (Stockholm, Sweden) treatment systems with x-ray imaging. Certain imaging parameters are required for processing the data including pixel spacing, source to isocenter distance, source to detector distance, detector offset, and imager angle. These imaging parameters can be attained from the image headers for some file types. Parameters that cannot be attained from the image headers must be entered by the user prior to proceeding. Additionally, the user can specify the quantity of images to be imported to manage time. The subset of images to be imported will be equispaced between the start and stop angles.

Once the files have been imported, the data are processed for data labelling. As the treatment planning structures are 3D, the structures must be adapted to the 2D space. The planning CT and selected structure are forward projected using the Reconstruction Toolkit[14] at the respective image angles to create digitally reconstructed radiographs (DRR). The forward projections are performed using CUDA Ray Casting for CUDA-enabled graphics processing units (GPU), and the Joseph method[15] if no CUDA-enabled GPU is available.

6.3.2 Data Labelling

Following data processing, the GUI displays the 2D contour on the intrafraction image and the user can rigidly shift the contour to the correct position (Fig. 6.2). The main GUI allows users to navigate between images, adjust image contrast, choose the confidence in alignment using a Likert scale, and export the aligned contours. There are several methods for contrast adjustment including manual, automatic, windowing based on the contour position, and windowing based on a user defined region of interest. Once the user is satisfied with the contour position, they can score their confidence in alignment using a 5-point Likert scale with the following levels: *not at all confident*, *not very confident*, *neither*, *fairly confident*, and *very confident*. The confidence scoring can be used when evaluating an AI model's performance on the images. Once all images have been labelled, the contours can be exported as binary PNG files.

Additional features are available in the menus to assist the user with contouring the images. The DRR Viewer allows the user to visualise the planning CT DRR with the structure contour overlayed. The DRR Viewer can guide the user in positioning the contour when it is difficult to differentiate the structure in the intrafraction images. Users can select different colours for the contour and change the contour display to a transparent fill. Finally, there are keyboard shortcuts for all main actions to improve usability and efficiency.

6.3.3 Development and Deployment

The Contour Alignment Tool was developed in close collaboration with medical physicists and radiation oncologists to ensure it met clinical and research needs. The tool was developed in-house, leveraging data with various projection file types to cover all expected treatment systems. The early development phase allowed for testing of the initial GUI, performance enhancements, and bug fixes. Demonstrations of the tool provided an opportunity for future users to give feedback on desired functionality and modifications.

Once the prototype was functional, it was deployed to one clinical site to label data for the ROCK-RT clinical trial (NCT05169177[16]). The prototype testing phase served to resolve bugs and refine functionality. Based on user feedback, shortcuts and GUI enhancements were implemented to improve usability. The tool was subsequently deployed to an additional two sites leading to further bug fixes and some minor GUI enhancements.

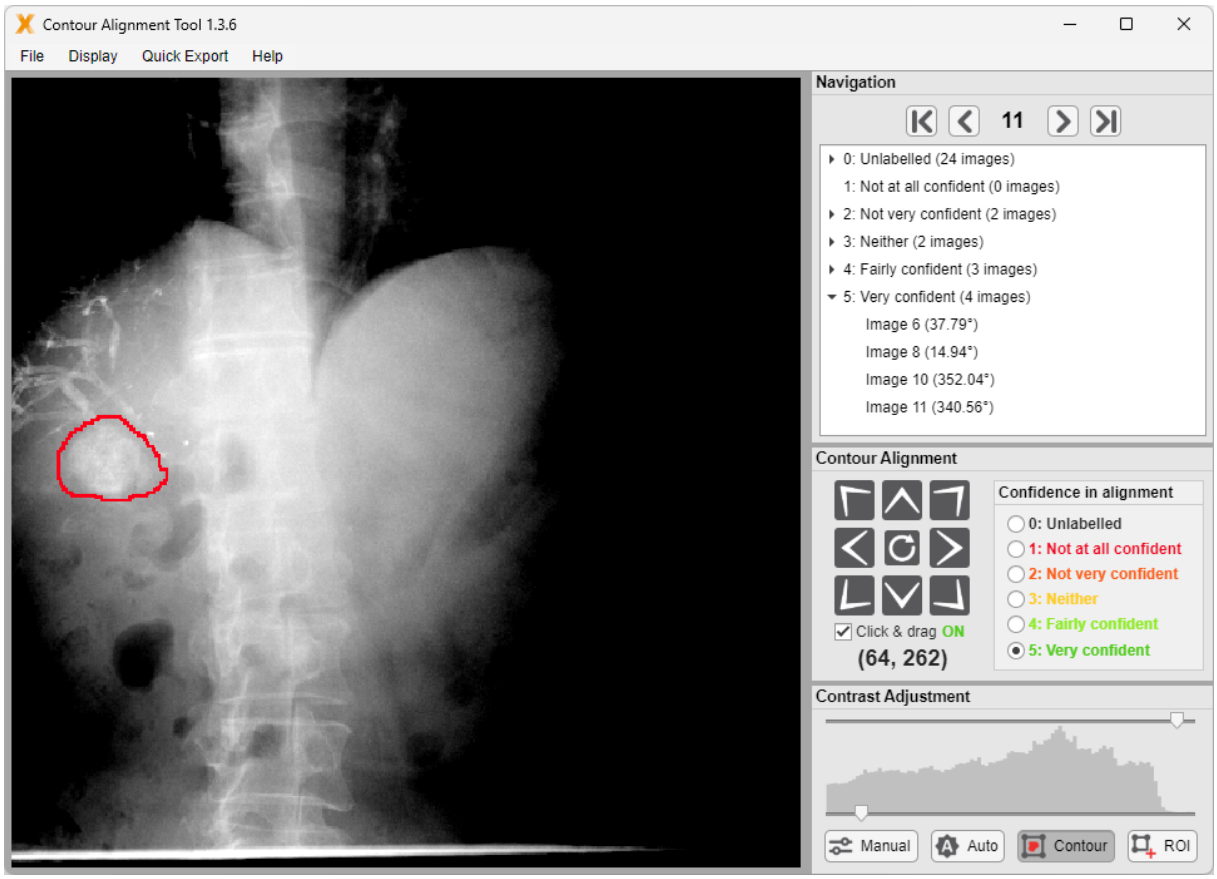


Fig. 6.2 | Data labelling user interface. Data labelling graphical user interface with an example of an abdominal kilovoltage image and liver contrast agent mass contour.

6.4 Results and Discussion

The Contour Alignment Tool is being used to contour x-ray images for two multi-centre imaging studies with a combined target accrual of 350 patients (NCT05169177[16] and NCT05184790[17]). The two studies will cover nine anatomical sites including brain, breast, head and neck, kidney, liver, pancreas, prostate, spine, and heart. The Contour Alignment Tool is being used for several image guidance research projects in addition to the clinical trials. The contour matching of a liver contrast agent mass took an average (± standard deviation) of 12.3 ± 8.7 seconds per image (N = 692 images). While the average time will vary depending on the site being contoured and the image quality, the time is acceptable for research purposes. For the aforementioned clinical trials, the expectation is to label 35 images per fraction giving an anticipated time of less than 10 minutes for the data labelling component.

The Contour Alignment Tool is a convenient, time efficient labelling tool adapted to the requirements for ground truth generation. The users do not require any coding experience to

contour the images. Therefore, the application can be used by students, radiation therapists, medical physicists, and radiation oncologists. The application creates time savings as it leverages treatment planning data to remove the need to manually draw the contour in each image. The Contour Alignment Tool and all supporting code used for development are free to use and are publicly accessible through GitHub. The standalone application can be run using MATLAB Runtime which does not require a MATLAB license. The App Designer code is available for users to contribute to the Contour Alignment Tool or adapt for their own research purposes. Any contributors should be aware of the software license requirements and is responsible for their code.

The application is not without limitations. Currently, there is only the ability to perform a rigid shift of the contour. It does not account for changes in volume or deformation in between the planning CT and intrafraction images. This limitation can be somewhat negated by using an updated contour from the pre-treatment CBCT. Future versions could provide the ability to deform the 2D contour to better match with the intrafraction image. Other developments could include the incorporation of the developed AI models to assist with labelling the structure. The tool could be used to score the AI performance in labelling the structure. Furthermore, the Contour Alignment Tool offers the ability to load a large range of 2D intrafraction image file types, serving a wide range of usage scenarios. The application could be adapted for other x-ray systems including CyberKnife and Radixact (Accuray Incorporated, Sunnyvale, CA, USA), as well as MRI systems. The application also has potential for other medical imaging applications outside of radiation therapy.

In conclusion, this open-source Contour Alignment Tool allows users to efficiently contour intrafraction images and can be used to create large, expertly labelled datasets. These datasets will be required to develop real time x-ray IGRT solutions which in turn can be used to improve beam–tumour targeting and ultimately patient outcomes.

6.5 References

1. Solberg, T. D. et al. Quality and safety considerations in stereotactic radiosurgery and stereotactic body radiation therapy: Executive summary. *Practical radiation oncology* **2**, 2–9 (2012).

2. Keall, P. J. et al. Review of real-time 3-dimensional image guided radiation therapy on

standard-equipped cancer radiation therapy systems: are we at the tipping point for the era of real-time radiation therapy? *International Journal of Radiation Oncology* Biology* Physics* **102**, 922–931 (2018).

3. Bertholet, J. et al. Real-time intrafraction motion monitoring in external beam radiotherapy. *Physics in medicine & biology* **64**, 15TR01 (2019).

4. Anastasi, G. et al. Patterns of practice for adaptive and real-time radiation therapy (POP-ART RT) part I: Intra-fraction breathing motion management. *Radiotherapy and Oncology* **153**, 79–87 (2020).

5. Mylonas, A., Booth, J. & Nguyen, D. T. A review of artificial intelligence applications for motion tracking in radiotherapy. *Journal of Medical Imaging and Radiation Oncology* **65**, 596–611 (2021).

6. Terunuma, T., Tokui, A. & Sakae, T. Novel real-time tumor-contouring method using deep learning to prevent mistracking in X-ray fluoroscopy. *Radiological physics and technology* **11**, 43–53 (2018).

7. Hirai, R., Sakata, Y., Tanizawa, A. & Mori, S. Real-time tumor tracking using fluoroscopic imaging with deep neural network analysis. *Physica Medica* **59**, 22–29 (2019).

8. Takahashi, W., Oshikawa, S. & Mori, S. Real-time markerless tumour tracking with patient-specific deep learning using a personalised data generation strategy: proof of concept by phantom study. *The British journal of radiology* **93**, 20190420 (2020).

9. Edmunds, D., Sharp, G. & Winey, B. Automatic diaphragm segmentation for real-time lung tumor tracking on cone-beam CT projections: a convolutional neural network approach. *Biomedical physics & engineering express* **5**, 035005 (2019).

10. Zhao, W. et al. Markerless pancreatic tumor target localization enabled by deep learning. *International Journal of Radiation Oncology* Biology* Physics* **105**, 432–439 (2019).

11. Zhou, D., Nakamura, M., Mukumoto, N., Yoshimura, M. & Mizowaki, T. Development of a deep learning-based patient-specific target contour prediction model for markerless tumor positioning. *Medical physics* **49**, 1382–1390 (2022).

12. Zhao, W. et al. Incorporating imaging information from deep neural network layers into image guided radiation therapy (IGRT). *Radiotherapy and Oncology* **140**, 167–174 (2019).

13. Gardner, M. et al. Realistic CT data augmentation for accurate deep-learning based segmentation of head and neck tumors in kV images acquired during radiation therapy. *Medical Physics* **50**, 4206–4219 (2023).

14. Rit, S. et al. The Reconstruction Toolkit (RTK), an open-source cone-beam CT reconstruction toolkit based on the Insight Toolkit (ITK). In *Journal of Physics: Conference Series*, vol. 489, 012079 (IOP Publishing, 2014).

15. Joseph, P. M. An improved algorithm for reprojecting rays through pixel images. *IEEE transactions on medical imaging* **1**, 192–196 (1982).

16. Plant, N. et al. Radio-opaque contrast agents for liver cancer targeting with KIM during radiation therapy (ROCK-RT): an observational feasibility study. *Radiation Oncology* **19**, 139 (2024).

17. University of Sydney. LEARN: Learning Environment for Artificial Intelligence in Radiotherapy New Technology (LEARN). `https://clinicaltrials.gov/study/NCT05184790` (2024).

7

Conclusions and Future Work

7.1 Summary

Increasing evidence points towards improved radiation therapy treatment outcomes and reduced radiation toxicity through dose escalation combined with reduced margins. However, the delivery of targeted radiation therapy remains a challenge due to intrafraction motion. Recent developments in the field such as image-guided adaptive radiation therapy have allowed for the precise targeting of the tumour and sparing of OARs. Real-time motion monitoring on a standard linac is difficult for organs with poor soft tissue contrast such as the prostate, pancreas, and liver. While fiducial markers can be implanted to visualise these targets, the implantation procedure is invasive and carries risks. Furthermore, there is a need world-wide for motion management strategies that are not costly and can be implemented into existing workflows.

AI has shown great potential for markerless tumour tracking in x-ray images.[1] However, existing studies for the prostate and pancreas are limited to evaluation on synthetic imaging data. Therefore, the overarching objective of this thesis was to develop the first real-time markerless tumour tracking framework for prostate, pancreatic, and liver radiation therapy. The specific aims of this thesis were:

(i) To conduct the first large-scale proof of principle for x-ray-based markerless prostate segmentation, suitable for clinical implementation on standard linacs.

(ii) To explore the potential of x-ray-based markerless pancreas head segmentation, focusing on accuracy and identifying technical challenges.

(iii) To investigate the feasibility of using residual contrast agents from chemotherapy for real-time markerless liver cancer tracking.

7.2 Conclusions and Contribution

This thesis presents a markerless deep learning framework for real-time motion management on a standard linac, eliminating the need for fiducial markers. The studies are an important step towards achieving real-time tracking on standard treatment machines without additional hardware, leading to more efficient radiation therapy for more patients. The main body of this thesis involved three studies focused on different cancer sites: prostate, pancreas, and liver. These three sites were chosen due to the challenges in visualising the target in intrafraction images. Markers are commonly used for these sites to act as a surrogate and improve tracking accuracy.

First, an algorithm was developed to track the prostate in a large, multi-institutional retrospective study in Chapter 3. The study addresses Aim 1 by conducting the first large-scale proof-of-principle of x-ray-based prostate markerless segmentation for standard linacs, advancing the field closer to a clinical implementation of this technology. The deep learning model was validated on patients with and without fiducial markers, demonstrating accuracy within several millimetres and no correlation between the observed motion and error, indicating its ability to detect cases of large motion.

Building on this, the framework was extended to pancreas head segmentation in Chapter 4 to address Aim 2. The added complexity of larger motion and deformation necessitated deformable data augmentation and angle-specific networks. Using these adaptations, the first algorithm for liver chemotherapy contrast agent segmentation during radiation therapy was developed. Aim 3 was achieved by demonstrating that the residual contrast agent could enable real-time markerless tracking for a cohort of patients. Finally, the open-source Contour Alignment Tool, crucial in ground truth generation for both the pancreas and liver studies, was introduced. This tool will also support additional clinical trials in the coming years.

7.3 Future Work

This thesis has demonstrated the ability of the deep learning framework for prostate, pancreas, and liver motion management. It is expected that other sites such as the lung, where there is higher tumour contrast, would similarly benefit from this deep learning framework. The framework has already been successfully applied to head and neck cancer using simulated 2D x-ray images (Appendix B). Future work will involve retrospective studies on other cancer sites through the Learning Environment for Artificial Intelligence in Radiotherapy New Technology (LEARN; NCT05184790[2]) clinical trial. LEARN aims to develop a whole-of-body markerless tracking method including brain, breast, head and neck, kidney, liver, pancreas, prostate, spine and cardiac anatomic sites. The deep learning framework presented in this thesis will be evaluated on 300 participants, with at least 30 individuals from each of the nine cohorts. The methods of this thesis will be crucial in the implementation of the model in the LEARN clinical trial.

While the deep learning framework has demonstrated promising results, further development is needed to improve its accuracy and robustness in clinical settings. As part of this, investigating explainable AI techniques such as activation maps will be important to better understand which image features the algorithm relies on for tracking. These analyses can guide refinements to the model and training process, as well as help assess whether the method relies on any residual information from the masking applied in Chapters 3 and 4. Additionally, in this thesis, the framework was tested on x-ray images acquired during patient setup, which do not contain MV scatter. Future evaluation is required to assess performance in scenarios involving MV scatter. However, the framework may be applied effectively in environments where MV scatter is mitigated through methods such as triggered imaging[3] and frame averaging[4].

The average inference time of the model was 10 ms per image, making it suitable for real-time motion monitoring. According to the AAPM Task Group 264, real-time systems are defined as having a total latency below 500 ms.[5] While the model meets this requirement at the segmentation level, further optimisation may include streamlining the data pre-processing pipeline, parallelising computations, and deploying the model on high-performance and dedicated hardware. The latency could be further reduced through model optimisations. Optimisation efforts will be focused on the generator, as the final application only uses the generator network for segmentation. These include model pruning and quantisation to reduce compu-

tational load,[6] as well as exploring lightweight generator architectures.

To enable patient-specific training, the deep learning framework must first be integrated with treatment planning systems so that it can be trained using each patient's planning data prior to treatment. However, training solely on the planning computed tomography (CT) data may not capture the full range of interfraction anatomical variation observed throughout a treatment course. Since daily cone-beam CTs (CBCTs) are not available at the time of the first fraction, the current approach uses data augmentation to simulate such variation. Therefore, daily CBCTs could be incorporated progressively by fine-tuning the model as treatment progresses. This adaptive approach could help the model learn from patient-specific anatomical changes over time and improve its robustness.

However, it is important to acknowledge the challenges associated with training generative adversarial networks (GAN), including non-convergence, mode collapse, and instability.[7] These issues can introduce uncertainty into the segmentation, potentially affecting its reliability in a clinical setting. Across the three studies, the model performance was found to be sensitive to the number of training epochs. Future work could focus on optimising the loss function to improve training stability and mitigate issues such as mode collapse. In addition, incorporating in-training validation strategies may help identify the optimal training duration, ultimately enhancing the robustness and clinical applicability of the system.

A GAN-based segmentation model was chosen for target segmentation due to its ability to model complex structures and generate realistic outputs. The training data consisted of digitally reconstructed radiographs (DRRs) of the entire anatomy, paired with ground truth DRRs showing only the target. The generator network was trained to produce realistic target-only DRRs that represent the segmentations. This approach is beneficial for soft tissue segmentation in x-ray images, where anatomical boundaries can be subtle or poorly defined. Compared to traditional CNN-based segmentation methods, GANs can produce sharper and more anatomically plausible results. Preliminary studies supported this, demonstrating that GAN-based segmentations appeared more anatomically realistic than those generated by a conventional UNet. However, given the rapidly evolving field of deep learning, alternative architectures such as transformer-based models should also be explored in future work to further improve performance and robustness.

The next step towards clinical implementation will involve incorporating the deep learning model into a real-time system that is integrated with clinical treatment systems. This

process is already underway, using an existing, clinically validated software solution for fiducial marker tracking known as Kilovoltage Intrafraction Monitoring (KIM). KIM has been used in multi-centre clinical trials for the prostate[8] and liver[9]. Once integrated, the deep learning real-time framework will be evaluated on a standard linac using phantom studies to assess its performance on real patient motion traces and extreme motion scenarios. Following the integration of the framework and successful phantom studies, an observational clinical trial will be conducted to compare delivered treatments without motion tracking to hypothetical treatments with the deep learning motion tracking.

7.4 Concluding Remarks

The work in this thesis has advanced the field closer to the clinical implementation of markerless tumour tracking on a standard linac. It has investigated the feasibility of the framework for three challenging cancer sites with poor soft tissue contrast, laying the foundation for a whole-of-body markerless tracking system. The deep learning framework has the potential to make real-time motion management treatments accessible to all patients, eliminating the need for expensive, dedicated equipment or fiducial marker implantation.

7.5 References

1. Mylonas, A., Booth, J. & Nguyen, D. T. A review of artificial intelligence applications for motion tracking in radiotherapy. *Journal of Medical Imaging and Radiation Oncology* **65**, 596–611 (2021).

2. University of Sydney. LEARN: Learning Environment for Artificial Intelligence in Radiotherapy New Technology (LEARN). `https://clinicaltrials.gov/study/NCT05184790` (2024).

3. Kaur, G., Lehmann, J., Greer, P. & Simpson, J. Assessment of the accuracy of truebeam intrafraction motion review (IMR) system for prostate treatment guidance. *Australasian Physical & Engineering Sciences in Medicine* **42**, 585–598 (2019).

4. Ng, J. A. et al. Kilovoltage intrafraction monitoring for prostate intensity modulated arc

therapy: first clinical results. *International Journal of Radiation Oncology* Biology* Physics* **84**, e655–e661 (2012).

5. Keall, P. J. et al. AAPM Task Group 264: The safe clinical implementation of MLC tracking in radiotherapy. *Medical physics* **48**, e44–e64 (2021).

6. Liang, T., Glossner, J., Wang, L., Shi, S. & Zhang, X. Pruning and quantization for deep neural network acceleration: A survey. *Neurocomputing* **461**, 370–403 (2021).

7. Saxena, D. & Cao, J. Generative adversarial networks (GANs) challenges, solutions, and future directions. *ACM Computing Surveys (CSUR)* **54**, 1–42 (2021).

8. Keall, P. et al. Real-time image guided ablative prostate cancer radiation therapy: results from the TROG 15.01 SPARK trial. *International Journal of Radiation Oncology* Biology* Physics* **107**, 530–538 (2020).

9. Sengupta, C. et al. The first clinical implementation of real-time 6 degree-of-freedom image-guided radiotherapy for liver SABR patients. *Radiotherapy and Oncology* **190**, 110031 (2024).

A

Appendix A: Radio-Opaque Contrast Agents for Liver Cancer Targeting with KIM During Radiation Therapy (ROCK-RT)

This appendix outlines the protocol of the ROCK-RT observational feasibility study which was the basis for Chapter 5. ROCK-RT aims to investigate the feasibility of using residual radiopaque contrast agents from chemotherapy for real-time image-guided radiation therapy. The study has received ethics approval and is registered to ClinicalTrials.gov (NCT05169177). This appendix includes initial results for three patients generated as part of this thesis using the deep learning framework. The appendix has been published as "Plant, N., **Mylonas, A.**, Sengupta, C., Nguyen, D. T., Silvester, S., Pryor, D., Greer, P., Lee, Y. Y. D., Ramachandran, P., Seshadri, V., Trada, Y., Khor, R., Wang, T., Hardcastle, N., & Keall, P. Radio-opaque contrast agents for liver cancer targeting with KIM during radiation therapy (ROCK-RT): An observational feasibility study. *Radiation Oncology*, **19**, 139 (2024)."

Plant *et al. Radiation Oncology* (2024) 19:139
https://doi.org/10.1186/s13014-024-02524-4

Radiation Oncology

STUDY PROTOCOL Open Access

Radio-opaque contrast agents for liver cancer targeting with KIM during radiation therapy (ROCK-RT): an observational feasibility study

Natalie Plant[1], Adam Mylonas[1], Chandrima Sengupta[1], Doan Trang Nguyen[1], Shona Silvester[1], David Pryor[2], Peter Greer[3], Yoo Young (Dominique) Lee[2], Prabhakar Ramachandran[2], Venkatakrishnan Seshadri[2], Yuvnik Trada[3], Richard Khor[4], Tim Wang[5], Nicholas Hardcastle[6] and Paul Keall[1*]

Abstract

Background This observational study aims to establish the feasibility of using x-ray images of radio-opaque chemoembolisation deposits in patients as a method for real-time image-guided radiation therapy of hepatocellular carcinoma.

Methods This study will recruit 50 hepatocellular carcinoma patients who have had or will have stereotactic ablative radiation therapy and have had transarterial chemoembolisation with a radio-opaque agent. X-ray and computed tomography images of the patients will be analysed retrospectively. Additionally, a deep learning method for real-time motion tracking will be developed. We hypothesise that: (i) deep learning software can be developed that will successfully track the contrast agent mass on two thirds of cone beam computed tomography (CBCT) projection and intra-treatment images (ii), the mean and standard deviation (mm) difference in the location of the mass between ground truth and deep learning detection are ≤ 2 mm and ≤ 3 mm respectively and (iii) statistical modelling of study data will predict tracking success in 85% of trial participants.

Discussion Developing a real-time tracking method will enable increased targeting accuracy, without the need for additional invasive procedures to implant fiducial markers.

Trial registration Registered to ClinicalTrials.gov (NCT05169177) 12th October 2021.

Keywords Neoplasms, Radiotherapy, Kilovoltage Intrafraction Monitoring, Image Guidance, Stereotactic ablative Radiation Therapy (SABR), Liver, Radio-opaque contrast, Transarterial chemoemobilisation (TACE)

*Correspondence:
Paul Keall
paul.keall@sydney.edu.au
[1]Image X Institute, University of Sydney, Suite 201, Biomedical Building (C81), 1 Central Ave, Eveleigh, NSW 2015, Australia
[2]Department of Radiation Oncology, Princess Alexandra Hospital, Brisbane, QLD, Australia
[3]Department of Radiation Oncology, Calvary Mater Newcastle, Newcastle, NSW, Australia
[4]Olivia Newton-John Cancer Wellness & Research Centre, Austin Health, Melbourne, VIC, Australia
[5]Department of Radiation Oncology, Crown Princess Mary Cancer Centre, Sydney, NSW, Australia
[6]Peter MacCallum Cancer Centre, Melbourne, VIC, Australia

Background

Liver cancer is a global health concern with increasing incidence worldwide [1–3]. It is the sixth most common cancer globally and the third leading cause of cancer-related deaths [1]. Hepatocellular carcinoma (HCC) is the most common form of primary liver cancer accounting for ~85% of cases [1]. In Australia, the incidence and mortality rate of liver cancer has increased more than any other cancer [4].

Stereotactic Ablative Radiation Therapy (SABR) is a technique used to deliver high-precision, ablative doses of radiation in a small number of fractions to an extracranial target [5]. It is utilised to treat a variety of malignancies including in the lung [6], liver [7] and spine [8].

One of the challenges of radiation therapy (RT) is the significant movement of tumours that can occur during treatment, particularly for tumours in the liver due to respiratory motion [9]. SABR is often utilised as an alternative to surgery or thermal ablation in patients with limited functional liver reserve or where tumours lie close to bowel, pericardium, or central biliary structures. If respiratory motion is not accounted for, there can be increased dose to surrounding structures and subsequent injury. Therefore, the delivery of a precise ablative dose is highly dependent on being able to verify the tumour position and any associated movement [10]. Real-time image-guided radiation therapy (IGRT) is a targeting method used to track the movement of tumour tissue during radiation therapy by tracking the motion of fiducial markers implanted in or near the tumour tissue [11]. One of these technologies, Kilovoltage Intrafraction Monitoring (KIM), has recently been applied to liver SABR [12]. The markers are implanted by an invasive procedure which can be associated with a risk of bleeding, infection, marker migration, additional treatment delays and may require a general anaesthetic [13, 14]. The current study seeks to develop and investigate a deep learning method that maintains the real-time tracking benefits of KIM, without the need to implant markers into patients.

In HCC patients, chemotherapy is often delivered via transarterial chemoembolisation (TACE), which involves the injection of chemotherapy drugs into the artery supplying the tumour combined with embolic particles to restrict blood supply and retain the therapeutic agent within the treatment zone, and a radio-opaque contrast agent. Radio-opaque contrast material within the tumour allows direct visualisation of the treated tumour at the time of TACE and is often retained in the tumour and visible on imaging for many months afterwards. The approach of treating HCC using TACE immediately before SABR is increasingly used, and there is the potential to utilise this retained radio-opaque contrast as a marker for non-invasive, real-time tracking of tumour motion [15–17].

The most common radio-opaque agent used in this setting is ethiodised oil, however, there are also some drug-eluting bead (DEB-TACE) formulations that incorporate iodine directly into the bead structure. These retained radio-opaque contrast agents are currently often used as a surrogate tumour marker to aid target verification before and during SABR.

This observational study aims to establish the feasibility of using residual radio-opaque contrast agent in patient images as a real-time guidance method for IGRT treatment of HCC by applying deep-learning to x-ray images obtained as standard of care during radiation therapy. Completion of this observational study will lead to a prospective use of real-time markerless KIM real-time IGRT for eligible liver SABR patients.

Methods/design

Aim

This study aims to provide a non-invasive alternative to implanting fiducial markers to track tumour movement in real-time during SABR treatment of HCC. To determine whether radio-opaque contrast agents in the radiological images used to guide HCC radiotherapy can be used for real-time tracking of tumour movement, this study will train a deep-learning model to segment residual radio-opaque agents in radiation therapy planning images then attempt to accurately detect the agents in images obtained during treatment, for use with motion management software (KIM). The deep learning real-time tracking process is shown in Fig. 1.

Design

This observational study will recruit patients who had or will have SABR treatment for HCC following TACE chemotherapy. Standard of care radiotherapy planning and in-treatment x-ray images will be collected, and analysis will occur offline. Following deep-learning and KIM algorithm adaptation, the detected location of the contrast agent mass by KIM will be compared with manual delineation (Fig. 2).

Hypotheses

We hypothesise that: (i) deep learning software can be developed that will successfully track the contrast agent mass on two thirds of cone beam computed tomography (CBCT) projection and intra-treatment images (ii), the mean and standard deviation (mm) difference in the location of the mass between ground truth and deep learning detection are ≤2 mm and ≤3 mm respectively and (iii) statistical modelling of study data will predict tracking success in 85% of trial participants.

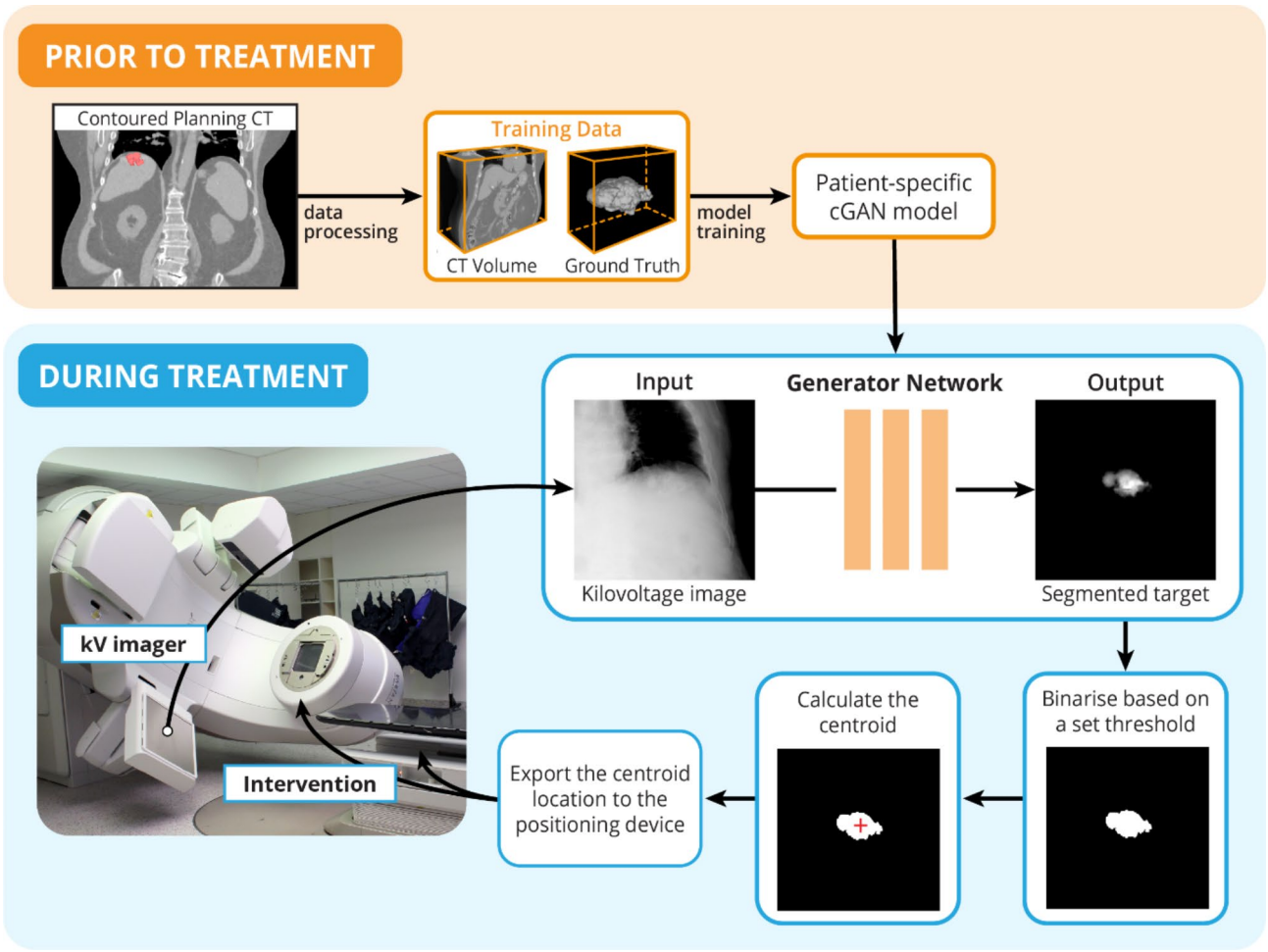


Fig. 1 The clinical workflow for automatic target tracking using residual contrast agent is comprised of two key components: prior to treatment and during treatment. A patient-specific network is trained prior to the patient's treatment using radiation therapy planning data. The generator network from the conditional generative adversarial network (cGAN) is used during the treatment to segment the target. The location of the segmented target can be used for motion management

Eligibility criteria

This study will recruit 50 participants who: will be ≥18 years of age; had or will have SABR for HCC; will/have residual radio-opaque contrast (e.g., ethiodised oil or drug-eluting beads containing iodine) from prior TACE chemotherapy visible within the imaging field on RT planning CT scans; will provide written informed consent or meet criteria for a waiver of consent; and will have the minimum image dataset available in the required format.

Participants

Participants will be recruited from four Australian sites that currently use SABR to treat primary liver cancer; Princess Alexandra Hospital in Queensland, Calvary Mater Newcastle Hospital and the Crown Princess Mary Cancer Centre in New South Wales, and the Austin Health in Victoria. The Image-X Institute at the University of Sydney will develop the KIM algorithm and software for ground truth delineation and provide central study coordination.

Consent/recruitment

Participants who had their radiation therapy prior to site activation will be recruited retrospectively and a waiver of the need for consent will be sought from an approved HREC (Human Research Ethics Committee) and governing State Health Data Custodian. Participants are otherwise recruited prospectively using a HREC-approved patient information sheet and consent form prior to starting RT.

Datasets

The minimum RT treatment dataset required from retrospectively recruited participants includes the 3D or 4D CT scans used for planning, contours, treatment plans and pre-treatment 3D or 4D CBCT reconstructed images in DICOM format, and 2D projection images.

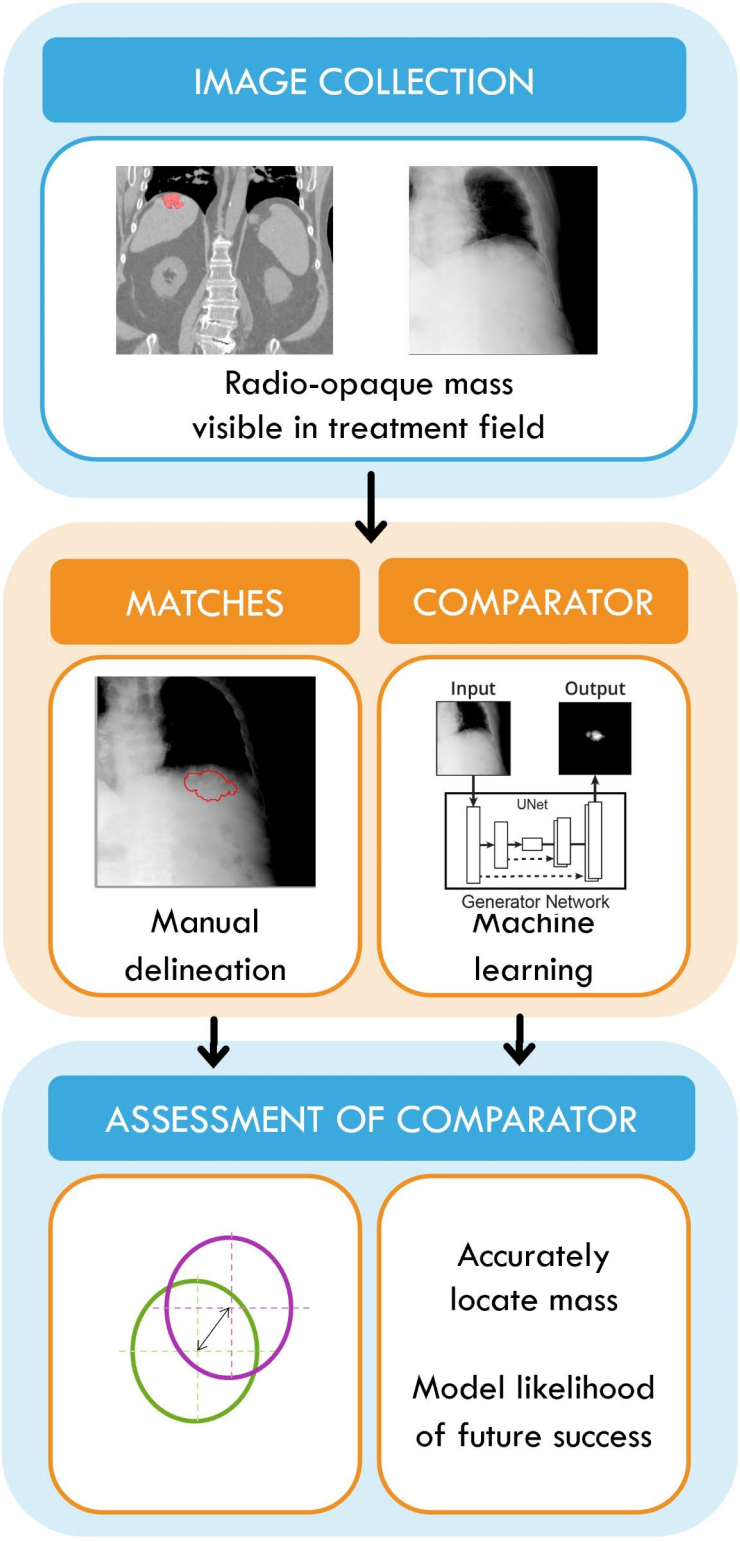


Fig. 2 Study Design. Standardly collected radiation therapy planning, CBCT projections and intra-treatment monitoring images will be used to created manually delineated matches. After improvement of existing mass-detecting software (KIM) through machine learning, the updated algorithm will be applied to un-delineated images to (i) detect the contrast mass, and (ii) locate its centre in comparison with manually delineated matches. If the contrast mass can be detected accurately, modelling for likelihood of KIM tracking success will be conducted

In-treatment x-ray images are mandatory for prospectively recruited participants and desirable from those retrospectively recruited and any additional images from screening or testing sessions are desirable but not mandatory from all participants.

In addition to RT treatment images, data will be collected on participant characteristics (demographics, medical history, diagnosis), TACE chemotherapy (type of contrast agent - ethiodised oil or drug eluting beads, date of procedure), and other RT treatment-related data (treatment centre and treating doctor, treatment device and version, imaging system type/model, motion management techniques such as free-breathing or breath hold, the use of abdominal compression, and breathing training).

Matching

Manual delineation of the contrast mass on the planning CT and pre-treatment CBCT images (3D or 4D) with a purpose-designed alignment tool using MATLAB Runtime R2021a (version 9.10) (Fig. 3) will provide a ground truth location of the centre of the contrast mass to which delineation by the KIM software can be compared. If more than one contrast mass is visible on the planning CT and pre-treatment CBCT images, these will be separately delineated, and then the two volumes will be combined into a single contrast mass structure. To ensure at least 100 labelled images will be available for each participant, the contour alignment tool will choose 35 projections per CBCT from each fraction that are equally spaced angularly over the scanning arc to represent a range of angles. Where the contrast agent mass cannot be equivocally identified manually, these images will not be used as the ground truth. Users of the alignment tool will give a confidence score for the alignment using a five-level Likert scale ranging from 'Not at all confident' to 'Very confident'.

Comparator

The adapted KIM software will be run on un-delineated copies of the acquired data used to determine the ground truth location of the contrast agent. The KIM program will output the 2D and 3D position of the centre of the contrast agent mass. The KIM algorithm may be adjusted

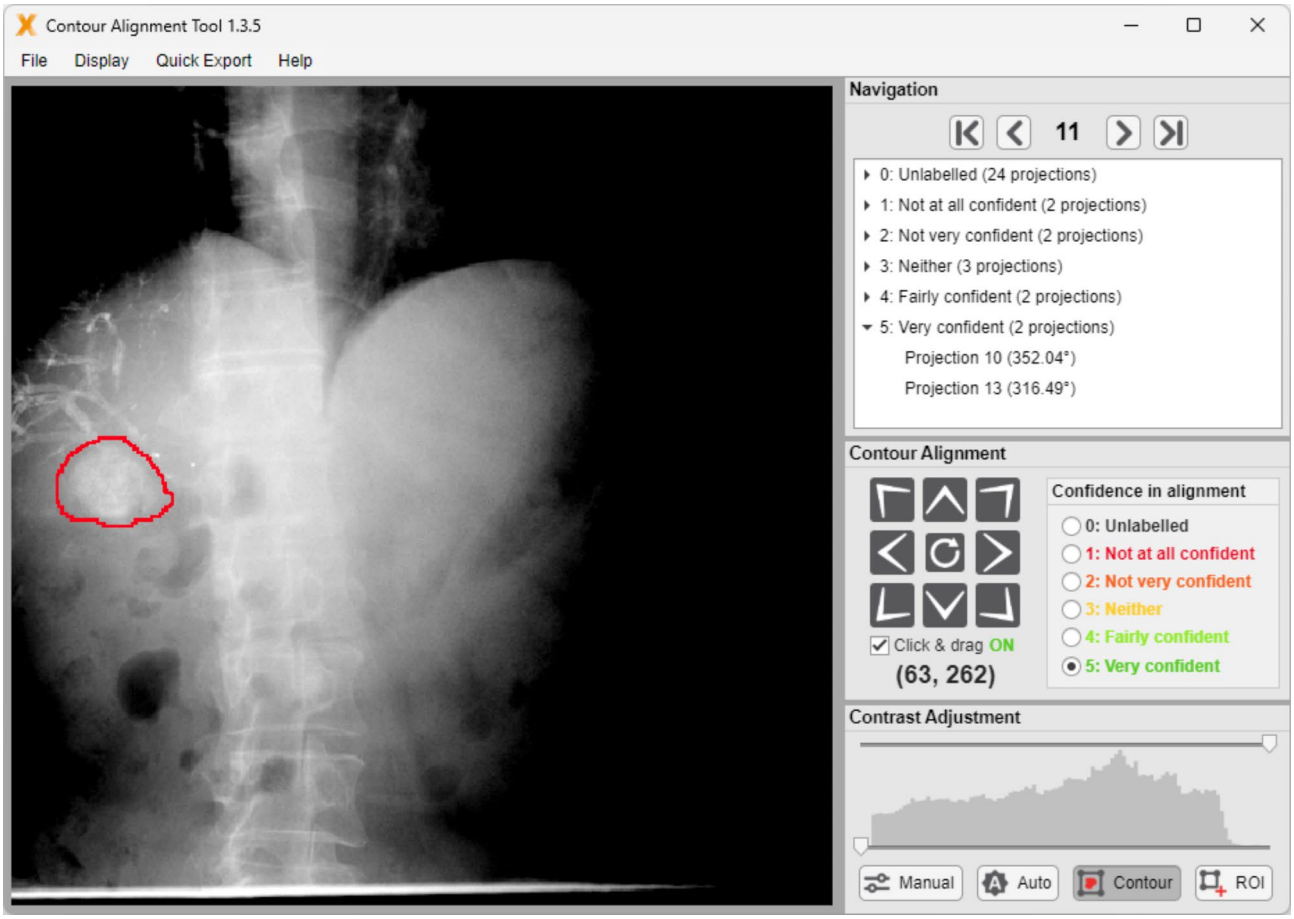


Fig. 3 The contour alignment tool graphical user interface with an unaligned contour. The red contour (liver contrast agent mass) can be repositioned by the user

as more patient data is acquired, and the same version of the software will be run across all patients for the analysis.

Data collection and transfer

An anonymisation tool will be applied to all images and datasets will be coded with participants' unique Study ID before transfer. Image data will be uploaded to a data-sharing platform created and maintained by the University of Sydney for this study via a site-specific unique link to study folders. To quantity the interobserver error, the image data will be independently manually delineated by a separate study site.

All other data will be obtained from the participant's medical record by delegated study personnel at the recruiting site and entered in a password-protected online database (REDCap 12.5.8, Vanderbilt University), coded by Study ID and year of birth.

Bias

The study design minimises bias/errors due to sampling (baseline characteristics), matching (accuracy of comparator delineation) and assessments. Baseline characteristics to be considered in final analyses include (i) type, size, shape, density, and location of the contrast agent, (ii) type of imaging and treatment machine, (iii) time since TACE, (iv) participant demographics, and (v) treatment site. Accuracy of the ground truth (manual contrast mass delineation on images) will be maximised by (i) providing sites with the same purpose-designed software to conduct delineation of contrast mass location on images, (ii) for the same projections, performing ground truth delineation by an independent observer who is a member of the research team from another site, and (iii) assessment by qualified and experienced medical physicists or radiation therapists. Final assessment of the KIM software (comparator) will be conducted (i) on images that have not been marked with the ground truth, (ii) by site study personnel who have not seen images marked with ground truth, (iii) by study personnel who are qualified and experienced medical physicists.

Outcome measures

(i) the proportion of CBCT projection and intra-treatment images in which the KIM software detects a contrast mass. (ii) the mean and standard deviation of the difference (mm) of location of the centre of the contrast mass detected on CBCT projection and intra-fraction images by KIM software compared with the ground truth in each of the horizontal and vertical directions. (iii) the mean and standard deviation of the centroid error between the segmentation and ground truth will be calculated, and DICE analysis will be performed to measure the similarity between the two delineation methods. Characteristics of the participants, chemotherapy, RT, and images, (e.g., treatment machine type, treatment site; contrast agent type, density, size, shape, and location; and patient size, age, cancer stage and sex), will be used to create a generalised linear model, or appropriate alternative, to identity univariate and multivariate patient or treatment features that contributed to the success or failure of the KIM tracking algorithm.

Preliminary data

To investigate the feasibility of the ROCK-RT protocol, data from three patients recruited into ROCK-RT have been analysed. A conditional generative adversarial network [18] was used to train a patient-specific model to track the contrast mass from the pre-treatment data including data augmentation (translation and rotation). This model was then applied to the data acquired during treatment, representing the clinical scenario of real-time target tracking. Figure 4 shows examples of the target tracking prediction compared with the ground truth. Figure 5 quantifies the centroid error and Dice similarity coefficient acquired from two fractions for three patients using 35 images per fraction. The mean and standard deviation of the centroid tracking error in the anterior-posterior/lateral and super-inferior directions were -0.4 ± 2.8 mm and -0.6 ± 1.4 mm respectively. The mean Dice similarity coefficient was 0.87 ± 0.08.

Discussion

The detection of residual radio-opaque contrast in the liver by KIM may lead to the development of technology that improves the accuracy, safety, and effectiveness of RT for liver cancer. The main challenges to this study are the availability of participants whose minimum image dataset is available, especially for retrospectively recruited participants, and the time commitment required by sites to contour images.

The hypotheses will be accepted if, in two thirds of participants (i), KIM detects the contrast mass; (ii) the mean and standard deviation (mm) difference in the location of the mass between ground truth and KIM detection are ≤ 2 mm and ≤ 3 mm, respectively and (iii) if a model of the success rate in the primary hypothesis and patient/treatment characteristics can predict KIM success in 85% of participants. Should the three hypotheses be accepted, there will be sufficient justification for the prospective use of KIM as a real-time image guided radiation therapy tool for liver cancer patients.

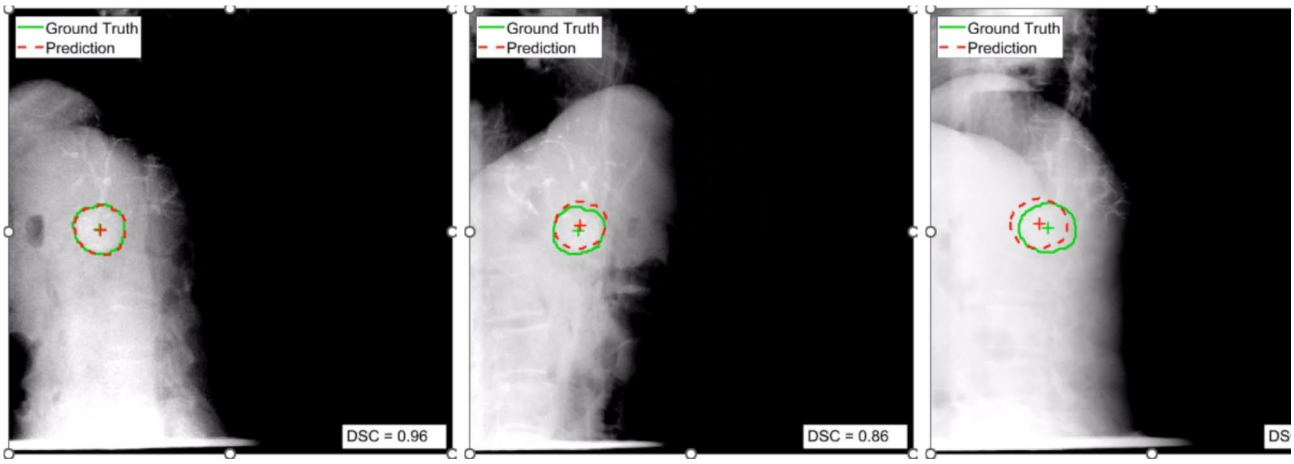


Fig. 4 Examples of the target tracking prediction compared with the ground truth on kV images from different imaging angles from one fraction of one patient

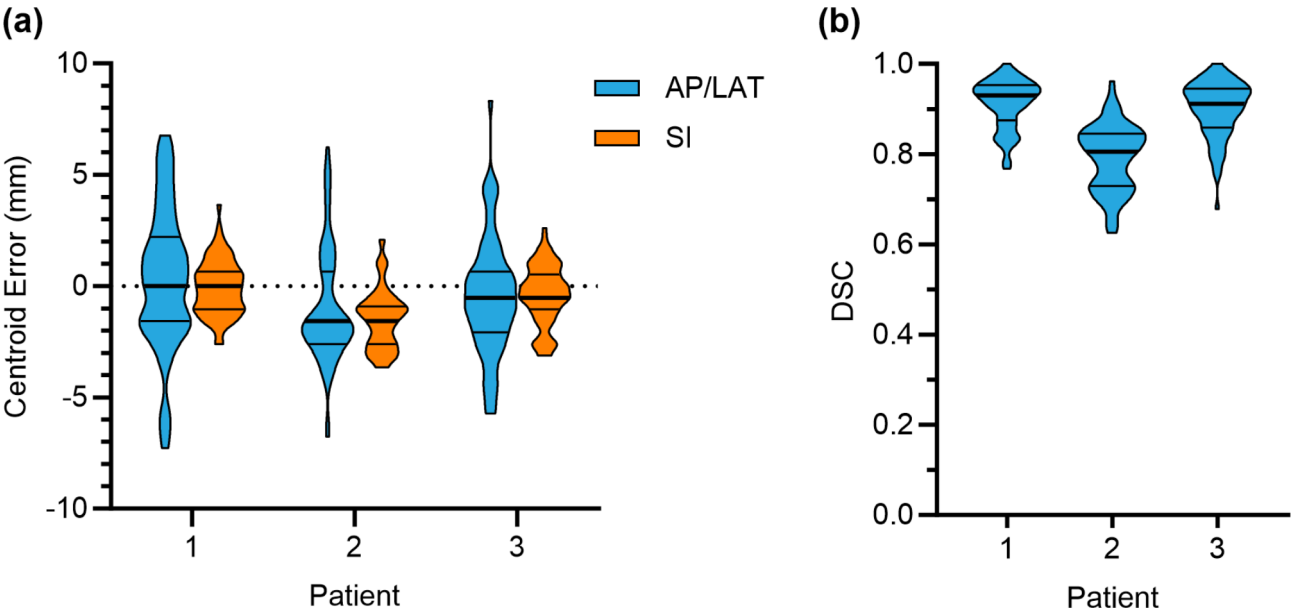


Fig. 5 The centroid error and Dice similarity coefficient acquired from two fractions for three patients. AP = Anterior-Posterior; LAT = Lateral; SI = Superior-Inferior; DSC = Dice similarity coefficient

Abbreviations

2D/3D/4D	Two-dimensional/three-dimensional/four-dimensional
CBCT	Cone beam computed tomography
DICE	Dice Similarity Coefficient
DICOM	Digital imaging and communications in medicine
HCC	Hepatocellular carcinoma
HREC	Human research ethics committee
IGRT	Image guided radiation therapy
KIM	Kilovoltage intrafraction monitoring
RT	Radiation therapy
SABR	Stereotactic ablative radiotherapy
TACE/DEB-TACE	Transarterial chemoembolisation/Drug-eluting bead TACE

Acknowledgements
The authors acknowledge the assistance of Julia Johnson with the figures and Dr Hilary Byrne and Dr Caterina Brighi for reviewing the draft manuscript.

Author contributions
N.P. wrote the main manuscript text and also, with C.S. and A.M. developed the data acquisition process. A.M. led the research for the preliminary results. All authors contributed to the study protocol development and reviewed the manuscript.

Funding
Research activities conducted at the University of Sydney will be supported by Cancer Institute NSW Translational Program Grant 2019/TPG2165 and Medical Research Future Fund NCRI000139. Paul Keall receives support from NHMRC APP 1194004. Research activities conducted by site personnel will be funded by site departmental funds.

Data availability
No datasets were generated or analysed during the current study.

Declarations

Ethics approval and consent to participate
The ethical aspects of this research project have been approved by the Metro South Hospital and Health Service Human Research Ethics Committee (Metro South HREC; registration EC00167). Victorian Specific Modules (VSM) for retrospective and prospective recruitment types, required for all Victorian sites, was also approved by Metro South HREC. Public Health Act (PHA) approval was provided by the Director of Health Innovation, Investment and Research Office, Queensland Health for Queensland sites for access to data via waiver of consent. Consenting procedures, including a waiver of the need to consent for retrospectively recruited participants, were approved by the Metro South HREC.

Consent for publication
Not applicable.

Competing interests
The University of Sydney, Prof Paul Keall, Adam Mylonas and Dr Doan Trang Nguyen are inventors on a filed patent on markerless tracking that is owned by SeeTreat Pty Ltd. Prof Paul Keall and Dr Doan Trang Nguyen are shareholders and directors of SeeTreat.

Received: 9 January 2024 / Accepted: 17 September 2024
Published online: 08 October 2024

References

1. Sung H, Ferlay J, Siegel RL, Laversanne M, Soerjomataram I, Jemal A, et al. Global Cancer statistics 2020: GLOBOCAN estimates of incidence and Mortality Worldwide for 36 cancers in 185 countries. CA Cancer J Clin. 2021;71(3):209–49.
2. Llovet JM, Kelley RK, Villanueva A, Singal AG, Pikarsky E, Roayaie S, et al. Hepatocellular carcinoma. Nat Rev Dis Primers. 2021;7(1):6.
3. Villanueva A, Hepatocellular Carcinoma. N Engl J Med. 2019;380(15):1450–62.
4. Cocker F, Chien Yee K, Palmer AJ, de Graaff B. Increasing incidence and mortality related to liver cancer in Australia: time to turn the tide. Aust N Z J Public Health. 2019;43(3):267–73.
5. Chang BK, Timmerman RD. Stereotactic body radiation therapy: a comprehensive review. Am J Clin Oncol. 2007;30(6):637–44.
6. Timmerman R, Paulus R, Galvin J, Michalski J, Straube W, Bradley J, et al. Stereotactic body radiation therapy for inoperable early stage lung cancer. JAMA. 2010;303(11):1070–6.
7. Hoyer M, Swaminath A, Bydder S, Lock M, Mendez Romero A, Kavanagh B, et al. Radiotherapy for liver metastases: a review of evidence. Int J Radiat Oncol Biol Phys. 2012;82(3):1047–57.
8. Hall WA, Stapleford LJ, Hadjipanayis CG, Curran WJ, Crocker I, Shu HK. Stereotactic body radiosurgery for spinal metastatic disease: an evidence-based review. Int J Surg Oncol. 2011;2011:979214.
9. Shirato H, Seppenwoolde Y, Kitamura K, Onimura R, Shimizu S. Intrafractional tumor motion: lung and liver. Semin Radiat Oncol. 2004;14(1):10–8.
10. Gargett M, Haddad C, Kneebone A, Booth JT, Hardcastle N. Clinical impact of removing respiratory motion during liver SABR. Radiat Oncol. 2019;14(1):93.
11. Keall PJ, Nguyen DT, O'Brien R, Caillet V, Hewson E, Poulsen PR, et al. The first clinical implementation of real-time image-guided adaptive radiotherapy using a standard linear accelerator. Radiother Oncol. 2018;127(1):6–11.
12. Sengupta C, Nguyen DT, Moodie T, Mason D, Luo J, Causer T et al. The first clinical implementation of real-time 6 degree-of-freedom image-guided Radiotherapy for Liver SABR patients. Radiother Oncol. 2023:110031.
13. Dutta D, Kataki KJ, George S, Reddy SK, Sashidharan A, Kannan R, et al. Prospective evaluation of fiducial marker placement quality and toxicity in liver CyberKnife stereotactic body radiotherapy. Radiat Oncol J. 2020;38(4):253–61.
14. Bertholet J, Knopf A, Eiben B, McClelland J, Grimwood A, Harris E, et al. Real-time intrafraction motion monitoring in external beam radiotherapy. Phys Med Biol. 2019;64(15):15TR01.
15. Yue J, Sun X, Cai J, Yin FF, Yin Y, Zhu J, et al. Lipiodol: a potential direct surrogate for cone-beam computed tomography image guidance in radiotherapy of liver tumor. Int J Radiat Oncol Biol Phys. 2012;82(2):834–41.
16. Buckstein M, Kim E, Ozbek U, Tabrizian P, Gunasekaran G, Facciuto M, et al. Combination Transarterial Chemoembolization and Stereotactic Body Radiation Therapy for Unresectable single large Hepatocellular Carcinoma: results from a prospective phase 2 trial. Int J Radiat Oncol Biol Phys. 2022;114(2):221–30.
17. Yoon SM, Ryoo BY, Lee SJ, Kim JH, Shin JH, An JH, et al. Efficacy and Safety of Transarterial Chemoembolization Plus External Beam Radiotherapy vs Sorafenib in Hepatocellular Carcinoma with Macroscopic Vascular Invasion: a Randomized Clinical Trial. JAMA Oncol. 2018;4(5):661–9.
18. Isola P, Zhu J-Y, Zhou T, Efros AA, editors. Image-to-image translation with conditional adversarial networks. Proceedings of the IEEE conference on computer vision and pattern recognition; 2017.

Publisher's note

B

Appendix B: Realistic CT Data Augmentation for Accurate Deep-Learning Based Segmentation of Head and Neck Tumours

Markerless tumour tracking is desired for head and neck cancer as surgically implanting fiducial markers is high risk in the head and neck region. Additionally, immobilisation masks can be distressing for patients and motion can still occur. The deep learning framework presented in this thesis was used to segment head and neck tumours. The study proposed a method for realistic data augmentation to improve the robustness of deep learning models, achieving millimetre accuracy. This appendix has been published as "Gardner, M., Bouchta, Y.B., **Mylonas, A.**, Mueller, M., Cheng, C., Chlap, P., Finnegan, R., Sykes, J., Keall, P.J. & Nguyen, D.T., Realistic CT data augmentation for accurate deep-learning based segmentation of head and neck tumors in kV images acquired during radiation therapy. *Medical Physics*, **50**, 4206-4219 (2023)."

Received: 13 May 2022 | Revised: 15 March 2023 | Accepted: 20 March 2023

DOI: 10.1002/mp.16388

RESEARCH ARTICLE

MEDICAL PHYSICS

Realistic CT data augmentation for accurate deep-learning based segmentation of head and neck tumors in kV images acquired during radiation therapy

Mark Gardner[1] | **Youssef Ben Bouchta**[1] | **Adam Mylonas**[1] | **Marco Mueller**[1] | **Chen Cheng**[1] | **Phillip Chlap**[2,3,4] | **Robert Finnegan**[3,5,6] | **Jonathan Sykes**[7] | **Paul J Keall**[1] | **Doan Trang Nguyen**[1,8]

[1]ACRF Image X Institute, The University of Sydney, Eveleigh, New South Wales, Australia

[2]South Western Sydney Clinical School, University of New South Wales, Liverpool, New South Wales, Australia

[3]Ingham Institute for Applied Medical Research, Liverpool, New South Wales, Australia

[4]Liverpool Cancer Therapy Centre, Liverpool Hospital, Liverpool, New South Wales, Australia

[5]Northern Sydney Cancer Centre, Royal North Shore Hospital, St Leonards, New South Wales, Australia

[6]Institute of Medical Physics, University of Sydney, Camperdown, New South Wales, Australia

[7]Blacktown Cancer & Haematology Centre, Blacktown Hospital, Sydney, New South Wales, Australia

[8]University of Technology, Sydney, New South Wales, Australia

Correspondence
Mark Gardner, ACRF Image X Institute, The University of Sydney, Eveleigh, New South Wales 2015, Australia.
Email: mark.gardner@sydney.edu.au

Funding information
Cancer Australia, Grant/Award Number: APP1180776

Abstract

Background: Using radiation therapy (RT) to treat head and neck (H&N) cancers requires precise targeting of the tumor to avoid damaging the surrounding healthy organs. Immobilisation masks and planning target volume margins are used to attempt to mitigate patient motion during treatment, however patient motion can still occur. Patient motion during RT can lead to decreased treatment effectiveness and a higher chance of treatment related side effects. Tracking tumor motion would enable motion compensation during RT, leading to more accurate dose delivery.

Purpose: The purpose of this paper is to develop a method to detect and segment the tumor in kV images acquired during RT. Unlike previous tumor segmentation methods for kV images, in this paper, a process for generating realistic and synthetic CT deformations was developed to augment the training data and make the segmentation method robust to patient motion. Detecting the tumor in 2D kV images is a necessary step toward 3D tracking of the tumor position during treatment.

Method: In this paper, a conditional generative adversarial network (cGAN) is presented that can detect and segment the gross tumor volume (GTV) in kV images acquired during H&N RT. Retrospective data from 15 H&N cancer patients obtained from the Cancer Imaging Archive were used to train and test patient-specific cGANs. The training data consisted of digitally reconstructed radiographs (DRRs) generated from each patient's planning CT and contoured GTV. Training data was augmented by using synthetically deformed CTs to generate additional DRRs (in total 39 600 DRRs per patient or 25 200 DRRs for nasopharyngeal patients) containing realistic patient motion. The method for deforming the CTs was a novel deformation method based on simulating head rotation and internal tumor motion. The testing dataset consisted of 1080 DRRs for each patient, obtained by deforming the planning CT and GTV at different magnitudes to the training data.

The accuracy of the generated segmentations was evaluated by measuring the segmentation centroid error, Dice similarity coefficient (DSC) and mean surface distance (MSD). This paper evaluated the hypothesis that when patient motion occurs, using a cGAN to segment the GTV would create a more accurate segmentation than no-tracking segmentations from the original contoured

GTV, the current standard-of-care. This hypothesis was tested using the 1-tailed Mann-Whitney U-test.

Results: The magnitude of our cGAN segmentation centroid error was (mean ± standard deviation) 1.1 ± 0.8 mm and the DSC and MSD values were 0.90 ± 0.03 and 1.6 ± 0.5 mm, respectively. Our cGAN segmentation method reduced the segmentation centroid error ($p < 0.001$), and MSD ($p = 0.031$) when compared to the no-tracking segmentation, but did not significantly increase the DSC ($p = 0.294$).

Conclusions: The accuracy of our cGAN segmentation method demonstrates the feasibility of this method for H&N cancer patients during RT. Accurate tumor segmentation of H&N tumors would allow for intrafraction monitoring methods to compensate for tumor motion during treatment, ensuring more accurate dose delivery and enabling better H&N cancer patient outcomes.

KEYWORDS
cGAN, radiotherapy, segmentation

1 | INTRODUCTION

Radiation therapy (RT) is indicated for 74% of head and neck (H&N) cancer patients.[1] H&N RT has a higher risk of adverse side effects than treatments to other sites since there are many important organs located near the planning target volume (PTV). Recent advances in RT, including intensity-modulated RT, which confirms the high dose to the complex shapes of the target volume and minimizes dose to organs at risk (OAR), has led to improved survival rates and reduced toxicities.[2–5] However, treatment-related toxicities still occur and can become a serious health risk if the dose received by organs close to the PTV exceeds certain thresholds.[4–6]

To minimize the dose delivered to healthy organs, H&N cancer patients are required to wear a skin-tight immobilization mask that minimizes patient motion.[7,8] However, despite the restrictive nature of immobilization masks, motion in the order of several millimeters of the tumor and surrounding tissue can still occur during and between treatment fractions.[9,10] This motion can be caused by a change in the mask's fit due to patient weight loss between fractions, imperfections in the mask manufacturing and fitting process, tumor shrinkage, or treatment-related oedema.[7] The current standard of care is to use PTV margins of 2–5 mm rather than motion tracking to account for motion and changes in the target volume.[11] This margin, combined with intrafraction motion, leads to increased dose to surrounding healthy tissue, as well as decreased dose to the target.[12]

Tracking the tumor position during treatment would enable the use of more accurate radiotherapy techniques such as gating or real-time beam adaptation technology and could lead to being able to reduce the PTV margins. One method used in other treatment sites such as prostate, lung and liver to track the tumor position during treatment involves surgically implanting fiducial markers into the tumor which can then be tracked using kV projection images acquired during treatment.[13–15] However, surgically implanting fiducial markers into H&N tumors has a high risk of complication and the large and complex shape of H&N tumors makes implanting markers difficult. If intrafraction tumor monitoring is to be implemented for H&N tumors, a markerless approach to detecting the tumor position in kV images needs to be implemented.

Markerless tumor detection methods have been previously proposed for the lungs,[16–22] liver,[22] prostate,[23,24] and pancreas.[25] As detecting the Gross Tumor Volume (GTV) in kV images is difficult due to the low contrast between the GTV and surrounding tissue, most markerless tracking methods are based on deep learning.[22,25] Template-matching or feature based registration has also been used for markerless tumor tracking, however these methods use high-gradient or high intensity features such as bone as a surrogate for tumor motion.[26,27]

4DCT is commonly used to train deep learning based markerless segmentations methods because the multiple volumes provide large amounts of training data showing how the images will change with motion.[22,23,28] However, 4DCTs are not routinely acquired for H&N RT, so in this paper we present a way of training a deep-learning network to segment H&N tumors from only the planning CT.

The aim of this paper was to develop a realistic data augmentation approach for H&N images which would allow a deep learning method to detect and segment the primary GTV for H&N cancer patients in kV images using only the planning CT as training data. The deep learning network used was a conditional generative adversarial network (cGAN). The effectiveness of our cGAN segmentation method was evaluated by testing the hypothesis that our cGAN segmentation method improves GTV segmentation accuracy when compared to the current standard of care in which no GTV tracking

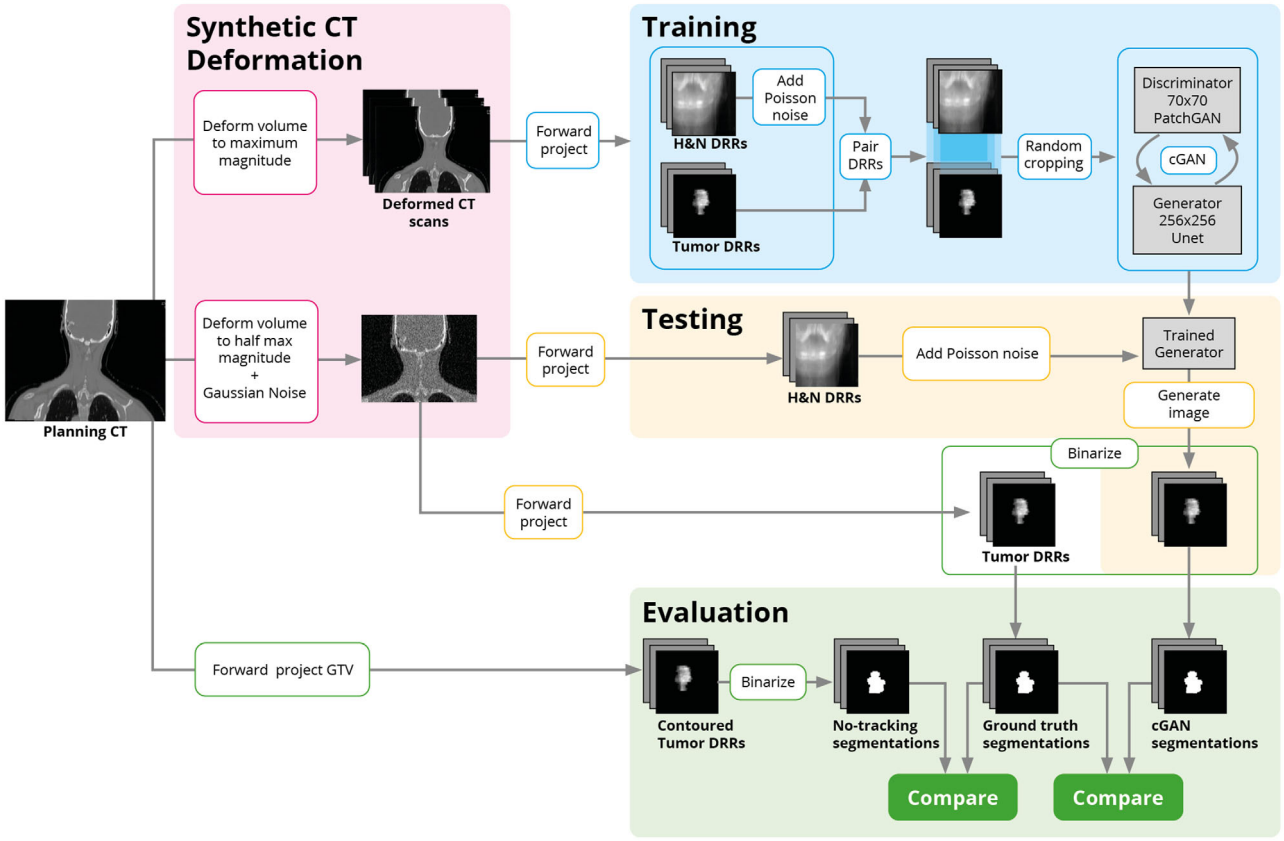


FIGURE 1 A flowchart showing the method used to evaluate the patient-specific deep learning method for one patient. First, the planning CT and contoured GTV is deformed multiple times and forward projected to create the training data. The same planning CT is deformed at half the magnitude of the training data to create the testing data. The original contoured GTV is also forward projected to create the no-tracking segmentations which assume no motion occurred in the testing data. DRR, digitally reconstructed radiograph; GTV, gross tumor volume.

is used. The data augmentation simulated realistic patient movement, which was achieved using a novel synthetic deformation method. To the authors knowledge, this paper describes the first implementation of markerless tumor detection of H&N tumors in kV images. This feasibility study is an important step in demonstrating the feasibility of markerless tracking of H&N tumors, and is an important step toward reducing the need for immobilization equipment during RT.

2 | METHODS

A flowchart detailing the training, testing and evaluation of our GTV tracking method is shown in Figure 1. The code used in this paper to generate these results is available on a public repository (https://github.com/ACRF-Image-X-Institute/MarkerlessHNGTVTracking). Most deep neural network-based markerless tumor tracking methods uses the large amount of data acquired in 4DCTs to train their neural network.[22,23] As 4DCTs are not routinely used for H&N treatments, our training dataset was generated from the planning CT by using a novel synthetic CT deformation method to deform each patient's planning CT to generate multiple CT volumes. From these multiple CT volumes, synthetic images in the form of digitally reconstructed radiographs (DRRs) were created and used to train a patient-specific cGAN to segment the GTV in the DRRs. To create the testing dataset, the planning CT volumes were again deformed by creating an additional realistic synthetic deformation. This additional deformation had different magnitudes to the deformations used to create the training data. The resultant deformed CT was then used to create a set of testing DRRs.

The cGANs used in this paper were trained using DRRs, which are simulated 2D fluoroscopy x-ray images created from a 3D CT volume.[2] Using a known projection geometry, DRRs can be created at different projection angles to simulate kV images acquired during RT. There are known differences between the noise properties and the image quality of kV images and DRRs,[29-31] however using DRRs to train the patient-specific cGANs allows for the networks to be trained without needing any additional images to be acquired. The use of DRRs for testing allows for the exact location of the ground truth GTV segmentations to be known in each projection and is a useful first step in evaluating the feasibility of our cGAN segmentation method described in this paper.

2.1 | Database and patient selection

The data for this paper involved 15 patients with head and neck squamous cell carcinoma (HNSCC) and were acquired from the HNSCC database[32-34] on The Cancer Imaging Archives (TCIA).[35] For each patient, the original data consisted of a planning CT and corresponding structure file, which contained the contoured primary GTV. The 15 patients were sequentially selected based on tumor location to ensure a range of primary tumor locations and to investigate the feasibility of the method described in this paper. The locations of the tumor were the oropharynx ($n = 5$), the larynx ($n = 5$), and the nasopharynx ($n = 5$). Patients with different tumor locations in the head and neck were selected to test the robustness of our patient-specific segmentation method.

2.2 | Synthetic CT deformations

Previous implementations of deep learning networks for markerless tumor detection and segmentation in kV images trained the network using data from 4DCT scans.[16,17,22,25] The advantage of using the 4DCT for training is that the network is trained on images showing how the tumor and surrounding tissue move and deform. H&N cancer treatment planning is typically done on a regular CT scan, which presents a challenge for training a patient-specific segmentation model because regular CT scans contain a single volume whereas 4DCTs contain multiple volumes. This reduces the training dataset and results in the network being less effective at detecting and segmenting the tumor when motion occurs.

To compensate for the lack of motion data, we developed a CT deformation-based data augmentation method that can be used to generate synthetic images showing realistic head motion. This data augmentation method allowed each patient's cGAN to be trained on a patient-specific dataset containing images of H&N motion without requiring additional CT scans. The assumption was made that two types of movements would be the primary sources of tumor motion during RT treatment: head rotation and internal tumor motion. A brief summary of the method used to create these synthetic deformations is included below, with a more detailed discussion provided in the Supplementary Materials.

2.2.1 | Head rotation

Six different training volumes representing the different types of head rotation were created: axial rotation (left-right rotation around the superior-inferior (SI) axis rotation), lateral bending (left-right rotation around the anterior-posterior (AP) axis), and flexion/extension (back-front rotation around the left-right (LR) axis). The head rotation deformation vector fields (DVFs) were created by rotating the volume around anatomical landmarks to replicate real head motion. A summary of the anatomical landmarks used as centres of rotation and realistic upper limits of the magnitude of patient head rotations during treatment is described in the Supplementary Materials. A head mask was created so the deformation would only be applied to the head and upper neck, with the inferior boundary on the neck being defined manually, as approximately halfway between the superior boundary of the shoulders, and the junction between the occipital bone in the neck and the top vertebrae.

2.2.2 | Tumor motion

To simulate tumor motion, the contoured GTV volume was rigidly shifted while the surrounding bones were kept stationary. The DVF used to shift the GTV had a Gaussian smoothing filter applied with a filter width $\sigma = 2\ mm$. The use of the smoothing filter created a more realistic deformation by ensuring that there was a gradual and continuous change in the DVF magnitude from the outside the boundary of the GTV to the centre of the GTV. Although σ is a heuristic value, the value chosen for this paper was based on similar instances of smoothing organ boundaries in CT scans.[36-38] The magnitudes of the applied tumor shifts were based on realistically large magnitudes of motion of the GTV during treatment and depended on the tumor location, as tumor motion during treatment is different for oropharyngeal, laryngeal, and nasopharyngeal tumors. A discussion on realistic magnitudes of tumor shifts during treatment is provided in the Supplementary Materials. Since minimal tumor motion has been observed in the LR direction the magnitude of these shifts was set to 0.[11,39] Similarly, since nasopharyngeal tumors have minimal tumor motion independent of the surrounding bony anatomy, the magnitude of the GTV shifts in all directions for nasopharyngeal tumors was set to 0. In total four volumes were created to simulate the tumor motion: an anterior and posterior GTV shift, and a superior and an inferior GTV shift.

2.2.3 | DVF creation

A flowchart showing how the final volumes were created is shown in Figure 2. For each deformed training volume, the head rotation and internal tumor motion DVFs were calculated separately and then combined to create the initial DVF estimation. For combined head rotation

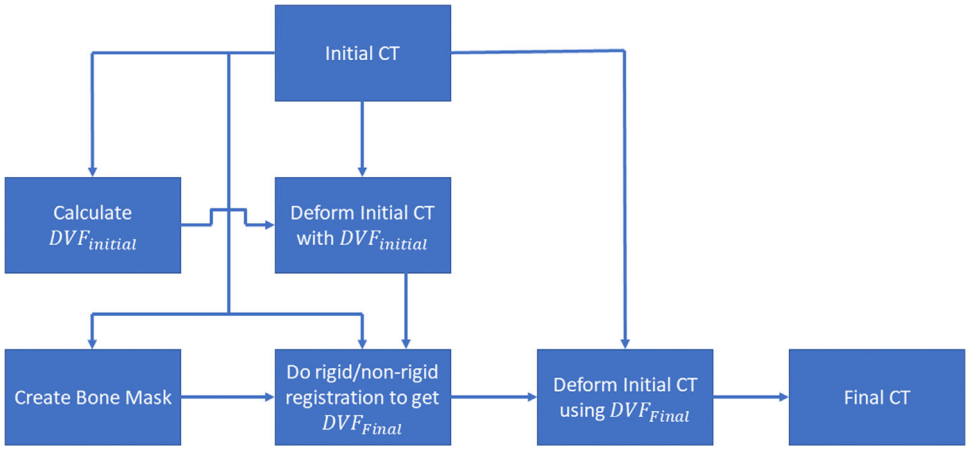


FIGURE 2 Flowchart for deforming the initial CT volume to simulate realistic motion.

motion (e.g., combination head flexion and internal tumor motion) the individual head rotation DVFs were similarly calculated for each type of motion separately (e.g., head flexion deformation calculated separately to internal tumor motion) and then combined to create the initial head motion DVF. The initial DVFs of the head rotation and the internal tumor motion were created using the platipy library for python.[40]

To create realistic DVFs that reduce boundary artefacts, a combined rigid/non-rigid registration method was used. The rigid/non-rigid registration method is a non-rigid registration method developed by Staring et al. which attempts to penalise non-rigid motion in user defined regions and is described in more detail in the Supplementary Material.[41] To implement this registration algorithm, we first applied the initial DVF generated by platipy to the planning CT to obtain an approximately deformed CT. We then created a bone mask from the planning CT and used it to define a region in which non-rigid motion was penalised. Finally, we obtained the final DVF by registering the original planning CT to the approximately deformed CT using the combined rigid/non-rigid registration method. The result was a more realistic transformation in which bones move rigidly and soft tissue non-rigidly. The registration process was completed using the elastix toolbox.[42] This synthetic deformation method was converted into a generalized software package for deforming the head and neck in CT scans and is available in a public repository (https://github.com/ACRF-Image-X-Institute/CTHeadDeformation).

2.2.4 | Synthetic deformations for creating training data

A table summarizing the different training volumes created (and the deformation magnitudes used to create those volumes) for each patient is shown in Table 1. For patients with oropharyngeal and laryngeal tumors, the training data for each patient was derived from all 11 volumes (as annotated in Table 1). However, for patients with nasopharyngeal tumors since no GTV shifts were added, only volumes 1–7 (as annotated in Table 1) were generated for each patient and used to produce the training data. To robustly demonstrate the effectiveness of the GTV segmentation method described in this paper, the same deformations cannot be used to create the training volume as well as the testing volume, but it is also unrealistic to have the motion of the testing volume larger than the training data. Hence, the head rotation and GTV shift magnitudes used for training (shown in Table 1) were twice as large as realistic upper limits of head and GTV motion (as described in the Supplementary Material), so the testing data could be created using realistic head and GTV deformations.

2.3 | cGAN segmentation method

To train a cGAN, two convolutional neural networks are trained simultaneously. A generator network *G* takes an input image *x* and creates a segmentation image *G*(*x*) based on the training data. The discriminator network *D* classifies whether the paired image *xy* came from the training set or the generator network, as shown in Figure 3.[43]

The cGAN was chosen for segmentation as the use of the discriminator network in the cGAN allows for the generation of unique and adaptive loss metrics.[44,45] The DRR images (and in future patient kV images) vary considerably for different patients depending on the tumor location, and even between patients depending on the projection angle. The ability of the cGAN to "learn" the loss function will ideally allow for the generation of patient-specific networks that can perform

TABLE 1 A summary of the training volumes created for each patient. The magnitude of the tumor motion depends on the tumor location and is shown in shown in more detail in the supplementary materials. No additional motion means that any tumor displacement resulting from the applied deformation is only as a result of the head moving, and not from any additional internal organ motion.

Volume number	Head rotation	Tumor motion	Motion magnitude (deg/mm)
1	None	No additional motion	0
2	Head rotation left	No additional motion	5.0°
3	Head rotation right	No additional motion	5.0°
4	Head tilt left	No additional motion	6.4°
5	Head tilt right	No additional motion	6.4°
6	Head nod up	No additional motion	6.2°
7	Head nod down	No additional motion	6.2°
8	None	Superior GTV shift	Oropharynx: 4 mm Larynx: 7.6 mm Nasopharynx: 0 mm
9	None	Inferior GTV shift	Oropharynx: 4 mm Larynx: 7.6 mm Nasopharynx: 0 mm
10	None	Anterior GTV shift	Oropharynx: 3.4 mm Larynx: 4.4 mm Nasopharynx: 0 mm
11	None	Posterior GTV shift	Oropharynx: 3.4 mm Larynx: 4.4 mm Nasopharynx: 0 mm

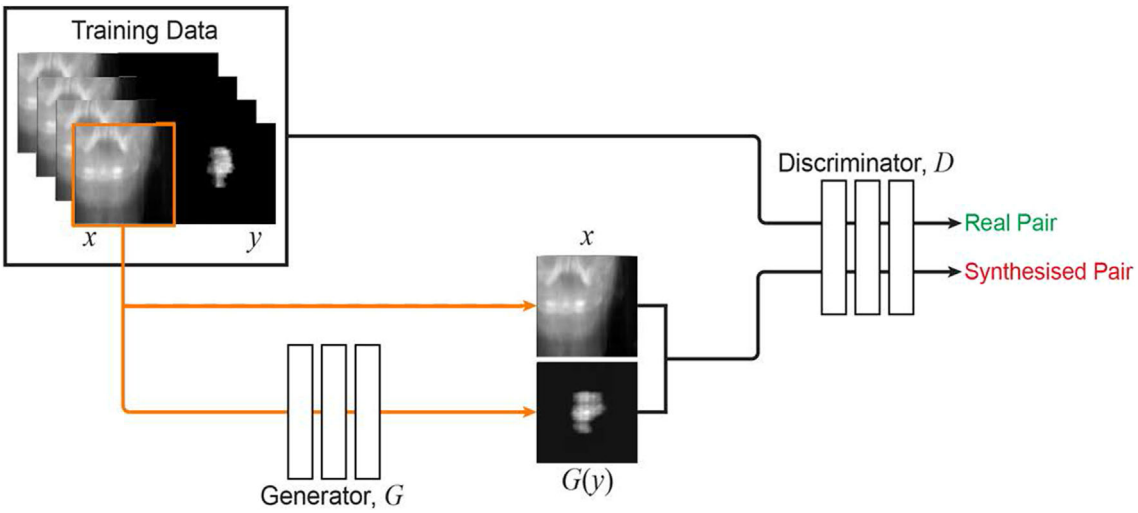


FIGURE 3 During training, the generator G creates image $G(x)$ from the image x that was passed to the discriminator D which aims to reject synthetic images and accept real segmentations y. This method efficiently trains the generator to produce accurate tumor segmentations from H&N DRRs.

accurate tumor segmentation of the unique and variable DRR images. In this paper, the aim of the cGAN was to allow for segmentation of the GTV in DRRs. However, to make the network optimization process more efficient, the cGAN was trained to generate an image showing a tumor DRR from a regular H&N DRR, as shown in Figure 3.[24] The output image generated by the cGAN was then binarized to create the GTV segmentation.

Our cGAN implementation used was based on the pix2pix method which provides a general solution to train a network to perform image-to-image translation (https://github.com/junyanz/pytorch-CycleGAN-and-pix2pix).[44] The cGAN was initialized with random parameters and trained to optimize the loss function

$$G^* = arg\min_G \max_D \mathcal{L}_{cGAN}(G, D) + \lambda \mathcal{L}_{L1}(G) \quad (1)$$

where λ is a constant (set to 100 for this implementation) and for all expected values of x (E_x):

$$\mathcal{L}_{\text{cGAN}}(G, D) = E_{x,y}[\log D(x, y)] + E_x[\log(1 - \mathrm{D}(\mathrm{x}, \mathrm{G}(\mathrm{x})))] \quad (2)$$

$$\mathcal{L}_{L1} = E_{x,y}\|y - G(x)\|;_1 \quad (3)$$

A custom data loader was used to allow 16-bit grayscale images to be used as input. A 70 × 70 PatchGAN[44,46] was used for the discriminator architecture and a 256 × 256 UNet[44,47] for the generator architecture (architecture shown in Supplementary Material). The training images of 550 × 550 pixels in size with a pixel width of 0.388 mm were randomly cropped to a size of 512 × 512 pixels during network training, and the testing images had a size of 512 × 512 pixels. Both training and testing images had random Poisson noise ($100 \leq \lambda \leq 200$) added to mimic the noise patterns of a kV image.[31]

Each patient-specific model was trained with a batch size of 8 and a learning rate of 0.0002 using the Adam optimiser. Each network was trained until the loss functions reached an equilibrium.[48,49] A plot of the loss functions for one patient is shown in the Supplementary Material. The models were trained on a desktop computer with two Intel Xeon Gold 6130 processors (2.1 GHz) with 64 GB RAM and a NIVIDIA Quadro P6000 Graphics Processing Unit (GPU). A patient-specific network was trained for each patient as training an effective patient-specific network requires less data than a similar generalized network and will be applicable to patients imaged using different imaging systems.

2.3.1 | Training data

The training data consisted of sets of paired images: a training H&N DRR and a corresponding training tumor DRR obtained from the 3D contour of the primary GTV. The training H&N and tumor DRRs were created by forward x-ray projecting each CT and GTV volume respectively using the RTK library.[50] For each CT volume, 3600 pairs of DRRs were generated evenly in a 360° arc around the volume. For each oropharyngeal and laryngeal patient, the training data were derived from 11 CT volumes (the original planning CT and 10 deformed CTs) and for nasopharyngeal patients seven CT volumes (the original planning CT and six deformed CTs). In total, the training data consisted of 39 600 paired DRRs for patients with oropharyngeal and laryngeal tumors and 25 200 paired DRRs for patients with nasopharyngeal tumors.

2.3.2 | Testing data

To generate the testing data for each patient, the planning CT and contoured GTV were deformed again three separate times, using the synthetic deformation method used to create the training data, but with different magnitudes of head rotations and GTV shifts. The magnitudes for the head rotations and GTV shifts used to create the three testing CT volumes are shown in Table 2. Additionally, the testing CT volumes had Gaussian noise ($\sigma = 10^{-5}\text{mm}^{-1}$) added to further differentiate the testing volume from the training CT volumes.[51] Each testing CT volume and GTV volume were used to create the testing dataset consisting of 360 paired DRRs evenly spaced around a 360° range leading to a total of 1080 testing DRRs for each patient.

For each testing H&N DRR created, the trained generator network produced an image $G(x)$ which was an estimation of the testing tumor DRR. To obtain segmentations from the predicted image $G(x)$ and the ground truth testing tumor DRR, both sets of images were normalized and then binarized using a threshold of 0.1. These binarized segmentations were then compared to evaluate the accuracy of our cGAN segmentation method.

2.3.3 | No-tracking segmentations

The original GTV contour from the planning CT was forward projected to create an additional set of testing segmentations (referred to as no-tracking segmentations), which assumed no motion occurs and hence does not track the GTV after the testing DVF is applied (as shown in Figure 1). The size of the no-tracking dataset was the same as the testing dataset. The no-tracking segmentations were used to test the hypothesis that the cGAN segmentation method improves the GTV segmentation accuracy when compared with the current standard of care (no GTV tracking). Hence the hypothesis is true if the accuracy of the cGAN segmentations is significantly greater than the accuracy of the no-tracking segmentations.

2.4 | Accuracy evaluation

To compare the accuracy of each segmentation, three metrics were used: the centroid error, Dice Similarity Coefficient (DSC), and mean surface distance (MSD). In this paper, we hypothesised that our cGAN segmentation method would improve the accuracy of GTV segmentations when compared with the no-tracking segmentations. The mean of each metric was calculated for every patient for both our cGAN segmentation method and the no-tracking segmentations. The

TABLE 2 The magnitude and direction of the applied head rotations and tumor shifts used to deform the planning CT and contoured GTV to create the testing data.

Testing volume number	Head rotation	Rotation magnitude	Tumor shift direction	Tumor shift magnitude
1 (25% of training magnitude)	Axial Rotation—left	1.25°	Superior	Oropharynx—1.0 mm Larynx—1.9 mm Nasopharynx—0 mm
	Lateral Bending—left	1.6°	Anterior	Oropharynx—0.85 mm Larynx—1.1 mm Nasopharynx—0 mm
	Head extension	1.55°	L/R	0 mm
2 (50% of training magnitude)	Axial Rotation—left	2.5°	Superior	Oropharynx—2.0 mm Larynx—3.8 mm Nasopharynx—0 mm
	Lateral Bending—left	3.2°	Anterior	Oropharynx—1.7 mm Larynx—2.2 mm Nasopharynx—0 mm
	Head extension	3.1°	L/R	0 mm
3 (75% of training magnitude)	Axial Rotation—left	3.75°	Superior	Oropharynx—3.0 mm Larynx—5.7 mm Nasopharynx—0 mm
	Lateral Bending—left	4.8°	Anterior	Oropharynx—2.55 mm Larynx—3.3 mm Nasopharynx—0 mm
	Head extension	4.65°	L/R	0 mm

averaged metric values were used to test the hypotheses using a paired 1-tail Mann-Whitney U test with a significance level of $\alpha = 0.05$.

The segmentation centroid error was used to measure the localization accuracy of our cGAN segmentation when compared to the ground truth segmentation. The centroid error was calculated in the kV imaging frame of reference where u is in the lateral direction and v is in the SI direction. The combined centroid error in the u and v directions was also calculated and is reported as the absolute centroid error magnitude.

The DSC was used to estimate how similar the shape of our cGAN and no-tracking segmentations were to the ground truth segmentations.[52] Similarly the MSD was also calculated, which measured the average distance (in mm) between the outline of the predicted and ground truth segmentations.[53] The MSD decreases as the accuracy of the segmentation increases.

3 | RESULTS

The centroid error, DSC and MSD of our cGAN segmentations for all testing data for all patients are shown in Table 3. Analysis of the change in segmentation accuracy due to the different magnitudes of head rotations and GTV shifts is shown in the Supplementary Material. For all patients the mean ± standard deviation cGAN segmentation centroid error in the u (lateral) and the v (vertical) directions was −0.0 ± 1.0 mm and 0.1 ± 0.9 mm, respectively. Additionally, for all patients, the 95th percentile centroid error in the u and v directions were [−1.4, 1.3] mm and [−1.4, 1.6] mm, respectively. The absolute centroid error of our cGAN segmentation 1.1 ± 0.8 mm was for all patients, with the 95th percentile error of [0.3, 2.3] mm. This is shown in comparison to the centroid error of the no-tracking segmentations in Figure 4. Using a paired 1-tail Mann-Whitney U test, our cGAN segmentation method significantly reduced the absolute segmentation centroid error when compared to the no-tracking segmentations ($p < 0.001$). Each patient-specific cGAN network took an average of 3 h to train.

For all patients the mean ± standard deviation DSC and MSD values for our cGAN segmentation were 0.90 ± 0.03 and 1.6 ± 0.5 mm, respectively, with the 95th percentile error for the DSC and MSD being [0.85, 0.94] and [0.9, 2.5] mm, respectively. The distribution of both the DSC and MSD values for our cGAN segmentation method are shown in comparison to the no-tracking segmentations in Figure 4. Using a paired 1-tail Mann-Whitney U test, our cGAN segmentation significantly reduced the MSD ($p = 0.031$) when compared to the no-tracking segmentations. The cGAN segmentation method did not significantly improve the DSC when compared to the no-tracking method ($p = 0.294221$) and only significantly improved the DSC for the Oropharynx tumors ($p < 0.0001$) but not for the Larynx ($p = 0.203$) and reduced the DSC for the Nasopharyx ($p < 0.0001$).

TABLE 3 The centroid error, DSC and MSD values for the predicted cGAN segmentations. All values are mean ± standard deviation. The best result for each tumor location group for each metric is highlighted by*.

Patient number	Centroid error—*v* (mm)	Centroid error—*u* (mm)	Absolute centroid error magnitude (mm)	DSC	MSD (mm)
Oropharynx					
1	0.7 ± 0.7	0.1 ± 0.9	1.2 ± 0.5	0.90 ± 0.01	1.5 ± 0.2
2	−0.3 ± 0.6*	0.2 ± 0.6	0.9 ± 0.4*	0.92 ± 0.01	1.1 ± 0.2*
3	−0.8 ± 0.8	−0.5 ± 0.7	1.2 ± 0.5	0.92 ± 0.01	1.6 ± 0.2
4	0.7 ± 0.5	−0.2 ± 0.7	1.0 ± 0.5	0.89 ± 0.03	1.6 ± 0.4
5	0.7 ± 0.7	−0.1 ± 0.7*	1.0 ± 0.5	0.92 ± 0.01	1.9 ± 0.3
All oropharynx	0.2 ± 0.9	−0.1 ± 0.7	1.1 ± 0.5	0.91 ± 0.02	1.5 ± 0.4
Larynx					
6	−0.6 ± 0.5	0.1 ± 0.8	1.0 ± 0.5	0.90 ± 0.02	1.3 ± 0.3*
7	0.7 ± 0.6	0.2 ± 0.6	1.0 ± 0.5	0.87 ± 0.02	1.6 ± 0.3
8	−0.2 ± 0.9*	0.0 ± 0.6*	1.0 ± 0.6	0.92 ± 0.02*	1.3 ± 0.3*
9	−0.7 ± 1.0	0.0 ± 0.6*	1.2 ± 0.6	0.86 ± 0.03	1.5 ± 0.3
10	−0.3 ± 0.4	0.1 ± 0.6	0.7 ± 0.4*	0.91 ± 0.02	1.3 ± 0.3*
All larynx	−0.2 ± 0.9	0.1 ± 0.7	1.0 ± 0.5	0.89 ± 0.03	1.4 ± 0.3
Nasopharynx					
11	0.9 ± 0.8	0.2 ± 1.5	1.7 ± 0.9	0.91 ± 0.02	2.3 ± 0.5
12	−0.3 ± 0.5*	0.0 ± 0.9	0.9 ± 0.5	0.90 ± 0.02	2.1 ± 0.4
13	0.9 ± 1.0	−0.2 ± 1.1	1.4 ± 1.0	0.92 ± 0.02	2.0 ± 0.6
14	−0.4 ± 0.5	0.0 ± 0.5*	0.7 ± 0.4*	0.94 ± 0.01*	1.4 ± 0.3*
15	0.1 ± 0.6	−0.3 ± 2.3	1.3 ± 2.0	0.89 ± 0.06	1.4 ± 0.3
All nasopharynx	0.3 ± 0.9	−0.1 ± 1.4	1.2 ± 1.2	0.91 ± 0.04	1.8 ± 0.7
Total					
All H&N patients	0.1 ± 0.9	−0.0 ± 1.0	1.1 ± 0.8	0.90 ± 0.03	1.6 ± 0.5

An example of our cGAN segmentation for several DRRs at different projection angles is shown in Figure 5 and video showing our cGAN segmentations for all projection angles is shown in Video S1. Figure 5 demonstrates an example of the typical accuracy of the cGAN segmentations achieved by this method for one H&N cancer patient.

4 | DISCUSSION

In this paper, we investigated the feasibility of a realistic data augmentation methods to assist in markerless tumor segmentation method for detecting and segmenting the primary GTV in DRRs for H&N cancer patients. We have demonstrated that we can use a synthetic CT deformation method to augment the training data by realistically deforming the planning CT. The results demonstrated that our cGAN segmentation method increased the accuracy of the GTV segmentation compared to the current standard of care. Our cGAN segmentation method detected the centroid of the GTV in DRRs with an accuracy of 1.1 ± 0.8 mm and segmented tumors in DRRs with DSC and MSD value of 0.90 ± 0.03 and 1.6 ± 0.5 mm, respectively.

Table 3 shows the magnitude of the mean centroid error for all patients was approximately 1 mm. This suggests that our cGAN segmentation method can accurately localize the GTV in DRR images. Accurate detection of the GTV is a crucial step in the implementation of image-guided RT methods.[54] Similarly, the mean DSC and MSD values of 0.90 and 1.6 mm, respectively suggest that our cGAN segmentation method can accurately segment the GTV in DRR images. The ability to adapt to the changing shape of the GTV would enable real-time adaptive RT that can not only modify the treatment beam location but also adapt to changes in the shape of the GTV volume. Future work will look to evaluate whether our cGAN segmentation method can be used to implement markerless image-guided RT or adaptive RT for H&N patients.

Most markerless tracking results described in other literature present the 3D localization error instead of the error in each 2D image, making direct accuracy

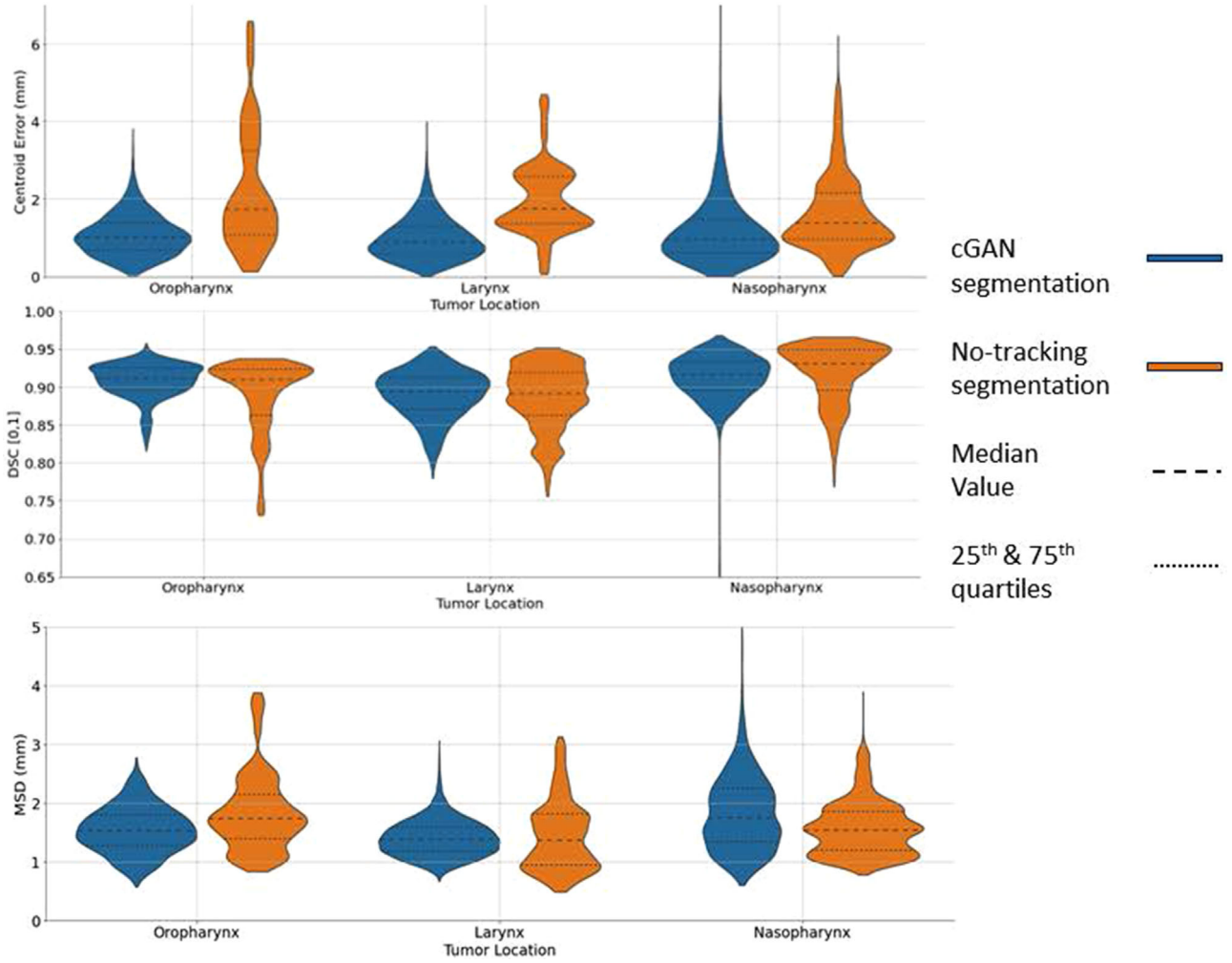
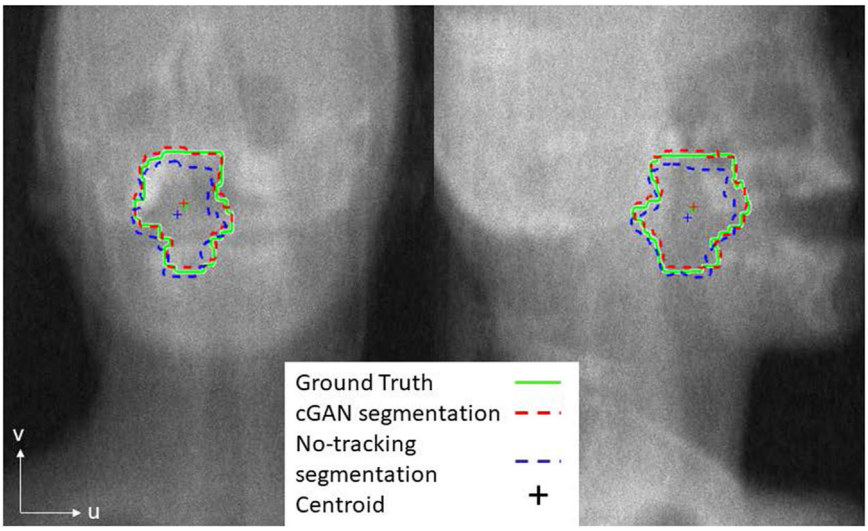


FIGURE 4 Violin plots showing the distribution of the accuracy metrics for our cGAN segmentation (blue) compared with the no-tracking segmentations (orange) for the different tumor locations. The metrics shown are the magnitude of the absolute centroid error (top), the Dice Similarity Coefficient (middle) and mean surface distance (bottom). The width of the violin plot at each *y* value corresponds to the frequency of that value.

FIGURE 5 Example of our cGAN segmentation (red) and the no-tracking segmentation (blue) in comparison to the ground truth segmentation (green) for different projection angles.



comparisons difficult. Zhao et al. developed a deep learning method for identifying a rectangular bounding box containing the prostate location in DRRs and acquired kV projections at three specific angles (0°, 90°, 135°).[55] They were able to estimate the tumor position in the generated DRRs with a mean accuracy of 1.58-1.67 mm, however this accuracy was only demonstrated for those three projection angles. A similar method was applied for pancreatic tumors with a mean absolute distance between the predicted and actual tumor centroid of less than 2.6 mm.[25] A deep-learning method for generating GTV contours in DRRs was applied for lung tumors, which resulted in DSC values of 0.81–0.98. These results show that our cGAN segmentation method has similar levels of accuracy to other deep learning based markerless GTV detection methods.

The time taken for the trained network to generate an estimate of the GTV segmentation was approximately 0.01 s per image. This demonstrates that this method has a small latency, which is a feature of deep learning methods as has been noted previously.[23] This short latency, as well as a mean absolute accuracy of 1.1 mm suggest that real-time tracking of H&N tumors, which requires a system latency of less than 0.5 s, is feasible.[56] The training of the cGAN prioritized accuracy over training time to prove that cGANs can be used to accurately segment DRR images. Future work can optimize the training process to further reduce the training time, by experimenting with the number of training images, as well as the training parameters such as step length.

The images that were used to train and test our cGAN segmentation method in this paper were DRR images. This is a limitation of the HNSCC dataset that was used, as this dataset has only one CT per patient, with no available kV images. However, the use of DRRs instead of real kV images in this paper ensures an accurate ground truth location for each test image, enabling an accurate quantification of the GTV segmentation effectiveness. In a clinical environment, our cGAN segmentation method would detect and segment the GTV in kV images acquired during RT fractions. However, intrafraction motion during treatment would make it difficult to establish an accurate GTV location in any kV images acquired during treatment.[9,10] We view this study as a necessary but not sufficient study toward the long-term goal of clinical implementation. Future work will focus on clinically acquired data for which there is no direct ground truth.

DRRs have a greater quality than kV images acquired during treatment. Even though Poisson noise was added to make the DRRs used in this study more similar to kV images, Poisson noise does not model all additional sources of artefacts and noise.[31] The difference in image quality between DRR and kV images means that a cGAN network trained on DRRs may decrease in accuracy when applied to kV images. In the future, the methodology presented in this paper will be used to evaluate our cGAN segmentation method on kV images[23,24] or by adding more realistic noise to the DRRs to attempt to replicate kV images.[29,30]

Similarly, the deformation method used to create the testing data approximated patient motion based on available data on H&N motion. The rigid/non-rigid registration method used to generate DVFs attempts to keep the bones rigid while all other tissue can be non-rigidly deformed. Adil Al-Mayah et al. used a biomechanical based finite element analysis to simulate deformation of the upper vertebrae and H&N,[57] which could be investigated in the future to potentially refine the generated deformations and allow for realistic deformation of all tissue. Pukala et al. developed a method for realistically deforming a cone-beam CT image to match a second cone-beam CT image of the same patient from a different treatment fraction.[58,59] However, this method only allows for the transformation from one scan to another scan, and cannot be used to make additional, realistic deformations. Hence there is scope to further optimise the deformation algorithms to generate more realistic synthetic deformations.

During RT for H&N cancer, the PTV often includes nearby lymph nodes as targeting these nodes this can prevent cancer recurrence.[60,61] While this work focused on tracking the primary GTV in DRRs, our cGAN segmentation method could be easily adapted to segment nearby organs of interest which are also likely to be affected by patient motion. This is demonstrated in this paper by the accurate detection and segmentation results for patients with tumors of different shapes, sizes, and locations. Future work could investigate multi-target tracking methods to allow for the tracking of the primary GTV as well as surrounding organs for either targeting or avoidance.

Our cGAN segmentation method described in this paper was demonstrated using DRRs. As radiographs are a planar imaging modality, our method can only estimate motion in the same plane as the DRR. In marker-based tumor tracking applications, methods have been developed for using the marker locations in kV images to estimate the 3D position and rotation of the tumor.[62,63] It is thought that this 3D position estimation method can be adapted such that markerless 3D position and motion estimation is possible, either by using the location of the segmentation centroid or points on the segmentation outline instead of the marker locations.[64] An accurate estimation of the 3D position and motion of the tumor would allow for improved dose delivery through the use of beam adaptation methods[65] or gated treatments.[63]

Currently for H&N cancer RT treatments, kV images are mainly acquired prior to treatment to assist in accurately positioning the patient.[7] For real-time tumor tracking of H&N tumors, additional kV images would have to be acquired during treatment at regular intervals.[14] The acquisition of these images would result in extra imaging dose to the patient. Future development of this

method should quantify the increase in dose caused by the acquisition of these additional kV images and evaluate the risk associated with this additional dose.

The immobilization masks that are used to restrict motion of H&N patients during RT cause anxiety and distress in a significant number of patients.[66,67] If mask-free RT can be achieved without a reduction in the accuracy of the dose delivery, then this would lead to better H&N cancer patient experience. While our cGAN segmentation method described in this paper has been developed for patients wearing the mask during treatment, future work will look to extend the markerless tumor detection framework described in this paper to allow for H&N cancer radiation therapy treatments without the need for an immobilization mask.

5 | CONCLUSION

In conclusion we have demonstrated the feasibility of our cGAN segmentation method in detecting and segmenting the primary GTV location in DRRs of H&N cancer patients. The centroid accuracy of our cGAN segmentation was -0.0 ± 1.0 mm in the u (Lateral) direction and 0.1 ± 0.9 mm and v (SI) directions. The DSC and MSD of our cGAN template segmentation was 0.90 ± 0.03 and 1.6 ± 0.5 mm, respectively when compared with the ground truth segmentation. When comparing the centroid error and MSD, the accuracy of our cGAN segmentation method was significantly greater than the no-tracking segmentations which assumed no motion. Accurate segmentation of the GTV can enable to tracking of the tumor during H&N RT treatment, leading to better dose delivery and better patient outcomes.

ACKNOWLEDGMENTS

The authors would like to thank Dr Helen Ball for her assistance with statistics, and Julia Johnson for her assistance with the figures. Finally, the authors would like to thank Dr Paul Liu and Dr Emily Hewson for helping to review the manuscript prior to submission.

CONFLICT OF INTEREST STATEMENT

Author Paul J Keall is an inventor on a patent application. A full list of disclosures is given at: https://imagex.sydney.edu.au/home/disclosures/. Authors Adam Mylonas, Marco Mueller, and Doan Trang Nguyen are inventors on a patent application for deep learning-based markerless prostate segmentation.

REFERENCES

1. Barton MB, Jacob S, Shafiq J, et al. Estimating the demand for radiotherapy from the evidence: a review of changes from 2003 to 2012. *Radiother Oncol*. 2014;112(1):140-144. doi:10.1016/j.radonc.2014.03.024
2. Lee AWM, Ng WT, Chan LLK, et al. Evolution of treatment for nasopharyngeal cancer - success and setback in the intensity-modulated radiotherapy era. *Radiother Oncol*. 2014;110(3):377-384. doi:10.1016/j.radonc.2014.02.003
3. Setton J, Caria N, Romanyshyn J, et al. Intensity-modulated radiotherapy in the treatment of oropharyngeal cancer: an update of the memorial sloan-kettering cancer center experience. *Int J Radiat Oncol Biol Phys*. 2012;82(1):291-298. doi:10.1016/j.ijrobp.2010.10.041
4. Nutting C, A'Hern R, Rogers MS, et al. First results of a phase III multicenter randomized controlled trial of intensity modulated (IMRT) versus conventional radiotherapy (RT) in head and neck cancer (PARSPORT: ISRCTN48243537; CRUK/03/005). *J Clin Oncol*. 2009;27(18 suppl):LBA6006-LBA6006. doi:10.1200/jco.2009.27.18_suppl.lba6006
5. Fang FM, Chien CY, Tsai WL, et al. Quality of life and survival outcome for patients with nasopharyngeal carcinoma receiving three-dimensional conformal radiotherapy vs. intensity-modulated radiotherapy-a longitudinal study. *Int J Radiat Oncol Biol Phys*. 2008;72(2):356-364. doi:10.1016/j.ijrobp.2007.12.054
6. Eisbruch A, Kim HM, Terrell JE, Marsh LH, Dawson LA, Ship JA. Xerostomia and its predictors following parotid-sparing irradiation of head-and-neck cancer. *Int J Radiat Oncol Biol Phys*. 2001;50(3):695-704. doi:10.1016/S0360-3016(01)01512-7
7. Leech M, Coffey M, Mast M, et al. ESTRO ACROP guidelines for positioning, immobilisation and position verification of head and neck patients for radiation therapists. *Tech Innov Patient Support Radiat Oncol*. 2017;1:1-7. doi:10.1016/j.tipsro.2016.12.001
8. Marsh R, Balter J, Evans VL, Eisbruch A. Design and analysis of an immobilization and repositioning system for treatment of neck malignancies. *Med Dosim*. 1997;22(4):293-297. doi:10.1016/S0958-3947(97)00102-7
9. Amelio D, Winter M, Habermehl D, Jäkel O, Debus J, Combs SE. Analysis of inter- and intrafraction accuracy of a commercial thermoplastic mask system used for image-guided particle radiation therapy. *J Radiat Res*. 2013;54(1):i69-i76. doi:10.1093/jrr/rrt038
10. Juan-Senabre XJ, López-Tarjuelo J, Conde-Moreno A, et al. Uncertainties and CTV to PTV margins quantitative assessment using cone-beam CT technique in clinical application for prostate, and head and neck irradiation tumours. *Clin Transl Oncol*. 2011;13(11):819-825. doi:10.1007/s12094-011-0740-8
11. Bruijnen T, Stemkens B, Terhaard CHJ, Lagendijk JJW, Raaijmakers CPJ, Tijssen RHN. Intrafraction motion quantification and planning target volume margin determination of head-and-neck tumors using cine magnetic resonance imaging. *Radiother Oncol*. 2019;130:82-88. doi:10.1016/j.radonc.2018.09.015
12. Neubauer E, Dong L, Followill DS, et al. Assessment of shoulder position variation and its impact on IMRT and VMAT doses for head and neck cancer. *Radiat Oncol*. 2012;7(1):19. doi:10.1186/1748-717X-7-19
13. Keall PJ, Ng JA, Juneja P, et al. Real-time 3D image guidance using a standard LINAC: measured motion, accuracy, and precision of the first prospective clinical trial of kilovoltage intrafraction monitoring-guided gating for prostate cancer radiation therapy. *Int J Radiat Oncol Biol Phys*. 2016;94(5):1015-1021. doi:10.1016/j.ijrobp.2015.10.009
14. Ng JA, Booth JT, Poulsen PR, et al. Kilovoltage intrafraction monitoring for prostate intensity modulated arc therapy: first clinical results. *Int J Radiat Oncol Biol Phys*. 2012;84(5):e655-e661. doi:10.1016/J.IJROBP.2012.07.2367
15. Keall PJ, Aun Ng J, O'Brien R, et al. The first clinical treatment with kilovoltage intrafraction monitoring (KIM): a real-time image guidance method. *Med Phys*. 2014;42(1):354-358. doi:10.1118/1.4904023
16. Mueller M, Zolfaghari R, Briggs A, et al. The first prospective implementation of markerless lung target tracking in an experimental quality assurance procedure on a standard linear accelerator. *Phys Med Biol*. 2020;65(2):025008. doi:10.1088/1361-6560/ab5d8b

17. Shieh CC, Caillet V, Dunbar M, et al. A Bayesian approach for three-dimensional markerless tumor tracking using kV imaging during lung radiotherapy. *Phys Med Biol*. 2017;62(8):3065-3080. doi:10.1088/1361-6560/aa6393
18. Mori S. Real-time image-processing algorithm for markerless tumour tracking using x-ray fluoroscopic imaging. *Br J Radiol*. 2014;87:20140001. doi:10.1259/bjr.20140001
19. Lewis JH, Li R, Watkins WT, et al. Markerless lung tumor tracking and trajectory reconstruction using rotational cone-beam projections: a feasibility study. *Phys Med Biol*. 2010;55(9):2505-2522. doi:10.1088/0031-9155/55/9/006
20. Hugo GD, Liang J, Yan D. Marker-free lung tumor trajectory estimation from a cone beam CT sinogram. *Phys Med Biol*. 2010;55(9):2637-2650. doi:10.1088/0031-9155/55/9/014
21. Fu D, Kahn R, Wang B, et al. Xsight lung tracking system: a fiducial-less method for respiratory motion tracking. *Treating Tumors that Move with Respiration*. Springer; 2007:265-282. doi:10.1007/978-3-540-69886-9_26
22. Hirai R, Sakata Y, Tanizawa A, Mori S. Real-time tumor tracking using fluoroscopic imaging with deep neural network analysis. *Phys Medica*. 2019;59:22-29. doi:10.1016/j.ejmp.2019.02.006
23. Zhao W, Han B, Yang Y, et al. Incorporating imaging information from deep neural network layers into image guided radiation therapy (IGRT). *Radiother Oncol*. 2019;140:167-174. doi:10.1016/j.radonc.2019.06.027
24. Mylonas A, Mueller M, Keall PJ, Booth JT, Nguyen DT. Towards real-time markerless prostate IGRT during VMAT treatment. In: American Association of Physicists in Medicine (AAPM). 2021.
25. Zhao W, Shen L, Han B, et al. Markerless pancreatic tumor target localization enabled by deep learning. *Int J Radiat Oncol Biol Phys*. 2019;105(2):432-439. doi:10.1016/j.ijrobp.2019.05.071
26. Xie Y, Xing L, Gu J, Liu W. Tissue feature-based intra-fractional motion tracking for stereoscopic x-ray image guided radiotherapy. *Phys Med Biol*. 2013;58(11):3615. doi:10.1088/0031-9155/58/11/3615
27. Lin T, Cervîo LI, Tang X, Vasconcelos N, Jiang SB. Fluoroscopic tumor tracking for image-guided lung cancer radiotherapy. *Phys Med Biol*. 2009;54(4):981. doi:10.1088/0031-9155/54/4/011
28. Shieh CC, Keall PJ, Kuncic Z, Huang CY, Feain I. Markerless tumor tracking using short kilovoltage imaging arcs for lung image-guided radiotherapy. *Phys Med Biol*. 2015;60(24):9437. doi:10.1088/0031-9155/60/24/9437
29. Staub D, Murphy MJ. A digitally reconstructed radiograph algorithm calculated from first principles. *Med Phys*. 2013;40(1):011902. doi:10.1118/1.4769413
30. Dhont J, Verellen D, Mollaert I, Vanreusel V, Vandemeulebroucke J. RealDRR – rendering of realistic digitally reconstructed radiographs using locally trained image-to-image translation. *Radiother Oncol*. 2020;153:213-219. doi:10.1016/j.radonc.2020.10.004
31. Strid KG. Significance of quantum fluctuations in roentgen imaging. *Acta Oncol (Madr)*. 1980;19(2):129-138. doi:10.3109/02841868009130144
32. Grossberg AJ, Elhalawani H, Mohamed ASR, et al. HNSCC Dataset. Published online 2020. doi:10.7937/k9/tcia.2020.a8sh-7363
33. Grossberg AJ, Mohamed ASR, Elhalawani H, et al. Imaging and clinical data archive for head and neck squamous cell carcinoma patients treated with radiotherapy. *Sci Data*. 2018;5(1):1-10. doi:10.1038/sdata.2018.173
34. Hesham E, Abdallah SRM, Aubrey LW, et al. Matched computed tomography segmentation and demographic data for oropharyngeal cancer radiomics challenges. *Sci Data*. 2017;4(1):1-14. doi:10.1038/sdata.2017.77
35. Clark K, Vendt B, Smith K, et al. The cancer imaging archive (TCIA): maintaining and operating a public information repository. *J Digit Imaging*. 2013;26:1045-1057. doi:10.1007/s10278-013-9622-7
36. Vercauteren T, Pennec X, Perchant A, Ayache N. Diffeomorphic demons: efficient non-parametric image registration. *Neuroimage*. 2009;45(1):S61-S72. doi:10.1016/J.NEUROIMAGE.2008.10.040
37. Eiben B, Bertholet J, Menten MJ, Nill S, Oelfke U, McClelland JR. Consistent and invertible deformation vector fields for a breathing anthropomorphic phantom: a post-processing framework for the XCAT phantom. *Phys Med Biol*. 2020;65(16):165005. doi:10.1088/1361-6560/ab8533
38. Papiez BW, Heinrich MP, Fehrenbach J, Risser L, Schnabel JA. An implicit sliding-motion preserving regularisation via bilateral filtering for deformable image registration. *Med Image Anal*. 2014;18(8):1299-1311. doi:10.1016/J.MEDIA.2014.05.005
39. Gurney-Champion OJ, McQuaid D, Dunlop A, et al. MRI-based assessment of 3D intrafractional motion of head and neck cancer for radiation therapy. *Int J Radiat Oncol Biol Phys*. 2018;100(2):306-316. doi:10.1016/j.ijrobp.2017.10.016
40. Finnegan R, Chlap P. PlatiPy (Processing Library and Analysis Toolkit for Medical Imaging in Python). GitHub. Accessed June 21, 2021. https://github.com/pyplati/platipy
41. Staring M, Klein S, Pluim JPW. A rigidity penalty term for nonrigid registration. *Med Phys*. 2007;34(11):4098-4108. doi:10.1118/1.2776236
42. Klein S, Staring M, Murphy K, Viergever MA, Pluim JPW. Elastix: a toolbox for intensity-based medical image registration. *IEEE Trans Med Imaging*. 2010;29(1):196-205. doi:10.1109/TMI.2009.2035616
43. Mirza M, Osindero S. Conditional Generative Adversarial Nets. arXiv preprint arXiv:1411.1784 (2014) http://arxiv.org/abs/1411.1784
44. Isola P, Zhu JY, Zhou T, Efros AA. Image-to-Image Translation with Conditional Adversarial Networks. Proceedings of 30th IEEE Conf Comput Vis Pattern Recognition, CVPR 2017. 2016; 5967-5976. Accessed April 19, 2021. http://arxiv.org/abs/1611.07004
45. Yi X, Walia E, Babyn P. Generative adversarial network in medical imaging: a review. *Med Image Anal*. 2019;58:101552. doi:10.1016/J.MEDIA.2019.101552
46. Li C, Wand M. Precomputed real-time texture synthesis with Markovian Generative Adversarial Networks. *Lecture Notes in Computer Science*. 2016:9907:702-716. doi:10.1007/978-3-319-46487-9_43
47. Ronneberger O, Fischer P, BroxT. U-Net: convolutional networks for biomedical image segmentation. *Lecture Notes in Computer Science*. 2015;9351:234-241. Accessed August 18, 2021. https://arxiv.org/abs/1505.04597v1
48. Goodfellow I, Pouget-Abadie J, Mirza M, et al. Generative adversarial nets. In: *Advances in Neural Information Processing Systems 27 (NIPS 2014)*. 2014.
49. Kessler DA, MacKay JW, Crowe VA, et al. The optimisation of deep neural networks for segmenting multiple knee joint tissues from MRIs. *Comput Med Imaging Graph*. 2020;86:101793. doi:10.1016/J.COMPMEDIMAG.2020.101793
50. Rit S, Vila Oliva M, Brousmiche S, Labarbe R, Sarrut D, Sharp GC. The Reconstruction Toolkit (RTK), an open-source cone-beam CT reconstruction toolkit based on the Insight Toolkit (ITK). In: *Journal of Physics: Conference Series*. Vol 489. Institute of Physics Publishing; 2014:012079. doi:10.1088/1742-6596/489/1/012079
51. Singhrao K, Kirby N, Pouliot J. A three-dimensional head-and-neck phantom for validation of multimodality deformable image registration for adaptive radiotherapy. *Med Phys*. 2014;41(12):121709. doi:10.1118/1.4901523
52. Shamir RR, Duchin Y, Kim J, Sapiro G, Harel N. Continuous dice coefficient: a method for evaluating probabilistic segmentations. *arXiv preprint arXiv:1906*. 2019:11031.

53. Chalana V, Kim Y. A methodology for evaluation of boundary detection algorithms on medical images. *IEEE Trans Med Imaging*. 1997;16(5):642-652. doi:10.1109/42.640755
54. Santos JDL, Popple R, Agazaryan N, et al. Image guided radiation therapy (IGRT) technologies for radiation therapy localization and delivery. *Int J Radiat Oncol Biol Phys*. 2013;87(1):33-45. doi:10.1016/J.IJROBP.2013.02.021
55. Zhao W, Shen L, Han B, et al. Deep learning approach for markerless pancreatic tumor target localization. *Int J Radiat Oncol*. 2019;105(1):S202-S203. doi:10.1016/j.ijrobp.2019.06.268
56. Keall PJ, Mageras GS, Balter JM, et al. The management of respiratory motion in radiation oncology report of AAPM Task Group 76a). *Med Phys*. 2006;33(10):3874-3900. doi:10.1118/1.2349696
57. Al-Mayah A, Moseley J, Hunter S, et al. Biomechanical-based image registration for head and neck radiation treatment. *Phys Med Biol*. 2010;55(21):6491-6500. doi:10.1088/0031-9155/55/21/010
58. Pukala J, Meeks SL, Staton RJ, Bova FJ, Mañon RR, Langen KM. A virtual phantom library for the quantification of deformable image registration uncertainties in patients with cancers of the head and neck. *Med Phys*. 2013;40(11):111703. doi:10.1118/1.4823467
59. Varadhan R, Karangelis G, Krishnan K, Hui S. A framework for deformable image registration validation in radiotherapy clinical applications. *J Appl Clin Med Phys*. 2013;14(1):192-213. doi:10.1120/jacmp.v14i1.4066
60. Barkley HT, Gilbert Fletcher TH, Richard Jesse TH, Robert Lindberg TD. Management of cervical lymph node metastases in squamous ceil carcinoma of the tonsillar fossa, base of tongue, supraglottic larynx, and hypopharynx. *Am J Surg*. 1972;124(4):462-467.
61. Lee N, Harris J, Garden AS, et al. Intensity-modulated radiation therapy with or without chemotherapy for nasopharyngeal carcinoma: radiation therapy oncology group phase II trial 0225. *J Clin Oncol*. 2009;27(22):3684-3690. doi:10.1200/JCO.2008.19.9109
62. Poulsen PR, Cho B, Langen K, Kupelian P, Keall PJ. Three-dimensional prostate position estimation with a single x-ray imager utilizing the spatial probability density. *Phys Med Biol*. 2008;53(16):4331-4353. doi:10.1088/0031-9155/53/16/008
63. Nguyen DT, O'Brien R, Kim JH, et al. The first clinical implementation of a real-time six degree of freedom target tracking system during radiation therapy based on Kilovoltage Intrafraction Monitoring (KIM). *Radiother Oncol*. 2017;123(1):37-42. doi:10.1016/j.radonc.2017.02.013
64. Roggen T, Bobic M, Givehchi N, Scheib SG. Deep learning model for markerless tracking in spinal SBRT. *Phys Medica*. 2020;74:66-73. doi:10.1016/J.EJMP.2020.04.029
65. Zhang L, LoSasso T, Zhang P, Hunt M, Mageras G, Tang G. Couch and multileaf collimator tracking: a clinical feasibility study for pancreas and liver treatment. *Med Phys*. 2020;47(10):4743-4757. doi:10.1002/mp.14438
66. Clover K, Oultram S, Adams C, Cross L, Findlay N, Ponman L. Disruption to radiation therapy sessions due to anxiety among patients receiving radiation therapy to the head and neck area can be predicted using patient self-report measures. *Psychooncology*. 2011;20(12):1334-1341. doi:10.1002/pon.1854
67. Molassiotis A, Rogers M. Symptom experience and regaining normality in the first year following a diagnosis of head and neck cancer: a qualitative longitudinal study. *Palliat Support Care*. 2012;10(3):197-204. doi:10.1017/S147895151200020X

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: Gardner M, Bouchta YB, Mylonas A, et al. Realistic CT data augmentation for accurate deep-learning based segmentation of head and neck tumors in kV images acquired during radiation therapy. *Med Phys*. 2023;50:4206–4219. https://doi.org/10.1002/mp.16388

C

Appendix C: CBCT-DRRs Superior to CT-DRRs for Target-Tracking Applications for Pancreatic SBRT

The markerless framework presented in this thesis uses a traditional digitally reconstructed radiograph (DRR) algorithm to generate the training data. The planning CT is forward projected using the DRR algorithm to create 2D stimulated intrafraction images. However, traditional DRR algorithms may not accurately represent intrafraction images. This appendix investigates DRR algorithms which incorporate physical modelling of the on-board imagers and utilises the daily CBCTs to improve DRR similarity to intrafraction images. This appendix has been published as "Madden, L., Ahmed, A., Stewart, M., Chrystall, D., **Mylonas, A.**, Brown, R., Nguyen, D.T., Keall, P. & Booth, J., CBCT-DRRs superior to CT-DRRs for target-tracking applications for pancreatic SBRT. *Biomedical Physics & Engineering Express*, **10**, 035039 (2024)."

Biomedical Physics & Engineering Express

PAPER

CBCT-DRRs superior to CT-DRRs for target-tracking applications for pancreatic SBRT

RECEIVED
16 February 2024

REVISED
14 March 2024

ACCEPTED FOR PUBLICATION
8 April 2024

PUBLISHED
26 April 2024

Levi Madden[1,2], **Abdella Ahmed**[1], **Maegan Stewart**[1,3], **Danielle Chrystall**[4], **Adam Mylonas**[5], **Ryan Brown**[1], **Doan Trang Nguyen**[5,6,7], **Paul Keall**[5] and **Jeremy Booth**[1,4]

[1] Northern Sydney Cancer Centre, Royal North Shore Hospital, St Leonards, NSW, 2065, Australia
[2] Centre for Medical Radiation Physics, Faculty of Engineering and Information Sciences, University of Wollongong, University of Wollongong, NSW, 2522, Australia
[3] School of Health Sciences, Faculty of Medicine and Health, University of Sydney, Camperdown, NSW, 2050, Australia
[4] School of Physics, Faculty of Science, University of Sydney, Camperdown, NSW, 2050, Australia
[5] ACRF Image X Institute, Faculty of Medicine and Health, University of Sydney, Eveleigh, NSW, 2015, Australia
[6] School of Biomedical Engineering, University of Technology Sydney, Ultimo, NSW, 2007, Australia
[7] SeeTreat Medical, Sydney, NSW, 2000, Australia

E-mail: Levi.Madden@health.nsw.gov.au

Keywords: radiographs, DRRs, inter-fraction variation, Cone Beam CT, markerless target-tracking

Supplementary material for this article is available online

Abstract

Objective. In current radiograph-based intra-fraction markerless target-tracking, digitally reconstructed radiographs (DRRs) from planning CTs (CT-DRRs) are often used to train deep learning models that extract information from the intra-fraction radiographs acquired during treatment. Traditional DRR algorithms were designed for patient alignment (*i.e.* bone matching) and may not replicate the radiographic image quality of intra-fraction radiographs at treatment. Hypothetically, generating DRRs from pre-treatment Cone-Beam CTs (CBCT-DRRs) with DRR algorithms incorporating physical modelling of on-board-imagers (OBIs) could improve the similarity between intra-fraction radiographs and DRRs by eliminating inter-fraction variation and reducing image-quality mismatches between radiographs and DRRs. In this study, we test the two hypotheses that intra-fraction radiographs are more similar to CBCT-DRRs than CT-DRRs, and that intra-fraction radiographs are more similar to DRRs from algorithms incorporating physical models of OBI components than DRRs from algorithms omitting these models.

Approach. DRRs were generated from CBCT and CT image sets collected from 20 patients undergoing pancreas stereotactic body radiotherapy. CBCT-DRRs and CT-DRRs were generated replicating the treatment position of patients and the OBI geometry during intra-fraction radiograph acquisition. To investigate whether the modelling of physical OBI components influenced radiograph-DRR similarity, four DRR algorithms were applied for the generation of CBCT-DRRs and CT-DRRs, incorporating and omitting different combinations of OBI component models. The four DRR algorithms were: a traditional DRR algorithm, a DRR algorithm with source-spectrum modelling, a DRR algorithm with source-spectrum and detector modelling, and a DRR algorithm with source-spectrum, detector and patient material modelling. Similarity between radiographs and matched DRRs was quantified using Pearson's correlation and Czekanowski's index, calculated on a per-image basis. Distributions of correlations and indexes were compared to test each of the hypotheses. Distribution differences were determined to be statistically significant when Wilcoxon's signed rank test and the Kolmogorov-Smirnov two sample test returned $p \leqslant 0.05$ for both tests.

Main results. Intra-fraction radiographs were more similar to CBCT-DRRs than CT-DRRs for both metrics across all algorithms, with all $p \leqslant 0.007$. Source-spectrum modelling improved radiograph-DRR similarity for both metrics, with all $p < 10^{-6}$. OBI detector modelling and patient material modelling did not influence radiograph-DRR similarity for either metric.

Significance. Generating DRRs from pre-treatment CBCT-DRRs is feasible, and incorporating CBCT-DRRs into markerless target-tracking methods may promote improved target-tracking accuracies. Incorporating source-spectrum modelling into a treatment planning system's DRR algorithms may reinforce the safe treatment of cancer patients by aiding in patient alignment.

1. Introduction

There exists significant demand for real-time motion-management in modern radiotherapy clinics. Magnetic resonance imaging linear accelerators address this demand, however their cost and scarcity render them inaccessible for many radiotherapy clinics. In a survey on the current patterns of practice in radiotherapy, 71 % of responding cancer centres wished to implement real-time motion management for at least one cancer site [1]. However, the greatest barrier to implementation of real-time motion management was a lack of widespread, commercially available real-time target-tracking technologies compatible with standard linear accelerators (linacs). Several real-time target-tracking methods compatible with standard linacs have been developed and clinically implemented in academic settings to overcome this barrier [2, 3]. Many of these methods track fiducial markers in intra-fraction radiographs during radiotherapy treatments [2]. However, marker implantation introduces additional risks, costs and treatment delays for patients. With regards to marker-based tracking, observed marker motions may not represent complex target motions and or deformations [4]. Alternatively, direct target-tracking methods that do not require fiducial markers (termed markerless target-tracking methods) can better represent such complex motions and or deformations whilst obviating the requirement for fiducial marker implantation.

Several different approaches to markerless target-tracking have been proposed for intra-fraction radiographs in the literature. Template-matching methods that align digitally reconstructed radiographs (DRRs) from planning CTs to intra-fraction radiographs have demonstrated successful 3D target-tracking capabilities during spinal and lung stereotactic body radiotherapy [5, 6]. However, template-matching methods have been reported to suffer reduced tracking accuracies when tumour visibility is poor [5, 6] and when DRRs have different image qualities to intra-fraction radiographs [7]. Deep-learning based markerless target-tracking models trained with DRRs from planning CTs have demonstrated promising results across a number of tumour sites in pre-clinical settings [8–10]. Similar to template-matching methods, deep-learning target-tracking models suffer reduced tracking accuracies when these models are trained on DRRs with dissimilar image qualities to the intra-fraction radiographs they are deployed on [11].

For template-matching and deep-learning based markerless target-tracking methods, it is common that DRRs are generated from planning CTs (CT-DRRs). Pre-treatment Cone-Beam CTs (CBCTs) may better represent intra-fractional anatomy than planning CTs, however, CBCTs have not been considered for the generation of DRRs for markerless target-tracking applications. In the scope of markerless target-tracking, incorporating DRRs generated from pre-treatment CBCTs (CBCT-DRRs) could also improve target-tracking by eliminating inter-fraction variation between DRRs and intra-fraction radiographs. Additionally, traditional DRR algorithms generate DRRs with dissimilar image qualities to intra-fraction radiographs. In the scope of markerless target-tracking, incorporating physical models of a linac's on-board-imager (OBI) into DRR algorithms should improve the modelling of radiographic image formation, and thereby improve target-tracking accuracies.

In this study, we investigate two hypotheses. Firstly, we investigate whether intra-fraction radiographs are more similar to CBCT-DRRs than planning CT-DRRs. Secondly, we investigate whether intra-fraction radiographs are more similar to DRRs generated by algorithms incorporating physical OBI component models than DRRs generated with algorithms omitting these component models. Specifically, we investigate whether source-spectrum modelling, detector modelling and patient material modelling affects radiograph-DRR similarity with statistical significance.

2. Methods

2.1. Data acquisition

Planning CT and daily CBCT images from 20 patients were used in this retrospective study. Patients had been treated on an ethics approved clinical trial for locally advanced pancreatic cancer patients receiving stereotactic body radiotherapy delivering 40 Gy in 5 fractions (ClinicalTrials.gov ID: NCT03505229). The characteristics of the patient cohort are reported in table A1. 19 of the 20 patients were treated in exhale and CBCT/CT-imaged during exhale; the remaining patient was treated in inhale and CBCT/CT-imaged during inhale. All planning CTs were contrast enhanced to improve the visibility of organ boundaries for contouring purposes. Contrast enhanced planning CTs were acquired with a Brilliance Big Bore CT scanner (Phillips, Netherlands) at a voxel size of $1.17 \times 1.17 \times 2.00\ mm^3$, with the same scanning parameters used for each patient's CT. For all planning CT acquisitions, the Real-time Position Management

(RPM) system (Varian Medical Systems, Palo Alto, USA) was used as an abdominal respiratory surrogate.

All patients were treated using a Truebeam linac (version 2.7, Varian Medical Systems, Palo Alto, USA). At the beginning of each treatment fraction, each patient received a pre-treatment CBCT acquired in half-fan beam geometry. All CBCTs were reconstructed using an iterative CBCT reconstruction algorithm with a voxel size of $0.908 \times 0.908 \times 2.00\ mm^3$. For all CBCT acquisitions, the Respiratory Gating for Scanners system (RGSC; Varian Medical Systems, Palo Alto, USA) was used as an abdominal respiratory surrogate. All CBCT scans throughout the study were acquired using the same scanning parameters for each patient.

During treatment, all patient's breathing amplitudes were monitored using the same Varian RGSC abdominal block as for their CBCT acquisition. Intra-fraction radiographs were acquired throughout each treatment session, triggered either at entry to the gating window or every 3 seconds. A total of 9307 planar radiographs were acquired throughout this study. All intra-fraction radiographs were acquired using a framegrabber [12] with a pixel size of $0.388 \times 0.388\ mm^2$ and with dimensions of 768×1024 pixels. Regarding the acquisition of intra-fraction radiographs, radiographic imaging fields were collimated to limit the exposure of healthy organs. All parameters related to radiographic image acquisition (*i.e.* projection angle, Source-Isocentre-Distance (SID), Source-Detector-Distance (SDD), detector pixel size, positions of source collimation and source potential) were recorded as header information with their intra-fraction radiographs.

2.2. DRR methods

All DRRs were generated by modelling the primary attenuation of an x-ray beam through CBCT and CT images. Four variations of the same ray-tracing algorithm were investigated, incorporating and omitting different combinations of physical OBI component models. Derivations of these four algorithms are detailed in appendix B. The four variations were:

1. A traditional DRR ray-tracing algorithm, *i.e.* modelling attenuation with a monoenergetic spectrum (referred to as Algorithm A, detailed in B.1).
2. A ray-tracing algorithm incorporating a physical model of the OBI source-spectrum, *i.e.* modelling attenuation with a polyenergetic spectrum (referred to as Algorithm AS, detailed in appendix B.2).
3. A ray-tracing algorithm incorporating physical models of the OBI source-spectrum and detector, *i.e.* Algorithm AS with detector modelling (referred to as Algorithm ASD, detailed in appendix B.3).
4. A ray-tracing algorithm incorporating physical models of the OBI source-spectrum and detector, as well as material modelling for CBCT and CT images, *i.e.* Algorithm ASD with patient material modelling (referred to as Algorithm ASDM, detailed in appendix B.4).

2.2.1. OBI component modelling

The Truebeam's OBI source was analytically modelled using the SpekPy package [13] with experimentally validated parameters from the literature [14]; these parameters are reported in table C1, appendix C. The bremsstrahlung x-ray source was assumed to be point-like with an isotropic fluence distribution given that imaging fields were typically collimated to small sizes [13]. No bowtie filter modelling was applied in this study as bowtie filtration was not used when acquiring intra-fraction radiographs. To best replicate each intra-fraction radiograph, the recorded source potential was used when modelling the x-ray source for the generation of corresponding DRRs.

The detector panel used in this work was a PaxScan4030 CB (Varex Imaging Corporation, USA), composed of a 0.6 mm thick CsI(Tl) crystal. Energy absorption in the crystal was analytically modelled following the experimentally validated model for this panel from FastCAT [14]. For the calculation of energy absorption, mass-energy attenuation coefficients were sourced from NIST [15]. The detector's anti-scatter grid was assumed to have a negligible effect on primary attenuation as the imaging fields were collimated to be small in size to promote healthy tissue sparing [14, 16]. The detector panel's point-spread-function was not modelled, assumed to have an insignificant effect on DRRs compared to the blurring introduced from ray-tracing through the comparatively large voxels of CBCT and CT images.

The CBCT scanner's and CT scanner's CT numbers were calibrated to relative electron densities through stoichiometric calibration [17] with an electron density phantom (CatPhan504, The Phantom Laboratory, USA). CBCT and CT images of the electron density phantom were acquired using the same CBCT and CT acquisition parameters as were used for the acquisition of patient data. Images of the electron density phantom were analysed to determine the CT numbers of the phantom's electron density plugs. The calibration curve for CT number to relative electron density was achieved by fitting a piecewise-linear model to data in the form of CT number versus relative electron density.

2.2.2. DRR generation

Prior to CBCT-DRR generation, each pre-treatment CBCT image was aligned to the treatment isocentre position using the corresponding couch shift applied at treatment. Prior to CT-DRR generation, each patient's planning CT image was aligned to the treatment isocentre position using the corresponding planned isocentre position from that patient's

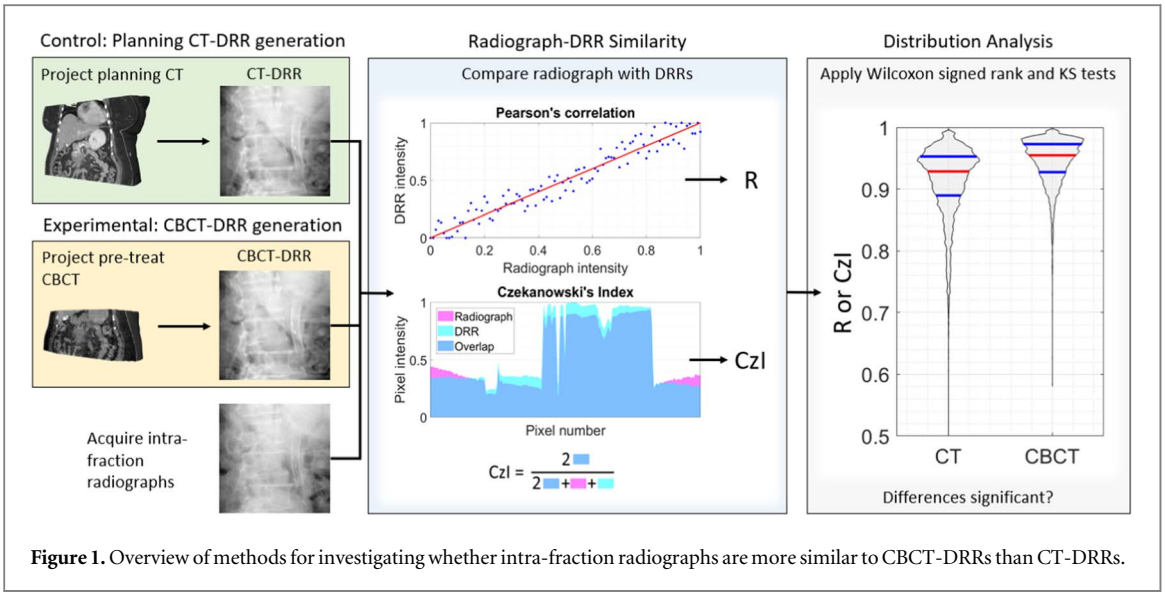


Figure 1. Overview of methods for investigating whether intra-fraction radiographs are more similar to CBCT-DRRs than CT-DRRs.

treatment plan. CBCT and CT images were upsampled to voxel sizes of $0.6 \times 0.6 \times 0.6\ \text{mm}^3$ using tri-cubic spline upsampling to mitigate systematic differences in radiograph-DRR similarity that may arise with differing voxel sizes of CBCT and CT images. The upsampled CBCT images and CT images were used for the generation of DRRs. OBI parameters such as projection angle, SDD and SID were used to replicate the projection geometry of each intra-fraction radiograph. The Reconstruction ToolKit's forward projection module [18] was applied for the calculation of water-equivalent path lengths through CBCT and CT images during DRR generation. All DRRs were generated with a pixel size of $0.388 \times 0.388\ \text{mm}^2$ and dimensions of 768×1024 pixels to match intra-fraction radiographs.

2.3. Radiograph-DRR comparison

Image similarity was assessed using Pearson's correlation (R) [19], it's associated probability of the null hypothesis (p_R), and Czekanowski's index (CzI) [20, 21]. Recorded OBI parameters for imaging field collimation were used to define regions of interest where the radiographs were uncollimated. Prior to analysis, intra-fraction radiographs and DRRs were cropped to the radiograph's uncollimated regions and min-max normalised so that each cropped radiograph's and DRR's minimum and maximum intensities were 0 and 1, respectively. Pearson's correlation was calculated between the image intensities of cropped radiographs and cropped DRRs on an image by image basis. Correlations between cropped radiographs and cropped DRRs were taken to be statistically significant when $p_R < 0.05$. Czekanowski's index, CzI, was calculated between a cropped radiograph and cropped DRR on an image by image basis:

$$\text{CzI} = \frac{2\sum_j \min\left(\text{Radiograph}_j, \text{DRR}_j\right)}{\sum_j\left(\text{Radiograph}_j + \text{DRR}_j\right)} \quad (1)$$

where Radiograph_j corresponds to the j^{th} pixel in a cropped radiograph and DRR_j corresponds to the j^{th} pixel in the cropped DRR. Czekanowski's index is defined such that $\text{CzI} = 0$ indicates no similarity, $\text{CzI} = 1$ indicates the radiograph and DRR are identical, and increasing CzI indicates increasing radiograph-DRR similarity. It should be noted that there exists no associated probability of the null hypothesis for Czekanowski's index.

It should be noted that Pearson's correlation and Czekanowski's index were chosen to assess radiograph similarity as they can be interpreted to meaningfully assess two independent aspects of radiograph-DRR similarity. Correlation traditionally assesses the goodness of fit between a linear model and data; calculated between pairs of matching pixels from radiographs and DRRs, this can be interpreted to assess the goodness of fit between the structures of projected anatomy in DRRs versus intra-fraction radiographs. Czekanowski's index assesses the similarity between pixel intensities from matching positions in radiographs and DRRs, indicative of how similar the radiograph and DRRs are in terms of brightness/darkness of radiographic features.

2.4. Distribution analysis

An overview of the workflow for testing whether intra-fraction radiographs are more similar to CBCT-DRRs than CT-DRRs is presented in figure 1. An overview of the workflow for testing whether source-spectrum modelling, detector modelling and patient material modelling affects radiograph-DRR similarity is presented in figure 2.

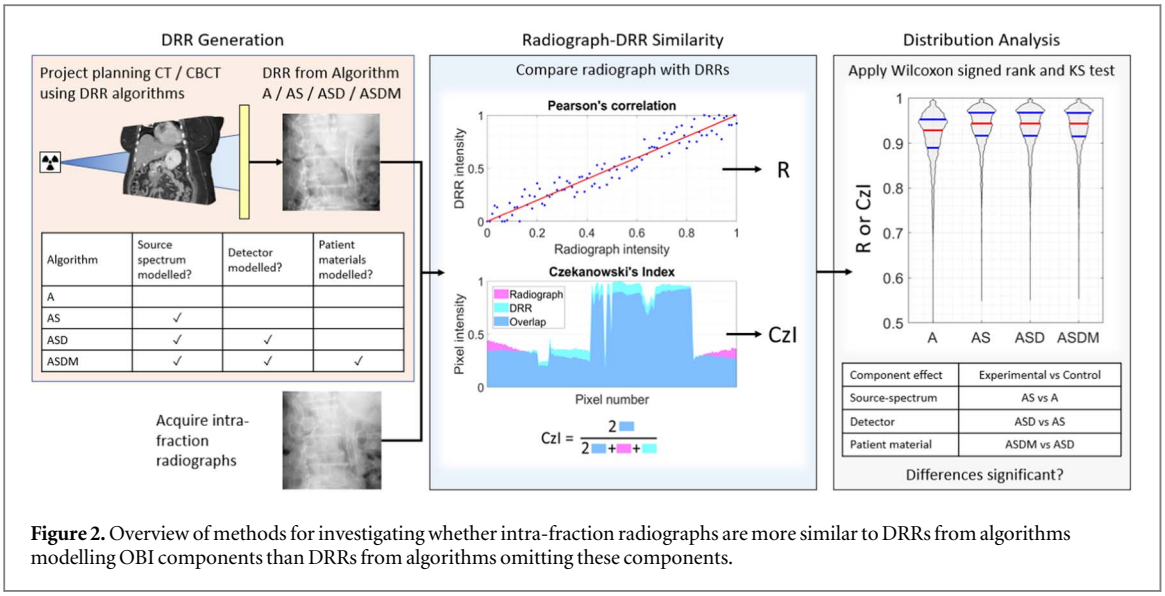


Figure 2. Overview of methods for investigating whether intra-fraction radiographs are more similar to DRRs from algorithms modelling OBI components than DRRs from algorithms omitting these components.

2.4.1. Pre-treatment CBCT versus planning CT

To determine whether intra-fraction radiographs were more similar to CBCT-DRRs than CT-DRRs, each CBCT-DRR distribution was compared with the CT-DRR distribution produced by the same algorithm by calculating differences between these distributions and taking the median difference (*e.g.* median of difference between CBCT-DRR from Algorithm A and CT-DRR from Algorithm A). To determine whether these median differences were statistically significant, Wilcoxon's signed rank test [22] and the Kolmogorov-Smirnov two sample test [23, 24] were applied to assess pairs of CBCT-DRR and CT-DRR distributions. Wilcoxon's signed rank test assessed the null hypothesis that the median difference between CBCT-DRR and CT-DRR distributions was 0, computing a probability that the calculated median difference had arisen randomly under this null hypothesis, p_W. The Kolmogorov-Smirnov two sample test assessed the null hypothesis that the CBCT-DRR and CT-DRR distributions were from the same population, computing a probability that differences in the distributions had arisen randomly, p_{KS}. Median differences were taken to be statistically significant when $p_W < 0.05$ and $p_{KS} < 0.05$.

2.4.2. Incorporation versus omission of OBI component models

To determine how each of the physical component models affected radiograph-DRR similarity, the same distributional comparisons were carried out between distributions from algorithms incorporating these components and distributions from algorithms omitting these components. The source-spectrum model's effect and its subsequent statistical significance was determined by calculating median differences, p_W and p_{KS} between distributions from Algorithm AS and from Algorithm A. Similarly, the detector model's effect and statistical significance was determined by comparing distributions from Algorithm ASD and Algorithm AS. Finally, patient material modelling's effect and statistical significance was determined by comparing distributions from Algorithm ASDM and Algorithm ASD.

2.5. Computational cost of incorporating OBI component models in DRR algorithms

To quantify the computational cost associated with incorporating each OBI component model in the DRR algorithm, the time required by each algorithm to generate the CT-DRRs from one randomly chosen fraction of data was recorded. This computational experiment was carried out in MATLAB (R2021b) using a workstation CPU. From the recorded distributions of times, the mean time per DRR and its standard deviation were calculated for each algorithm. The computational cost of incorporating source-spectrum modelling was determined by taking the difference between the mean times per DRR from algorithms AS and A. Similarly, the cost of incorporating detector modelling was quantified as the difference between mean times per DRR from algorithms ASD and AS, and the cost of incorporating patient material modelling was quantified as the difference between mean times per DRR from algorithms ASDM and ASD. Computational costs were determined to be insignificant when the aforementioned differences were less than the combined standard deviation from the two contributing algorithms.

3. Results

3.1. Stoichiometric calibration

Piecewise-linear models were fit to the measured CT numbers and given relative electron densities to calibrate the CBCT and CT scanners to relative electron density. Results from the stoichiometric calibration of CT number to relative electron density

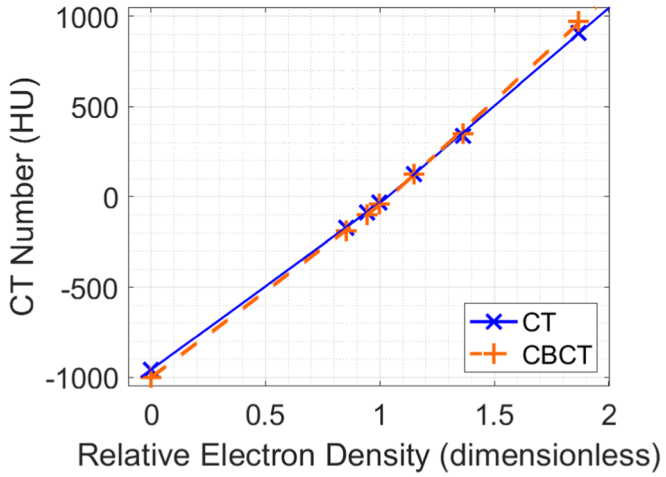


Figure 3. Results from the stoichiometric calibration of CT number to relative electron density. In the legend, the orange pluses are measured data points for CBCT calibration, the orange dashed-line correspond to the fitted CBCT calibration curve, the blue crosses are measured data points for CT calibration, and the blue line corresponds to the fitted CT calibration curve.

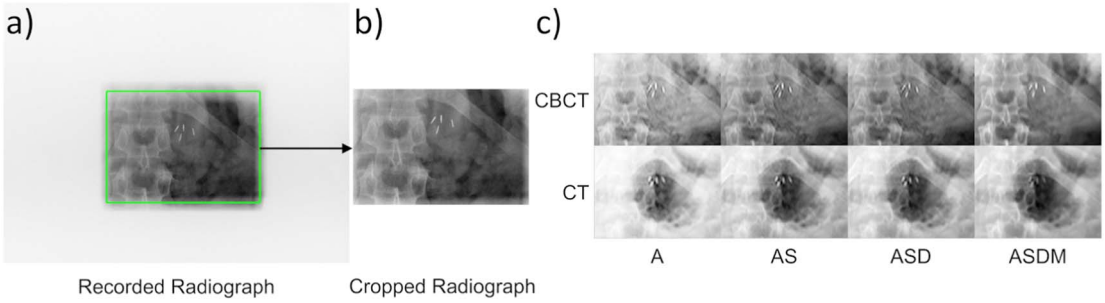


Figure 4. Examples of a recorded intra-fraction radiograph, its corresponding cropped radiograph and matching CBCT-DRRs and CT-DRRs. In (a) is the recorded intra-fraction radiograph with the cropping region of interest drawn in green, (b) the cropped radiograph, and (c) four cropped CBCT-DRRs and CT-DRRs are shown with CBCT-DRRs in the top row (labelled as CBCT) and CT-DRRs shown in the bottom row (labelled CT). With regards to columns in (c), A, AS, ASD and ASDM correspond to the DRRs from algorithms A, AS, ASD and ASDM, respectively.

are presented in figure 3. Each calibration curve was fit to the corresponding data with $R^2 \geqslant 0.999$. The presented calibration curves were applied only in the DRR algorithm incorporating patient material modelling, Algorithm ASDM.

3.2. DRR generation

An example of a recorded intra-fraction radiograph, cropped radiograph, and matched CBCT-DRRs and CT-DRRs are presented in figure 4. It should be noted that the four white structures near the centre of the radiographs and DRRs in figure 4 are gold fiducial markers. In figure 4 (c), the CBCT-DRRs present image qualities and anatomy that appears qualitatively similar to the matched intra-fraction radiograph. Conversely, for the example CT-DRRs in figure 4 (c), image qualities and anatomy appears different to the intra-fraction radiograph. We attribute the significant visual differences between CT-DRRs and intra-fraction radiographs to three contributing factors. First, there is a significant stomach gas-bubble present in the planning CT, visualised as the dark spot near the centre of the CT-DRRs in figure 4. However, this gas-bubble is absent in the pre-treatment CBCT and intra-fraction radiographs. Second, contrast agent was administered for the acquisition of the planning CT, but not for the acquisition of intra-fraction radiographs. Third, the CT scanner's couch is thicker than the linac's couch, increasing the water-equivalent path length along anterior-posterior projection angles (as in figure 4) and contributing to a reduction in image contrast relative to intra-fraction radiographs.

For further comparison of radiographs and DRRs, absolute pixel intensity differences were calculated between the cropped radiograph and matching cropped DRRs presented in figure 4. The resultant absolute difference images are presented in figure 5. In these difference images, black pixels show where radiograph and DRR pixel intensities are equal, and increasing brightness indicates increasing discrepancies between the radiograph and matched DRR. When comparing difference images from the same algorithm, CBCT-DRR difference images are visibly darker than corresponding CT-DRR difference

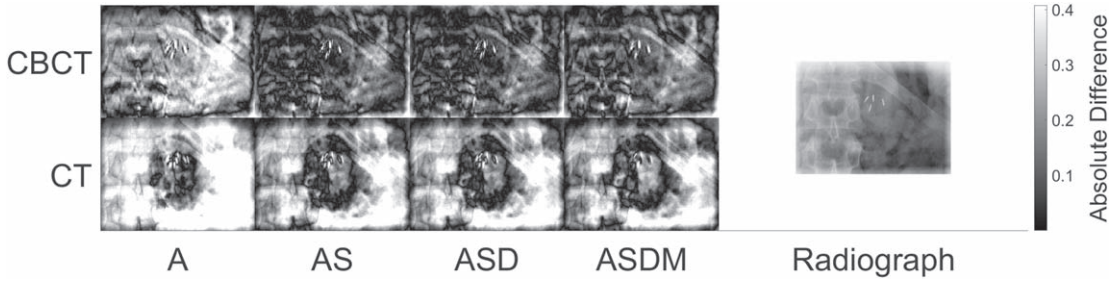


Figure 5. Absolute pixel intensity differences calculated between the cropped radiograph and matching cropped DRRs presented in figure 4. Right: The cropped radiograph from figure 4, shown as a reference for the difference images. Top row, left four columns: absolute differences between the radiograph and CBCT-DRRs. Bottom row, left four columns: absolute differences between the radiograph and CT-DRRs. The column labels A, AS, ASD and ASDM are defined as in figure 4. For visualisation of these absolute difference images, the colour bar limits were set to the 0th and 90th percentile value from the set of these eight difference images.

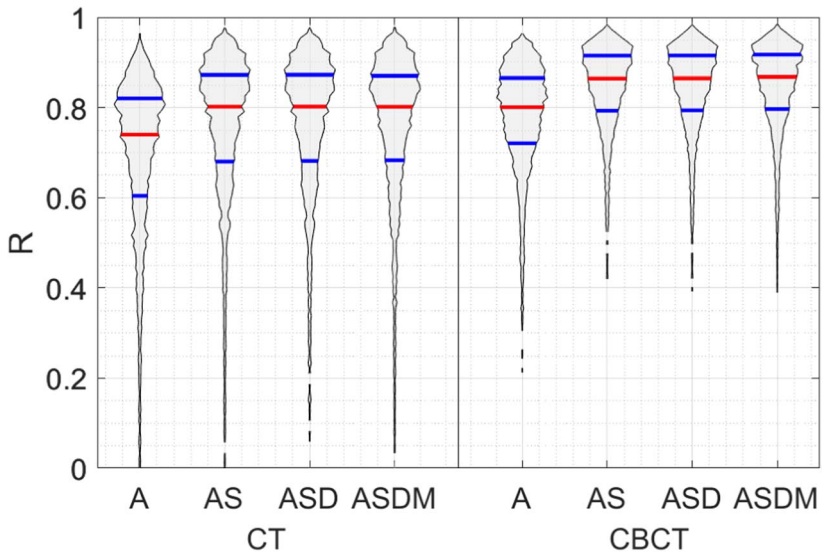


Figure 6. Distribution of correlations between intra-fraction radiographs and matching DRRs. Along the *x*-axis, A, AS, ASD and ASDM corresponds to distributions from algorithms A, AS, ASD and ASDM, respectively. CBCT and CT corresponds to CBCT-DRRs and CT-DRRs, respectively. In the violin plot, red lines correspond to the distribution's median, blue lines correspond to the distribution's 25th and 75th percentile values, and the width of the plot corresponds to the relative frequency in the distribution.

images. Comparing the CBCT-DRR difference images from different algorithms, the difference image from Algorithm A is visibly brighter than that of the other algorithms, with the difference images from algorithms AS, ASD and ASDM appearing similar; the same trend is observed for the difference images for CT-DRR difference images.

3.3. Distribution analysis

Distributions of correlations and indexes are presented in figures 6 and 7, respectively. Each of the 8×9307 radiograph-DRR correlation coefficients were statistically significant with all $p_R \leqslant 0.001$. For each of the distributions in figure 6, there are a number of data-points with poor, yet statistically significant correlations, attributed to arise from two main sources. First, mismatches in image quality (*e.g.* when radiographs were noisy, or when DRRs did not replicate the image features of radiographs due to a lack of scatter modelling) were responsible for approximately 35 % of all the correlations with $R < 0.4$ in each CBCT-DRR distribution and more than 60 % of correlations with $R < 0.4$ in each CT-DRR distribution (refer to figures S1-S4 in the Supplementary Material). It should be noted that mismatches between radiographs and DRRs were frequent for one specific patient; from their 5 treatment fractions, more than 25 % of all the correlations with $R < 0.4$ had occurred for CBCT-DRR distributions, and more than 50 % for all CT-DRR distributions (refer to figures S1-S2 in the Supplementary Material). Second, 6 of the 20 patients radiographs were acquired using time-based triggering, with all 6 patients imaged and treated in breath-hold. Across these 6 patients treatment fractions, there were a number of radiographs acquired between breath-holds, causing systematic variations between the anatomy of intra-fraction radiographs and the anatomy of DRRs (refer to figures S5-S8 in the

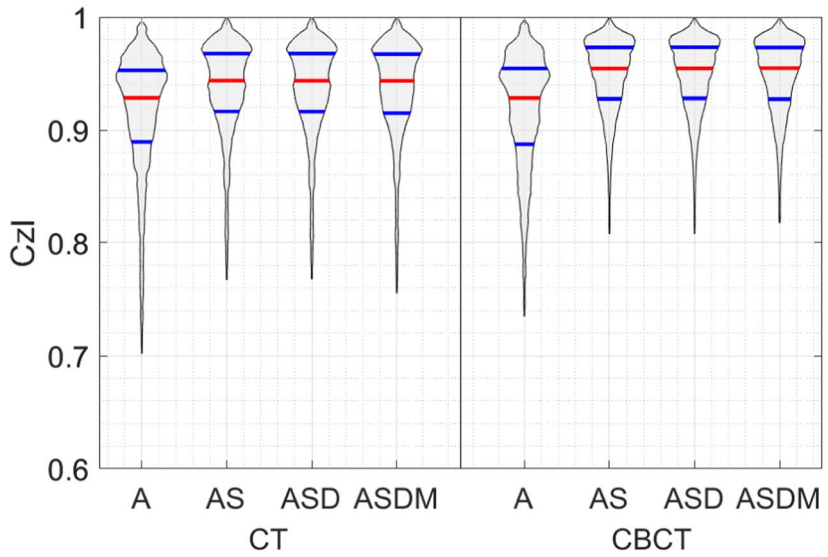


Figure 7. Distribution of Czekanowski's index between intra-fraction radiographs and matching DRRs. A, AS, ASD, ASDM, CT and CBCT are defined as in figure 6. Red lines correspond to the distribution median, blue lines correspond to the distribution's 25th and 75th percentile values, and the width of the plot corresponds to the relative frequency in the distribution.

Table 1. Table reporting impact of pre-treatment CBCT vs planning CT (top) and impact of incorporation versus omission of each component model (bottom). In the top row, Experimental and Control list combinations of distributions analysed to test each of the hypotheses, MD corresponds to median difference between Experimental and Control distributions, and p_W and p_{KS} are probabilities of observing the null hypothesis between the Experimental and Control distributions for the Wilcoxon and Kolmogorov-Smirnov tests, respectively. In the leftmost two columns, A, AS, ASD, ASDM, CT and CBCT defined as in figure 6.

Distribution analysis							
Experimental	Control	R			CzI		
		MD	p_W	p_{KS}	MD	p_W	p_{KS}
CBCT, A	CT, A	0.056	$<10^{-6}$	$<10^{-6}$	0.002	$<10^{-6}$	0.007
CBCT, AS	CT, AS	0.053	$<10^{-6}$	$<10^{-6}$	0.007	$<10^{-6}$	$<10^{-6}$
CBCT, ASD	CT, ASD	0.053	$<10^{-6}$	$<10^{-6}$	0.007	$<10^{-6}$	$<10^{-6}$
CBCT, ASDM	CT, ASDM	0.056	$<10^{-6}$	$<10^{-6}$	0.007	$<10^{-6}$	$<10^{-6}$
CT, AS	CT, A	0.047	$<10^{-6}$	$<10^{-6}$	0.021	$<10^{-6}$	$<10^{-6}$
CBCT, AS	CBCT, A	0.048	$<10^{-6}$	$<10^{-6}$	0.028	$<10^{-6}$	$<10^{-6}$
CT, ASD	CT, AS	0	$<10^{-6}$	>0.999	0	$<10^{-6}$	>0.999
CBCT, ASD	CBCT, AS	0.001	$<10^{-6}$	0.999	0	$<10^{-6}$	>0.999
CT, ASDM	CT, ASD	−0.002	$<10^{-6}$	0.727	0	$<10^{-6}$	0.419
CBCT, ASDM	CBCT, ASD	0.002	$<10^{-6}$	0.166	0	$<10^{-6}$	0.936

Supplementary Material). From these 6 patients, intra-fraction motions were responsible for approximately 60 % of all correlations with $R < 0.4$ in each CBCT-DRR distribution and approximately 30 % in each CT-DRR distribution.

3.3.1. Pre-treatment CBCT versus planning CT

From the distributional analysis reported in table 1, intra-fraction radiographs were more strongly correlated with CBCT-DRRs than CT-DRRs for all 4 algorithms with all corresponding median differences statistically significant (all $p_W < 10^{-6}$ and all $p_{KS} < 10^{-6}$). Similarly, from the distributional analysis of Czekanowski's index distributions, intra-fraction radiographs were more similar to CBCT-DRRs than CT-DRRs for all 4 algorithms; all corresponding median differences were statistically significant with all $p_W < 10^{-6}$ and all $p_{KS} \leqslant 0.007$. All presented results support that intra-fraction radiographs were more similar to CBCT-DRRs than CT-DRRs with statistical significance.

There are several factors that may contribute to the superior radiograph-DRR similarity of CBCT-DRRs over corresponding CT-DRRs. Firstly, contrast agent was administered for acquisition of the planning CT, but not the acquisitions of CBCTs or intra-fraction radiographs. Secondly, the CT scanner's couch is significantly thicker than the linac's couch, increasing the water-equivalent path lengths along anterior-posterior and oblique projections through the planning anatomy when compared to treatment anatomy. Third, the CT scanner uses an abdominal block that is

significantly different to that of the linac. Finally, inter-fraction variation is significant in abdominal tumour sites. Given that each of these factors cannot be decoupled from the dataset, it is currently not known how much each factor contributes to the differences in CBCT-DRR and CT-DRR radiograph-DRR similarity.

3.3.2. Incorporation versus omission of OBI component models

With regards to source-spectrum modelling, intra-fraction radiographs were more strongly correlated with DRRs from algorithm AS versus DRRs from algorithm A; corresponding median differences were statistically significant with both $p_W < 10^{-6}$ and both $p_{KS} < 10^{-6}$. Similarly, intra-fraction radiographs more similar to DRRs from algorithm AS than DRRs from algorithm A in terms of indexes; corresponding median differences were also statistically significant with both $p_W < 10^{-6}$ and both $p_{KS} < 10^{-6}$. These results support that the incorporation of source-spectrum modelling in DRR algorithms improved radiograph-DRR similarity with statistical significance.

With regards to detector modelling, there were no statistically significant differences in radiograph-DRR correlation between DRRs from algorithm ASD and algorithm AS, with both $p_{KS} \geqslant 0.999$. Similarly, there were no statistically significant differences in indexes between DRRs from algorithm ASD and algorithm AS, with both $p_{KS} > 0.999$. From these distributional analyses, the evidence suggests that the incorporation of detector modelling did not affect radiograph-DRR similarity.

With regards to patient material modelling, differences between radiograph-DRR correlation from algorithm ASDM and algorithm ASD were not statistically significant with both $p_{KS} \geqslant 0.166$. Similarly, there were no statistically significant differences between radiograph-DRR indexes from algorithm ASDM and algorithm ASD, with both $p_{KS} \geqslant 0.419$. From these distributional analyses, the evidence suggests that the incorporation of patient material modelling did not affect radiograph-DRR similarity.

3.4. Computational cost of incorporating OBI component models in DRR algorithms

The mean times required by each DRR algorithm to generate a DRR and their associated standard deviations are presented in table 2. Incorporating source-spectrum modelling into the DRR algorithm increased the computational time required to generate a DRR by (8 ± 16)%, deemed statistically insignificant. Incorporating detector modelling into the DRR generation algorithms did not cause significant increases in computation time. However, incorporating patient material modelling into the DRR algorithm increased the computation time required to generate a DRR by approximately (130 ± 15)%.

Table 2. Table reporting the time required by each DRR algorithm to generate one DRR. Algorithm A, AS, ASD and ASDM are as defined in figure 6. In the middle and right columns, Mean corresponds to the mean time required to produce a DRR, and St. Dev. corresponds to the standard deviation of the mean time required to produce a DRR.

Computation time comparison		
Algorithm	Mean (s/DRR)	St. Dev. (s/DRR)
A	2.5	0.4
AS	2.7	<0.1
ASD	2.7	0.1
ASDM	6.1	0.4

4. Discussion

4.1. Key findings

This study aimed to test the two hypotheses: that intra-fraction radiographs were more similar to CBCT-DRRs than CT-DRRs, and that intra-fraction radiographs were more similar to DRRs from algorithms incorporating physical modelling of OBI components than DRRs from traditional DRR algorithms omitting these models. From distributional analysis, intra-fraction radiographs were determined to be more similar to CBCT-DRRs than CT-DRRs, statistically significant for both similarity metrics. Similarly, intra-fraction radiographs were determined to be more similar to DRRs from algorithms incorporating source-spectrum modelling versus DRRs from algorithms omitting this modelling, also statistically significant for both similarity metrics. However, it was determined that OBI detector modelling and patient material modelling had a statistically insignificant effect on radiograph-DRR similarity. The study as a whole is significant as it not only demonstrates that generating CBCT-DRRs is feasible, but these CBCT-DRRs may have superior radiograph-DRR similarity over planning CT-DRRs. Additionally, the study is significant as it informs end users and developers of DRR algorithms of the benefits and requirements for incorporating OBI component modelling in DRR algorithms, *e.g.* incorporating source-spectrum modelling into a treatment planning systems DRR algorithm could clinically aid in patient alignment by improving generated the DRRs radiographic image quality.

4.2. Strengths and limitations

This study has a number of strengths. First, the set of intra-fraction radiographs was comprised of 9307 radiographs, providing sufficient statistical power when assessing the two hypotheses. Second, two unrelated radiograph-DRR similarity metrics were employed to assess radiograph-DRR similarity, and two unrelated statistical tests were used to independently assess whether observed distributional differences were statistically significant. Third, this study was able to investigate two hypotheses over a range of controlled conditions to determine whether effects were dependent on controlled variables, *e.g.* whether

CBCT-DRRs had superior radiograph-DRR similarity for some algorithms, or whether the modelling of individual OBI components caused significant effects for CT-DRRs but insignificant effects for CBCT-DRRs (or vice versa). Finally, no such investigations of radiograph-DRR similarity have been published in the literature for radiograph-DRR similarity of CBCT-DRRs versus CT-DRRs.

One important caveat in this study was that radiograph-DRR similarity was investigated for only the pancreas tumour site, with all radiographs collected using the same OBI and all patient CTs acquired with the same CT scanner. For other tumour sites, it is unknown how radiograph-DRR similarity will be impacted by inter-fraction variation and the incorporation versus omission of physical OBI component models into DRR algorithms. Hypothetically, for tumour sites where material densities vary widely from water (*e.g.* thorax or head and neck), material modelling could impact radiograph-DRR similarity more significantly than was observed in this study. For different OBIs and CT scanners, the impact and significance of each physical OBI component model in DRR algorithms may vary from the results obtained in this study. In cases where planning CT acquisitions use different immobilisation devices, patient support devices or abdominal compression different from those used during treatment, it is expected that inter-fraction variation will be influenced by these systematic setup differences. With these points in mind, we recommend that radiograph-DRR similarity and markerless target-tracking accuracies should be assessed to determine the combination of CBCT-DRR versus CT-DRR and DRR algorithm that optimises markerless target-tracking accuracy.

A second caveat in this study are the systematic, non-anatomical differences between planning CTs versus pre-treatment CBCTs and intra-fraction radiographs. All planning CTs were acquired with contrast enhancement, with a different couch to that of the linac, and with an abdominal block shaped differently to that used during treatment. Conversely, pre-treatment CBCTs and intra-fraction radiographs were acquired without contrast enhancement, and with the same couch and abdominal block. These differences likely reduce the correlations and indexes of CT-DRRs, and it is unclear to what degree any of the improvements in radiograph-DRR similarity are related to inter-fraction variation. For clinical cases where systematic, non-anatomical differences do not exist, we recommend that radiograph-DRR similarity and markerless target-tracking accuracies should be assessed for CBCT-DRRs and CT-DRRs to determine which DRRs are more beneficial for markerless target-tracking.

A final caveat in this study regards the implementation of the traditional DRR algorithm. In Algorithm A, the effective attenuation coefficient was calculated at an effective energy equal to one third of the OBI source's known potential [25]. For the example radiograph and DRRs shown in figure 4, it was shown that the one-third rule was a poor estimate of the source's effective energy. In practice, it is plausible that end users of traditional DRR algorithms could tune the effective attenuation coefficient to optimise radiograph-DRR similarity for markerless target-tracking purposes. In this case, it is not clear whether the statistical significance of incorporation versus omission of physical OBI source-spectrum modelling into the DRR algorithm will translate to other end users of traditional DRR algorithms. With the study's results in mind, we recommend that end users incorporating traditional DRR algorithms with tuned attenuation coefficients assess radiograph-DRR similarity and markerless target-tracking accuracy with respect to DRR algorithms incorporating physical modelling of OBI component.

4.3. Comparisons with prior studies

With regards to the implications of findings from this study, the studies by Dhont *et al* [7] and Unberath *et al* [11] provides evidence that supports that markerless target-tracking accuracies are improved when these target-tracking methods incorporated DRRs with superior radiograph-DRR similarity [7, 11]. Additionally, contrapositive results have been reported for clinical implementations of markerless target-tracking methods, where reduced target-tracking accuracies were observed when radiographic image quality degraded [3, 5]. With these studies in mind, there is mounting evidence to support that the incorporation of CBCT-DRRs into markerless target-tracking methods could benefit these methods accuracies. Future studies investigating the impact of inter-fraction variation on target-tracking accuracy are required to verify this hypotheses.

4.4. Implications and future work

4.4.1. CBCT-DRRs

The pancreas-SBRT protocol followed throughout this study aligns the patient such that the pancreas GTV is at isocentre. Intra-fraction radiographs are acquired with the centre of the OBI detector panel colinear with the OBI source and treatment isocentre. Consequently, the GTV, whole pancreas and proximal dose-limiting OARs (duodenum, small bowel and stomach) are frequently visualised in the intra-fraction radiographs and matched DRRs, supporting that intra-fraction variation contributes significantly to the improved radiograph-DRR similarity of CBCT-DRRs over CT-DRRs.

CBCT artifacts should be considered when evaluating whether CBCT-DRRs should be incorporated into target-tracking methods as the artifact-ridden CBCTs will produce artifact-ridden CBCT-DRRs. It should be noted that the CBCT images acquired during this study did not require artifact corrections, and so the findings of this study are specific to the case

when CBCTs are not affected by significant artifacts *e.g.* metal and/or gas artifact. A number of strategies could be applied to post-correct CBCT artifacts. Manual correction of affected voxel CT numbers or utilisation of alternate image reconstruction algorithms could be viable means of artifact reduction. Transforming CBCTs into synthetic CTs may also mitigate CBCT artifacts. Currently, superior CBCT reconstruction algorithms that reduce the severity and incidence of CBCT artifacts are being developed and rolled out by vendors [26–28]. We recommend that target-tracking practitioners evaluate whether CBCT-DRRs benefit their target-tracking algorithms and whether artifact mitigation is required, effective and reproducible when incorporating CBCT-DRRs.

Extrapolating from the result that intra-fraction radiographs were more similar to DRRs from pre-treatment CBCTs than DRRs from planning CTs, there are additional hypothetical implications for real-time target-tracking methods. First, 4D-CBCT can model a patient's anatomy driven by respiratory induced intra-fraction motions. Hypothetically, intra-fraction radiographs may be more similar to DRRs from 4D-CBCTs than DRRs from 3D-CBCTs. Second, for image-guided radiotherapy treatments that acquire a CBCT during treatments for patient position correction, these mid-treatment CBCTs provide an updated model of intra-fractional anatomy. Hypothetically, intra-fraction radiographs acquired post-position correction may be more similar to DRRs from the mid-treatment CBCT than DRRs from the pre-treatment CBCT. Future studies should investigate these opportunities for further improved radiograph-DRR similarity and the subsequent results of incorporating such DRRs into markerless target-tracking methods.

A barrier that may prohibit CBCT-DRR generation are the limited fields of view (FOVs) that are associated with pre-treatment CBCTs. For CBCT-DRRs to present adequate anatomy, the CBCT's FOV diameter must encompass the entire patient cross-section in each transverse slice. For CBCT-DRR generation, the CBCT must also span an adequate length along the patient's superior-inferior axis to replicate intra-fraction radiographs. When either of these conditions are not met, the anatomy presented in DRRs would systematically differ from the anatomy presented in intra-fraction radiographs. With retrospective datasets, limited FOVs may prohibit the generation of CBCT-DRRs.

4.4.2. Incorporating OBI modelling into DRR algorithms

In this study, the incorporation of source-spectrum modelling was demonstrated to significantly improve radiograph-DRR similarity. Incorporating source-spectrum modelling could benefit image registration applications of DRR algorithms given the increased radiograph-DRR similarity of algorithm AS over algorithm A. With regards to incorporating source-spectrum modelling into DRR algorithms, there exists experimentally validated open source software for modelling x-ray bremsstrahlung spectra [13]; the parameters required by this software are available from radiograph metadata (source kVp) and technical specifications (anode material, anode angle, inherent filter materials and their thicknesses). From the computational cost study, it was demonstrated that the incorporation of source-spectrum modelling into DRR algorithms did not significantly increase the computation of time for DRRs. With these points in mind, source-spectrum modelling could be incorporated into vendor-implemented DRR algorithms with minimal computational or resourcing costs, and may aid in patient alignment.

Cost functions for radiograph-DRR template-matching algorithms are relatively robust with respect to differences between radiographs and DRRs. However, studies investigating template-matching algorithms have reported that target-tracking algorithms suffer from reduced accuracies when radiograph-DRR similarity decreases [5–7] . Given this study's findings that DRR algorithm influenced radiograph-DRR similarity, direct evaluations of template-matching accuracy are required to determine if, when and to what degree template-matching accuracy depends on the DRR algorithm used to produce DRRs.

4.5. Further discussion

4.5.1. Clinical applications

Comparisons of 2D image metrics such as Pearson's correlation and Czekanowski's index cannot indicate whether target-tracking algorithms and models will benefit from the incorporation of CBCT-DRRs or physical OBI component modelling into DRR algorithms. It is intuitive that improved agreement between DRRs and intra-fraction radiographs should benefit the performance of deep-learning models by bridging the translation gap between training and testing data, however the degree of improvement can only be determined through direct evaluations of target-tracking performance with different training datasets. Given the vast range of setups used to treat different tumour sites, we recommend that direct evaluation be performed by end users to determine whether their target-tracking algorithm benefits from incorporation of CBCT-DRRs or physical OBI component modelling into DRR algorithms.

The CBCT and CT data acquired throughout this study were acquired exclusively in breath-hold or gated-free-breathing. However, real-time tracking applications are also relevant for patients imaged and treated in free-breathing without gating. In such cases, CT anatomy becomes blurred, and CBCTs in particular can present significant motion artifacts. For such motion-affected CTs and CBCTs, it is inappropriate to consider generating DRRs from 3D image sets in which there is poor anatomical representation of structures.

Real-time tracking may be considered for motion management during free-breathing treatments for thoracic and abdominal tumour sites. Respiratory motion is relevant for these tumour sites, and likely reduces radiograph-DRR similarity. From the result that intra-fraction radiographs were more similar to CBCT-DRRs than CT-DRRs, it is plausible that radiographs acquired during free-breathing may be more similar to DRRs generated from 4DCBCTs or 4DCTs at the relevant phase than DRRs generated from corresponding averaged 3DCBCTs or 3DCTs. Direct target-tracking evaluations are required to determine the best DRR generation practices for real-time tracking applications in the scope of free-breathing treatments.

Manual correction of the CT numbers of voxels affected by contrast agent could be used to correct the systematic difference caused by contrast agent, and thereby improve the radiograph-DRR similarity of these CT-DRRs. However, this process could introduce artificial structures into training DRRs that arent anatomically plausible or present in intra-fraction radiographs. From a clinical standpoint, end users should consider generating DRRS from CBCTs and CTs that were acquired without contrast to prevent DL models from learning artificial features that will be absent during their deployment on intra-fraction radiographs.

4.5.2. Generating DRRs from noisy images

A derivation of the relationship between the signal-noise ratio (SNR) of voxels in the CBCT and pixels in the DRR is detailed in appendix D. A simple example was considered, and it was shown that the SNR of CBCT-DRRs was superior to the SNR of CBCT voxels. This result arises as some of each voxel's noise cancels out with the other voxel's noise when ray-tracing through CBCT images. This theoretical result suggests that CBCT noise is mitigated significantly when generating CBCT-DRRs, and supports that CBCT-DRRs can present adequate visualisations of patient anatomies for target-tracking applications.

5. Conclusions

In this study, it was determined that intra-fraction radiographs were more similar to CBCT-DRRs than CT-DRRs, and also that the incorporation of physical source-spectrum modelling in DRR algorithms improved radiograph-DRR similarity. However, OBI detector and patient material modelling were determined to have insignificant impacts on radiograph-DRR similarity. The findings of this study suggests that incorporating CBCT-DRRs into markerless target-tracking methods may promote improved target-tracking accuracies, and that the incorporation of OBI source-spectrum modelling in treatment planning system DRR algorithms may reinforce the safe treatment of cancer patients by aiding in patient alignment.

Acknowledgments

The SPAN-C-SBRT (Trial ID: NCT03505229) is funded by Varian Medical Systems. This research was partially funded by a Cancer Institute NSW Translational Program Grant. AM receives scholarship from the University of Sydney Postgraduate Award and the Cancer Institute NSW Translational Program Supplementary Scholarship. PK is funded by an Australasian Government NHMRC Investigator Grant. We would like to thank Jericho O'Connell and Magdalena Bazalova-Carter for their guidance in implementing FastCAT CBCT simulator's source and detector models. We would like to thank the many staff from Northern Sydney Cancer Centre and The Image X Institute who supported this research.

Data availability statement

The data cannot be made publicly available upon publication because they contain sensitive personal information. The data that support the findings of this study are available upon reasonable request from the authors.

Appendix A. Patient data characteristics

Table A1. Table reporting the patient cohort characteristics. All patients were locally advanced pancreatic cancer patients with either borderline-resectable or unresectable gradings. No. patients corresponds to the number of patients in the cohort. With regards to gating, EBH denotes patients imaged and treated in Exhale-Breath-Hold, FB corresponds to patients imaged during exhale and treated in Free-Breathing gated-at-exhale, and IBH denotes patients imaged and treated in Inhale-Breath-Hold. With regards to the Diagnosis, Body, Endocrine, Head, Tail and Unspecified corresponded to patients presenting malignant neoplasms of the pancreas body, pancreas endocrine, pancreas head, pancreas tail, or an unspecified portion of the pancreas, respectively.

Cohort characteristics		
Characteristic		Quantity
Age (years)	Mean [Range]	68.1 [44 − 83]
GTV volume (cm^3)	Mean [Range]	25.7 [3.8 − 79.8]
Gender (No. patients)	Female	10
	Male	10
Gating (No. patients)	EBH	15
	FB	4
	IBH	1
Diagnosis (No. patients)	Body	1
	Endocrine	1
	Head	7
	Tail	1
	Unspecified	10

Appendix B. Derivation of DRR calculations

A point-like bremsstrahlung x-ray source at position, $\mathbf{r}_0$, with energy spectrum, $\mathbf{E}$, has fluence spectrum described by $\Phi(\mathbf{r}_0, \mathbf{E})$. A set of N pixels from a detector panel are located at $(\mathbf{r}_1, \mathbf{r}_2, \ldots, \mathbf{r}_N)$. There exists some arbitrary distribution of materials between the source and detector whose material's linear attenuation coefficients (LACs) are distributed according $\mu(\mathbf{r}, \mathbf{E})$. When this distribution of materials is imaged by a CT scanner and CBCT scanner, the material's CT numbers are distributed according to $C(\mathbf{r})$. The fluence of the attenuated x-ray beam incident upon the pixel, $\mathbf{r}_i$, at energy E, $\Phi(\mathrm{E}, \mathbf{r}_i)$, is given by:

$$\Phi(\mathbf{r}_i, \mathrm{E}) = \frac{\Phi(\mathbf{r}_0, \mathrm{E}) \exp\left(-\int_{\mathbf{r}_0}^{\mathbf{r}_i} \mu(\mathbf{r}, \mathrm{E}) d\mathbf{r}\right)}{||\mathbf{r}_i - \mathbf{r}_0||^2} \quad \text{(B1)}$$

where $\mathbf{r}$ corresponds to locations along the line segment between $\mathbf{r}_0$ and $\mathbf{r}_i$.

B.1. Traditional DRR algorithm: Algorithm A

In this case, only the bremsstrahlung source's potential and distribution of CT numbers are known. The bremsstrahlung source outputs x-rays with intensity $\mathrm{I}(\mathbf{r}_0)$:

$$\mathrm{I}(\mathbf{r}_0) = \int_0^{\mathrm{E}_{\max}} \Phi(\mathbf{r}_0, \mathrm{E}) d\mathrm{E} \quad \text{(B2)}$$

Assuming the bremsstrahlung spectrum has an effective energy, $\mathrm{E}_{\mathrm{eff}}$ one third the source potential [25], equation (B1) simplifies to:

$$\mathrm{I}(\mathbf{r}_i) = \frac{\mathrm{I}(\mathbf{r}_0) \exp\left(-\int_{\mathbf{r}_0}^{\mathbf{r}_i} \mu(\mathbf{r}, \mathrm{E}_{\mathrm{eff}}) d\mathbf{r}\right)}{||\mathbf{r}_i - \mathbf{r}_0||^2} \quad \text{(B3)}$$

where $\mathrm{I}(\mathbf{r}_1)$ is the x-ray intensity incident upon the detector pixel at $\mathbf{r}_i$, and $\mu(\mathbf{r}, \mathrm{E}_{\mathrm{eff}})$ is the material linear attenuation coefficient (LAC) at location, $\mathbf{r}$, for effective energy, $\mathrm{E}_{\mathrm{eff}}$. Without knowledge of the calibration of the CT or CBCT scanner that imaged $C(\mathbf{r})$, the relationship between CT number and LAC, $\mu(\mathbf{r}, \mathrm{E})$, is approximated by:

$$\mu(\mathbf{r}, \mathrm{E}) \approx \mu_{\mathrm{water}}(\mathrm{E}) \left(\frac{C(\mathbf{r})}{1000} + 1\right) \quad \text{(B4)}$$

where $C(\mathbf{r})$ is in Hounsfield units. For this case, DRRs are calculated:

$$\mathrm{DRR}(\mathbf{r}_i) = \frac{\exp\left(-\mu_{\mathrm{water}}(\mathrm{E}_{\mathrm{eff}}) \int_{\mathbf{r}_0}^{\mathbf{r}_i} \left(\frac{C(\mathbf{r})}{1000} + 1\right) d\mathbf{r}\right)}{||\mathbf{r}_i - \mathbf{r}_0||^2} \quad \text{(B5)}$$

with the integral, $\int_{\mathbf{r}_0}^{\mathbf{r}_i} \left(\frac{C(\mathbf{r})}{1000} + 1\right) d\mathbf{r}$ calculated using the Reconstruction ToolKit's forward projection module [18].

B.2. Ray-tracing algorithm with source-spectrum modelling: Algorithm AS

In this case, the fluence spectrum, $\Phi(\mathbf{r}_0, \mathbf{E})$ is known. The fluence spectrum incident upon the detector pixel $\mathbf{r}_i$ for x-ray energy E, $\Phi(\mathbf{r}_i, \mathrm{E})$, is then given by:

$$\Phi(\mathbf{r}_i, \mathrm{E}) = \frac{\Phi(\mathbf{r}_0, \mathrm{E}) \exp\left(-\mu_{\mathrm{water}}(\mathrm{E}) \int_{\mathbf{r}_0}^{\mathbf{r}_i} \left(\frac{C(\mathbf{r})}{1000} + 1\right) d\mathbf{r}\right)}{||\mathbf{r}_i - \mathbf{r}_0||^2} \quad \text{(B6)}$$

With bremsstrahlung source spectrum known, DRRs are calculated:

$$\mathrm{DRR}(\mathbf{r}_i) = \frac{\int_0^{\mathrm{E}_{\max}} \Phi(\mathbf{r}_i, \mathrm{E}) d\mathrm{E}}{\mathrm{I}(\mathbf{r}_0)} \quad \text{(B7)}$$

where $\mathrm{E}_{\max}$ is the maximum x-ray energy. As for the traditional DRR ray-tracing algorithm, the integral $\int_{\mathbf{r}_0}^{\mathbf{r}_i} \left(\frac{C(\mathbf{r})}{1000} + 1\right) d\mathbf{r}$ was calculated using the Reconstruction ToolKit's forward projection module [18]. *N.B.* Given the energy independence of this integral, the CT and CBCT image need only be ray-traced through once.

B.3. Ray-tracing algorithm with source-spectrum and detector modelling: Algorithm ASD

The fluence incident upon detector pixel, $\mathbf{r}_i$ described by equation (B6) remains valid. Radiographic images are formed through energy depositions in detector crystals. For a detector crystal with thickness, l, the energy deposited by an x-ray with energy, E, into the pixel at $\mathbf{r}_i$, is given by:

$$\mathrm{E}_{\mathrm{dep}}(\mathbf{r}_i, \mathrm{E}) = \mathrm{E}\left[1 - \exp\left(-\mu_{\mathrm{en}}(\mathrm{E}) \frac{l}{\cos(\theta)}\right)\right] \quad \text{(B8)}$$

where $\mu_{\mathrm{en}}(\mathrm{E})$ is the linear energy absorption coefficient for the crystal at energy, E, and θ is the angle of incidence for the ray drawn from $\mathbf{r}_0$ to $\mathbf{r}_i$ onto the detector pixel. With source and detector modelled, DRRs are calculated:

$$\mathrm{DRR}(\mathbf{r}_i) = \frac{\int_0^{\mathrm{E}_{\max}} \Phi(\mathbf{r}_i, \mathrm{E}) \mathrm{E}_{\mathrm{dep}}(\mathbf{r}_i, \mathrm{E}) d\mathrm{E}}{\mathrm{I}(\mathbf{r}_0)} \quad \text{(B9)}$$

B.4. Ray-tracing algorithm with source-spectrum, detector and patient material modelling: Algorithm ASDM

A material's energy-dependent attenuation coefficients have complex dependences on x-ray energy and material composition [29]. When the Compton effect is the dominant interaction process in an unknown material at a given x-ray energy, the unknown material's attenuation coefficient can be approximated using [30, 31]:

$$\mu_{\mathrm{water}}(\mathbf{r}, \mathrm{E}) \approxeq \mu_{\mathrm{water}}(\mathrm{E}) \rho_{\mathrm{e}}(\mathbf{r}) \quad \text{(B10)}$$

where E corresponds to the x-ray energy, $\mu_{\mathrm{water}}(\mathbf{r}, \mathrm{E})$ is the attenuation coefficient of the unknown material, $\mu_{\mathrm{water}}(\mathrm{E})$ is the attenuation coefficient of water at energy E, and $\rho_{\mathrm{e}}(\mathbf{r})$ is the relative electron density of the unknown material [30, 31]. For soft tissues, the Compton effect is expected to dominate x-ray interactions above 30keV [32]. For bony materials, the Compton effect dominates x-ray interactions above energies of approximately 50 keV [32]. For such bony materials, attenuation coefficient may be better approximated using:

$$\mu_{\text{bone}}(\mathbf{r},\ \text{E}) \approx \mu_{\text{bone}}(\text{E})\frac{\rho_{\text{e}(\mathbf{r})}}{\rho_{\text{e}_{\text{bone}}}} \tag{B11}$$

where $\mu_{\text{bone}}(\mathbf{r},\ \text{E})$ is the LAC of the unknown bony material, $\mu_{\text{bone}}(\text{E})$ is the attenuation coefficient of some reference bone material, and $\rho_{\text{e}_{\text{bone}}}$ is the known relative electron density of the reference bone material. It should be noted that a scanner's CT number can be calibrated to electron density [17]. With an appropriately calibrated CT scanner or CBCT scanning system, the CT number of a given voxel, $C(\mathbf{r})$, is converted to relative electron density, $\rho_{\text{e}}(\mathbf{r})$, using the derived CT number to relative electron density calibration curve. Incorporating CT and CBCT scanner system modelling, each voxel's LAC, $\mu(\mathbf{r}, \text{E})$ is given by:

$$\mu(\mathbf{r},\ \text{E}) = \begin{cases} \mu_{\text{water}}(\mathbf{r},\ \text{E}) & \text{if } C(\mathbf{r}) \leqslant \tau_1 \\ \mu_{\text{bone}}(\mathbf{r},\ \text{E}) & \text{if } C(\mathbf{r}) \geqslant \tau_2 \\ a\,\mu_{\text{water}}(\mathbf{r},\ \text{E}) + (1-a)\,\mu_{\text{bone}}(\mathbf{r},\ \text{E}) & \text{otherwise} \end{cases} \tag{B12}$$

where τ_1 and τ_2 are thresholds defining CT numbers in which materials are assumed to be comprised solely of water-like materials ($C(\mathbf{r}) \leqslant \tau_1$) and bone-like materials ($C(\mathbf{r}) \geqslant \tau_2$). For materials with CT numbers between τ_1 and τ_2, linear interpolation is used to approximate a mixture of water-like and bone-like materials, where a corresponds to the interpolation coefficient. In this work, the thresholds were set to $\tau_1 = 200$ HU and $\tau_2 = 600$ HU, matching those used by Acuros CTS [26, 33]. With OBI source and detector modelled, and the materials in CBCT and CT images modelled DRRs are calculated:

$$\text{DRR}(\mathbf{r}_i) = \int_0^{\text{E}_{\max}} \frac{\Phi(\mathbf{r}_0,\ \text{E})\exp\left(-\int_{\mathbf{r}_0}^{\mathbf{r}_i}\mu(\mathbf{r},\ \text{E})d\mathbf{r}\right)\text{E}_{\text{dep}}(\mathbf{r}_i,\ \text{E})d\text{E}}{\text{I}(\mathbf{r}_0)\|\mathbf{r}_i - \mathbf{r}_0\|^2} \tag{B13}$$

With the integral, $\int_{\mathbf{r}_0}^{\mathbf{r}_i}\mu(\mathbf{r},\ \text{E})d\mathbf{r}$ calculated using the Reconstruction ToolKit's forward projection module [18].

Appendix C. Spekpy source model parameters

Table C1. OBI source-spectrum model parameters for SpekPy [13] (from FastCAT CBCT simulator [14]).

Source model components			
Component	Material	Angle	Thickness
Anode	Tungsten	14°	—
Filter 1	Aluminium	—	2.70 mm
Filter 2	Titanium	—	0.89 mm
Filter 3	Air	—	900 mm
Source modelling parameters			
Source potential		Set to match radiograph	
Energy increment		1 keV	

Appendix D. Derivation of relationship between signal-to-noise ratio of CBCTs and their subsequent CBCT-DRRs

Let X_i be an independent Gaussian random variable corresponding to the measured attenuation coefficient of a CBCT image's i^{th} voxel. Let the true attenuation coefficient for this voxel be μ_i and the standard deviation for the measurement process be σ_i; *N.B.* μ_i corresponds to X_i's distribution mean, and σ_i corresponds to X_i's distribution standard deviation. The signal-noise ratio (SNR) of the CBCT's i^{th} voxel is then given by:

$$\text{SNR}_i = \frac{\mu_i}{\sigma_i} \tag{D1}$$

Ray-tracing through the CBCT image as during DRR generation, the water-equivalent path length along that ray is proportional to the linear combination of the measured attenuation coefficients from the voxels intercepted by the ray [34]. The linear combination of independent Gaussian random variables is described by a single new Gaussian random variable [35], X_{tot}, such that:

$$X_{tot} = \sum_i a_i X_i \tag{D2}$$

where a_i is the i^{th} voxel's weighting factor in this linear combination. The random variable X_{tot} has a distribution mean of $\mu_{tot} = \sum_i a_i\mu_i$ and a distribution standard deviation of $\sigma_{tot} = \sqrt{\sum_i a_i\sigma_i^2}$. *N.B.* the true value of the line integral along the ray corresponds to μ_{tot}, and the standard deviation from the measurement process of all voxels along the ray corresponds to σ_{tot}. The SNR for the line integral along the ray is then:

$$\text{SNR}_{tot} = \frac{\mu_{tot}}{\sigma_{tot}} = \frac{\sum_i a_i\mu_i}{\sqrt{\sum_i a_i\sigma_i^2}} \tag{D3}$$

As a simple example, consider that a uniform medium is imaged. Each voxel in the CBCT has a distribution mean of μ, and the distribution standard deviation for each voxel is σ. From equation D1, each voxel has an SNR of $\frac{\mu}{\sigma}$. For simplicity, consider that a ray being traced through the CBCT such that all weighting factors in the linear combination are 1, and let there be N voxels intercepted by this ray. The distribution mean for the line integral is $\mu_{tot} = \sum_{i=1}^{N}\mu = N\mu$, and the distribution standard deviation for the line integral is $\sigma_{tot} = \sqrt{\sum_{i=1}^{N}\sigma^2} = \sqrt{N}\sigma$. From equation D3, the SNR of the line integral along the ray is then $\sqrt{N}\frac{\mu}{\sigma}$.

ORCID iDs

Levi Madden https://orcid.org/0000-0002-7754-0756

References

[1] Anastasi G, Bertholet J, Poulsen P, Roggen T, Garibaldi C, Tilly N, Booth J T, Oelfke U, Heijmen B and Aznar M C 2020 Patterns of practice for adaptive and real-time radiation therapy (POP-ART RT) part I: Intra-fraction breathing motion management *Radiother. Oncol.* **153** 79–87

[2] Keall P J, Nguyen D T, O'Brien R, Zhang P, Happersett L, Bertholet J and Poulsen P R 2018 Review of real-time 3-dimensional image guided radiation therapy on standard-equipped cancer radiation therapy systems: Are we at the tipping point for the era of real-time radiation therapy? *International Journal of Radiation Oncology, Biology, Physics* **102** 922–31
[3] Dahele M and Verbakel W 2019 In regard to keall et al *International Journal of Radiation Oncology, Biology, Physics* **103** 282–3
[4] Hardcastle N *et al* 2021 Quantification of the geometric uncertainty when using implanted markers as a surrogate for lung tumor motion *Med. Phys.* **48** 2724–32
[5] Hazelaar C, Verbakel W F, Mostafavi H, van der Weide L, Slotman B J and Dahele M 2018 First experience with markerless online 3d spine position monitoring during SBRT delivery using a conventional LINAC *International Journal of Radiation Oncology, Biology, Physics* **101** 1253–8
[6] de Bruin K, Dahele M, Mostafavi H, Slotman B J and Verbakel W F 2021 Markerless real-time 3-dimensional kv tracking of lung tumors during free breathing stereotactic radiation therapy *Advances in Radiation Oncology* **6** 100705
[7] Dhont J, Verellen D, Mollaert I, Vanreusel V and Vandemeulebroucke J 2020 RealDRR—Rendering of realistic digitally reconstructed radiographs using locally trained image-to-image translation *Radiother. Oncol.* **153** 213–9
[8] Shen L, Zhao W and Xing L 2019 Patient-specific reconstruction of volumetric computed tomography images from a single projection view via deep learning *Nature Biomedical Engineering* **3** 880–8
[9] Zhao W, Han B, Yang Y, Buyyounouski M, Hancock S L, Bagshaw H and Xing L 2019 Incorporating imaging information from deep neural network layers into image guided radiation therapy (IGRT) *Radiother. Oncol.* **140** 167–74
[10] Terunuma T, Sakae T, Hu Y, Takei H, Moriya S, Okumura T and Sakurai H 2023 Explainability and controllability of patient-specific deep learning with attention-based augmentation for markerless image-guided radiotherapy *Med. Phys.* **50** 480–94
[11] Unberath M, Zaech J-N, Lee S C, Bier B, Fotouhi J, Armand M and Navab N 2018 DeepDRR—A Catalyst for Machine Learning in Fluoroscopy-guided Procedures *Medical Image Computing and Computer Assisted Intervention – MICCAI 2018 Lecture Notes in Computer Science* vol 11073 (Springer) p 98–106
[12] Fledelius W, Keall P J, Cho B, Yang X, Morf D, Scheib S and Poulsen P R 2011 Tracking latency in image-based dynamic MLC tracking with direct image access *Acta Oncol.* **50** 952–9
[13] Poludniowski G, Omar A, Bujila R and Andreo P 2021 Technical Note: SpekPy v2.0—a software toolkit for modeling x-ray tube spectra *Med. Phys.* **48** 3630–7
[14] O'Connell J, Lindsay C and Bazalova-Carter M 2021 Experimental validation of Fastcat kV and MV cone beam CT (CBCT) simulator *Med. Phys.* **48** 6869–80
[15] Hubbell J and Seltzer s 1995 *Tables of X-Ray Mass Attenuation Coefficients and Mass Energy-Absorption Coefficients 1 keV to 20 MeV for Elements Z = 1 to 92 and 48 Additional Substances of Dosimetric Interest. NIST Standard Reference Database 126* National Institute of Standards and Technology (Available from: http://nist.gov/pml/data/xraycoef/index.cfm) (https://doi.org/10.18434/T4D01F)
[16] Wiegert J, Bertram M, Schaefer D, Conrads N, Timmer J, Aach T and Rose G 2004 Performance of standard fluoroscopy antiscatter grids in flat-detector-based cone-beam CT *Medical Imaging 2004: Physics of Medical Imaging, Vol. 5368 of Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series* ed M J Yaffe and M J Flynn 67–78
[17] Schneider U, Pedroni E and Lomax A 1996 The calibration of CT Hounsfield units for radiotherapy treatment planning *Phys. Med. Biol.* **41** 111
[18] Rit S, Oliva M V, Brousmiche S, Labarbe R, Sarrut D and Sharp G C 2014 The Reconstruction Toolkit (RTK), an open-source cone-beam CT reconstruction toolkit based on the Insight Toolkit (ITK) *J. Phys. Conf. Ser.* **489** 012079
[19] Pearson K 1895 Note on Regression and Inheritance in the Case of Two Parents *Proceedings of the Royal Society of London* 58 240–2
[20] Bray J R and Curtis J T 1957 An Ordination of the Upland Forest Communities of Southern Wisconsin *Ecological Monographs* **27** 325–49
[21] Field J and Mcfarlane G 1968 Numerical Methods in Marine Ecology *Zoologica Africana* **3** 119–37
[22] Wilcoxon F 1945 Individual comparisons by ranking methods *Biometrics Bulletin* **1** 80–3
[23] Kolmogorov A 1933 Sulla Determinazione Empirica di Una Legge di Distribuzione *Giornale dell'Istituto Italiano degli Attuari* **4** 83–91
[24] Smirnov N 1948 Table for estimating the goodness of fit of empirical distributions *The Annals of Mathematical Statistics* **19** 279–81
[25] Bushberg J 2002 Chapter 3: Interaction of radiation with matter *The Essential Physics of Medical Imaging* (Lippincott Williams & Wilkins) 2nd ednp 51
[26] Wang A *et al* 2018 Acuros CTS: a fast, linear Boltzmann transport equation solver for computed tomography scatter—Part II: System modeling, scatter correction, and optimization *Med. Phys.* **45** 1914–25
[27] Gardner S J, Mao W, Liu C, Aref I, Elshaikh M, Lee J K, Pradhan D, Movsas B, Chetty I J and Siddiqui F 2019 Improvements in cbct image quality using a novel iterative reconstruction algorithm: a clinical evaluation *Advances in Radiation Oncology* **4** 390–400
[28] Robar J L, Cherpak A, MacDonald R L, Yashayaeva A, McAloney D, McMaster N, Zhan K, Cwajna S, Patil N and Dahn H 2023 Novel technology allowing cone beam computed tomography in 6 seconds: A patient study of comparative image quality *Practical Radiation Oncology* 1–10 (In Press)
[29] Rutherford R, Pullan B and Isherwood I 1976 Measurement of effective atomic number and electron-density using an EMI scanner *Neuroradiology* **11** 15–21
[30] Watanabe Y 1999 Derivation of linear attenuation coefficients from CT numbers for low-energy photons *Phys. Med. Biol.* **44** 2201
[31] Seco J and Evans P M 2006 Assessing the effect of electron density in photon dose calculations *Med. Phys.* **33** 540–52
[32] Hubbell J H 1969 *Photon Cross-Sections, Attenuation Coefficients, and Energy Absorption Coefficients from 10-KeV to 100-GeV* NSRDS-NBS 29 United states department of commerce. National bureau of standards (https://doi.org/10.6028/nbs.nsrds.29)
[33] Maslowski A, Wang A, Sun M, Wareing T, Davis I and Star-Lack J 2018 Acuros CTS: a fast, linear Boltzmann transport equation solver for computed tomography scatter—Part I: Core algorithms and validation, J. Star-Lack, Acuros CTSJ. Star-Lack, Acuros CTS *Med. Phys.* **45** 1899–913
[34] Siddon R L 1985 Fast calculation of the exact radiological path for a three-dimensional ct array *Med. Phys.* **12** 252–5
[35] Lemons D and Langevin P 2002 *An Introduction to Stochastic Processes in Physics* (Johns Hopkins University Press)