Biomedical Physics & Engineering Express

Endorsed by IPEM Institute of Physics and Engineering in Medicine

**ACCEPTED MANUSCRIPT**

# Evaluation of deep learning based implanted fiducial markers tracking in pancreatic cancer patients

To cite this article before publication: Abdella M Ahmed *et al* 2023 *Biomed. Phys. Eng. Express* in press https://doi.org/10.1088/2057-1976/acb550

# Evaluation of deep learning based implanted fiducial markers tracking in pancreatic cancer patients

Abdella M. Ahmed[1,2], Maegan Gargett[1,4], Levi Madden[1,2], Adam Mylonas[2], Danielle Chrystall[1,3], Ryan Brown[1], Adam Briggs[5], Trang Nguyen[2], Paul Keall[2], Andrew Kneebone[1,6], George Hruby[1,6] and Jeremy Booth[1,3]

[1.] Northern Sydney Cancer Centre, Royal North Shore Hospital, St Leonards, NSW, Australia

[2.] ACRF Image X Institute, Faculty of Medicine and Health, The University of Sydney, NSW Australia

[3.] Institute of Medical Physics, School of Physics, The University of Sydney, NSW, Australia

[4.] School of Health Sciences, Faculty of Medicine and Health, University of Sydney, Australia

[5.] Shoalhaven Cancer Care Centre, Shoalhaven District Memorial Hospital, Nowra, NSW, Australia

[6.] Northern Clinical School, Sydney Medical School, University of Sydney, NSW, Australia

## Abstract

Real-time target position verification during pancreas stereotactic body radiation therapy (SBRT) is important for the detection of unplanned tumour motions. Fast and accurate fiducial marker segmentation is a requirement of real-time marker-based verification. Deep learning (DL) segmentation techniques are ideal because they don't require additional learning imaging or prior marker information (e.g., shape, orientation). In this study, we evaluated three DL frameworks for marker tracking applied to pancreatic cancer patient data.

The DL frameworks evaluated were 1) a convolutional neural network (CNN) classifier with sliding window, 2) a pretrained you-only-look-once (YOLO) version-4 architecture, and 3) a hybrid CNN-YOLO. Intrafraction kV images collected during pancreas SBRT treatments were used as training data (44 fractions, 2017 frames). All patients had 1-4 implanted fiducial markers. Each model was evaluated on unseen kV images (42 fractions, 2517 frames). The ground truth was calculated from manual segmentation and triangulation of markers in orthogonal paired kV/MV images. The sensitivity, specificity, and area under the precision-recall curve (AUC) were calculated. In addition, the mean-absolute-error (MAE), root-mean-square-error (RMSE) and standard-error-of-mean (SEM) were calculated for the centroid of the markers predicted by the models, relative to the ground truth.

The sensitivity and specificity of the CNN model were 99.41% and 99.69%, respectively. The AUC was 0.9998. The average precision of the YOLO model for different values of recall was 96.49%. The MAE of the three models in the left-right, superior-inferior, and anterior-posterior directions were under 0.88±0.11 mm, and the RMSE were under 1.09±0.12 mm. The detection times per frame on a GPU were 48.3, 22.9, and 17.1 milliseconds for the CNN, YOLO, and CNN-YOLO, respectively.

The results demonstrate submillimeter accuracy of marker position predicted by DL models compared to the ground truth. The marker detection time was fast enough to meet the requirements for real-time application.

SPAN-C trial ID: NCT03505229

## 1. Introduction

Pancreatic cancer patients typically have poor prognosis, with a 5% 5-year survival rate (Ilic and Ilic 2016). Neoadjuvant stereotactic body radiation therapy (SBRT) is playing an increasing role in the management of patients with inoperable disease, to promising effect (Petrelli et al 2017). SBRT characteristically involves the delivery of large doses in few fractions, placing paramount importance on target localisation accuracy in order to achieve tumour control and avoid toxicity to nearby radiosensitive organs (bowel, stomach). Accurate targeting of the tumour volume during treatment is complicated by the presence of intrafraction motion due to respiration and digestive processes, both of which can cause variation in target position and organ deformation. Monitoring the target position in real time is therefore desirable for intrafraction motion management purposes. This has been most accurately and practicably achieved by employing image guidance to monitor the position of implanted fiducial markers, as a surrogate for the tumour (Imura *et al* 2005, Balter *et al* 1995). Fast and accurate segmentation of these markers in online imaging is of paramount importance for such applications.

Several methods have been developed and implemented for real-time marker segmentation in kV planar imaging based on template matching algorithms (Regmi *et al* 2014, Fledelius *et al* 2011, 2014). Wan et al (2014) developed a combination of dynamic programming and template matching for improved marker localisation. Campbell et al (2017) developed a method to automatically generate marker templates from cone beam computed tomography (CBCT) imaging that enabled tracking of a cluster of markers in kV images. Kilovoltage intrafraction monitoring (KIM) employs a template matching method in conjunction with a 3D probability density function to report 3D marker positions from 2D imaging (Nguyen *et al* 2017, Hewson *et al* 2019, Kim *et al* 2018). These template matching techniques by nature require prior knowledge of marker properties. To determine the properties of the marker, an additional learning period is needed, exposing the patient to additional radiation dose and increasing time on the treatment couch (Bertholet *et al* 2017). Moreover, in template matching, high-contrast objects in an image are used as landmarks. If fiducial markers are obscured by high-density material such as bony anatomy and other high-density materials (such as surgical clips and stents), the template matching process becomes more uncertain. In addition, when using coil markers, the template matching process becomes more complicated due to their arbitrary shape and deformation during implantation and treatment (due to organ motion). Therefore, there is a growing interest to address the issue of target tracking with deep learning (DL) based methods. DL methods have been shown to provide fast and accurate marker tracking (Mylonas *et al* 2019) and do not require prior knowledge of marker properties, negating the need for a learning period to generate a template.

The success of DL models to learn from a training dataset by automatically extracting features has led researchers to apply it in the field of medical physics, with the goal to improve patient outcomes (Edmunds *et al* 2019, Cui *et al* 2020, Liang *et al* 2020, Mylonas *et al* 2021, Motley *et al* 2022). Amarsee et al (2021) reported automatic detection of markers using you-only-look-once (YOLO) version-2 applied to three prostate cancer patients. The accuracy of the marker detection was within 1 mm in 98% of the kV images. Previously, we developed a convolutional neural network (CNN) model with four building blocks for automatic detection of cylindrical and arbitrarily shaped fiducial markers in kV images for prostate cancer patients (Mylonas *et al* 2019). However, the above DL methods were trained and evaluated on a small cohort of prostate cancer patients.

In this study, we developed and evaluated three DL frameworks applied to pancreatic cancer patients. The DL frameworks evaluated were 1) a CNN with sliding window classification, 2) a pretrained YOLO version-4 architecture, and 3) a hybrid CNN-YOLO method. We aim to achieve better marker detection results (in both accuracy and latency) by improving the methods involved in individual models and combining them into a hybrid detection system.

## 2. Methods

### 2.1 Patient Data Collection

This study was conducted using 2D kV images acquired from an ethics approved clinical trial for pancreas SBRT (SPAN-C; ID: NCT03505229). All patients had 1-4 implanted fiducial markers (EchoTip® Ultra, Cook Medical) and were treated using a conventional linear accelerator (TrueBeam, Varian, Palo Alto CA, USA). The treatment protocol utilised respiratory gating for motion management with the treatment beam gated at the end-exhale phase, either during free-breathing or breath hold. Intrafraction kV imaging was acquired within the gating window. The source-detector distance (SDD) and source-isocenter distance (SID) were 150 and 100 cm, respectively. The pixel sizes at the detector and isocentre were 0.388×0.388 mm and 0.258×0.258 mm, respectively. The imaging parameters varied from 80 kV to120 kV, and from 100 mA to 300 mA.

### 2.2 Deep Learning Methods

The main steps involved in this study are illustrated in Figure 1. The collimated area of each kV image was removed and the area containing the markers and the surrounding region was used to generate the training dataset. The training dataset contained a total of 44 fractions from 13 patients (a total of 2017 frames). Annotated training (90%) and validation datasets (10%) were generated from these images. Rotation-based data augmentation was performed. All the software implementations and analysis were performed using MATLAB 2021b (The Mathworks Inc. 2021). The training was done on a GPU (NVIDIA GeForce RTX 3090).

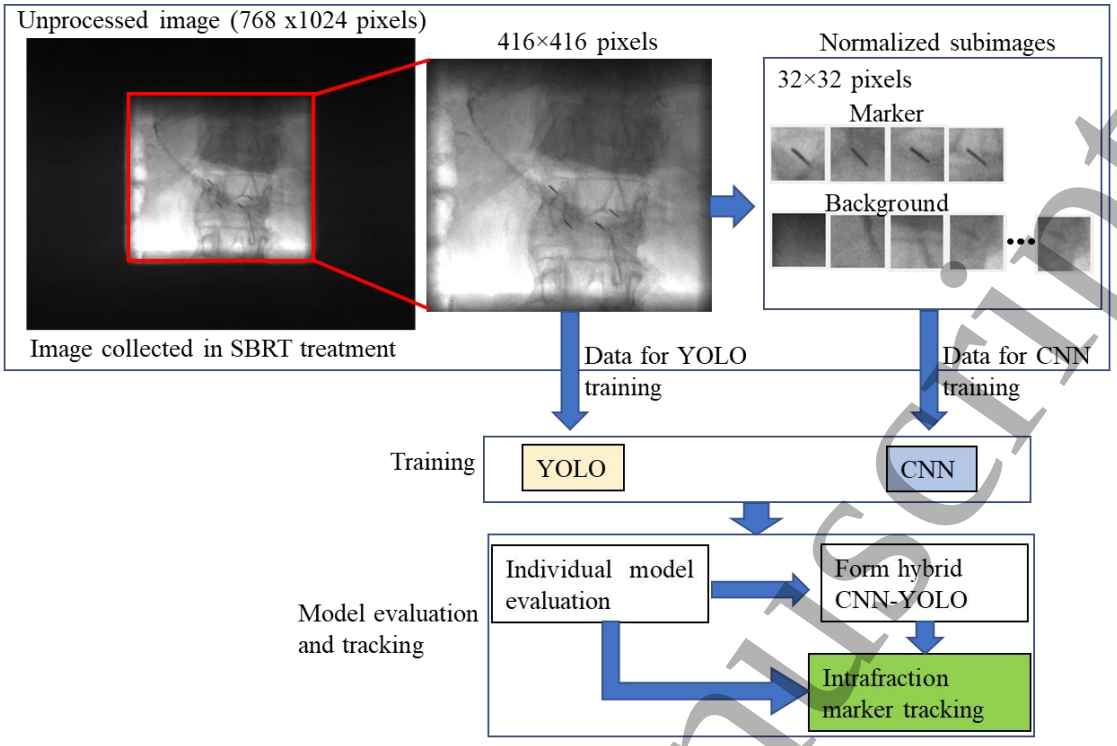


Figure 1. Overview of data processing, training, and model implementation.

### 2.2.1 CNN

CNNs are a class of deep neural networks with a specific architecture that are very effective when dealing with image data. A CNN is composed of several building blocks with the convolutional layers being the core part of the architecture. In this study, the CNN was designed to classify 32×32 pixel subimages either as marker or background (no marker present). Prior to training, each subimage was normalised so that the minimum and maximum pixel values were 0 and 1, respectively. The architecture of the CNN is shown in Figure 2. The first and the third building blocks lack the batch-normalisation layer to minimize the computation time of the training. A stochastic gradient descent (SGD) with momentum was used for fast convergence. The size of the mini batch was 128. The learning rate and the momentum were set as 0.0001 and 0.9, respectively. Data shuffling was performed prior to each epoch. Early stopping regularization was used to avoid overfitting. The trained CNN model was then incorporated into an automatic tracking system which is uses a sliding window technique to determine 2D marker position (Mylonas *et al* 2019) (Figure 3). The size of the search area was large enough to include intrafraction motion. The search area of each marker for the first frame is estimated using the information from planning CT (by projecting the planning positions of the markers into the kV image) (Poulsen *et al* 2008). The position of the markers from patient coordinates (from planning CT) $r_p = (x_p; y_p; z_p)$ were transformed into the

kV imager coordinates $r = (x; y; z)$. The transformation consists of a rotation followed by a projection as given in eq. 1.

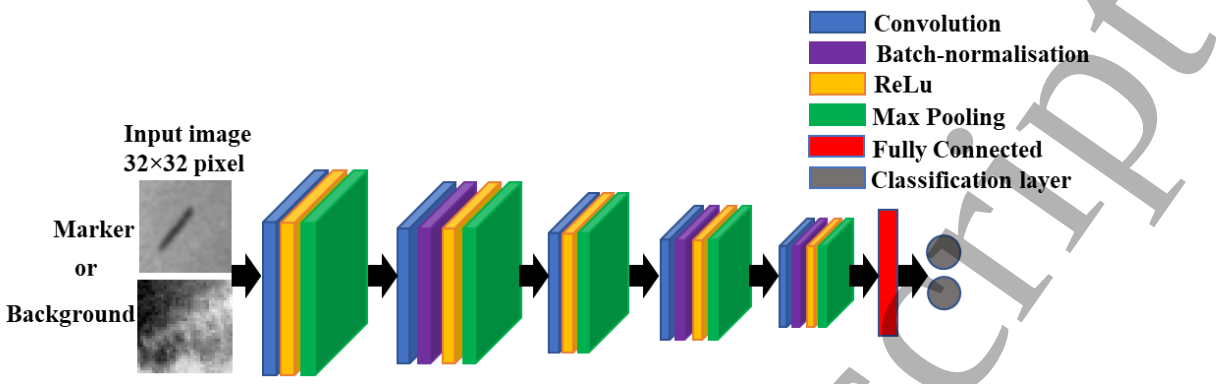


Figure 2. Schematic illustration of the CNN architecture. The input image is a normalised 32 x 32 image and passed to the first building blocks.

$$r = \begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} \frac{SDD}{SID - z} & 0 & 0 \\ 0 & \frac{SDD}{SID - z} & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} cos\,\alpha & 0 & -sin\,\alpha \\ 0 & 1 & 0 \\ sin\,\alpha & 0 & cos\,\alpha \end{pmatrix} r_p \qquad 1$$

Finally, the projected position on the kV imager is given by eq. 2.

$$\begin{pmatrix} x \\ y \end{pmatrix} = \frac{SDD}{SID - z} \begin{pmatrix} x_p\,cos\,\alpha - z\,sin\,\alpha \\ y_p \end{pmatrix} \qquad 2$$

where α is the gantry angle. However, the projected marker positions are not always accurate due to uncertainties which arise from patient positioning, organ motion (Wysocka *et al* 2010)and/or marker migration (Motley *et al* 2022). In instances where the marker is not present within the CNN's initial search window, the search area was progressively expanded until all markers were found. The search area for consecutive frames depended on the location of the markers on the previous frame. If the markers could not be found within the specified search area (such as when large projection angle change (> $10^{o}$) between consecutive frames occurred), the information from the planning CT was used for that specific frame at the given projection angle. The number of planning CT position usages during tracking was calculated.

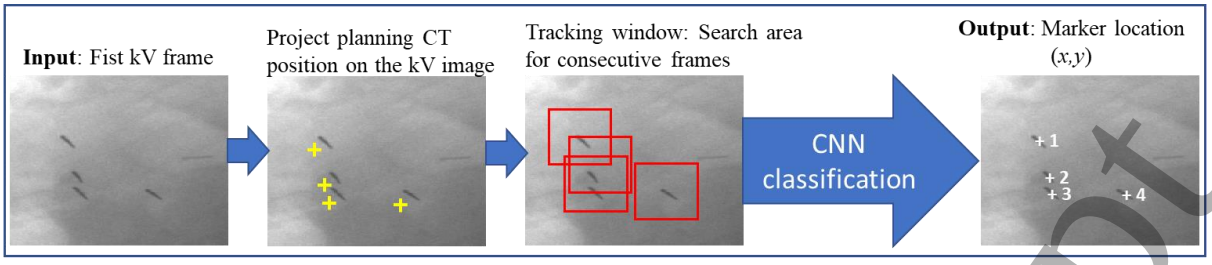


Figure 3. Tracking window technique for improved efficiency of the CNN marker classification. The centre of the search area is estimated based on the projected position of each marker from the planning CT (yellow crosses). Once the search area is initialised (red boxes), the centre of the search area for consecutive frames is calculated from the location of the markers on the previous frame. The white crosses are the predicted marker positions by the CNN.

### 2.2.2 YOLO

YOLO object detection consists of two steps: locating and classifying the object in an image. First invented by Redmon et al (2016), it has gone through several improvements especially for its stability and computation time to be able to detect multiple objects at once (Redmon and Farhadi 2017, 2018, Bochkovskiy *et al* 2020). YOLO uses a series of CNN and regression methods to classify and locate objects in an image. The input image is divided into $S \times S$ grid cells (Figure 4). Each cell predicts $B$ bounding boxes and a class probability. The box with the highest probability will be chosen as having the object (in our case the marker). The bounding box coordinate is defined as ($x$, $y$, $w$, $h$), where ($x$, $y$) are the top-left coordinates of the box, and $w$ and $h$ are its width and height, respectively. During data preparation, the ground truth is labelled manually (a bounding box that contains the marker) and provided to YOLO. This ground truth is used to optimise the weights and the biases during training. For each bounding box, intersection-over-union (IoU) is calculated. The IoU gives the confidence score by measuring the overlap between a predicted bounding box and the ground truth. When predicting more than two boxes (i.e., detecting one marker multiple times), Non-Max Suppression (NMS) technique is used to remove boxes with lower confidence score.

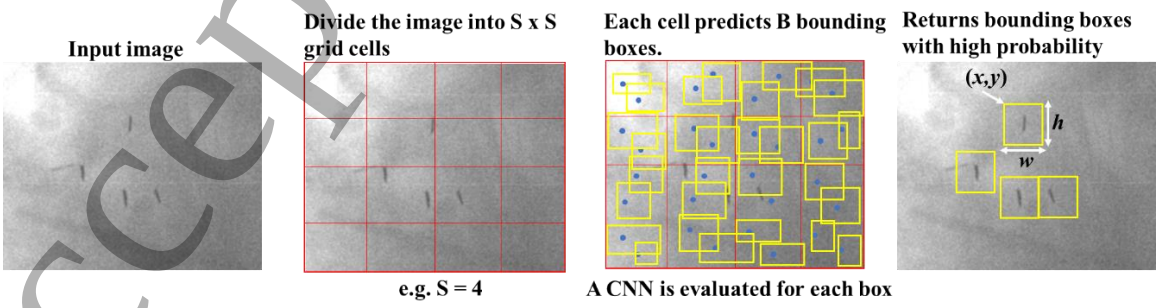

Figure 4. Example of YOLO technique, with a single kV input image to the prediction of the marker with a confidence score for each bounding box (in this study, a threshold of 0.85 was used).

For YOLO training, the size of the input image was 416×416 pixels to ensure all markers implanted near the pancreas gross tumour volume (GTV) were included. SGD with momentum was chosen as an optimiser. The learning rate and the L2 regularisation factor were set to 0.0001 and 0.0005, respectively. The number of epochs was 100. The trained model was then implemented to track fiducial markers. The centre of the bounding box was considered as the position of the marker. A common challenge with marker detection in radiotherapy is overlapping markers (or markers placed near to each other) being detected as one marker. A bi-detection technique was employed (Figure 5), whereby the image is cropped near to the centroid of the detected markers and passed to the YOLO detector again when fewer than the expected markers are detected. This improves the detection of individual markers by preventing two or more closely spaced markers being detected as one.

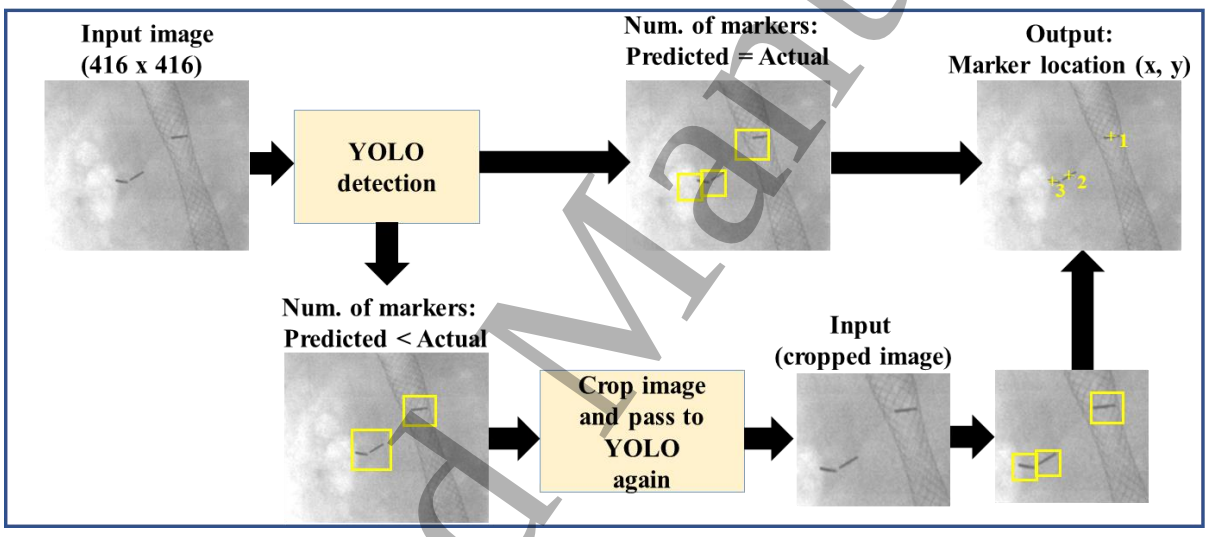


Figure 5. YOLO bi-detection technique. First, the entire image is passed as an input. The model returns bounding boxes that contain the markers with the highest probability. If fewer markers are detected than the expected number of markers, then the YOLO bi-detection is applied.

### 2.2.3 CNN-YOLO

The CNN-based marker tracking requires the information from the planning CT to initialise the search area. In instances where the intrafraction marker position varies from the planned position, the CNN's search area will expand until all the markers are found. This increases the computation time for marker detection. To address this problem, we designed a hybrid DL model from individual CNN and YOLO models previously trained (the model is named as CNN-YOLO). In this method, a larger cropped area will be passed to the YOLO to initialise the position of the markers. The location of the markers (predicted with the highest confidence) will be passed to the

CNN and the CNN will track the marker for consecutive frames. If the marker is not found by the CNN, the YOLO will take over. Subsequently, the positions of the markers from YOLO will be passed to CNN again. This process continues until all the kV images are processed. In doing so, the number of YOLO overtakes was calculated. The hybrid CNN-YOLO process is illustrated in Figure 6.

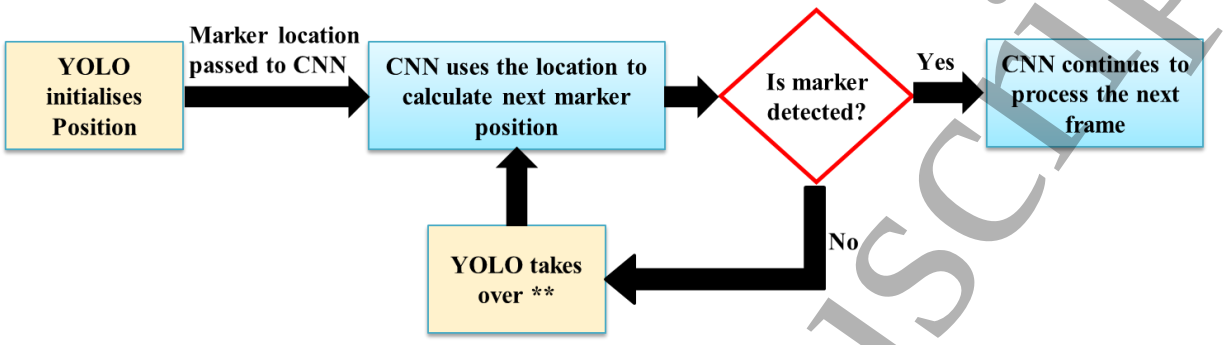


Figure 6. Illustration of the hybrid CNN-YOLO. The location of the marker is first initialised by the YOLO and the locations are passed to the CNN to calculate the location of the marker position for the consecutive frames. If the CNN fails to detect the markers, then the YOLO will take over.

**The YOLO bi-detection is also applied if the number of predicted markers are less than that of the actual (as in Figure 5).

### 2.2.4 DL Model evaluation

For the CNN, the sensitivity (the ability of the model to correctly predict subimages with markers), specificity (the ability of the model to correctly classify background subimages), the precision-recall curve (PRC) and the Area Under PRC (AUC) plot were evaluated. For YOLO, the evaluation of precision (accurate classification of the markers by the model) and recall (the ability to find all markers in kV images) are slightly different from the evaluation of these metrics in CNN since we have only one class (which is marker). Thus, average precision for different values of recall was evaluated.

The tracking accuracy of the models was evaluated against a ground truth derived from manual segmentation of the markers in the planar images (acquired from 15 patients (42 fractions, 2517 frames)). The 3D ground truth marker position was calculated from triangulation of marker positions in orthogonally paired kV/MV images (Hewson *et al* 2019). Due to low visibility of the markers in the MV field of view, calculation of triangulated positions was not possible for all kV images. In addition, the MV and kV images should be in synchronisation within 0.5 seconds. In the absence of triangulated marker positions, 3D positions were reconstructed from manually segmented 2D positions using 3D probability density function (3D-PDF) (Poulsen *et al* 2008). The

2D marker positions predicted by the models were also converted into 3D positions using 3D-PDF. For each fraction, the mean absolute error (MAE), root-mean-square-error (RMSE) and standard-error-of-mean (SEM) of the DL models were calculated. In addition, the 5$^{th}$ and 95$^{th}$ percentiles of the errors were also evaluated. The mean Euclidian distance, *d*, between the centroid of the marker positions of the ground truth and that of the predictions was also calculated.

## 3. Results

The sensitivity and specificity of the CNN model was 99.41% and 99.69%, respectively. The AUC of the CNN was 0.9998. The average precision of the YOLO was 96.49 %. These results indicate the ability of the models to detect the markers with high accuracy. The MAE and the RMSE of marker tracking for all the fraction is shown in Figure 7. The overall MAE and RMSE with SEM is given in Table 1. In the case of stand-alone CNN classification, the planning CT positions were used 36.20% of the time, and for the hybrid CNN-YOLO, YOLO took over 32.20% of the time. The statistical significance (p-value estimated from correlation coefficient) of each model in the LR, SI, and AP directions was 0.017, 0.001, and 0.011, respectively.

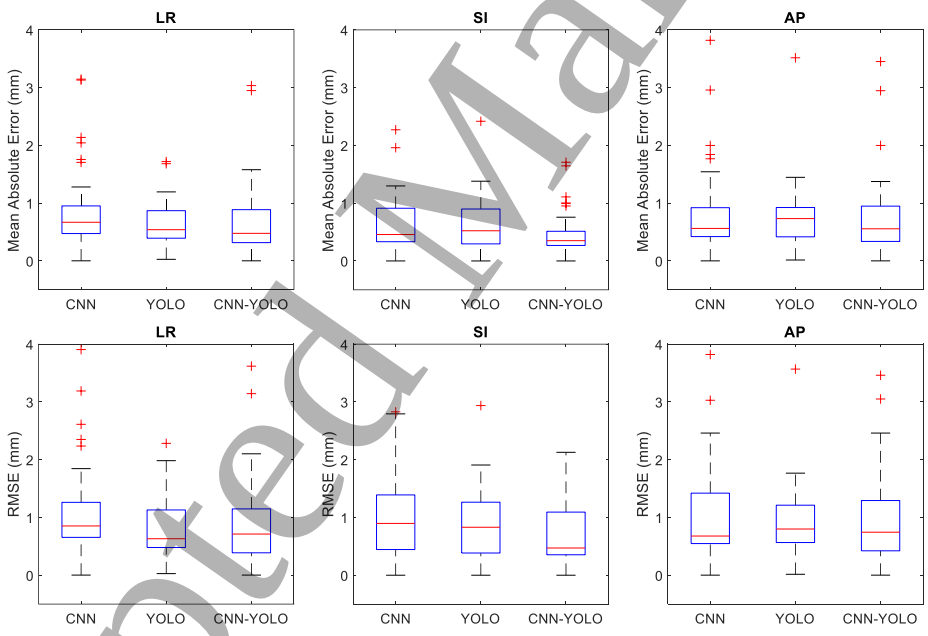


Figure 7. Marker tracking error with respect to the ground truth in LR, SI, and AP direction of the MAE and RMSE. The error was averaged for each fraction.

Table 1. Mean absolute errors and RMSE averaged for all the fractions

| | Mean error (mm) | | | Mean RMSE (mm) | | |
|---|---|---|---|---|---|---|
| | LR | SI | AP | LR | SI | AP |
| **CNN** | 0.88 ± 0.11 | 0.65 ± 0.07 | 0.84 ± 0.11 | 1.09 ± 0.12 | 1.04 ± 0.11 | 1.07 ± 0.13 |
| **YOLO** | 0.74 ± 0.11 | 0.65 ± 0.07 | 0.77 ± 0.08 | 0.95 ± 0.15 | 0.91 ± 0.09 | 0.95 ± 0.09 |
| **CNN-YOLO** | 0.81 ± 0.14 | 0.47 ± 0.06 | 0.75 ± 0.10 | 1.01 ± 0.17 | 0.73 ± 0.08 | 0.95 ± 0.12 |

The 5$^{th}$ and the 95$^{th}$ percentiles of the errors are given in Table 2. Although all the DL models have similar performances, the CNN has the lowest error in the SI direction, and YOLO has the lowest error in the AP direction for the 5$^{th}$ percentile. For the 95$^{th}$ percentile, CNN-YOLO has the lowest error in the SI direction.

Table 2. The 5th and 95th percentiles of the errors calculated with respect to the ground truth and average Euclidian distance measured between the centroid of the predicted and ground truth marker location.

| | 5$^{th}$ percentile of the errors (mm) | | | 95$^{th}$ percentile of the errors (mm) | | | Ave. |
|---|---|---|---|---|---|---|---|
| | LR | SI | AP | LR | SI | AP | Euclidian |
| **CNN** | -1.34 | -0. 79 | -0.90 | 1.05 | 1.62 | 1.45 | 1.58 |
| **YOLO** | -1.30 | -1.09 | -0.87 | 1.17 | 1.00 | 1.33 | 1.48 |
| **CNN-YOLO** | -1.30 | -1.00 | -1.19 | 1.14 | 0.96 | 1.13 | 1.39 |

An example of tracking with two implanted fiducial markers in the pancreas GTV is shown in Figure 8. All the models have tracked the markers successfully and the median of the error is close to zero as shown in the boxplot given in Figure 9.

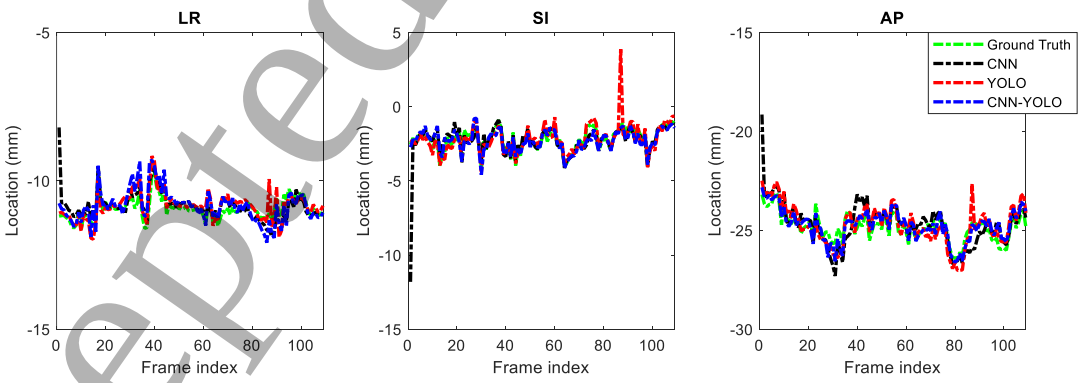


Figure 8. Marker tracking using the three DL methods in 3D in comparison to the ground truth in the (a) LR, (b) SI and (c) AP directions.

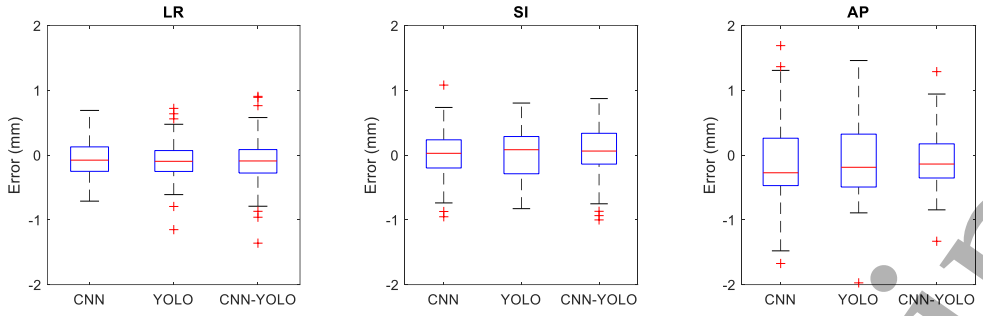


Figure 9. Error of marker tracking for the three models with respect to the ground truth in the (a) LR, (b) SI and (c) AP directions.

The computation time of the models to detect the markers on a single frame was computed on CPU (AMD Ryzen 9 5950X 16-core processor) and GPU (NVIDIA GeForce RTX 3090) hardware. Detection time of the marker was calculated from the ratio of the total time of the computation to the number of frames processed. A patient with four markers (with 140 kV image frames) implanted in the pancreas GTV was chosen. Time measurement was evaluated 6 times and the average result was reported. The CNN with sliding window has 54.2 and 48.3 milliseconds on CPU and GPU, respectively, while the detection time for YOLO was 35.9 and 22.9 milliseconds. The CNN-YOLO has shown the best performance when the CNN-part is run on a CPU and the YOLO-part run on GPU, and the computation time was 17.1 milliseconds.

## 4. Discussion

In this study, we investigated three types of DL frameworks for marker classification and detection for real-time radiotherapy treatment of the pancreas. The performance of a CNN with a sliding window, a pretrained standalone YOLO and a hybrid CNN-YOLO have demonstrated excellent results in both geometric accuracy and computation time. The geometric accuracy of marker tracking for the three models was sub-millimetre as quantified by the MAE and RMSE. This indicates that clinical implementation of this marker localisation method will enable patient setup within the recommended accuracy requirement of 1 mm (Klein et al 2009, Willoughby et al 2012).

In conventional marker tracking methods such as template matching, the presence of high-density materials near the treatment volume can make the template matching approach difficult. Tang et al (2007) developed a system that integrates the template matching approach with a multiple object tracking process to avoid marker confusion with non-marker materials. The system had excellent marker detection ability when breathing pattern information (BPI) was incorporated, whereby the breathing pattern is correlated with abdominal organ motion (pancreas, lung, etc) to provide predicted tumor trajectory information. In the absence of BPI, the failure of marker detection reached up to 12%. In our case, the DL models successfully detected the markers in the presence of high-density

materials without provision of BPI. This shows the generalisability of the DL method compared to template matching methods, where prior information such as marker characteristics and patient specific information such as BPI are not required.

The presented work has shown that DL methods produce localisation accuracy appropriate for stereotactic applications. However, the markers are used as a surrogate for tumour localisation; if there is a related surrogacy error, for example due to marker migration, the tumour localisation accuracy will be reduced even if the marker localisation accuracy is high. DL methods may be able to be employed to track the tumour volume directly, however this presents challenges in relation to tumour visibility in intrafraction imaging.

In our previous work, we reported a CNN with a compact design (4 building blocks) for marker segmentation in the prostate (Mylonas *et al* 2019). The same compact CNN architecture was trained and evaluated on the SPAN-C data, resulting in the sensitivity and specificity of 95.60% and 99.90%, respectively. In order to improve the sensitivity, we designed the CNN with a deeper architecture having six building blocks without increasing the number of trainable parameters (hence, similar computation time) by reducing the batch-normalisation layer in the first and third building blocks (Figure 2). The model achieved high sensitivity and high specificity of 99.41% and 99.69%, respectively. These results imply the ability of the model to classify the positive and the negative images with high accuracy even with images that contain high-density non-marker materials. The CNN showed slightly higher specificity compared to its sensitivity. For real-time monitoring, high specificity is desirable because the system should not identify the background as a marker and send an incorrect message to the system monitoring which instructs interventions in the treatment delivery.

For real-time IGRT, both accuracy and detection time is important. The detection time (per frame with four markers) by the three models favourably compares with the detection time reported in previous results (Mylonas *et al* 2019). For a real-time tracking in clinical use, the latency should be kept below 500 milliseconds (Keall *et al* 2006). Given the computation time of all the three models in 2D, the models have the potential to be as short as 100 milliseconds including the conversion from 2D to 3D positions with 3D-PDF. Of all the three models, the CNN-YOLO has the best temporal performance (17.1 milliseconds per frame per four markers). This is because, in the case of CNN, the search area automatically changes until the marker is found and this relatively increased the computation time. In the case of YOLO, we pass the entire cropped image as an input (dimension of 416×416 pixels) for each frame and the marker is searched for the entire input image. However, in the case of CNN-YOLO, the YOLO-part is used instead of planning CT information and the markers were usually found with the first step. Then the CNN continues the detection for

consecutive frames and the search area is usually small. However, if the CNN could not find the marker and YOLO fails to recover it, then this would be one limitation. In this case, marker position from previous frame or from planning CT could be used.

## 5. Conclusion

Three DL approaches were implemented to classify and track implanted fiducial markers in pancreatic cancer patient data. The performance of the models was evaluated on unseen data. The accuracy of marker position prediction by the DL models from the ground truth was sub-millimetre, and detection time was fast enough to meet the requirements for online application. Specifically, the hybrid model from the CNN and YOLO (CNN-YOLO) has achieved faster detection of fiducial markers (17.1 milliseconds). Therefore, the models could be deployed as part of a real-time intrafraction motion monitoring software.

## References

Amarsee K, Ramachandran P, Fielding A, Lehman M, Noble C, Perrett B and Ning D 2021 Automatic detection and tracking of marker seeds implanted in prostate cancer patients using a deep learning algorithm *J Med Phys* **46** 80–7

Balter J M, Lam K L, Sandler H M, Littles J F, Bree R L and ten Haken R K 1995 Automated localization of the prostate at the time of treatment using implanted radiopaque markers: Technical feasibility *Int J Radiat Oncol Biol Phys* **33**

Bertholet J, Wan H, Toftegaard J, Schmidt M L, Chotard F, Parikh P J and Poulsen P R 2017 Fully automatic segmentation of arbitrarily shaped fiducial markers in cone-beam CT projections *Phys Med Biol* **62**

Bochkovskiy A, Wang C-Y and Liao H-Y M 2020 YOLOv4: Optimal Speed and Accuracy of Object Detection Online: http://arxiv.org/abs/2004.10934

Campbell W G, Miften M and Jones B L 2017 Automated target tracking in kilovoltage images using dynamic templates of fiducial marker clusters: *Med Phys* **44** 364–74

Cui S, Tseng H H, Pakela J, ten Haken R K and el Naqa I 2020 Introduction to machine and deep learning for medical physicists *Medical Physics* vol 47

Edmunds D, Sharp G and Winey B 2019 Automatic diaphragm segmentation for real-time lung tumor tracking on cone-beam CT projections: A convolutional neural network approach *Biomed Phys Eng Express* **5**

Fledelius W, Worm E, Elstrøm U v., Petersen J B, Grau C, Høyer M and Poulsen P R 2011 Robust automatic segmentation of multiple implanted cylindrical gold fiducial markers in cone-beam CT projections *Med Phys* **38**

Fledelius W, Worm E, Høyer M, Grau C and Poulsen P R 2014 Real-time segmentation of multiple implanted cylindrical liver markers in kilovoltage and megavoltage x-ray images *Phys Med Biol* **59**

Hewson E A, Nguyen D T, O'Brien R, Kim J H, Montanaro T, Moodie T, Greer P B, Hardcastle N, Eade T, Kneebone A, Hruby G, Hayden A J, Turner S, Siva S, Tai K H, Hunter P, Sams J, Poulsen P R, Booth J T, Martin J and Keall P J 2019 The accuracy and precision of the KIM motion monitoring system used in the multi-institutional TROG 15.01 Stereotactic Prostate Ablative Radiotherapy with KIM (SPARK) trial *Med Phys* **46** 4725–37

Ilic M and Ilic I 2016 Epidemiology of pancreatic cancer *World J Gastroenterol* **22**

Imura M, Yamazaki K, Shirato H, Onimaru R, Fujino M, Shimizu S, Harada T, Ogura S, Dosaka-Akita H, Miyasaka K and Nishimura M 2005 Insertion and fixation of fiducial markers for setup and tracking of lung tumors in radiotherapy *Int J Radiat Oncol Biol Phys* **63**

Keall P J, Mageras G S, Balter J M, Emery R S, Forster K M, Jiang S B, Kapatoes J M, Low D A, Murphy M J, Murray B R, Ramsey C R, van Herk M B, Vedam S S, Wong J W and Yorke E 2006 The management of respiratory motion in radiation oncology report of AAPM Task Group 76 *Med Phys* **33**

Kim J H, Nguyen D T, Booth J T, Huang C Y, Fuangrod T, Poulsen P, O'Brien R, Caillet V, Eade T, Kneebone A and Keall P 2018 The accuracy and precision of Kilovoltage Intrafraction Monitoring (KIM) six degree-of-freedom prostate motion measurements during patient treatments *Radiotherapy and Oncology* **126** 236–43

Klein E E, Hanley J, Bayouth J, Yin F F, Simon W, Dresser S, Serago C, Aguirre F, Ma L, Arjomandy B, Liu C, Sandin C and Holmes T 2009 Task group 142 report: Quality assurance of medical acceleratorsa *Med Phys* **36**

Liang Z, Zhou Q, Yang J, Zhang L, Liu D, Tu B and Zhang S 2020 Artificial intelligence-based framework in evaluating intrafraction motion for liver cancer robotic stereotactic body radiation therapy with fiducial tracking *Med Phys* **47**

Motley R, Ramachandran P and Fielding A 2022 A feasibility study on the development and use of a deep learning model to automate real-time monitoring of tumor position and assessment of interfraction fiducial marker migration in prostate radiotherapy patients *Biomed Phys Eng Express* **8**

Mylonas A, Booth J and Nguyen D T 2021 A review of artificial intelligence applications for motion tracking in radiotherapy *J Med Imaging Radiat Oncol* **65**

Mylonas A, Keall P J, Booth J T, Shieh C C, Eade T, Poulsen P R and Nguyen D T 2019 A deep learning framework for automatic detection of arbitrarily shaped fiducial markers in intrafraction fluoroscopic images *Med Phys* **46** 2286–97

Nguyen D T, O'Brien R, Kim J H, Huang C Y, Wilton L, Greer P, Legge K, Booth J T, Poulsen P R, Martin J and Keall P J 2017 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

Petrelli F, Comito T, Ghidini A, Torri V, Scorsetti M and Barni S 2017 Stereotactic Body Radiation Therapy for Locally Advanced Pancreatic Cancer: A Systematic Review and Pooled Analysis of 19 Trials *Int J Radiat Oncol Biol Phys* **97**

Poulsen P R, Cho B, Langen K, Kupelian P and Keall P J 2008 Three-dimensional prostate position estimation with a single x-ray imager utilizing the spatial probability density *Phys Med Biol* **53** 4331–53

Redmon J, Divvala S, Girshick R and Farhadi A 2016 You only look once: Unified, real-time object detection *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition* vol 2016-December

Redmon J and Farhadi A 2018 YOLOv3: An incremental improvement *arXiv preprint*

Redmon J and Farhadi A 2017 YOLO9000: Better, faster, stronger *Proceedings - 30th IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2017* vol 2017-January

Regmi R, Lovelock D M, Hunt M, Zhang P, Pham H, Xiong J, Yorke E D, Goodman K A, Rimner A, Mostafavi H and Mageras G S 2014 Automatic tracking of arbitrarily shaped implanted markers in kilovoltage projection images: A feasibility study *Med Phys* **41**

Tang X, Sharp G C and Jiang S B 2007 Fluoroscopic tracking of multiple implanted fiducial markers using multiple object tracking *Phys Med Biol* **52**

The Mathworks Inc. 2021 MATLAB 2021b. Natick, Massachusetts:

Wan H, Ge J and Parikh P 2014 Using dynamic programming to improve fiducial marker localization *Phys Med Biol* **59**

Willoughby T, Lehmann J, Bencomo J A, Jani S K, Santanam L, Sethi A, Solberg T D, Tomé W A and Waldron T J 2012 Quality assurance for nonradiographic radiotherapy localization and positioning systems: Report of Task Group 147 *Med Phys* **39**

Wysocka B, Kassam Z, Lockwood G, Brierley J, Dawson L A, Buckley C A, Jaffray D, Cummings B, Kim J, Wong R and Ringash J 2010 Interfraction and Respiratory Organ Motion During Conformal Radiotherapy in Gastric Cancer *Int J Radiat Oncol Biol Phys* **77**