

Medical Image Segmentation under Data-Efficient Scenarios

CANRAN LI

Doctor of Philosophy (Computer Science)



THE UNIVERSITY OF
SYDNEY

Supervisor: A/Prof. Weidong Cai

A thesis submitted in fulfilment of
the requirements for the degree of
Doctor of Philosophy

School of Computer Science
Faculty of Engineering
The University of Sydney
Australia

28 February 2025

Declaration

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.

Canran Li

28 February 2025

Abstract

Medical image analysis plays a crucial role in modern healthcare, contributing to more accurate diagnoses, treatment planning, and patient management. Among the various tasks in this field, medical image segmentation is particularly important for clinical analysis, as it involves identifying anatomical structures and abnormalities. Deep learning techniques have recently achieved remarkable success in medical image segmentation due to their ability to learn complex patterns from large datasets. However, training deep neural networks typically requires the availability of large-scale labeled data, which is often expensive and time-consuming to obtain in medical imaging. Moreover, variations in clinical environments and differences in imaging equipment across institutions bring challenges to the generalized ability of traditional supervised deep learning models. Therefore, transfer learning methods, especially deep domain adaptation methods, have been developed to improve model performance by bridging the distributional differences between domains. These advanced techniques open new possibilities for reducing reliance on large labeled medical image datasets. Despite these advancements, this field still faces significant challenges, including semantic gaps in medical image categories, limitations of adversarial learning for domain adaptation, and lack of generalization with unavailable target data. This thesis introduces various deep learning-based domain adaptation methods for medical image segmentation under data-efficient scenarios, with the goal of addressing the challenges associated with limited data availability. The main contributions of this work are as follows:

The first part of this thesis introduces CAPL-Net, a novel deep unsupervised domain adaptation (UDA) approach specifically designed for cross-domain nuclei instance segmentation and classification using category-aware feature alignment and pseudo-labeling. In histopathology images, substantial semantic gaps exist between different types of nuclei. To address this, we propose a category-aware feature alignment module that enhances knowledge transfer for intra-class features across domains while reducing negative transfer for inter-class

features. Next, we introduce a self-supervised learning stage that leverages nuclei-level feature prototypes to improve model performance on unlabeled target data. CAPL-Net addresses the limitations of previous class-agnostic UDA methods by learning more comprehensive domain-specific features across various nuclei categories.

The second contribution presents an advanced dynamic diffusion-based framework (DDF-UDA) for cross-domain optic disc (OD) and optic cup (OC) segmentation in fundus images. DDF-UDA includes an adaptive module based on feature- and pixel-level diffusion processes, along with a Nash equilibrium-based adjustment strategy to reduce cross-domain discrepancies. This approach enables the transfer of domain-specific information from the target data to the source data, thereby facilitating effective domain adaptation. The framework demonstrates superior performance under imbalanced data conditions, outperforming state-of-the-art (SOTA) methods.

Finally, we propose a novel zero-shot domain adaptation (ZSDA) paradigm, named LMaD, to address the challenge of requiring access to target images during training. LMaD utilizes a language model as a feature disentangler for ZSDA medical image segmentation tasks, aiming to eliminate interference caused by domain-specific information, resulting in superior performance. Specifically, we first introduce a graph learning mechanism to construct a visual graph that captures relationships between visual features. Next, we leverage a pre-trained language model to obtain task-specific and domain-specific feature embeddings. By disentangling features within the visual graph, we aim to eliminate interference from domain-specific information. Additionally, a graph convolutional network is employed to infer higher-order relationships among task-specific features within the visual graph, further enriching the contextual information in the images.

Overall, this thesis explores novel domain adaptation methods specifically designed for medical image segmentation in data-efficient scenarios, addressing the challenge of data scarcity, and the need for robust, generalizable models. The proposed approaches advance the latest developments in UDA and ZSDA, offering practical solutions for improving segmentation performance across various medical imaging tasks.

Acknowledgements

I am sincerely grateful to all the individuals and organizations who supported and encouraged me throughout my PhD journey. This thesis would not have been possible without their valuable contributions.

First and foremost, I extend my deepest thanks to my supervisor, Associate Professor Weidong (Tom) Cai. His patient guidance, unwavering support, and continuous encouragement have been the cornerstone of this research. His mentorship has greatly influenced both my academic growth and personal development. I also wish to sincerely thank my associate supervisor, Dr. Dongnan Liu, for his expertise and thoughtful advice. His professional feedback and ongoing support significantly enhanced the quality of this work. I feel incredibly fortunate to have had such dedicated mentors guiding me through my PhD studies.

I am also thankful to my colleagues and friends, Jianan Fan, Zhiyuan Li, Yuxiang An, Hao Xu and Yui Lo for their friendship and collaboration. They made this journey both productive and enjoyable. Our many discussions, shared experiences, and moments of laughter brought light to each day.

A special acknowledgment goes to my dear friends in China: Yuqing Zhang, Yangbai Qi, Mengjie Cui, Muzhang Jin, Keyu Chen, and Haohao Zhang. Their support and companionship have been a constant source of strength. Despite the distance, their presence in my life has brought great comfort and joy. The bonds we share have been essential to my well-being throughout this journey. I would also like to extend my heartfelt thanks to my friends in Australia, Jiayi Zhou and Jiaqi Wang. Whether through thoughtful advice or shared laughter, their companionship has been a bright light throughout my academic pursuit.

My deepest gratitude is extended to my parents and family in China. I am forever thankful to my wonderful parents for their unconditional love and support. Their unwavering belief in me has been my greatest motivation and has made my dreams a reality.

Special thanks to my husband, Haoran Li, whose love and support have been my anchor throughout this process. His encouragement, patience, and understanding made this journey much smoother. I also want to express my appreciation to our beloved cat, Pami, for being a comforting presence during the long hours of research.

I dedicate this thesis to all of them.

Publications

- [1] **Canran Li**, Dongnan Liu, Jianan Fan, and Weidong Cai, “Unlocking the Potential of LLMs for Medical Domain Adaptation: A Novel Paradigm in Medical Computer Vision”, *IEEE Transactions on Medical Imaging (TMI)*. (pre-submission)
- [2] **Canran Li**, Dongnan Liu, Jianan Fan, and Weidong Cai, “Diffusion is All You Need for Domain Adaptive Medical Image Segmentation”, *International Journal of Computer Vision (IJCV)*. (pre-submission) (invite submission to BMVC2024 Special Issue)
- [3] **Canran Li**, Dongnan Liu, Jianan Fan, and Weidong Cai, “Language Model as Disentangler: Rethinking Zero-shot Domain Adaptation for Medical Image Segmentation”, *International Conference on Computer Vision (ICCV2025)*, October 19–23, Hawai’i, USA, 2025. (under review)
- [4] **Canran Li**, Dongnan Liu, and Weidong Cai, “Rethinking Domain Adaptive Optic Disc and Cup Segmentation in Fundus Image through Dynamic Diffusion Flow”, *The 35th British Machine Vision Conference (BMVC2024)*, November 25-28, Glasgow, UK, 2024.
- [5] Jianan Fan, Dongnan Liu, **Canran Li**, Hang Chang, Heng Huang, Filip Braet, Mei Chen, and Weidong Cai, “Revisiting Adaptive Cellular Recognition Under Domain Shifts: A Contextual Correspondence View”, *The 18th European Conference on Computer Vision (ECCV2024)*, September 29-October 4, Milano, Italy, 2024.
- [6] **Canran Li**, Dongnan Liu, Haoran Li, Zheng Zhang, Guangming Lu, Xiaojun Chang, and Weidong Cai, “Domain Adaptive Nuclei Instance Segmentation and Classification via

Category-aware Feature Alignment and Pseudo-labelling”, *The 25th International Conference on Medical Image Computing and Computer-Assisted Intervention* (MICCAI2022), September 18-22, Singapore, 2022.

Authorship Attribution Statement

Chapter 3 of this thesis is based on the paper “Domain Adaptive Nuclei Instance Segmentation and Classification via Category-aware Feature Alignment and Pseudo-labelling” [160] at MICCAI 2022.

I designed the study, developed the algorithm, implemented the experiments, analyzed the data, and wrote the manuscript with the co-authors.

Chapter 4 of this thesis is based on the paper “Rethinking Domain Adaptive Optic Disc and Cup Segmentation in Fundus Image through Dynamic Diffusion Flow” at BMVC 2024 and the pre-submission paper “Diffusion is All You Need for Domain Adaptive Medical Image Segmentation” at IJCV.

I designed the study, developed the algorithm, implemented the experiments, analyzed the data, and wrote the manuscript with the co-authors.

Chapter 5 of this thesis is based on the under review paper “Language Model as Disentangler: Rethinking Zero-shot Domain Adaptation for Medical Image Segmentation” at ICCV 2025 and the pre-submission paper “Unlocking the Potential of LLMs for Medical Domain Adaptation: A Novel Paradigm in Medical Computer Vision” at TMI.

I designed the study, developed the algorithm, implemented the experiments, analyzed the data, and wrote the manuscript with the co-authors.

Attest Authorship Attribution Statement

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.

Student name: Canran Li

Signature:

Date: 28 February 2025

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

Supervisor name: Weidong Cai

Signature:

Date: 28 February 2025

Contents

Declaration	ii
Abstract	iii
Acknowledgements	v
Publications	vii
Authorship Attribution Statement	ix
Attest Authorship Attribution Statement	x
Contents	xi
List of Figures	xvi
List of Tables	xxi
Chapter 1 Introduction	1
1.1 Introduction to Medical Image Segmentation	1
1.1.1 Traditional Image Segmentation Methods	2
1.1.2 Deep Learning-based Image Segmentation Methods	2
1.1.3 Medical Imaging Segmentation Applications	4
1.2 Problem Statement	5
1.3 Existing Challenges	8
1.3.1 Semantic Gaps in Medical Image Categories	8
1.3.2 Limitations of Adversarial Learning in UDA	9
1.3.3 Lack of Generalization with Unavailable Target Data	9
1.4 Thesis Contributions	10
1.4.1 UDA using Category-aware Feature Alignment and Pseudo-labelling	10

1.4.2	Dynamic Diffusion Flow for Unsupervised Domain Adaptation	11
1.4.3	Feature Disentanglement via Language Model for Domain Adaptation . . .	12
1.5	Thesis Structure	13
Chapter 2	Literature Review	15
2.1	Introduction to Deep Learning	15
2.2	Deep Learning-based Medical Image Segmentation	16
2.2.1	Nuclei Segmentation in Histopathology Images	19
2.2.2	Optic Disc and Cup Segmentation in Fundus Images	23
2.2.3	Prostate MRI Segmentation	24
2.3	Transfer Learning	26
2.3.1	Introduction to Transfer Learning	26
2.3.2	Applications of Transfer Learning	27
2.4	Domain Adaptation	29
2.4.1	Problem Definition	29
2.4.2	Traditional Domain Adaptation Algorithms	30
2.4.3	Unsupervised Domain Adaptation	32
2.4.3.1	Problem Definition	32
2.4.3.2	Deep-learning based UDA Methods	32
2.4.4	Zero-shot Domain Adaptation	35
2.4.5	Domain Adaptation in Medical Image Segmentation	37
2.4.5.1	UDA in Nuclei Segmentation	38
2.4.5.2	UDA in Fundus Image Segmentation	40
2.4.5.3	UDA in Prostate MRI Segmentation	41
2.5	Summary	42
Chapter 3	Unsupervised Domain Adaptation via Category-aware Feature Alignment and Pseudo-label	44
3.1	Introduction	45
3.1.1	Motivation	45
3.1.2	Contributions	46

3.2	Methods	47
3.2.1	Overview	47
3.2.2	Segmentation Architecture	48
3.2.3	Category-aware Feature Alignment	50
3.2.4	Nuclei-level Prototype Pseudo-labelling	52
3.3	Experiments and Results	53
3.3.1	Dataset Description	53
3.3.1.1	DigestPath (Dpath) Dataset	54
3.3.1.2	CRAG Dataset	55
3.3.1.3	GlaS Dataset	55
3.3.2	Evaluation Metrics	56
3.3.3	Implementation Details	56
3.3.4	Comparison Experiments	56
3.3.4.1	Evaluation on Nucleus Instance Segmentation	58
3.3.4.2	Evaluation on Nucleus Classification	58
3.3.5	Ablation Studies	59
3.3.6	Visualization Results	63
3.4	Discussion	63
3.4.1	Graph Search Mechanism	63
3.4.1.1	Graph-based Search Mechanism Methods	64
3.4.1.2	Experiment Results for Graph Search Mechanism	64
3.4.2	Limitations and Future Work	66
3.5	Summary	67
Chapter 4	Dynamic Diffusion Flow for Unsupervised Image Segmentation	68
4.1	Introduction	69
4.1.1	Motivation	69
4.1.2	Contributions	71
4.2	Methods	73
4.2.1	Overview of the Framework	73
4.2.2	Domain Diffusion Processing	76

4.2.3	Optimization Processing	77
4.2.4	Nash Equilibrium	78
4.2.4.1	Proof of Convergence Conditions	78
4.2.4.2	Dynamic Nash Equilibrium Strategy	80
4.3	Experiments and Results	82
4.3.1	Dataset Description	82
4.3.1.1	RIM-ONE-r3 Dataset	82
4.3.1.2	DRISHTI-GS Dataset	83
4.3.1.3	REFUGE Dataset	83
4.3.2	Experiment Details	84
4.3.2.1	Experiment Setting	84
4.3.2.2	Evaluation Metrics	84
4.3.2.3	Implementation Details	84
4.3.3	Performance of the Proposed Method	85
4.3.3.1	Results on Imbalanced UDA Settings	85
4.3.3.2	Results on Traditional UDA Settings	87
4.3.4	Ablation Study	88
4.3.5	Visualization Results	89
4.4	Discussion	92
4.4.1	Domain-specific Information Extraction	92
4.4.2	Limitations and Future Work	94
4.5	Summary	94

Chapter 5 Feature Disentanglement via Language Model for Zero-shot Domain

Adaptation 96

5.1	Introduction	97
5.1.1	Motivation	97
5.1.2	Related Work	98
5.1.2.1	Zero-shot Domain Adaptation	98
5.1.2.2	Disentangling Representation in UDA	100
5.1.3	Contributions	100

5.2	Methods	101
5.2.1	Problem Definition.....	101
5.2.2	Overview	102
5.2.3	Graph Construction	103
5.2.4	Feature Disentanglement Module (FDM)	104
5.2.5	Graph Reasoning	106
5.3	Experiments and Results	108
5.3.1	Dataset Description	108
5.3.1.1	Fundus Datasets	108
5.3.1.2	Prostate MRI Datasets	108
5.3.2	Experimental Setup	110
5.3.2.1	Evaluation Metric	110
5.3.2.2	Baseline Methods	110
5.3.2.3	Implementation Details	110
5.3.3	Comparisons with State-of-the-art Methods	112
5.3.4	Ablation Studies	114
5.3.4.1	Component Effectiveness	114
5.3.4.2	Hyperparameter Analysis	115
5.3.5	Visualization Results	116
5.4	Discussion	116
5.4.1	LLMs for Medical Domain Adaptation	118
5.4.2	Limitations and Future Work	119
5.5	Summary	119
Chapter 6	Conclusions and Future Directions	121
6.1	Conclusions	121
6.2	Future Directions	123
Bibliography		124
Appendix		160
Abbreviations		160

List of Figures

- 1.1 Some medical image segmentation examples [309]. 2
- 1.2 An example of different types of deep learning-based image segmentation (semantic segmentation, instance segmentation, and panoptic segmentation) [270]. 4
- 1.3 Visualization of significant domain shifts between source and target modalities in three different applications. Top: Prostate cancer histopathology whole slide images (WSIs) from TCGA (A) [133] and RCINJ (B). WSIs from different institutions show varying glandular distributions and staining appearances [238]. Middle: Image slices (top) and corresponding histograms (bottom) of normalized T1-weighted (a, b) and T2-weighted (c, d) MRIs from different scanners. Despite sharing similar high-level information, there are substantial differences in intensity and contrast, which segmentation algorithms are particularly sensitive [135]. Bottom: Two cardiac images from MRI and CT, along with their segmentation ground-truth. The cardiac structures—AA, LVC, and MYO—are shown in blue, purple, and brown, respectively [15]. 6
- 1.4 An overview of our contributions. In Chapter 3, we aim to solve the class-level imbalance within the target domain dataset. In Chapter 4, our goal is to address the issue of sample imbalance between different datasets. In Chapter 5, we aim to directly eliminate the phenomenon of data imbalance in domain adaptation. 10
- 2.1 The U-Net architecture (example for 32x32 pixels at the lowest resolution) is illustrated with blue boxes representing multi-channel feature maps. The number of channels is indicated above each box, while the x-y dimensions are shown at the bottom left corner. White boxes indicate copied feature maps, and the arrows depict various operations performed within the network [242]. 17
- 2.2 The architecture of DeepLabv3+. It builds on DeepLabv3 by incorporating an encoder-decoder structure. The encoder captures multi-scale contextual information

	through atrous convolution at various scales, while the straightforward yet efficient decoder refines segmentation results, particularly along object boundaries [34].	18
2.3	The architecture of Mask R-CNN for instance segmentation [103].	19
2.4	Example of nuclei segmentation. Left: Sample images from the CPM dataset. Right: Ground truth annotations, marked in green, on the corresponding images. Top row: Head and neck squamous cell carcinoma nuclei. Bottom row: Lower grade glioma [234].	20
2.5	The architecture of Hover-Net. (a) Pre-activated residual unit, (b) Dense unit. The variable m represents the number of feature maps in each residual unit. The yellow square within the input highlights the region being considered in the output. When classification labels are unavailable, only the up-sampling branches within the dashed box are used [91].	21
2.6	An example of a zoomed-in (a) normal and (b) glaucoma OD [98].	23
2.7	Example of T2-weighted transverse prostate MRI images showing apex, center, and base slices. The reference standard segmentation is highlighted in yellow, while the second observer's segmentation is shown in red [172].	25
2.8	Transfer learning categories [214].	27
2.9	The architecture of DANN consists of a deep feature extractor (green) and a deep label predictor (blue), forming a standard feed-forward network. Unsupervised domain adaptation is achieved by introducing a domain classifier (red), connected to the feature extractor through a gradient reversal layer. This layer multiplies the gradient by a negative constant during backpropagation, while the rest of the training follows the standard process, minimizing the label prediction loss for source data and the domain classification loss for all samples. The gradient reversal forces the feature distributions across domains to become as similar as possible, making them indistinguishable to the domain classifier and resulting in domain-invariant features [80].	33
2.10	An intuitive example of widely used ZSDA methods involves the use of ToI for digital image analysis and IrT for letter image analysis. The source domain consists of grayscale images, while the target domain includes color images. To train a	

model for the unseen MNIST-M dataset (the target domain data in ToI), traditional ZSDA methods introduce IrT to learn domain transformations and then transfer this knowledge to ToI [302].	35
2.11 Example image regions sampled from each of the six data sources show significant morphological differences. [92].	39
2.12 Fundus images in different datasets (REFUGE, RIM-ONE-r3 and Drishti-GS dataset).	41
2.13 Prostate MRI images in different datasets [181].	42
3.1 Overview of the proposed category-aware prototype pseudo-labelling network. Our framework is optimized through a two-stage process to address domain shift issues in nuclei instance segmentation and classification. In the first stage, we incorporate class-level feature alignment modules, which are designed to reduce the domain gap by aligning features of similar classes across the source and target domains, thereby enhancing the model’s ability to learn consistent representations. In the second stage, we introduce a pseudo-labelling process that is enhanced by nuclei-level prototypes. This approach generates pseudo labels based on these prototypes, enabling self-supervised learning on the unlabelled target images. This two-stage optimization strategy allows the model to effectively leverage unlabelled data, improving its adaptability and performance on the target domain.	49
3.2 Visualization predictions for the UDA experiment under the $D_{path} \rightarrow GlaS$ setting. Red: Eosinophil, Green: Epithelial, Yellow: Lymphocyte, Blue: Plasma, Magenta: Neutrophil, Cyan: Connective tissue. (Best to be viewed in color and zoomed-in)	61
3.3 Visualization predictions for the ablation experiments under $D_{path} \rightarrow CRAG$ setting.	62
4.1 Schematic diagram of unbalanced source and target domain samples in the UDA method. (a): Existing UDA methods align source and target distributions under the balanced data distribution scenario. (b): Existing UDA methods may result in subpar solutions with indistinct decision boundaries when the source and target	

	domain samples are imbalanced. (c): Our method can correct the decision boundary by diffusing the domain information of target samples into the source ones.	70
4.2	Overview of our proposed DDF-UDA framework. We use different colors to represent student and teacher models (best views in color).	73
4.3	Diffusion adaptation processing. We use both the feature-level target domain information I_{tN}^f and the pixel-level target domain information I_{tN}^p to update image x_i^s during the diffusion process.	75
4.4	Qualitative results of OD (green) and OC (blue) segmentation, which adapted from Drishti-GS to RIM-ONE-r3.	90
4.5	Visualization of OD (green) and OC (blue) segmentation results for the ablation study conducted from Drishti-GS to RIM-ONE-r3. (a): original images, (b): BL: Baseline, (c): BL+DM, (d): BL+DM+NE, (e): Ours.	91
5.1	Illustration of the differences between previous ZSDA methods and our approach. (a): Previous ZSDA methods introduce IrT to learn domain transformations and transfer this knowledge to the ToI. (b): Our LMaD uses a language model as a disentangler to eliminate the interference caused by domain information.	99
5.2	Overview of our proposed LMaD framework. LMaD includes two branches: a visual branch produces visual graphs and final segmentation results, and a text branch generates domain-specific and task-specific information for feature disentanglement. Best view in color. It includes three major components: visual graph construction, feature disentanglement module, and graph reasoning. The task prompt can be any sentence describing the task, i.e., “This is a fundus image for cup and disc segmentation”, and the domain prompt can be any sentence containing the domain information, i.e., “This is a fundus image from the source dataset”.	102
5.3	Proposed Feature Disentanglement Module (FDM). The visual graph is divided into K sub-graphs to measure the similarity with domain-specific and task-specific information, separately.	105
5.4	Visualization of segmentation results for different comparison algorithms (REFUGE $\rightarrow$ Drishti-GS and A $\rightarrow$ C). In fundus images, green for OD	

segmentation, and blue for OC segmentation. In MRI task, green for prostate segmentation.

List of Tables

3.1 Details of the number of nuclei under each class in the datasets we used, DigestPath, CRAG, and GlaS.	54
3.2 Experimental results on UDA nuclei instance segmentation on Dpath $\rightarrow$ CRAG task.	57
3.3 Experimental results on UDA nuclei instance segmentation on Dpath $\rightarrow$ GlaS task.	57
3.4 Experimental results on UDA nuclei classification on Dpath $\rightarrow$ CRAG task. F_c^1 , F_c^2 , F_c^3 , F_c^4 , F_c^5 and F_c^6 denote the F1 classification score for the Eosinophil, Epithelial, Lymphocyte, Plasma, Neutrophil and Connective tissue, respectively. F_{avg} denotes the average of all the F1-score for the classification under each category.	58
3.5 Experimental results on UDA nuclei instance segmentation by transferring from <i>Dpath</i> to the <i>Glas</i> dataset.	59
3.6 Computational complexity analysis, which indicates our proposed modules can introduce performance gain but bring negligible auxiliary cost.	60
3.7 Graph-based method experimental results on UDA nuclei instance segmentation on Dpath $\rightarrow$ CRAG task.	64
3.8 Graph-based method experimental results on UDA nuclei instance segmentation on Dpath $\rightarrow$ GlaS task.	65
3.9 Graph-based method experimental results on UDA nuclei classification on Dpath $\rightarrow$ CRAG task.	66
3.10 Graph-based method experimental results on nuclei classification by transferring from Dpath $\rightarrow$ Glas dataset.	66
4.1 Statistics of three fundus datasets used for the proposed method.	82
4.2 Comparison OD/OC segmentation results when adapted from Drishti-GS to RIM-ONE-r3.	85

4.3 Comparison OD/OC segmentation results when adapted from RIM-ONE-r3 to Drishti-GS.	85
4.4 Quantitative comparison OD/OC segmentation results when adapted from REFUGE to Drishti-GS and adapted from REFUGE to RIM-ONE-r3.	86
4.5 Ablation study results of adapting Drishti-GS to RIM-ONE-r3 and adapting REFUGE to Drishti-GS.	87
4.6 Additional ablation study results of adapting Drishti-GS to RIM-ONE-r3.	88
4.7 Experiment results of adapting Drishti-GS to RIM-ONE-r3 and adapting REFUGE to Drishti-GS of different domain-specific information extraction methods.	93
5.1 Statistics of five prostate MRI datasets used for the proposed method.	109
5.2 Comparison with state-of-the-art methods on the OD/OC Segmentation task when adapted from RIM-ONE-r3 to Drishti-GS.	111
5.3 Comparison with state-of-the-art methods on the OD/OC Segmentation task when adapted from Drishti-GS to RIM-ONE-r3.	111
5.4 Comparison with state-of-the-art methods on the OD/OC Segmentation task When adapted from REFUGE to Drishti-GS.	112
5.5 Comparison with state-of-the-art methods on the OD/OC Segmentation task When adapted from REFUGE to RIM-ONE-r3.	113
5.6 Comparison with state-of-the-art methods on the ZSDA prostate MRI segmentation tasks (Site A as source domain).	113
5.7 Comparison with state-of-the-art methods on the ZSDA prostate MRI segmentation tasks (Site B as source domain).	114
5.8 Ablation study of key components on REFUGE $\rightarrow$ Drishti-GS adaptation task.	114
5.9 Ablation study for graph node number on REFUGE $\rightarrow$ Drishti-GS adaptation task.	115
5.10 Ablation study for N on REFUGE $\rightarrow$ Drishti-GS adaptation task.	116
5.11 Ablation study for λ on REFUGE $\rightarrow$ Drishti-GS adaptation task.	116

Introduction

1.1 Introduction to Medical Image Segmentation

Image segmentation is a fundamental task in computer vision, involving the division of an image into meaningful regions or segments [77, 341]. By simplifying image representation, segmentation facilitates easier analysis and interpretation. In medical imaging, segmentation is particularly critical, as it isolates essential anatomical structures, tumors, or regions of interest, enabling precise measurements and playing a key role in diagnosis and treatment planning [221, 250]. In addition, medical image segmentation is also crucial to the success of modern Computer-Aided Diagnosis (CAD) systems, which have transformed medical imaging by offering advanced tools for analyzing medical data and supporting clinical decision-making [59, 276].

Figure 1.1 shows some medical image segmentation examples. In the datasets presented in Figure 1.1, KITS [105] (Kidney Tumor Segmentation), LITS [18] (Liver Tumor Segmentation), MSD [262] (Medical Segmentation Decathlon), TNSCUI [354] (Thyroid Nodule Segmentation Challenge in Ultrasound Images), STARE [113] (Structured Analysis of the Retina), ISIC [96] (International Skin Imaging Collaboration), CHAOS [138] (Combined Healthy Abdominal Organ Segmentation), and BRATS [197, 11, 12] (Brain Tumor Segmentation) are widely utilized in the field of medical image segmentation. These datasets encompass a diverse range of anatomical structures and imaging modalities, including computed tomography (CT), magnetic resonance imaging (MRI), ultrasound, fundus photography, and dermatological imaging, thereby serving as critical benchmarks for the development and evaluation of automated segmentation methodologies.

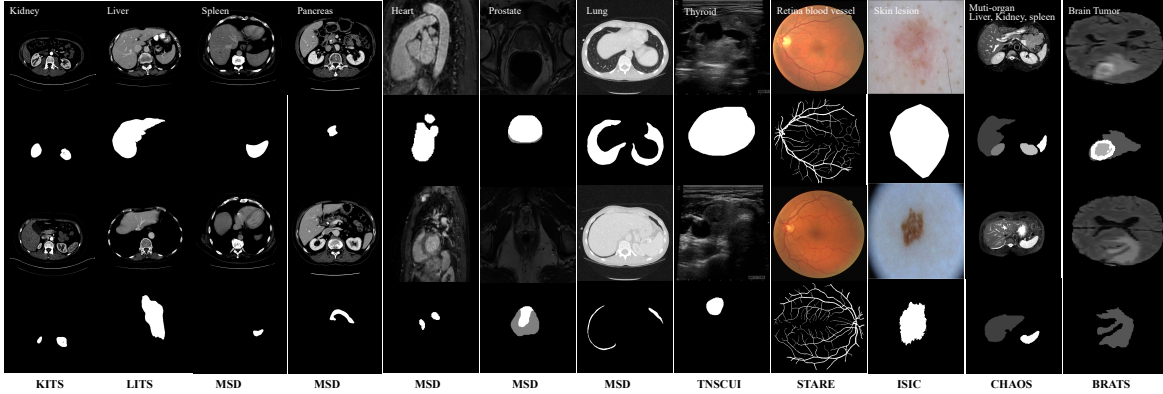


FIGURE 1.1: Some medical image segmentation examples [309].

1.1.1 Traditional Image Segmentation Methods

Traditional segmentation methods, including techniques such as thresholding, region growing, and edge detection, formed the foundation of early medical image analysis [348, 217]. These methods relied on fundamental principles to identify boundaries within images based on pixel intensity, texture, or gradients. The thresholding method involves setting a specific pixel intensity value to distinguish foreground from background, which makes it useful for segmenting regions with distinct intensity levels. Region growing starts from a seed point and expands to include neighboring pixels that share similar properties, making it effective for grouping connected regions. Edge detection algorithms focus on identifying discontinuities in intensity, helping to outline the boundaries of anatomical structures or objects within the image [356].

While these traditional methods are critical in advancing early medical image segmentation, they often exhibit performance limitations. Their effectiveness is heavily dependent on the quality of the input data, and they are particularly sensitive to image noise and variations in contrast or lighting conditions [64, 224].

1.1.2 Deep Learning-based Image Segmentation Methods

Recently, deep learning techniques have been widely adopted in image segmentation, offering effective solutions to handle the rapidly growing volume of data [255, 46, 78]. These

techniques have driven significant progress across various applications [164, 164, 54, 326, 344, 329, 165, 231, 84] in computer vision tasks. In the medical field, where high precision and expertise are essential, deep learning-based segmentation has demonstrated significant improvements [321, 275].

Currently, deep learning-based image segmentation methods are generally classified into three categories: semantic segmentation [7, 151], instance segmentation [31, 36], and panoptic segmentation [345, 167]. Figure 1.2 shows an example of different types of deep image segmentation.

Semantic segmentation assigns a predefined class to each pixel in an image, making it useful for tasks such as identifying overall structures within an image. In medical imaging, semantic segmentation is applied to segment organs, tissues, or other anatomical structures by labeling each pixel according to specific body parts or regions [183, 224]. Instance segmentation extends semantic segmentation by not only classifying each pixel but also distinguishing between individual objects of the same class. This approach is valuable in medical imaging for isolating multiple instances of similar structures, such as different lesions or tumors, thus providing more detailed and clinically relevant information [195, 289]. Panoptic segmentation combines the strengths of both semantic and instance segmentation by offering a comprehensive understanding, which labels all pixels while also distinguishing distinct objects. This method is especially beneficial in complex medical images, where accurately segmenting overlapping or closely situated anatomical features is crucial [169].

By leveraging advanced algorithms, deep learning techniques have significantly improved the accuracy, speed, and efficiency of medical image segmentation. These improvements allow for more precise identification of anatomical structures and abnormalities, which are crucial for both diagnostic and therapeutic applications [198, 65]. As a result, the integration of deep learning into medical image segmentation is leading to the development of more reliable solutions, improving patient care and clinical outcomes [81, 204].

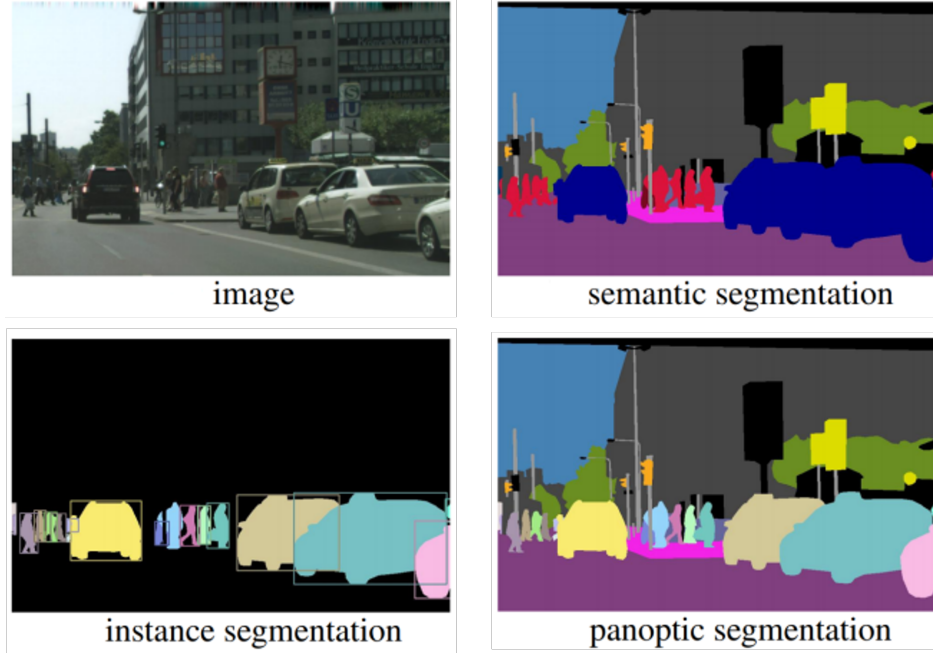


FIGURE 1.2: An example of different types of deep learning-based image segmentation (semantic segmentation, instance segmentation, and panoptic segmentation) [270].

1.1.3 Medical Imaging Segmentation Applications

Medical image segmentation is an important component in numerous clinical and research applications, encompassing the identification and delineation of anatomical structures, pathological regions, and regions of interest. As discussed earlier, traditional image segmentation methods such as thresholding, edge detection, and region-based segmentation form the basis for early developments in medical imaging [341, 270]. However, these methods face limitations in handling the complexity and variability often encountered in medical data. Deep learning-based approaches have significantly advanced segmentation accuracy and robustness, addressing these limitations by leveraging large datasets and powerful computational models.

The advent of deep learning has marked a significant paradigm shift in the field of medical image segmentation. Convolutional neural networks (CNNs) have become widely adopted for their ability to model intricate patterns and relationships within medical images. In particular, the U-Net architecture and its derivatives have proven highly effective [242, 50, 210]. Some other advanced architectures such as DeepLab [33], Mask R-CNN [103] and

Segment Anything Model (SAM) [145]. These deep learning-based methods offer superior performance across a range of segmentation tasks. A detailed analysis of these methods is provided in Chapter 2.2.

This thesis focuses on three key segmentation applications, including nuclei segmentation in histopathology images, optic disc (OD) and optic cup (OC) segmentation in fundus images, and prostate MRI segmentation. Nuclei segmentation is crucial for cancer diagnosis, tackling the challenges of staining variability and complex cell morphology [111]. OD and OC segmentation is essential for glaucoma assessment, addressing difficulties such as variations in illumination and anatomy [3]. Prostate MRI segmentation aids in prostate cancer diagnosis, where anatomical variability and imaging artifacts complicate the task [63]. These three applications are then discussed in detail in Chapter 2.2.1, Chapter 2.2.2 and Chapter 2.2.3.

1.2 Problem Statement

Deep learning methods are primarily based on supervised learning, which requires training models using large labeled datasets. In this approach, each input is matched with its correct output, and the model learns by minimizing errors in its predictions [245]. Supervised learning has achieved great success in many fields, particularly in image segmentation, due to its ability to capture detailed patterns and make accurate predictions [294, 323]. These models are highly effective when sufficient labeled data is available, allowing them to perform well in medical image segmentation. However, the main challenge with supervised learning is its reliance on large amounts of labeled data. Collecting these datasets is often time-consuming and costly, especially in specialized areas like medical imaging, where expert annotations are required [369, 223, 184]. In addition, there are significant differences between medical images from different domains (see Figure 1.3). When attempting to apply pre-trained deep learning models directly to new medical datasets with different distributions, the models often experience a performance drop due to domain bias. This reliance on large labeled datasets limits the generalization ability of supervised learning in medical fields, driving interest in approaches that reduce the need for labeled data while still offering reliable results.

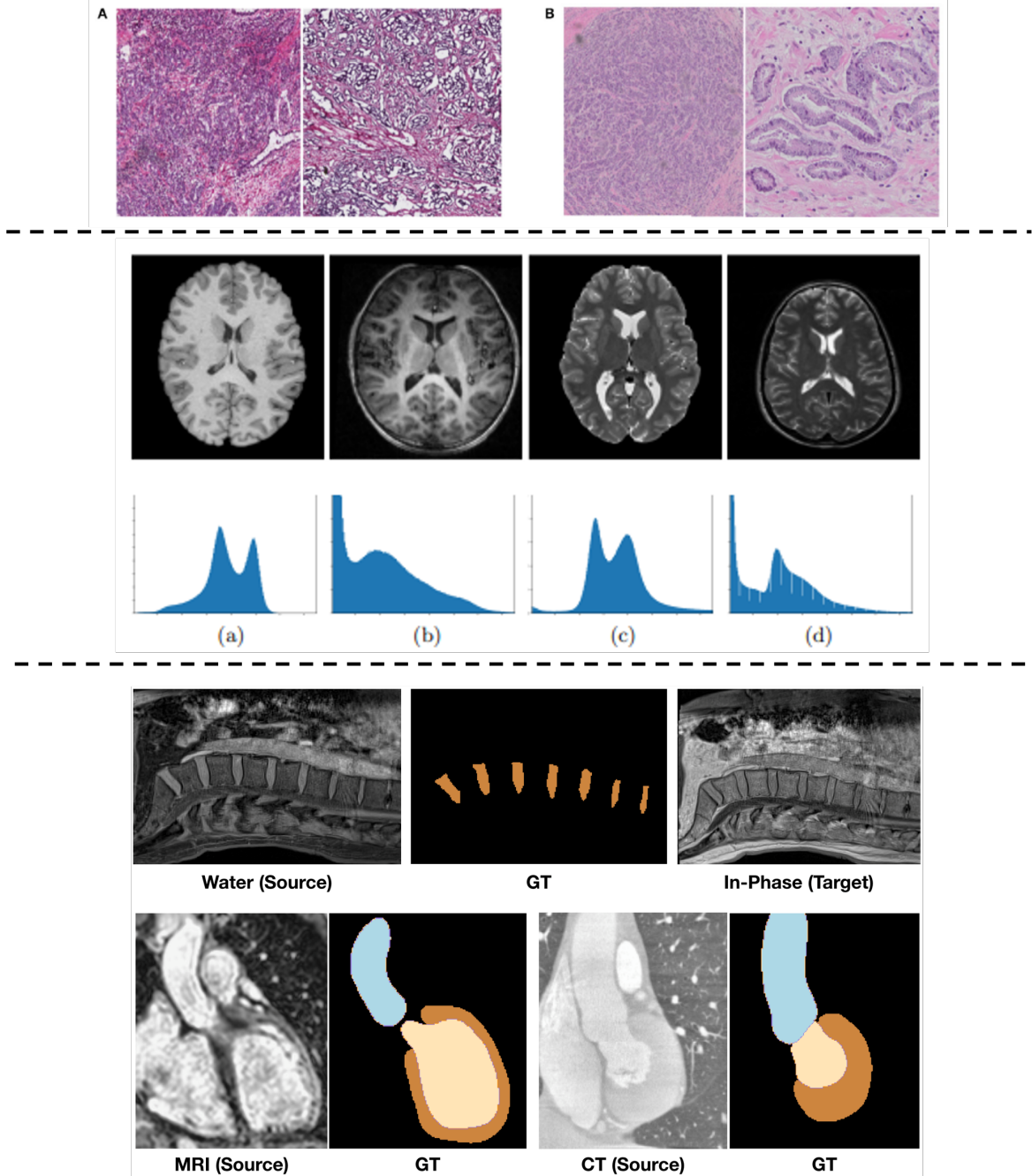


FIGURE 1.3: Visualization of significant domain shifts between source and target modalities in three different applications. Top: Prostate cancer histopathology whole slide images (WSIs) from TCGA (A) [133] and RCINJ (B). WSIs from different institutions show varying glandular distributions and staining appearances [238]. Middle: Image slices (top) and corresponding histograms (bottom) of normalized T1-weighted (a, b) and T2-weighted (c, d) MRIs from different scanners. Despite sharing similar high-level information, there are substantial differences in intensity and contrast, which segmentation algorithms are particularly sensitive [135]. Bottom: Two cardiac images from MRI and CT, along with their segmentation ground-truth. The cardiac structures—AA, LVC, and MYO—are shown in blue, purple, and brown, respectively [15].

To address the challenges of data scarcity, deep transfer learning methods have been introduced [122, 367]. Transfer learning allows models to leverage knowledge gained from one domain and apply it to another related domain, reducing the need for extensive labeled data in the new domain [53, 290]. As one of the most important subareas of transfer learning, domain adaptation aims to reduce the distribution differences between different domains. It specifically focuses on adapting a model trained in one domain (the source domain) to perform well in a different but related domain (the target domain) without extensive retraining [305, 150]. This is especially important in medical image segmentation, where variations in data distributions, such as differences in imaging devices, patient populations, or anatomical regions, can significantly affect model performance. By effectively adapting models to these new domains, domain adaptation techniques help overcome the challenges associated with data scarcity and variability, enhancing the robustness and generalizability of deep learning models in medical image segmentation tasks [61, 349, 99].

Common deep domain adaptation methods can be broadly classified into four main categories: statistic-based approaches, adversarial learning approaches, reconstruction-based approaches, and self-learning approaches.

Statistic-based approaches [97, 229] aim to reduce the domain shift by aligning the statistical distributions between the source and target domains. These methods typically rely on metrics such as Maximum Mean Discrepancy (MMD) or other statistical measures to minimize discrepancies between feature distributions, thereby facilitating more effective knowledge transfer. Adversarial learning approaches [19, 351, 191, 318] leverage the power of adversarial networks to train a model that becomes domain-invariant. These methods typically introduce a domain discriminator that is trained to distinguish between source and target domain features, while the feature extractor is optimized to confuse the discriminator by generating domain-agnostic features. Reconstruction-based approaches [28, 332, 118] utilize reconstruction tasks to enforce structural consistency between the source and target domains. By requiring the model to accurately reconstruct domain-specific data, these approaches encourage the learning of domain-shared representations that are robust to domain shifts, facilitating more effective adaptation. Self-learning approaches [285, 112] rely on the iterative refinement of

model predictions in the target domain. These methods generate pseudo-labels for unlabeled target domain samples, which are then used as supervision to further optimize the model. Through iterative updates, the model progressively improves its performance on the target domain, even in the absence of ground-truth annotations, by leveraging its own predictions as a form of weak supervision. Details of these deep domain adaptation methods can be seen in Chapter 2.4.3.2.

1.3 Existing Challenges

While existing deep learning-based domain adaptation methods have shown remarkable results in medical image segmentation, they also face significant challenges. We highlight three challenges faced when employing deep domain adaptation approaches in medical image segmentation.

1.3.1 Semantic Gaps in Medical Image Categories

In medical image segmentation, a key challenge faced by UDA models is the significant semantic differences between various categories or regions within the images. Although several methods have recently been proposed for UDA segmentation tasks in medical images [175, 177, 281, 334, 232], their adaptation strategies are class-agnostic. In medical imaging, whether from histopathology, MRI, or other modalities, exhibit substantial variability across different structures, such as organs, tissues, or even distinct nuclei categories. This is especially evident in histopathology images, where different categories of nuclei can display considerable semantic variation [175, 334]. The large semantic gaps between categories make it difficult to generalize UDA models to different domains for tasks like segmentation or classification. As a result, developing UDA models that can effectively overcome these semantic differences and perform accurately across various types of medical images remains a challenging and unresolved problem.

1.3.2 Limitations of Adversarial Learning in UDA

Additionally, adversarial learning strategies are the most widely used approaches in UDA tasks for medical image segmentation [327, 142, 209, 304]. Most existing deep domain adaptation methods leverage adversarial learning to generate domain-invariant features, aiming to bridge the domain gap and achieve impressive results [311, 158, 338]. However, these adversarial UDA methods heavily rely on maintaining a balance between the source and target datasets. When this balance is disrupted, the effectiveness of UDA methods can be compromised [160]. If the source dataset is significantly larger than the target dataset, the model may develop a bias toward the source samples due to the inherent strong inductive bias in deep learning models. This bias can result in poor generalization to the target domain, where adaptation is most needed. On the other hand, if the source dataset is much smaller than the target dataset, the model might suffer from undertraining, making it difficult to achieve reliable representation and performance in the target domain [119, 146]. Therefore, it is crucial to design a domain adaptation approach that does not depend on the size balance between datasets, ensuring robust performance across domains regardless of dataset size discrepancies.

1.3.3 Lack of Generalization with Unavailable Target Data

Moreover, UDA methods still require specific target domain data during training to achieve good performance in that domain [361, 256]. In real-world scenarios, predicting which data will form the target domain can be challenging, limiting the generalizability of these methods. Additionally, in certain medical environments, obtaining target domain images can be difficult due to privacy concerns or data scarcity [183, 309]. Recently, various zero-shot domain adaptation (ZSDA) methods have been introduced for natural images, aiming to allow models to adapt to unseen target domains that are not accessible during training [219, 302, 127, 66]. However, these ZSDA methods face significant limitations in medical imaging, where data is often scarce. Additionally, existing methods do not demonstrate strong generalization ability for completely unseen target domains. These challenges highlight the need for more robust and adaptable methods for medical image segmentation in data-efficient scenarios.

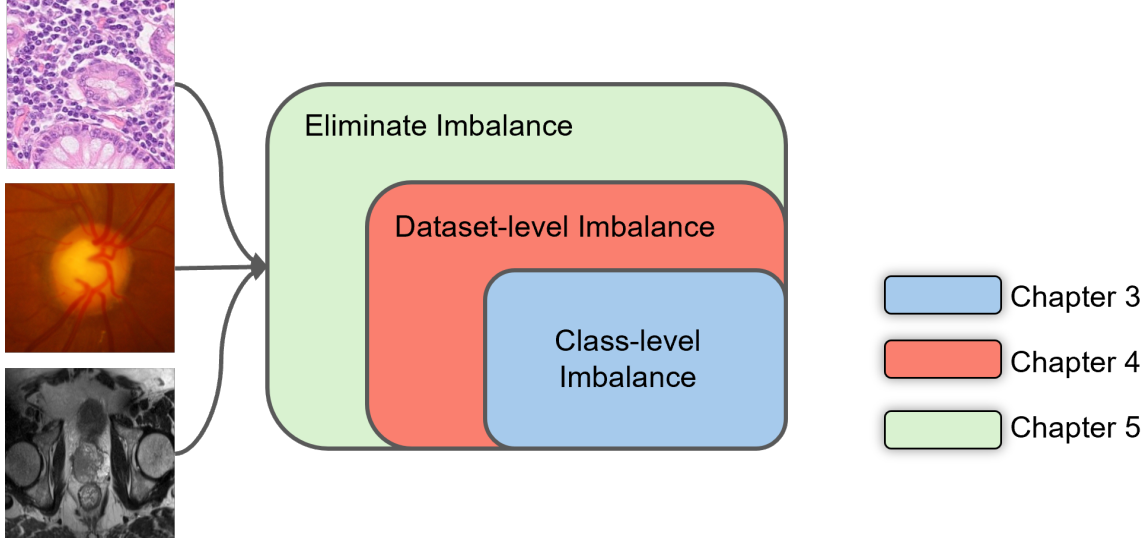


FIGURE 1.4: An overview of our contributions. In Chapter 3, we aim to solve the class-level imbalance within the target domain dataset. In Chapter 4, our goal is to address the issue of sample imbalance between different datasets. In Chapter 5, we aim to directly eliminate the phenomenon of data imbalance in domain adaptation.

1.4 Thesis Contributions

In this thesis, we introduce three novel cross-domain deep learning methods proposed for different medical image segmentation tasks under data-efficient scenarios. An overview of our contributions is shown in Figure 1.4. The main contributions of this thesis are summarized as follows.

1.4.1 UDA using Category-aware Feature Alignment and Pseudo-labelling

Unlike natural images, histopathology images exhibit significant semantic gaps between different categories of nuclei. The challenge of developing generalized unsupervised domain adaptation (UDA) models capable of achieving precise nuclei instance segmentation or classification across diverse datasets remains under-explored. Our main contributions (Chapter 3) introduce a novel deep neural network named the Category-Aware Feature Alignment and Pseudo-Labeling Network (CAPL-Net), specifically designed for UDA in nuclei instance

segmentation and classification. Our approach tackles the significant challenges associated with domain shifts and category variations inherent in histopathology images. To address these issues, we first propose a category-level feature alignment module that employs dynamically learnable trade-off weights, enabling balanced alignment of both shared and category-specific features across domains. This module effectively reduces domain shifts by aligning features at the category-specific level. Additionally, we enhance model performance on the target domain data through self-supervised training using pseudo labels derived from nuclei-level prototype features. This pseudo-labelling strategy allows the model to learn from unlabeled target data, thereby improving its generalization capabilities. Extensive experiments on cross-domain nuclei instance segmentation and classification tasks validate the effectiveness of CAPL-Net, showing that our approach significantly outperforms existing state-of-the-art UDA methods and achieves outstanding performance improvements. These results underscore the potential of CAPL-Net as a robust solution for UDA in the diverse field of histopathology image analysis.

1.4.2 Dynamic Diffusion Flow for Unsupervised Domain Adaptation

Most existing UDA approaches for fundus image segmentation rely on adversarial learning, which depends heavily on maintaining a balance between the source and target datasets to align features effectively. Existing UDA methods perform poorly when there is an imbalance in the data volume between the source and target domains. We propose a diffusion-based framework called Dynamic Diffusion Flow Unsupervised Domain Adaptation (DDF-UDA) for cross-domain optic disc (OD) and optic cup (OC) segmentation in fundus images. This framework is designed to address the challenges of domain shifts between source and target datasets by introducing a novel adaptation module that utilizes diffusion processes at both feature and pixel levels. This dual-level diffusion approach helps to mitigate cross-domain discrepancies by aligning features and pixel distributions between domains more effectively. To further refine this alignment, we introduce a Nash equilibrium adjustment strategy that dynamically modifies the diffusion steps. This strategy allows for the adaptation of domain-specific information in the source images while minimizing alterations to their underlying

content, ensuring that the essential structural information is preserved. By carefully balancing the diffusion steps, our approach achieves a more precise domain adaptation, leading to improved segmentation outcomes. Experimental evaluations on public datasets demonstrate that DDF-UDA effectively leverages unlabeled target data to enhance segmentation performance, achieving state-of-the-art results in OD and OC segmentation tasks.

1.4.3 Feature Disentanglement via Language Model for Domain Adaptation

Cross-domain deep learning models are crucial for computer-aided diagnosis. Recent advances in UDA have sought to address this challenge by utilizing transfer learning techniques. While these methods are generally efficient and effective, they still rely on access to target domain images during training, which is an often impractical requirement in real-world medical scenarios due to limited data availability. Some approaches have introduced zero-shot learning strategies for domain adaptation, aiming to transfer the knowledge acquired from unrelated tasks to unseen target domains. However, zero-shot domain adaptation (ZSDA) faces significant challenges in medical imaging, where data scarcity is common. Moreover, current methods struggle to generalize effectively to entirely unseen domains.

In this contribution (Chapter 5), we propose a novel framework called the Language Model as a Feature Disentangler (LMaD) designed for zero-shot domain adaptation (ZSDA) in medical image segmentation tasks. This framework leverages the power of language models to address the challenges of domain shifts without requiring access to target domain data during training. Specifically, we first introduce a graph learning mechanism to construct a visual graph that captures the relationships among visual features extracted from medical images. This visual graph serves as a foundation for understanding the complex interactions between different image features. We then utilize a pre-trained language model to generate task-specific and domain-specific feature embeddings. These embeddings enable us to perform feature disentanglement using the visual graph, effectively eliminating interference caused by domain-specific information and allowing the model to focus on task-specific features. Furthermore, we integrate a graph convolutional network (GCN) to infer higher-order

relationships among task-specific features within the visual graph. This GCN enhances the contextual understanding of the images, providing a more robust representation that captures intricate details and dependencies. We conduct extensive experiments on OD/OC fundus image segmentation and prostate MRI segmentation to validate the effectiveness of our LMaD framework. The experimental results demonstrate that our approach significantly improves the generalization ability of ZSDA models, achieving state-of-the-art performance across these tasks. These findings underscore the strong potential of LMaD for cross-domain knowledge transfer in medical image segmentation, offering a promising solution for overcoming the limitations associated with domain adaptation in data-scarce scenarios.

1.5 Thesis Structure

This thesis is organized into six detailed chapters, focusing on medical image segmentation under data-efficient scenarios.

Chapter 1 introduces the challenges of medical image segmentation in data-efficient scenarios and summarizes the major contributions of this thesis. It also provides an overview of the thesis structure, guiding readers through the subsequent chapters.

Chapter 2 presents a comprehensive literature review, exploring the impact of deep learning on various medical image segmentation tasks and detailing the main topics of this thesis, including the definition of domain adaptation and the existing algorithms. It also reviews the current methods for applying domain adaptation to different medical image segmentation tasks across various datasets.

Chapter 3 introduces a novel category-level adaptation framework specifically designed for cross-domain nuclei instance segmentation and classification. This chapter includes an introduction, the proposed method, experimental setup, results, a discussion of further exploration of this work and limitations, and concludes with a summary of the findings.

Chapter 4 presents a cross-domain information fusion strategy through a diffusion model. It follows a similar structure, starting with an introduction, followed by the method, experimental results, detailed discussion, and a chapter summary.

Chapter 5 focuses on a novel framework that utilizes a language model as a feature disentangler for zero-shot domain adaptation in medical image segmentation tasks. It follows the same structural format, covering the introduction, methodological details, experimental results, and an in-depth discussion with an extended method based on this work, concluding with a chapter summary.

Chapter 6 summarizes the key findings of this thesis and discusses possible future directions for further research.

Literature Review

2.1 Introduction to Deep Learning

Deep learning is a subset of machine learning, and has significantly impacted various fields [156], including computer vision [55, 188, 236, 284], natural language processing [292, 57, 178], speech recognition [100, 1, 5, 10], and multimodal analysis [207, 13, 299].

The concept of deep learning was proposed by Hinton *et al.* through their work on Deep Belief Networks (DBNs) [107], which demonstrated that multilayer neural networks could learn complex representations of data using a layer-by-layer unsupervised learning approach. It uses neural networks with multiple layers to learn patterns from large datasets, thereby automating the feature extraction process. This contrasts with traditional machine learning, where feature engineering often requires manual effort and domain expertise [149, 131]. The core strength of deep learning relies on its ability to learn hierarchical representations from data, making it to perform complex image processing tasks such as image classification [274, 266, 296], object detection [239, 52, 279], and segmentation [254, 188, 357, 42, 145].

In recent years, various deep learning architectures, such as Convolutional Neural Networks (CNNs) [152, 261, 102, 93], Recurrent Neural Networks (RNNs) [246, 273, 9], and more advanced models like Generative Adversarial Networks (GANs) [89, 225] and Transformers [62, 24], have been developed and applied across numerous image processing tasks. Among these, CNNs are particularly remarkable in image processing due to their ability to efficiently capture spatial hierarchies in visual data. CNNs exploit spatial hierarchies in images through convolutional layers, which apply filters to capture local patterns [152].

CNNs-based architectures like AlexNet [152], VGGNet [261], and ResNet [102] have demonstrated significant advancements in image recognition tasks. AlexNet [152] is a breakthrough CNN architecture that used five convolutional layers and three fully connected layers, introducing ReLU activation functions and dropout for regularization, which greatly improved training speed and accuracy. VGGNet [261] takes a different approach by using smaller convolution filters and increasing the network depth to 16-19 layers, which allows it to capture more complex features but also increases computational demands. ResNet [102] tackled the difficulties of training very deep networks by using skip connections, which allow gradients to flow directly through the network, thus preventing the vanishing gradient problem and enabling the training of extremely deep networks with improved performance.

As the availability of computational resources and annotated datasets has increased, deep learning models have become more powerful in fields that rely heavily on image analysis, such as medical imaging [251, 235, 298]. Deep learning has been applied to a wide range of tasks in medical imaging, including classification [22, 312, 230], recognition [237], counting [315, 347], and segmentation [265, 282]. Among these, medical image segmentation is one of the most critical tasks in the field of medical imaging, as it plays a pivotal role in computer-aided diagnosis and treatment planning [200, 275]. Accurate segmentation of medical images enables precise recognition of anatomical structures, pathological regions, and other critical features, significantly improving the effectiveness of various clinical applications [26, 268]. It facilitates early detection and monitoring of diseases, improves surgical planning, and supports the development of personalized treatment strategies. By enabling more reliable and reproducible image analysis, medical image segmentation has a profound impact on the overall quality and efficiency of computer-aided medicine [171, 21].

2.2 Deep Learning-based Medical Image Segmentation

Medical image segmentation involves the partitioning of medical images into different regions or structures, such as tissues, organs, or lesions [201, 155, 362]. This process is essential for numerous clinical applications, including disease diagnosis, treatment planning, and

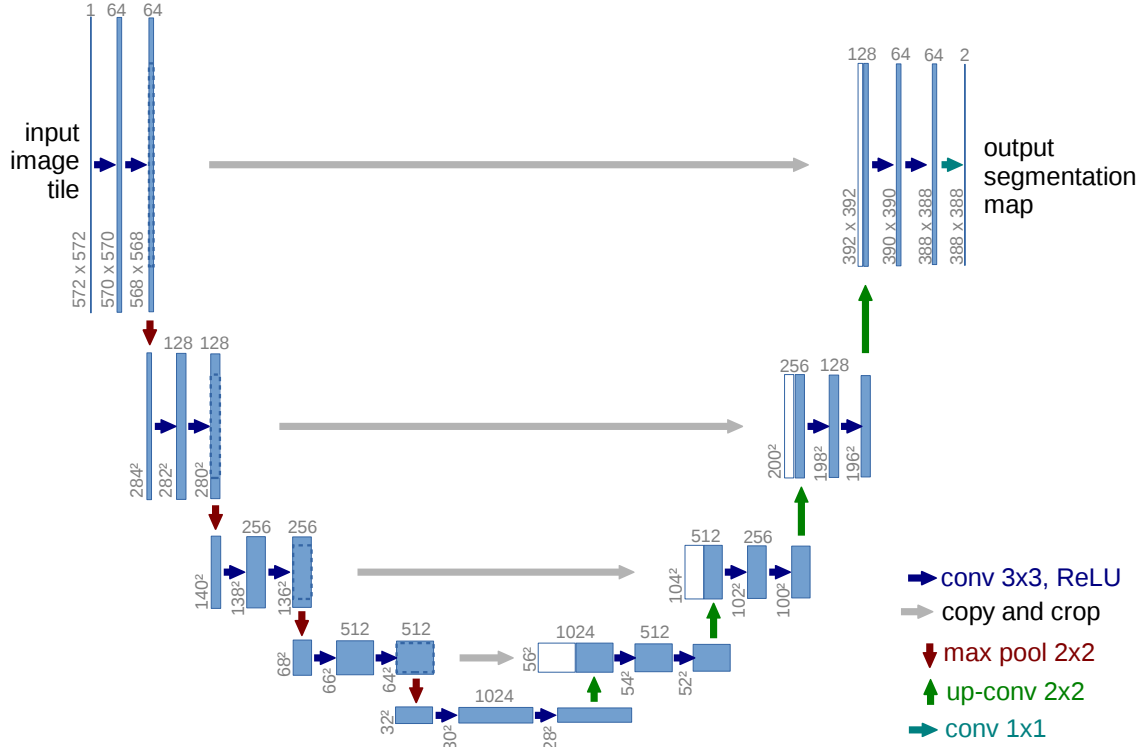


FIGURE 2.1: The U-Net architecture (example for 32x32 pixels at the lowest resolution) is illustrated with blue boxes representing multi-channel feature maps. The number of channels is indicated above each box, while the x-y dimensions are shown at the bottom left corner. White boxes indicate copied feature maps, and the arrows depict various operations performed within the network [242].

monitoring [185, 124]. Traditional segmentation methods, such as thresholding [248, 73], region growing [222, 352], and active contour models [182, 41], often struggle with the complexity and variability of medical images, making deep learning-based approaches highly attractive for this task [199].

Deep learning-based medical image segmentation primarily utilizes CNNs due to their excellent performance in extracting spatial features from images [153]. A landmark architecture in this field is the U-Net [242], which is specifically designed for biomedical image segmentation. The architecture of U-Net is shown in Figure 2.1. U-Net features an encoder-decoder structure with symmetric skip connections, allowing it to capture both local and global features of an image. This architecture has been widely adopted and modified for various segmentation tasks, demonstrating robust performance across different imaging modalities.

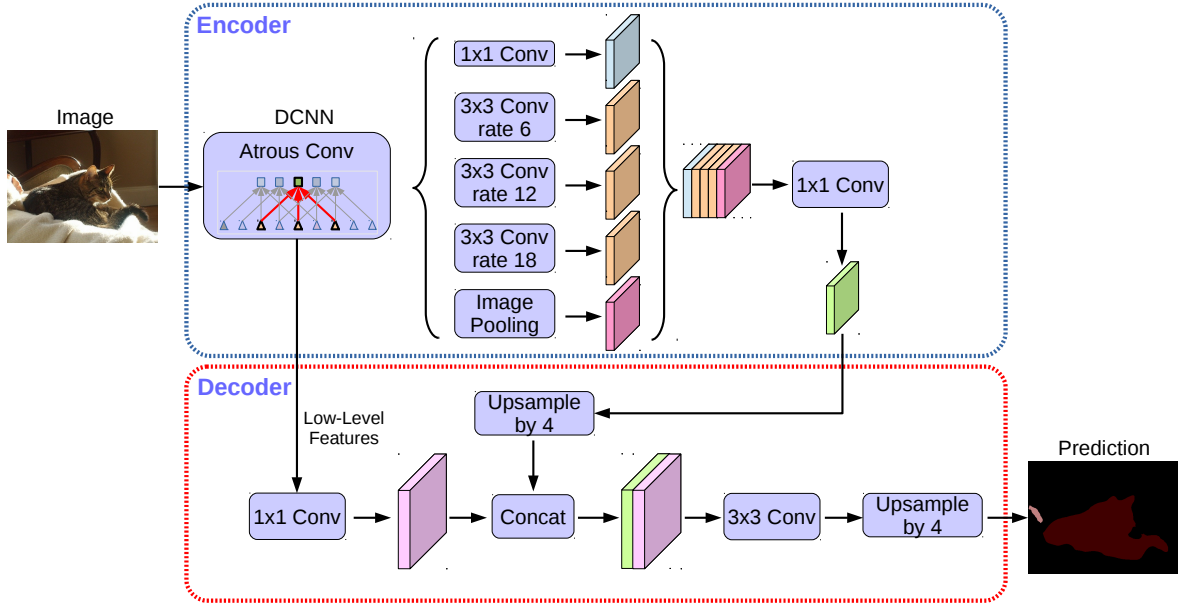


FIGURE 2.2: The architecture of DeepLabv3+. It builds on DeepLabv3 by incorporating an encoder-decoder structure. The encoder captures multi-scale contextual information through atrous convolution at various scales, while the straightforward yet efficient decoder refines segmentation results, particularly along object boundaries [34].

Several variations of U-Net [260], such as 3D U-Net [50], ResU-Net [58], and Attention U-Net [210], have been developed to address specific challenges in medical image segmentation. For instance, 3D U-Net [50] extends the U-Net architecture to three-dimensional data, making it suitable for volumetric segmentation tasks like brain MRI analysis. ResU-Net [58] integrates residual connections inspired by the ResNet architecture to improve the training of deeper networks, enhancing segmentation accuracy. Attention U-Net [210] incorporates attention mechanisms to enable the model to focus on the most relevant regions of the image, which is particularly useful in medical imaging where important structures can be small or have low contrast.

In addition to U-Net variants, other advanced architectures such as DeepLab [33] and Mask R-CNN [103] have also been adapted for medical image segmentation. DeepLab [33] employs atrous convolutions and pyramid pooling to capture multi-scale contextual information, which helps in accurately segmenting objects of varying sizes (see Figure 2.2). Mask R-CNN [103] is originally designed for object detection and segmentation in natural images and has been

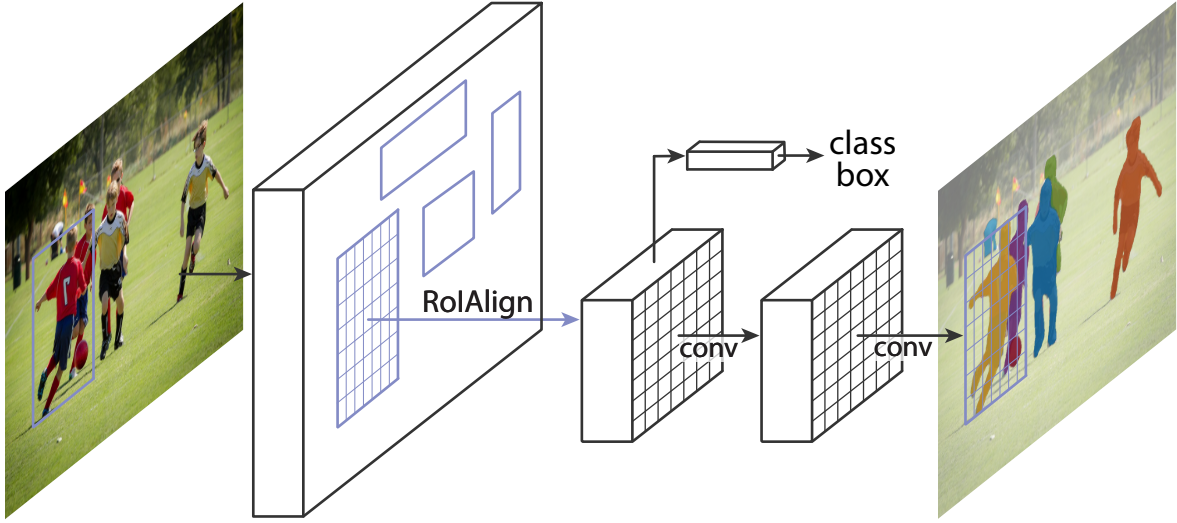


FIGURE 2.3: The architecture of Mask R-CNN for instance segmentation [103].

effectively applied to medical images for tasks requiring instance segmentation, such as identifying individual lesions or anatomical structures (see Figure 2.3). Recently, the Segment Anything Model (SAM) [145] also has gained attention in the field of medical segmentation due to its ability to generalize across a wide variety of image types with minimal manual intervention.

In this section, we discuss three different types of medical image segmentation applications used in this thesis, including nuclei segmentation in histopathology images, optic disc and cup segmentation in fundus images, and prostate MRI segmentation.

2.2.1 Nuclei Segmentation in Histopathology Images

Nuclei segmentation in histopathology images is a crucial task in medical image analysis, particularly in the diagnosis and treatment of various diseases, including cancer [293, 111]. Histopathology involves the microscopic examination of tissue samples to identify abnormal cell structures that may indicate disease. The accurate identification and segmentation of cell nuclei are fundamental for pathologists to assess the tissue architecture, cellular morphology, and spatial relationships within the tissue. This information is essential for diagnosing conditions, grading cancer severity, and evaluating the effectiveness of treatments [306, 101].

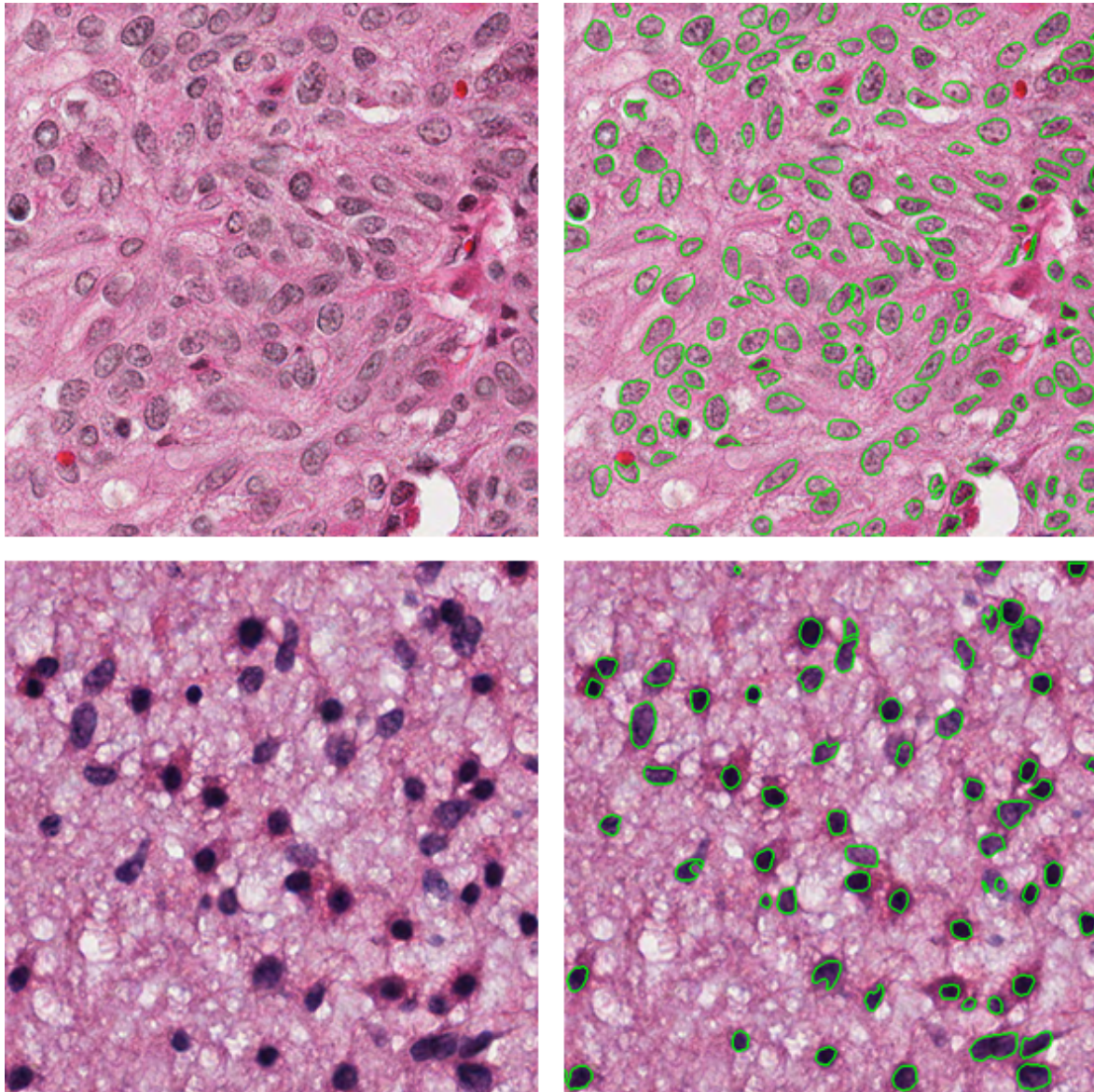


FIGURE 2.4: Example of nuclei segmentation. Left: Sample images from the CPM dataset. Right: Ground truth annotations, marked in green, on the corresponding images. Top row: Head and neck squamous cell carcinoma nuclei. Bottom row: Lower grade glioma [234].

Nuclei segmentation enables the extraction of quantitative features, such as nuclear size, shape, density, and texture, which are critical biomarkers in many diagnostic and prognostic applications. Automated nuclei segmentation can significantly reduce the workload of pathologists by providing rapid and consistent analysis, thus enhancing the accuracy and efficiency of clinical workflows. However, the task is challenging due to the variability in nuclear size,

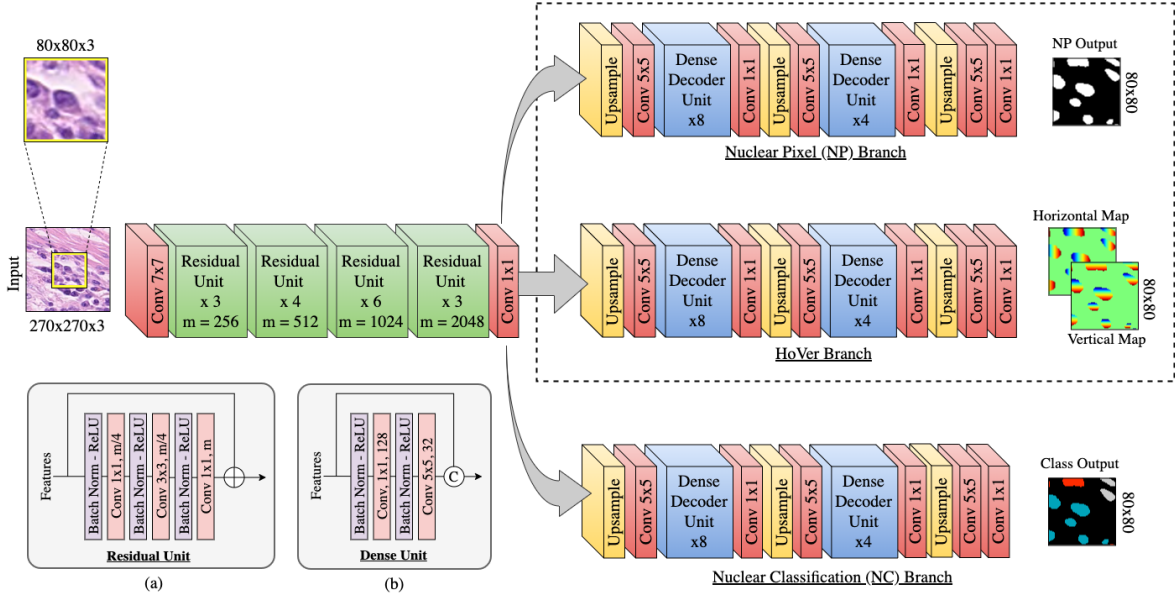


FIGURE 2.5: The architecture of Hover-Net. (a) Pre-activated residual unit, (b) Dense unit. The variable m represents the number of feature maps in each residual unit. The yellow square within the input highlights the region being considered in the output. When classification labels are unavailable, only the up-sampling branches within the dashed box are used [91].

shape, and staining intensity, as well as the presence of overlapping nuclei and artifacts in histopathology images [101]. Figure 2.4 shows an example of nuclei segmentation.

Traditional methods for nuclei segmentation, such as thresholding [258], watershed algorithms [335], and active contour models [202], often struggle with the complexities of histopathology images. These methods require extensive parameter tuning and are sensitive to image quality and variations in staining, leading to inconsistent results. Therefore, there is a need for more robust and automated approaches to accurately segment nuclei in histopathology images.

Deep learning has emerged as a powerful tool for nuclei segmentation in histopathology images, providing significant improvements over traditional methods [205, 154]. CNNs have demonstrated exceptional performance in image segmentation tasks due to their ability to learn hierarchical features directly from raw pixel data [173]. One of the most widely used CNN-based architectures for nuclei segmentation is U-Net [4, 355]. By utilizing multi-resolution input and bypassing certain pooling layers, Micro-Net [234] effectively retains

both high-level contextual information and detailed localization of U-Net, making it robust to noise and variations in object size and intensity. Xing *et al.* [327] proposes a robust, automatic framework for nucleus segmentation using a deep convolutional neural network to generate probability maps and a novel segmentation algorithm that combines a sparse shape model with a local repulsive deformable model. Some methods leverage boundary information of the nuclei within the network for instance segmentation to address the challenge of overlapping nuclei. Other methods use distance maps of nuclei instead of considering boundaries. For example, Naylor *et al.* [206] used regression of the nuclei distance map to detect high-quality markers for the watershed. Hover-Net [91] improves nuclei segmentation by encoding instance-specific information through the vertical and horizontal distances from each nuclei pixel to its centroid. Figure 2.5 shows the architecture of Hover-Net.

Another effective approach in deep learning-based nuclei segmentation is the use of Mask R-CNN [295, 192], which has been successfully adapted for nuclei segmentation, providing robust instance segmentation capabilities that are particularly useful for distinguishing individual nuclei in dense and complex tissue samples. Recent advancements in deep learning have also explored the use of Generative Adversarial Networks (GANs) [216, 121] and Transformer-based models [104, 114] for nuclei segmentation. GANs have been used to generate synthetic histopathology images to augment training datasets, which helps improve model robustness and generalization [216]. Transformer-based models, known for their ability to capture long-range dependencies in data, are being investigated for their potential to further enhance segmentation performance by effectively modeling spatial relationships between nuclei [114].

Despite the successes of deep learning-based methods, challenges remain in achieving consistent and highly accurate segmentation across diverse histopathology datasets. Variability in staining protocols, imaging equipment, and tissue types can affect model performance, necessitating the use of techniques like domain adaptation and transfer learning to improve generalization. Additionally, the requirement for large annotated datasets for training deep learning models poses a challenge, as manual annotation of nuclei is time-consuming and requires expert knowledge [175].

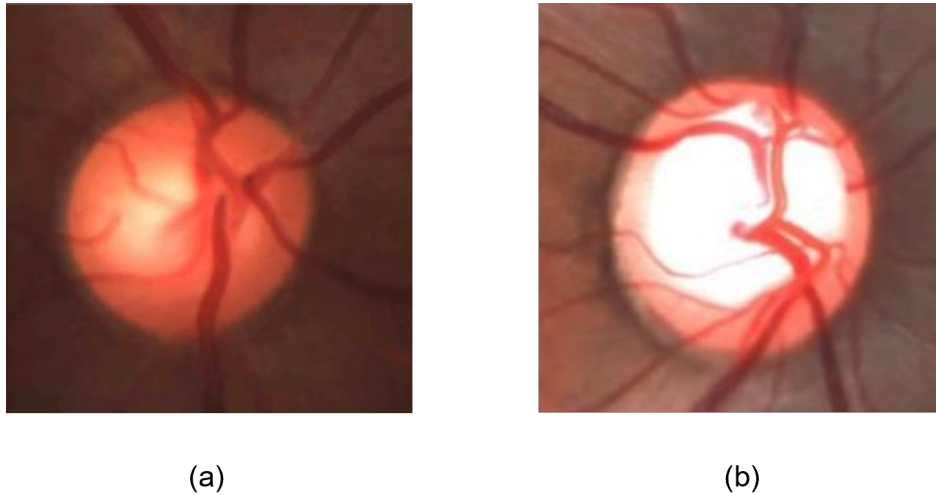


FIGURE 2.6: An example of a zoomed-in (a) normal and (b) glaucoma OD [98].

2.2.2 Optic Disc and Cup Segmentation in Fundus Images

Optic disc and cup segmentation in fundus images are important for the diagnosis and management of glaucoma, a leading cause of irreversible blindness worldwide. The optic disc (OD) is the bright, circular area where the optic nerve fibers exit the retina, and the optic cup (OC) is a depression within the optic disc. The relationship between the optic disc and cup, often measured by the cup-to-disc ratio (CDR), is a critical indicator for detecting glaucoma. A higher CDR suggests an increased risk of glaucoma, as it indicates the loss of retinal nerve fibers and subsequent enlargement of the optic cup [47, 3]. Figure 2.6 shows an example of healthy and glaucoma OD.

Accurate segmentation of the optic disc and cup is essential for quantifying the CDR and other relevant metrics that help clinicians assess the progression of glaucoma and make informed treatment decisions. Manual segmentation by ophthalmologists is time-consuming and subject to inter-observer variability, leading to inconsistencies in diagnosis and management. Automated segmentation methods can provide consistent and objective measurements, reducing the workload by clinicians and improving the accuracy of glaucoma screening and monitoring [137].

In recent years, CNN-based deep learning methods have achieved impressive results in fundus image segmentation. Artem [249] proposed a deep learning method based on the U-Net structure for segmenting the optic disc and the optic cup in fundus images, demonstrating superior performance in terms of experimental results and prediction time. M-Net [76] is a multi-label deep network structure that has been used for OD and OC segmentation. This methodology incorporates a Polar Coordinate System through the M-Net to generate Multi-Label Probability Maps for OC regions. In addition, some methods that classify pixels using deep learning networks [359] and methods based on the U-Net [339] structure are used for optic disc and cup segmentation. Traditional techniques like CNNs process input data through local windows and struggle to establish long-distance connections in global images. In recent years, the Visual Transformer model (ViTs), which can capture long-distance dependencies, has been widely used in image segmentation tasks. Transformers have also achieved good results in OD/OC segmentation [320].

2.2.3 Prostate MRI Segmentation

As prostate MRI is increasingly used for clinical examinations of prostate cancer, prostate MRI image segmentation has become a popular research area. Accurate segmentation of the prostate gland and associated structures, such as the peripheral zone, transition zone, and tumor regions, is essential for several clinical applications. These include the precise targeting of biopsies, radiation therapy planning, surgical guidance, and the assessment of tumor progression or response to treatment [63]. Figure 2.7 shows the example of T2-weighted transverse prostate MRI images.

Magnetic Resonance Imaging (MRI) is widely considered the gold standard for prostate evaluation due to its superior soft-tissue contrast, offering better visualization of the prostate anatomy than techniques like ultrasound or CT. However, manual segmentation of prostate MRI is both time-consuming and labor-intensive, requiring expert knowledge and often subject to variability between observers. The complexity arises from the irregular shape of prostate, its heterogeneous appearance across different regions, and the presence of noise and artifacts in MRI images, which can obscure critical boundaries and make segmentation

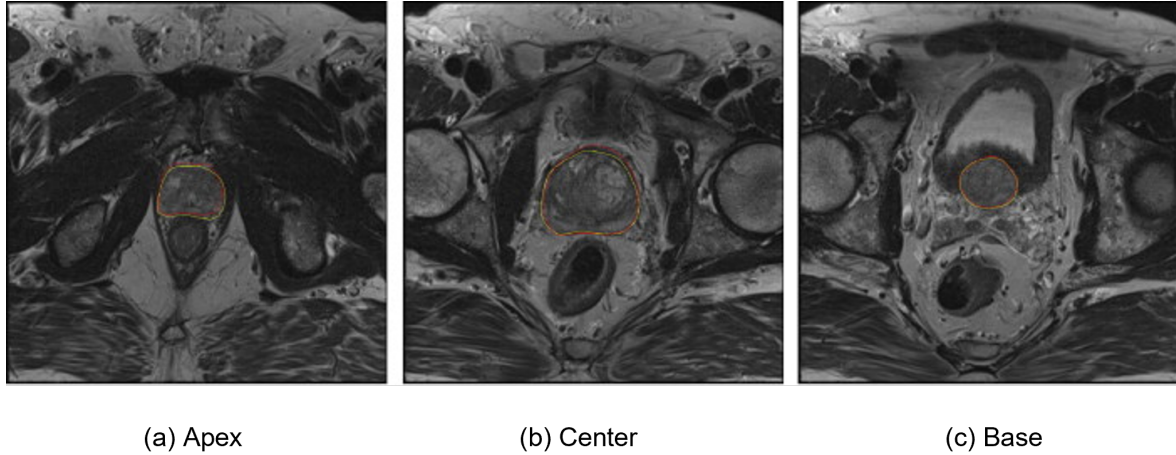


FIGURE 2.7: Example of T2-weighted transverse prostate MRI images showing apex, center, and base slices. The reference standard segmentation is highlighted in yellow, while the second observer's segmentation is shown in red [172].

challenging. These factors not only increase the risk of errors but also limit the efficiency and consistency of manual segmentation [203].

Deep learning approaches for prostate MRI segmentation have offered significant improvements in accuracy and automation. For example, Michelle *et al.* [14] introduces a deep learning approach using three U-Net models to segment the prostate, transition zone (TZ), and peripheral zone (PZ) from MRI images, achieving high segmentation accuracy. ResNet-based models [286], which introduce residual connections to facilitate the training of deep networks, have also been successfully applied to prostate MRI segmentation. The use of residual connections helps overcome the vanishing gradient problem, allowing the network to be trained effectively even with a large number of layers. Another prominent architecture in deep learning-based prostate MRI segmentation is DeepLab [140, 141], which incorporates atrous convolutions and fully connected Conditional Random Fields (CRFs) to improve segmentation accuracy by capturing multi-scale contextual information.

However, challenges remain in deploying deep learning-based segmentation models in clinical practice. One major issue is the scarcity of labeled medical data, as annotating medical images is time-consuming and requires expert knowledge [300]. To address this, techniques such as data augmentation [257, 49, 49, 83], transfer learning [148, 340], especially unsupervised

learning [331, 25, 233, 51] are often employed. Additionally, ensuring that models can generalize effectively across different patient populations and imaging devices is essential but challenging. The variability in medical images, influenced by factors such as patient demographics, imaging protocols, and equipment differences, makes it difficult for models to maintain consistent performance across diverse clinical settings [180].

2.3 Transfer Learning

2.3.1 Introduction to Transfer Learning

Transfer learning is a machine learning approach that leverages knowledge from one domain or task to improve learning performance in a different but related domain or task [337, 106, 316]. It addresses situations where data distributions or features differ between training and target environments. Transfer learning addresses this challenge by allowing models to transfer learned features from a source domain with abundant data to a target domain with limited data, thus enhancing the performance and efficiency of the model on the target task [214, 278].

Following [214], transfer learning can be divided into three main categories based on how tasks and domains relate (see Figure 2.8). The first category is inductive transfer learning, where the source and target tasks differ but are related. This approach often involves scenarios like multi-task learning, where multiple related tasks are learned simultaneously, or self-taught learning, which utilizes labeled data from the source domain alongside unlabeled data from the target domain to improve the learning process.

The second category is transductive transfer learning, where the source and target tasks remain the same, but the domains differ. In this setting, methods such as domain adaptation are used to adjust models so they perform well across different domains, even when the data distributions differ. Another common scenario in this category involves addressing sample selection bias, where data from the source and target domains do not share the same distribution, necessitating careful adjustments to the learning model.

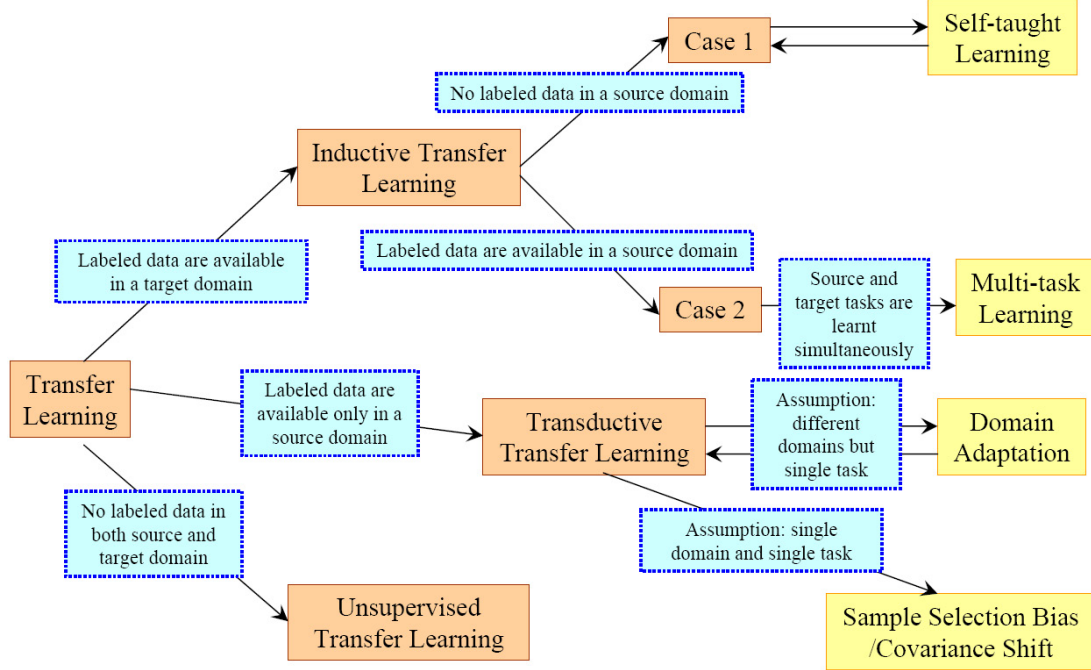


FIGURE 2.8: Transfer learning categories [214].

Lastly, unsupervised transfer learning deals with unsupervised tasks where no labels are available in either the source or target domains. This category often includes techniques like self-taught clustering, which leverages additional unlabeled data from the source domain to enhance clustering tasks or transferred dimensionality reduction, where knowledge from related domains helps in reducing the dimensionality of data. These categorizations guide researchers in selecting the right transfer learning strategy by examining the relationships between tasks and domains.

2.3.2 Applications of Transfer Learning

Transfer learning has been extensively explored in many critical real-world applications, including computer vision [90, 166], bioinformatics [88, 74], and natural language processing [115, 227]. In the field of computer vision, transfer learning has made a significant impact by effectively addressing challenges such as image classification [60, 143], object detection [211, 277], and image segmentation [290, 85, 75]. This approach leverages the

power of pre-trained models that have already been trained on large-scale datasets like ImageNet [55], which contain millions of images across thousands of categories. By transferring the learned features and knowledge from these pre-trained models to new tasks, researchers can significantly reduce the amount of labeled data required for training, as well as the time and computational resources needed. This is particularly beneficial because it allows models to generalize well to new tasks even when only limited labeled data is available.

In computer vision, the use of transfer learning enables models to capture complex patterns and features from large datasets and apply them effectively to related but distinct tasks. For example, features learned from natural images in ImageNet [55], such as edges, textures, and shapes, can be highly useful for other image recognition tasks, even in domains with different types of images, such as medical or industrial images. This ability to generalize across different visual domains makes transfer learning a powerful tool for tackling a wide range of vision-related challenges, from everyday object recognition to specialized applications like autonomous driving and robotics [255].

One of the most significant applications of transfer learning in computer vision is in medical image analysis, where labeled data is often scarce and costly to obtain due to the need for expert annotation [48]. Medical images require precise labeling by radiologists or pathologists, which is both time-consuming and expensive. Transfer learning addresses this issue by enabling the adaptation of models pre-trained on more general datasets to medical imaging tasks with minimal additional training. This approach has been particularly effective in improving the performance of models for tasks such as tumor detection, organ segmentation, and disease classification, where high accuracy is critical [228, 143].

Overall, the strategic use of transfer learning in computer vision demonstrates its profound ability to bridge the gap between research and practical application, enabling the development of robust, high-performing models that can tackle a wide array of visual tasks efficiently and effectively [290].

In this work, we primarily focus on one of the most common challenges in transfer learning, which is domain adaptation. We then delve into the practical applications of domain adaptation, specifically in medical image segmentation, as discussed in the following sections.

2.4 Domain Adaptation

Domain adaptation is essential in transfer learning because it addresses the challenge of domain shift across different domains. Domain adaptation helps align the feature spaces of the source and target domains, reducing discrepancies and ensuring that the transferred knowledge remains effective. This process is crucial for building models that can generalize well across different but related domains [215, 288, 35].

In this section, we first introduce the definition of domain adaptation, followed by an overview of some traditional domain adaptation algorithms. Next, we discuss the two types of domain adaptation relevant to our work: unsupervised domain adaptation (UDA) and zero-shot domain adaptation (ZSDA). Finally, we present some of the recent domain adaptation methods that use deep learning algorithms for medical image segmentation tasks.

2.4.1 Problem Definition

Domain adaptation addresses the challenge of applying a model trained on one domain, known as the source domain $\mathcal{D}_S$, to a different domain, called the target domain $\mathcal{D}_T$. This scenario is common in machine learning when the source domain contains abundant labeled data, but the target domain has limited or no labeled data available. The main difficulty arises from the distribution shift between the source and target domains, denoted as $P_S(X, Y) \neq P_T(X, Y)$. This distribution mismatch, often referred to as domain shift, can significantly degrade the performance of a model trained on the source domain when applied to the target domain.

In domain adaptation, the goal is to minimize this discrepancy by aligning the feature distributions of the source and target domains, enabling the model to generalize effectively to the target domain. Domain adaptation generally includes several sub-fields, such as supervised

domain adaptation, semi-supervised domain adaptation, and unsupervised domain adaptation. Supervised domain adaptation involves using labeled data from both the source and target domains, while semi-supervised domain adaptation uses labeled data from the source domain and a small amount of labeled data from the target domain. Unsupervised domain adaptation (UDA), which is the focus of this work, uses labeled data from the source domain but only unlabeled data from the target domain. UDA aims to bridge the gap between domains by aligning their feature distributions without relying on target domain labels [86, 243].

2.4.2 Traditional Domain Adaptation Algorithms

In domain adaptation, following [70], traditional methods can be broadly classified into two categories: instance-based adaptation [120, 269] and feature-based adaptation [215, 16, 271, 252]. Each of these categories uses different strategies to reduce the discrepancy between the source and target domains.

Instance-based adaptation focuses on re-weighting or selecting instances from the source domain to make them more representative of the target domain. This approach assumes that certain source instances are more relevant and should be given more importance during training.

- **Kernel Mean Matching (KMM):** KMM [120] is an instance-based technique that adjusts the weights of source domain samples so that their mean in the feature space aligns with the mean of the target domain samples. By directly matching the data distributions in the feature space, KMM reduces the mismatch between domains, which is particularly useful when labeled target data is scarce.
- **Kullback-Leibler Importance Estimation Procedure (KLIEP):** KLIEP [269] estimates the importance weights of source instances by directly minimizing the KL-Divergence between the reweighted source distribution and the target distribution. Unlike methods that estimate the full distributions separately, KLIEP focuses solely on estimating the density ratio, making it an efficient and effective method for domain adaptation.

Feature-based adaptation aims to find a common feature space where the distributions of the source and target domains are aligned. This approach typically involves learning new feature representations or transforming existing ones to minimize the discrepancy between domains.

- **Maximum Mean Discrepancy (MMD):** MMD [215] is a widely used metric that measures the difference between the means of two distributions in a reproducing kernel Hilbert space (RKHS). By minimizing MMD, the source and target domain distributions can be aligned in the feature space, reducing the domain shift and improving model generalization. One of the advantages of MMD is its flexibility, as it can capture complex non-linear relationships between the domains through the use of appropriate kernel functions, making it a powerful tool for domain adaptation tasks.
- **Correlation Alignment (CORAL):** CORAL [271] aligns the second-order statistics, specifically the covariance, of source and target feature representations. This method operates by transforming the source domain features so that their covariance closely matches that of the target domain, thereby reducing the domain shift. By minimizing the statistical discrepancy between domains, CORAL enables more effective adaptation in the shared feature space, ultimately enhancing the model's generalization performance across different domains.
- **Wasserstein Distance:** Wasserstein Distance [252] measures the cost of transforming one distribution into another by considering both the distance between samples and their distributions. This approach provides a robust measure of domain discrepancy, particularly in complex scenarios where distributions have different supports or are not absolutely continuous.
- **KL-Divergence:** KL-Divergence [16] is a widely used metric for quantifying the divergence between two probability distributions. In the context of domain adaptation, it serves as a tool to measure the difference between the feature distributions of the source and target domains. By minimizing the KL-Divergence between these distributions, models can better align the features, potentially improving their performance on the target domain. However, KL-Divergence is asymmetric and sensitive to zero probabilities, which can limit its practical applications.

2.4.3 Unsupervised Domain Adaptation

Unsupervised Domain Adaptation (UDA) deals with scenarios where labeled data is available in the source domain, but only unlabeled data is accessible in the target domain. UDA aims to align the source and target domains to mitigate the impact of domain shift, thereby improving the model's performance on the target domain without requiring labeled target data [317].

2.4.3.1 Problem Definition

Unsupervised domain adaptation (UDA) can be formally defined using the following notation: Let $\mathbf{D}_S = \{(\mathbf{x}_i^S, y_i^S)\}_{i=1}^{N_S}$ denote the labeled source domain data, where $\mathbf{x}_i^S$ represents the input samples and y_i^S are the corresponding labels. The target domain is represented by $\mathbf{D}_T = \{\mathbf{x}_j^T\}_{j=1}^{N_T}$, which consists of only unlabeled data, where $\mathbf{x}_j^T$ are the input samples. The goal of UDA is to learn a predictive model using both $\mathbf{D}_S$ and $\mathbf{D}_T$, such that the model performs well on the target domain data $\mathbf{D}_T$. The key challenge in UDA is to mitigate the domain shift between $\mathbf{D}_S$ and $\mathbf{D}_T$ to ensure that the model generalizes effectively from the source to the target domain, despite the absence of target domain labels.

2.4.3.2 Deep-learning based UDA Methods

Deep learning-based UDA methods leverage various strategies to reduce the domain shift, thereby improving the generalization capabilities of models on the target domain [6]. This section explores key categories of deep learning-based UDA methods, including statistic-based approaches, adversarial learning approaches, reconstruction-based approaches, and self-learning approaches.

- **Statistic-based Approaches:** Statistic-based approaches focus on minimizing the statistical discrepancies between source and target domains. These methods typically align distributions by matching the mean and variance of feature representations. A well-known example is the Maximum Mean Discrepancy (MMD) loss, which measures the distance between source and target feature distributions. Longr *et al.* [189] introduces the Deep Adaptation Network (DAN) that utilizes MMD to

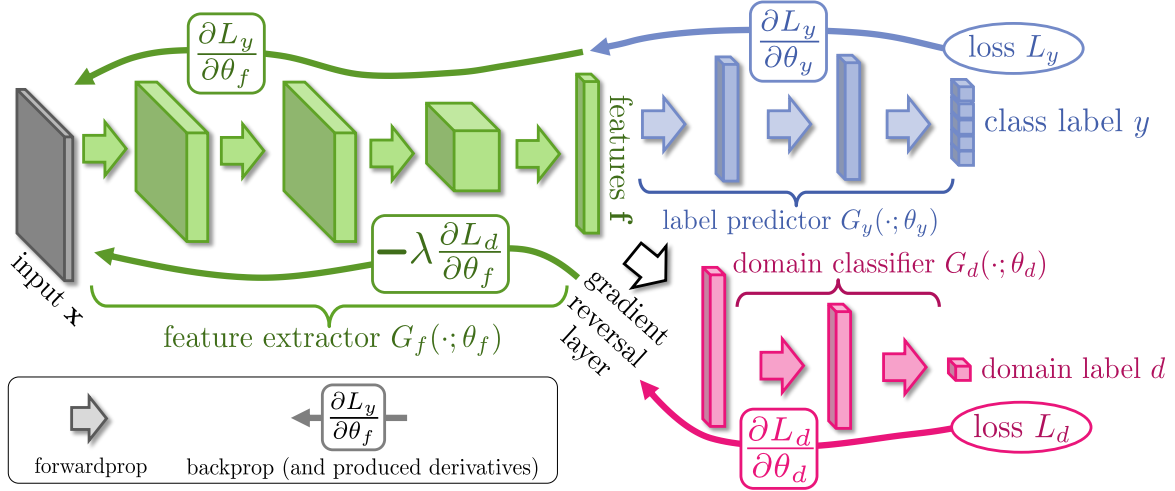


FIGURE 2.9: The architecture of DANN consists of a deep feature extractor (green) and a deep label predictor (blue), forming a standard feed-forward network. Unsupervised domain adaptation is achieved by introducing a domain classifier (red), connected to the feature extractor through a gradient reversal layer. This layer multiplies the gradient by a negative constant during backpropagation, while the rest of the training follows the standard process, minimizing the label prediction loss for source data and the domain classification loss for all samples. The gradient reversal forces the feature distributions across domains to become as similar as possible, making them indistinguishable to the domain classifier and resulting in domain-invariant features [80].

align both marginal and conditional distributions across domains. Some Correlation alignment (CORAL) based methods also have been proposed to compute the distance between different domains [272, 313]. These approaches effectively reduce domain shifts by aligning statistical properties, making them robust for various tasks such as image classification and segmentation.

- **Adversarial Learning Approaches:** Adversarial learning methods leverage the power of generative adversarial networks (GANs) to encourage domain invariance by training models that can not distinguish between source and target domain features. A seminal work in this category is DANN [80] (See Figure 2.9), which integrates adversarial loss to promote indistinguishable domain features. Another remarkable method is CycleGAN [365], which enables image-to-image translation between domains without paired examples, thereby reducing domain discrepancies

in a visually meaningful way. These adversarial techniques are highly effective in scenarios where visual characteristics between domains differ significantly.

- Reconstruction-based Approaches:** Reconstruction-based methods aim to generate or reconstruct target domain data that closely resemble source domain data, thereby learning domain-invariant features. One approach is the use of autoencoders to reconstruct input data, aligning features by minimizing reconstruction errors. Bousmalis *et al.* [19] introduced the PixelDA framework, which uses a generator to reconstruct target domain images in the style of the source domain while preserving semantic content. Another example is Cycle-Consistent Adversarial Domain Adaptation (CyCADA) [109], which combines adversarial and reconstruction losses to enforce consistency in feature representations. Recently, Some work predicated on the diffusion generation model has addressed the UDA problem by utilizing image generation to reduce the domain gap [336, 328, 218]. These methods leverage reconstruction to align domains, ensuring that the extracted features retain relevant semantic information while being domain-invariant.
- Self-learning Approaches:** Self-learning approaches utilize pseudo-labeling strategies to iteratively refine model predictions on target domain data. These methods start with an initial model trained on the source domain, which is then used to generate pseudo-labels for the target domain. Shu *et al.* [259] introduced the Self-ensembling approach, where a student model learns from a teacher model's output on unlabeled target data, progressively refining its predictions through consistency regularization. Zou *et al.* [368] proposed the Confidence Regularized Self-training (CRST) method, which utilizes confidence scores to filter reliable pseudo-labels for self-training. Self-learning methods are advantageous in scenarios where explicit domain alignment is challenging, allowing the model to adapt by iteratively improving its predictions on the target domain.

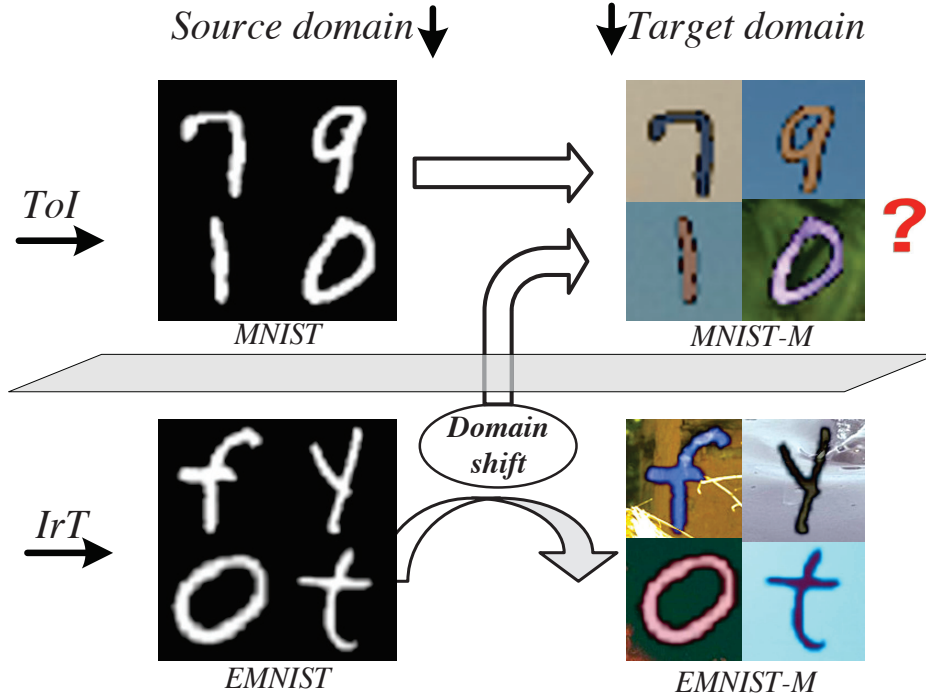


FIGURE 2.10: An intuitive example of widely used ZSDA methods involves the use of ToI for digital image analysis and IrT for letter image analysis. The source domain consists of grayscale images, while the target domain includes color images. To train a model for the unseen MNIST-M dataset (the target domain data in ToI), traditional ZSDA methods introduce IrT to learn domain transformations and then transfer this knowledge to ToI [302].

2.4.4 Zero-shot Domain Adaptation

Zero-shot domain adaptation (ZSDA) tackles the more challenging scenario where the target domain is completely unseen during the training phase, with no access to any target data, whether labeled or unlabeled [147, 307]. This approach is particularly important in situations where obtaining data from the target domain is either impractical or impossible, such as in the case of rare diseases where patient data is scarce, or in specific imaging modalities that are not readily accessible due to cost, availability, or ethical constraints. ZSDA enables the development of robust models that can generalize effectively to new domains without prior exposure, making it a critical technique for expanding the applicability of machine learning models in diverse and data-limited medical contexts [127, 301].

ZSDA relies on leveraging prior knowledge or domain generalization techniques to create models that are inherently robust to domain shifts. Recently, various ZSDA methods have been introduced for natural images, aiming to adapt models to unseen target domains that are inaccessible during training [219, 127].

Existing ZSDA methods mainly focus on introducing irrelevant tasks (IrT) to learn the ability of knowledge transfer between inter-domains within these tasks and then apply this ability to the task of interest (ToI) (see Figure 2.10). These methods use adversarial networks to capture inter-domain transformations, aligning target domains in ToI and achieving strong performance [219, 302, 186].

One common strategy in zero-shot domain adaptation is synthetic data augmentation, where the source data is enhanced with artificially generated variations that simulate potential target domain conditions [322, 17, 303]. For instance, style transfer methods can be employed to modify the appearance of images to resemble different domains while maintaining their original content, thereby creating synthetic, target-like data for training [253, 82, 280]. Notably, ZDDA [219] proposes a unique approach to domain adaptation without the need for target domain data by utilizing task-irrelevant dual-domain pairs to learn generalized representations of the source domain. Similarly, ZSDA [302] leverages adversarial learning to transfer inter-domain transformations learned in Intermediate Tasks (IrT) to the Task of Interest (ToI).

Large-scale pre-trained models, such as Vision Transformers (ViTs)[62, 187] and language models[226], have also shown significant promise for ZSDA due to their extensive generalization capabilities. These models, trained on vast and diverse datasets spanning multiple domains, can act as robust baselines for ZSDA, utilizing their rich and diverse feature representations to tackle novel domains without explicit adaptation [66, 333]. Among these, PODA [66] introduces an innovative method that uses natural language prompts to guide zero-shot domain adaptation. Meanwhile, ULDA [333] suggests converting natural language descriptions into guiding information, facilitating domain adaptation without the need for target domain images.

Despite these advancements, existing ZSDA methods still encounter significant challenges when applied to medical image tasks, primarily due to the unique and complex nature of medical images. The specificity, variability, and intricate details inherent in medical data make it difficult for general ZSDA approaches to perform effectively. Although some studies [17, 2] have explored the application of ZSDA to medical image segmentation, these methods often struggle in scenarios where no target images are available, including those from IrT tasks. This underscores the need for continued innovation in ZSDA techniques specifically tailored to meet the demanding requirements of the medical domain.

2.4.5 Domain Adaptation in Medical Image Segmentation

In medical image segmentation, it is common to face data scarcity scenarios where the availability of large, annotated datasets is limited [330, 95]. This scarcity is primarily due to the high costs and specialized expertise required for the manual labeling of medical images. Furthermore, ethical and privacy concerns often restrict data sharing between institutions, further intensifying the challenge of data scarcity [217, 28]. These constraints underscore the urgent need for domain adaptation techniques that can utilize existing labeled data from a source domain to enhance performance on a target domain with little or no labeled data. By effectively transferring knowledge from one domain to another, domain adaptation not only reduces the reliance on extensive manual annotation but also improves the generalizability of models, making them more applicable for deployment in real-world clinical settings. Among various approaches, domain adaptation stands out as one of the most crucial methods for addressing the challenges of medical image segmentation under data-efficient scenarios [349, 67].

Domain adaptation in medical image segmentation addresses the critical challenge of applying machine learning models trained on one dataset (source domain) to another dataset (target domain) with different characteristics [220, 325]. This challenge is particularly relevant in medical imaging, where significant variability arises from diverse imaging modalities (e.g., MRI [366, 43], CT [130, 128], ultrasound [32]), variations in equipment settings, patient demographics, and differing clinical protocols [224]. These variations can result in substantial

differences in image appearance and quality, making it difficult for models trained on one dataset to perform effectively on another. Domain adaptation aims to bridge these gaps, enabling models to generalize across diverse domains and thereby enhancing their robustness and clinical applicability [136, 69].

The rise of deep learning has introduced more advanced methods for domain adaptation, leveraging the capacity of neural networks to learn complex and high-dimensional feature representations. In medical image segmentation, domain adaptation techniques have been applied across various tasks, such as tumor segmentation [85, 87], lesion segmentation [132, 287], and fundus image segmentation [310, 72], among others [267, 94]. These applications demonstrate the versatility of domain adaptation in addressing the variability and challenges inherent in different medical imaging tasks, making it a valuable approach for enhancing model performance across diverse clinical scenarios.

By integrating domain adaptation strategies, medical image segmentation models can achieve higher performance in data-efficient scenarios, significantly enhancing their utility in clinical practice where annotated data is often limited. This makes domain adaptation not only a practical solution but also a pivotal advancement in the quest for data-efficient medical imaging technologies [126].

In the following section, we will focus on the review of three types of medical image UDA segmentation mentioned in our paper, including histopathology images, fundus images, and prostate MRI images. This overview will highlight the specific challenges and recent advancements in unsupervised domain adaptation for each imaging type, providing a comprehensive understanding of their applications and the strategies employed to address domain shifts in medical image segmentation.

2.4.5.1 UDA in Nuclei Segmentation

Several methods have recently been proposed for unsupervised domain adaptive nuclei instance segmentation in histopathology images [175, 334]. Inspired by CyCADA [109], PDAM [175] first synthesizes target-like images and uses a nuclei inpainting mechanism to

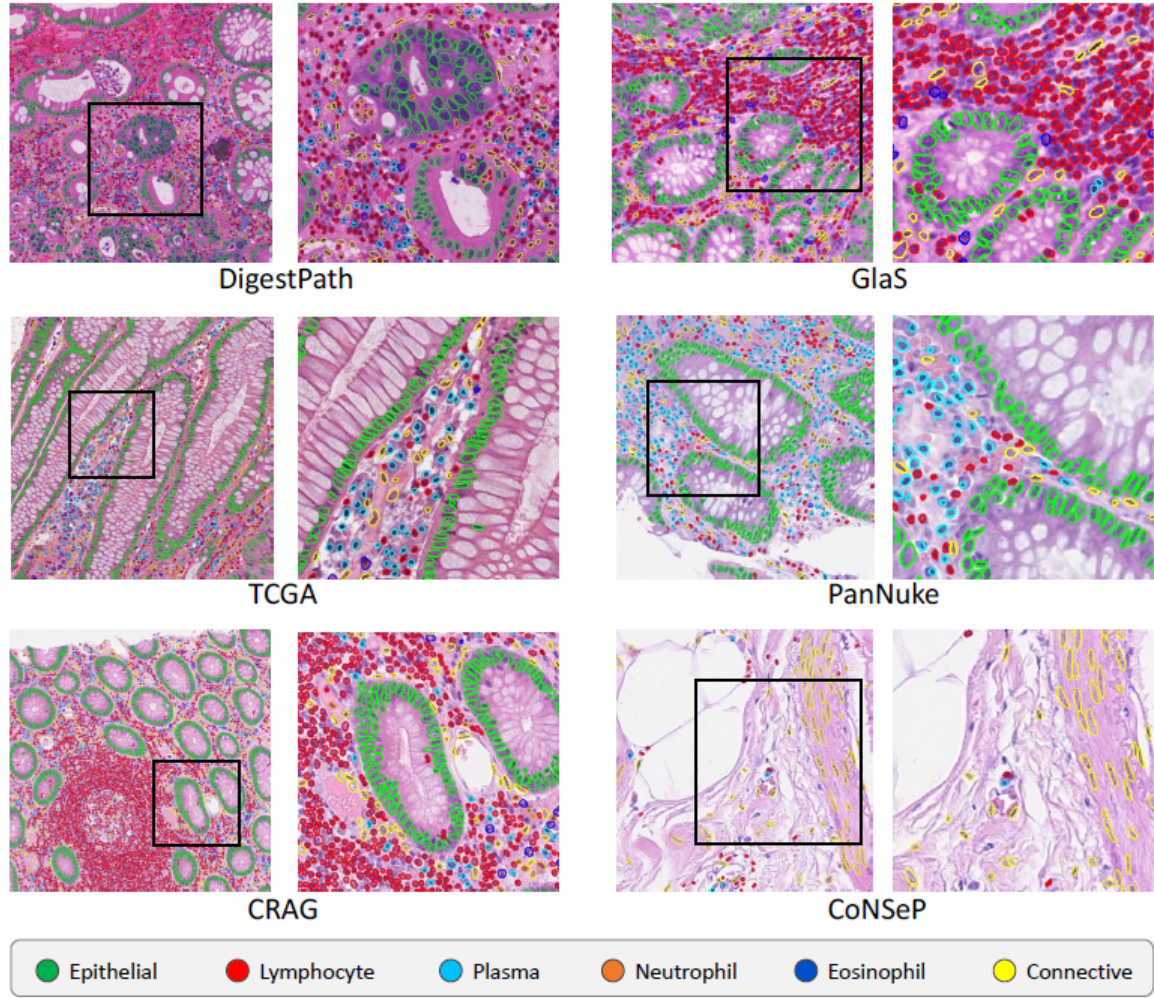


FIGURE 2.11: Example image regions sampled from each of the six data sources show significant morphological differences. [92].

remove the incorrectly synthesized nuclei. The adversarial training strategies are then used separately at the image-, semantic- and instance-level with a task re-weighting mechanism. However, this work can only address the instance segmentation for nuclei within the same class. In Yang *et al.*'s work [334], firstly, local features are aligned by an adversarial domain discriminator, and then a pseudo-labelling self-training approach is used to further induce the adaptation. However, the performance improvement of this work relies on weak labels and fails to get good training results without target domain labels. In addition, although this work is validated on cross-domain nuclei segmentation and classification, their adaptation strategies are class-agnostic. Fan *et al.* [67] proposes a novel two-stage disentanglement framework

for heterogeneity-aware domain adaptive nuclei segmentation, addressing the intra-domain discrepancies in histopathology images by introducing open compound domain adaptation (OCDA). Figure 2.11 shows an example of image regions in different data sources. From the figure, we can see that the nuclei have overlapping morphological features, and there are significant differences between different domains and nuclei types.

In the real clinical, the nuclei objects in the histopathology images belong to various classes, and the characteristics of the nuclei in different classes are also distinct [123]. In addition, the number of objects within each category is also imbalanced in the histopathology datasets [92]. To this end, previous UDA methods are limited for cross-domain nuclei segmentation and classification due to the lack of analysis of the nuclei classes when transferring knowledge.

2.4.5.2 UDA in Fundus Image Segmentation

In the task of fundus image segmentation, adversarial learning is frequently employed to enhance the model's generalization ability of the model. BEAL [310] uses adversarial training to strengthen the boundary prediction of the target domain image and further proposes a discriminator for entropy-driven adversarial learning to minimize the uncertainty in the segmentation. The pOSAL approach [311] first extracts the Region of Interest (ROI) by roughly segmenting the optic disc through the U-Net network and then uses the patch-based Output Space Adversarial Learning framework (pOSAL) to robustly segment OD and OC. Feng *et al.* [158] propose an unsupervised domain adaptation framework based on category-level regularization, which regularizes category distributions from three perspectives. Target-like query images are generated using Generative Adversarial Networks (GANs) and boundary information, and content and style feature alignment (CSFA) is used to ensure feature consistency. RDR-Net [44] uses the variational auto-encoder (VAE) to reconstruct images and utilizes dynamic convolution specific to entropy-driven adversarial learning for low-level feature refinement and prediction map alignment, respectively. TriLA [45] introduces the Triple-Level Alignment (TriLA) model for unsupervised domain adaptation in joint segmentation of optic disc and optic cup on fundus images, addressing domain shift by aligning at the input, feature, and output levels.

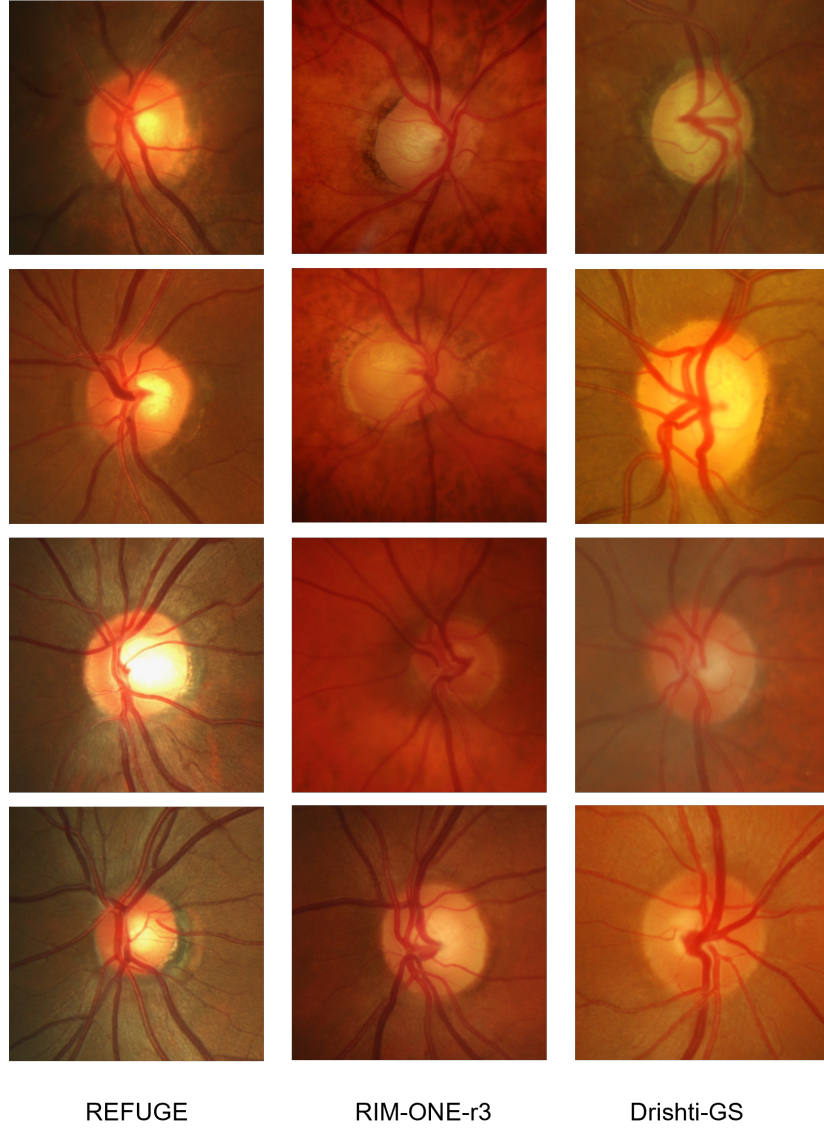


FIGURE 2.12: Fundus images in different datasets (REFUGE, RIM-ONE-r3 and Drishti-GS dataset).

However, these methods have their limitations as they fail to consider the issue of data imbalance, as well as the inherent limitations of adversarial learning strategies.

2.4.5.3 UDA in Prostate MRI Segmentation

In recent years, several deep learning-based methods have been proposed to address the domain gap between different prostate MRI datasets. HarmoFL [129] proposes a framework that addresses both local and global drifts in federated learning for heterogeneous medical

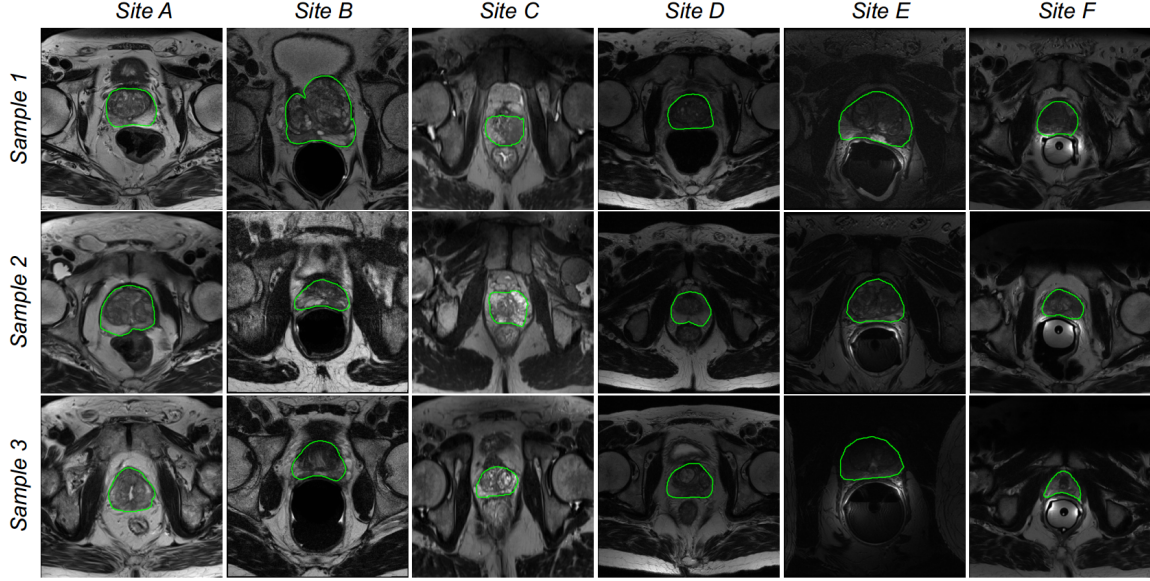


FIGURE 2.13: Prostate MRI images in different datasets [181].

images by harmonizing features and applying weight perturbations. Liu *et al.* [179] presents a shape-aware meta-learning framework designed to improve the generalization of prostate MRI segmentation models to unseen domains with different imaging protocols and scanners. By incorporating shape compactness and boundary smoothness constraints into a meta-learning process, the proposed method enhances the model's ability to maintain complete and accurate segmentation shapes under domain shifts. Cheng *et al.* [213] introduce a causality-inspired data augmentation approach for single-source domain generalization in medical image segmentation, addressing domain shifts caused by variations in image acquisition processes. The approach uses randomly-weighted shallow networks for global intensity transformations and interventional pseudo-correlation augmentation to reduce spurious correlations, thereby enhancing the robustness of segmentation models to unseen domains.

2.5 Summary

This chapter introduces a review of the background and related literature relevant to this thesis. First, we introduce the concept of deep learning, along with deep learning-based medical image segmentation methods. Next, we focus on the main topic of this thesis, domain

adaptation, which is an essential subfield of transfer learning. Following that, we provide a detailed explanation of the definition of domain adaptation, traditional approaches, and various UDA methods across different domain adaptation scenarios. Lastly, we review UDA methods applied to various medical segmentation tasks. More specifically, we discuss the application of UDA methods to three medical segmentation tasks related to the datasets used in this thesis, including nuclei segmentation, OD/OC segmentation in fundus images, and prostate MRI segmentation.

Unsupervised Domain Adaptation via Category-aware Feature Alignment and Pseudo-label

Nuclei segmentation and classification in histopathology images is a prerequisite for accurate tumor analysis by computer. However, due to variations in staining, imaging conditions, and the inherent complexity of histopathological samples, it is challenging to achieve robust performance across different datasets. Unsupervised domain adaptation (UDA) methods have been widely employed to enhance the adaptability of models in general computer vision tasks. Nevertheless, unlike natural images, histopathology images present significant semantic gaps among nuclei of different categories, making it difficult to generalize UDA models effectively for nuclei instance segmentation or classification.

In this chapter, we propose a novel deep neural network, termed Category-Aware Feature Alignment and Pseudo-Labeling Network (CAPL-Net), specifically designed for UDA in nuclei instance segmentation and classification. Our approach addresses the unique challenges posed by the substantial domain shifts and category variations in histopathology images. Firstly, we introduce a category-level feature alignment module that utilizes dynamic learnable trade-off weights to balance the alignment of shared and category-specific features across domains. This module helps in reducing the domain shift by aligning features at the category-specific level. Secondly, we enhance the performance of the model on target domain data through self-supervised training using pseudo labels, which are generated based on nuclei-level prototype features. This pseudo-labelling strategy enables the model to learn from the unlabeled target data, further improving its generalization capability. Extensive experiments on cross-domain nuclei instance segmentation and classification tasks demonstrate the effectiveness of our proposed CAPL-Net. The results show that our approach significantly

outperforms state-of-the-art UDA methods, achieving a notable improvement in performance. This highlights the potential of CAPL-Net as a robust solution for UDA in the complex and diverse field of histopathology image analysis.

3.1 Introduction

3.1.1 Motivation

Automatic nuclei instance segmentation and classification are crucial for digital pathology with various application scenarios, such as tumor classification and cancer grading [196]. However, manual labelling is limited by high subjective, low reproducibility, and resource-intensive [157, 206]. Although deep learning-based methods can achieve impressive nuclei recognition performance, they require sufficient labelled data for training [206, 363, 37, 91, 174]. By directly adopting the off-the-shelf deep learning models to a new histopathology dataset with a distinct distribution, the models suffer from performance drop due to domain bias [175, 334]. Recently, unsupervised domain adaption methods have been proposed to tackle this issue [29, 364, 176], and enable the learning models to transfer the knowledge from one labelled source domain to the other unlabelled target domain [134, 189, 350].

Several methods have recently been proposed for unsupervised domain adaptive nuclei instance segmentation in histopathology images [175, 334]. Inspired by CyCADA [109], Liu *et al.* [175] first synthesize target-like images and use a nuclei inpainting mechanism to remove the incorrectly synthesized nuclei. The adversarial training strategies are then used separately at the image-, semantic- and instance-level with a task re-weighting mechanism. However, this work can only address the instance segmentation for nuclei within the same class. In Yang *et al.*'s work [334], firstly, local features are aligned by an adversarial domain discriminator, and then a pseudo-labelling self-training approach is used to further induce the adaptation. However, the performance improvement of this work relies on weak labels and fails to get good training results without target domain labels. In addition, although this work is validated on cross-domain nuclei segmentation and classification, their adaptation strategies

are class-agnostic. In real clinical settings, the nuclei objects in the histopathology images belong to various categories, and the characteristics of the nuclei in different classes are also distinct [123]. In addition, the number of objects within each category is often imbalanced in the histopathology datasets [92].

The motivation for this work is that previous UDA methods are limited for cross-domain nuclei segmentation and classification due to the lack of analysis of the nuclei categories when transferring the knowledge. We believe that distinguishing representations for different nuclei categories can help capture more comprehensive domain features. Additionally, pseudo-labelling approaches have been shown to be useful for domain adaptation, but previous methods [334] have not been specifically designed for the unique characteristics of histopathology images. Simply applying pseudo-labelling for training may fail to capture key feature representations from the target domain effectively.

3.1.2 Contributions

In this chapter, we investigate the problem of cross-domain nuclei instance segmentation and classification through a novel category-level adaptation framework specifically for histopathology images. This framework is designed to address the significant semantic gaps that exist among nuclei of different categories across various datasets, which is challenging for UDA methods in nuclei segmentation and classification tasks. To overcome these challenges, we introduce a category-aware feature alignment module that aims to enhance knowledge transfer for intra-class features across domains while mitigating the risk of negative transfer between inter-class features. This module dynamically adjusts the alignment of features, allowing the model to effectively distinguish between different nuclei categories, thus improving the overall segmentation and classification performance. Additionally, we introduce a self-supervised learning stage that leverages nuclei-level feature prototypes. This stage focuses on improving the model’s performance on unlabelled target data by generating pseudo labels based on the learned prototypes. This self-supervised approach allows the model to further refine its understanding of the target domain, thus enhancing its adaptability and generalization capabilities. Extensive experiments validate the efficacy of our proposed framework, demonstrating its

superior performance over state-of-the-art UDA methods on cross-domain nuclei instance segmentation and classification tasks. Significantly, our approach not only outperforms existing UDA methods but also achieves performance that is comparable to or even surpasses the fully supervised upper bound under various evaluation metrics, highlighting the potential of our framework in practical applications.

Our contribution is summarized as follows:

- We introduce a category-aware feature alignment module that enhances knowledge transfer by aligning intra-class features across domains while avoiding negative transfer between inter-class features.
- We develop a self-supervised learning stage that utilizes nuclei-level feature prototypes to generate pseudo labels for the unlabelled target data.
- Our proposed framework has been extensively evaluated, demonstrating superior performance compared to state-of-the-art UDA methods in nuclei instance segmentation and classification tasks.

3.2 Methods

3.2.1 Overview

The architecture of our proposed model is illustrated in Figure 3.1. Our UDA framework is optimized through a two-stage process designed to address the challenges of domain shift in nuclei instance segmentation and classification. In the first stage, we introduce category-level feature alignment modules, which are specifically designed to reduce the domain gap at the feature level by aligning features of the same class across the source and target domains. This helps the model learn consistent and robust feature representations, making it more adaptable to variations in the data from different domains. In the second stage, we further enhance the model’s adaptability on the unlabelled target images through a pseudo-labelling process. This process is guided by nuclei-level prototypes, which serve as reference points for generating pseudo labels. By using these prototypes, the model can refine its feature representations

and improve its performance on the target domain through self-supervised learning. This approach enables the model to leverage the unlabelled data effectively, enhancing its ability to generalize to unseen images from the target domain.

3.2.2 Segmentation Architecture

Our proposed model is based on Hover-Net [91], a state-of-the-art method for fully-supervised nuclei instance segmentation and classification. Hover-Net proposes a deep learning framework specifically designed to tackle the challenges of nuclei instance segmentation and classification in histology images. It uses horizontal and vertical distance maps, which encode the pixel-wise distances from nuclei centers of mass. Specifically, for each pixel within a nucleus, it computes its horizontal and vertical distances from the center of mass of the nucleus, thereby generating the corresponding distance maps. These distance maps enable the model to effectively separate closely clustered nuclei, which is a common problem in histopathology images where nuclei often overlap or are densely packed. Hover-Net significantly improves segmentation performance by using distance maps for instance separation, especially in cases with overlapping nuclei.

The framework is constructed by three branches sharing the same encoder for feature extraction and using three decoders for different tasks: 1) Nuclear pixel (NP) branch to perform binary classification of a pixel (nuclei or background); 2) Hover (HV) branch to predict the horizontal and vertical distances of nuclei pixels to their centroid; 3) Nuclear classification (NC) branch to classify the nuclei types of pixels. The framework of Hover-Net can be seen in Figure 2.5.

The supervised Hover-Net loss function of our model is defined as $\mathcal{L}_F$:

$$\mathcal{L}_F = \mathcal{L}_{np} + \mathcal{L}_{hover} + \mathcal{L}_{nc}, \quad (3.1)$$

where $\mathcal{L}_{np}$, $\mathcal{L}_{hover}$ and $\mathcal{L}_{nc}$ represent the loss with respect to the output at the NP, HV and NC branch, respectively. For the NP and the NC branches, the loss is calculated by adding

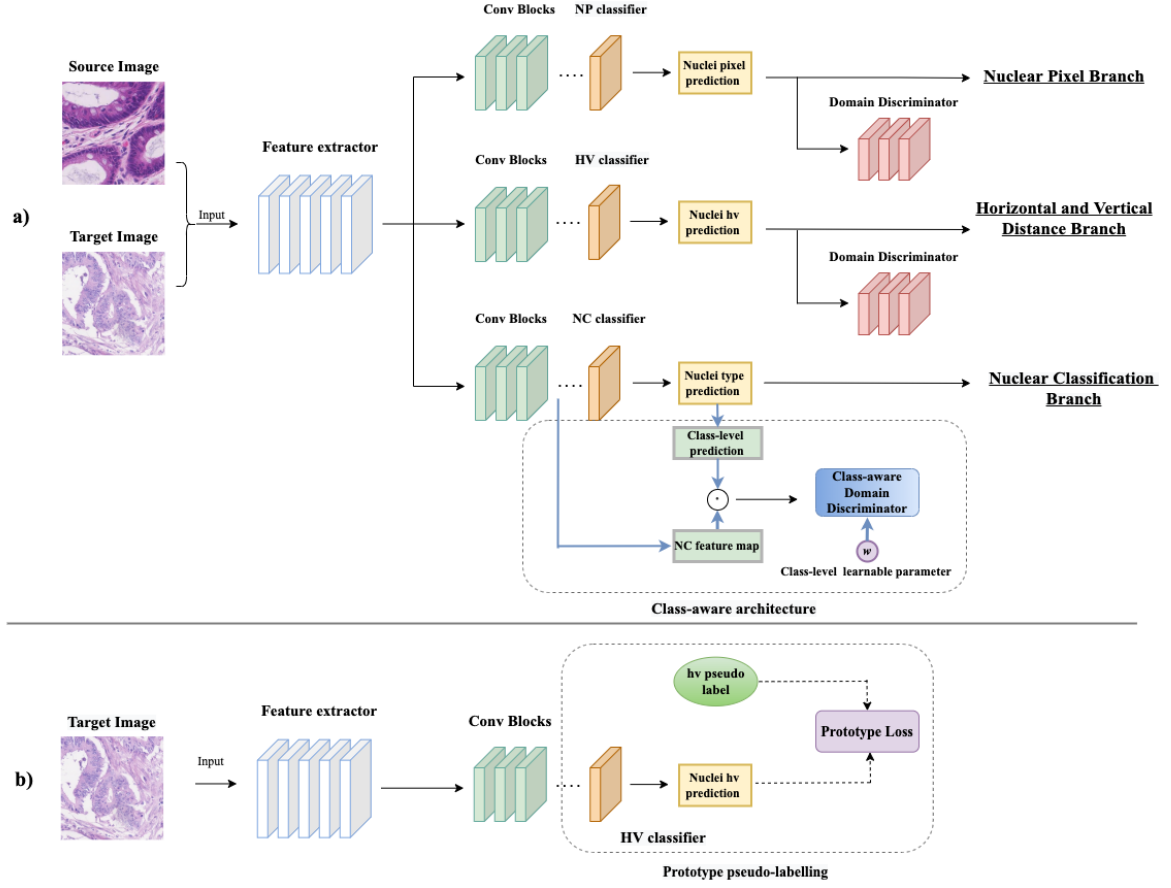


FIGURE 3.1: Overview of the proposed category-aware prototype pseudo-labelling network. Our framework is optimized through a two-stage process to address domain shift issues in nuclei instance segmentation and classification. In the first stage, we incorporate class-level feature alignment modules, which are designed to reduce the domain gap by aligning features of similar classes across the source and target domains, thereby enhancing the model’s ability to learn consistent representations. In the second stage, we introduce a pseudo-labelling process that is enhanced by nuclei-level prototypes. This approach generates pseudo labels based on these prototypes, enabling self-supervised learning on the unlabelled target images. This two-stage optimization strategy allows the model to effectively leverage unlabelled data, improving its adaptability and performance on the target domain.

cross-entropy loss and dice loss:

$$\mathcal{L}_{np} = \mathcal{L}_{np}^{CE} + \mathcal{L}_{np}^{dice}, \quad (3.2)$$

$$\mathcal{L}_{nc} = \mathcal{L}_{nc}^{CE} + \mathcal{L}_{nc}^{dice}. \quad (3.3)$$

The cross-entropy and dice losses are defined as:

$$\mathcal{L}^{CE} = -\frac{1}{N} \sum_{i=1} [p_i \log(q_i)], \quad (3.4)$$

$$\mathcal{L}^{dice} = 1 - \frac{2(q_i * p_i)}{q_i + p_i}. \quad (3.5)$$

where N is the total number of pixels, p_i represents the predicted probability for pixel i , and q_i is the corresponding ground truth label.

For the HV branch, we denote $\mathcal{L}^{mqe}$ as the mean squared error loss, and define the $\mathcal{L}^{hv}$ loss as the mean squared error between the horizontal and vertical gradients and the corresponding gradients of the ground truth:

$$\mathcal{L}_{hover} = \mathcal{L}^{mqe} + \mathcal{L}^{hv}, \quad (3.6)$$

$$\mathcal{L}^{mqe} = \frac{1}{N} \sum_{i=1} (p_i - q_i)^2, \quad (3.7)$$

$$\mathcal{L}^{hv} = \frac{1}{M} \sum_{i=1} (p_{i,hor} - q_{i,hor})^2 - \frac{1}{M} \sum_{i=1} (p_{i,ver} - q_{i,ver})^2. \quad (3.8)$$

where M is the total number of nuclear pixels. $p_{i,hor}$ and $p_{i,ver}$ represent the predicted horizontal and vertical gradients of the distance map at pixel i . $q_{i,hor}$ and $q_{i,ver}$ denotes the corresponding ground truth gradients.

3.2.3 Category-aware Feature Alignment

In the histopathology datasets with multi-class nuclei, there is a very large gap in the number of nuclear in each category. This may lead to many images in the dataset having only a few nuclear categories, and some nuclear types are absent from an image. In addition, the features

of the nuclei from different classes also vary. Under this situation, the typical class-agnostic feature alignment strategies may lead to the negative transfer of the features in different categories. To tackle this issue, for the NC branch on nuclei classification, we propose to conduct feature alignment for the cross-domain features within each class separately. Particularly, an adversarial domain discriminator D_c is introduced to adapt the features under the class c . Compared with directly employing a class-agnostic domain discriminator for the NC branch, the class-aware domain discriminators can avoid the misalignment across different classes and further encourage the knowledge transfer in classification learning.

The detailed paradigm is shown in Figure 3.1 (a). First, we denote the features of the NC branch in the typical Hover-Net as F_{nc} , and the N-class prediction as P_{nc} . Note that P_{nc} contains N channels, and the predictions in each channel represent the classification results for each specific class. For each category c , we formulate the P_{nc} into a binary class prediction map P_{nc}^c , where the pixel value is set to 1 if it belongs to the nuclei in this category, otherwise set to 0. To incorporate the class-aware information into the feature alignment, we propose to generate the prototype features F_{pt}^c for the class c by dot-multiplying F_{nc} with the binary P_{nc}^c . In addition, if any P_{nc}^c is empty, we will not perform subsequent training on the nuclei in this class. In other words, we only deal with the nuclear types that exist in the input images. In each domain, the prototype features F_{pt}^c for the class c pass through the corresponding adversarial discriminators D_c for class-aware adaption at the feature level.

To avoid manually finetuning the trade-off weights for the adversarial loss in N categories of the NC branch, we let the overall framework automatically learn these weight parameters during training. The learnable weighted discriminator loss is formulated as follows:

$$\mathcal{L}_{NC}^{ca} = \omega_c^L \sum_{c=1} \mathcal{L}_c^{adv}, \quad (3.9)$$

where the $\mathcal{L}_c^{adv}$ denotes the adversarial training loss of D_c for class c , and ω_c^L is its corresponding learnable loss weight. Concretely, we define the adversarial training loss as:

$$\mathcal{L}^{adv} = -\frac{1}{N} \sum_{i=1} [y_i \log(p_i) + (1 - y_i) \log(1 - p_i)]. \quad (3.10)$$

The overall domain discriminator loss function of our model is defined as:

$$\mathcal{L}_{dis} = \mathcal{L}_{NC}^{ca} + \mathcal{L}_{NP}^{adv} + \mathcal{L}_{HV}^{adv}, \quad (3.11)$$

where $\mathcal{L}_{NP}^{adv}$ and $\mathcal{L}_{HV}^{adv}$ are the feature adaptation loss functions in the NP and HV branches, respectively. Particularly, we utilize adversarial domain discriminators on the cross-domain output features of the NP and HV branches for adaptation.

With the above loss terms, the overall loss function of the first stage approach can be written as:

$$\mathcal{L}_{s1} = \mathcal{L}_F + \mathcal{L}_{dis}. \quad (3.12)$$

3.2.4 Nuclei-level Prototype Pseudo-labelling

Although the category-aware feature alignment modules can reduce the domain gaps, the lack of supervised optimization on the target images still limits the model's segmentation and classification performance. Therefore, in the second stage, we use the output of the first stage model as pseudo labels for self-supervised learning to further improve the model performance on the target images. The detailed training process can be referred to Figure 3.1 (b).

Different from the traditional self-training process with the pseudo labels, we only train the target domain of the HV branch during the second stage pseudo-labelling process. In the extensive experiments, we noticed that the performance of the classification predictions on the target images from the first stage model is limited. In addition, the binary segmentation predictions lack object-wise information. Moreover, the model trained with all pseudo labels may not perform well due to the low quality of some pseudo-labelling classes.

To avoid the disturbance from the less accurate and representative pseudo labels, we no longer consider classification and binary segmentation branches in the second stage but particularly focus on the predictions from the HV branch, where the feature maps describe the distance from each pixel to the nuclei's centre point.

Therefore, the features of each nuclear object can be regarded as a prototype at the object level, which contains morphological information such as the shape and size of the specific nuclear. By self-supervised learning with the pseudo labels on the predictions of the HV branch, the bias between nuclei objects can be further reduced. The overall loss function for the second stage is as follows:

$$\mathcal{L}_p = \frac{1}{N^p} \sum_{i=1}^{N^p} |x_i^p - \hat{y}_i^p|^2, \quad (3.13)$$

where x_i^p is the predicted features from the HV branch for each nuclear object p in the second stage and $\hat{y}_i^p$ is the object features generated by the pseudo labels.

3.3 Experiments and Results

3.3.1 Dataset Description

We conduct experiments on three datasets from the Lizard database [92], a large-scale colon tissue histopathology dataset collected at 20x objective magnification, specifically designed for nuclei instance segmentation and classification. It contains nearly half a million labeled nuclei from histology images of colon tissue, making it the largest dataset of its kind in computational pathology. The dataset was generated using a multi-stage annotation pipeline, which combines automated and semi-automated methods with pathologist-in-the-loop refinement steps to ensure high-quality and accurate annotations. The Lizard dataset includes full segmentation and classification labels for various types of nuclei, such as epithelial, connective tissue, lymphocytes, plasma, neutrophils, and eosinophils. By focusing on nuclei from colon tissue across a wide range of normal, inflammatory, dysplastic, and cancerous conditions, the dataset enhances the potential for models trained on it to generalize well to unseen samples. The diversity of conditions represented in the dataset increases its utility for profiling the tumor micro-environment and developing downstream computational pathology models.

TABLE 3.1: Details of the number of nuclei under each class in the datasets we used, DigestPath, CRAG, and GlaS.

Class	DigestPath	CRAG	GlaS
Epithelial	70,789	99,124	31,986
Lymphocyte	49,932	27,634	9,763
Plasma	11,352	9,363	2,349
Neutrophil	2,262	1,673	90
Eosinophil	1,349	1,255	286
Connective	32,826	49,994	10,890
Total	168,510	189,043	55,364

In this study, we use DigestPath (Dpath), CRAG, and GlaS datasets from the Lizard collection. For all three datasets, we split the images into training, testing, and validation sets, with 2/3 of the images allocated for training and the remaining 1/3 for testing and validation. Specifically, we use Dpath as the source domain, utilizing 46 images for training and 22 images for validation. The CRAG and GlaS datasets serve as target domains, with 42 images used for training and 21 for testing in the CRAG dataset, and 40 images for training and 20 for testing in the GlaS dataset. Training images are randomly cropped into 256×256 patches, and data augmentation techniques, including flipping, rotating, Gaussian blur, and median blur, are applied to enhance model robustness. The detailed number of nuclei in each class across the datasets is shown in Table 3.1.

3.3.1.1 DigestPath (Dpath) Dataset

The DigestPath dataset is designed for the segmentation and classification of cells in histopathological images, focusing primarily on digestive system tissues. This dataset includes a variety of biopsy samples collected from four different Chinese hospitals, providing a diverse set of images with varying staining techniques and tissue conditions. DigestPath includes images of colonic tissues, where different types of cells, such as epithelial, inflammatory, and connective tissue cells, are segmented and classified. The dataset aims to support the development of robust algorithms for digestive tract cancer diagnosis, offering a wide range of images that reflect real-world clinical scenarios. This dataset is particularly valuable for

training models to handle the complex and heterogeneous nature of histopathological images from digestive tissues, making it a critical resource for research in computational pathology.

3.3.1.2 CRAG Dataset

The Colorectal Adenocarcinoma Gland (CRAG) dataset is specifically focused on the segmentation of glandular structures in colorectal cancer histology images. CRAG provides high-resolution images extracted from whole-slide images (WSIs) at University Hospitals Coventry and Warwickshire (UHCW). The dataset includes annotations for glands in colorectal adenocarcinoma tissues, which are critical for understanding tumor morphology and aiding in the diagnosis and grading of cancer. CRAG's detailed gland segmentation allows researchers to develop models that can accurately delineate gland boundaries, even in complex scenarios where glands are fragmented or surrounded by dense tumor cells. This dataset is a benchmark for evaluating the performance of gland segmentation algorithms and supports the development of automated tools for colorectal cancer diagnosis.

3.3.1.3 GlaS Dataset

The Gland Segmentation in Colon Histology Images (GlaS) dataset is another significant dataset used for the segmentation of glandular structures in histological images of colorectal tissues. This dataset was created to facilitate the development of algorithms that can accurately segment glands, which are crucial for assessing colorectal cancer progression. GlaS contains images sourced from whole-slide images (WSIs) at UHCW, with precise annotations provided by expert pathologists. The dataset includes both benign and malignant tissue samples, offering a comprehensive representation of gland morphology in various stages of colorectal cancer. GlaS is widely used in computational pathology to benchmark segmentation methods, particularly those aimed at distinguishing between normal and abnormal glandular structures in cancerous tissues.

3.3.2 Evaluation Metrics

For evaluation, we use the same metrics as those employed by Hover-Net [91] to comprehensively assess the performance of our model on nuclei instance segmentation classification tasks. The metrics used for nuclei instance segmentation include Dice coefficient, Aggregated Jaccard Index (AJI), Detection Quality (DQ), Segmentation Quality (SQ), and Panoptic Quality (PQ). These metrics provide a detailed understanding of how well the model segments and identifies individual nuclei instances within the images. Additionally, to evaluate the performance of nuclei detection and classification, we utilize F1-scores at classification levels.

3.3.3 Implementation Details

We utilize the Hover-Net framework with ResNet50 [102] pre-trained weights on ImageNet as our base architecture. In the first stage of our adaptation process, the model is trained in two steps following the Hover-Net [91]. In the first step, only the decoders are trained 50 epochs. In the second step, all layers are trained for another 50 epochs. We use Adam optimization in both steps, with an initial learning rate of $1e-4$, which is then reduced to $1e-5$ after 25 epochs. In the self-training pseudo-labelling part, the Adam optimizer with a learning rate of $1e-4$ was used to train 50 epochs with a batch size of 20. Experiments were conducted on one NVIDIA GeForce 3090 GPU and implemented using PyTorch.

3.3.4 Comparison Experiments

We conduct a series of comparative experiments to compare the performance. There are four comparative experiments in our work, including Source Only [91], PDAM [175], Yang *et al.* [334] and Fully-supervised. The details are as follows:

- Source Only [91]: original Hover-Net without adaptation.
- PDAM [175]: a UDA nuclei instance segmentation method with pixel-level and feature-level adaptation. Since it can only be used for binary classification for objects, we only compare the results of nuclei instance segmentation with this method.

TABLE 3.2: Experimental results on UDA nuclei instance segmentation on Dpath $\rightarrow$ CRAG task.

Methods	Dpath $\rightarrow$ CRAG				
	<i>Dice</i> $\uparrow$	<i>AJI</i> $\uparrow$	<i>DQ</i> $\uparrow$	<i>SQ</i> $\uparrow$	<i>PQ</i> $\uparrow$
Source Only [91]	0.378	0.201	0.336	0.768	0.259
PDAM [175]	0.596	0.323	0.467	0.676	0.316
Yang <i>et al.</i> [334]	0.766	0.494	0.648	0.765	0.496
Baseline	0.750	0.455	0.604	0.759	0.458
Baseline+CA	0.772	0.502	0.661	0.773	0.510
Baseline+CA+PL	0.781	0.517	0.675	0.772	0.522
Proposed	0.785	0.519	0.681	0.769	0.524
Full-supervised	0.778	0.526	0.683	0.783	0.535

TABLE 3.3: Experimental results on UDA nuclei instance segmentation on Dpath $\rightarrow$ GlaS task.

Methods	Dpath $\rightarrow$ GlaS				
	<i>Dice</i> $\uparrow$	<i>AJI</i> $\uparrow$	<i>DQ</i> $\uparrow$	<i>SQ</i> $\uparrow$	<i>PQ</i> $\uparrow$
Source Only	0.494	0.264	0.359	0.732	0.262
PDAM	0.571	0.298	0.400	0.667	0.267
Yang <i>et al.</i>	0.639	0.294	0.377	0.735	0.275
Baseline	0.645	0.288	0.374	0.739	0.275
Proposed	0.651	0.296	0.385	0.734	0.281
Full-supervised	0.721	0.423	0.539	0.765	0.411

- Yang *et al.* [334]: a UDA framework is proposed based on global-level feature alignment for nuclei classification and nuclear instance segmentation. In addition, a weakly-supervised DA method is also proposed using weak labels for the target images such as nuclei centroid. We only compare with the UDA method in this work for a fair comparison.
- Fully-supervised: fully supervised training on the labelled target images with original Hover-Net, as the upper bound.

TABLE 3.4: Experimental results on UDA nuclei classification on Dpath $\rightarrow$ CRAG task. F_c^1 , F_c^2 , F_c^3 , F_c^4 , F_c^5 and F_c^6 denote the F1 classification score for the Eosinophil, Epithelial, Lymphocyte, Plasma, Neutrophil and Connective tissue, respectively. F_{avg} denotes the average of all the F1-score for the classification under each category.

Methods	Dpath $\rightarrow$ CRAG							
	$Det \uparrow$	$F_c^1 \uparrow$	$F_c^2 \uparrow$	$F_c^3 \uparrow$	$F_c^4 \uparrow$	$F_c^5 \uparrow$	$F_c^6 \uparrow$	$F_{avg} \uparrow$
Source Only [91]	0.490	0.022	0.389	0.324	0.195	0.038	0.161	0.188
Yang <i>et al.</i> [334]	0.736	0.037	0.670	0.330	0.371	0.017	0.428	0.309
Baseline	0.702	0.044	0.687	0.292	0.381	0.155	0.475	0.339
Baseline+CA	0.731	0.128	0.697	0.351	0.400	0.084	0.498	0.360
Baseline+CA+PL	0.772	0.110	0.725	0.327	0.383	0.352	0.558	0.409
Proposed	0.775	0.111	0.714	0.389	0.402	0.377	0.535	0.421
Full-supervised	0.748	0.167	0.724	0.388	0.419	0.428	0.545	0.445

3.3.4.1 Evaluation on Nucleus Instance Segmentation

A comparison of the segmentation performance between our model and state-of-the-art methods is reported in Table 3.2 and Table 3.3. From Table 3.2, it can be observed that our instance segmentation effect is better than the two existing models. Compared with the source-only method, PDAM [175] has a performance improvement by aligning features at the panoptic level. The UDA method of Yang *et al.* [334] achieves good segmentation performance. Our method achieves the highest scores among all methods, with Dice and AJI being 2.3% and 1.6% higher than the previous methods, respectively. In comparison with the full-supervised model, the performance of our proposed UDA architecture is close to it. In particular, our Dice is higher than the results of the full-supervised model. From Table 3.3, it can be observed that our method achieves good results in the task of adapting from Dpath to domain GlaS. This demonstrates the effectiveness of our approach in handling the complexities involved in this task.

3.3.4.2 Evaluation on Nucleus Classification

Table 3.4 and Table 3.5 report the performance comparison of our method with other works on nuclei classification. The values appear as 0 due to their extremely small magnitudes.

TABLE 3.5: Experimental results on UDA nuclei instance segmentation by transferring from *Dpath* to the *Glas* dataset.

Methods	Dpath $\rightarrow$ Glas							
	$Det \uparrow$	$F_c^1 \uparrow$	$F_c^2 \uparrow$	$F_c^3 \uparrow$	$F_c^4 \uparrow$	$F_c^5 \uparrow$	$F_c^6 \uparrow$	$F_{avg} \uparrow$
Source Only	0.512	0.022	0.467	0.206	0.068	0.215	0.186	0.194
Yang <i>et al.</i>	0.565	0.000	0.621	0.141	0.152	0.018	0.240	0.195
Baseline	0.566	0.000	0.617	0.145	0.175	0.063	0.249	0.208
Proposed	0.572	0.000	0.647	0.131	0.192	0.168	0.323	0.244
Full-supervised	0.675	0.000	0.775	0.207	0.191	0.000	0.440	0.269

Although Yang *et al.* [334] can achieve cross-domain nuclei classification and segmentation based on Hover-Net, their feature alignment modules are class-agnostic, which incurs the misalignment issues for the cross-domain features from different categories and further limits the performance. On the other hand, our proposed class-aware UDA framework has outperformed [334] by a large margin. Our model achieves a 4% improvement in the F1 detection score in the detection metric. In addition, our method improves the classification performance for all categories. Moreover, our method achieves significant improvements in classes with sparse samples like Neutrophil and Eosinophil, around 10% and 34%, respectively.

3.3.5 Ablation Studies

To test the validity of each component of our proposed model, we conducted ablation experiments. Firstly, based on our architecture, we kept only the feature-level domain discriminator on the three branches as our baseline method. Secondly, we kept only the class-aware structure, removing the pseudo labels based on the nuclei-level prototype (Baseline+CA). In addition, we also compared our nuclei-prototype pseudo-labelling process with the traditional one, which directly trains the model on the target images with all predictions in the first stage as the pseudo labels (Baseline+CA+PL).

Table 3.2 and Table 3.4 show the instance segmentation and classification performance of all the ablation methods. From these tables, it is clear that the class-aware structure significantly enhances classification performance, especially in categories with sparse samples

TABLE 3.6: Computational complexity analysis, which indicates our proposed modules can introduce performance gain but bring negligible auxiliary cost.

Methods	Dpath $\rightarrow$ GlaS	
	Time Cost	Number of Parameters
Source Only	35 s/iter	3.7M
Baseline	45 s/iter	3.7M
Proposed	45 s/iter	3.7M

(*e.g.*, Neutrophil and Eosinophil), showing more than a 10% improvement in the classification F1-score. This improvement highlights the effectiveness of class-aware adaptation in facilitating knowledge transfer across multi-class datasets, particularly in scenarios where certain classes are underrepresented. The substantial gain in classification accuracy demonstrates how focusing on class-specific features can bridge gaps caused by sample imbalances between different categories. Furthermore, the class-aware structure also provides a notable boost of approximately 2% in nuclei segmentation performance.

Self-supervised training plays a pivotal role in further improving both instance segmentation and classification. The introduction of prototype loss yields more than a 4% improvement in F1-score for nuclei detection, while also delivering better overall segmentation results. This improvement suggests that by using class prototypes, the model can more accurately capture key features that are essential for distinguishing between different nuclei types. Additionally, our proposed nuclei prototype pseudo-labeling process surpasses traditional pseudo-labeling approaches.

Computational complexity analysis shown in Table 3.6 offers an in-depth evaluation of the efficiency of our proposed method. This table presents key computational metrics, including time cost, memory usage, and the number of parameters, facilitating a thorough comparison with other existing methods. These metrics provide valuable insights into the resource requirements and operational feasibility of each approach. The results clearly demonstrate that our method strikes an optimal balance between performance and computational demand. Despite its effectiveness in achieving high-quality segmentation results, our model operates within reasonable computational limits, with time efficiency and parameter count comparable

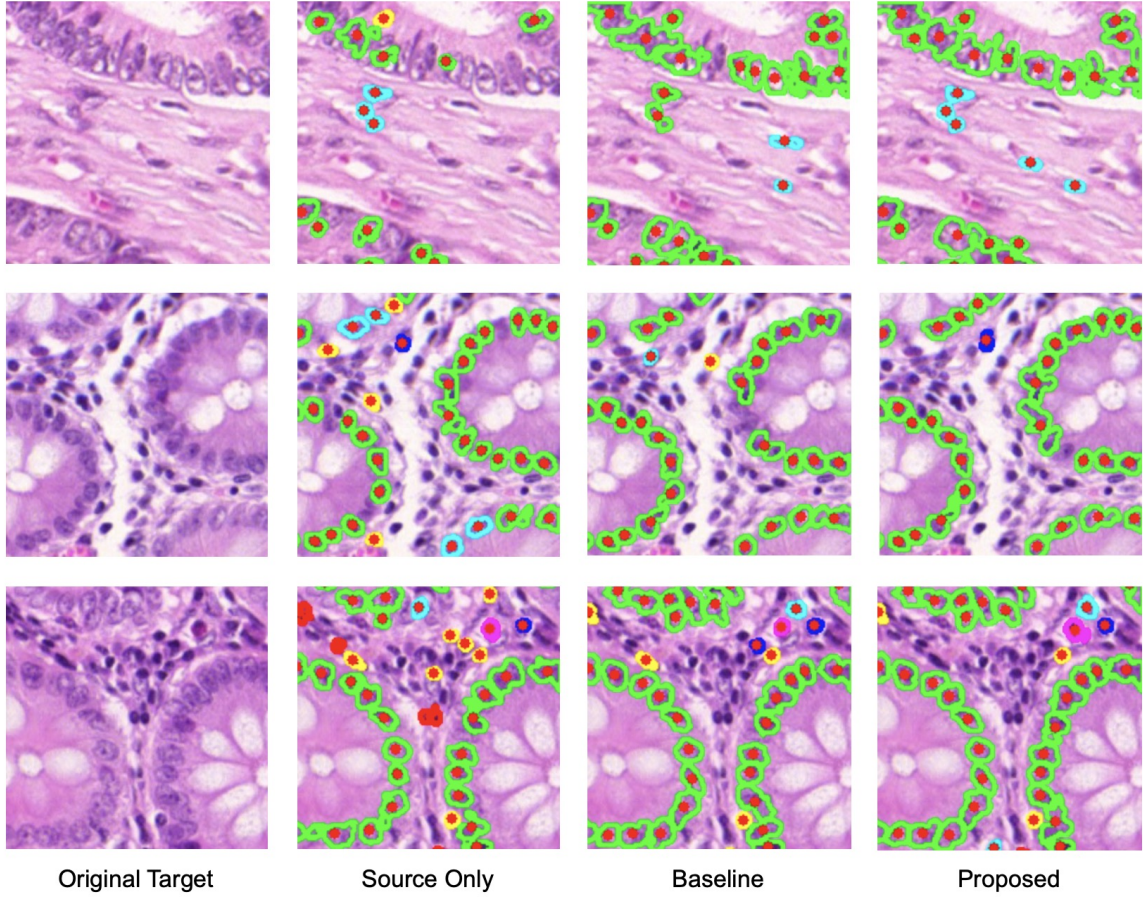


FIGURE 3.2: Visualization predictions for the UDA experiment under the $Dpath \rightarrow GlaS$ setting. Red: Eosinophil, Green: Epithelial, Yellow: Lymphocyte, Blue: Plasma, Magenta: Neutrophil, Cyan: Connective tissue. (Best to be viewed in color and zoomed-in)

to, or better than, alternative techniques. This efficient design ensures that our approach is well-suited for practical applications, particularly in resource-constrained environments such as clinical settings where computational power and time are often limited.

Moreover, the analysis highlights the scalability of our method, as it maintains consistent performance across various datasets without significantly increasing computational overhead. This makes our approach not only a powerful solution from a technical standpoint but also a viable candidate for real-world deployment, where both accuracy and efficiency are critical.

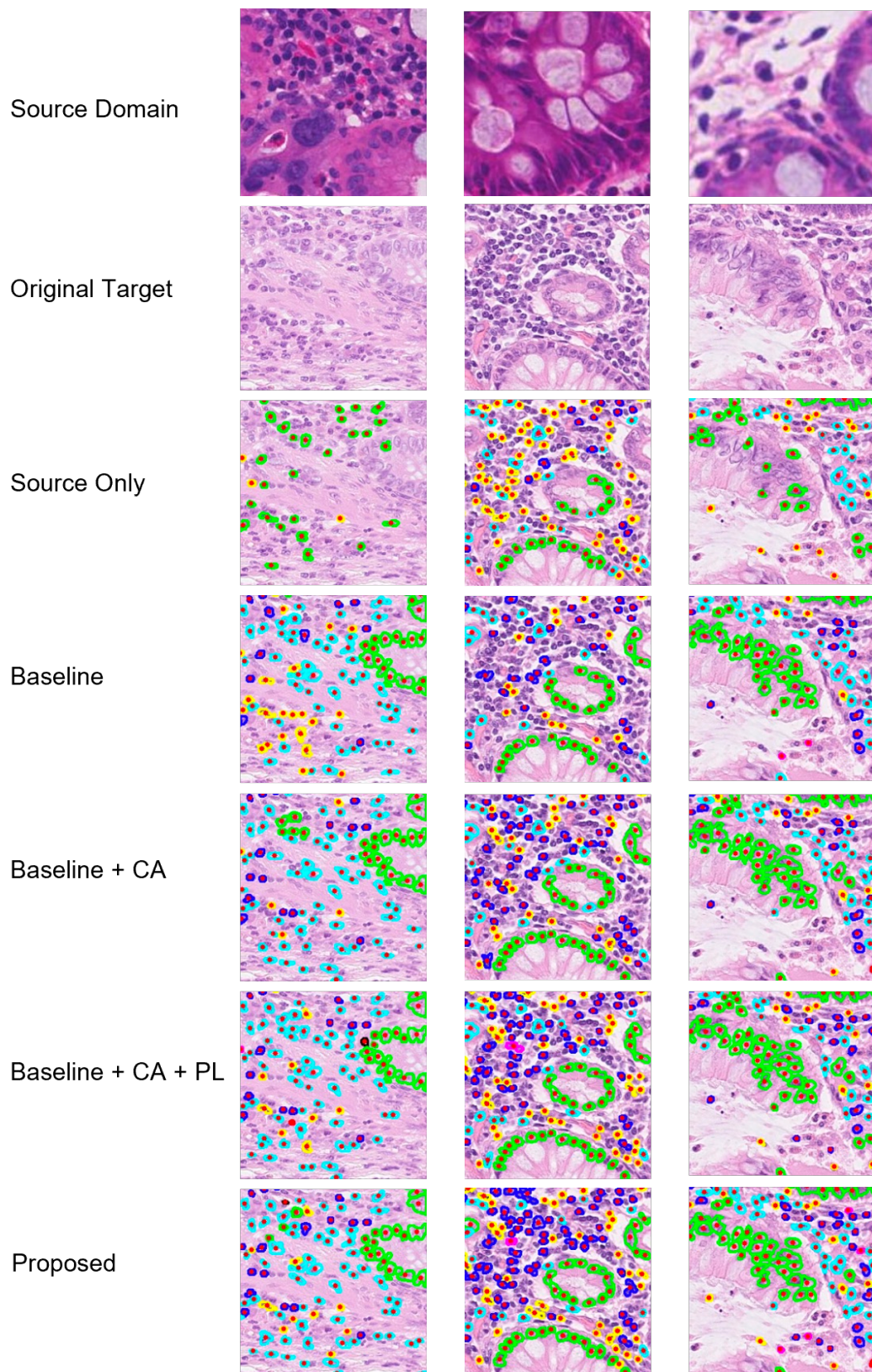


FIGURE 3.3: Visualization predictions for the ablation experiments under $Dpath \rightarrow CRAG$ setting.

3.3.6 Visualization Results

Visualization examples of the ablation studies are shown in Figure 3.2 and Figure 3.3. These visual comparisons highlight the superior performance of our proposed method over the alternatives. In particular, our approach demonstrates more precise segmentation, capturing finer details and reducing errors where other methods struggle. The clear distinction in the quality of the output showcases the robustness of our model in handling complex cases, where the alternatives either fail to capture critical structures or produce less reliable results. These findings underscore the effectiveness of our approach, as it consistently yields more accurate and dependable outcomes, reinforcing the insights gained from the quantitative evaluations.

3.4 Discussion

In this work, we propose an unsupervised domain adaptation framework, named CAPL-Net, for cross-domain nuclei instance segmentation and classification. To address the challenges of class-level adaptation, a category-aware feature alignment module is adopted to facilitate effective knowledge transfer for intra-class features while minimizing negative transfer between classes. Additionally, CAPL-Net incorporates a self-supervised learning stage using nuclei-level feature prototypes to enhance performance on unlabelled target data. The performance of CAPL-Net was evaluated on multiple histopathology datasets from diverse domains.

In addition, we conducted further exploration of CAPL-Net to enhance its capabilities in addressing the challenges of cross-domain nuclei segmentation and classification.

3.4.1 Graph Search Mechanism

To explore whether certain class-specific features have a more significant impact on domain adaptation, we introduce a graph search mechanism that builds on CAPL-Net. This mechanism is designed to assess the importance of individual class-aware features by evaluating how each feature contributes to the adaptation process. By analyzing the role of specific features, we aim

TABLE 3.7: Graph-based method experimental results on UDA nuclei instance segmentation on Dpath $\rightarrow$ CRAG task.

Methods	Dpath $\rightarrow$ CRAG				
	<i>Dice</i> $\uparrow$	<i>AJI</i> $\uparrow$	<i>DQ</i> $\uparrow$	<i>SQ</i> $\uparrow$	<i>PQ</i> $\uparrow$
Source Only [91]	0.378	0.201	0.336	0.768	0.259
PDAM [175]	0.596	0.323	0.467	0.676	0.316
Yang <i>et al.</i> [334]	0.766	0.494	0.648	0.765	0.496
Graph-based	0.765	0.514	0.672	0.766	0.513
Proposed	0.785	0.519	0.681	0.769	0.524
Full-supervised	0.778	0.526	0.683	0.783	0.535

to gain deeper insights into which elements are most crucial for successful domain adaptation, particularly in addressing cross-domain variations and enhancing model performance in diverse medical imaging tasks.

3.4.1.1 Graph-based Search Mechanism Methods

In this approach, we construct feature graphs where pixels that share similar class-related features are represented as nodes, with edges formed based on the degree of feature similarity between these nodes. The goal is to identify which specific features have the most substantial impact on domain adaptation tasks. We hypothesize that emphasizing these key class-aware features can significantly enhance the adaptation process by directing the model’s focus toward the most influential aspects of the data. Instead of treating all features equally, this approach aims to selectively prioritize features that are critical to class distinction, thus improving the model’s ability to generalize to unseen domains. By highlighting the most relevant features, we expect the model to perform better in handling domain shifts, particularly in challenging medical imaging scenarios where variations across domains can hinder performance.

3.4.1.2 Experiment Results for Graph Search Mechanism

For our experiments, this graph-based approach is applied to nuclei segmentation and classification tasks, aiming to determine whether class-specific features can boost performance. The

TABLE 3.8: Graph-based method experimental results on UDA nuclei instance segmentation on Dpath $\rightarrow$ GlaS task.

Methods	Dpath $\rightarrow$ GlaS				
	<i>Dice</i> $\uparrow$	<i>AJI</i> $\uparrow$	<i>DQ</i> $\uparrow$	<i>SQ</i> $\uparrow$	<i>PQ</i> $\uparrow$
Source Only	0.494	0.264	0.359	0.732	0.262
PDAM	0.571	0.298	0.400	0.667	0.267
Yang <i>et al.</i>	0.639	0.294	0.377	0.735	0.275
Graph-based	0.647	0.292	0.370	0.736	0.279
Proposed	0.651	0.296	0.385	0.734	0.281
Full-supervised	0.721	0.423	0.539	0.765	0.411

results, detailed in Table 3.7, Table 3.8, Table 3.9 and Table 3.10, represent this extended class-aware UDA method, referred to as the “Graph-based” method in the tables. The outcomes are compared with those achieved by our proposed global feature-based method. As illustrated by the results, the majority of the graph-based results show less favorable performance compared to our method in both nuclei classification and segmentation tasks.

This observation suggests that focusing on isolated class-specific features does not yield substantial improvements over global feature representations in the context of UDA for nuclei tasks. These findings imply that no individual class-aware feature is inherently more critical for domain adaptation than the overall feature set. While the graph-based mechanism provides a novel analytical framework for examining feature importance, it does not lead to enhanced performance. This highlights the importance of maintaining a holistic approach, where the alignment of global features across all classes is crucial for robust domain adaptation.

In complex scenarios like nuclei segmentation and classification, where a broad range of features must be captured to handle variability across domains, global feature alignment emerges as a more effective strategy. The results reinforce the idea that domain adaptation benefits more from a comprehensive feature representation that spans the entire feature space, rather than focusing narrowly on specific class-aware features.

TABLE 3.9: Graph-based method experimental results on UDA nuclei classification on Dpath $\rightarrow$ CRAG task.

Methods	Dpath $\rightarrow$ CRAG							
	$Det \uparrow$	$F_c^1 \uparrow$	$F_c^2 \uparrow$	$F_c^3 \uparrow$	$F_c^4 \uparrow$	$F_c^5 \uparrow$	$F_c^6 \uparrow$	$F_{avg} \uparrow$
Source Only [91]	0.490	0.022	0.389	0.324	0.195	0.038	0.161	0.188
Yang <i>et al.</i> [334]	0.736	0.037	0.670	0.330	0.371	0.017	0.428	0.309
Graph-based	0.759	0.101	0.708	0.375	0.392	0.379	0.515	0.420
Proposed	0.775	0.111	0.714	0.389	0.402	0.377	0.535	0.421
Full-supervised	0.748	0.167	0.724	0.388	0.419	0.428	0.545	0.445

TABLE 3.10: Graph-based method experimental results on nuclei classification by transferring from Dpath $\rightarrow$ Glas dataset.

Methods	Dpath $\rightarrow$ Glas							
	$Det \uparrow$	$F_c^1 \uparrow$	$F_c^2 \uparrow$	$F_c^3 \uparrow$	$F_c^4 \uparrow$	$F_c^5 \uparrow$	$F_c^6 \uparrow$	$F_{avg} \uparrow$
Source Only	0.512	0.022	0.467	0.206	0.068	0.215	0.186	0.194
Yang <i>et al.</i>	0.565	0.000	0.621	0.141	0.152	0.018	0.240	0.195
Graph-based	0.562	0.000	0.645	0.121	0.188	0.156	0.309	0.217
Proposed	0.572	0.000	0.647	0.131	0.192	0.168	0.323	0.244
Full-supervised	0.675	0.000	0.775	0.207	0.191	0.000	0.440	0.269

3.4.2 Limitations and Future Work

Despite the outstanding performance of our proposed framework compared to previous state-of-the-art methods, there are still some limitations. First, the pseudo labels generated for the nuclei prototypes still contain noise, which limits the overall effectiveness of the self-supervised learning process. This noise can lead to inaccurate feature representations, ultimately affecting the model’s ability to generalize well to unseen domains. Second, the current framework may exhibit performance drop in certain challenging cross-domain nuclei instance segmentation and classification tasks. For example, adaptation from Glas to Dpath, which involves significant domain shifts, presents a particular challenge where our method’s effectiveness diminishes. Furthermore, we have identified that our approach struggles in situations where data imbalance is a factor. When the source dataset is significantly larger than the target dataset, the model may become biased toward the source samples due to the

strong inductive bias inherent in deep learning models. Conversely, if the source dataset is much smaller than the target domain, the model may suffer from insufficient training, making it challenging to achieve a reliable representation of the target domain.

In future research, we aim to address these challenges by refining the pseudo-labelling process to reduce noise, improving the model’s robustness in diverse cross-domain scenarios, and developing strategies to handle imbalanced data more effectively. These efforts will help enhance the overall reliability and adaptability of our framework across various domain adaptation tasks.

3.5 Summary

In this paper, we proposed a category-aware prototype pseudo-labelling architecture for unsupervised domain adaptive nuclear instance segmentation and classification. In our two-stage framework, category-aware feature alignment with learnable trade-off loss weights is proposed to tackle the class-imbalance issue and avoid misalignment during the cross-domain study. In addition, we proposed a nuclei-level prototype loss to correct the deviation in the second stage pseudo-labelling training, which further improves the segmentation and classification performance on the target images by introducing auxiliary self-supervision. Comprehensive results on various cross-domain nuclei instance segmentation and classification tasks demonstrate the prominent performance of our approach. Given the appealing performance of our method on the UDA nuclei instance segmentation tasks, we suggest that future directions can focus on the cross-domain multi-class object recognition tasks for other medical and general computer vision scenarios.

Dynamic Diffusion Flow for Unsupervised Image Segmentation

Impacted by the domain shift issue across varying fundus image datasets collected from different medical centres and devices, the performance of a well-trained optic segmentation network is usually affected when applied to another dataset with different distributions. To handle this issue, the unsupervised domain adaptation (UDA) strategy is widely used to improve the generalization ability of deep learning networks by using unlabeled data. In the last chapter (Chapter 3), we explored the limitations of adversarial learning in UDA, specifically its dependency on the balance between the source and target datasets for effective feature alignment.

However, existing UDA approaches for optic disc and cup segmentation tasks are also mostly adversarial learning-based, which heavily rely on the balance between the source and target datasets to align the features. This dependency can lead to ineffective feature alignment when datasets are imbalanced or significantly different, thus limiting the adaptability and performance of the models. In this regard, we propose a diffusion-based framework, named Dynamic Diffusion Flow Unsupervised Domain Adaptation (termed DDF-UDA), for the cross-domain optic disc (OD) and optic cup (OC) segmentation in fundus images. Specifically, we propose an adaption module based on diffusion procedure at both feature and pixel levels to alleviate the cross-domain gaps. In order to modify the domain information of the source image while minimizing changes to its content, we further propose an adjustment strategy based on Nash equilibrium, which could dynamically modify the diffusion steps. Experimental results on public datasets demonstrate that our DDF-UDA can effectively leverage unlabeled data to achieve state-of-the-art performance in OD/OC segmentation.

4.1 Introduction

4.1.1 Motivation

Glaucoma is a chronic eye disease that ranks as the second leading cause of blindness worldwide, just behind cataracts [358]. One of the key biomarkers for diagnosing and analyzing glaucoma is the cup-to-disc ratio (CDR), which measures the ratio of the optic cup (OC) diameter to the optic disc (OD) diameter in fundus images [283]. Therefore, accurately segmenting the OC and OD is essential for obtaining precise CDR values, which are critical for clinical diagnosis [139, 311, 162]. In recent years, various deep learning-based methods have been developed for OD/OC segmentation, achieving significant success in this field [311, 359]. However, these methods often struggle when applied to new datasets with similar content due to domain gaps, which can severely limit their performance [16]. To address these challenges, unsupervised domain adaptation (UDA) methods have been introduced as a promising solution [240, 175, 68].

In UDA tasks for OD/OC segmentation, adversarial learning is commonly employed to minimize the feature gap between different domains, showing impressive results [311, 158, 338]. For instance, BEAL [310] utilizes entropy-driven adversarial training strategies to enhance boundary prediction for target domain images and reduce feature uncertainty. Similarly, RDR-Net [44] employs a variational auto-encoder (VAE) for image reconstruction and uses dynamic convolution specific to entropy-driven adversarial learning for low-level feature refinement and alignment of prediction maps.

Despite the success of these adversarial learning-based UDA methods, they are highly dependent on maintaining a balance between source and target datasets. The breakdown of this balance will affect the effectiveness of the UDA methods [160]. If the source dataset is significantly larger than the target dataset, the model may develop a bias towards the source samples due to the strong inductive biases inherent in deep learning models. This can lead to poor generalization to the target domain, where the desired adaptation should occur. On the other hand, if the source dataset is much smaller than the target dataset, the model may

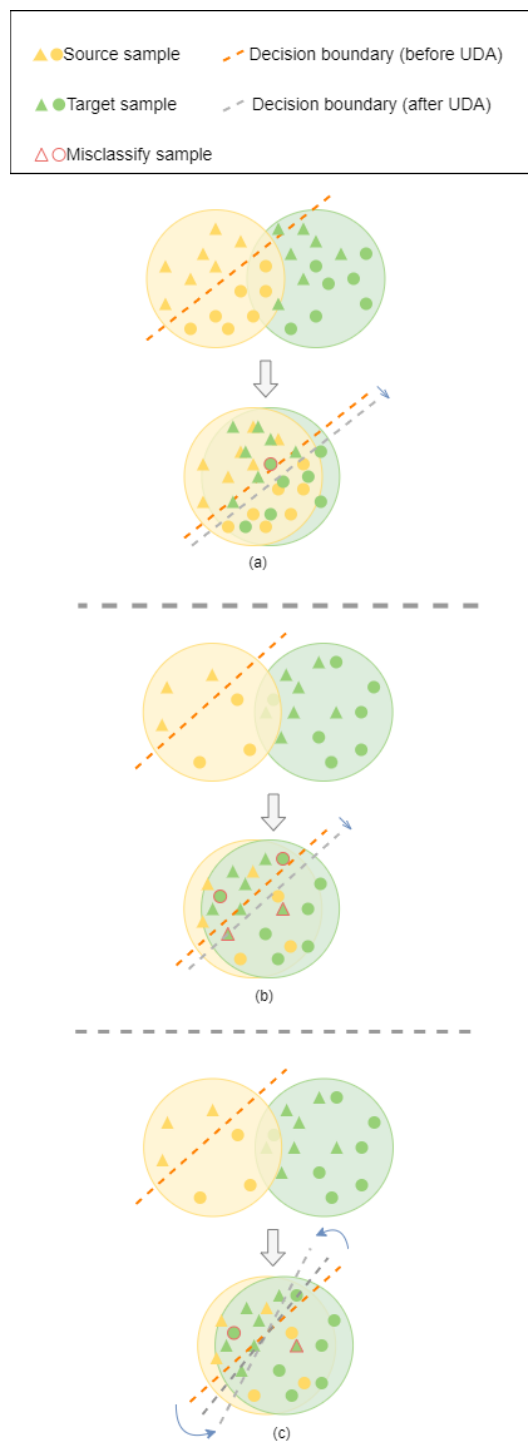


FIGURE 4.1: Schematic diagram of unbalanced source and target domain samples in the UDA method. (a): Existing UDA methods align source and target distributions under the balanced data distribution scenario. (b): Existing UDA methods may result in subpar solutions with indistinct decision boundaries when the source and target domain samples are imbalanced. (c): Our method can correct the decision boundary by diffusing the domain information of target samples into the source ones.

suffer from insufficient training, making it difficult to achieve reliable representation and performance on the target domain [119, 147] (see Figure 4.1 (b)). These challenges highlight the importance of developing UDA strategies that are less reliant on dataset size balance, ensuring effective adaptation and accurate segmentation across diverse datasets.

4.1.2 Contributions

Hence, as illustrated in Figure 4.1 (c), instead of using feature alignment or image reconstruction to reduce cross-domain bias, we propose a cross-domain information fusion strategy through the diffusion model. The diffusion model is a latent variable model that has recently been applied to generative models and gets impressive outcomes [108]. Diffusion models consist of a forward diffusion process and a reverse generation process, which can obtain a trajectory containing a sample by repetitively adding Gaussian noise T times to any initial sample and create new data by reverse traversing the Markov chain [241, 38].

Some work based on the diffusion generation model has addressed the UDA problem by utilizing image generation to reduce the domain gap [336, 328, 218]. Unlike these diffusion UDA methods for synthesized images, our approach mainly focuses on latent feature generation. Specifically, we propose a novel end-to-end UDA network based on the guided diffusion model, which transfers the domain information from the target domain into the source domain through forward diffusion processes.

Previous research [27] has shown that the Batch Normalization (BN) layer can reflect domain-specific information. The output of the BN layer could be seen as a Gaussian distribution with expectation equal to 0 and variance equal to 1. Inspired by recent research [39] which proves the superiority of utilizing the diffusion model from a feature perspective, we conduct diffusion on the feature-level.

Accordingly, we define the output of the encoder’s final BN layer and the decoder’s final BN layer as feature-level and pixel-level domain-specific information, respectively. Then, the target domain’s feature-level and pixel-level features are diffused into the source domain dataset as directional noise (which can be seen as guidance). Furthermore, when applying

the diffusion strategy to transform the source data, we aim to maintain the content of the source domain images to the greatest extent possible during the transition to the target domain. Therefore, we propose a Nash equilibrium strategy to adjust the diffusion process. This approach ensures that the source domain data is effectively transformed into the target domain while preventing over-transformation (which may lead to the original source domain labels being unusable), thus augmenting the dataset. By maintaining a balance, our method effectively augments the dataset, enhancing the overall performance and robustness of the UDA network in cross-domain optic disc and optic cup segmentation tasks.

The main contributions of this paper are summarized as follows:

- We propose a novel dynamic diffusion flow unsupervised domain adaptation framework, named as DDF-UDA, which employs domain diffusion learning to tackle the UDA task associated with joint OD/OC segmentation across diverse fundus image datasets.
- Instead of using feature alignment to bridge the gap between the source and target domains, DDF-UDA employs a diffusion process at both the feature and pixel levels. This approach enables the transfer of domain information from the target data to the source data, thus establishing domain adaptation.
- Given that the transformation from the source domain to the target domain, as well as the maintaining of the data content inherent to the source, represent a contradictory game theory problem, we propose a Nash equilibrium strategy to dynamically adjust the diffusion process to achieve an optimal solution.
- We conduct extensive experiments on two kinds of UDA scenarios: imbalance UDA (where a significant difference in data quantity exists between the source and target domains) and traditional UDA. Experimental results demonstrate the effectiveness of the proposed DDF-UDA compared with existing SOTA methods.

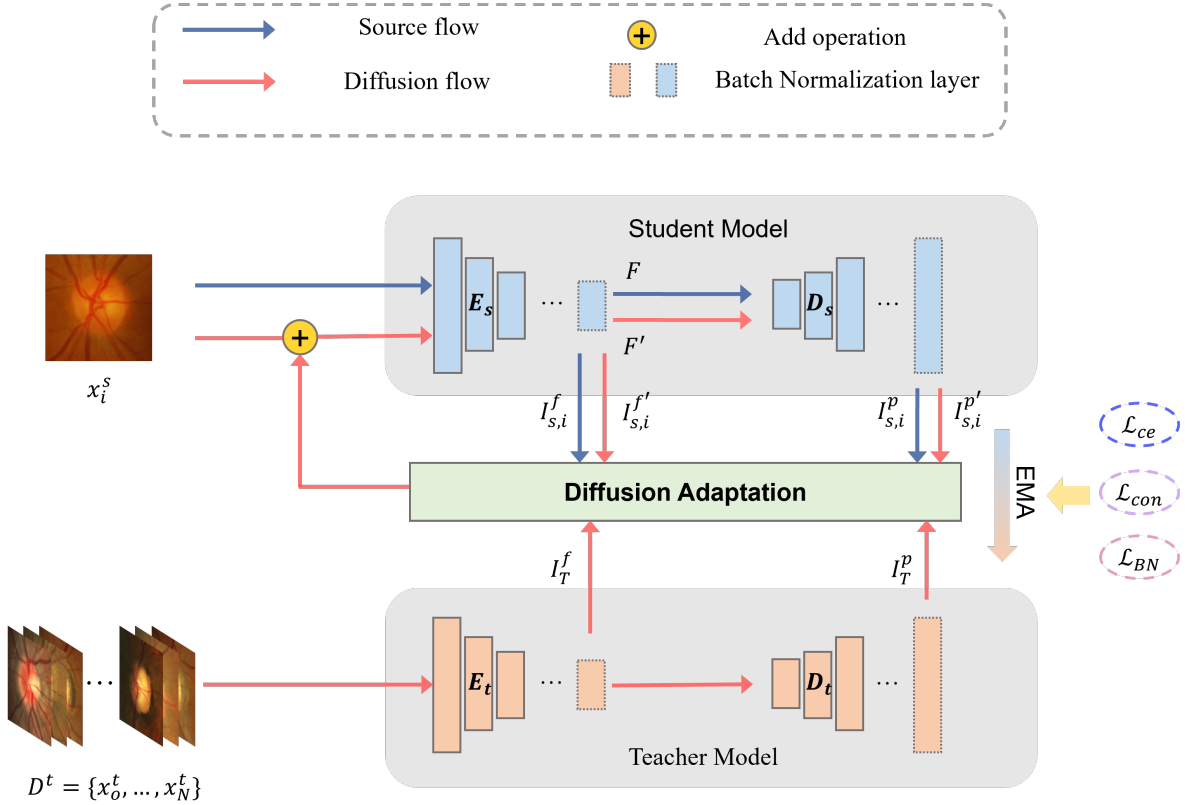


FIGURE 4.2: Overview of our proposed DDF-UDA framework. We use different colors to represent student and teacher models (best views in color).

4.2 Methods

4.2.1 Overview of the Framework

The purpose of UDA is to use the source domain data $X^S = \{x_i^s, y_i^s\}_{i=1}^{N_s}$ and the unannotated target domain data $X^T = \{x_i^t\}_{i=1}^{N_t}$ to optimize the segmentation model. Figure 4.2 shows an overview of the proposed approach. We propose a novel UDA paradigm based on the diffusion model to improve the model’s generalization ability in the target domain dataset. Specifically, our model adapts to different domain distributions by diffusing the domain information from the target dataset into the source domain which has supervised signals.

In detail, our framework employs a dual-stream structure. Given a student network and a teacher network, we use $E_s \& D_s$ and $E_t \& D_t$ to represent the encoder&decoder from different networks, respectively. When x_i^s is input into the student network $E_s \& D_s$, N images from

the target dataset are passed to the teacher network $E_t \& D_t$ simultaneously. Firstly, the source domain feature map F of x_i^s could be extracted by E_s and the output from the BN layer of E_s is obtained as the feature-level source domain information $I_{s,i}^f$. Then, with F passed to D_s , the pixel-level source domain information $I_{s,i}^p$ could be obtained through the BN layer of D_s .

As we once send N samples (which can be seen as a subset) from the target dataset to the framework during each source data training step, we could treat this subset of the target dataset as the target data stream $X_N^T = \{x_n^t\}_{n=1}^N$. The target data stream is input into the teacher model to obtain a set of feature-level and pixel-level target domain-specific information, denoted as $I_T^f = \{I_{t,n}^f\}_{n=1}^N$ and $I_T^p = \{I_{t,n}^p\}_{n=1}^N$, respectively. We regard I_T^f and I_T^p as a type of directional noise for conducting domain-level information diffusion on the source input image to achieve updated $x_{i,N}^{s'}$. $x_{i,N}^{s'}$ is then resent to the student network, and the feature map F' and the feature-level domain-specific information $I_{s,i}^{f'}$ are both obtained through E_s . F' is further passed to D_s to obtain the pixel-level domain-specific information $I_{s,i}^{p'}$. As for the updated $x_{i,N}^{s'}$, the initial label y_i^s is still used as the supervision signal. Hence, one goal of the optimization is to make sure $F \simeq F'$. Furthermore, from the UDA perspective, we strive to ensure that $I_{s,i}^{f'} \simeq I_{t,i}^f$, as well as $I_{s,i}^{p'} \simeq I_{t,i}^p$, to secure the domain information transformation. Following [264], we use exponential moving average (EMA) to update the teacher-student framework. The teacher model, which guides the learning of the student model, often benefits from parameter updates using EMA. By averaging the weights of the teacher model over time, EMA produces a more stable. By leveraging EMA, the teacher model maintains stability and reduces the variance that can arise due to the domain gap. The student network is updated through optimizing, while the teacher network is updated through EMA:

$$\theta_t^{(i)} = \omega \theta_t^{(i-1)} + (1 - \omega) \theta_s^{(i)}, \quad (4.1)$$

where $\theta_t^{(i)}$ denotes the teacher network parameters at i_{th} , $\theta_s^{(i)}$ represents the student network parameters at the same step, and ω is the EMA weight hyperparameter ($\omega = 0.999$ in our case).

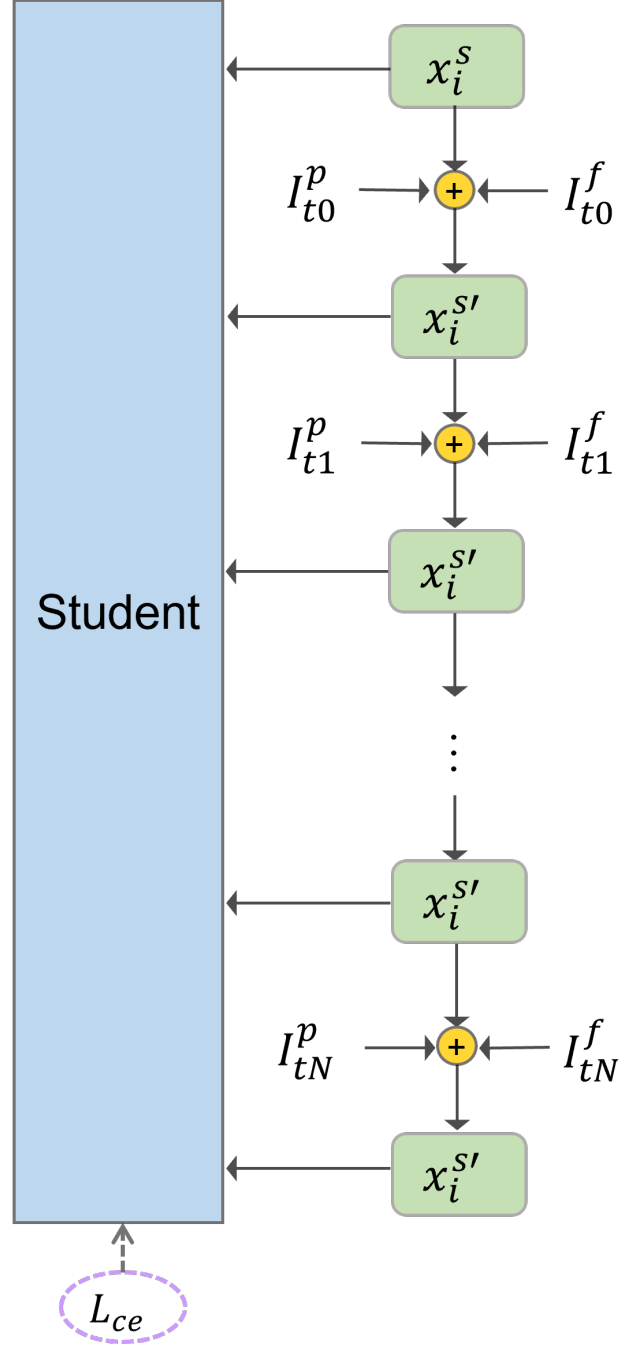


FIGURE 4.3: Diffusion adaptation processing. We use both the feature-level target domain information I_{tN}^f and the pixel-level target domain information I_{tN}^p to update image x_i^s during the diffusion process.

4.2.2 Domain Diffusion Processing

Inspired by existing diffusion models [108, 38, 218], we adapt the forward process used for generating diffusion models into a directed process (see Figure 4.3). In the context of diffusion models, the forward process is a key component that plays a fundamental role in understanding how these models generate data. The forward process gradually transforms the original data into a highly noisy version, effectively creating a sequence of intermediate states that transition from the data distribution to a noise distribution, typically Gaussian noise. This process is characterized by the progressive addition of noise over a series of steps, making the data increasingly unrecognizable as it moves forward through the sequence.

Given a real source domain image sample $x_i^s \sim q(x)$, the domain information transferring process could be formulated as follows [108]:

$$q\left(x_{i,1:N}^{s'} \mid x_i^s\right) = \prod_{n=1}^N \mathcal{N}\left(x_{i,n}^{s'}; \sqrt{1 - \beta_n} x_{i,n-1}^{s'}, \beta_n I_{t,n}^f\right), \quad (4.2)$$

$$\alpha_n = 1 - \beta_n, \bar{\alpha}_n = \prod_{i=1}^n \alpha_i, \quad (4.3)$$

$$x_{i,N}^{s'} = \sqrt{\bar{\alpha}_N} x_i^s + \sum_{n=0}^N \sqrt{1 - \bar{\alpha}_n} I_{t,n}^f, \quad (4.4)$$

where $x_{i,N}^{s'}$ denotes the updated x_i^s . The product $\prod_{n=1}^N$ indicates that the distribution is factorized over n steps, and $\mathcal{N}$ represents a Gaussian distribution. Both α_n and β_n are adjustable hyperparameters.

Additionally, except $I_{t,i}^f$, we also introduce pixel-level domain information for the diffusion process. For the pixel-level features, we aim to eliminate semantic information as much as possible while maintaining domain information. Hence, we propose an assumption that the image contains only two types of information: semantic level and domain level. We introduce a Pixel-level Feature Extraction Threshold (PFT). Suppose after passing through the dual-stream model, the segmentation results in a poor agreement between the two models (below a specified threshold) while still maintaining similar performance (which can be seen

through the evaluation metrics). We assume that under this scenario, the output of the features by the two decoders mostly consists of domain information [314]. Therefore, we further adapt Equation 4.5 as follows:

$$x_{i,N}^{s'} = \sqrt{1 - \bar{\alpha}_n - \bar{\gamma}_n} x_i^s + \sum_{n=0}^N \sqrt{\bar{\alpha}_n} I_{t,n}^f + \sum_{n=0}^N \sqrt{\bar{\gamma}_n} I_{t,n}^p, \quad (4.5)$$

where α_n and γ_n both denote adjustable hyperparameters.

4.2.3 Optimization Processing

For optimization, we use the cross-entropy loss as our supervision loss for segmentation, which can be denoted as $\mathcal{L}_{ce}$:

$$\mathcal{L}_{ce} = -\frac{1}{M} \sum_{i=1}^M y_i^s \log(p_{y_i^s}). \quad (4.6)$$

Note that $\mathcal{L}_{ce}$ is applied to both the original source input x_i^s and the updated source $x_{i,N}^{s'}$. Similarly, the updated segmentation loss $\mathcal{L}_{ce'}$ are defined as:

$$\mathcal{L}_{ce'} = -\frac{1}{M} \sum_{i=1}^M y_i^s \log(p'_{y_i^s}), \quad (4.7)$$

where M is the total number of pixels, and y_i^s is the ground-truth label. $p_{y_i^s}$ and $p'_{y_i^s}$ are the predicted probability of x_i^s and $x_{i,N}^{s'}$, respectively.

As mentioned in Sec. 4.2.1, we aim to induce the similarity between domain information from both source and target data for effective adaptation. Additionally, the domain characteristics of $I_{s,i}^{f'}$ and $I_{t,i}^f$, as well as $I_{s,i}^{p'}$ and $I_{t,i}^p$, need to be extremely similar to guarantee that the domain state of the image x_i^s transitions entirely from source to target. Therefore, we propose the BN loss $\mathcal{L}_{BN}$ related to domain information and the consistency loss $\mathcal{L}_{con}$ to ensure the consistency of the source domain supervision model to combine the supervision signal to effectively facilitate model convergence. $\mathcal{L}_{con}$ is calculated by Kullback-Leibler divergence as follows:

$$\mathcal{L}_{con} = \sigma \left[\sum_{i=1}^N F \cdot (\log F - \log F') \right], \quad (4.8)$$

where σ denotes the activation function. $\mathcal{L}_{BN}$ is calculated by cosine similarity as follows:

$$\mathcal{L}_{BN} = \mathcal{L}_{cos}(I_{s,i}^p, I_t^p) + \mathcal{L}_{cos}(I_{s,i}^{f'}, I_t^f), \quad (4.9)$$

$$\mathcal{L}_{cos} = \frac{1}{N} \sum_{i=1}^N \left(1 - \frac{I_{s,i}' \cdot I_t}{\|I_{s,i}'\| \cdot \|I_t\|} \right). \quad (4.10)$$

Hence, the total loss of the DDF-UDA framework can be formulated as

$$\mathcal{L}_{total} = \mathcal{L}_{ce} + \mathcal{L}'_{ce} + \mathcal{L}_{con} + \mathcal{L}_{BN}. \quad (4.11)$$

Note that the dual-stream structure and the diffusion process are only employed during the training phase. For inference, the test images are directly sent to the student network to obtain the final prediction masks.

4.2.4 Nash Equilibrium

4.2.4.1 Proof of Convergence Conditions

As mentioned in Chapter 4.1, due to the opposite convergence conditions for domain transformation and the preservation of source domain data's inherent content, it is difficult to find the optimal solution for model parameters during the training process.

Given an input source image x_i^s and a network f parameterized by θ , we can formulate the prediction result as $f(x_i^s; \theta)$. Moreover, the task gradient of each training process can be expressed as:

$$\mathbf{g}_t = \frac{\partial \mathcal{L}(f(x_i^s; \theta), y_i^s)}{\partial \theta}, \quad (4.12)$$

where $\mathcal{L}$ denotes the loss function. Similarly, given the updated x_i^s , the task gradient of each training process should be:

$$\mathbf{g}_t' = \frac{\partial \mathcal{L}(f(x_N^s; \theta), y_i^s)}{\partial \theta} \quad (4.13)$$

$$= \frac{\partial \mathcal{L}(f(\sqrt{1 - \bar{\alpha}_n} - \bar{\gamma}_n x_i^s + \sqrt{\bar{\alpha}_n} I_T^f + \sqrt{\bar{\gamma}_n} I_T^p; \theta), y_i^s)}{\partial \theta}, \quad (4.14)$$

where $\bar{\alpha}_n$ and $\bar{\gamma}_n$ denote adjustable hyper-parameters, and I_T^f and I_T^p denote feature-level and pixel-level target domain-specific information, respectively. For the domain gradient, we use f_d to represent the network to avoid confusion:

$$\mathbf{g}_d = \frac{\partial \mathcal{L}_{BN}(I_{s,i}^f, I_t^f)}{\partial \theta} + \frac{\partial \mathcal{L}_{BN}(I_{s,i}^p, I_t^p)}{\partial \theta} \quad (4.15)$$

$$= \frac{\partial \mathcal{L}_{BN}(f_d(\sqrt{1 - \bar{\alpha}_n - \bar{\gamma}_n} x_i^s + \sqrt{\bar{\alpha}_n} I_T^f + \sqrt{\bar{\gamma}_n} I_T^p; \theta), I_T)}{\partial \theta}, \quad (4.16)$$

where I_T denotes the target domain-specific information from both feature-level and pixel-level. To preserve source domain data's inherent characteristics, each input image x_i^s is a fixed value, which can be seen as a constant term. Hence, the variables of Equation 4.14 are I_T^f and I_T^p , which are decided by the number of target input data stream images N . To preserve the data's inherent characteristics, the following conditions should be satisfied:

$$\mathbf{g}_t' \rightarrow \mathbf{g}_t. \quad (4.17)$$

Based on Equation 4.12 and Equation 4.14, Equation 4.17 can be also written as:

$$f(\sqrt{1 - \bar{\alpha}_n - \bar{\gamma}_n} x_i^s + \sqrt{\bar{\alpha}_n} I_T^f + \sqrt{\bar{\gamma}_n} I_T^p; \theta) \rightarrow f(x_i^s; \theta). \quad (4.18)$$

Since $\bar{\alpha}_n, \bar{\gamma}_n$ and x_i^s are all constant terms, the convergence condition of Equation 4.18 is:

$$I_T^f, I_T^p \rightarrow 0. \quad (4.19)$$

For the domain transformation of source image x_i^s , the domain gradient $\mathbf{g}_d$ needs to satisfy $\mathbf{g}_d \rightarrow 0$. As shown in the main paper, the $\mathcal{L}_{BN}$ used in our method is cosine similarity loss. Hence, to satisfy $\mathbf{g}_d \rightarrow 0$, the following condition should be satisfied:

$$f_d(\sqrt{1 - \bar{\alpha}_n - \bar{\gamma}_n} x_i^s + \sqrt{\bar{\alpha}_n} I_T^f + \sqrt{\bar{\gamma}_n} I_T^p; \theta) \rightarrow I_T. \quad (4.20)$$

Similar to Equation 4.18, since $\bar{\alpha}_n, \bar{\gamma}_n$ and x_i^s are all constant terms, the convergence condition of Equation 4.20 is:

$$I_T^f, I_T^p \rightarrow I_T. \quad (4.21)$$

As mentioned in Chapter 4.2.1, I_T is obtained through the target data stream, which is a non-zero value. Therefore, according to Equation 4.19 and Equation 4.21, the convergence

condition of domain transformation and the preservation of source domain data's inherent characteristics are opposite, which makes it impossible for the training process to satisfy the convergence conditions of both simultaneously. Hence, we propose a Nash equilibrium strategy to dynamically modify the diffusion steps to achieve a local optimal for our proposed network for the best performance.

4.2.4.2 Dynamic Nash Equilibrium Strategy

Hence, instead of setting a fixed number of target input data stream images N , we set it as a dynamic one based on the training loss to force the training process to reach a local optimum for the best performance of our method.

In detail, when N target samples are proceeded to transfer the domain information, which will facilitate $\mathcal{L}_{BN}$ reduction and convergence. However, this would significantly alter the original information of the source image. Thus, $\mathcal{L}_{con}$ will deteriorate. To this end, for the $\mathcal{L}_{BN}$ and the $\mathcal{L}_{con}$, their convergence conditions are inconsistent, making it challenging to obtain a global optimal solution [40]. This situation presents a game theory problem: the mathematical modeling of rational behavior in interdependent situations. Therefore, our goal is no longer to attain a globally optimal solution but to achieve local optimal situations for each objective function.

Nash equilibrium is a game theory strategy, which can dynamically adjust training strategies to achieve local optimal solutions while ensuring the stability of the training process [117, 247]. Drawing inspiration from the Nash equilibrium method, after assigning an initial value to the number of target images, we use the $\mathcal{L}_{BN}$ and $\mathcal{L}_{con}$ values obtained in each diffusion process to correct the number of target images N . During the diffusion process, the $\mathcal{L}_{BN}/\mathcal{L}_{con}$ ratio is noted as $\mathcal{R}_1$. If $\mathcal{R}_1$ is greater than 1, indicating that $\mathcal{L}_{BN}$ is greater than $\mathcal{L}_{con}$, it is necessary to increase N to decrease the $\mathcal{L}_{BN}$. Conversely, we need to reduce the value of N . To this end, the training process will keep establishing a dynamic competition until a state of equilibrium (Nash equilibrium) is achieved by both parties. The following formula shows the

Algorithm 1 DDF-UDA Algorithm

Require: Set of source domain data X^S and set of target domain data X^T .

Student model parameters θ , learning rate η ;

for x_i^s in X^S **do**

1:**Input** image x_i to the student model;

2:**Produce** feature map F ;

3:**Produce** predicted mask y_p ;

4:**Compute** $\mathcal{L}_{ce} = \mathcal{L}_{CE}(y, y_p)$;

5:**Update** $\theta = \theta - \eta \nabla_{\theta} \mathcal{L}_{ce}$;

6:**Sample** N samples from X^T denote as X_N^T ;

for x_n^t in X_N^T **do**

7:**Input** image x_n^t to the teacher model;

8:**Produce** feature-level target domain specific information $I_{t,n}^f$;

9:**Produce** pixel-level target domain specific information $I_{t,n}^p$;

10:**Update** $x'_{i,N} = \sqrt{1 - \hat{\alpha}_n - \hat{\gamma}_n} x_i^s + \sum_{n=0}^N \hat{\alpha}_n I_{t,n}^f + \sum_{n=0}^N \hat{\gamma}_n I_{t,n}^p$;

11:**Input** $x'_{i,N}$ to the student model;

12:**Produce** updated feature map F' ;

13:**Produce** predicted mask y'_p ;

14:**Produce** feature-level domain specific information $I_{t,n}^{f'}$;

15:**Produce** pixel-level domain specific information $I_{t,n}^{p'}$;

16:**Compute** denoised segmentation loss $\mathcal{L}'_{ce} = \mathcal{L}_{CE}(y, y'_p)$;

17:**Compute** $\mathcal{L}_{BN} = \mathcal{L}_{cos}(I_{s,i}^{p'}, I_t^p) + \mathcal{L}_{cos}(I_{s,i}^{f'}, I_t^f)$;

18:**Compute** $\mathcal{L}_{con} = \sigma[\sum_{i=1}^N F \cdot (\log F - \log F')]$;

19:**Compute** $\mathcal{L} = \mathcal{L}_{ce} + \mathcal{L}'_{ce} + \mathcal{L}_{con} + \mathcal{L}_{BN}$;

20:**Update** $\theta = \theta - \eta \nabla_{\theta} \mathcal{L}$;

end for

21:**Update** Teacher model with θ using EMA;

end for

Ensure: Model parameters θ .

objective of the Nash equilibrium strategy:

$$u_1(\mathcal{L}_{BN}^*, \mathcal{L}_{con}^*) \geq u_0(\mathcal{L}_{BN}, \mathcal{L}_{con}), \quad (4.22)$$

$$u_N(\mathcal{L}_{BN}^*, \mathcal{L}_{con}^*) \geq u_{N-1}(\mathcal{L}_{BN}^*, \mathcal{L}_{con}^*), \quad (4.23)$$

where u_0 denotes the first strategy, and u_1 denotes the second strategy. u_N and u_{N-1} denote the strategy for the N th and $(N-1)$ th optimization, respectively. $\mathcal{L}_{BN}$ and $\mathcal{L}_{con}$ are the losses without adjustment, and $\mathcal{L}_{BN}^*$ and $\mathcal{L}_{con}^*$ are the adjusted losses.

TABLE 4.1: Statistics of three fundus datasets used for the proposed method.

Dataset	Resolution	Camera	Number of samples	Release year
DRISHTI-GS [263]	2047×1760	Unknown	50 Train + 51 Test	2014
RIM-ONE-r3 [79]	1072×1424	Canon EOS 5D	99 Train + 60 Test	2015
REFUGE [212]	2124×2056	Zeiss Viscucam 50	320 Train + 80 Test	2018

For a better understanding of our proposed approach, we provide the pseudo-algorithm in Algorithm 1 (DDF-UDA Algorithm).

4.3 Experiments and Results

4.3.1 Dataset Description

We conduct experiments on three public glaucoma diagnostic datasets: RIM-ONE-r3 [79], DRISHTI-GS [263] and REFUGE [212]. The detailed statistics of the datasets are shown in Table 4.1.

4.3.1.1 RIM-ONE-r3 Dataset

The RIM-ONE-r3 dataset [79] is part of the RIM-ONE series, which is specifically designed for glaucoma assessment through the segmentation of the optic disc and optic cup in fundus images. This dataset contains images that are primarily sourced from clinical environments, including different types of glaucomatous and healthy eyes. RIM-ONE-r3 consists of high-resolution fundus images with varying degrees of quality and contrast, reflecting real-world clinical conditions. The dataset provides manual annotations for the optic disc and optic cup, making it suitable for training and validating segmentation algorithms aimed at glaucoma detection. The diversity in image quality and the presence of various ocular conditions in RIM-ONE-r3 pose significant challenges for segmentation models, highlighting the need for robust algorithms that can generalize well across different cases.

4.3.1.2 DRISHTI-GS Dataset

The DRISHTI-GS dataset [263] is another benchmark dataset specifically proposed for optic disc and optic cup segmentation in glaucoma diagnosis. This dataset consists of 101 high-quality fundus images, which are split into training and test sets. Each image comes with expert annotations for both the optic disc and optic cup, providing accurate ground truth data for evaluating segmentation algorithms. The images in DRISHTI-GS are notable for their consistent resolution and clarity, ensuring that the optic disc and cup are well-defined, which is crucial for precise segmentation. The value of the DRISHTI-GS dataset lies not only in the quality of its annotations but also in its inclusion of both normal and glaucomatous eyes. This diversity enables a thorough evaluation of segmentation models across a range of clinical conditions, making it highly relevant for practical applications. Furthermore, the careful curation of this dataset makes it an ideal resource for benchmarking the performance and robustness of segmentation models, ensuring they are capable of accurately identifying key anatomical structures under varying conditions.

4.3.1.3 REFUGE Dataset

The REFUGE (Retinal Fundus Glaucoma Challenge) dataset [212] is introduced as part of a challenge aimed at advancing the development of automated methods for glaucoma detection through optic disc and optic cup segmentation. REFUGE includes a larger collection of fundus images compared to RIM-ONE-r3 and DRISHTI-GS, with 1,200 images sourced from various populations and clinical settings. The dataset is divided into training, validation, and test sets, with manual annotations provided for the optic disc and cup segmentation. The images in REFUGE vary in quality, with some containing artifacts, varying illumination, and different levels of disease progression, which mimic real-world clinical conditions. This diversity makes REFUGE a comprehensive dataset for assessing the robustness and generalizability of segmentation algorithms across different demographic and clinical contexts. The dataset is also used to evaluate the performance of models in detecting glaucoma based on segmented features, making it a valuable resource for both segmentation and diagnostic tasks.

4.3.2 Experiment Details

4.3.2.1 Experiment Setting

We conduct extensive experimental scenarios on OD/OC segmentation for two types of UDA: imbalanced UDA and traditional UDA. For imbalanced UDA, there is a substantial difference in the quantity of data between the source and target domains. We select adapting Drishti-GS to RIM-ONE-r3 and RIM-ONE-r3 to Drishti-GS settings in an imbalanced scenario. In the case of traditional UDA, we adopt the same settings as in the previous works, utilizing the REFUGE dataset as the source domain and the Drishti-GS and RIM-ONE-r3 datasets as the target domains.

4.3.2.2 Evaluation Metrics

Various commonly used metrics are used to evaluate the experimental results. These include the Dice Similarity Coefficient (Dice), which offers pixel-level accuracy evaluation. We also employ Mean Intersection over Union (MIoU) to further compare the segmentation performances of different approaches.

4.3.2.3 Implementation Details

Due to the image size variance across different fundus image datasets, we initially crop the original image to yield a 512×512 Region of Interest (ROI) centered on the OD. For computational efficiency, we subsequently resize the cropped image to a 256×256 format as the input of the network. This study employs standard data augmentation techniques to increase the sample size, incorporating strategies such as elastic transformation, Gaussian noise injection, random erasing, as well as random scaling and cropping. For the segmentation network, we apply DeepLabv3+ with a pre-trained MobileNetV2 as the backbone. The model is implemented on a workstation with one NVIDIA GeForce RTX 3090Ti GPU. For optimization, we set the batch size to 8 with the Adam optimizer using “poly” learning rate policy. The initial learning rate is set to 0.001, with a pixel-level feature extraction threshold set to 0.75. For the adjustable hyper-parameters, both $\bar{\alpha}_n$ and $\bar{\gamma}_n$ are set to 0.01.

TABLE 4.2: Comparison OD/OC segmentation results when adapted from Drishti-GS to RIM-ONE-r3.

Method	Drishti-GS $\rightarrow$ RIM-ONE-r3			
	DI_{disc}	DI_{cup}	$MIoU_{disc}$	$MIoU_{cup}$
ρ OSAL [311]	0.9153	0.8159	0.8957	0.8383
BEAL [310]	0.9208	0.8001	0.9003	0.8345
ISFA [158]	0.9252	0.8220	0.9059	0.8471
RDR-Net [44]	0.9215	0.8282	0.9028	0.8512
Ours	0.9572	0.8324	0.9551	0.8583

TABLE 4.3: Comparison OD/OC segmentation results when adapted from RIM-ONE-r3 to Drishti-GS.

Method	RIM-ONE-r3 $\rightarrow$ Drishti-GS			
	DI_{disc}	DI_{cup}	$MIoU_{disc}$	$MIoU_{cup}$
ρ OSAL [311]	0.9096	0.8303	0.8614	0.8325
BEAL [310]	0.9293	0.8054	0.8927	0.8116
ISFA [158]	0.9324	0.8385	0.8961	0.8413
RDR-Net [44]	0.9434	0.8430	0.9124	0.8455
Ours	0.9265	0.8478	0.9186	0.8655

4.3.3 Performance of the Proposed Method

4.3.3.1 Results on Imbalanced UDA Settings

We report the results of our model under imbalanced UDA settings in Table 4.2 and Table 4.3. Specifically, we first select DRISHTI-GS and RIM-ONE-r3 as the source and target domains, respectively. We compare the proposed method with four state-of-the-art OD/OC segmentation UDA methods, i.e., ρ OSAL [311], BEAL [310], ISFA [158] and RDR-Net [44]. Table 4.2 and Table 4.3 report the segmentation results of adapting Drishti-GS to RIM-ONE-r3 and the segmentation performance when adapting from RIM-ONE-r3 to DRISHTI-GS, with the best results highlighted in bold. For the convenience of description, the dice, and mIoU values of OD/OC are expressed as DI_{disc} , DI_{cup} , $MIoU_{disc}$, and $MIoU_{cup}$.

TABLE 4.4: Quantitative comparison OD/OC segmentation results when adapted from REFUGE to Drishti-GS and adapted from REFUGE to RIM-ONE-r3.

Method	Drishti-GS		RIM-ONE-r3	
	DI_{disc}	DI_{cup}	DI_{disc}	DI_{cup}
No Adapt	0.930	0.854	0.822	0.703
Upper bound	0.977	0.910	0.969	0.877
Hoffman <i>et al.</i> [110]	0.959	0.851	0.852	0.755
Javanmardi <i>et al.</i> [125]	0.961	0.849	0.853	0.779
ρ OSAL [311]	0.965	0.858	0.865	0.787
BEAL [310]	0.961	0.862	0.898	0.810
ISFA [158]	0.966	0.892	0.908	0.822
RDR-Net [44]	0.971	0.893	0.918	0.840
Ours	0.969	0.893	0.962	0.850

Table 4.2 illustrates that in adaptive OD/OC segmentation tasks with imbalanced data, our DDF-UDA consistently yields competitive results across various scenarios. As the results are shown in Table 4.2, which are performed on the Drishti-GS to RIM-ONE-r3 adaptation task, our method achieves state-of-the-art performance on all evaluation metrics. Specifically, DDF-UDA outperforms the prior work in terms of Dice of OD segmentation of 3.57%, improving from 92.15% (by RDR-Net) to 95.72%. Additionally, our proposed method exhibits remarkable performance in MIoU metrics for OD/OC segmentation. In particular, for $MIoU_{disc}$, there were respective increases of 5.23% compared to RDR-Net. In addition, we also report the results of adapting from RIM-ONE-r3 to Drishti-GS.

As shown in Table 4.3, under the setting of adapting RIM-ONE-r3 to Drishti-GS, our proposed framework performs better than existing methods on 3 out of 4 evaluation metrics. The above results validate the effectiveness of DDF-UDA, which leverages information diffusion to bridge the domain gap between the source and target domains. Notably, when the target dataset is larger than the source one, our method proves especially effective in adaptation tasks by diffusing target domain information, surpassing the performance of adversarial learning methods which rely on domain alignment.

TABLE 4.5: Ablation study results of adapting Drishti-GS to RIM-ONE-r3 and adapting REFUGE to Drishti-GS.

Method	Baseline	DM	NE	PFT	DI_{disc}	DI_{cup}	$MIoU_{disc}$	$MIoU_{cup}$
Drishti-GS $\rightarrow$ RIM-ONE-r3								
No Adapt	✓				0.8952	0.7425	0.8938	0.8074
	✓	✓			0.9530	0.8203	0.9510	0.8504
	✓	✓	✓		0.9580	0.8234	0.9558	0.8520
	✓	✓	✓	✓	0.9572	0.8324	0.9551	0.8583
REFUGE $\rightarrow$ Drishti-GS								
	✓				0.9298	0.8542	0.9201	0.8700
	✓	✓			0.9543	0.8891	0.9611	0.8890
	✓	✓	✓		0.9628	0.8814	0.9629	0.8865
	✓	✓	✓	✓	0.9686	0.8931	0.9640	0.9010

4.3.3.2 Results on Traditional UDA Settings

To assess the performance of our method on traditional datasets, we employ the REFUGE dataset as the source domain, and use Drishti-GS and RIM-ONE-r3 as the target domains separately. For the results in Table 4.4, we compare our proposed method with six SOTA UDA methods and two baselines. In addition to the four previously mentioned SOTA OD/OC segmentation UDA approaches, we also compare the globally aligned method proposed by Hoffman *et al.* [110] and the adversarial learning-based UDA algorithm by Javanmardi *et al.* [125]. For the baselines, we also train a DeepLabV3+ with source data only (“No Adapt”) and with annotated target data only (“Upper bound”), respectively.

The results in Table 4.4 show that DDF-UDA achieves highly competitive results under the traditional UDA settings. On the Drishti-GS dataset, the Dice scores for OD/OC segmentation are nearly equivalent to the SOTA methods. In the adaptation from REFUGE to RIM-ONE-r3, DDF-UDA achieves superior segmentation performance. Specifically, our method achieves an impressive improvement on Dice score for OD segmentation ($91.8\% \rightarrow 96.2\%$), indicating a significant 4.4% gap and approaching the performance of fully supervised results (96.9%). These reported results demonstrate that in the scenarios where the source domain contains

TABLE 4.6: Additional ablation study results of adapting Drishti-GS to RIM-ONE-r3.

Methods	Baseline	DM	NE	PFT	DI_{disc}	DI_{cup}	ASD_{disc}	ASD_{cup}
Drishti-GS $\rightarrow$ RIM-ONE-r3								
No Adapt	✓				0.8952	0.7425	13.0813	12.4975
	✓	✓			0.9530	0.8203	4.9301	8.4319
	✓	✓	✓		0.9580	0.8234	3.6613	7.6471
Ours	✓	✓	✓	✓	0.9572	0.8324	3.7097	7.0776
Dis & PL					0.9013	0.8146	6.1226	8.9716

more data than the target domain, our proposed DDF-UDA consistently delivers advanced UDA segmentation performance through dynamically diffusing target domain information.

4.3.4 Ablation Study

To assess the impact of various modules within our proposed model, we conduct ablation experiments to evaluate the effectiveness of each module. We conduct these experiments using Drishti-GS as the source domain and RIM-ONE-r3 as the target domain (imbalanced UDA setting), and then using REFUGE as the source domain and Drishti-GS as the target domain (traditional UDA setting) for clarity and consistency. Each module is designated as follows: DM for the Diffusion Module, NE for Nash Equilibrium, and PFT for the Pixel-level Feature Extraction Threshold. The outcomes of the different module implementations are visually presented in Table 4.5. We utilized a model trained only on the source domain as our baseline.

The results illustrate that as the DM, NE, and PFT modules are incrementally integrated into the segmentation model, all evaluation metrics exhibited continuous improvement. It is notable that in the absence of domain adaptation, the baseline displayed relatively poor performance, with the lowest Dice scores for OD and OC in both two tasks. In contrast, the DM module leads to an increase in Dice scores by 5.78% & 7.78% (imbalanced UDA) and 2.45% & 3.49% (traditional UDA) when compared to the baseline. It is important to highlight that the DM module significantly contributes to improving segmentation accuracy. This is

attributed to the DM module dynamically diffusing domain information from the target domain data stream into the source domain training model, thereby improving domain adaptation effects. When the NE module is introduced into the network, segmentation performance also shows varying degrees of improvement. Finally, experimental results demonstrate that the addition of the PFT module can improve both OD/OC segmentation results.

To provide a more comprehensive comparison of the experimental results, we conducted additional experiments on the ablation settings of adapting Drishti-GS to RIM-ONE-r3 in Tab 4.6. In our analysis, we included the Average Symmetric Surface Distance (ASD) for the ablation experiments. We also added the adversarial learning-based UDA modules in our base segmentation model. Specifically, we added a discriminator to the student model and also followed the teacher-student framework to apply pseudo-labeling to the student model as an extra supervision signal. Based on the results, our diffusion-based method has a stronger adaption ability than existing adversarial learning-based UDA methods.

4.3.5 Visualization Results

Figure 4.4 presents the qualitative results of optic disc (OD) and optic cup (OC) segmentation, where OD is highlighted in green and OC in blue. These results demonstrate the effectiveness of our proposed approach in adapting from the Drishti-GS dataset to the RIM-ONE-r3 dataset. The ability to maintain precise segmentation across domains highlights the robustness of our adaptation strategy in handling domain shifts commonly encountered in fundus image segmentation tasks.

To further assess the effectiveness of our method, we conduct an ablation study under the same adaptation setting and visualize the corresponding results in Figure 4.5. These visual comparisons illustrate the impact of different network configurations on segmentation performance. Specifically, we observe that removing key components of our framework leads to degraded segmentation quality, with less precise boundaries and increased misclassification in OD and OC regions. In contrast, our proposed method consistently achieves the most accurate and reliable segmentation results, demonstrating superior domain adaptation capability.

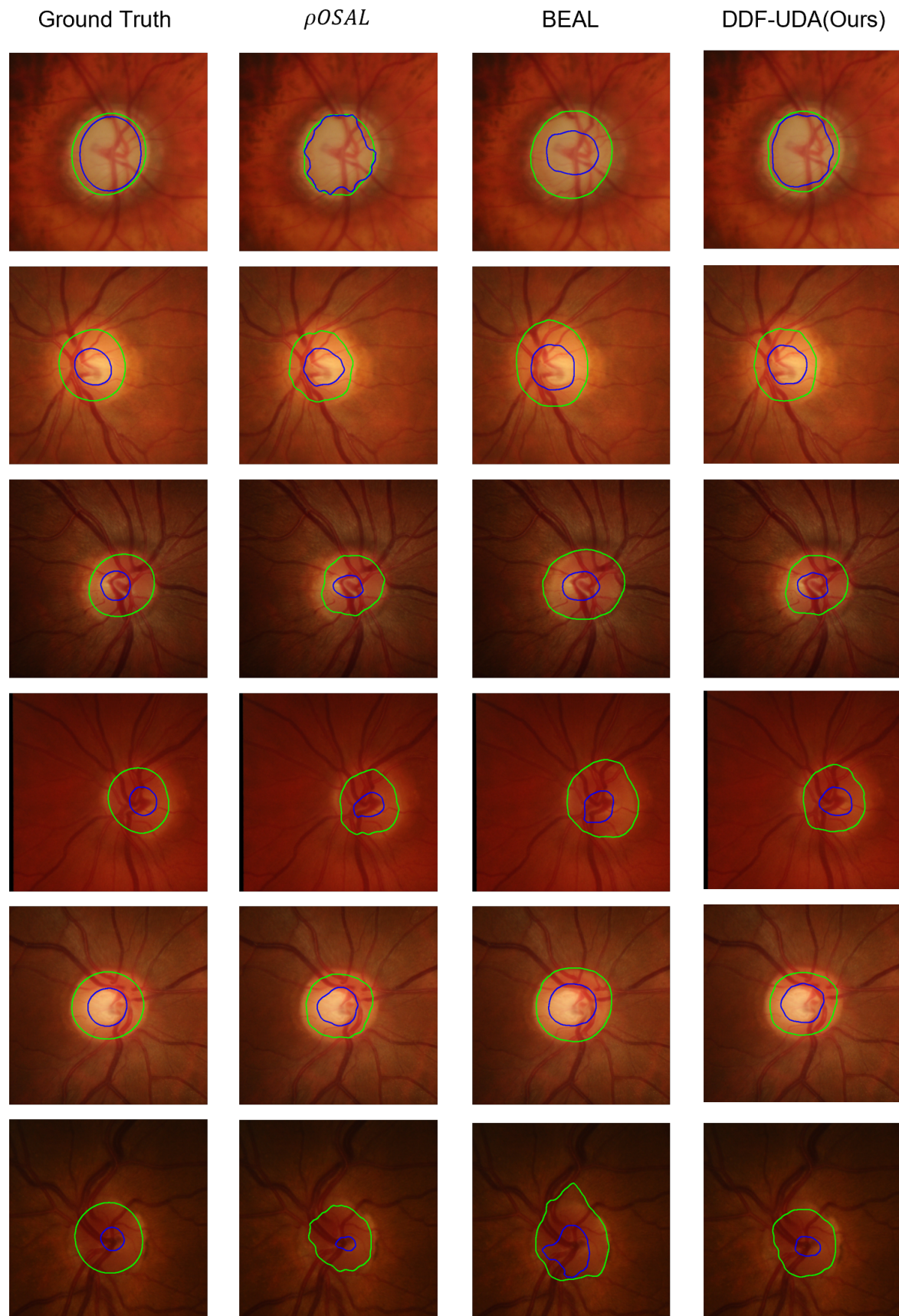


FIGURE 4.4: Qualitative results of OD (green) and OC (blue) segmentation, which adapted from Drishti-GS to RIM-ONE-r3.

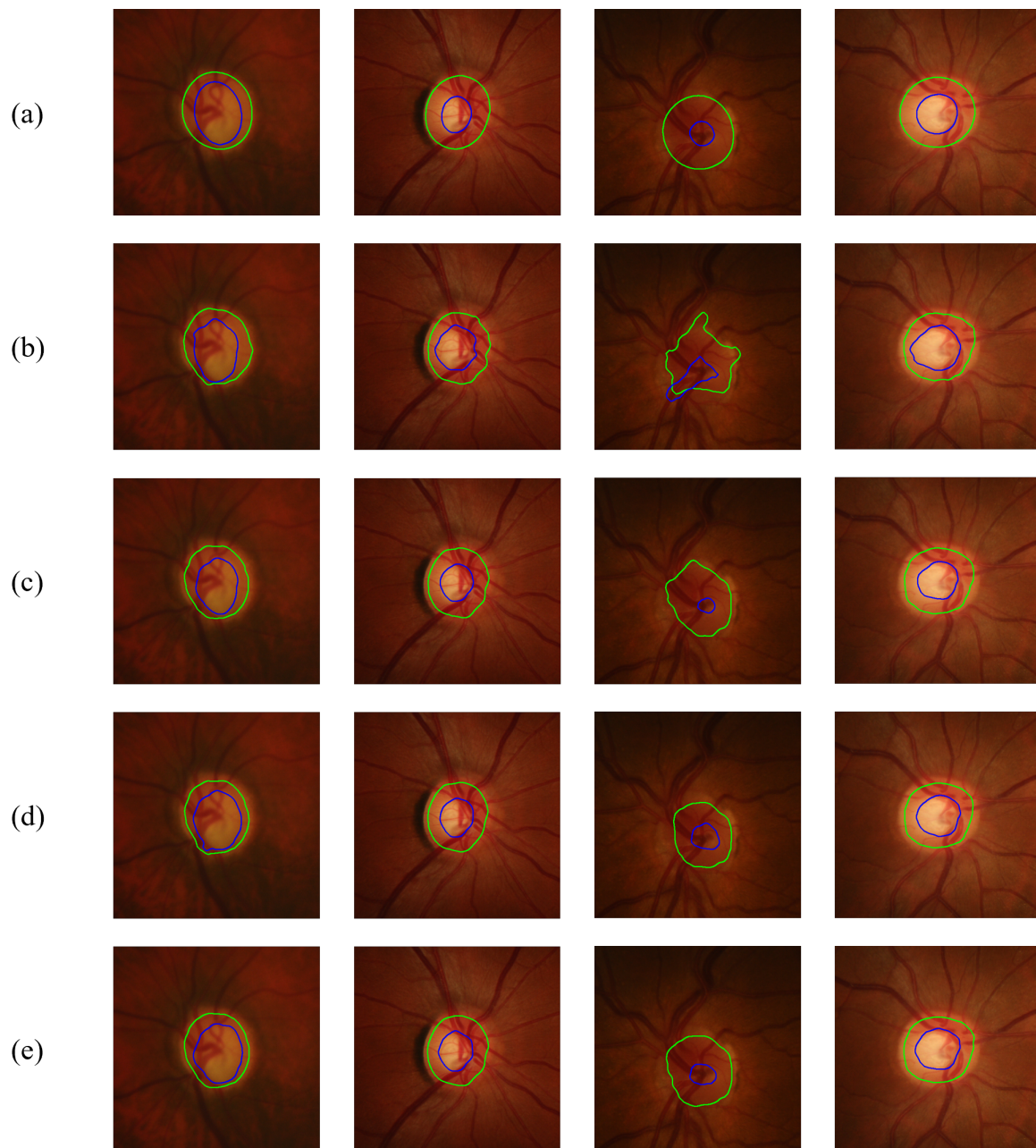


FIGURE 4.5: Visualization of OD (green) and OC (blue) segmentation results for the ablation study conducted from Drishti-GS to RIM-ONE-r3. (a): original images, (b): BL: Baseline, (c): BL+DM, (d): BL+DM+NE, (e): Ours.

4.4 Discussion

The DDF-UDA study represents a significant advancement in the field of unsupervised domain adaptation (UDA) for medical image segmentation, particularly addressing the challenges of cross-domain segmentation tasks. This research introduces a novel dynamic diffusion-based framework that not only surpasses traditional feature alignment methods but also improves the accuracy of joint optic disc (OD) and optic cup (OC) segmentation. By implementing a diffusion process at both the feature and pixel levels, the framework effectively transfers domain-specific information across diverse fundus image datasets, which are typically marked by significant variations in imaging conditions.

The deep learning framework developed in this research has demonstrated superior performance, achieving outstanding results on multiple fundus datasets, both in imbalance UDA, where there is a significant difference in data distribution, and traditional UDA scenarios. This achievement marks a novel contribution to cross-domain medical image analysis by offering a more flexible and scalable approach to handling domain shifts. This study achieves the best performance across these diverse scenarios, highlighting the robustness and adaptability of the proposed deep learning approach. This innovative strategy not only addresses the domain gap more effectively but also allows for a deeper understanding of domain-specific information, providing a comprehensive solution for UDA tasks in medical imaging.

Building upon this research, we further explored the concentration of domain-specific information in different layers. Following that, we discussed the limitations of this work and outlined directions for future research.

4.4.1 Domain-specific Information Extraction

In Chapter 4.1.2, previous studies [27] have demonstrated that the Batch Normalization (BN) layer plays a critical role in capturing domain-specific information, making it a valuable tool in domain adaptation tasks. The output of the BN layer can be modeled as a Gaussian distribution with a mean of 0 and a variance of 1. This characteristic allows the BN layer to

TABLE 4.7: Experiment results of adapting Drishti-GS to RIM-ONE-r3 and adapting REFUGE to Drishti-GS of different domain-specific information extraction methods.

Method	DI_{disc}	DI_{cup}	$MIOU_{disc}$	$MIOU_{cup}$
Drishti-GS $\rightarrow$ RIM-ONE-r3				
w/o BN	0.8879	0.7581	0.8964	0.8153
VarBN	0.9384	0.8093	0.9426	0.8376
BN	0.9572	0.8324	0.9551	0.8583
REFUGE $\rightarrow$ Drishti-GS				
w/o BN	0.9203	0.8477	0.9235	0.8713
VarBN	0.9481	0.8640	0.9413	0.8647
BN	0.9686	0.8931	0.9640	0.9010

normalize features across different domains, implicitly capturing domain-specific variations. In our approach, we leverage the features extracted from the BN layer and treat them as key indicators of domain-specific information.

To investigate where this domain-specific information is most effectively concentrated, we design a series of experiments aimed at understanding the impact of different feature extraction methods. First, we utilize the output from the encoder/decoder's final BN layer, using these features directly within our model for domain adaptation. This method is based on the hypothesis that the BN layer's normalization process effectively captures critical domain-specific differences. We refer to this method as "BN". Next, we test the model's final output feature to compare its effectiveness against the BN layer's output (w/o BN). Finally, we explored an alternative approach by extracting only the variance component from the BN layer's output and generating Gaussian noise using this variance, aiming to assess the significance of the variance alone in capturing domain-specific traits. We abbreviated this approach as "VarBN."

The experiments (see Table 4.7) show that using the output from the encoder/decoder's final BN layer, as proposed in our method, resulted in the highest performance. This outcome highlights the idea that the BN layer's features gain essential domain-specific information, contributing significantly to the model's ability to adapt to new domains. By focusing on these

BN-layer features, our method enhances the overall effectiveness of domain adaptation in medical image segmentation tasks, highlighting the critical role of the BN layer in addressing domain shifts.

4.4.2 Limitations and Future Work

Although our method demonstrates significant performance improvements, several limitations need to be discussed. First, while DDF-UDA applies to optic disc and cup segmentation tasks, it is not explicitly tailored or optimized for the specific characteristics of the OD/OC segmentation dataset. The lack of dataset-specific adjustments may limit the method from reaching its full potential in these tasks. A more fine-tuned approach that considers the unique properties of the dataset could enhance performance further.

Moreover, DDF-UDA is currently evaluated in a specific task, and there is significant potential to explore its application across a wider range of datasets and medical imaging tasks. It could extend its application to other datasets to assess its generalizability across different domains. By testing DDF-UDA on a wider variety of datasets, the goal is to demonstrate its robustness and adaptability in diverse medical imaging scenarios.

Our aim is to develop DDF-UDA into a flexible, plug-and-play domain adaptation module that can be easily integrated into various segmentation tasks and datasets. In future work, we plan to refine its design to ensure it can be simply applied across different contexts without the need for extensive modifications. This will enable DDF-UDA to become a more universally applicable tool, enhancing its utility and effectiveness in addressing a wide range of domain adaptation challenges in medical image segmentation.

4.5 Summary

In this paper, we propose a novel DDF-UDA method to tackle the UDA for OD/OC segmentation under both balanced and imbalanced data conditions. Instead of relying on feature alignment to bridge the gap between the source and target domains, our method implements a

diffusion process at both feature and pixel levels to gradually transfer domain information from the target to the source for adaptation. Considering that the transformation from the source domain to the target domain, and the maintenance of the data content of the source domain, is a self-contradictory game theory problem, we introduce a Nash equilibrium to govern the diffusion process and reach an optimal solution. We conduct extensive experiments on different UDA OD/OC segmentation settings, including traditional UDA settings and unbalanced UDA settings, to demonstrate the effectiveness of our method.

Feature Disentanglement via Language Model for Zero-shot Domain Adaptation

Cross-domain deep learning models are crucial for computer-aided diagnosis. Recent advances in unsupervised domain adaptation (UDA) aim to tackle this challenge by leveraging transfer learning techniques. In Chapter 3 and Chapter 4, we respectively explore UDA methods for medical image segmentation under class imbalance and data imbalance scenarios. Although these approaches remain efficient and effective, they still require access to target images during training, which is not always feasible in real-world medical scenarios due to data availability issues.

Recently, some approaches have been proposed to achieve domain adaptation in a zero-shot learning way, hoping to transfer the domain adaptation capability learned from irrelevant tasks to the unseen target domains. However, these zero-shot domain adaptation (ZSDA) paradigms face significant limitations in medical imaging, where data is often scarce. Additionally, existing methods do not demonstrate strong generalization ability for completely unseen target domains.

To address these challenges, we propose a novel framework that uses the language model as the feature disentangler (named LMaD) for the ZSDA medical image segmentation task. Specifically, we first introduce a graph learning mechanism to construct the visual graph that captures the relationships between visual features. We then utilize a pre-trained language model to obtain task-specific and domain-specific feature embedding. By performing feature disentanglement with the visual graph, we aim to eliminate the interference caused by domain information. Furthermore, a graph convolutional network is introduced to infer higher-order

relationships among task-specific features within the visual graph, further enhancing the contextual information within images. We conduct extensive experiments on fundus image segmentation and prostate MRI segmentation. The experimental results demonstrate that our method significantly improves the generalization ability of ZSDA models and achieves state-of-the-art performance, highlighting its strong potential for cross-domain knowledge transfer.

5.1 Introduction

5.1.1 Motivation

Recent deep learning models have achieved significant success in medical image segmentation [190, 360, 319]. However, due to the distribution shifts [28, 95] caused by different data sources (i.e., diverse equipment, medical institutions, imaging protocols, and patient populations), the performance of a model trained on the source domain usually declines on the target domain [310, 71]. While fine-tuning the new domain is a simple solution, manual annotation of these datasets by clinicians is extremely expensive and time-consuming [353]. To address the domain shift problem, unsupervised domain adaptation (UDA) methods have gained attention for their ability to bridge the distribution gap between domains [16, 324].

Nevertheless, these methods still require specified target domain data for training to achieve good performance in the corresponding domain [80, 310]. In real-world scenarios, predicting which data will become the target domain is challenging, limiting the generalization ability of these methods. Besides, in some medical scenarios, obtaining target domain images can be difficult due to medical privacy concerns or data scarcity.

Recently, various zero-shot domain adaptation (ZSDA) methods have been introduced for natural images, aiming to adapt models to unseen target domains that are inaccessible during training [219, 127]. As shown in Figure 5.1 (a), existing ZSDA methods mainly focus on introducing irrelevant tasks (IrT) to learn the ability of knowledge transfer between inter-domains within these tasks and then apply this ability to the task of interest (ToI). These

methods use adversarial networks to capture inter-domain transformations, aligning target domains in ToI and achieving strong performance [219, 302, 186]. Inspired by generative-based zero-shot learning methods [20, 8], recent ZSDA methods utilize text features obtained from CLIP to simulate target domain visual representations, modifying prompt descriptions to adjust source domain features to match target domain representations [66, 333]. Due to the lower data requirements of these methods, they have also been applied in the medical field to enhance the model’s flexibility [17, 2].

However, the aforementioned approaches are all based on the assumption that the inter-domain discrepancies are shared across various tasks, which further brings inherent issues: (1) Due to the scarcity of medical data and the constraints of privacy protocols, it is not always possible to find suitable IrT to assist in domain transfer; (2) There is also a domain gap between the IrT and the ToI, making it challenging to achieve complete feature alignment during the transfer, which can result in performance fluctuation in the target domain.

5.1.2 Related Work

5.1.2.1 Zero-shot Domain Adaptation

Zero-shot domain adaptation (ZSDA) assumes that no target images are available during the training phase [219]. A common approach involves using irrelevant tasks (IrT) to learn the inter-domain transformation within these tasks and then transfer this knowledge to the task of interest (ToI) [219, 170, 302, 186].

ZDDA [219] proposes a method to achieve domain adaptation without access to target domain data by using task-irrelevant dual-domain pairs to learn generalized source domain representations. ZSDA [302] introduces a method using adversarial learning to transfer inter-domain transformations learned in IrT to ToI. Recent works have achieved significant success in ZSDA tasks by leveraging vision-language pre-trained models [66, 333]. Among these, PODA [66] introduces a method that uses natural language prompts to guide zero-shot domain adaptation. ULDA [333] proposes a method that converts natural language descriptions into guiding information, enabling domain adaptation without target domain images. Despite

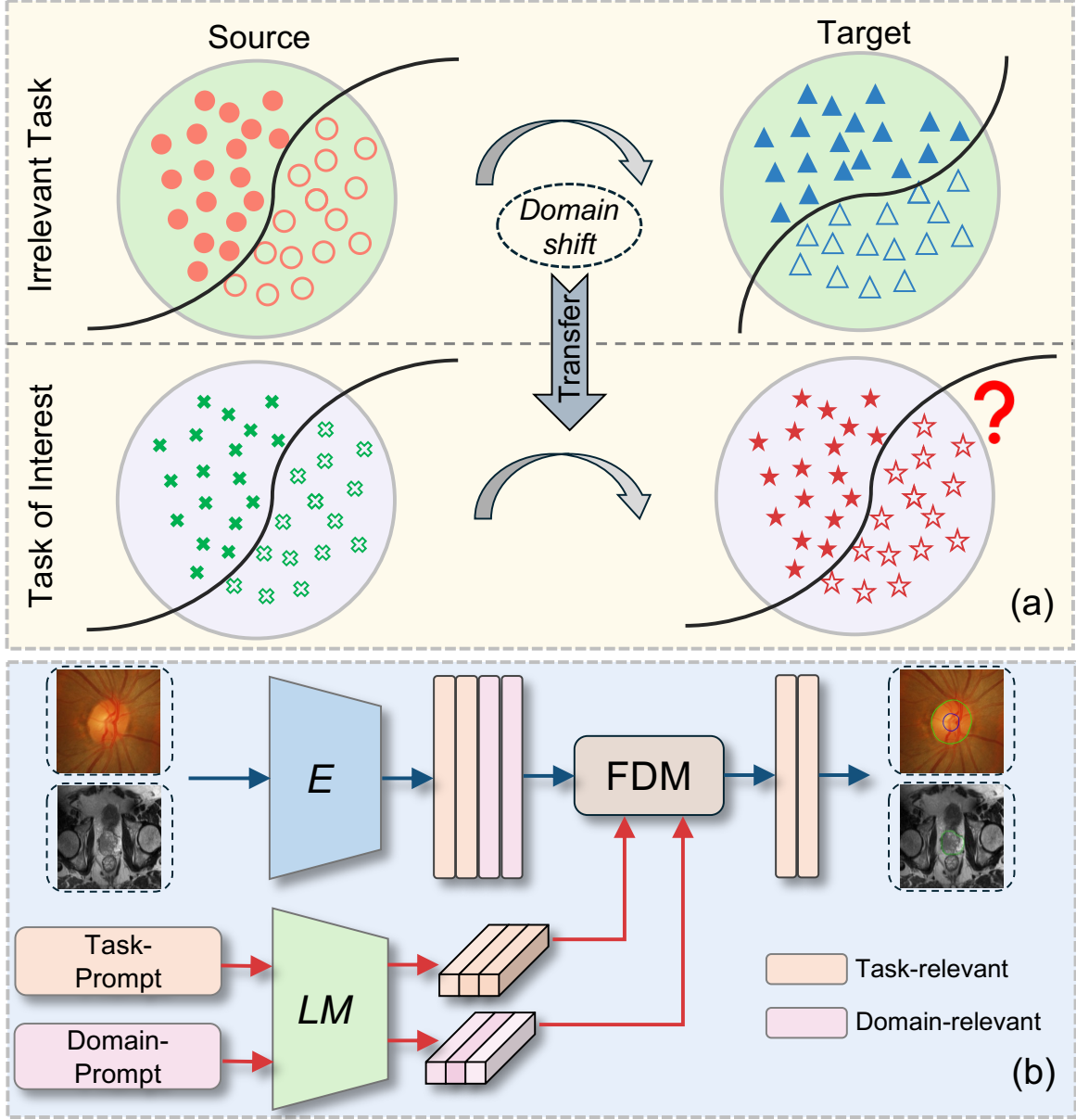


FIGURE 5.1: Illustration of the differences between previous ZSDA methods and our approach. (a): Previous ZSDA methods introduce IrT to learn domain transformations and transfer this knowledge to the ToI. (b): Our LMaD uses a language model as a disentangler to eliminate the interference caused by domain information.

those from IrT tasks, are available. As mentioned in the Introduction, it is crucial to find a method that does not rely on the IrT transfer.

5.1.2.2 Disentangling Representation in UDA

The concept of disentangled representations has been utilized in unsupervised domain adaptation to decompose features into domain-invariant and domain-specific components [23]. Recent works have employed adversarial learning [343], variational autoencoders (VAE) [56], and graph neural networks (GNNs) [194, 193] to learn disentangled representations.

In addressing UDA problems using GNN and disentangled learning techniques, Mahapatra [194] proposed a UDA method for medical image classification that leverages feature disentanglement and graph convolutional networks (GCN). Similarly, GCAN [193] introduced a graph convolutional adversarial network for unsupervised domain adaptation, which transfers features on graphs and aligns and disentangles source and target domain features through adversarial learning. However, these methods are not directly applicable to zero-shot domain adaptation because they require access to target domain data during training. In contrast, our approach focuses on disentangling task-specific and domain-specific features without relying on any target domain data during the training phase, making it more suitable for ZSDA tasks.

5.1.3 Contributions

Hence, in this chapter, we propose a novel language-driven feature disentanglement paradigm for medical image segmentation ZSDA task, named LMaD (see Figure 5.1 (b)). Unlike existing ZSDA methods, our paradigm no longer relies on IrT during training to pre-learn inter-domain correlations. Instead, we use feature disentanglement learning to retain task-specific information while removing domain-specific information from the source domain. LMaD consists of a graph learning mechanism and a feature disentanglement module (FDM). In detail, the graph learning mechanism projects visual features into graph representations to capture the inter-relationships. The FDM leverages the pre-trained language model (BERT [57] in our cases) to obtain task-specific and domain-specific features through different text prompts.

These features interact with the visual feature graph to disentangle visual task-specific and domain-specific features, thereby eventually removing the interference caused by domain-specific information. Additionally, a graph convolutional network (GCN) is utilized to reason the remaining task-specific visual graph to infer the contextual information within the images.

In a nutshell, our contributions are summarized as follows:

- We propose a language-driven feature disentanglement paradigm, termed LMaD, which uses a language model as the disentangler for ZSDA tasks in medical image segmentation.
- We introduce a Feature Disentanglement Module (FDM) that uses a language model for feature disentanglement and works with the source domain visual feature graph to remove domain-specific information. Graph reasoning is designed to enhance the inference of contextual information in source images, resulting in improved segmentation performance.
- We evaluate the performance of the proposed algorithm on the joint optic disc (OD) and optic cup (OC) fundus image segmentation, as well as prostate MRI segmentation, demonstrating competitive performance that surpasses state-of-the-art comparison algorithms.

5.2 Methods

5.2.1 Problem Definition

ZSDA focuses on leveraging labeled data from a source domain to enhance the generalization performance of a model in a target domain where no labeled data is available. Formally, the problem formulation can be defined as follows:

Given a source domain dataset with its corresponding annotation denoted as $\mathcal{D}_s = \{(x_i^s, y_i^s) \mid x_i^s \in \mathbb{R}^{H \times W \times 3}, y_i^s \in \{0, 1\}^{H \times W}\}_{i=1}^{N_s}$, where x_i^s represents the i -th image in the source domain, y_i^s represents the corresponding segmentation label for the i -th image, and N_s is

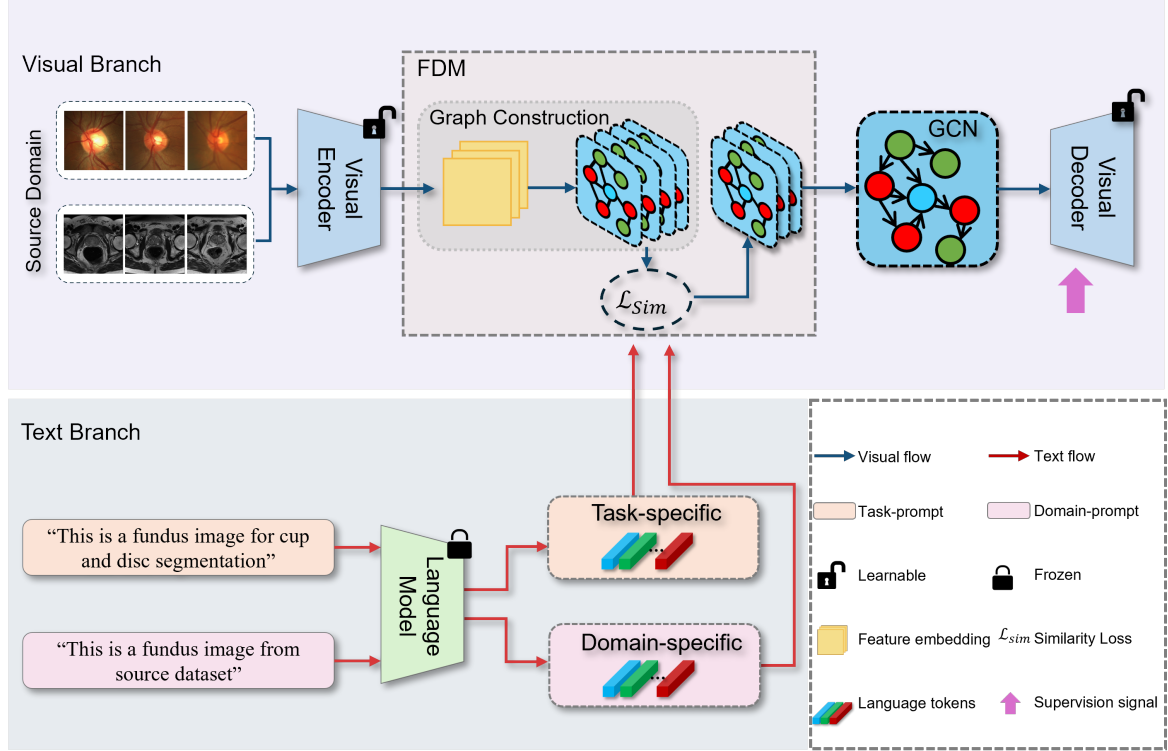


FIGURE 5.2: Overview of our proposed LMaD framework. LMaD includes two branches: a visual branch produces visual graphs and final segmentation results, and a text branch generates domain-specific and task-specific information for feature disentanglement. Best view in color. It includes three major components: visual graph construction, feature disentanglement module, and graph reasoning. The task prompt can be any sentence describing the task, i.e., “This is a fundus image for cup and disc segmentation”, and the domain prompt can be any sentence containing the domain information, i.e., “This is a fundus image from the source dataset”.

the number of labeled samples. We train a segmentation model M using the DeepLabv3+ architecture on the source domain dataset only. During inference, an unseen target test dataset $\mathcal{D}_t = \{x_i^t \mid x_i^t \in \mathbb{R}^{H \times W \times 3}\}$ is given, where x_i^t represents the i -th image in the target domain. Our goal is to adapt the trained model M to this target domain to improve its performance.

5.2.2 Overview

Figure 5.2 illustrates the framework of our LMaD, which includes three major components: visual graph construction, feature disentanglement module, and graph reasoning.

Visual Graph Construction. Given an input image x_i^s , the visual encoder extracts visual feature embedding $f \in \mathbb{R}^{h \times w \times d}$. The feature f is then transformed into the source-dependent semantic graph $G = (V, E)$ using graph projection and the K-means algorithm. In this visual graph, the vertices $V \in \mathbb{R}^{D \times K}$, where $D = \frac{h \times w \times d}{K}$, and the vertices are partitioned into K clusters.

FDM. To extract semantic representations of textual descriptions, we employ a feature extraction pipeline using a pre-trained BERT model. We define two input sentences (task and domain prompts) to distinguish task-specific and domain-specific features. The language model converts these sentences into high-dimensional feature vectors $\mathbf{L}_T$ and $\mathbf{L}_D$, representing the task-specific and domain-specific text embedding, respectively. Then, $\mathbf{L}_T$, $\mathbf{L}_D$ and G are sent to the FDM to disentangle visual task-specific and domain-specific features. We can further obtain the updated $V'_K \in \mathbb{R}^{D \times K}$ with the domain information removed.

Graph Reasoning. In this stage, graph reasoning is conducted to obtain enhanced graph representations $\hat{V}_K$ by graph convolution network [144]. The enhanced $\hat{V}_K$ is then projected back into the original embedding space to get the updated visual feature f' . Finally, f' is fed into the decoder of M to produce the predicted segmentation result.

5.2.3 Graph Construction

In our method, the graph construction process begins with inputting x_i^s into a visual encoder to extract visual feature embedding $f \in \mathbb{R}^{h \times w \times d}$, where f has a spatial resolution of $h \times w$. To remove domain information from f , we construct a graph that organizes its information. There are two common ways to construct graphs: pixel as nodes [116, 163] and K-means as nodes [168, 342]. Since our goal is to use the graph to organize the internal relationships of visual features for subsequent feature disentanglement, instead of using pixel-wise features as nodes, we employ the K-means algorithm to cluster the features. In detail, following [346, 168], we introduce a learnable graph projection $W_{proj} \in \mathbb{R}^{K \times D}$ with each column c_k as the clustering center of the k -th node.

For each feature vector f_i located at coordinates i in the feature map, we calculate its soft assignment to the k -th node using the following formula:

$$Q_k^i = \frac{\exp\left(-\frac{\|f_i - c_k\|^2}{2\sigma^2}\right)}{\sum_{m=1}^K \exp\left(-\frac{\|f_i - c_m\|^2}{2\sigma^2}\right)}, \quad (5.1)$$

where σ is a scaling factor that controls the spread of the clusters, and K is the total number of clusters. The feature representation of each graph node is then derived by aggregating the features weighted by their assignments:

$$V_K = \frac{\sum_i Q_k^i f_i}{\sum_i Q_k^i}. \quad (5.2)$$

Here, V_K is the feature vector of the k -th node, capturing the collective properties of the features assigned to it. We can obtain the visual graph nodes $V \in \mathbb{R}^{D \times K}$, where V consists of K clusters, each of which can also be viewed as a sub-graph. Hence, the visual graph can be also represented as $V = \{V_1, V_2, \dots, V_K\}$. This graph projection process translates high-dimensional image features into a structured graph format, which facilitates subsequent analysis and reasoning tasks by encapsulating spatial and semantic relationships within the image.

5.2.4 Feature Disentanglement Module (FDM)

We use a pre-trained BERT [57] as the language model for feature disentanglement. Specifically, we design two different prompts: task prompt and domain prompt, and send these prompts into the language model separately to obtain the task-specific feature $\mathbf{L}_T \in \mathbb{R}^{m \times 1}$ and domain-specific feature $\mathbf{L}_D \in \mathbb{R}^{m \times 1}$, where m is the feature dimension. The task prompt can be any sentence describing the task, i.e., “This is a fundus image for cup and disc segmentation”, and the domain prompt can be any sentence containing the domain information, i.e., “This is a fundus image from the source dataset”.

To remove domain-specific information while retaining task-specific information, we propose a Feature Disentanglement Module (FDM). As shown in Figure 5.3, the visual graph is divided into K sub-graphs. To eliminate the interference caused by domain information, we

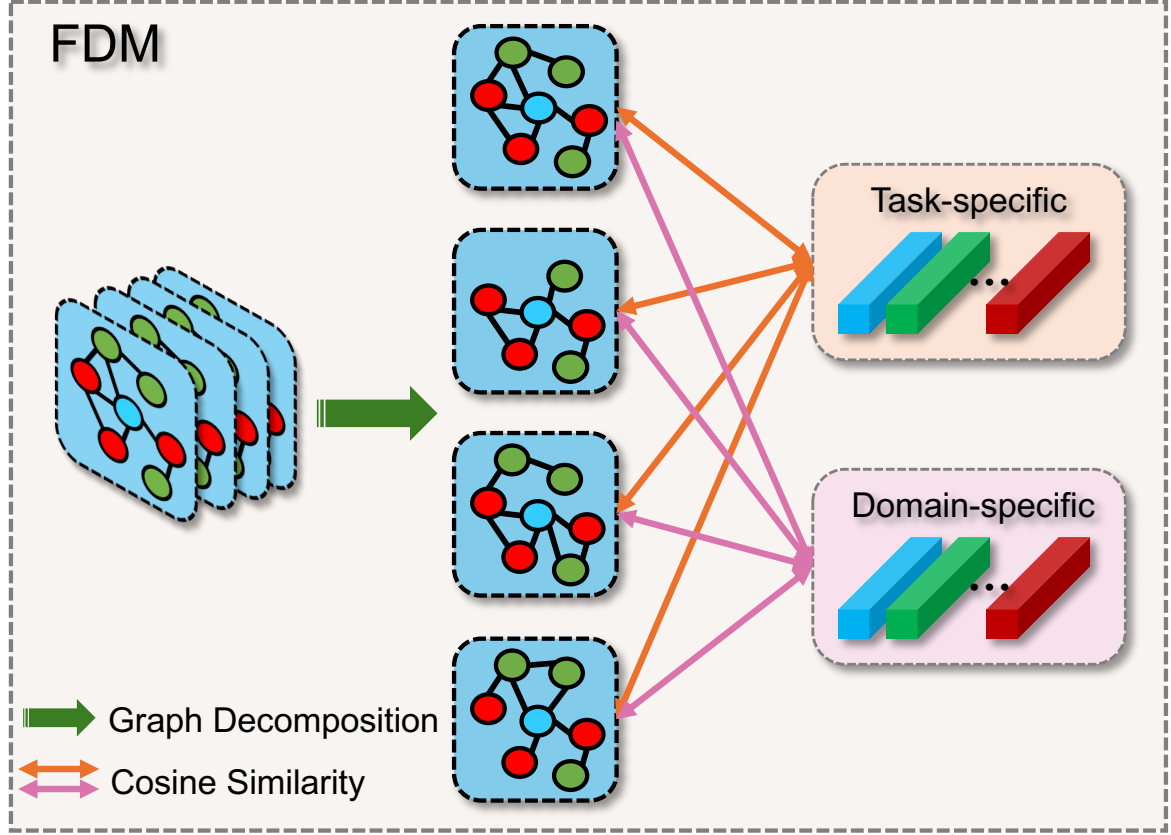


FIGURE 5.3: Proposed Feature Disentanglement Module (FDM). The visual graph is divided into K sub-graphs to measure the similarity with domain-specific and task-specific information, separately.

calculate the cosine similarity between each sub-graph $V_K \in \mathbb{R}^{D \times 1}$ and the two text features: task-specific $\mathbf{L}_T \in \mathbb{R}^{m \times 1}$ and domain-specific $\mathbf{L}_D \in \mathbb{R}^{m \times 1}$. Note that $\mathbf{L}_T$ and $\mathbf{L}_D$ are both passed through a linear layer and resized from $\mathbb{R}^{m \times 1} \rightarrow \mathbb{R}^{D \times 1}$ to ensure that V_K , $\mathbf{L}_T$ and $\mathbf{L}_D$ have the same dimensions. The cosine similarity between V_K and $\mathbf{L}_T$ is computed as:

$$\mathcal{L}_{\cos}^T = 1 - \frac{V_K \cdot \mathbf{L}_T}{\|V_K\| \|\mathbf{L}_T\|}. \quad (5.3)$$

Similarly, the cosine similarity for V_K and $\mathbf{L}_D$ is given by:

$$\mathcal{L}_{\cos}^D = 1 - \frac{V_K \cdot \mathbf{L}_D}{\|V_K\| \|\mathbf{L}_D\|}. \quad (5.4)$$

By performing Equation 5.3 and Equation 5.4, we can obtain two lists of similarity scores: the task-specific similarity scores, represented as $S^T = \{S_1^T, S_2^T, \dots, S_K^T\}$, and the domain-specific similarity scores, represented as $S^D = \{S_1^D, S_2^D, \dots, S_K^D\}$, where S denotes the scores. Since each sub-graph may simultaneously contain both task and domain information, we cannot simply remove those sub-graphs that are strongly correlated with the domain. To better eliminate the interference caused by domain information, we propose a node information update strategy based on conditional judgments.

For the two lists S_K^T and S_K^D , we first introduce the hyperparameter N ($0 < N < K$), which is defined as the number of top N pieces of information considered most relevant, with a total of K results. Hence, two subsets S_N^T and S_N^D of S_K^T and S_K^D can be obtained, with each subset containing their first N values. The hyperparameter λ is used to adjust the control rate of the updated V_K' . We then update V_K using Algorithm 2 to obtain the updated visual graph $V_K' \in \mathbb{R}^{D \times K}$, where domain information is effectively removed while task information is retained.

5.2.5 Graph Reasoning

After getting the updated visual graph V_K' from the FDM, we first compute its adjacency matrix using:

$$A = V_K' \times V_K'^T, \quad (5.5)$$

where $V_K'^T$ denotes the transpose matrix of V_K' .

We then introduce the graph convolution network (GCN) to infer higher-order relationships between task-specific features in the updated visual graph and further enhance the contextual information in the image:

$$\hat{V}_K = \sigma(AV_K'W), \quad (5.6)$$

where $\hat{V}_K$ denotes the enhanced visual graph, $\sigma(\cdot)$ denotes the activation function, and W denotes the learnable parameters.

We then conduct graph re-projection to merge the refined graph features back into the original feature space:

$$f' = \hat{V}_K^T \times Q_k^i, \quad (5.7)$$

where $f' \in \mathbb{R}^{H \times W \times d}$ is the enhanced feature map, which is further fed into the decoder of network M to obtain the final segmentation result $\hat{y}_i^s$.

For optimization, we only use the cross-entropy loss $\mathcal{L}_{ce}$ as the target function for segmentation.

Therefore, the overall loss for training is expressed as:

$$\mathcal{L} = \mathcal{L}_{ce}(y_i^s, \hat{y}_i^s). \quad (5.8)$$

Algorithm 2 Node Information Update

Require: Lists of similarity scores S_K^T and S_K^D , subsets S_N^T and S_N^D of the similarity scores, N, λ

```

1: for each  $K$  do
2:   if  $S_K^T \in S_N^T$  and  $S_K^D \notin S_N^D$  then
3:      $V'_K = V_K \times 1$ 
4:   else if  $S_K^T \notin S_N^T$  and  $S_K^D \in S_N^D$  then
5:      $V'_K = V_K \times 0$ 
6:   else if  $S_K^T \in S_N^T$  and  $S_K^D \in S_N^D$  then
7:     if  $S_K^T \geq S_K^D$  then
8:        $V'_K = V_K \times \lambda$ 
9:     else
10:       $V'_K = V_K \times (1 - \lambda)$ 
11:    end if
12:   else if  $S_K^T \notin S_N^T$  and  $S_K^D \notin S_N^D$  then
13:     if  $S_K^T \geq S_K^D$  then
14:        $V'_K = V_K \times 1$ 
15:     else
16:        $V'_K = V_K \times 0$ 
17:     end if
18:   end if
19: end for
Ensure:  $V'_K$ 

```

5.3 Experiments and Results

5.3.1 Dataset Description

5.3.1.1 Fundus Datasets

To evaluate the effectiveness of our proposed method, we conduct experiments on three widely used public OD/OC segmentation datasets: the REFUGE dataset [212], the Drishti-GS dataset [263], and the RIM-ONE-r3 dataset [79]. Detailed statistics for these three datasets are provided in Chapter 4.3.1.

For a more comprehensive evaluation, we evaluate our method on four challenging zero-shot domain adaptation scenarios: RIM-ONE-r3 $\rightarrow$ Drishti-GS, Drishti-GS $\rightarrow$ RIM-ONE-r3, REFUGE $\rightarrow$ Drishti-GS and REFUGE $\rightarrow$ RIM-ONE-r3.

5.3.1.2 Prostate MRI Datasets

We use five multi-site public prostate MRI datasets for comparison experiments. Specifically, RUNMC (site A) and BMC (site B) from NCI-ISBI 2013 dataset are used as two separate source datasets [208]. HCRUDb (site C) from I2CVB dataset [159], UCL (site D) and BIDMC (site E) from PROMISE12 dataset [172] are used as the target datasets. Table 5.1 shows the details of five prostate MRI datasets used for the proposed method.

- NCI-ISBI 2013 Dataset:** The RUNMC dataset is sourced from Radboud University Nijmegen Medical Centre, and the BMC dataset originates from Boston Medical Center. Both datasets are part of the NCI-ISBI 2013 Challenge, designed to promote the development of automated segmentation algorithms for prostate MRIs. These datasets include multi-parametric MRI scans with T2-weighted imaging, which is widely used in clinical practice for prostate cancer diagnosis and management. RUNMC and BMC datasets offer variations in imaging quality, scanner models, and acquisition protocols, providing a comprehensive challenge for algorithm robustness and generalization across different medical imaging setups.

TABLE 5.1: Statistics of five prostate MRI datasets used for the proposed method.

Dataset	Institution	Num of samples	Field Strength	Endorectal Coil	Camera
Site A	RUNMC	30	3 T	Surface	Siemens
Site B	BMC	30	1.5 T	Endorectal	Philips
Site C	HCRUDB	19	3 T	No	Siemens
Site D	UCL	13	1.5 and 3 T	No	Siemens
Site E	BIDMC	12	3 T	Endorectal	GE

- **I2CVB Dataset:** The HCRUDB dataset is derived from the I2CVB dataset, curated by the Initiative for Collaborative Computer Vision Benchmarking. This dataset includes high-resolution T2-weighted MRI scans, providing a detailed anatomical view of the prostate gland. HCRUDB is notable for its rigorous annotation standards, which involve expert radiologists to ensure high accuracy in the delineation of prostate boundaries. The dataset is particularly valuable for research into the segmentation of the prostate in complex cases where the gland’s appearance may be obscured by surrounding anatomical structures or pathology.
- **PROMISE12 Dataset:** The UCL dataset comes from University College London, while the BIDMC dataset is provided by Beth Israel Deaconess Medical Center. Both datasets are part of the PROMISE12 challenge, a benchmark competition aimed at advancing the field of prostate MRI segmentation. These datasets contain T2-weighted MRIs and are characterized by their inclusion of diverse patient populations and imaging conditions, which are crucial for developing algorithms that can perform reliably across different clinical environments. The PROMISE12 dataset emphasizes the importance of model validation in multi-center studies, where variability in image acquisition and patient demographics can significantly impact the performance of segmentation algorithms.

Detailed statistics for these three datasets are followed [181]¹. We set six zero-shot domain adaptation scenarios: $A \rightarrow \{C, D, E\}$ and $B \rightarrow \{C, D, E\}$.

¹<https://liuquande.github.io/SAML/>

5.3.2 Experimental Setup

5.3.2.1 Evaluation Metric

For the ZSDA OD/OC segmentation task, we use the Dice similarity coefficient (Dice) and the average symmetric surface distance (ASD) as evaluation metrics for segmentation performance. Dice measures the overlap between predicted and ground-truth masks, effectively indicating segmentation accuracy. ASD assesses the average distance between the surfaces of the masks, capturing boundary precision. A higher Dice value denotes better performance, while a lower ASD value is preferable. For convenience, the Dice and ASD values of OD/OC are expressed as DI_{disc} , DI_{cup} , ASD_{disc} , and ASD_{cup} , respectively.

For the ZSDA prostate MRI segmentation task, we only use the Dice similarity coefficient (Dice) as the evaluation metric.

5.3.2.2 Baseline Methods

To comprehensively evaluate our proposed method, we compared it with several baselines, including:

- Following the settings of previous domain adaptation methods [310], we set the benchmark of our model to “upper-bound”. The “upper-bound” setting involves fully supervised model training on the annotated target data.
- The proposed method is compared with five state-of-the-art methods for the zero-shot domain adaptation task, *i.e.*, ZDDA [219], CoCoGAN [170], ZSDA [302], PØDA [66], and Liu *et al.* [186].

5.3.2.3 Implementation Details

For the segmentation network, we use DeepLabv3+ [34] with a pre-trained MobileNetV2 [244] as the backbone. The pre-trained BERT [57] is used for feature embedding generation in the language model. All experiments are conducted on a single GeForce RTX 3090Ti GPU with 24GB memory. For training, the Adam optimizer is used with a “poly” learning rate

TABLE 5.2: Comparison with state-of-the-art methods on the OD/OC Segmentation task when adapted from RIM-ONE-r3 to Drishti-GS.

Methods	RIM-ONE-r3 $\rightarrow$ Drishti-GS			
	$DI_{disc} \uparrow$	$DI_{cup} \uparrow$	$ASD_{disc} \downarrow$	$ASD_{cup} \downarrow$
Upper-bound	97.7	91.0	-	-
ZDDA [219]	82.6	58.0	8.91	31.4
CoCoGAN [170]	87.1	67.2	7.49	11.45
ZSDA [302]	87.1	79.5	7.81	7.56
PØDA [66]	89.1	80.1	6.41	7.18
Liu <i>et al.</i> [186]	92.5	83.6	5.59	5.84
LMaD (Ours)	93.6	84.3	3.91	5.46

TABLE 5.3: Comparison with state-of-the-art methods on the OD/OC Segmentation task when adapted from Drishti-GS to RIM-ONE-r3.

Methods	Drishti-GS $\rightarrow$ RIM-ONE-r3			
	$DI_{disc} \uparrow$	$DI_{cup} \uparrow$	$ASD_{disc} \downarrow$	$ASD_{cup} \downarrow$
Upper-bound	96.7	87.7	-	-
ZDDA [219]	86.7	68.6	6.17	7.32
CoCoGAN [170]	88.8	66.1	5.20	5.22
ZSDA [302]	89.9	72.0	6.67	4.99
PØDA [66]	91.9	73.3	5.49	5.32
Liu <i>et al.</i> [186]	92.3	79.6	5.48	3.89
LMaD (Ours)	94.5	80.9	4.72	7.19

policy. The initial learning rate is set to 0.001, and the model is trained for 200 epochs with the batch size of 8. For the hyperparameters, we set the number of nodes to 8, N to 4, and λ to 0.8. A detailed analysis and ablation studies of these hyperparameters can be found in the Supplementary Materials.

For data pre-processing, we set different strategies for different tasks:

- **Fundus datasets:** Given the varying image sizes across different fundus image datasets, we first crop the original images to obtain a 512×512 Region of Interest (ROI) centered on the OD. For computational efficiency, we then resize the cropped

TABLE 5.4: Comparison with state-of-the-art methods on the OD/OC Segmentation task When adapted from REFUGE to Drishti-GS.

Methods	REFUGE $\rightarrow$ Drishti-GS			
	$DI_{disc} \uparrow$	$DI_{cup} \uparrow$	$ASD_{disc} \downarrow$	$ASD_{cup} \downarrow$
Upper-bound	97.7	91.0	-	-
ZDDA [219]	86.6	76.8	7.99	8.51
CoCoGAN [170]	92.5	81.6	3.84	4.84
ZSDA [302]	93.6	86.0	5.37	8.48
PØDA [66]	93.0	85.3	4.03	5.67
Liu <i>et al.</i> [186]	94.8	84.6	3.49	5.18
LMaD (Ours)	96.6	87.0	2.04	4.18

images to 256×256 for network input. To enhance the sample size, we apply standard data augmentation techniques, including Gaussian noise injection, elastic transformation, random erasing, and random scaling and cropping.

- **Prostate MRI datasets:** We follow the same data pre-processing setting as [129]. Following the approach of [181], we augment the images by applying random rotations with angles of 0, 90, 180, or 270 degrees, and by flipping them either horizontally or vertically.

5.3.3 Comparisons with State-of-the-art Methods

Table 5.2 and Table 5.3 show the segmentation performance on the two domain adaptive tasks: RIM-ONE-r3 $\rightarrow$ Drishti-GS and Drishti-GS $\rightarrow$ RIM-ONE-r3. In the RIM-ONE-r3 $\rightarrow$ Drishti-GS task, LMaD outperforms all baseline approaches under four evaluation metrics. We observe that the Dice values of our method outperform Liu *et al.* by a margin of 1.1% on disc segmentation, with a notable improvement of 1.68 for ASD_{disc} . In the Drishti-GS $\rightarrow$ RIM-ONE-r3 adaptation task, LMaD achieves 2.2% and 1.3% performance improvements on DI_{disc} and DI_{cup} , respectively, compared to the respective state-of-the-art techniques.

Table 5.4 and Table 5.5 present the experimental results on REFUGE $\rightarrow$ Drishti-GS and REFUGE $\rightarrow$ RIM-ONE-r3 adaptation tasks. It is worth noting that our method significantly

TABLE 5.5: Comparison with state-of-the-art methods on the OD/OC Segmentation task When adapted from REFUGE to RIM-ONE-r3.

Methods	REFUGE $\rightarrow$ RIM-ONE-r3			
	$DI_{disc} \uparrow$	$DI_{cup} \uparrow$	$ASD_{disc} \downarrow$	$ASD_{cup} \downarrow$
Upper-bound	96.9	87.7	-	-
ZDDA [219]	73.0	55.0	13.69	7.39
CoCoGAN [170]	87.6	60.4	5.77	7.94
ZSDA [302]	87.0	67.2	7.49	11.45
PØDA [66]	88.4	77.3	6.81	8.56
Liu <i>et al.</i> [186]	89.1	76.0	6.41	8.18
LMaD (Ours)	90.4	78.6	4.56	7.78

TABLE 5.6: Comparison with state-of-the-art methods on the ZSDA prostate MRI segmentation tasks (Site A as source domain).

Methods	Prostate MRI Segmentation (Dice %)		
	Site A $\rightarrow$ Site C	Site A $\rightarrow$ Site D	Site A $\rightarrow$ Site E
Upper-bound	90.8	80.4	91.8
ZDDA [219]	77.2	69.5	78.8
CoCoGAN [170]	79.7	71.2	81.1
ZSDA [302]	82.6	73.2	83.8
PØDA [66]	83.3	74.8	85.9
Liu <i>et al.</i> [186]	86.1	76.7	87.7
LMaD (Ours)	88.2	78.8	89.7

outperforms other irrelevant task-based ZSDA methods. In the REFUGE $\rightarrow$ Drishti-GS task, it is important to observe that our method is able to obtain segmentation results in DI_{disc} that is closer to the upper-bound compared to other comparison algorithms, with improvements of 2.4% in DI_{cup} . In the REFUGE $\rightarrow$ RIM-ONE-r3 task, our method achieves the best scores among all methods, with Dice being 1.3% and 2.6% improvement on disc and cup segmentation, respectively. In addition, LMaD demonstrates a 1.85 performance improvement on ASD_{disc} .

We also report the segmentation results on the prostate MRI segmentation task in Table 5.6 and Table 5.7. As shown in the tables, our proposed LMaD yields more accurate segmentation

TABLE 5.7: Comparison with state-of-the-art methods on the ZSDA prostate MRI segmentation tasks (Site B as source domain).

Methods	Prostate MRI Segmentation (Dice %)		
	Site B $\rightarrow$ Site C	Site B $\rightarrow$ Site D	Site B $\rightarrow$ Site E
Upper-bound	90.8	80.4	91.8
ZDDA [219]	78.3	69.5	77.0
CoCoGAN [170]	81.4	71.4	83.0
ZSDA [302]	82.6	72.6	84.3
P $\emptyset$ DA [66]	84.4	74.4	86.5
Liu <i>et al.</i> [186]	86.8	76.5	87.0
LMaD (Ours)	87.9	78.7	89.9

TABLE 5.8: Ablation study of key components on REFUGE $\rightarrow$ Drishti-GS adaptation task.

Methods	REFUGE $\rightarrow$ Drishti-GS			
	$DI_{disc} \uparrow$	$DI_{cup} \uparrow$	$ASD_{disc} \downarrow$	$ASD_{cup} \downarrow$
w/o FDM	93.0	85.4	8.72	10.37
w/o GCN	95.2	84.7	2.78	5.15
LMaD (Ours)	96.6	87.0	2.04	4.18

results over existing SOTA methods (i.e., 86.1% $\rightarrow$ 88.2% for $A \rightarrow C$ and 76.5% $\rightarrow$ 78.7% for $B \rightarrow D$).

5.3.4 Ablation Studies

5.3.4.1 Component Effectiveness

We study the effectiveness of two key components (FDM and GCN) in our method, with results shown in Table 5.8. To evaluate the contribution of each component, we conduct the ablation study on REFUGE $\rightarrow$ Drishti-GS adaptation task. “w/o FDM” denotes directly using the language model and visual feature for feature disentanglement without employing the graph. As shown in Table 5.8, we can observe that without using the graph, the results for all metrics are worse than our LMaD. Specifically, DI_{disc} is 3.6% lower than our method.

TABLE 5.9: Ablation study for graph node number on REFUGE $\rightarrow$ Drishti-GS adaptation task.

Nodes number	$DI_{disc} \uparrow$	$DI_{cup} \uparrow$	$ASD_{disc} \downarrow$	$ASD_{cup} \downarrow$
6	94.2	81.1	3.35	6.27
10	96.0	84.1	2.3	5.4
Ours (8)	96.6	87.0	2.04	4.18

ASD_{disc} and ASD_{cup} show significant differences from LMaD, with discrepancies of 6.68 and 6.19, respectively. This highlights the importance of constructing a graph to organize visual features. “w/o GCN” denotes the updated visual graph V'_K is directly sent to the visual decoder instead of using a GCN for reasoning. It can be seen that without the GCN to enhance image context features, the performance is lower than our method, with a 2.3% decrease in DI_{cup} . This denotes the effectiveness of graph reasoning. It can be concluded that all modules positively contributed to our method, with the complete framework achieving the best overall performance.

5.3.4.2 Hyperparameter Analysis

Using the REFUGE $\rightarrow$ Drishti-GS on fundus image segmentation task as an example, we investigated the performance of LMaD under various parameter settings. First, we examined the sensitivity of the number of nodes during graph construction.

Table 5.9 shows that the model’s performance improves as the number of nodes increases from 6 to 8. However, increasing the number of nodes to 10 leads to a decline in performance, indicating an optimal range for node quantity.

Additionally, Table 5.10 illustrates the impact of selecting different values for N on the results. We can observe that when N is set to 4, the model’s performance significantly improves, suggesting that this selection enhances the model’s ability to focus on the most relevant features. The control rate λ reflects the effect of different update weights on removing domain-specific information.

TABLE 5.10: Ablation study for N on REFUGE $\rightarrow$ Drishti-GS adaptation task.

N	$DI_{disc} \uparrow$	$DI_{cup} \uparrow$	$ASD_{disc} \downarrow$	$ASD_{cup} \downarrow$
2	91.3	79.0	5.67	7.59
3	92.0	86.0	5.21	4.81
Ours (4)	96.6	87.0	2.04	4.18

TABLE 5.11: Ablation study for λ on REFUGE $\rightarrow$ Drishti-GS adaptation task.

λ	$DI_{disc} \uparrow$	$DI_{cup} \uparrow$	$ASD_{disc} \downarrow$	$ASD_{cup} \downarrow$
0.9	95.4	82.7	2.55	5.77
0.7	94.8	86.0	3.66	4.60
Ours (0.8)	96.6	87.0	2.04	4.18

As shown in Table 5.11, setting λ too low (0.7) or too high (0.9) adversely affects the model’s performance, indicating that a balanced setting is crucial for optimal results. This analysis underscores the importance of carefully tuning these parameters to maximize the effectiveness of our approach.

5.3.5 Visualization Results

Figure 5.4 shows the visualization results of REFUGE $\rightarrow$ Drishti-GS (fundus image) and A $\rightarrow$ C (prostate MRI image) adaptation tasks. From these visualizations, it is clear that our method outperforms other approaches, demonstrating superior adaptation performance in both tasks. The results highlight our model’s ability to effectively handle domain shifts in ZSDA task, maintaining higher accuracy compared to existing methods.

5.4 Discussion

The presented LMaD framework marks an important advancement in the field of Zero-Shot Domain Adaptation (ZSDA) for medical image segmentation, offering an innovative approach with a novel representation of feature disentanglement using pre-trained language models. The experimental results of LMaD have been impressive, demonstrating its capability to

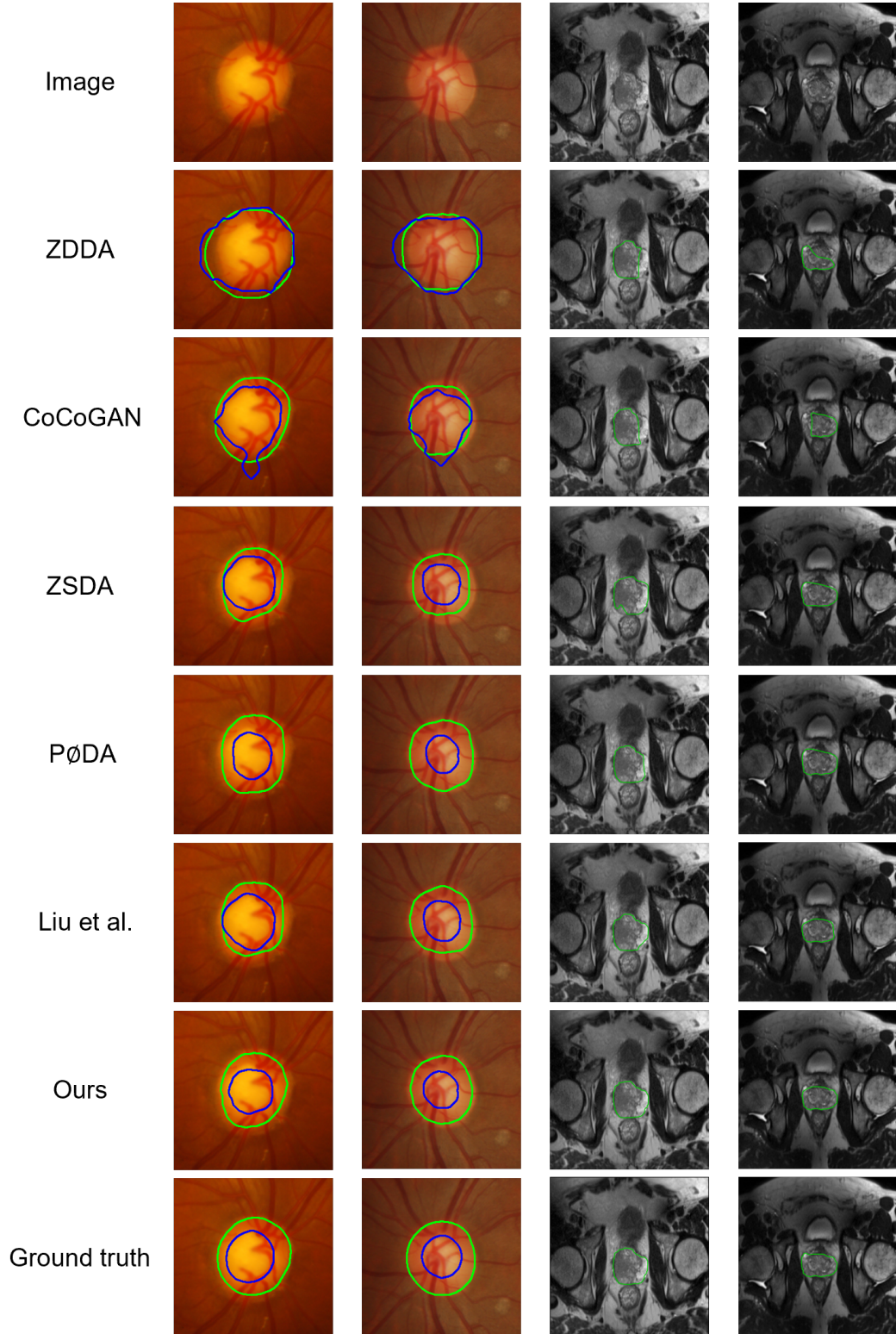


FIGURE 5.4: Visualization of segmentation results for different comparison algorithms (REFUGE $\rightarrow$ Drishti-GS and A $\rightarrow$ C). In fundus images, green for OD segmentation, and blue for OC segmentation. In MRI task, green for prostate segmentation.

achieve efficient and consistent segmentation across fundus images and prostate MRI datasets. This efficiency is further complemented by the framework’s ability to generalize to unseen domains, making it particularly useful for medical scenarios with limited or unavailable target domain data. One of the key strengths of LMaD lies in its use of a language model to disentangle task-specific and domain-specific features. By disentangling task-specific from domain-specific features, the framework eliminates domain interference, thereby improving segmentation accuracy. This feature can be beneficial in various cross-domain medical imaging tasks, illustrating the framework’s broad applicability in medical image analysis.

5.4.1 LLMs for Medical Domain Adaptation

Building on this work, we expanded LMaD to explore the potential of large language models (LLMs) for medical domain adaptation in a new study. This extended research goes beyond segmentation tasks to examine the application of LLMs in both classification and segmentation within medical computer vision. Specifically, we introduced the use of text prompts as a plug-and-play mechanism, enabling LLMs to adapt to various medical imaging tasks without requiring any target domain training. Leveraging large language models [57, 161] may provide a deeper understanding of textual information, allowing the extraction of more relevant details. These LLMs could also offer enhanced comprehension of the differences between images from various domains, making a more complex domain adaptation process. By integrating the LLMs with our existing network, it could improve the overall adaptation process, leading to better generalization across diverse datasets.

This approach also addresses domain generalization, allowing the model to perform effectively across different domains without the need for labeled target data. With the ability to handle different medical image processing tasks, this approach opens new avenues for applying LLMs to medical image analysis, offering a comprehensive solution to domain adaptation challenges in medical data. The use of LLMs holds promise for significantly improving the accuracy and efficiency of medical diagnosis and analysis, positioning it as a promising direction for future advancements in the field.

5.4.2 Limitations and Future Work

While our current work achieved considerable success, there are several areas where improvements can be made in future research to further enhance the model’s performance and adaptability. First, since we incorporated a language model BERT [57], the training time and complexity of the model have increased due to the higher computational requirements and the large number of parameters involved. This added complexity can lead to longer training periods, making the model less efficient in scenarios where computational resources are limited. Future work could focus on optimizing the language model to reduce both the training time and the number of parameters without compromising performance, thereby improving the overall efficiency of the approach.

In addition to enhancing the language model, there is substantial potential to refine the text branch of our approach by replacing the current static text prompts with the dynamic or context-specific prompts. While the current prompts are functional, they could be augmented by incorporating external sources of information that provide detailed descriptions related to the images. This could include domain-specific knowledge or real-time updates to the prompts based on the content/description of the image, making a more flexible and adaptive interaction between the text and image features. Such improvements would allow the model to generate richer and more contextually informed representations, thereby improving performance.

These enhancements would contribute to a more robust and versatile system capable of handling diverse and complex visual scenarios. All of these could be interesting future directions in this area.

5.5 Summary

In this paper, we propose a novel language-driven feature disentanglement paradigm for the OD/OC segmentation ZSDA task. Specifically, we introduce a graph learning mechanism to project visual features into graph representations to capture the inter-relationships. We use a pre-trained language model to obtain task-specific and domain-specific feature embeddings

and perform feature disentanglement with the visual graph to eliminate domain interference. Additionally, a graph convolutional network infers higher-order relationships among task-specific features and enhances contextual information in the images. Extensive experiments on joint segmentation of OD/OC and prostate MRI segmentation demonstrate that our approach effectively enhances ZSDA model performance.

Conclusions and Future Directions

6.1 Conclusions

In this thesis, we explore novel domain adaptive deep learning frameworks for medical image segmentation under data-efficient scenarios. Our goal is to develop methods that tackle the challenge of limited annotated data, which is a common challenge in deep learning for medical imaging. This comprehensive work consists of three studies, each addressing different tasks in data-efficient settings, and collectively contributing to advancements in domain adaptation for medical image segmentation. In the first study, we introduce CAPL-Net, an unsupervised domain adaptation framework for nuclei segmentation and classification, which reduces domain bias through a category-aware feature alignment module and a pseudo-labeling network. The second study (DDF-UDA) proposes a dynamic diffusion-based framework that enhances cross-domain segmentation performance for fundus images. Our third study, LMaD, focuses on zero-shot domain adaptation for medical image segmentation, leveraging language models to disentangle domain information and improve model generalization when the target data is unseen during training.

More specifically, in the first study (Chapter 3), we introduce a novel deep network called the Category-Aware Feature Alignment and Pseudo-Labeling Network (CAPL-Net), specifically designed for unsupervised domain adaptation (UDA) in nuclei instance segmentation and classification. CAPL-Net is motivated by two key factors. First, nuclei in histopathology images belong to different categories and vary in quantity, with distinct features across these categories. To address this, we propose a category-level feature alignment module with dynamically learnable balancing weights, enabling balanced alignment between shared

domain features and category-specific features. Additionally, a nuclei-level prototype pseudo-labeling mechanism further enhances the generalization capability of the model. Extensive experiments on cross-domain nuclei segmentation and classification tasks demonstrate state-of-the-art performance, validating the effectiveness of the proposed approach.

Most previous UDA methods rely on adversarial learning, which heavily depends on the balance of the source and target datasets to align features. In Chapter 4, we develop DDF-UDA, a dynamic diffusion-based framework to address this issue. Instead of using adversarial learning, DDF-UDA introduces an adaptive module based on feature- and pixel-level diffusion processes to reduce cross-domain discrepancies. To adjust the domain information of the source images while preserving their inherent content, we propose a Nash equilibrium adjustment strategy that dynamically modifies the diffusion steps. Experimental results for OD/OC segmentation demonstrate that the proposed framework enhances prediction performance and outperforms traditional adversarial learning-based UDA methods.

In Chapter 5, we present the third contribution, a framework called Language Model as a Disentangler (LMaD), specifically designed for zero-shot domain adaptation (ZSDA) in medical image segmentation tasks. This framework leverages language models to disentangle domain features, addressing the challenges of domain information transfer without requiring access to target domain data during training. We first introduce a graph learning mechanism to construct a visual graph that captures the relationships between visual features. The novelty of this work lies in the use of a pre-trained language model to obtain task-specific and domain-specific feature embeddings. By disentangling the features in the visual graph, we aim to eliminate interference from domain-specific information. Additionally, a graph convolutional network is introduced to infer higher-order relationships between task-specific features in the visual graph, further enhancing contextual information within the images. Extensive experiments on retinal fundus image segmentation and prostate MRI segmentation validate the effectiveness of our LMaD framework. The results show that our approach significantly improves the generalization ability of ZSDA models, achieving SOTA performance in these tasks.

6.2 Future Directions

In this thesis, while our methods successfully address three key challenges for medical image segmentation under data-efficient scenarios, there are still potential research directions that warrant further exploration. For future research, we propose several potential directions:

First, test-time domain adaptation for medical segmentation would be an intriguing area of future research. Currently, most domain adaptation techniques require retraining models when new target domain data is introduced [28, 297, 15]. Test-time adaptation focuses on enabling models to adjust dynamically during inference using only the data encountered at test time [291, 30, 308]. This approach could significantly improve model adaptability and performance, especially in clinical environments where imaging data is constantly evolving. Investigating this direction could lead to more efficient and flexible solutions that better meet the demands of real-world medical applications [136].

Multi-modality domain adaptation, such as combining medical images with large language models (LLMs) or voice data, has the potential to greatly enhance the versatility of segmentation models. Our recent work, LMaD, explored domain adaptation by incorporating language models, but we limited our experiments to using only BERT [57] as the language model for extracting textual features. Future research could investigate the use of larger and more advanced language models to further improve performance. Leveraging cutting-edge LLMs [57, 161] could offer a deeper understanding of textual information, enabling the extraction of more relevant details. Additionally, these LLMs could improve the model's ability to comprehend differences between images from diverse domains, further strengthening domain adaptation and contributing to more robust medical image segmentation solutions. This direction aims to leverage complementary information from different modalities, providing more comprehensive and robust solutions for complex medical imaging tasks.

These potential directions outline promising avenues for future research to further advance domain adaptation in medical image segmentation, addressing ongoing challenges and expanding the applicability of deep learning techniques in data-efficient scenarios.

Bibliography

- [1] Ossama Abdel-Hamid et al. ‘Convolutional neural networks for speech recognition’. In: *IEEE/ACM Transactions on audio, speech, and language processing* 22.10 (2014), pp. 1533–1545.
- [2] Sidra Aleem et al. ‘Test-Time Adaptation with SaLIP: A Cascade of SAM and CLIP for Zero-shot Medical Image Segmentation’. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 2024, pp. 5184–5193.
- [3] Ahmed Almazroa et al. ‘Optic disc and optic cup segmentation methodologies for glaucoma image detection: a survey’. In: *Journal of ophthalmology* 2015.1 (2015), p. 180972.
- [4] Md Zahangir Alom et al. ‘Nuclei segmentation with recurrent residual convolutional neural networks based U-Net (R2U-Net)’. In: *NAECON 2018-IEEE National Aerospace and Electronics Conference*. IEEE. 2018, pp. 228–233.
- [5] Dario Amodei et al. ‘Deep speech 2: End-to-end speech recognition in english and mandarin’. In: *International conference on machine learning*. PMLR. 2016, pp. 173–182.
- [6] Yuxiang An, Dongnan Liu and Weidong Cai. ‘Unsupervised Domain Adaptation for Neuron Membrane Segmentation based on Structural Features’. In: *2023 IEEE International Conference on Multimedia and Expo (ICME)*. IEEE. 2023, pp. 948–953.
- [7] Saeid Asgari Taghanaki et al. ‘Deep semantic segmentation of natural and medical images: a review’. In: *Artificial Intelligence Review* 54 (2021), pp. 137–178.
- [8] Donghyeon Baek, Youngmin Oh and Bumsub Ham. ‘Exploiting a joint embedding space for generalized zero-shot semantic segmentation’. In: *Proceedings of the IEEE/CVF International Conference on Computer Vision*. 2021, pp. 9536–9545.

- [9] Dzmitry Bahdanau. ‘Neural machine translation by jointly learning to align and translate’. In: *arXiv preprint arXiv:1409.0473* (2014).
- [10] Zhongxin Bai and Xiao-Lei Zhang. ‘Speaker recognition based on deep learning: An overview’. In: *Neural Networks* 140 (2021), pp. 65–99.
- [11] Spyridon Bakas et al. ‘Advancing the cancer genome atlas glioma MRI collections with expert segmentation labels and radiomic features’. In: *Scientific data* 4.1 (2017), pp. 1–13.
- [12] Spyridon Bakas et al. ‘Identifying the best machine learning algorithms for brain tumor segmentation, progression assessment, and overall survival prediction in the BRATS challenge’. In: *arXiv preprint arXiv:1811.02629* (2018).
- [13] Tadas Baltrušaitis, Chaitanya Ahuja and Louis-Philippe Morency. ‘Multimodal machine learning: A survey and taxonomy’. In: *IEEE transactions on pattern analysis and machine intelligence* 41.2 (2018), pp. 423–443.
- [14] Michelle Bardis et al. ‘Segmentation of the prostate transition zone and peripheral zone on MR images with deep learning’. In: *Radiology: Imaging Cancer* 3.3 (2021), e200024.
- [15] Mathilde Bateson et al. ‘Constrained domain adaptation for image segmentation’. In: *IEEE Transactions on Medical Imaging* 40.7 (2021), pp. 1875–1887.
- [16] Shai Ben-David et al. ‘A theory of learning from different domains’. In: *Machine learning* 79 (2010), pp. 151–175.
- [17] Cheng Bian et al. ‘Domain adaptation meets zero-shot learning: an annotation-efficient approach to multi-modality medical image segmentation’. In: *IEEE Transactions on Medical Imaging* 41.5 (2021), pp. 1043–1056.
- [18] Patrick Bilic et al. ‘The liver tumor segmentation benchmark (lits)’. In: *Medical Image Analysis* 84 (2023), p. 102680.
- [19] Konstantinos Bousmalis et al. ‘Unsupervised pixel-level domain adaptation with generative adversarial networks’. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2017, pp. 3722–3731.
- [20] Maxime Bucher et al. ‘Zero-shot semantic segmentation’. In: *Advances in Neural Information Processing Systems* 32 (2019).

- [21] Victor Ion Butoi et al. ‘Universeg: Universal medical image segmentation’. In: *Proceedings of the IEEE/CVF International Conference on Computer Vision*. 2023, pp. 21438–21451.
- [22] Lei Cai, Jingyang Gao and Di Zhao. ‘A review of the application of deep learning in medical image classification and segmentation’. In: *Annals of translational medicine* 8.11 (2020).
- [23] Ruichu Cai et al. ‘Learning disentangled semantic representation for domain adaptation’. In: *IJCAI: proceedings of the conference*. Vol. 2019. NIH Public Access. 2019, p. 2060.
- [24] Nicolas Carion et al. ‘End-to-end object detection with transformers’. In: *European conference on computer vision*. Springer. 2020, pp. 213–229.
- [25] Mathilde Caron et al. ‘Deep clustering for unsupervised learning of visual features’. In: *Proceedings of the European conference on computer vision (ECCV)*. 2018, pp. 132–149.
- [26] Heang-Ping Chan et al. ‘Deep learning in medical image analysis’. In: *Deep learning in medical image analysis: challenges and applications* (2020), pp. 3–21.
- [27] Woong-Gi Chang et al. ‘Domain-specific batch normalization for unsupervised domain adaptation’. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 2019, pp. 7354–7362.
- [28] Cheng Chen et al. ‘Synergistic image and feature adaptation: Towards cross-modality domain adaptation for medical image segmentation’. In: *Proceedings of the AAAI conference on artificial intelligence*. Vol. 33. 01. 2019, pp. 865–872.
- [29] Cheng Chen et al. ‘Unsupervised bidirectional cross-modality adaptation via deeply synergistic image and feature alignment for medical image segmentation’. In: *IEEE transactions on medical imaging* 39.7 (2020), pp. 2494–2505.
- [30] Dian Chen et al. ‘Contrastive test-time adaptation’. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 2022, pp. 295–305.
- [31] Hao Chen et al. ‘DCAN: Deep contour-aware networks for object instance segmentation from histology images’. In: *Medical image analysis* 36 (2017), pp. 135–146.

- [32] Hao Chen et al. ‘Iterative multi-domain regularized deep learning for anatomical structure detection and segmentation from ultrasound images’. In: *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2016: 19th International Conference, Athens, Greece, October 17-21, 2016, Proceedings, Part II* 19. Springer. 2016, pp. 487–495.
- [33] Liang-Chieh Chen et al. ‘Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs’. In: *IEEE transactions on pattern analysis and machine intelligence* 40.4 (2017), pp. 834–848.
- [34] Liang-Chieh Chen et al. ‘Encoder-decoder with atrous separable convolution for semantic image segmentation’. In: *Proceedings of the European Conference on Computer Vision*. 2018, pp. 801–818.
- [35] Lin Chen et al. ‘Deliberated domain bridging for domain adaptive semantic segmentation’. In: *Advances in Neural Information Processing Systems* 35 (2022), pp. 15105–15118.
- [36] Long Chen, Martin Strauch and Dorit Merhof. ‘Instance segmentation of biomedical images with an object-aware embedding learned with local constraints’. In: *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer. 2019, pp. 451–459.
- [37] Shengcong Chen, Changxing Ding and Dacheng Tao. ‘Boundary-assisted region proposal networks for nucleus segmentation’. In: *Medical Image Computing and Computer Assisted Intervention–MICCAI 2020: 23rd International Conference, Lima, Peru, October 4–8, 2020, Proceedings, Part V* 23. Springer. 2020, pp. 279–288.
- [38] Shoufa Chen et al. ‘Diffusionnet: Diffusion model for object detection’. In: *Proceedings of the IEEE/CVF International Conference on Computer Vision*. 2023, pp. 19830–19843.
- [39] Xinlei Chen et al. ‘Deconstructing Denoising Diffusion Models for Self-Supervised Learning’. In: *arXiv preprint arXiv:2401.14404* (2024).
- [40] Xinyang Chen et al. ‘Transferability vs. discriminability: Batch spectral penalization for adversarial domain adaptation’. In: *International Conference on Machine Learning*. PMLR. 2019, pp. 1081–1090.

- [41] Xu Chen et al. ‘Learning active contour models for medical image segmentation’. In: *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. 2019, pp. 11632–11640.
- [42] Yuqian Chen et al. ‘CellTrack R-CNN: A novel end-to-end deep neural network for cell segmentation and tracking in microscopy images’. In: *2021 IEEE 18th International Symposium on Biomedical Imaging (ISBI)*. IEEE. 2021, pp. 779–782.
- [43] Yuqian Chen et al. ‘TractGraphCNN: anatomically informed graph CNN for classification using diffusion MRI tractography’. In: *2023 IEEE 20th International Symposium on Biomedical Imaging (ISBI)*. IEEE. 2023, pp. 1–5.
- [44] Ziyang Chen, Yongsheng Pan and Yong Xia. ‘Reconstruction-Driven Dynamic Refinement Based Unsupervised Domain Adaptation for Joint Optic Disc and Cup Segmentation’. In: *IEEE Journal of Biomedical and Health Informatics* (2023).
- [45] Ziyang Chen et al. ‘TriLA: Triple-Level Alignment based Unsupervised Domain Adaptation for Joint Segmentation of Optic Disc and Optic Cup’. In: *IEEE Journal of Biomedical and Health Informatics* (2024).
- [46] Jie-Zhi Cheng et al. ‘Computer-aided diagnosis with deep learning architecture: applications to breast lesions in US images and pulmonary nodules in CT scans’. In: *Scientific reports* 6.1 (2016), p. 24454.
- [47] Jun Cheng et al. ‘Superpixel classification based optic disc and optic cup segmentation for glaucoma screening’. In: *IEEE transactions on medical imaging* 32.6 (2013), pp. 1019–1032.
- [48] Veronika Cheplygina, Marleen De Bruijne and Josien PW Pluim. ‘Not-so-supervised: a survey of semi-supervised, multi-instance, and transfer learning in medical image analysis’. In: *Medical image analysis* 54 (2019), pp. 280–296.
- [49] Phillip Chlap et al. ‘A review of medical image data augmentation techniques for deep learning applications’. In: *Journal of Medical Imaging and Radiation Oncology* 65.5 (2021), pp. 545–563.

- [50] Özgün Çiçek et al. ‘3D U-Net: learning dense volumetric segmentation from sparse annotation’. In: *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2016: 19th International Conference, Athens, Greece, October 17-21, 2016, Proceedings, Part II* 19. Springer. 2016, pp. 424–432.
- [51] Gabriela Csurka, Riccardo Volpi and Boris Chidlovskii. ‘Unsupervised domain adaptation for semantic image segmentation: a comprehensive survey’. In: *arXiv preprint arXiv:2112.03241* (2021).
- [52] Jifeng Dai et al. ‘R-fcn: Object detection via region-based fully convolutional networks’. In: *Advances in neural information processing systems* 29 (2016).
- [53] Wenyan Dai et al. ‘Translated learning: Transfer learning across different feature spaces’. In: *Advances in neural information processing systems* 21 (2008).
- [54] Bob D De Vos et al. ‘ConvNet-based localization of anatomical structures in 3-D medical images’. In: *IEEE transactions on medical imaging* 36.7 (2017), pp. 1470–1481.
- [55] Jia Deng et al. ‘Imagenet: A large-scale hierarchical image database’. In: *2009 IEEE conference on computer vision and pattern recognition*. Ieee. 2009, pp. 248–255.
- [56] Wanxia Deng et al. ‘Informative feature disentanglement for unsupervised domain adaptation’. In: *IEEE Transactions on Multimedia* 24 (2021), pp. 2407–2421.
- [57] Jacob Devlin. ‘Bert: Pre-training of deep bidirectional transformers for language understanding’. In: *arXiv preprint arXiv:1810.04805* (2018).
- [58] Foivos I Diakogiannis et al. ‘ResUNet-a: A deep learning framework for semantic segmentation of remotely sensed data’. In: *ISPRS Journal of Photogrammetry and Remote Sensing* 162 (2020), pp. 94–114.
- [59] Kunio Doi. ‘Computer-aided diagnosis in medical imaging: historical review, current status and future potential’. In: *Computerized medical imaging and graphics* 31.4-5 (2007), pp. 198–211.
- [60] Jeff Donahue et al. ‘Decaf: A deep convolutional activation feature for generic visual recognition’. In: *International conference on machine learning*. PMLR. 2014, pp. 647–655.

- [61] Jiahua Dong et al. ‘What can be transferred: Unsupervised domain adaptation for endoscopic lesions segmentation’. In: *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. 2020, pp. 4023–4032.
- [62] Alexey Dosovitskiy. ‘An image is worth 16x16 words: Transformers for image recognition at scale’. In: *arXiv preprint arXiv:2010.11929* (2020).
- [63] Frank-Jan H Drost et al. ‘Prostate MRI, with or without MRI-targeted biopsy, and systematic biopsy for detecting prostate cancer’. In: *Cochrane Database of Systematic Reviews* 4 (2019).
- [64] Ahmed Elnakib et al. ‘Medical image segmentation: a brief survey’. In: *Multi Modality State-of-the-Art Medical Image Segmentation and Registration Methodologies: Volume II* (2011), pp. 1–39.
- [65] Andre Esteva et al. ‘Deep learning-enabled medical computer vision’. In: *NPJ digital medicine* 4.1 (2021), p. 5.
- [66] Mohammad Fahes et al. ‘Poda: Prompt-driven zero-shot domain adaptation’. In: *Proceedings of the IEEE/CVF International Conference on Computer Vision*. 2023, pp. 18623–18633.
- [67] Jianan Fan et al. ‘Learning to Generalize over Subpartitions for Heterogeneity-Aware Domain Adaptive Nuclei Segmentation’. In: *International Journal of Computer Vision* (2024), pp. 1–24.
- [68] Jianan Fan et al. ‘Revisiting Adaptive Cellular Recognition Under Domain Shifts: A Contextual Correspondence View’. In: *Proceedings of the European Conference on Computer Vision*. 2024.
- [69] Jianan Fan et al. ‘Taxonomy adaptive cross-domain adaptation in medical imaging via optimization trajectory distillation’. In: *Proceedings of the IEEE/CVF International Conference on Computer Vision*. 2023, pp. 21174–21184.
- [70] Abolfazl Farahani et al. ‘A brief review of domain adaptation’. In: *Advances in data science and information engineering: proceedings from ICDATA 2020 and IKE 2020* (2021), pp. 877–894.

- [71] Wei Feng et al. ‘Unsupervised domain adaptation for medical image segmentation by selective entropy constraints and adaptive semantic alignment’. In: *Proceedings of the AAAI Conference on Artificial Intelligence*. Vol. 37. 2023, pp. 623–631.
- [72] Wei Feng et al. ‘Unsupervised domain adaptive fundus image segmentation with category-level regularization’. In: *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer. 2022, pp. 497–506.
- [73] Yuncong Feng et al. ‘A multi-scale 3D Otsu thresholding algorithm for medical image segmentation’. In: *Digital Signal Processing* 60 (2017), pp. 186–199.
- [74] Emilio Fenoy, Alejandro A Edera and Georgina Stegmayer. ‘Transfer learning in proteins: evaluating novel protein learned representations for bioinformatics tasks’. In: *Briefings in Bioinformatics* 23.4 (2022), bbac232.
- [75] Max Ferguson et al. ‘Detection and segmentation of manufacturing defects with convolutional neural networks and transfer learning’. In: *Smart and sustainable manufacturing systems* 2.1 (2018), pp. 137–164.
- [76] Huazhu Fu et al. ‘Joint optic disc and cup segmentation based on multi-label deep network and polar transformation’. In: *IEEE Transactions on Medical Imaging* 37.7 (2018), pp. 1597–1605.
- [77] King-Sun Fu and JK Mui. ‘A survey on image segmentation’. In: *Pattern recognition* 13.1 (1981), pp. 3–16.
- [78] Hiroshi Fujita. ‘AI-based computer-aided diagnosis (AI-CAD): the latest review to read first’. In: *Radiological physics and technology* 13.1 (2020), pp. 6–19.
- [79] Francisco Fumero et al. ‘RIM-ONE: An open retinal image database for optic nerve evaluation’. In: *2011 24th International Symposium on Computer-based Medical Systems (CBMS)*. IEEE. 2011, pp. 1–6.
- [80] Yaroslav Ganin and Victor Lempitsky. ‘Unsupervised domain adaptation by back-propagation’. In: *International conference on machine learning*. PMLR. 2015, pp. 1180–1189.
- [81] Junfeng Gao et al. ‘Computer vision in healthcare applications’. In: *Journal of healthcare engineering* 2018 (2018).

- [82] Stephan J Garbin et al. ‘High resolution zero-shot domain adaptation of synthetically rendered face images’. In: *Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part XXVIII* 16. Springer. 2020, pp. 220–236.
- [83] Fabio Garcea et al. ‘Data augmentation for medical imaging: A systematic literature review’. In: *Computers in Biology and Medicine* 152 (2023), p. 106391.
- [84] Loveleen Gaur et al. ‘Medical image-based detection of COVID-19 using deep convolution neural networks’. In: *Multimedia systems* 29.3 (2023), pp. 1729–1738.
- [85] Mohsen Ghafoorian et al. ‘Transfer learning for domain adaptation in mri: Application in brain lesion segmentation’. In: *Medical Image Computing and Computer Assisted Intervention- MICCAI 2017: 20th International Conference, Quebec City, QC, Canada, September 11-13, 2017, Proceedings, Part III* 20. Springer. 2017, pp. 516–524.
- [86] Muhammad Ghifary et al. ‘Deep reconstruction-classification networks for unsupervised domain adaptation’. In: *Computer Vision–ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, October 11–14, 2016, Proceedings, Part IV* 14. Springer. 2016, pp. 597–613.
- [87] Edoardo Giacomello, Daniele Loiacono and Luca Mainardi. ‘Brain MRI tumor segmentation with adversarial networks’. In: *2020 International Joint Conference on Neural Networks (IJCNN)*. IEEE. 2020, pp. 1–8.
- [88] John M Giorgi and Gary D Bader. ‘Transfer learning for biomedical named entity recognition with neural networks’. In: *Bioinformatics* 34.23 (2018), pp. 4087–4094.
- [89] Ian Goodfellow et al. ‘Generative adversarial nets’. In: *Advances in neural information processing systems* 27 (2014).
- [90] Kasthurirangan Gopalakrishnan et al. ‘Deep convolutional neural networks with transfer learning for computer vision-based data-driven pavement distress detection’. In: *Construction and building materials* 157 (2017), pp. 322–330.
- [91] Simon Graham et al. ‘Hover-net: Simultaneous segmentation and classification of nuclei in multi-tissue histology images’. In: *Medical image analysis* 58 (2019), p. 101563.

- [92] Simon Graham et al. ‘Lizard: A large-scale dataset for colonic nuclear instance segmentation and classification’. In: *Proceedings of the IEEE/CVF international conference on computer vision*. 2021, pp. 684–693.
- [93] Jiuxiang Gu et al. ‘Recent advances in convolutional neural networks’. In: *Pattern recognition* 77 (2018), pp. 354–377.
- [94] Ran Gu et al. ‘CDDSA: Contrastive domain disentanglement and style augmentation for generalizable medical image segmentation’. In: *Medical Image Analysis* 89 (2023), p. 102904.
- [95] Hao Guan and Mingxia Liu. ‘Domain adaptation for medical image analysis: a survey’. In: *IEEE Transactions on Biomedical Engineering* 69.3 (2021), pp. 1173–1185.
- [96] David Gutman et al. ‘Skin lesion analysis toward melanoma detection: A challenge at the international symposium on biomedical imaging (ISBI) 2016, hosted by the international skin imaging collaboration (ISIC)’. In: *arXiv preprint arXiv:1605.01397* (2016).
- [97] Philip Haeusser et al. ‘Associative domain adaptation’. In: *Proceedings of the IEEE international conference on computer vision*. 2017, pp. 2765–2773.
- [98] Yuki Hagiwara et al. ‘Computer-aided diagnosis of glaucoma using fundus images: A review’. In: *Computer methods and programs in biomedicine* 165 (2018), pp. 1–12.
- [99] Xiaoting Han et al. ‘Deep symmetric adaptation network for cross-modality medical image segmentation’. In: *IEEE transactions on medical imaging* 41.1 (2021), pp. 121–132.
- [100] A Hannun. ‘Deep Speech: Scaling up end-to-end speech recognition’. In: *arXiv preprint arXiv:1412.5567* (2014).
- [101] Tomohiro Hayakawa et al. ‘Computational nuclei segmentation methods in digital pathology: a survey’. In: *Archives of Computational Methods in Engineering* 28 (2021), pp. 1–13.
- [102] Kaiming He et al. ‘Deep residual learning for image recognition’. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2016, pp. 770–778.
- [103] Kaiming He et al. ‘Mask r-cnn’. In: *Proceedings of the IEEE international conference on computer vision*. 2017, pp. 2961–2969.

- [104] Zhenqi He et al. ‘Transnuseg: A lightweight multi-task transformer for nuclei segmentation’. In: *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer. 2023, pp. 206–215.
- [105] Nicholas Heller et al. ‘The kits19 challenge data: 300 kidney tumor cases with clinical context, ct semantic segmentations, and surgical outcomes’. In: *arXiv preprint arXiv:1904.00445* (2019).
- [106] Geoffrey Hinton. ‘Distilling the Knowledge in a Neural Network’. In: *arXiv preprint arXiv:1503.02531* (2015).
- [107] Geoffrey E Hinton, Simon Osindero and Yee-Whye Teh. ‘A fast learning algorithm for deep belief nets’. In: *Neural computation* 18.7 (2006), pp. 1527–1554.
- [108] Jonathan Ho, Ajay Jain and Pieter Abbeel. ‘Denoising diffusion probabilistic models’. In: *Advances in Neural Information Processing Systems* 33 (2020), pp. 6840–6851.
- [109] Judy Hoffman et al. ‘Cycada: Cycle-consistent adversarial domain adaptation’. In: *International conference on machine learning*. Pmlr. 2018, pp. 1989–1998.
- [110] Judy Hoffman et al. ‘Fcns in the wild: Pixel-level adversarial and constraint-based adaptation’. In: *arXiv preprint arXiv:1612.02649* (2016).
- [111] Reka Hollandi et al. ‘Nucleus segmentation: towards automated solutions’. In: *Trends in Cell Biology* 32.4 (2022), pp. 295–310.
- [112] Jin Hong, Simon Chun-Ho Yu and Weitian Chen. ‘Unsupervised domain adaptation for cross-modality liver segmentation via joint adversarial learning and self-learning’. In: *Applied Soft Computing* 121 (2022), p. 108729.
- [113] AD Hoover, Valentina Kouznetsova and Michael Goldbaum. ‘Locating blood vessels in retinal images by piecewise threshold probing of a matched filter response’. In: *IEEE Transactions on Medical imaging* 19.3 (2000), pp. 203–210.
- [114] Fabian Hörst et al. ‘Cellvit: Vision transformers for precise cell segmentation and classification’. In: *Medical Image Analysis* 94 (2024), p. 103143.
- [115] Neil Houlsby et al. ‘Parameter-efficient transfer learning for NLP’. In: *International conference on machine learning*. PMLR. 2019, pp. 2790–2799.

- [116] Hanzhe Hu et al. ‘Class-wise dynamic graph convolution for semantic segmentation’. In: *Proceedings of the European Conference on Computer Vision*. Springer. 2020, pp. 1–17.
- [117] Junling Hu and Michael P Wellman. ‘Nash Q-learning for general-sum stochastic games’. In: *Journal of Machine Learning Research* 4.Nov (2003), pp. 1039–1069.
- [118] Shishuai Hu, Zehui Liao and Yong Xia. ‘Domain specific convolution and high frequency reconstruction based unsupervised domain adaptation for medical image segmentation’. In: *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer. 2022, pp. 650–659.
- [119] Jiaying Huang et al. ‘Category contrast for unsupervised domain adaptation in visual tasks’. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 2022, pp. 1203–1214.
- [120] Jiayuan Huang et al. ‘Correcting sample selection bias by unlabeled data’. In: *Advances in neural information processing systems* 19 (2006).
- [121] Jinjie Huang et al. ‘Segmentation of cervical cell images based on generative adversarial networks’. In: *IEEE Access* 9 (2021), pp. 115415–115428.
- [122] Mahbub Hussain, Jordan J Bird and Diego R Faria. ‘A study on CNN transfer learning for image classification’. In: *Advances in Computational Intelligence Systems: Contributions Presented at the 18th UK Workshop on Computational Intelligence, September 5-7, 2018, Nottingham, UK*. Springer. 2019, pp. 191–202.
- [123] Humayun Irshad et al. ‘Methods for nuclei detection, segmentation, and classification in digital histopathology: a review—current status and future potential’. In: *IEEE reviews in biomedical engineering* 7 (2013), pp. 97–114.
- [124] V Durga Prasad Jasti et al. ‘Computational technique based on machine learning and image processing for medical image analysis of breast cancer diagnosis’. In: *Security and communication networks* 2022.1 (2022), p. 1918379.
- [125] Mehran Javanmardi and Tolga Tasdizen. ‘Domain adaptation for biomedical image segmentation using adversarial training’. In: *2018 IEEE 15th International Symposium on Biomedical Imaging*. 2018, pp. 554–558.

- [126] Jiwoong J Jeong et al. ‘Systematic review of generative adversarial networks (GANs) for medical image classification and segmentation’. In: *Journal of Digital Imaging* 35.2 (2022), pp. 137–152.
- [127] Won Young Jhoo and Jae-Pil Heo. ‘Collaborative learning with disentangled features for zero-shot domain adaptation’. In: *Proceedings of the IEEE/CVF international conference on computer vision*. 2021, pp. 8896–8905.
- [128] Haozhe Jia et al. ‘A convolutional neural network with pixel-wise sparse graph reasoning for COVID-19 lesion segmentation in CT images’. In: *Computers in biology and medicine* 155 (2023), p. 106698.
- [129] Meirui Jiang, Zirui Wang and Qi Dou. ‘Harmofl: Harmonizing local and global drifts in federated learning on heterogeneous medical images’. In: *Proceedings of the AAAI Conference on Artificial Intelligence*. Vol. 36. 1. 2022, pp. 1087–1095.
- [130] Yizhang Jiang et al. ‘A novel negative-transfer-resistant fuzzy clustering model with a shared cross-domain transfer latent space and its application to brain CT image segmentation’. In: *IEEE/ACM transactions on computational biology and bioinformatics* 18.1 (2020), pp. 40–52.
- [131] Michael I Jordan and Tom M Mitchell. ‘Machine learning: Trends, perspectives, and prospects’. In: *Science* 349.6245 (2015), pp. 255–260.
- [132] Konstantinos Kamnitsas et al. ‘Unsupervised domain adaptation in brain lesion segmentation with adversarial networks’. In: *Information Processing in Medical Imaging: 25th International Conference, IPMI 2017, Boone, NC, USA, June 25-30, 2017, Proceedings* 25. Springer. 2017, pp. 597–609.
- [133] Cyriac Kandoth et al. ‘Mutational landscape and significance across 12 major cancer types’. In: *Nature* 502.7471 (2013), pp. 333–339.
- [134] Guoliang Kang et al. ‘Contrastive adaptation network for unsupervised domain adaptation’. In: *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. 2019, pp. 4893–4902.
- [135] Neerav Karani et al. ‘A lifelong learning approach to brain MR segmentation across scanners and protocols’. In: *International conference on medical image computing and computer-assisted intervention*. Springer. 2018, pp. 476–484.

- [136] Neerav Karani et al. ‘Test-time adaptable neural networks for robust medical image segmentation’. In: *Medical Image Analysis* 68 (2021), p. 101907.
- [137] S Kavitha, S Karthikeyan and K Duraiswamy. ‘Early detection of glaucoma in retinal images using cup to disc ratio’. In: *2010 Second International conference on Computing, Communication and Networking Technologies*. IEEE. 2010, pp. 1–5.
- [138] Ali Emre Kavur et al. ‘CHAOS-combined (CT-MR) healthy abdominal organ segmentation challenge data’. In: (*No Title*) (2019).
- [139] Noor Elaiza Abdul Khalid, Noorhayati Mohamed Noor and Norharyati Md Ariff. ‘Fuzzy c-means (FCM) for optic cup and disc segmentation with morphological operation’. In: *Procedia Computer Science* 42 (2014), pp. 255–262.
- [140] Zia Khan et al. ‘Evaluation of deep neural networks for semantic segmentation of prostate in T2W MRI’. In: *Sensors* 20.11 (2020), p. 3183.
- [141] Zia Khan et al. ‘Recent automatic segmentation algorithms of MRI prostate regions: a review’. In: *IEEE Access* 9 (2021), pp. 97878–97905.
- [142] Naji Khosravan et al. ‘Pan: Projective adversarial network for medical image segmentation’. In: *Medical Image Computing and Computer Assisted Intervention–MICCAI 2019: 22nd International Conference, Shenzhen, China, October 13–17, 2019, Proceedings, Part VI* 22. Springer. 2019, pp. 68–76.
- [143] Hee E Kim et al. ‘Transfer learning for medical image classification: a literature review’. In: *BMC medical imaging* 22.1 (2022), p. 69.
- [144] Thomas N Kipf and Max Welling. ‘Semi-supervised classification with graph convolutional networks’. In: *arXiv preprint arXiv:1609.02907* (2016).
- [145] Alexander Kirillov et al. ‘Segment anything’. In: *Proceedings of the IEEE/CVF International Conference on Computer Vision*. 2023, pp. 4015–4026.
- [146] Elyor Kodirov et al. ‘Unsupervised domain adaptation for zero-shot learning’. In: *Proceedings of the IEEE International Conference on Computer Vision*. 2015, pp. 2452–2460.
- [147] Elyor Kodirov et al. ‘Unsupervised domain adaptation for zero-shot learning’. In: *Proceedings of the IEEE international conference on computer vision*. 2015, pp. 2452–2460.

- [148] Padmavathi Kora et al. ‘Transfer learning techniques for medical image analysis: A review’. In: *Biocybernetics and Biomedical Engineering* 42.1 (2022), pp. 79–107.
- [149] Sotiris B Kotsiantis, Ioannis Zaharakis, P Pintelas et al. ‘Supervised machine learning: A review of classification techniques’. In: *Emerging artificial intelligence applications in computer engineering* 160.1 (2007), pp. 3–24.
- [150] Wouter M Kouw and Marco Loog. ‘A review of domain adaptation without target labels’. In: *IEEE transactions on pattern analysis and machine intelligence* 43.3 (2019), pp. 766–785.
- [151] M Krithika Alias AnbuDevi and K Suganthi. ‘Review of semantic segmentation of medical images using modified architectures of UNET’. In: *Diagnostics* 12.12 (2022), p. 3064.
- [152] Alex Krizhevsky, Ilya Sutskever and Geoffrey E Hinton. ‘Imagenet classification with deep convolutional neural networks’. In: *Advances in neural information processing systems* 25 (2012).
- [153] Sapna Singh Kshatri and Deepak Singh. ‘Convolutional neural network in medical image analysis: a review’. In: *Archives of Computational Methods in Engineering* 30.4 (2023), pp. 2793–2810.
- [154] Shyam Lal et al. ‘NucleiSegNet: Robust deep learning architecture for the nuclei segmentation of liver cancer histopathology images’. In: *Computers in Biology and Medicine* 128 (2021), p. 104075.
- [155] Jahanzaib Latif et al. ‘Medical imaging using machine learning and deep learning algorithms: a review’. In: *2019 2nd International conference on computing, mathematics and engineering technologies (iCoMET)*. IEEE. 2019, pp. 1–5.
- [156] Yann LeCun, Yoshua Bengio and Geoffrey Hinton. ‘Deep learning’. In: *nature* 521.7553 (2015), pp. 436–444.
- [157] Hansang Lee and Junmo Kim. ‘Segmentation of overlapping cervical cells in microscopic images with superpixel partitioning and cell-wise contour refinement’. In: *Proceedings of the IEEE conference on computer vision and pattern recognition workshops*. 2016, pp. 63–69.

- [158] Haijun Lei et al. ‘Unsupervised domain adaptation based image synthesis and feature alignment for joint optic disc and cup segmentation’. In: *IEEE Journal of Biomedical and Health Informatics* 26.1 (2021), pp. 90–102.
- [159] Guillaume Lemaitre et al. ‘Computer-aided detection and diagnosis for prostate cancer based on mono and multi-parametric MRI: a review’. In: *Computers in Biology and Medicine* 60 (2015), pp. 8–31.
- [160] Canran Li et al. ‘Domain adaptive nuclei instance segmentation and classification via category-aware feature alignment and pseudo-labelling’. In: *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer. 2022, pp. 715–724.
- [161] Chunyuan Li et al. ‘Llava-med: Training a large language-and-vision assistant for biomedicine in one day’. In: *Advances in Neural Information Processing Systems* 36 (2024).
- [162] Ge Li et al. ‘Region focus network for joint optic disc and cup segmentation’. In: *Proceedings of the AAAI Conference on Artificial Intelligence*. Vol. 34. 01. 2020, pp. 751–758.
- [163] Haoran Li et al. ‘Zero-shot camouflaged object detection’. In: *IEEE Transactions on Image Processing* (2023).
- [164] Qing Li et al. ‘Medical image classification with convolutional neural network’. In: *2014 13th international conference on control automation robotics & vision (ICARCV)*. IEEE. 2014, pp. 844–848.
- [165] Xiaomeng Li et al. ‘Transformation-consistent self-ensembling model for semisupervised medical image segmentation’. In: *IEEE transactions on neural networks and learning systems* 32.2 (2020), pp. 523–534.
- [166] Xuhong Li et al. ‘Transfer learning in computer vision tasks: Remember where you come from’. In: *Image and Vision Computing* 93 (2020), p. 103853.
- [167] Yanwei Li et al. ‘Fully convolutional networks for panoptic segmentation’. In: *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. 2021, pp. 214–223.

- [168] Yin Li and Abhinav Gupta. ‘Beyond grids: Learning graph representations for visual recognition’. In: *Advances in Neural Information Processing Systems* 31 (2018).
- [169] Zhiqi Li et al. ‘Panoptic segformer: Delving deeper into panoptic segmentation with transformers’. In: *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. 2022, pp. 1280–1289.
- [170] Chieh Hubert Lin et al. ‘Coco-gan: Generation by parts via conditional coordinating’. In: *Proceedings of the IEEE/CVF International Conference on Computer Vision*. 2019, pp. 4512–4521.
- [171] Geert Litjens et al. ‘A survey on deep learning in medical image analysis’. In: *Medical image analysis* 42 (2017), pp. 60–88.
- [172] Geert Litjens et al. ‘Evaluation of prostate segmentation algorithms for MRI: the PROMISE12 challenge’. In: *Medical image analysis* 18.2 (2014), pp. 359–373.
- [173] Dongnan Liu et al. ‘Nuclei segmentation via a deep panoptic model with semantic feature fusion’. In: *Proceedings of the 28th International Joint Conference on Artificial Intelligence*. 2019, pp. 861–868.
- [174] Dongnan Liu et al. ‘Panoptic feature fusion net: a novel instance segmentation paradigm for biomedical and biological images’. In: *IEEE Transactions on Image Processing* 30 (2021), pp. 2045–2059.
- [175] Dongnan Liu et al. ‘Pdam: A panoptic-level feature alignment framework for unsupervised domain adaptive instance segmentation in microscopy images’. In: *IEEE Transactions on Medical Imaging* 40.1 (2020), pp. 154–165.
- [176] Dongnan Liu et al. ‘Unsupervised instance segmentation in microscopy images via panoptic domain adaptation and task re-weighting’. In: *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. 2020, pp. 4243–4252.
- [177] Hong Liu, Jianmin Wang and Mingsheng Long. ‘Cycle self-training for domain adaptation’. In: *Advances in Neural Information Processing Systems* 34 (2021), pp. 22968–22981.
- [178] Pengfei Liu et al. ‘Pre-train, prompt, and predict: A systematic survey of prompting methods in natural language processing’. In: *ACM Computing Surveys* 55.9 (2023), pp. 1–35.

- [179] Quande Liu, Qi Dou and Pheng-Ann Heng. ‘Shape-aware meta-learning for generalizing prostate MRI segmentation to unseen domains’. In: *Medical Image Computing and Computer Assisted Intervention–MICCAI 2020: 23rd International Conference, Lima, Peru, October 4–8, 2020, Proceedings, Part II* 23. Springer. 2020, pp. 475–485.
- [180] Quande Liu et al. ‘Feddg: Federated domain generalization on medical image segmentation via episodic learning in continuous frequency space’. In: *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. 2021, pp. 1013–1023.
- [181] Quande Liu et al. ‘MS-Net: multi-site network for improving prostate segmentation with heterogeneous MRI data’. In: *IEEE Transactions on Medical Imaging* 39.9 (2020), pp. 2713–2724.
- [182] Tingting Liu et al. ‘Medical Image Segmentation Based on a Hybrid Region-Based Active Contour Model’. In: *Computational and mathematical methods in medicine* 2014.1 (2014), p. 890725.
- [183] Xiangbin Liu et al. ‘A review of deep-learning-based medical image segmentation methods’. In: *Sustainability* 13.3 (2021), p. 1224.
- [184] Xiaofeng Liu et al. ‘Deep unsupervised domain adaptation: A review of recent advances and perspectives’. In: *APSIPA Transactions on Signal and Information Processing* 11.1 (2022).
- [185] Xiaoqing Liu et al. ‘Advances in deep learning-based medical image analysis’. In: *Health Data Science 2021* (2021).
- [186] Yabo Liu et al. ‘Fine-grained Representation Alignment for Zero-shot Domain Adaptation’. In: *IEEE Transactions on Multimedia* (2024).
- [187] Ze Liu et al. ‘Swin transformer: Hierarchical vision transformer using shifted windows’. In: *Proceedings of the IEEE/CVF international conference on computer vision*. 2021, pp. 10012–10022.
- [188] Jonathan Long, Evan Shelhamer and Trevor Darrell. ‘Fully convolutional networks for semantic segmentation’. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2015, pp. 3431–3440.

- [189] Mingsheng Long et al. ‘Learning transferable features with deep adaptation networks’. In: *International conference on machine learning*. PMLR. 2015, pp. 97–105.
- [190] Xiangde Luo et al. ‘Semi-supervised medical image segmentation through dual-task consistency’. In: *Proceedings of the AAAI Conference on Artificial Intelligence*. Vol. 35. 2021, pp. 8801–8809.
- [191] Yawei Luo et al. ‘Taking a closer look at domain shift: Category-level adversaries for semantics consistent domain adaptation’. In: *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. 2019, pp. 2507–2516.
- [192] Guofeng Lv et al. ‘Nuclei R-CNN: improve mask R-CNN for nuclei segmentation’. In: *2019 IEEE 2nd International Conference on Information Communication and Signal Processing (ICICSP)*. IEEE. 2019, pp. 357–362.
- [193] Xinhong Ma, Tianzhu Zhang and Changsheng Xu. ‘Gcan: Graph convolutional adversarial network for unsupervised domain adaptation’. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 2019, pp. 8266–8276.
- [194] Dwarikanath Mahapatra et al. ‘Unsupervised domain adaptation using feature disentanglement and GCNs for medical image classification’. In: *Proceedings of the European Conference on Computer Vision*. Springer. 2022, pp. 735–748.
- [195] Priyanka Malhotra et al. ‘[Retracted] Deep Neural Networks for Medical Image Segmentation’. In: *Journal of Healthcare Engineering* 2022.1 (2022), p. 9580991.
- [196] Mike May. ‘A better lens on disease’. In: *Scientific American* 302.5 (2010), pp. 74–77.
- [197] Bjoern H Menze et al. ‘The multimodal brain tumor image segmentation benchmark (BRATS)’. In: *IEEE transactions on medical imaging* 34.10 (2014), pp. 1993–2024.
- [198] Fausto Milletari, Nassir Navab and Seyed-Ahmad Ahmadi. ‘V-net: Fully convolutional neural networks for volumetric medical image segmentation’. In: *2016 fourth international conference on 3D vision (3DV)*. Ieee. 2016, pp. 565–571.
- [199] Shervin Minaee et al. ‘Image segmentation using deep learning: A survey’. In: *IEEE transactions on pattern analysis and machine intelligence* 44.7 (2021), pp. 3523–3542.

- [200] Pim Moeskops et al. ‘Deep learning for multi-task medical image segmentation in multiple modalities’. In: *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2016: 19th International Conference, Athens, Greece, October 17-21, 2016, Proceedings, Part II* 19. Springer. 2016, pp. 478–486.
- [201] Aliasghar Mortazi and Ulas Bagci. ‘Automatically designing CNN architectures for medical image segmentation’. In: *Machine Learning in Medical Imaging: 9th International Workshop, MLMI 2018, Held in Conjunction with MICCAI 2018, Granada, Spain, September 16, 2018, Proceedings* 9. Springer. 2018, pp. 98–106.
- [202] Aymen Mouelhi et al. ‘Automatic image segmentation of nuclear stained breast tissue sections using color active contour model and an improved watershed method’. In: *Biomedical Signal Processing and Control* 8.5 (2013), pp. 421–436.
- [203] Gillian Murphy et al. ‘The expanding role of MRI in prostate cancer’. In: *American Journal of Roentgenology* 201.6 (2013), pp. 1229–1238.
- [204] Ghulam Murtaza et al. ‘Deep learning-based breast cancer classification through medical imaging modalities: state of the art and research challenges’. In: *Artificial Intelligence Review* 53 (2020), pp. 1655–1720.
- [205] Peter Naylor et al. ‘Nuclei segmentation in histopathology images using deep neural networks’. In: *2017 IEEE 14th international symposium on biomedical imaging (ISBI 2017)*. IEEE. 2017, pp. 933–936.
- [206] Peter Naylor et al. ‘Segmentation of nuclei in histopathology images by deep regression of the distance map’. In: *IEEE transactions on medical imaging* 38.2 (2018), pp. 448–459.
- [207] Jiquan Ngiam et al. ‘Multimodal deep learning’. In: *Proceedings of the 28th international conference on machine learning (ICML-11)*. 2011, pp. 689–696.
- [208] Bloch Nicholas et al. ‘Nci-proc. ieee-isbi conf. 2013 challenge: Automated segmentation of prostate structures’. In: *The Cancer Imaging Archive* 5 (2015).
- [209] Dong Nie and Dinggang Shen. ‘Adversarial confidence learning for medical image segmentation and synthesis’. In: *International journal of computer vision* 128.10 (2020), pp. 2494–2513.

- [210] Ozan Oktay et al. ‘Attention u-net: Learning where to look for the pancreas’. In: *arXiv preprint arXiv:1804.03999* (2018).
- [211] Maxime Oquab et al. ‘Learning and transferring mid-level image representations using convolutional neural networks’. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2014, pp. 1717–1724.
- [212] José Ignacio Orlando et al. ‘Refuge challenge: A unified framework for evaluating automated methods for glaucoma assessment from fundus photographs’. In: *Medical Image Analysis* 59 (2020), p. 101570.
- [213] Cheng Ouyang et al. ‘Causality-inspired single-source domain generalization for medical image segmentation’. In: *IEEE Transactions on Medical Imaging* 42.4 (2022), pp. 1095–1106.
- [214] Sinno Jialin Pan and Qiang Yang. ‘A survey on transfer learning’. In: *IEEE Transactions on knowledge and data engineering* 22.10 (2009), pp. 1345–1359.
- [215] Sinno Jialin Pan et al. ‘Domain adaptation via transfer component analysis’. In: *IEEE transactions on neural networks* 22.2 (2010), pp. 199–210.
- [216] Siddharth Pandey, Pranshu Ranjan Singh and Jing Tian. ‘An image augmentation approach using two-stage generative adversarial network for nuclei image segmentation’. In: *Biomedical Signal Processing and Control* 57 (2020), p. 101782.
- [217] Dinesh D Patil and Sonal G Deore. ‘Medical image segmentation: a review’. In: *International Journal of Computer Science and Mobile Computing* 2.1 (2013), pp. 22–27.
- [218] Duo Peng et al. ‘Diffusion-based Image Translation with Label Guidance for Domain Adaptive Semantic Segmentation’. In: *Proceedings of the IEEE/CVF International Conference on Computer Vision*. 2023, pp. 808–820.
- [219] Kuan-Chuan Peng, Ziyan Wu and Jan Ernst. ‘Zero-shot deep domain adaptation’. In: *Proceedings of the European Conference on Computer Vision*. 2018, pp. 764–781.
- [220] Christian S Perone et al. ‘Unsupervised domain adaptation for medical imaging segmentation with self-ensembling’. In: *NeuroImage* 194 (2019), pp. 1–11.
- [221] Dzung L Pham, Chenyang Xu and Jerry L Prince. ‘Current methods in medical image segmentation’. In: *Annual review of biomedical engineering* 2.1 (2000), pp. 315–337.

- [222] Regina Pohle and Klaus D Toennies. ‘Segmentation of medical images using adaptive region growing’. In: *Medical Imaging 2001: Image Processing*. Vol. 4322. SPIE. 2001, pp. 1337–1346.
- [223] Guo-Jun Qi and Jiebo Luo. ‘Small data challenges in big data era: A survey of recent progress on unsupervised and semi-supervised methods’. In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* 44.4 (2020), pp. 2168–2187.
- [224] Imran Qureshi et al. ‘Medical image segmentation using deep semantic-based methods: A review of techniques, applications and emerging trends’. In: *Information Fusion* 90 (2023), pp. 316–352.
- [225] Alec Radford. ‘Unsupervised representation learning with deep convolutional generative adversarial networks’. In: *arXiv preprint arXiv:1511.06434* (2015).
- [226] Alec Radford et al. ‘Learning transferable visual models from natural language supervision’. In: *International conference on machine learning*. PMLR. 2021, pp. 8748–8763.
- [227] Colin Raffel et al. ‘Exploring the limits of transfer learning with a unified text-to-text transformer’. In: *Journal of machine learning research* 21.140 (2020), pp. 1–67.
- [228] Maithra Raghu et al. ‘Transfusion: Understanding transfer learning for medical imaging’. In: *Advances in neural information processing systems* 32 (2019).
- [229] Mohammad Mahfujur Rahman et al. ‘On minimum discrepancy estimation for deep domain adaptation’. In: *Domain Adaptation for Visual Understanding* (2020), pp. 81–94.
- [230] R Joshua Samuel Raj et al. ‘Optimal feature selection-based medical image classification using deep learning model in internet of medical things’. In: *IEEE Access* 8 (2020), pp. 58006–58017.
- [231] Meghavi Rana and Megha Bhushan. ‘Machine learning and deep learning approach for medical image analysis: diagnosis to detection’. In: *Multimedia Tools and Applications* 82.17 (2023), pp. 26731–26769.
- [232] Harsh Rangwani et al. ‘A closer look at smoothness in domain adversarial training’. In: *International conference on machine learning*. PMLR. 2022, pp. 18378–18399.

- [233] Khalid Raza and Nripendra K Singh. ‘A tour of unsupervised deep learning for medical image analysis’. In: *Current Medical Imaging* 17.9 (2021), pp. 1059–1077.
- [234] Shan E Ahmed Raza et al. ‘Micro-Net: A unified model for segmentation of various objects in microscopy images’. In: *Medical image analysis* 52 (2019), pp. 160–173.
- [235] Muhammad Imran Razzak, Saeeda Naz and Ahmad Zaib. ‘Deep learning for medical image processing: Overview, challenges and the future’. In: *Classification in BioApps: Automation of decision making* (2018), pp. 323–350.
- [236] Joseph Redmon et al. ‘You only look once: Unified, real-time object detection’. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2016, pp. 779–788.
- [237] Aasia Rehman, Muheet Ahmed Butt and Majid Zaman. ‘A survey of medical image analysis using deep learning approaches’. In: *2021 5th International Conference on Computing Methodologies and Communication (ICCMC)*. IEEE. 2021, pp. 1334–1342.
- [238] Jian Ren et al. ‘Unsupervised domain adaptation for classification of histopathology whole-slide images’. In: *Frontiers in bioengineering and biotechnology* 7 (2019), p. 102.
- [239] Shaoqing Ren et al. ‘Faster R-CNN: Towards real-time object detection with region proposal networks’. In: *IEEE transactions on pattern analysis and machine intelligence* 39.6 (2016), pp. 1137–1149.
- [240] Ringwald, Tobias and Stiefelhagen Rainer. ‘Unsupervised domain adaptation by uncertain feature alignment’. In: *British Machine Vision Conference*. 2020.
- [241] Robin Rombach et al. ‘High-resolution image synthesis with latent diffusion models’. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 2022, pp. 10684–10695.
- [242] Olaf Ronneberger, Philipp Fischer and Thomas Brox. ‘U-net: Convolutional networks for biomedical image segmentation’. In: *Medical image computing and computer-assisted intervention–MICCAI 2015: 18th international conference, Munich, Germany, October 5-9, 2015, proceedings, part III* 18. Springer. 2015, pp. 234–241.

- [243] Kuniaki Saito et al. ‘Maximum classifier discrepancy for unsupervised domain adaptation’. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2018, pp. 3723–3732.
- [244] Mark Sandler et al. ‘Mobilenetv2: Inverted residuals and linear bottlenecks’. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 2018, pp. 4510–4520.
- [245] Jürgen Schmidhuber. ‘Deep learning in neural networks: An overview’. In: *Neural networks* 61 (2015), pp. 85–117.
- [246] Mike Schuster and Kuldip K Paliwal. ‘Bidirectional recurrent neural networks’. In: *IEEE transactions on Signal Processing* 45.11 (1997), pp. 2673–2681.
- [247] Dale Schuurmans and Martin A Zinkevich. ‘Deep learning games’. In: *Advances in Neural Information Processing Systems* 29 (2016).
- [248] N Senthilkumaran and S Vaithegi. ‘Image segmentation by using thresholding techniques for medical images’. In: *Computer Science & Engineering: An International Journal* 6.1 (2016), pp. 1–13.
- [249] Artem Sevastopolsky. ‘Optic disc and cup segmentation methods for glaucoma detection with modification of U-Net convolutional neural network’. In: *Pattern Recognition and Image Analysis* 27 (2017), pp. 618–624.
- [250] Neeraj Sharma and Lalit M Aggarwal. ‘Automated medical image segmentation techniques’. In: *Journal of medical physics* 35.1 (2010), pp. 3–14.
- [251] Dinggang Shen, Guorong Wu and Heung-Il Suk. ‘Deep learning in medical image analysis’. In: *Annual review of biomedical engineering* 19.1 (2017), pp. 221–248.
- [252] Jian Shen et al. ‘Wasserstein distance guided representation learning for domain adaptation’. In: *Proceedings of the AAAI conference on artificial intelligence*. Vol. 32. 1. 2018.
- [253] Lu Sheng et al. ‘Avatar-net: Multi-scale zero-shot style transfer by feature decoration’. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2018, pp. 8242–8250.
- [254] Jianbo Shi and Jitendra Malik. ‘Normalized cuts and image segmentation’. In: *IEEE Transactions on pattern analysis and machine intelligence* 22.8 (2000), pp. 888–905.

- [255] Hoo-Chang Shin et al. ‘Deep convolutional neural networks for computer-aided detection: CNN architectures, dataset characteristics and transfer learning’. In: *IEEE transactions on medical imaging* 35.5 (2016), pp. 1285–1298.
- [256] Hyungseob Shin et al. ‘SDC-UDA: volumetric unsupervised domain adaptation framework for slice-direction continuous cross-modality medical image segmentation’. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 2023, pp. 7412–7421.
- [257] Connor Shorten and Taghi M Khoshgoftaar. ‘A survey on image data augmentation for deep learning’. In: *Journal of big data* 6.1 (2019), pp. 1–48.
- [258] Jie Shu et al. ‘Segmenting overlapping cell nuclei in digital histopathology images’. In: *2013 35th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*. IEEE. 2013, pp. 5445–5448.
- [259] Rui Shu et al. ‘A dirt-t approach to unsupervised domain adaptation’. In: *arXiv preprint arXiv:1802.08735* (2018).
- [260] Nahian Siddique et al. ‘U-net and its variants for medical image segmentation: A review of theory and applications’. In: *IEEE access* 9 (2021), pp. 82031–82057.
- [261] Karen Simonyan and Andrew Zisserman. ‘Very deep convolutional networks for large-scale image recognition’. In: *arXiv preprint arXiv:1409.1556* (2014).
- [262] Amber L Simpson et al. ‘A large annotated medical image dataset for the development and evaluation of segmentation algorithms’. In: *arXiv preprint arXiv:1902.09063* (2019).
- [263] Jayanthi Sivaswamy et al. ‘Drishti-gs: Retinal image dataset for optic nerve head (onh) segmentation’. In: *2014 IEEE 11th International Symposium on Biomedical Imaging*. IEEE. 2014, pp. 53–56.
- [264] Kihyuk Sohn et al. ‘Fixmatch: Simplifying semi-supervised learning with consistency and confidence’. In: *Advances in Neural Information Processing Systems* 33 (2020), pp. 596–608.
- [265] Yang Song et al. ‘Cell nuclei segmentation in fluorescence microscopy images using inter-and intra-region discriminative information’. In: *2013 35th Annual International*

- Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*. IEEE. 2013, pp. 6087–6090.
- [266] Yang Song et al. ‘Supervised intra-embedding of fisher vectors for histopathology image classification’. In: *Medical Image Computing and Computer Assisted Intervention-MICCAI 2017: 20th International Conference, Quebec City, QC, Canada, September 11-13, 2017, Proceedings, Part III* 20. Springer. 2017, pp. 99–106.
 - [267] Zixian Su et al. ‘Rethinking data augmentation for single-source domain generalization in medical image segmentation’. In: *Proceedings of the AAAI Conference on Artificial Intelligence*. Vol. 37. 2. 2023, pp. 2366–2374.
 - [268] S Suganyadevi, V Seethalakshmi and Krishnasamy Balasamy. ‘A review on deep learning in medical image analysis’. In: *International Journal of Multimedia Information Retrieval* 11.1 (2022), pp. 19–38.
 - [269] Masashi Sugiyama, Matthias Krauledat and Klaus-Robert Müller. ‘Covariate shift adaptation by importance weighted cross validation.’ In: *Journal of Machine Learning Research* 8.5 (2007).
 - [270] Farhana Sultana, Abu Sufian and Paramartha Dutta. ‘Evolution of image segmentation using deep convolutional neural network: A survey’. In: *Knowledge-Based Systems* 201 (2020), p. 106062.
 - [271] Baochen Sun, Jiashi Feng and Kate Saenko. ‘Correlation alignment for unsupervised domain adaptation’. In: *Domain adaptation in computer vision applications* (2017), pp. 153–171.
 - [272] Baochen Sun and Kate Saenko. ‘Deep coral: Correlation alignment for deep domain adaptation’. In: *Computer Vision–ECCV 2016 Workshops: Amsterdam, The Netherlands, October 8-10 and 15-16, 2016, Proceedings, Part III* 14. Springer. 2016, pp. 443–450.
 - [273] I Sutskever. ‘Sequence to Sequence Learning with Neural Networks’. In: *arXiv preprint arXiv:1409.3215* (2014).
 - [274] Christian Szegedy et al. ‘Going deeper with convolutions’. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2015, pp. 1–9.

- [275] Nima Tajbakhsh et al. ‘Embracing imperfect datasets: A review of deep learning solutions for medical image segmentation’. In: *Medical image analysis* 63 (2020), p. 101693.
- [276] Ryohei Takahashi and Yuya Kajikawa. ‘Computer-aided diagnosis: A survey with bibliometric analysis’. In: *International journal of medical informatics* 101 (2017), pp. 58–67.
- [277] Jonti Talukdar et al. ‘Transfer learning for object detection using state-of-the-art deep neural networks’. In: *2018 5th international conference on signal processing and integrated networks (SPIN)*. IEEE. 2018, pp. 78–83.
- [278] Chuanqi Tan et al. ‘A survey on deep transfer learning’. In: *Artificial Neural Networks and Machine Learning–ICANN 2018: 27th International Conference on Artificial Neural Networks, Rhodes, Greece, October 4-7, 2018, Proceedings, Part III* 27. Springer. 2018, pp. 270–279.
- [279] Mingxing Tan, Ruoming Pang and Quoc V Le. ‘Efficientdet: Scalable and efficient object detection’. In: *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. 2020, pp. 10781–10790.
- [280] Hao Tang et al. ‘Master: Meta style transformer for controllable zero-shot and few-shot artistic style transfer’. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 2023, pp. 18329–18338.
- [281] Korawat Tanwisuth et al. ‘A prototype-oriented framework for unsupervised domain adaptation’. In: *Advances in Neural Information Processing Systems* 34 (2021), pp. 17194–17208.
- [282] Afaf Tareef et al. ‘Multi-pass fast watershed for accurate segmentation of overlapping cervical cells’. In: *IEEE transactions on medical imaging* 37.9 (2018), pp. 2044–2059.
- [283] Yih-Chung Tham et al. ‘Global prevalence of glaucoma and projections of glaucoma burden through 2040: a systematic review and meta-analysis’. In: *Ophthalmology* 121.11 (2014), pp. 2081–2090.
- [284] Chunwei Tian et al. ‘Deep learning on image denoising: An overview’. In: *Neural Networks* 131 (2020), pp. 251–275.

- [285] Lei Tian et al. ‘Domain adaptation by class centroid matching and local manifold self-learning’. In: *IEEE Transactions on Image Processing* 29 (2020), pp. 9703–9718.
- [286] Zhiqiang Tian et al. ‘PSNet: prostate segmentation on MRI based on a convolutional neural network’. In: *Journal of medical imaging* 5.2 (2018), pp. 021208–021208.
- [287] Philipp Tschandl, Christoph Sinz and Harald Kittler. ‘Domain-specific classification-pretrained fully convolutional network encoders for skin lesion segmentation’. In: *Computers in biology and medicine* 104 (2019), pp. 111–116.
- [288] Eric Tzeng et al. ‘Adversarial discriminative domain adaptation’. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2017, pp. 7167–7176.
- [289] Eric Upschulte et al. ‘Contour proposal networks for biomedical instance segmentation’. In: *Medical image analysis* 77 (2022), p. 102371.
- [290] Annegreet Van Opbroek et al. ‘Transfer learning improves supervised image segmentation across imaging protocols’. In: *IEEE transactions on medical imaging* 34.5 (2014), pp. 1018–1030.
- [291] Thomas Varsavsky et al. ‘Test-time unsupervised domain adaptation’. In: *Medical Image Computing and Computer Assisted Intervention–MICCAI 2020: 23rd International Conference, Lima, Peru, October 4–8, 2020, Proceedings, Part I* 23. Springer. 2020, pp. 428–436.
- [292] A Vaswani. ‘Attention is all you need’. In: *Advances in Neural Information Processing Systems* (2017).
- [293] Mitko Veta et al. ‘Automatic nuclei segmentation in H&E stained breast cancer histopathology images’. In: *PloS one* 8.7 (2013), e70221.
- [294] Athanasios Voulodimos et al. ‘Deep learning for computer vision: A brief review’. In: *Computational intelligence and neuroscience* 2018.1 (2018), p. 7068349.
- [295] Aarno Oskar Vuola, Saad Ullah Akram and Juho Kannala. ‘Mask-RCNN and U-net ensembled for nuclei segmentation’. In: *2019 IEEE 16th international symposium on biomedical imaging (ISBI 2019)*. IEEE. 2019, pp. 208–212.

- [296] Fei Wang et al. ‘Residual attention network for image classification’. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2017, pp. 3156–3164.
- [297] Haoran Wang et al. ‘Classes matter: A fine-grained adversarial approach to cross-domain semantic segmentation’. In: *European conference on computer vision*. Springer. 2020, pp. 642–659.
- [298] Heng Wang et al. ‘Multiscale kernels for enhanced U-shaped network to improve 3D neuron tracing’. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*. 2019, pp. 0–0.
- [299] Heng Wang et al. ‘V2a-mapper: A lightweight solution for vision-to-audio generation by connecting foundation models’. In: *Proceedings of the AAAI Conference on Artificial Intelligence*. Vol. 38. 14. 2024, pp. 15492–15501.
- [300] Jian Wang et al. ‘A review of deep learning on medical image analysis’. In: *Mobile Networks and Applications* 26.1 (2021), pp. 351–380.
- [301] Jinghua Wang, Ming-Ming Cheng and Jianmin Jiang. ‘Domain shift preservation for zero-shot domain adaptation’. In: *IEEE Transactions on Image Processing* 30 (2021), pp. 5505–5517.
- [302] Jinghua Wang and Jianmin Jiang. ‘Adversarial learning for zero-shot domain adaptation’. In: *Computer Vision—ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part XXI 16*. Springer. 2020, pp. 329–344.
- [303] Jinghua Wang and Jianmin Jiang. ‘Learning across tasks for zero-shot domain adaptation from a single source domain’. In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* 44.10 (2021), pp. 6264–6279.
- [304] Kun Wang et al. ‘GSAL: Geometric structure adversarial learning for robust medical image segmentation’. In: *Pattern Recognition* 140 (2023), p. 109596.
- [305] Mei Wang and Weihong Deng. ‘Deep visual domain adaptation: A survey’. In: *Neurocomputing* 312 (2018), pp. 135–153.
- [306] Pin Wang et al. ‘Automatic cell nuclei segmentation and classification of breast cancer histopathology images’. In: *Signal Processing* 122 (2016), pp. 1–13.

- [307] Qian Wang, Penghui Bu and Toby P Breckon. ‘Unifying unsupervised domain adaptation and zero-shot visual recognition’. In: *2019 International Joint Conference on Neural Networks (IJCNN)*. IEEE. 2019, pp. 1–8.
- [308] Qin Wang et al. ‘Continual test-time domain adaptation’. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 2022, pp. 7201–7211.
- [309] Risheng Wang et al. ‘Medical image segmentation using deep learning: A survey’. In: *IET image processing* 16.5 (2022), pp. 1243–1267.
- [310] Shujun Wang et al. ‘Boundary and entropy-driven adversarial learning for fundus image segmentation’. In: *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer. 2019, pp. 102–110.
- [311] Shujun Wang et al. ‘Patch-based output space adversarial learning for joint optic disc and cup segmentation’. In: *IEEE Transactions on Medical Imaging* 38.11 (2019), pp. 2485–2495.
- [312] Weibin Wang et al. ‘Medical image classification using deep learning’. In: *Deep learning in healthcare: paradigms and applications* (2020), pp. 33–51.
- [313] Yifei Wang et al. ‘Deep domain adaptation by geodesic distance minimization’. In: *Proceedings of the IEEE International Conference on Computer Vision Workshops*. 2017, pp. 2651–2657.
- [314] Yuchao Wang et al. ‘Semi-supervised semantic segmentation using unreliable pseudo-labels’. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 2022, pp. 4248–4257.
- [315] Zuhui Wang and Zhaozheng Yin. ‘Annotation-efficient cell counting’. In: *Medical Image Computing and Computer Assisted Intervention–MICCAI 2021: 24th International Conference, Strasbourg, France, September 27–October 1, 2021, Proceedings, Part VIII* 24. Springer. 2021, pp. 405–414.
- [316] Karl Weiss, Taghi M Khoshgoftaar and DingDing Wang. ‘A survey of transfer learning’. In: *Journal of Big data* 3 (2016), pp. 1–40.

- [317] Garrett Wilson and Diane J Cook. ‘A survey of unsupervised deep domain adaptation’. In: *ACM Transactions on Intelligent Systems and Technology (TIST)* 11.5 (2020), pp. 1–46.
- [318] Garrett Wilson, Janardhan Rao Doppa and Diane J Cook. ‘CALDA: Improving multi-source time series domain adaptation with contrastive adversarial learning’. In: *IEEE transactions on pattern analysis and machine intelligence* (2023).
- [319] Junde Wu et al. ‘Medsegdiff-v2: Diffusion-based medical image segmentation with transformer’. In: *Proceedings of the AAAI Conference on Artificial Intelligence*. Vol. 38. 2024, pp. 6030–6038.
- [320] Junde Wu et al. ‘SeATrans: learning segmentation-assisted diagnosis model via transformer’. In: *International Conference on Medical Image Computing and Computer-Assisted Intervention*. 2022, pp. 677–687.
- [321] Yicheng Wu et al. ‘Vessel-Net: Retinal vessel segmentation under multi-path supervision’. In: *Medical Image Computing and Computer Assisted Intervention–MICCAI 2019: 22nd International Conference, Shenzhen, China, October 13–17, 2019, Proceedings, Part I* 22. Springer. 2019, pp. 264–272.
- [322] Haifeng Xia and Zhengming Ding. ‘Hgnet: Hybrid generative network for zero-shot domain adaptation’. In: *Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part XXVII* 16. Springer. 2020, pp. 55–70.
- [323] Tiange Xiang et al. ‘BiO-Net: learning recurrent bi-directional connections for encoder-decoder architecture’. In: *Medical Image Computing and Computer Assisted Intervention–MICCAI 2020: 23rd International Conference, Lima, Peru, October 4–8, 2020, Proceedings, Part I* 23. Springer. 2020, pp. 74–84.
- [324] Qingsong Xie et al. ‘Unsupervised domain adaptation for medical image segmentation by disentanglement learning and self-training’. In: *IEEE Transactions on Medical Imaging* 43.1 (2022), pp. 4–14.
- [325] Xiaozheng Xie et al. ‘A survey on incorporating domain knowledge into deep learning for medical image analysis’. In: *Medical Image Analysis* 69 (2021), p. 101985.

- [326] Yutong Xie et al. ‘Knowledge-based collaborative deep learning for benign-malignant lung nodule classification on chest CT’. In: *IEEE transactions on medical imaging* 38.4 (2018), pp. 991–1004.
- [327] Fuyong Xing, Yuanpu Xie and Lin Yang. ‘An automatic learning-based framework for robust nucleus segmentation’. In: *IEEE transactions on medical imaging* 35.2 (2015), pp. 550–566.
- [328] Zhaohu Xing et al. ‘Diff-UNet: A Diffusion Embedded Network for Volumetric Segmentation’. In: *arXiv preprint arXiv:2303.10326* (2023).
- [329] Xuanang Xu et al. ‘Efficient multiple organ localization in CT image using 3D region proposal network’. In: *IEEE transactions on medical imaging* 38.8 (2019), pp. 1885–1898.
- [330] Hong Yan et al. ‘Development and validation of an unsupervised feature learning system for leukocyte characterization and classification: A multi-hospital study’. In: *International Journal of Computer Vision* 129 (2021), pp. 1837–1856.
- [331] Jianwei Yang, Devi Parikh and Dhruv Batra. ‘Joint unsupervised learning of deep representations and image clusters’. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2016, pp. 5147–5156.
- [332] Jinyu Yang et al. ‘Label-driven reconstruction for domain adaptation in semantic segmentation’. In: *Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part XXVII* 16. Springer. 2020, pp. 480–498.
- [333] Senqiao Yang et al. ‘Unified Language-driven Zero-shot Domain Adaptation’. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 2024, pp. 23407–23415.
- [334] Siqi Yang et al. ‘Minimizing labeling cost for nuclei instance segmentation and classification with cross-domain images and weak labels’. In: *Proceedings of the AAAI Conference on Artificial Intelligence*. Vol. 35. 1. 2021, pp. 697–705.
- [335] Xiaodong Yang, Houqiang Li and Xiaobo Zhou. ‘Nuclei segmentation using marker-controlled watershed, tracking using mean-shift, and Kalman filter in time-lapse microscopy’. In: *IEEE Transactions on Circuits and Systems I: Regular Papers* 53.11 (2006), pp. 2405–2414.

- [336] Yijun Yang et al. ‘DiffMIC: Dual-Guidance Diffusion Network for Medical Image Classification’. In: *arXiv preprint arXiv:2303.10610* (2023).
- [337] Jason Yosinski et al. ‘How transferable are features in deep neural networks?’ In: *Advances in neural information processing systems* 27 (2014).
- [338] Qianbi Yu et al. ‘Unsupervised domain adaptive fundus image segmentation with few labeled source data’. In: *British Machine Vision Conference*. 2022.
- [339] Shuang Yu et al. ‘Robust optic disc and cup segmentation with deep learning for glaucoma detection’. In: *Computerized Medical Imaging and Graphics* 74 (2019), pp. 61–71.
- [340] Xiang Yu et al. ‘Transfer learning for medical images analyses: A survey’. In: *Neuro-computing* 489 (2022), pp. 230–254.
- [341] Nida M Zaitoun and Musbah J Aqel. ‘Survey on image segmentation techniques’. In: *Procedia Computer Science* 65 (2015), pp. 797–806.
- [342] Qiang Zhai et al. ‘Mutual graph learning for camouflaged object detection’. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 2021, pp. 12997–13007.
- [343] Fangneng Zhan and Changgong Zhang. ‘Spatial-aware gan for unsupervised person re-identification’. In: *2020 25th international conference on pattern recognition (ICPR)*. IEEE. 2021, pp. 6889–6896.
- [344] Donghao Zhang et al. ‘Nuclei instance segmentation with dual contour-enhanced adversarial network’. In: *2018 IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018)*. IEEE. 2018, pp. 409–412.
- [345] Donghao Zhang et al. ‘Panoptic segmentation with an end-to-end cell r-cnn for pathology image analysis’. In: *Medical Image Computing and Computer Assisted Intervention–MICCAI 2018: 21st International Conference, Granada, Spain, September 16-20, 2018, Proceedings, Part II 11*. Springer. 2018, pp. 237–244.
- [346] Hang Zhang, Jia Xue and Kristin Dana. ‘Deep ten: Texture encoding network’. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 2017, pp. 708–717.

- [347] Jiawei Zhang et al. ‘A comprehensive review of image analysis methods for microorganism counting: from classical image processing to deep learning approaches’. In: *Artificial Intelligence Review* (2022), pp. 1–70.
- [348] Yu-Jin Zhang. ‘A survey on evaluation methods for image segmentation’. In: *Pattern recognition* 29.8 (1996), pp. 1335–1346.
- [349] Ling Zhang et al. ‘Generalizing deep learning for medical image segmentation to unseen domains via deep stacked transformation’. In: *IEEE transactions on medical imaging* 39.7 (2020), pp. 2531–2540.
- [350] Qiming Zhang et al. ‘Category anchor-guided unsupervised domain adaptation for semantic segmentation’. In: *Advances in neural information processing systems* 32 (2019).
- [351] Weichen Zhang et al. ‘Collaborative and adversarial network for unsupervised domain adaptation’. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2018, pp. 3801–3809.
- [352] Xiaoli Zhang, Xiongfei Li and Yuncong Feng. ‘A medical image segmentation algorithm based on bi-directional region growing’. In: *Optik* 126.20 (2015), pp. 2398–2404.
- [353] Yifan Zhang et al. ‘Collaborative unsupervised domain adaptation for medical image diagnosis’. In: *IEEE Transactions on Image Processing* 29 (2020), pp. 7834–7844.
- [354] Yiwen Zhang, Haoran Lai and Wei Yang. ‘Cascade UNet and CH-UNet for thyroid nodule segmentation and benign and malignant classification’. In: *Segmentation, Classification, and Registration of Multi-modality Medical Imaging Data: MICCAI 2020 Challenges, ABCs 2020, L2R 2020, TN-SCUI 2020, Held in Conjunction with MICCAI 2020, Lima, Peru, October 4–8, 2020, Proceedings* 23. Springer. 2021, pp. 129–134.
- [355] Bingchao Zhao et al. ‘Triple U-net: Hematoxylin-aware nuclei segmentation with progressive dense feature aggregation’. In: *Medical Image Analysis* 65 (2020), p. 101786.
- [356] Feng Zhao and Xianghua Xie. ‘An overview of interactive medical image segmentation’. In: *Annals of the BMVA* 2013.7 (2013), pp. 1–22.

- [357] Hengshuang Zhao et al. ‘Pyramid scene parsing network’. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2017, pp. 2881–2890.
- [358] Rongchang Zhao et al. ‘EGDCL: An adaptive curriculum learning framework for unbiased glaucoma diagnosis’. In: *Proceedings of the European Conference on Computer Vision*. Springer. 2020, pp. 190–205.
- [359] Rongchang Zhao et al. ‘Weakly-supervised simultaneous evidence identification and segmentation for automated glaucoma diagnosis’. In: *Proceedings of the AAAI Conference on Artificial Intelligence*. Vol. 33. 2019, pp. 809–816.
- [360] Ziyuan Zhao et al. ‘LE-UDA: Label-efficient unsupervised domain adaptation for medical image segmentation’. In: *IEEE Transactions on Medical Imaging* 42.3 (2022), pp. 633–646.
- [361] Ziyuan Zhao et al. ‘Mt-uda: Towards unsupervised cross-modality medical image segmentation with limited source labels’. In: *Medical Image Computing and Computer Assisted Intervention—MICCAI 2021: 24th International Conference, Strasbourg, France, September 27–October 1, 2021, Proceedings, Part I* 24. Springer. 2021, pp. 293–303.
- [362] S Kevin Zhou, Hayit Greenspan and Dinggang Shen. *Deep learning for medical image analysis*. Academic Press, 2023.
- [363] Yanning Zhou et al. ‘Cia-net: Robust nuclei instance segmentation with contour-aware information aggregation’. In: *Information Processing in Medical Imaging: 26th International Conference, IPMI 2019, Hong Kong, China, June 2–7, 2019, Proceedings* 26. Springer. 2019, pp. 682–693.
- [364] Yi Zhou et al. ‘CCT-Net: category-invariant cross-domain transfer for medical single-to-multiple disease diagnosis’. In: *Proceedings of the IEEE/CVF International Conference on Computer Vision*. 2021, pp. 8260–8270.
- [365] Jun-Yan Zhu et al. ‘Unpaired image-to-image translation using cycle-consistent adversarial networks’. In: *Proceedings of the IEEE international conference on computer vision*. 2017, pp. 2223–2232.
- [366] Qikui Zhu, Bo Du and Pingkun Yan. ‘Boundary-weighted domain adaptive neural network for prostate MR image segmentation’. In: *IEEE transactions on medical imaging* 39.3 (2019), pp. 753–763.

- [367] Fuzhen Zhuang et al. ‘A comprehensive survey on transfer learning’. In: *Proceedings of the IEEE* 109.1 (2020), pp. 43–76.
- [368] Yang Zou et al. ‘Confidence regularized self-training’. In: *Proceedings of the IEEE/CVF international conference on computer vision*. 2019, pp. 5982–5991.
- [369] Yang Zou et al. ‘Unsupervised domain adaptation for semantic segmentation via class-balanced self-training’. In: *Proceedings of the European conference on computer vision (ECCV)*. 2018, pp. 289–305.

Appendix

Abbreviations

3D: Three-dimensional

AJI: Aggregated Jaccard Index

ASD: Average Symmetric Surface Distance

BN: Batch Normalization

BERT: Bidirectional Encoder Representations from Transformers

CAD: Computer-Aided Diagnosis

CDR: Cup-to-disc Ratio

CNNs: Convolutional Neural Networks

CORAL: Correlation Alignment

CRAG: Colorectal Adenocarcinoma Gland

CRFs: Conditional Random Fields

CT: Computed Tomography

DAN: Deep Adaptation Network

DBNs: Deep Belief Networks

Dice: Dice Similarity Coefficient

Dpath: DigestPath

DQ: Detection Quality

DCNN: Deep Convolutional Neural Network

GANs: Generative Adversarial Networks

GCN: Graph Convolutional Network

GlaS: Gland Segmentation in Colon Histology Images

GNNs: Graph Neural Networks

GlaS: Gland Segmentation in Colon Histology Images

GPU: Graphics Processing Unit

HV: Hover

IrT: Irrelevant Tasks

KLIEP: Kullback-Leibler Importance Estimation Procedure

KMM: Kernel Mean Matching

LLMs: Large Language Models

LM: Language Models

MMD: Maximum Mean Discrepancy

MIoU: Mean Intersection over Union

MRI: Magnetic Resonance Imaging

NC: Nuclear Classification

NP: Nuclear Pixel

NLP: Natural Language Processing

OC: Optic Cup

OD: Optic Disc

PQ: Panoptic Quality

PZ: Peripheral Zone

REFUGE: Retinal Fundus Glaucoma Challenge

ReLU: Rectified Linear Unit

RKHS: Kernel Hilbert Space

RNNs: Recurrent Neural Networks

SAM: Segment Anything Model

SOTA: State-Of-The-Art

SQ: Segmentation Quality

ToI: Target of Interest

TZ: Transition Zone

UDA: Unsupervised Domain Adaptation

VAE: Variational Auto-encoder

ViTs: Vision Transformers

VGG: Visual Geometry Group

WSIs: Whole-slide Images

ZSDA: Zero-shot Domain Adaptation