Transfer Learning for Domain Adaptation

Xiaoquan Liu
Interdisciplinary Data Science
x.liu@duke.edu

Yuanjing Zhu
Interdisciplinary Data Science
yuanjing.zhu@duke.edu

Abstract

Transfer learning is an effective method for medical image segmentation, particularly with limited training data availability. This study focuses on optimizing pre-trained semantic segmentation models for fluid segmentation in optical coherence tomography (OCT) images, specifically focusing on diabetic macular edema (DME). Utilizing a public OCT dataset, we first demonstrate that the improvements in segmentation accuracy are highly task-dependent in transfer learning. Initial results without fine-tuning show low mean intersection over Union(mIoU) scores, indicating the need for further model adjustments. Then, we explore various fine-tuning techniques and employ data augmentation strategies and loss functions to enhance model performance. The results demonstrate significant improvements in segmentation accuracy, as indicated by substantial increases in both mIoU and mean Dice Coefficient (mDice) scores. These experiments underscore the potential of domain-adapted network tuning to achieve high accuracy and reliability in medical image segmentation, even with limited representative training data.

1 Introduction

Semantic segmentation classifies pixels to isolate specific items in images. While many existing segmentation models are initially developed for tasks like segmenting objects in autonomous driving images, it is now gaining traction in healthcare. Accurate image segmentation is fundamental for diagnosing and monitoring a wide range of medical conditions, including the detection of tumors, lesions, or abnormalities within organs and tissues. It enables precise quantification of anatomical structures and their volumetric analysis, facilitating treatment planning, surgical interventions, and radiation therapy targeting. Moreover, segmentation is essential for tracking disease progression over time and assessing treatment outcomes, offering valuable insights into the effectiveness of medical interventions. In essence, medical image segmentation empowers healthcare professionals with the tools needed for early diagnosis, personalized treatment, and improved patient care, ultimately leading to better clinical decision-making and patient outcomes.

A technique called transfer learning, where a pre-trained neural network model, often trained on a large dataset for a related task, is fine-tuned or adapted to a specific dataset, is usually used in semantic segmentation. This approach is favored for overcoming limited labeled data, transferring knowledge from pre-trained models, accelerating training convergence, enhancing model accuracy, reducing overfitting risk, and enabling reusable models for related tasks.

Our work, centered on optimizing established pre-trained segmentation models through fine-tuning, holds great promise for advancing ophthalmological diagnostics, especially for patients diagnosed with diabetic macular edema (DME). By tailoring segmentation models to address the intricacies of the eye's structure in DME cases, our work aims to significantly enhance the accuracy and reliability of retinal layer segmentation, making it particularly beneficial for applications in the field of medical image analysis of optical coherence tomography (OCT).

Preprint. Under review.

2 Related Works

In medical image processing, localization—that is, assigning a class name to each pixel—is part of the intended output. Among the myriad of segmentation architectures, U-Net has established itself as a widely adopted and effective solution. As shown in Figure 1, U-Net's architecture, proposed by Ronneberger et al. in 2015, is characterized by a U-shaped network that combines the strengths of convolutional neural networks (CNNs) for feature extraction and fully convolutional networks for precise pixel-wise segmentation [1]. Its success lies in its ability to capture both local and global contextual information through expansive paths and shortcut connections, facilitating the accurate delineation of object boundaries.

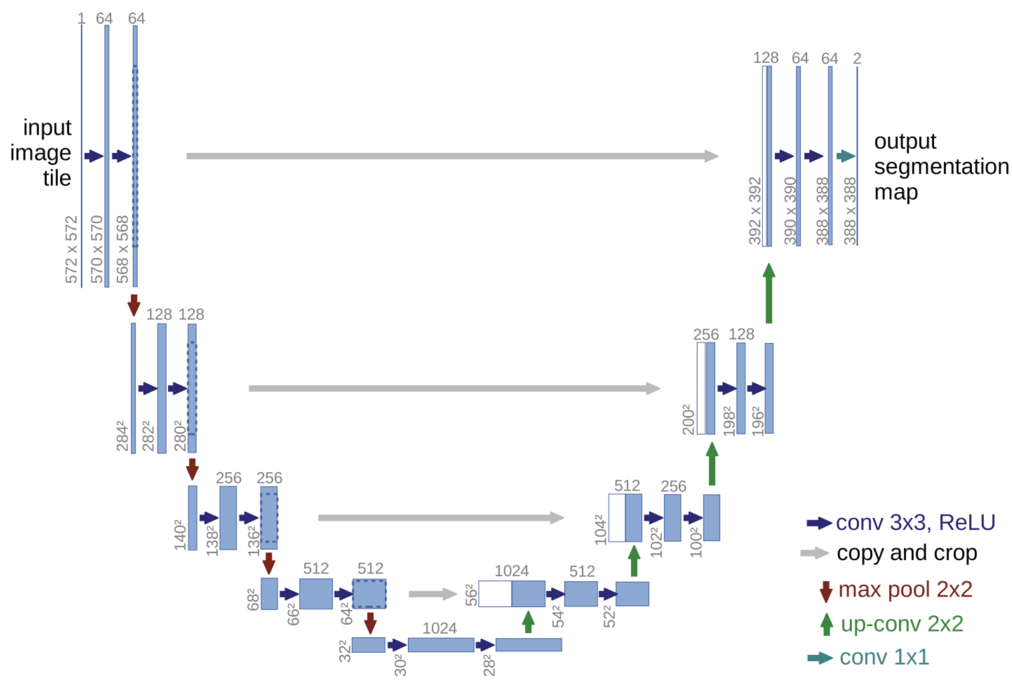

Figure 1: U-Net architecture (example for 32x32 pixels in the lowest resolution)

With the increasing need for usability of the trained models in real-world practice, transfer learning has been studied and widely used for medical image segmentation. Fine-tuned pre-trained CNN models for brain tumor classification [2] and brain lesion segmentation [3] have achieved state-of-the-art performance. Given U-Net's intrinsic design, which renders it highly effective for tasks requiring high-resolution segmentation, especially in applications where preserving spatial details is paramount, our study concentrates on investigating various decoder heads within the U-Net architecture to enhance model performance in medical image segmentation through the application of transfer learning techniques.

3 Methodology

3.1 Data

Our study utilizes a publicly available medical OCT dataset for segmentation of optical coherence tomography images with diabetic macular edema. It contains 610 images, of which 110 are manually annotated, providing comprehensive true class for each pixel. Our primary focus was on these 110 annotated images, as they offer a rich resource for accurate segmentation analysis. Figure 1 illustrates an example of an original OCT image with its corresponding annotated label mask, where diabetic macular edema is not present. In these instances, the retinal layers appear flat or concave, and there is an absence of fluid accumulation. In contrast, Figure 2 presents an OCT image and its label mask where diabetic macular edema is evident, distinctly revealing fluid-filled regions in the label mask.

In our analysis, we consolidated the 13 distinct fluid labels into one single label for fluid presence. The original images, at 468x796 resolution, were resized to 256x256 pixels for compatibility with our chosen processing tool, the MMSegmentation library. After resizing, we converted the images to .png format. Finally, we split the dataset into training and testing sets in an 80:20 ratio, ensuring ample data for model training and a sufficient subset for performance evaluation.

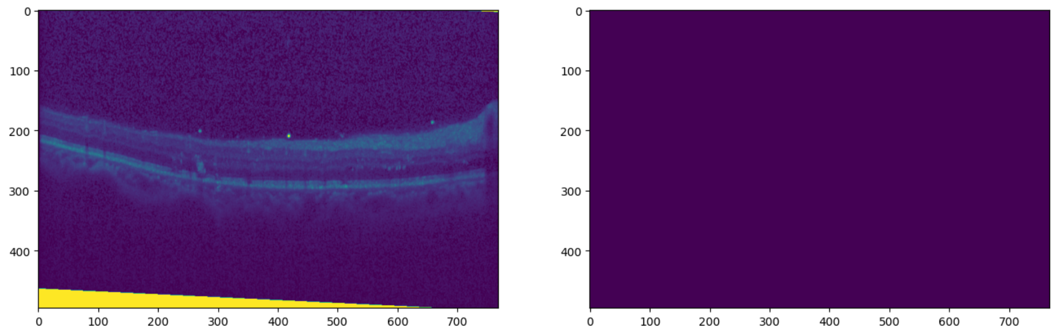

Figure 2: OCT image and its corresponding annotations in absence of diabetic macular edema. (Left) Original OCT image. (Right) Ground truth annotation.

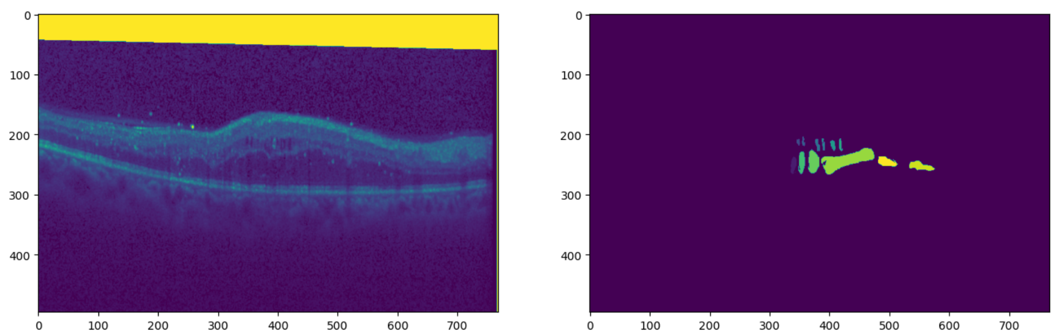

Figure 3: OCT images and corresponding annotations displaying diabetic macular edema. (Left) Original OCT image. (Right) Ground truth annotation.

3.2 Segmentation Process Workflow

Figure 3 outlines the complete workflow of our segmentation study for OCT images with diabetic macular edema. We started by preprocessing the dataset, including image resizing and train test splitting. We largely leveraged the MMSegmentation library, a versatile toolkit for semantic segmentation that contains high-quality implementations of popular semantic segmentation methods and datasets.

For model selection, we conducted preliminary inferences using pre-trained models from the MMSegmentation library. The Unet architecture was selected as our backbone due to its robust performance in segmentation tasks. We then assessed the inference capabilities of different decoder heads: FCN, PSPNet, and DeepLabV3. The decoder head yielding the highest performance on the OCT dataset was chosen for further fine-tuning.

During fine-tuning, we applied several data augmentation strategies such as random cropping, horizontal flipping, and random rotation to enhance the model's generalization ability. We also experimented with different loss functions, including cross-entropy loss, Dice loss, and focal loss, to find the most effective approach. Additionally, we teseted with the SGD and Adam optimizers, making necessary adjustments to the regularization parameters to fine-tune the model for optimal performance in segmenting the OCT images.

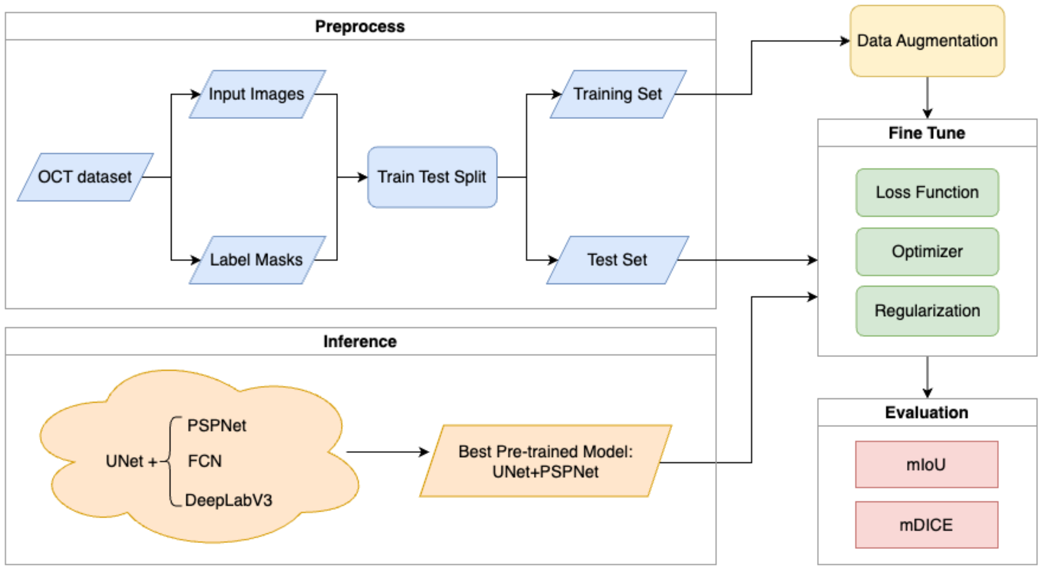

Figure 4: Flowchart showing the semantic segmentation process. This includes (1) data preprocessing, (2) making inferences with pre-trained model, (3) data augmentation, (4) fine-tuning pre-trained model on current dataset, (5) model evaluation

3.3 Evaluation Metrics

In contrast to standard classification tasks that rely on accuracy, recall, and precision for evaluation, segmentation tasks require metrics that can assess pixel-level accuracy. In this study, we chose the mean Intersection over Union (mIOU) and mean Dice Coefficient (mDice) as our primary metrics.

3.3.1 mIOU

The Intersection over Union (IoU) score, also known as the Jaccard Index, is a commonly used metric to evaluate the performance of image segmentation algorithms. It quantifies how well the predicted segmentation aligns with the ground truth by calculating the ratio between the area of overlap and the area of union of the two regions. The mean IoU (mIOU) is calculated by taking the average of the IoU scores for each class. It ranges from 0 to 1, where higher mIoU scores suggest better segmentation performance.

$$IoU = \frac{Area\ of\ Intersection}{Area\ of\ Union}$$

$$mIoU = \frac{1}{N}\sum_{i=1}^{N} IoU_i$$

3.3.2 mDICE

The mean Dice Coefficient (mDICE), also known as the Dice Similarity Coefficient, is another evaluation metric that is commonly used in semantic segmentation. Like mIoU, the dice score also ranges between 0 and 1, with a score of 1 signifying a perfect match between the predicted segmentation and the ground truth. However, the distinction lies in their sensitivity: IoU tends to penalize both under- and over-segmentation more harshly compared to the mDICE.

$$DICE = 2 * \frac{Area\ of\ Overlap}{Total\ Area}$$

4 Experiments

4.1 Preliminary Model Evaluation without Fine-tuning

We first conducted inference using three pre-existing models available in the MMSegmentation library. All three models share a common architecture, utilizing the U-Net framework with distinct decoder heads pre-trained on the High-Resolution Fundus (HRF) dataset. As depicted in Table 1, all models exhibit suboptimal mIoU scores when applied to the OCT dataset, indicating a notable performance gap in segmentation tasks.

Table 1: Model configurations and evaluation results

Name	Decoder Head	mIoU
Model1	PSPNet	0.0022
Model2	FCN	0.0008
Model3	DeepLabV3	0.0016

Figure 5 presents the outcomes of the model inference on a OCT image using the PSPNet decoder head. It is evident from the results that the models, initially designed for retinal vessel segmentation, fall short in achieving desirable performance for retinal layer segmentation. The low mIoU scores obtained across all models underscore the need for further refinement through fine-tuning.

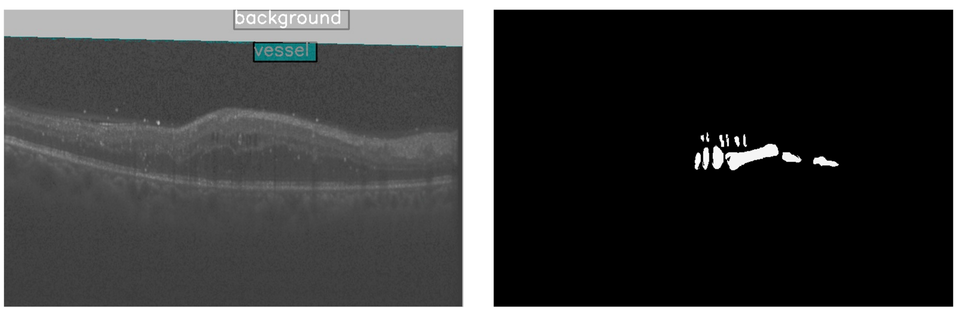

Figure 5: Model inference result with PSPNet decoder head. (Left) Preliminary inference result. (Right) Ground truth.

4.2 Fine-tuning

To enhance the model's performance, we undertook a systematic exploration, incrementally refining key training parameters. Recognizing the potential for improved convergence, we augmented the number of training iterations. This allowed the model to iteratively refine its parameters and achieve a more nuanced understanding of the data. Data augmentation techniques including leveraging random horizontal flipping and varying random crop sizes are applied to address the necessity for a robust model that could handle diverse input conditions. Acknowledging the class imbalance within the OCT dataset, particularly concerning fluid accumulation areas, we incorporated specialized loss functions. Dice loss and focal loss, renowned for their effectiveness in handling class imbalance, were applied to optimize segmentation results. Additionally, we tested different optimizer and applied l2-regularization to avoid overfitting. Table 2 summarizes the outcomes of our experimentation, highlighting the configurations that yielded the optimal model performance.

Our iterative refinement process culminated in the optimal configuration, where random crop training images to 320x320 without random flipping, utilizing cross-entropy loss for 400 iterations, SGD optimizer, and applying 0.0001 L2 regularization, achieved superior performance with an mIoU score of 0.7600 and an mDice score of 0.8430. Figure 6 illustrates the results of model inference on a consistent OCT image, comparing the outcomes between the preliminary model and the optimized version achieved through fine-tuning.

Table 2: Model configurations and evaluation results (fine-tuning)

# of iteration	Crop Size	Random Flip	Loss Function	Optimizer	L_2 reg.	mDice	mIoU
200	False	False	Cross Entropy	SGD	0.0005	0.7721	0.6855
400	False	False	Cross Entropy	SGD	0.0005	0.7173	0.6371
400	False	Horizontal	Cross Entropy	SGD	0.0005	0.7040	0.6263
400	$192 * 192$	False	Cross Entropy	SGD	0.0005	0.6951	0.6190
400	$224 * 224$	False	Cross Entropy	SGD	0.0005	0.7106	0.6299
400	$320 * 320$	False	Cross Entropy	SGD	0.0005	0.8401	0.7567
400	$320 * 320$	False	Dice Loss	SGD	0.0005	0.8395	0.7560
400	$320 * 320$	False	Focal Loss	SGD	0.0005	0.8280	0.7429
400	$320 * 320$	False	Cross Entropy	Adam	0.0005	0.4981	0.4963
400	$\mathbf{320 * 320}$	**False**	**Cross Entropy**	**SGD**	**0.0001**	**0.8430**	**0.7600**
400	$320 * 320$	False	Cross Entropy	SGD	0.001	0.8312	0.7464

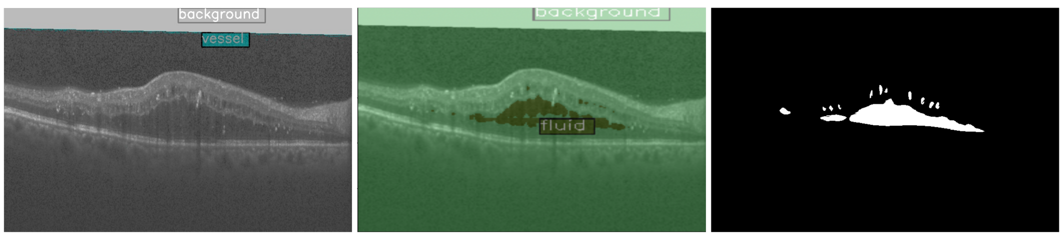

Figure 6: Comparative model inference on a OCT Image employing the preliminary and fine-tuned models.(Left) Preliminary inference result. (Middle) Fine-tuned inference result. (Right) Ground truth.

5 Conclusion

Our study has delved into the realm of transfer learning for medical image segmentation, with a specific focus on optimizing pre-trained semantic segmentation models for fluid segmentation in OCT images, addressing the nuances of DME. Through rigorous experimentation, we showcased that transfer learning, while a potent tool, requires meticulous customization to the target task.

Our initial findings revealed the limitations of off-the-shelf pre-trained models, emphasizing the necessity for fine-tuning to bridge the performance gap. We significantly enhanced segmentation accuracy as evaluated by the mIoU and mDice scores by utilizing a comprehensive set of strategies such as data augmentation, varied loss functions, and iterative model adjustments.

The significance of our work extends to ophthalmological diagnostics, where the accurate segmentation of retinal layers is crucial. The optimized model, fine-tuned for diabetic macular edema cases, exhibits promising results in segmenting fluid accumulation areas. This holds immense potential for advancing medical image analysis in optical coherence tomography, empowering clinicians with precise tools for early diagnosis and personalized treatment planning.

While our study contributes valuable insights, it is essential to acknowledge its limitations. The original dataset class imbalance, particularly concerning fluid accumulation areas, warrants further investigation. Future research avenues could explore the combination of cross-entropy and Dice loss by assigning different weights to them, aiming to strike a balance that accommodates the inherent data distribution challenges. Such endeavors would contribute to refining the model's robustness and generalizability across diverse medical imaging scenarios.

Link to code: https://github.com/nogibjj/yj_xq_661_final_project

References

[1] Ronneberger, Olaf, Philipp Fischer, & Thomas Brox *U-net: Convolutional networks for biomedical image segmentation.* Medical Image Computing and Computer-Assisted Intervention–MICCAI 2015: 18th International Conference, Munich, Germany, October 5-9, 2015, Proceedings, Part III 18. Springer International Publishing, 2015.

[2] J. Amin, M. Sharif, M. Yasmin, T. Saba, M. A. Anjum & S. L. Fernandes *A New Approach for Brain Tumor Segmentation and Classification Based on Score Level Fusion Using Transfer Learning.* Journal of Medical Systems, vol. 43, no. 11, Oct. 2019

[3] M. Ghafoorian et al., *Transfer Learning for Domain Adaptation in MRI: Application in Brain Lesion Segmentation.* arXiv:1702.07841 [cs], vol. 10435, pp. 516–524, 2017