Domain Transfer and Generalization of Image Segmentation Tasks Based on 2D Fourier Transform

Zhehui Zhang 12232381, Zhou Zhou 12332482, Junhong Cai 12332479, Yuwei Huang 12332473

Abstract—This paper implements a pioneering approach to image segmentation, focusing on domain transfer and generalization through 2D Fourier Transformations and federated learning methodologies. We introduce a three-fold strategy: (1) enhancing traditional image segmentation techniques using a 2D-UNet model trained on a specific domain and tested across multiple domains, (2) implementing domain visualization and style transfer via t-SNE and FedDG methodologies, emphasizing the 2D Fourier Transform, and (3) combining segmentation with style transfer to train a robust model capable of handling varied domains without direct access to their labels. This report outlines our systematic approach to these challenges, adhering to a structured project framework. Our focus is on implementing existing methodologies in a cohesive manner to understand their effectiveness in varying scenarios. The project is segmented into three interconnected sub-projects, each targeting a specific aspect of image segmentation and domain transfer, supported by comprehensive performance metrics including Dice coefficient, HD95, and ASSD.

Index Terms—Image Segmentation, Domain Transfer,Domain Generalization,2D Fourier Transform ,Federated Learning

I. Introduction

The rapidly evolving field of computer vision has seen significant advancements in image segmentation, domain transfer, and style transfer, greatly influenced by deep learning techniques. Our research presents an integrative approach that combines these elements, aiming to address the challenges in domain generalization and transfer for image segmentation tasks. The study is divided into three interrelated sub-projects, each focusing on a unique aspect of this challenge: enhanced image segmentation, domain visualization and style transfer, and the application of federated learning principles.

A. Image Segmentation

The issue with our project lies in the image segmentation task conducted on the FAZ dataset. We tested two different models and designed several distinct evaluation metrics. The problem is divided into two parts: firstly, using images from the 'train' directory of domain1 as training samples and masks as labels, a model is trained using supervised learning methods, and its segmentation performance is assessed on test samples from all domains 1-5. Secondly, an image segmentation model is trained using images and labels from domain1 and those obtained through domain transfer, and its performance is similarly evaluated on test samples from all domains 1-5. We tested two different models: Unet++ and nnUnet, and designed several different evaluation metrics, including the Dice coefficient, HD95, and ASSD.

B. Domain Visualization and Style Transfer

The second facet of our research delves into domain visualization and style transfer. Here, we utilize t-SNE for visualizing the relative distribution of multiple domains. The adoption of FedDG methodology allows for the execution of 2D Fourier Transforms on training samples across domains1-5, focusing on the amplitude and phase spectrum. This step is crucial in understanding the inherent differences and similarities between domains, thus aiding in effective style transfer. Our approach also includes parameter selection for transferring source images from domain1 to other domains, followed by an assessment of inter-domain distances.

C. Federated Learning

The concept of federated learning is intertwined in our approach, particularly in the combination of segmentation and style transfer. We train an image segmentation model using images and labels from domain1 and the transferred images obtained in the previous step. Notably, labels from domains2-5 are not utilized, adhering to the principles of federated learning where data privacy and locality are of paramount importance. This methodology demonstrates the feasibility and effectiveness of federated learning in multi-domain image segmentation tasks.

D. Summary

The core focus of our project is on the segmentation of fundus images, a critical task in medical imaging and ophthalmology. We have implemented a novel approach by applying 2D Fourier Transformations for transferring images from the source domain to target domains. This method is central to our study, as it enables us to explore the efficiency of domain adaptation in image segmentation.

A key aspect of our work is the verification of the model's improved performance after training with this domain transfer approach. By utilizing the 2D Fourier Transformations, we aim to enhance the model's generalization capabilities and robustness. This is particularly crucial in medical imaging, where the ability to accurately segment images across various domains (i.e., different types of fundus images) is essential for reliable diagnosis and treatment planning.

In addition to improving generalization and robustness, our project also incorporates the principles of federated learning. By doing so, we address the growing concern for data privacy in medical imaging. Federated learning allows us to train our models on decentralized data, ensuring that the local data's privacy is maintained. This is particularly important in medical contexts where patient data sensitivity is paramount.

Through this integrated approach, our project not only contributes to the advancement of image segmentation techniques in medical imaging but also demonstrates a practical application of federated learning in ensuring data privacy. The combination of domain transfer via 2D Fourier Transformations and federated learning presents a comprehensive strategy for addressing some of the key challenges in the field of medical image analysis.

II. Method

In this section, we will introduce our experimental methodology, which includes how we design image segmentation tasks, how we design domain transfer tasks, and so on. In this part, we will explore these issues in detail.

A. Image Segmentation Task

In the image segmentation task, we trained different models to test and compare their segmentation performance. Multiple evaluation metrics were designed and implemented to measure the effectiveness of the models. Among these, to better calculate the evaluation metrics, we also compared different binarization methods, solely to pursue the most suitable binarization effect.

a) Models Used:

Unet++

Unet++ is an advanced neural network architecture designed specifically for biomedical image segmentation. It is an expansion of the traditional U-Net architecture and includes a series of nested, densely connected U-Net structures. In this design, Unet++ introduces a series of skip connections from the upsampling path to the downsampling path, aimed at improving the gradient flow of the feature maps, reducing information loss, and enhancing the precision of segmentation. As can be seen in the image, Unet++ has multiple levels of feature representation, each connected by downsampling (pooling) or upsampling (deconvolution). These connections allow the network to capture details at different resolutions and maintain the integrity of high-resolution feature maps through skip connections. This type of network is typically used in medical image analysis that requires precise contour detection, such as tumor segmentation or organ segmentation. The architecture of Unet++ is shown in figure 1

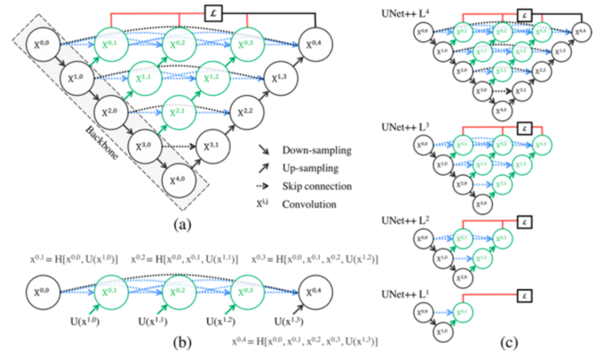

Fig. 1. This figure shows the architecture of Unet++.

nnU-Net

nnU-Net (no-new-Net) is an adaptive neural network framework specifically designed for medical image segmentation tasks. It automatically configures all critical aspects of the U-Net architecture, including preprocessing, network architecture, training strategy, and post-processing steps. nnU-Net uses the "data fingerprint" of the training data to derive the best network configuration and processing strategies, obtained through heuristic rules and inferred parameters, such as image resampling, normalization, batch size, and patch size. It also includes cross-validation for training the network and a set of empirical parameters, such as post-processing and ensemble strategies. Through this approach, nnU-Net is able to generate specific, optimized prediction models for different datasets, thereby achieving leading performance in various medical image segmentation challenges. The architecture of nnU-Net is shown in figure 2

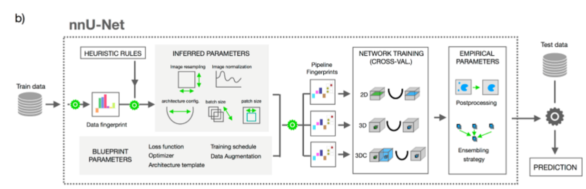

Fig. 2. This figure shows the architecture of nnU-Net.

b) Evaluation Metrics:

Our evaluation metrics are implemented by referencing the MedPy library. The reason for referencing instead of directly calling this library is that it is outdated and incompatible with our version of NumPy. We chose not to downgrade our NumPy version to avoid a chain reaction of compatibility issues. Therefore, by referencing the MedPy library, we made some modifications and implemented the following evaluation metrics. After modification, the methods we implemented can directly take two tensors on GPUs as input. In our implementation, we have also included the conversion and binarization of tensors. The metrics are as below.

$$\text{Dice} = \frac{2 * |X \bigcap Y|}{|X| + |Y|} \tag{1}$$

In the field of image segmentation, the Dice coefficient (also known as Dice similarity coefficient) is a commonly used evaluation metric for measuring the similarity between two samples. X and Y represent two datasets, typically denoting the segmentation results of an algorithm and the true segmentation, respectively. The Dice coefficient is used to quantify the degree of overlap between the predicted segmentation results and the true segmentation. A value of 1 indicates perfect overlap (i.e., the prediction is completely accurate), while a value of 0 indicates no overlap at all (i.e., the prediction is completely inaccurate). This metric is particularly important in medical image segmentation, as it effectively quantifies the performance of segmentation algorithms.

$$\text{D} = d(a, B) \bigcup d(b, A) \tag{2}$$

$$\text{HD95} = \text{95th percentile of D} \tag{3}$$

The term "d(a,B)" refers to calculating the distance from each point a in set A to the nearest point in set B. The "95th percentile of D" indicates the distance value at the 95th percentile in the list of distances D. This means that 95% of the points in set A have a distance to the nearest point in set B that is equal to or less than this value, making it a useful statistic for understanding the overall distribution of distances in a dataset.

$$\text{AD(A, B)} = \frac{1}{|A|} \sum_{a \in A} \min_{b \in B} d(a, b) \tag{4}$$

$$\text{ASSD} = \frac{AD(A, B) + AD(BA)}{2} \tag{5}$$

A and B are respectively the boundaries of two segmentations, then AD(A, B) represents the one-way average distance from surface A to surface B. ASSD works by calculating the average distance from each point on the two surfaces to the nearest point on the opposite surface. It is a symmetrical metric because it takes into account the distance from A to B as well as from B to A. The smaller the ASSD value, the closer the two surfaces are, indicating a more accurate image segmentation result.

$$\text{Jaccard(A,B)} = \frac{|A \bigcap B|}{|A \bigcup B|} \tag{6}$$

$$\text{Specificity} = \frac{TN}{TN + FP} \tag{7}$$

$$\text{Recall} = \frac{TP}{TP + FN} \tag{8}$$

$$\text{Precision} = \frac{TP}{TP + FP} \tag{9}$$

In the context of evaluation metrics, especially in classification tasks like image segmentation, the terms T, P, F, N typically stand for True, False, Positive, and Negative, respectively. They are used to denote the outcomes of a binary classification and are part of a confusion matrix.TP(True Positive) represents the instances correctly identified as positive, and the pattern continues likewise. The Jaccard coefficient is a statistical metric used to measure the similarity and diversity between sample sets. Specificity is used to measure the ability of a classifier or diagnostic test to correctly identify negative instances. Recall measures the ability of a classifier to correctly identify positive instances. Precision is a metric that measures the accuracy of a classifier or diagnostic test in correctly identifying positive instances.

c) Binarization Methods:

Most metrics require the input images to be binarized before calculation, and the choice of binarization threshold is very important. We have tried different threshold selection methods, including a manually set fixed threshold, mean thresholding, median thresholding, and Otsu's thresholding method. In the end, we adopted Otsu's thresholding method.

$$\sigma_b^2(t) = w_0(t) \cdot w_1(t) \cdot [\mu_0(t) - \mu_1(t)]^2 \tag{10}$$

$$t^* = argmax_t \sigma_b^2(t) \tag{11}$$

The Otsu thresholding method, also known as Otsu's algorithm, aims to find a threshold value t^*that maximizes the inter-class variance between the two categories of pixel values on either side of t.Here, $w_0$and $w_1$are the probabilities of the two categories, while $\mu_0$and $\mu_1$are the respective class means within each category.

B. Domain Transfer and Generalization

a) t-SNE Visualization of Domain Distributions: t-Distributed Stochastic Neighbor Embedding (t-SNE) is a powerful technique for dimensionality reduction and visualization of high-dimensional datasets. In the context of image segmentation, t-SNE can be used to visualize the relative distribution of multiple domains either in the image space or feature space. [3]This visualization aids in understanding the differences and similarities between various domains, which is crucial for domain adaptation and transfer learning in image segmentation tasks. And the visualization between different domain distribution is implemented by two separate features. The Fourier Transform domain offers a unique perspective on the data by focusing on the frequency components. By applying t-SNE to features obtained from the Fourier Transform of the images, it's possible to visualize and understand the frequency-based characteristics of different domains. This approach is particularly useful for identifying and differentiating domains based on their spectral properties.The results are as follows3 and 4.

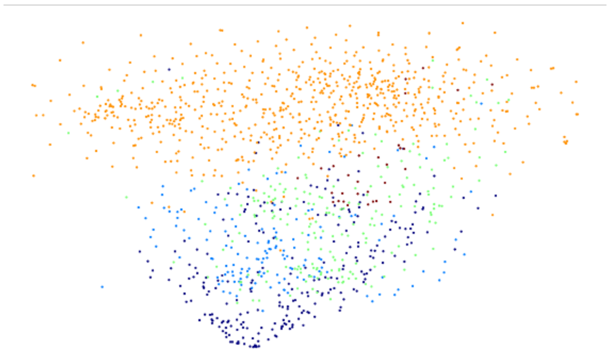

Fig. 3. This image visualizes the difference in the image space between domain distribution.

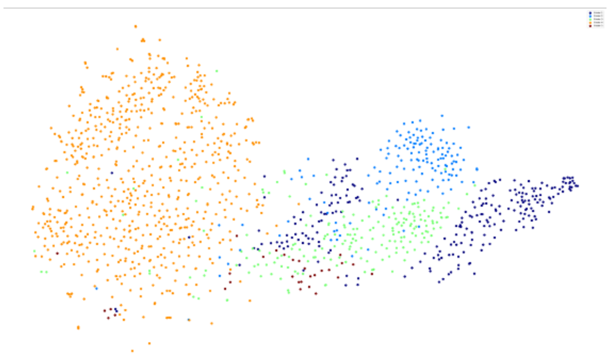

Fig. 4. This image visualizes the difference in the 2-D Fourier space between domain distribution.

b) FedDG and 2D Fourier Transformation on Training Samples: Applying the FedDG methodology combined with 2D Fourier Transformations on training samples from domains 1-5 can provide insightful information about the image's spectral properties. The Fourier Transform decomposes an image into its sine and cosine components, offering a view into the frequency domain. This process involves transforming the images to their amplitude and phase spectra, followed by gray-level adjustments and spectral centering. These transformations are critical in understanding the underlying patterns and features that may not be visible in the spatial domain. The application of such techniques can reveal unique characteristics of each domain, aiding in the development of more robust and domain-adaptive image segmentation models.The methodolody is just an implementation in [4]. The results is as follows.5.

c) Domain Transfer of Source Images from Domain1: Domain transfer involves adapting a source image from one domain (e.g., Domain1) to various target domains. This process requires careful selection of parameters to ensure effective style transfer while preserving the essential characteristics of the source images. The challenge lies in maintaining the integrity of the original image content while adapting to the stylistic elements of the target domains. This step is crucial for training a model that

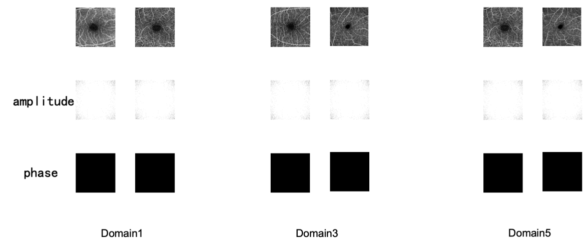

Fig. 5. Visualizes FedDG and 2D Fourier transformation on training samples

can generalize across different domains without having direct access to their labels. The selection of appropriate parameters for domain transfer is a key step in ensuring the success of this process.6.

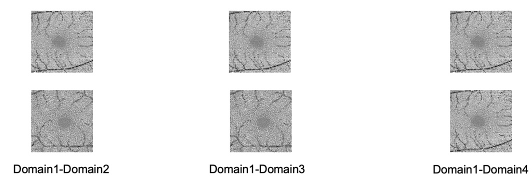

Fig. 6. Visualizes domain transfer from domain1 to other domains.

III. Experiments

To compare the effects of the models before and after applying domain transfer results for training, we designed comparative experiments. We trained two models, one before and one after incorporating the domain transfer results, and compared their performance across different domains. We designed comparative experiments for both Unet++ and nnU-Net. In terms of hyperparameter selection, we kept them consistent: we used the Adam optimizer, the loss function was BCE (Binary Cross-Entropy), and we set the number of training epochs to 60. And the result is shown in figure 7 and 8.

The domain transfer process was a critical aspect of our experimental setup. We applied domain transfer to our dataset with specific parameters, α and λ, set to 0.5 and 0.6 respectively. These values were chosen after

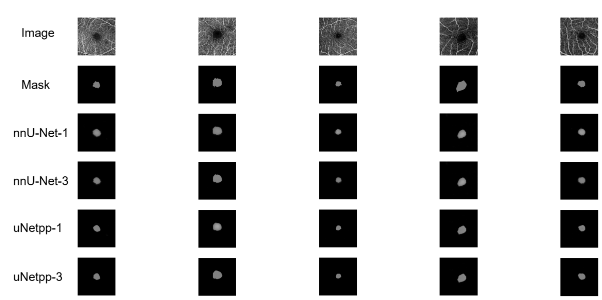

Fig. 7. This table shows the average evaluation metrics for the two models before and after incorporating domain transfer results, across five domains.

model	task	dice	hd95	assd	jaccard	specificity	recall	precision
nnU-Net	1	0.852768	8.211695	1.2175	0.753632	0.995979	**0.846175**	0.89444
	3	**0.866521**	**1.974904**	**0.333049**	**0.779254**	**0.997656**	0.843241	**0.923446**
uNet++	1	0.849591	2.415148	0.485448	0.75026	**0.998661**	0.785911	**0.952572**
	3	**0.877338**	**2.215453**	**0.410318**	**0.791241**	0.99817	**0.837083**	0.941061

Fig. 8. This image presents a comparison of the image segmentation effects of different models under varying conditions.

several trials, as they yielded the best results in terms of model performance across different domains. The domain transfer effectively augmented the training data, providing the models with a broader range of features and scenarios, which is evident in the improved segmentation outcomes. As can be seen, after the addition of domain transfer results, both types of models show a certain improvement in the performance of image segmentation tasks. Furthermore, during the experiment, we attempted to divide the training set into a 4:1 ratio of training and validation sets and designed an early stopping mechanism to monitor the training process of the model. However, due to the small size of the dataset, there was considerable fluctuation during training, regardless of which evaluation metric was chosen as the monitoring indicator. Therefore, we ultimately decided to remove this part from the final implementation.

IV. Conclusion

This study successfully demonstrates the integration of 2D Fourier Transformations and federated learning in enhancing image segmentation across various domains. By applying t-SNE for domain visualization and FedDG methodology for domain adaptation, significant improvements in segmentation performance were observed with Unet++ and nnU-Net models. Our approach, which aligns with the need for data privacy, shows promising potential in medical imaging, offering a robust solution for domain generalization and transfer learning. This research not only advances image analysis techniques but also holds substantial implications for medical diagnostics and treatment planning.

References

[1] Zhou, Z. et al. (2018) 'UNet++: A Nested U-Net Architecture for Medical Image Segmentation', in D. Stoyanov et al. (eds) Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support. Cham: Springer International Publishing (Lecture Notes in Computer Science), pp. 3–11. Available at: https://doi.org/10.1007/978-3-030-00889-5-1.

[2] Isensee, F. et al. (2021) 'nnU-Net: a self-configuring method for deep learning-based biomedical image segmentation', Nature Methods, 18(2), pp. 203–211. Available at: https://doi.org/10.1038/s41592-020-01008-z.

[3] Liu, Quande, et al. "Feddg: Federated domain generalization on medical image segmentation via episodic learning in continuous frequency space." Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2021.

[4] Zhuang, F., Qi, Z., Duan, K., Xi, D., Zhu, Y., Zhu, H., ... & He, Q. (2020). A Comprehensive Survey on Transfer Learning. \textit{Proceedings of the IEEE}, 109(1), 43-76. https://doi.org/10.1109/JPROC.2020.3004555

[5] Sun, C., Shrivastava, A., Singh, S., & Gupta, A. (2017). Revisiting Unreasonable Effectiveness of Data in Deep Learning Era. In \textit{Proceedings of the IEEE International Conference on Computer Vision} (pp. 843-852). https://doi.org/10.1109/ICCV.2017.97