Dual-Spatial Domain Generalization for Fundus Lesion Segmentation in Unseen Manufacturer's OCT Images

Shichen Liao, Tao Peng, Haoyu Chen, Tian Lin, Weifang Zhu, Fei Shi, Xinjian Chen, *Senior Member, IEEE*, and Dehui Xiang, *Member, IEEE*

Abstract*—*Objective: **Optical Coherence Tomography (OCT) images can provide non-invasive visualization of fundus lesions; however, scanners from different OCT manufacturers largely vary from each other, which often leads to model deterioration to unseen OCT scanners due to domain shift. *Methods:* To produce the T-styles of the potential target domain, an Orthogonal Style Space Reparameterization (OSSR) method is proposed to apply orthogonal constraints in the latent orthogonal style space to the sampled marginal styles. To leverage the high-level features of multi-source domains and potential T-styles in the graph semantic space, a Graph Adversarial Network (GAN) is constructed to align the generated samples with the source domain samples. To align features with the same label based on the semantic feature in the graph semantic space, Graph Semantic Alignment (GSA) is performed to focus on the shape and the morphological differences between the lesions and their surrounding regions. *Results:* Comprehensive experiments have been performed on two OCT image datasets. Compared to state-of-the-art methods, the proposed method can achieve better segmentation. *Conclusion:* The proposed fundus lesion segmentation method can be trained with labeled OCT images from multiple manufacturers' scanners and be tested on an unseen manufacturer's scanner with better domain generalization. *Significance:* The proposed method can be used in routine clinical occasions when an unseen manufacturer's OCT image is available for a patient.**

***Index Terms*—OCT image segmentation, domain generalization, orthogonal style space, graph semantic space.**

Manuscript received 12 March 2024; revised 8 April 2024; accepted 17 April 2024. Date of publication 25 April 2024; date of current version 22 August 2024. This work was supported in part by the National Nature Science Foundation of China under Grant 62371328, Grant 61971298, Grant 61771326, Grant 81871352, and Grant U20A20170, and in part by the National Key R&D Program of China under Grant 2018YFA0701700. *(Shichen Liao, Tao Peng, and Haoyu Chen equally contributed to this work.) (Corresponding author: Dehui Xiang.)*

Shichen Liao, Weifang Zhu, Fei Shi, and Xinjian Chen are with the School of Electronic and Information Engineering, Soochow University, China.

Tao Peng is with the School of Future Science and Engineering, Soochow University, China.

Haoyu Chen and Tian Lin are with Joint Shantou International Eye Center, Shantou University and the Chinese University of Hong Kong, China.

Dehui Xiang is with the School of Electronic and Information Engineering, Soochow University, Suzhou 215006, China (e-mail: xiangdehui@suda.edu.cn).

Digital Object Identifier 10.1109/TBME.2024.3393453

I. Introduction

OCT [1] has emerged as a powerful imaging modality for non-invasive visualization of retinal structures. It can provide high-resolution cross-sectional images of the retina for early detection and diagnosis of various ocular diseases. Due to the development of medical image segmentation [2], [3], [4], the segmentation of fundus lesions such as intraretinal cysts and subretinal fluid [5], [6] has been highly improved, as it plays a crucial role in quantitative assessment of ocular diseases.

Deep learning-based segmentation methods[7], [8] have achieved remarkable success by leveraging large-scale annotated datasets. However, the performance of these models often deteriorates when applied to unseen medical domains due to domain shift [9], [10] such as appearance and characteristics, physiological differences, temporal shift and acquisition protocols. Scanners vary across different hospitals or imaging centers from different manufacturers (shown in Fig. 1). Large domain shift hampers the generalization of segmentation models in real-world clinical practice.

To improve generalization, significant research efforts have been devoted to Unsupervised Domain Adaptation (UDA) and Domain Generalization (DG). UDA aims to mitigate the decline in generalization caused by distribution variations between labeled source domain data and unlabeled target domain data. UDA has been widely studied in the literature [11], [12], [13], and they employed adversarial learning to align the distributions between the source and target domains. Synthetic target domain images were generated to train a segmentation model and further mitigate the domain shift [14], [15]. Some reseachers [16], [17] proposed a feature-disentanglement style-transfer module to synthesize the target-like source images to mitigate domain shift. Though UDA methods have shown promising performance, their clinical applicability is limited due to the necessity of accessing to target domain data.

To overcome the limitation of UDA, researchers have introduced DG methods that solely rely on the source domain data. Some DG methods used randomization-based strategies [18], [19] to generate augmented input data by applying random transformations to the image-space, frequency-space, or feature space. Adversarial-based techniques [20], [21] have also been developed to maximize data diversity while simultaneously constrain its reliability. Additionally,

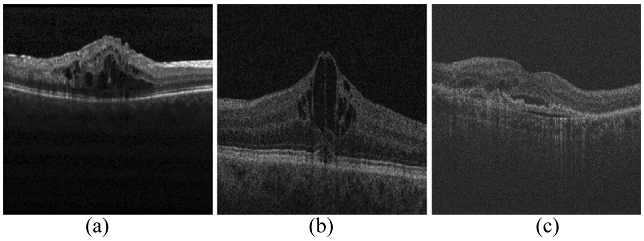

Fig. 1. Examples of OCT images from the RETOUCH dataset, captured by three different manufacturers' scanners (domains). (a) Cirrus; (b) Spectralis; (c) Topcon. The images exhibit significant variations, i.e., domain shift.

normalization-based methods [22], [23], commonly employed for specific tasks like pathological images, have shown effectiveness in improving generalization. Besides, at feature denoteation level, some well-generalized methods focused on invariant feature representation [24], [25], [26] and feature disentanglement [27], [28], [29], [30] to decompose the features of input samples into domain-invariant and domain-specific components. The objective of robust generalization models was to concentrate exclusively on the domain-invariant feature components while disregarde the domain-specific ones. Learning strategies have garnered considerable attention in addressing DG issues across diverse domains. These strategies could be broadly categorized into ensemble learning [31], [32], meta-learning [33], [34], [35], and self-supervised learning [36], which utilized generic learning paradigms to enhance the generalization performance.

In this paper, we propose a novel domain generalization method that combines style augmentation and semantic alignment based on dual-space constraints named OSSR and Graph Semantic Space Alignment (GSSA), respectively. OSSR takes into account that the target domain styles mostly exist in the marginal distribution of the source domain data. A marginal style sampling strategy is introduced and the marginal style is subsequently mapped to an orthogonal style space for reparameterization. This maximizes the difference between the generated pseudo-target domain style and the source domain style to recognize style-invariant features and overcome the limitation of traditional style transfer methods that mainly rely on the typical styles of source domain. Furthermore, we leverage the generated T-styles samples which have different styles but the same semantic content with source domain in GSSA. GSSA maps and aligns the labels and features to the latent graph semantic space to encourage the segmentation network to focus on shape of lesions. In addition, adversarial learning of graph convolutional networks is proposed to capture the underlying relationships among various lesions. The main contributions of our work are summarized as follows,

- An OSSR data augmentation method is proposed to apply orthogonal constraints in latent orthogonal style space to the sampled marginal styles for producing closely approximate the style of the potential target domain, such that domain generalization of the network can be improved in unseen target domains.
- A GAN is constructed to align the generated samples with the source domain samples in a graph semantic space for leveraging the high-level features of multi-source domains and potential T-styles samples in graph semantic space.
- GSA is performed to align features with the same label based on the semantic feature in the graph semantic space. A Feature Mapping Module (FMM) is constructed to map the features of the samples to the graph semantic space, and a Label Mapping Module (LMM) is also constructed to map the labels to the same graph semantic space. This allows us to effectively constrain the backbone to focus on the shape and the morphological differences between the lesions and their surrounding regions.
- Comprehensive experimental results have demonstrated the superiority of our method over state-of-the-art domain generalization techniques on two OCT image segmentation tasks.

II. Related Work

A. Domain Generalization for Medical Image Analysis

Unlike the domain adaptation task [12], [15], [17] which needs the images of the target domain, the domain generalization task [19], [23], [26] requires good generalization on the unseen target domain. The domain generalized segmentation methods aim to learn a model from a single or multiple source domains for unseen target domains. Existing DG methods can be roughly categorized into strategy-based methods, feature-based methods and data-based methods.

A typical strategy-based method is meta-learning [37]. Khandelwal et al. [33] employed few-shot learning to adapt the generalized model with very few examples from the unseen domain to new unseen data distribution. Kim et al. [38] presented a memory-guided domain generalization method that learned how to memorize a domain-agnostic and distinct information of classes. Wang et al. [35] designed a meta-sampling strategy to simulate the source/target domain shift and then developed a style-invariant model for image segmentation.

Feature-based methods usually deal with domain generalization by learning domain-invariant features [24], [27], [39]. Lai et al. [26] learned image features with knowledge originating from multi-source domains and handled the intra-domain variation by individually modeling the pixel and region relations within an image. Hu et al. [40] used the domain-discriminative feature embedded in the encoder to generate the domain code of each input image, which established the relationship between multiple source domains and the unseen target domain. However, it was challenging to distinguish domain-invariant features from domain-specific data.

Data-based generalized methods were usually based on different data augmentation strategies. Fick et al. [20] used CycleGAN [41] to enrich the training samples by transforming images to another style. Su et al. [19] sampled from a linear combination of random variables, which are location-scale distribution at the class level, to generate fused images for data augmentation. Some other linear-dependency generalized methods [28], [42], [43] were proposed in the feature space. However, the

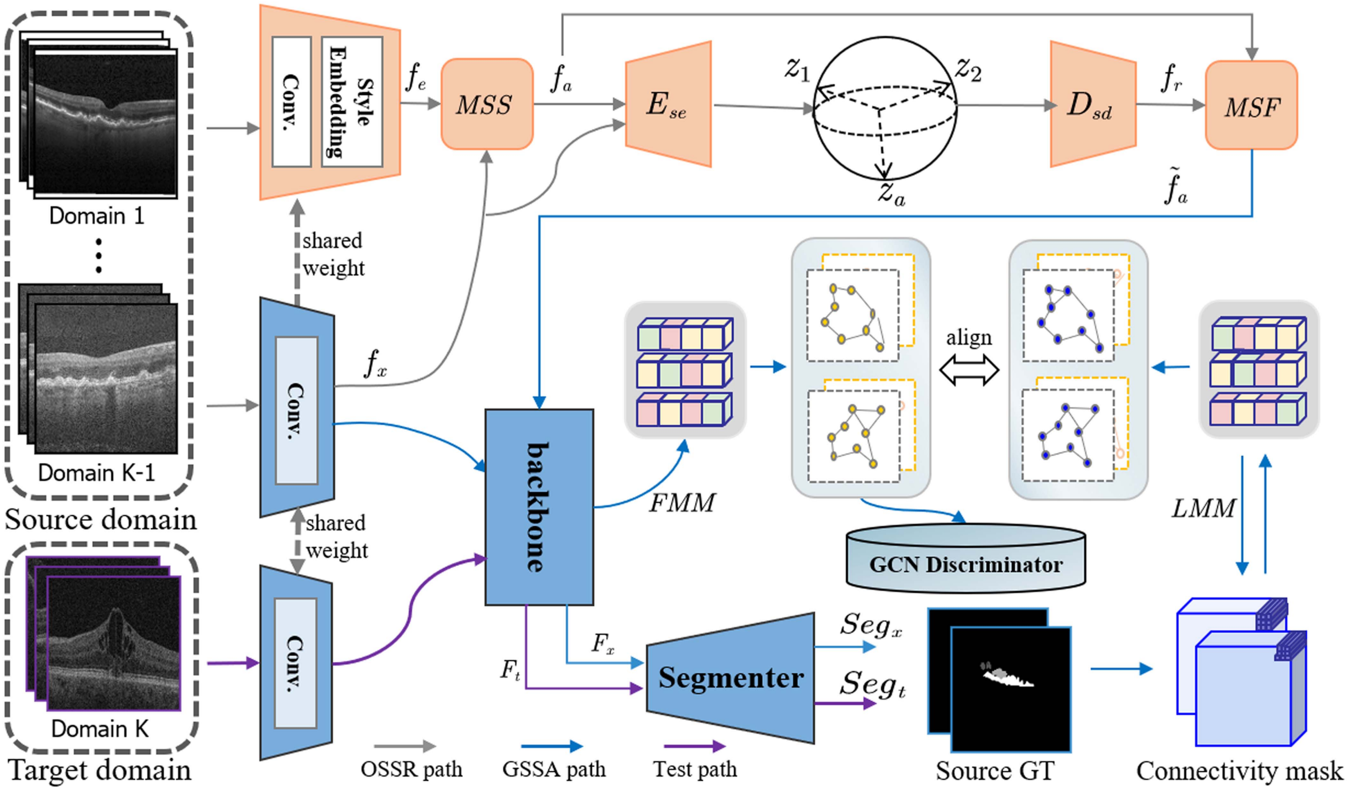

Fig. 2. Overview of our proposed method. The OSSR path serves as our auxiliary branch for enhancing marginal orthogonal style transfer in our framework, then the generated samples $\tilde{f}_a$ is used to perform graph semantic space alignment in the GSSA path with source domain samples.

effectiveness of these strategies depended on the ability to fit the distribution of the target domain. If the style of the target domain differed significantly from the source domain, it was difficult to achieve satisfactory results with random combinations.

B. Semantic Content Consistency

Semantic content consistency is a crucial aspect of domain generalization in medical image segmentation to ensure the semantic consistency of images across different domains. For instance, some data augmentation methods [21], [28] aimed to enhance the diversity of data and improve the model's generalization and semantic content consistency across different data domains. These methods utilized various data augmentation techniques such as rotation, scaling, mirroring, elastic deformation, etc., to transform training data and generate more domain variations, and to help the model learn more robust feature representations.

Some methods aimed to achieve semantic content consistency by aligning semantic features across different data domains. Rahman et al. [44] employed adversarial training frameworks using generative adversarial networks [45] to minimize the discrepancy between the generated images and synthetic images [30], [46], [47], which rely on external images to incorporate more diverse styles, and leverage content consistency across them. Wang et al. [48] designed a content consistency loss for balancing the adversarial relationship between the encoder and the auxiliary predictor to make the content features more style irrelevant. Kundu et al. [49] mitigated the inherent shift across domains through adversarial learning and explicitly imposed content consistency on the adapted target denoteation. To the best of our knowledge, all of these methods are based on convolutional neural networks.

III. Method

A. Problem Definition and Method Overview

For multiple-source domain generalized OCT image segmentation, K datasets are collected from different manufacturers' OCT scanners. The multiple-source domain dataset D_s consists of $K-1$ datasets, defined as $D_s = \{(x_i^k, y_i^k)_{i=1}^{N_k}\}_{k=1}^{K-1}$, where x_i^k denotes the i-th image from the k-th source domain and y_i^k denotes the corresponding label. The target domain $D_t = \{x_i\}_{i=1}^{N_K}$ is the K-th domain and is not visible during the training process.

The proposed dual-spatial constraints domain generalization model is shown in Fig. 2. In our OSSR module, a style augmentation strategy is presented based on a simple encoder-decoder structure in our Marginal Style Sampling (MSS) module at the first convolutional layer to enhance the encoding process by obtaining more discriminative styles across different source domains. In order to generate T-styles samples that may be close to the potential target domain, an orthogonal constraint is imposed to the diversified style features in a latent style space. The T-styles samples which incorporate the maximum difference in style from the source domains are fed into the segmentation network to augment the training samples. These T-styles samples encourage the backbone to reach a low inter-domain shift and improve the model's generalization in capturing domain-invariant features.

In the GSSA module, we combine graph adversarial learning and graph semantic alignment so that the changed style samples and the original source samples are supposed to share the same semantic content. To be specific, FMM is proposed to map the features extracted by the backbone to the graph semantic space and LMM is proposed to map the labels to the graph semantic space. Semantic features are then transformed into graph nodes by aggregating the information of the graph nodes through a graph network to sharpen the differences between samples with various styles by training a graph discriminator with a graph adversarial loss and to enhance the feature extraction capability of the encoder. In the following sections, we will elaborate the dual-space constraints for fundus lesion segmentation of unseen manufacturer's OCT images.

B. Orthogonal Style Space Reparameterization

1) Marginal Style Fusion: Previous style transfer methods mainly focused on transforming typical styles between existing source domains, but transferring to the style of the unseen target domain was challenging because the styles of the multiple-source domains largely differ from those of the unseen target domain. A small portion of the source domain samples may have similar styles to those of the unseen target domain, i.e., the style distribution of the target domain samples lies at the margins of the style distribution of the source domain samples. To address this issue, MSS is presented to generate new styles, called potential T-styles, which are potentially similar to those of the unseen target domain.

Style code $f_{i,e}^{k}$ is extracted from each source domain by embedding the output of the convolutional layer before the backbone. These style codes $f_{i,e}^{k}$ are then stored in a style bank with a capacity of size C. During each epoch update, the sampled marginal style code $f_{i,s}^{k}$, which exhibits the largest difference in the current epoch's input samples, is identified in the style bank as

$$f_{i,s}^{k} = MSS(f_{i,e}^{k}), \tag{1}$$

where $f_{i,e}^{k}$ is saved in style bank, then cosine similarity is used to calculate the marginal style $f_{i,s}^{k}$, which exhibits the largest difference between the other style codes in style bank.

After the marginal style codes f_s are sampled, Marginal Style Fusion (MSF) module is used to get the potential T-styles samples f_a which contain the style of f_s but the same semantic content of the source feature as

$$f_a = MSF(f_s, f_x) = \sigma(f_s)\left(\frac{f_x - \mu(f_x)}{\sigma(f_x)}\right) + \mu(f_s), \tag{2}$$

In $MSF(\cdot,\cdot)$ module, style transfer is performed by using AdaIN [50], and f_x denotes the feature map extracted by a convolutional layer, μ and σ denote the mean and standard deviation, respectively. This sampling strategy may produce some styles of the target domain samples. However, merely sampling the source samples would result in the same outcome as traditional style transfer methods. This is because the subsequent style transfer still relies on the styles present in the source domain and cannot generate most styles of the target domain that do not exist in the source domain samples.

Note that the style codes f_s obtained through resampling in the current epoch still denotes the styles presented in the source domain. Its sampled styles are close to the marginal distribution and lead to effective data augmentation for maximizing differences in style. This idea is similar to Mixup for style transfer across domains, which is randomly transferring samples between domains. However, a drawback of Mixup is that style transfer is only performed on samples from different source domains within the current epoch. The style obtained by any sample in different source domains is limited by the diversity of styles inherent in the source domains. When there is a large difference in style between an unseen target domain and the source domains, styles from the source domains may differ significantly from those of an unseen target domain. To address this issue, we sample from a Dirichlet distribution to generate potential styles not present in the known source domains, such that domain generalization of the network can be improved in unseen target domains. Furthermore, we choose to assign weights to the style codes f_s from a Dirichlet distribution. The weights $[\alpha_1, \ldots, \alpha_S]$ are defined as W, and the number of sampled style is defined as S. Eq. (2) can be re-expressed as

$$f_a = MSF\left(W \cdot Dirichlet\left(\frac{1}{S}\right) \cdot f_s, f_x\right), \tag{3}$$

where the operation of $MSF(\cdot,\cdot)$ is the same as Eq. (2), and $\frac{1}{S}$ is a concentration parameter. The Dirichlet distribution is capable of generating diverse style codes because it is a multidimensional distribution and each dimension represents a category. This makes it particularly useful for generating data with multiple discrete attributes.

2) Orthogonal Style Reparameterization: To enhance the diversity of potential T-styles, style reparameterization is further performed in a high-level orthogonal style space. When two vectors are orthogonal, they contain the maximum amount of independent information, and can result in diverse potential T-styles. Thus, by introducing an orthogonal constraint to the potential T-styles, we can transform the task of generating approximate styles for unseen target domain into an orthogonal constraint problem within the high-level orthogonal style space. Therefore, an encoder-decoder structure is designed to impose the orthogonal constraint on marginal style sampling.

Incorporating features with new styles is more beneficial for extracting feature from the mapping vectors in the high-level orthogonal style space, rather than solely relying on style codes to be fed into the style encoder E_{se}. Therefore, f_a and f_x are both fed into the E_{se} to obtain high-level features $z_{i,a}^{k} \in Z_a$ and $z_{i,x}^{k} \in Z_x$, where k represents k-th source domain, and i represents the i-th sample. The mapped features $z_{i,a}^{k}$ and $z_{i,x}^{k}$ are imposed to maximize the differences between style vectors of different domains. Therefore, $z_{i,a}^{1}, z_{i,a}^{2}, \ldots, z_{i,a}^{K-1}, z_{i,x}^{1}, z_{i,x}^{2}, \ldots, z_{i,x}^{K-1}$ are stacked into a matrix $\hat{\mathbf{Z}}_i \in \mathbb{R}^{(2K-2)\times L}$, where L is the length of each style vector. Since the learned style codes need to preserve the distinction, these style codes should be well-separated in the orthogonal style space, and the distance between styles of different domains should be maximized. To achieve this, a margin loss is proposed to constrain positive correlations between style codes. Each style codes extracted from a sample in the source domain forms a pair with the style codes extracted from another sample in a different source domain. The orthogonal loss of E_{se} can be expressed as

$$\mathcal{L}_{style} = \sum_{i=1}^{B} \left\| \frac{\hat{\mathbf{Z}}_{\mathbf{i}}^{T} \hat{\mathbf{Z}}_{\mathbf{i}}}{\|\hat{\mathbf{Z}}_{\mathbf{i}}^{T}\| \cdot \|\hat{\mathbf{Z}}_{\mathbf{i}}\|} - \boldsymbol{I}_{L\times L} \right\|^2 + \left\| \frac{\hat{\mathbf{Z}}_{\mathbf{i}} \hat{\mathbf{Z}}_{\mathbf{i}}^{T}}{\|\hat{\mathbf{Z}}_{\mathbf{i}}\| \cdot \|\hat{\mathbf{Z}}_{\mathbf{i}}^{T}\|} - \boldsymbol{I}_{(2K-2)\times(2K-2)} \right\|^2, \tag{4}$$

where $\boldsymbol{I}_{L\times L}$ denotes the identity matrix of size $L \times L$. The first term encourages orthogonality of the style codes, thereby maximizing the inter-class distance of distinctive style codes and the second term promotes independence of the style codes to reduce redundant information and maximize the information capacity of the high-level style space.

To ensure the effectiveness of style encoding, the codes produced by E_{se} are fed into our style decoder D_{sd}. D_{sd} discerns the distinctive style characteristics of different domains and reconstructs these high-level style codes into low-level style features of the source domain samples, which are then fed back into the input network. This ensures that E_{se} does not generate randomly high-dimensional style vectors in the high-level space. Consequently, further constraints are imposed on the preceding style generator, and the style consistency loss is defined as

$$\mathcal{L}_{sc} = \frac{1}{K-1} \sum_{k=1}^{K-1} \left(f_{i,e}^{k} - D_{sd}\left(z_{i,x}^{k}\right) \right)^2, \tag{5}$$

where D_{sd} denotes the style decoder in orthogonal latent space. It is important to note that the gradients in Eq. (5) need to be backpropagated to both the style encoder and the style decoder. In contrast, the gradients from the aforementioned orthogonal style loss in Eq. (4) are only backpropagated to the style encoder.

Our style consistency loss only operates on the styles of the source domain features and does not consider the potential T-styles f_a, because we aim to fine-tune f_a through the orthogonal constraints in the high-level orthogonal style space, to promote diverse potential T-styles and ensure distinctiveness among styles from different source domains. The entire style orthogonality module serves as a style transfer-based data augmentation strategy. Subsequently, the reparameterized orthogonally mixed style feature f_r is generated by feeding Z_a into the style decoder D_{sd} to map it back to the low-level features as $f_{i,r}^1, f_{i,r}^2, f_{i,r}^{K-1} \in Z_r$. Meanwhile, the style bank with $f_{i,r}^k$ is updated as a style code to enrich the variety of styles contained in the style bank. Finally, style fusion is employed to merge the reparameterized style feature code f_r with its source domain sample as

$$\hat{f}_a = MSF(f_r, f_a) \tag{6}$$

The source feature f_x and the marginal reparameterized orthogonally style mixed feature $\hat{f}_a$ are jointly fed into the backbone network, then the segmenter generates prediction map from pairs of samples $f_{i,x}^k$ and $\hat{f}_{i,a}^k$ with different styles but the same semantic labels. This strategy effectively enhances the backbone's ability to generalize well to samples with diverse domain shifts.

C. Graph Semantic Space Alignment

Our orthogonal space style reparameterization data augmentation strategy generates samples that closely resemble the style of the target domain. This greatly improves the semantic coding ability of the network. However, this improvement overly relies on the unequipped backbone's ability to capture style differences and lacks constraints on sample pairs with the same semantic content. Therefore, our GSSA module is proposed to help our backbone identify the semantic content of samples of different styles. Taking advantage of the style combination of different domains with the same content, our GSSA module is composed of two parts, one is graph convolutional adversarial learning that benefits from the powerful information aggregation capability of graph convolutional network, which is used to align the high-level features extracted by the backbone, and the other is the graph semantic alignment to connect the high-level features with the labels.

1) Graph Adversarial Network: For domain generalization tasks, neither the backbone nor the segmenter have encountered the target domain during training. In previous approaches, adversarial learning with convolutional neural network was applied to the whole high-level features, but it still lacked sufficient utilization of different channels of the high-level features and their relationships. The emergence of graph convolutional networks can overcome this drawback, as they can leverage the high-level features of multi-source domains and potential T-styles samples for domain generalization tasks.

In our GAN, an undirected fully connected graph $\Gamma = (\mathcal{V} \in \mathbb{R}^{C\cdot 2B\cdot(K-1)\times d}, \mathcal{E} \in \mathbb{R}^{C\cdot 2B\cdot(K-1)\times C\cdot 2B\cdot(K-1)}, \mathcal{A} \in \mathbb{R}^{C\cdot 2B\cdot(K-1)\times C\cdot 2B\cdot(K-1)})$ for all graph nodes is constructed in a batch, where C denotes the number of each sample's nodes, B is the batch size, d is the size of each node and $\mathcal{V}$ denotes the graph nodes, $\mathcal{E}$ denotes graph edges, $\mathcal{A}$ denotes the adjacency matrix. Specifically, for each feature $F_{i,m} \in F_m$ generated by our FMM, each $F_{i,m}$ is semantically mapped to a vector in the latent graph semantic space, denoted as node $\mathbf{v}_i \in \mathcal{V}$ in Γ. The edge $e_{i,j} \in \mathcal{E}$ denotes a connection between nodes v_i and v_j. The semantic similarity score $a_{i,j}$ in the adjacency matrix A corresponds to the pair of nodes (v_i, v_j).

For the first layer of the GAN, each graph node $\mathbf{v}_i \in \mathcal{V}$ is initialized with $F_{i,m} \in \mathbb{R}^{C\times H\times W}$ extracted by FMM. Each channel is flattened to form a node, resulting in a total of C nodes for a single sample input to the graph network. The semantic similarity scores $a_{i,j}^{(l)} \in (0, 1)$ are computed for all node pairs $(v_i, v_j) \in \mathcal{E}$ at l-th layer as follows,

$$a_{i,j}^{(l)} = f_{\text{edge}}^{(l)}\left(\mathbf{v}_i^{(l-1)}, \mathbf{v}_j^{(l-1)}\right) \tag{7}$$

where $f_{\text{edge}}^{(l)}(\cdot)$ denotes the similarity, and $\mathbf{v}_i^{(l-1)}$ denotes the features of node v_i in $l-1$th layer of the GAN.

Self-connections are added to the nodes in the graph and the obtained similarity scores are normalized as follows,

$$\mathcal{A}^{(l)} = M^{-\frac{1}{2}} \left(\hat{\mathcal{A}}^{(l)} + I \right) M^{-\frac{1}{2}} \tag{8}$$

where M denotes the degree matrix, I denotes the identity matrix, and $\hat{\mathcal{A}}$ denotes the unnormalized adjacency matrix.

By multiplying the obtained adjacency matrix $\mathcal{A}$ with $\mathcal{V}$, the aggregated node features $F_{\text{agg}} \in \mathbb{R}^{C \cdot 2B \cdot (K-1) \times HW}$ that integrate features from all samples can be obtained by

$$F_{i,\text{agg}}^{(l)} = \sum_{j=1}^{n} a_{ij}^{(l)} \cdot f_{v_j}^{(l-1)} \tag{9}$$

where $F_{i,\text{agg}}^{(l)}$ denotes the aggregated feature of the i-th node v_i, which is connected to all other nodes. Therefore, it is multiplied by the adjacency matrix coefficients $a_{ij}^{(l)}$ between the other nodes and this node, and $f_{v_j}^{(l-1)}$ denotes the feature of v_j from the previous layer.

Finally, the graph nodes mapping from F_m with the graph nodes from the aggregated feature F_{agg} to update $\hat{\mathcal{V}}$ are concatenated and fed into the graph convolutional discriminator. A binary classification loss is defined to enable the graph convolutional discriminator to distinguish whether the input graph nodes come from the source domain or from the potential target domain generated by style transfer as

$$\begin{aligned}\mathcal{L}_{gcn} = -\sum_{v_i, v_j \in \mathcal{V}} &(\log\left(D_{gcn}(cat(f_{v_i}, F_{i,agg}))\right) \\ &- \log\left(1 - D_{gcn}(cat(f_{v_j}, F_{j,agg}))\right)) \end{aligned} \tag{10}$$

where $D_{\text{gcn}}(\cdot)$ denotes the graph convolutional discriminator, $\text{cat}(\cdot,\cdot)$ denotes the concatenation operation, v_i denotes the graph nodes from the source domain, and v_j denotes the graph nodes mapped from samples with the potential T-styles. On the other hand, an adversarial loss is also defined to deceive the graph convolutional discriminator as

$$\mathcal{L}_{adv} = -\sum_{v_j \in \mathcal{V}} \log\left(1 - D_{gcn}(cat(f_{v_j}, F_{j,agg}))\right) \tag{11}$$

Note that this loss is used to optimize the first convolutional layer, backbone, and FMM, not the graph convolutional discriminator.

2) Graph Semantic Alignment: To constrain the semantic mapping and enhance the ability of GAN to capture semantics of graph nodes during the process of graph adversarial learning, GSA is performed to align features with the same label based on the semantic feature in F_x. It can be solved by two reversible processes, one process transforms the high-level feature to connectivity label Y, while the other process performs the reverse mapping from connectivity label to the high-level feature as $F_x \rightleftharpoons Y$. Due to a significant difference of Y and F_x, an intermediate graph semantic space V is constructed, and thus the reversible process becomes $F_x \rightleftharpoons V$, followed by the process $V \rightleftharpoons Y$, such that the feature of lesions can be utilized through the mutual conversion of Y and the graph semantic space V.

In GSA module, FMM is implemented as a downsampling network, which consists of a combination of convolution and downsampling operations to map F_x to a graph semantic vector V_1, instead of directly mapping F_x to the label, because it is not feasible to map the label back to the features F_x extracted by the backbone in this reversible process, and the dimensions of F_x and the label are significantly different. While the high-level feature F_x contains rich information, the label only provide such as shape-related features. Therefore, we aim to fully utilize the information of the label and extract shape-related content from the features F_x. This allows us to effectively constrain the backbone to focus on the shape and the morphological difference between lesions and their surrounding tissues. Consequently, it enables accurate lesion segmentation, without being limited by domain shift. Furthermore, LMM consists of an autoencoder. The encoder of LMM maps to a graph semantic vector V_2 and the decoder of LMM is used to reconstruct the label from V_2. $l_{2,1}$ is imposed as prior of structured sparsity in matrices to make the semantic feature mapped by the autoencoder more comprehensive. Therefore, a graph semantic reconstruction loss can be defined as

$$\mathcal{L}_{gsr} = ||Y_r' - Y_c||_2 + ||V_2||_{2,1}, \tag{12}$$

where Y_r' is the reconstructed connectivity map, and Y_c is the connectivity map in which the number of channels is eight times of segmentation categories, i.e. each category has eight corresponding channels to denote its neighboring regions.

Simultaneously, leveraging the matrix V_2 for its effectiveness, to align the generated V_1 from the FMM with V_2, Jensen-Shannon divergence (JSD) between the corresponding posterior probabilities P and Q of matrices V_1 and V_2 are used as the loss for graph semantic consistency,

$$\mathcal{L}_{lsc} = JSD\left(P; R\right) = \frac{1}{2}\left(D_{\text{KL}}[P||Q] + D_{\text{KL}}[R||Q]\right) \tag{13}$$

where $Q = (P+R)/2$ denotes the average probability between the original samples and the stylized samples. D_{KL} denotes Kullback-Leibler divergence between the posterior probabilities P, R and Q. JSD constrains the invariant semantic feature between the two graph semantic spaces.

Dice loss and cross-entropy loss are used to train the framework,

$$\mathcal{L}_{seg} = \mathcal{L}_{mse}(Y_c', Y_c) + \mathcal{L}_{ce}(\tilde{P}, Y) + \mathcal{L}_{dice}(\tilde{P}, Y), \tag{14}$$

where $\mathcal{L}_{mse}$, $\mathcal{L}_{ce}$ and $\mathcal{L}_{dice}$ are the mean squared error loss, the cross-entropy loss and the dice loss, respectively. Y_c' is the predicted connectivity map. $\tilde{P}$ is the predicted segmented result. Y is the segmentation label. In line with [51], the bilateral voting module and the region-guided channel aggregation module are used to get the segmentation prediction,

$$\tilde{P}(x, y) = \max\{Y_{c,i}'(x, y) \times Y_{c,7-i}'(x + a, y + b)\}_{i=0}^{7}, \tag{15}$$

where i is the i-th channel of the Bicon map, $a, b \in \{0, \pm 1\}$ denote the location offsets of neighboring pixels.

TABLE I
DETAILS ABOUT THE TWO DATASETS

Dataset	Domain No.	Size	Volume(Slices)	Scanners/Clinical Centers	Class

Fluid	Domain 1
	Domain 2
	Domain 3
		512 × 1024
		512 × 496
		512 × 650
			24(1568)
			24(711)
			22(1106)
				Cirrus, Zeiss
				Spectralis, Heidelberg
				T-1000 and T-2000, Topcon
					IRF,SRF,PED

Drusen	Domain 1
	Domain 2
	Domain 3
		512 × 496
		1000 × 512
		1024 × 1177
			100(1474)
			100(1337)
			100(825)
				Spectralis, Heidelberg/Z-Lab
				Bioptigen/VIP
				3DOCT-2000, Topcon/JSIEC
					Drusen

Overall, the total loss is defined as

$$\mathcal{L}_{total} = \mathcal{L}_{style} + \mathcal{L}_{sc} + \mathcal{L}_{gcn} + \mathcal{L}_{adv} + \mathcal{L}_{gsr} + \mathcal{L}_{lsc} + \mathcal{L}_{seg} \tag{16}$$

IV. EXPERIMENTS

In this section, we tested the performance of the fundus lesion segmentation framework by performing comprehensive experiments on two benchmark datasets. Experiments and results were reported as follows.

A. Dataset

The proposed method was evaluated on two OCT image datasets: a publicly fundus fluid segmentation dataset RETOUCH [52], and a drusen segmentation dataset from three clinical centers. The descriptions of the datasets are shown in Table I. The fundus image datasets were collected from different clinical centers. The heterogeneity of fundus OCT images across different domains was primarily due to different manufacturer scanners.

1) Fluid Segmentation Dataset: OCT images were collected from three different manufacturers' OCT scanners, which were regarded as three domains. This dataset consisted of 70 OCT volumes, where 24 volumes were acquired with the Cirrus scanner (Zeiss), 24 volumes were acquired with the Spectralis scanner (Heidelberg), and 22 volumes were acquired with the T-1000 and T-2000 scanners (Topcon). For each volume, there were 128, 49, and 128 B-scans with size of 512 × 1024, 512 × 496, and 512 × 650 for Cirrus, Spectralis and Topcon, respectively. In this dataset, three different fluid types, i.e., the intraretinal fluid (IRF), subretinal (SRF), and PED (pigment epithelial detachments), were manually annotated. All B-scans were randomly cropped around the lesion with size of 512 × 512.

2) Drusen Segmentation Dataset: OCT images were also collected from three distinct clinical centers, in which OCT scanners were made from three different manufacturers and therefore regarded as three domains. Each domain consists of 100 patients, with domain 1 images acquired from Z-lab [53] (Spectralis, Heidelberg), domain 2 images with age-related macular degeneration acquired from Vision and Image Processing (VIP) Laboratory of Duke University [54] (Bioptigen Inc.), and domain 3 images acquired from Joint Shantou International Eye Center (JSIEC), Shantou University and the Chinese University of Hong Kong (3DOCT-2000, Topcon). OCT volumes were with 128, 49, and 128 B-scans with sizes of 512 × 496, 1000 × 512, and 1024 × 1177 for domain 1, domain 2, and domain 3, respectively. All B-scans were randomly cropped around the lesion with size of 512 × 512. Manual annotation was performed by an experienced eye doctors.

B. Implementation Details

All models were tested on one NVIDIA GTX 3090 GPU. The entire network was trained for 200 epochs, with a learning rate of 0.001. The Adam optimizer was used to train the segmentation model. The batch size was determined to 1. E_{se} consists of convolutional blocks, each accompanied by a down-sampling layer to decrease the resolution and the output of E_{se} was flattened to obtain a latent style code with a length of 256. D_{sd} consisted of three fully connected layers with respective sizes of 256, 512, 512 and 64. In our FMM module, the number of nodes as well as the output channel number of FMM for each sample was set to 90, and the length of feature vector of each node was set to 256. The autoencoder of LMM consisted of convolutional layers, followed by ReLU activation, and downsampling in the encoder/upsampling in the decoder. Data-augmentation was performed based BigAug [55].

C. Experimental Results and Analysis

1) Comparative Methods and Evaluation Metrics: The lower bound (Inter-domain) refers to training the encoder-decoder segmentation network with all source domains and then directly testing on the unseen target domain, while the performance was poor due to the distribution discrepancy between the source (train) data and target (test) data. The upper bound (Intra-domain) means that each single domain was respectively trained and tested using 4-fold cross-validation. The proposed method was compared with several state-of-the-art methods. Data manipulation-based methods include **Mixup** [56] and **Cutmix** [57], which operated at the image and pixel level, respectively. **DualNorm** [18], **CDDSA** [28] and **MixStyle** [58] transformed image to another style. **DCAC** [40] used dynamic convolutions to train the model. **DoFE** [27] and **UniSeg** [59] learned domain-invariant features using prior of different domains and the specific features by integrating the universal prompt of each domain, respectively. **DCANet** [42] and **TriD** [60] were based on feature decomposition and recomposition. **RobustNet** [61] and **SANSAW** [62] aligned the distributions of the different domains. The segmentation performance was evaluated using the Dice similarity coefficient (Dice).

2) Experiments on Fluid Segmentation: Table II presents the evaluation results in terms of Dice scores for IRF, SRF, PED

TABLE II
QUANTITATIVE COMPARISON OF DIFFERENT METHODS IN FLUID SEGMENTATION DATASET

Methods	Domain 1	Domain 2	Domain 3	Avg(Dice)
Upper bound	61.45	66.65	58.53	62.21
Lower bound	21.79	13.36	23.98	19.71
DCAC [40]	44.47	55.12	55.94	51.84‡
DoFE [27]	41.83	53.66	61.85	52.45‡
TriD [60]	42.08	56.15	59.41	52.55‡
UniSeg [59]	46.01	55.59	58.99	53.53‡
DCANet [42]	48.63	55.35	57.14	53.71‡
CDDSA [28]	50.83	55.83	58.88	55.18‡
DualNorm [18]	53.57	58.24	56.51	56.11‡
SANSAW [62]	56.25	49.65	64.60	56.83‡
MixStyle [58]	55.93	56.47	63.78	58.73‡
Mixup [56]	56.10	57.10	63.09	58.76‡
Cutmix [57]	56.88	56.86	63.30	59.01‡
RobustNet [61]	56.88	57.07	64.32	59.42‡
Ours	**57.55**	**59.74**	**64.73**	**60.67**

‡ denotes $p < 0.001$ of paired t-test.
The best results are highlighted using bold and the second results are highlighted using underline, respectively.

TABLE III
QUANTITATIVE COMPARISON OF DIFFERENT METHODS IN DRUSEN SEGMENTATION DATASET

Methods	Domain 1	Domain 2	Domain 3	Avg(Dice)
Upper bound	80.98	82.58	71.61	78.39
Lower bound	69.55	73.12	60.17	67.18
DCAC [40]	78.68	76.26	47.21	67.38‡
TriD [60]	74.98	79.98	58.44	71.13‡
DCANet [42]	79.16	74.03	64.41	72.53‡
DualNorm [18]	79.67	79.12	59.35	72.71‡
RobustNet [61]	78.01	79.32	65.91	74.41‡
DoFE [27]	77.73	81.15	65.84	74.91‡
MixStyle [58]	78.33	80.37	66.17	74.96‡
CDDSA [28]	76.06	81.08	69.24	75.46‡
UniSeg [59]	78.63	81.61	66.06	75.43‡
Mixup [56]	79.31	81.43	66.73	75.82‡
Cutmix [57]	78.59	80.61	69.22	76.14‡
SANSAW [62]	77.89	80.95	70.11	76.32‡
Ours	**79.80**	**82.06**	**70.14**	**77.33**

‡ denotes $p < 0.001$ of paired t-test.
The best results are highlighted using bold and the second results are highlighted using underline, respectively.

segmentation. The upper bound achieved the highest performance, with an average Dice score of 62.21 across the three domains. In contrast, the lower bound only achieved an average Dice score of 19.71 due to the large domain shift. Compared to previous methods, our approach achieved a performance improvement of 1.25 over RobustNet, indicating that it effectively improved performance in both data augmentation strategies and semantic alignment strategies. Fig. 3 shows a visual comparison of our proposed method with 12 previous methods on three target domains. The results demonstrated that our proposed method achieved superior segmented results with better boundaries to ground truth. In contrast, the previous DG methods exhibited a higher occurrence of over-segmented and under-segmented regions. $p < 0.001$ of paired t-test showed that the superiority of our method for fluid segmentation was statistically significant.

3) Experiments on Drusen Segmentation: Table III reports the segmentation performance on the drusen segmentation dataset. The upper bound demonstrated the highest performance and achieved an average Dice score of 78.39 across the three domains, and the lower bound only reached an average Dice score of 67.18. The performance decline was more than 10 in terms of average Dice score, indicating a significant domain shift among the different domains. Among the previous methods, some DG methods achieved great improvements compared with the lower bound, and data augmentation strategies, such as CutMix and Mixup, yielded improvements of 1.19 and 0.87, respectively. SANSAW surpassed all other previous methods with a Dice score of 76.32. CDDSA and UniSeg had a similar segmentation performance with average Dice score of 75.46 and 75.43, respectively. Compared to SANSAW, the Dice score of our method improved by 1.01. $p < 0.001$ of paired t-test showed that the superiority of our method for drusen segmentation was statistically significant.

4) Analysis of Different Methods: Note that the performance of each previous method varied across different target domains due to differences in model generalization capabilities. For example, although SANSAW achieved good results on the drusen segmentation dataset, its performance significantly declined on the second domain of the fluid segmentation dataset. Due to significant intra-domain variations, DoFE resulted in unreliable domain-invariant knowledge extraction, and it led to poor performance on the first domain of the fluid segmentation dataset. RobustNet demonstrated notable improvements in whitening operations for multi-class tasks, but performed less effectively on binary segmentation tasks compared to data augmentation techniques. While traditional Mixup and Cutmix performed admirably across different domains, they were inferior to our method based on dual-spatial constraints.

D. Ablation Study

The proposed method consists of three primary components: Orthogonal Style Space Reparameterization (OSSR), Graph Adversarial Network (GAN), and Graph Semantic Alignment (GSA). Ablation experiments were performed on the fluid segmentation dataset to evaluate each component. Baseline was DeepLabv3+[63] with ResNet101 but without batch normalization.

1) Analysis of OSSR: Compared to the lower bound, the baseline achieved a noticeable improvement in the network's generalization and the average Dice score improved by 37.63. As shown in the 4th row in Table IV, all the Dice scores of the three domains were improved and the average Dice score improved from 57.34 to 58.84. This indicates that OSSR applied in different source domains yielded potential T-styles samples for domain generalization.

2) Analysis of GAN and GSA: Building upon the generation of new samples using OSSR, we conducted ablation experiments on GAN and GSA. The 5th and 6th rows in Table IV denote the quantitative results using graph adversarial network and graph semantic alignment, respectively. On the basis of OSSR, the average Dice scores for GAN and GSA improved

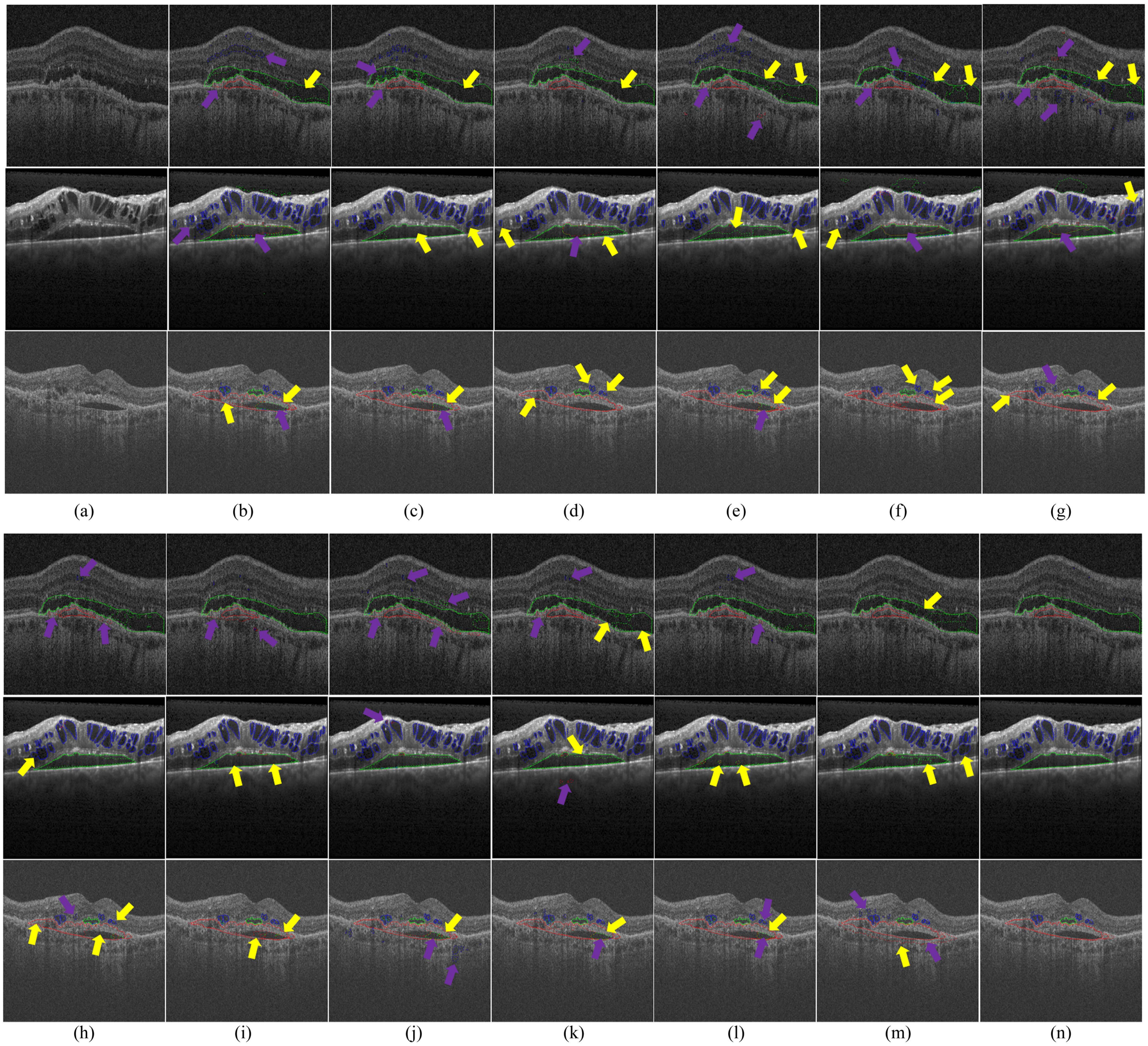

Fig. 3. Visual comparison in three domains between different DG methods for fluid segmentation. The solid line denotes ground truth, and the dash line denotes the prediction. The yellow arrow points to under-segmentation. The purple arrow points to over-segmentation. (a) Original image; (b) DCAC; (c) DoFE; (d) TriD; (e) UniSeg; (f) DCANet; (g) CDDSA; (h) DualNorm; (i) SANSAW; (j) MixStyle; (k) Mixup; (l) Cutmix; (m) RobustNet; (n) ours.

TABLE IV
ABLATION EXPERIMENTS OF THE FLUID DATASET

Methods	Domain 1	Domain 2	Domain 3	Avg(Dice)
Upper bound	61.45	66.65	58.53	62.21
Lower bound	21.79	13.36	23.98	19.71
Baseline	52.30	57.63	62.11	57.34
+OSSR	55.09	58.26	63.18	58.84
+OSSR+GAN	56.64	58.56	63.78	59.66
+OSSR+GSA	57.06	59.52	64.24	60.27
Ours	57.55	59.74	64.73	60.67

from 58.84 to 59.66 and 60.27, respectively, across the three target domains. The results indicated that through the auxiliary training of the two modules, the network was able to capture the semantic information of low-dimensional features that had different styles but the same semantics and also extracted the domain-invariant features. In addition, the fusion of GAN and GSA in the graph semantic space also improved the Dice score of each domain and the average Dice score reached 60.27.

V. CONCLUSION

To achieve generalization to unseen domains, a dual-spatial constrained segmentation network is proposed for fundus lesion segmentation. OSSR generates source domain samples with potential T-styles in the orthogonal style space to simulate the potential target domain styles, to mitigate domain shift. GAN and GSA can capture intra-domain semantic features and leverage cross-domain intrinsic relationships within the label. This guides the prediction process and improves segmentation performance. Our method has achieved better results on two benchmark datasets for fundus OCT image segmentation. In future work, we plan to collect more OCT images from more manufacturers' scanners and focus on multi-domain generalization of fundamental models to benefit our OCT image segmentation task, such that a universal OCT image segmentation model can be applied to a wider range of OCT image segmentation scenarios.

References

[1] D. Huang et al., "Optical coherence tomography," *Science*, vol. 254, no. 5035, pp. 1178–1181, 1991.

[2] D. Shen, G. Wu, and H.-I. Suk, "Deep learning in medical image analysis," *Annu. Rev. Biomed. Eng.*, vol. 19, pp. 221–248, 2017.

[3] G. Litjens et al., "A survey on deep learning in medical image analysis," *Med. Image Anal.*, vol. 42, pp. 60–88, 2017.

[4] X. Xie et al., "A survey on incorporating domain knowledge into deep learning for medical image analysis," *Med. Image Anal.*, vol. 69, 2021, Art. no. 101985.

[5] M. Ritter et al., "Intraretinal cysts are the most relevant prognostic biomarker in neovascular age-related macular degeneration independent of the therapeutic strategy," *Brit. J. Ophthalmol.*, vol. 98, no. 12, pp. 1629–1635, 2014.

[6] M. Veckeneer et al., "Persistent subretinal fluid after surgery for rhegmatogenous retinal detachment: Hypothesis and review," *Graefe's Archive Clin. Exp. Ophthalmol.*, vol. 250, pp. 795–802, 2012.

[7] J. Hao et al., "Uncertainty-guided graph attention network for parapneumonic effusion diagnosis," *Med. Image Anal.*, vol. 75, 2022, Art. no. 102217.

[8] M. Wang et al., "Self-guided optimization semi-supervised method for joint segmentation of macular hole and cystoid macular edema in retinal OCT images," *IEEE Trans. Biomed. Eng.*, vol. 70, no. 7, pp. 2013–2024, Jul. 2023.

[9] Y. Luo et al., "Taking a closer look at domain shift: Category-level adversaries for semantics consistent domain adaptation," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2019, pp. 2507–2516.

[10] K. Stacke et al., "Measuring domain shift for deep learning in histopathology," *IEEE J. Biomed. Health Informat.*, vol. 25, no. 2, pp. 325–336, Feb. 2021.

[11] Y. Sun, D. Dai, and S. Xu, "Rethinking adversarial domain adaptation: Orthogonal decomposition for unsupervised domain adaptation in medical image segmentation," *Med. Image Anal.*, vol. 82, 2022, Art. no. 102623.

[12] Z. Zheng, R. Li, and C. Liu, "Learning robust features alignment for cross-domain medical image analysis," *Complex Intell. Syst.*, vol. 10, pp. 2717–2731, 2023.

[13] T. Ilyas et al., "Enhancing medical image analysis with unsupervised domain adaptation approach across microscopes and magnifications," *Comput. Biol. Med.*, vol. 170, 2024, Art. no. 108055.

[14] X. Chen et al., "Anatomy-regularized representation learning for cross-modality medical image segmentation," *IEEE Trans. Med. Imag.*, vol. 40, no. 1, pp. 274–285, Jan. 2021.

[15] H. Liu et al., "Learning site-specific styles for multi-institutional unsupervised cross-modality domain adaptation," 2023, *arXiv:2311.12437*.

[16] X. Chen et al., "Diverse data augmentation for learning image segmentation with cross-modality annotations," *Med. Image Anal.*, vol. 71, 2021, Art. no. 102060.

[17] Z. Su et al., "Mind the gap: Alleviating local imbalance for unsupervised cross-modality medical image segmentation," *IEEE J. Biomed. Health Informat.*, vol. 27, no. 7, pp. 3396–3407, Jul. 2023.

[18] Z. Zhou et al., "Generalizable cross-modality medical image segmentation via style augmentation and dual normalization," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2022, pp. 20856–20865.

[19] Z. Su et al., "Rethinking data augmentation for single-source domain generalization in medical image segmentation," in *Proc. AAAI Conf. Artif. Intell.*, 2023, pp. 2366–2374.

[20] R. H. Fick et al., "Domain-specific cycle-gan augmentation improves domain generalizability for mitosis detection," in *Proc. Int. Conf. Med. Image Comput. Comput.- Assist. Intervention*, 2021, pp. 40–47.

[21] J. Lyu et al., "AADG: Automatic augmentation for domain generalization on retinal image segmentation," *IEEE Trans. Med. Imag.*, vol. 41, no. 12, pp. 3699–3711, Dec. 2022.

[22] J. Xiong et al., "Improve unseen domain generalization via enhanced local color transformation," in *Proc. 23rd Int. Conf. Med. Image Comput. Comput.–Assist. Interv.*, 2020, pp. 433–443.

[23] M. Salvi et al., "Generative models for color normalization in digital pathology and dermatology: Advancing the learning paradigm," *Expert Syst. Appl.*, vol. 245, 2024, Art. no. 123105.

[24] K. Muandet, D. Balduzzi, and B. Schölkopf, "Domain generalization via invariant feature representation," in *Proc. Int. Conf. Mach. Learn.*, 2013, pp. 10–18.

[25] H. Li et al., "Domain generalization for medical imaging classification with linear-dependency regularization," in *Proc. Adv. Neural Inf. Process. Syst.*, 2020, pp. 3118–3129.

[26] H. Lai et al., "Domain-aware dual attention for generalized medical image segmentation on unseen domains," *IEEE J. Biomed. Health Informat.*, vol. 27, no. 5, pp. 2399–2410, May 2023.

[27] S. Wang et al., "DoFE: Domain-oriented feature embedding for generalizable fundus image segmentation on unseen datasets," *IEEE Trans. Med. Imag.*, vol. 39, no. 12, pp. 4237–4248, Dec. 2020.

[28] R. Gu et al., "CDDSA: Contrastive domain disentanglement and style augmentation for generalizable medical image segmentation," *Med. Image Anal.*, vol. 89, 2023, Art. no. 102904.

[29] Y. Bi et al., "MI-SegNet: Mutual information-based us segmentation for unseen domain generalization," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assisted Interv.*, 2023, pp. 130–140.

[30] H. Kim, Y. Shin, and D. Hwang, "DiMix: Disentangle-and-mix based domain generalizable medical image segmentation," in *Int. Conf. Med. Image Comput. Comput.- Assist. Interv.*, 2023, pp. 242–251.

[31] Q. Liu et al., "MS-Net: Multi-site network for improving prostate segmentation with heterogeneous MRI data," *IEEE Trans. Med. Imag.*, vol. 39, no. 9, pp. 2713–2724, Sep. 2020.

[32] J. Hu et al., "Mixture of calibrated networks for domain generalization in brain tumor segmentation," *Knowl.-Based Syst.*, vol. 270, 2023, Art. no. 110520.

[33] P. Khandelwal and P. Yushkevich, "Domain generalizer: A few-shot meta learning framework for domain generalization in medical imaging," in *Proc. 2nd MICCAI Workshop Domain Adapt. Representation Transfer Distrib. Collaborative Learn.*, 2020, pp. 73–84.

[34] Q. Liu et al., "FedDG: Federated domain generalization on medical image segmentation via episodic learning in continuous frequency space," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2021, pp. 1013–1023.

[35] C. Wang et al., "MetaScleraSeg: An effective meta-learning framework for generalized sclera segmentation," *Neural Comput. Appl.*, vol. 35, no. 29, pp. 21797–21826, 2023.

[36] H. Li et al., "Frequency-mixed single-source domain generalization for medical image segmentation," in *Proc. Int. Conf. Medical Image Comput. Comput.–Assisted Intervention*, 2023, pp. 127–136.

[37] H. Oliveira et al., "Domain generalization in medical image segmentation via meta-learners," in *Proc. 35th SIBGRAPI Conf. Graph. Patterns Images*, 2022, pp. 288–293.

[38] J. Kim et al., "Pin the memory: Learning to generalize semantic segmentation," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2022, pp. 4350–4360.

[39] Y. Li et al., "Domain generalization via conditional invariant representations," in *Proc. AAAI Conf. Artif. Intell.*, 2018, pp. 3580–3587.

[40] S. Hu et al., "Domain and content adaptive convolution based multi-source domain generalization for medical image segmentation," *IEEE Trans. Med. Imag.*, vol. 42, no. 1, pp. 233–244, Jan. 2023.

[41] J.-Y. Zhu et al., "Unpaired image-to-image translation using cycle-consistent adversarial networks," in *Proc. IEEE Int. Conf. Comput. Vis.*, 2017, pp. 2223–2232.

[42] R. Gu et al., "Domain composition and attention for unseen-domain generalizable medical image segmentation," in *Proc. 24th Int. Conf. Med. Image Comput. Comput.–Assist. Interv.*, Strasbourg, France, 2021, pp. 241–250.

[43] Y. Zhao et al., "Style-hallucinated dual consistency learning: A unified framework for visual domain generalization," *Int. J. Comput. Vis.*, vol. 132, pp. 837–85, 2023.

[44] M. M. Rahman et al., "Multi-component image translation for deep domain generalization," in *Proc. IEEE Winter Conf. Appl. Comput. Vis.*, 2019, pp. 579–588.

[45] I. Goodfellow et al., "Generative adversarial nets," in *Proc. Adv. Neural Inf. Process. Syst.*, 2014, pp. 2672–2680.

[46] Q. Zhao et al., "A multi-modality ovarian tumor ultrasound image dataset for unsupervised cross-domain semantic segmentation," 2022, *arXiv:2207.06799*.

[47] Y. Chen et al., "Generating and weighting semantically consistent sample pairs for ultrasound contrastive learning," *IEEE Trans. Med. Imag.*, vol. 42, no. 5, pp. 1388–1400, May 2023.

[48] C. Wang et al., "Dynamic style transferring and content preserving for domain generalization," in *Int. Conf. Mobile Comput. Appl. Serv.*, 2022, pp. 298–315.

[49] J. N. Kundu et al., "AdaDepth: Unsupervised content congruent adaptation for depth estimation," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2018, pp. 2656–2665.

[50] X. Huang and S. Belongie, "Arbitrary style transfer in real-time with adaptive instance normalization," in *Proc. IEEE Int. Conf. Comput. Vis.*, 2017, pp. 1501–1510.

[51] Z. Yang and S. Farsiu, "Directional connectivity-based segmentation of medical images," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2023, pp. 11525–11535.

[52] H. Bogunović et al., "RETOUCH: The retinal OCT fluid detection and segmentation benchmark and challenge," *IEEE Trans. Med. Imag.*, vol. 38, no. 8, pp. 1858–1874, Aug. 2019.

[53] D. S. Kermany et al., "Identifying medical diagnoses and treatable diseases by image-based deep learning," *Cell*, vol. 172, no. 5, pp. 1122–1131, 2018.

[54] S. Farsiuet al., "Quantitative classification of eyes with and without intermediate age-related macular degeneration using optical coherence tomography," *Ophthalmol.*, vol. 121, no. 1, pp. 162–172, 2014.

[55] L. Zhang et al., "Generalizing deep learning for medical image segmentation to unseen domains via deep stacked transformation," *IEEE Trans. Med. Imag.*, vol. 39, no. 7, pp. 2531–2540, Jul. 2020.

[56] H. Zhang et al., "mixup: Beyond empirical risk minimization," 2017, *arXiv:1710.09412*.

[57] S. Yun et al., "CutMix: Regularization strategy to train strong classifiers with localizable features," in *Proc. IEEE/CVF Int. Conf. Comput. Vis.*, 2019, pp. 6023–6032.

[58] K. Zhou et al., "MixStyle neural networks for domain generalization and adaptation," *Int. J. Comput. Vis.*, vol. 132, pp. 822–836, 2023.

[59] Y. Ye et al., "UniSeg: A prompt-driven universal segmentation model as well as a strong representation learner," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assisted Interv.*, 2023, pp. 508–518.

[60] Z. Chen et al., "Treasure in distribution: A domain randomization based multi-source domain generalization for 2D medical image segmentation," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assisted Interv.*, 2023, pp. 89–99.

[61] S. Choi et al., "Robustnet: Improving domain generalization in urban-scene segmentation via instance selective whitening," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2021, pp. 11580–11590.

[62] D. Peng et al., "Semantic-aware domain generalized segmentation," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2022, pp. 2594–2605.

[63] L.-C. Chen et al., "Encoder-decoder with atrous separable convolution for semantic image segmentation," in *Proc. Eur. Conf. Comput. Vis.*, 2018, pp. 801–818.