Kurtosis Transformation for Multi-source Domain Generalization Segmentation of OCT Images from Multi-Manufacturers' Devices

Feiyu Sun, Ling Gao, Haoyu Chen, Bei Tian, Tao Peng, Weifang Zhu, Fei Shi, Ronghan Wu, Xinjian Chen, *Senior Member, IEEE*, Dehui Xiang, *Member, IEEE*

***Abstract*— Objective: Optical coherence tomography (OCT) images can visualize retinal layers and fundus lesions. Retinal structure segmentation is of great significance in early lesion detection and treatment guidance. However, devices from different OCT manufacturers are largely different from each other, which often leads to degraded results in image segmentation. Methods: To enrich the diversity of multi-source domains in intensity distributions and image contrasts of multi-manufacturers' OCT images, a kurtosis transformation method is proposed to generate a kurtosis-transformed image. To make the generated potential T-styles as different as possible from the source domain styles, a mean style contrastive learning method is proposed to maximize the distance between style features of a kurtosis-transformed image and multi-source domain images. To improve the diversity and independence of potential T-styles in a high-level orthogonal space, variance style orthogonalization is proposed to impose an orthogonal constraint on the reparametrized variance styles. Mean style features and variance style features are combined to modulate an input image for the training of the segmentation network. 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 segmentation method can be trained on labeled OCT images from multi-manufacturers' devices and can be tested on unseen manufacturer's device, and has good domain generalization performance in both retinal layer and lesions segmentation tasks. Significance: The proposed method can be used in routine clinical settings, when OCT images from multi-manufacturers' devices are available.**

***Index Terms*— OCT Image Segmentation, Domain Generalization, Retina Segmentation, Kurtosis, Orthogonalization, Contrastive Learning.**

This study was supported in part by National Nature Science Foundation of China (62371328, 61971298, 61771326, 62376167, 81871352) and in part by Natural Sciences Funding Project of Hunan Province (2023JJ70017) and in part by Peak Climbing Project of Eye Hospital Affiliated to Wenzhou Medical University and in part by Natural Science Foundation of the Jiangsu Higher Education Institutions of China (24KJB510042) and in part by Jiangsu Provincial Association for Science and Technology Young Science and Technology Talent Support Project (JSTJ-2024-434) and in part by the Priority Academic Program Development of Jiangsu Higher Education Institutions. Feiyu Sun, Ling Gao, Haoyu Chen, Bei Tian, Tao Peng and Ronghan Wu contributed this work equally. Dehui Xiang is the corresponding author. Email: xiangdehui@suda.edu.cn.

This work involved human subjects in its research. Approvals of all ethical and experimental procedures and protocols were granted by the Ethics Committee of Eye Hospital of Wenzhou Medical University under Application No. 2023-060-K-49-02, Ethics Committee of Joint Shantou International Eye Center of Shantou University and the Chinese University of Hong Kong under Application No. EC20240626(5)-P05, and Laboratory Animal Welfare Ethics Committee of Beijing Tongren Hospital under Application No. TREC2024-KY008.

Feiyu Sun, Weifang Zhu, Fei Shi, Xinjian Chen, Dehui Xiang are with the School of Electronic and Information Engineering, Soochow University, Suzhou, Jiangsu, 215006, China.

Ling Gao is with National Clinical Research Center for Ocular Diseases, Eye Hospital, Wenzhou Medical University, Wenzhou, 325027, China; National Engineering Research Center of Ophthalmology and Optometry, Eye Hospital, Wenzhou Medical University, Wenzhou, 325027, China; State Key Laboratory of Ophthalmology, Optometry and Visual Science, Eye Hospital, Wenzhou Medical University, Wenzhou, 325027, China; Department of Ophthalmology, the Second Xiangya Hospital, Central South University, Changsha, 410011, China.

Haoyu Chen is with Joint Shantou International Eye Center, Shantou University and the Chinese University of Hong Kong, Shantou, 515041, China.

Bei Tian is with Beijing Tongren Eye Center, Beijing Tongren Hospital, Capital Medical University, Beijing, China; Beijing Key Laboratory of Ophthalmology and Visual Sciences, Beijing, 100730, China.

Tao Peng is with the School of Future Science and Engineering, Soochow University, Suzhou, Jiangsu, 215222, China.

Ronghan Wu is with the National Clinical Research Center for Ocular Diseases, Eye Hospital, Wenzhou Medical University, Wenzhou, 325027, China.

I. Introduction

OCT [1] has emerged as a pivotal imaging modality for non-invasive visualization of retinal structures. It can provide high-resolution cross-sectional images and therefore it has become an indispensable tool for the early detection and monitoring of various ocular diseases. With the development of medical image segmentation [2]–[4], the segmentation of OCT images plays a critical role in the quantitative assessment and management of these ocular diseases. In addition, it is equally vital to perform segmentation of retinal layer and lesion [5], [6] in OCT images, as it can provide critical insights for clinical diagnosis of various ocular and systemic diseases.

Deep learning-based segmentation methods [7], [8] have achieved remarkable success in medical image analysis, due to their ability to learn complex patterns from large-scale annotated datasets. However, OCT images acquired from different manufacturers' devices or using different imaging protocols often exhibit substantial variability, and its distribution characteristics also have diversity (shown in Fig.1), which leads to significant domain shifts [9], [10]. This variability poses

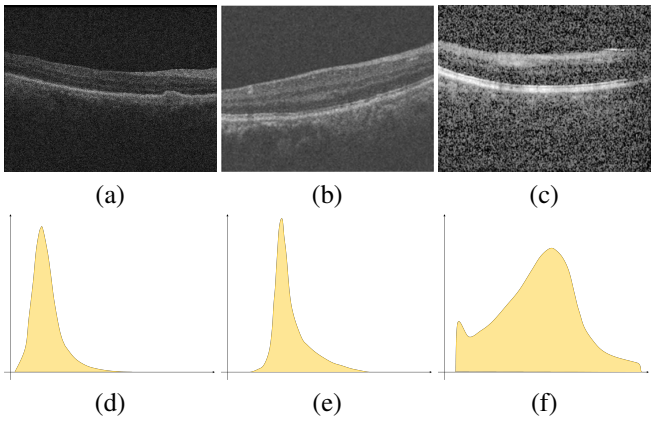

Fig. 1. Different intensity distributions and image contrasts of multi-manufacturers' OCT images. OCT images captured by three different manufacturers' devices (domains), (a) Cirrus; (b) Topcon; (c) Spectralis. Varied histograms of different domain images, (d) Cirrus; (e) Topcon; (f) Spectralis.

a major challenge to the generalizability and robustness of segmentation models, as models trained on one domain often fail to perform well on unseen domains.

To address these issues, research has focused on domain adaptation (DA) [11]–[13] and domain generalization (DG) [14]–[16] methods. DA methods attempted to align the distributions of source and target domains. The synthetic target domain images were generated to train a segmentation model and further alleviate the domain shift [17], [18]. They often used adversarial learning [19], [20] or style-transfer techniques [21], [22]. These approaches required access to target domain data during training, which limited their applicability in clinical scenarios where target data might not be available. In contrast, DG methods aimed to improve model performance on unseen domains by leveraging only source domain data. Some DG methods used randomization-based strategies [23], [24] to generate augmented data by applying random transformations. Normalization-based methods [25], [26] were also a common strategy in DG, and had achieved significant improvements in multiple domain generalization tasks. Additionally, at the level of feature representation, some generalized methods focused on invariant feature representation [27]–[29] and feature disentanglement [30]–[33], decomposing the features of input samples into domain invariant and domain specific components. There were also some commonly used strategies such as ensemble learning [34], [35], meta-learning [36]–[38], and self-supervised learning [39], which could enhance the generalization ability to unseen domains.

In this paper, we propose a novel domain generalization method to increase the diversity of kurtosis in the histogram of the samples in the multi-source domains. Kurtosis is used to describe the distribution of image intensities. OCT images with high histogram kurtosis may have a concentrated distribution that makes it easier for the model to capture the main features of the retina. On the other hand, OCT images with low histogram kurtosis have a more uniform distribution of image intensities, and the intensities appear relatively balanced, which is more conducive to preventing the model from focusing too much on local features and ignoring overall features of the retina. Therefore, kurtosis transformation is applied to OCT images from multi-source domains such that generalization of the proposed segmentation framework can be improved. A Mean Style Contrastive Learning (MSCL) method is proposed to widen the distance between the generated potential target domain styles (T-styles) and the multi-source domain styles. A Variance Style Orthogonalization (VSO) method is proposed to further improve the diversity and independence of potential T-styles. Comprehensive experiments have been performed with one in-house multi-source domain OCT image dataset, and a publicly available multi-source domain OCT image dataset. The experimental results have demonstrated the superiority of our framework over state-of-the-art domain generalization methods. The main contributions of our work are summarized as follows,

- A kurtosis transformation method is proposed to generate a kurtosis-transformed image for enriching the diversity of source domains in intensity distributions and image contrast. The interval with the highest kurtosis is selected, assuming that its peak falls within the interval, and an intensity mapping curve is chosen for the change of intensity distributions and contrast of retina, such that intensity distributions and contrast of retinal structures can be modified for different manufacturers' OCT images.
- A mean style contrastive learning method is proposed to guide the generation of the desirable T-styles of the unseen target domain. In order to make the generated potential T-styles as different as possible from the source domain styles, the contrastive representation learns the style features of images by minimizing the similarity between a kurtosis-transformed image and multi-source domain images while the contrastive representation learns the style features of images by maximizing the similarity between kurtosis-transformed images in the same batch.
- Variance style orthogonalization is proposed to improve the diversity and independence of potential T-styles in a high-level orthogonal space. To improve the diversity of style codes, a random style following the normal distribution is sampled 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. By introducing an orthogonal constraint to the reparametrized variance style, the task of generating T-styles for unseen target domain is transformed into an orthogonal constraint problem within the high-level orthogonal style space.

Comprehensive experiments on both the OCT retinal layer segmentation dataset and the OCT retinal lesion segmentation dataset have been performed, and the experimental results showed that our method outperformed state-of-the-art domain generalization methods on two OCT image segmentation tasks.

II. Related Work

A. Domain Generalization for Medical Image Analysis

DG [24], [26], [29] has gained significant attention in recent years, particularly for its ability to improve model performance on unseen target domains without requiring access to

target domain data. This was particularly relevant in medical image segmentation tasks, where domain shifts could occur due to variations in imaging protocols, equipment, or patient populations. Many methods have been proposed to address these issues, including feature learning [40]–[43], adversarial training [19], [20], data augmentation [44]–[46], etc.

Many methods have focused on learning domain-invariant features [27], [30], [47] to mitigate domain shifts. The goal was to extract features that were not specific to any particular domain, allowing the model to generalize better to new domains. Lai et al. [29] proposed a domain-aware dual attention network to alleviate the severe distribution shift between the source and target domain. Hu et al. [48] employed the domain-discriminative features embedded within the encoder to produce the domain code for each input image. This process established a connection between multiple source domains and the unseen target domain. Zhang et al. [49] used shape-invariant learning to perform implicit shape distribution alignment at the domain-level to learn shape-invariant representation. Zhang et al. [50] unified the multi-source domains into a global inter-source domain via a novel feature statistics update mechanism. Although these methods have shown a promising progress, they often struggled to balance the trade-off between domain invariance and task-specific feature learning, and therefore, they tended to produce suboptimal performance on unseen target domains.

Adversarial training has proven effective in many domain generalization tasks, where the goal was to align the feature distributions between the source and target domains. Fick et al. [51] used Cycle-GAN [52] to enrich the samples by transforming images to another style. In our previous study [53], adversarial learning of graph convolutional networks was used to capture the underlying relationships among various lesions.

B. Data Augmentation for Domain Generalization

Data augmentation has always been an effective method for domain generalization. Traditional augmentation strategies, such as rotation, scaling, and flipping, were commonly used to increase the diversity of the training data [54]. For instance, methods like Mixup [55], MixStyle [56] and Cutmix [57] have been shown to improve model generalization by mixing images or pixels from different domains to create hybrid samples. Qi et al. [58] proposed a normalization-based method, called NormAUG, which used normalization based on samples from different domains to enhance the diversity of training data. Guo et al. [59] transformed token features by perturbing local edge cues while preserving global shape features to enhance the model learning of shape information and the generalization ability. However, while these methods increased diversity, they might still fail to capture the complex domain-specific variations presented in medical images, especially when there was a large gap between source and target domains.

To better simulate the target domain, some methods explicitly aimed to generate target-like data. For example, contrastive domain disentanglement and style augmentation (CDDSA) [31] disentangled style and content representations, allowing the model to generate new styles based on a combination of learned style features. Similarly, dynamic style transfer [60] dynamically altered the style of source images to mimic potential target styles and bridge the gap between domains. Zhou et al. [23] enhanced the diversity of source domain via generating source-similar and source-dissimilar images based on Bezier Curves [61] and developed a dual-normalization network for effective exploitation. However, these methods often assumed that the target domain style lay within the distribution of source domains, which might not always hold true for medical images.

Advanced data augmentation strategies have sought to explicitly sample styles that mimic potential target domain distributions. For instance, Wang et al. [62] used a sampling-based method named CrossSmooth to acquire semantic directions from inter-class, and then used CrossVariance to obtain the styles of different domains by sampling semantic directions. In our previous study [53], Dirichlet sampling has been applied to create pseudo-target domain samples by interpolating marginal style features of the source domain. This technique enhanced the diversity of generated styles and improved the generalization. Marginal style sampling methods further identified marginal styles within the source domain distributions, which were more likely to approximate unseen target domain styles.

III. Method

A. Problem Definition and Method Overview

For multi-source domain OCT image segmentation, K subsets of OCT images are collected from different manufacturers' OCT devices. The multi-source domain dataset D^s consists of $K-1$ subsets, defined as $D^s = \left\{ D^{s^k} \right\}_{k=1}^{K-1}$, where D^{s^k} denotes the k-th source domain dataset. $D^{s^k} = \left\{ \left(x_i^k, y_i^k \right) \right\}_{i=1}^{N_k}$, where x_i^k denotes the i-th image from the k-th source domain and y_i^k denotes the corresponding segmentation label. N_k denotes the number of images in k-th source domain. $D^t = \left\{ \left(x_i^{K-1}, y_i^{K-1} \right) \right\}_{i=1}^{N_K}$ denotes the target domain dataset and N_K denotes the number of images in target domain. D^t is not visible in the training stage.

The proposed multi-source domain generalization framework for OCT image segmentation is shown in Fig.2. To enrich the diversity of multi-source domains in intensity distributions and image contrasts of different manufacturers' OCT images, a kurtosis transformation method is proposed to generate a kurtosis-transformed image. The sub-histogram with the highest kurtosis is treated as the prominent interval of image intensities, and an intensity mapping curve is chosen to change intensity distributions and image contrasts, such that intensity distributions and contrast of retinal layers or lesions can be modified for different manufacturers' OCT images. In order to make the generated potential T-styles as different as possible from the source domain styles, a mean style contrastive learning method is proposed to maximize the distance between style features of a kurtosis-transformed image and multi-source domain images. In order to improve the diversity and independence of potential T-styles in a high-level orthogonal space, variance style orthogonalization is proposed to impose

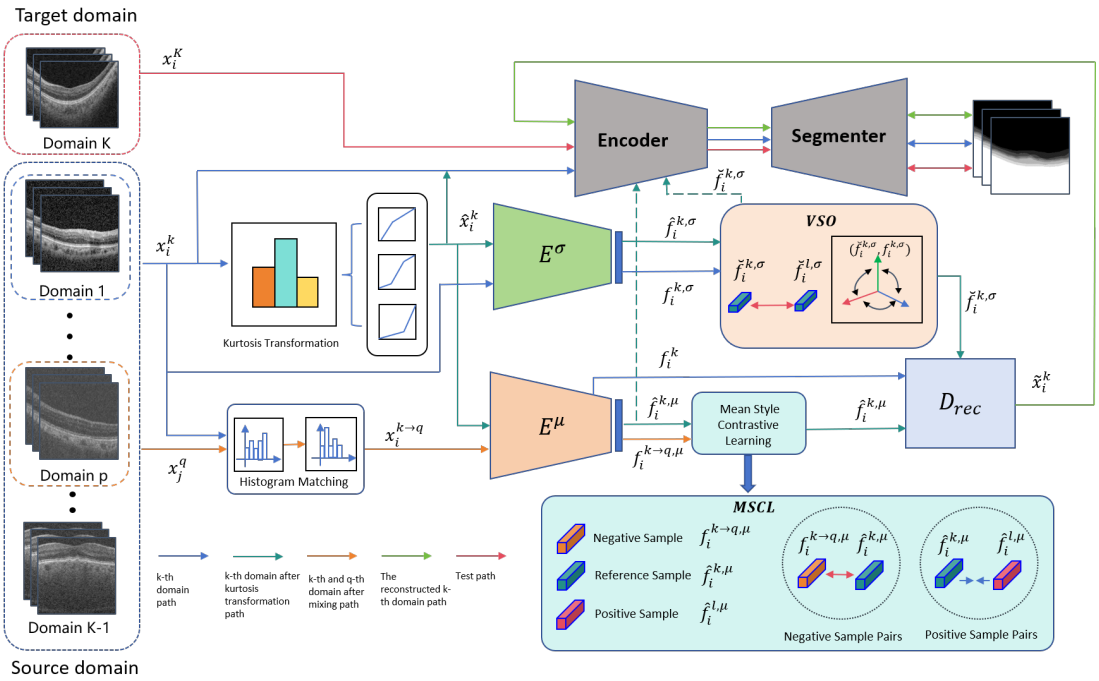

Fig. 2. Overview of our proposed framework. The source domain sample $\boldsymbol{x_i^k}$ is fed into Kurtosis Transformation (KT) to obtain $\hat{\boldsymbol{x}}_{\boldsymbol{i}}^{\boldsymbol{k}}$ and generate $\boldsymbol{x_i^{k\to q}}$ through histogram matching. Then, the transformed sample $\hat{\boldsymbol{x}}_{\boldsymbol{i}}^{\boldsymbol{k}}$ and the source domain sample $\boldsymbol{x_i^k}$ are fed into the variance style extractor for Variance Style Orthogonalization (VSO) to obtain the variance style features $\breve{\boldsymbol{f}}_{\boldsymbol{i}}^{\boldsymbol{k,\sigma}}$, which are used to modulate the encoder with the mean style features $\hat{\boldsymbol{f}}_{\boldsymbol{i}}^{\boldsymbol{k,\mu}}$. The $\hat{\boldsymbol{x}}_{\boldsymbol{i}}^{\boldsymbol{k}}$ and $\boldsymbol{x_i^k}$ are fed into the mean style encoder for the Mean Style Contrastive Learning (MSCL) to obtain the $\hat{\boldsymbol{f}}_{\boldsymbol{i}}^{\boldsymbol{k,\mu}}$. Finally, $\breve{\boldsymbol{f}}_{\boldsymbol{i}}^{\boldsymbol{k,\sigma}}$ and $\hat{\boldsymbol{f}}_{\boldsymbol{i}}^{\boldsymbol{k,\mu}}$ are fed into the reconstruction decoder to obtain the final reconstructed image $\tilde{\boldsymbol{x}}_{\boldsymbol{i}}^{\boldsymbol{k}}$, which is then fed into the segmentation network along with the source domain sample $\boldsymbol{x_i^k}$ for training.

an orthogonal constraint on the reparametrized variance styles. Mean style features and variance style features are combined to modulate an input image for the training of the segmentation network. In the following sections, we will elaborate the proposed constraints for retinal layer/lesion segmentation of unseen manufacturer's OCT images.

B. Kurtosis Transformation

As shown in Fig.1, images acquired from different manufacturers' OCT scanners present significant differences in image appearance and contrast since they are captured with different cameras, different light sources and settings. The corresponding histograms also show significant differences in intensity distributions. Such distribution differences lead to domain shift between sources and target domains and can greatly degrade the performance of retinal structure segmentation in fundus OCT images, which is more severe when target domain is unseen. To mitigate domain shift and improve the generalization of the retina segmentation model, it is advisable to enrich the diversity of source domains in intensity distributions and image contrasts. In our study, we find that retinal structures present a concentrated distribution in the histogram. In statistics, kurtosis is an indicator that describes the degree of steepness of a distribution pattern; and therefore, histogram kurtosis can be used to measure the concentration of retinal structures in fundus OCT images, such that intensity distributions and contrasts of retinal structures can be modified for different manufacturers' OCT images.

For a fundus OCT image, intensities of retinal structure often locate in an interval with high kurtosis of histogram. When a new intensity mapping function is applied, the kurtosis of an interval of histogram can be modified. A prominent intensity interval detection method for the OCT image is proposed to modify the kurtosis of the histogram distribution. For a selected source domain image $x_i^k \in D^{s_k}$, intensities are normalized to a fixed range $[0, E_M^{s_i^k}]$, which is then divided into M isometric sub-intervals, $[0, E_1^{s_i^k})$, $[E_1^{s_i^k}, E_2^{s_i^k})$, $\cdots$, $[E_{M-1}^{s_i^k}, E_M^{s_i^k}]$. To improve the random sampling and the diversity of intensity transformation, the start and end of each interval are added with a random value that is uniformly sampled from an interval $[-\xi, \xi]$. For all pixels in each interval, their kurtosis of histogram can be computed by

$$K_m^{s_i^k} = \frac{N_m^{s_i^k}(N_m^{s_i^k}+1)}{(N_m^{s_i^k}-1)(N_m^{s_i^k}-2)(N_m^{s_i^k}-3)} \sum_{j=1}^{N_m^{s_i^k}} \left(\frac{p_j-\overline{p}}{\tilde{p}}\right)^4 - \frac{3(N_m^{s_i^k}-1)^2}{(N_m^{s_i^k}-2)(N_m^{s_i^k}-3)}, \quad (1)$$

where $N_m^{s_i^k}$ denotes the number of pixels in the m-th interval of x_i^k, and p_j denotes the j-th intensity. $\overline{p}$ denotes the mean intensity and $\tilde{p}$ denotes the standard deviation of the intensities

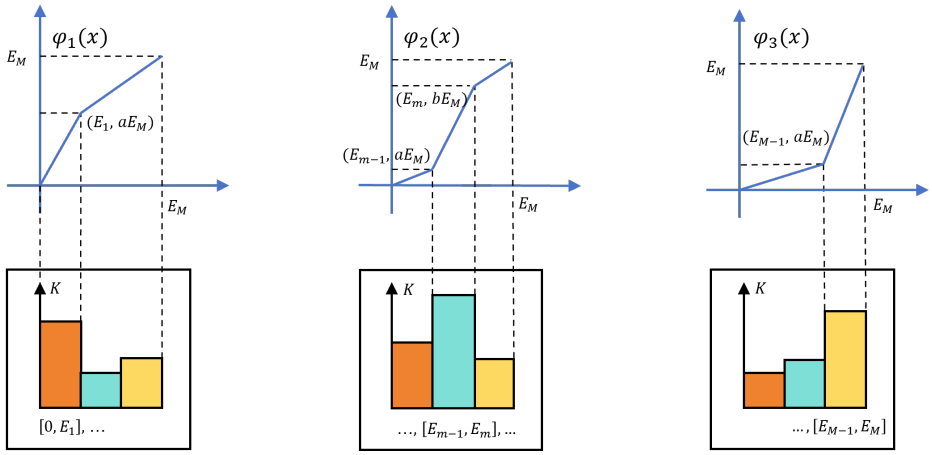

Fig. 3. Kurtosis transformation with multiple mapping curves.

in the interval. Eq.(1) can be approximated by

$$K_m^{s_i^k} \approx \frac{1}{N_m^{s_i^k}} \sum_{j=1}^{N_m^{s_i^k}} \left(\frac{p_j - \overline{p}}{\tilde{p}}\right)^4 - 3 \tag{2}$$

In order to better adapt the model to unseen domains, we need to enrich the diversity of kurtosis of the source images. Therefore, a kurtosis transformation method is proposed to generate a kurtosis-transformed image $\hat{x}_i^k$ with a kurtosis different from the original image x_i^k. As is shown in Fig.3, the interval with the highest kurtosis is selected, assuming that its peak falls within the interval, and an intensity mapping curve is chosen for the change of intensity distributions and contrasts of retinal structures. This process can be described as

$$\hat{x}_i^k = \sum_{m=1}^{M} M_m * \varphi_m(x_i^k), \tag{3}$$

$$M_m = \begin{cases} 1, & \text{if } \max\{K_m^{s_i^k}\} \\ 0, & \text{otherwise} \end{cases}, m = 1, 2, \cdots, M, \tag{4}$$

where φ_m denotes the m-th intensity mapping curve, which can be defined by added some control points as

(1) If $[0, E_1^{s_i^k})$ is the interval with the highest kurtosis, a piecewise linear function $\varphi_1(p_j)$ is applied to transform the intensity distribution. Besides the start point $(0, 0)$ and the end point $\left(E_M^{s_i^k}, E_M^{s_i^k}\right)$, a control point $\left(E_1^{s_i^k}, aE_M^{s_i^k}\right)$ can be added a two-piece linear function $\varphi_1(p_j)$, where a is the slope of the first piece linear function. For pixel values in $[0, E_1^{s_i^k})$, intensity distributions and image contrasts can be widened if $a > E_1^{s_i^k}/E_M^{s_i^k}$ or else they become narrow. For pixel values in $[E_{M-1}^{s_i^k}, E_M^{s_i^k}]$, intensity distributions and image contrasts can be widened if $a < (M-1)/M$ or else they become narrow.

(2) If $[E_{m-1}^{s_i^k}, E_m^{s_i^k}]$ is the interval with the highest kurtosis, a piecewise linear function $\varphi_2(p_j)$ is applied to transform the intensity distribution. Besides the start point $(0, 0)$ and the end point $\left(E_M^{s_i^k}, E_M^{s_i^k}\right)$, two control points $\left(E_{m-1}^{s_i^k}, aE_M^{s_i^k}\right)$ and $\left(E_m^{s_i^k}, bE_M^{s_i^k}\right)$ can be added a three-piece linear function $\varphi_2(p_j)$, where a is the slope of the first piece linear function and b is the slope of the third piece linear function. If $(b-a)\, E_M^{s_i^k} > E_m^{s_i^k} - E_{m-1}^{s_i^k}$, intensity distributions and image contrasts can be widened or else they become narrow for pixel intensities in $[E_{m-1}^{s_i^k}, E_m^{s_i^k}]$.

(3) If $[E_{M-1}^{s_i^k}, E_M^{s_i^k}]$ is the interval with the highest kurtosis, a piecewise linear function $\varphi_3(p_j)$ is applied to transform the intensity distribution. Besides the start point $(0, 0)$ and the end point $\left(E_M^{s_i^k}, E_M^{s_i^k}\right)$, a control point $\left(E_{M-1}^{s_i^k}, aE_M^{s_i^k}\right)$ can be added a two-piece linear function $\varphi_3(p_j)$, where a is the slope of the first piece linear function. For pixel values in $[E_{M-1}^{s_i^k}, E_M^{s_i^k}]$, intensity distributions and image contrasts can be widened if $a < E_{M-1}^{s_i^k}/E_M^{s_i^k}$ or else they become narrow.

Besides the above piecewise linear functions, other functions such as exponential functions, logarithmic functions, Bezier functions and B-spline functions can also be used as the intensity mapping curves to modify intensity distributions and image contrasts.

C. Mean Style Contrastive Learning

Our goal is to develop a generalized retinal structure segmentation framework that can segment an unseen manufacturer' OCT image by capturing and transferring the overall style of a multi-source OCT image. The above kurtosis transformation method mainly focuses on transforming and generating potential target domain images from existing source domains, but it is challenging to fully randomly generate the desirable style of the unseen target domain because the styles of the multi-source domains largely differ from those of unseen target domains. A key component is to find a suitable style representation that can be used to distinguish different styles and further guide the generation of the desirable styles of an unseen target domain.

To this end, an mean style feature extractor is designed. Instead of using features from a specific layer or a fusion of multiple layers, our feature extractor projects mean features of the last layer into a latent style space to encode style cues. Specifically, the encoder of UNet is used as our feature

extractor and an average pooling is used to project the kurtosis-transformed image $\hat{x}_i^k$ into an L-dimensional latent style code. A branch of the contrastive learning scheme is potential T-style representation learning. The mean style feature extractor needs to be trained to obtain a reasonable style representation that is in the form of the style code. However, the ground-truth style code of unseen target domain is unavailable for supervised training. In order to make the generated potential T-styles as different as possible from the source domain styles, it is necessary to widen the distance between the generated potential T-styles and the multi-source domain styles. To achieve this goal, a style contrastive learning strategy is proposed, and a new contrastive style loss is designed as an implicit measurement for the mean style feature extractor training.

When training the mean style feature extractor for the task of multi-source domain generalization, histogram matching is used to transform the cross styles between multi-source domain images such that domain shift between $K-1$ source domains can be mitigated and the feature extractor is allowed to extract desirable styles of the generated target domain different from those of multi-source domains. Specifically, for a selected source domain image $x_i^k \in D^{s^k}$, another source domain image $x_j^q \in D^{s^k}$ is randomly selected, and their intensities are normalized to a fixed range $[0, E_M^{s_i^k}]$. The histogram-matched image $x_i^{k\to q}$ can be obtained by applying the histogram matching method. The kurtosis-transformed image $\hat{x}_i^k$ and the histogram-matched image $x_i^{k\to q}$ are respectively fed into the encoder of the mean style feature extractor. The mean style feature of $\hat{x}_i^k$ can be obtained as $\hat{f}_i^{k,\mu} = E^{\mu}(\hat{x}_i^k)$ and the mean style feature of $x_i^{k\to q}$ can be obtained as $f_i^{k\to q,\mu} = E^{\mu}(x_i^{k\to q})$. $\hat{f}_i^{k,\mu}$ from the k-th domain is treated as the reference style feature and $\hat{f}_i^{l,\mu}$ from the l-th domain is treated as the positive style feature. The contrastive representation learns the style features of images by maximizing the similarity between $\hat{f}_i^{k,\mu}$ and $\hat{f}_i^{l,\mu}$ such that the generated potential T-styles in each iteration can be aligned. $f_i^{k\to q,\mu}$ is treated as the negative style feature. The contrastive representation learns the style features of images by minimizing the similarity between $\hat{f}_i^{k,\mu}$ and $f_i^{k\to q,\mu}$ such that the generated potential T-styles in each iteration can be different from those of $K-1$ source domains and make the kurtosis-transformed image $\hat{x}_i^k$ present potential T-styles, which are potentially similar to those of the unseen target domains. The contrastive loss is defined to train our mean style feature extractor as

$$\mathcal{L}_c^{\mu} = -log \frac{\sum_{k=0}^{K-1}\sum_{l=k+1}^{K-1} e^{sim(\hat{f}_i^{k,\mu},\hat{f}_i^{l,\mu})/\tau}}{\sum_{k=0}^{K-1}\sum_{l=k+1}^{K-1} e^{sim(\hat{f}_i^{k,\mu},\hat{f}_i^{l,\mu})/\tau} + \sum_{k=0}^{K-1} e^{sim(f_i^{k\to q,\mu},\hat{f}_i^{k,\mu})/\tau}} \tag{5}$$

where $sim(\cdot,\cdot)$ is the cosine similarity, and τ is the temperature scaling parameter.

D. Variance Style Orthogonalization

The above contrastive learning aims to maximize the distance between the mean style features of kurtosis-transformed images and multi-source domain images. Variance style reparameterization is further performed in a high-level orthogonal space to further improve the diversity of potential T-styles. To this end, a variance style feature extractor E^{σ} is designed to project the kurtosis-transformed image $\hat{x}_i^k$ into an L-dimensional latent style feature as the mean style feature extractor with a channel-wise variance layer, i.e., the variance style feature of $\hat{x}_i^k$ can be obtained as $\hat{f}_i^{k,\sigma} = E^{\sigma}\left(\hat{x}_i^k\right)$. Note that the variance style code $\hat{f}_i^{k,\sigma}$ obtained through transforming in the current epoch may be the similar to the styles presented in the source domain. To improve the diversity of style codes, a random style following to the normal distribution $\tilde{f} \sim N(0,1)$ is sampled to generate potential styles not presented in the known source domains, such that domain generalization of the network can be improved in unseen target domains. The variance style feature is mixed as

$$\breve{f}_i^{k,\sigma} = \alpha \hat{f}_i^{k,\sigma} + (1-\alpha)\,\tilde{f}, \alpha \in [0,1], \tag{6}$$

where α is a uniformly random coefficient.

For each iteration in the above contrastive learning, the mean style feature $\hat{x}_i^k$ of kurtosis-transformed images is treated as the reference. As can be seen in Eq.(1), the standard deviation of the intensities is also an important factor to affect intensity distributions and contrasts of retinal structures. To enhance the diversity of potential T-styles, $\breve{f}_i^{k,\sigma}$ from the k-th domain is treated as the reference style feature and $\breve{f}_i^{l,\sigma}$ from the l-th domain is treated as the negative style feature. A inter-domain variance style function is designed to diversify the visual styles of images by maximizing the distance between $\breve{f}_i^{k,\sigma}$ and $\breve{f}_i^{l,\sigma}$ such that the generated potential T-styles in each iteration can be different between kurtosis-transformed multi-source images. The inter-domain variance style function is defined as

$$L_{te}^{\sigma} = log \sum_{k=0}^{K-1}\sum_{l=k+1}^{K-1} e^{sim(\breve{f}_i^{k,\sigma},\breve{f}_i^{l,\sigma})/\tau} \tag{7}$$

To further improve the diversity of potential T-styles, variance style reparameterization is further performed in a high-level orthogonal style space. If the two features are orthogonal, more independent variance styles are probably produced and diverse potential T-styles can be hence produced. Therefore, by introducing an orthogonal constraint to the reparametrized variance style, we can transform the task of generating T-styles for unseen target domain into an orthogonal constraint problem within the high-level orthogonal style space. Therefore, the orthogonal constraint is imposed on variance style sampling. The variance style $f_i^{k,\sigma}$ of the source domain image x_i^k is fed into the variance style feature extractor E^{σ}, i.e., $f_i^{k,\sigma} = E^{\sigma}\left(x_i^k\right)$, to obtain high-level features $f_i^{k,\sigma} \in F^{\sigma}$. After $\breve{f}_i^{k,\sigma} \in \breve{F}^{\sigma}$ is obtained as Eq.(6), the mapped features $f_i^{k,\sigma}$ and $\hat{f}_i^{k,\sigma}$ are imposed to maximize the differences between style vectors of different domains, reparametrized and original styles. $f_i^{1,\sigma}, f_i^{2,\sigma}, \cdots, f_i^{K-1,\sigma}, \breve{f}_i^{1,\sigma}, \breve{f}_i^{2,\sigma}, \cdots, \breve{f}_i^{K-1,\sigma}$ are stacked into a matrix $\hat{F}_i^{\sigma} \in R^{(2K-2)\times L}$. These variance styles should be well-separated in the orthogonal style space, and the

distance between styles of different domains, reparametrized and original styles, should be maximized such that the learned styles can be distinguishable. Each reparametrized variance style extracted from a sample in the source domain forms a bundle of styles with the original styles extracted from some samples in other source domains. The orthogonal loss of $\mathcal{L}_{or}^{\sigma}$ can be expressed as

$$L_{or}^{\sigma} = \left\| \frac{\hat{F}_i^{\sigma} \left(\hat{F}_i^{\sigma}\right)^T}{\|\hat{F}_i^{\sigma}\| \cdot \left\|\left(\hat{F}_i^{\sigma}\right)^T\right\|} - I_{(2K-2)\times(2K-2)} \right\| + \left\| \frac{\left(\hat{F}_i^{\sigma}\right)^T \hat{F}_i^{\sigma}}{\left\|\left(\hat{F}_i^{\sigma}\right)^T\right\| \cdot \|\hat{F}_i^{\sigma}\|} - I_{L\times L} \right\|, \tag{8}$$

where $I_{(2K-2)\times(2K-2)}$ denotes the identity matrix of size $(2K-2)\times(2K-2)$ and $I_{L\times L}$ denotes the identity matrix of size $L \times L$. The first term promotes independence of the variance styles to reduce redundant information and maximize the information capacity of the high-level style space and the second term encourages orthogonality of the variance styles, thereby maximizing the inter-class distance of distinctive variance styles.

E. Style Transfer for Generalized Segmentation

Many previous domain generalization methods mainly focused on data augmentation from source domains. Apart from the image-based method, style transfer is also performed in the encoder of our segmentation network such that the model is probably to see more reparametrized and orthogonal potential T-styles. To this end, style transfer is applied to the features of the original multi-source domain image in the encoder of our segmentation network by using AdaIN [63] as

$$f_i^{k,a} = \breve{f}_i^{k,\sigma} \frac{f_i^k - \mu\left(f_i^k\right)}{\sigma\left(f_i^k\right)} + \hat{f}_i^{k,\mu} \tag{9}$$

where $\mu(\cdot)$ denotes the mean operation and $\sigma(\cdot)$ denotes the standard deviation operation.

To ensure the effectiveness of style encoding, the codes produced by E^{μ} and E^{σ} are fed into our style decoder D_{re}. D_{re} discerns the distinctive style characteristics of different domains and reconstructs the generated target domain images. This ensures that E^{μ} and E^{σ} do not generate randomly high-dimensional style vectors in the high-level space. Consequently, further constraints are imposed on the preceding mean style feature extractor and variance style feature extractor, and the reconstruction consistency loss is defined as

$$L_{rc} = \left\| \hat{x}_i^k - \tilde{x}_i^k \right\|, \tag{10}$$

where the style feature $\hat{f}_i^k$ of $\hat{x}_i^k$ is obtained as $\hat{f}_i^k = E^{\mu}(\hat{x}_i^k)$ without applying the average pooling. $\tilde{x}_i^k$ denotes the reconstructed image as $\tilde{x}_i^k = D_{re}\left(f_i^k, f_i^{k,a}\right)$. After the variance style $\breve{f}_i^{k,\sigma}$ are sampled with the orthogonalization constraint, style transfer is performed by using AdaIN [63] as

$$\tilde{f}_i^{k,a} = \breve{f}_i^{k,\sigma} \frac{\hat{f}_i^k - \hat{f}_i^{k,\mu}}{\sigma\left(\hat{f}_i^k\right)} + \hat{f}_i^{k,\mu} \tag{11}$$

It is important to note that the gradients in Eq.(10) need to be backpropagated to the mean style encoder, the variance style encoder and the style decoder. In contrast, the gradients from the aforementioned contrastive loss in Eq.(5) are only backpropagated to the mean style feature extractor. To preserve the semantic content of the source domain in reconstructed image, $\tilde{x}_i^k$ is fed into our segmentation network and the segmentation loss is defined as

$$L_{rc}^{seg} = L_{DSC}\left(\tilde{y}_i^k, y_i^k\right) + L_{CE}\left(\tilde{y}_i^k, y_i^k\right), \tag{12}$$

where L_{DSC} denotes Dice similarity loss and L_{CE} denotes cross-entropy loss. $\tilde{y}_i^k$ denotes the segmentation prediction of $\tilde{x}_i^k$. For each image in the multi-source domain and its transformed image, they are also fed into our segmentation network and the segmentation loss is defined as

$$L_{or}^{seg} = L_{DSC}\left(\bar{y}_i^k, y_i^k\right) + L_{CE}\left(\bar{y}_i^k, y_i^k\right) + L_{DSC}\left(\hat{y}_i^k, y_i^k\right) + L_{CE}\left(\hat{y}_i^k, y_i^k\right), \tag{13}$$

where $\bar{y}_i^k$ denotes the segmentation prediction of x_i^k and $\hat{y}_i^k$ denotes the segmentation prediction of $\hat{x}_i^k$.

Overall, the total loss is defined as

$$L_{total} = L_{or}^{seg} + L_{rc}^{seg} + \lambda_c^{\mu} L_c^{\mu} + \lambda_{te}^{\sigma} L_{te}^{\sigma} + \lambda_{or}^{\sigma} L_{or}^{\sigma} + \lambda_{rc} L_{rc}, \tag{14}$$

where λ_c^{μ}, λ_{te}^{σ}, λ_{or}^{σ} and λ_{rc} are the weights.

IV. Experiments

A. Dataset

The proposed method was evaluated on two OCT image datasets scanned by multi-manufacturers' OCT devices: a retinal layer segmentation dataset from three clinical centers and a publicly available retinal fluid segmentation dataset (RETOUCH) [64]. The descriptions of the datasets were shown in Table I. The fundus image datasets were collected from different clinical centers.

1) Retinal Layer Segmentation Dataset: OCT images were collected from four distinct clinical centers, in which OCT scanners were scanned by four different manufacturers' OCT devices and therefore regarded as four domains. This dataset consisted of 394 OCT volumes, where 94 volumes of 94 patients with age-related macular degeneration were acquired with the Cirrus scanner (Zeiss) in Joint Shantou International Eye Center, Shantou University and the Chinese University of Hong Kong. 100 volumes of 100 patients with diabetic retinopathy were acquired with the Spectralis scanner (Heidelberg) in fundus disease clinical center, school of ophthalmology and optometry and eye hospital of Wenzhou Medical University. 100 volumes of 100 patients with myopia were acquired with the VG200D scanner (SVision) in Beijing Tongren Hospital. 100 volumes of 100 patients with age-related macular degeneration were acquired with the Topcon 2000 scanner (Topcon) in Joint Shantou International Eye Center, Shantou University and the Chinese University of Hong Kong. For each volume, there were 128, 19, 1240, and 128 B-scans with the size of 1389×926, 512×496, 1536×3936, and 1024×1177 for Cirrus, Spectralis, Svision and Topcon, respectively. The retinal layers were manually annotated by an experienced expert. Besides vitreous and choroid, retinal layers such as nerve fiber layer (NFL), ganglion cell layer and

TABLE I
DETAILS ABOUT THE TWO DATASETS.

Dataset	Domain No.	Size	Volume(B-scans)	Scanners	Class
Retinal Layer	Domain 1	1389×926	94(709)	Cirrus	Vitreous, Choroid, NFL, GCL-IPL, INL, OPL, ONL-ELM, MZ-RPE
	Domain 2	512×496	100(1000)	Spectralis	
	Domain 3	1536×3936	100(1000)	Svision	
	Domain 4	1024×1177	100(1000)	Topcon	
Retinal Fluid	Domain 1	512×1024	24(1568)	Cirrus	IRF, SRF, PED
	Domain 2	512×496	24(711)	Spectralis	
	Domain 3	512×650	22(1106)	T-1000 and T-2000	

inner plexiform layer (GCL-IPL), inner nuclear layer (INL), outer plexiform layer (OPL), outer nuclear layer and external limiting membrane (ONL-ELM), myoid zone to retinal pigment epithelium (MZ-RPE) were annotated. Due to the heavy annotation of retinal layers, about 10 B-scans were randomly annotated from each OCT volume. All B-scans were resized to the size of 512×512.

2) Retinal 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 resized to the size of 512×512.

B. Implementation Details

All models were tested on one NVIDIA RTX 3090 GPU. The networks were trained for 120 epochs, with an initial learning rate of 0.001. The Adam optimizer was used. The batch size was set to 1. DeepLabv3+ [65] with ResNet101 was used as the main framework for our segmentation network. Intensities in each B-scan were normalized to a fixed range $[0, 1]$. Three different intervals were set for kurtosis transformation ($M = 3$). The slop a of the main interval the highest kurtosis was set to a random number in $[2, 10]$ while the slop b of other intervals was set to a random number in $[0.1, 0.5]$. ξ, τ, λ_c^{μ}, λ_{te}^{σ}, λ_{or}^{σ} and λ_{rc} were set to 0.1. Data-augmentation was performed based on BigAug [54].

C. Experimental Results and Analysis

1) Comparative Methods and Evaluation Metrics: The lower bound (Inter-domain) referred to training the encoder-decoder segmentation network using all source domains, followed by testing on an unseen target domain. This approach often resulted in degraded performance due to the significant distribution discrepancy between the source (training) data and the target (testing) data. Conversely, the upper bound (Intra-domain) represented a scenario where each individual domain was separately trained and evaluated through 4-fold cross-validation without patient overlapping. The proposed method was compared with several state-of-the-art approaches. Data manipulation-based techniques, such as **Mixup** [55] and **Cutmix** [57], operated at the image and pixel levels, respectively. Style transformation methods, including **DualNorm** [23], **CDDSA** [31], and **MixStyle** [56], modified the style of images to improve domain generalization. Dynamic convolution-based approaches, such as **DCAC** [48], leveraged dynamic convolutions during training to adapt the model. Methods like **DoFE** [30] focused on learning domain-invariant features by utilizing prior knowledge of different domains. Feature decomposition and recomposition served as the foundation for **DCANet** [43]. **DSC** [53] sampled the styles of potential target domains and aligned the graph semantic space samples. **UniAda** [50] learned a domain-aware base model and adapted it to out-of-distribution domains during inference. **MRFP** [66] achieved domain generalization by facilitating the learning of domain-agnostic semantic features. Additionally, **NormAUG** [58], **STEA** [59] and **CDSA** [62] focused on the correlation and features between multiple domains for data augmentation. The segmentation performance of these methods was evaluated using the Dice similarity coefficient (Dice). Two-tailed paired Wilcoxon test of Dice was performed to compare the difference between our method and related methods, and $p < 0.05$ was considered to be statistically significant.

TABLE II
THE QUANTITATIVE COMPARISON (DICE) OF DIFFERENT METHODS IN RETINAL LAYER SEGMENTATION DATASET.

Methods	Domain 1	Domain 2	Domain 3	Domain 4	Avg
Upper bound	82.04	89.45	85.78	94.18	87.86
Lower bound	45.92	45.97	47.18	41.77	45.21
DCAC [48]	68.79‡	71.06‡	69.29‡	70.05‡	69.80‡
CDDSA [31]	67.33‡	75.37‡	60.16‡	77.71‡	70.14‡
DoFE [30]	72.55‡	71.84‡	69.04‡	77.98‡	72.85‡
DCANet [43]	71.54‡	74.85‡	70.38‡	79.73‡	74.13‡
Mixup [55]	73.40‡	76.83‡	70.00‡	80.03‡	75.06‡
MixStyle [56]	72.08‡	76.80‡	70.09‡	82.01‡	75.24‡
DualNorm [23]	73.53‡	73.79‡	73.31‡	81.34‡	75.49‡
CDSA [62]	74.23‡	74.72‡	70.53‡	82.71‡	75.55‡
Cutmix [57]	73.27‡	76.63‡	71.78‡	82.40‡	76.02‡
DSC [53]	74.22‡	79.52‡	69.12‡	81.32‡	76.04‡
MRFP [66]	74.28‡	77.92‡	72.44‡	81.33‡	76.49‡
SETA [59]	73.86‡	76.16‡	72.94‡	83.06‡	76.51‡
NormAUG [58]	74.34‡	78.93‡	73.83‡	81.32‡	77.11‡
UniAda [50]	74.20‡	79.64‡	73.42‡	81.41‡	77.17‡
Ours	**74.98**	**80.59**	**74.56**	**83.17**	**78.32**

‡ denotes $p < 0.001$ of two-tailed paired Wilcoxon test.

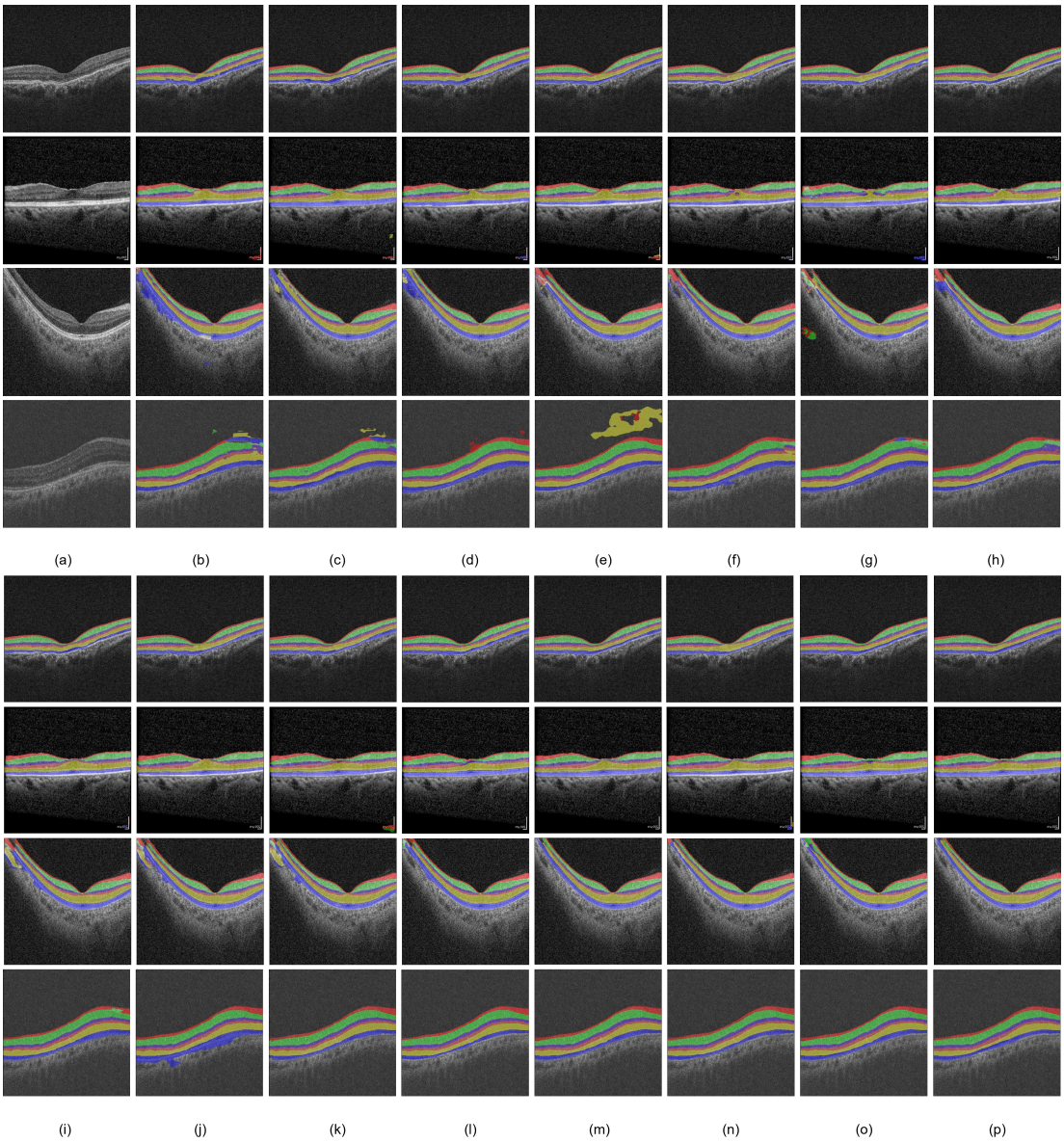

Fig. 4. Visual comparison in four domains between different DG methods for retinal layer segmentation. (a) Original image; (b) DCAC; (c) CDDSA; (d) DoFE; (e) DCANet; (f) Mixup; (g) MixStyle; (h) DualNorm; (i) CDSA; (j) Cutmix; (k) DSC; (l) MRFP; (m) SETA; (n) NormAUG; (o) UniAda; (p) Ours;

2) Experiments on Retinal Layer Segmentation: Table II reported the segmentation performance on the retinal layer segmentation dataset. The upper bound showed the highest performance and achieved an average Dice score of 87.86 across the three domains, and the lower bound only reached an average Dice score of 45.21. Due to the style differences between different domains, a gap of 40+ was generated, but our method improved the Dice score to 78.32. Compared to other traditional data augmentation methods such as CutMix [57] and Mixup [55], our method outperformed them by 3.26 and 2.30, respectively. For the recent enhancement method NormAUG [58], Dice score of our method was improved by 1.21. $p < 0.001$ of two-tailed paired Wilcoxon test showed that the superiority of our method for retinal layer segmentation was statistically significant. Fig.4 showed a visual comparison of our proposed method with 12 previous methods on four target domains. This proved that our method outperformed other DG methods in the task of retinal layer segmentation.

3) Experiments on Retinal Fluid Segmentation: Table III presented the evaluation results in terms of Dice scores for IRF, SRF, PED 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 CutMix [57] and Mixup [55], our method achieved an improvement of 2.87 and 3.12 respectively. Compared to our previous method, the proposed method achieved a performance improvement of 1.21 over DSC [53]. $p < 0.001$ of two-tailed paired Wilcoxon test showed that the superiority of our method for retinal fluid segmentation was statistically significant. The experiments demonstrated that achieved better segmentation results.

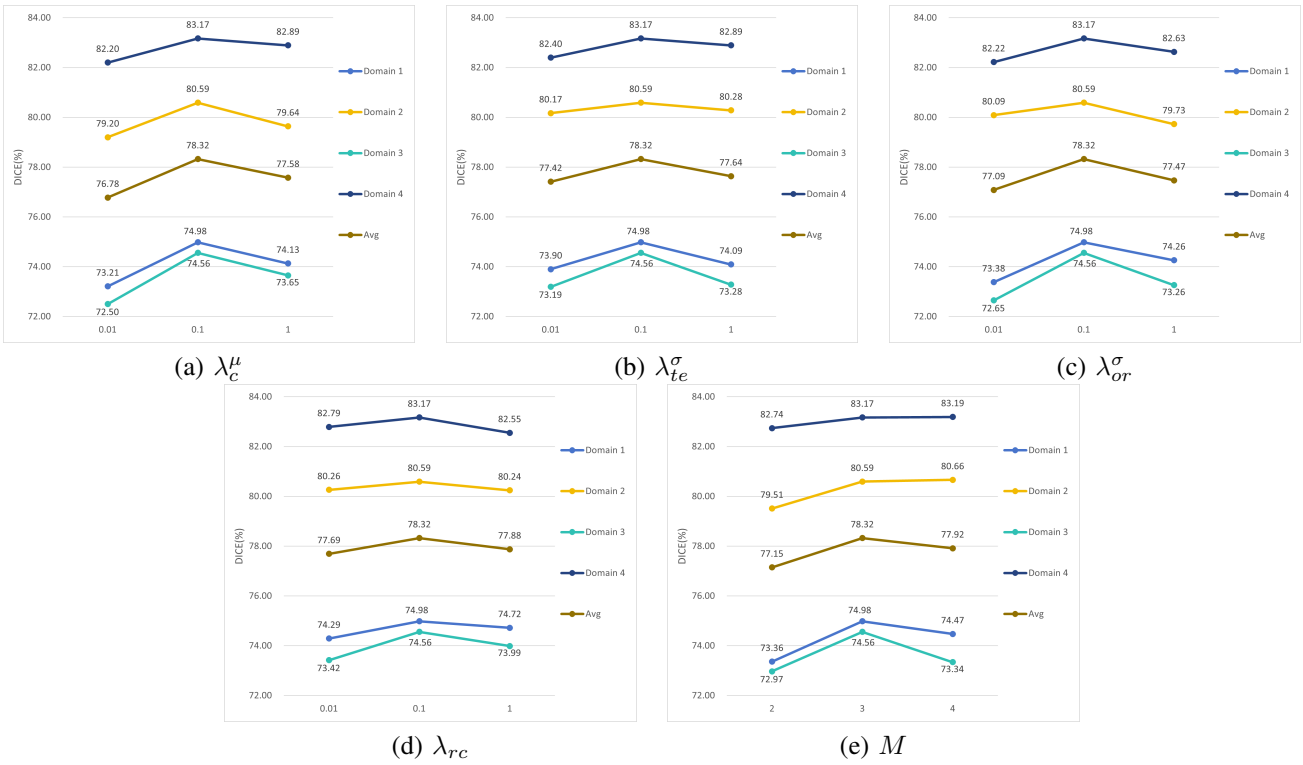

Fig. 5. Analysis of five parameters($\boldsymbol{\lambda_c^\mu}, \boldsymbol{\lambda_{te}^\sigma}, \boldsymbol{\lambda_{or}^\sigma}, \boldsymbol{\lambda_{rc}}, \boldsymbol{M}$) on Retinal Layer Segmentation.

TABLE III
THE QUANTITATIVE COMPARISON (DICE) OF DIFFERENT METHODS IN RETINAL FLUID SEGMENTATION DATASET.

Methods	Domain 1	Domain 2	Domain 3	Avg
Upper bound	61.45	66.65	58.53	62.21
Lower bound	21.79	13.36	23.98	19.71
DCAC [48]	44.47‡	55.12‡	55.94‡	51.84‡
DoFE [30]	41.83‡	53.66‡	61.85‡	52.45‡
DCANet [43]	48.63‡	55.35‡	57.14‡	53.71‡
CDDSA [31]	50.83‡	55.83‡	58.88‡	55.18‡
DualNorm [23]	53.57‡	58.24‡	56.51‡	56.11‡
NormAUG [58]	50.17‡	58.34‡	60.96‡	56.49‡
SETA [59]	58.09‡	54.76‡	57.94‡	56.93‡
CDSA [62]	56.36‡	57.44‡	59.05‡	57.62‡
MixStyle [56]	55.93‡	56.47‡	63.78‡	58.73‡
Mixup [55]	56.10‡	57.10‡	63.09‡	58.76‡
MRFP [66]	56.52‡	58.25‡	61.92‡	58.89‡
Cutmix [57]	56.88‡	56.86‡	63.30‡	59.01‡
UniAda [50]	57.24‡	60.33‡	63.62‡	60.39‡
DSC [53]	57.55‡	59.74‡	64.73‡	60.67‡
Ours	**60.18**	**60.46**	**65.00**	**61.88**

‡ denotes $p < 0.001$ of two-tailed paired Wilcoxon test.

TABLE IV
ABLATION EXPERIMENTS OF THE RETINAL LAYER SEGMENTATION DATASET (DICE).

Methods	Domain 1	Domain 2	Domain 3	Domain 4	Avg
Upper bound	82.04	89.45	85.78	94.18	87.86
Lower bound	45.92	45.97	47.18	41.77	45.21
Baseline	72.52	75.29	69.89	80.89	74.65
+KT	73.21	77.72	72.04	81.44	76.10
+MSCL	72.83	77.63	72.25	81.81	76.13
+VSO	72.65	77.72	71.75	81.22	75.84
+KT+MSCL	74.05	78.87	73.13	82.64	77.17
+KT+VSO	74.19	79.66	73.42	82.32	77.40
+MSCL+VSO	74.24	79.87	73.26	82.48	77.46
Ours	**74.98**	**80.59**	**74.56**	**83.17**	**78.32**

4) Analysis of Different Methods: Kindly note that the performance of each previous method varied across different target domains, due to differences in model generalization capabilities. For example, although NormAUG [58] performed better in retinal layer segmentation, it degraded in lesion segmentation, while DSC [53] had good results in lesion segmentation, but did not have significant advantages in retinal layer segmentation. MRFP [66] and UniAda [50] performed well in two datasets. Traditional data augmentation methods such as CutMix [57] and Mixup [55] have shown relatively stable performance on both datasets, which might be attributed to their significant data augmentation. Other methods such as CDDSA [31] and DualNorm [23] did not show significant advantages on either dataset. In summary, due to the different semantic information of different tasks and differences in datasets, as well as the different principles and focuses of different methods, it was difficult for each method to achieve significant advantages on both datasets.

D. Ablation Study

The proposed method consisted of three primary components: Kurtosis Transformation (KT), Mean Style Contrastive Learning (MSCL), and Variance Style Orthogonalization (VSO). Ablation experiments were performed on the

retinal layer segmentation dataset to evaluate each component. Baseline was DeepLabv3+ [65] with ResNet101.

1) Analysis of KT: Compared to the lower bound, the baseline achieved a noticeable improvement in the network's generalization and the average Dice score improved by 29.44. As shown in the 4th row in Table IV. After adding KT compared to the baseline, it has been improved by 1.45. This indicated that KT could simulate potential target domains and improve the model's generalization ability by generating kurtosis-transformed samples that were different from the source domain.

2) Analysis of MSCL: The 5th row of Table IV showed the impact of adding only mean style contrastive learning. Compared to the baseline, the average Dice score has been improved by 1.48. The 7th row of Table IV showed the results of adding MSCL after adding KT and the average Dice score has been further improved by 1.04. Compared to just adding KT, the average Dice score increased by 1.07. This indicated that MSCL was able to generate mean style features that are different from the source domains but more uniform to facilitate the generation of more diverse styles.

3) Analysis of VSO: The 6th row of Table IV showed the impact of adding only variance style orthogonalization. Compared to the baseline, the average Dice score has been improved by 1.19. The 8th row of Table IV showed the results of adding VSO after adding KT. Compared to just adding KT, the average Dice score increased by 1.30. The 9th row of Table IV showed the effect of adding MSCL and VSO, which showed a significant improvement compared to using them separately. This indicated that VSO improved the diversity and independence of potential T-styles in a high-level orthogonal space by imposing an orthogonal constraint on the reparametrized variance styles. After combining KT, a more significant improvement was achieved to a Dice score of 78.32. This further demonstrated that each component of our method was effective and proved the significant effectiveness of our method in retinal layer segmentation tasks.

4) Analysis of different parameters: In order to further explore the scalability of the method, we conducted a series of experiments on several major hyper-parameters. Fig.5 shows the experimental results on retinal layer segmentation after adjusting several major parameters:λ_c^μ, λ_{te}^σ, λ_{or}^σ and λ_{rc}, and M. Initially set to $(\lambda_c^\mu = 0.1, \lambda_{te}^\sigma = 0.1, \lambda_{or}^\sigma = 0.1, \lambda_{rc} = 0.1, M = 3)$, each parameter was increased or decreased while keeping the others. The results indicated that the optimal parameters for domain 1 were $(\lambda_c^\mu = 0.1, \lambda_{te}^\sigma = 0.1, \lambda_{or}^\sigma = 0.1, \lambda_{rc} = 0.1, M = 3)$, while the optimal parameters for domains 2, 3, 4 and all domains were $(\lambda_c^\mu = 0.1, \lambda_{te}^\sigma = 0.1, \lambda_{or}^\sigma = 0.1, \lambda_{rc} = 0.1, M = 4)$, $(\lambda_c^\mu = 0.1, \lambda_{te}^\sigma = 0.1, \lambda_{or}^\sigma = 0.1, \lambda_{rc} = 0.1, M = 3)$, $(\lambda_c^\mu = 0.1, \lambda_{te}^\sigma = 0.1, \lambda_{or}^\sigma = 0.1, \lambda_{rc} = 0.1, M = 4)$, $(\lambda_c^\mu = 0.1, \lambda_{te}^\sigma = 0.1, \lambda_{or}^\sigma = 0.1, \lambda_{rc} = 0.1, M = 3)$, respectively. Therefore, for parameters λ_c^μ, λ_{te}^σ, λ_{or}^σ and λ_{rc}, independently increasing or decreasing them had a certain degree of negative impact on the segmentation results, which may be due to the imbalance of each component in the total loss caused by the amplification or decline of a single loss function. In terms of setting the number of intervals M, theoretically, a larger M will lead to a more refined interval kurtosis detection, and we attempted to increase and decrease M. From the results, it can be seen that reducing M significantly degraded the segmentation results, while increasing M showed that larger M performed better in domains 2 and 4, but performed less well in domains 1 and 3. So finer interval kurtosis detection may not necessarily achieve better results due to the diversity of domain distributions, and larger M further increased training time and computational burden. Therefore, under normal conditions, the parameters were $(\lambda_c^\mu = 0.1, \lambda_{te}^\sigma = 0.1, \lambda_{or}^\sigma = 0.1, \lambda_{rc} = 0.1, M = 3)$ may be better.

V. Conclusion

Different manufacturers' OCT devices produce the similar fundus OCT images but deep learning models often fail to cope with the different image appearances and contrasts since distribution differences exist between these different manufacturers' OCT images. The degeneration of deep learning models becomes more severe when target domain is unseen. To produce the T-styles of the potential target domain, we need to generate more samples with richer style features for deep learning models. Kurtosis transformation is proposed to modify the distributions of multi-source domains and produce diverse style features to improve the generalization performance of the model in unseen domains. Then, MSCL is used to generate mean style features that are different from the source domains but more uniform to facilitate the generation of more diverse styles. VSO is used to generate variance style features with rich semantic information. With the help of more uniform mean style features, reconstruction can be used to obtain samples with more diverse intensity distributions and image contrasts that are different from the source domains. Finally, by modulating the segmentation network encoder with mean and variance style features, and feeding diverse samples into the network, the overall generalization performance of the model can be therefore improved. Our method has not only achieved significant improvements over traditional lesion segmentation, but also achieved excellent results in the task of retinal layer segmentation. In future work, we will further analyze the distribution characteristics of samples from different domains, and delve deeper into the potential connections and differences between different domains to improve domain generalization.

References

[1] D. Huang, E. A. Swanson, C. P. Lin, J. S. Schuman, W. G. Stinson, W. Chang, M. R. Hee, T. Flotte, K. Gregory, C. A. Puliafito *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," *Annual review of biomedical engineering*, vol. 19, pp. 221–248, 2017.

[3] G. Litjens, T. Kooi, B. E. Bejnordi, A. A. A. Setio, F. Ciompi, M. Ghafoorian, J. A. Van Der Laak, B. Van Ginneken, and C. I. Sánchez, "A survey on deep learning in medical image analysis," *Medical image analysis*, vol. 42, pp. 60–88, 2017.

[4] X. Xie, J. Niu, X. Liu, Z. Chen, S. Tang, and S. Yu, "A survey on incorporating domain knowledge into deep learning for medical image analysis," *Medical Image Analysis*, vol. 69, p. 101985, 2021.

[5] Q. Li, S. Li, Z. He, H. Guan, R. Chen, Y. Xu, T. Wang, S. Qi, J. Mei, and W. Wang, "Deepretina: layer segmentation of retina in oct images using deep learning," *Translational vision science & technology*, vol. 9, no. 2, pp. 61–61, 2020.

[6] K. Gopinath, S. B. Rangrej, and J. Sivaswamy, "A deep learning framework for segmentation of retinal layers from oct images," in *2017 4th IAPR Asian conference on pattern recognition (ACPR)*. IEEE, 2017, pp. 888–893.

[7] J. Hao, J. Liu, E. Pereira, R. Liu, J. Zhang, Y. Zhang, K. Yan, Y. Gong, J. Zheng, J. Zhang *et al.*, "Uncertainty-guided graph attention network for parapneumonic effusion diagnosis," *Medical Image Analysis*, vol. 75, p. 102217, 2022.

[8] M. Wang, T. Lin, Y. Peng, W. Zhu, Y. Zhou, F. Shi, K. Yu, Q. Meng, Y. Liu, Z. Chen *et al.*, "Self-guided optimization semi-supervised method for joint segmentation of macular hole and cystoid macular edema in retinal oct images," *IEEE Transactions on Biomedical Engineering*, 2023.

[9] Y. Luo, L. Zheng, T. Guan, J. Yu, and Y. Yang, "Taking a closer look at domain shift: Category-level adversaries for semantics consistent domain adaptation," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2019, pp. 2507–2516.

[10] K. Stacke, G. Eilertsen, J. Unger, and C. Lundström, "Measuring domain shift for deep learning in histopathology," *IEEE journal of biomedical and health informatics*, vol. 25, no. 2, pp. 325–336, 2020.

[11] Y. Sun, D. Dai, and S. Xu, "Rethinking adversarial domain adaptation: Orthogonal decomposition for unsupervised domain adaptation in medical image segmentation," *Medical Image Analysis*, vol. 82, p. 102623, 2022.

[12] Z. Zheng, R. Li, and C. Liu, "Learning robust features alignment for cross-domain medical image analysis," *Complex & Intelligent Systems*, pp. 1–15, 2023.

[13] T. Ilyas, K. Ahmad, D. M. S. Arsa, Y. C. Jeong, and H. Kim, "Enhancing medical image analysis with unsupervised domain adaptation approach across microscopes and magnifications," *Computers in Biology and Medicine*, p. 108055, 2024.

[14] K. Zhou, Z. Liu, Y. Qiao, T. Xiang, and C. C. Loy, "Domain generalization: A survey," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 45, no. 4, pp. 4396–4415, 2022.

[15] J. Wang, C. Lan, C. Liu, Y. Ouyang, T. Qin, W. Lu, Y. Chen, W. Zeng, and S. Y. Philip, "Generalizing to unseen domains: A survey on domain generalization," *IEEE transactions on knowledge and data engineering*, vol. 35, no. 8, pp. 8052–8072, 2022.

[16] F. Qiao, L. Zhao, and X. Peng, "Learning to learn single domain generalization," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2020, pp. 12 556–12 565.

[17] X. Chen, C. Lian, L. Wang, H. Deng, T. Kuang, S. Fung, J. Gateno, P.-T. Yap, J. J. Xia, and D. Shen, "Anatomy-regularized representation learning for cross-modality medical image segmentation," *IEEE transactions on medical imaging*, vol. 40, no. 1, pp. 274–285, 2020.

[18] H. Liu, Y. Fan, Z. Xu, B. M. Dawant, and I. Oguz, "Learning site-specific styles for multi-institutional unsupervised cross-modality domain adaptation," *arXiv e-prints*, pp. arXiv–2311, 2023.

[19] D. Lowd and C. Meek, "Adversarial learning," in *Proceedings of the eleventh ACM SIGKDD international conference on Knowledge discovery in data mining*, 2005, pp. 641–647.

[20] B. H. Zhang, B. Lemoine, and M. Mitchell, "Mitigating unwanted biases with adversarial learning," in *Proceedings of the 2018 AAAI/ACM Conference on AI, Ethics, and Society*, 2018, pp. 335–340.

[21] Y. Jing, Y. Yang, Z. Feng, J. Ye, Y. Yu, and M. Song, "Neural style transfer: A review," *IEEE transactions on visualization and computer graphics*, vol. 26, no. 11, pp. 3365–3385, 2019.

[22] A. Selim, M. A. Elgharib, and L. Doyle, "Painting style transfer for head portraits using convolutional neural networks." *ACM Trans. Graph.*, vol. 35, no. 4, pp. 129–1, 2016.

[23] Z. Zhou, L. Qi, X. Yang, D. Ni, and Y. Shi, "Generalizable cross-modality medical image segmentation via style augmentation and dual normalization," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2022, pp. 20 856–20 865.

[24] Z. Su, K. Yao, X. Yang, K. Huang, Q. Wang, and J. Sun, "Rethinking data augmentation for single-source domain generalization in medical image segmentation," in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 37, no. 2, 2023, pp. 2366–2374.

[25] J. Xiong, A. W. He, M. Fu, X. Hu, Y. Zhang, C. Liu, X. Zhao, and Z. Ge, "Improve unseen domain generalization via enhanced local color transformation," in *Medical Image Computing and Computer Assisted Intervention–MICCAI 2020: 23rd International Conference, Lima, Peru, October 4–8, 2020, Proceedings, Part II 23*. Springer, 2020, pp. 433–443.

[26] M. Salvi, F. Branciforti, F. Molinari, and K. M. Meiburger, "Generative models for color normalization in digital pathology and dermatology: Advancing the learning paradigm," *Expert Systems with Applications*, vol. 245, p. 123105, 2024.

[27] K. Muandet, D. Balduzzi, and B. Schölkopf, "Domain generalization via invariant feature representation," in *International conference on machine learning*. PMLR, 2013, pp. 10–18.

[28] H. Li, Y. Wang, R. Wan, S. Wang, T.-Q. Li, and A. Kot, "Domain generalization for medical imaging classification with linear-dependency regularization," *Advances in neural information processing systems*, vol. 33, pp. 3118–3129, 2020.

[29] H. Lai, Y. Luo, B. Li, G. Zhang, and J. Lu, "Domain-aware dual attention for generalized medical image segmentation on unseen domains," *IEEE Journal of Biomedical and Health Informatics*, 2023.

[30] S. Wang, L. Yu, K. Li, X. Yang, C.-W. Fu, and P.-A. Heng, "Dofe: Domain-oriented feature embedding for generalizable fundus image segmentation on unseen datasets," *IEEE Transactions on Medical Imaging*, vol. 39, no. 12, pp. 4237–4248, 2020.

[31] R. Gu, G. Wang, J. Lu, J. Zhang, W. Lei, Y. Chen, W. Liao, S. Zhang, K. Li, D. N. Metaxas *et al.*, "Cddsa: Contrastive domain disentanglement and style augmentation for generalizable medical image segmentation," *Medical Image Analysis*, vol. 89, p. 102904, 2023.

[32] Y. Bi, Z. Jiang, R. Clarenbach, R. Ghotbi, A. Karlas, and N. Navab, "Mi-segnet: Mutual information-based us segmentation for unseen domain generalization," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, 2023, pp. 130–140.

[33] H. Kim, Y. Shin, and D. Hwang, "Dimix: Disentangle-and-mix based domain generalizable medical image segmentation," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, 2023, pp. 242–251.

[34] Q. Liu, Q. Dou, L. Yu, and P. A. Heng, "Ms-net: Multi-site network for improving prostate segmentation with heterogeneous mri data," *IEEE Transactions on Medical Imaging*, vol. 39, no. 9, pp. 2713–2724, 2020.

[35] J. Hu, X. Gu, Z. Wang, and X. Gu, "Mixture of calibrated networks for domain generalization in brain tumor segmentation," *Knowledge-Based Systems*, vol. 270, p. 110520, 2023.

[36] P. Khandelwal and P. Yushkevich, "Domain generalizer: A few-shot meta learning framework for domain generalization in medical imaging," in *Domain Adaptation and Representation Transfer, and Distributed and Collaborative Learning: Second MICCAI Workshop, DART 2020, and First MICCAI Workshop, DCL 2020, Held in Conjunction with MICCAI 2020, Lima, Peru, October 4–8, 2020, Proceedings 2*. Springer, 2020, pp. 73–84.

[37] Q. Liu, C. Chen, J. Qin, Q. Dou, and P.-A. Heng, "Feddg: Federated domain generalization on medical image segmentation via episodic learning in continuous frequency space," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2021, pp. 1013–1023.

[38] C. Wang, H. Li, W. Ma, G. Zhao, and Z. He, "Metascleraseg: an effective meta-learning framework for generalized sclera segmentation," *Neural Computing and Applications*, vol. 35, no. 29, pp. 21 797–21 826, 2023.

[39] H. Li, H. Li, W. Zhao, H. Fu, X. Su, Y. Hu, and J. Liu, "Frequency-mixed single-source domain generalization for medical image segmentation," in *Medical Image Computing and Computer Assisted Intervention – MICCAI 2023*, H. Greenspan, A. Madabhushi, P. Mousavi, S. Salcudean, J. Duncan, T. Syeda-Mahmood, and R. Taylor, Eds. Springer Nature Switzerland, pp. 127–136.

[40] Y. Bengio, A. C. Courville, and P. Vincent, "Unsupervised feature learning and deep learning: A review and new perspectives," *CoRR, abs/1206.5538*, vol. 1, no. 2665, p. 2012, 2012.

[41] J. Donahue, P. Krähenbühl, and T. Darrell, "Adversarial feature learning," *arXiv preprint arXiv:1605.09782*, 2016.

[42] Y. Wen, K. Zhang, Z. Li, and Y. Qiao, "A discriminative feature learning approach for deep face recognition," in *Computer vision–ECCV 2016: 14th European conference, amsterdam, the netherlands, October 11–14, 2016, proceedings, part VII 14*. Springer, 2016, pp. 499–515.

[43] R. Gu, J. Zhang, R. Huang, W. Lei, G. Wang, and S. Zhang, "Domain composition and attention for unseen-domain generalizable medical image segmentation," in *Medical Image Computing and Computer Assisted Intervention–MICCAI 2021: 24th International Conference, Strasbourg, France, September 27–October 1, 2021, Proceedings, Part III 24*. Springer, 2021, pp. 241–250.

[44] C. Shorten and T. M. Khoshgoftaar, "A survey on image data augmentation for deep learning," *Journal of big data*, vol. 6, no. 1, pp. 1–48, 2019.

[45] P. T. Jackson, A. A. Abarghouei, S. Bonner, T. P. Breckon, and B. Obara, "Style augmentation: data augmentation via style randomization." in *CVPR workshops*, vol. 6, 2019, pp. 10–11.

[46] A. Antoniou, "Data augmentation generative adversarial networks," *arXiv preprint arXiv:1711.04340*, 2017.
[47] Y. Li, M. Gong, X. Tian, T. Liu, and D. Tao, "Domain generalization via conditional invariant representations," in *Proceedings of the AAAI conference on artificial intelligence*, vol. 32, 2018, pp. 3580–3587.
[48] S. Hu, Z. Liao, J. Zhang, and Y. Xia, "Domain and content adaptive convolution based multi-source domain generalization for medical image segmentation," *IEEE Transactions on Medical Imaging*, vol. 42, no. 1, pp. 233–244, 2022.
[49] Y. Zhang, S. Tian, M. Liao, G. Hua, W. Zou, and C. Xu, "Learning shape-invariant representation for generalizable semantic segmentation," *IEEE Transactions on Image Processing*, vol. 32, pp. 5031–5045, 2023.
[50] Z. Zhang, Y. Chen, H. Yu, Z. Wang, S. Wang, F. Fan, H. Shan, and Y. Zhang, "Uniada: Domain unifying and adapting network for generalizable medical image segmentation," *IEEE Transactions on Medical Imaging*, 2024.
[51] R. H. Fick, A. Moshayedi, G. Roy, J. Dedieu, S. Petit, and S. B. Hadj, "Domain-specific cycle-gan augmentation improves domain generalizability for mitosis detection," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, 2021, pp. 40–47.
[52] J.-Y. Zhu, T. Park, P. Isola, and A. A. Efros, "Unpaired image-to-image translation using cycle-consistent adversarial networks," in *Proceedings of the IEEE international conference on computer vision*, 2017, pp. 2223–2232.
[53] S. Liao, T. Peng, H. Chen, T. Lin, W. Zhu, F. Shi, X. Chen, and D. Xiang, "Dual-spatial domain generalization for fundus lesion segmentation in unseen manufacturer's oct images," *IEEE Transactions on Biomedical Engineering*, 2024.
[54] L. Zhang, X. Wang, D. Yang, T. Sanford, S. Harmon, B. Turkbey, B. J. Wood, H. Roth, A. Myronenko, D. Xu *et al.*, "Generalizing deep learning for medical image segmentation to unseen domains via deep stacked transformation," *IEEE transactions on medical imaging*, vol. 39, no. 7, pp. 2531–2540, 2020.
[55] H. Zhang, M. Cisse, Y. N. Dauphin, and D. Lopez-Paz, "mixup: Beyond empirical risk minimization," *arXiv preprint arXiv:1710.09412*, 2017.
[56] K. Zhou, Y. Yang, Y. Qiao, and T. Xiang, "Mixstyle neural networks for domain generalization and adaptation," *International Journal of Computer Vision*, pp. 1–15, 2023.
[57] S. Yun, D. Han, S. J. Oh, S. Chun, J. Choe, and Y. Yoo, "Cutmix: Regularization strategy to train strong classifiers with localizable features," in *Proceedings of the IEEE/CVF international conference on computer vision*, 2019, pp. 6023–6032.
[58] L. Qi, H. Yang, Y. Shi, and X. Geng, "Normaug: Normalization-guided augmentation for domain generalization," *IEEE Transactions on Image Processing*, vol. 33, pp. 1419–1431, 2024.
[59] J. Guo, L. Qi, Y. Shi, and Y. Gao, "Seta: Semantic-aware edge-guided token augmentation for domain generalization," *IEEE Transactions on Image Processing*, 2024.
[60] W. Xu, C. Long, and Y. Nie, "Learning dynamic style kernels for artistic style transfer," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2023, pp. 10 083–10 092.
[61] M. E. Mortenson, *Mathematics for computer graphics applications*. Industrial Press Inc., 1999.
[62] M. Wang, Y. Liu, J. Yuan, S. Wang, Z. Wang, and W. Wang, "Inter-class and inter-domain semantic augmentation for domain generalization," *IEEE Transactions on Image Processing*, 2024.
[63] X. Huang and S. Belongie, "Arbitrary style transfer in real-time with adaptive instance normalization," in *Proceedings of the IEEE international conference on computer vision*, 2017, pp. 1501–1510.
[64] H. Bogunović, F. Venhuizen, S. Klimscha, S. Apostolopoulos, A. Bab-Hadiashar, U. Bagci, M. F. Beg, L. Bekalo, Q. Chen, C. Ciller *et al.*, "Retouch: The retinal oct fluid detection and segmentation benchmark and challenge," *IEEE transactions on medical imaging*, vol. 38, no. 8, pp. 1858–1874, 2019.
[65] L.-C. Chen, Y. Zhu, G. Papandreou, F. Schroff, and H. Adam, "Encoder-decoder with atrous separable convolution for semantic image segmentation," in *Proceedings of the European conference on computer vision (ECCV)*, 2018, pp. 801–818.
[66] S. Udupa, P. Gurunath, A. Sikdar, and S. Sundaram, "Mrfp: Learning generalizable semantic segmentation from sim-2-real with multi-resolution feature perturbation," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2024, pp. 5904–5914.