

# Fundus Vascular Segmentation Based on Data Enhancement and Invariant Feature Extraction

Lei Cheng<sup>a</sup>, Yumeng Li<sup>b</sup>, Jingyi Han<sup>c</sup>

Beijing Jiaotong University

<sup>a</sup> 21231092@bjtu.edu.cn, <sup>b</sup> 21231109@bjtu.edu.cn, <sup>c</sup> 21271062@bjtu.edu.cn

Project 20231000411221 supported by National Training Program of Innovation and Entrepreneurship for Undergraduates

**Abstract:** Deep neural networks have emerged as the predominant method for medical image segmentation, owing to their robust feature learning capabilities, enabling accurate automatic segmentation of target structures within intricate medical images. Fundus images, being crucial in diagnosing ophthalmic diseases, underscore the importance of effective segmentation techniques. However, fundus vascular images pose challenges due to their high complexity and subtle individual differences, necessitating improvement in existing segmentation methodologies for enhanced disease classification accuracy. This paper introduces a fundus blood vessel segmentation model, employing data augmentation and invariant feature extraction, to systematically tackle the core challenges in medical image processing, particularly in fundus blood vessel segmentation. These challenges include limited source domain samples and inadequate model domain generalization. The model adopts a dual-dimensional strategy. Firstly, it delves into data augmentation technology to enhance the diversity and representativeness of samples within the finite source domain. This is achieved through an image enhancement module based on Fourier transform, mitigating the impact of data scarcity on model training effectiveness. Secondly, the research focuses on interdomain invariant feature extraction, aiming to extract feature representations that consistently characterize fundus blood vessel structure and pathology across different data distributions, thereby enhancing model generalization performance in unfamiliar domains. Specifically, the paper designs a fundus image enhancement module based on Fourier transform in the data augmentation dimension. In the feature extraction dimension, it proposes a normalization module based on uncertainty theory, departing from traditional normalization methods. Experimental results demonstrate the efficacy of the proposed method, showcasing superior generalization performance compared to existing techniques in retinal blood vessel and OD/OC segmentation tasks. Experience validates the model-agnostic nature of the learned strategy, indicating its potential for seamless transferability to other models, thereby offering robust support for advancing medical image segmentation research and applications.

**Key words:** deep learning; image segmentation; fundus blood vessels; convolutional neural network

## 1. Introduction

Medical image processing and analysis plays an important role in contemporary medical diagnosis. Segmentation of medical images can be used to extract the structure or lesion region of interest in the image, helping doctors to accurately diagnose and quantitatively analyze diseases, such as tumors and cardiovascular lesions. Fundus vascular image segmentation refers to the precise pixel-level division of vascular structures in fundus images to separate vascular regions from the fundus background and other tissue structures. This process is an important step in the diagnosis, disease monitoring, treatment planning and prognosis assessment of ophthalmic diseases, especially in the management of retinal vascular diseases such as glaucoma, diabetic retinopathy, hypertensive retinopathy and others. With the rise of deep learning methods, many advanced techniques based on deep learning have emerged in the field of medical image segmentation, and deep learning models such as Convolutional Neural Network (CNN), Full Convolutional Networks (FCN), U-Net, DeepLab, etc., are commonly used to learn the vascular segmentation mask directly from the original image in an end-to-end manner. Many scholars have already conducted research on the



above models, Olaf Ronneberger et al. studied the segmentation performance of U-net network on small targets and complex structures [1]. Jonathan Long et al. proposed the concept of FCN for the first time [2], which applies CNN directly to pixel-level image segmentation task without the need of a fully-connected layer, and end-to-end pixel-level prediction was realized. Scholars such as Kaiming He have also considered the impact that Residual Network (ResNet) structures [3] have had on the field of image segmentation. On these foundations, George Papandreou, Liang-Chieh Chen and others have proposed an Encoder-Decoder architecture (DeepLabv3+) using Atrous Convolution and Separable Convolution [4]. The DeepLab family of models has achieved excellent performance on several semantic segmentation benchmark datasets and has become one of the dominant models in the field. Neural networks are able to capture subtle structural changes and signs of abnormality when processing high-dimensional, high-resolution medical images, which helps to achieve early detection of diseases.

Fundus images are characterized by high complexity and small individual differences, making the accuracy of existing segmentation techniques in fundus disease image classification tasks still insufficient. Although CNN have been widely and effectively used in automated OD/OC, retinal vascular and lesion segmentation tasks, testing usually assumes a similar distribution of training and test data, an assumption that often does not hold in reality. Test fundus images vary significantly in appearance, contrast, quality, and field of view due to different scanners and settings, posing a challenge to well-trained CNN models, and performance may be significantly degraded. If the differences between the source and target domains are too large, the model may focus too much on the features of the source domain, leading to overfitting or underfitting problems on the target domain, which may result in inaccurate predictions or poor generalization of the model on the target domain. In order to deal with the domain transfer problem, researchers have proposed a series of Domain Transfer (DT) methods and techniques, such as instance-based migration [5], feature-based migration [6], and model-based migration [7]. These methods aim to reduce the differences between domains by extracting the shared knowledge or feature representations of the source and target domains, but DT may be more advantageous when the source and target domains have high task similarity. On this basis, Domain Adaptation (DA) is more compatible with training on multi-style data. It adjusts the model of the source domain to fit the data distribution of the target domain by learning the differences between the source and target domains. However, in the case of deep learning and complex modeling, training and adjusting the model to fit the data distribution of different domains may require a lot of time and computational power, which is inappropriate for practical application scenarios. In contrast, domain generalization (DG) [8] has a broader focus, which does not rely on a specific target domain, but emphasizes the model's ability to generalize to arbitrary unknown domains, allowing the model to maintain good performance even on new data. Early researchers began to focus on the model's generalization performance on different data distributions, and proposed some preliminary theoretical frameworks and methods, such as regularization techniques in transfer learning [9] and feature selection [10]. With the breakthrough progress of Deep Neural Networks (DNN) in image recognition, natural language processing and other fields, researchers began to explore how to use deep learning models for DG. Initial attempts such as deep CNN trained on multiple source domains to improve the generalization ability appeared in this period [11, 12]. Specialized theoretical frameworks and algorithms for the DG problem began to appear. Examples include meta-learning-based DG methods [13], methods that use adversarial training to model domain bias [14], and strategies based on data augmentation and simulation [15]. Recent research further combines multidisciplinary theories such as information theory, graph theory, statistics, etc., to propose more refined feature learning and regularization strategies, such as information theory-inspired feature untangling methods [16].

Therefore, in this paper, we consider two aspects of data enhancement [17, 18] and feature learning [19, 20], which enable the network to learn more invariant features by increasing the diversity and number of training samples, and further improve the generalization ability of the model by extracting more abstract and generalized feature representations from the enhanced data.



The performance of the depth model is closely related to the diversity of samples. Compared with natural images, the data volume of medical images is extremely limited, and the limited data volume and single distribution of fundus vascular image samples are not conducive to the improvement of model segmentation performance. If the original data are transformed and expanded, more diversified data samples can be generated, which can help the model learn the features that are invariant to different transformations and perturbations, and reduce the overfitting phenomenon of the model. According to the characteristics of the magnitude spectrum and phase spectrum of the image, as shown in Fig. 1, we envision to retain the semantic features of the phase spectrum, design different magnitude transformation methods to transform the original magnitude spectrum, and finally combine the original phase spectrum and the transformed magnitude spectrum to obtain semantically consistent and stylistically diverse augmented images. Using multiple augmented images as network inputs can improve the segmentation capability of the model from the perspective of sample diversity. There have been many magnitude spectrum transformation methods that have achieved significant results in different tasks and domains. The magnitude spectrum represents the intensity or amplitude distribution of different spatial frequencies in an image. It provides information about the energy distribution of an image in the frequency domain. In image classification tasks, data samples are often enhanced by geometric transformations such as random rotation, translation, scaling or adding noise [21]. Inspired by the above, we perform linear interpolation between the magnitude spectra of two images in any source domain. With the Fourier transform, we can transform the images from the original null domain to the frequency domain and perform enhancement operations in the frequency domain. By adjusting the magnitude spectra of different frequency components, we can generate images with rich diversity and style. Such an image enhancement method will help to improve the performance of fundus vascular image segmentation model on different styles of data.

In some machine learning and deep learning methods, large differences in the range of feature values may lead to overfitting of the model during training. By normalizing the feature values, the risk of overfitting can be reduced and the generalization ability of the model can be improved to better adapt to different image segmentation tasks. Existing normalization methods usually extract domain-invariant features by simple methods such as IN [22], BN [23] or a combination of both, which leads to a single distribution of features extracted by the network and insufficient feature expression ability. Some studies have extracted the mean and variance features of images by normalizing them [24, 25] and used them for image processing tasks. When the training data and test data come from different distributions, the model may face the DT problem in the testing phase, and the difference in the probability distribution between the training data and the test data due to the DT, this difference dissimilarity may introduce uncertainty that affects the model's generalization ability and performance stability. The predictions of deep learning models usually give only deterministic results, but in practical applications, understanding the uncertainty of model predictions can provide more comprehensive information. Therefore, there is a lot of research devoted to developing methods to estimate model uncertainty. For example, Bayesian deep learning, [26] Monte Carlo inference [27], Dropout [28], and other methods are used to estimate model prediction uncertainty. To address this problem, in this paper, we enhance the diversity and robustness of feature extraction by introducing uncertainty based on the mean and variance of feature extraction previously studied. Specifically, we make full use of the variation range of features by adding uncertainty to the normalization layer. By introducing uncertainty, we can better reflect the network's confidence level in different features and use it to adjust the representation of features. This uncertainty-based normalization module will help to improve the richness and diversity of features and enable the network to better express different classes of features.



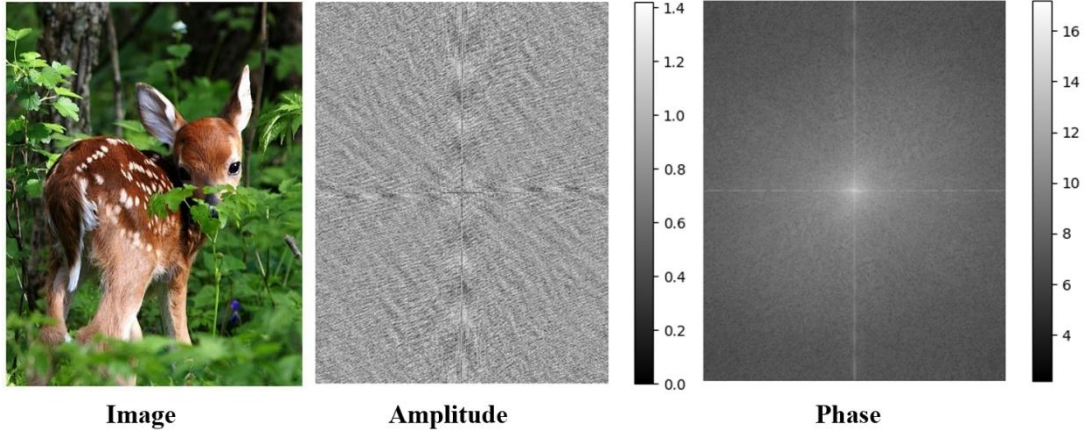


Fig.1. The amplitude spectrum and phase spectrum obtained by Fourier transform

Aiming at the problems of insufficient source domain sample size and low DG performance, the innovation of this paper is mainly reflected in the fact that the generalization performance of fundus image segmentation model is effectively improved through two aspects of data augmentation and inter-domain invariant feature extraction, which provides a new idea and method for solving the problems of insufficient source domain sample size and low DG performance.

## 2. Method

Drawing inspiration from the semantic preservation attributes elucidated in the works concerning Fourier phase components [29, 30], we posit that models accentuating phase information manifest superior cross-domain generalization capacities. Consequently, we propose a data augmentation framework (*Aug-DSU*) within the context of segmentation networks, embracing uncertainty considerations, as depicted in Fig. 2. This framework comprises two integral constituents: the Fourier-based data augmentation module and the uncertainty normalization module. Herein, we delineate the core constituents of this methodology, specifically the Fourier-based data augmentation and the uncertainty-based normalization module.

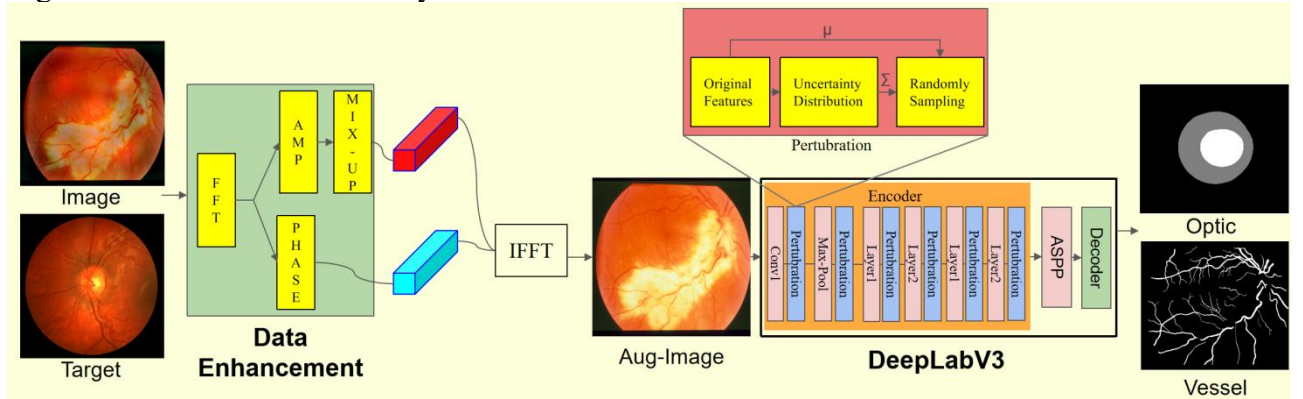


Fig.2. The framework of the proposed *Aug-DSU*. Our framework contains two key components, namely Fourier-based data augmentation and uncertainty normalization modules, emphasized in bold. The process commences with the original phase and transformed amplitude spectrum, which undergo inverse Fourier transform, yielding an enhanced image. Subsequently, integration of a normalization module grounded in uncertainty is introduced into the DeepLabV3 network.

### 2.1 Data enhancement based on Fourier transform

For single-channel images (grayscale images), the formula for the Fourier transform is the two-dimensional discrete Fourier transform (2DDFT). Suppose the image is  $f(x, y)$ , where  $x$  and  $y$  represent the position in the row and column directions of the image, respectively, and the image



size is  $M \times N$  (the number of rows and columns of the image, respectively). The two-dimensional discrete Fourier transform converts the image from the spatial domain to the frequency domain, and the resulting spectral image is represented as  $F(u, v)$ , where  $u$  and  $v$  are coordinates in the frequency domain. The specific formula is as follows:

$$F(u, v) = \sum_{x=0}^{M-1} \sum_{y=0}^{N-1} f(x, y) \cdot e^{-j2\pi(\frac{ux}{M} + \frac{vy}{N})} \quad (1)$$

Where  $f(x, y)$  is the pixel value of the image's position  $(x, y)$  in the spatial domain, and  $F(u, v)$  is the spectral value of the image's frequency coordinate  $(u, v)$  in the frequency domain. The exponential function represents a complex rotation. Given the spectrum image  $F(u, v)$ , the inverse transformation of the original image  $f(x, y)$  can be obtained by the formula:

$$f(x, y) = \frac{1}{MN} \sum_{u=0}^{M-1} \sum_{v=0}^{N-1} F(u, v) \cdot e^{j2\pi(\frac{ux}{M} + \frac{vy}{N})} \quad (2)$$

Among them, the division by  $MN$  is to maintain the equivalence of the positive and inverse transformation and ensure that the original image can be accurately restored. For the complex spectrum:

$$F(u, v) = A(u, v) \cdot e^{j\phi(u, v)} \quad (3)$$

Where  $A(u, v)$  is the amplitude and  $\phi(u, v)$  is the phase. The amplitude spectrum and phase spectrum can be extracted from the spectrum  $F(u, v)$  by the following formula:

$$A(u, v) = |F(u, v)| = \{\text{Re}[F(u, v)]^2 + \text{Im}[F(u, v)]^2\}^{1/2} \quad (4)$$

$$\phi(u, v) = \arctan\left(\frac{\text{Im}[F(u, v)]}{\text{Re}[F(u, v)]}\right) \quad (5)$$

Where  $\text{Re}(F(u, v))$  and  $\text{Im}(F(u, v))$  are the real and imaginary parts of  $F(u, v)$ , respectively, and  $|\cdot|$  represents the module (absolute value) of the complex number. For a color image, because it contains multiple color channels (usually red, green, and blue, i.e. RGB), each channel can be regarded as an independent single-channel grayscale image. Therefore, the process of Fourier transforming a color image and calculating the amplitude spectrum and phase spectrum is actually the same operation for each color channel separately. For the spectrum  $F(u, v)$  of each channel, its amplitude spectrum  $A(u, v)$  and phase spectrum  $\phi(u, v)$  are calculated separately. Finally, the three amplitude spectra (or phase spectra) are stacked in the order of the original RGB channels to form a new three-dimensional array.

The Fourier transform serves to transition the image representation from the spatial domain to the frequency domain, where the amplitude spectrum delineates the intensity dispersion across various frequency components, thereby elucidating the textural intricacies, finer details, and other salient features. Conversely, the phase spectrum encapsulates the spatial configuration of these frequency constituents, dictating the overarching structural layout and contour of the image. Leveraging the semantic preservation attributes inherent in Fourier phase components, we endeavor to inject target image attributes while introducing perturbations to simulate real-world variability, all while accentuating the significance of phase information. This augmentation strategy aims to cater to the requirements of machine learning frameworks for enhanced data diversity and bolstered model generalization. Inspired by the Mix-up methodology [31], our approach involves incorporating a target image, exchanging the amplitude characteristics of the central region between the source and target images, and amalgamating select frequency components from the target image—such as high-frequency details and textural motifs—into the source image. This process imbues the source image with novel visual elements, achieved through linearly interpolating the amplitude spectra of two images originating from any source domain.

$$\hat{A}_{mixed}(x_{ki}) = (1 - \lambda)A(x_{ki}) + \lambda A(x_{k'i'}) \quad (6)$$

Where  $\lambda$  is a random variable uniformly sampled from the interval  $[0, 1]$ , which is the hyperparameter controlling the intensification. The mixed amplitude spectrum is then combined with the original phase spectrum to form a new Fourier representation:

$$F(\hat{x}_i^k)(u, v) = \hat{A}_{mixed}(x_i^k)(u, v) \cdot e^{-j\phi(x_i^k)(u, v)} \quad (7)$$

Subsequently, the enhanced image is derived through the inverse Fourier transform following the process of linear interpolation. This Fourier-based enhancement technique, termed amplitude



mixing (AM), has garnered recognition. Upon obtaining the mixed images, the model proceeds to generate predictions. The loss function quantifies the disparity between the model-derived predictions and the mixed labels, leveraging the enhanced images alongside their corresponding original labels for classification training. Standard cross-entropy serves as the metric for calculating the loss.

$$L_{cls}^{aug} = -y_i^k \log(\sigma(f(\hat{x}_i^k; \theta))) \quad (9)$$

In this equation,  $\sigma$  denotes the SoftMax activation function. Simultaneously, during training with the original images, the definition of classification loss parallels the aforementioned formula.

## 2.2 Normalization of uncertainty

The computation of feature statistics within Convolutional Neural Networks (CNNs) holds paramount importance, primarily aimed at quantifying the characteristics of the feature representation embedded within the network. This endeavor serves manifold purposes, including facilitating comprehension and elucidation of model behavior, aiding in network design, optimization, and diagnosis. Computational statistics, such as mean and variance, find application in regularization techniques like batch normalization, which accelerates the training process, mitigates internal covariate shifts, and potentially enhances model generalization by standardizing inputs across layers. Within each layer, Batch Normalization (BN) computes the mean and variance for the activation values ( $x_i$ ) of each neuron within a mini-batch, particularly in the context of convolutional layers, where operations are performed along the channel dimension.

These mean and variance metrics serve as characteristic statistics encapsulating domain-specific attributes present in the training data. While these features may not directly contribute to the task's objective, they play a pivotal role in distinguishing between various data sources, such as diverse photo styles, lighting conditions, or shooting environments. The mean of an image corresponds to the average of all pixel values constituting the image, reflecting its overall brightness or grayscale level. It serves as a pivotal metric for gauging the central tendency of the image's brightness distribution, thereby facilitating brightness correction and contrast adjustment during image preprocessing. On the other hand, image variance emerges as a critical indicator for assessing image contrast, texture complexity, and overall image quality, including noise levels.

For a single-channel grayscale image with dimensions  $M \times N$  and pixel values denoted by  $I(x, y)$ , the calculation of image means and variance proceeds as follows:

$$\mu = \frac{1}{MN} \sum_{x=0}^{M-1} \sum_{y=0}^{N-1} I(x, y) \quad (10)$$

$$\sigma^2 = \frac{1}{MN} \sum_{x=0}^{M-1} \sum_{y=0}^{N-1} (I(x, y) - \mu)^2 \quad (11)$$

In practical scenarios, ensuring robust model performance amidst disparities between the distributions of training and test data, commonly referred to as domain shifting, is imperative. Variations in mean and variance serve as direct indicators of the statistical disparities among features across different domains, encapsulating the essence of "statistical transfer of features." This phenomenon stands as a primary contributor to the deterioration of model performance when confronted with out-of-distribution data. To address this challenge, modeling the uncertainty associated with mean and variance becomes essential in capturing potential domain shifts. Specifically, it is postulated that characteristic statistics conform to a multivariate Gaussian distribution, accounting for inherent uncertainty. This uncertainty is characterized by the assignment of a probability distribution to the mean and variance, rather than relying on singular deterministic values. The practical realization of this concept hinges upon utilizing variance to gauge the extent of dispersion of values within the feature space relative to the mean. Estimating the uncertainty of characteristic statistics entails calculating the variance of the mean and standard deviation:

$$\Sigma_{\mu}^2(x) = \frac{1}{B} \sum_{b=1}^B (\mu_{bc}(x) - E_b[\mu_{bc}(x)])^2 \quad (12)$$

$$\Sigma_{\sigma}^2(x) = \frac{1}{B} \sum_{b=1}^B (\sigma_{bc}(x) - E_b[\sigma_{bc}(x)])^2 \quad (13)$$



These variances reflect the degree to which the feature statistics vary between instances within a batch, i.e. their uncertainty.

Assuming that the characteristic statistics, factoring in inherent uncertainty, conform to a multivariate Gaussian distribution, the original statistical value serves as the "center" of this distribution. Utilizing the previously computed variance, we delineate the magnitude of variation, essentially representing the "range" of potential domain shifts. This conceptual framework enables the translation of characteristic statistic uncertainty into a probabilistic distribution form. Drawing upon this probabilistic distribution, the generation of new values for characteristic statistics, namely the mean ( $\beta(x)$ ) and standard deviation ( $\gamma(x)$ ), involves a process of random sampling:

$$\beta(x) = \mu(x) + \varepsilon_\mu \sum_\mu(x), \varepsilon_\mu \sim N(0,1) \quad (14)$$

$$\gamma(x) = \sigma(x) + \varepsilon_\sigma \sum_\sigma(x), \varepsilon_\sigma \sim N(0,1) \quad (15)$$

Here,  $\varepsilon_\mu$  and  $\varepsilon_\sigma$  represent the original mean and variance, respectively, following a standard Gaussian distribution. Through this formulation, an array of new characteristic statistics emerges, embodying diverse directions and intensity combinations. In essence, the original characteristic statistics undergo replacement by randomly sampled counterparts, accounting for the uncertainty inherent in domain transition.

### 3. Experiment

In this section, we obtained visualization images and relevant parameters through experimental analysis, which conspicuously demonstrate a substantial enhancement in the segmentation performance of the network. In order to verify the effectiveness of the proposed method in improving the generalization ability of the network, we conducted experiments on a wide range of tasks, including cross-dataset generalization, noise resistance, light adaptation, background suppression, and labeling uncertainty processing, in which the training set and the test set have different distributional shifts, such as style shifts and scene changes.

**Setting and implementation details:** Firstly, the enhanced images were obtained through linear interpolation, as illustrated in Fig. 3. Our study comprehensively assessed the efficacy of *Aug-DSU* in retinal vessel segmentation, utilizing retinal fundus images. To assess the performance of our approach, we carried out experimental evaluations on four publicly accessible datasets, namely STARE [32], HRF [33], DRIVE [34], and CHASEDB1 [35]. These datasets encompassed distinct sample sizes, specifically 20, 45, 40, and 28, respectively. Additionally, we verified the efficacy of *Aug-DSU* in segmenting OD/OC using retinal fundus images sourced from Drishti-GS [36], RIM-ONE-r3 [37], REFUGE-train, and REFUGE-val [38]. These datasets contained 101, 159, 400, and 400 samples, respectively, providing a comprehensive evaluation of our method's performance.

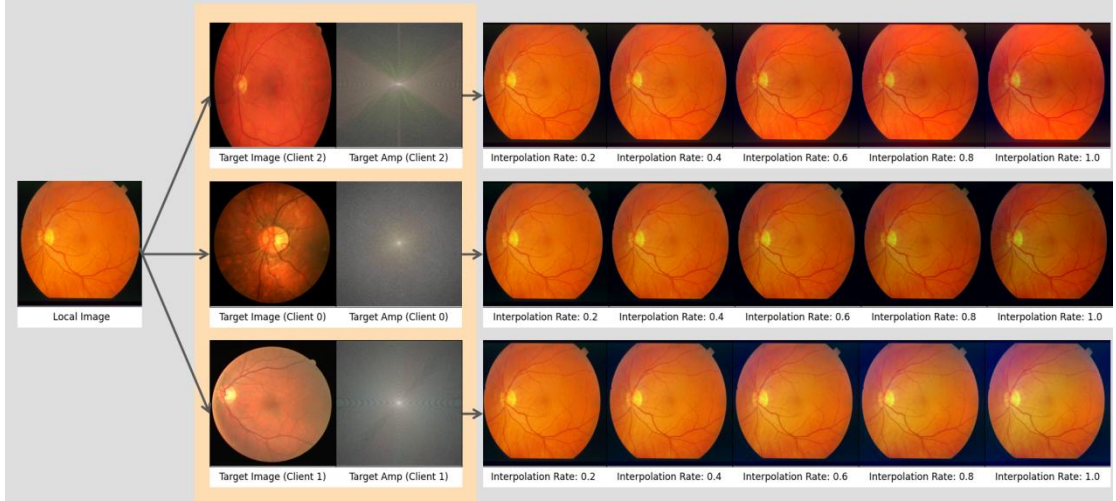


Fig. 3. Schematic diagram of amplitude mixing strategy



These datasets are widely acknowledged as standard benchmarks, offering researchers a unified platform for testing. They encompass fundus images sourced from diverse origins, age demographics (e.g., children, adults), and pathological conditions (e.g., healthy, diabetic retinopathy, etc.), facilitating an assessment of the algorithm's generalization capabilities across a spectrum of diversity and complexity. Notably, discrepancies exist within the dataset annotations provided by experts, underscoring the subjective nature and inherent uncertainty of clinical annotations. Additionally, variations in lighting conditions, contrast levels, vessel morphology (e.g., curvature, diameter fluctuations, branching structures), and background intricacies are observed across different datasets. These diversities contribute to the adaptability and robustness assessment of algorithms, thereby validating their clinical utility in real-world scenarios. For segmentation, we employed DeepLabv3+ with Mobilenet\_v2 as the backbone, configuring the batch size to 10. To train our policy controller, we utilize the proximal policy optimization (PPO) algorithm. During the training process, we apply a tanh constant of 2.5 and a SoftMax temperature of 2 to the logits of the controller, as per the recommendations [39]. Unless otherwise specified, we adopt default values of  $R = 10$ ,  $S = 5$ , and  $L = 2$ . For OD/OC segmentation tasks, we set  $E = 150$ , while for retinal vessel and lesion segmentation, due to the limited sample size, we increase  $E$  to 300. Prior to training, we preprocess the images for OD/OC segmentation by centering them on the OD and resizing them to  $256 \times 256$  pixels. Conversely, for vessel and lesion segmentation, we adjust the image dimensions to  $512 \times 512$  pixels. Additionally, to enrich our training dataset, we employ Fourier-based data augmentation techniques for all tasks. By incorporating these modifications and enhancements, we aim to improve the performance and generalization capabilities of our policy controller.

Dice similarity coefficient (DSC) and accuracy (ACC) were used as evaluation metrics for the *Aug-DSU* method for the task of vascular segmentation of fundus images on all segmentation tasks.

**Experimental results:** In our experimental setup, we adhere to established methodologies by evaluating the effectiveness of our approach through the adoption of a leave-one-out strategy. Specifically, we train models on three unknown domains and test performance on another domain. It is noteworthy that the training of the baseline model does not incorporate any specific generalization techniques; instead, it amalgamates multiple source domains in a straightforward manner. Tables 1 and 2 provide a quantitative assessment comparing the baseline model with *Aug-DSU*, with a focus on retinal vessel and OD/OC segmentation performance metrics. Notably, the proposed *Aug-DSU* method demonstrates statistically significant improvements over the baseline in terms of DSC. These findings underscore the superior generalization capability of the *Aug-DSU* approach.

Table 1. Comparison of DSC on OD/OC segmentation. Top 1 results are highlighted in bold.

Method	OD $\uparrow$				Average	OC $\uparrow$				Average
	A	B	C	D		A	B	C	D	
Baseline [40]	<b>94.96</b>	89.69	89.33	90.09	91.02	77.03	78.21	80.28	<b>84.74</b>	80.07
<i>Aug-DSU</i> (ours)	92.77	<b>90.99</b>	<b>92.78</b>	<b>92.29</b>	<b>92.21</b>	<b>86.76</b>	<b>79.41</b>	<b>86.76</b>	84.31	<b>84.31</b>

Table 2. Comparison of fundus vascular segmentation. Top 1 results are highlighted in bold.

Method	DSC $\uparrow$				Average	ACC $\uparrow$				Average
	A	B	C	D		A	B	C	D	
Baseline [40]	<b>76.32</b>	72.23	<b>76.27</b>	75.71	75.13	<b>94.10</b>	<b>94.46</b>	92.07	93.45	92.52
<i>Aug-DSU</i> (ours)	74.13	<b>75.56</b>	73.11	<b>78.37</b>	<b>75.29</b>	93.04	94.04	<b>95.17</b>	<b>97.13</b>	<b>94.84</b>

In addition to quantitative assessments, qualitative comparisons are presented in Fig. 4 to provide further insights into the performance of the proposed *Aug-DSU* approach across different domains. The visualizations clearly demonstrate the robustness of *Aug-DSU* in handling diverse domain



variations. Specifically, compared to the baseline network, which exhibits signs of overfitting during testing, the segmented images generated by *Aug-DSU* depict improved clarity and structural coherence. Conversely, the baseline network's output suffers from blurring and lacks detailed information, indicating its limited adaptability across domains. Furthermore, the incorporation of the *DSU* algorithm and Fourier data enhancement strategy enhances the visual quality of the segmented images. This is evidenced by the clearer texture and more explicit structural features, aligning more closely with the requirements for ideal image perception and interpretation.

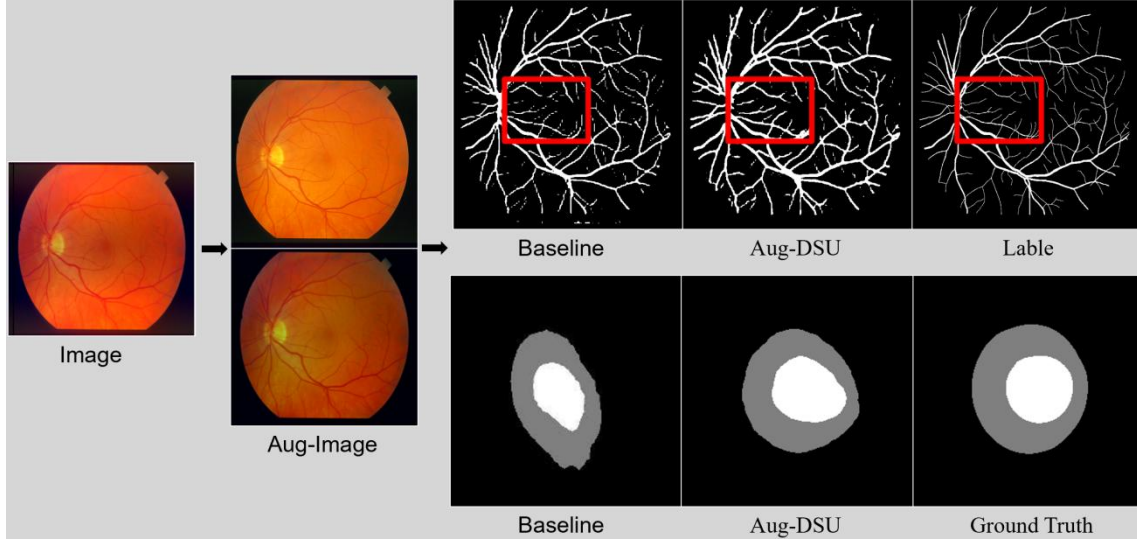


Fig. 4. Schematic diagram of qualitative comparison

#### 4. Conclusion

In this paper, we introduced a novel approach, *Aug-DSU*, for retinal fundus image processing, achieving significant advancements in image segmentation. By synthesizing feature statistics during training to model uncertainty in domain transfer, *Aug-DSU* simulates potential displacements and lays a solid foundation for image processing. Leveraging Fourier transformation, we performed linear interpolation on the amplitude spectrum, introduced noise to augment the sample set, and fed them into the network for training.

Extensive experiments on retinal vessel and OD/OC segmentation tasks demonstrate *Aug-DSU*'s effectiveness and superiority. Its ability to precisely segment retinal vessels and accurately identify OD/OC validates its practical value. Furthermore, our analysis revealed the importance of phase information in enhancing domain generalization and clarified the role of uncertainty in domain transfer.

We believe that with further research and technological advancements, *Aug-DSU* will play a pivotal role in retinal fundus image processing, enabling enhanced medical diagnosis and treatment.

#### References

- [1] Ronneberger, Olaf, Philipp Fischer and Thomas Brox. "U-Net: Convolutional Networks for Biomedical Image Segmentation." ArXiv abs/1505.04597 (2015): n. pag.
- [2] E. Shelhamer, J. Long and T. Darrell, "Fully Convolutional Networks for Semantic Segmentation," in IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 39, no. 4, pp. 640-651, 1 April 2017, doi: 10.1109/TPAMI.2016.2572683.
- [3] K. He, X. Zhang, S. Ren and J. Sun, "Deep Residual Learning for Image Recognition," 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 2016, pp. 770-778, doi: 10.1109/CVPR.2016.90.
- [4] Chen, LC., Zhu, Y., Papandreou, G., Schroff, F., Adam, H. (2018). Encoder-Decoder with Atrous Separable Convolution for Semantic Image Segmentation. In: Ferrari, V., Hebert, M., Sminchisescu, C.,



- Weiss, Y. (eds) Computer Vision – ECCV 2018. ECCV 2018. Lecture Notes in Computer Science (), vol 11211. Springer, Cham.
- [5] Zhou, P., Liu, Y., Li, G., & Guo, Y. (2013/4). A Comprehensive Survey on Transfer Learning. IEEE Transactions on Knowledge and Data Engineering (TKDE).
  - [6] Tzeng, Eric, Judy Hoffman, N. Zhang, Kate Saenko and Trevor Darrell. "Deep Domain Confusion: Maximizing for Domain Invariance." ArXiv abs/1412.3474 (2014): n. pag.
  - [7] S. Jindal and S. Singh, "Image sentiment analysis using deep convolutional neural networks with domain specific fine tuning," 2015 International Conference on Information Processing (ICIP), Pune, India, 2015, pp. 447-451, doi: 10.1109/INFOP.2015.7489424.
  - [8] Krikamol Muandet, David Balduzzi, and Bernhard Schölkopf. Domain generalization via invariant feature representation. in International Conference on Machine Learning, pages 10-18, 2013. 1, 2
  - [9] Long, Mingsheng, Yue Cao, Jianmin Wang and Michael I. Jordan. "Learning Transferable Features with Deep Adaptation Networks." ArXiv abs/1502.02791 (2015): n. pag.
  - [10] Ganin, Yaroslav and Victor S. Lempitsky. "Unsupervised Domain Adaptation by Backpropagation." International Conference on Machine Learning (2014).
  - [11] Tzeng, Eric, Judy Hoffman, N. Zhang, Kate Saenko and Trevor Darrell. "Deep Domain Confusion: Maximizing for Domain Invariance." ArXiv abs/1412.3474 (2014): n. pag.
  - [12] Ganin, Yaroslav, E. Ustinova, Hana Ajakan, Pascal Germain, H. Larochelle, François Laviolette, Mario Marchand and Victor S. Lempitsky. "Domain-Adversarial Training of Neural Networks." Journal of machine learning research (2015).
  - [13] Muandet, Krikamol, David Balduzzi and Bernhard Schölkopf. "Domain Generalization via Invariant Feature Representation." International Conference on Machine Learning (2013).
  - [14] E. Tzeng, J. Hoffman, K. Saenko and T. Darrell, "Adversarial Discriminative Domain Adaptation," 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, USA, 2017, pp. 2962-2971, doi: 10.1109/CVPR.2017.316.
  - [15] Miyato, Takeru, Shin-ichi Maeda, Masanori Koyama and Shin Ishii. "Virtual Adversarial Training: A Regularization Method for Supervised and Semi-Supervised Learning." IEEE Transactions on Pattern Analysis and Machine Intelligence 41 (2017): 1979-1993.
  - [16] Yu, Xi, Huan-Hsin Tseng, Shinjae Yoo, Haibin Ling and Yuewei Lin. "INSURE: An Information Theory Inspired Disentanglement and Purification Model for Domain Generalization." ArXiv abs/2309.04063 (2023): n. pag.
  - [17] Xu, Qinwei, Ruipeng Zhang, Ya Zhang, Yanfeng Wang and Qi Tian. "A Fourier-based Framework for Domain Generalization." 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) (2021): 14378-14387.
  - [18] E. D. Cubuk, B. Zoph, D. Mané, V. Vasudevan and Q. V. Le, "AutoAugment: Learning Augmentation Strategies From Data," 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Long Beach, CA, USA, 2019, pp. 113-123, doi: 10.1109/CVPR.2019.00020.
  - [19] Li, Xiaotong, Yongxing Dai, Yixiao Ge, Jun Liu, Ying Shan and Ling-Yu Duan. "Uncertainty Modeling for Out-of-Distribution Generalization." ArXiv abs/2202.03958 (2022): n. pag.
  - [20] Krizhevsky, Alex, Ilya Sutskever and Geoffrey E. Hinton. "ImageNet classification with deep convolutional neural networks." Communications of the ACM 60 (2012): 84 - 90.
  - [21] Gidaris, Spyros, Praveer Singh and Nikos Komodakis. "Unsupervised Representation Learning by Predicting Image Rotations." ArXiv abs/1803.07728 (2018): n. pag.
  - [22] Ulyanov, Dmitry, Andrea Vedaldi and Victor S. Lempitsky. "Instance Normalization: The Missing Ingredient for Fast Stylization." ArXiv abs/1607.08022 (2016): n. pag.
  - [23] Sergey Ioffe, Christian Szegedy Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift. Proceedings of the 32nd International Conference on Machine Learning, PMLR 37:448-456, 2015.
  - [24] S. Saeedi, M. Amini, S. Mozaffari. A novel feature extraction method using statistical moments for image classification. Journal of Visual Communication and Image Representation, 2017
  - [25] D. Torricione, J. Portilla, T. Simoncelli. Texture analysis and synthesis with grayscale random fields and steerable pyramid transforms. IEEE Transactions on Image Processing, 2008
  - [26] Blundell, Charles, et al. "Weight uncertainty in neural networks." International Conference on Machine Learning (ICML), 2015, pp. 1613-1622.
  - [27] Gal, Yarin, and Zoubin Ghahramani. "A theoretically grounded application of dropout in recurrent neural networks." Advances in Neural Information Processing Systems (NeurIPS), 2016, pp. 1019-1027



- [28] Yarin Gal, Zoubin Ghahramani. "Dropout as a Bayesian Approximation: Representing Model Uncertainty in Deep Learning", 2016 International Conference on Machine Learning (ICML), 1050-1059.
- [29] Oppenheim, A., Lim, J., Kopec, G., & Pohlig, S. C. (1979). Phase in Speech and Pictures. In Proceedings of ICASSP'79: IEEE International Conference on Acoustics, Speech, and Signal Processing, Volume 4, pp. 632-637. IEEE.
- [30] Oppenheim, A. V., & Lim, J. S. (1981). The Importance of Phase in Signals. Proceedings of the IEEE, 69(5), 529-541.
- [31] Zhang, H., Cisse, M., Dauphin, Y. N., & Lopez-Paz, D. (2018). Mixup: Beyond Empirical Risk Minimization. In Proceedings of the International Conference on Learning Representations.
- [32] D. Hoover, V. Kouznetsova, and M. Goldbaum, "Locating blood vessels in retinal images by piecewise threshold probing of a matched filter response," IEEE Trans. Med. Imag. vol. 19, no. 3, pp. 203-210, Mar. 2000.
- [33] Budai, R. Bock, A. Maier, J. Hornegger, and G. Michelson, "Robust vessel segmentation in fundus images," Int. J. Biomed. Imag. vol. 2013, Dec. 2013, Art. no. 154860.
- [34] J. Staal, M. D. Abramoff, M. Niemeijer, M. A. Viergever, and B. van Ginneken, "Ridge-based vessel segmentation in color images of the retina," IEEE Trans. Med. Imag. vol. 23, no. 4, pp. 501-509, Apr. 2004.
- [35] M. M. Fraz et al. "An ensemble classification-based approach applied to retinal blood vessel segmentation," IEEE Trans. Biomed. Eng. vol. 59, no. 9, pp. 2538-2548, Sep. 2012.
- [36] J. Sivaswamy et al. "A comprehensive retinal image dataset for the assessment of glaucoma from the optic nerve head analysis," JSM Biomed. Imag. Data Papers, vol. 2, no. 1, p. 1004, Mar. 2015.
- [37] F. Fumero, S. Alayon, J. L. Sanchez, J. Sigut, and M. Gonzalez-Hernandez, "RIM-ONE: An open retinal image database for optic nerve evaluation," in Proc. 24th Int. Symp. Comput.-Based Med. Syst. (CBMS), Jun. 2011, pp. 1-6.
- [38] J. I. Orlando et al. "REFUGE challenge: A unified framework for evaluating automated methods for glaucoma assessment from fundus photographs," Med. Image Anal. vol. 59, Jan. 2020, Art. no. 101570.
- [39] Bello, H. Pham, Q. V. Le, M. Norouzi, and S. Bengio, "Neural combinatorial optimization with reinforcement learning," 2016, arXiv:1611.09940.
- [40] Junyan Lyu, Yiqi Zhang, Yijin Huang, Li Lin, Pujin Cheng, Xiaoying Tang, "AADG: Automatic Augmentation for Domain Generalization on Retinal Image Segmentation[J]," IEEE Transactions on Medical Imaging, 2022, 41(12):3699-3711