> TPAMI-2024-09-2423 <

Task Augmentation-Based Meta-Learning Segmentation Method for Retinopathy

Jingtao Wang, Muhammad Mateen, *Member, IEEE*, Dehui Xiang, *Member, IEEE,* Weifang Zhu, Fei Shi, Jing Huang, Kai Sun, Jun Dai, Jingchen Xu, Su Zhang, and Xinjian Chen, *Senior Member, IEEE*

***Abstract*—Deep learning (DL) requires large amounts of labeled data, which is extremely time-consuming and labor-intensive to obtain for medical image segmentation tasks. Meta-learning focuses on developing learning strategies that enable quick adaptation to new tasks with limited labeled data. However, rich-class medical image segmentation datasets for constructing meta-learning multi-tasks are currently unavailable. In addition, data collected from various healthcare sites and devices may present significant distribution differences, potentially degrading model's performance. In this paper, we propose a task augmentation-based meta-learning method for retinal image segmentation (TAMS) to meet labor-intensive annotation demand. A retinal Lesion Simulation Algorithm (LSA) is proposed to automatically generate multi-class retinal disease datasets with pixel-level segmentation labels, such that meta-learning tasks can be augmented without collecting data from various sources. In addition, a novel simulation function library is designed to control generation process and ensure interpretability. Moreover, a generative simulation network (GSNet) with an improved adversarial training strategy is introduced to maintain high-quality representations of complex retinal diseases. TAMS is evaluated on three different OCT and CFP image datasets, and comprehensive experiments have demonstrated that TAMS achieves superior segmentation performance than state-of-the-art models.**

***Index Terms*—Meta-Learning, Task Augmentation, Lesion Simulation, Retinal Image Segmentation.**

I. Introduction

RETINAL diseases can damage the retina and result in severe vision loss or even blindness [1]. Recent studies have shown that deep learning (DL) models can detect and diagnose various retinal diseases by interpreting ocular data from different diagnostic modalities, including color fundus photography (CFP) and optical coherence tomography (OCT) [2]. The integration of DL algorithms into ophthalmology could play a crucial role in the screening and diagnosis of eye diseases in the future [3]. Specifically, medical-aided diagnosis segmentation algorithms based on computer vision can rapidly extract the shape, size, location, and optical density value, which provide dependable and accurate quantitative information for diagnosis and treatment [4, 5]. An explicit description of retinal imaging is given in the Appendix for Choroid Neovascularization (CNV) [6-8] in OCT and Retinitis Pigmentosa (RP) [9] in CFP.

Conventional deep learning requires a substantial amount of labeled data, especially for CFP and OCT images with abnormal retinal pathologies, where expert annotators with specialized knowledge are necessary. Particularly, acquiring pixel-wise annotated data in rare diseases, such as RP [9], is challenging due to limited sample availability. Furthermore, the data collected from different healthcare sites or image acquisition devices vary in distribution, significantly degrading the model's performance. Therefore, it is especially challenging to automatically segment retinal diseases with limited pixel-wise annotated training data.

Recently, meta-learning has garnered considerable attention to address the issues associated with limited resources, scarcity of labeled samples, and domain generalization. Meta-learning [10, 11] focuses on developing learning strategies that enable quick adaptation to new tasks based on prior experiences, similar to human intelligence. In this paper, we introduce a novel task augmentation-based meta-learning method for retinal image segmentation, as illustrated in Fig. 1. A novel retinal Lesion Simulation Algorithm (LSA) is proposed for meta-task augmentation to improve the model's segmentation performance by attaining higher accuracy, faster convergence rates, and reduced dependence on annotated samples. This unique feature is particularly effective for segmenting rare diseases, such as RP [9], where it is impractical to obtain large-scale training data with pixel-level labels.

The specific contributions are summarized as follows:

1. **Novel Task Augmentation-Based Meta-Learning Method**: This method aims to address the scarcity of retinal disease data with pixel-level segmentation labels.
2. **Lesion Simulation Algorithm (LSA)**: The core of the LSA consists of a new simulation function library for lesion feature customization and a new generative simulation network (GSNet) with an improved adversarial training strategy to optimize simulated images, capable of synthesizing multi-class lesion images with pixel-level segmentation labels in batches.

Jingtao Wang, Muhammad Mateen, Dehui Xiang, Weifang Zhu, Fei Sh, Kai Sun, Jun Dai, Jingchen Xu, Su Zhang, and Xinjian Chen are with the School of Electronic and Information Engineering, Soochow University, Suzhou, Jiangsu 215006, China (e-mail: xjchen@suda.edu.cn).

Jing Huang is with the School of Basic Medical Science, Soochow University, China.

Kai Sun is with the College of Medical Information and Artificial Intelligence, Shandong First Medical University in China.

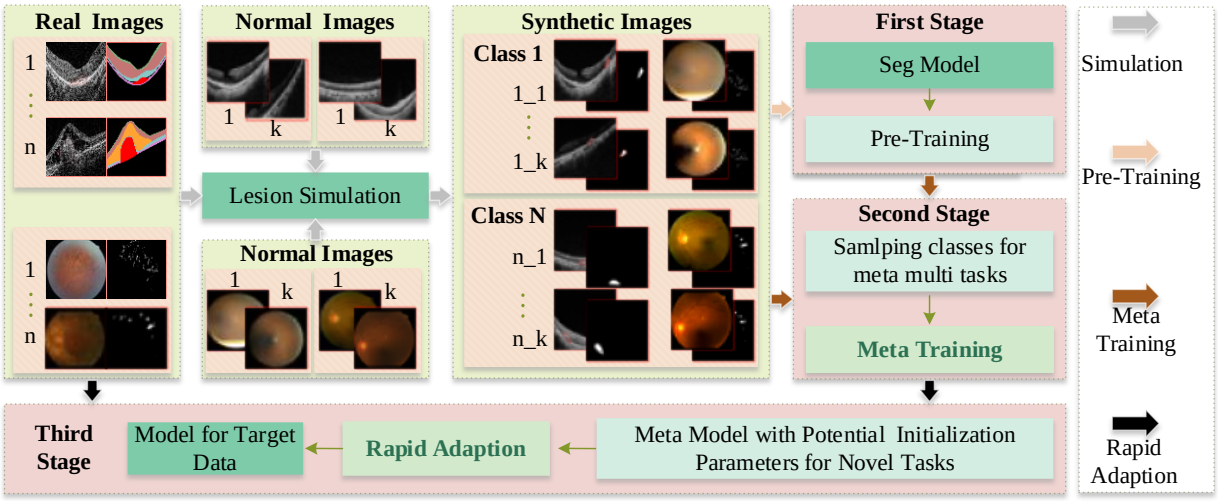


Fig. 1. Three-Stage Meta-Learning training strategy based on task augmentation. Multi-class synthetic data with segmentation labels are generated by the LSA for model pre-training and meta multi-task construction. The meta model, acquired through meta-training, subsequently exhibits rapid adaptation to novel tasks with limited real data (target data). *K* images with simulated lesions and corresponding segmentation labels can be synthesized using LSA based on a real retinal image with lesions and K normal images. *N* categories of synthetic data with simulated lesions can be generated based on *N* real images with lesions, each category containing *K* synthetic images, resulting in a total of $N \times K$ synthetic images. Details on "Meta Training" and "Rapid Adaptation" are provided in Fig. 3 in the Appendix.

3. **Three-Stage Meta-Learning Training Scheme**: This scheme improves meta-learning capabilities for tackling novel tasks, while facilitating faster and more effective adaptation to target data distributions.
4. **Robust Extensibility**: Our method improves the model's segmentation performance for CNV and RP retinal diseases while accommodating images captured by different manufacturers' devices (e.g., Heidelberg, BV-1000), and different imaging modalities (e.g., OCT and CFP). LSA can also be extended to pigment epithelial detachment (PED) and Drusen, demonstrating potential for more retinal diseases, various imaging modalities, and independent of the scanner specifications.

II. Related Work

Deep learning demands extensive labeled data. However, pixel-wise annotation is labor-intensive. Automated segmentation of retinal diseases with limited pixel-wise annotated data poses a particular challenge. To overcome the requirement of large-scale and well-annotated training data, researchers have developed advanced DL approaches, including transfer learning [12-19], data augmentation [20, 21], incremental learning [22-28], multitask learning [29-32], domain adaptation [33-36], image simulation [37-40], and meta-learning [41-48].

Numerous recent studies [14-18, 22-24, 29, 33, 34, 41-44] predominantly focus on classification with limited labeled samples, a much simpler task than segmentation. Pixel-level annotation is far more challenging than classification, as it requires creating a 2D map to separate the target object from the background. Precise segmentation annotation requires the expertise of experienced doctors and specialists, due to the smaller lesions and the blurred boundaries in retinal images. Furthermore, segmentation networks are typically more complex than classification networks. For instance, MAML [41], iMAML [43] and Reptile [42] utilize only a four-layer CNN for classification, making them more susceptible to overfitting in segmentation tasks. In contrast, our method automatically generates pixel-level segmentation labels. Moreover, within the framework of meta-learning, we utilize a deep network specifically suitable for segmentation tasks to avoid shallow network's overfitting, thereby enhancing segmentation performance for retinal diseases.

A. Transfer Learning

Several studies [14-19] employed transfer learning based on non-medical datasets like ImageNet. However, the effectiveness of transfer learning largely depends on the similarity between pre-trained and target data distributions, which is a major challenge in medical applications. Natural datasets, despite their easy accessibility, exhibit a significant domain gap with medical data. In stark contrast, our synthetic data is fully derived from the same distribution as the target real retinal data, ensuring accurate lesion simulation. This close alignment with the target data distribution makes it far more suitable for transfer learning.

B. Common Data Augmentation

Common hand-tuned data augmentation techniques, such as random image rotations, translations, flipping, cropping, noise injection, color space transformations, jitter, and nonlinear deformations, are often used to expand the training sets [49-52]. However, these methods fundamentally produce highly correlated images, which limits the ability to emulate real lesion variations. This suggests a potential constraint in accurately segmenting retinal lesions.

C. Incremental Learning

Incremental learning is commonly employed in multi-class classification [23] or segmentation tasks [24, 27]. Recently, few-shot class-incremental learning (FSCIL) has posed a

significant challenge for deep neural networks, requiring them to learn new classes from only a few labeled samples without forgetting previously learned ones [25]. Hassan et al. [23] presented an incremental cross-domain adaptation framework, enabling a classification model to progressively learn to identify abnormal retinal pathologies in OCT and CFP via few-shot training. Vu et al. [28] introduced a data-adaptive loss function for incremental learning in image segmentation tasks with limited labeled samples. Incremental learning relies on acquiring multi-class data with a consistent distribution to accumulate prior knowledge, which is then utilized for learning new categories with limited labeled samples. However, procuring multi-class data with consistent distribution poses significant challenges.

D. Multitask Learning

Chen et al. [32] enhanced segmentation performance in target tasks by utilizing multiple auxiliary tasks. Similarly, Wang et al. [29] developed a multi-task framework with various subnetworks to diagnose retinal diseases. Another multi-task learning network could simultaneously segment the left ventricle and track its motion across multiple time frames [31]. Zhao et al. [30] constructed diverse tasks, such as classification and registration, to accumulate prior knowledge that could then be applied to target segmentation tasks, thereby reducing the reliance on labeled data. However, a prerequisite for any multi-task learning method is the collection of a comprehensive multi-task dataset.

E. Domain Adaption

Domain adaptation methods mitigate domain shift between source and target domains [53], allowing prior knowledge from the source to compensate for limited annotated samples in the target domain. Ju et al. [34] employed adversarial domain adaptation to diagnose retinal diseases by managing cross-domain shifts between normal and ultra-wide-field fundus photography. Chen et al. [35] proposed RDR-Net, an unsupervised domain adaptation network for optic disc and cup segmentation, showing strong generalization capabilities. Additionally, He et al. [36] leveraged the retinal layer structure in a multi-task framework to facilitate domain alignment and enhance cross-domain OCT fluid segmentation. Although these domain adaptation methods improved generalization to target domain datasets, they typically require source data sharing similar pathologies with the target domain, even if derived from different devices or centers. In contrast, our method does not require strict source domain data collection. By strategically employing limited source data, we are able to significantly enhance segmentation performance on target domain data.

F. Generative Model

GAN [54], cGAN [55], CycleGAN [56], and diffusion models [57], along with their various variations, are commonly used to mitigate data scarcity. However, these methods generate retinal lesion images from noise or other image styles, making the simultaneous generation of pixel-level labels challenging. Additionally, it is challenging to train generative models to synthesize retinal images with small and diverse lesions, which hinders model convergence and degrades image quality.

In this paper, we first generate retinal lesion images using the proposed Lesion Simulation Function Library, then optimize them with our generative model, GSNet. Unlike the opaque nature of deep learning models, our simulation functions are transparent and interpretable to enable the automatic generation of pixel-level labels. Additionally, we utilize an improved adversarial training strategy to train a self-developed generative model dedicated solely to image refinement. This approach not only enhances the quality of the synthetic images but also avoids the convergence and quality challenges associated with training generative models directly to produce lesion images.

G. Meta-learning

In view of the great potential of meta-learning to address issues such as limited resources, scarce labeled data, and domain generalization, we are motivated to apply meta-learning algorithms to retinal disease segmentation. Specifically, we make efforts on meta-task augmentation based on gradient-based meta-learning algorithms [41]. To the best of our knowledge, no previous study has explored this integrated approach.

IV. METHODS

In this study, we introduce a task augmentation-based meta-learning approach to address the issues of retinal disease image scarcity and annotation difficulties. This method facilitates meta-task augmentation by generating high-quality meta-tasks, thereby enhancing performance in retinal disease segmentation. Algorithm 1 provides a detailed description of our TAMS.

A. Meta-Learning Algorithms

Meta-learning focuses on developing learning strategies that enable quick adaptation to new tasks based on prior experiences, similar to human intelligence [13, 58]. Typically, it involves a meta-learner that extracts common meta-knowledge from a known distribution of tasks to solve novel tasks [59, 60]. Recent studies have shown its effectiveness in few-shot learning, where the learner must adapt to new tasks with limited training samples [10, 11, 59, 61-65]. Algorithm 1 in the Appendix provides a detailed description of meta-learning.

Meta-learning consists of two phases: meta-train and meta-test. In the meta-training phase, a task τ is sampled from the task distribution $p(\tau)$. This task, referred to as an episode, includes a training split $\tau^{(tr)}$ (support set) to optimize the base-learner, and a test split $\tau^{(te)}$ (query set) to optimize the meta-learner. The goal during meta-training is to learn common prior knowledge from multiple episodes τ. Meta-training involves two stages of optimization within each episode:

1. **Base-Learning:** Given an episode $\tau \in p(\tau)$, the base-learner θ is learned from episode's support set $\tau^{(tr)}$ and its corresponding loss $L_\tau\left(\theta, \tau^{(tr)}\right)$. After optimizing this loss, the base-learner's parameters are updated to ϕ_τ.
2. **Meta-Learning:** The test loss $L_\tau\left(\phi_\tau, \tau^{(te)}\right)$ is used to optimize the parameters of the meta-learner. After meta-training on all episodes, the meta-learner is optimized by the test losses $\left\{L_\tau\left(\phi_\tau, \tau^{(te)}\right)\right\}_{\tau \in p(\tau)}$, the meta-learner's parameters are updated to θ'.

Algorithm 1: Task Augmentation-Based Meta-Learning

Input: Real disease images and labels: $R(1,2,\dots,n)$ $L(1,2,\dots,n)$, Normal images $Nor(1,2,\dots)$

Output: Potential initial parameters for novel tasks: θ', Optimal parameters for novel tasks: θ''

Generate multi-class synthetic data with pixel-level segmentation labels:
$Syn \leftarrow LSA(R, L, Nor)$
Obtain pre-trained parameters based on Syn:
$\theta \leftarrow DeepLearning(Syn)$
Build meta-training tasks $p(\tau)$ based on Syn
Build target task T with limited R
for iteration = 1,2… **do**
  Sample batch of tasks $\tau \sim p(\tau)$
  # All episodes
  for each τ **do**
    # one episode of meta training
    if *alg* == *MAML* **then**
      Compute ϕ_τ with respect to Support Set
      $\phi_\tau = \theta - \alpha \nabla_\theta L_\tau^{train}(\theta)$
    else if *alg* == *iMAML* **then**
      Compute ϕ_τ with respect to Support Set
      $\phi_\tau = \underset{\phi}{\operatorname{argmin}}(L_\tau^{train}(\theta) + \frac{\lambda}{2} \| \phi - \theta \|^2)$
    end if
    Evaluate $\nabla_\theta L(\theta)$ with respect to Query Set:
    $\nabla_\theta L(\theta) = \sum_{p(\tau)} \nabla_\theta L_\tau^{test}(\phi_\tau)$
  end for
  $\theta' \leftarrow \theta - \beta \nabla_\theta L(\theta)$
  # Evaluate θ' on target segmentation task
  for iteration = 1,2… **do**
    Sample batch of real samples $(r, l) \sim T$
    for each (r, l) **do**
      Evaluate BCE and Dice Loss
    end for
    Update $\theta'' \leftarrow \theta'$
  end for
end for
return θ', θ''

Note: α and β denote the learning rate, λ indicates the regularization term

The meta-test phase aims to evaluate the performance of the trained meta-learner for rapid adaptation to unseen tasks. Given an unseen task τ_{unseen}, the meta-learner θ' guides the base-learner $\theta_{\tau_{unseen}}$ to adapt to τ_{unseen}, for example, through initialization [10]. The performance on the test split $\tau_{unseen}^{(te)}$ is used to evaluate the meta-learning approach. If there are multiple unseen tasks $\{\tau_{unseen}\}$, the average performance on $\left\{\tau_{unseen}^{(te)}\right\}$ is used as the final evaluation metric.

In this study, we construct test tasks (unseen target tasks) using real retinal data, referred to as $\{\tau_{unseen}\}$. The meta-learner, trained on $p(\tau)$ from multi-class synthetic data, guides the base-learner, a lesion segmentation model, to quickly adapt to these unseen tasks. In this paper, two typical meta-learning algorithms, MAML [41] and iMAML [43, 47], are integrated into our TAMS method to improve the retinal disease segmentation model, named mTAMS and iTAMS, respectively. The performance of the Reptile[42] algorithm in our TAMS method is not satisfactory.

B. Lesion Simulation Algorithm

Currently, no multi-class medical dataset is available for constructing high-quality meta-tasks. To address these issues, we propose LSA to generate multi-class data with pixel-level segmentation labels (Algorithm 2 in the Appendix).

LSA's architecture, depicted in Fig. 2, comprises two main components: lesion feature generation and lesion forgery. The first module, named "F" in Fig. 2 (A), extracts lesion features from real retinal disease images. These features are then fed into the "Aug" module for augmentation in batches. The Lesion Forgery Module, shown in Fig. 2 (B), utilizes a trained U-Net to extract OCT image masks and CFP optic disc and boundary masks from normal retinal images. These masks, along with generated features and real retinal images, are then processed by the lesion simulation functions to produce high-resolution synthetic images (CFP size: 565×584, OCT size: 1024×512). Finally, the synthetic images are further refined using the trained GSNet. Due to hardware resource constraints, the final size of the images generated by GSNet is limited to 256×256.

B.1 Lesion Feature Generation

First, LSA extracts lesion features from real retinal disease images. These features are then passed to the feature augmentation module (referred to as "Aug" Fig. 2 (A)), which applies affine and elastic transformations to augment the lesion features in batches. By extracting and augmenting real lesion features to simulate new ones, this process ensures both authenticity and interpretability, while also increases the diversity of the simulated lesions.

Fig. 6 in the Appendix outlines the lesion feature generation process:

1) Feature Extraction:
- Contour: Lesion boundary.
- Size: Pixel count within the lesion contour.
- Position: Coordinates of the lesion's top-left and bottom-right corners.

2) Feature Augmentation: Affine and elastic transformations are applied to augment lesion features.

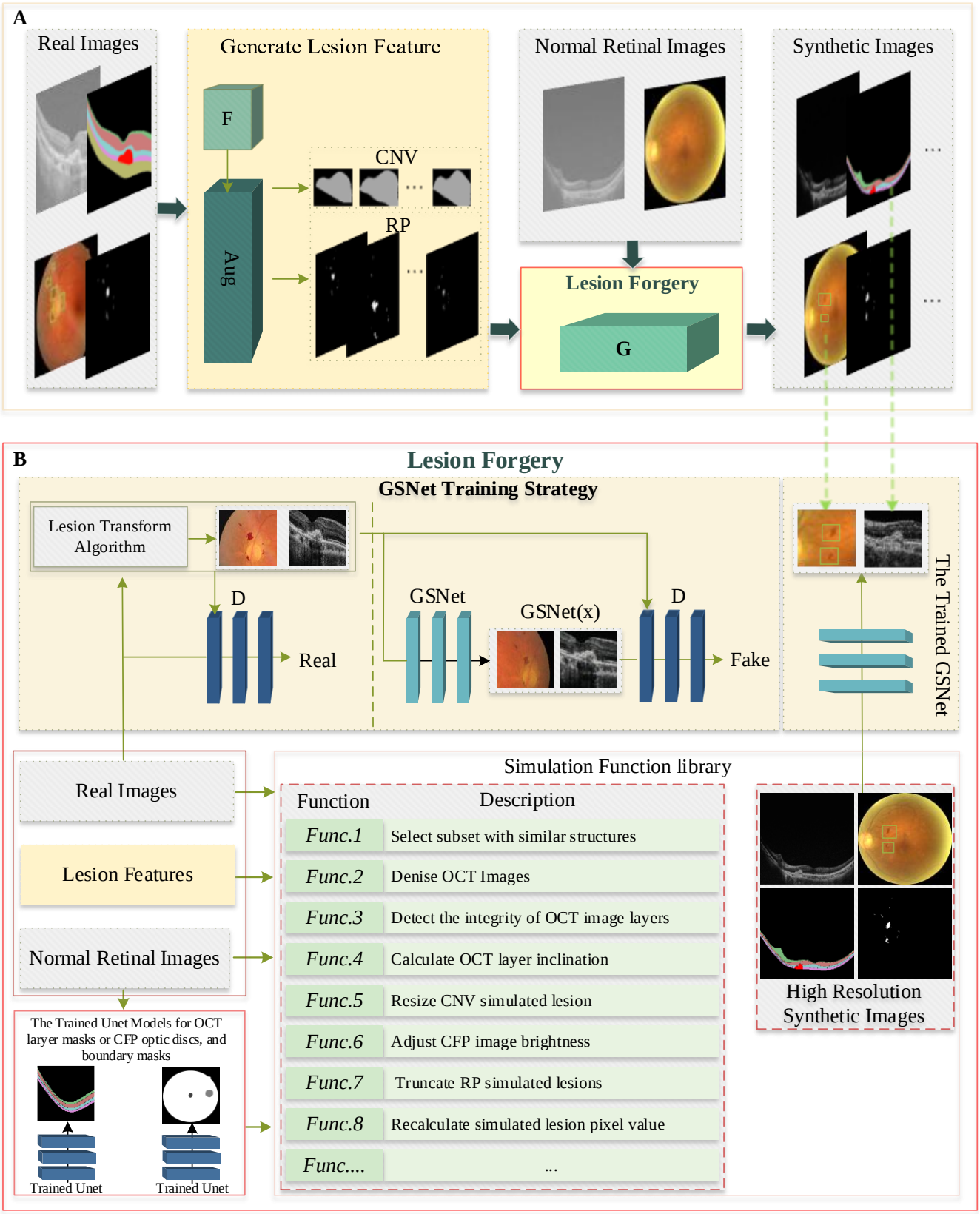


Fig. 2 Architecture of LSA. (A): "Aug" is the feature augmentation module, "F" is lesion features extraction module;" G" is the lesion forgery module. (B): Details of lesion forgery.

B.2 Lesion Forgery

The Lesion Forgery Module simulates lesions on normal retinal images and consists of two components: the Simulation Function Library and the generative simulation network (GSNet), as illustrated in Fig. 2 (B). The Simulation Function Library customizes lesions on normal retinal images (referred to as background images) based on specific disease characteristics, and generates high-resolution synthetic images. For OCT CNV, the simulation sequence is: Func.2 -> Func.3 -> Func.4 -> Func.5 -> Func.8. For CFP RP, the sequence is: Func.1 -> Func.6 -> Func.7 -> Func.8. For a comprehensive description of the simulation functions, please refer to the Appendix.

B.3 The Trained UNet Models

To simulate specific retinal diseases, images without such lesions are required. These easily accessible images are referred to as normal or background images in this paper. Trained UNet [66] models are used to segment OCT layers, as well as CFP optic discs and boundary masks from these normal images, thereby providing more feature information for lesion forgery. Moreover, training such UNet models is a relatively straightforward task.

B.4 Simulation Function Library

We develop a Simulation Function Library to customize specific lesions on normal retinal images. Unlike the opaque learning and inference processes of DL models, every operation in our simulation functions is transparent and interpretable. Since every step of the simulation process is known, we can generate corresponding segmentation labels by applying the same operations. Our simulation functions are primarily from the "cv2" package, which requires minimal hardware resources and enables the generation of high-resolution images. For detailed algorithms please refer to the Appendix.

B.5 Generative Simulation Network (GSNet)

In the process of customizing synthetic images using simulation functions, subtle details between lesion features and the background, which are difficult to detect with the naked eye, are often overlooked. To address this issue, we design a generative simulation network (GSNet) with an improved confrontation training strategy, as shown in Fig.7 in the Appendix, to further improve the high-resolution synthetic images.

GSNet adopts the UNet [66] encoder-decoder architecture, where the encoder extracts image features through downsampling, and the decoder restores these features into images via upsampling. Additionally, MAE [67] can reconstruct missing pixels, and Zhou et al. [68] has released a feature extractor trained on large-scale OCT and CFP datasets. To fully utilize MAE's powerful pixel reconstruction capability, we design a multi-scale feature fusion module. After reshaping and pooling the features extracted by the MAE, the resulting multi-scale features are fused with the features from the CNN. Finally, these fused features are skip-connected to the decoder for image restoration.

Algorithm 2: GSNet Training Strategy

Input: D: Discriminator,
G: GSNet,
Real disease images and labels: $R(1,2,\dots,n)$, $L(1,2,\dots,n)$

Output: The trained GSNet

Load pre-trained MAE feature extractor parameters for GSNet
While not done **do**
  Sample batch of data $(r,l)\sim(R,L)$
  for each (r,l) **do**
    Randomly alter the pixel values in the lesion
    The transformed whole image is given by $r_t = HueSaturationValue\big(r(l \neq 0)\big) \oplus r$, where $r(l \neq 0)$ denotes the lesion area
    # Optimize discriminator
    Evaluate $Loss(D) = [MSELoss(D(G(r_t), r_t), zeros) + MSELoss(D(r, r_t), ones)] \times 0.5$
    # Optimize GSNet
    Evaluate $Loss(G) = [MSELoss(D(G(r_t), r_t), ones) + L1Loss(G(r_t), r)] \times \lambda$
  end for
  Updating D and G: $optim(D, Loss(D), lr), optim(G, Loss(G), lr)$
End while
return the trained GSNet

Note: The Adam optimizer is used with a learning rate of 0.001 for CFP images and 0.0002 for OCT images. $\lambda = 100$

We simulate the lesion pixel calculation process described in Func.8 using the "HueSaturationValue" function from the "Albumentations" package. This function randomly modifies the pixel values of the lesion area in real images. GSNet is then used to restore these altered images to their original pixel distribution. GSNet is trained using an enhanced Pix2Pix adversarial network strategy, incorporating a Markovian (PatchGAN) discriminator [69, 70]. The trained GSNet can automatically adjust the pixel values of the simulated lesions in synthetic images, compensating for details overlooked in the customization process. For details on training GSNet, refer to Algorithm 2. The loss function incorporates both MSE loss and L1 loss:

$$MSELoss = \frac{1}{n}\sum_{i=1}^{n}(\hat{y}_i - y_i)^2 \quad (1)$$

$$L1Loss = \frac{1}{n}\sum_{i=1}^{n}\left|\hat{y}_i - y_i\right| \quad (2)$$

TABLE I
DATASET INFORMATION

Index	Size	Imaging Type	Number
D1	1024 × 512	OCT	94 eyes, 176 images with CNV; 6061 images without CNV
D2	3100 × 2848 × 3	CFP	1 image ×219 patients with RP; 800 normal images
D3	1440 × 2160 × 3	CFP	5 images × 2 eyes × 4 patients × 3 sessions
D4	Variable	CFP	5000 patients, 2817 normal images.
D5	Variable	CFP	40
D6	384 × 288 × 3	Colonoscopy	612
D7	384 × 512 × 3	Dermoscopy	2596
D8	768 × 560 × 3	Dermoscopy	200
D9	Variable	Colonoscopy	1000
D10	Variable	Capsule Endoscopy	55
D11	Variable	/	1000 classes

Datasets: D1 (BV1000), D2(RPHS) D3(RIPS [71]), D4(ODIR-5K [72]), D5(Drive [73]), D6(CVC-612 [74]), D7(ISIC-2018 [8]), D8(PH2 [75]), D9(Kvasir-SEG [76]), D10(KvasirCapsuleSEG [77]), D11(FSS-1000 [78, 79]). D1 and D2 are private, the others are public.

C. Segmentation Networks for Retinal Lesions

We employ UNet in our TAMS framework for retinal lesion segmentation, which is specifically designed for medical images. Our study also compares it with UNet++[80], TransUNet[81], and MANet [82], demonstrating its superior performance.

Retinal image lesion areas are typically small and irregular (e.g., CNV, RP), requiring deep neural networks (DNNs) for accurate segmentation (e.g., UNet [66], nnUNet [83], and ResUNet). However, DNNs are prone to overfitting with a few samples, whereas meta-learning often utilizes shallow neural networks (SNNs), which limits its effectiveness [44]. To address this issue, we first pre-train UNet using our synthetic large-scale data. Subsequently, we employ synthetic multi-class data to construct meta-tasks for meta-training UNet.

D. Meta-task Generation Scheme

Fig. 8 in the Appendix illustrates the process of generating meta-tasks. Each real sample is used to produce K synthetic samples as one class, resulting in N-classes of data. These N classes synthetic data can then be used to construct meta-tasks. For instance, if we aim to train a base model on the 2ways-5shot-5query set using the meta-learning algorithm, we can sample two classes from N classes to form each task, with each class containing five synthetic images for the support set and another five for the query set. This approach allows us to generate a sufficient number of similar meta-tasks for meta-training, based on the multi-class synthetic data.

IV. EXPERIMENT SETUP

A. Datasets

TABLE II
TARGET DATASET SPLIT FOR TRAINING AND TESTING

	Training Data (eyes/images)	Test Data (eyes/images)
BV1000 (D1)	75/141	19/35
RPHS (D2)	176/176	43/43
RIPS (D3)	8/120	-

Table I shows the datasets used in this paper. D1 (BV1000) is an OCT dataset, while D2 (RPHS) and D3 (RIPS) [71] are CFP datasets, which are the target datasets for segmentation. D1, D2, and D3 cover two types of retinal images and two types of retinal diseases, demonstrating the universal and high clinical applicability of our approach. Dataset D4 is used for D3 lesion simulation because D3 does not contain normal CFP images. Dataset D5 [73] is used for training a UNet model for optic disc and boundary segmentation. Datasets D6 [74], D7 [8], D8 [75], D9 [76], and D10 [77] are utilized for constructing multi-tasks, and comparison experiments were conducted using the multi-class synthetic data generated by our LSA. The D6 to D10 datasets have a significant domain gap and only a few classes (five classes), which adversely affects the meta-learning algorithm. Although FSS-1000 has a rich variety of categories, it also negatively impacts the meta-learning algorithm due to the large domain gap between natural datasets and target segmentation retinal data. Detailed information on D1-D10 can be found in the Appendix, and the target segmentation dataset division can be found in Table II.

B. Implementation Details

The implementation of our proposed method is based on the public platform PyTorch 2.3.1 with CUDA 12.0 parallel computing library and GeForce RTX A6000 GPU with 48-GB memory. Details of the experimental settings can be found in Table I in the Appendix. Evaluation metrics and the loss function can also be found in the Appendix.

V. RESULTS

A. Quantitative Comparison Results

We compared our method, iTAMS, and mTAMS, with several state-of-the-art image segmentation models, as listed in Table III, including DL-trained UNet [66], UNet++ [80], ResUNet, TransUNet [81], MANet [82], nnUNet [83], SwinUNetr [84], and LiteMedSAM [85]. Our method achieved the highest scores across all four metrics. Details of the qualitative analysis are provided in the Appendix.

TABLE III
QUANTITATIVE RESULTS ON BV1000, RPHS, AND RIPS

Target Datesets	Method	Recall (%)	Precision (%)	IoU (%)	DSC (%)
BV1000	D-ResUNet	50.89	29.96	25.76	36.36
	D-SwinUNetr	55.59	54.46	35.10	49.40
	D- MANet	59.86	56.52	38.30	55.14
	D- UNet++	61.59	56.63	38.50	55.34
	D- TransUNet	61.75	60.64	41.30	58.23
	D-LiteMedSAM	61.94	62.49	45.69	60.83
	D-nnUNet	61.99	69.95	51.11	62.61
	D-UNet	63.54	67.50	46.55	63.14
	iTAMS(Ours)	72.80	**78.25**	59.59	74.27
	mTAMS(Ours)	**76.55**	74.14	**60.01**	**74.73**
RPHS	D-ResUNet	41.55	57.43	26.1	36.0
	D-SwinUNetr	46.48	50.23	29.40	43.53
	D-LiteMedSAM	50.17	56.41	36.44	51.73
	D-MANet	40.52	56.09	35.62	52.30
	D-UNet++	39.89	58.87	37.03	53.63
	D-nnUNet	52.86	63.76	40.61	55.09
	D-TransUNet	43.57	57.68	38.48	55.38
	D-UNet	48.06	63.53	43.47	60.17
	mTAMS(**Ours**)	52.85	**64.30**	47.53	64.06
	iTAMS(**Ours**)	**54.08**	63.88	**47.93**	**64.41**
RIPS	D- MANet	45.59	42.40	27.26	42.73
	D-LiteMedSAM	54.96	44.32	30.20	44.47
	D-UNet++	47.14	46.81	29.50	45.34
	D-SwinUNetr	58.51	44.50	30.32	45.51
	D-TransUNet	44.00	52.23	30.20	46.27
	D-ResUNet	48.26	51.47	37.75	48.34
	D-nnUNet	41.35	54.50	38.01	50.89
	D-UNet	53.37	52.83	36.16	53.00
	mTAMS(Ours)	56.56	55.90	39.25	56.21
	iTAMS(Ours)	**58.76**	**56.03**	**39.92**	**56.85**

Four-fold cross-validation was conducted on the training dataset. Our method achieved the highest scores across all four metrics respectively.

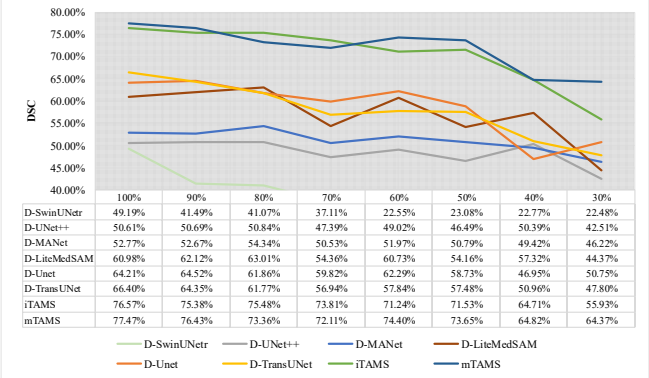
For the CNV segmentation task on BV1000 OCT dataset, our iTAMS improved the four metrics (DSC, IoU, Recall, and Precision) by 11.13%, 13.04%, 9.26%, and 10.74%, respectively, compared to the second-best method, the D-UNet. In the rare retinal disease RP segmentation task on the RPHS color fundus photography dataset, the improvements were 4.24%, 4.46%, 6.02%, and 0.35%, respectively. On the publicly available RIPS color fundus photography dataset for RP segmentation, the improvements were 3.85%, 3.77%, 5.38%, and 3.21%, respectively. The improvements of our mTAMS, on BV1000, are 11.59%, 13.46%, 13.01%, 6.64%; on RPHS are 3.89%, 4.06%, 4.79%, 0.77%; on RIPS is 3.21%, 3.09%, 3.19%, 3.07%. Our mTAMS performed better than iTAMS on the BV1000 dataset, showing greater improvements in DSC, IoU, and Recall. However, on the CFP dataset (RPHS and RIPS), iTAMS showed better performance. Additionally, both nnUNet and UNet exhibited competitive performance across all three datasets, outperforming the other evaluated advanced segmentation networks.

Notably, our method demonstrated remarkable improvements on BV1000 and RPHS datasets (e.g., DSC improvements were 11.14% and 4.24%, respectively, compared to the second-best method), whereas the improvement on the RIPS dataset was smaller (3.85%). This is likely due to the limited size of the RIPS dataset, which contains data from only eight eyes, while the BV1000 and RPHS datasets contain data from 94 and 219 eyes,

B. Results with Limited Annotated Training Data

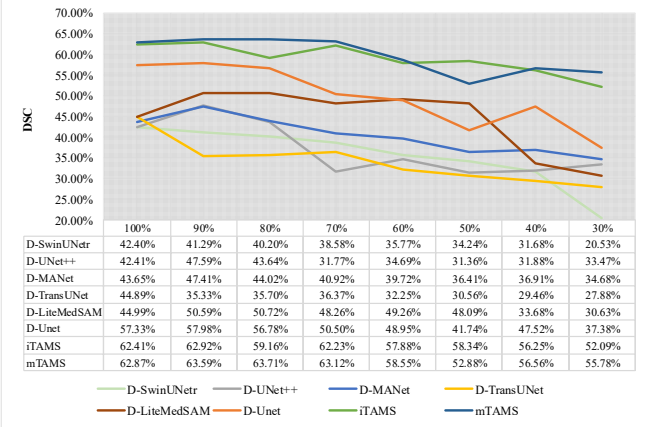
The model was trained with a limited amount of annotated training data and evaluated on the same test data. The experimental results, presented in Fig. 3, demonstrate that our method achieved the highest score despite the limited annotated data, indicating a reduced dependence on annotated data. On the BV1000 dataset, our mTAM method achieved a DSC score of 64.37% with 30% training data The annotated training data, whereas the second-best method, DL-based UNet, required 100% of the data to achieve a comparable score of 64.21%. A similar trend was observed on the RPHS dataset, where our iTAMS method achieved a DSC score of 56.25% using 40% of the annotated training data, just 1.07% lower than the DL-based UNet's score of 57.32% with 100% of the data. Furthermore, our iTAMS method achieved a DSC score of 58.34% using 50% of the labeled training data, which is 1.02% higher than the DL-based UNet's score with full data utilization.

In summary, as the amount of annotated training data decreases, the segmentation accuracy of all methods shows a downward trend. However, our method not only consistently maintains high segmentation accuracy but also experiences a slower decline, a performance unmatched by other methods.



	100%	90%	80%	70%	60%	50%	40%	30%
D-SwinUNetr	49.19%	41.49%	41.07%	37.11%	22.55%	23.08%	22.77%	22.48%
D-UNet++	50.61%	50.69%	50.84%	47.39%	49.02%	46.49%	50.39%	42.51%
D-MANet	52.77%	52.67%	54.34%	50.53%	51.97%	50.79%	49.42%	46.22%
D-LiteMedSAM	60.98%	62.12%	63.01%	54.36%	60.73%	54.16%	57.32%	44.37%
D-Unet	64.21%	64.52%	61.86%	59.82%	62.29%	58.73%	46.95%	50.75%
D-TransUNet	66.40%	64.35%	61.77%	56.94%	57.84%	57.48%	50.96%	47.80%
iTAMS	76.57%	75.38%	75.48%	73.81%	71.24%	71.53%	64.71%	55.93%
mTAMS	77.47%	76.43%	73.36%	72.11%	74.40%	73.65%	64.82%	64.37%

(A) BV1000 OCT images CNV segmentation



	100%	90%	80%	70%	60%	50%	40%	30%
D-SwinUNetr	42.40%	41.29%	40.20%	38.58%	35.77%	34.24%	31.68%	20.53%
D-UNet++	42.41%	47.59%	43.64%	31.77%	34.69%	31.36%	31.88%	33.47%
D-MANet	43.65%	47.41%	44.02%	40.92%	39.72%	36.41%	36.91%	34.68%
D-TransUNet	44.89%	35.33%	35.70%	36.37%	32.25%	30.56%	29.46%	27.88%
D-LiteMedSAM	44.99%	50.59%	50.72%	48.26%	49.26%	48.09%	33.68%	30.63%
D-Unet	57.33%	57.98%	56.78%	50.50%	48.95%	41.74%	47.52%	37.38%
iTAMS	62.41%	62.92%	59.16%	62.23%	57.88%	58.34%	56.25%	52.09%
mTAMS	62.87%	63.59%	63.71%	63.12%	58.55%	52.88%	56.56%	55.78%

(B) RPHS CFP images RP segmentation

Fig. 3 Quantitative results on the testing dataset with limited training data. Despite using limited annotated training data, our method achieved superior segmentation performance, demonstrating a reduced dependence on annotated data. In the BV1000 dataset, our method achieved the same performance as the second-best approach while requiring only 30% of the training data. This proportion was approximately 40% in the RPHS dataset.

C. Results by Different Meta Training Tasks

Several public datasets are utilized as meta datasets to construct multiple tasks for meta-training, and compared with multi-class data generated by our LSA. Crucially, regardless of whether the medical datasets D6-D10 or the category-rich natural dataset FS-1000, the performance remains inferior to our synthetic data, as shown in Fig. 4. Specifically, our best-performing method (TAMS) on BV1000 improved the DSC metric by 14.38%, compared to the highest score (63.09%) of "D6-D10", and by 11.53% compared to "FS-1000" with the highest score 65.94%. Furthermore, on the RPHS dataset, this improvement is 5.5% and 4.04%, respectively.

It is also worth noting that using D6-D10 and FS-1000 for meta-training yields very limited performance improvements compared to deep learning methods. Specifically, on the BV1000 dataset, the performance of D6-D10 (iMAML) on metric decreased by 0.52% compared to D-UNet, while FSS-1000 (MAML) showed a marginal increase of 1.73%. Similarly, on the RPHS dataset, D6-D10 (iMAML) improved by only 0.04% compared to D-UNet, and FSS-1000 (MAML) improved by just 1.5%.

In addition, the performance of the iMAML meta-learning algorithm on the meta dataset D6-D10 is better than that of MAML, whereas the performance on the meta dataset FS-1000 is reversed. The D6~D10 dataset consists of medical images, while FS-1000 comprises natural images. D6~D10 is more similar to our target segmentation datasets, OCT and CFP, but FS-1000 has a greater number of classes (D6-D10 has only 5 classes, while FS-1000 has 1000 classes). This suggests that the MAML algorithm is more suitable for data with diverse categories, while the iMAML algorithm is more effective for data where the training and target categories are more similar in domain.

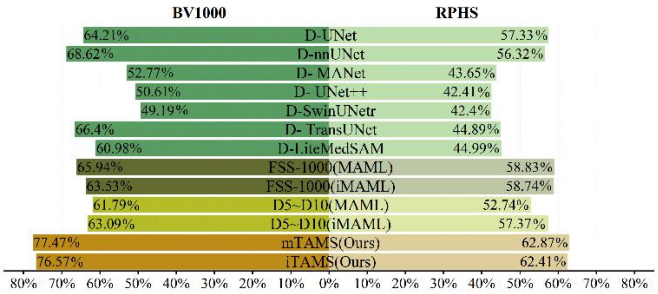


Fig. 4 "D6-D10(iMAML)" and "D6-D10(MAML)" refer to constructed multi-tasks utilizing datasets D6, D7, D8, D9, and D10 for meta-training, with each dataset representing a distinct class, totaling five classes. "FS-1000(iMAML) and FS-1000(MAML)" denote the construction of meta multi-tasks utilizing FS-1000, which consists of 1000 classes.

In summary, the synthetic data automatically generated by our LSA not only possesses rich categories but also ensures appropriate inter-domain differences between the multi-class data used for meta-training and the target data to be segmented. TAMS facilitates the application of meta-learning algorithms to lesion segmentation tasks, significantly enhancing generalization capabilities for novel tasks, thereby substantially improving lesion segmentation performance.

D. Results by Different Data Augmentation Methods

We used common data augmentation methods during the DL training phase, including horizontal flip, vertical flip, random rotation, and random adjustments to brightness and contrast. It is evident that these common data augmentation methods do not generally improve the model's performance, as shown in Fig. 5 (A) and Fig. 5 (B). On the BV1000 dataset, TransUNet, SwinUNetr, and LiteMedSAM showed improvement, while UNet, UNet++, and MANet did not. Similarly, on the RPHS dataset, MANet and SwinUNetr showed gains, with no improvement for UNet, UNet++, LiteMedSAM, or TransUNet.

Furthermore, we applied common data augmentation methods during meta-training. However, as shown in Fig. 5 (C), they had no significant impact on the meta-learning algorithm. In summary, these methods increase image quantity, but fail to enhance lesion diversity, resulting in limited model improvements. Meta-learning, in particular, requires an appropriate domain gap between different classes of data, which common data augmentation methods cannot provide. In contrast, our method not only automatically generates multiple categories of data but also enhances lesion diversity, achieving task augmentation for meta-learning.

TABLE IV
ABLATION EXPERIMENTS ON GSNET

Target Datasets	Method	GSNet	Recall (%)	Precision (%)	IoU (%)	DSC (%)
BV1000	iTAMS		72.14	**80.11**	60.15	74.95
	iTAMS	√	76.38	76.59	62.27	76.57
	mTAMS		72.12	79.30	59.30	74.24
	mTAMS	√	**76.45**	76.94	**63.70**	**77.47**
RPHS	iTAMS		52.82	48.44	41.47	58.60
	iTAMS	√	**52.36**	**53.71**	45.36	62.41
	mTAMS		51.88	47.93	40.39	57.54
	mTAMS	√	50.44	55.84	**45.86**	**62.87**

TABLE V
RESULTS BY DIFFERENT BACKBONES AS SEGMENTATION NETWORKS

Target Datasets	Meta-Algorithm	Backbone	Recall (%)	Precision (%)	IoU (%)	DSC (%)	Imp-Ave (%)
BV1000	mTAMS	SwinUNetr	64.08	56.31	39.22	54.56	6.07
		UNet++	69.29	56.87	44.35	60.13	7.68
		TransUNet	75.65	55.96	44.83	60.52	-4.55
		MANet	71.36	64.34	47.45	62.55	9.79
		UNet	**76.45**	**76.94**	**63.70**	**77.47**	**11.92**
	iTAMS	SwinUNetr	60.24	57.88	38.21	53.59	5.0
		UNet++	58.12	64.69	41.65	56.35	5.22
		TransUNet	71.77	55.02	43.90	59.35	-6.275
		MANet	72.46	67.05	49.95	64.68	**11.9**
		UNet	**76.38**	**76.59**	**62.27**	**76.57**	**11.23**
RPHS	mTAMS	MANet	46.55	45.57	27.72	41.93	3.52
		UNet++	47.73	45.55	28.95	43.25	4.83
		SwinUNetr	47.94	50.31	30.21	44.38	2.21
		TransUNet	45.07	**56.90**	31.43	46.33	**6.74**
		UNet	**50.44**	55.84	**45.86**	**62.87**	5.01
	iTAMS	UNet++	46.15	46.68	28.84	43.37	4.72
		MANet	46.97	47.04	29.41	44.15	4.97
		SwinUNetr	49.87	48.87	30.21	44.31	2.32
		TransUNet	42.07	**55.77**	30.79	45.52	**5.35**
		UNet	**52.36**	53.71	**45.36**	**62.41**	4.72

Imp-Ave is defined as the mean enhancement of TAMS compared to DL in four metrics: DSC, IoU, Recall, and Precision.

E. Results of Ablation Study on GSNet

To optimize high-resolution synthetic images generated by the Simulation Function Library and address invisible details missed during lesion customization, we designed GSNet. As shown in Table IV, GSNet significantly improved iTAMS metrics on the BV1000 dataset, with increases of 1.62% in DSC, 2.12% in IoU, and 4.24% in Recall. For the rare retinal disease RP segmentation task on the RPHS CFP dataset, the improvements were 3.81% in DSC, 3.89% in IoU, and 5.27% in Precision, respectively.

Additionally, our mTAMS method, when using GSNet, showed improvements of 3.23% in DSC, 4.4% in IoU, and 4.32% in Recall on the BV1000, and 5.33% in DSC, 5.47% in IoU, and 7.91% in Precision on RPHS. Overall, GSNet-optimized images significantly enhanced performance across three out of four metrics on both datasets.

F. Comparison with TL and SSL

TL is frequently employed in scenarios where sample resources are limited, while SSL typically addresses data labeling challenges. However, both methods necessitate additional data collection. In the absence of a disease dataset with a similar distribution as the target task, we directly use LSA to generate synthetic data. For TL, the synthetic data is first used for pre-training, then fine-tuned on the target task. For SSL, we apply the advanced MAE method. RETFound [68], which leverages MAE for self-supervised learning on large-scale OCT and CFP images, serves as a powerful feature extractor. We enhance RETFound with a SAM-like decoder to produce segmentation masks and perform SSL training on synthetic data.

As shown in Fig. 5 (A)&(B), the TL method boosts performance on six networks (UNet, UNet++, TransUNet, MANet, SwinUNetr, LiteMedSAM) on the BV1000 dataset, and it improves four of six networks On the RPHS dataset ("TL-CA" vs. "D-CA"). The SSL method on synthetic data also enhances performance on both datasets, demonstrating the effectiveness of our synthetic data for both TL and SSL. However, both methods lag behind the proposed TAMS, highlighting the need for innovation via meta-learning.

Notably, the SSL method achieves the lowest performance among all approaches. While SSL tackles data labeling challenges, it requires a large volume of unlabeled data [86]. The generation or collection of sufficient unlabeled data also poses significant challenges. Moreover, when all training data are labeled, SSL underperforms fully supervised methods. In summary, SSL is unsuitable for disease segmentation tasks without a large volume of unlabeled disease data. Even with

our synthetic dataset of a few thousand disease images, improvements are limited.

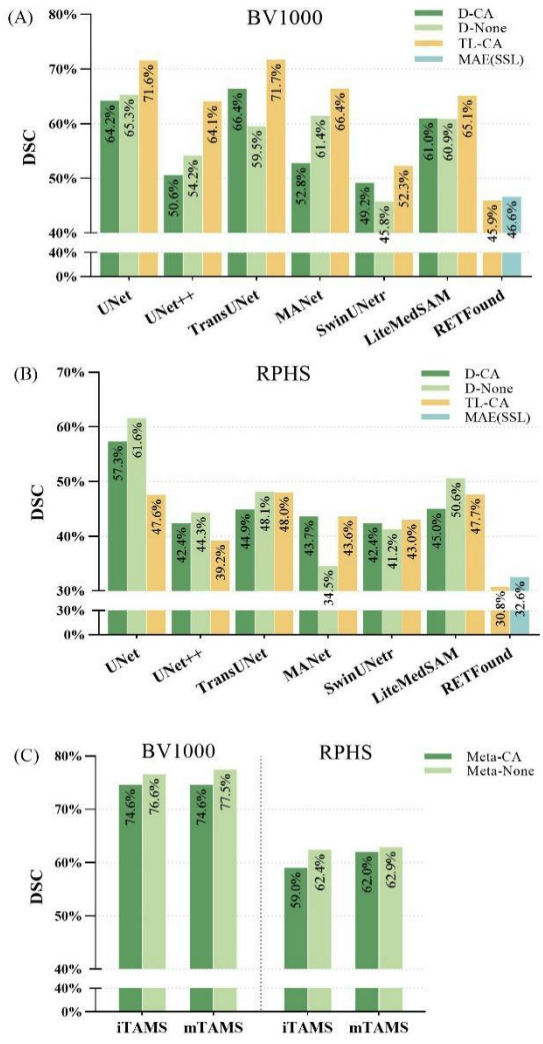


Fig. 5 Results of DL, TL, SSL, and common data augmentation methods. "CA" indicates that common data augmentation methods were used during DL, TL or meta-training, while "None" indicates that no data augmentation methods were used.

G. Result by Different Backbones as Segmentation Networks

In addition to UNet, we have integrated UNet++, MANet, SwinUNetr, and TransUNet into our TAMS framework (Table V). Compared with fully supervised DL training paradigms, TAMS achieved significant performance enhancements for four out of five architectures (excluding TransUNet) on the BV1000, while exhibiting universal optimization across all networks on RPHS. Despite not having the highest relative gains in all scenarios, UNet with TAMS maintained optimal performance. Notably, TAMS demonstrates stable enhancement effects on homogeneous architectures (e.g., pure CNN-based UNet or pure Transformer-based SwinUNet), yet requires robustness improvements for hybrid architectures like TransUNet.

VI DISCUSSIONS

A. Challenge in Lager Scale Data Collection and Annotation

Deep learning's strong dependence on large-scale data has severely hindered its application in the medical field, especially for the segmentation of rare retinal diseases. Firstly, collecting enough samples is challenging. Secondly, labeling these samples is difficult. Thirdly, the variety of medical data and different scanner specifications significantly limit the performance of segmentation models. TL [14-19], domain adaptation [34-36, 53], multitask learning [29-31, 87], and incremental learning[23-25, 27, 28], these methods struggle with the challenges of collecting similar datasets. Common data augmentation methods, such as rotation, noise injection, and cropping et al., only produce highly correlated images, limit the ability to emulate real lesion variations [49-52, 88]. Moreover, although the synthetic data generated by our LSA can be directly utilized to enhance TL and SSL methods, it is less effective than our TAMS, as shown in Fig. 5 (A)&(B), underscoring the value of our innovative meta-learning-based approach.

Generative models, such as CycleGAN [56], GAN [54] and Diffusion methods [57], face difficulties in controlling the generation process and present several issues: 1) lack of interpretability and difficulty in control during generation; 2) challenge in simultaneously simulating images and segmenting labels; 3) poor quality of complex retinal images, especially those with lesions, which fail to meet the requirements of segmentation models. To address these problems, we customize characteristics for specific diseases, then use models to optimize, compensating for factors not addressed in the customization process. Our approach enhances the interpretability of the simulation algorithm, ensures the quality of synthetic images, and generates corresponding segmentation labels simultaneously.

Our TAMS has demonstrated great potential in addressing challenges associated with limited resources and scarcity of labeled samples. TAMS leverages common knowledge learned from multiple meta-training tasks, constructed using multi-class synthetic images, to quickly adapt to lesion segmentation in target tasks.

B. Rich Multi-Class Data for Constructing Meta-Tasks

The effectiveness of meta-learning critically depends on the availability of rich multi-class data for constructing meta-tasks. Although there are natural image datasets for few-shot segmentation studies, such as FSS-1000 [78], there are no similar datasets in the medical field with rich categories and sufficient samples. Moreover, to fully realize the potential of meta-learning, the domain gap between tasks must be carefully calibrated.

The domain gap between meta-training tasks must be balanced—not too large or too small. R. Khadga et al. [47, 48] utilized five medical datasets as distinct classes, randomly sampling data from these five classes to construct multi-tasks. However, the large domain gap between meta-training tasks hindered the meta-learning algorithm's ability to learn useful

common prior knowledge from the known task distribution, as shown in Fig. 4. Conversely, if the distributions of different classes are too similar (e.g., using random rotations or deformations to augment the data), meta-learning struggles to extract useful common knowledge to generalize to new tasks, failing to enhance meta-tasks, as shown in Fig. 5 (c).

Therefore, it is necessary to simultaneously generate new samples and classes. Similarly, MetaMed [89] uses CutOut, MixUp, and CutMix to augment the data for meta-tasks. Although these methods improve model performance, they are mainly effective for classification tasks where labels are easily obtainable.

The domain gap between meta-training tasks and meta-testing tasks (target tasks) must not be too large. Even if the multi-tasks in meta-training share the same distributions, meta-level overfitting can still occur if the meta-training and meta-testing tasks are sampled from different task distributions or have a large domain gap [90]. For example, using multi-tasks constructed from the class-enriched FSS-1000 [78] dataset, which has a large domain gap from the target retinal dataset, makes it impossible to learn prior knowledge that can be used to segment retinal diseases, as shown in Fig. 4.

Rather than simply overlaying lesions from real RP or CNV images onto a background, the algorithm ensures a thorough analysis of various factors, concerning the relationship between the simulated lesion and the background. Specifically, it considers:

- Brightness Relationship: The brightness correlation between the simulated lesion and the background image.
- Pixel Distribution: The pixel distribution around the simulated lesion.
- Background Suitability: Whether the background image is suitable for generating the simulated lesions.
- Location Appropriateness: Determining the optimal placement of the simulated lesions.

In summary, the synthetic data automatically generated by our LSA not only possesses rich categories, but also ensures appropriate inter-domain differences between the multi-class data used for meta-training and the target data. Moreover, LSA has been applied to both common and rare retinal diseases (CNV, PED, Drusen, and RP), and is applicable to both OCT and CFP, independent of scanner specifications. It is also potentially extendable to more rare retinal diseases (e.g., Punctate Inner Choroidopathy, Primary Vitreoretinal Lymphoma).

C. TAMS Improvement on Retinal Disease Segmentation

We evaluated our method on two retinal diseases, CNV and RP, using three scanner specifications and two imaging types (CFP and OCT). Our method demonstrated superior lesion segmentation performance across various datasets. Specifically, on the BV1000 dataset, our method outperformed the second-best method, improving DSC, IoU, Recall, and Precision by 11.59%, 13.46%, 13.01%, and 6.64%, respectively. The improvements on the RPHS dataset were 4.24%, 4.46%, 6.02%, and 0.35%, respectively, while on the RIPS dataset, the improvements were 3.85%, 3.77%, 5.38%, and 3.21%, respectively. For the BV1000 dataset, our method required only 30% of the annotation data to achieve the same performance as the second-best methods using all the annotation data, and this proportion was 40% on the RPHS dataset. By utilizing common meta-task knowledge built on multi-class synthetic data, our method reaches the optimal solution with fewer model updates (Fig.13 in the Appendix). Furthermore, compared to TL, SSL, common data and task augmentation methods (Fig. 5), and various segmentation backbones (Table V), our approach demonstrates overall superior performance and demonstrated perfect extension on PED and Drusen (Fig. 12 in the Appendix).

D. Future work

Currently, hyperparameters involved in lesion simulation functions still need to be manually adjusted, which is highly sensitive. In the future, we plan to use DL models to play a greater role in the lesion customization process of LSA, thereby reducing manual intervention and increasing the method's universality. Additionally, future work will explore incorporating more advanced transformer-based models to extend TAMS's potential, especially medical-image pretrained larger models.

VII. Conclusion

In this paper, we proposed a task augmentation-based meta-learning method for retinal image segmentation (TAMS). Specifically, we introduced a retinal Lesion Simulation Algorithm (LSA) to automatically generate multi-class synthetic data and corresponding pixel-level segmentation labels for building meta-training tasks. LSA consists of a Simulation Function Library for lesion feature customization and a generative simulation network (GSNet) to optimize synthetic images. TAMS integrates two model-agnostic meta-learning algorithms namely iMAML and MAML, based on LSA's meta-training tasks to augment the meta-learning process at the task level. TAMS improves the model's segmentation performance by achieving higher accuracy, faster convergence rates, and reduced reliance on annotated samples; it is independent of scanner specifications and has potential for extension to additional retinal diseases and imaging modalities.

Acknowledgment

This work was supported in part by National Key R&D Program of China (2018YFA0701700), National Nature Science Foundation of China (U20A20170, 62271337, 62371328, 62371326, 82203273); Natural Science Foundation of Jiangsu Province of China (BK20211308, BK20231310, BK20220495); Natural Science Foundation of the Jiangsu Higher Education Institutions of China (No. 21KJB320018).

CODE AVAILABILITY

Source code available at https://github.com/wangjingtao93/TAMS.

DATA AVAILABILITY

The BV1000 and RPSH dataset supporting the findings of the current study are not publicly available due to the confidentiality policy of the National Health Commission of China and institutional patient privacy regulation. But we have released the synthetic data available at https://drive.google.com/drive/folders/19v-idbsfs8amRPfcQiMvbPrjD96x83JQ?usp=drive_link.

REFERENCES

[1] A. Guo, L. Fang, M. Qi, and S. Li, "Unsupervised denoising of optical coherence tomography images with nonlocal-generative adversarial network," *IEEE Transactions on Instrumentation and Measurement*, vol. 70, pp. 1-12, 2020.

[2] P. S. Grewal, O. Faraz, R. Uriel, and M. Tennant, "Deep learning in ophthalmology: a review," *Canadian Journal of Ophthalmology*, vol. 53, pp. 309-313, 2018.

[3] D. J. Fu, L. Faes, S. K. Wagner, G. Moraes, and P. A. Keane, "Predicting Incremental and Future Visual Change in Neovascular Age-Related Macular Degeneration Using Deep Learning," *Ophthalmology Retina*, 2021.

[4] J. Su, X. Chen, Y. Ma, W. Zhu, and F. Shi, "Segmentation of choroid neovascularization in OCT images based on convolutional neural network with differential amplification blocks," in *Medical Imaging 2020: Image Processing*, 2020, vol. 11313, pp. 491-497

[5] W. M. Abdelmoula, S. M. Shah, and A. S. Fahmy, "Segmentation of Choroidal Neovascularization in Fundus Fluorescein Angiograms," *IEEE Transactions on Bio-medical Engineering*, vol. 60, no. 5, pp. 1439-1445, 2013.

[6] P. Lopez, "Pathologic Features of Surgically Excised Subretinal Neovascular Membranes in Age-related Macular Degeneration," *American Journal of Ophthalmology*, vol. 112, no. 6, p. 647, 1991.

[7] A. Tewari, M. S. Dhalla, and R. S. Apte, "Intravitreal bevacizumab for treatment of choroidal neovascularization in pathologic myopia," *Retina*, vol. 26, no. 9, pp. 1093-1094, 2006.

[8] V. Rotemberg *et al.*, "A patient-centric dataset of images and metadata for identifying melanomas using clinical context," *Sci Data*, vol. 8, no. 1, p. 34, Jan. 2021, doi: 10.1038/s41597-021-00815-z.

[9] D. T. Hartong, E. L. Berson, and T. P. Dryja, "Retinitis pigmentosa," *The Lancet*, no. 9549, p. 368, 2006.

[10] A. A. Rusu *et al.*, "Meta-learning with latent embedding optimization," *arXiv preprint arXiv:1807.05960*, 2018.

[11] Y. Liu *et al.*, "Learning to propagate labels: Transductive propagation network for few-shot learning," in *International Conference on Learning Representations*, 2019,

[12] S. J. Pan, I. W. Tsang, J. T. Kwok, and Q. Yang, "Domain adaptation via transfer component analysis," *IEEE transactions on neural networks*, vol. 22, no. 2, pp. 199-210, 2010.

[13] S. Thrun and L. Pratt, "Learning to Learn: Introduction and Overview," in *Learning to learn*. Boston: Springer US, 1998, ch. Learning to Learn: Introduction and Overview, pp. 3-17.

[14] G. C. Y. Chan, A. Muhammad, S. A. A. Shah, T. B. Tang, C.-K. Lu, and F. Meriaudeau, "Transfer learning for diabetic macular edema (DME) detection on optical coherence tomography (OCT) images," in *2017 IEEE International Conference on Signal and Image Processing Applications (ICSIPA)*, 2017, pp. 493-496

[15] S. P. K. Karri, D. Chakraborty, and J. Chatterjee, "Transfer learning based classification of optical coherence tomography images with diabetic macular edema and dry age-related macular degeneration," *Biomedical Optics Express*, vol. 8, no. 2, pp. 579-592, 2017.

[16] M. Chaabane, A. Chehri, H. Chaibi, A. Elrharras, and R. Saadane, "Glaucoma Retinal Image Classification Based on Multichannel Gabor Filtering and Transfer Learning," in *2023 IEEE 97th Vehicular Technology Conference (VTC2023-Spring)*, June 20-23, 2023, pp. 1-6

[17] M. S. B. Mostafa, D. Bal, K. A. Sathi, and M. A. Hossain, "Diagnosis of Glaucoma from Retinal Fundus Image Using Deep Transfer Learning," in *2022 First International Conference on Artificial Intelligence Trends and Pattern Recognition (ICAITPR)*, March 10-12, 2022, pp. 1-4

[18] S. R, M. A. K, S. S. A, K. S. N, and C. S, "Retinal Disease Diagnosis Reimagined: A Deep Transfer Learning and Multi-Scale Attention Network Algorithm for Precision Ophthalmology," in *2023 International Conference on Energy, Materials and Communication Engineering (ICEMCE)*, Dec. 14-15, 2023, pp. 1-6

[19] E. Parra-Mora, A. Cazañas-Gordon, and L. A. d. S. Cruz, "Transfer Learning Based Retinal Layer and Fluid Segmentation using Fully Convolutional Networks," in *2022 10th European Workshop on Visual Information Processing (EUVIP)*, Sept. 11-14, 2022, pp. 1-5

[20] A. Zhao, G. Balakrishnan, F. Durand, J. V. Guttag, and A. V. Dalca, "Data Augmentation Using Learned Transformations for One-Shot Medical Image Segmentation," in *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, June 15-20 2019, pp. 8535-8545

[21] K. Chaitanya *et al.*, "Semi-supervised task-driven data augmentation for medical image segmentation," *Med Image Anal*, vol. 68, Feb. 2021.

[22] Q. Meng and S. Shin'Ichi, "ADINet: Attribute Driven Incremental Network for Retinal Image Classification," in *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2020,

[23] T. Hassan, B. Hassan, M. U. Akram, S. Hashmi, A. H. Taguri, and N. Werghi, "Incremental cross-domain adaptation for robust retinopathy screening via Bayesian deep learning," *IEEE Transactions on Instrumentation and Measurement*, vol. 70, pp. 1-14, 2021.

[24] F. Akmel, F. Meng, Q. Wu, S. Chen, R. Zhang, and M. Assefa, "Class similarity weighted knowledge distillation for few shot incremental learning," *Neurocomputing*, p. 127587, 2024, doi: https://doi.org/10.1016/j.neucom.2024.127587.

[25] S. Tian, L. Li, W. Li, H. Ran, X. Ning, and P. Tiwari, "A survey on few-shot class-incremental learning," *Neural Networks*, vol. 169, pp. 307-324, 2024, doi: https://doi.org/10.1016/j.neunet.2023.10.039.

[26] W. J. He *et al.*, "Incremental learning for exudate and hemorrhage segmentation on fundus images," *Information Fusion*, vol. 73, pp. 157-164, Sep. 2021, doi: 10.1016/j.inffus.2021.02.017.

[27] J. Y. Chen, E. Frey, and Y. Du, "Class-incremental learning for multi-organ segmentation," *Journal of Nuclear Medicine*, vol. 63, Jun. 2022. [Online]. Available: <Go to ISI>://WOS:000893739700335.

[28] M. H. Vu, G. Norman, T. Nyholm, and T. Löfstedt, "A Data-Adaptive Loss Function for Incomplete Data and Incremental Learning in Semantic Image Segmentation," *IEEE Transactions on Medical Imaging*, vol. 41, no. 6, pp. 1320-1330, 2022, doi: 10.1109/TMI.2021.3139161.

[29] X. Wang, L. Ju, X. Zhao, and Z. Ge, "Retinal Abnormalities Recognition Using Regional Multitask Learning," in *MICCAI 2019*, Shenzhen China, October 13–17, 2019, pp. 30-38

[30] B. Wang, H. Zhao, X. Li, M. Tian, B. Huang, and F. Yang, "Multi-Task Self-Supervised Learning for Medical Image Segmentation," in *ICASSP 2024 - 2024 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2024, pp. 1751-1755

[31] K. Ta *et al.*, "Multi-task Learning for Motion Analysis and Segmentation in 3D Echocardiography," *IEEE Transactions on Medical Imaging*, pp. 1-1, 2024, doi: 10.1109/tmi.2024.3355383.

[32] Y. Chen, C. Wu, S. Liu, and G. Cao, "Calibrated Uncertainty-Guided Multi-task Framework for Medical Image Segmentation," in *2023 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*, 2023, pp. 1060-1067

[33] D. Mahapatra and Z. Ge, "Training data independent image registration using generative adversarial networks and domain adaptation," *Pattern Recognition*, vol. 100, p. 107109, 2020.

[34] L. Ju, X. Wang, Q. Zhou, H. Zhu, and Z. Ge, "Bridge the Domain Gap Between Ultra-wide-field and Traditional Fundus Images via Adversarial Domain Adaptation," *arXiv preprint arXiv:2003.10042*, 2020.

[35] Z. Y. Chen, Y. S. Pan, and Y. Xia, "Reconstruction-Driven Dynamic Refinement Based Unsupervised Domain Adaptation for Joint Optic Disc and Cup Segmentation," *Ieee J Biomed Health*, vol. 27, no. 7, pp. 3537-3548, Jul. 2023, doi: 10.1109/Jbhi.2023.3266576.

[36] X. X. He, Z. Zhong, L. Y. Fang, M. He, and N. Sebe, "Structure-Guided Cross-Attention Network for Cross-Domain OCT Fluid Segmentation," *Ieee T Image Process*, vol. 32, pp. 309-320, 2023, doi: 10.1109/Tip.2022.3228163.

[37] H. C. Shin *et al.*, "Medical Image Synthesis for Data Augmentation and Anonymization Using Generative Adversarial Networks," in *Springer, Cham*, 2018,

[38] A. F. Frangi, S. A. Tsaftaris, and J. L. Prince, "Simulation and Synthesis in Medical Imaging," *IEEE Transactions on Medical Imaging*, vol. 37, no. 3, pp. 673-679, 2018, doi: 10.1109/TMI.2018.2800298.

[39] L. Kreitner *et al.*, "Synthetic optical coherence tomography angiographs for detailed retinal vessel segmentation without human annotations," *IEEE Transactions on Medical Imaging*, pp. 1-1, 2024, doi: 10.1109/TMI.2024.3354408.

[40] S. Amirrajab, Y. Al Khalil, C. Lorenz, J. Weese, J. Pluim, and M. Breeuwer, "Label-informed cardiac magnetic resonance image synthesis through conditional generative adversarial networks," *Computerized Medical Imaging and Graphics*, vol. 101, p. 102123, 2022, doi: https://doi.org/10.1016/j.compmedimag.2022.102123.

[41] C. Finn, P. Abbeel, and S. Levine, "Model-agnostic meta-learning for fast adaptation of deep networks," in *International Conference on Machine Learning*, 2017, pp. 1126-1135

[42] A. Nichol, J. Achiam, and J. Schulman, "On First-Order Meta-Learning Algorithms," *arXiv preprint arXiv:1803.02999*, 2018.

[43] A. Rajeswaran, C. Finn, S. M. Kakade, and S. Levine, "Meta-learning with implicit gradients," *Advances in Neural Information Processing Systems*, vol. 32, 2019.

[44] Q. Sun, Y. Liu, T. S. Chua, and B. Schiele, "Meta-Transfer Learning for Few-Shot Learning," in *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2019,

[45] S. Luo, Y. Li, P. Gao, Y. Wang, and S. Serikawa, "Meta-seg: A survey of meta-learning for image segmentation," *Pattern Recognition*, vol. 126, p. 108586, 2022, doi: https://doi.org/10.1016/j.patcog.2022.108586.

[46] P. Tian, Z. Wu, L. Qi, L. Wang, Y. Shi, and Y. Gao, "Differentiable meta-learning model for few-shot semantic segmentation," in *Proceedings of the AAAI Conference on Artificial Intelligence*, 2020, vol. 34, no. 07, pp. 12087-12094

[47] R. Khadka *et al.*, "Meta-learning with implicit gradients in a few-shot setting for medical image segmentation," *Computers in Biology and Medicine*, vol. 143, p. 105227, 2022.

[48] R. Khadga, J. H. A. Debesh, and A. L. I. Sharib, "Few-shot segmentation of medical images based on meta-learning with implicit gradients," *arXiv preprint arXiv:2106.03223*, 2021.

[49] A. Oliveira, S. Pereira, and C. A. Silva, "Augmenting data when training a CNN for retinal vessel segmentation: How to warp?," in *2017 IEEE 5th Portuguese Meeting on Bioengineering (ENBENG)*, Feb. 16-18, 2017, pp. 1-4

[50] P. Seeböck, J. I. Orlando, M. Michl, J. Mai, U. Schmidt-Erfurth, and H. Bogunović, "Anomaly guided segmentation: Introducing semantic context for lesion segmentation in retinal OCT using weak context supervision from anomaly detection," *Med Image Anal*, vol. 93, p. 103104, 2024, doi: https://doi.org/10.1016/j.media.2024.103104.

[51] Y. Zhou *et al.*, "CF-Loss: Clinically-relevant feature optimised loss function for retinal multi-class vessel segmentation and vascular feature measurement," *Med Image Anal*, vol. 93, p. 103098, 2024, doi: https://doi.org/10.1016/j.media.2024.103098.

[52] T. Jahnavi and D. Vasundhara, "Segmentation of medical images using U-Net++," in *2022 4th International Conference on Advances in Computing, Communication Control and Networking (ICAC3N)*, Dec.16-17, 2022, pp. 801-807

[53] J. Wang, M. M. Cheng, and J. Jiang, "Domain Shift Preservation for Zero-Shot Domain Adaptation," *IEEE Trans Image Process*, vol. 30, pp. 5505-5517, 2021, doi: 10.1109/tip.2021.3084354.

[54] I. Goodfellow *et al.*, "Generative adversarial networks," *Communications of the ACM*, vol. 63, no. 11, pp. 139-144, 2020.

[55] M. Mirza and S. Osindero, "Conditional generative adversarial nets," *arXiv preprint arXiv:1411.1784*, 2014.

[56] 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

[57] A. Q. Nichol and P. Dhariwal, "Improved denoising diffusion probabilistic models," in *International Conference on Machine Learning*, 2021, pp. 8162-8171

[58] J. Schmidhuber, "Evolutionary principles in self-referential learning," *Genetic Programming*, 1987.

[59] O. Vinyals, C. Blundell, T. Lillicrap, K. Kavukcuoglu, and D. Wierstra, "Matching Networks for One Shot Learning," *Advances in Neural Information Processing Systems*, vol. 29, 2016.

[60] S. Larochelle, "Optimization as a model for few-shot learning," in *International Conference on Learning Representations*, 2017,

[61] Y. Lee and S. Choi, "Gradient-based meta-learning with learned layerwise metric and subspace," in *International Conference on Machine Learning*, 2018, pp. 2927-2936

[62] K. Lee, S. Maji, A. Ravichandran, and S. Soatto, "Meta-Learning with Differentiable Convex Optimization," in *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2019, pp. 10657-10665

[63] J. Snell, K. Swersky, and R. Zemel, "Prototypical networks for few-shot learning," *Advances in neural information processing systems*, vol. 30, 2017.

[64] F. Sung, Y. Yang, Z. Li, X. Tao, and T. M. Hospedales, "Learning to Compare: Relation Network for Few-Shot Learning," in *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2018,

[65] B. N. Oreshkin, A. Lacoste, and P. Rodriguez, "TADAM: Task dependent adaptive metric for improved few-shot learning," *Advances in Neural Information Processing Systems*, vol. 31, 2018.

[66] W. Weng and X. Zhu, "INet: Convolutional Networks for Biomedical Image Segmentation," *IEEE Access*, vol. PP, no. 99, pp. 1-1, 2021.

[67] K. He, X. Chen, S. Xie, Y. Li, P. Dollár, and R. Girshick, "Masked autoencoders are scalable vision learners," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2022, pp. 16000-16009

[68] Y. Zhou *et al.*, "A foundation model for generalizable disease detection from retinal images," *Nature*, vol. 622, no. 7981, pp. 156-163, 2023.

[69] C. Li and M. Wand, "Precomputed real-time texture synthesis with markovian generative adversarial networks," in *Computer Vision–ECCV 2016: 14th European Conference*, Amsterdam Netherlands, October 11-14, 2016, pp. 702-716

[70] P. Isola, J.-Y. Zhu, T. Zhou, and A. A. Efros, "Image-to-image translation with conditional adversarial networks," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2017, pp. 1125-1134

[71] N. Brancati *et al.*, "Learning-based approach to segment pigment signs in fundus images for retinitis pigmentosa analysis," *Neurocomputing*, vol. 308, pp. 159-171, 2018.

[72] G. Challenge. 2019, *Peking University International Competition on Ocular Disease Intelligent Recognition (ODIR-2019)*. [Online]. Available: https://github.com/talhaanwarch/ODIR2019

[73] 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 Imaging*, vol. 23, no. 4, pp. 501-9, Apr. 2004, doi: 10.1109/TMI.2004.825627.

[74] Bernal *et al.*, "WM-DOVA maps for accurate polyp highlighting in colonoscopy: Validation vs. saliency maps from physicians," *Computerized Medical Imaging and Graphics: The Official Jounal of the Computerized Medical Imaging Society*, 2015.

[75] T. Mendona, P. M. Ferreira, J. S. Marques, A. R. S. Marcal, and J. Rozeira, "PH 2-A dermoscopic image database for research and benchmarking," in *2013 35th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, 2013, pp. 5437-5440

[76] D. Jha *et al.*, "Kvasir-seg: A segmented polyp dataset," in *MultiMedia Modeling: 26th International Conference*, Daejeon South Korea, January 5-8, 2020, pp. 451-462

[77] D. Jha *et al.*, "NanoNet: Real-Time Polyp Segmentation in Video Capsule Endoscopy and Colonoscopy," in *IEEE 34th International Symposium on Computer-Based Medical Systems (CBMS)*, Aveiro, Portugal, 2021, pp. 37-43

[78] X. Li, T. Wei, Y. P. Chen, Y.-W. Tai, and C.-K. Tang, "Fss-1000: A 1000-class dataset for few-shot segmentation," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2020, pp. 2869-2878

[79] G. Staurenghi, S. Sadda, U. Chakravarthy, and R. F. Spaide, "Proposed Lexicon for Anatomic Landmarks in Normal Posterior Segment Spectral-Domain Optical Coherence Tomography: The IN•OCT Consensus," *Ophthalmology*, vol. 121, no. 8, pp. 1572-1578, 2014, doi: https://doi.org/10.1016/j.ophtha.2014.02.023.

[80] Z. Zhou, M. M. R. Siddiquee, N. Tajbakhsh, and J. Liang, "Unet++: Redesigning skip connections to exploit multiscale features in image

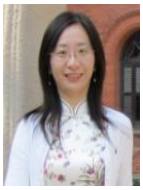
segmentation," *IEEE Transactions on Medical Imaging*, vol. 39, no. 6, pp. 1856-1867, 2019.
[81] S. Guo, S. Sheng, Z. Lai, and S. Chen, "Trans-U: Transformer Enhanced U-Net for Medical Image Segmentation," presented at the 3rd International conference on computer vision, image and deep learning & international conference on computer engineering and applications (CVIDL & ICCEA), 2022.
[82] K. Jiang, J. Liu, W. Zhang, F. Liu, and L. Xiao, "MANet: An Efficient Multidimensional Attention-Aggregated Network for Remote Sensing Image Change Detection," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 61, pp. 1-18, 2023, doi: 10.1109/TGRS.2023.3328334.
[83] F. Isensee *et al.*, "nnu-net: Self-adapting framework for u-net-based medical image segmentation," *arXiv preprint arXiv:1809.10486*, 2018.
[84] A. Hatamizadeh, V. Nath, Y. Tang, D. Yang, H. R. Roth, and D. Xu, "Swin unetr: Swin transformers for semantic segmentation of brain tumors in mri images," in *International MICCAI brainlesion workshop*, 2021, pp. 272-284
[85] J. Ma, Y. He, F. Li, L. Han, C. You, and B. Wang, "Segment anything in medical images," *Nature Communications*, vol. 15, no. 1, p. 654, 2024.
[86] Y. Chen, M. Mancini, X. Zhu, and Z. Akata, "Semi-Supervised and Unsupervised Deep Visual Learning: A Survey," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 46, no. 3, pp. 1327-1347, 2024, doi: 10.1109/TPAMI.2022.3201576.
[87] Y. Chen, C. Wu, S. Liu, and G. Cao, "Calibrated Uncertainty-Guided Multi-task Framework for Medical Image Segmentation," 2023.[Online]. Available: https://dx.doi.org/10.1109/bibm58861.2023.10385430
[88] Z. Eaton-Rosen, F. J. S. Bragman, S. Ourselin, and M. J. Cardoso, "Improving Data Augmentation for Medical Image Segmentation," 2018,
[89] H. Zhao *et al.*, "MetaMed: Linking Microbiota Functions with Medicine Therapeutics," *mSystems*, vol. 4, no. 5, Oct 8 2019, doi: 10.1128/mSystems.00413-19.
[90] W. Y. Chen, Y. C. Liu, Z. Kira, Y. Wang, and J. B. Huang, "A Closer Look at Few-shot Classification," in *arxiv:1904.04232*, 2019,

Jingtao Wang received the BSc degree in Information and Communication Engineering from Xidian University in 2016. Currently, he is pursuing the PhD in the School of Electronic and Information Engineering, Soochow University, China. His research interests include medical image analysis, meta learning.

Muhammad Mateen (Member, IEEE) received the PhD degree in Software Engineering from Chongqing University, China in 2020. Currently, he is working as Postdoctoral Researcher in the School of Electronic and Information Engineering, Soochow University, China. He has published more than 15 articles in prestigious journals and conferences. His research interests include medical image analysis and deep learning.

Dehui Xiang (Member, IEEE) received the BSc degree in automation from Sichuan University, China in 2007, and received the PhD degree from State Key Lab of Management and Control for Complex Systems, Institute of Automation, Chinese Academy of Sciences, China in 2012. Currently, he is working as Associated Professor in the School of Electronic and Information Engineering, Soochow University. His research interests include volume rendering, medical image analysis, computer vision, and pattern recognition.

Weifang Zhu received the BSc degree from Xi'an Jiaotong University and PhD degree from Suzhou University, China. Currently, she is working as Associated Professor in the School of Electronic and Information Engineering, Soochow University, China. Her research interests include image processing, pattern recognition, and machine learning.



Fei Shi received the BSc degree in Information and Electronics Engineering from Zhejiang University. She received the PhD degree in Electrical Engineering from Polytechinic University, Brooklyn, New York in 2006. Currently, she is working as Associated Professor in the School of Electronic and Information Engineering, Soochow University, China. Her research interests include image processing, pattern recognition, and machine learning.

Jing Huang received the BSc degree in clinical medicine from Huazhong University of Science and Technology, China in 2014, and the PhD degree in cell biology in Capital Medical University, China in 2020. Currently, she is working as a lecturer in the School of Basic Medical Science, Soochow University, China. Her research interests include epigenetic of liver tumor genesis.

Sun Kai received the PhD degree in Information and Communication Engineering from Xidian University in 2022. Currently, he is working as a lecturer at the College of Medical Information and Artificial Intelligence, Shandong First Medical University in China. His research focuses on radiomics and brain imaging analysis.

Dai Jun received the BSc degree from East China Jiaotong University, China in 2022. Currently, he is pursuing his master degree in the School of Electronic and Information Engineering, Soochow University, China. Her research interests include medical image analysis, meta learning.

Jingcheng Xu received the BSc degree and master degree from School of Electronic and Information Engineering, Soochow University, China in 2022. Her research interests include medical image processing, analysis and visualization

Su Zhang received the BSc degree from Nanjing University of Information Science & Technology, China in 2022. Currently, she is pursuing her master degree in the School of Electronic and Information Engineering, Soochow University, China. Her research interests include medical image processing, analysis and visualization.

Dr. Xinjian Chen (Senior Member, IEEE) received the Ph.D. degree from the Institute of Automation, Chinese Academy of Sciences, China, in 2006. He was a research fellow with University of Pennsylvania, National Institutes of Health (NIH), University of Lowa., from 2003 to 2010. Currently, he is working as Professor in the School of Electronic and Information Engineering, Soochow University. He is the Director of Medical Image Processing, Analysis and Visualization Laboratory. His research interests include medical image processing and analysis, pattern recognition and machine learning. He is an Associate Editor of *IEEE Transactions on Medical Imaging and IEEE Journal of Translational Engineering in Health and Medicine.*