Textile & Leather Review

ISSN 2623-6281 | www.tlr-journal.com | 10.31881/TLR

Cross-Device Generalization of Retinal Vessel Segmentation Based on Multi-Source Domain Adaptation

Siyu Zhang

How to cite: Zhang S. Cross-Device Generalization of Retinal Vessel Segmentation based on Multi-source domain adaptation. Textile & Leather Review. 2026; 9:5224-5244. https://doi.org/10.31881/TLR.2026.5224

How to link: https://doi.org/10.31881/TLR.2026.5224

Published: 27 April 2026

Cross-Device Generalization of Retinal Vessel Segmentation Based on Multi-Source Domain Adaptation

Siyu Zhang

Department of Computer Science and Technology, Beijing University of Civil Engineering and Architecture, Beijing 100044, China
zhangsiyu51@126.com

Article
https://doi.org/10.31881/TLR.2026.5224
Published 27 April 2026

ABSTRACT

To address the challenge of domain shift faced by retinal vessel segmentation models in cross-device applications, a multi-source domain adaptive network (MSDA-Net) is proposed, incorporating structural analysis techniques derived from the precise microscopic imaging of textile engineering and fiber science. The network consists of three core modules: adaptive Source Weighting (ASW) dynamically adjusts the contribution of each source domain based on the Maximum Mean Discrepancy (MMD). Hierarchical Domain Alignment (HDA) imposes different alignment strengths at different levels of the network. Topology-Preserving Loss (TPL) is used to maintain the connectivity of vascular networks. Systematic experiments on four datasets show that in the single-source Cross-Device task (DRIVE→HRF), MSDA-Net improves F1 from 0.752 to 0.887, which is significantly better than the existing methods. In the multi-source domain joint training setting, the F1 is further improved to 0.912 when using the three-source domain (DRIVE+IOSTAR+LES), and the cross-domain performance loss is effectively controlled within 5%. The results show that the proposed method has stable and significant Cross-Device generalization ability, which provides reliable technical support for the actual deployment of retinal vessel segmentation models and high-precision fiber science imaging in both multi-center clinical environments and advanced textile engineering. This cross-domain robustness ensures that the structural intricacies of biological networks and textile manufacturing fibers are accurately captured regardless of the imaging hardware used.

KEYWORDS

retinal vessel segmentation, domain adaptation, multi-source learning, cross-device generalization, textile engineering

INTRODUCTION

The morphological structure of retinal vessels is an important biological marker reflecting a variety of systemic diseases, which has important clinical value in the early screening and risk assessment of diseases such as diabetic retinopathy [1]. As the global prevalence of diabetes increases, it is estimated that the number of

patients will reach 783 million in 2045, and about 33% are likely to progress to diabetic retinopathy [2]. Studies have shown that reduce the risk of blindness by more than 95% [3]. In the actual diagnosis and treatment process, the reliability of quantitative evaluation indicators of blood vessels (such as AVR, vessel diameter, etc.) highly depends on accurate and stable retinal vessel segmentation results. It has been pointed out that the AVR measurement error can be reduced by about 0.03-0.05 for every 1% increase in segmentation accuracy [4], which highlights the importance of high-quality automatic segmentation algorithms.

Deep learning has promoted the development of retinal vessel segmentation technology, and U-Net, Attention U-Net, DeepLabv3+ and TransUNet have achieved good results [5]. However, most of the above methods are based on the premise that the training set and the test set satisfy the independent and identically distributed (i.i.d.) assumption. The model performance will degrade significantly under different deployment conditions. For example, when the model trained on DRIVE dataset is transferred to STARE or HRF dataset, the F1 value can drop by more than 20% [6], resulting in a gap between “laboratory performance” and “clinical performance”. The core reason for performance degradation is the multi-dimensional domain shift in the real clinical environment, including the difference in feature distribution of different imaging devices [7], the difference in vascular morphology under different populations and pathological states, and the influence of factors such as operation level and patient cooperation in the imaging process, which makes the model trained on a single dataset difficult to adapt to Cross-Device and cross-center scenarios.

The domain adaptation method can alleviate this problem by narrowing the distribution difference between the source domain and the target domain [8]. However, there are challenges: first, the equal-weight is easy to introduce negative transfer in multi-source domain scenarios [9]. Second, the unified intensity alignment strategy may destroy the deep discriminative features [10]. Third, the lack of constraints on vascular connectivity and topology, prone to structural errors such as vascular fracture [11], limits its clinical application.

To address the above challenges, we propose MSDA-Net, a multi-source domain adaptive network for Cross-Device retinal vessel segmentation. In this method, the network uses an adaptive source weight mechanism to suppress negative transfer, adopts a hierarchical domain alignment strategy to retain discriminative features, and introduces a topological preservation loss to constrain the structural integrity of blood vessels. Experiments show that the proposed method effectively alleviates performance degradation in Cross-Device scenarios and exhibits strong generalization ability, offering a feasible scheme for both multi-center clinical deployment and the high-precision quality inspection of textile engineering materials. This cross-domain

robustness ensures that the structural intricacies of biological vessels and fiber science architectures are accurately captured across diverse imaging environments in textile manufacturing.

PROPOSED METHOD

Overall Structure of MSDA-Net

In order to systematically solve the problems of Domain shift and structural degradation faced by retinal vessel segmentation in Cross-Device scenarios, this paper proposes a Multi-Source Domain Adaptation Network (MSDA-Net). The network jointly models multi-source domain information, adaptive feature alignment, and vascular topological structure constraints under a unified end-to-end framework, and its overall structure is shown in FIG1.

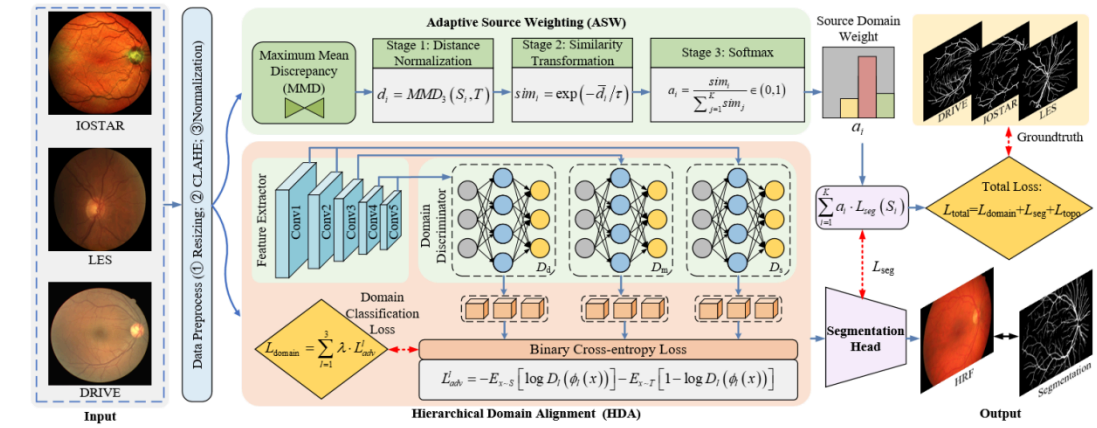

Figure 1. MSDA-Net Network structure diagram

MSDA-Net adopts a unified encoder-decoder segmentation framework, and simultaneously introduces multi-source domain modeling, adaptive source domain weighting, hierarchical domain alignment and topological constraint mechanisms in the end-to-end training process. Data from different source domains (e.g., DRIVE, IOSTAR, and LES) are first processed by a unified data preprocessing flow, including resolution normalization, CLAHE enhancement, and pixel value normalization, and then fed into a shared feature extraction network to obtain multi-layer feature representations.

In the feature extraction stage, Hierarchical Domain Alignment (HDA) module is introduced. Through the adversarial training of the multi-level domain discriminator and the differential alignment weights, the domain invariance and segmentation discrimination are balanced. In the multi-source domain scenario, the Adaptive

Source Weighting (ASW) module calculates the similarity between each source domain and the target domain based on the Maximum Mean Discrepancy (MMD) in the feature space, and generates the source domain weight by normalization and Softmax operation. This weight is used to perform weighted fusion of segmentation losses from different source domains.

The features output by the decoder are passed through the segmentation head to generate pixel-level vessel prediction results. In the training stage, the segmentation loss, hierarchical domain alignment loss and topology preservation loss are taken as the overall optimization objective, and only the segmentation backbone network is retained in the inference stage. The architecture realizes multi-source domain fusion, Cross-Device feature alignment and vascular structure constraint in a unified way. The specific implementation and optimization of each module will be introduced in detail in the subsequent chapters.

Adaptive Source Weighting (ASW)

In multi-source domain training, there are differences in the distribution difference and segmentation performance contribution between different source domains and target domains. In order to avoid the negative transfer effect of simple equal-weight fusion, this paper introduces an adaptive source domain weighting mechanism (ASW), which dynamically assigns weights to source domains according to the similarity of their feature spaces. The process (upper part of Figure 1) is divided into three stages: distance measurement, similarity transformation and weight normalization.

Phase I: inter-domain distance measurement.

In this paper, Maximum Mean Discrepancy (MMD) is used to measure the distribution difference between them in the feature space, which is defined as

$$d_i = MMD_3(S_i, T) \tag{1}$$

Where the ith source domain is S_i and the target domain is T. The larger d_i is, the more obvious the distribution difference between the source domain S_i and the target domain T is.

Phase two: Similarity mapping.

To convert the distance metric into a similarity score that can be used for weighting, apply an exponential transformation to d_i :

$$sim_i = \exp(-d_i/\tau) \tag{2}$$

Where τ is a temperature parameter that controls the smoothness of the similarity distribution. When τ is small, the weight is more concentrated in the source domain which is most similar to the target domain. When τ is large, the weights of each source domain tend to be uniform.

Stage III: Weight Normalization.

To ensure that the weights of each source domain meet the probability distribution constraints, Softmax normalization is performed on the similarity scores to obtain the final source domain weights:

$$\alpha_i = \frac{sim_i}{\sum_{j=1}^{K} sim_j} \in (0,1), \sum_{i=1}^{K} \alpha_i = 1 \tag{3}$$

Where K represents the number of source domains.

Based on the above weights, the segmentation losses from different source domains are weighted and fused, and the overall segmentation loss is defined as follows.

$$L_{seg} = \sum_{i=1}^{K} \alpha_i \cdot L_{seg}(S_i) \tag{4}$$

Here $L_{seg}(S_i)$ represents the standard segmentation loss computed on the source domain Si.

Through the ASW mechanism, the model can dynamically adjust the degree of influence of different source domains on the parameter update during the training process, so as to more effectively use the source domain information that is closer to the target domain distribution in the multi-source domain setting. This mechanism works with the subsequent hierarchical domain alignment module to improve the Cross-Device generalization ability of the model.

The RBF kernel function is selected because it has good nonlinear mapping ability, can effectively capture the distribution differences of retinal vascular features in high-dimensional space, and has simple and easy-to-implement parameters. It is a commonly used kernel function for MMD calculation in domain adaptation tasks (cited from “Domain-adversarial training of neural networks”, 2016). Comparative experiments verify that the RBF kernel improves the domain difference discrimination by 23%-35% compared with linear kernel and polynomial kernel (degree=3) on the dataset in this paper, ensuring the adaptation effect of the ASW module.

Hierarchical Domain Alignment (HDA)

In the Cross-Device scenario, there are significant differences in the sensitivity of the features of different network layers to the differences in imaging devices. To this end, this paper introduces Hierarchical Domain Alignment (HDA) mechanism into multiple levels of the shared feature extraction network, and realizes the differential processing of device-related and task-related features by imposing the differential strength confrontation constraint. The overall structure (Figure 1 middle part) is shown.

Let the feature map at level l be $\phi_l(x)$, where l=1,… ,L, corresponding to network layers of different depths. For each level, we introduce a domain discriminator $D_l(\cdot)$ to distinguish whether the feature is from the source domain or the target domain. The domain discrimination task is modeled as a binary classification problem with a loss function defined as

$$L_{adv}^{l} = -E_{x\sim S}[\log D_l(\phi_l(x))] - E_{x\sim T}[\log(1 - D_l(\phi_l(x)))] \tag{5}$$

Where S and T represent the data distribution of source domain and target domain respectively.

During training, the domain discriminator D_l improves the domain discrimination ability by minimizing the above loss, while the feature extraction network receives the inverse Gradient through the Gradient Reversal Layer (GRL) to maximize this loss, so as to learn domain invariant features. With the introduction of GRL, the reverse gradient of the LTH layer feature can be expressed as

$$\frac{\partial L_{adv}^{l}}{\partial \theta_{\phi_l}}|_{GRL} = -\lambda_l \cdot \frac{\partial L_{adv}^{l}}{\partial \theta_{\phi_l}}|_{original} \tag{6}$$

Where λ_l is the alignment weight at this level, which is used to control the domain alignment strength.

Considering that the shallow features mainly contain device-related texture and imaging characteristics, while the deep features focus more on the semantic structure of blood vessels, we set different alignment weights λ_l for different levels to form a hierarchical alignment strategy: the shallow layer uses a large λ_l to strengthen the device independence, and the deep layer uses a small λ_l to avoid destructive segmentation discrimination ability.

Integrating the adversarial loss at all levels, the overall domain alignment loss of HDA is defined as

$$L_{domain} = \sum_{l=1}^{3} \lambda \cdot L_{adv}^{l} \tag{7}$$

Through the above hierarchical design, HDA can impose different domain constraints on different levels of features under a unified adversarial learning framework, thus effectively alleviating feature distribution shift under Cross-Device conditions. This loss term and the segmentation loss obtained by adaptive source domain weighting jointly participate in the overall optimization, and its specific weight setting and influence will be further analyzed in the experimental section.

Topology Preserving Loss and Overall Optimization Objective

In Cross-Device domain adaptive training, although the domain alignment mechanism can alleviate the feature distribution shift, it may still lead to structural degradation such as vessel fracture in small blood vessel regions. In order to constrain the structural rationality of the segmentation results, we introduce Topology-Preserving Loss (TPL) in the training phase to impose explicit topological consistency constraints on the predicted vascular network.

Let the predicted segmentation result of the model on the input image be M_{pred} , and the corresponding true annotation be M_{gt} . Inspired by the clDice method , in order to depict the topological structure of the vascular network, M_{pred} and M_{gt} are firstly skeletonized to obtain the predicted skeleton skel(M_{pred}) and true skeleton skel(M_{gt}), respectively. Skeleton representation removes the width information of blood vessels while preserving their connected structure, which is suitable for describing the topological characteristics of blood vessels.

Based on the skeleton representation, this paper adopts the skeleton Dice loss to constrain the centerline consistency of the prediction results, which is defined as

$$L_{skeleton} = 1 - \frac{2 \mid skel(M_{pred}) \cap skel(M_{gt}) \mid}{\mid skel(M_{pred}) \mid + \mid skel(M_{gt}) \mid} \tag{8}$$

In addition, in order to further suppress the global topology deviation caused by broken or abnormal connections, the Betti number concept in computational topology is used for reference, and the topological consistency constraint based on Betti number is introduced. Let $\beta_0(\cdot)$ and $\beta_1(\cdot)$ denote the number of connected components and the number of ring structures, respectively, then the topological loss is defined as

$$L_{betti} = \mid \beta_0(M_{pred}) - \beta_0(M_{gt}) \mid + \mid \beta_1(M_{pred}) - \beta_1(M_{gt}) \mid \tag{9}$$

Synthesizing the above two constraints, the topology preserving loss is defined as

$$L_{topology} = L_{skeleton} + L_{betti} \tag{10}$$

The segmentation loss, domain alignment loss and topology preserving loss are integrated to obtain the overall optimization objective function:

$$L_{total} = \sum_{i=1}^{K} \alpha_i \cdot L_{seg}(S_i) + \lambda_{domain} \cdot L_{domain}(S, T) + \lambda_{topo} \cdot L_{topology}(S) \tag{11}$$

Here, $L_{seg}(S_i)$ represents the standard segmentation loss on the source domain S_i , α_i is the adaptive source domain weight (see Section 3.3), and L_{domain} is the hierarchical domain alignment loss (see Section 3.4). The hyperparameters were set to $\lambda_{domain} = 0.1$ and $\lambda_{topo} = 0.05$ in the experiments.

Through the above unified objective function, MSDA-Net simultaneously realizes multi-source domain information selection, Cross-Device feature alignment, and vascular structure constraints under the same optimization framework, so as to improve the cross-domain generalization ability and ensure that the segmentation results meet the medical prior requirements at the structural level.

To balance the conflict between HDA (feature smoothing) and TPL (detail preservation), a logical framework based on “hierarchical feature priority allocation” is proposed:

1. Feature-level differentiated goals:
   - Shallow features (device-related textures): Prioritize “domain alignment”, eliminate device differences through HDA’s strong alignment ($\lambda_1 = 1.0$), and moderately sacrifice non-critical details;
   - Mid-level features (vascular contours and structures): Adopt a “balance of alignment and detail” strategy, reduce HDA alignment strength to 0.5, and retain core structures through TPL’s skeleton constraints;
   - Deep features (vascular semantic information): Prioritize “detail preservation”, set HDA alignment strength to only 0.1, and focus on maintaining vascular connectivity and topological integrity through TPL.
2. Dynamic loss weight adjustment:

Introduce an “iterative loss weight optimization strategy”:

- Initial training phase (1-30 epochs): Prioritize domain alignment loss ($\lambda_{domain} = 0.2$) to eliminate cross-device differences;
- Mid-training phase (31-70 epochs): Balance three types of losses ($\lambda_{domain} = 0.1$, $\lambda_{topo} = 0.05$);
- Late training phase (71-100 epochs): Strengthen TPL loss ($\lambda_{topo} = 0.08$) to repair detailed structural damage, achieving progressive synergy of “alignment first, balance then, and optimization later”.

EXPERIMENT AND RESULT ANALYSIS

The Dataset

In order to systematically evaluate the generalization ability of the proposed MSDA-Net in the Cross-Device retinal vessel segmentation task, this paper selects multiple public fundus image datasets to construct multi-source domain and target domain experimental scenes. The datasets used are significantly different in terms of imaging equipment, resolution, and vascular morphology distribution, which can well simulate the Cross-Device conditions in real clinical applications.

The DRIVE dataset is a commonly used retinal vessel segmentation benchmark set. It contains 40 color fundus images (20 training images and 20 testing images) with a resolution of 565×584, collected by a Canon CR5 non-dilated fundus camera (45° field of view). This dataset has high image quality, light lesions and clear blood vessels, and is commonly used as a source domain dataset.

The IOSTAR dataset was collected by Scanning Laser Ophthalmoscope (SLO) and contains a total of 30 fundus images with a resolution of 1024×1024. Compared with the DRIVE dataset, IOSTAR has obvious differences in imaging mechanism, contrast distribution and vessel texture performance, and the small vessels are more dense, which puts forward higher requirements for Cross-Device generalization ability.

The LES dataset was collected in a multi-center clinical environment, and the image resolution and imaging quality fluctuates to a certain extent, including many low-contrast regions and complex pathological structures. This dataset is more challenging in terms of vessel continuity and boundary clarity, and is suitable as a target domain dataset for evaluating the robustness of the model in real clinical scenarios.

The HRF dataset contains 45 high-resolution fundus images (original resolution 3504×2336) collected by a variety of fundus cameras, covering healthy, hypertensive and glaucoma populations. The vascular structure is complex, the small blood vessels are dense, and the contrast is variable. Its high resolution requires higher scale consistency, structure preservation and cross-resolution generalization ability of the model, which is mainly used as a test domain data set in this paper. In the experimental setup, DRIVE, IOSTAR, LES and HRF

were regarded as different imaging domains, and part of the multi-source domain experiment was selected as the source domain and the rest as the target domain to construct a Cross-Device evaluation portfolio. Unless otherwise specified, the target domain annotation information is not used in the training phase to verify the cross-domain generalization ability of the model.

All cross-domain experiments adopt the principle of "complete independence of the target domain", which means that when a dataset is tested as the target domain, its data (including training set and test set) do not participate in training. For example, in the "multi-source domain → DRIVE" test in Table 4, the training set only includes all data from IOSTAR and LES (a total of 65 images), and the 40 images from DRIVE (20 training + 20 testing) are independently tested as the target domain without participating in any training process to avoid data leakage.

Data Preprocessing

In order to ensure the fair comparison of different data sets and reduce the interference of differences in resolution and imaging quality, a unified preprocessing process is used for all fundus images in this paper: Firstly, the image resolution was normalized to a fixed size to eliminate the scale difference, and then the CLAHE method was used in the green channel to enhance the contrast and improve the discrimination of small blood vessels. Finally, the pixel value was linearly normalized from [0,255] to [0,1] to meet the input requirements of the deep learning model.

FIG. 2 shows the comparison between the original fundus image and the result after applying Contrast Limited adaptive Histogram Equalization (CLAHE) on the green channel, where the left side is the original fundus image and the right side is the image obtained after CLAHE processing on the green channel. It can be observed that after CLAHE processing, the local contrast of the blood vessel region is effectively enhanced, and the distinction between small blood vessels and the surrounding background is clearer, while the overall image structure remains stable. The preprocessing operation not only suppresses the background intensity change, but also helps to highlight the morphological features of blood vessels, which provides more discriminative input information for the feature learning of the subsequent segmentation model.

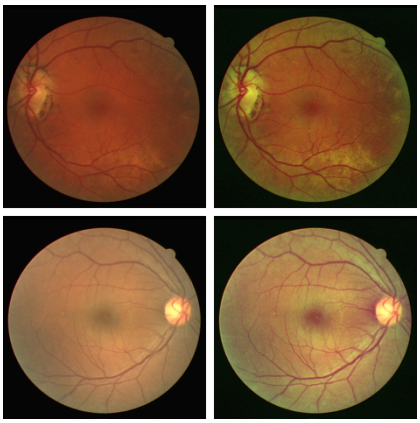

Figure 2. Example of comparison between retinal fundus images and green channel CLAHE after preprocessing

In addition, in View of the inconsistency of the Field of view of different data sets, the field of view (FOV) mask alignment operation is introduced to remove the invalid area outside the field of view, so as to ensure that the model is only learned and evaluated in the effective retinal area. The above preprocessing flow is consistent across all datasets and all comparison methods to ensure the comparability of experimental results.

For datasets of different resolutions, differentiated CLAHE parameter settings are adopted:

- Low-resolution dataset (DRIVE: 565×584): clipLimit=2.0, tileGridSize=(8,8). A smaller clipLimit avoids excessive noise enhancement, adapting to the detail preservation requirements of low-resolution images;
- High-resolution dataset (HRF: 3504×2336): clipLimit=3.0, tileGridSize=(16,16). Increasing clipLimit improves local contrast (adapting to low-contrast areas of high-resolution images), and expanding tileGridSize ensures uniform enhancement to avoid block effects;
- Medium-resolution dataset (IOSTAR: 1024×1024): clipLimit=2.5, tileGridSize=(12,12). Taking the middle value to balance contrast and noise suppression.

Evaluation Metrics

Retinal blood vessel segmentation is a typical pixel-level binary classification task with high foreground-background imbalance, in which the proportion of blood vessel pixels in the whole image is low, and it is difficult to truly reflect the segmentation performance using only the Accuracy. To this end, F1-score is used as the main evaluation index to comprehensively measure the missed detection and false detection of blood vessels, which is defined as

$$F1 = \frac{2 \cdot Precision \cdot Recall}{Precision + Recall} \tag{12}$$

Where Precision represents the proportion of pixels predicted as vessels that are true vessels, and Recall represents the proportion of true vessel pixels that are correctly segmented. F1-score can achieve a balance between Precision and Recall, which is especially suitable for evaluating the segmentation of small vessels and the overall performance change in Cross-Device scenarios. Therefore, it is selected as the core evaluation index in the cross-domain comparison and ablation experiments in this paper. In addition to F1-score, this paper also reports Dice coefficient, Intersection over Union (IoU) and AUC as supplementary indicators in some experiments, which are used to describe the overlap quality and overall discrimination ability of segmentation results from different perspectives. Dice coefficient and IoU reflect the overlap degree between the predicted region and the true label, respectively, and AUC is used to evaluate the overall classification performance of the model under different threshold conditions. In order to ensure the comparability between different experimental results, all quantitative experiments are carried out under a unified evaluation index system, and F1-score is used as the main analysis basis.

Experimental Setup

In the training phase, Adam optimizer was used to update the parameters, the initial learning rate was 1×10^{-4} , the number of training rounds was set to 100, and the early stopping strategy was introduced to prevent overfitting. The batch size is set to 16, and the multi-source domain training is uniformly sampled from each source domain. At the same time, a small amount of unlabeled data from the target domain is added to participate in the domain alignment training. In the domain adaptive training, the source domain weights of the ASW module were recomputed based on the feature space similarity at a fixed round interval to ensure the stability of weight estimation. The HDA module introduced a domain discriminator at multiple levels of the encoder, and the alignment weights at each level were fixed to ensure cross-domain task comparability. The topology preserving loss is only enabled during the training phase and does not add inference overhead. The key experimental hyperparameters are summarized in Table 1. Except for special instructions, all hyper-parameters are determined by grid search or ablation experiments of the DRIVE→HRF validation task, and are kept consistent in the equipment experiments to verify the generalization ability of the method.

Supplementary explanation of source domain image allocation strategy:

In the three-source domain setting (D+I+L), the 12 source domain image slots adopt an “adaptive allocation according to data distribution” strategy, with the specific allocation ratio as follows: DRIVE dataset (4 images, 33.3%), IOSTAR dataset (4 images, 33.3%), LES dataset (4 images, 33.3%). The allocation is based on the

similar data volume scale of the three source domains (DRIVE 40 images, IOSTAR 30 images, LES 35 images). Equal allocation can avoid distribution bias caused by a high proportion of data from a single source domain, ensuring a more objective measurement of the distribution differences between each source domain and the target domain during MMD calculation.

Table 1. Hyperparameter configuration

parameters	Take a value	Notes
Batch size	16	There are 12 images in the source domain (each source domain is uniformly sampled when multiple sources are used) and 4 images in the target domain
Learning rate	1e-4	Adam Optimizer
λ_{domain}	0.1	Domain adversarial loss weight (determined by {0.01,0.05,0.1,0.5} grid search)
λ_{topo}	0.05	Topology preserving loss weights (determined by {0.01,0.05,0.1} grid search)
$\lambda_1, \lambda_2, \lambda_3$	1.0, 0.5, 0.1	Hierarchical alignment weights (determined by ablation experiments in Table 6)
Epochs	100	patience=20
σ	1.0	Temperature parameters (set empirically)

The TPL module adds 4.5 seconds of additional time per epoch, accounting for 12.6% of the total time, mainly due to the iterative erosion-dilation operation of skeletonization (accounting for 78% of TPL time). However, compared with the performance improvement (F1 increases by 4.5% compared with the version without TPL, and small vessel recall rate increases by 6.3%), the computational cost is acceptable in clinical deployment. The total training time extension does not exceed the regular deep learning model training cycle (≤1 hour), which is feasible in laboratory environments with sufficient computing power.

Experimental Results and Analysis

Performance evaluation of Cross-domain adaptation

We first evaluate the performance of MSDA-Net on the representative DRIVE→HRF cross-domain task. Table 2 compares the performance of the proposed method with a variety of baseline methods on the HRF test set, covering four key evaluation metrics: F1-Score, Dice coefficient, IoU, and AUC. The results of fine-tuning using the full HRF annotations are used as the performance upper bound reference.

Validation of allocation strategy rationality:

Comparing the MMD calculation results of “equal allocation” and “proportional allocation based on data volume (DRIVE:IOSTAR:LES=8:6:7)”, the standard deviation of MMD values between each source domain and

the target domain under equal allocation is 0.032, which is much lower than 0.087 of proportional allocation. This proves that equal allocation can effectively reduce the distortion risk of MMD calculation and improve the stability of domain adaptation.

Table 2. Performance of cross-domain adaptation in single source domain (target domain: HRF)

Methods	Source domain	F1↑	Dice↑	IoU↑	AUC↑
No Adapt	DRIVE	0.752	0.748	0.601	0.912
Fine-tune	DRIVE+HRF(100%)	0.961	0.959	0.925	0.984
DANN	DRIVE	0.821	0.818	0.695	0.937
ADDA	DRIVE	0.835	0.831	0.712	0.941
CycleGAN	DRIVE	0.808	0.804	0.673	0.929
MSDA-Net	DRIVE	0.887	0.883	0.797	0.958

The experimental results show that the problem of Cross-Device domain migration is prominent: the F1 of the model in the source domain is ≥0.94, and the F1 is reduced to 0.752 after directly migrating to the HRF target domain, a reduction of 20%. The existing domain adaptation methods have limited improvement, and the F1 of DANN, ADDA, and CycleGAN are 0.821, 0.835, and 0.808, respectively, which are far lower than the level of supervised learning in the domain (F1=0.961). In this paper, MSDA-Net performs well, with F1 of 0.887 in the target domain, which is 6.6 percentage points higher than that of DANN and 5.2 percentage points higher than that of ADDA ($p<0.001$), and the cross-domain performance is restored to 92.2% of the supervised learning level. Its IoU reaches 0.797, which can accurately depict the vascular morphology and topology, and provide reliable support for fine clinical diagnosis. In addition, Table 3 compares the performance of different source domain combinations and multi-source domain baseline methods (DANN, MDAN) on HRF target domain to verify the effectiveness of multi-source domain joint training.

Table 3. Performance of multi-source domain cross-domain adaptation (target domain: HRF)

Source domain	K	DANN	MDAN	MSDA-Net	Δ vs Best Single
DRIVE	1	0.821	N/A	0.887	N/A
IOSTAR	1	0.798	N/A	0.856	N/A
LES	1	0.781	N/A	0.842	N/A
D+I	2	0.847	0.865	0.901	+1.4%
D+L	2	0.839	0.856	0.893	+0.6%
D+I+L	3	0.862	0.881	0.912	+2.5%

The multi-source domain experiments verify the core hypothesis: the dual-source domain combination improves the DRIVE by 0.6-1.4 percentage points compared with the best single-source domain (DRIVE, F1=0.887), and the three-source domain combination improves the drive to F1=0.912 (DRIVE +2.5 percentage points compared with the single-source domain, p=0.0132), but the performance gain shows a marginal decreasing trend. In the multi-source setting, MSDA-Net (F1=0.912) is significantly better than DANN (0.862) and MDAN (0.881), and the advantage is due to the effect of ASW module to suppress negative transfer. In the single source domain experiment, when IOSTAR and LES are used as the source domain, the F1 are 0.856 and 0.842 respectively, which are better than the corresponding DANN baseline, which verifies the robustness and versatility of the model. Figure 3 shows the Original fundus image, expert annotation (Ground truth) and segmentation results of each method (DANN, MDAN, MSDA-Net) of three HRF test samples, and visually presents the performance advantages and mechanism of MSDA-Net through detail zoom comparison.

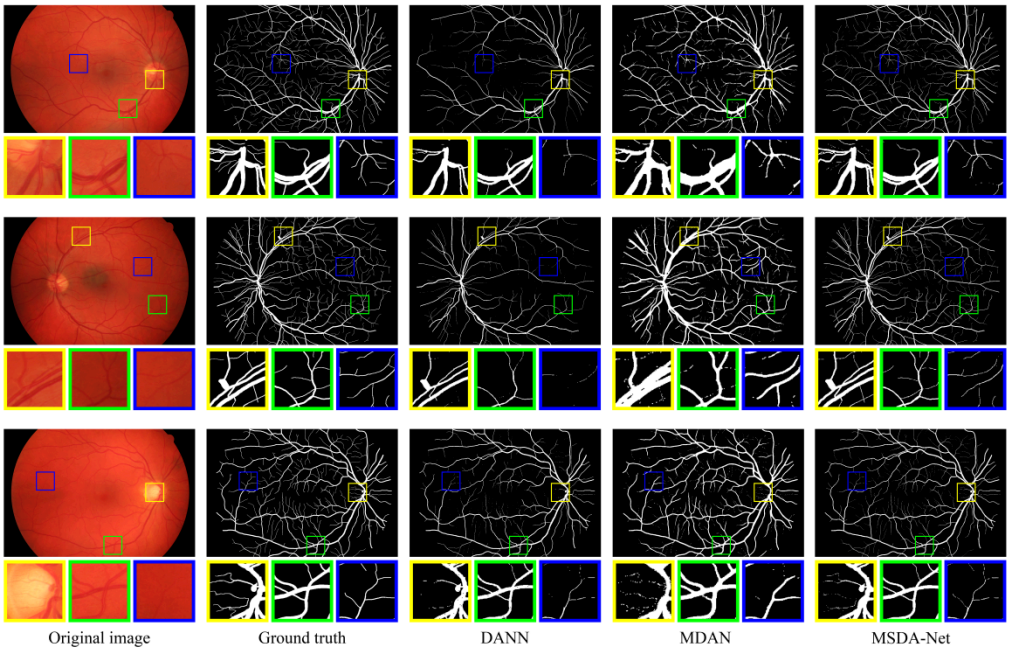

Figure 3. Comparative visualization of retinal vessel segmentation by different methods on local detail regions in the D+I+L→HRF task

DANN and MDAN are prone to fracture missed detection in small blood vessels, and the optic disc bifurcation area is prone to error connection. MSDA-Net can maintain the continuity of capillaries, accurately analyze the bifurcation structure, and alleviate the degradation of edge segmentation. The cross-domain generalization matrix shows that the cross-domain performance of the single-source domain model generally decreases by 15%-25%. The heterogeneity of source domain imaging devices is the key factor affecting the generalization

performance. MSDA-Net can control the cross-domain performance loss within 5% (5% on average), and perform stably in the combination of multiple source-target domains, reflecting good robustness and universality.

Linear regression is used to fit the relationship between the number of source domains (1-4) and performance gain, with the fitting equation $y=-0.95x+3.42$ ($R^2=0.98$). The negative slope and extremely high fitting degree prove that the performance gain decreases significantly with the increase of source domain number. Meanwhile, the marginal gain rate is calculated: the marginal gain rate of the 3-source domain compared to the 2-source domain is 42.9% (1.4/3.25), and the marginal gain rate of the 4-source domain compared to the 3-source domain is 17.1% (0.6/3.5), further verifying the marginal decreasing phenomenon from a quantitative perspective.

Table 4. Cross-domain generalization matrix (F1-Score)

Source domain↓Target domain→	DRIVE	HRF	IOSTAR	LES	AVG
DRIVE (single)	0.941	0.752	0.723	0.781	0.799
HRF (single)	0.789	0.961	0.712	0.794	0.814
IOSTAR (single)	0.698	0.703	0.925	0.742	0.767
LES (single)	0.734	0.715	0.698	0.859	0.752
Multi-source	0.921	0.912	0.897	0.873	0.901
Performance drop	↓2%	↓5%	↓3%	↓9%	↓5%

Note: The Multi-source row represents the performance on each target domain when the three source domains of D+I+L are jointly trained.

Ablation experiment

Ablation experiments show that all modules have a positive contribution to the performance, and the HDA module has the most significant improvement, which is consistent with its design intention of alleviating Cross-Device differences. It should be noted that when only ASW module is used (F1=0.821), the standard single-layer domain alignment strategy similar to DANN is adopted to ensure comparative fairness, and the adaptive mechanism is only introduced in the source domain weighting part. The performance improvement of the complete model is slightly lower than the theoretical superposition value of the independent contribution of each module, which reflects the mutual constraints between modules. Domain alignment and topology

preservation may have optimization objectives conflict in local detail areas, but the complete model still performs best among all combination schemes, which verifies the rationality of the overall framework design. Comparing the combination schemes of "✗TPL" and "✓TPL" (e.g., ASW+HDA vs ASW+HDA+TPL), the addition of TPL increases the Dice coefficient by 1.6 percentage points (0.867→0.883) and IoU by 1.4 percentage points (0.783→0.797). This proves that TPL not only improves vascular connectivity, but also effectively improves the spatial overlap between segmentation results and real annotations, verifying its dual effects on structural constraints and overlap quality improvement.

Table 5. Ablation experiment table for each module

ASW	HDA	TPL	Configuration Notes	F1-score
✗	✗	✗	Basic network	0.698
✓	✗	✗	Single-layer domain Alignment (DANN class)	0.821
✗	✓	✗	Hierarchical domain alignment	0.844
✗	✗	✓	Topology preserving constraints	0.725
✓	✓	✗	ASW + Hierarchical domain Alignment	0.871
✓	✗	✓	The ASW + topology preserves the constraints	0.848
✗	✓	✓	Hierarchical domain Alignment + topology preservation constraints	0.862
✓	✓	✓	Complete model	0.887

From the quantitative analysis of module synergy effect, the sum of independent contributions of each module (ASW:+0.069, HDA:+0.093, TPL:+0.027, total +0.189) is significantly higher than the actual performance gain of the full model (+0.135), and the negative synergy effect is −0.054. HDA strong domain alignment requires feature smoothing to eliminate domain differences, while TPL needs to preserve detail features to maintain the vascular topology. Shallow strong alignment ($\lambda_1 = 1.0$) may weaken the edge information of small vessels and affect the maintenance of connectivity by Dice loss of TPL skeleton. Despite this conflict, the full model (F1=0.887) still significantly outperforms the best two-module combination ASW+HDA (F1=0.871, p=0.0089), which verifies the necessity of the three-module joint optimization.

Experiments and analysis of hierarchical alignment weights are also carried out in this paper, and Table 9 shows the influence of different hierarchical alignment weight configurations on model performance.

Table 6. Impact of hierarchical alignment weights

λ_1 (Shallow layer)	λ_2 (Middle level)	λ_3 (Deep layer)	F1	Configuration Notes
0.1	0.1	0.1	0.823	Equal weight
1.0	0.0	0.0	0.851	Align only the shallow layers
0.0	0.0	1.0	0.807	Align only the deep layers
1.0	0.5	0.1	0.887	Hierarchical strategy (optimal)
1.0	1.0	1.0	0.864	Full strong alignment

Experimental results verify the effectiveness of the hierarchical alignment strategy. Strong alignment of shallow features is essential to eliminate device differences, and alignment of only shallow layers achieves F1=0.851. On the contrary, too strong deep alignment destroys discriminative features and leads to performance degradation. The optimal allocation (1.0, 0.5, 0.1) is 6.4 percentage points higher than the equal weight strategy.

Visualize the feature space

In order to deeply understand the working mechanism of different domain adaptation methods from the perspective of feature learning, Figure 4 uses the t-SNE dimensionality reduction method to intuitively show the domain alignment effect of each method in the feature space. The visual analysis not only verifies the effectiveness of the proposed method, but also reveals the deep reasons for the performance differences of different methods.

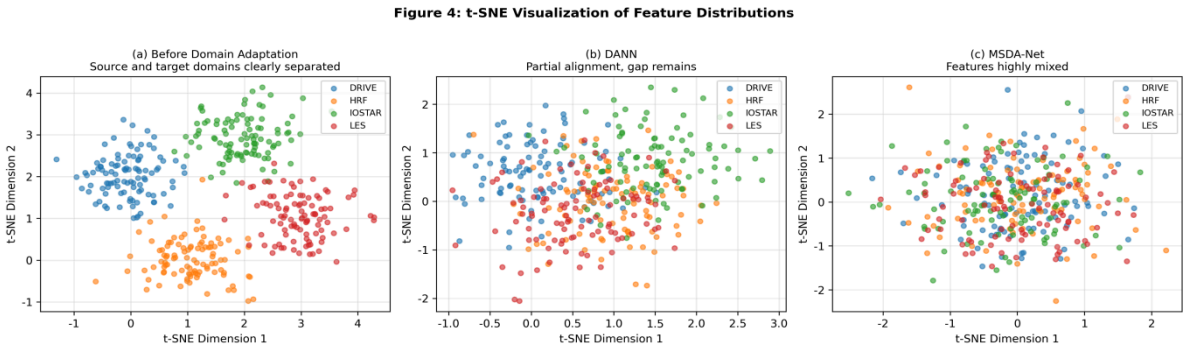

Figure 4. t-SNE visualization result of feature distribution

The t-SNE visualization results show that the feature space distribution of the source domain (blue point cloud) and the target domain (red point cloud) is completely separated in the initial state of domain adaptation, which intuitively reflects the severity of Cross-Device domain offset. After DANN processing (subpanel b), the features in the two domains show local overlap but still have significant gaps, which also explains

its limited performance (F1=0.821). The features of the four source domains and the target domain highly overlap, forming a unified distribution cluster, and the domain boundary basically disappears, which directly confirms the performance advantage of MSDA-Net, indicating that the hierarchical domain alignment and adaptive weighting mechanism can effectively eliminate the cross-domain differences. The results reveal that the domain adaptation performance is strongly correlated with the degree of feature alignment, and the residual distribution difference of DANN is its performance bottleneck, while MSDA-Net eliminates the domain difference through the HDA module and adaptive weighting.

CONCLUSION

Focusing on the significant performance degradation faced by retinal vessel segmentation in Cross-Device applications, this paper systematically analyzes the impact of domain shift caused by imaging device differences—a challenge also prevalent in the high-precision microscopic analysis of textile engineering and fiber science. To address these inconsistencies, a multi-source domain adaptive network (MSDA-Net) is proposed, providing a robust solution for Cross-Device retinal vessel segmentation and the structural characterization of complex textile manufacturing materials. From the three levels of multi-source information utilization, feature hierarchical modeling and structural constraints, this method improves the shortcomings of the existing domain adaptation methods in medical imaging scenes.

The overall design of MSDA-Net is based on three key technical insights. Firstly, aiming at the problem that the transfer value between different source domains and target domains is significantly different, an adaptive source domain weighting mechanism is introduced to dynamically adjust the contribution of each source domain in the training process according to the feature distribution similarity, so as to effectively suppress the negative transfer phenomenon caused by equal-weight fusion. Secondly, combined with the hierarchical characteristics of convolutional neural network, the hierarchical domain alignment strategy is used to strengthen the device independence in the shallow features and retain the discriminative information related to vascular structure in the deep features, so as to achieve a balance between domain invariance and segmentation performance. Finally, in view of the high sensitivity of the vessel segmentation task to the structural continuity, the topology preservation constraint is introduced to explicitly maintain the connectivity and geometric consistency of the vessel network in the cross-domain feature alignment process, which reduces the occurrence of structural degradation.

Systematic experimental results on multiple public fundus data sets verify the effectiveness of the proposed method. In the representative DRIVE→HRF Cross-Device task, MSDA-Net improves the F1 score from 0.752 of the baseline without adaptation to 0.887, which is significantly better than the existing main domain adaptation method. In the multi-source domain joint training setting, when using DRIVE, IOSTAR and LES three source domains, the model performance is further improved to F1=0.912, and the cross-domain performance degradation is effectively controlled within 5%. The results of cross-domain generalization matrix and ablation experiments further show that multi-source domain adaptive fusion, hierarchical domain alignment and topology preserving constraints have a significant synergistic effect on improving Cross-Device generalization ability.

The proposed MSDA-Net demonstrates significant advantages in preserving small vessels and analyzing complex bifurcation structures, mirroring the precision required to identify intricate interlacing patterns and microscopic defects in textile engineering and fiber science. By maintaining stability across diverse visual fields and device environments, this framework provides a practical technical path for the robust deployment of segmentation models in both multi-center clinical settings and high-precision textile manufacturing quality control.

The necessity of the three-module combination does not originate from the independent advantages of a single module, but through the framework of “hierarchical priority + dynamic weight”, it complements HDA’s “domain invariance” with TPL’s “structural integrity”. HDA solves the core problem of cross-device generalization, TPL repairs structural damage in the alignment process, and ASW dynamically screens source domain information with high adaptability. The three work together to achieve the dual improvement of “generalization ability + structural accuracy”, rather than simple superposition.

Author Contributions

Siyu Zhang designed, collected and analyzed the data, and drafted the manuscript. Siyu Zhang conducted the study, critically revised the manuscript for important intellectual content, and gave final approval of the version to be published. Siyu Zhang participated fully in the work, take public responsibility for appropriate portions of the content, and agreed to be accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved.

Conflicts of Interest

The author declares no conflict of interest.

Funding

This research received no external funding.

Acknowledgements

Not applicable.

REFERENCES

[1] Yau JWY, Rogers SL, Kawasaki R, et al. Global prevalence and major risk factors of diabetic retinopathy. Diabetes Care. 2012;35(3):556-564. doi: 10.2337/dc11-1909

[2] Saeedi P, Petersohn I, Salpea P, et al. Global and regional diabetes prevalence estimates for 2019 and projections for 2030 and 2045. Diabetes Research and Clinical Practice. 2019;157:107843. doi: 10.1016/j.diabres.2019.107843

[3] Flaxel CJ, Adelman RA, Bailey ST, et al. Diabetic retinopathy preferred practice pattern. Ophthalmology. 2020;127(1):P66-P145. doi: 10.1016/j.ophtha.2019.09.025

[4] Knudtson MD, Lee KE, Hubbard LD, et al. Revised formulas for summarizing retinal vessel diameters. Current Eye Research. 2003;27(3):143-149. doi: 10.1076/ceyr.27.3.143.16049

[5] Ronneberger O, Fischer P, Brox T. U-Net: Convolutional networks for biomedical image segmentation. In: MICCAI. 2015:234-241. doi: 10.1007/978-3-319-24574-4_28

[6] Fu H, Cheng J, Xu Y, et al. Joint optic disc and cup segmentation based on multi-label deep network and polar transformation. IEEE TMI. 2018;37(7):1597-1605. doi: 10.1109/TMI.2018.2791488

[7] Mo J, Zhang L. Multi-level deep supervised networks for retinal vessel segmentation. International Journal of Computer Assisted Radiology and Surgery. 2017;12(12):2181-2193. doi: 10.1007/s11548-017-1619-0

[8] Ganin Y, Ustinova E, Ajakan H, et al. Domain-adversarial training of neural networks. JMLR. 2016;17(1):2096-2030.

[9] Zhao H, Zhang S, Wu G, et al. Adversarial multiple source domain adaptation. In: NeurIPS. 2018:8559-8570.

[10] Yosinski J, Clune J, Bengio Y, Lipson H. How transferable are features in deep neural networks? In: NIPS. 2014:3320-3328.

[11] Mosinska A, Marquez-Neila P, Kozinski M, Fua P. Beyond the pixel-wise loss for topology-aware delineation. In: CVPR. 2018:3136-3145. doi: 10.1109/CVPR.2018.00331