International Journal of Research in Engineering and Science (IJRES)
ISSN (Online): 2320-9364, ISSN (Print): 2320-9356
www.ijres.org Volume 13 Issue 12 ǁ December 2025 ǁ PP. 131-136

A Multi-Normalization Framework for Single-Source Domain Generalization in Medical Image Segmentation

Ruibing Fu
*[*1]School of Computer Science and Technology, Tongji University, Shanghai, China*
Corresponding Author: Ruibing Fu

Abstract

Medical image segmentation models often suffer significant performance degradation during deployment due to distributional differences between the training source domain and the clinical target domain. Single-Source Domain Generalization (SSDG) aims to train a model robust to any unknown domain using only data from a single source domain, possessing significant clinical value. Existing normalization-based SSDG methods typically rely on a single or fixed normalization layer, making it difficult to adequately adapt to complex and varied domain shifts and failing to achieve an optimal balance between feature style stripping and semantic content preservation. To address this, this paper proposes a novel multi-normalization framework. The core contributions of this framework are twofold: first, it designs a multi-normalization layer structure that integrates diverse normalization operations, which can be dynamically combined during training to learn more domain-invariant deep feature representations; second, it introduces a test-time adaptive mechanism that can select the most suitable normalization computation path in real time based on the features of the input samples, thereby achieving immediate adaptation to unknown target domains. Comprehensive experiments on multiple publicly available cross-domain medical image segmentation datasets demonstrate that our proposed method significantly outperforms existing mainstream SSDG methods, validating the effectiveness of the proposed framework in improving model generalization performance and deployment robustness.

Keywords: *Multi-Normalization Layers, Single-Source Domain Generalization, Medical Image Segmentation.*

Date of Submission: 07-12-2025 Date of acceptance: 19-12-2025

I. INTRODUCTION

In the field of medical image segmentation, while deep neural network models have achieved remarkable results, their performance often heavily relies on the assumption that training and testing data follow the same distribution. However, in actual clinical applications, medical image data from different hospitals, scanning devices, imaging protocols, or patient groups typically exhibit significant inter-domain differences, known as "domain shift." This mismatch in data distribution leads to a sharp decline in the model's generalization ability in unknown target domains, severely limiting the model's clinical deployment and application prospects.

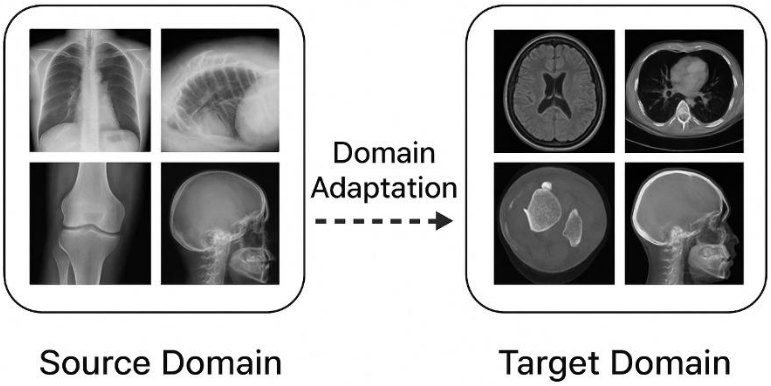

Figure1: Cross-domain medical image segmentation diagram

To address this challenge, research aimed at improving the cross-domain performance of models has evolved along two main technical routes: domain adaptation[1,2] and domain generalization[3,4]. Domain adaptation (DA) primarily focuses on adapting the knowledge of the source domain model to a specific target domain using partially or fully unlabeled target domain data during the training phase. Unsupervised domain adaptation (UDA) is a hot research topic in medical image analysis, assuming that unlabeled target domain samples are available during training. Mainstream methods align the distributions of the two domains by minimizing the feature distribution differences between the source and target domains or by using generative models to perform image or feature-level domain transformations. Although UDA has achieved significant results, its reliance on the assumption of target domain data availability means its generalization ability remains limited when dealing with unknown domains that are completely unseen during training.

Another approach is Domain Generalization (DG), which aims for a more thorough solution—training a model using only one or more source domain data so that it can directly generalize to any unknown target domain. Multi-Source Domain Generalization (MSDG) methods utilize labeled source domain data from multiple disparate distributions, providing a natural diversity foundation for the model to learn domain-invariant features. However, collecting and labeling multiple high-quality, diverse medical image domains is extremely costly and impractical in many real-world clinical scenarios.

In contrast, Single-Source Domain Generalization (SSDG) has become a more challenging yet practical research direction due to its more stringent and realistic setting of limited data (training using only a single source domain). Recently, feature normalization-based methods have shown great potential in SSDG. The core idea of these methods is to implicitly remove domain-related style information from features by normalizing the statistics of intermediate feature maps in the network, thereby encouraging the model to learn more domain-invariant semantic representations. However, current mainstream segmentation networks typically employ a single, fixed normalization layer (such as batch normalization). This "single normalization" paradigm has inherent limitations: first, fixed normalization statistics are difficult to adequately adapt to the complex and varied domain shift patterns in medical images; second, a single operation cannot achieve an optimal balance between feature style removal and discriminative semantic preservation, potentially leading to over-normalization or under-normalization, thus impairing the model's generalization performance.

In this paper, we propose a multi-normalization framework, aiming to promote the development of SSDG in medical image segmentation from the following two aspects:

1. **A novel multi-normalization layer structure:** We designed a modular normalization unit that integrates a set of normalization operations with distinct characteristics. This unit adaptively selects and combines the most suitable normalization strategies for feature maps at different levels and locations in the network. This structure allows the network to simulate diverse feature statistics in a single forward propagation, thereby more thoroughly stripping away domain-specific styles while preserving discriminative semantic structure.

2. **Test-Time Adaptive Normalization Path Selection:** Unlike the traditional paradigm of fixing network behavior after training, we propose a lightweight test-time adaptive mechanism. This mechanism can select the optimal normalization computation path for each test sample in real time based on the shallow features of the input image. This not only achieves immediate adaptation to features of unknown target domains, but also enables the model to make the best trade-off between "style invariance" and "content discriminative" according to the characteristics of different inputs, greatly enhancing the model's deployment robustness in real-world complex scenarios.

II. Method

In this section, I will introduce our proposed multi-normalization framework from the perspectives of training and testing.

2.1 Training Method.

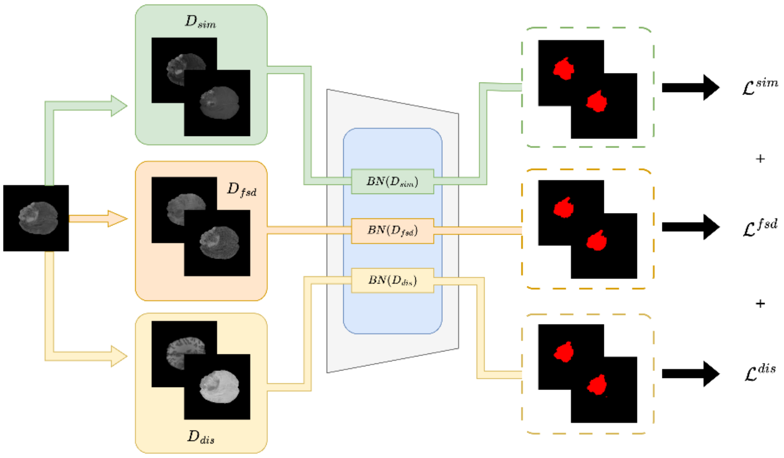

Figure 2: Training flowchart

Figure 2 illustrates the training process of the multi-normalization framework, which can be divided into two parts: image style enhancement and multi-normalization layer training. To explicitly simulate complex domain shifts during training, we need to transform the original source domain data into multiple enhancement domains with distinct styles, and then perform multi-path collaborative training based on these. Specifically, for a batch of original source domain images I_{src}, we generate three enhancement domains using image enhancement methods: the source-similar style enhancement domain D^{sim}, the source-dissimilar style enhancement domain D^{dis}, and the fused style enhancement domain D^{fsd}. The source-similar style enhancement domain primarily simulates weak domain shifts caused by slight changes in imaging parameters, enabling the network to learn stable feature extraction capabilities for small style changes during training. The source-dissimilar style enhancement domain aims to generate enhancement domains that are significantly different from the original image by applying strong brightness, contrast, and local style perturbations to local regions of the image. Large domain shifts are a major challenge in complex domain generalization tasks. The introduction of the source-dissimilar style enhancement domain allows the model to encounter extreme but realistically observable domain shifts during training, thereby improving its adaptability to strong distributional differences. The fusion style enhancement domain combines the two, forming a "hybrid domain" that lies between small and large offsets. In complex domain generalization scenarios, real-world data often exhibits hybrid and multi-scale offsets. By combining style features with different offset intensities, the fused image provides the model with a cross-domain "intermediate sample," helping to mitigate the angle differences between multi-domain distributions and improve the consistency of the shared feature space learned by the model.

To explicitly enable the model to learn and adapt to various data distributions, we construct a multi-path training framework. The core of this framework is to equip each augmentation domain with a dedicated batch normalization (BN) layer. These dedicated BN layers collectively constitute the multi-normalization layer module in our method. Specifically, for an input feature f, its augmentation domain d determines which normalization path it is processed by. The operation of the multi-normalization layer can be formally represented as:

$$MN(f;d) = \gamma_d \frac{f-\mu_d}{\sqrt{{\sigma_d}^2+\epsilon}} + \beta_d, \tag{1}$$

where $d \in \{sim, dis, fsd\}$, γ_d and β_d are affine parameters specific to domain d, (μ_d, σ_d^2) are the mean and variance of the input features of domain d, and ϵ is a small constant for numerical stability. In each training iteration, a batch of original images I_{src} is fed in parallel into three independent processing paths, corresponding to three augmentation domains. The dedicated BN layer for each path independently estimates and updates the statistics of its corresponding domain. All three paths share the weights of the backbone network, but each has its own independent dedicated BN layer. Through this collaborative training mechanism, the model is forced to learn a set of shared feature representations that can be applied to different style variations. At the same time, each dedicated multi-normalization layer accurately captures the distribution information of a specific augmentation domain. These distribution statistics which are solidified during training, provide key discriminative criteria for intelligent path selection during the testing phase.

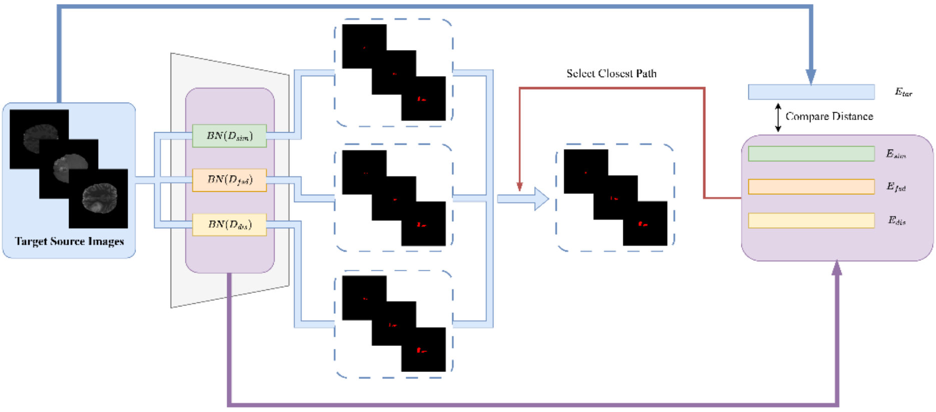

Figure 3: Testing flowchart

2.2 Testing Method.

During the testing phase, the model will face images from an unknown target domain D_t, thus requiring an adaptive inference mechanism that can automatically select the most suitable normalization path based on the style of the input image. To this end, this study constructs an inference process dependent on statistical similarity (as shown in Figure 3) based on the multi-normalization layer structure obtained during the training phase, in order to reduce the potential performance loss caused by domain shift.

For a target domain image I_t, the model first performs forward propagation on three parallel paths, corresponding to the three augmentation domains D^{sim}, D^{dis}, and D^{fsd}, respectively, and normalizes the features using the normalized statistics obtained during training. Specifically, the path $d \in \{sim, dis, fsd\}$ provides a fixed set of statistics for all normalization layers in the network after training, which can be represented as a statistical embedding:

$$\mathcal{E}_d = \left\{ \left(\mu_d^{(l)}, \left(\sigma_d^{(l)}\right)^2\right) \middle| \, l = 1, \dots, L\right\}, \tag{2}$$

where l represents the l-th normalization layer in the network, and L is the total number of normalization layers. Each path generates an independent segmentation prediction result M_d based on this set of statistics. To select the path that best suits the target domain style from the three candidate paths, we construct a target domain statistics embedding from the feature representation of the input image during the inference process. Once the instance statistic embedding of the target domain is determined, the similarity of the target domain sample I_t to the three augmentation domains can be measured by calculating the distance between $\mathcal{E}_t$ and $\mathcal{E}_d$. We use a symmetric distance function that satisfies the triangle inequality and choose Euclidean distance as the method for calculating the distance. Therefore, the embedding distance of the l-th layer can be written as:

$$\mathcal{W}\left(\mathcal{E}_t^l, \mathcal{E}_d^l\right) = \left\|\mu_t^{(l)} - \mu_d^{(l)}\right\|_2^2 + \left\|\left(\sigma_t^{(l)}\right)^2 - \left(\sigma_d^{(l)}\right)^2\right\|_2^2. \tag{3}$$

Once the distance to each augmentation domain is calculated, the nearest augmentation domain statistic embedding and affine parameters γ_d and β_d can be selected to normalize the input features f_t of the target domain:

$$d^* = \arg\min_d D(t, d). \tag{4}$$

Finally, the predicted mask M_{d^*} obtained from the path d^* is used as the final segmentation result of the target domain image. By embedding the mean and variance across layers in the entire network into a statistical measure, this method can comprehensively characterize the deep feature structure of the image style distribution, making path selection more stable and detailed, thereby significantly improving the model's adaptability in complex domain generalization scenarios.

III. EXPERIMENTS

To verify the generalization ability of the proposed multimodal normalization framework in complex cross-domain scenarios, we selected the multimodal brain tumor segmentation dataset Brats as the evaluation dataset. This dataset contains multimodal MRI data of the brain acquired from multiple centers and multiple scanners, and the inherent differences in equipment and protocols naturally constitute significant domain shifts. The dataset provides four modalities of MRI sequences for each subject: T1, T2, T1CE, and FLAIR. These sequences reflect the characteristics of brain tissue from different physiological and physical perspectives, providing complementary information for comprehensive segmentation of brain tumors and their subregions. In terms of image preprocessing, we first normalize the intensity values of each 2D slice for each modality, scaling its pixel value range to the [-1, 1] interval. Subsequently, all 2D slices are uniformly scaled to a resolution of 256 × 256 pixels to meet the fixed size requirements of the model input. During the model training phase, we apply a series of random data augmentations to the source domain data, including random cropping, random rotation, and random scaling.

In this study, we employ the classic U-Net architecture paired with a ResNet-50 encoder as the segmentation backbone. To embed our proposed method, we replace all the original batch normalization layers in the network with our designed multi-normalization layer module, which contains three dedicated BN layers corresponding to the D^{sim}, D^{dis}, and D^{fsd} augmentation domains, respectively. The model is trained uniformly for 50 epochs on all datasets to ensure sufficient convergence. The batch size is set to 64, and in multi-GPU training, data is automatically distributed across GPUs for parallel processing. We use the Adam[5] optimizer to update model parameters, with an initial learning rate lr_0 set to 4×10^{-3}. To stabilize the training process and improve model performance, we employ a multinomial learning rate decay strategy. This strategy keeps the learning rate high in the early stages of training to accelerate convergence, and gradually decreases it in the later stages of training to facilitate model fine-tuning and smoothly approach the optimal solution.

To objectively evaluate the performance of our method, we compare it with state-of-the-art domain generalization methods. The methods included in the comparison are: DoFE[6], SAML[7], Dual-Norm[8], DCAC[9], RAM-DSIR[10], and CCSDG[11]. All comparison methods were reproduced using the authors' publicly available code under the above training hyperparameter settings and were compared fairly under the same training-test data split. We use two widely used evaluation metrics: the Dice similarity coefficient and the Hausdorff distance (HD).

Table 1: Comparison of our method with other single-source domain generalization methods on the Brats dataset.

Method	Source Domain: T2 Dice（%）↑/HD（mm）↓			
	T1	T1CE	FLAIR	Average
DoFE	42.15/20.12	45.02/19.34	48.33/21.01	45.17/20.16
SAML	49.77/16.88	46.22/18.90	52.11/15.33	49.37/17.04
Dual-Norm	47.90/18.50	48.80/15.60	51.45/17.80	49.38/17.30
DCAC	62.80/10.12	68.00/9.02	**67.50**/9.88	66.10/9.67
RAM-DSIR	63.10/10.50	67.10/9.20	66.50/8.20	65.57/9.30
CCSDG	64.50/**9.90**	68.20/8.60	66.00/**8.00**	66.23/**8.83**
Ours	**64.99**/10.53	**68.49**/**8.75**	67.01/8.04	**66.83**/9.11

Table 1 shows the performance comparison between our proposed method and other single-source domain generalization methods on the Brats dataset when using T2 as the source domain. Overall, our proposed method achieves the best comprehensive performance, with an average Dice coefficient of 66.83%, slightly higher than the CCSDG method (66.23%), while also performing well in the average Hausdorff distance metric. Analysis of cross-domain generalization ability shows that our proposed method maintains stable performance across the three different target domains. Of particular note is that in the most challenging T1 target domain, our proposed method achieves a Dice coefficient of 64.99%, outperforming all compared methods; in the FLAIR target domain, our proposed method achieves the lowest HD distance, indicating a significant advantage in boundary segmentation accuracy. Compared to recently high-performing methods such as DCAC, RAM-DSIR, and CCSDG, our proposed method achieves the best Dice coefficient in the T1CE target domain while maintaining competitiveness in the HD distance metric.

These results demonstrate that the proposed multi-normalization path framework effectively improves the model's generalization ability across different target domains. Compared to earlier methods which exhibit significant performance degradation in cross-domain tasks, our method demonstrates stronger robustness, comparable to recent state-of-the-art methods. Particularly when handling cross-domain scenarios with significant modal differences, our method achieves more stable feature representations and more accurate segmentation boundaries through a multi-path normalization structure and adaptive selection mechanism, validating the effectiveness of the proposed approach.

REFERENCES

[1]. Zhao Z, Xu K, Li S, et al. Mt-uda: Towards unsupervised cross-modality medical image segmentation with limited source labels[C]//Medical Image Computing and Computer Assisted Intervention–MICCAI 2021: 24th International Conference, Strasbourg, France, September 27–October 1, 2021, Proceedings, Part I 24. Springer Inter-national Publishing, 2021: 293-303.

[2]. Chen C, Dou Q, Chen H, et al. Unsupervised bidirectional cross-modality adaptation via deeply synergistic image and feature alignment for medical image segmentation[J]. IEEE transactions on medical imaging, 2020, 39(7): 2494-2505.

[3]. Bi Q, Yi J, Zheng H, et al. Learning Generalized Medical Image Segmentation from Decoupled Feature Queries[C]//Proceedings of the AAAI Conference on Artificial Intelligence. 2024, 38(2): 810-818.

[4]. Liu Q, Chen C, Qin J, et al. Feddg: Federated domain generalization on medical image segmentation via episodic learning in continuous frequency space[C]// Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2021: 1013-1023.

[5]. Kingma D P. Adam: A method for stochastic optimization[J]. arXiv preprint arXiv:1412.6980, 2014.

[6]. Wang S, Yu L, Li K, et al. Dofe: Domain-oriented feature embedding for generalizable fundus image segmentation on unseen datasets[J]. IEEE Transactions on Medical Imaging, 2020, 39(12): 4237-4248.

[7]. Liu Q, Dou Q, Heng P A. Shape-aware meta-learning for generalizing prostate MRI segmentation to unseen domains[C]//Medical Image Computing and Computer Assisted Intervention–MICCAI 2020: 23rd International Conference, Lima, Peru, October 4–8, 2020, Proceedings, Part II 23. Springer International Publishing, 2020: 475-485.

[8]. Zhou Z, Qi L, Yang X, et al. Generalizable cross-modality medical image segmentation via style augmentation and dual normalization[C]//Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2022: 20856-20865.

[9]. Hu S, Liao Z, Zhang J, et al. Domain and content adaptive convolution based multi-source domain generalization for medical image segmentation[J]. IEEE Transactions on Medical Imaging, 2022, 42(1): 233-244.

[10]. Zhou Z, Qi L, Shi Y. Generalizable medical image segmentation via random amplitude mixup and domain-specific image restoration[C]//European Conference on Computer Vision. Cham: Springer Nature Switzerland, 2022: 420-436.

[11]. Hu S, Liao Z, Xia Y. Devil is in channels: Contrastive single domain generalization for medical image segmentation[C]//International Conference on Medical Image Computing and Computer-Assisted Intervention. Cham: Springer Nature Switzerland, 2023: 14-23.