HDAug: Optimizing Medical Image Segmentation through Hyperparameter-Driven Brightness and Contrast Augmentation for Deep Learning Models

Yun Yang[1], Haijian Shao[1,*], Xing Deng[1] and Yingtao Jiang[2]

[1] School of Computer, Jiangsu University of Science and Technology, Zhenjiang, 212003, China

[2] Department of Electrical and Computer Engineering, University of Nevada, Las Vegas, NV 89154, USA

INFORMATION

Keywords:
Hyperparameter-driven Data Augmentation (HDAug)
overfitting mitigation
cross-modality image segmentation

DOI: 10.23967/j.rimni.2024.10.57908

SCIPEDIA Revista Internacional de Métodos Numéricos para Cálculo y Diseño en Ingeniería

RIMNI Revista Internacional Métodos numéricos para cálculo y diseño en ingeniería

HDAug: Optimizing Medical Image Segmentation through Hyperparameter-Driven Brightness and Contrast Augmentation for Deep Learning Models

Yun Yang[1], **Haijian Shao**[1,*], **Xing Deng**[1] **and Yingtao Jiang**[2]

[1]School of Computer, Jiangsu University of Science and Technology, Zhenjiang, 212003, China

[2]Department of Electrical and Computer Engineering, University of Nevada, Las Vegas, NV 89154, USA

ABSTRACT

Traditional data augmentation methods, which employ static hyperparameters, often lead to model overfitting. To address this limitation, a novel hyperparameter-driven data augmentation approach (HDAug) is introduced in this study. Training images that simulate a plethora of lighting and exposure conditions are synthesized by HDAug through the stochastic sampling of augmentation hyperparameters, within predefined ranges. Additionally, HDAug does not rely on prior knowledge of specific datasets, endowing it with superior generalization capabilities. The Dice coefficient was utilized as the primary evaluation metric. Experimental results demonstrate that HDAug achieves significant performance improvements in two challenging cross-modality medical image segmentation datasets, with average Dice coefficients of 86.77%, 88.08%, and 84.11%, respectively. The superiority of HDAug lies in its ability to substantially enhance model robustness across diverse imaging conditions while circumventing the overfitting issues inherent in conventional methods. Furthermore, HDAug is computationally efficient and is integrated into existing medical image segmentation workflows.

OPEN ACCESS

Received: 30/08/2024

Accepted: 22/10/2024

Published: 07/04/2025

DOI
10.23967/j.rimni.2024.10.57908

Keywords:
Hyperparameter-driven Data Augmentation (HDAug)
overfitting mitigation
cross-modality image segmentation

1 Introduction

In the domain of medical image segmentation [1–3], high-capacity models are prone to overfitting during training, particularly in scenarios where data is scarce [4]. To efficaciously develop deep neural networks (DNNs) with robust generalization capabilities, substantial temporal, fiscal, and human resources are demanded for the acquisition of high-quality reference standards [5]. Furthermore, manually annotated datasets often suffer class imbalance issues [6]. This insufficiency is disproportionately impactful on particular medical image types or lesion categories, thereby impeding the model's learning capacity and overall efficacy [7].

Data augmentation techniques serve as an effective solution to this problem [8,9], enhancing model robustness and precision in organ segmentation through techniques like deformation [10],

*Correspondence: Haijian Shao (jsj_shj@just.edu.cn).

blurring [11], and noise [12]. Notably, certain augmentation techniques may engender information loss or distortion [13], thereby impinging upon the accuracy of the ensuing segmentation outcomes. In this study, data augmentation techniques for medical image segmentation are focused on. While existing data augmentation technologies can alleviate the limitations of image segmentation models to some extent, fixed enhancement strategies are often employed, which fail to fully capture the complexity and diversity of medical images. These limitations lead to suboptimal model performance on data outside the training distribution. Particularly when dealing with specific types of medical images or lesion categories, class imbalance and information loss may not be adequately addressed by current data augmentation methods. Furthermore, the specific characteristics of medical images, such as the complexity of tissue structures and the subtlety of pathological changes, are typically not considered by existing data augmentation methods. Therefore, there is a need to develop more refined and adaptable data augmentation techniques to better simulate the variability and uncertainty of medical images.

The goal of this research is to enhance precision and mitigate the overfitting phenomena in the domain of medical image segmentation by refining data augmentation strategies. This endeavor aims to bolster the clinical viability of the models, thereby facilitating their extensive deployment in real-world medical contexts.

In this study, accordingly, improvements have been innovatively made based on the Saliency-balancing Location-scale Augmentation (SLAug) framework proposed by Su et al. [14]. The diversity of images was increased through Bézier transformations by the original global enhancement method. However, it was found in our application to specific medical image segmentation tasks that further adjusting the brightness and contrast of the enhanced images can better simulate images obtained under different devices and conditions. Therefore, a random enhancement adjustment module was introduced into the global enhancement step, which can dynamically adjust the brightness and contrast levels of the images to more accurately match the image characteristics of the target domain. In summary, contributions are made by this study in the following aspects:

1). To enhance model robustness and circumvent dataset-specific overfitting, a novel image quality enhancement strategy for random data augmentation (HDAug) is introduced, which simulates various lighting and exposure conditions through hyperparameter tuning for medical image segmentation.

2). Innovative refinement of critical hyperparameters fine-tuning, along with meticulous investigation into their segmentation impact and the design of a rigorous hyperparameter adjustment system. This system facilitates precise manipulation of image visual properties, enhancing diversity, realism, and adaptability across varied clinical scenarios, thereby augmenting model generalization capabilities.

3). Within the context of medical image segmentation, an evaluation of the method's performance has been conducted on two tasks that span different domains in the medical imaging field. Notably, its versatility extends to datasets of diverse modalities such as MRI (Magnetic Resonance Imaging) and CT (Computed Tomography) scans.

The remaining structure of this article is as follows:

The pertinent literature on data augmentation in the field of medical imaging is reviewed in the second section of the manuscript. The Methodological Framework for Hyperparameter-driven Data Augmentation is presented in the third section, providing a detailed exposition of the proposed approach. The datasets relevant to the study are discussed in the fourth section of the manuscript, while the fifth section delves into the training configuration and experimental details. The results of comparative experiments are introduced and analyzed in the sixth section of the manuscript. Ablation studies are conducted and visualizations of the experimental outcomes are provided in the seventh

section of the manuscript. Finally, the eighth section of the manuscript is dedicated to discussing and summarizing the experimental findings.

2 Related Work

In supervised learning tasks such as image recognition and classification [15,16], the application of data augmentation techniques has become a key strategy for enhancing model performance [17,18]. The conventional approach involves the use of simple parametric transformation techniques, such as rotation, scaling and flipping, to expand the content of images in the training dataset. The fundus images were enhanced in the study by Kaushik et al. [19] through the use of flipping and filling. Similarly, in the study by Cangöz et al. [20], the impact of five most commonly used data augmentation techniques—random rotation, horizontal flip, vertical flip, random crop and translation—on the need for precise anatomical division in the diagnosis and treatment of diseases was individually assessed. Furthermore, Isensee et al. [21] designed a data augmentation pipeline that includes a series of operations such as rotation, scaling, Gaussian noise injection and Gaussian blurring. Although these traditional data augmentation methods are straightforward and effective, their process design often relies on experience and may not achieve the optimal diversity enhancement effect [22,23].

New horizons have been opened up by the application of Generative Adversarial Networks (GANs) [24] in the field of medical image data augmentation. Chen et al. [25] proposed a novel end-to-end primary-auxiliary dual generative adversarial network (PadGAN) for enhancing macaque brain diffusion magnetic resonance imaging (dMRI) data. Through the collaborative operation of the primary and auxiliary generators, the quality of low b-value image generation was effectively enhanced. A novel framework was designed by Motamed et al. [26] for enhancing chest X-ray images. Golhar et al. [27] proposed a data augmentation model based on Generative Adversarial Network (GAN) inversion to address the issue of limited annotated data in the classification task of polyps during colonoscopy examinations. Moreover, the Data Pair Generative Adversarial Network (DPGAN) [28] utilizes pairs of data to train the generative model, producing new and balanced pairs of images and segmentation masks. However, they face challenges such as instability and mode collapse [29]. Addressing these, diffusion models are proving to be competitive with GANs as an alternative for medical image data augmentation [30]. Pioneering works introduced the Diffusion Probabilistic Model (DPDM) by Ho et al. [31], while Moghadam et al. [32] focused on morphology with their Diffusion Probabilistic Model (DPM), successfully generating histopathological images. Pinaya et al. [33] leveraged latent diffusion models for data augmentation and the synthesis of high-resolution 3D brain imaging images. Furthermore, Shao et al. [34] introduced a novel feature enhancement framework, AugDiff, which integrates a diffusion model (DM) into multiple instance learning (MIL) to efficiently enhance features of whole-slide images. Although diffusion models may confront challenges such as relatively slower sampling speeds and higher computational overheads, comprehensive distribution coverage, ease of training and scalability are excelled in by diffusion models.

In light of the previously mentioned shortcomings, learning-based methods have emerged as a complementary approach, introducing an innovative approach to learning data augmentation transformation strategies directly from the data. Two labeled images were combined by Zhang et al. [35] to generate new synthetic training samples, specifically targeting the segmentation of brain lesions based on convolutional neural networks. The Sinkhorn distance was utilized by Lyu et al. [36] to maximize diversity among different data augmentation strategies. Dabouei et al. [37] optimized the combination of samples through a supervised mixed data augmentation process by calculating salient

regions and performing data augmentation based on these areas. The learning of polyp features by the YOLOv4 model was enhanced by Chou et al. [38] by integrating a module that combines discrete wavelet transform (DWT) to extract polyp texture features, and employs a style-based GAN2 generative adversarial network to augment image data.

3 Methodological Framework for Hyperparameter-Driven Data Augmentation (HDAug)

As depicted in Fig. 1, the HDAug framework is composed of two stages. The Random Data Augmentation component functions by stochastically generating enhancement hyperparameters, specifically gain (α) and bias (β), within pre-defined parameter intervals, thereby instilling variability into the dataset. In contrast, the Location-Scale Augmentation component is designed to meticulously augment specific regions or features within the imagery, highlighting critical information that promotes refined model discernment and acquisition of local nuances.

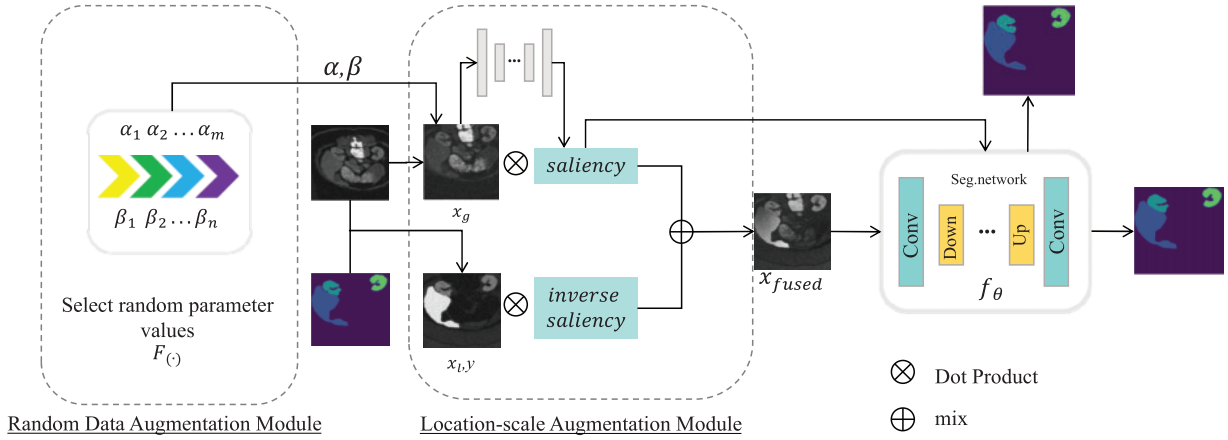


Figure 1: Workflow of the proposed data augmentation approach: Given a segmentation network f_θ, an input image x and two lists of random hyperparameters $\alpha\colon \alpha_1, \alpha_2 \ldots \alpha_m$ $(m \geq 1)$ and $\beta\colon \beta_1, \beta_2 \ldots \beta_n$ $(n \geq 1)$, firstly in we randomly select hyperparameters α and β are randomly selected based on the given hyperparameter list and then linearly transform them with the pixel values of the input data to get the processed new data

3.1 Random Data Augmentation

3.1.1 Setting Enhancement Parameters

Initially, let α and β be two sets comprising multiple potential factors, one representing gains and the other biases. Subsequently, randomly select the i factor from the set α, denoted as α_i, where i is an index uniformly and randomly drawn from the set $\{1, 2, \ldots, length\,(\alpha)\}$. Similarly, randomly select the j factor from the set β, denoted as β_j, where j is an index uniformly and randomly drawn from the set $\{1, 2, \ldots, length\,(\beta)\}$. These coefficients will be used as adjustment parameters in the subsequent data augmentation process.

3.1.2 Linear Transformation Enhancement

For each pixel in the image $\mathcal{X}$, the following steps are performed, utilizing the two hyperparameters previously extracted:

$$F(x) = \alpha_i * G(x) + \beta_j \tag{1}$$

where α is coefficient that is randomly selected to amplify disparities between light and dark areas within the image, and β is a randomly selected brightness adjustment coefficient that elevates image luminance levels. $G(x)$ is the input original image data and $F(x)$ is the obtained enhanced image. The original images and the images after enhancement are depicted in Fig. 2.

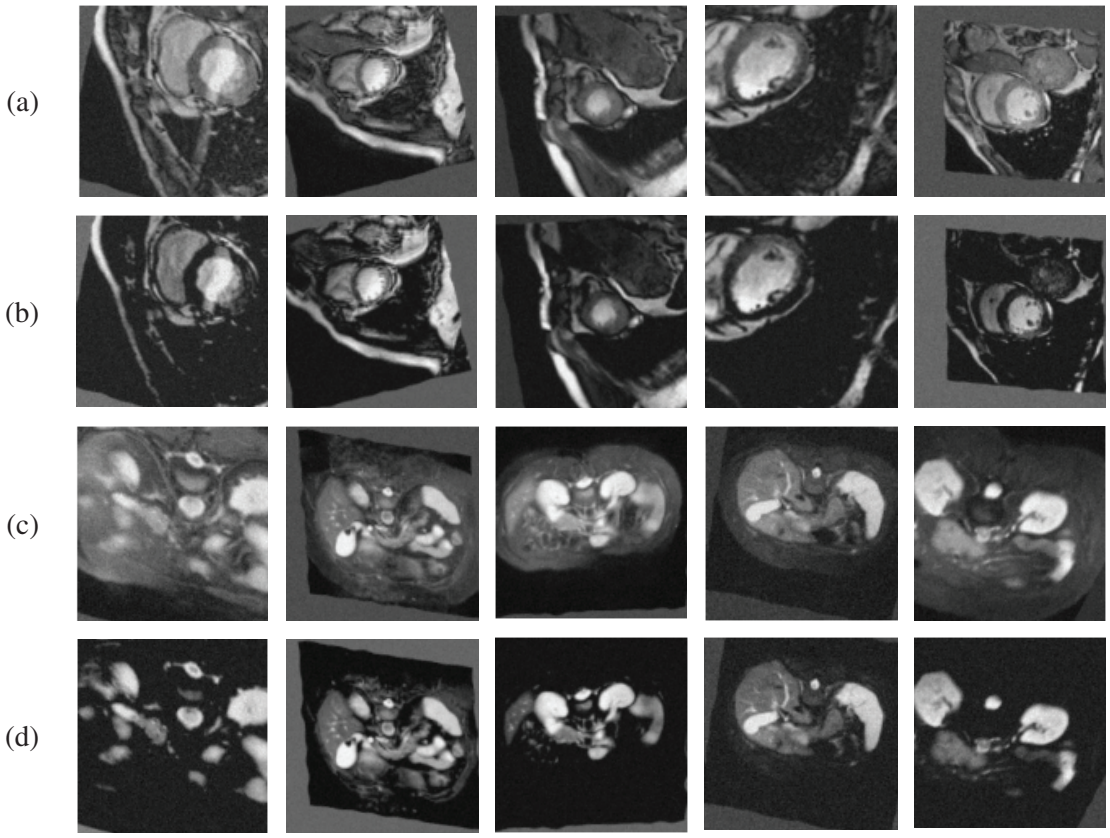


Figure 2: Examples of original and adjusted images: (a) Original images from the cross-sequence cardiac dataset, (b) Generated adjusted samples from the cardiac dataset, (c) Original images from the cross-modality abdominal dataset and (d) Generated adjusted samples from the abdominal dataset

By virtue of the randomized selection of hyperparameters, distinct image variations are introduced by each data augmentation iteration, effectively preventing the model from succumbing to overfitting on any singular dataset. This attribute enables robust generalization across disparate datasets, fostering enhanced adaptability and performance across varied domains. The data generation process is illustrated in Fig. 3.

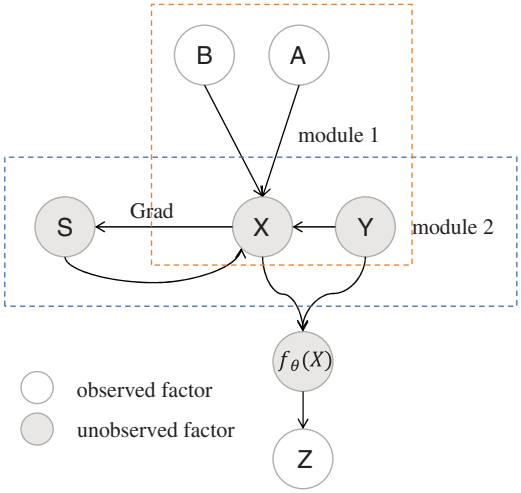


Figure 3: Data generation process of medical images. In this study, image is modeled as being influenced by two independent factors: B, which represents the control factor for image gain; and A, which governs the image bias factor. S is the gradient information derived from X and the ground truth Y, processed through a network $f_\theta(X)$. Our method simulates different acquisition processes for A and B, yielding the predicted outcome Z

3.2 Location-Scale Augmentation Image Fusion

In this study, a cubic Bezier curve was employed for data augmentation, leveraging its smoothness and monotonicity for precise, non-disruptive image enhancement through adjustable control points. The curve is determined by the control points P_0, P_1, P_2 and P_3, where P_0 and P_3 are the start and end points of the Bézier curve. P_1 and P_2 are the control points that govern the shape of the curve.

Global Location-scale Augmentation enhances the data by offsetting the global distribution of the entire image, which can be expressed as:

$$G(x) = (\mu \mathrm{H}_0(x) + \sigma) \tag{2}$$

where $\mathrm{H}_0(x)$ is the nonlinear transformation function, $\mu \sim \mathrm{TN}(1, \omega_1)$ and $\sigma \sim \mathrm{TN}(0, \omega_2)$ are the position-scaling factors, respectively, where ω_1 and ω_2 represent the standard deviations of their respective truncated normal distributions.

The core of global scale transformation lies in the application of specific transformation function H_0 to adjust the pixel intensity values of images, thereby covering a variety of intensity range variations that may occur in the actual application target scenarios. Subsequently, our images undergo further secondary transformations based on stochastically selected hyperparameters. In this way, the model learns how to accurately recognize and segment images with varying intensity distributions.

Due to the fact that the high-frequency components of an image, such as edges and details, typically exhibit greater contrast than the low-frequency components, such as smooth areas, enhancing the high-frequency parts can elevate the local contrast of an image, making the details more pronounced. The core of Local Location-scale Augmentation is to perform localized enhancement on various components of an image. Distinct from global enhancement, Local Location-scale Augmentation utilizes semantic masks to identify regions of interest (ROIs), segments the image into its constituting categories using binary masks, and then processes each category-specific area individually.

Local Location-scale Augmentation, which is designed to model the distributional offsets specific to different categories within an image. For a given image x and its corresponding mask m, the image is decomposed into multiple class regions x^c:

$$x^c = m_c \odot x \tag{3}$$

where m_c is the binary mask for class c, and $\odot$ denotes element-wise multiplication. For each class c region x^c, it undergoes a monotonic nonlinear transformation represented by a Bézier curve H_{pc}. The enhancement process to be represented as:

$$L(x, y) = \left(\sum_{c=1}^{c} \gamma_c H_{pc}(x^c) + \delta_c \right) \tag{4}$$

where x^c is the image region belonging to category c, γ_c and δ_c are the scaling factor and position factor of category c, respectively. $H_{pc}(x^c)$ is the nonlinear transformation function of category C, which is defined using cubic Bezier curves.

This meticulous adjustment yields an enhanced image wherein the visual characteristics are tailored to the specific features of interest. Following these tailored transformations, the fusion of the two sets of images is facilitated by integrating gradient information to enrich the informativeness of the augmented samples.

By computing the gradient map of the image, the model can identify key areas within the image, such as edges and textured regions, preserving or emphasizing these areas during the enhancement process. Initially, the gradient is derived by calculating the gradient of the loss function of a pre-trained model with respect to the input image, denoted as x_g.

$$Grad = \nabla_{x_g'} Loss\left(f_\theta\left(x_g'\right), y'\right) \tag{5}$$

The computed gradient is processed by taking the squared sum across the input channels to determine the magnitude of the gradient through the l_2 norm. The gradient map is then down-sampled to a smaller grid size and interpolated using a quadratic spline kernel to obtain a smoothed gradient map λ. Utilizing this gradient map, sensitive areas in the globally enhanced image (i.e., regions with higher gradients) are identified and preserved.

Subsequently, the gradient is smoothed through adaptive average pooling and transposed convolution with a B-spline kernel, generating a preliminary saliency map. The saliency map is constructed by summing the squares of the gradients and applying a smoothing filter.

$$\lambda = normalize\,(smooth\,(|Grad|)) \tag{6}$$

In the saliency map, areas with higher values (close to 1) are considered bright segments, indicating regions that are more sensitive to model predictions. In the fused image, these areas will primarily retain the characteristics of the global-enhanced image. Conversely, areas in the saliency map with lower values (close to 0) are considered dark segments, representing regions that are more stable in model predictions, and the information in these areas has a lesser impact on the model's final decision-making. In the fused image, these areas will primarily adopt the characteristics of the local-enhanced image to introduce new visual effects.

Then, information from the gradient map is employed to integrate the global Location-scale image with the Local Location-scale Augmentation image. In regions where the gradient is minimal, the portions enhanced by Local Location-scale Augmentation are substituted, thereby introducing

greater variability while maintaining essential information. This fusion generates a composite image that seamlessly integrates the optimal visual attributes and informative characteristics. This process can be represented as follows:

$$M(x,y) = \lambda \cdot F(x) + (1-\lambda) \cdot L(x,y)) \tag{7}$$

where $M(x,y)$ is the pixel value of the hybrid image, λ is the weight determined by the normalized value of the gradient map, Through this approach, we aim for the model to simultaneously learn the ability to accurately segment objects at various scales and to robustly recognize the model under adverse conditions. We present the algorithm of HDAug in Algorithm 1.

Algorithm 1: Process for applying a data augmentation policy

Input: *Training images and annotations* $I(x,y)$, *Enhanced hyperparameter list* α (*gain*), β (*bias*), *network* f_θ, *loss function Loss*, *global location − scale augmentation* $G(x)$, *local location − scale augmentation* $L(x,y)$, *common augmentation* F.

Output: *Augmented training data* $I\left(x'_{fused}, y\right)$

if $(len(\alpha) > 0 \&\& len(\beta) > 0)$ **then**
- $i \leftarrow random.randrange(len(\alpha))$
- $\alpha_i \leftarrow \alpha[i]$
- $j \leftarrow random.randrange(len(\beta))$
- $\beta_j \leftarrow \beta[j]$
- $x_g \leftarrow clip\left(\alpha_i * G(x) + \beta_j, 0, 255\right)$

else:
- $x_g = x$

$x_l \leftarrow L(x,y)$

$x'_g, x'_l, y' \leftarrow F\left(x_g, x_l, y\right)$

$Calculate\ gradient\ Grad = \nabla_{x'_g} Loss\left(f_\theta\left(x'_g\right), y'\right)$

$S \leftarrow normalize(smooth(|Grad|))$

$x'_{fused} \leftarrow s \otimes x'_g + (1-s) \otimes x'_l$

return x'_{fused}

end

4 Datasets

The dataset partitioning and preprocessing were conducted in accordance with the protocols specified in [14]. Each 3D volume is reformatted into 2D slices, and each slice is cropped and resized to a size of 192 × 192 pixels.

Cross-sequence Cardiac Dataset: The dataset is composed of Balanced Steady-State Free Precession (bSSFP) and Late Gadolinium Enhanced (LGE) MRI images, totaling 45 volumes for each modality.

Cross-modality Abdominal Dataset: This dataset, comprising 30 volumes of CT (Computed Tomography) scans and 20 volumes of T2-SPIR MRI (Magnetic Resonance Imaging) images.

5 Training Configuration and Experimental Details

In the experimental configuration, a U-Net architecture that uses EfficientNet-b2 as the backbone network. This network utilizes a series of convolutional layers to extract features, with each convolutional layer employing a 3×3 kernel to capture local features [39].

The training of the segmentation network employed the Adam optimizer with a default learning rate of 3×10^{-4}. The weight decay is set to 3×10^{-5}. The learning rate remains constant for the first 50 epochs and then decays linearly over the next 1950 epochs until it reaches zero.

The loss function was based on cross-entropy, and all models were trained for 2000 epochs with a batch size of 16. The training and testing processes were executed on a Nvidia GeForce GTX 2080Ti GPU with 11 GB of memory.

6 Quantitative and Qualitative Results

To substantiate the superiority of our proposed HDAug, its performance was compared with a series of state-of-the-art methods, including Empirical Risk Minimization (ERM), Cutout [40], Random Sampling of Complement (RSC) [41], MixStyle Neural Networks (Mixstyle) [42], Adversarial Bias (AdvBias) [12], Random Convolution (RandConv) [43], Consistency-Sensitive Domain Generalization (CSDG) [44], Saliency-Balancing Location-Scale Augmentation (SlAug) [14] and Semi-Supervised Learning for Domain Generalization (SSL-DG) [45]. Tables 1 to 4 represent the quantitative performance of different methods on the two segmentation tasks, respectively. The best results are indicated in bold with red font for the second best. The Dice metric being utilized for evaluation.

The efficacy of diverse data augmentation methodologies, when applied to the cardiac dataset, is meticulously delineated in Tables 1 and 2. Specifically, Table 1 illustrates dice scores of 90.15%, 82.13%, 81.57% and 82.60% for the myocardium, left ventricle, and right ventricle, respectively, an average Dice score of 86.77%. Significant enhancements were observed in the segmentation of the myocardium and right ventricle, outperforming the previously superior SLAug method by 0.29% and 0.05%, respectively. Moreover, Table 2 illustrates Dice scores of 81.85%, 92.02%, and 90.36% for the myocardium, left ventricle, and right ventricle, respectively. These scores surpass those of the top-performing SLAug method by 0.36%, 0.04%, and 0.75%, respectively. The overall segmentation efficacy reached an impressive 88.08%, particularly notable for improvements in myocardium and right ventricle segmentation, highlighting the method's adeptness in augmenting cardiac MRI image data and bolstering model precision in delineating critical anatomical structures. This underscores the method's exceptional domain generalization capabilities when applied to varied medical image distributions that exhibit divergent inter-domain characteristics.

Table 1: Segmentation results on the cross-sequence cardiac dataset, where the model was trained on the source domain bSSFP and tested on the target domain LGE

Method	Myocardium	L-ventricle	R-ventricle	Average
ERM	66.98	86.06	74.94	75.99
Cutout [40]	69.06	88.35	79.19	78.87

(Continued)

Table 1 (continued)

Method	Myocardium	L-ventricle	R-ventricle	Average
RSC [41]	69.77	87.06	75.69	77.51
Mixstyle [42]	64.23	85.78	75.61	75.21
AdvBias [12]	70.29	88.23	80.32	79.62
RandConv [43]	75.60	89.88	85.70	83.73
CSDG [44]	77.82	90.35	86.87	85.01
SlAug [14]	80.65	**91.53**	87.90	86.69
SSL-DG [45]	80.11	91.18	86.33	85.87
Proposed	**80.94**	91.48	**87.95**	**86.77**

Transitioning to the abdominal cross-modality experiments, the comparative performance of various data augmentation methods across different organs is presented in Tables 3 and 4. In the cross-modality medical image segmentation task from the MRI to the CT dataset, the proposed method exhibited exemplary performance, achieving Dice values of 90.15%, 82.13%, 81.57%, and 82.60% for the liver, right kidney, left kidney, and spleen, respectively, culminating in an average segmentation efficacy of 84.11%. Notably, the method surpassed the best-performing SLAug method by 0.89% and 1.15% in the segmentation of the liver, and right kidney, respectively, further corroborating the robustness and generalization capabilities of the approach across heterogeneous datasets. While the performance in the segmentation tasks of the left kidney and spleen did not surpass that of the state-of-the-art method, the approach presented in this study ranked third and second in these tasks, respectively, demonstrating a strong competitive edge. Furthermore, in the cross-modality medical image segmentation task from the CT dataset to the MRI dataset, the method achieved an average Dice score of 88.44% for all organs, yet did not surpass previous studies. This observation suggests that the features of our method may significantly augment effectiveness for certain tasks but may not be universally applicable, underscoring the complexities encountered when addressing dataset-specific peculiarities.

Table 2: Segmentation results on the cross-sequence cardiac dataset, where the model was trained on the source domain LGE and tested on the target domain bSSFP

Method	Myocardium	L-ventricle	R-ventricle	Average
ERM	78.59	90.16	87.04	85.26
Cutout [40]	79.14	90.88	87.74	85.92
RSC [41]	78.63	90.21	87.96	85.60
Mixstyle [42]	79.64	91.22	88.16	86.34
AdvBias [12]	79.50	91.20	88.10	86.27
RandConv [43]	80.92	91.98	88.83	87.24
CSDG [44]	80.43	91.37	89.16	86.99
SlAug [14]	81.49	91.92	89.61	87.67

(Continued)

Table 2 (continued)

Method	Myocardium	L-ventricle	R-ventricle	Average
SSL-DG [45]	80.02	90.92	85.05	85.33
Proposed	**81.85**	**92.02**	**90.36**	**88.08**

Table 3: Segmentation results on the Cross-modality Abdominal Dataset, where the model was trained on the source domain MRI and tested on the target domain CT

Method	Liver	R-Kidney	L-Kidney	Spleen	Average
ERM	87.90	40.44	65.17	55.90	62.35
Cutout [40]	86.99	63.66	73.74	57.60	70.50
RSC [41]	88.10	46.60	75.94	53.61	66.07
Mixstyle [42]	86.66	48.26	65.20	55.68	63.95
AdvBias [12]	87.63	52.48	68.28	50.95	64.84
RandConv [43]	84.14	76.81	77.99	67.32	76.56
CSDG [44]	85.62	80.02	80.42	75.56	80.40
SlAug [14]	89.26	80.98	82.05	79.93	83.05
SSL-DG [45]	88.28	80.41	**85.92**	**85.27**	**84.97**
Proposed	**90.15**	**82.13**	81.57	82.60	84.11

Table 4: Segmentation results on the Cross-modality Abdominal Dataset, where the model was trained on the source domain CT and tested on the target domain MRI

Method	Liver	R-Kidney	L-Kidney	Spleen	Average
ERM	78.03	78.11	78.45	74.65	77.31
Cutout [40]	79.80	82.32	82.14	76.24	80.12
RSC [41]	76.40	75.79	76.60	67.56	74.09
Mixstyle [42]	77.63	78.41	78.03	77.12	77.80
AdvBias [12]	78.54	81.70	80.69	79.73	80.17
RandConv [43]	73.63	79.69	85.89	83.43	80.66
CSDG [44]	86.62	87.48	86.88	84.27	86.31
SlAug [14]	90.08	**89.23**	**87.54**	**87.67**	**88.63**
SSL-DG [45]	87.59	85.75	85.69	84.93	85.99
Proposed	**90.83**	88.98	87.02	86.91	88.44

In this study, the model ceases training at the 2000th epoch. The convergence curve of the model's loss function is depicted in Fig. 4. The loss function of the model exhibits a significant decline within the first 800 epochs on both datasets, after which it plateaus. This phenomenon indicates that the model rapidly learns the primary features of the datasets during the initial phase and begins to stabilize after

800 epochs. Although training continues until the 2000th epoch, further reduction in the loss function is minimal, suggesting that the model has converged. This rate of convergence is acceptable as it allows the model to achieve high performance within a reasonable timeframe. However, it is acknowledged that faster convergence might be achievable tuning and optimization of training strategies, which will be a focus of future work.

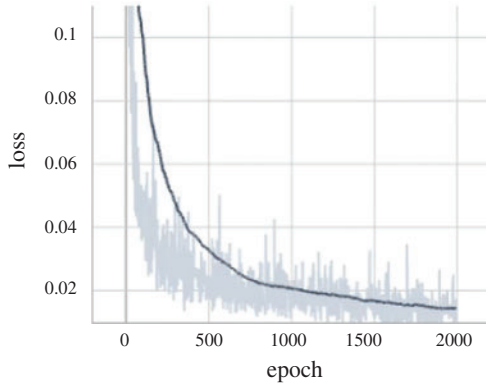


Figure 4: Model convergence curve. The vertical axis represents the loss function (loss), and the horizontal axis represents the epoch

Fig. 5 illustrates the performance of various data augmentation methods on segmentation tasks using a radar chart. The radial position of each data point corresponds to the score or efficacy within a particular evaluative dimension, while the connecting lines delineate the overarching trend. Each spoke on the radar chart symbolizes a distinct data augmentation technique (e.g., ERM, Cutout, etc.).

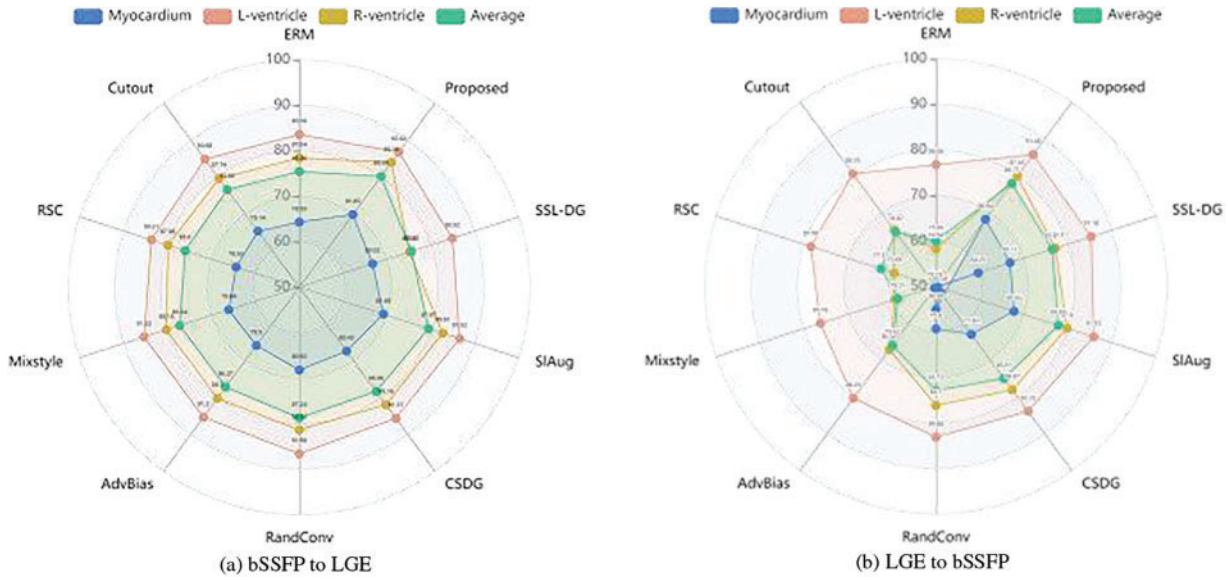
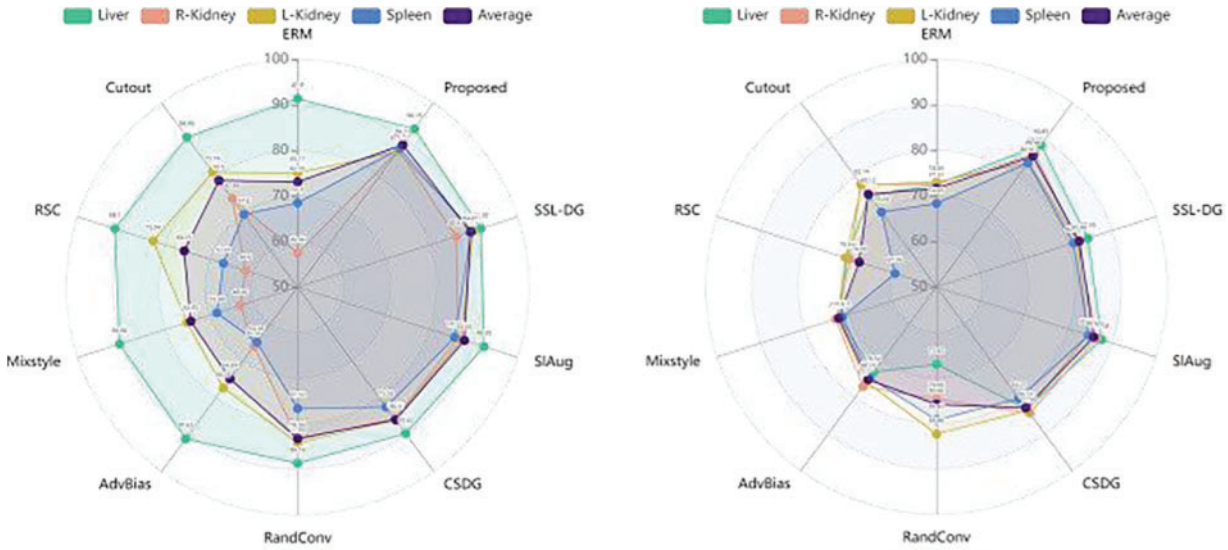


Figure 5: (Continued)



(c) MR to CT (d) CT to MR

Figure 5: The comprehensive performance of different data augmentation methods across various modality dataset segmentation tasks. (a) illustrates the performance comparison of several data augmentation methods on the task of mapping from balanced steady-state free precession (bSSFP) to late gadolinium enhancement (LGE) images. (b) depicts the performance comparison of the same data augmentation methods when applied to the reverse task, i.e., mapping from LGE to bSSFP images. (c) represents the performance comparisons of several data augmentation methods on the magnetic resonance (MR) to computed tomography (CT) segmentation task. (d) represents the performance comparisons of several data augmentation methods on the computed tomography (CT) to magnetic resonance (MR) segmentation task

(a) and (b) present a compilation of the Dice coefficients from preceding methodologies for disparate cardiac tissues (myocardium, left ventricle, right ventricle), as well as the mean Dice coefficient for each cardiac tissue type. Subfigure (a) illustrates the performance comparison of several data augmentation methods on the task of mapping from balanced steady-state free precession (bSSFP) to late gadolinium enhancement (LGE) images. Subfigure (b), on the other hand, depicts the performance comparison of the same data augmentation methods when applied to the reverse task, i.e., mapping from LGE to bSSFP images. Our method consistently achieves higher Dice scores, demonstrating better consistency and precision across all cardiac tissues, including the challenging left atrium.

(c) and (d) show the Dice coefficients for segmenting various abdominal tissues (Liver, R-Kidney, L-Kidney, Spleen) using different methodologies. Subfigures (c) and (d) represent the performance comparisons of several data augmentation methods on the magnetic resonance (MR) to computed tomography (CT) segmentation task and the CT to MR segmentation task, respectively. Comparatively, our method showed the most significant improvement, indicating superior segmentation quality and robustness in abdominal image segmentation.

7 Ablation Studies on the Impact of Hyperparameters α and β

7.1 Hyperparametric Evaluation

To demonstrate our approach's effectiveness, the influence of the hyperparameters α and β on two datasets is scrutinized. These studies assessed the performance metrics of models trained with individual hyperparameters, ensuring equitable comparison with the pre-existing U-Net model. Tables 5 to 8 showcase the ablation analysis of HDAug in the context of the cross-sequence segmentation challenge for the cardiac dataset and the cross-modality abdominal segmentation task.

Table 5: Ablation study on bSSFP to LGE

	α	β	Myocardium	L-ventricle	R-ventricle	Average
SLAug			80.65	91.53	87.90	86.69
Variant 1		Yes	80.92	91.03	86.80	86.25
Variant 2	Yes		79.10	90.74	85.55	85.13
HDAug	Yes	Yes	80.94	91.48	87.95	86.77

Notably, a marked decline was observed in model performance when either α and β was individually omitted, indicating the adverse impact on efficiency and accuracy in the absence of these factors. Crucially, when α and β were combined and allowed to interact within the model, not only a significant improvement in performance but also an enhancement surpassing the original model's capabilities was noted. Ultimately, HDAug demonstrated the best performance, highlighting the indispensable role of both α and β components in achieving advanced domain generalization and their synergistic effect, wherein their combined action complements each other, fortifying the model's performance.

Table 6: Ablation study on LGE to bSSFP

	α	β	Myocardium	L-ventricle	R-ventricle	Average
SLAug			81.49	91.92	89.61	87.67
Variant 1		Yes	75.32	83.28	84.53	81.04
Variant 2	Yes		77.11	86.46	84.65	82.74
HDAug	Yes	Yes	81.85	92.02	90.36	88.08

Table 7: Ablation study on CT to MRI

	α	β	Liver	R-Kidney	L-Kidney	Spleen	Average
SLAug			89.26	80.98	82.05	79.93	83.05
Variant 1		Yes	87.63	79.41	80.73	80.07	81.96
Variant 2	Yes		87.26	80.05	87.25	77.98	81.63
HDAug	Yes	Yes	90.15	82.13	81.57	82.60	84.11

Table 8: Ablation study on MRI to CT

	α	β	Liver	R-idney	L-idney	Spleen	Average
SLAug			90.08	89.23	87.54	87.67	88.63
Variant 1		Yes	90.42	88.40	88.14	85.08	88.01
Variant 2	Yes		90.33	89.23	86.44	85.62	87.90
HDAug	Yes	Yes	90.83	88.98	87.02	86.91	88.44

7.2 Visual Evaluation

To emphasize the efficacy of our proposed data augmentation methodology, Fig. 6 prominently features the comparative findings of our visual analysis, with boxed areas highlighting intricate details within the images. Fig. 7 presents a visual assessment that exclusively focuses on the extreme parameter settings from these two datasets, aligning with the image names in Fig. 6. This selection is made to meticulously examine the distinctions between various parameter configurations. Through this unequivocal visual comparison, it became evident that our enhancement technique substantially improved the overall quality of the images. This improvement was evident not only in the amplification of image details but also in the refinement of contrast and the minimization of noise.

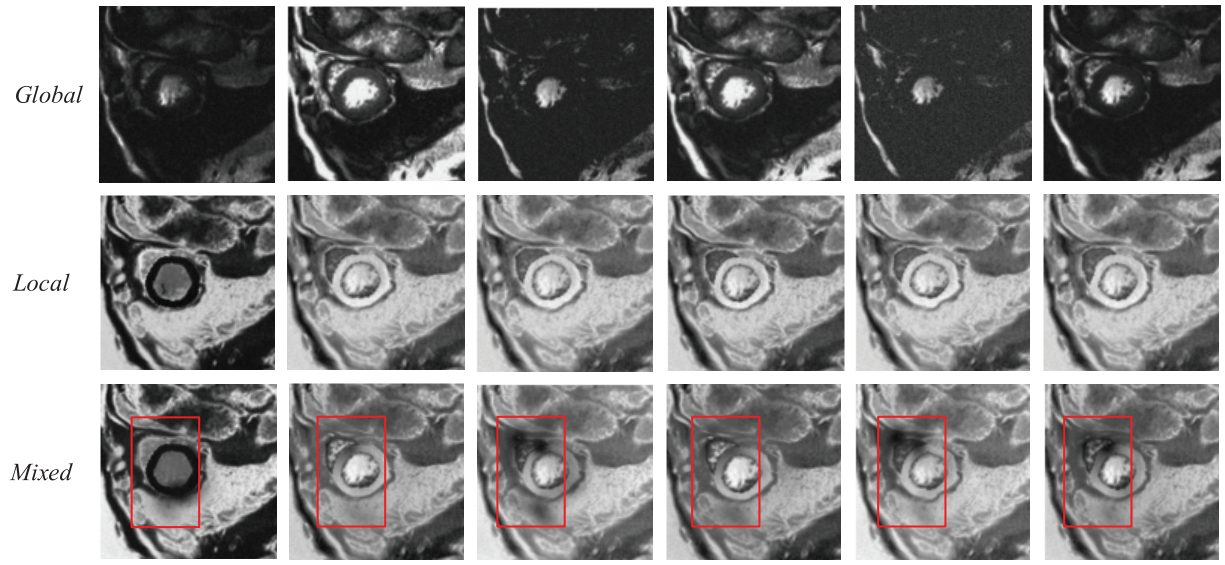


Figure 6: (Continued)

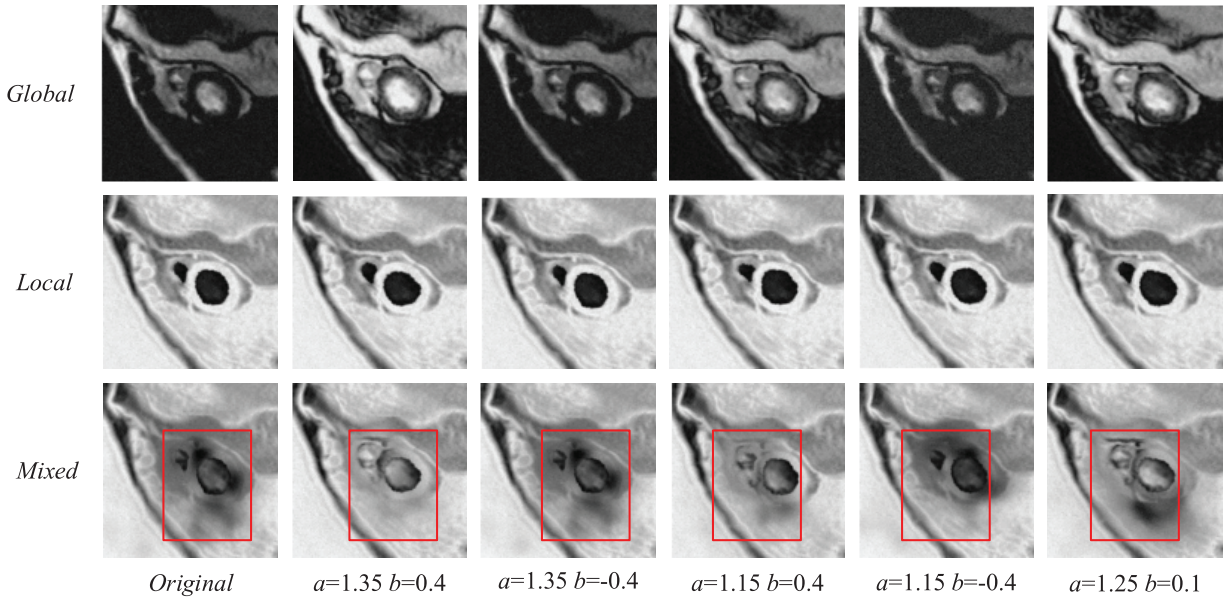


Figure 6: Visual evaluation of cross-sequence cardiac datasets. The initial column showcased the original image, while the succeeding columns illustrated the effects of distinct values of α and β on image quality. The first and fourth rows displayed globally enhanced images; the second and fifth rows depicted locally enhanced images; and the third and sixth rows presented composite images, derived from both global and local enhancements

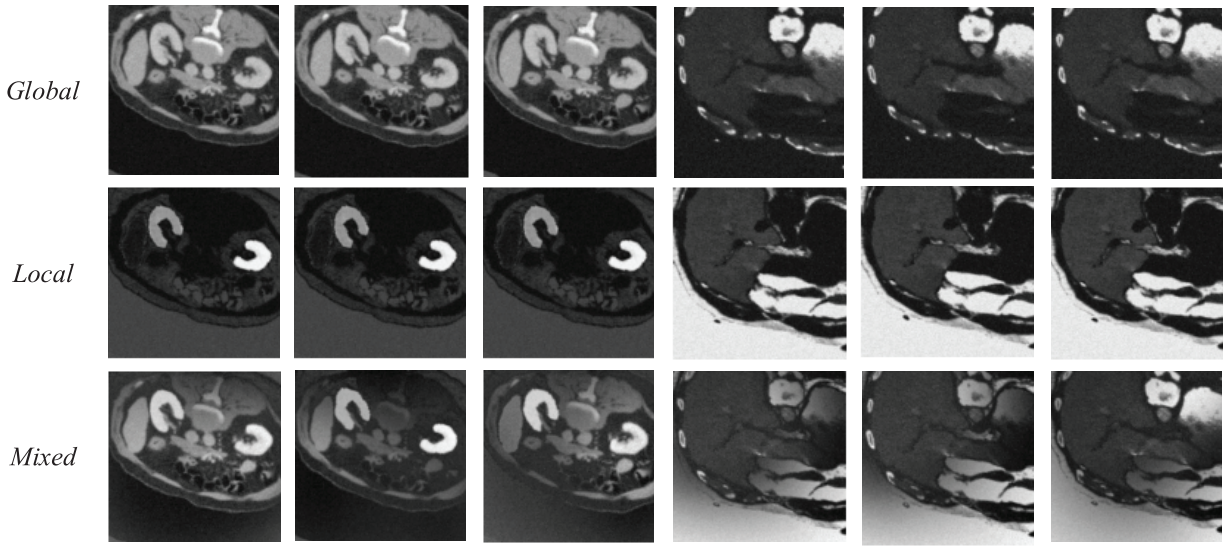


Figure 7: (Continued)

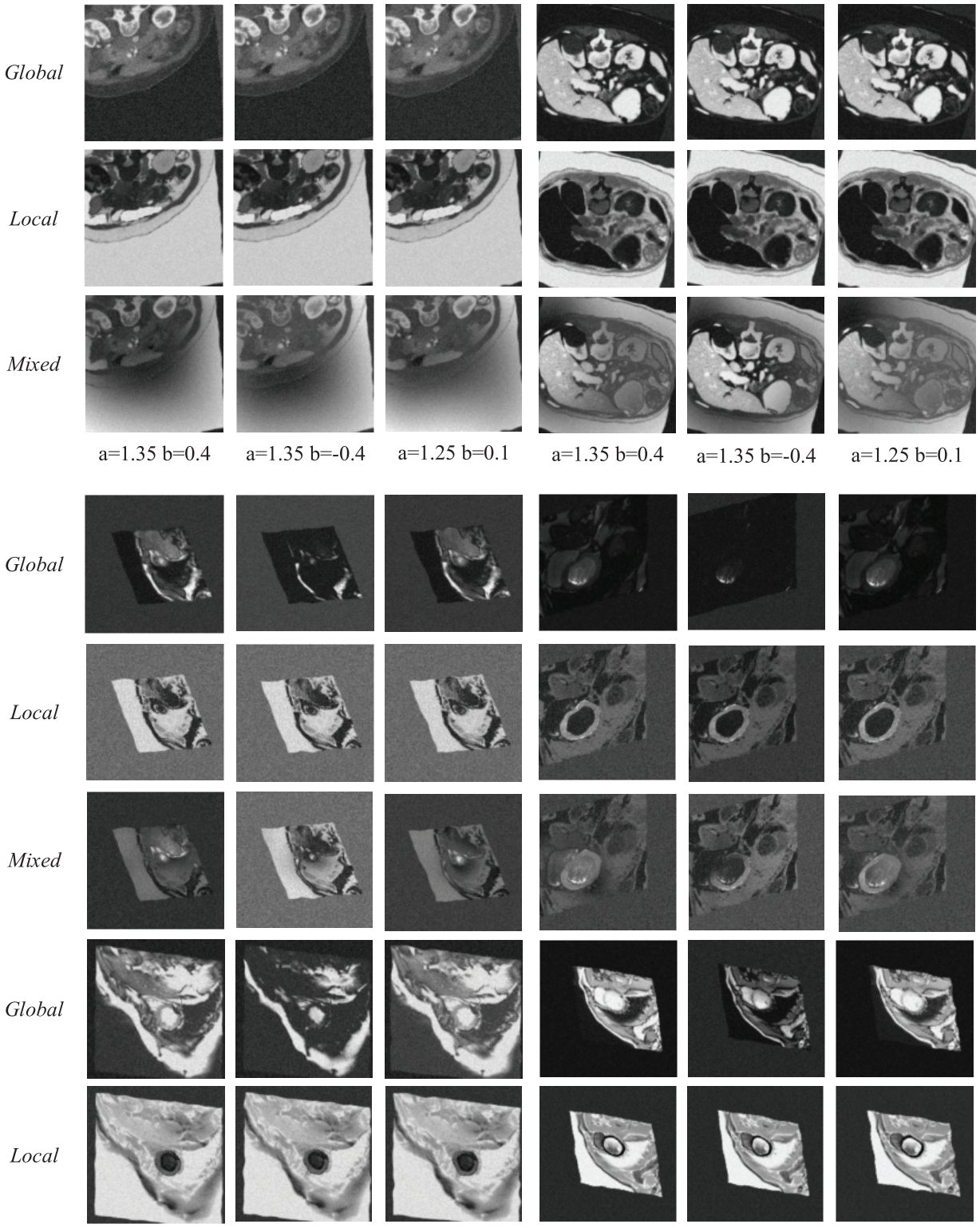


Figure 7: (Continued)

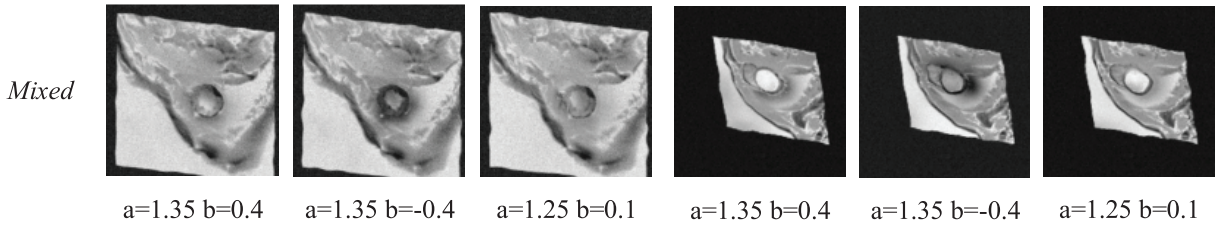


Figure 7: Visual evaluation of cross-modality abdominal datasets. The first and fourth rows displayed globally enhanced images; the second and fifth rows depicted locally enhanced images; and the third and sixth rows presented composite images, derived from both global and local enhancements

8 Conclusion

Our research introduces a novel stochastic data augmentation technique for medical image segmentation, dynamically generating enhancement hyperparameters to overcome the limitations of static settings and reduce overfitting. Compared to the traditional static hyperparameter configuration, our dynamic approach significantly enhances the model's generalization capability on unseen data, effectively mitigating the overfitting phenomenon. The study employed the widely recognized performance metric, Dice, to comprehensively evaluate our model. This research conducted training and experimental validation on two cross-modality datasets to compare and analyze the effectiveness and performance of the proposed HDAug framework. The experimental results demonstrate that our model has achieved significant improvements over the state-of-the-art techniques. Our method not only enhances image quality and segmentation accuracy in both qualitative and quantitative metrics but also validates the effectiveness of the proposed approach in improving segmentation quality through visual segmentation results. By visualizing the convergence curves, the increased stability of the model can be intuitively observed. Furthermore, ablation studies were conducted to verify the contribution of each component within the model. It is believed that these objective experimental results not only substantiate the effectiveness of the model but also provide valuable references for future research.

Acknowledgement: The authors would like to thank the editors and anonymous reviewers for their thoughtful comments and efforts toward improving our manuscript.

Funding Statement: This paper is supported by the National Natural Science Foundation of China (NSFC) (No. 61902158).

Author Contributions: The authors confirm contribution to the paper as follows: study conception and design: Yun Yang, Haijian Shao; data collection: Yun Yang, Xing Deng; analysis and interpretation of results: Haijian Shao, Yingtao Jiang; draft manuscript preparation: Yun Yang, Haijian Shao. All authors reviewed the results and approved the final version of the manuscript.

Availability of Data and Materials: The datasets generated and/or analyzed during the current study are available in the Github repository, https://drive.google.com/file/d/1WlXGt3Nffzu1bn6coqaidHjqWH51smU/view (accessed on 18 October 2024).

Ethics Approval: Not applicable.

Conflicts of Interest: The authors declare no conflicts of interest to report regarding the present study.

References

1. Sun N, Yuan Y, Zhang L, Han Z. Convolutional neural network image segmentation of Alzheimer's disease based on multi-order 3D U-Net. Electro Commun Netw. IOS Press, 2024;182–8. doi:10.1016/j.cogsys.2018.12.015.
2. Kushibar K, Salem M, Valverde S, Rovira À, Salvi J, Oliver A, et al. Transductive transfer learning for domain adaptation in brain magnetic resonance image segmentation. Front Neurosci. 2021;15:608808. doi:10.3389/fnins.2021.608808.
3. Lu Y, Chen Y, Zhao D, Liu B, Lai Z, Chen J. CNN-G: convolutional neural network combined with graph for image segmentation with theoretical analysis. IEEE Trans Cogn Dev Syst. 2020;13(3):631–44. doi:10.1109/TCDS.2020.2998497.
4. Jiani LEl, Filali SEl, Benlahmer EH. Overcome medical image data scarcity by data augmentation techniques: a review. In: 2022 International Conference on Microelectronics (ICM), 2022; Casablanca, Morocco; p. 21–4. doi:10.1109/ICM56065.2022.10005544.
5. Litjens G, Kooi T, Bejnordi BE, Setio AAA, Ciompi F, Ghafoorian M, et al. A survey on deep learning in medical image analysis. Med Image Anal. 2017;42:60–88. doi:10.1109/ICM56065.2022.10005544.
6. Bria A, Marrocco C, Tortorella F. Addressing class imbalance in deep learning for small lesion detection on medical images. Comput Biol Med. 2020;120(132):103735. doi:10.1016/j.compbiomed.2020.103735.
7. Ma L, Dong M, Wu Z, Yu X, Du H, Zhang Y. View planes recognition for fetal echocardiography with unbalanced samples. In: 2023 IEEE 6th International Conference on Computer and Communication Engineering Technology (CCET), 2023; IEEE; p. 127. doi:10.1109/CCET59170.2023.10335131.
8. Rama J, Nalini C, Kumaravel A. Image pre-processing: enhance the performance of medical image classification using various data augmentation technique. ACCENTS Trans Image Process Comput Vis. 2019;5(14):7. doi:10.19101/TIPCV.413001.
9. Jones C, Castro DC, De Sousa Ribeiro F, Oktay O, McCradden M, Glocker B. A causal perspective on dataset bias in machine learning for medical imaging. Nat Mach Intell. 2024;6(2):138–46. doi:10.1038/s42256-024-00797-8.
10. He W, Zhang C, Dai J, Liu L, Wang T, Liu X, et al. A statistical deformation model-based data augmentation method for volumetric medical image segmentation. Med Image Anal. 2024;91(27):102984. doi:10.1016/j.media.2023.102984.
11. Oksuz I, Ruijsink B, Puyol-Antón E, Clough JR, Cruz G, Bustin A, et al. Automatic CNN-based detection of cardiac MR motion artefacts using k-space data augmentation and curriculum learning. Med Image Anal. 2019;55(6):136–47. doi:10.1016/j.media.2019.04.009.
12. Chen C. Realistic adversarial data augmentation for mr image segmentation. In: Medical Image Computing and Computer Assisted Intervention-MICCAI 2020: 23rd International Conference, 2020 Oct 4–8; Lima, Peru: Springer; p. 667–77. doi:10.1007/978-3-030-59710-8.
13. Lennartz J, Schultz T. Segmentation distortion: quantifying segmentation uncertainty under domain shift via the effects of anomalous activations. In: International Conference on Medical Image Computing and Computer-Assisted Intervention, 2023; Springer; p. 316–25. doi:10.1007/978-3-031-43898-1_31.
14. Su Z, Yao K, Yang X, Huang K, Wang Q, Sun J. Rethinking data augmentation for single-source domain eneralization in medical image segmentation. Proc AAAI Conf Artif Intell. 2023;37(2):2366–74. doi:10.1609/aaai.v37i2.25332.
15. Chhabra M, Ravulakollu KK, Kumar M, Sharma A, Nayyar A. Improving automated latent fingerprint detection and segmentation using deep convolutional neural network. Neural Comput Appl. 2023;35(9):6471–97. doi:10.1007/s00521-022-07894-y.

16. Venkatesh C, Chinna Babu J, Kiran A, Nagaraju CH, Kumar M. A hybrid model for lung cancer prediction using patch processing and deep learning on CT images. Multimed Tools Appl. 2024;83(15):43931–52. doi:10.1007/s11042-023-17349-8.
17. Lei S, Deng X, Shao H, Xia W. Improved Res2Net based on multi-scale attention mechanism for breast cancer image classification. In: International Conference on Computer Engineering and Networks, 2022; Springer; p. 597–611. doi:10.1007/978-981-19-6901-0_62.
18. Zhao H, Deng X, Shao H, Jiang Y. COVID-19 diagnostic prediction on chest CT scan images using hybrid quantum-classical convolutional neural network. J Biomol Struct Dyn. 2024;42(7):3737–46. doi:10.1080/07391102.2023.2226215.
19. Kaushik H, Singh D, Kaur M, Alshazly H, Zaguia A, Hamam H. Diabetic retinopathy diagnosis from fundus images using stacked generalization of deep models. IEEE Access. 2021;9:108276–92. doi:10.1109/ACCESS.2021.3101142.
20. Cangöz GB, Güney S. The effects of the traditional data augmentation techniques on long bone fracture detection. Bilge Int J Sci Technol Res. 2023;7(1):63–9. doi:10.30516/bilgesci.1128622.
21. Isensee F, Jaeger PF, Kohl SA, Petersen J, Maier-Hein KH. Maier-Hein, nnU-Net: a self-configuring method for deep learning-based biomedical image segmentation. Nat Methods. 2021;18(2):203–11. doi:10.1038/s41592-020-01008-z.
22. Sun D, Dornaika F, Charafeddine J. LCAMix: local-and-contour aware grid mixing based data augmentation for medical image segmentation. Inf Fusion. 2024;110(1):102484. doi:10.1016/j.inffus.2024.102484.
23. Kumar A, Sharma A, Singh AK, Singh SK, Saxena S. Data augmentation for medical image classification based on gaussian laplacian pyramid blending with a similarity measure. IEEE J Biomed Health Inform. 2023;1–8. doi:10.1109/JBHI.2023.3307216.
24. Yang H, Sun J, Carass A, Zhao C, Lee J, Prince JL, et al. Unsupervised MR-to-CT synthesis using structure-constrained cyclegan. IEEE Trans Med Imaging. 2020;39(12):4249–61. doi:10.1109/TMI.2020.3015379.
25. Chen Y, Zhang L, Xue X, Lu X, Li H, Wang Q. PadGAN: an end-to-end dMRI data augmentation method for macaque brain. Appl Sci. 2024;14(8):3229. doi:10.3390/app14083229.
26. Motamed S, Rogalla P, Khalvati F. Data augmentation using generative adversarial networks (GANs) for GAN-based detection of pneumonia and COVID-19 in chest x-ray images. Inform Med Unlocked. 2021;27:100779. doi:10.48550/arXiv.2006.03622.
27. Golhar MV, Bobrow TL, Ngamruengphong S, Durr NJ. Gan inversion for data augmentation to improve colonoscopy lesion classification. IEEE J Biomed Health Inform. 2024;1–10. doi:10.48550/arXiv.2205.02840.
28. Chai L, Wang Z, Chen J, Zhang G, Alsaadi FE, Alsaadi FE, et al. Synthetic augmentation for semantic segmentation of class imbalanced biomedical images: a data pair generative adversarial network approach. Comput Biol Med. 2022;150(8):105985. doi:10.1016/j.compbiomed.2022.105985.
29. Tronchin L, Sicilia R, Cordelli E, Ramella S, Soda P. Evaluating GANs in medical imaging. In: Deep Generative Models, and Data Augmentation, Labelling, and Imperfections, 2021 Oct 1; Strasbourg, France: Springer; vol. 10, p. 112–21. doi:10.1007/978-3-030-88210-5.
30. Dhariwal P, Nichol A. Diffusion models beat GANs on image synthesis, Advances in neural information processing systems, 2021;34:8780–94. doi:10.48550/arXiv.2105.05233.
31. Ho J, Jain A, Abbeel P. Denoising diffusion probabilistic models. Adv Neural Inf Process Syst. 2020;33:6840–51. doi:10.48550/arXiv.2006.11239.
32. Moghadam PA, Dalen SVan, Martin KC, Lennerz J, Yip S, Farahani H, et al. A morphology focused diffusion probabilistic model for synthesis of histopathology images. In: Proceedings of the IEEE/CVF winter conference on applications of computer vision, 2023; p. 2000–9. doi:10.48550/arXiv.2209.13167.
33. Pinaya WHL, Tudosiu P-D, Dafflon J, Da Costa PF, Fernandez V, Nachev P, et al. Brain imaging generation with latent diffusion models. In: MICCAI Workshop on Deep Generative Models. Springer; 2022; p. 117–26. doi:10.1007/978-3-031-18576-2_12.

34. shao Z, Dai L, Wang Y, Wang H, Zhang Y. AugDiff: diffusion based feature augmentation for multiple instance learning in whole slide image. IEEE Trans Artif Intell. 2024;1–12. doi:10.1109/TAI.2024.3454591.
35. Zhang X, Liu C, Ou N, Zeng X, Zhuo Z, Duan Y, et al. CarveMix: a simple data augmentation method for brain lesion segmentation. NeuroImage. 2023;271:120041. doi:10.1016/j.neuroimage.2023.120041.
36. Lyu J, Zhang Y, Huang Y, Lin L, Cheng P, Tang X. AADG: automatic augmentation for domain generalization on retinal image segmentation. IEEE Trans Med Imaging. 2022;41(12):3699–711. doi:10.1109/TMI.2022.3193146.
37. Dabouei A, Soleymani S, Taherkhani F, Nasrabadi NM. Supermix: supervising the mixing data augmentation. In: 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2021; Nashville, TN, USA; p. 13789–98. doi:10.1109/CVPR46437.2021.01358.
38. Chou Y-C, Chen C-C. Improving deep learning-based polyp detection using feature extraction and data augmentation. Multimed Tools Appl. 2023;82(11):16817–37. doi:10.1007/s11042-022-13995-6.
39. Srinivasu PN, Lakshmi GJ, Narahari SC, Shafi J, Choi J, Ijaz MF. Enhancing medical image classification via federated learning and pre-trained model. Egypt Inform J. 2024;27(47):100530. doi:10.1016/j.eij.2024.100530.
40. Mumuni A, Mumuni F. Data augmentation with automated machine learning: approaches and performance comparison with classical data augmentation methods. 2024. doi:10.48550/arXiv.2403.08352.
41. Huang Z, Wang H, Xing EP, Huang D. Self-challenging improves cross-domain generalization. In: Computer vision-ECCV 2020: 16th European Conference, 2020 Aug 23–28; Glasgow, UK: Springer; p. 124–40. doi:10.48550/arXiv.2007.02454.
42. Zhou K, Yang Y, Qiao Y, Xiang T. Domain generalization with mixstyle. 2021. doi:10.48550/arXiv.2104.02008.
43. Xu Z, Liu D, Yang J, Raffel C, Niethammer M. Robust and generalizable visual representation learning via random convolutions. 2020. doi:10.48550/arXiv.2007.13003.
44. Ouyang C, Chen C, Li S, Li Z, Qin C, Bai W, et al. Causality-inspired single-source domain generalization for medical image segmentation. IEEE Trans Med Imaging. 2022;42(4):1095–106. doi:10.1109/TMI.2022.3224067.
45. Ye Z. SSL-DG: rethinking and fusing semi-supervised learning and domain generalization in medical image segmentation, 2023. doi:10.48550/arXiv.2311.02583.