Hallucinated Style Distillation for Single Domain Generalization in Medical Image Segmentation

Jingjun Yi[2], Qi Bi[2], Hao Zheng[2✉], Haolan Zhan[2], Wei Ji[3], Yawen Huang[2], Shaoxin Li[1], Yuexiang Li[4✉], Yefeng Zheng[2], and Feiyue Huang[1]

[1] Shanghai Digital Medicine Innovation Center, Ruijin Hospital, Shanghai, China
[2] Jarvis Research Center, Tencent Youtu Lab, Shenzhen, China
[3] School of Medicine, Yale University, New Haven, USA
[4] Medical AI ReSearch (MARS) Group, Guangxi Medical University, Nanning, China
howzheng@tencent.com, yuexiang.li@sr.gxmu.edu.cn

Abstract. Single domain generalization (single-DG) for medical image segmentation aims to learn a style-invariant representation, which can be generalized to a variety unseen target domains, with the data from a single source. However, due to the limitation of sample diversity in the single source domain, the robustness of generalized features yielded by existing methods is still unsatisfactory. In this paper, we introduce a novel single-DG framework, namely Hallucinated Style Distillation (HSD), to generate style-invariant features with consistent contents under style variations within an expanded representation space. Specifically, our HSD firstly enhances the style diversity of the single source domain via hallucinating the samples with random channel statistics. Given that out-of-distribution input impacts both the activation value statistics and activated locations, we further propose a decorrelated representation expansion method to indirectly simulate the latter scenario by broadening the representation space. Finally, a hallucinated cross-style distillation paradigm is proposed to distill the style-invariant knowledge between the original and style-hallucinated features, thereby promoting the extraction of structural information. Extensive experiments on two standard domain generalized medical image segmentation datasets show the superior performance of our HSD.

Keywords: Single Domain Generalization · Medical Image Segmentation · Style Invariance.

1 Introduction

Most existing medical image segmentation methods assume that the training and testing images follow the independent and identical distribution. Unfortunately, such an assumption is difficult to fulfill in realistic scenes. In practice, medical images are often collected from different hospitals using different types of scanners, which inevitably lead to the problem of *domain gap*. To this end, lots of domain generalization methods [4,9,10,19,21,22], which adopt medical images from multiple source domains, have been recently proposed to tackle this challenge. However, due to the privacy issue of medical data, the access to multiple sources, *i.e.,* the basic assumption for conventional domain generalization methods, can be infeasible and unpractical.

In this paper, we focus on the more challenging yet practical scenario, *i.e.,* single domain generalization (single-DG) [2,14,17,18], to deal with the domain gap and increase the model generalization, where there is only a single source domain available for model training. Existing single-DG methods either exploited image augmentation to enrich the style diversity [6,21] or learned shape-invariant features as a prior [11]. In contrast, our objective is to introduce random styles that are integrated in the learning pipeline and constrain their similarity, which is able to learn more generalized medical representation despite domain variation. Specifically, we simultaneously expand the diversity of single source domain and extract style-invariant feature representation, which is established based on two key observations:

- Medical images from different domains contain the similar organ/pathological information, *i.e.,* the structural/semantic information.
- Medical images from different domains are usually acquired by different instruments under different imaging conditions. Thus, the style information, such as color, image contrast and illumination, may dramatically vary.

We propose a novel **H**allucinated **S**tyle **D**istillation (HSD) framework for single domain generalized medical image segmentation. Specifically, to enrich the style diversity of single domain, we first randomly sample styles from a large style space $\mathcal{S}$ for hallucination. This process mainly changes the statistics of feature maps. However, in practice, the consequence of out-of-distribution inputs not only includes the changes of activation values, but also the activated locations. To address this, we propose a decorrelated representation expansion (DRE) method which pushes the redundant channels to explore new activation patterns. Then, a novel hallucinated cross-style distillation (HCD) scheme is proposed for the extraction of style-invariant features. It is assumed that domain generalized features include consistent contents after style hallucination in the expanded representation space. To learn such features, HCD incorporates the knowledge distillation paradigm into the domain generalization perspective, which distills the commonly-shared information between the original and style-hallucinated features. Extensive experiments on two domain generalized fundus and prostate benchmarks show its state-of-the-art performance.

2 Methodology

Given a number of unseen target domains $\mathcal{D}_1$, $\cdots$, $\mathcal{D}_K$ and a single source domain $\mathcal{D}_{K+1}$, where domain $\mathcal{D}_k$ $(k = 1, 2, \cdots, K+1)$ has a total number of N_k samples, for single domain generalized medical image segmentation, the objective is to learn a segmentation model $F_\theta : x \rightarrow y$ using only the source domain $\mathcal{D}_{K+1} = \{(x_n^{(K+1)}, y_n^{(K+1)})\}_{n=1}^{N_{K+1}}$, and the trained model F_θ is expected to show good generalization on all the unseen target domains $\mathcal{D}_1 = \{(x_n^{(1)})\}_{n=1}^{N_1}$, $\cdots$,

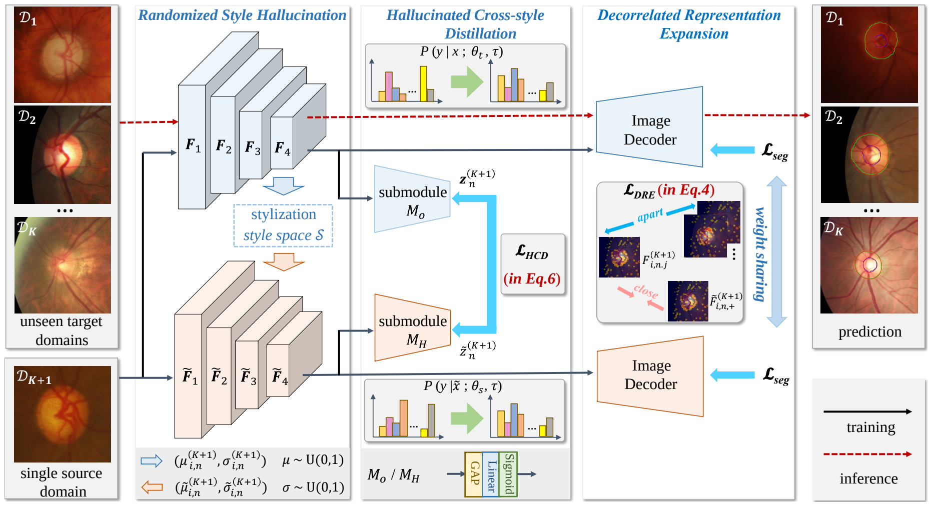

Fig. 1. Framework overview of the proposed Hallucinated Style Distillation (HSD) learning scheme. After feature extraction with an image encoder, it consists of three key components, namely, randomized style hallucination (RSH, in Sec. 2.1), decorrelated representation expansion (DRE, in Sec. 2.2, presented in Eq. 4) and hallucinated cross-style distillation (HCD, in Sec. 2.3, presented in Eq. 6).

$\mathcal{D}_K = \{(x_n^{(K)})\}_{n=1}^{N_K}$. Here, $\{(x_n^{(k)}, y_n^{(k)})\}_{n=1}^{N_k}$ denotes the medical image and its corresponding segmentation label. Our HSD aims to increase the style diversity of $\mathcal{D}_{K+1}$ via style hallucination and decorrelated representation expansion, and learn the style-invariant feature representation from the augmented $\mathcal{D}_{K+1}$.

2.1 Randomized Style Hallucination

The style hallucination technique is introduced to enrich the style diversity under the setting of single-DG. Particularly, we adopt a randomized style hallucination (RSH) scheme. Following the existing definition of styles, the channel-wise mean and standard deviation of an image feature is used to represent the style [5]. Its key idea is to firstly sample styles from a style space $\mathcal{S} \subset \mathbb{R}^{[0,1]\times[0,1]}$ without the aid of any prior knowledge, and then hallucinate intermediate features with the sampled styles.

As shown in Fig. 1, given a MiT-b3 Transformer encoder, from shallow to deep, its architecture can be separated into four feature blocks. For a certain batch of medical images $\{x_n^{(K+1)}\}_{n=1}^{B}$ from the source domain $\mathcal{D}_{K+1}$, where B denotes the batch size, each feature block outputs a feature representation $\mathbf{F}_{i,n}^{(K+1)} \in \mathbb{R}^{B\times C_i\times(W_i\cdot H_i)}$, where $i = 1, 2, 3, 4$, and W_i, H_i and C_i denote the width, height and channel size of feature map, respectively. Then, the channel-wise mean $\boldsymbol{\mu}_{i,n}^{(K+1)} \in \mathbb{R}^{B\times C_i}$ and standard deviation $\boldsymbol{\sigma}_{i,n}^{(K+1)} \in \mathbb{R}^{B\times C_i}$, reflecting the style of a medical image in domain $\mathcal{D}_{K+1}$, can be computed via:

$$\boldsymbol{\mu}_{i,n}^{(K+1)} = \frac{1}{H_i W_i} \mathbf{F}_{i,n}^{(K+1)}, \boldsymbol{\sigma}_{i,n}^{(K+1)} = \sqrt{\frac{1}{H_i W_i} \sum_{h,w \in H_i, W_i} (\mathbf{F}_{i,n}^{(K+1)} - \boldsymbol{\mu}_{i,n}^{(K+1)})^2}. \quad (1)$$

In single-DG, we can only access the style $(\boldsymbol{\mu}_i^{(K+1)}, \boldsymbol{\sigma}_i^{(K+1)})^{\mathrm{T}}$ of the single source domain, which decreases the generalization of learnt features. Due to the normalization layers within the backbone, the mean and standard deviation of almost all channels are between 0 and 1. Therefore, we sample the channel-wise mean and standard deviation randomly from $\mathcal{S} \subset \mathbb{R}^{[0,1]\times[0,1]}$ to hallucinate the style changes for generalized feature learning. The random sampling process can be written as:

$$\widetilde{\boldsymbol{\mu}}_{i,n}^{(K+1)} \sim [0,1], \quad \widetilde{\boldsymbol{\sigma}}_{i,n}^{(K+1)} \sim [0,1]. \quad (2)$$

Then, we scale $(\widetilde{\boldsymbol{\mu}}_{i,n}^{(K+1)}, \widetilde{\boldsymbol{\sigma}}_{i,n}^{(K+1)})^{\mathrm{T}}$ from the style space $\mathcal{S}$ to the feature space. The style-hallucinated counterpart $\widetilde{\mathbf{F}}_i^{(K+1)}$ of the feature representation $\mathbf{F}_i^{(K+1)}$ is generated by AdaIN [5], where the hallucinated style (in Eq. 2) is injected into the original features via:

$$\widetilde{\mathbf{F}}_{i,n}^{(K+1)} = \widetilde{\boldsymbol{\sigma}}_{i,n}^{(K+1)} \boldsymbol{\sigma}_{i,n}^{(K+1)} \cdot \frac{\mathbf{F}_{i,n}^{(K+1)} - \boldsymbol{\mu}_{i,n}^{(K+1)}}{\boldsymbol{\sigma}_{i,n}^{(K+1)}} + \widetilde{\boldsymbol{\mu}}_{i,n}^{(K+1)} \boldsymbol{\mu}_{i,n}^{(K+1)}. \quad (3)$$

2.2 Decorrelated Representation Expansion

The practice of altering statistical values is a prevalent method employed to emulate style variations. However, within the feature level, out-of-distribution inputs not only cause changes in not only the activation values, but also the activated locations. The former scenario can be produced through style hallucination, but the latter is challenging to replicate due to the irregular shape and location of intermediate feature activation. To address this, we propose an indirect approach that compels the network to independently learn diverse features at each encoding level, thereby capturing a wider range of activation patterns and broadening the representation space.

A crucial prerequisite for this method is the inherent presence of redundant features in the intermediate layers of deep neural networks. These features capture similar information and activate at identical locations. By minimizing this feature redundancy, the network can be encouraged to learn a variety of representations. This can be accomplished by penalizing channel-wise feature similarity, effectively decorrelating the feature channels. However, it is important to note that while redundant channels activate at the same locations, the distribution of activation values can vary. Consequently, we propose to penalize the cross-channel similarity between original and hallucinated channels.

Particularly, we adopt the dual form of the contrastive learning paradigm, and leverage the exponent form instead of the logarithm to highlight the impact of them during the learning process. Features from all the four blocks of the image encoder are taken into account. Given the medical image $x_n^{(K+1)}$, and its original and style-hallucinated feature maps $\mathbf{F}_{i,n}^{(K+1)} \in \mathbb{R}^{(W_i \cdot H_i) \times C_i}$ and $\widetilde{\mathbf{F}}_{i,n}^{(K+1)} \in \mathbb{R}^{(W_i \cdot H_i) \times C_i}$, respectively, the proposed decorrelated representation expansion loss $\mathcal{L}_{DRE}$ is computed as:

$$\mathcal{L}_{DRE} = \sum_{i=1}^{4} (\frac{1}{C_i} \sum_{j=1}^{C_i} (-\exp(\mathcal{F}_{i,n,j}^{(K+1)\top} \cdot \widetilde{\mathcal{F}}_{i,n,j}^{(K+1)} / \tau') + \sum_{k=1}^{C_i} \exp(\mathcal{F}_{i,n,j}^{(K+1)\top} \cdot \widetilde{\mathcal{F}}_{i,n,k}^{(K+1)} / \tau'))), \tag{4}$$

where τ' is a temperature scaling parameter, and $\mathcal{F}_{i,n,j}$ denotes the feature map from the j^{th} channel in $\mathcal{F}_{i,n}$. Note that all the feature maps have been resized into the form of vector embedding. With this loss function, the redundant channels are enforced to explore new patterns and the representation space is extended.

2.3 Hallucinated Cross-style Distillation

In this paper, we assume that domain generalized features include consistent contents after style hallucination in the expanded representation space. To learn such features, we add a knowledge distillation module at the end of encoder. As shown in Fig. 1, the knowledge distillation module contains two mapping submodules which extract the semantic content from the output of the last encoding layer. Specifically, mapping submodules M_O and M_H project the original feature $\mathbf{F}_{4,n}^{(K+1)}$ and hallucinated feature $\widetilde{\mathbf{F}}_{4,n}^{(K+1)}$ into a low-dimensional latent space, respectively. We denote the projected representations of $\mathbf{F}_{4,n}^{(K+1)}$ and $\widetilde{\mathbf{F}}_{4,n}^{(K+1)}$ as $\mathbf{z}_n^{(K+1)} \in \mathbb{R}^{1 \times M}$ and $\widetilde{\mathbf{z}}_n^{(K+1)} \in \mathbb{R}^{1 \times M}$. Here M refers to the dimension of projected embedding. Then this process can be written as:

$$\mathbf{z}_n^{(K+1)} = h_{M_O}(\mathrm{GAP}(\mathbf{F}_{4,n}^{(K+1)})), \quad \widetilde{\mathbf{z}}_n^{(K+1)} = h_{M_H}(\mathrm{GAP}(\widetilde{\mathbf{F}}_{4,n}^{(K+1)})), \tag{5}$$

where $h(.)$ denote fully connected layers and $\mathrm{GAP}(\cdot)$ is the global average pooling (GAP) operation.

After acquiring the latent embeddings $\mathbf{z}_n^{(K+1)}$ and $\widetilde{\mathbf{z}}_n^{(K+1)}$, we adopt the Kullback-Leibler (KL) divergence (D_{KL}) as the learning objective ($\mathcal{L}_{HCD}$) for domain-invariant feature distillation:

$$\mathcal{L}_{HCD} = \sum_{n=1}^{N_{K+1}} D_{KL}(\mathbf{z}_n^{(K+1)} || \widetilde{\mathbf{z}}_n^{(K+1)}). \tag{6}$$

This objective requires the features to record more information through the structural activation patterns instead of numerical activation values. Structural patterns encompass the semantic content necessary for segmentation, while numerical values are more susceptible to style variations. In the context of medical images, anatomical structure patterns remain relatively stable, whereas the style can be significantly influenced by imaging conditions. Therefore, prioritizing structural information enhances out-of-distribution generalization.

2.4 Optimization & Implementation Details

In the proposed HSD, the segmentation loss $\mathcal{L}_{seg}$ directly follows the default setting of existing methods [19,22], *i.e.,* a combination between the cross-entropy loss and Dice loss. Then, our overall loss $\mathcal{L}$ can be computed as:

$$\mathcal{L} = \mathcal{L}_{seg} + \lambda \cdot (\mathcal{L}_{HCD} + \mathcal{L}_{DRE}), \tag{7}$$

where λ is a hyper-parameter to balance the loss functions, empirically setting to be 0.01. Other hyper-parameter settings directly follow the prior work [22] without any additional modification. Adam optimizer is used for training. On the fundus dataset, the model was trained 400 epochs with an initial learning rate 5×10^{-4}. On the prostate dataset, the model was trained 200 epochs with an initial learning rate 3×10^{-4}.

3 Experiments

3.1 Datasets & Evaluation Metrics

Following prior single-DG medical image segmentation works [6,11], Dice coefficient (denoted as Dice, in percentage) and Hausdorff Distance (denoted as HD, in pixels) are used for evaluation, which measure the errors in whole object and surface, respectively. The mean and standard deviation of five independent runs are reported. The detail of two benchmarks is provided as follows.

DG Fundus Image Segmentation Dataset consists of retinal images from four different domains, namely, REFUGE (train) [13], REFUGE (val) [13], DrishtiGS [16] and RIM-ONE-r3 [3]. We denote them as Domain-A to Domain-D.

DG Prostate Image Segmentation Dataset consists of magnetic resonance imaging (MRI) samples from six domains out of three prostate datasets, namely, NCI-ISBI13 [1], I2CVB [7], and PROMISE12 [8]. For simplicity, we denote them as Domain-A to Domain-F in the following text. All the images are resized to 384×384 pixels in the axial plane, and the intensities are normalized to a distribution with a zero mean and a unit standard deviation.

3.2 Comparison with State-of-the-Art

Experimental Settings. The experiments on the fundus dataset use Domain-A as the singe source domain, and rest three as unseen target domains. The experiments on the prostate dataset use Domain-A as the single source domain, and the rest five as unseen target domains. For results on the fundus dataset, the performance of optic disc and optic cup is averaged in the report [11]. The proposed HSD is compared with 1) existing single domain generalized medical image segmentation methods, namely, JiGen [2], M-ADA [14], TTT [17], TTST [6], BigAug [21], Tent [18], and TASD [11], with a residual U-Net [12] under empirical risk minimization (ERM) as baseline, where the results are directly cited from [11]; 2) two Vision Transformer (ViT) based segmentation models,

Table 1. Performance comparison between the proposed HSD and existing methods on the fundus dataset. The results of all pairs are significantly different at p= 0.01.

Unseen Domain	B	C	D	Avg.	B	C	D	Avg.
	Dice Coefficient (Dice, mean±std) ↑				**Hausdorff Distance (HD, mean±std) ↓**			
CNN based:								
Baseline [12]	83.2±8.2	76.0±14.8	88.8±5.7	82.7	27.4±15.3	36.8±23.8	21.5±10.3	28.6
M-ADA [14]	85.9±9.3	77.4±13.7	90.6±4.8	84.6	22.7±11.2	30.4±17.1	13.8±7.5	22.3
TTT [17]	84.6±5.7	77.3±11.9	89.0±4.1	83.7	22.2±8.6	31.4±19.1	17.4±7.1	23.7
TTST [6]	85.5±9.0	76.5±13.4	91.0±4.7	84.3	22.0±10.2	33.6±19.3	**12.1±8.1**	22.6
Tent [18]	85.1±8.1	77.1±13.0	89.2±4.9	83.8	23.1±11.7	35.2±20.9	17.8±8.5	25.4
BigAug [21]	84.7±7.5	78.0±13.5	90.7±4.6	84.5	27.1±13.2	30.3±21.6	14.8±7.9	24.1
JiGen [2]	84.5±5.8	77.5±12.0	88.5±4.1	83.5	23.4±9.2	34.3±19.4	20.4±10.6	26.0
TASD [11]	87.6±8.0	78.5±12.6	**91.3±4.2**	85.8	19.8±9.5	29.4±18.0	12.3±6.2	20.5
HSD (Ours)	86.7±4.2	80.1±8.0	90.5±3.8	85.8	18.7±4.6	27.8±9.6	12.4±5.2	19.6
ViT based:								
SegFormer [20]	84.7±2.4	86.3±3.9	80.1±3.0	83.7	27.7±5.5	33.8±4.2	38.2±4.8	33.2
FeedFormer [15]	87.6±2.1	86.0±3.7	80.5±1.4	84.7	23.6±7.5	30.2±7.3	25.2±4.5	26.4
HSD (Ours)	**89.9±1.5**	**89.7±1.0**	88.4±1.6	**89.3**	**13.4±1.0**	**22.3±4.2**	18.8±2.1	**18.2**

Table 2. Performance comparison between the proposed HSD and existing methods on the prostate dataset. The results of all pairs are significantly different at p= 0.01.

Unseen Domain	B	C	D	E	F	Avg.	B	C	D	E	F	Avg.
	Dice Coefficient (Dice, mean±std) ↑						**Hausdorff Distance (HD, mean±std) ↓**					
CNN based:												
Baseline [12]	83.8±5.3	73.3±11.1	72.6±7.0	65.5±29.5	78.7±7.8	74.8	40.9±34.9	59.0±29.4	59.5±21.2	61.2±42.6	37.2±20.9	51.6
JiGen [2]	83.2±6.1	70.8±14.7	74.0±7.9	71.5±10.2	80.3±6.2	75.9	29.3±23.2	64.5±23.3	50.4±25.9	50.6±26.0	24.3±10.7	43.8
M-ADA [14]	86.2±4.4	74.7±9.1	80.9±4.9	69.7±12.2	79.5±9.3	78.2	19.1±21.1	46.1±28.1	53.9±19.3	54.2±19.6	31.9±26.7	41.0
TTT [17]	83.5±5.9	73.1±17.5	75.3±7.8	67.5±11.1	81.5±5.9	76.2	26.4±22.1	55.4±22.3	54.8±25.5	53.0±22.2	21.8±19.2	42.3
TTST [6]	86.0±3.7	74.8±10.5	81.0±3.9	74.0±8.4	80.9±9.2	79.3	20.5±20.7	47.5±28.1	41.4±19.7	51.4±26.1	34.5±25.5	39.1
BigAug [21]	84.2±5.0	73.9±14.1	73.3±7.7	74.7±9.7	79.0±6.8	77.0	35.9±26.3	49.1±20.7	53.8±22.0	44.5±18.7	28.9±14.4	42.4
Tent [18]	84.5±4.7	74.2±13.9	76.4±8.1	67.1±10.1	80.1±9.6	76.5	27.2±24.7	50.3±22.7	45.7±23.5	49.6±30.9	29.8±20.1	40.5
TASD [11]	87.1±2.5	76.4±6.1	82.5±5.2	76.0±6.6	83.2±6.7	81.1	19.3±21.3	39.1±17.5	38.7±12.2	43.4±14.2	21.0±17.5	32.3
HSD (Ours)	85.7±3.2	78.9±5.6	84.5±3.0	78.6±5.1	82.3±4.1	82.0	10.1±5.6	22.8±9.2	16.3±6.5	21.6±8.9	16.2±7.0	17.4
ViT based:												
SegFormer [20]	83.2±1.8	69.0±4.1	83.5±2.7	66.5±2.5	86.3±1.8	77.7	4.1±0.3	19.6±4.4	3.8±0.7	15.7±1.1	3.8±1.4	9.4
FeedFormer [15]	87.2±1.3	79.2±3.7	86.6±2.0	76.4±3.5	87.7±1.5	83.4	4.2±0.4	18.6±5.0	4.0±0.8	20.3±5.4	3.9±1.3	10.2
HSD (Ours)	**88.7±2.2**	**85.0±2.5**	**87.7±1.9**	**87.2±1.7**	**90.3±2.1**	**87.8**	**3.3±0.9**	**14.5±4.1**	**3.7±0.9**	**10.8±2.0**	**3.1±1.3**	**7.1**

namely, SegFormer [20] and FeedFormer [15], under ERM, where the results are implemented by us under all default hyper-parameter settings.

Results on Fundus Dataset. Table 1 reports the performance of the proposed HSD and existing methods. When embedded into the same convolutional neural network (CNN) based segmentation model, the proposed HSD achieves a competitive performance against TASD, and significantly outperforms the other methods, yielding an average Dice of 85.8% and HD of 19.6 pixels. When embedded into existing ViT based segmentation model, it achieves the state-of-the-art performance with an average of Dice of 89.3% and HD of 18.2 pixels.

Results on Prostate Dataset. Table 2 reports the performance of the proposed HSD and existing methods. When embedded into existing CNN segmentation model, the proposed HSD outperforms all the compared methods by an average of 0.9% improvement in Dice, respectively. When embedded into the Vision Transformer segmentation model, it achieves the state-of-the-art performance, with an average Dice of 87.8%. Notably, the standard deviation of the proposed HSD is much smaller than existing methods, indicating its robustness.

Table 3. Ablation study on each component in the proposed HSD. The results of all pairs are significantly different at $p = 0.01$.

Component			B	C	D	Avg.	B	C	D	Avg.
RSH	HCD	DRE	**Dice Coefficient (Dice, mean±std) ↑**				**Hausdorff Distance (HD, mean±std) ↓**			
✗	✗	✗	87.6±2.1	86.0±3.7	80.5±1.4	84.7	23.6±7.5	30.2±7.3	25.2±4.5	26.4
✓	✗	✗	88.0±2.3	86.7±1.9	82.6±1.5	85.8	21.2±6.8	28.3±5.9	23.1±2.8	24.2
✓	✓	✗	88.8±1.3	88.1±1.4	86.0±1.7	87.6	16.0±3.2	27.6±3.5	21.9±2.0	21.8
✓	✓	✓	**89.9±1.5**	**89.7±1.0**	**88.4±1.6**	**89.3**	**13.4±1.0**	**22.3±4.2**	**18.8±2.1**	**18.2**

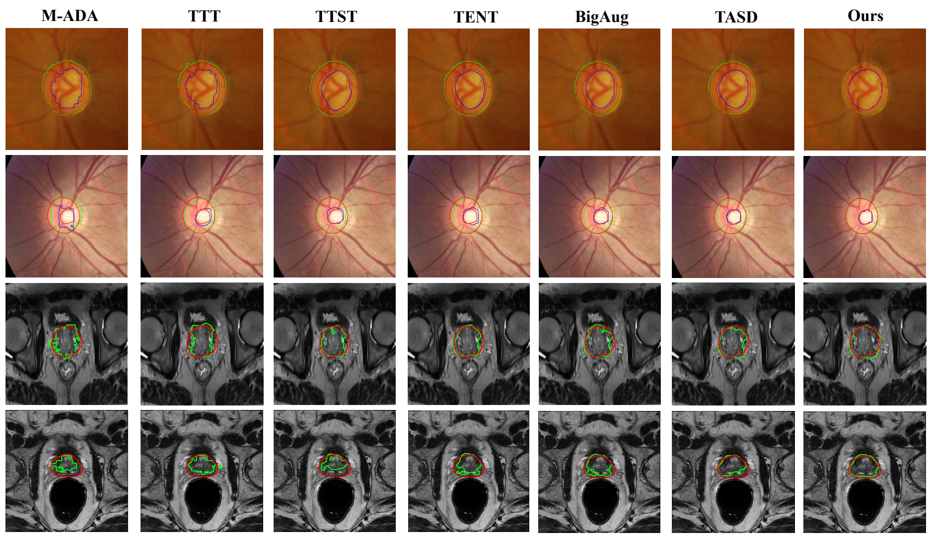

Fig. 2. Exemplar domain generalized segmentation results of the proposed method and the state-of-the-art methods. The first and second rows are results from the fundus benchmark. The third and forth rows are results from the prostate benchmark. Ideally, the green and blue segmentation predictions should coincide with the red ground truth.

3.3 Ablation Studies

Table 3 studies the impact of each component in the proposed HSD, namely, RSH, HCD and DRE. When none of these components are available, the proposed HSD degrades into a FeedFormer baseline. A naive use of RSH leads to an average improvement of 1.1% in Dice and 2.2 pixels in HD. The proposed HCD and DRE both lead to a clear performance gain, by 1.8% and 1.7% in Dice, respectively, while the combination of HCD and DRE can achieve further improvement.

3.4 Visualization of Segmentation Results

Fig. 2 shows visual segmentation results of the proposed HSD and existing state-of-the-art methods on the fundus dataset and prostate dataset. The proposed HSD shows better prediction with smooth boundary on unseen target domains.

4 Conclusion

In this paper, we focused on the most challenging but practical case for domain generalized medical image segmentation, when there is only a single source domain available. We proposed a novel Hallucinated Style Distillation (HSD) learning scheme, which aims to extract features with consistent contents under style variations within an expanded representation space. Extensive experiments on two standard domain generalized medical image segmentation datasets showed its state-of-the-art performance. Notably, the proposed HSD is applicable to both CNN and ViT based segmentation models.

Acknowledgments. This study was funded by the Science and Technology Major Project of Guangxi (AA22096030 and AA22096032), and National Key R&D Program of China under Grant (2020AAA0109500 and 2020AAA0109501).

Disclosure of Interests. The authors have no competing interests to declare that are relevant to the content of this article.

References

1. Bloch, N., et al.: NCI-ISBI 2013 challenge: automated segmentation of prostate structures. The Cancer Imaging Archive **370** (2015)
2. Carlucci, F.M., D'Innocente, A., Bucci, S., Caputo, B., Tommasi, T.: Domain generalization by solving jigsaw puzzles. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 2229–2238 (2019)
3. Fumero, F., et al.: RIM-ONE: An open retinal image database for optic nerve evaluation. In: 24th International Symposium on Computer-Based Medical Systems. pp. 1–6 (2011)
4. Hu, S., Liao, Z., Zhang, J., Xia, Y.: Domain and content adaptive convolution based multi-source domain generalization for medical image segmentation. IEEE Transactions on Medical Imaging **42**(1), 233–244 (2023)
5. Huang, X., Belongie, S.: Arbitrary style transfer in real-time with adaptive instance normalization. In: Proceedings of the IEEE international conference on computer vision. pp. 1501–1510 (2017)
6. Karani, N., Erdil, E., Chaitanya, K., Konukoglu, E.: Test-time adaptable neural networks for robust medical image segmentation. Medical Image Analysis **68**, 101907 (2021)
7. Lemaître, G., et al.: Computer-aided detection and diagnosis for prostate cancer based on mono and multiparametric MRI: a review. Computers in Biology and Medicine **60**, 8–31 (2015)
8. Litjens, G., et al.: Evaluation of prostate segmentation algorithms for MRI: the PROMISE12 challenge. Medical Image Analysis **18**(2), 359–373 (2014)
9. Liu, Q., Chen, C., Qin, J., Dou, Q., Heng, P.A.: FedDG: Federated domain generalization on medical image segmentation via episodic learning in continuous frequency space. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 1013–1023 (2021)
10. Liu, Q., Dou, Q., Heng, P.A.: Shape-aware meta-learning for generalizing prostate MRI segmentation to unseen domains. In: International Conference on Medical Image Computing and Computer Assisted Intervention. pp. 475–485 (2020)

11. Liu, Q., Chen, C., Dou, Q., Heng, P.A.: Single-domain generalization in medical image segmentation via test-time adaptation from shape dictionary. In: Proceedings of the AAAI Conference on Artificial Intelligence. vol. 36, pp. 1756–1764 (2022)
12. Liu, Q., Dou, Q., Yu, L., Heng, P.A.: MS-Net: multi-site network for improving prostate segmentation with heterogeneous MRI data. IEEE Transactions on Medical Imaging **39**(9), 2713–2724 (2020)
13. Orlando, J.I., et al.: REFUGE challenge: A unified framework for evaluating automated methods for glaucoma assessment from fundus photographs. Medical Image Analysis **59**, 101570 (2020)
14. Qiao, F., Zhao, L., Peng, X.: Learning to learn single domain generalization. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 12556–12565 (2020)
15. Shim, J.h., Yu, H., Kong, K., Kang, S.J.: FeedFormer: Revisiting Transformer decoder for efficient semantic segmentation. In: Proceedings of the AAAI Conference on Artificial Intelligence. vol. 37, pp. 2263–2271 (2023)
16. Sivaswamy, J., Krishnadas, S., Chakravarty, A., Joshi, G., Tabish, A., et al.: A comprehensive retinal image dataset for the assessment of glaucoma from the optic nerve head analysis. JSM Biomedical Imaging Data Papers **2**(1), 1004 (2015)
17. Sun, Y., Wang, X., Liu, Z., Miller, J., Efros, A., Hardt, M.: Test-time training with self-supervision for generalization under distribution shifts. In: International Conference on Machine Learning. pp. 9229–9248 (2020)
18. Wang, D., Shelhamer, E., Liu, S., Olshausen, B., Darrell, T.: Fully test-time adaptation by entropy minimization. In: International Conference on Learning Representations (2021)
19. Wang, S., Yu, L., Li, K., Yang, X., Fu, C.W., Heng, P.A.: DoFE: Domain-oriented feature embedding for generalizable fundus image segmentation on unseen datasets. IEEE Transactions on Medical Imaging **39**(12), 4237–4248 (2020)
20. Xie, E., Wang, W., Yu, Z., Anandkumar, A., Alvarez, J.M., Luo, P.: SegFormer: Simple and efficient design for semantic segmentation with Transformers. Advances in Neural Information Processing Systems **34**, 12077–12090 (2021)
21. Zhang, L., Wang, X., Yang, D., Sanford, T., Harmon, S., Turkbey, B., Wood, B.J., Roth, H., Myronenko, A., Xu, D., et al.: Generalizing deep learning for medical image segmentation to unseen domains via deep stacked transformation. IEEE Transactions on Medical Imaging **39**(7), 2531–2540 (2020)
22. Zhou, Z., Qi, L., Shi, Y.: Generalizable medical image segmentation via random amplitude mixup and domain-specific image restoration. In: European Conference on Computer Vision. pp. 420–436 (2022)