

# Medical Image Segmentation via Single-Source Domain Generalization with Random Amplitude Spectrum Synthesis

Qiang Qiao<sup>1,2</sup> , Wenyu Wang<sup>1,2</sup> , Meixia Qu<sup>1,2</sup> , Kun Su<sup>1,2</sup> ,  
Bin Jiang<sup>1,2</sup> , and Qiang Guo<sup>3,4</sup> 

<sup>1</sup> School of Mechanical, Electrical & Information Engineering,  
Shandong University, Weihai, China  
[jiangbin@sdu.edu.cn](mailto:jiangbin@sdu.edu.cn)

<sup>2</sup> Shenzhen Research Institute of Shandong University, Shenzhen, China

<sup>3</sup> School of Computer Science and Technology,  
Shandong University of Finance and Economics, Jinan, China  
[guoqiang@sdufe.edu.cn](mailto:guoqiang@sdufe.edu.cn)

<sup>4</sup> Shandong Provincial Key Laboratory of Digital Media Technology, Jinan, China

**Abstract.** The field of medical image segmentation is challenged by domain generalization (DG) due to domain shifts in clinical datasets. The DG challenge is exacerbated by the scarcity of medical data and privacy concerns. Traditional single-source domain generalization (SSDG) methods primarily rely on stacking data augmentation techniques to minimize domain discrepancies. In this paper, we propose Random Amplitude Spectrum Synthesis (RASS) as a training augmentation for medical images. RASS enhances model generalization by simulating distribution changes from a frequency perspective. This strategy introduces variability by applying amplitude-dependent perturbations to ensure broad coverage of potential domain variations. Furthermore, we propose random mask shuffle and reconstruction components, which can enhance the ability of the backbone to process structural information and increase resilience intra- and cross-domain changes. The proposed Random Amplitude Spectrum Synthesis for Single-Source Domain Generalization (RAS<sup>4</sup>DG) is validated on 3D fetal brain images and 2D fundus photography, and achieves an improved DG segmentation performance compared to other SSDG models. The source code is available at: <https://github.com/qintianjian-lab/RAS4DG>.

**Keywords:** Single-source domain generalization · Semantic segmentation · Frequency spectrum · Mask shuffle.

## 1 Introduction

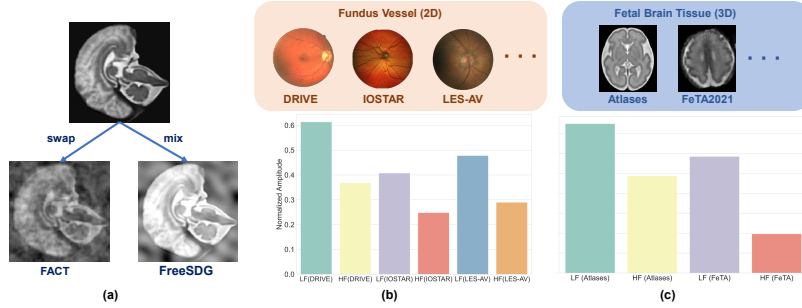
Benefited by deep learning, medical image segmentation has witnessed considerable progress in the clinical practice of computer-aided diagnosis and treatment

[24]. However, segmentation performance drops dramatically in the face of domain shifts such as different imaging modalities (e.g., CT and MR) or center data. To mitigate this, researchers explore domain adaptation (DA) and multi-source domain generalization (MSDG), which aim to transfer knowledge from annotated source domains to unlabeled target domains [26].

MSDG utilizes diversity from multiple source domains for generalization [31]. However, the success of data-driven approaches in the field of medical imaging is often hindered by the lack of sufficient data. This is because 1) medical images are difficult to collect; and 2) manual labeling of fine-grained data from different domains is time-consuming and labor-intensive. Thus, training in an SSDG setting from only a single source domain faces serious challenges [8]. To deal with this problem, meta-learning [10] divides the source domain to simulate domain migration to unknown target domains [14]. In the context of medical image segmentation, SSDG is constrained by its reliance on data from a single domain, which is insufficient for capturing the varied distribution characteristics present across multiple domains. Recent studies have concentrated on enhancing the dataset to emulate the characteristics of target domains with uneven distributions [17,30,32]. For example, RandConv [28] leverages stochastic convolution as an innovative data augmentation method designed to maintain the integrity of shape and local texture information. Unlike existing DG methods that standardize data to a uniform distribution, SAN-SAW [19] encourages both intra-category compactness and inter-category separability. From the perspective of gradient information, SLAug [22] employs saliency-balancing to guide enhancement of global and local regions. In addition, generating adversarial samples is also employed to bolster the generalization capabilities of models [23,7].

Analyzing medical images in the frequency domain provides a more profound understanding of the underlying patterns. Specifically, the Fourier transform of images produces both amplitude and phase spectra [13]. We identify two key insights from this transform: the phase spectrum captures local details through contours, while the amplitude spectrum reflects low-level information such as texture. There exist some works that focus on considering the amplitude spectrum in different images. Chattopadhyay et.al [2] discovered that introducing variable perturbations to the amplitude information in natural images can effectively mitigate domain shifts. The amplitude spectra of any two different images in a single source domain are swapped and mixed in FACT [27] and FreeSDG [11], respectively. When applying them in the Atlases dataset, we observe instabilities: FACT produces artifacts and FreeSDG causes significant color changes, as shown in Fig. 1(a). Besides, we analyze 2D and 3D image datasets from various domains, and reveal that high-frequency bands of the amplitude spectrum show less variation than low-frequency bands (see Fig. 1(b) and (c)).

Based on these observations, we explore the frequency bands from the amplitude and propose a novel medical SSDG method named Random Amplitude Spectrum Synthesis for Single-Source Domain Generalization (RAS<sup>4</sup>DG). During the RASS process, high-frequency bands are subjected to more pronounced perturbations than low-frequency bands. But images from different domains are



**Fig. 1.** Visualization of various results. (a) Results of different methods on the Atlases dataset. (b) 2D datasets and (c) 3D datasets used in our experiments and the statistical analysis of the image amplitude for high frequency (HF) and low frequency (LF).

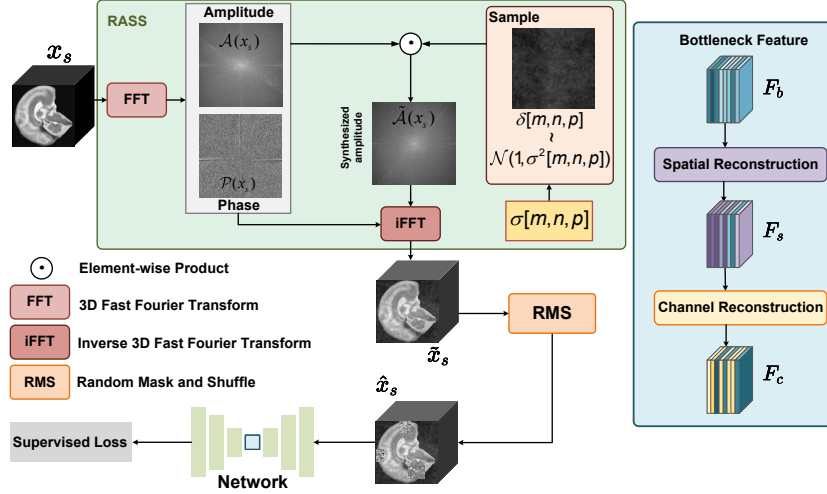
heterogeneous. It is difficult for the backbone network to capture structural information among images only by reducing the stylistic differences. Thus, we resort to make the backbone mine inter-image structural information to increase its context extraction capability. Specifically, local regions of the image are randomly masked with structural information, and the pixels within these regions are shuffled. To reduce feature redundancy and capture key information, a plug-and-play reconstruction convolution is further integrated into the backbone.

## 2 Method

Let  $\mathcal{S}$  and  $\mathcal{T}$  represent source and target domains, respectively, which share the same label space. The training set  $\mathcal{S} = \{(x_s, y_s)\}_{s=1}^N$  contains  $N$  training pairs, where  $x_s$  is the  $s$ -th image from the source domain, and  $y_s$  is the corresponding ground truth label. The data from the target domain  $\mathcal{T}$  is unknown and is not involved in the training process. Our RAS<sup>4</sup>DG framework is depicted in Fig. 2. The amplitudes obtained by applying the Fourier transform on  $x_s$  are randomly perturbed, and the inverse Fourier transform of the perturbed amplitudes and original phases is performed to recover the image  $\tilde{x}_s$ . After performing the style transformation, structural information is interfered by using random mask and shuffle (RMS) to obtain  $\hat{x}_s$ . Finally, the reconstruction design (RSD) at the bottleneck layer of the network eliminates redundant features.

### 2.1 Random Amplitude Spectrum Synthesis

From Fig. 1 (b) and (c), we can observe that the variance of low-frequency bands of the amplitudes is larger than that of high-frequency bands in both 2D and 3D medical images. Meanwhile, there is an inconsistency in the variance of the high- and low-frequency bands of these amplitudes. Specifically, the high-frequency bands exhibit relative consistency across domains, whereas low-frequency bands display significant disparities. This suggests that high-frequency bands should



**Fig. 2.** Pipeline of our RAS<sup>4</sup>DG. The bottleneck of the network is shown in right side.

be subjected to more variation than low-frequency bands, which help the model to extract domain-invariant features and prevent overfitting to a single domain. Thus, our method differentially perturbs the amplitude spectrum of images, with a particular focus on amplifying disturbances within the high-frequency range.

Given a 3D medical image  $x_s$ , we can extract its amplitude  $\mathcal{A}(x_s)$  and phase  $\mathcal{P}(x_s)$  by the Fourier transform. To minimize domain disparity and simulate potential changes in the unknown domain  $\mathcal{T}$ , we inject perturbations into the amplitude spectrum of source domain images, and apply the perturbation magnitude  $\sigma[m, n, p]$  uniformly across the frequency spectrum, where  $m, n, p$  denote spatial frequencies. Nonetheless, it is incapable of discerning the unique attributes of various frequency bands, which may lead to ignoring detailed information on high-frequency components vital for distinguishing domain differences.

To address the intrinsic complexity of medical images and exploit the nuanced differences between high- and low-frequency components, we propose Random Amplitude Spectrum Synthesis (RASS) to proportionally perturb the amplitude spectrum based on spatial frequency. Our perturbation function  $g(\cdot)$  is then applied to the amplitude spectrum  $\mathcal{A}(x_s)$  to achieve the desired perturbation. For each spatial frequency  $[m, n, p]$ , we introduce a random element  $\delta[m, n, p]$  sampled from a Gaussian distribution (i.e.,  $\delta[m, n, p] \sim \mathcal{N}(1, \sigma^2[m, n, p])$ ) to modulate the spectral components, which can be formalized as:

$$g(\mathcal{A}(x_s))[m, n, p] = \delta[m, n, p]\mathcal{A}(x_s)[m, n, p]. \quad (1)$$

This frequency-dependent perturbation magnitude  $\sigma[m, n, p]$  varies with frequency and allows a controlled increase in the perturbation amplitude of the high-frequency component. It is tailored to emphasize the high-frequency content that is often more susceptible to variations in medical imaging environments

and devices. The magnitude of perturbation for spatial frequency is defined as:

$$\sigma[m, n, p] = \left( 2\alpha \sqrt{\frac{m^2 + n^2 + p^2}{H^2 + W^2 + D^2}} \right)^\gamma + \beta, \quad (2)$$

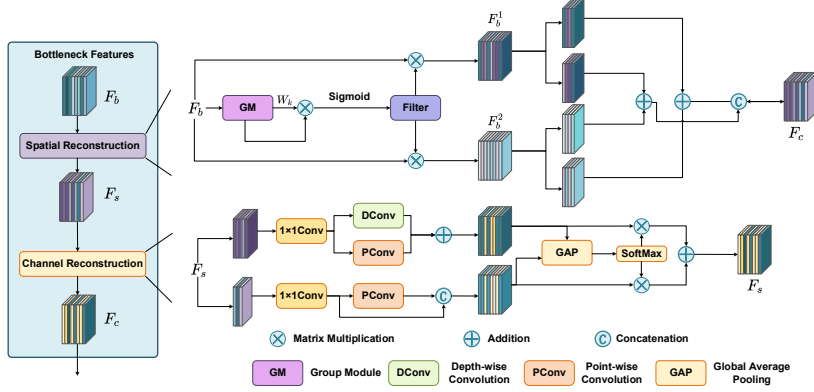
where  $\alpha$  controls the overall amplitude of the perturbation,  $\gamma$  determines the rate at which the perturbation increases with frequency, and  $\beta$  ensures that a minimal level of perturbation is uniformly applied, providing a baseline for all frequencies. This polynomial function ensures dynamic interactions within the amplitude spectra. In other words, higher frequencies are subject to a more pronounced perturbation in comparison to lower frequencies. The workflow of the RASS is summarized in Fig. 2. After obtaining the randomly synthesized amplitude, we combine it and its original phase to recover the image  $\hat{x}_s$ . Fine amplitude processing can simulate more realistic domain-varying medical images. Thus, RASS enhances the ability of model to capture domain-invariant features in the image domain by applying different levels of frequency perturbation.

## 2.2 Random Mask Shuffle and Reconstruction

RASS has been presented to address with artifacts and unknown domain distributions in this paper. However, current SSDG methods still face limitations in learning from heterogeneous structures. PatchShuffle [9] shows that random patch shuffling is more robust to noise and local image changes. To handle local variations present in medical images and inject randomness, we employ a strategy of randomly masking image regions and shuffling the pixels within these sections. This RMS strategy not only addresses the issue of local discrepancies but also compels our model to derive more robust feature representations by embracing randomness. Specifically, random positions and sizes are determined from the image  $\hat{x}_s$  after RASS, which aims to select the regions to be masked in the image. Then pixels within each designated region are shuffled and filled to their original locations. This pixels shuffling enables weights to be shared between neighboring pixels, which increases the complexity of the backbone. This is useful for higher level feature mapping. At last, we obtain the image  $\hat{x}_s$ . The injection of stochasticity allows our RASS and RMS strategies to enhance the adaptability of the network to various unforeseen changes in medical images.

Besides, feature redundancy hampers effective feature learning and diminishes the capacity of the network for feature representation. Thus, a plug-and-play feature reconstruction design tailored for segmentation models is proposed. RSD is used to enhance the capture of stylistic information after RASS and the feature representation within the region shuffled after RMS. We further optimize the features from the spatial and channel dimensions, as detailed in Fig. 3.

For the spatial dimension, we employ a strategy of ‘separation and reconstruction’ for the bottleneck layer features  $F_b$ , where we utilize trainable scaling factors from group normalization to evaluate the significance of individual features and measure the variability in spatial information. The features  $F_b$  are categorized into high-information  $F_b^1$  and low-information  $F_b^2$  based on weights

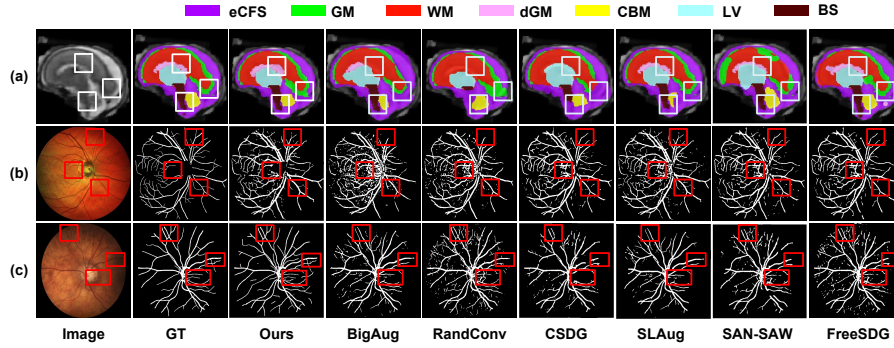


**Fig. 3.** Detail view of RSD.  $F_b$  is reconstructed from spatial and channel dimensions.

and cross-fused. This nuanced differentiation of features based on their information content allows for strategic recombination of a diverse set of features. This not only preserves the feature space but also enriches the representational power of the features. Furthermore, to refine and eliminate redundant features, we implement a strategy of ‘divide-transformation-merging’ along the channel dimension. Specifically, channels are divided, with one part enriched via depth-wise and point-wise convolutions, and the other part supplemented with point-wise convolutions. The enriched and supplemented features are merged by using the SKNet [12]. This process merges the richly learned features with supplementary ones to refine the channel features. By merging these features, we enhance the ability of our network to capture complex patterns and nuances within medical images, which is crucial for accurate segmentation.

### 3 Experiments and Results

**Dataset.** To demonstrate the versatility of RAS<sup>4</sup>DG, we evaluated it using 3D and 2D medical image segmentation tasks. For the 3D datasets, the source domain dataset of fetal brain tissue is obtained from Atlases, which contains 47 samples [4,5,25]. The target domain dataset is composed of 80 MRI volumes from the Fetal Brain Tissue Annotation and Segmentation Challenge (FeTA) 2021 dataset [18]. In addition to the background, both the source and target domains included seven distinct categories: external cerebrospinal fluid (eCSF), grey matter (GM), white matter (WM), deep grey matter (dGM), cerebellum (CBM), lateral ventricles (LV), and brainstem (BS). We further integrated 2D fundus vessel images from three different public datasets. The source domain images are derived from the DRIVE [21] fundus vascular dataset, which includes images of 20 patients and annotations. To assess the generalization capabilities of RAS<sup>4</sup>DG, we utilized the IOSTAR [1] and LES-AV [16] datasets as target domains, containing images from 30 and 22 patients, respectively.



**Fig. 4.** Visualization of segmentation masks predicted by different methods in different datasets. (a) FeTA2021, (b) IOSTAR, and (c) LES-AV.

**Implementation Details.** All experiments are trained on NVIDIA A100 GPU (80G Memory) using the PyTorch framework. The results are evaluated using the Dice similarity coefficient (DSC) scores. We use SegResNet [15] as the backbone model. The training epoch is set to 300, and the initial learning rate is set to 0.0001. The network parameters is handled by the Adam optimizer. Furthermore, we implement a ‘Ploy’ learning strategy to dynamically adjust the learning rate. All 3D images are resized to  $144 \times 144 \times 144$  voxels, while 2D images are resized to  $512 \times 512$  pixels. Our RAS<sup>4</sup>DG is trained from scratch with early stopping. In RASS, hyperparameters  $\alpha$ ,  $\beta$ , and  $\gamma$  are set to 3.0, 0.25, and 2.0, respectively. More details can be found in the supplementary materials.

**Comparison with the State-of-the-arts.** To demonstrate the effectiveness of the proposed RAS<sup>4</sup>DG, we report its average performance against state-of-the-art (SOTA) methods across three training rounds on 3D and 2D medical image datasets. We compare RAS<sup>4</sup>DG against two benchmark methods: the ‘Intra-Domain’ (an upper bound by training and testing within the same target domain) and the ‘w/o SSDG’ (a lower bound by training in the source domain and testing in the target domain). We also explore two DA methods (CutMix [29] and DSA [6]), two Fourier-transform-based DG techniques (FedDG [13] and FreeSDG [11]), and five SSDG methods (BigAug [32], RandConv [28], CSDG [17], SLAug [22], and SAN-SAW [19]). All competitive methods are re-implemented based on their official repositories and utilize the same backbone network as our method. The experimental results are listed in Table 1. Our method outperforms the two DA methods as well as the SSDG methods in terms of performance. Compared with the baseline model (w/o SSDG), our RAS<sup>4</sup>DG achieves an average improvement of 8.15% in DSC for the FeTA2021 dataset, 8.43% for the IOSTAR dataset, and 13.71% for the LES-AV dataset. Additionally, the proposed method narrows the DSC gap with supervised training to 4.80% in FeTA2021, 8.38% in IOSTAR, and 4.41% in LES-AV. It is worth noting that FedDG is a Fourier based method for MSDG. The experiments on the 3D dataset do not fulfill the

**Table 1.** Comparisons with SOTA methods using DSC scores (%). The three columns for FeTA2021 results are abnormal, normal, and average DSC. Best scores are in **bold**.

| Methods                   | 3D                               |                                  |                                  | 2D                               |                                  |
|---------------------------|----------------------------------|----------------------------------|----------------------------------|----------------------------------|----------------------------------|
|                           | FeTA2021                         |                                  |                                  | IOSTAR                           | LES-AV                           |
| Intra-Domain              | 76.98 $\pm$ 0.17                 | 85.74 $\pm$ 0.23                 | 81.36 $\pm$ 0.21                 | 74.24 $\pm$ 0.13                 | 77.29 $\pm$ 0.31                 |
| CutMix [29]               | 74.10 $\pm$ 0.13                 | 79.70 $\pm$ 0.50                 | 76.20 $\pm$ 0.20                 | 64.82 $\pm$ 0.31                 | 71.48 $\pm$ 0.21                 |
| DSA [6]                   | 73.44 $\pm$ 0.61                 | 79.91 $\pm$ 0.42                 | 75.93 $\pm$ 0.52                 | 65.58 $\pm$ 0.19                 | 71.03 $\pm$ 0.11                 |
| FedDG [13]                | -                                | -                                | -                                | 65.30 $\pm$ 0.33                 | 71.33 $\pm$ 0.71                 |
| w/o SSDG                  | 66.78 $\pm$ 0.14                 | 70.04 $\pm$ 0.18                 | 68.41 $\pm$ 0.14                 | 57.43 $\pm$ 0.21                 | 59.17 $\pm$ 0.21                 |
| BigAug [32]               | 70.90 $\pm$ 0.41                 | 79.27 $\pm$ 0.31                 | 74.15 $\pm$ 0.26                 | 62.43 $\pm$ 0.13                 | 66.87 $\pm$ 0.14                 |
| RandConv [28]             | 70.03 $\pm$ 0.21                 | 76.81 $\pm$ 0.35                 | 73.42 $\pm$ 0.22                 | 62.01 $\pm$ 0.16                 | 67.98 $\pm$ 0.21                 |
| CSDG [17]                 | 69.38 $\pm$ 0.21                 | 74.88 $\pm$ 0.13                 | 72.13 $\pm$ 0.15                 | 61.94 $\pm$ 0.14                 | 68.30 $\pm$ 0.12                 |
| SLAug [22]                | 71.52 $\pm$ 0.13                 | 77.04 $\pm$ 0.32                 | 74.28 $\pm$ 0.27                 | 62.36 $\pm$ 0.28                 | 68.79 $\pm$ 0.29                 |
| SAN-SAW [19]              | 68.73 $\pm$ 0.38                 | 72.53 $\pm$ 0.32                 | 70.63 $\pm$ 0.34                 | 63.21 $\pm$ 0.12                 | 62.92 $\pm$ 0.31                 |
| FreeSDG [11]              | 68.57 $\pm$ 0.19                 | 71.59 $\pm$ 0.32                 | 70.08 $\pm$ 0.29                 | 59.72 $\pm$ 0.13                 | 63.53 $\pm$ 0.19                 |
| RAS <sup>4</sup> DG(ours) | <b>73.52<math>\pm</math>0.17</b> | <b>81.32<math>\pm</math>0.14</b> | <b>76.56<math>\pm</math>0.23</b> | <b>65.86<math>\pm</math>0.12</b> | <b>72.88<math>\pm</math>0.07</b> |

**Table 2.** Ablation study on different components of the proposed method on the FeTA2021 dataset. The baseline is ‘w/o SSDG’ method.

| Methods     | RASS | RMS | RSD | DSC                              |
|-------------|------|-----|-----|----------------------------------|
| Baseline    |      |     |     | 68.41 $\pm$ 0.14                 |
| RASS (only) | ✓    |     |     | 73.23 $\pm$ 0.19                 |
| RMS (only)  |      | ✓   |     | 70.92 $\pm$ 0.12                 |
| RSD (only)  |      |     | ✓   | 69.41 $\pm$ 0.31                 |
| w/o RASS    |      | ✓   | ✓   | 73.33 $\pm$ 0.12                 |
| w/o RMS     | ✓    |     | ✓   | 73.56 $\pm$ 0.18                 |
| w/o RSD     | ✓    | ✓   |     | 74.48 $\pm$ 0.25                 |
| Ours        | ✓    | ✓   | ✓   | <b>76.56<math>\pm</math>0.23</b> |

multi-source condition, so the results are not provided. Fig. 4 show some segmentation results. As evident in the boxed area, our RAS<sup>4</sup>DG enhances local details compared to other methods. Our results match the ground truth most accurately.

**Ablation Study** The ablation study is presented in Table 2. We adopt the trained baseline model as a lower bound. These three components improve DSC performance by 4.82%, 2.51%, and 1.00%, respectively. Specifically, RASS yields the greatest performance improvement by sophisticated synthesis of the amplitude spectrum, while the combination of RMS and RSD better facilitates the learning of structural information. All the components ultimately derive our RAS<sup>4</sup>DG. The reader is referred to the Supplementary Material for more information on architecture and hyperparametric ablation.

## 4 Conclusion

In this paper, a framework called RAS<sup>4</sup>DG is proposed for performing DG segmentation of medical images from the frequency domain perspective, which includes RASS, RMS, and RSD. In contrast to standard DG learning schemes, SSDG aims to ensure out-of-domain generalization by using data from only one source dataset. RASS is to introduce perturbations in the amplitude spectrum of medical images to simulate inter-domain variability. We also integrate RMS during training and reconstruction design into the network bottleneck to further improve segmentation performance. Our experimental results on 2D and 3D segmentation tasks show that RAS<sup>4</sup>DG significantly outperforms some recent SSDG methods and DA methods.

**Acknowledgments.** This work was supported by the Shenzhen Science and Technology Program (JCYJ20230807094104009), in part by the National Natural Science Foundation of China (61873145).

**Disclosure of Interests.** The authors have no competing interests in the paper.

## References

1. Abbasi-Sureshjani, S., Smit-Ockeloen, I., Bekkers, E., Dashtbozorg, B., ter Haar Romeny, B.: Automatic detection of vascular bifurcations and crossings in retinal images using orientation scores. In: *Proceeding of the IEEE International Symposium on Biomedical Imaging (ISBI)*, pp. 189–192 (2016)
2. Chattopadhyay, P., Sarangmath, K., Vijaykumar V., Hoffman, J.: PASTA: proportional amplitude spectrum training augmentation for syn-to-real domain generalization. In: *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, pp. 19288–19300 (2023)
3. Falk, T., et al.: U-net: deep learning for cell counting, detection, and morphometry. *Nat. Methods* **16**(1), 67–70 (2019)
4. Fidon, L., et al.: A spatio-temporal atlas of the developing fetal brain with spina bifida aperta. *Open Res. Eur.* **1**, 123 (2022)
5. Gholipour, A.: A normative spatiotemporal MRI atlas of the fetal brain for automatic segmentation and analysis of early brain growth. *Sci. Rep.* **7**(1), 476 (2017)
6. Han, X., et al.: Deep symmetric adaptation network for cross-modality medical image segmentation. *IEEE Trans. Med. Imaging* **41**(1), 121–132 (2021)
7. Hendrycks, D., Mu, N., Cubuk, E., D., Zoph, B., Gilmer, J., and Lakshminarayanan, B.: AugMix: a simple data processing method to improve robustness and uncertainty. In: *International Conference on Learning Representations (ICLR)* (2020)
8. Hu, S., Liao, Z., Zhang, J., Xia, Y.: Domain and content adaptive convolution based multi-source domain generalization for medical image segmentation. *IEEE Trans. Med. Imaging* **42**(1), 233–244 (2022)
9. Kang G., Dong X., Zheng L., and Yang Y.: Patchshuffle regularization. *arXiv preprint arXiv:1707.07103* (2017)
10. Li, D., Yang, Y., Song, Y.Z., Hospedales, T.M.: Learning to generalize: meta-learning for domain generalization. In: *Proceedings of the AAAI Conference on Artificial Intelligence (AAAI)*, pp. 3490–3497 (2018)

11. Li, H., et al.: Frequency-mixed single-source domain generalization for medical image segmentation. In: Greenspan, H., Madabhushi, A., Mousavi, P., Salcudean, S., Duncan, J., Syeda-Mahmood, T., Taylor, R. (eds.) MICCAI 2023. LNCS, vol. 14225, pp. 127–136. Springer, Cham (2023).
12. Li, X., Wang, W., Hu, X., Yang, J.: Selective kernel networks. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), pp. 510–519 (2019)
13. 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 (CVPR), pp. 1013–1023 (2021)
14. Liu, X., Thermos, S., O’Neil, A., Tsaftaris, S.A.: Semi-supervised meta-learning with disentanglement for domain-generalised medical image segmentation. In: de Bruijne, M., Cattin, P.C., Cotin, S., Padoy, N., Speidel, S., Zheng, Y., Essert, C. (eds.) MICCAI 2021. LNCS, vol. 12902, pp. 307–317. Springer, Cham (2021).
15. Myronenko, A.: 3D MRI brain tumor segmentation using autoencoder regularization. In: Crimi, A., Bakas, S., Kuijff, H., Keyvan, F., Reyes, M., van Walsum, T. (eds.) MICCAI Brainlesion Workshop. LNCS, vol. 11384, pp. 311–320. Springer, Cham (2019)
16. Orlando, J.I., Barbosa Breda, J., Van Keer, K., Blaschko, M.B., Blanco, P.J., Bulant, C.A.: Towards a glaucoma risk index based on simulated hemodynamics from fundus images. In: Frangi, A.F., Schnabel, J.A., Davatzikos, C., Alberola-López, C., Fichtinger, G. (eds.) MICCAI 2018. vol. 11071, pp. 65–73. Springer, Cham (2018)
17. Ouyang, C., et al.: Causality-inspired single-source domain generalization for medical image segmentation. *IEEE Trans. Med. Imaging* **42**(4), 1095–1106 (2022)
18. Payette, K., et al.: An automatic multi-tissue human fetal brain segmentation benchmark using the fetal tissue annotation dataset. *Sci. Data* **8**(1), 167 (2021)
19. Peng, D., Lei, Y., Hayat, M., Guo, Y., Li, W.: Semantic-aware domain generalized segmentation. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), pp. 2594–2605 (2022)
20. Roy, S., et al.: MedNeXt: Transformer-driven scaling of convnets for medical image segmentation. In: Greenspan, H., Madabhushi, A., Mousavi, P., Salcudean, S., Duncan, J., Syeda-Mahmood, T., Taylor, R. (eds.) MICCAI 2023. LNCS, vol. 14223, pp. 405–415. Springer, Cham (2023).
21. Staal, J., Abràmoff, M.D., Niemeijer, M., Viergever, M.A., Van Ginneken, B.: Ridge-based vessel segmentation in color images of the retina. *IEEE Trans. Med. Imaging* **23**(4), 501–509 (2004)
22. Su, Z., Yao, K., Yang, X., Huang, K., Wang, Q., Sun, J.: Rethinking data augmentation for single-source domain generalization in medical image segmentation. In: Proceedings of the AAAI Conference on Artificial Intelligence (AAAI), pp. 2366–2374 (2023)
23. Volpi, R., Namkoong, H., Sener, O., Duchi, J.C., Murino, V., Savarese, S.: Generalizing to unseen domains via adversarial data augmentation. In: Advances in Neural Information Processing Systems (NIPS), pp. 5339–5349 (2018)
24. Wang, J., et al.: Generalizing to unseen domains: a survey on domain generalization. *IEEE Trans. Knowl. Data Eng.* **35**(8), 8052–8072 (2022)
25. Wu, J., et al.: Age-specific structural fetal brain atlases construction and cortical development quantification for Chinese population. *Neuroimage* **241**(1), 118412 (2021)

26. Xie, X., Niu, J., Liu, X., Chen, Z., Tang, S., Yu, S.: A survey on incorporating domain knowledge into deep learning for medical image analysis. *Med. Image Anal.* **69**, 101985 (2021).
27. Xu, Q., Zhang, R., Zhang, Y., Wang, Y., Tian, Q.: A Fourier-based framework for domain generalization. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 14383–14392 (2021)
28. Xu, Z., Liu, D., Yang, J., Raffel, C., Niethammer, M.: Robust and generalizable visual representation learning via random convolutions. In: *International Conference on Learning Representations (ICLR)* (2021)
29. Yun, S., Han, D., Oh, S.J., Chun, S., Choe, J., Yoo, Y.: CutMix: regularization strategy to train strong classifiers with localizable features. In: *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, pp. 6023–6032 (2019)
30. Zhang, L., et al.: Generalizing deep learning for medical image segmentation to unseen domains via deep stacked transformation. *IEEE Trans. Med. Imaging* **39**(7), 2531–2540 (2020)
31. Zhou, K., Liu, Z., Qiao, Y., Xiang, T., Loy, C.C.: Domain generalization: A survey. *IEEE Trans. Pattern Anal. Mach. Intell.* **45**(4), 4396–4415 (2022)
32. Zhou, Z., Qi, L., Yang, X., Ni, D., Shi, Y.: Generalizable cross-modality medical image segmentation via style augmentation and dual normalization. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 20856–20865 (2022)