ADA: An Adaptive Augmentation Framework for Single-Source Domain Generalization in Medical Image Segmentation

Runlin Huang[1], Hongmin Cai[3], Weipeng Zhuo[1,2], Shangyan Cai[1], Haowei Lin[1], Wentao Fan[1,2], and Weifeng Su[⋆1,2]

[1] Division of Science and Technology, Beijing Normal-Hong Kong Baptist University, Zhuhai 519087, China
[2] Guangdong Provincial Key Laboratory of Interdisciplinary Research and Application for Data Science, Beijing Normal-Hong Kong Baptist University, Zhuhai 519087, China
[3] School of Computer Science and Engineering, South China University of Technology, Guangzhou 510006, China

Abstract. In medical image analysis, significant challenges arise from domain shifts. Models trained on one dataset often struggle to generalize to unseen domains, limiting their clinical utility. To overcome this challenge, recent advancements have tried to increase the diversity of training data with data augmentation, in which the augmentation rules are pre-set before training commences and remain unchanged throughout the training process. Previous methods do not augment according to the unique characteristics of individual samples. As a result, they fail to cover the full diversity of unseen domains. To tackle this problem, we propose a learnable framework, the Adaptive Augmentation Framework (ADA), which can adaptively augment data catering to each individual sample. It has three operators for different purposes: 1) the Learnable Bezier Remap operator dynamically adjusts parameters to do the augmentation according to its content features. 2) the Channel Shift Control operator dynamically tunes shift and scale parameters for each color channel. By capturing fine-grained variations and improving spectral detail representation. 3) The Gradient-guided Feature Weaken operator dynamically reduces the influence of high-impact features to improve the model's ability to generalize. Extensive experiments conducted on seven medical segmentation datasets demonstrate that adaptive augmentation is more likely to cover large diversity in the unseen domain.

Keywords: Domain Generalization · Segmentation · Medical Image.

1 Introduction

Despite the excellent performance of deep-learning models in medical image analysis, a pivotal challenge remains the domain shift phenomenon, where models

⋆ Co-corresponding author: wfsu@uic.edu.cn

Static Image Augmentation	Dynamic Image Augmentation
Fixed rules, predefined before training	Adaptive rules, adjusted based on sample characteristics
Uniform transformations applied to all samples	Personalized transformations based on individual sample
Limit to cover the full distribution of unseen domains	Integrating diverse inputs into a consistent feature space

Table 1. Comparison of Static and Dynamic Image Augmentation. Previous methods like CISDG, SLAUG, and CCSDG use static augmentation techniques that lack the flexibility to dynamically adapt to the specific characteristics of individual samples. Our method (ADA) combines both static and dynamic augmentation methods, enhancing the model's generalization capability.

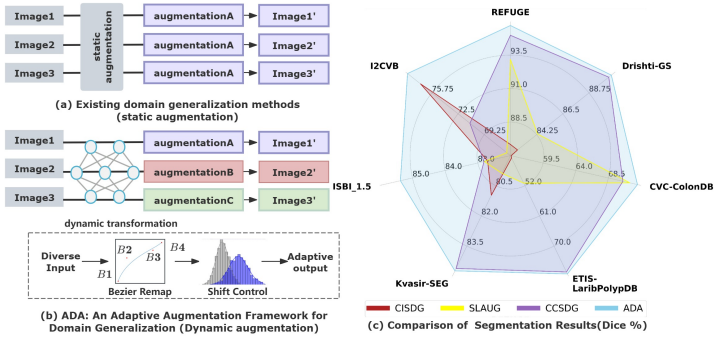

Fig. 1. Comparative analysis of ADA versus previous domain generalization methods. (a) Previous domain generalization methods enhance diversity by static augmentations to improve model generalization. (b) ADA employs a novel approach by integrating static and dynamic augmentations. It adapts to diverse input through learnable Bezier curve remapping and controlled distribution shifts, enabling personalized transformations based on individual samples, thus enhancing generalization. (c) ADA outperforms state-of-the-art methods on seven medical image segmentation datasets.

trained on data from one domain struggle to generalize effectively to unseen domains. The domain shift poses a substantial challenge in clinical practice, given the broad variability in medical images. This variability stems from differences in imaging equipment, protocols, and patient demographics. Consequently, the model's inability to generalize can severely limit its clinical utility, as its performance may degrade when applied to images from new centers or devices not represented in the training data.

To alleviate this problem, such methods using ensemble learning [8] for brain MRI lesion segmentation or utilizing meta-learning[4] for swift adaptation in brain MRI tasks, have made strides in refining learning strategies and architectures. Nonetheless, these approaches often require extensive computational re-

sources and retraining efforts, which are impractical for real-world applications with limited computational capabilities and diverse medical imaging datasets.

Different from previous methods, existing works propose increasing the diversity of input and utilizing consistency strategies to extract domain-invariant features [16, 15, 10]. For instance, SLAUG [13] introduces a saliency-balanced fusion approach that integrates both global and local image augmentation. C2SDG[5] enhances image diversity using the StyleAug method and uses contrastive feature disentanglement to constrain model consistency. These methods have primarily focused on enhancing the diversity of training images to combat this issue. However, these methods are inadequate to cover the full diversity of unknown domain distributions. These approaches only explore static image augmentation and consistency strategies in domain generalization (**Figure 1a**), while dynamic adaptation remains an ignored topic in the literature.

Existing methods primarily delve into static image augmentation for domain generalization (**Table 1**), in which the augmentation rules are set before training commences and remain immutable throughout the training process. Moreover, a uniform set of transformations is applied to all samples, without taking into account the distinct characteristics or variations of each individual sample. In contrast, our work introduces dynamic adaptation methods through the Adaptive Augmentation framework (ADA), which dynamically adapts transformation parameters based on each sample's unique characteristics (**Figure 1b**), thereby enhancing generalization across diverse and unseen domains. This is achieved through a dual-module framework: the Learnable Bezier Remap operator and the Channel Shift Control operator.

The Learnable Bezier Remap operator is designed to minimize domain disparity by dynamically adjusting parameters for each sample, thereby mapping it into a unified feature space. Concurrently, the Channel Shift Control operator refines the process by adjusting shift and scale parameters for each RGB channel individually, thereby capturing both global channel characteristics and local details within the images. Moreover, to enhance the model's generalization capability across varied domains, we have integrated a Gradient-guided Feature Weaken operator. This module weakens the influence of high-impact features, enhancing the model's ability to generalize by simulating diverse domain samples. This strategy helps in reducing overfitting to domain-specific features.

Our comprehensive experiments, conducted on seven distinct medical segmentation datasets from different domains, have demonstrated the effectiveness of ADA (**Figure 1c**). The empirical evidence clearly shows that our method not only surpasses previous methods in terms of performance but also significantly contributes to the advancement of single-source domain generalization in medical image segmentation.

In summary, this paper makes the following key contributions to the field of medical image segmentation:(1) We propose ADA, a novel adaptive augmentation framework that improves generalization in medical image segmentation by dynamically adapting augmentations to each sample's unique characteristics. Through adaptive feature adjustments, ADA effectively enhances robustness

against unseen domain variations. (2) We introduce three novel operators: the Learnable Bezier Remap operator for dynamic sample remapping, the Channel Shift Control operator for fine-tuning RGB channel parameters, and the Gradient-guided Feature Weaken operator for reducing high-impact features. Together, these operators enhance generalization, robustness, and adaptability in medical image segmentation across diverse domains. (3) Extensive experiments conducted on seven distinct medical segmentation datasets from different domains demonstrate the superiority of our ADA framework over existing methods, demonstrating its effectiveness in improving model generalization and addressing domain shift challenges in medical image segmentation.

2 Method

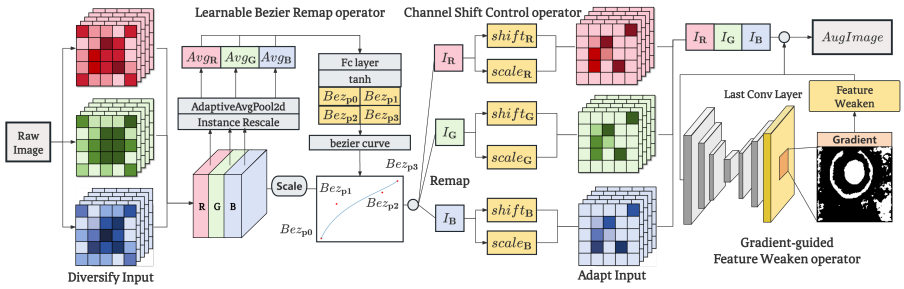

Fig. 2. Overview of our Adaptive Augmentation Framework (ADA) Structure. The framework consists of three core modules: (1) The Learnable Bezier Remap operator, which maps samples from different domains into a unified feature space by dynamically adjusting transformation parameters. (2) The Channel Shift Control operator, which refines spectral detail representation by tuning shift and scale parameters for each color channel. (3) The Gradient-guided Feature Weaken operator, which reduces the influence of domain-specific features, further improving the model's adaptability and robustness across unseen domains.

2.1 Learnable Bezier Remap operator

We introduce the Learnable Bezier Remap operator, which leverages the principles of Bezier curves to perform image enhancement. Given a batch of images $X \in \mathbb{R}^{N\times3\times512\times512}$, where N is the batch size, the module adaptively computes Bezier control points based on the image content and applies the transformation to enhance image contrast and sharpness.

The module first employs an adaptive average pooling layer, compressing each image in the batch to its average color intensities, yielding $A \in \mathbb{R}^{N\times3}$.

Subsequently, a linear transformation, defined by the weight matrix $W \in \mathbb{R}^{3\times4}$ and bias $b \in \mathbb{R}^4$, maps A to a set of control points for the Bezier curves:

$$\begin{bmatrix} P_{1b} \\ P_{2b} \end{bmatrix} = tanh(W(A_{b,c}) + b) \tag{1}$$

for $b = 1, \ldots, B$ and $c = 1, \ldots, C$. P_{1b},P_{2b} represents the intermediate control points, which define the dynamic control points of the Bezier curve for each image.

$$A_{b,c} = \frac{1}{H \times W} \sum_{i=1}^{H} \sum_{j=1}^{W} X_{b,c,i,j} \tag{2}$$

for $i = 1, \ldots, 512$ and $j = 1, \ldots, 512$. Here, $P \in \mathbb{R}^{B\times C}$ is the result of applying adaptive average pooling, reducing each spatial dimension of the image to a single scalar per channel.

The Bezier curve for each image and channel is computed using the fixed points $P_0 = (0, 0)$ and $P_3 = (1, 1)$, and the dynamically estimated P_1 and P_2. For a parameter $t \in [0, 1]$, the Bezier curve is defined as:

$$\text{Bezier}(t) = (1-t)^3 P_0 + 3(1-t)^2 t P_1 + 3(1-t)t^2 P_2 + t^3 P_3 \tag{3}$$

z:

$$X_{(i,c)}^{\text{enhanced}} = \text{Bezier}(X_{(i,c)}^{\text{norm}}) \tag{4}$$

where i and c index into the batch and color channels, respectively.

2.2 Channel Shift Control operator

After the raw image, X is processed by the Learnable Bezier Remap operator, given an enhanced image $E \in \mathbb{R}^{N\times C\times H\times W}$, where N denotes the batch size, C is the number of channels (3 for RGB images), and H, W represent the height and width of the image, respectively.

The module first employs an adaptive average pooling layer, compressing each image in the batch to its average color intensities, yielding $A \in \mathbb{R}^{N\times3}$. Function F is represented by a fully connected layer, which maps the pooled features to scale and shift parameters. For the input A, we obtain scale (S) and shift (T) parameters as follows:

$$\begin{bmatrix} S_{b,c} \\ T_{b,c} \end{bmatrix} = tanh(F(A_{b,c}) + b) \tag{5}$$

for $b = 1, \ldots, B$ and $c = 1, \ldots, C$. $S, T \in \mathbb{R}^{B\times C}$ are the scale and shift parameters for each channel, respectively.

The adjusted image $\hat{E}$ is then computed by applying the scale and shift parameters to each pixel of the original image:

$$\hat{E}_{b,c,i,j} = E_{b,c,i,j} \cdot (1 + S_{b,c}) + T_{b,c} \tag{6}$$

for all $b = 1, \ldots, B$, $c = 1, \ldots, C$, $i = 1, \ldots, H$, and $j = 1, \ldots, W$.

Each channel of the image is independently adjusted with its own scale and shift parameters, which are derived from the global characteristics of the channel captured by the adaptive average pooling.

2.3 Gradient-guided Feature Weaken operator

Let $x_s \in \mathbb{R}^{C' \times H' \times W'}$ denote the input tensor to the last convolutional layer, where C' represents the number of channels and H', W' are the spatial dimensions of the height and width, respectively. Given a binary cross-entropy loss function , the loss L with respect to the ground truth gt is:

$$L = -\sum(gt \cdot \log(y) + (1 - gt) \cdot \log(1 - y)) \tag{7}$$

The gradient of L with respect to x_s, denoted as ∇L, is computed using backpropagation. Define $g \in \mathbb{R}^{H' \times W' \times C'}$ as the gradient tensor ∇L. The threshold θ for mask generation is set based on the specified percentile of the gradient magnitudes. The weaken tensor x_w is defined as:

$$x_w = \begin{cases} x_s \odot \alpha, & \text{if } |\nabla L_{h,w,c}| \geq \theta \\ x_s, & \text{otherwise} \end{cases} \tag{8}$$

x_w represents the features processed by the Gradient-guided Feature Weaken operator, which reduces the impact of high-gradient features. It improves the model's ability to generalize by focusing on less dominant, more robust features, making the model more adaptable to new, unseen data.

Metric	$\text{REFUGE}_{\text{CUP}}$	$\text{REFUGE}_{\text{DISC}}$	$\text{REFUGE}_{\text{MEAN}}$	$\text{Drishti-GS}_{\text{CUP}}$	$\text{Drishti-GS}_{\text{DISC}}$	$\text{Drishti-GS}_{\text{MEAN}}$
ERM	70.67%	92.88%	81.77%	83.18%	96.49%	89.83%
RSC[6]	75.81%	90.82%	83.31%	84.16%	95.87%	90.01%
CISDG[10]	67.45%	86.83%	77.14%	77.00%	88.23%	82.61%
SLAUG[13]	77.98%	93.17%	85.58%	72.32%	96.34%	84.33%
CCSDG[5]	83.60%	94.98%	89.29%	84.34%	**96.55%**	90.45%
ADA	**85.80%**	**95.71%**	**90.76%**	**86.74%**	94.59%	**90.67%**

	CVC-ColonDB		ETIS-LaribPolypDB		Kvasir-SEG		ISBI_1.5		I2CVB	
Metric	Dice (%)	IoU (%)	Dice (%)	IoU (%)	Dice (%)	IoU (%)	Dice (%)	IoU (%)	Dice (%)	IoU (%)
ERM	65.11	48.81	65.02	51.82	82.87	71.61	82.95	71.18	64.46	48.07
RSC[6]	68.72	54.04	45.34	29.63	79.59	67.74	83.47	71.85	73.86	58.87
CISDG[10]	55.10	38.91	43.71	30.28	80.97	68.77	82.70	70.73	77.16	63.10
SLAUG[13]	71.46	55.75	51.22	38.43	79.88	67.52	82.80	70.85	66.44	50.49
CCSDG[5]	70.51	55.65	75.42	61.97	84.61	74.15	82.81	71.02	71.03	55.76
ADA	**72.51**	**57.22**	**76.08**	**63.43**	**84.73**	**74.93**	**85.37**	**74.67**	**78.77**	**65.26**

Table 2. Dice and IoU results for different datasets and models (in percentages)

3 Experiments and results

3.1 Experimental setup

Single-Source Domain Generalization for segmentation: The segmentation model is trained on a single source domain and directly evaluated on other target (unseen) domains. Performance is measured using Dice and IoU metrics.

Dataset Our study uses three medical datasets: Polyp, Fundus, and Prostate. The source domains include 488 colonoscopy images from CVC-ClinicDB[1] for polyp segmentation, 400 fundus images from REFUGE[9] for optic disc and cup segmentation, and 162 MRI images from UCL for prostate segmentation. The target domains for Polyp are CVC-ColonDB[14], ETIS[2], and Kvasir-SEG[7]. For Fundus, the targets are Drishti-GS[12] and REFUGE. For Prostate, the targets are I2CVB and ISBI_1.5.

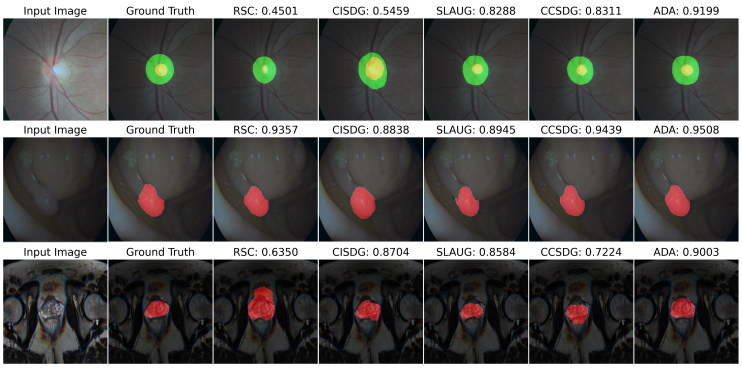

Fig. 3. Comparison of masks predicted by ADA and previous competing methods.

Experimental Settings Our segmentation network leverages an architecture based on CNN structure[11] and DeepLabv3+[3], specifically adapted for segmentation. mage augmentations employed, following the settings in [5], include Gaussian noise, contrast adjustment, etc. The model implementation was carried out using PyTorch on a single NVIDIA GeForce RTX 4090 GPU.

3.2 Comparison with SOTA Methods

The quantitative evaluation in **Table 2** highlights the superiority of our proposed method, ADA, over existing techniques across various medical image segmentation tasks, showcasing its enhanced generalization ability. These results clearly indicate that ADA's dynamic adaptation approach is more effective than static image augmentation strategies employed by previous methods. The visualization of different methods is presented in **Fig. 3**, where ADA demonstrates fewer misclassified predictions in the unseen target domain, further illustrating its superior domain generalization capability.

	Bezier	ConShift	Weaken	Avg Dice
ERM	-	-	-	82.95
Variant 1	✓			83.88
Variant 2		✓		83.61
Variant 3			✓	81.90
Variant 4		✓	✓	83.68
Variant 5	✓	✓		84.69
Variant 6	✓		✓	83.89
Variant 7	✓	✓	✓	85.37

Table 3. Ablation study on the effect of Bezier, ConShift, and Weaken.

	Cup	Disc	Avg
Negative Grad	83.41	96.45	89.93
Positive Grad	86.74	94.59	90.67
Abs Grad	86.14	95.65	90.90
	Cup	Disc	Avg
Quartic	83.64	93.78	88.38
Cubic	85.80	95.71	90.76
Quadratic	83.87	93.71	88.82

Table 4. Impact of Component Strategies on Performance Metrics.

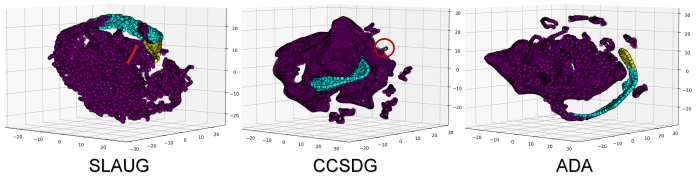

Fig. 4. t-SNE Visualization of Target Pixel Features. The plots show feature distributions, with purple for the background, cyan for the optic disc, and yellow for the optic cup. In SLAUG, the red arrow highlights regions of poor class distinction. In CCSDG, the red circle indicates confusion between cup and background features. In contrast, ADA clearly separates disc and cup features from the background, demonstrating superior segmentation performance.

3.3 Analytical Experiments

Ablation Study. We conduct ablation studies on the three main ADA components: Learnable Bezier Remap (Bezier), Channel Shift Control (ConShift), and Gradient-guided Feature Weaken (Weaken). Table 3 shows that each component is crucial for enhancing model performance, with their combined use leading to superior generalization. **Gradient-Guided Strategy.** In Table 4 Upper, we evaluate the impact of different gradient manipulation strategies on segmentation performance. The Negative Grad strategy, which weakens negative gradients, underperforms, highlighting the importance of these gradients. In contrast, both the Positive Grad and Abs Grad strategies improve performance by reducing overfitting and promoting better generalization. **Bezier Curve Complexity.** In Table 4 Lower, we evaluate Bezier curve complexity in the Global Bezier Remap Module. Testing quadratic, cubic, and quartic curves shows that a single control point lacks flexibility, while three lead to overfitting. The cubic Bezier curve strikes the best balance, offering improved generalization by avoiding both underfitting and overfitting.

4 Conclusion

We introduce the Adaptive Augmentation Framework (ADA), which adapts augmentations to individual samples, improving generalization across unseen domains. Experiments on seven medical segmentation datasets show ADA outperforms existing methods, enhancing robustness and clinical relevance.

Acknowledgments. Our work was supported in part by the Guangdong Provincial Key Laboratory of IRADS (2022B1212010006) and in part by Guangdong Higher Education Upgrading Plan (2021-2025) and in part by by Guangdong and Hong Kong Universities "1+1+1" Joint Research Collaboration Scheme and in part by the National Key Research and Development Program of China (2024YFF1206600), the National Natural Science Foundation of China (U21A20520, 62325204) and the Key-Area Research and Development Program of Guangzhou City (2023B01J1001).

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

References

1. Bernal, J., Sánchez, F.J., Fernández-Esparrach, G., Gil, D., Rodríguez, C., Vilariño, F.: Wm-dova maps for accurate polyp highlighting in colonoscopy: Validation vs. saliency maps from physicians. Computerized medical imaging and graphics **43**, 99–111 (2015)
2. Bernal, J., Tajkbaksh, N., Sanchez, F.J., Matuszewski, B.J., Chen, H., Yu, L., Angermann, Q., Romain, O., Rustad, B., Balasingham, I., et al.: Comparative validation of polyp detection methods in video colonoscopy: results from the miccai 2015 endoscopic vision challenge. IEEE transactions on medical imaging **36**(6), 1231–1249 (2017)
3. Chen, L.C., Zhu, Y., Papandreou, G., Schroff, F., Adam, H.: Encoder-decoder with atrous separable convolution for semantic image segmentation. In: Ferrari, V., Hebert, M., Sminchisescu, C., Weiss, Y. (eds.) Computer Vision – ECCV 2018. pp. 833–851. Springer International Publishing, Cham (2018)
4. Dou, Q., Castro, D.C., Kamnitsas, K., Glocker, B.: Domain generalization via model-agnostic learning of semantic features. Curran Associates Inc., Red Hook, NY, USA (2019)
5. Hu, S., Liao, Z., Xia, Y.: Devil is in channels: Contrastive single domain generalization for medical image segmentation. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. Springer (2023)
6. Huang, Z., Wang, H., Xing, E.P., Huang, D.: Self-challenging improves cross-domain generalization. In: Computer Vision – ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part II. p. 124–140. Springer-Verlag, Berlin, Heidelberg (2020). https://doi.org/10.1007/978-3-030-58536-5_8
7. Jha, D., Smedsrud, P.H., Riegler, M.A., Halvorsen, P., De Lange, T., Johansen, D., Johansen, H.D.: Kvasir-seg: A segmented polyp dataset. In: MultiMedia modeling: 26th international conference, MMM 2020, Daejeon, South Korea, January 5–8, 2020, proceedings, part II 26. pp. 451–462. Springer (2020)

8. Kamraoui, R.A., Ta, V.T., Tourdias, T., Mansencal, B., Manjon, J.V., Coupé, P.: Deeplesionbrain: Towards a broader deep-learning generalization for multiple sclerosis lesion segmentation. Medical Image Analysis **76**, 102312 (2022). https://doi.org/https://doi.org/10.1016/j.media.2021.102312, https://www.sciencedirect.com/science/article/pii/S1361841521003571
9. Orlando, J.I., Fu, H., Barbosa Breda, J., van Keer, K., Bathula, D.R., Diaz-Pinto, A., Fang, R., Heng, P.A., Kim, J., Lee, J., Lee, J., Li, X., Liu, P., Lu, S., Murugesan, B., Naranjo, V., Phaye, S.S.R., Shankaranarayana, S.M., Sikka, A., Son, J., van den Hengel, A., Wang, S., Wu, J., Wu, Z., Xu, G., Xu, Y., Yin, P., Li, F., Zhang, X., Xu, Y., Bogunović, H.: Refuge challenge: A unified framework for evaluating automated methods for glaucoma assessment from fundus photographs. Medical Image Analysis **59**, 101570 (2020). https://doi.org/https://doi.org/10.1016/j.media.2019.101570, https://www.sciencedirect.com/science/article/pii/S1361841519301100
10. Ouyang, C., Chen, C., Li, S., Li, Z., Qin, C., Bai, W., Rueckert, D.: Causality-inspired single-source domain generalization for medical image segmentation. IEEE Transactions on Medical Imaging **42**(4), 1095–1106 (2023). https://doi.org/10.1109/TMI.2022.3224067
11. Sandler, M., Howard, A., Zhu, M., Zhmoginov, A., Chen, L.C.: Mobilenetv2: Inverted residuals and linear bottlenecks. In: 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 4510–4520 (2018). https://doi.org/10.1109/CVPR.2018.00474
12. Sivaswamy, J., Chakravarty, A., Joshi, G.D., Syed, T.A.: A comprehensive retinal image dataset for the assessment of glaucoma from the optic nerve head analysis (2015), https://api.semanticscholar.org/CorpusID:29553360
13. 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 Thirty-Seventh AAAI Conference on Artificial Intelligence and Thirty-Fifth Conference on Innovative Applications of Artificial Intelligence and Thirteenth Symposium on Educational Advances in Artificial Intelligence. AAAI'23/IAAI'23/EAAI'23, AAAI Press (2023). https://doi.org/10.1609/aaai.v37i2.25332, https://doi.org/10.1609/aaai.v37i2.25332
14. Vázquez, D., Bernal, J., Sánchez, F.J., Fernández-Esparrach, G., López, A.M., Romero, A., Drozdzal, M., Courville, A.: A benchmark for endoluminal scene segmentation of colonoscopy images. Journal of healthcare engineering **2017**(1), 4037190 (2017)
15. 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 (2021), https://openreview.net/forum?id=BVSM0x3EDK6
16. Zhou, K., Yang, Y., Qiao, Y., Xiang, T.: Domain generalization with mixstyle. In: International Conference on Learning Representations (2021), https://openreview.net/forum?id=6xHJ37MVxxp