# Generalizable Weakly-Supervised Medical Image Segmentation with Inner-Outer Random Augmentation

No Author Given

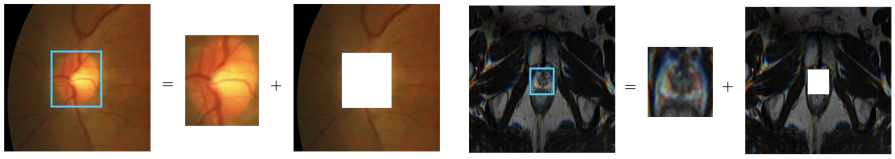
No Institute Given

**Abstract.** Domain generalization aims to reduce domain shift between domains to achieve better performance on the unseen target domain. However, most of the existing methods for domain generalization are based on fully-labeled source domains, which requires lots of manual operations for precisely labeling works. In this paper, we introduce the concept of weakly-supervised segmentation to domain generalization, which only bounding-box annotations of source domains to train the model. To address this task, we introduce a model with dual segment networks, each network is in charge of handing different augmentation strategies of random scaling and random flipping. The prediction from one of the networks are set to teach other for learning . We evaluate our model on several medical image segmentation tasks and achieve competitive results compared to the fully-supervised domain generalization methods.

**Keywords:** Weakly-supervised segmentation · Domain generalization · Inner-outer random augmentation

## 1 Introduction

Medical image segmentation has aroused lots of research interest. Recently, a new paradigm named domain generalization is becoming a hot topic for medical image segmentation. Domain generalization (DG) aims to solve the issue that the current deep network model trained for specific domains may not perform well when introduced to unseen target domains due to unexpected domain shifts. In a nutshell, the goal of domain generalization is to train a model with source domains to generalized well on targets. Many current works of domain generalization [8, 16, 18, 24] have achieved good performance in different aspects, such as medical images [14, 30]. However, for medical image segmentation tasks, it is labor-consuming to manually assign each pixel of an image a accurate label, and most current works of domain generalization are based on fully-supervised methods. In terms of segmentation, weakly-supervised segmentation (WSS) [19, 23] which significantly reduces labeling costs by using image-level labels or bounding boxes, is getting increasing attention. Therefore, we propose a new setting that draws on WSS's merit to offset the weakness of current DG methods.



(a) Fundus Image with Bounding Box, Foreground and Background

(b) Prostate Image with Bounding Box, Foreground and Background

**Fig. 1.** We can separate the image's inner part from its outer part with bounding box.

Traditional DG methods [3, 12, 29] mainly focus on how to extract the most domain-invariant features from existing domains. They believe that the domain-invariant features are general and transferable to different target domains. Furthermore, the information learned from source domains is limited, but performing simple and cheap augmentation operations to source domains can enhance the diversity of existing domains. Current DG methods have adopted some simple yet effective prepossessing-based augmentation methods such as MixUp [26], CutMix [25], or other random proceedings. We combine the augmentation with WSS and propose a novel augmentation strategy.

For WSS, several methods learned with image-level [13, 22, 23] labels mainly generate pseudo-labels with class activation map (CAM) and train a supervised segmentation model with the pseudo-labels. Some works based on bounding boxes [10, 19] rely on the regional proposal created by MCG [1] and GrabCut [17]. In this paper, we adopt a simple projection loss from BoxInst [19] to ensure the bounding boxes tightly enclose our model's predictions.

We propose a simple but effective weakly-supervised domain generalization strategy for medical image segmentation. For the WSS aspect, we adopt a projection loss from BoxInst [19], which helps ensure that the model limits the prediction's location. For the DG aspect, considering the advantage provided by bounding boxes, we can easily depart the inner and outer parts of an image with its bounding box. Therefore, we perform a random augmentation for the inner part and outer part of the image, respectively. We randomly adjust its brightness, contrast and saturation within a predetermined threshold. Therefore, we get an image of different styles of background and foreground. The proposed augmentation method greatly enriches the variety of the source domains, which triggers the model to capture the common features while ignoring the differences between augmented foreground and background.

In addition, the original images and augmented images will be trained together, and we compose a consistency loss to their predictions, which guarantees their prediction should be the same. And then, we introduce a module for decoding the feature generated by the backbone of the model. The decoder aims to reconstruct the original images from the features of both original images and augmented images because only the model with the ability to extract the

domain-invariant features can enable the decoder to restore the image from its feature.

In conclusion, our contribution can be summarized as follows:

1. We introduce the concept of weakly-supervised segmentation for domain generalization, which greatly saves the labor of manual labeling for traditional domain generalization methods while keeping the accuracy of segmentation.

2. We propose a novel augmentation strategy that fully uses the bounding boxes' characteristics by performing random augmentation for the foreground and background of the image, respectively, and a decoder to further promote the model's ability to extract domain-invariant features.

3. We conduct experiments on several multi-source domains datasets and achieve results comparable to state-of-art fully-supervised domain generalization methods with a significant reduction on the annotation.

## 2 Related Work

**Weakly-Supervised Segmentation.** Most of the works of WSS mainly use image-level labels [13, 22, 23] or boxes [7, 9, 10, 19] to supervise their training. For image-level labels, the most popular methods are learning representations from class activation maps and training a supervised model with generated pseudo-labels from CAM. For bounding boxes, traditional methods generate pseudo-labels by unsupervised methods like MCG [1] or GrabCut [17], such as Box-sup [5] gets the pseudo-labels from MCG. Box2seg [10] uses GrabCut to generate pseudo-labels and generate class specific attention maps to improve the prediction. The bounding box is also popular with weakly-supervised instance segmentation. Traditional methods like SDI [9]still use MCG to guide the model. Recent research BoxInst [19] uses projection loss to restrict the prediction and eliminate the need for sampling. Due to its lower annotation cost, we introduce the strategy of WSS for DG that requires extensive manual labeling.

**Domain Generalization.** Domain Generalization [8, 16, 18, 21, 24] aims to train a model with existing source domains and generalize the model on unseen target domains without extra training. Data augmentation [11, 18, 24] and learning domain-invariant features [3, 6, 12, 29] are the most popular procedure. Basic augmentation methods include flipping, random-scaling, and rotating. Recent works have proposed other effective augmentation methods, such as MixUp [26] and CutMix [25], which are already adopted by some works of DG. Current works of DG propose strategies such as generating the augmented image trough

the generative adversarial network [4] or conducting augmentation on the frequency spectrum [30]. We propose our novel augmentation strategy based on the setting of our work. For learning domain-invariant features, traditional works try to align the features of source domains by using MMD or adopting adversarial methods, which extract features from source domains and deceive discriminators in order to enhance the ability to extract domain-invariant features.

## 3 Approach

The overall framework of our proposed method is shown in Figure 2. We adopt projection loss to supervise the model with the bounding box, and we propose a new augmentation strategy that greatly enriches the variability of the source domains while introducing a shared decoder to reconstruct the source domains within the feature extracted by the shared encoder, which promotes the model's ability to extract domain-invariant features. The symbols are summarized in Table 1. We describe the framework in detail as follows:

### 3.1 Preliminary

**Projection Loss for Weakly-Supervised Segmentation.** For the weakly-ddsupervised aspect, we need to get predictions based on the bounding boxes which only provide us with location information. We adopt the projection loss from BoxInst [19]. The main function of the bounding box is to help us supervise the images' horizontal and vertical projection of the predicted mask. Therefore, the predicted masks can tightly match the bounding boxes. It is obvious that the ground-truth mask and bounding box mask should be the same when they project onto the x-axis and y-axis as shown in Figure 3, which means that the $x_1 = x_3(y_1 = y_3)$. Thus, by computing loss between the projection onto the x-axis and y-axis of ground-truth mask and the predicted mask, we can promise

**Table 1.** Summary of Symbols.

| Symbol | Notation |
|---|---|
| $S, S_1, S_2$ | Image set of source domains and their augmented forms |
| $B$ | Ground-truth bounding box mask set of $S$ |
| $P, P_1, P_2$ | Predictions of $S, S_1, S_2$ |
| $E$ | Shared Encoder |
| $F, F_1, F_2$ | Features of $S, S_1, S_2$ extracted by $E$ |
| $D$ | Decoder |
| $R, R_1, R_2$ | Reconstructions of $S, S_1, S_2$ decoded by $D$ |
| $C$ | Classifier |

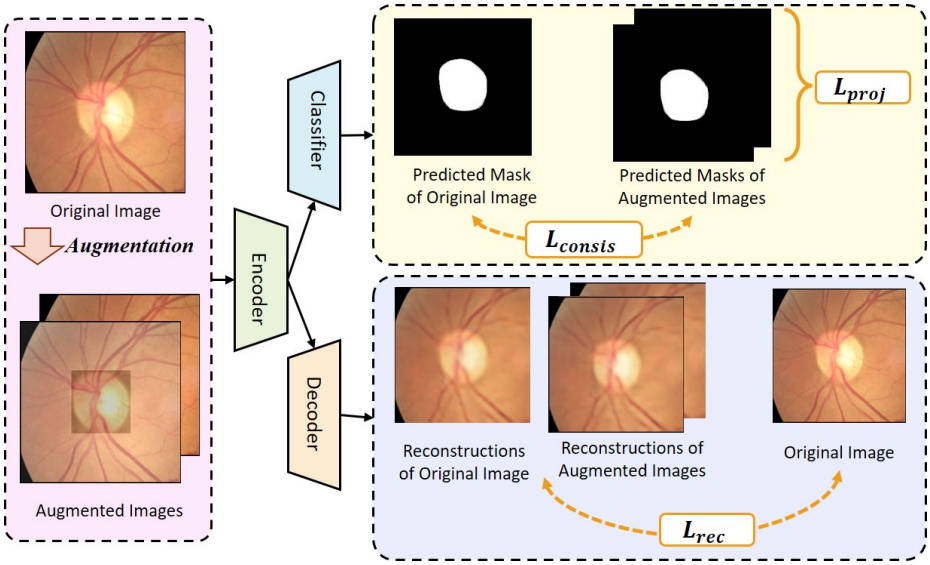



**Fig. 2.** The overall framework of our proposed method contains a shared encoder for extracting features, a classifier for prediction, and a decoder for reconstruction.

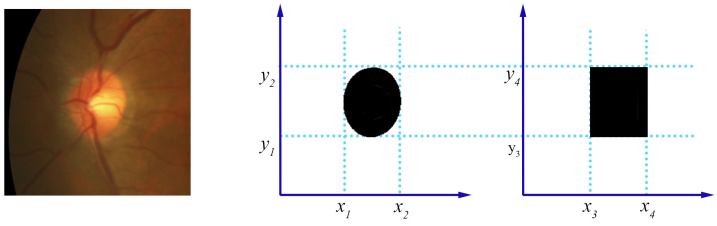


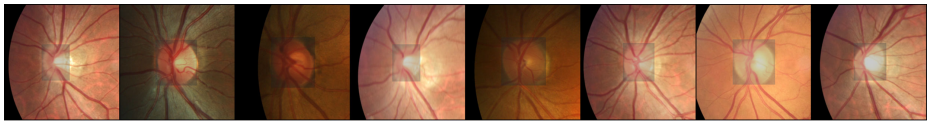

**Fig. 3.** The description of the projection loss. The ground-truth mask and the mask of the bounding box should have the same projection on the x and y axis.

that prediction will be enclosed by the bounding box. The projection loss can be summarized as:

$$L_{pro}^{S} = L_{Dice}(Proj_x(P), Proj_x(B)) + L_{Dice}(Proj_y(P), Proj_y(B)) \quad (1)$$

### 3.2 Inner-Outer Random Augmentation

We know that the bounding box can provide us with the projection information of the ground-truth mask for the image. Moreover, it easily helps us divide the image into the inner and outer parts. Therefore, we conduct a random augmentation strategy on the inner and outer part, respectively.



(a) Augmented Images

(b) Predictions of Preliminary Experiment

**Fig. 4.** There is a distinct boundary in augmented images. The predictions of the preliminary experiment show that the model almost ignores the influence of the boundary.

In detail, we randomly scale the brightness, contrast, and saturation of the source domain images within a certain threshold, which are the essential attribute of the image. After the augmentation, we get the augmented image of varying augmented degrees for its inner and outer parts. As shown in Figure 1, there is an obvious boundary between its foreground and background. We assume that the network may produce predicted masks that match such boundary because such prediction may have the same projection as the ground-truth mask, which will make the projection loss ineffective. However, We conduct a prior experiment with the augmentation strategy. The results show us that the model seems to ignore the existence of the boundary and stays more focused on the information, as the results shown in Figure 4. Therefore, we believe that the model with projection loss is able to discriminate part from the foreground of the image, and the method will trigger the model to extract the invariant feature from different augmented images.

With the random augmentation, we get plenty of the augmented images with different augmented degrees. Due to its randomness, we generate completely different augmented images every iteration of the training, which greatly enhances the richness of the source domains and pushes the model to learn the general feature of an image within its different versions.

For the experiment, we use a progressive strategy for our model. At the beginning of the training, we limit the threshold of the augmentation rate and gradually increase the rate. Therefore, the model will not be misled by the strong augmented image. For each batch of the images, we generate two different augmented images. Both the batch of original images and its two batches of augmented images will be sent to the model and computed their projection loss. Additionally, we add consistency loss between the prediction of the original image and its augmented image, which aims to ensure that the prediction of the augmented image is as same as the original one.

The overall loss can be described as:

$$L_{pro} = L_{pro}^{S} + \lambda_{aug}(L_{pro}^{S_1} + L_{pro}^{S_2}), \tag{2}$$

$$L_{consis} = L_{dis}(P, P_1) + L_{dis}(P, P_2), \tag{3}$$

where $L_{dis}$ is L1 norm.

### 3.3 Shared Decoder

Although our novel augmentation strategy helps promote the ability of the model to extract the invariant features from numerous augmented features, this is not enough to reach a satisfactory result. Therefore, we propose a new module of our network to increase the ability to learn domain-invariant features further. We assume that if the feature we learned from the augmented image is domain-invariant, it can be reconstructed to its original image. Therefore, we introduce a shared decoder to recover the original image.

In the last section, we have manipulated the original image to generate its two augmented forms. Each of them will be sent into the model to obtain its feature. Then we send these features into a decoder composed of a series of upsampling convolution layers. Finally, we add reconstruction loss between the reconstructed image of its augmented image and its original image.

$$L_{rec} = L_{dis}(S, R) + L_{dis}(S, R_1) + L_{dis}(S, R). \tag{4}$$

The overall loss item can be summarized as follows, the projection loss of the prediction of the original image and its augmented images, consistency loss between the prediction of the original image and its augmented ones, and the reconstruction loss between the original image and its reconstructed image from the feature of its augmented form.

$$L_{=}\lambda_{pro}L_{pro} + \lambda_{consis}L_{consis} + \lambda_{rec}L_{rec}, \tag{5}$$

where $\lambda_{pro}, \lambda_{consis}, \lambda_{rec}$ are hyper-parameters.

## 4 Experiment

In this section, we evaluate the effectiveness of our proposed methods from both qualitative and quantitative perspectives. Specifically, we compare our method with several state-of-the-art methods and performed several experiments to prove the effectiveness of our proposed module on model performance.

**Fig. 5.** Some qualitative results of our proposed method.

### 4.1 Datasets

Our experiment is conducted on the retinal fundus images (Fundus) [20] and Prostate T2 weighted magnetic resonance imaging (MRI) dataset (Prostate) [15]. The Fundus dataset for the segmentation of disc and cup is from four different clinical centers. The prostate dataset has six domains from different sources.

### 4.2 Implantation Details

All the settings of hyper-parameters of different generalization directions are kept the same. The default batch size is set to 8, the learning rate is set to 0.0001 and gradually reduced in training. The $\lambda_{aug}$ is set as 0.5, and other hyper-parameters are set to default 1. We adopt Deeplabv3 with Resnet50 as the backbone for our model. For Fundus, it takes 1.5 G of GPU memory and about 2 hours per training for one generalization direction. For Prostate, it takes about 4 GPUs and 3 hours per training.

### 4.3 Evaluation Metrics

To evaluate our experimental results, we use two commonly used metrics: Dice similarity coefficient (Dice) and average surface distance (ASD) to quantitatively evaluate the segmentation results of the entire object region and surface shape, respectively. Dice is used to calculate how similar the predicted results are to the actual labels. The higher the Dice value means better segmentation performance. ASD is introduced to measure the average distance between the predicted results and the actual labels. The lower the ASD value means better performance.

**Table 2.** Comparison of Dice of different fully-supervised domain generalization methods and weakly-supervised settings on Fundus dataset.

| Type | Task | Optic Cup | | | | | Optic Disc | | | | | Overall |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Target | A | B | C | D | Avg. | A | B | C | D | Avg. | |
| WSS | Baseline | 76.01 | 74.80 | 84.87 | 78.97 | 78.66 | 93.57 | 85.45 | 93.48 | 88.24 | 90.18 | 84.42 |
| | Normal Aug. | 77.92 | **76.74** | 84.24 | **80.80** | 79.93 | 94.34 | **87.50** | 94.21 | **90.10** | **91.53** | 85.73 |
| | Ours | **79.87** | 76.54 | **85.32** | 80.51 | **80.56** | **95.52** | 86.31 | **94.50** | 89.40 | 91.43 | **85.99** |
| Upperbound | | | | | | | | | | | | |
| FSS | JiGen[2] | 80.81 | 79.46 | 82.65 | 84.30 | 81.81 | 95.03 | 90.47 | 91.94 | 91.06 | 92.13 | 86.97 |
| | BigAug[28] | 81.62 | 69.46 | 82.64 | 84.51 | 79.55 | 93.49 | 86.18 | 92.09 | 93.67 | 91.36 | 85.46 |
| | DST[27] | 75.63 | 80.80 | 84.32 | 86.24 | 81.75 | 92.20 | 90.77 | 94.02 | 90.66 | 91.91 | 86.83 |
| | FedDG[14] | 84.13 | 71.88 | 83.94 | 85.51 | 81.37 | 95.37 | 87.52 | 93.37 | 94.50 | 92.69 | 87.03 |

**Table 3.** Comparison of ASD of different fully-supervised domain generalization methods and weakly-supervised settings on Fundus dataset.

| Type | Task | Optic Cup | | | | | Optic Disc | | | | | Overall |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Target | A | B | C | D | Avg. | A | B | C | D | Avg. | |
| WSS | Baseline | 24.30 | 19.10 | 10.60 | 14.43 | 17.11 | 11.24 | 21.71 | 9.33 | 14.47 | 14.19 | 15.65 |
| | Normal Aug. | 23.24 | **17.15** | 10.93 | **11.74** | 15.77 | 9.97 | **18.72** | **8.31** | **11.76** | **12.19** | 13.97 |
| | Ours | **21.23** | 17.58 | **10.36** | 12.07 | **15.31** | **8.13** | 20.32 | 8.35 | 12.83 | 12.41 | **13.86** |
| Upperbound | | | | | | | | | | | | |
| FSS | DST[27] | 24.42 | 12.89 | 10.91 | 7.05 | 13.82 | 13.24 | 14.00 | 8.52 | 10.05 | 11.45 | 12.31 |
| | JiGen[2] | 19.56 | 13.99 | 11.90 | 8.92 | 13.60 | 8.55 | 14.09 | 11.35 | 12.57 | 11.64 | 12.62 |

### 4.4 Comparison with State-of-art Fully-Supervised Domain Generalization Methods

In our experiment, we follow the other fully-supervised domain generalization methods' setting that training on k-1 distributed source domains, and finally testing on the excluded invisible target domain. Therefore, the Fundus segmentation task has 4 generalization directions, and the prostate MRI segmentation task has 6 generalization directions. Results of other methods in this paper are collected in FedDG [14].

We compare our methods with the following recent fully-supervised domain generalization approach: JiGen [2]: Learning domain invariant representation information through self-supervised learning. BigAug [28] : A method to regularize general representation learning through mass data transformation. FedDG [14]: An improvement in federated learning enables it to be applied to domain generalization segmentation tasks. DST [15]: Uses a series of data augmentation strategies for domain generalization, including random sharpening, blurring, noise, brightness adjustment, contrast change, disturbance, rotation, scaling, deformation and cropping. SAML [15]: Meta-learning based on gradient explicitly simulates domain transfer through the use of virtual meta-training and meta-testing in the training process.

**Table 4.** Comparison of dice of different fully-supervised domain generalization methods and weakly-supervised settings on Prostate Dataset.

| Type | Task | Prostate MRI Segmentation | | | | | | Overall |
|---|---|---|---|---|---|---|---|---|
| | Target | A | B | C | D | E | F | |
| WSS | Baseline | 84.64 | 89.07 | 83.47 | 87.56 | 86.75 | 87.65 | 86.52 |
| | Normal Aug. | 84.38 | **89.71** | **86.56** | 85.27 | 87.22 | **89.27** | 87.06 |
| | Ours | **85.16** | 89.29 | 85.59 | **89.44** | **88.23** | 88.64 | **87.73** |
| Upperbound | | | | | | | | |
| FSS | JiGen[2] | 89.95 | 85.81 | 84.06 | 87.34 | 81.32 | 89.11 | 86.26 |
| | BigAug[28] | 88.62 | 86.22 | 83.76 | 87.35 | 85.53 | 85.83 | 86.21 |
| | SAML[15] | 89.66 | 87.53 | 84.43 | 88.67 | 87.37 | 88.34 | 87.67 |
| | FedDG[14] | 90.19 | 87.17 | 85.26 | 88.23 | 83.02 | 90.47 | 87.39 |

**Table 5.** Comparison of dice of different fully-supervised domain generalization methods and weakly-supervised settings on Prostate Dataset.

| Type | Task | Prostate MRI Segmentation | | | | | | Overall |
|---|---|---|---|---|---|---|---|---|
| | Target | A | B | C | D | E | F | |
| WSS | Baseline | 1.29 | 0.84 | 4.10 | 0.92 | 1.75 | 1.62 | 1.75 |
| | Normal Aug. | 1.18 | **0.75** | **3.40** | 1.90 | 2.26 | **0.77** | 1.70 |
| | Ours | **1.12** | 0.85 | 4.33 | **0.75** | **1.61** | 0.85 | **1.58** |
| Upperbound | | | | | | | | |
| FSS | BigAug[28] | 1.70 | 1.56 | 2.72 | 1.98 | 1.90 | 1.75 | 1.93 |
| | SAML[15] | 1.38 | 1.46 | 2.07 | 1.56 | 1.77 | 1.22 | 1.58 |

In Table 2 and Table 3, we report the results of the optic cup and optic disc segmentation tasks on the Fundus dataset. Although we still have small gap on some generalization with fully-supervised methods in some directions, in specific directions such as optic disc segmentation for domain A and D we reach dice of 95.52% and 94.50% respectively, and for optic cup segmentation for domain C achieve the dice of 85.32%. The performance of ASD is about the same as Dice. In general, our weakly-supervised method is only about 1 percent of Dice behind the latest fully-supervised method, which only uses the bounding box for supervision instead of ground-truth masks. Compared to other experiments of weakly-supervised settings, both the performance of dice and ASD are ahead of other WSS settings on average.

Our performance on the Prostate dataset is better as the results shown in Table 4 and Table 5. The performance on most of the generalization directions is almost equal to even higher than the SOTA fully-supervised methods, especially for domain B and E, which reaches Dice of 89.29% and 88.23%. The average performance is higher than current methods both for Dice and ASD, 87.73% and 1.54 voxels, respectively. Our method shows a great advantage compared to other experiments of weakly-supervised settings on the Prostate dataset, reaching dice of 87.73% and ASD of 1.58%.

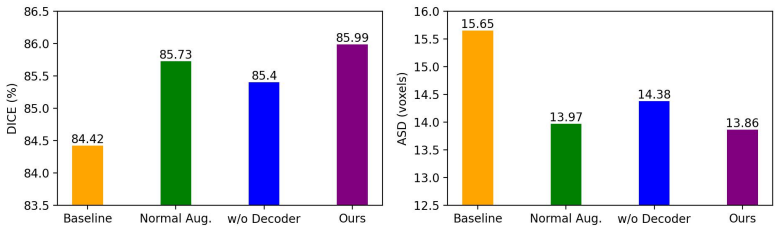



**Fig. 6.** The results of different settings for Fundus dataset.

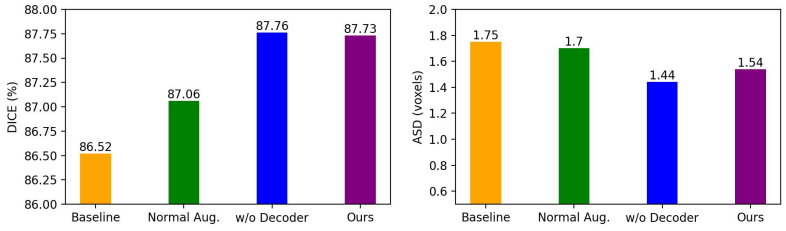



**Fig. 7.** The results of different settings for Prostate dataset.

## 5 Conclusion

In this paper, we propose a novel concept for current research of domain generalization and combine it with weakly-supervised segmentation. We introduce a novel data augmentation strategy that perfectly matches the bounding box attribute. Moreover, we introduce a shared decoder to help the model extract domain-invariant feature. Through a series of experiment, we prove the effectiveness of all the modules and show significant advantages compared to other labor-consuming fully-supervised domain generalization methods.

## References

1. Arbeláez, P., Pont-Tuset, J., Barron, J.T., Marques, F., Malik, J.: Multiscale combinatorial grouping. In: IEEE Conference on Computer Vision and Pattern Recognition. pp. 328–335 (2014)
2. Carlucci, F.M., D'Innocente, A., Bucci, S., Caputo, B., Tommasi, T.: Domain generalization by solving jigsaw puzzles. In: IEEE Conference on Computer Vision and Pattern Recognition. pp. 2229–2238 (2019)
3. Choi, S., Jung, S., Yun, H., Kim, J.T., Kim, S., Choo, J.: Robustnet: Improving domain generalization in urban-scene segmentation via instance selective whitening. In: IEEE Conference on Computer Vision and Pattern Recognition. pp. 11580–11590 (2021)
4. Creswell, A., White, T., Dumoulin, V., Arulkumaran, K., Sengupta, B., Bharath, A.A.: Generative adversarial networks: An overview. IEEE Signal Processing Magazine **35**(1), 53–65 (2018)
5. Dai, J., He, K., Sun, J.: Boxsup: Exploiting bounding boxes to supervise convolutional networks for semantic segmentation. In: IEEE International Conference on Computer Vision. pp. 1635–1643 (2015)
6. Gong, R., Li, W., Chen, Y., Gool, L.V.: Dlow: Domain flow for adaptation and generalization. In: IEEE Conference on Computer Vision and Pattern Recognition. pp. 2477–2486 (2019)
7. Hsu, C.C., Hsu, K.J., Tsai, C.C., Lin, Y.Y., Chuang, Y.Y.: Weakly supervised instance segmentation using the bounding box tightness prior. In: Advances in Neural Information Processing Systems. vol. 32 (2019)

8. Huang, J., Guan, D., Xiao, A., Lu, S.: Fsdr: Frequency space domain randomization for domain generalization. In: IEEE Conference on Computer Vision and Pattern Recognition. pp. 6891–6902 (2021)
9. Khoreva, A., Benenson, R., Hosang, J., Hein, M., Schiele, B.: Simple does it: Weakly supervised instance and semantic segmentation. In: IEEE Conference on Computer Vision and Pattern Recognition. pp. 876–885 (2017)
10. Kulharia, V., Chandra, S., Agrawal, A., Torr, P., Tyagi, A.: Box2seg: Attention weighted loss and discriminative feature learning for weakly supervised segmentation. In: European Conference on Computer Vision. pp. 290–308 (2020)
11. Li, P., Li, D., Li, W., Gong, S., Fu, Y., Hospedales, T.M.: A simple feature augmentation for domain generalization. In: IEEE International Conference on Computer Vision. pp. 8886–8895 (2021)
12. Li, Y., Tian, X., Gong, M., Liu, Y., Liu, T., Zhang, K., Tao, D.: Deep domain generalization via conditional invariant adversarial networks. In: European Conference on Computer Vision. pp. 624–639 (2018)
13. Li, Y., Kuang, Z., Liu, L., Chen, Y., Zhang, W.: Pseudo-mask matters in weakly-supervised semantic segmentation. In: IEEE International Conference on Computer Vision. pp. 6964–6973 (2021)
14. 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: IEEE Conference on Computer Vision and Pattern Recognition. pp. 1013–1023 (2021)
15. 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)
16. Robey, A., Pappas, G.J., Hassani, H.: Model-based domain generalization. In: Advances in Neural Information Processing Systems. vol. 34, pp. 20210–20229 (2021)
17. Rother, C., Kolmogorov, V., Blake, A.: " grabcut" interactive foreground extraction using iterated graph cuts. ACM transactions on graphics **23**(3), 309–314 (2004)
18. Shu, Y., Cao, Z., Wang, C., Wang, J., Long, M.: Open domain generalization with domain-augmented meta-learning. In: IEEE Conference on Computer Vision and Pattern Recognition. pp. 9624–9633 (2021)
19. Tian, Z., Shen, C., Wang, X., Chen, H.: Boxinst: High-performance instance segmentation with box annotations. In: IEEE Conference on Computer Vision and Pattern Recognition. pp. 5443–5452 (2021)
20. 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)
21. Wang, Z., Wang, Z., Yu, Z., Deng, W., Li, J., Gao, T., Wang, Z.: Domain generalization via shuffled style assembly for face anti-spoofing. In: IEEE Conference on Computer Vision and Pattern Recognition. pp. 4123–4133 (2022)
22. Wu, T., Huang, J., Gao, G., Wei, X., Wei, X., Luo, X., Liu, C.H.: Embedded discriminative attention mechanism for weakly supervised semantic segmentation. In: IEEE Conference on Computer Vision and Pattern Recognition. pp. 16765–16774 (2021)
23. Xu, L., Ouyang, W., Bennamoun, M., Boussaid, F., Sohel, F., Xu, D.: Leveraging auxiliary tasks with affinity learning for weakly supervised semantic segmentation. In: IEEE International Conference on Computer Vision. pp. 6984–6993 (2021)
24. Xu, Q., Zhang, R., Zhang, Y., Wang, Y., Tian, Q.: A fourier-based framework for domain generalization. In: IEEE Conference on Computer Vision and Pattern Recognition. pp. 14383–14392 (2021)

25. Yun, S., Han, D., Oh, S.J., Chun, S., Choe, J., Yoo, Y.: Cutmix: Regularization strategy to train strong classifiers with localizable features. In: IEEE International Conference on Computer Vision. pp. 6023–6032 (2019)
26. Zhang, H., Cisse, M., Dauphin, Y.N., Lopez-Paz, D.: mixup: Beyond empirical risk minimization. In: International Conference on Learning Representations (2018)
27. Zhang, L., Wang, X., Yang, D., Sanford, T., Harmon, S., Turkbey, B., Roth, H., Myronenko, A., Xu, D., Xu, Z.: When unseen domain generalization is unnecessary? rethinking data augmentation. arXiv preprint arXiv:1906.03347 (2019)
28. 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)
29. Zhao, S., Gong, M., Liu, T., Fu, H., Tao, D.: Domain generalization via entropy regularization. In: Advances in Neural Information Processing Systems. vol. 33, pp. 16096–16107 (2020)
30. Zhou, Z., Qi, L., Yang, X., Ni, D., Shi, Y.: Generalizable cross-modality medical image segmentation via style augmentation and dual normalization. In: IEEE Conference on Computer Vision and Pattern Recognition. pp. 20856–20865 (2022)