Generalizable Weakly-Supervised Medical Image Segmentation via Mutual Supervision

Huaize Ye, Lei Qi, and Yinghuan Shi

Nanjing University

Abstract. Domain generalization, which aims to reduce domain shift between domains to achieve promising performance on the unseen target domain, has been widely practiced in medical image segmentation. Although this technique has made considerable progress, existing methods rely on fully-supervised methods with pixel-wise labels, which consume significant manpower for annotation. In this paper, we introduce the concept of generalizable weakly-supervised segmentation and train the model with bounding-box annotations of source domains only. To address this task, we propose a model with a dual-branch augmentation and mutual supervision structure. The source images are augmented in two different strategies and sent to two segmentation networks, respectively, which enrich the source domains and force the network to learn different views of augmented features. The prediction from the network is used to teach the other one through mutual supervision for knowledge sharing, promoting the performance of generalization. We evaluate our model on several medical image segmentation tasks and achieve competitive results compared to its upper-bound, *i.e.*, fully-supervised domain generalization methods.

Keywords: Weakly-supervised segmentation · Weakly supervised domain generalization · Mutual supervision · Dual-branch augmentation

1 Introduction

Medical image segmentation is critical for computer-aided diagnosis and treatment planning. However, the current deep network models trained for specific domains may not perform well when introduced to new unseen target domains because of unpredictable domain shift. To address this challenge, domain generalization (DG) [9, 17, 20, 26] has become a popular research topic in various segmentation tasks. The objective of DG is to train a model with source domains that can generalize well on unseen target domains. Several recent works have achieved good performance in medical image segmentation [8, 15, 16, 28, 29], however, for medical image segmentation tasks, fully-supervised methods are supervised by pixel-level annotation and making these labels is labor-intensive and time-consuming, which limits the scalability and applicability of DG.

To mitigate the limitations of fully-supervised methods, weakly-supervised segmentation (WSS) [21, 22, 25] has gained increasing attention in recent years.

WSS significantly reduces labeling costs by using weakly-supervised label like image-level labels or bounding-boxes. Therefore, we propose a new setting that draws on WSS's merit to offset the weakness of current DG methods for medical image segmentation.

Data augmentation [12, 20, 26] and learning domain-invariant features [3, 5, 13, 27] are the most popular procedure for DG. So far, some works have applied domain generalization on medical image segmentation. For instance, Liu *et al.* [16] proposed a methods based on meta-learning by encouraging the shape compactness and shape smoothness for prostate segmentation. Wang *et al.* [23] developed a Domain Knowledge Pool to learn and memorize the prior information extracted from multi-source domain, which helps to augment original image. Hu *et al.* [8]designed a domain adaptive convolution module and content adaptive convolution module and incorporate both into an encoder-decoder backbone. Zhou *et al.* [29] introduced a domain-specific restoration module for regularization and Random Amplitude Mixup module with low-level frequency knowledge. However, the field of bounding weakly-supervised segmentation and domain generalization combined has hardly been explored. Hence, we propose the concept of generalizable weakly-supervised segmentation.

For WSS, most works mainly use image-level labels [14, 24, 25] or bounding-boxes [7, 10, 11, 22, 21] for training. For image-level labels, the most popular methods are learning representations from class activation maps (CAM) 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 [18], such as Dai *et al.* [4] got the pseudo-labels from MCG. Kulharia *et al.* [11] utilized GrabCut to generate pseudo-labels and generate class specific attention maps to improve the prediction. The bounding-box is also popular in instance segmentation. For example, Tian *et al.* [21] proposed projection loss to restrict the prediction and eliminate the need for sampling. For medical image, Wang *et al.* [22] introduced a generalized MIL formulation and smooth maximum approximation to integrate the bounding-box tightness prior into the network for medical images. Given the limited research on WSS for medical image segmentation and the lack of investigation into weakly-supervised domain generalization, it is challenging to develop a model that combines these two approaches.

In this study, we propose a novel architecture for learning generalizable knowledge in segmentation using bounding-boxes. We identify two main directions of augmentation that benefit performance, scaling and flipping, which simulate size variations and symmetry feature in medical images, respectively. We introduce the dual-branch augmentation strategy, which leverages these two types of augmentation to improve generalization without additional parameters. We also present a weakly-supervised mutual supervision model that effectively matches the dual-branch augmentation. Specifically, augmented images are segmented by two networks, and each network's prediction is restricted by weakly-supervised segmentation loss and supervised by the prediction from the other network. Our proposed approach is evaluated on two segmentation tasks and

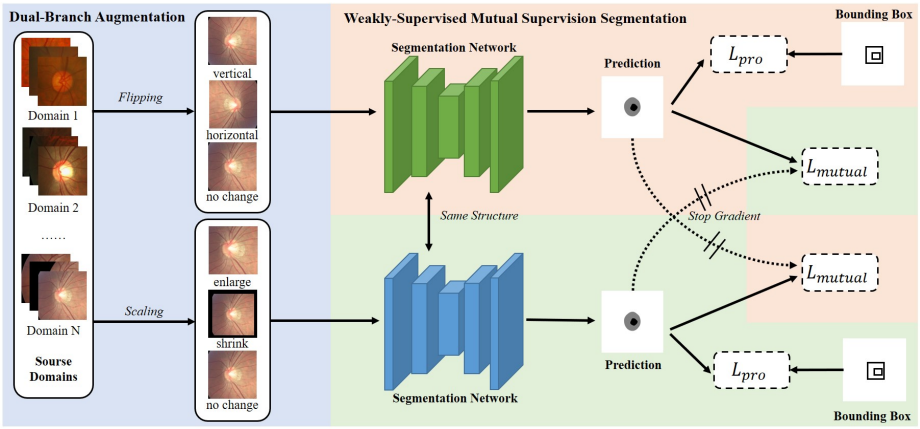

Fig. 1. The overall framework of our proposed method contains three main steps: a) Dual-branch Augmentation, b) Dual Segmentation Networks, C) Weakly-supervised Mutual Supervision. First, we augment the images of source domains in two strategy: flipping and scaling. Each group of augmentation is sent to its network, respectively. Then we get two predictions containing different knowledge, they are supervised by bounding-boxes and the other prediction.

achieve the competitive performance compared to its upper-bound (*i.e.* fully-supervised methods), despite using only bounding-box annotations.

2 Methodology

The methodology of our approach comprises three main components, namely, dual-branch augmentation, segmentation, and mutual supervision, as depicted in Fig. 1. To begin with, we employ two distinct strategies on the original images to achieve dual-branch augmentation. The first strategy involves flipping the images vertically or horizontally or keeping them unchanged with different possibilities, inspired by the symmetry commonly found in medical images. To simulate the size variation, the second strategy involves randomly shrinking, enlarging, or leaving the images unchanged, akin to the first strategy but different in terms of augmentation form. Then we have two identical networks for segmentation that both models have same structure but they are responsible for segmenting images of different augmentation during training. After segmentation, two predictions from the segmentation networks are supervised by bounding-boxes and the other prediction. We describe the framework in detail as follows:

Projection Loss for Weakly-Supervised Segmentation. For the weakly-supervised aspect, we obtain predictions based on the bounding-boxes, which only provide us with location information. To achieve this, we adopt the projection loss from BoxInst [21]. The main function of the bounding-box is to help us supervise the 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. 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 that prediction will be enclosed by the bounding-box. The projection loss can be described as:

$$\mathcal{L}_{pro}^{M} = \mathcal{L}_{Dice}(Proj_x(P_M), Proj_x(B)) + \mathcal{L}_{Dice}(Proj_y(P_M), Proj_y(B)), \quad (1)$$

where P_M is the prediction from segmentation network M, B is the bounding-box, $Proj_x$ is the projection an the x-axis, $\mathcal{L}_{Dice}$ is Dice loss. In this paper, the predictions from both segmentation networks are required to be supervised by bounding-boxes. The overall projection loss item can be summarized as:

$$\mathcal{L}_{pro} = \mathcal{L}_{pro}^{M} + \mathcal{L}_{pro}^{N}. \quad (2)$$

Dual-branch Augmentation. We introduce a novel augmentation strategy. We divide the original images from source domains into two branches by applying different augmentation strategy. This operation not only enhances the richness of the source domains but also compels the network to learn distinct features from diverse augmentations. In this paper, we choose scaling and flipping for the augmentation due to their simplicity.

The first strategy is randomly flipping the images. Given that most medical images exhibit axial symmetry, flipping enhances the features of the original images and improves the model's generalization ability. Each batch of images from the source domains may undergo one of three actions with varying probabilities: vertical flipping, horizontal flipping, or remaining unchanged. The details of these probabilities are presented in the experiment section.

The second strategy involves randomly scaling the images. Medical images may differ in size due to varying sampling distances or organ sizes. By randomly scaling the images, we can simulate these difference and improve the network's robustness when dealing with various sizes of segmentation parts. For actual operation, we center-crop the image and resize it to its original size as the shrink action, and pad the original image and resize it as the enlarge action. Similar to the first action, the images are enlarged, shrunk, or remain the same with varying possibilities.

Both types of augmentation are straightforward and have been proven effective in previous studies [19]. They do not introduce additional parameters and promote the generalization ability of the model effectively.

Segmentation via Mutual Supervision. Inspired by the work on semi-supervised segmentation by Chen *et al*. [2], which employs two networks with identical architectures and enforces constraints such that the outputs of the two networks for the same example are similar, we propose a mutual supervision model with cross-teaching procedure.

For the segmentation task, we use two identical networks with the same structure but different initializations. Each network is responsible for the images

augmented by one type of augmentation, respectively. The two networks produce two predictions, which are supervised by the projection loss and also by each other via cross-entropy loss. This mutual supervision operation helps the models to break the limited feature zones and absorb new information from each other. The two networks teach each other to improve their performance while promoting their ability to generalize.

The mutual-supervision loss is defined as follows:

$$\mathcal{L}_{mutual} = \mathcal{L}_{CE}(Pred_M(P), Pred_N^*(P)) + \mathcal{L}_{CE}(Pred_N(P), Pred_M^*(P)), \quad (3)$$

where $\mathcal{L}_{CE}$ is the cross-entropy loss, $Pred_M$ is the prediction from network M, and * indicates that the gradient is not included.

The overall mutual supervision loss is defined as:

$$\mathcal{L} = \mathcal{L}_{mutual} + \mathcal{L}_{pro}, \quad (4)$$

where, $\mathcal{L}_{pro}$ is the projection loss described in the first section of our methodology. The combination of these two losses enables our model to achieve robust and accurate segmentation results.

3 Experiment

Dataset. We evaluate our method on two public DG medical image segmentation datasets: Fundus [23] and Prostate [16]. Both of them have been widely adopted in fully-supervised DG methods [29, 8]. The Fundus dataset contains 4 domains of retinal fundus images from 4 different medical centers, and it has two segmentation task: cup segmentation and disc segmentation. For preprocessing, We adopt the same pre-processing operation as in RAM [29]. The Prostate dataset contains 6 domains of T2-weighted MRI prostate images which are from 6 different data sources. The images from prostate dataset have been cropped to 3D prostate region and 2D slices in axial plane have been resized to 384×384.We extract the bounding-box information from it ground-truth labels for both datasets to supervise our model.

Implantation Details and Evaluation Metrics. In the implementation details, we keep all the settings of hyper-parameters of different generalization directions the same. The default batch size is set to 8, and we employ the Adam optimizer with an initial learning rate of 0.0001 to optimize our model. All the coefficients of losses are set to default 1. We adopt Deeplabv3 with Resnet50 [6] as the backbone for our model, and the training epoch for Fundus was 100, and for Prostate, it was 50. The possibilities for choosing different augmentations are 40% for each augmentation, 20% for no processing, except for scaling in Fundus, which was 20% for enlarging, 60% for shrinking, and 20% for no processing. For Fundus, it took 8 GB of GPU memory and about 2 hours per training for one generalization direction, while for Prostate, it took about 11 GB and 3 hours per training.

Table 1. Comparison of DICE of different fully-supervised domain generalization methods and weakly-supervised settings on Fundus dataset. * means replace its segmentation loss with projection loss.

Task	Optic Cup/Disc Segmentation (Dice)				Avg.
Unseen Site	Domain 1	Domain 2	Domain 3	Domain 4	
Weakly-Supervised					
Baseline	80.87/94.62	78.63/86.84	86.24/93.08	81.46/90.63	86.54
RAM*[29]	82.29/90.58	60.33/81.16	84.81/88.65	79.34/85.15	81.54
DoFE*[23]	78.98/95.26	**81.81/91.40**	85.88/92.79	**86.86**/89.43	87.80
Ours	**85.43/95.65**	81.79/87.70	**86.59/94.33**	85.67/**93.49**	**88.83**
Fully-Supervised (Upper-bound)					
FedDG[15]	81.72/95.62	77.87/88.71	83.96/94.83	81.90/93.37	87.25
DoFE[23]	84.17/94.96	81.03/89.29	86.54/91.67	87.28/93.04	88.50
DCAC[8]	81.43/96.54	77.72/87.85	86.80/94.28	87.68/95.40	88.47
RAM[29]	85.48/95.75	78.82/89.43	87.44/94.67	85.84/94.10	88.94

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

Task	Optic Cup/Disc Segmentation (ASD)				Avg.
Unseen Site	Domain 1	Domain 2	Domain 3	Domain 4	
Weakly-Supervised					
Baseline	20.24/10.30	14.90/19.08	9.71/9.88	11.14/11.15	13.30
RAM* [29]	18.89/16.45	37.35/27.68	10.92/15.98	13.01/18.71	19.88
DoFE* [23]	21.83/8.23	**12.48/12.45**	10.09/10.35	**7.24**/11.51	11.77
Ours	**15.04/7.64**	12.51/18.15	**9.58/8.12**	8.37/**7.35**	**10.85**
Fully-Supervised (Upper-Bound)					
FedDG[15]	18.57/7.69	15.87/16.93	11.09/7.28	10.23/7.51	11.90
DoFE[23]	16.07/7.18	13.44/17.06	10.12/10.75	8.14/7.29	11.26
DCAC[8]	19.20/6.35	17.15/18.28	9.14/8.11	7.12/5.20	11.32
RAM[29]	16.05/7.12	14.01/13.86	9.02/7.11	8.29/7.06	10.32

To evaluate experimental results, we adopt two commonly used metrics: Dice 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, the better the segmentation performance. ASD is introduced to measure the average distance between the predicted results and the actual labels. The lower the ASD value, the better the performance.

Comparison with Other Methods. In our experiment, we followed the setting of other fully-supervised domain generalization methods [23, 29] by training on k-1 distributed source domains and finally testing on the excluded invisible target domain. As a result, the Fundus segmentation task had 4 generalization directions, and the prostate MRI segmentation task had 6 generalization directions. The results of other methods in this paper were collected in RAM [29].

Table 3. Comparison of Dice of different fully-supervised domain generalization methods and weakly-supervised settings on Prostate dataset.

Task	Prostate MRI Segmentation (Dice)						Avg.
Unseen Site	Domain 1	Domain 2	Domain 3	Domain 4	Domain 5	Domain 6	
Weakly-Supervised							
Baseline	86.55	89.10	83.99	89.20	86.83	89.41	87.51
RAM* [29]	86.17	89.30	84.56	89.30	85.46	88.79	87.26
Ours	**87.85**	**90.33**	**86.00**	**89.39**	**89.02**	**89.57**	**88.69**
Fully-Supervised (Upper-Bound)							
SAML [16]	86.35	90.18	85.03	88.20	86.97	87.69	87.40
FedDG [15]	86.43	89.59	85.30	88.95	85.93	87.39	87.27
DoFE [23]	89.64	87.56	85.08	89.06	86.15	87.03	87.42
DCAC [8]	91.76	90.51	86.30	89.13	83.39	90.56	88.61
RAM [29]	87.56	90.20	86.92	88.72	87.17	87.93	88.08

Table 4. Comparison of ASD of different fully-supervised domain generalization methods and weakly-supervised settings on Prostate dataset.

Task	Prostate MRI Segmentation (ASD)						Avg.
Unseen Site	Domain 1	Domain 2	Domain 3	Domain 4	Domain 5	Domain 6	
Weakly-Supervised							
Baseline	0.99	**0.76**	4.30	0.75	1.68	0.73	1.53
RAM* [29]	1.07	0.83	3.48	0.75	1.73	0.77	1.44
Ours	**0.86**	0.81	**2.84**	**0.75**	**1.43**	**0.68**	**1.22**
Fully-Supervised (Upper-Bound)							
SAML [16]	1.09	1.54	2.52	1.41	2.01	1.77	1.72
FedDG [15]	1.30	1.67	2.36	1.37	2.19	1.94	1.81
DoFE [23]	0.92	1.49	2.74	1.46	1.89	1.53	1.68
DCAC [8]	0.98	0.89	1.77	1.53	2.46	0.85	1.41
RAM [29]	1.04	0.81	2.23	1.16	1.81	1.15	1.37

We compared our approach with five state-of-the-art fully-supervised DG methods as the reference of our upper-bound. SAML [16] and FedDG [15] are based on meta-learning; DoFE [23] and DCAC [8] are based on domain-invariant feature learning. RAM [29] is an augmentation-based method. We further replaced the segmentation loss in RAM [29] and DoFE [23] with the projection loss as a comparison for the weakly-supervised setting. Meanwhile, we trained the model only with the projection loss as our baseline model.

In Table 1 and Table 2, we report the results of the optic cup and optic disc segmentation tasks on the Fundus dataset. We almost eliminated the gap with the fully-supervised methods, as our method reached a Dice coefficient of 88.83%, and the best result of fully-supervised DG was 88.94%. Compared to other weakly-supervised settings, we achieved the best performance on most unseen domains. On average, our performance was 2.3%, 7.3%, and 1% higher than the baseline, RAM [29], and DoFE [23], respectively. The outcome of ASD

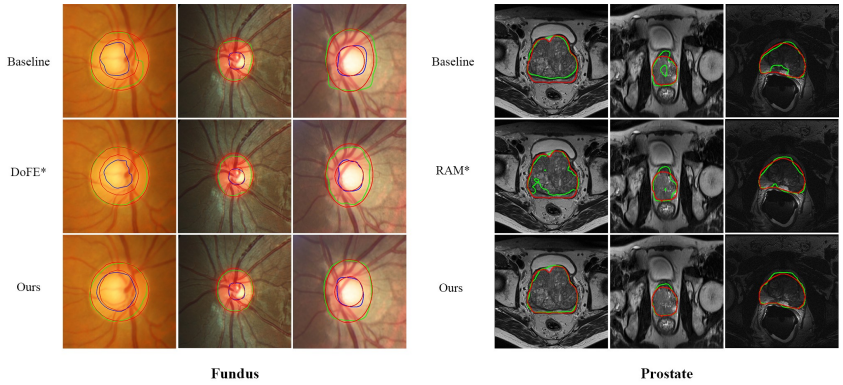

Fig. 2. Visualization on segmentation results of different settings on Fundus and Prostate datasets. The red contours indicate the boundaries of ground truths while the green and blue contours are predictions.

was about the same as Dice, which achieved consistent improvement compared to the baseline model.

The performance on the Prostate dataset is shown in Table 3 and Table 4. The average results for both Dice and ASD were even higher than the best outcome of fully-supervised methods, as we achieved 88.69% for Dice and 1.22 for ASD. Compared to other weakly-supervised settings, we achieved the best performance in all of the unseen domains for Dice coefficient and most of the unseen domains for ASD.

Figure 2 presents a comparison of different settings' visualization results on the prediction on unseen domains. It is clear to notice that our method accurately overlapped the ground-truth label, while others may not precisely segment the unseen domain images. In general, our approach's performance is on the bar with the latest fully-supervised methods, while we only used the bounding-box for supervision instead of ground-truth masks. Compared to other experiments of weakly-supervised settings, both the results of Dice coefficient and ASD were ahead of other WSS settings on average.

4 Conclusion

In conclusion, we have presented a novel mutual supervision learning network with a dual-branch augmentation strategy that integrates DG and WSS. We demonstrated the effectiveness of all the proposed modules and evaluated our approach on the Fundus and Prostate datasets. Our experiments showed that our method can achieve performance levels that match or even exceed those of fully-supervised methods in specific domains on the datasets we trained. Our method has significant advantages over fully-supervised DG methods as it can be applied to a considerable number of medical images that lack pixel-wise labels. For the future directions, we plan to investigate more complex augmentation strategies to further enhance the model's performance.

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. Chen, X., Yuan, Y., Zeng, G., Wang, J.: Semi-supervised semantic segmentation with cross pseudo supervision. In: IEEE Conference on Computer Vision and Pattern Recognition. pp. 2613–2622 (2021)
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. 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)
5. 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)
6. He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. In: IEEE Conference on Computer Vision and Pattern Recognition. pp. 770–778 (2016)
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. 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 (2022)
9. 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)
10. 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)
11. 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)
12. 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)
13. 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)
14. 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)
15. 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)

16. 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)
17. Robey, A., Pappas, G.J., Hassani, H.: Model-based domain generalization. In: Advances in Neural Information Processing Systems. vol. 34, pp. 20210–20229 (2021)
18. Rother, C., Kolmogorov, V., Blake, A.: " grabcut" interactive foreground extraction using iterated graph cuts. ACM transactions on graphics **23**(3), 309–314 (2004)
19. Shorten, C., Khoshgoftaar, T.M.: A survey on image data augmentation for deep learning. Journal of Big Data **6**(1), 1–48 (2019)
20. 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)
21. 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)
22. Wang, J., Xia, B.: Bounding box tightness prior for weakly supervised image segmentation. In: Medical Image Computing and Computer Assisted Intervention. pp. 526–536. Springer (2021)
23. 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)
24. 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)
25. 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)
26. 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)
27. 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)
28. 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. Springer (2022)
29. 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)