D-CAM: Learning Generalizable Weakly-Supervised Medical Image Segmentation from Domain-invariant CAM

Jingjun Yi [2,*], Qi Bi [3,*], Hao Zheng[2], Haolan Zhan[2], Wei Ji[4], Huimin Huang[2], Yuexiang Li[5], Shaoxin Li[1], Xian Wu[2], Yefeng Zheng[3], and Feiyue Huang [1,†]

[1] Shanghai Digital Medicine Innovation Center, Ruijin Hospital, Shanghai, China
[2] Tencent Jarvis Lab, Shenzhen, China
[3] Medical Artificial Intelligence Laboratory, Westlake University, Hangzhou, China
[4] School of Medicine, Yale University, New Haven, USA
[5] Medical AI ReSearch (MARS) Group, Guangxi Medical University, Nanning, China
rsjingjuny@whu.edu.cn, q.bi@ieee.org, hfy30711@rjh.com.cn

Abstract. Weakly-supervised medical image segmentation with only image-level annotation is particularly challenging to infer precise pixel-wise predictions. Existing works are usually highly restricted by the assumption that the medical images for training and testing are under the same distribution. However, a robust weakly-supervised segmentation model needs to show accurate inference on medical images from unseen distributions. Different feature distributions can lead to a dramatic shift in the feature activation and class activation map (CAM), which in turn leads to the degradation of pseudo labels. In this paper, we aim to learn generalizable weakly-supervised medical image segmentation by focusing on enhancing the domain invariance for pseudo labels. A novel domain-invariant CAM learning scheme (D-CAM) is proposed, in which the content and style are decoupled during training. By inferring domain-invariant pseudo labels, the supervision of a segmentation model is more generalizable to different target domains. Extensive experiments under multiple generalized medical image segmentation settings show the state-of-the-art performance of our D-CAM. Source code is available at https://github.com/JingjunYi/D-CAM.

Keywords: Domain Generalization · Medical Image Segmentation · Image-level Weak Supervision.

1 Introduction

The pixel-wise annotation for medical images usually requires high-level expertise knowledge and is more difficult to collect than natural images. Thus, weakly-supervised image segmentation [3,23], which involves less expert labor to annotate, demonstrates its necessity in the area of medical image analysis. Among a

†Corresponding author

variety of weak supervisions, category label is the most challenging one, as there is no pixel-level auxiliary annotations (*e.g.,* points and line scribes); therefore, the quality of the pseudo label is difficult to guarantee [2]. In the past few years, great effort has been made in this field. The quality of pseudo labels from only category-label supervision has been significantly improved [4,6,13,16,26,7], and it has become practically and clinically flexible to be adapted under the medical context such as histopathology images.

However, these algorithms usually assume that the medical images for training and for inference are under the independent and identical distribution (*i.i.d.*), which is difficult to fulfill in realistic scenario. As the medical images can be collected from different scanners and different hospitals, the imaging factors (*e.g.,* contrast, color distribution, and resolution), which can be termed as *style*, may greatly vary. Consequently, the so-called domain gap widely rests between the source domain for training and the arbitrary unseen target domains for inference. From a machine learning perspective, domain generalization aims to enhance the generalization ability of a model to arbitrary unseen target domains when only trained by the source domain [18,20,9,10]. Although domain generalized medical image segmentation has been systematically studied in the past few years [22,11,27,17,1], those works are under the fully-supervised paradigm, where precise pixel-wise annotations are available for training. To the best of our knowledge, learning domain generalized weakly-supervised medical image segmentation with only category labels remains unexplored.

Domain generalized weakly-supervised medical image segmentation holds unique challenges. Usually, image-level annotation based weakly-supervised segmentation is a two-stage process. **First stage:** A classification network is trained with image-level annotations and generates the pseudo segmentation label; **Second stage:** The pseudo segmentation labels are further used to train a segmentation model. The second stage is crucial for refining coarse CAM-based masks into accurate segmentation boundaries [7,16]Due to the limited accuracy of image-level labels, the segmentation pseudo-labels generated in the first stage often only cover the approximate regions of interest and fail to capture precise boundaries. Consequently, the segmentation model trained on these pseudo-labels is prone to boundary errors. When there is a significant style difference between the target domain and the source domain, these boundary errors are further amplified, affecting the practical utility of the segmentation model. In practice, when deploying the model to a new hospital, we can quickly collect a batch of images from the target hospital. If the pseudo-label generation model from the first stage has sufficient generalization capability, we can use it to generate segmentation pseudo-labels for the target domain based on the collected images. Subsequently, we can retrain the segmentation model using the target domain images and pseudo-labels to enhance its performance in this specific domain.

In this paper, we aim to push the frontier of domain generalized weakly-supervised medical image segmentation. Our objective is to allow the pseudo labels generated in the first stage to be as robust as possible to unseen target domains, so that the segmentation model can be retrained in the target domain

with unlabeled images. To realize this objective, we turn to the frequency space, where the style variation mainly reflects in the amplitude component and the content information rests more in the phase component [17,12]. Removing the style information contained in the amplitude component allows the reconstructed spatial features to be more domain-invariant, thus the calculated class activation maps (CAMs) can be used to generate pseudo labels for unseen target domains. By decoupling the impacts of frequency components, the generation of pseudo labels can be more robust to the cross-domain style variation.

Our contributions can be summarized as follows: **i)** We make an early exploration to learn weakly-supervised medical image segmentation with image-level annotations in source domain and unlabeled images in target domain; **ii)** To address the key challenge that the cross-domain style variation degrades the quality of pseudo segmentation labels, we propose a domain-invariant class activation map (D-CAM) learning scheme; **iii)** It leverages the frequency space to decouple the style and content information, and generate domain-invariant pseudo segmentation labels for target domains; **iv)** Experiments on three pathological image datasets under multiple cross-domain settings show its state-of-the-art performance.

2 Methodology

Given a number of medical images x with category annotation y from N domains, denoted as $\mathcal{D}^{(K_1)} = \{(x_n^{(K_1)}, y_n^{(K_1)})\}_{n=1}^{N^{(K_1)}}, \cdots, \mathcal{D}^{(K_N)} = \{(x_n^{(K_N)}, y_n^{(K_N)})\}_{n=1}^{N^{(K_N)}}$, where $\mathcal{D}^{(K_i)}$ refers to the dataset of the i-th domain and $N^{(K_i)}$ denotes the number of samples in that domain, the task of generalizable weakly-supervised medical image segmentation is formulated as follows: In stage one, a classification network f_{cls} with parameters θ_{cls} is trained only on a source domain $\mathcal{D}^{(K_1)}$ to generate the dense pixel-level pseudo mask p, given by $f_{cls}(x_n^{(K_1)}, y_n^{(K_1)}; \theta_{cls}) \rightarrow p^{(K_1)}$. In stage two, the segmentation model f_{seg} is trained by the unlabeled target domain training images $x_n^{(K_n)}$ and the corresponding pixel-level pseudo masks $p^{(K_n)}$, generated by $f_{cls}(x_n^{(K_n)}; \theta_{cls}) \rightarrow p^{(K_n)}$. Our objective is to achieve robust cross-domain pseudo label generation based on the classification network f_{cls} trained on source domain. Consequently, supervised by the pseudo label p, the segmentation model f_{seg} can perform stably on the test set of target domains $\mathcal{D}^{(K_2)}, \cdots, \mathcal{D}^{(K_N)}$.

2.1 Frequency Space Decoupling

To address the domain differences primarily reflected in style diversity, it is essential to decouple the style representation from the content representation before further processing. From a frequency perspective, style information is typically found in the amplitude component, whereas content information is generally located in the phase component [15,25]. Thus, we first consider separating the style and content from the frequency space. Given a certain medical image x, assume

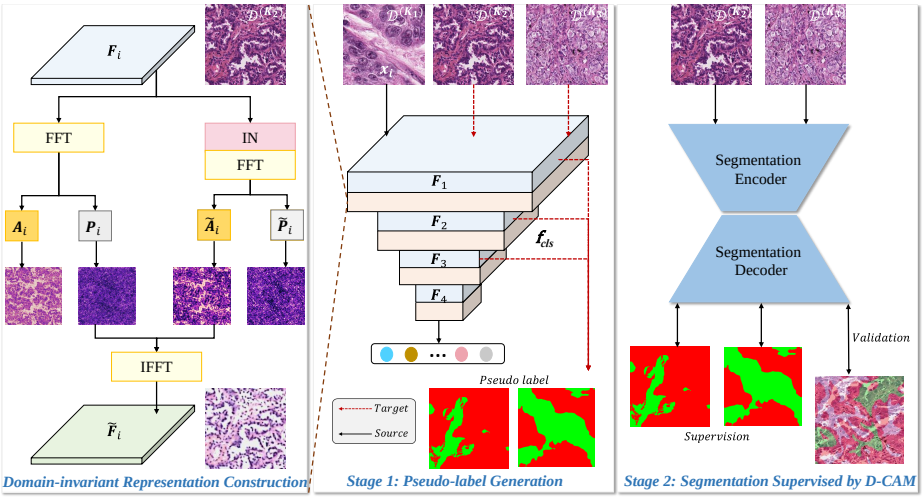

Fig. 1. Overview of the proposed domain-invariant class activation map (D-CAM) learning scheme for generalizable weakly-supervised medical image segmentation.

$\mathbf{F}_i$ is the corresponding feature map from the classification network f_{cls} of the i^{th} block. We employ the Fast Fourier Transform (FFT) to map the feature $\mathbf{F}_i$ to the frequency space, given by

$$\mathcal{F}_i(u,v) = \sum_{w=1}^{W}\sum_{h=1}^{H} \mathbf{F}_i(w,h) e^{-j2\pi(wu/W+hv/H)}, \tag{1}$$

where H and W denote the height and width of the feature map, respectively. Then, after transforming $\mathcal{F}_i(u,v)$ into the polar coordinate, the phase $\mathbf{P}_i$ and amplitude $\mathbf{A}_i$ components are computed as

$$\mathbf{P}_i(u,v) = \phi(u,v), \quad \mathbf{A}_i(u,v) = |\mathcal{F}_i(u,v)|. \tag{2}$$

2.2 Domain-invariant Representation Constructing

Frequency space uncoupling is a preliminary step that helps us separate style and content. However, how to obtain a domain-invariant feature representation under cross-domain style variations requires further investigation. To address this issue, we aim to learn a domain-invariant representation $\tilde{\mathbf{F}}_i$. Ideally, the domain-invariant representation $\tilde{\mathbf{F}}_i$ should contain the same content representation as $\mathbf{F}_i$ does, while maintaining a consistent style representation across domains. In contrast, $\mathbf{F}_i$ contains rich style-related information that varies according to the source. To obtain the domain-invariant representation $\tilde{\mathbf{F}}_i$, we apply instance normalization, a commonly-used style decoupling operation, on the original domain-specific representation $\mathbf{F}_i$. It is mathematically computed as

$$\tilde{\mathbf{F}}_i^{(W\cdot H),c} = \frac{\mathbf{F}_i^{(W\cdot H),c} - \mu}{\sigma + \epsilon} \cdot \gamma + \beta, \tag{3}$$

where $\gamma, \beta \in \mathbb{R}^C$ are affine transformation parameters that scale and shift the normalized features, as described in instance normalization [19].

$$\mu = \frac{1}{C}\sum_{c=1}^{C}\mathbf{F}_i^{(W\cdot H),c}, \sigma = \sqrt{\frac{1}{C}\sum_{i=1}^{C}(\mathbf{F}_i^{(W\cdot H),c} - \mu)^2}, \tag{4}$$

where $c = 1, 2, \cdots, C$ refers to the channel of the feature map. After that, the normalized representation $\tilde{\mathbf{F}}_i$ is processed using the FFT to obtain the phase component $\tilde{\mathbf{P}}_i$ and the amplitude component $\tilde{\mathbf{A}}_i$ by Eqs. 1 and 2. In comparison, this amplitude component $\tilde{\mathbf{A}}_i$ encapsulates style information that is intended to be invariant to domain variation, based on the findings of prior works [15,25] which suggest that amplitude reflects style. Then, we construct the domain invariant representation of medical image x by fusing its original phase component $\mathbf{P}_i$ and the domain-normalized amplitude component $\tilde{\mathbf{A}}_i$. This is implemented by the Inverse Fast Fourier Transformation (IFFT), given by

$$\tilde{\mathbf{F}}_i' = \text{IFFT}([\mathbf{P}_i, \tilde{\mathbf{A}}_i]). \tag{5}$$

Here, instance normalization removes style noise, while FFT separates style and content, making their combination effective for domain-invariant representation.

2.3 Domain-invariant CAM Generation

For the feature map $\mathbf{F}_i$ $(i = 1, 2, 3, 4)$ from each block of the classification network f_{cls}, we apply Eq. 5 to construct the domain-invariant feature $\tilde{\mathbf{F}}_i'$. Afterward, the outputted feature map $\tilde{\mathbf{F}}_4'$ from the last block of f_{cls} is utilized to generate the domain-invariant class activation map (D-CAM). We upsample $\tilde{\mathbf{F}}_4'$ to the input image resolution before thresholding to obtain pixel-level pseudo labels.

Specifically, we adopt Gradient-weighted Class Activation Mapping (Grad-CAM) [21] to generate the activation map based on $\tilde{\mathbf{F}}_4'$,

$$\tilde{\mathbf{C}} = \text{GradCAM}(\tilde{F}_4'). \tag{6}$$

$\tilde{\mathbf{C}}$ is then thresholded to obtain the segmentation pseudo-label. For a target domain with unlabeled images, we generate the domain-invariant class activation maps as pseudo segmentation labels based on the classification network f_{cls} trained on source domain. Then, the corresponding segmentation model can be trained with these images and pseudo labels. In the following experiments, for each target domain, we use the images in the training set as unlabeled images, and validate the segmentation model trained with pseudo labels on the test set.

3 Experiments

The proposed method is compared with eleven image-label weakly-supervised medical image segmentation methods, namely, HistoSegNet [2], SEAM [23], SC-CAM [3], C-CAM [4], WSSS-Tissue [7], OEEM [16], PistoSeg [6], HAMIL [26],

Table 1. Comparison with state-of-the-art image-label weakly-supervised medical image segmentation methods. BCSS as source domain.

Method	Publication	BCSS → Hist					BCSS → WSSS				
		Tumor	Stroma	mIoU	FwIoU	ACC	Tumor	Stroma	mIoU	FwIoU	ACC
Baseline		39.40	48.94	44.17	44.21	61.66	31.87	**46.44**	39.15	39.69	57.17
HistoSegNet [2]	ICCV'2019	41.62	46.94	44.28	44.30	61.50	36.43	44.04	40.23	40.51	57.63
SEAM [23]	CVPR2020	40.87	47.52	44.19	44.22	61.49	39.84	41.58	40.71	40.77	57.87
SC-CAM [3]	CVPR'2020	41.57	46.76	44.16	44.19	61.39	39.53	42.50	41.02	41.12	58.21
C-CAM [4]	CVPR'2022	41.22	47.67	44.44	44.47	61.72	39.94	42.14	41.04	41.12	58.22
WSSS-Tissue [7]	MIA'2022	38.85	**50.52**	44.68	44.74	62.35	38.55	43.66	41.11	41.29	58.37
OEEM [16]	MICCAI'2022	42.54	46.88	44.71	44.73	61.87	39.40	43.55	41.48	41.63	58.71
PistoSeg [6]	AAAI'2023	41.52	48.19	44.85	44.89	62.12	40.09	44.73	42.09	42.58	59.35
HAMIL [26]	TMI'2023	40.26	49.53	44.89	44.94	62.34	43.16	43.17	43.17	43.17	60.30
S2C [14]	CVPR'2024	41.35	45.67	43.51	43.62	61.46	39.98	44.06	42.02	42.19	58.94
CoSA [24]	ECCV'2024	40.98	50.01	45.50	45.87	62.09	44.35	42.48	43.42	43.80	60.18
PathMamba [5]	MICCAI'2024	41.58	45.36	43.47	44.08	61.50	41.06	44.18	42.62	42.79	59.07
D-CAM (*ours*)	MICCAI'2025	**44.63**	49.56	**47.09**	**47.12**	**64.14**	**47.41**	45.85	**46.63**	**46.57**	**63.61**

Table 2. Comparison with state-of-the-art image-label weakly-supervised medical image segmentation methods. Hist as source domain.

Method	Publication	Hist→ BCSS					Hist→ WSSS				
		Tumor	Stroma	mIoU	FwIoU	ACC	Tumor	Stroma	mIoU	FwIoU	ACC
Baseline		66.03	64.29	65.16	65.16	78.92	45.36	64.68	55.02	55.73	72.69
HistoSegNet [2]	ICCV'2019	66.34	65.21	65.77	65.77	79.36	47.73	64.98	56.35	56.99	73.47
SEAM [23]	CVPR'2020	66.61	65.04	65.83	65.83	79.41	46.44	64.62	55.53	56.20	72.93
SC-CAM [3]	CVPR'2020	67.79	65.05	66.42	66.42	79.86	49.01	65.39	57.20	57.80	74.03
C-CAM [4]	CVPR'2022	66.99	65.04	66.02	66.01	79.55	49.06	64.94	57.00	57.58	73.79
WSSS-Tissue [7]	MIA'2022	67.45	65.43	66.44	66.44	79.86	47.72	64.98	56.35	56.98	73.46
OEEM [16]	MICCAI'2022	67.91	66.44	67.18	67.17	80.38	49.23	64.76	56.99	57.56	73.73
PistoSeg [6]	AAAI'2023	68.44	66.15	67.30	67.30	80.48	49.85	65.84	57.85	58.44	74.50
HAMIL [26]	TMI'2023	68.25	67.00	67.63	67.63	80.70	**52.29**	65.22	**58.75**	**59.23**	74.82
S2C [14]	CVPR'2024	64.95	63.28	64.12	64.24	78.41	43.91	62.08	53.00	53.24	70.19
CoSA [24]	ECCV'2024	66.48	65.17	65.83	65.95	80.64	44.50	62.97	53.74	53.92	75.06
PathMamba [5]	MICCAI'2024	68.14	67.57	67.86	67.99	81.02	46.85	65.30	56.08	56.17	72.80
D-CAM (*ours*)	MICCAI'2025	**70.51**	**69.08**	**69.80**	**69.80**	**82.23**	50.88	**66.04**	58.46	59.02	**74.88**

S2C [14], CoSA [24] and PathMamba [5], under the cross-domain setting. Among them, WSSS-Tissue [7] serves as our baseline. Since benchmarking the domain generalized WSSS is a new task where only the common categories across domains are involved, all the compared methods are re-implemented by the default hyper-parameters and configurations and the results only include the common categories.

Datasets: Three publicly-available tissue segmentation datasets, *i.e.,* **LUAD-HistoSeg** [7], **BCSS-WSSS** [7] and **WSSS4LUAD** [8], are involved for cross-domain experiments. For simplicity, we denote them as Hist, BCSS, and WSSS, respectively, in the following text. Two cross-domain common categories among these three datasets, namely, *tumor* and *stroma*, are used for validation.

Evaluation Metrics: Following existing weakly-supervised medical image segmentation works [7], the mean IoU (mIoU), frequency weighted IoU (FwIoU) and pixel-level accuracy (ACC) of both categories, are used for evaluation.

Implementation Details: The proposed D-CAM keeps the same baseline and hyper-parameter settings of [7]. In the first stage, the classification network f_{cls} uses a ResNet-38 with pre-trained weights on ImageNet as the backbone and then finetunes on the corresponding tissue dataset under the default hyper-

Table 3. Ablation study on how each frequency component (P, A, $\tilde{P}$, $\tilde{A}$) impacts the weakly-supervised medical image segmentation performance on unseen target domains. BCSS as source domain. Evaluation metrics are presented in percentage (%).

Component	BCSS → Hist					BCSS → WSSS				
	Tumor	Stroma	mIoU	FwIoU	ACC	Tumor	Stroma	mIoU	FwIoU	ACC
P+A (baseline)	39.40	48.94	44.17	44.21	61.66	31.87	**46.44**	39.15	39.69	57.17
$\tilde{P}$+$\tilde{A}$	39.68	48.87	44.29	44.33	61.75	38.98	43.70	42.16	42.35	58.89
$\tilde{P}$+A	37.19	**51.76**	44.48	44.54	62.48	41.97	43.31	43.38	43.52	59.97
P	37.21	48.69	42.95	43.00	60.65	43.31	43.53	44.11	44.47	61.35
$\tilde{P}$	42.20	51.21	46.70	46.74	64.02	45.27	44.18	44.23	46.17	62.30
P+$\tilde{A}$ (ours)	**44.63**	49.56	**47.09**	**47.12**	**64.14**	**47.41**	45.85	**46.63**	**46.57**	**63.61**

parameter settings. The classification activation map (CAM) is generated from the classification network f_{cls} and converted into the pseudo label following all the default operations in [7]. In the second stage, a DeepLab-V3 segmentation model is trained under the supervision of the generated pseudo label by the proposed method, under the same and default hyper-parameter settings of [7].

3.1 Comparison with State-of-the-art

BCSS as Source Domain: The results, when generalized to Hist and WSSS, are listed in Table 1. The proposed D-CAM improves the mIoU metric by up to 2.20% and the accuracy metric by 1.79% on the Hist target domain. Additionally, when embedded into other CAM-based methods like C-CAM [4] and S2C [14], our amplitude normalization strategy consistently boosts generalization performance on unseen domains. The improvement of tumor category is up to 2.09% in mIoU. Similarly, D-CAM improves the mIoU by 3.40% and the accuracy by 3.31% on the WSSS target domain. The improvement of tumor category is up to 4.25% in mIoU.

Hist as Source Domain: The results, when generalized to BCSS and WSSS target domains, are reported in Table 2. The proposed D-CAM improves the mIoU metric by 2.17% and the accuracy metric by 1.53% on the BCSS target domain. Notably, the improvement of tumor category is up to 2.07% in mIoU. On the other hand, the proposed D-CAM yields an mIoU of 58.46% and an accuracy of 74.88%, outperforming all the rest state-of-the-art and achieving very close performance to HAMIL. In general, the proposed D-CAM especially allows the generated pseudo labels to be domain invariant and shows more robustness than existing methods on unseen target domains.

3.2 Ablation Study

The proposed D-CAM is plug and play when embedded into the first-stage classification network. Our ablation study focuses on the impact of each frequency component and its interaction. Especially we focus on the following cases: 1) baseline (i.e., original classification network), where the phase and amplitude components both come from the original medical image (denoted as P+A); 2)

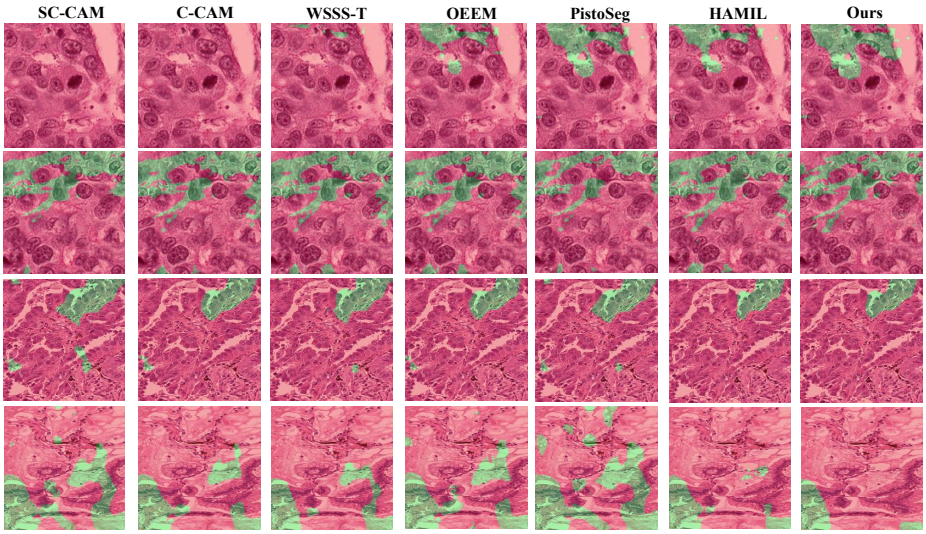

Fig. 2. Visual segmentation result comparison between the proposed method and the state-of-the-art methods on the BCSS (first and second rows) and WSSS (third and fourth rows) target domains. Hist is used as the source domain.

using the features from the normalized medical image (denoted as $\tilde{P}+\tilde{A}$); 3) using the phase component from the normalized medical image and the amplitude component from the original medical image (denoted as $\tilde{P}+A$); 4) only using the phase component from the original medical image (denoted as P); 5) only using the phase component from the normalized medical image (denoted as $\tilde{P}$). Notice that, the proposed method can be regarded as using the phase component from the original medical image and the amplitude component from the normalized medical image (denoted as $P+\tilde{A}$).

The results are listed in Table 3. Naively using the phase component from the original (P) or normalized ($\tilde{P}$) image shows varying performance depending on the domain pair. Using the original medical image ($P+A$) only leverages the information from the source domain, and the performance is consequently inferior to the other settings. On the other hand, combining the normalized amplitude and the original phase ($P+\tilde{A}$) shows a superior performance, as the domain-dependent style information is removed. Additionally, removing Stage 2 leads to a 3–5% mIoU drop on target domains, confirming the necessity of refinement.

3.3 Visualization of Segmentation Results

Fig. 2 shows visual results of the proposed method and existing state-of-the-art methods. Hist dataset is used as the source domain. BCSS (first and second rows) and WSSS (third and fourth rows) datasets are used as the unseen target domains, respectively. The proposed method shows the best visual results.

4 Conclusion

In this paper, we made an initial exploration to learn domain generalized weakly-supervised medical image segmentation when only under the image label supervision. Compared with domain generalized medical image segmentation under the conventional fully-supervised conditions, the key challenge of this task mainly lies in that the quality of pseudo label is easily degraded by the style information from different domains. To address this issue, we proposed a domain invariant class activation map learning scheme dubbed as D-CAM. Its key idea is to decouple the content that is stable among different domains and the style that is domain dependent in the frequency space. Extensive experiments on multiple cross-domain settings showed the state-of-the-art performance of D-CAM. We hope our research could foster further exploration of the generalization ability of weakly-supervised segmentation in the medical community.

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

References

1. Bi, Q., Yi, J., Zheng, H., Ji, W., Huang, Y., Li, Y., Zheng, Y.: Learning generalized medical image segmentation from decoupled feature queries. In: Proceedings of the AAAI Conference on Artificial Intelligence. vol. 38, pp. 810–818 (2024)
2. Chan, L., Hosseini, M.S., Rowsell, C., Plataniotis, K.N., Damaskinos, S.: HistoSegNet: Semantic segmentation of histological tissue type in whole slide images. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 10662–10671 (2019)
3. Chang, Y.T., Wang, Q., Hung, W.C., Piramuthu, R., Tsai, Y.H., Yang, M.H.: Weakly-supervised semantic segmentation via sub-category exploration. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 8991–9000 (2020)
4. Chen, Z., Tian, Z., Zhu, J., Li, C., Du, S.: C-CAM: Causal CAM for weakly supervised semantic segmentation on medical image. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 11676–11685 (2022)
5. Fan, J., Lv, T., Di, Y., Li, L., Pan, X.: PathMamba: Weakly supervised state space model for multi-class segmentation of pathology images. In: International Conference on Medical Image Computing and Computer Assisted Intervention. pp. 500–509 (2024)
6. Fang, Z., Chen, Y., Wang, Y., Wang, Z., Ji, X., Zhang, Y.: Weakly-supervised semantic segmentation for histopathology images based on dataset synthesis and feature consistency constraint. In: Proceedings of the AAAI Conference on Artificial Intelligence. vol. 37, pp. 606–613 (2023)
7. Han, C., Lin, J., Mai, J., et al.: Multi-layer pseudo-supervision for histopathology tissue semantic segmentation using patch-level classification labels. Medical Image Analysis **80**, 102487 (2022)
8. Han, C., Pan, X., Yan, L., et al.: WSSS4LUAD: Grand challenge on weakly-supervised tissue semantic segmentation for lung adenocarcinoma. arXiv preprint arXiv:2204.06455 (2022)

9. Harary, S., Schwartz, E., Arbelle, A., et al.: Unsupervised domain generalization by learning a bridge across domains. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 5280–5290 (2022)
10. Hu, C., Lee, G.H.: Feature representation learning for unsupervised cross-domain image retrieval. In: European Conference on Computer Vision. pp. 529–544 (2022)
11. Hu, S., Liao, Z., Zhang, J., Xia, Y.: Domain and content adaptive convolution based multi-source domain generalization for medical image segmentation. IEEE Transactions on Medical Imaging **42**(1), 233–244 (2023)
12. Huang, J., Guan, D., Xiao, A., Lu, S.: FSDR: Frequency space domain randomization for domain generalization. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 6891–6902 (2021)
13. Kuang, Z., Yan, Z., Zhou, H., Yu, L.: Cluster-re-supervision: Bridging the gap between image-level and pixel-wise labels for weakly supervised medical image segmentation. IEEE Journal of Biomedical and Health Informatics **27**(10), 4890–4901 (2023)
14. Kweon, H., Yoon, K.J.: From SAM to CAMs: Exploring segment anything model for weakly supervised semantic segmentation. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 19499–19509 (2024)
15. Lee, S., Bae, J., Kim, H.Y.: Decompose, adjust, compose: Effective normalization by playing with frequency for domain generalization. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp. 11776–11785 (2023)
16. Li, Y., Yu, Y., Zou, Y., Xiang, T., Li, X.: Online easy example mining for weakly-supervised gland segmentation from histology images. In: International Conference on Medical Image Computing and Computer Assisted Intervention. pp. 578–587 (2022)
17. Liu, Q., Chen, C., Qin, J., Dou, Q., Heng, P.A.: FedDG: Federated domain generalization on medical image segmentation via episodic learning in continuous frequency space. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 1013–1023 (2021)
18. Mahajan, D., Tople, S., Sharma, A.: Domain generalization using causal matching. In: International Conference on Machine Learning. pp. 7313–7324 (2021)
19. Nam, H., Kim, H.E.: Batch-instance normalization for adaptively style-invariant neural networks. Advances in Neural Information Processing Systems **31** (2018)
20. Segu, M., Tonioni, A., Tombari, F.: Batch normalization embeddings for deep domain generalization. Pattern Recognition **135**, 109115 (2023)
21. Selvaraju, R.R., Cogswell, M., Das, A., Vedantam, R., Parikh, D., Batra, D.: Grad-CAM: Visual explanations from deep networks via gradient-based localization. In: Proceedings of the IEEE International Conference on Computer Vision. pp. 618–626 (2017)
22. 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)
23. Wang, Y., Zhang, J., Kan, M., Shan, S., Chen, X.: Self-supervised equivariant attention mechanism for weakly supervised semantic segmentation. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 12275–12284 (2020)
24. Yang, X., Rahmani, H., Black, S., Williams, B.M.: Weakly supervised co-training with swapping assignments for semantic segmentation. In: European Conference on Computer Vision. pp. 459–478 (2024)

25. Yi, J., Bi, Q., Zheng, H., Zhan, H., Ji, W., Huang, Y., Li, Y., Zheng, Y.: Learning spectral-decomposited tokens for domain generalized semantic segmentation. In: Proceedings of the 32nd ACM International Conference on Multimedia. pp. 8159–8168 (2024)
26. Zhong, L., Wang, G., Liao, X., Zhang, S.: HAMIL: High-resolution activation maps and interleaved learning for weakly supervised segmentation of histopathological images. IEEE Transactions on Medical Imaging **42**(10), 2912–2923 (2023)
27. Zhou, Z., Qi, L., Shi, Y.: Generalizable medical image segmentation via random amplitude mixup and domain-specific image restoration. In: European Conference on Computer Vision. pp. 420–436 (2022)