Uncertainty-aware meta-weighted optimization framework for domain-generalized medical image segmentation

Seok-Hwan Oh[1⋆], Guil Jung[1⋆], Sang-Yun Kim[1], Myeong-Gee Kim[2], Young-Min Kim[1], Hyeon-Jik Lee[1], Hyuk-Sool Kwon[3], and Hyeon-Min Bae[1]

[1] Electrical Engineering Department, KAIST, 34141 Daejeon, South Korea
joseph9337@kaist.ac.kr
[2] Barreleye Inc., 06211 Seoul, South Korea
[3] Department of Emergency Medicine, SNUBH, 13620 Seong-nam, South Korea

Abstract. Accurate segmentation of echocardiograph images is essential for the diagnosis of cardiovascular diseases. Recent advances in deep learning have opened a possibility for automated cardiac image segmentation. However, the data-driven echocardiography segmentation schemes suffer from domain shift problems, since the ultrasonic image characteristics are largely affected by measurement conditions determined by device and probe specification. In order to overcome this problem, we propose a domain generalization method, utilizing a generative model for data augmentation. An acoustic content- and style-aware diffusion probabilistic model is proposed to synthesize echocardiography images of diverse cardiac anatomy and measurement conditions. In addition, a meta-learning-based spatial weighting scheme is introduced to prevent the network from training unreliable pixels of synthetic images, thereby achieving precise image segmentation. The proposed framework is thoroughly evaluated using both in-distribution and out-of-distribution echocardiography datasets and demonstrates outstanding performance compared to state-of-the-art methods. Code is available at https://github.com/Seokhwan-Oh/MLSW

Keywords: Image segmentation · domain generalization · meta-learning.

1 Introduction

Echocardiography (Echo) image segmentation is an essential diagnostic procedure for identification of cardiovascular disease [1]. The cardiac segment provides clinically significant bio-markers such as Left-Ventricle (LV) thickness, LV volume, and ejection fraction [1, 5]. Manual cardiac image segmentation, which is the current gold standard, is a laborious task that requires several working hours of expert radiologists [25]. Hence, technological development has been made to automate cardiac image segmentation [26, 11, 20, 7].

⋆ SH. Oh and G. Jung contributed equally

Recently, deep learning has proven its efficacy in image segmentation across diverse applications. In cardiac image segmentation, fully automated image segmentation provided by deep Neural Networks (NN) has contributed to effective and prompt diagnosis. Therefore, significant advances have been made in the field of deep learning-based Echo segmentation. However, such data-driven cardiac image segmentation method suffers from a domain shift problem. The performance of the NN is significantly degraded when the network is applied to the data that exhibit different distribution from the training data [19, 8]. In medical ultrasound (US), the features of the images are highly dependent on the device manufacturers, specification of probes, and frequency of the US pulse. Considering that the performance of deep NNs is dependent on the operating environment, such image variation poses a significant concern in cardiovascular disease diagnosis, where inaccurate diagnosis significantly impacts patient mortality and prognosis. Consequently, there is a growing demand for robust domain-generalized semantic segmentation in cardiac image analysis.

Domain Generalization (DG) aims to configure a versatile model that can perform well to out-of-distribution data [8]. Recently, data augmentation methods have been proposed to enhance the generalizability of the NN [23]. Especially, there exists an increasing interest in adopting a generative model to diversify the data distribution. However, the generative model-based data augmentations suffer from a fundamental drawback in that the fidelity of the synthesized image is not guaranteed. Specifically, in generative data augmentation for semantic segmentation, the generative model employs the image segment as a condition to synthesize corresponding images. Then the generated images and the corresponding image segment pairs, are employed as the input and ground truth for training the segmentation NN. However, the precision of such synthesized data is compromised compared to when the annotation is made by an expert radiologist, making precise parameterization of the segmentation network challenging.

In order to address the problem, we proposed a spatial uncertainty-aware meta-optimization method for generalized medical image segmentation. In the realm of noisy label training [15], there have been recent approaches suggest to utilize meta-learning for recognizing and filtering inaccurate data. We propose to extend the meta-learning-based noisy-label training techniques into the spatial domain. We configured a meta-net to capture unreliable spatial areas within synthesized data by utilizing precise human-annotated data as a validation sample. Then the segmentation network is optimized by employing only reliable pixels provided by the meta-net. We demonstrate that the proposed spatial meta-learning methods contribute to the DG of the network. In addition, we introduce a generative model, which is an acoustic content- and style-aware diffusion probabilistic model (DDPM) [4], that can synthesize Echo images, accurately depicting a wide range of cardiac anatomy and measurement conditions. The proposed generative model helps improving the versatility of the segmentation network by providing a broad spectrum of data distribution.

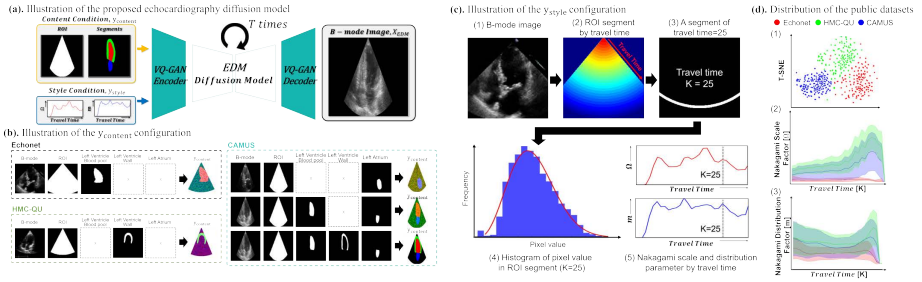

Fig. 1. Configuration of (a) EDM, (b) y_{content}, and (c) y_{style} are provided. In Fig.1-(d), distribution of the datasets are analyzed, the shaded area denotes 90% quantile region.

2 Echocardiography denoising diffusion model

In this section, we propose a universal Echo generative model designed to generate cardiac images that can provide a wide range of Echo features. The Echo image characteristics are largely determined by content- and style-based features. The content-based features refer to elements that determine the spatial configuration of the image. Shape of the heart and the observed Region Of Interest (ROI) are considered as content features. Even when examining a patient with identical content features, variations in images can arise based on the specifications of the measurement devices. Style-based features govern the characteristics related to the US equipment manufacturer, the frequency of US used, and the type of the probe employed to measure the US signals. Fig. 1 illustrates an overall configuration of the proposed Echocardiography denoising Diffusion Model (EDM). The EDM is implemented based on the DDPM which employs an iterative denoising process for precise image generation. The EDM utilizes content y_{content} and style y_{style} conditions as inputs for the NN, enabling the user to adjust image generation conditions and produce Echo images simulating a broad range of cardiac anatomy and US device characteristics.

Content condition. The y_{content} is composed of an Echo ROI $y_{\text{content,ROI}}$, and anatomical feature $y_{\text{content,SEG}}$. Echo images typically exhibit a sector shape configuration. $y_{\text{content,ROI}}$ is configured with a center angle of $60° < \theta < 150°$ to cover general ROI commonly employed by radiologists. Clinically important cardiac components, which are, the LV blood pool $y_{\text{content,LVBP}}$, LV wall $y_{\text{content,LVW}}$, and Left Atrium (LA) $y_{\text{content,LA}}$ are considered as $y_{\text{content,SEG}}$. The compositions of the $y_{\text{content,SEG}}$ can vary depending on the dataset characteristics (described in section 2.1 dataset). As presented in Fig. 1-(b), the $y_{\text{content,SEG}}$ are configured under diverse composition of the cardiac segments, allowing it to produce robust Echo images under various content conditions.

Style Condition. The y_{style} represents characteristics of the image such as brightness, contrast, and scattering, which are determined by US device specifications. To precisely quantify the style condition of the US image, we formulated y_{style} in the form of Nakagami distribution [14]. The Nakagami distribution is a

statistical model for representing the distribution of US signal and is denoted as

$$f(r) = \frac{2m^m r^{2m-1}}{\Gamma(m)\Omega^m} \exp\left(-\frac{m}{\Omega}r^2\right) U(r), \tag{1}$$

where Γ and U denote gamma and unit-step functions respectively. The Nakagami distribution is governed by two parameters, which are the Nakagami scale parameter $\Omega = E(R^2)$ and Nakagami distribution parameter $m = \frac{[E(R^2)]^2}{E[R^2 - E(R^2)]^2}$. E denotes the statistical mean. We segmented the US images into K(=30) partitions according to the US signal travel time. Subsequently, we quantify the Ω and m of the US signals within each subsection to construct the style feature (Fig. 1-(c)). Fig. 1-(d) presents Ω and m distribution plots of Echo images from three institutes using different US machines. As demonstrated in Fig. 1-(d), disparate Ω and m parameters are configured according to the US measurement environment, proving the proposed Nakagami parameter is a proper method for representing a wide range of US-style features.

Implementation details. The iterative denoising scheme requires considerable computational cost. In order to enhance efficiency of the EDM, we employed a latent diffusion model framework [12], that applies a denoising scheme in the latent domain instead of the image domain. The learning object of the EDM ϕ_{EDM} is described as follows:

$$L \coloneqq E_{\phi_{\text{EDM}}, \epsilon \sim N(0,1)} \left[|\epsilon - \phi_{\text{EDM}}(y_{\text{content}}, y_{\text{style}}, t)| \right], \tag{2}$$

where $t = 1, \ldots, T$ denotes the time steps of denoising.

2.1 Dataset

In order to evaluate Echo images of diverse measurement conditions and study cohorts, three Echo datasets, Echonet-Dynamic [10], HMC-QU [2], and CAMUS [6] from different institutes are employed. The Echonet-Dynamic dataset comprises 10,030 videos capturing an Apical 4-Chamber (A4C) heart view, with corresponding blood pool segmentation label. The HMC-QU dataset consists of 2,349 Echo images, which include a heart A4C view and LV wall labels. The CAMUS dataset includes 1,000 Echo images from 500 patients. The dataset comprises Apical 2-Chamber (A2C) and A4C views. The LV blood pool, LV wall, and the LA are annotated as segmentation labels. The detailed train and test split of each dataset is described in Supplementary A.

2.2 Experiments

Echocardiography image synthesis. In this section, assessments of the EDM are presented. The Echo image is generated through the denoising diffusion implicit model [16] with $T = 200$ denoising steps. Fig. 2-(a) presents Echo images generated under a distinct combination of y_{content} and y_{style} conditions. The EDM generates realistic Echo images, successfully expressing anatomic structure given in content condition. Furthermore, the study demonstrates that the

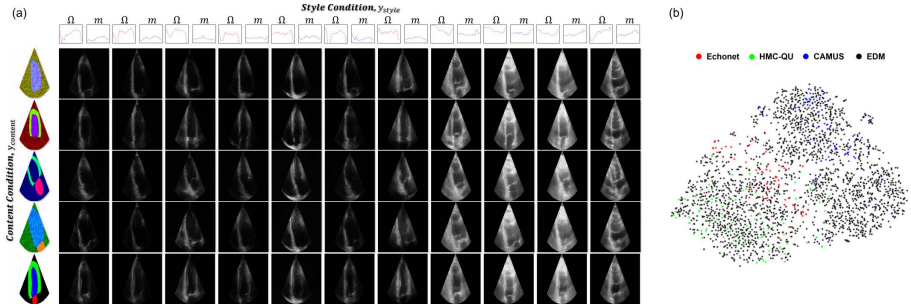

Fig. 2. (a) Generated Echo images with diverse content and style condition. (b) t-SNE analysis generated Echo samples together with the public datasets.

Ω and m parameters are an effective way of quantifying echo image style. By employing the Nakagami condition, the characteristic of Echo can easily be managed, thereby enabling the simulation of diverse US signal measurement environments. Fig. 2-(b) introduces a t-SNE plot of the generated dataset analyzed together with three public Echo images. The study shows that EDM provides a diversified generation of Echo data encompassing the distribution of the source dataset. In quantitative assessment, the EDM demonstrates FID scores of 16.86 for Echonet-dynamic, 17.45 for HMC-QU, and 10.22 for CAMUS. The model achieves competitive performance with the state-of-the-art (SoTA) echocardiography generative models showing 75.36 [17] and 50.83 [18] FID.

3 Meta-Learned Spatial Weighting (MLSW)

In this section, a domain generalized semantic segmentation network θ_{seg} that implements a MLSW method is proposed (Fig. 3). Through the use of EDM, numerous pairs of Echo images, and cardiac segments $\{X_{\text{EDM}}, Y_{\text{EDM}}\} \sim D_{\text{EDM}}$ involving diverse Echo image conditions are provided. X and Y denote the input image and the corresponding ground truth segmentation, respectively. The machine-generated dataset D_{EDM} is inherently less precise compared to the dataset $\{X_{\text{HM}}, Y_{\text{HM}}\} \sim D_{\text{HM}}$ that are manually segmented by expert radiologists. Therefore, the MLSW is proposed to provide a solution to the following question: *Is there a way to learn diversity from machine-made data, and learn fine details from man-made ones?* The MLSW scheme employs a meta-net θ_{meta} that captures inaccurate regions in the input X and the corresponding label Y to exclude the defective sub-pixels during training.

MLSW optimization. Algorithm 1 describes overall procedure of the proposed MLSW. The meta-net θ_{meta} is parameterized based on the meta-learning idea. Specifically, we employed a clean and precise $\mathcal{D}_{\text{HM}}$ for the meta-optimization of the θ_{meta}. The optimal θ_{meta} is obtained by minimizing the following loss:

$$\theta_{\text{meta}} = \arg\min_{\theta_{\text{meta}}} \mathbb{E}\left[L(Y_{\text{HM}}, \theta'_{\text{seg}}(X_{\text{HM}}))\right], \tag{3}$$

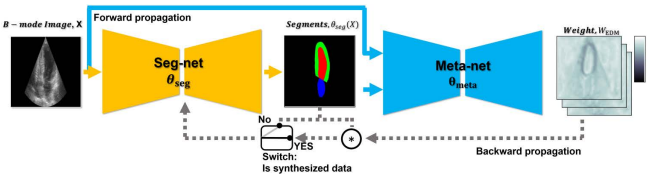

Fig. 3. Configuration of the proposed MLSW.

Algorithm 1 MLSW optimization

Input: $\{X_{\mathrm{HM}}, Y_{\mathrm{HM}}\} \sim \mathcal{D}_{\mathrm{HM}}$ and $\{X_{\mathrm{EDM}}, Y_{\mathrm{EDM}}\} \sim \mathcal{D}_{\mathrm{EDM}}$
Initialize: Seg-net parameter: θ_{seg}, Meta-net parameter: θ_{meta}
Require: Hyperparameters: $\mathrm{lr}_{\mathrm{base}} = 1e-4$, $\mathrm{lr}_{\mathrm{meta}} = 1e-4$

for it in iterations **do**
  Output image $O_{\mathrm{EDM}} = \theta_{\mathrm{seg}}(X_{\mathrm{EDM}})$
  Meta-weight $W_{\mathrm{EDM}} = \theta_{\mathrm{meta}}(\theta_{\mathrm{seg}}(X_{\mathrm{EDM}}), Y_{\mathrm{EDM}})$
  Evaluate weighted loss with machine-generated data: $\nabla_{\theta_{\mathrm{seg}}} W_{\mathrm{EDM}} L(Y_{\mathrm{EDM}}, O_{\mathrm{EDM}})$
  Compute adapted parameter: $\theta'_{\mathrm{seg}} = \theta_{\mathrm{seg}} - \mathrm{lr}_{\mathrm{base}} \nabla_{\theta_{\mathrm{seg}}} W_{\mathrm{EDM}} L(Y_{\mathrm{EDM}}, O_{\mathrm{EDM}})$
  Compute meta-test loss with adapted parameter: $\nabla_{\theta_{\mathrm{meta}}} L(Y_{\mathrm{HM}}, \theta'_{\mathrm{seg}}(X_{\mathrm{HM}}))$
  Update $\theta_{\mathrm{meta}} \leftarrow \theta_{\mathrm{meta}} - \mathrm{lr}_{\mathrm{meta}} \nabla_{\theta_{\mathrm{meta}}} L(Y_{\mathrm{HM}}, \theta'_{\mathrm{seg}}(X_{\mathrm{HM}}))$
  Update $\theta_{\mathrm{seg}} \leftarrow \theta_{\mathrm{seg}} - \mathrm{lr}_{\mathrm{base}}(\nabla_{\theta_{\mathrm{seg}}} W_{\mathrm{EDM}} L(Y_{\mathrm{EDM}}, O_{\mathrm{EDM}}) + \nabla_{\theta_{\mathrm{seg}}} L(Y_{\mathrm{HM}}, \theta_{\mathrm{seg}}(X_{\mathrm{HM}})))$
end for

where $\theta'_{\mathrm{seg}} = \theta_{\mathrm{seg}} - \mathrm{lr}_{\mathrm{base}} \nabla_{\theta_{\mathrm{seg}}} W_{\mathrm{EDM}} L(Y_{\mathrm{EDM}}, \theta_{\mathrm{seg}}(X_{\mathrm{EDM}}))$, $W_{\mathrm{EDM}} = \theta_{\mathrm{meta}}(X_{\mathrm{EDM}})$. Note that $L(x, y)$ denotes the cross-entropy loss between x and y. Namely, θ_{meta} is trained to configure a spatial weight matrix W_{EDM}, so that the filtered training $W_{\mathrm{EDM}} L(Y_{\mathrm{EDM}}, O_{\mathrm{EDM}})$ contributes to the accuracy of θ_{seg}, when it is validated with clean data D_{HM}. The structure of θ_{meta} is implemented based on U-net [13]. θ_{meta} employs the Echo image X, inferred cardiac segment $\theta_{\mathrm{seg}}(X)$, and spatial loss $L(Y, \theta'_{\mathrm{seg}}(X))$, as inputs for generating W_{EDM}. The MLSW requires only 31MB of additional parameters during neural network training.

Segmentation network optimization. The θ_{seg} is optimized with the following learning objective:

$$\theta_{\mathrm{seg}} = \mathrm{argmin}_{\theta_{\mathrm{seg}}} \mathbb{E}[W_{\mathrm{EDM}} L(Y_{\mathrm{EDM}}, O_{\mathrm{EDM}}) + \nabla_{\theta_{\mathrm{seg}}} L(Y_{\mathrm{HM}}, \theta_{\mathrm{seg}}(X_{\mathrm{HM}}))]. \quad (4)$$

Both $\mathcal{D}_{\mathrm{EDM}}$ and $\mathcal{D}_{\mathrm{HM}}$ are employed for the parameterization of θ_{seg}. During training with machine-made $\mathcal{D}_{\mathrm{EDM}}$, the θ_{seg} is optimized using reliable pixels filtered by W_{EDM}. On the other hand, when utilizing radiologist annotated clean data $\mathcal{D}_{\mathrm{HM}}$, θ_{seg} learns the details of $\mathcal{D}_{\mathrm{HM}}$ without filtration, ultimately achieving domain-generalized and precise semantic segmentation. The U-net is employed as the baseline architecture of the θ_{seg}.

4 Experiments

4.1 Semantic segmentation dataset

The dataset, employed for the Echo segmentation NN, is categorized into two distinct groups, in-distribution and out-of-distribution dataset.

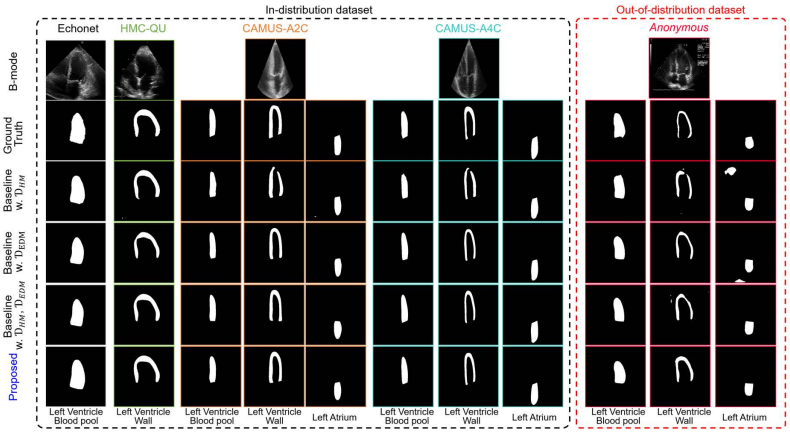

Fig. 4. Qualitative assessment of the proposed network with baseline models (Extended assessment of the comparative models is provided in supplementary material.)

In-distribution (ID) dataset. The ID dataset comprises three public datasets $\mathcal{D}_{\text{HM}}$, which are Echonet, HMC-QU, and CAMUS, and an EDM-generated dataset. The ID test set is configured to validate the performance of the NN in Echo segmentation without the domain shift. Furthermore, an ECHO-EDM dataset, denoted as $\mathcal{D}_{\text{EDM}}$, is configured by using the generative model presented in Section 2. A total of 570K synthetic Echo images and corresponding labels are generated under diverse combinations of y_{content} and y_{style}.

Out-of-distribution (OOD) dataset. The OOD test dataset is composed of unseen data that present a distinct distribution from the trained dataset, and is employed to assess the robustness of the model to the data with novel attributes. In particular, the OOD dataset demonstrates different US measurement setups and patient cohorts from the ID dataset. The OOD dataset provides Echo images of 237 randomly selected patients who visited the anonymous Hospital over the past decade. The dataset was captured using eight different US machines, which demonstrates diverse specification ranging from portable to high-performance console-type US devices. Three expert sonographers annotated the LV blood pool, LV wall, and LA of the Echo image. The data acquisition is conducted under institutional review board (IRB Number: B-2301-807-108) of Seoul national university bundang hospital.

4.2 Result and Discussion

Ablation study. Ablation studies are conducted to validate the efficacy of the proposed DG framework. The scheme is validated with conventional baseline NN that is trained with $\mathcal{D}_{\text{HM}}$ and $\mathcal{D}_{\text{EDM}}$, respectively, and the one utilizing both of $\mathcal{D}_{\text{HM}}$ and $\mathcal{D}_{\text{EDM}}$ without MLSW optimization. The qualitative assessment is presented in Fig. 4. The baseline NN trained only with $\mathcal{D}_{\text{HM}}$ demonstrates limited generalization capability, showing significant defaults such as disconnected LV

Table 1. Quantitative assessment of the network with baseline and SoTA models.

Ablation study													
MIOU			In-distribution								Out-of-distribution		
			Echo	HMC	CAMUS A2C			CAMUS A4C			Anonymous		
Model	Dataset	MLSW	LV pool	LV pool	LV pool	LV wall	LA	LV pool	LV wall	LA	LV pool	LV wall	LA
Baseline	$\mathcal{D}_{\text{HM}}$		0.8904	0.8772	0.8246	0.5713	0.8083	0.7679	0.5769	0.7655	0.7502	0.3804	0.4987
	$\mathcal{D}_{\text{EDM}}$		0.8448	0.8155	0.8649	0.7282	0.8487	0.8254	0.7168	0.7634	0.7792	0.3335	0.5750
	$\mathcal{D}_{\text{HM}} + \mathcal{D}_{\text{EDM}}$		0.8593	0.9067	0.8616	0.7205	0.8438	0.8240	0.7131	0.8085	0.7553	0.3020	0.5837
Proposed	$\mathcal{D}_{\text{HM}} + \mathcal{D}_{\text{EDM}}$	✓	**0.9378**	**0.9113**	**0.8822**	**0.7639**	**0.8571**	**0.8507**	**0.7460**	**0.8237**	**0.8066**	**0.3885**	**0.6512**
Comparative study													
DSU [9]			0.9151	0.8743	0.8389	0.6649	0.8365	0.7845	0.6690	0.7674	0.7787	0.4075	0.5952
Cutout [3]			0.9326	0.8632	0.8545	0.6921	0.8181	0.8075	0.6787	0.7644	0.7929	0.3586	0.5878
Mixup [22]			0.8629	0.7726	0.8303	0.5768	0.5820	0.7988	0.5836	0.5328	0.7730	0.3626	0.3566
MLDG [8]			0.8488	0.8807	0.8808	0.7531	0.8389	0.8441	0.7380	0.8162	0.8089	0.3829	0.5366
Mixstyle [24]			0.9198	0.8834	0.8606	0.6846	0.8252	0.8027	0.6760	0.7687	0.7765	0.3985	0.5588
FACT [21]			0.8870	0.8350	0.8722	0.7467	0.8273	0.8354	0.7236	0.8162	0.7332	0.3885	0.5466

wall, and artifact in LA, when it is applied to the OOD data. The baseline NN trained with numerous $\mathcal{D}_{\text{EDM}}$ shows enhanced generalization capability, however demonstrates limitation in precise segmentation of the cardiac image. The proposed framework demonstrates a reliable image segmentation in both the ID and OOD datasets, proving the MLSW scheme not only contributes to the DG but also offers fine-grained semantic segmentation.

Quantitative assessments are introduced in Table 1. By utilizing the $\mathcal{D}_{\text{EDM}}$, the baseline network, trained with conventional optimization scheme, achieves 3.5% Mean Intersection Over Union (MIOU) improvement in the OOD dataset assessment. The proposed framework, which employs MLSW optimization, outperforms the baseline NN in the ID dataset. The effectiveness of MLSW scheme is emphasized when it is compared with an OOD dataset. The proposed network demonstrates a 11.8% enhancement in the MIOU metric in average, compared to the conventional baseline NN trained with $\mathcal{D}_{\text{HM}}$. Extended ablation studies applying diverse backbone structure are presented in the supplementary material, demonstrating that the proposed framework can universally be applied to the NNs, and contributes to the segmentation accuracy.

Comparative study. Comparative studies are conducted by assessing the proposed scheme with diverse SoTA DG methodologies. The comparisons are conducted with a meta-learning approach (MLDG [8]), an uncertainty modeling method (DSU [9]), style (Mixstyle [24]) and Fourier-based (FACT [21]) augmentation schemes, and other representative data augmentation-based (CutOut [3] and MixUp [22]) DG schemes. The DSU and CutOut provide average accuracy improvements of 0.050 and 0.036 MIOU in the OOD dataset, respectively. The DSU presents excellent performance among the comparative models; however, it still shows insufficient precision when compared to the proposed NN. This is attributed to the fact that Echo image DG requires a comprehensive understanding of the cardiac anatomy and US characteristics. However, the comparative models lack such semantic interpretation, resulting in a limited performance increase. On the other hand, the proposed framework employs EDM model, a generative model that can represent a range of US measurement conditions, with a proper meta-weighted optimization methodology. The proposed scheme results in 0.022 higher MIOU compared to the best-performing SoTA method.

5 Conclusion

In this paper, we proposed a learning framework that contributes to the DG of cardiac image segmentation. A generative model, based on the DDPM, is proposed and offers cardiac images of diverse combinations of heart anatomy and US measurement conditions. The meta-learning-based spatial weighting scheme is introduced to prevent the network from training unreliable pixels of synthetic images and achieve precise image reconstruction. Qualitative and quantitative assessments are performed using both ID and OOD data from diverse institutes using varied US devices. The experiments demonstrate that the proposed framework contributes to domain robust and precise semantic segmentation, showing outstanding performance in comparison to the state-of-the-art DG methods.

Acknowledgement. This work was supported by the Korea Medical Device Development Fund grant funded by the Korea government (Project Number: RS2020-KD000007).

Disclosure of Interests. S-H. Oh and Y-M. Kim are under contracts with Ministry of Korea National Defense. Y-M. Kim, H-J. Lee and S-Y. Kim are under scholarship from the Korea Government Scholarship Program.

References

1. Chen, C., Qin, C., Qiu, H., Tarroni, G., Duan, J., Bai, W., Rueckert, D.: Deep learning for cardiac image segmentation: a review. Frontiers in Cardiovascular Medicine **7**, 25 (2020)
2. Degerli, A., Zabihi, M., Kiranyaz, S., Hamid, T., Mazhar, R., Hamila, R., Gabbouj, M.: Early detection of myocardial infarction in low-quality echocardiography. IEEE Access **9**, 34442–34453 (2021)
3. DeVries, T., Taylor, G.W.: Improved regularization of convolutional neural networks with cutout. arXiv preprint arXiv:1708.04552 (2017)
4. Ho, J., Jain, A., Abbeel, P.: Denoising diffusion probabilistic models. Advances in neural information processing systems **33**, 6840–6851 (2020)
5. Lang, R.M., Badano, L.P., Mor-Avi, V., Afilalo, J., Armstrong, A., Ernande, L., Flachskampf, F.A., Foster, E., Goldstein, S.A., Kuznetsova, T., et al.: Recommendations for cardiac chamber quantification by echocardiography in adults: an update from the american society of echocardiography and the european association of cardiovascular imaging. European Heart Journal-Cardiovascular Imaging **16**(3), 233–271 (2015)
6. Leclerc, S., Smistad, E., Pedrosa, J., Østvik, A., Cervenansky, F., Espinosa, F., Espeland, T., Berg, E.A.R., Jodoin, P.M., Grenier, T., et al.: Deep learning for segmentation using an open large-scale dataset in 2d echocardiography. IEEE transactions on medical imaging **38**(9), 2198–2210 (2019)
7. Li, C., Xu, C., Gui, C., Fox, M.D.: Distance regularized level set evolution and its application to image segmentation. IEEE transactions on image processing **19**(12), 3243–3254 (2010)

8. Li, D., Yang, Y., Song, Y.Z., Hospedales, T.: Learning to generalize: Meta-learning for domain generalization. In: Proceedings of the AAAI conference on artificial intelligence. vol. 32 (2018)
9. Li, X., Dai, Y., Ge, Y., Liu, J., Shan, Y., Duan, L.Y.: Uncertainty modeling for out-of-distribution generalization. arXiv preprint arXiv:2202.03958 (2022)
10. Ouyang, D., He, B., Ghorbani, A., Yuan, N., Ebinger, J., Langlotz, C.P., Heidenreich, P.A., Harrington, R.A., Liang, D.H., Ashley, E.A., et al.: Video-based ai for beat-to-beat assessment of cardiac function. Nature **580**(7802), 252–256 (2020)
11. Rezaee, M.R., Van der Zwet, P.M., Lelieveldt, B., Van der Geest, R.J., Reiber, J.H.: A multiresolution image segmentation technique based on pyramidal segmentation and fuzzy clustering. IEEE transactions on image processing **9**(7), 1238–1248 (2000)
12. Rombach, R., Blattmann, A., Lorenz, D., Esser, P., Ommer, B.: High-resolution image synthesis with latent diffusion models. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp. 10684–10695 (2022)
13. Ronneberger, O., Fischer, P., Brox, T.: U-net: Convolutional networks for biomedical image segmentation. In: Medical Image Computing and Computer-Assisted Intervention–MICCAI 2015: 18th International Conference, Munich, Germany, October 5-9, 2015, Proceedings, Part III 18. pp. 234–241. Springer (2015)
14. Shankar, P.M.: Ultrasonic tissue characterization using a generalized nakagami model. IEEE transactions on ultrasonics, ferroelectrics, and frequency control **48**(6), 1716–1720 (2001)
15. Song, H., Kim, M., Park, D., Shin, Y., Lee, J.G.: Learning from noisy labels with deep neural networks: A survey. IEEE Transactions on Neural Networks and Learning Systems (2022)
16. Song, J., Meng, C., Ermon, S.: Denoising diffusion implicit models. arXiv preprint arXiv:2010.02502 (2020)
17. Stojanovski, D., Hermida, U., Lamata, P., Beqiri, A., Gomez, A.: Echo from noise: synthetic ultrasound image generation using diffusion models for real image segmentation. arXiv preprint arXiv:2305.05424 (2023)
18. Tiago, C., Snare, S.R., Šprem, J., McLeod, K.: A domain translation framework with an adversarial denoising diffusion model to generate synthetic datasets of echocardiography images. IEEE Access **11**, 17594–17602 (2023)
19. Wang, M., Deng, W.: Deep visual domain adaptation: A survey. Neurocomputing **312**, 135–153 (2018)
20. Xu, C., Prince, J.L.: Snakes, shapes, and gradient vector flow. IEEE Transactions on image processing **7**(3), 359–369 (1998)
21. Xu, Q., Zhang, R., Zhang, Y., Wang, Y., Tian, Q.: A fourier-based framework for domain generalization. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 14383–14392 (2021)
22. Zhang, H., Cisse, M., Dauphin, Y.N., Lopez-Paz, D.: mixup: Beyond empirical risk minimization. arXiv preprint arXiv:1710.09412 (2017)
23. Zhou, K., Liu, Z., Qiao, Y., Xiang, T., Loy, C.C.: Domain generalization: A survey. IEEE Transactions on Pattern Analysis and Machine Intelligence (2022)
24. Zhou, K., Yang, Y., Qiao, Y., Xiang, T.: Domain generalization with mixstyle. arXiv preprint arXiv:2104.02008 (2021)
25. Zhou, S.K., Rueckert, D., Fichtinger, G.: Handbook of medical image computing and computer assisted intervention. Academic Press (2019)
26. Zuluaga, M.A., Biffi, B., Taylor, A.M., Schievano, S., Vercauteren, T., Ourselin, S.: Strengths and pitfalls of whole-heart atlas-based segmentation in congenital

heart disease patients. In: Reconstruction, Segmentation, and Analysis of Medical Images: First International Workshops, RAMBO 2016 and HVSMR 2016, Held in Conjunction with MICCAI 2016, Athens, Greece, October 17, 2016, Revised Selected Papers 1. pp. 139–146. Springer (2017)