DA-GAN: LEARNING STRUCTURED NOISE REMOVAL IN ULTRASOUND VOLUME PROJECTION IMAGING FOR ENHANCED SPINE SEGMENTATION

Zixun Huang[1,†], *Rui Zhao*[1,†,⋆], *Frank H. F. Leung*[1], *Kin-Man Lam*[1], *Sai Ho Ling*[3], *Juan Lyu*[4], *Sunetra Banerjee*[3], *Timothy Lee*[2], *De Yang*[2], *Yong-Ping Zheng*[2]

[1] Department of Electronic and Information Engineering, [2]Department of Biomedical Engineering, The Hong Kong Polytechnic University, Hong Kong
[3]School of Biomedical Engineering, University of Technology Sydney, NSW, Australia
[4]College of Information and Communication Engineering, Harbin Engineering University, China

ABSTRACT

Ultrasound volume projection imaging (VPI) has shown to be appealing from a clinical perspective, because of its harmlessness, flexibility, and efficiency in scoliosis assessment. However, the limitations in hardware devices degrade the resultant image content with strong structured noise. Owing to the unavailability of reference data and the unpredictable degradation model, VPI image recovery is a challenging problem. In this paper, we propose a novel framework to learn the structured noise removal from unpaired samples. We introduce the attention mechanism into the generative adversarial network to enhance the learning by focusing on the salient corrupted patterns. We also present a dual adversarial strategy and integrate the denoiser with a segmentation model to produce the task-oriented noiseless estimation. Experimental results show that the proposed method can greatly improve both the visual quality and the segmentation accuracy on spine images.

***Index Terms*—** Ultrasound image restoration, Spine segmentation, Unpaired learning.

1. INTRODUCTION

In clinical scoliosis diagnosis, experts need to view hundreds of frames in an ultrasound sequence of a whole spine column, which is time-consuming and tedious [1]. To simplify the measurement, Volume Projection Imaging (VPI) was proposed to synthesize coronal 2D images based on the intensity of the voxels in the ultrasound sequence [2, 3]. However, owing to the fast movement of the probe and the noise in the collected spatial information, ultrasound VPI images usually suffer from a significant degradation by structured noise, which not only affects the performance of automatic pathological analysis, but also poses a challenge to doctors for accurate diagnosis. As presented in Fig. 1, structured noise, different from random noise, shows high spatial correlation, and only appears in some regions in the image. The existence of structured noise degrades the discriminative patterns in the ultrasound image, and consequently, confuses the deep network when performing classification, detection, or segmentation. VPI image restoration is an open problem, where the ground-truth data is generally inaccessible. Moreover, the structured noise varies with the ultrasound operators, probes, and empirical imaging parameters, which makes the degradation hard to model. Hence, it is also impractical to synthesize the paired noisy and noiseless samples for learning.

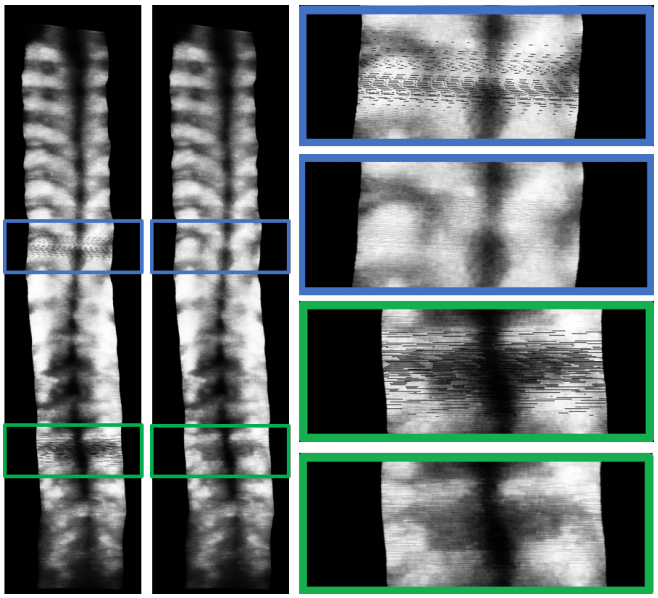

Fig. 1. An example of structured noise removal in the ultrasound VPI. Left to right: the original observation, the recovered image, and the upscaled details.

Recently, reference-free image restoration has been widely studied. NTGAN [4] was proposed to learn a deep denoiser without clean reference. However, it requires the prior knowledge of the degradation model, which limits its performance on ultrasound images. N2V [5] and N2S [6] presented the

† authors equally contribute to this paper.
⋆ Corresponding author: rick10.zhao@connect.polyu.hk

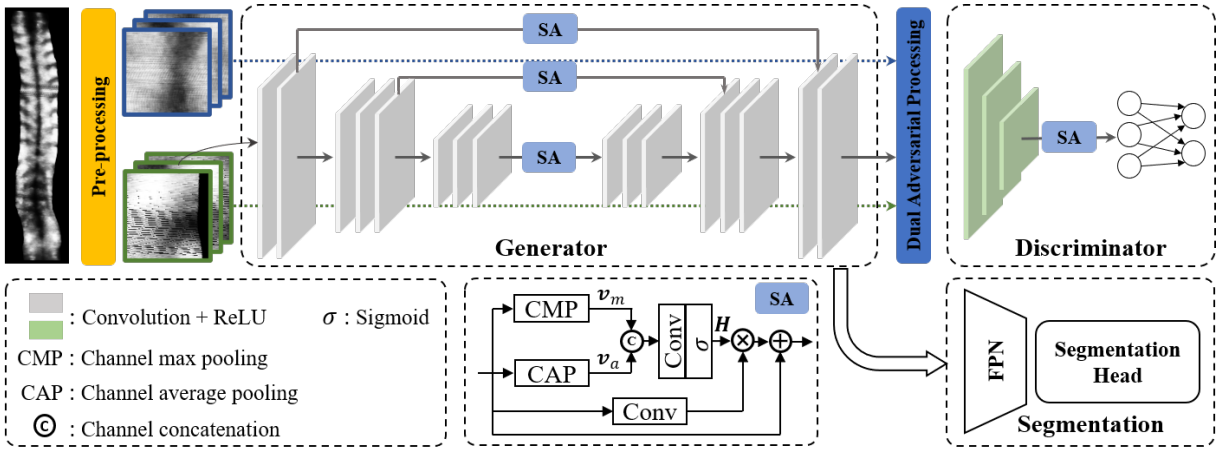

Fig. 2. Overview of the proposed framework for enhanced spine segmentation from ultrasound volume projection images.

self-supervision strategy for image restoration. However, they were shown to perform poorly on structured noise. On the other hand, generative adversarial networks (GANs) provide an alternative to solving the restoration problem in a weakly supervised manner. Hou et al. [7] proposed the cycle adversarial learning to reconstruct the image appearance for enhancing the segmentation on CT images. However, Liu et al. [8] showed that GAN-based methods would create artefacts in images. To address this issue, they presented a wavelet correction transfer network (WaveCT) to eliminate the appearance shift. However, the spectral-based supervision in [8] is not satisfactory for structured noise removal, and other regularization-based methods, e.g., the total variation penalty in [9], fail to preserve the segmentation details. To tackle these problems, we propose the dual attentive generative adversarial network (DA-GAN) to learn the structured noise removal without paired supervision. Specifically, we introduce the spatial attention mechanism into both the generator and the discriminator to facilitate the framework in localizing the corrupted patterns. By this means, the resultant model can better retain the image content while recovering the noisy regions. We also present the dual adversarial learning strategy, which serves as an online augmentation approach to learning the structured noise distributions. To achieve the task-oriented restoration, we integrate DA-GAN with a segmentation network, and form an end-to-end framework for spine segmentation. The segmentation task introduces the additional regularization to promote the restoration task. The main contributions of this paper can be concluded as follows:

- We propose the dual-attentive generative adversarial network, based on the spatial attention mechanism and the dual adversarial strategy, to learn the VPI image restoration under the unpaired supervision.
- We integrate the proposed DA-GAN with a segmentation network to form an end-to-end framework for enhanced spine segmentation.
- We conduct various experiments to evaluate DA-GAN on both the visual quality and the segmentation accuracy.

2. METHODOLOGY

In this section, we introduce the proposed DA-GAN in detail. Fig. 2 gives an overview of the proposed framework, including a preprocessing module, a restoration processor (generator), and a segmentation head. The preprocessing step prepares the unpaired image patches for learning. The main component, i.e. the generator in DA-GAN, follows a simple UNet design with spatial attention (SA) units to facilitate the learning of structured noise removal. The segmentation head introduces additional supervision from the ground-truth segments for the task-oriented restoration. **It is worth noting that the preprocessing module and the discriminator only affect the training stage, and will be detached from the model in inference, which bring no burden to the applications.**

2.1. Patch splitting preprocessing

The preprocessing step aims to extract the patches from the input observation, and automatically split them into two groups, i.e. the corrupted domain and the noise-free domain, for adversarial domain-transfer learning. To this end, we employ an edge detector to divide the extracted patches into the positive and the negative groups by their averaged vertical variations, as follows:

$$f(\boldsymbol{x}) = \begin{cases} 0, & \text{if } g(\boldsymbol{x}) \geq \beta_n, \\ 1, & \text{if } g(\boldsymbol{x}) \leq \beta_p, \end{cases} \quad \text{with } g(\boldsymbol{x}) = \frac{1}{MN} \sum_{i}^{M} \sum_{j}^{N} |h(\boldsymbol{x})_{i,j}|, \tag{1}$$

where $\boldsymbol{x}$ denotes the extracted observation patch. β_n and β_p account for the negative and positive thresholds for selecting the corrupted and clean patches, respectively. $h(\cdot)$ represents the edge detector in the vertical direction. M and N are height and width, respectively, of the image patches. By this means, we synthesize the domain-transfer pairs from a single

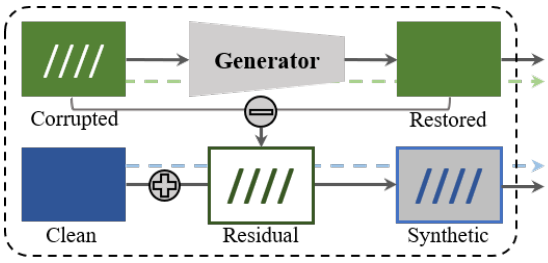

Fig. 3. Illustration of the proposed dual adversarial learning.

observation, so that the restoration processor can effectively penalize the detail leakage in learning.

2.2. Spatial attention unit

The attention mechanism plays an important role in the restoration learning. As shown in Fig. 1, the structured noisy degradation only exists in some regions of the corrupted patches. Thus, both the generator and the discriminator need to focus on those salient areas, to better recover the discriminative patterns and to provide higher confidence on the real/fake prediction, respectively. For this purpose, we introduce the spatial attention (SA) units into the framework to facilitate the restoration learning. The detailed structure of the spatial attention unit is presented in Fig. 2. Given a feature map $\boldsymbol{v} \in \mathbb{R}^{c\times h\times w}$, the SA unit first compresses it by using the average pooling and the max pooling in the channel dimension. The two compact representations, i.e. $\boldsymbol{v}_a \in \mathbb{R}^{1\times h\times w}$ and $\boldsymbol{v}_m \in \mathbb{R}^{1\times h\times w}$, respectively, are then aggregated by the channel concatenation. We employ a non-linear module with the sigmoid activation to further compress the aggregated representation and generate the heatmap $\boldsymbol{H} \in \mathbb{R}^{1\times h\times w}$. We utilize $\boldsymbol{H}$ to rescale the input signal along the spatial dimension to achieve spatial attention. The residual connection is established to stabilize the training process.

2.3. Dual adversarial learning

As the amount of the corrupted patches and the clean patches is highly imbalanced, we further propose the dual adversarial learning strategy to serve as an online augmentation in learning the structured noise distribution. Given an observation $\boldsymbol{x}$, the generator $\mathcal{G}$ produces a noise-free estimation, denoted as $\hat{\boldsymbol{x}} = \mathcal{G}(\boldsymbol{x})$. We compute the residual between $\boldsymbol{x}$ and $\hat{\boldsymbol{x}}$ as $\boldsymbol{r} = \boldsymbol{x} - \hat{\boldsymbol{x}}$, which represents the degradation pattern leading to the prediction of the discriminator. Therefore, if we add the residual back to another clean patch $\boldsymbol{z}$, the discriminator should predict the synthetic observation $\hat{\boldsymbol{z}} = \boldsymbol{z} + \boldsymbol{r}$ as a corrupted sample. The illustration of this procedure is shown in Fig. 3. We employ this strategy as the additional regularization in learning the restoration, and thus the dual adversarial loss $\mathcal{L}_{\text{adv}}$ is formulated as follows:

$$\begin{aligned}\mathcal{L}_{\text{adv}} =& \mathbb{E}_{\boldsymbol{z}}[\log \mathcal{D}_{\text{cln}}(\boldsymbol{z})] + \mathbb{E}_{\hat{\boldsymbol{x}}}[\log(1-\mathcal{D}_{\text{cln}}(\hat{\boldsymbol{x}})] + \\ & \mathbb{E}_{\boldsymbol{x}}[\log \mathcal{D}_{\text{crp}}(\boldsymbol{x})] + \mathbb{E}_{\hat{\boldsymbol{z}}}[\log(1-\mathcal{D}_{\text{crp}}(\hat{\boldsymbol{z}})],\end{aligned} \tag{2}$$

where $\mathcal{D}_{\text{cln}}$ and $\mathcal{D}_{\text{crp}}$ denote the discriminators for the clean and corrupted domains, respectively. We also employ the reconstruction loss $\mathcal{L}_{\text{rec}}$, defined by the pixel-wise distance between the input observation $\boldsymbol{x}$ and its output estimation $\hat{\boldsymbol{x}}$, to preserve the image content as follows:

$$\mathcal{L}_{\text{rec}} = ||\boldsymbol{x} - \hat{\boldsymbol{x}}||_2^2. \tag{3}$$

The task-oriented supervision $\mathcal{L}_{\text{seg}}$ from the ground-truth segments is defined by the pixel-wise Cross Entropy loss, as follows:

$$\mathcal{L}_{\text{seg}} = -\sum_{\#\text{cls}} y_{\text{true}} \log(y_{\text{pred}}), \tag{4}$$

where y_{true} and y_{pred} are the ground-truth and the predicted label for the pixel, respectively. #cls refers to the number of classes. Therefore, the overall objective function is formulated as:

$$\mathcal{L} = \mathcal{L}_{\text{adv}} + \lambda_1 \mathcal{L}_{\text{seg}} + \lambda_2 \mathcal{L}_{\text{rec}}, \tag{5}$$

where λ_1 and λ_2 are the hyperparameters, controlling the trade-off between the loss terms.

3. EXPERIMENT

3.1. Dataset

The dataset used in this paper is collected from 3D ultrasound scanning of the whole spine region. Then, the ultrasound VPI technique is utilized to generate the projected 2D images. Ultrasound VPI images from 109 subjects, with different degrees of scoliosis, were selected. The bone features are labelled by medical experts to serve as the ground-truth segmentation masks. We further divided the collected data into a training set and a testing set with 80 and 29 samples, respectively, based on the identity information. Since the size of the ultrasound images varies with the patients, we resize all the images to 256×1024 in the training set, and extract the patches with a size of 96×96 from the images. Random rotation and mirroring are applied to the patches for augmentation. In preprocessing, the two thresholds, i.e. β_n and β_p, are empirically set to 4 and 2, respectively. In the testing phase, a query sample is first denoised and resized to the same resolution, i.e. 256×1024, and then fed to the segmentation model for predicting the segment masks, which are then rescaled back to the original resolution.

3.2. Implementation details

We establish the proposed framework based on PyTorch [10] and MMSegment [11]. In DA-GAN, all the convolutional kernels are of size 3×3, with padding 1, except for those in SA, where the kernel size is 1×1 with padding 0. The number of convolutional filters is set to $[64, 128, 256]$. The segmentation head is built using the default setting in [12]. During training, we build a mini-batch of 16 samples. We adopt

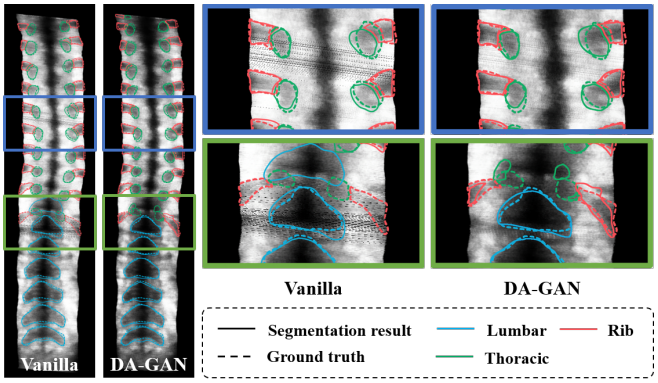

Fig. 4. Visualization of the DA-GAN restored image for the quality and segmentation assessment.

Adam [13] to minimize the objective function defined in Eq. (5), with λ_1 and λ_2 empirically set to 0.5 and 0.01, respectively. We initialize the learning rate to 10^{-4}, and employ the cosine annealing strategy [14] to decrease it to 10^{-6} within 100 epochs. We train DA-GAN on a Nvidia GEFORCE GTX 2080 Ti GPU, and it takes about 5 hours to train up the model.

3.3. Results

As a segmentation-oriented restoration method, we evaluate DA-GAN based on the visual quality of the produced VPI images, as well as the comparison with the other state-of-the-art segmentation algorithms.

Visualization: We first evaluate DA-GAN by visualizing the restored images for both the quality and the segmentation assessment. Fig. 4 reports an example. To make a fair comparison, we enlarge the kernel size and the network depth of the comparing vanilla model (only FPN), to make its capacity equal to, or larger than "DA-GAN + FPN". Apparently, DA-GAN provides a better visual quality, as the structured noise are effectively reduced. In terms of segmentation, we can clearly observe that our proposed algorithm accurately locates the first lumbar by eliminating the confusing patterns resulting from the structured noise, while the vanilla model predicts a false alarm, and misses the last thoracic and rib.

Quantitative segmentation results: To validate the benefit from DA-GAN for the spine segmentation, we compare the results with the other state-of-the-art segmentation algorithms, i.e. the vanilla FPN model, UNet [15], RSNU [9], WaveCT [8], and PPMU [16] . We also replace DA-GAN with the other weakly supervised denoisers, i.e. N2V [5], N2S [6] and CycleGAN [7], to verify its superiority in improving segmentation performance. The results are tabulated in Table 1. All the comparing methods are based on the same settings in Sec. 3.2, and are established with an equal, or larger, model capacity. It can be seen that the proposed DA-GAN outperforms the other weakly supervised denoisers by a large margin. In addition, the proposed method also surpasses those state-of-the-art segmentation algorithms (Vanilla–PPMU in Table 1) by about 1.5% in the three metrics. Among the competitors, RSNU [9] and WaveCT [8] are the recently proposed methods that are especially designed for appearance-corrupted segmentation. The proposed DA-GAN achieves the Dice scores of about 87.7%, 77.8%, and 79.7% for Lumbar, Thoracic, and Rib, respectively, which are much higher than the other comparing algorithms. Thus, DA-GAN is desirable for the VPI image enhancement and spine segmentation in clinical applications.

Table 1. Quantitative segmentation results, where D: Dice score(%), J: Jaccard index(%), and P: Pixel accuracy(%).

Methods	Lumbar			Thoracic			Rib			*Ave.*		
	D	J	P	D	J	P	D	J	P	D	J	P
Cycle [7]	86.81	76.96	87.90	76.71	62.47	80.74	78.54	65.00	80.46	80.67	68.14	83.03
N2S [6]	80.46	76.97	88.75	76.60	62.32	80.28	77.96	64.21	81.29	80.46	67.84	**83.44**
N2V [5]	87.28	77.71	88.50	77.51	63.50	75.84	78.87	65.32	78.40	81.22	68.84	80.91
Vanilla	85.69	75.29	85.57	76.42	62.12	74.12	78.02	64.24	76.45	80.04	67.21	78.72
UNet [15]	82.21	70.26	83.28	74.70	59.94	72.39	77.37	63.46	76.37	78.09	64.56	77.35
RSNU [9]	85.85	75.52	88.24	77.45	63.39	77.30	79.26	65.92	**80.28**	80.86	68.28	81.94
WaveCT [8]	86.59	76.58	87.91	75.91	61.36	72.34	78.49	64.82	79.72	80.33	67.58	79.99
PPMU [16]	84.58	73.68	85.11	76.55	62.22	75.22	78.21	64.48	78.30	79.78	66.79	79.55
∼w/o SA	87.23	77.65	89.07	77.85	63.94	77.33	79.24	65.84	79.78	81.44	69.14	82.06
∼w/o DAL	87.44	77.97	**89.32**	**77.98**	**64.11**	77.28	79.11	65.64	78.97	81.51	69.24	81.86
Ours	**87.72**	**78.46**	88.73	77.81	63.90	**77.70**	**79.68**	**66.42**	79.54	**81.73**	**69.59**	81.99

Ablation study: To investigate the effectiveness of the different designs in DA-GAN, we perform the ablation study. As listed in Table 1, we explore the effect of the SA units and the dual adversarial learning (DAL). We also establish the competitors with the same capacity as the proposed model. It is obvious that both SA and DAL contribute greatly to the segmentation, as the average performance is degraded by about 0.3%, if either of them is removed. Thus, the proposed strategies are beneficial to spine segmentation.

More visual comparisons, runtime results, and ablation studies can be found in the provided supplementary material.

4. CONCLUSION

In this paper, we have studied an enhanced method for spine segmentation by recovering the structured noisy patterns in ultrasound VPI images. We introduce the spatial attention mechanism into the GAN-based framework to force the model to concentrate on the corrupted patterns for learning the restoration in a weakly supervised manner. The dual adversarial learning strategy is further proposed to facilitate the memorization on the structured noise distribution. We aggregate the recovering model with the segmentation network to perform task-oriented restoration for improving the segmentation on spine images. Extensive experiments have shown that the proposed algorithm produces appealing results, in terms of both visual quality and spine segmentation, which makes it a potential solution to clinical applications.

5. REFERENCES

[1] Tamas Ungi, Franklin King, Michael Kempston, Zsuzsanna Keri, Andras Lasso, Parvin Mousavi, John

Rudan, Daniel P Borschneck, and Gabor Fichtinger, "Spinal curvature measurement by tracked ultrasound snapshots," *Ultrasound in medicine & biology*, vol. 40, no. 2, pp. 447–454, 2014.

[2] Chung-Wai James Cheung, Guang-Quan Zhou, Siu-Yin Law, Tak-Man Mak, Ka-Lee Lai, and Yong-Ping Zheng, "Ultrasound volume projection imaging for assessment of scoliosis," *IEEE transactions on medical imaging*, vol. 34, no. 8, pp. 1760–1768, 2015.

[3] S. Banerjee, S. H. Ling, J. Lyu, S. Su, and Y. Zheng, "Automatic segmentation of 3d ultrasound spine curvature using convolutional neural network," in *Annual International Conference of Engineering in Medicine Biology Society (EMBC)*. IEEE, 2020, pp. 2039–2042.

[4] R. Zhao, D. P. K. Lun, and K. Lam, "Ntgan: Learning blind image denoising without clean reference," in *British Machine Vision Conference (BMVC)*, Sep. 2020.

[5] Alexander Krull, Tim-Oliver Buchholz, and Florian Jug, "Noise2void-learning denoising from single noisy images," in *International Conference on Computer Vision and Pattern Recognition (CVPR)*, 2019, pp. 2129–2137.

[6] Joshua Batson and Loic Royer, "Noise2self: Blind denoising by self-supervision," *arXiv preprint arXiv:1901.11365*, 2019.

[7] Yuankai Huo, Zhoubing Xu, Shunxing Bao, Albert Assad, Richard G Abramson, and Bennett A Landman, "Adversarial synthesis learning enables segmentation without target modality ground truth," in *International Symposium on Biomedical Imaging (ISBI)*. IEEE, 2018, pp. 1217–1220.

[8] Z. Liu, X. Yang, R. Gao, S. Liu, H. Dou, S. He, Y. Huang, Y. Huang, H. Luo, Y. Zhang, Y. Xiong, and D. Ni, "Remove appearance shift for ultrasound image segmentation via fast and universal style transfer," in *International Symposium on Biomedical Imaging (ISBI)*. IEEE, 2020, pp. 1824–1828.

[9] Zixun Huang Li-Wen Wang, Frank H. F. Leung, Sunetra Banerjee, De Yang, Timothy Lee, Juan Lyu, Sai Ho Ling, and Yong-Ping Zheng, "Bone feature segmentation in ultrasound spine image with robustness to speckle and regular occlusion noise," in *International Conference on Systems, Man, and Cybernetics (SMC)*. IEEE, 2020.

[10] Adam Paszke, Sam Gross, Soumith Chintala, Gregory Chanan, Edward Yang, Zachary DeVito, Zeming Lin, Alban Desmaison, Luca Antiga, and Adam Lerer, "Automatic differentiation in PyTorch," in *Neural Information Processing Systems (NIPS) Workshop*, 2017.

[11] Jiarui Xu et al., "Mmsegmentation," in *https://github.com/open-mmlab/mmsegmentation*, 2020.

[12] Alexander Kirillov, Ross Girshick, Kaiming He, and Piotr Dollar, "Panoptic feature pyramid networks," in *International Conference on Computer Vision and Pattern Recognition (CVPR)*. IEEE, Jun 2019.

[13] Diederik P. and Jimmy Ba Kingma, "Adam: A method for stochastic optimization," in *International Conference on Learning Representations (ICLR)*, 2015.

[14] Ilya Loshchilov and Frank Hutter, "Sgdr: Stochastic gradient descent with warm restarts," in *International Conference on Learning Representations (ICLR)*, 2017.

[15] Olaf Ronneberger, Philipp Fischer, and Thomas Brox, "U-net: Convolutional networks for biomedical image segmentation," in *International Conference on Medical Image Computing and Computer Assisted Intervention (MICCAI)*. Springer, 2015, pp. 234–241.

[16] Ahmed H Shahin, Karim Amer, and Mustafa A Elattar, "Deep convolutional encoder-decoders with aggregated multi-resolution skip connections for skin lesion segmentation," in *International Symposium on Biomedical Imaging (ISBI)*. IEEE, 2019, pp. 451–454.