



# Ultrasound Variational Style Transfer to Generate Images Beyond the Observed Domain

Alex Ling Yu Hung<sup>(✉)</sup>  and John Galeotti 

Carnegie Mellon University, Pittsburgh, PA 15213, USA  
{lingyuh, jgaleotti}@andrew.cmu.edu

**Abstract.** The use of deep learning in medical image analysis is hindered by insufficient annotated data and the inability of models to generalize between different imaging settings. We address these problems using a novel variational style-transfer neural network that can sample various styles from a computed latent space to generate images from a broader domain than what was observed. We show that using our generative approach for ultrasound data augmentation and domain adaptation during training improves the performance of the resulting deep learning models, even when tested within the observed domain.

**Keywords:** Style transfer · Variational autoencoder · Generative models · Data augmentation · Domain generalization

## 1 Introduction

Deep learning in medical imaging has immense potential but the challenge is that images taken by different machines or in different settings follow different distributions [4]. As a result, models are usually not generalized enough to be used across datasets. Especially in ultrasound, images captured by different imaging settings on different subjects can be so different that the model trained on one dataset could completely fail on another dataset. Besides, since annotation of medical images is laborious work which requires experts, the number of annotated images is limited. It is extremely difficult to get a large dataset of annotated medical images that are from different distributions. Therefore, we propose a continuous neural style transfer algorithm, which is capable of generating ultrasound images with known content from unknown style latent space.

In the first attempt at style transfer with convolutional neural networks (CNNs), the content and style features were extracted using pretrained VGGs [21] and used to iteratively optimize the output image on-the-fly [6]. Feed-forward frameworks were then proposed to get rid of the numerous iterations during inference [10, 23]. The gram matrix, which describes the style information in images, were further explained in [14]. To control the style of the output, ways to manipulate the spatial location, color information and spatial scale were introduced in

[7]. Adaptive Instance Normalization (AdaIN) was proposed as a way to perform arbitrary style transfer in real-time [8]. Further improvements were made where the image transformation network was directly generated from a style image via meta network [20]. Some works are dedicated to using style transfer as a data augmentation tool [17, 25], which showed that style transfer on natural images can improve the performance of classification and segmentation.

In medical image analysis, [22] directly applied style transfer to fundus images for augmentation, while [3] analyzed the style of ultrasound images encoded by VGG encoders. [16] improved the segmentation results of cardiovascular MR images by style transfer. [5] built their network upon StyleGAN [11] to generate high-resolution medical images. [24] showed that generated medical images based on style transfer can improve the results of semantic segmentation on CT scans. [15] proposed a method to do arbitrary style transfer on ultrasound images.

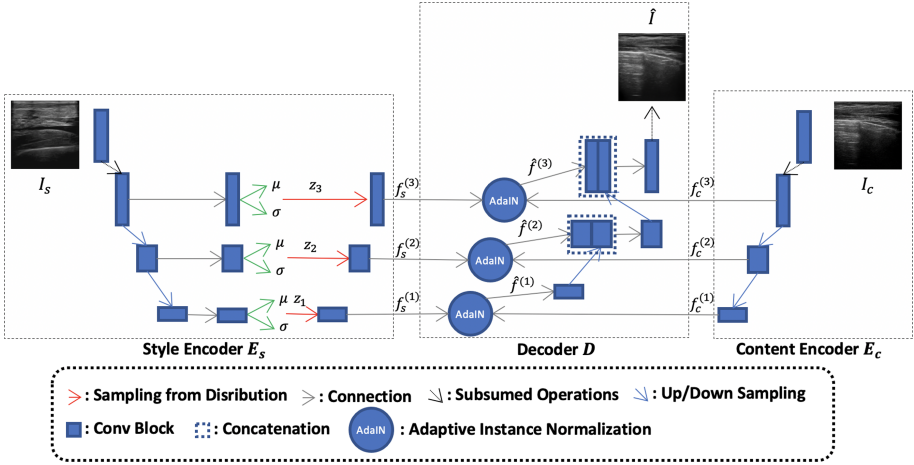
However, current works can only generate one result given a content and a style image, and few can sample the style from a latent space. Works in medical domain are mostly tested on images with similar styles, limiting the ability of generalization. As the style should follow a certain distribution instead of some specific values, we intent to generate multiple plausible output images given a content and a style image. Furthermore, we also want to sample the style in a continuous latent space so that we would be able to generate images from unseen styles. We propose a variational style transfer approach on medical images, which has the following contribution: (1) To the best of our knowledge, our method is the first variational style transfer approach that to explicitly sample the style from a latent space without giving the network a style reference image. (2) Our approach can be used to augment the data in ultrasound images, which results in better segmentation. (3) The method that we propose is able to transform the ultrasound images taken in one style to an unobserved style.

## 2 Methods

Our style transfer network consists of three parts: style encoder  $E_s$ , content encoder  $E_c$ , and decoder  $D$ . The network structure is shown in Fig. 1, where  $I_s$  is the style image,  $I_c$  is the content image and  $\hat{I}$  is the output image. During training, the decoder  $D$  learns to generate  $\hat{I}$  with the Gaussian latent variables  $\mathbf{z}$  conditioned on the content image  $I_c$ , while the style encoder  $E_s$  learns the distribution of  $\mathbf{z}$  given style image  $I_s$ . Before putting the three parts together, we pre-train the style encoder and the content encoder separately to provide better training stability. When generating images, our method can either use certain given style, or sample the style from the latent space.

### 2.1 Style Encoder

The style encoder  $E_s$  is the encoder part of a U-Net [19] based Variational Autoencoder (VAE) [13]. The difference between our VAE and traditional ones



**Fig. 1.** The network structure of the proposed method, which consists of two encoders that process the style image and content image separately, and a decoder. The style encoder generates distributions for some latent variables given the style image  $q_\phi(\mathbf{z}|I_s)$ , to create style that could change continuously.

is that our latent variables are at different scales to generate images with better resolutions. The style encoder  $E_s$  approximates the distribution of latent variables at different scale. In other words, it learns distributions  $q_\phi(z_i|I)$  that approximates the intractable true distribution  $p_\theta(z_i|I)$ , where  $\phi$  is the variational parameters, while  $\theta$  is the generative model parameters. Therefore the variational lower bound could be written as:

$$\mathcal{L}(\theta, \phi; I_s) = -KL(q_\phi(\mathbf{z}|I_s)||p_\theta(\mathbf{z})) + \mathbb{E}_{q_\phi(\mathbf{z}|I_s)} \log p_\theta(I_s|\mathbf{z}) \quad (1)$$

where  $KL(\cdot)$  is the Kullback–Leibler (KL) divergence between two distributions, and we further assume  $\mathbf{z}|\theta \sim \mathcal{N}(0, 1)$ .

The encoder is further incorporated into the style transfer network while the decoder here is only used during initial training. The structure of the style encoder is shown in Fig. 1, while the decoder part is a traditional U-Net decoder.

## 2.2 Content Encoder

The structure of the content encoder  $E_c$  is shown in Fig. 1, and like the style encoder  $E_s$ , it is the encoder of a U-Net autoencoder, where the decoder is only used in the initial training with the exact opposite structure of the encoder.

## 2.3 Decoder

The decoder takes in the encoded style and content before generating a new image, and is also designed based on a traditional U-Net. The structure is shown

in Fig. 1. Denote the content features as  $f_c = E_c(I_c)$  and the style feature as  $f_s = E_s(I_s)$ . Also denote the content and style features at scale  $i$  as  $f_c^{(i)}$  and  $f_s^{(i)}$  respectively. The features of the output images at scale  $i$ ,  $\hat{f}^{(i)}$ , can be treated as the input into the decoder of a normal U-Net. Inspired by [8], we utilize AdaIN to perform style transfer in feature space at each scale to calculate  $\hat{f}^{(i)}$  based on  $f_c^{(i)}$  and  $f_s^{(i)}$ . The calculation can be expressed as:

$$\hat{f}^{(i)} = \text{AdaIN}(f_c^{(i)}, f_s^{(i)}) = \sigma(f_s^{(i)}(\frac{f_c^{(i)} - \mu(f_c^{(i)})}{\sigma(f_c^{(i)})})) + \mu(f_s^{(i)}) \quad (2)$$

## 2.4 Loss Functions

The training objective of the network is to generate an output image  $\hat{I}$  containing the contents in the content image  $I_c$  and having the style of the style image  $I_s$ , all while maximizing the variational lower bound. Therefore, the loss function is made up of three parts: perceptual loss, style loss, and KL divergence loss.

Perceptual and style losses, are based on high level features extracted by pre-trained VGGs [21] and were first utilized in [6]. Perceptual loss is the difference between two feature maps encoded by a CNN. Since spatial correlation is considered in perceptual loss  $L_p$ , it is deemed as an expression of the content similarity between two images, which can be expressed as follows:

$$L_p = \sum_{i=1}^{N_{VGG}} w_i^p \|\psi_i(I_c) - \psi_i(\hat{I})\| \quad (3)$$

where  $\psi_i(x)$  extracts the layer  $i$  of VGG from  $x$ ,  $w_i^p$  is the weight at layer  $i$  for perceptual loss, and  $N_{VGG}$  as the total number of layers in VGG.

Denote the number of channels, height, and width of the  $i$  th layer of the feature map as  $C_i$ ,  $H_i$ , and  $W_i$  respectively. We also denote the  $C_i \times C_i$  gram matrix of  $i$  th layer of the feature map of image  $x$  as  $G_i(x)$ . The gram matrix can be described as:

$$G_i(x)(u, v) = \frac{\sum_{h=1}^{H_i} \sum_{w=1}^{W_i} \psi_i(x)(h, w, u) \psi_i(x)(h, w, v)}{C_i H_i W_i} \quad (4)$$

Since the gram matrix only records the relationship between different channels rather than the spatial correlation, it is considered to be the representation of the general textures and patterns of an image. Let  $w_i^s$  be the weight of the loss at layer  $i$  for style loss, the style loss  $L_s$  is the distance of two gram matrices:

$$L_s = \sum_{i=1}^{N_{VGG}} w_i^s \|G_i(I_s) - G_i(\hat{I})\| \quad (5)$$

To maximize the variational lower bound, we need to minimize the KL divergence between  $q_\phi(\mathbf{z}|I_s)$  and  $p_\theta(\mathbf{z})$ . Under the assumption that all the latent variables are i.i.d., we calculate the KL divergence on each scale and sum over all the

KL divergence to get the KL loss. Since we already assume that  $\mathbf{z}|\theta \sim \mathcal{N}(0, 1)$ , we can derive the KL divergence at each scale as:

$$KL(q_\phi(\mathbf{z}|I_s)||p_\theta(\mathbf{z})) = \frac{1}{2} \left( \frac{\mu^2}{\sigma^2} + \sigma^2 - \log \sigma^2 - 1 \right) \quad (6)$$

where we assume  $\mathbf{z}|I_s; \phi \sim N(\mu, \sigma)$ .

## 2.5 Implementation Details

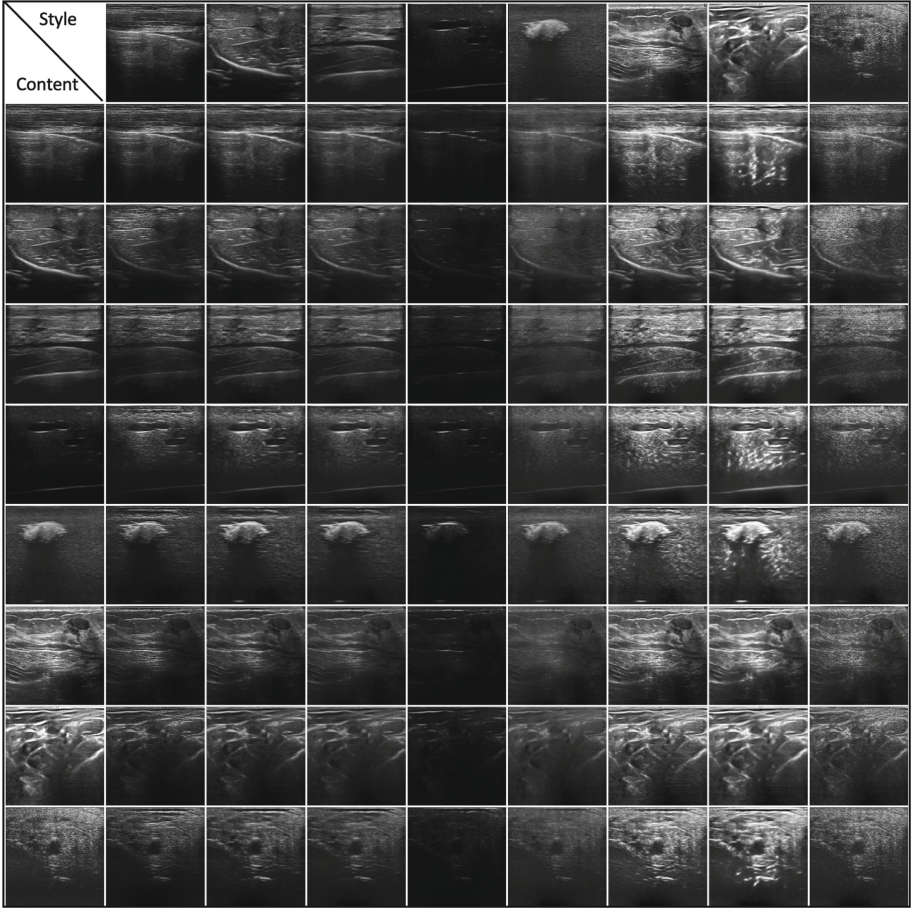
Encoders and decoders in the network follow the architecture below. There are 2 convolutional layers in each ConvBlock, followed by batch normalization [9] and swish activation [18]. We use 5 ConvBlocks in the encoder side which are followed by a max pooling layer, except for the last one. The number of convolutional filters are 64, 128, 256, 512, 1024 respectively. The decoders follow the inverse structure of the encoder. Note that in the style encoder, shown in Fig. 1, there is an additional ConvBlock before generating the distribution  $q_\phi(z_i|I_s)$  after the normal U-Net [19] encoder at each scale. There is also another ConvBlock after sampling  $z_i$  from  $q_\phi(z_i|I_s)$  to calculate  $f_s^{(i)}$  at each scale. The weights for perceptual and style loss are set to 0, 0, 0, 0, 0.01 and 0.1, 0.002, 0.001, 0.01, 10 for *block1\_conv1*, *block2\_conv1*, *block3\_conv1*, *block4\_conv1*, *block5\_conv1* of VGG respectively. The network is optimized via Adam optimizer [12] with a learning rate of  $5 \times 10^{-5}$  and trained for 20 epochs on a single Nvidia Titan RTX GPU.

## 3 Experiments

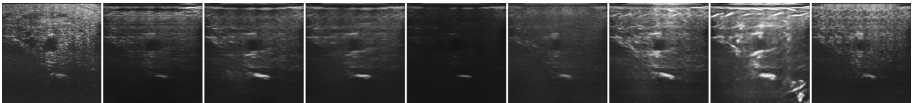
The images in the experiments are a combination of numerous datasets: (1) lung images on clinical patients by Sonosite ultrasound machine with HFL38xp linear transducer, (2) chicken breast images by UF-760AG Fukuda Denshi using a linear transducer with 51 mm scanning width (FDL), (3) live-pig images by FDL, (4) blue-gel phantom images by FDL, (5) leg phantom images by FDL, (6) Breast Ultrasound Images Dataset [2], (7) Ultrasound Nerve Segmentation dataset from Kaggle [1], (8) arteries and veins in human subjects by aVisualsonics Vevo 2100 UHFUS machine with ultrahigh frequency scanners. In total, there are 18308, 2285, 2283 images for training, validation, and testing respectively in the combined dataset. During training, we randomly select a pair of content and style images from the combined dataset without any additional restrictions.

### 3.1 Qualitative Results

In the first experiment, we directly generate the outputs given the content and style images. We transfer the style of images across (1)–(8). Shown in Fig. 2, the visual results are good in each combination of content and style images. Moreover, all the results still have the anatomy in the content images, including but not limited to vessels and ligaments, while looking like the style images. On



**Fig. 2.** Results of the algorithm given style and content images. We transfer content images from the first column to the style in the corresponding image in the first row. The content images from top to bottom and the style images from left to right are images from dataset (1)–(8).

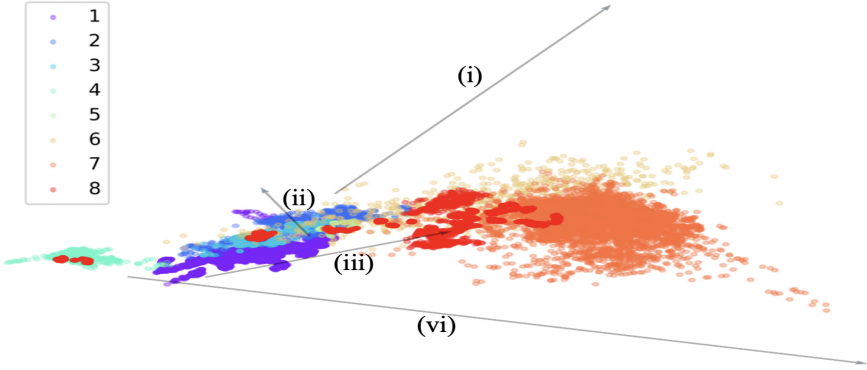


**Fig. 3.** Results of the method by Huang et al. [8]. The left is the content image (same as the last one in Fig. 2), whose style is transferred to styles of (1)–(8) (right 8 images).

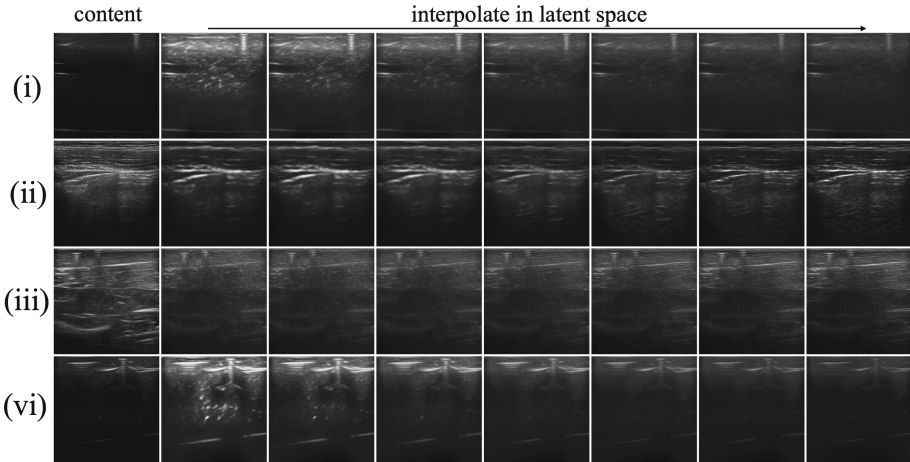
the contrary, shown in Fig. 3, method proposed by Huang et al. [8], is not able to capture the fine details in the content and generate realistic textures.

Another qualitative experiment is that we directly sample the style from the latent space without giving the model a style image. Figure 4 shows the





**Fig. 4.** The distribution (flattened to fit the paper) of styles in the latent space.



**Fig. 5.** Results of directly sampling from the latent space and interpolation. The first column are the content image. We sampled the second and the last column (starting point and ending point of the arrows in Fig. 4), then interpolate in between in the latent space in the direction of the corresponding arrows.

distribution of our training style images in the latent space. We then randomly sampled two styles (at the end points of the arrow in Fig. 4) from the latent space and interpolate between the two, and the results can be found in Fig. 5. We observe that even without a style image, the model can still generate visually reasonable ultrasound images given the content image without losing significant details while sampling from latent space that is not covered by the training data.

### 3.2 Quantitative Results

To show that our approach is effective in augmentation, assume that we only have limited live-pig data while having more leg-phantom data for veins and

**Table 1.** Comparison between style-transfer-based augmentation and the traditional augmentation

	Veins				Arteries			
	Dice	IoU	Pre.	Rec.	Dice	IoU	Pre.	Rec.
<i>vt-ph/pig-99/ aug.</i>	0.96	0.92	0.93	0.99	0.93	0.86	0.87	<b>0.99</b>
<i>vt-ph/pig-99</i>	<b>0.97</b>	<b>0.95</b>	<b>0.96</b>	<b>0.99</b>	<b>0.95</b>	<b>0.91</b>	<b>0.92</b>	<b>0.99</b>
<i>dt-ph/pig-99/ aug.</i>	0.94	0.90	0.92	0.97	0.92	0.84	0.88	0.96
<i>dt-ph/pig-99</i>	0.89	0.81	0.82	0.98	0.83	0.71	0.71	<b>0.99</b>
<i>pig-99/ aug.</i>	0.87	0.77	0.84	0.90	0.86	0.75	0.81	0.91
<i>pig-99</i>	0.83	0.71	0.73	0.96	0.81	0.68	0.70	0.95
<i>ph/pig-99/ aug.</i>	0.07	0.04	0.04	0.59	0.02	0.01	0.01	0.97
<i>ph/pig-99</i>	0.05	0.03	0.36	0.03	0.14	0.07	0.13	0.14
<i>vt-ph/pig-77/ aug.</i>	<b>0.95</b>	<b>0.91</b>	<b>0.95</b>	0.96	<b>0.93</b>	<b>0.87</b>	0.91	0.95
<i>vt-ph/pig-77</i>	0.93	0.88	0.91	0.96	0.90	0.81	<b>0.95</b>	0.90
<i>dt-ph/pig-77/ aug.</i>	0.93	0.88	0.89	<b>0.98</b>	0.87	0.77	0.78	0.98
<i>dt-ph/pig-77</i>	0.94	0.90	0.92	0.97	0.91	0.84	0.85	<b>0.99</b>
<i>pig-77/ aug.</i>	0.92	0.86	0.87	<b>0.98</b>	0.91	0.84	0.85	0.98
<i>pig-77</i>	0.88	0.78	0.90	0.86	0.68	0.51	0.71	0.65
<i>ph/pig-77/ aug.</i>	0.07	0.04	0.04	0.55	0.03	0.02	0.02	0.06
<i>ph/pig-77</i>	0.38	0.23	0.68	0.25	0.21	0.12	0.22	0.19
<i>vt-ph/pig-55/ aug.</i>	0.74	0.59	0.80	0.69	<b>0.79</b>	<b>0.65</b>	<b>0.73</b>	<b>0.87</b>
<i>vt-ph/pig-55</i>	<b>0.80</b>	<b>0.66</b>	0.87	<b>0.73</b>	0.59	0.41	0.65	0.54
<i>dt-ph/pig-55/ aug.</i>	0.70	0.53	0.71	0.69	0.53	0.36	0.59	0.48
<i>dt-ph/pig-55</i>	0.76	0.61	<b>0.88</b>	0.67	0.66	0.60	0.56	0.81
<i>pig-55/ aug.</i>	0.67	0.50	0.87	0.54	0.43	0.27	0.37	0.27
<i>pig-55</i>	0.15	0.01	0.09	0.43	0.01	0	0.01	0
<i>ph/pig-55/ aug.</i>	0.04	0.02	0.02	0.07	0.02	0.02	0.01	0.81
<i>ph/pig-55</i>	0.22	0.12	0.45	0.14	0.1	0.05	0.05	0.53
<i>vt-ph/pig-33/ aug.</i>	<b>0.54</b>	<b>0.37</b>	0.63	0.47	0.37	0.23	0.42	0.33
<i>vt-ph/pig-33</i>	0.43	0.28	0.38	<b>0.49</b>	<b>0.56</b>	<b>0.39</b>	<b>0.43</b>	<b>0.82</b>
<i>dt-ph/pig-33/ aug.</i>	0.11	0.06	0.13	0.10	0.06	0.03	0.03	0.37
<i>dt-ph/pig-33</i>	0.53	<b>0.37</b>	<b>0.68</b>	0.44	0.34	0.20	0.36	0.32
<i>pig-33/ aug.</i>	0.15	0.08	0.09	0.66	0.01	0	0	0.01
<i>pig-33</i>	0.06	0.03	0.06	0.06	0	0	0.01	0
<i>ph/pig-33/ aug.</i>	0.07	0.04	0.04	0.58	0.02	0.01	0.01	0.37
<i>ph/pig-33</i>	0.01	0	0.02	0	0	0	0	0

arteries (405 images). Denote *vt-ph*, *dt-ph* and *ph* as variational-style-transferred (ours), deterministic-style-transferred [8] and original phantom images respectively, *aug.* as traditional augmentation methods including gamma transform, Gaussian blurring, and image flipping. We further evaluate the effects of the number of the real images have on the final results. We transfer the leg-phantom images into the style of pig images. For live-pig data, we set 99 images as training set while having 11 images in each of the validation set and test set. In the experiments, we train U-Nets with the same network implementation and training settings as [19], on 99, 77, 55, 33 live-pig images denoted as *pig-99*,



*pig-77* etc. to show that our method works with very limited number of images. Note that, we balance the phantom data and pig data as roughly 1:1 ratio in each epoch. Shown in Table 1, the segmentation performs better by the networks trained on variational-style-transferred phantom images and pig images than the ones trained on other images in general. Additional augmentation on top of our variational-style-transfer augmentation sometimes improve the performance but our variational-style-transfer augmentation is an upgrade over the traditional augmentation and deterministic style transfer by Huang et al. [8]. Besides, when the number of real live-pig data is really limited, only style-transfer augmentation can produce a decent result. In any case, it can be seen that style-transfer-based approach has a significant improvement over traditional methods and our variational approach is superior than Huang et al.’s method [8].

## 4 Conclusion

We demonstrated that our method is capable of transferring the style of one ultrasound image to another style, e.g. from the style of phantom data to that of pig data, from the style of normal ultrasound machines to the style of high frequency ultrasound machines, etc.. Besides, it is also able to sample arbitrary and continuous style from a latent space. Our method can generate ultrasound images from both observed and unobserved domains, which helps address the insufficiency of data and labels insufficiency in medical imaging.

## References

1. Ultrasound Nerve Segmentation: Identify nerve structures in ultrasound images of the neck (2016). <https://www.kaggle.com/c/ultrasound-nerve-segmentation>
2. Al-Dhabyani, W., Gomaa, M., Khaled, H., Fahmy, A.: Dataset of breast ultrasound images. Data Brief **28**, 104863 (2020)
3. Byra, M.: Discriminant analysis of neural style representations for breast lesion classification in ultrasound. Biocybern. Biomed. Eng. **38**(3), 684–690 (2018)
4. Ching, T., et al.: Opportunities and obstacles for deep learning in biology and medicine. J. R. Soc. Interface **15**(141), 20170387 (2018)
5. Fetty, L., et al.: Latent space manipulation for high-resolution medical image synthesis via the styleGAN. Z. Med. Phys. **30**(4), 305–314 (2020)
6. Gatys, L.A., Ecker, A.S., Bethge, M.: A neural algorithm of artistic style. arXiv preprint [arXiv:1508.06576](https://arxiv.org/abs/1508.06576) (2015)
7. Gatys, L.A., Ecker, A.S., Bethge, M., Hertzmann, A., Shechtman, E.: Controlling perceptual factors in neural style transfer. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 3985–3993 (2017)
8. Huang, X., Belongie, S.: Arbitrary style transfer in real-time with adaptive instance normalization. In: Proceedings of the IEEE International Conference on Computer Vision, pp. 1501–1510 (2017)
9. Ioffe, S., Szegedy, C.: Batch normalization: accelerating deep network training by reducing internal covariate shift. arXiv preprint [arXiv:1502.03167](https://arxiv.org/abs/1502.03167) (2015)

10. Johnson, J., Alahi, A., Fei-Fei, L.: Perceptual losses for real-time style transfer and super-resolution. In: Leibe, B., Matas, J., Sebe, N., Welling, M. (eds.) ECCV 2016. LNCS, vol. 9906, pp. 694–711. Springer, Cham (2016). [https://doi.org/10.1007/978-3-319-46475-6\\_43](https://doi.org/10.1007/978-3-319-46475-6_43)
11. Karras, T., Laine, S., Aila, T.: A style-based generator architecture for generative adversarial networks. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 4401–4410 (2019)
12. Kingma, D.P., Ba, J.: Adam: a method for stochastic optimization. arXiv preprint [arXiv:1412.6980](https://arxiv.org/abs/1412.6980) (2014)
13. Kingma, D.P., Welling, M.: Auto-encoding variational bayes. arXiv preprint [arXiv:1312.6114](https://arxiv.org/abs/1312.6114) (2013)
14. Li, Y., Wang, N., Liu, J., Hou, X.: Demystifying neural style transfer. arXiv preprint [arXiv:1701.01036](https://arxiv.org/abs/1701.01036) (2017)
15. Liu, Z., et al.: Remove appearance shift for ultrasound image segmentation via fast and universal style transfer. In: 2020 IEEE 17th International Symposium on Biomedical Imaging (ISBI), pp. 1824–1828. IEEE (2020)
16. Ma, C., Ji, Z., Gao, M.: Neural style transfer improves 3D cardiovascular MR image segmentation on inconsistent data. In: Shen, D., et al. (eds.) MICCAI 2019. LNCS, vol. 11765, pp. 128–136. Springer, Cham (2019). [https://doi.org/10.1007/978-3-030-32245-8\\_15](https://doi.org/10.1007/978-3-030-32245-8_15)
17. Mikołajczyk, A., Grochowski, M.: Style transfer-based image synthesis as an efficient regularization technique in deep learning. In: 2019 24th International Conference on Methods and Models in Automation and Robotics (MMAR), pp. 42–47. IEEE (2019)
18. Ramachandran, P., Zoph, B., Le, Q.V.: Swish: a self-gated activation function. arXiv preprint [arXiv:1710.05941](https://arxiv.org/abs/1710.05941) 7 (2017)
19. Ronneberger, O., Fischer, P., Brox, T.: U-net: convolutional networks for biomedical image segmentation. In: Navab, N., Hornegger, J., Wells, W.M., Frangi, A.F. (eds.) MICCAI 2015. LNCS, vol. 9351, pp. 234–241. Springer, Cham (2015). [https://doi.org/10.1007/978-3-319-24574-4\\_28](https://doi.org/10.1007/978-3-319-24574-4_28)
20. Shen, F., Yan, S., Zeng, G.: Neural style transfer via meta networks. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 8061–8069 (2018)
21. Simonyan, K., Zisserman, A.: Very deep convolutional networks for large-scale image recognition. arXiv preprint [arXiv:1409.1556](https://arxiv.org/abs/1409.1556) (2014)
22. Singh, K., Drzewicki, D.: Neural style transfer for medical image augmentation. [https://scholar.google.com/scholar?hl=en&as\\_sdt=0%2C5&q=Singh%2C+K.%2C+Drzewicki%2C+D.%3A+Neural+style+transfer+for+medical+image+augmentation&btnG=](https://scholar.google.com/scholar?hl=en&as_sdt=0%2C5&q=Singh%2C+K.%2C+Drzewicki%2C+D.%3A+Neural+style+transfer+for+medical+image+augmentation&btnG=)
23. Ulyanov, D., Lebedev, V., Vedaldi, A., Lempitsky, V.S.: Texture networks: feed-forward synthesis of textures and stylized images. In: ICML, vol. 1, p. 4 (2016)
24. Xu, Y., Li, Y., Shin, B.-S.: Medical image processing with contextual style transfer. *Hum.-Cent. Inf. Sci.* **10**(1), 1–16 (2020). <https://doi.org/10.1186/s13673-020-00251-9>
25. Zheng, X., Chalasani, T., Ghosal, K., Lutz, S., Smolic, A.: STaDA: style transfer as data augmentation. arXiv preprint [arXiv:1909.01056](https://arxiv.org/abs/1909.01056) (2019)