UNIVERSITY OF LEEDS

This is a repository copy of *Remove Appearance Shift for Ultrasound Image Segmentation via Fast and Universal Style Transfer*.

Proceedings Paper:

White Rose
university consortium
Universities of Leeds, Sheffield & York

eprints@whiterose.ac.uk
https://eprints.whiterose.ac.uk/

REMOVE APPEARANCE SHIFT FOR ULTRASOUND IMAGE SEGMENTATION VIA FAST AND UNIVERSAL STYLE TRANSFER

Zhendong Liu[1,2,†], *Xin Yang*[1,2,†], *Rui Gao*[1,2], *Shengfeng Liu*[1,2], *Haoran Dou*[1,2], *Shuangchi He*[1,2], *Yuhao Huang*[1,2], *Yankai Huang*[3], *Huanjia Luo*[3], *Yuanji Zhang*[3], *Yi Xiong*[3], *Dong Ni*[1,2,*]

[1]National-Regional Key Technology Engineering Laboratory for Medical Ultrasound, Guangdong Key Laboratory for Biomedical Measurements and Ultrasound Imaging, School of Biomedical Engineering, Health Science Center, Shenzhen University, Shenzhen, China
[2]Medical UltraSound Image Computing (MUSIC) Lab, Shenzhen University, Shenzhen, China
[3]Department of Ultrasound, Luohu People's Hosptial, Shenzhen, China

ABSTRACT

Deep Neural Networks (DNNs) suffer from the performance degradation when image appearance shift occurs, especially in ultrasound (US) image segmentation. In this paper, we propose a novel and intuitive framework to remove the appearance shift, and hence improve the generalization ability of DNNs. Our work has three highlights. First, we follow the spirit of universal style transfer to remove appearance shifts, which was not explored before for US images. Without sacrificing image structure details, it enables the arbitrary style-content transfer. Second, accelerated with Adaptive Instance Normalization block, our framework achieved real-time speed required in the clinical US scanning. Third, an efficient and effective style image selection strategy is proposed to ensure the target-style US image and testing content US image properly match each other. Experiments on two large US datasets demonstrate that our methods are superior to state-of-the-art methods on making DNNs robust against various appearance shifts.

Index Terms— Style transfer, Image segmentation, Ultrasound image, Appearance shift, Generalization ability
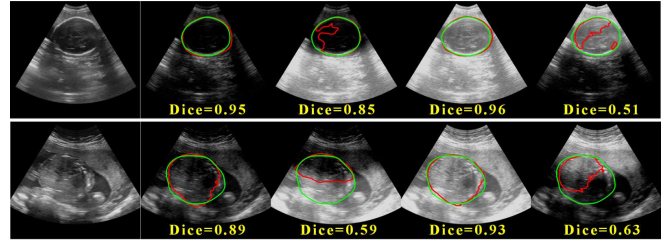
1. INTRODUCTION

In recent years, Deep Neural Networks (DNNs) have dominated the field of medical image analysis [1]. However, DNNs have the performance degradation when image appearance shift exists between the model training and testing phases. Therefore, DNNs cannot guarantee the accuracy in real clinical scenarios [2, 3].

Fig. 1. Original US images and four variants with different TGC. Two groups US images for the fetal head (top) and fetal abdomen (bottom). Yellow digits denote the Dice. Green and red curves represent ground truth and the segmentation results, respectively.

The situation becomes more severe for DNNs based ultrasound (US) image segmentation. As shown in Fig.1, there are two groups of US images from the prenatal scan. With different Time Gain Compensation (TGC) settings, tissues intensity in different depths are affected. The original US image changes into variants with obvious and typical appearance shift. A well-trained DNN then presents dramatically different segmentation results on the variants. Therefore, making DNNs robust against appearance shift on US images is highly desired. However, this task is non-trivial. First, DNN models in clinical US analysis are facing an open scenario, where DNNs only have access to limited source domain images/labels, and are blinded to target domain samples. Second, varitions in acoustic attenuation, scan operators, US machines, probes and empirical imaging parameters often make appearance shift unpredictable and hard to model.

Recently, Domain Adaptation (DA) was frequently used to remove image appearance shift. Kamnitsas et al. [4] proposed to align domain features, but it needs a large number of images and labels from the target domain, which is infeasible in clinic scenario. Based on Cycle-GAN [5], translating the image appearance among domains [6] and shape guided

image translation [7] were proposed. However, GAN-based methods often introduce artifacts into the original image and make the image distorted. In addition, DA often limits itself among a fixed number of domains with limited appearance shift. This is not suitable for the open scenario with unpredictable image appearance shift. By revisiting the style transfer [8], Ma et al. built an online scheme to remove appearance shift in the cardiovascular magnetic resonance (MR) image segmentation [9]. However, their style transfer method is unstable and may destroy image details. In [10], a new style transfer method, called STROTSS, was proposed to preserve image details for any style-content pairs during transfer. STROTSS is promising but it is time-consuming. It also outputs random results due to the resampling of hypercolumns.

In this paper, on the basis of Wavelet Corrected Transfer network (WaveCT) [11] and Adaptive Instance Normalization (AdaIN) [12], we propose a novel and general style transfer framework to remove US image appearance shift (denoted as *WaveCT-AIN*). Our framework has three highlights: (a) It is the first work to explore universal style transfer for US image segmentation. WaveCT-AIN preserves the image structure details and enables arbitrary style-content transfer between images. Unpredictable US image appearance shifts are consistently tackled under this framework. (b) Compared to STROTSS, our WaveCT-AIN is lightweight and satisfies the real-time requirements. With AdaIN block for acceleration, it only takes 0.07s for an inference. (c) An efficient and effective style image selection strategy is also adopted to retrieve suitable and appearance-invariant US style image for optimized style transfer. Experiments on two US datasets demonstrate that the proposed WaveCT-AIN obtained a superior performance than the state-of-the-art methods.

2. METHODOLOGY

The proposed framework is shown in Fig. 2. It consists of a style selection module S_c, a style transfer module S_t (WaveCT-AIN-D) and a segmentor S_g. For a testing content image, S_c retrieves an US style image from source domain as a reference and inputs it to S_t. Then, S_t instantly removes the appearance shift in content image and makes it suitable for S_g to segment.

2.1. Universal and High Quality Style Transfer

Style transfer can remove appearance shift to approach robust US image analysis. To meet clinical US scanning requirements, following aspects should be considered: (a) The method should enable *universal* transfer between any style-content image pairs. (b) Transfer results should be *stable* rather than fluctuant [9, 10]. (c) Image structure details should be highly preserved. (d) Transfer process should be *fast* to avoid high latency[10]. (e) The method should provide *style selection* strategy to provide case-specific style transfer.
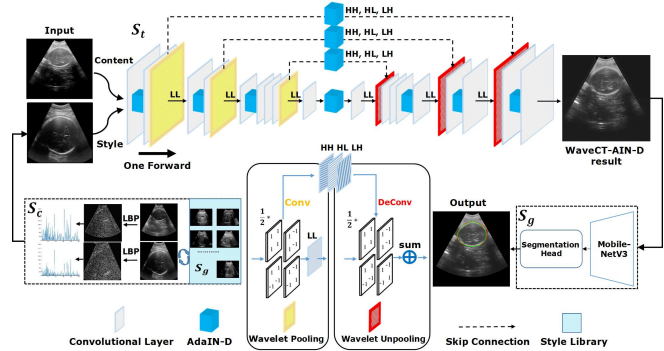

Fig. 2. Overview of the proposed framework.

In order to meet all the above needs, we adopt the WaveCT network recently used in WCT^2 [11]. As shown in Fig. 2, style image and content image are hierarchically processed by the encoder-decoder network in WaveCT. At several sites, multi-scale style information is distilled from the feature maps of style image. Style information is then transferred to modulate the feature maps of content image by style transform block. With the modulated features, decoder reconstructs the stylized content image.

Different from previous online style transfer methods which may distort image details [9], the proposed WaveCT has two main advantages. *First*, WaveCT replaces vanilla max-pooling/unpooling layers with the Haar wavelet pooling/unpooling layers, which can exactly reconstruct the stylized image with minimal structure loss. *Second*, WaveCT splits the features into low-frequency and high-frequency components via Haar wavelet pooling. The low-frequency captures smooth textures while the high-frequency extracts edge-like features. WaveCT only passes the low-frequency in the main network and skips the high-frequency to the decoder. Due to this layout, WaveCT with style transform block can achieve universal and high quality stylization as we expected.

2.2. Accelerating the Style Transfer

WaveCT associated with the style transform block, i.e., whitening and coloring transforms (WCT) [13], is the workhorse of WCT^2[11]. Although effective, WCT^2 has high computation latency. Specifically, WCT firstly applies whitening transformation to erase the style information in the content feature map, and then a coloring transformation to render the texture of style image to the whitening result. Since these two steps involve heavy computation among multiple large matrices, WCT^2 needs about 4.58 seconds for an inference, which is not acceptable in US image analyses.

To accelerate the transfer, we propose to upgrade all the WCT block in WCT^2 to a new block, AdaIN [12]. The modified system, i.e., WaveCT with AdaIn, is denoted as *WaveCT-AIN*. AdaIN is a variant of Instance Normalization. It only needs to align the channel-wise mean μ and standard devia-

tion σ of content feature maps x to match those of target-style feature maps y:

$$AdaIN(x,y) = \sigma(y)\left(\frac{x - \mu(x)}{\sigma(x)}\right) + \mu(y) \qquad (1)$$

AdaIN is lightweight, which significantly reduces the computation cost. Considering the fact that TGC changing along the depth is the main reason for the appearance shift in US image, we further modify the AdaIN into a depth-encoded version, *AdaIN-D*. Specially, as illustrated in Fig.3, where a window sliding along the depth is employed (Fig.3(a)). The bandwidth and stride are set to two-thirds and one third height of content image respectively. Then we perform AdaIN transform for each region x_i (i=1,2) and average the overlap. WaveCT with AdaIN-D is denoted as WaveCT-AIN-D (see Fig.2).

(a) Content image   (b) Style image   (c) WaveCT-AIN-D result

Fig. 3. (a) A content image split into two regions x_1 and x_2 with a half overlap. (b) A style image y. (c) Our WaveCT-AIN-D result.

2.3. Case-specific Style Image Selection

Due to the unpredictable appearance shift in testing images, style image should be adaptively determined to provide case-specific guidance. To achieve this, we propose a Local Binary Patterns (LBP) feature [14] based histogram matching strategy to improve the selection of style image. LBP is superior in capturing image textures and invariant to the changes in grayscale and rotation.

Based on the style image library from the source domain, the proposed strategy includes three steps. First, we obtain the histograms of testing content image c and all the style images in the library according to the corresponding LBP feature spectrum. Second, the top-10 relevant style images are retrieved via the LBP histogram correlation. Third, to further eliminate the gap between the testing image and the retrieved style images, we localize the final style image with the smallest Euclidean distance.

Since AdaIN is sensitive to the mean and variance of style image, in the third step above, we require the mean and variance of the target style image S_T to be as close as possible to those of content image c instead.

$$S_T = \underset{s_i \in \{s_1,...,s_{10}\}}{\arg\min} \|\mu(s_i) - \mu(c)\|_2 + \|\sigma(s_i) - \sigma(c)\|_2 \quad (2)$$

where s_i is each retrieved style image from the second step. μ and σ are the mean and variance of the whole image.

Our proposed style selection strategy is effective and not only applicable to our WaveCT-AIN-D, but also to the classic method in [8]. In Fig. 4, we show the sorted segmentation results on an testing content image which is transferred from 222 style library images. Among all the results, our strategy hits a proper style image and generates superior segmentation results (red diamond), which indicates that our selected style can properly help remove the appearance shift.
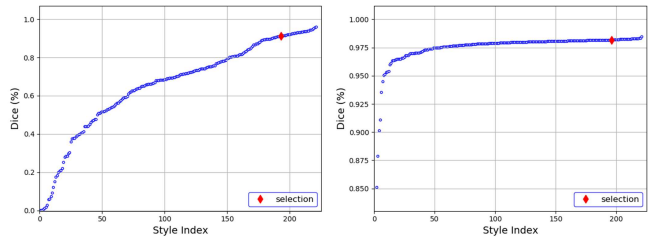
Fig. 4. Our style selection method in (a) Gatys et al.[8] and (b) WaveCT-AIN-D for stylization results of a content image. Y-axis is the sorted Dice with corresponding style image index.

3. EXPERIMENTAL RESULTS

3.1. Datasets and Implementation Details

Datasets. Our methods were tested on 180 groups of fetal head (FH) and 198 groups of fetal abdomen (FA) US images. Each group consists of one original US image and four variants with different TGC (see Fig.1). Approved by local IRB, all the images were anonymized and obtained by experienced experts using a Siemens system. Testing images cover the gestational age from 20 to 30 weeks. An expert provided delineation ground truth for all the images. Furthermore, 222 FH and 200 FA US images acquired by a GE scanner serve as the style image library.

Implementation Details. We freeze the encoder of WaveCT-AIN-D with VGG encoder weights. The decoder is firstly pre-trained on Microsoft COCO dataset and then fine-tuned with 3,000 US images, minimizing the L2 reconstruction loss. We use Adam optimizer with learning rate 10^{-3}. In the style selection, we use LBP with uniform pattern and set the neighbour points to 8, the radius to 3. For the segmentation model, we choose the MobileNetV3-Large with a segmentation head as our segmentor backbone. Adam with learning rate 10^{-4} is adopted to train segmentor using an augmented source domain training set of 10,000 US images. Segmentor is pre-trained on the source domain without TGC change and frozen during testing.

3.2. Results

The quantitative comparisons between WaveCT-AIN-D and other methods on FA and FH ultrasound (US) images are summarized in Table 1 and 2, respectively. Specifically, "SS" denotes the method with our style selection, "fix" denotes the method evaluated on the average of all the style library images based style transfer results.

Table 1. Quantitative comparison on FA US segmentation.

Methods	Dice(%)	Hdb(pixel)	Jaccard(%)	SSIM(%)	PSNR	Time(s)
No processing	85.60	21.83	79.27	/	/	/
HE	80.20	27.26	73.01	59.21	11.20	0.002
Cyclegan[5]	87.37	15.42	82.51	36.70	21.21	0.003
STROTSS[10]	88.71±1.22	15.34±1.88	83.21±0.83	41.34±2.55	20.26±1.41	181
Gatys et.al[8](fix)	70.13±2.31	42.32±3.50	65.56±1.42	24.23±3.30	16.10±2.03	3.19
Gatys et.al(SS)	88.62±0.87	15.32±1.55	83.25±0.51	62.50±1.42	28.90±1.20	3.20
WCT^2[11](SS)	90.50	14.21	83.79	74.51	25.42	4.58
WaveCT-AIN(SS)	93.42	**7.94**	89.44	78.55	28.23	**0.07**
WaveCT-AIN-D(fix)	92.34	9.83	88.52	79.43	29.34	0.10
WaveCT-AIN-D(SS)	**94.05**	8.32	**90.09**	**79.81**	**30.63**	0.11

Table 2. Quantitative comparisons on FH US segmentation.

Methods	Dice(%)	Hdb(pixel)	Jaccard(%)	SSIM(%)	PSNR	Time(s)
No processing	91.10	16.28	86.39	/	/	/
HE	85.03	20.44	80.49	59.15	11.04	0.002
Cyclegan[5]	93.36	9.02	90.05	36.66	21.14	0.003
STROTSS[10]	92.53±0.83	10.75±1.61	87.79±0.39	40.94±2.86	19.88±1.58	179
Gatys et.al[8](fix)	82.17±1.52	30.02±2.69	76.08±1.02	24.45±3.10	16.15±2.04	3.18
Gatys et.al(SS)	93.55±0.35	12.51±1.76	91.57±0.22	61.30±1.04	28.70±1.12	3.19
WCT^2[11](SS)	95.65	8.78	92.46	74.11	25.53	4.55
WaveCT-AIN(SS)	96.83	**6.91**	93.99	78.63	28.34	**0.08**
WaveCT-AIN-D(fix)	96.12	8.72	93.54	79.25	29.35	0.10
WaveCT-AIN-D(SS)	**97.31**	7.54	**94.50**	**79.65**	**30.44**	0.11

Runtime comparison. The last columns in Table 1 and 2 show the runtime comparisons among different methods using a single GTX 1080 Ti. We used the same hyperparameter settings in [9] for Gatys *et al.* [8] and followed the original settings in STROTSS [10]. The online style transfer methods above require much time to iteratively reconstruct a fluctuating stylized result, which goes against the clinical requirements. Because AdaIN [12] is much simpler than WCT [13], the proposed WaveCT-AIN has improved the speed around 60 times when compared to WCT^2 [11].

Visual quality comparison. To evaluate the visual quality, we calculated the structural similarity (SSIM) and Peak Signal to Noise Ratio (PSNR) between original images and stylized results. Based on the SSIM and PSNR in Table 1 and 2, we can see that the visual quality of our method is superior to other methods. Moreover, in Gatys *et al.* [8], the visual quality of stylized results with average style transfer (fix) is much worse than the results based on style selection (SS) method. In contrast, the proposed WaveCT-AIN-D is robust to the style change.

Segmentation performance comparison. As shown in Table 1 and 2, histogram equalization (HE) still suffer from appearance shift, but the proposed WaveCT-AIN-D (SS) obtain the highest Dice of 94.05% and 97.31% for FA and FH, respectively, which are also higher than without processing (91.0%). Compared to other methods, our proposed methods have also achieved significant improvement in terms of Hausdorff distance boundary (Hdb) and Jaccard. Fig.5 demonstrates the effectiveness of WaveCT-AIN-D.

(a) Origin, FH   (b) WaveCT-AIN-D, FH   (c) Origin, FA   (d) WaveCT-AIN-D, FA
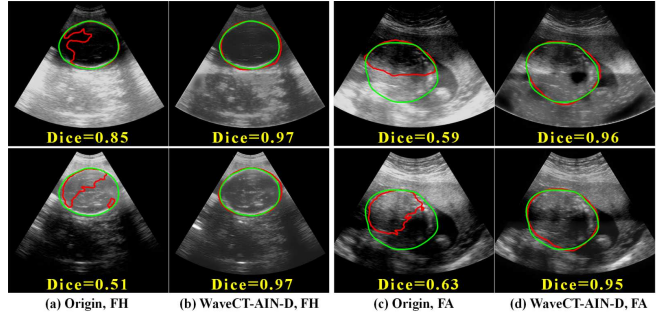
Fig. 5. Examples of segmentation results using WaveCT-AIN-D. (a) and (c) are FH and FA ultrasound images with different TGC, respectively, while (b) and (d) are corresponding WaveCT-AIN-D results.

4. CONCLUSIONS

In this paper, we proposed a novel style transfer framework, called WaveCT-AIN-D, to eliminate image appearance shift in US image segmentation. The proposed method is faster, more stable and efficient than the state-of-the-art methods. In addition, we explored a style selection method to match style-content pairs appropriately. Future study may focus on validating the proposed framework on other modalities and clinical practice.

5. REFERENCES

[1] Shengfeng Liu, Yi Wang, Xin Yang, Baiying Lei, Li Liu, Shawn Xiang Li, Dong Ni, and Tianfu Wang, "Deep learning in medical ultrasound analysis: A review," *Engineering*, vol. 5, no. 2, pp. 261 – 275, 2019.

[2] William W Stead, "Clinical implications and challenges of artificial intelligence and deep learning," *JAMA*, vol. 320, no. 11, pp. 1107–1108, 2018.

[3] Xin Yang, Haoran Dou, Ran Li, Xu Wang, Cheng Bian, Shengli Li, Dong Ni, and Pheng-Ann Heng, "Generalizing deep models for ultrasound image segmentation," in *MICCAI*. Springer, 2018, pp. 497–505.

[4] Konstantinos Kamnitsas et al., "Unsupervised domain adaptation in brain lesion segmentation with adversarial networks," in *MICCAI*. Springer, 2017, pp. 597–609.

[5] Jun-Yan Zhu, Taesung Park, et al., "Unpaired image-to-image translation using cycle-consistent adversarial networks," in *ICCV*, 2017, pp. 2223–2232.

[6] Yuankai Huo, Zhoubing Xu, et al., "Adversarial synthesis learning enables segmentation without target modality ground truth," in *ISBI*. IEEE, 2018, pp. 1217–1220.

[7] Cheng Chen, Qi Dou, Hao Chen, and Pheng-Ann Heng, "Semantic-aware generative adversarial nets for unsupervised domain adaptation in chest x-ray segmentation," *arXiv preprint arXiv:1806.00600*, 2018.

[8] Leon A Gatys, Alexander S Ecker, and Matthias Bethge, "Image style transfer using convolutional neural networks," in *CVPR*, 2016, pp. 2414–2423.

[9] Chunwei Ma, Zhanghexuan Ji, and Mingchen Gao, "Neural style transfer improves 3d cardiovascular mr image segmentation on inconsistent data," *arXiv preprint arXiv:1909.09716*, 2019.

[10] Nicholas Kolkin, Jason Salavon, et al., "Style transfer by relaxed optimal transport and self-similarity," in *CVPR*, 2019, pp. 10051–10060.

[11] Jaejun Yoo, Youngjung Uh, Sanghyuk Chun, et al., "Photorealistic style transfer via wavelet transforms," *arXiv preprint arXiv:1903.09760*, 2019.

[12] Xun Huang and Serge Belongie, "Arbitrary style transfer in real-time with adaptive instance normalization," in *ICCV*, 2017, pp. 1501–1510.

[13] Yijun Li, Chen Fang, et al., "Universal style transfer via feature transforms," in *NeurIPS*, 2017, pp. 386–396.

[14] Timo Ojala, Matti Pietikäinen, and Topi Mäenpää, "Multiresolution gray-scale and rotation invariant texture classification with local binary patterns," *IEEE TPAMI*, , no. 7, pp. 971–987, 2002.