2020 25th International Conference on Pattern Recognition (ICPR)
Milan, Italy, Jan 10-15, 2021

Attentional Wavelet Network for Traditional Chinese Painting Transfer

Rui Wang[†]*, Huaibo Huang[†], Aihua Zheng*, Ran He[†]
[†]Institute of Automation, Chinese Academy of Sciences
Email:{rui.wang, huaibo.huang}@cripac.ia.ac.cn, rhe@nlpr.ia.ac.cn
*School of Computer Science and Technology, Anhui University
Email:ahzheng214@foxmail.com

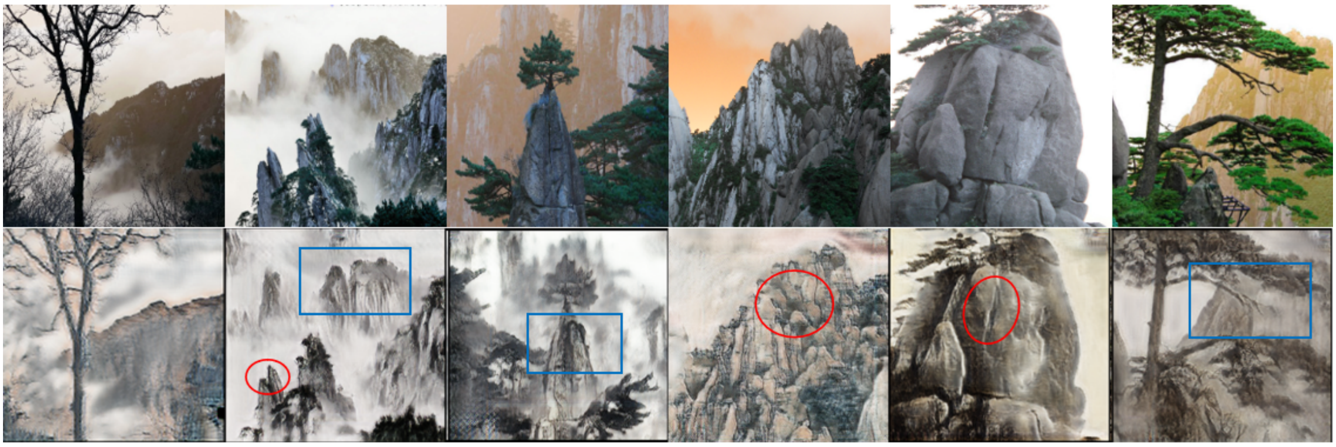


Fig. 1: Results of our AWNet. The first and the second rows represent the photos and the transferred traditional Chinese paintings, respectively. Traditional Chinese painting[1] contains the arts of 'Gongbi'[2] and 'Xieyi'[3] , as marked with red circles and blue boxes respectively.

Abstract—**Traditional Chinese paintings pay more attention to 'Gongbi' and 'Xieyi' in artworks, which raises a challenging task to generate Chinese paintings from photos. 'Xieyi' creates high-level conception for paintings, while 'Gongbi' refers to portraying local details in paintings. This paper proposes an attentional wavelet network for photo to Chinese painting transferring. We first introduce wavelets to obtain high-level conception and local details in Chinese paintings via 2-D haar wavelet transform. Moreover, we design high-level transform stream and local enhancement stream to dispose high frequencies and low frequency respectively. Furthermore, we exploit self-attention mechanism to compatibly pick up high-level information which is used to remedy the missing details when reconstructing the Chinese painting. To advance our experiment, we set up a new dataset named P2ADataset, with diverse photos and Chinese paintings on famous mountains around China. Experimental results comparing with the state-of-the-art style transferring algorithms verify the effectiveness of the proposed method. We will release the codes and data to the public.**

I. Introduction

Despite of the vast majority of style transfer tasks on western artistics (e.g. oils painting, abstract painting), the photo to Chinese painting transferring emerges due to its long history with charming artistic conception [1], [2], [3]. Different from the flexible use of color in western artworks, traditional Chinese paintings tend to convey a profound artistic conception(high-level information) and portray the details meticulously in the paintings which called 'Xieyi' and 'Gongbi' respectively. Therefore, 'Gongbi' and 'Xieyi' are the two core indicators to reflect the prospect of Chinese paintings. However, it is frustrated to depict the mental conceptions of 'Xieyi' and 'Gongbi' Chinese paintings via computer vision methods and machine learning algorithms, which brings the key challenge to photo to Chinese painting transferring task. Recent efforts have made great progress on photo to Chinese painting transferring. [1] propose a novel 'MXDoG-guided' filter base on 'eXtended Difference-of-Gaussians', which captures style information (corresponding to 'Xieyi' in this paper) and introduces three constraints to train the network. However, they can only capture the sub-styles information, such as 'line', 'ink wash' and 'black-leaving' etc. [2] propose a ChipGAN which simulates the three commonly used techniques in Chinese paintings, i.e. 'void', 'brush stroke', and 'ink wash tone and diffusion'. Specifically, to obtain style information (corresponding to 'Xieyi' in this paper), they create 'void' constraint via adversarial loss [4], [5], [6] between the generated and the real paintings. As for local details (corresponding to 'Gongbi' in this paper), they design a 'brush stroke' constraint by comparing the edges between the photo and the generated painting. In addition, they add 'ink wash' constraint to preserve

[1] https://en.wikipedia.org/wiki/Chinese_painting

[2] means 'meticulous' which uses delicate brushstrokes to portray the details in stylized images

[3] means exploiting concise lines and link to depict the expression of things

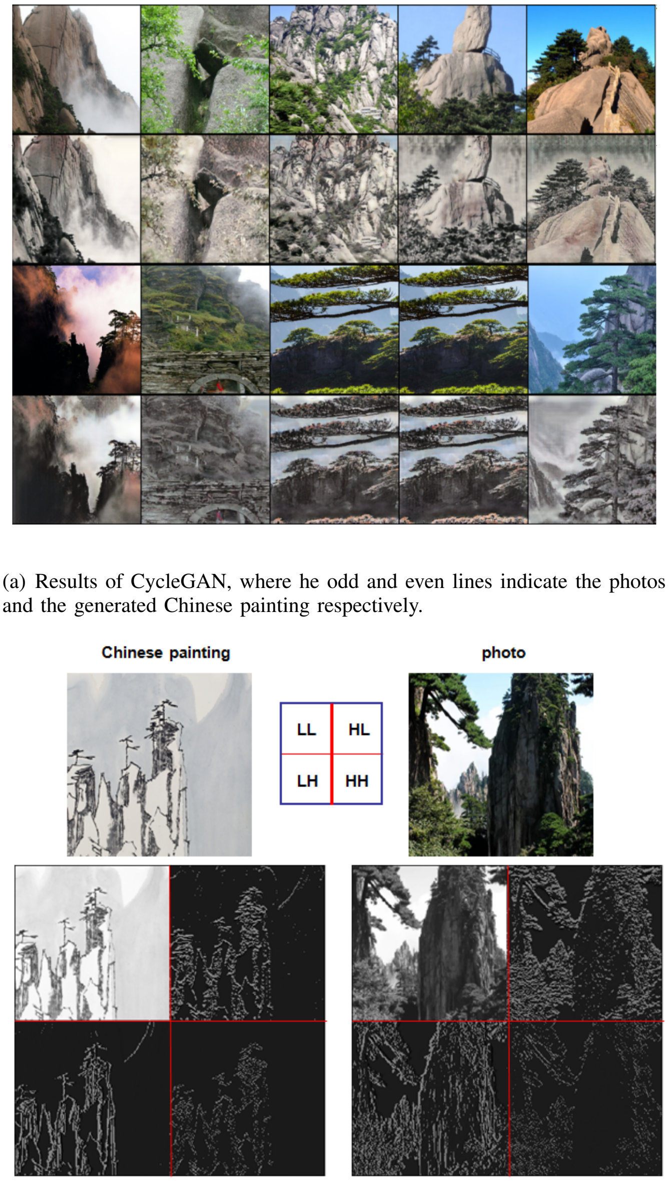
2020 25th International Conference on Pattern Recognition (ICPR) | 978-1-7281-8808-9/21/$31.00 ©2021 IEEE | DOI: 10.1109/ICPR48806.2021.9413063

’ink wash’ in the generated painting. Different from these two works that transfer from the photo, [3] investigate transferring from sketches and propose a multi-scale GAN for Chinese painting transferring, which explain the importance of details (correspond to ’Gongbi’ in this paper) in photos to Chinese paintings transferring task to some extent. However, it is often challenging to simultaneously take into account ’Gongbi’ and ’Xieyi’ in the Chinese painting. Furthermore, the spatial domain transfer methods always bring ’over grayscale’ results, as shown in Fig. 2(a). Recently, Yoo *et al.* [7] propose a WCT^2 model and use the low frequency and high frequencies to capture/reconstruct surface and detail information respectively during style transfer. However, they exploit wavelets via replacing pooling and unpooling layers with ’wavelet filter’. We observe that the low frequency in Chinese paintings always reflects meaningful high-level information i.e. ’Xieyi’ while high frequencies of photos contain rich local details i.e. ’Gongbi’ as shown in Fig. 2(b). In addition, ’Gongbi’ information normally scatters in the different features for each layer, in order to integrate the scattered ’Gongbi’ information layer by layer during the reconstruction, we introduce multi-scale self-attention mechanism. Beyond that, traditional Chinese painting dataset [1] is relatively small and partly watermarked, which lead to unsatisfied results when we train our model on this small-scale unclear dataset. Herein, we propose a new and large unpaired photo to Chinese painting transfer dataset named P2ADataset, for better evaluation. Based on above discussion, we propose a novel Attentional Wavelet Network (AWNet) that consists of high-level transform stream and local enhancement stream for Chinese painting transfer. The former is in charge of transforming photos to Chinese paintings while the latter disposes local information. In particular, we obtain ’over grayscale’ stylized images as shown in Fig. 2(a) when we feed photo into our base model. To improve it, we add low frequency to obtain more poetic Chinese paintings. To enhance details of stylized image, we propose a local enhancement stream to capture local details. Moreover, considering the diverse contribution of high frequencies, we introduce a multi-scale self-attention modules to assign different weights to each feature. Our contributions are summarized as follows:

- We propose a novel AWNet for photos-Chinese paintings transferring task, which can capture high-level information and local prospects simultaneously.
- To better portray the local prospects, we introduce a multi-scale self-attention mechanism to select details scattered in features of each layer.
- We propose a new large dataset, named P2ADataset contains unpaired photos and traditional Chinese paintings for photo-Chinese painting transferring task.

II. Related Work

a) Style Transfer.: Style transfer is an classical task in computer vision. [8] propose a optimization-based model to minimize style loss and content loss between content and style. [9] combine them into a perceptual loss. Recent efforts [10], [11], [12] devote to novel Instance normalization methods to substitute the traditional Batch normalization. However, above methods are only suitable for specific styles. To achieve arbitrary style transfer, [10] introduce an ’AdaIN’ module to approximate mean and variance between the content and the style image. Some works [13], [7], [14] introduce a Whitening and coloring transform (WCT) to decouple images to match the distributions between the content features and style images. To speed up the network, [15] introduced MRFs loss and ’inverse Net’ to produce stylized image base on the patches of features. Arbitrary style transfer task is flexible but unable to transfer a whole style such as Chinese painting.



(a) Results of CycleGAN, where he odd and even lines indicate the photos and the generated Chinese painting respectively.

(b) Results of haar wavelet transform

Fig. 2: (a) The ’over grayscale’ results of CycleGAN, i.e lack of ’Xieyi’ information, (b) The results of haar wavelet transform, where the image enhancement is used for better display.

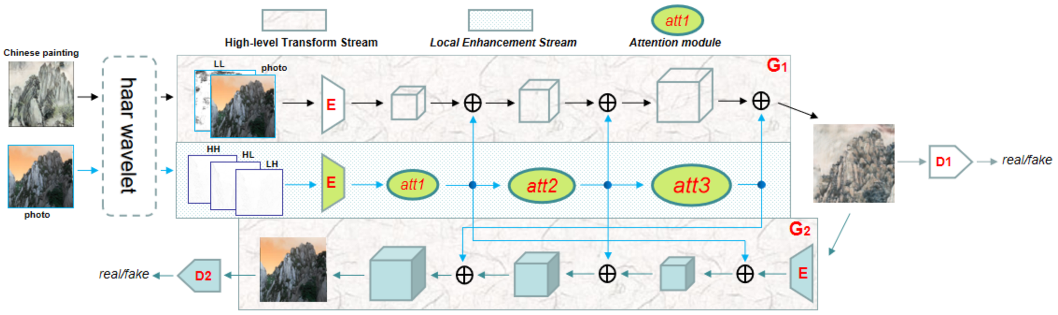


Fig. 3: Overview of our AWNet structure. Our model consists of two generators and discriminators. Each generator has 9 residual blocks while the discriminator compose of full convolutional layers. We feed photos and Chinese paintings to generators and discriminators to determine whether the output of generator is true or not. In order to capture the local details, we introduce a local enhancement stream and multi-scale self-attention modules to fuse them.

Moreover, a specific style and content produce a unmodifiable stylized image even if it is bad.

b) Wavelet Transform.: Wavelet Transform has recently been widely applied in Generative models. The information contained in the spectrogram can effectively assist the generation. [16] propose a Super-Resolution CliqueNet (SR-CliqueNet) to reconstruct a single image with its super-resolution version. [17] exploit wavelet packet transform to obtain multi-scale wavelet coefficients as auxiliary information and propose a wavelet-based GAN for face aging task. [18] propose a novel wavelet CNN consisting of several paths to get various wavelet levels for image classification. [19] propose to mix up spectral analysis and CNN for texture classification. They regard pooling layers as 'limited form of spectral analysis'. [20] exploit wavelet transform to carry out multi-scale face super resolution task. [21] utilize sub-network to deal with specific information. Inspired by them, we use the low and high frequencies information of wavelet in this paper.

III. Proposed Approach

Our objective is capturing high-level information and local details for Chinese paintings, First, we utilize haar wavelet transform algorithm [17], [16], [7] to decompose specific images into various frequencies due to its in dealing with approximate or detail information of images. In particular, LL denotes low frequency (**l**ow frequency pass through **l**ow-pass filter), while HL, LH and HH depict horizontal, vertical, diagonal high frequency respectively [16], [7]. Then, we propose to capture the poetic high-level information in the low frequency (LL) domain of **Chinese painting** via a high-level transform stream, followed by a local enhancement stream in three high frequency domains (LH, HL, HH) of **photo** to enforce the detail information. To emphasize the different contributions in feature maps, we further introduce a self-attention module to weight each channel of the high frequency feature map. We shall elaborate each component in the following three sections.

A. Haar Wavelet Transform

We exploit the haar wavelet transform to decompose a specific image to one overall information contained in LL domain and three detail information existing in LH, HL, HH domains as shown in Fig. 2(b). Given the image $x \in X$, $X \in R^{H\times W\times 1}$, we first calculate the horizontal low frequency $L \in R^{H\times(W/2)\times 1}$ and horizontal high frequency $H \in R^{H\times(W/2)\times 1}$ via 1D haar wavelet transform by rows. After obtaining the image $x_a \in R^{H\times W\times 1}$ consisting of L and H, we secondly conduct 1D haar wavelet transform by columns to obtain LL, LH, HL and $HH \in R^{(H/4)\times(W/4)\times 1}$ respectively, i.e. 1D wavelet transform for columns on L, H simultaneously. Specifically, we can obtain L and H via low-frequency filter $L_f{=}\frac{1}{\sqrt{2}}[1,1]$ and high-frequency filter $H_f{=}\frac{1}{\sqrt{2}}[1,-1]$ respectively, $L{=}L_f{\otimes}x$, $H{=}H_f{\otimes}x$ and then we can calculate $LL{=}\frac{1}{2}{L_f}^T{\otimes}L$, $LH{=}\frac{1}{2}{H_f}^T{\otimes}L$, $HL{=}\frac{1}{2}{L_f}^T{\otimes}H$ and $HH{=}\frac{1}{2}{H_f}^T{\otimes}H$ where $\otimes$ means convolution operation as shown in Fig. 4.

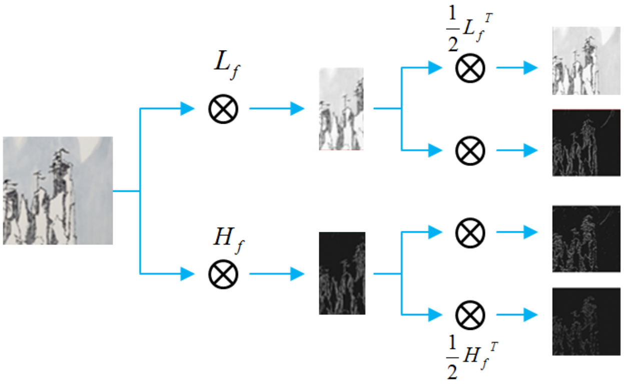


Fig. 4: Haar transform via convolution operation.

B. High-level Transform Stream

The main transformation framework is similar to Cycle-GAN [5] as shown in Fig. 3. It consists of two generators (G_1 and G_2) and two discriminators (D_1 and D_2) to constitute a ring structure. We focus on transferring photos to Chinese paintings while preserving Chinese paintings to photos loop as an auxiliary information. Given the photo $x \in X$, and the Chinese painting $y \in Y$, we obtain: G_1: $\hat{x} = G_1(x \oplus y_{LL})$,

G_2: $\check{x} = G_2(y \oplus x_{LL})$, D_1: $\hat{z} = D_1(\hat{x})$. In the same manner, D_2: $\check{z} = D_2(\check{x})$, where x_{LL}, y_{LL} represent the low frequency of x and y respectively,and $\oplus$ denotes cat operation. Notably, $\hat{x}$ is the final generated Chinese paining, while G_2 ensures the uniqueness of mappings. D_1 is to distinguish whether $\hat{z} \in Y$ or not. G_1 has the same structure as G_2 while D_1, D_2 has the same structure.

To integrate features between layers, we design a manifold residual structure consisting of n residual blocks followed by 3 convolution layers. In this paper we set $n = 9$. Inspired by previous works [10], [12], [11], we replace Batch normalization layers with Instance normalization layers which can not only accelerate model convergence, but also avoid the mutual interference of various images in the batch. We feed G_2 as stylized image.

$$sty_{img} = (x + x \otimes \beta) \otimes \delta \tag{1}$$

where δ, β represent the parameters of decoder and encoder respectively, and $\otimes$ represents convolution operation.

To preserve the structural consistency between the photo and the Chinese painting, we introduce the cycle-consistency loss as following:

$$L_{cyc_1} = ||G_2(G_1(x), x_{LL}) - x||_1 \tag{2}$$

$$L_{cyc_2} = ||G_1(G_2(y), x_{LL}) - y||_1 \tag{3}$$

The transferred stylized image normally present 'over grayscale' effect and lack of artistic conception. This is mainly caused by the emphasis on consistency of photo while ignoring the diversity of the styles during transferring. Herein, we introduce the low frequency (LL $\in R^{H \times W \times 1}$ where H, W represent height and weight respectively) of the Chinese paintings, which contains rich high-level information. Specifically, we add it as a channel of input, allowing the model to capture more high-level information and fusing them with photo features adequately.

Hence, we can generate the stylized image via playing the 'maximum and minimum game' for traditional adversarial loss and cycle-consistency loss [13], [22], [23], [24]:

$$L_{GAN1} = \mathbb{E}_{x \sim P_x}[\log D_1(x)] + \mathbb{E}_{\hat{x} \sim P_{\hat{x}}}[\log(1 - D_1(\hat{x}))] \tag{4}$$

$$L_{GAN2} = \mathbb{E}_{y \sim P_y}[\log D_2(y)] + \mathbb{E}_{\hat{y} \sim P_{\hat{y}}}[\log(1 - D_2(\hat{y}))] \tag{5}$$

The final loss can be written as:

$$L = \alpha L_{GAN1} + \gamma L_{GAN2} + \delta(L_{cyc_1} + L_{cyc_2}) \tag{6}$$

where α, γ, δ represent hyper-parameters.

C. Local Enhancement Stream

Compared to the low-frequency of image which contains more high-level information, the high frequencies contain rich local details. Herein, we introduce a local enhancement stream as an auxiliary structure for the high-level transform stream which also consists of a number of residual blocks and convolution layers. We combine LH, HL, and HH $\in R^{H \times W \times 1}$ as the input $HF = (LH \oplus HL \oplus HH)$, and then we obtain feature:

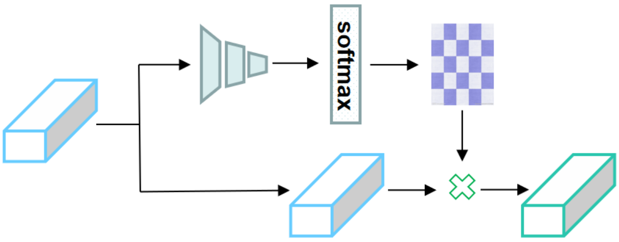


Fig. 5: Our attention module. The input of attention module is multi-channel feature, we can obtain a 1D vector after passing by several convolution and avgpooling layers and obtain the final feature via multiplying by input feature.

$$H_f = HF + HF \otimes \theta \tag{7}$$

$H_f \in R^{H \times W \times C}$ where C means the number of channels after encoder, $\otimes$ presents convolution with parameter θ. Therefore, H_f mixes all the local information.

Multi-scale self-attention. The detail information generally locate variously in different feature maps, while on a particular feature, detail information are scattered erratically. To pay more attention to these meaningful feature maps, we propose to endow each feature map a specific weight. By learning this weight, we can pick out the features that we're interested in i.e. features which contain more local details. Herein, we introduce a multi-scale self-attention mechanism [25], [26] and put forward a self-attention module att_{module} to emphasize the 'Gongbi' information as shown in Fig. 5, we can acquire the i^{th} attention map as,

$$A_i = (H_f \boxtimes \eta_i) \otimes \theta_i \tag{8}$$

Then we can obtain the attention map $Att = F(A_1, A_2, ..., A_n)$, for each weighted feature:

$$fea_i = A_i * H_{f_i} \tag{9}$$

The final feature $att_f = fea_1 \oplus fea_2 \oplus ... \oplus fea_n$, where i indicates the i^{th} channel. $\boxtimes$ and $\otimes$ represent avgpooling and convolution operation corresponding to the parameters η and θ respectively. F(●) stands for weighting, which is implemented by softmax function:

$$e(\varepsilon) = \frac{e^{\varepsilon}}{\sum_{k=1}^{n} e^k} \tag{10}$$

Different from previous method [25], we replace flatten operations with full convolutions to better capture global information.

The weighted high-frequency feature map can be regarded as an auxiliary local information for 'high-level transform stream'. Instead of directly adding to the generator, it is more effective to fuse them in the process of reconstruction which can further avoid missing many high frequencies while passing through generators. Therefore, the input of decoder is re-written as,

$$D_{in} = (x + x \otimes \beta) \oplus att_f \tag{11}$$

Meanwhile, in order to exploit the multi-scale information in high frequency features, we set up three attention modules to capture the high frequency features from different layers, which is then incorporated into the corresponding layers of the decoder in high-level transform stream, as shown in Fig. 3.

IV. Experimental Results

A. Implementation Details

Existing dataset [1] is relatively small and partly watermarked, training our model with this small-scale unclear dataset leads to poor results as shown in Fig. 6. Therefore, we establish a new dataset named P2ADataset, for photo to Chinese painting transferring. P2ADataset contains totally 5,348 unpaired images of the well-known Chinese mountains, with 2,563 Chinese paintings and 2,785 photos. The resolution is resized up to 256×256. We train our model via Adam optimizer with a learning rate of $1e$-3. To enhance the details in Chinese paintings, we exploit 3 attention modules and fuse weighted feature into different layers of each generator's decoder respectively. For generator encoder, we employ the residual network, where each block consists of 2 convolution layers. Each discriminator outputs a value in [0,1] on behalf of fake or real.

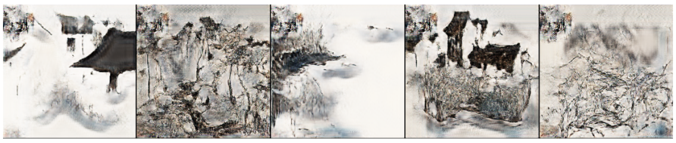


Fig. 6: Transferred results of our model on dataset [1].

B. Qualitative Evaluation

a) Ablation Studies: To evaluate the contribution of haar wavelets to our model, we design the following evaluation variants as ablation studies i.e., 1) baseline model (CycleGAN), 2) baseline model with low frequency (LL) of Chinese painting (baseline + LL), and 3) baseline model with both low-frequency information of Chinese paintings and high-frequencies of photos (Ours). As shown in Fig. 7, the results of baseline model are generally 'over grayscale' which lack of high-level information. After introducing the low-frequency of **Chinese paintings**, we can get more 'Xieyi' results, the price is that they loss many detail information. Our final model can combine high-level information and local details for results by incorporating both low-frequency of **Chinese paintings** and high frequencies of **photos**. The results also demonstrate that our 'Local Enhancement stream' in favour of enhancing the details and 'High-level Transform stream' can transform the high-level information effectively.

b) Comparison with Other Methods: To demonstrate the effectiveness of our model, we compare our model with classic style transfer methods: AdaIN [10], WCT [14], Style-Swap [15] and Gatys et al [8]. Fig. 9 demonstrates the results of each model on photos to Chinese paintings transfer.

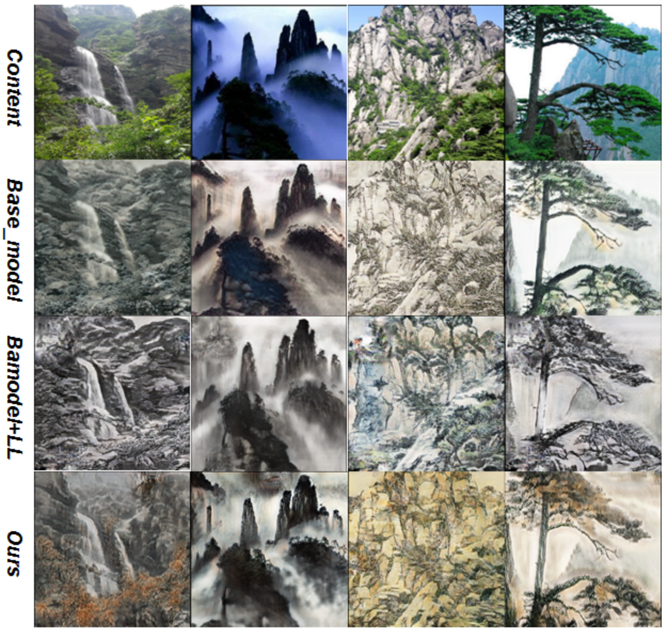


Fig. 7: Ablation Studies. The baseline model generally results in 'over grayscale'. The results of baseline+LL contains more "Xieyi' information while lacking details. Our results achieve more graceful results compared with them.

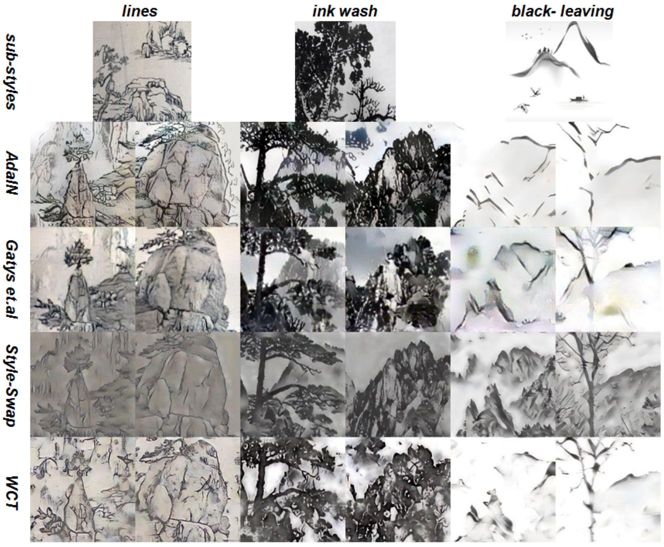


Fig. 8: Results of different sub-styles. Arbitrary style transfer task just aims at a special style rather than a whole style, when we give different sub-styles in Chinese paintings, they obtain various results.

c) Sub-styles: While giving different forms of Chinese painting, the compared style transfer methods achieves diametrically different generations as shown in Fig. 8. In fact, these methods can only transform one of the specific sub-style such as 'ink wash', 'blank-leaving', *etc*, while our method is insensitive to the styles. This is also one of the key capability of our method.

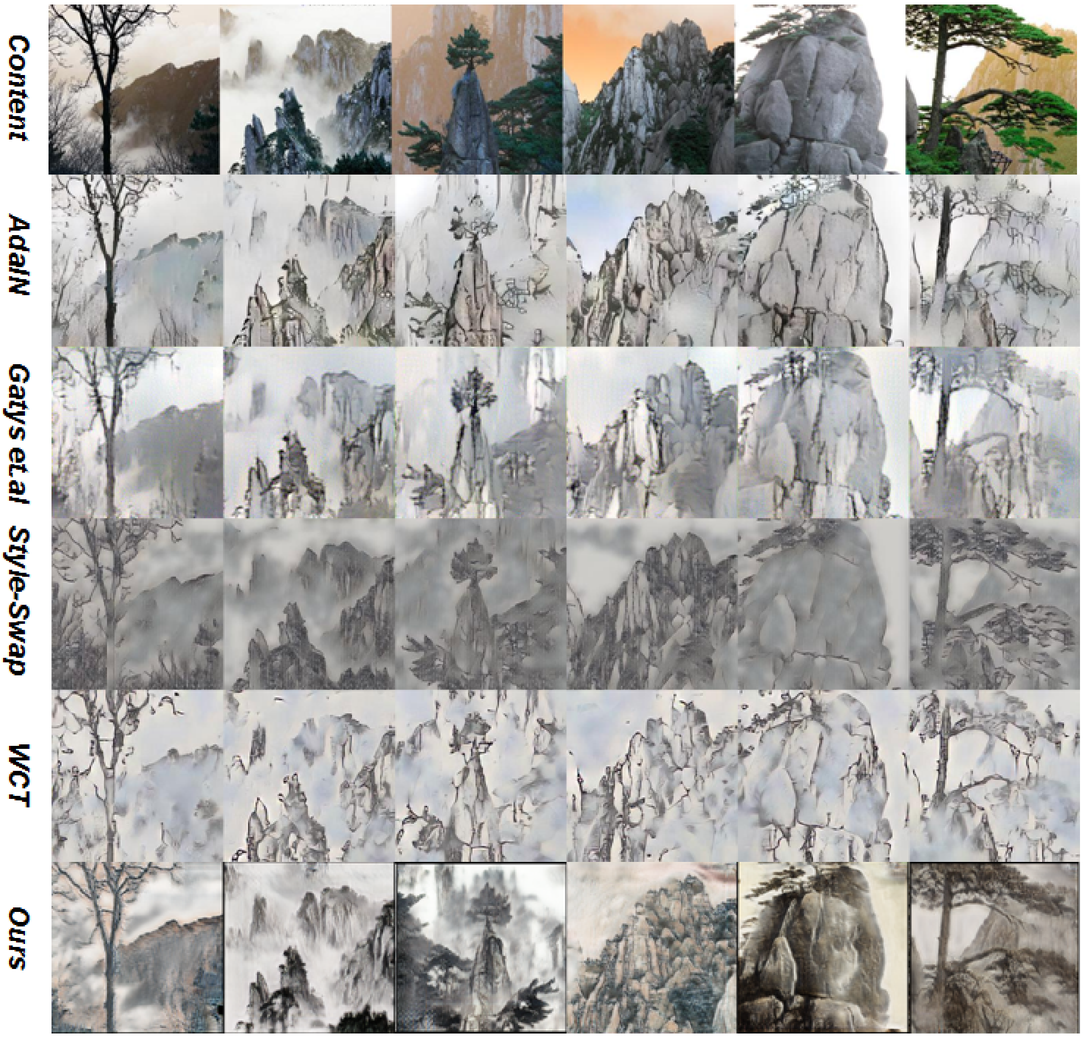


Fig. 9: Our results compared with classic style transfer algorithms. AdaIN losses some content information, Gatys et.al generates blurry results, Style-Swap looks too dark and WCT's results in fragments. Moreover, none of them can reflect 'Xieyi' prospect. Our method achieves more attractive results compared with others.

C. Quantitative evaluation

a) Consistency study: In Table I, we exploit SSIM and PSNR to quantitatively evaluate our model against the prevalent methods i.e AdaIN, Gatys et al, Style-Swap and WCT. We randomly select 6 photos and Chinese paintings for them, and compute the average of the 36 stylized images refer to [14] under different indicators. It is clear to see that our method outperforms the others on all the three metrics which verifies the effectiveness of our method. The only exception is that our PSNR is lower than Style-Swap due to we just evaluate content consistency without thinking style.

b) User Study: In our user study, we investigate the two key arts, i.e. high-level information and local details in Chinese paintings. For high-level information, we ask the participants which result is the most faithful to the details of the original image while for local details, the participants are asked to vote the result with the best artistic conception of traditional Chinese paintings. We obtain totally 300 votes for both prospects from 50 participants, each of whom contributes 6 votes. As shown in Fig. 10. It is obvious from the figure that whether high-level information or local details, our approach gain more votes.

Methods	Evaluation on P2ADataset	
	SSIM ↑	PSNR ↑
AdaIN	0.27	10.07
Gatys et.al	0.34	9.14
Style-Swap	0.36	**13.08**
WCT	0.18	8.67
Ours	**0.42**	11.04

TABLE I: Quantitative comparisons between ours and and the prevalent style transferring methods.

c) Failure Cases: We further provide five failure cases in Fig. 11, to discover the limitation of the proposed AWNet.

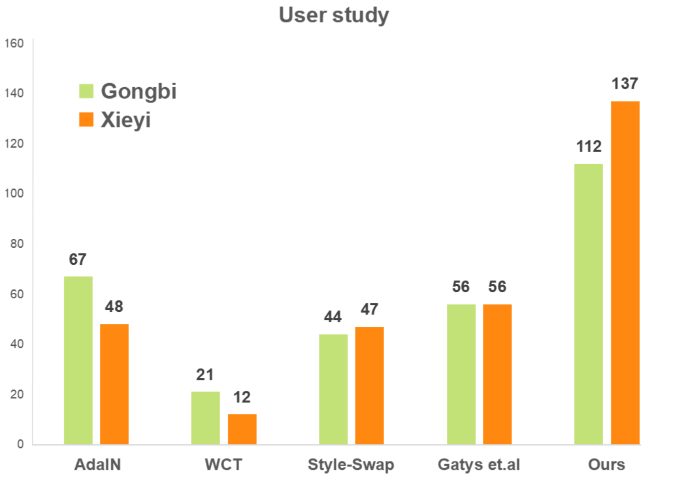


Fig. 10: Our user study. Horizontal axis shows different methods while vertical axis represents the votes on 'Gongbi' and 'Xieyi'.

Note that there is a large area of blur in the background with mystery ghosting along the edges of the images and the consistency of the results is weak. The possible reason is there is no constraints to the foreground or background in our model. It may be helpful to introduce attention mechanism in low-frequency transformation procedure. More intelligent high-frequency information fusing strategy could also relieve this issue.

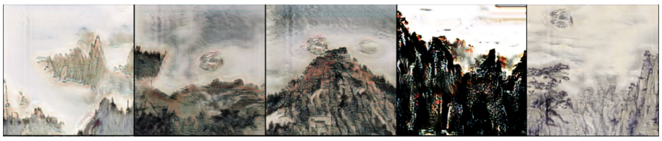


Fig. 11: Failure cases.

V. Conclusions

We propose a novel Attentional Wavelet Network (AWNet) to preserve both high-level information and local prospects in photo to Chinese painting transfer. By considering the low-frequency and high-frequency information after wavelet transform, AWNet can significantly relieve the 'over grayscale' effect in photo to Chinese painting transfer. Furthermore, we introduce a self-attention mechanism to select pivotal detail information for high-level transform stream. In addition, we set up a new dataset for photo to Chinese painting transfer and related communities. Extensive experiments on the proposed dataset demonstrate the effectiveness of the proposed AWNet comparing to prevalent style transfer methods.

VI. Acknowledgement

This research is supported in part by the National Natural Science Foundation of China (61976002, 62076004 and 61860206004), the Open Project Program of the National Laboratory of Pattern Recognition (NLPR) (201900046) and Beijing Natural Science Foundation (Grant No.JQ18017).

References

[1] B. Li, C. Xiong, T. Wu, Y. Zhou, L. Zhang, and R. Chu, "Neural abstract style transfer for chinese traditional painting," in *ACCV*, 2018.

[2] B. He, F. Gao, D. Ma, B. Shi, and L.-Y. Duan, "Chipgan: A generative adversarial network for chinese ink wash painting style transfer," in *ACM*, ser. MM '18, 2018, pp. 1172–1180.

[3] Lin, D. Wang, Y. Xu, G. Li, J. Fu, and K, "Transform a simple sketch to a chinese painting by a multiscale deep neural network," in *Algorithms*, 2018.

[4] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, "Generative adversarial nets," in *NIPS*, 2014.

[5] J.-Y. Zhu, T. Park, P. Isola, and A. A. Efros, "Unpaired image-to-image translation using cycle-consistent adversarial networks," in *ICCV*, 2017.

[6] J. Hoffman, E. Tzeng, T. Park, J. Zhu, P. Isola, K. Saenko, A. A. Efros, and T. Darrell, "Cycada: Cycle-consistent adversarial domain adaptation," *In ICLR*, 2017.

[7] J. Yoo, Y. Uh, S. Chun, B. Kang, and J. Ha, "Photorealistic style transfer via wavelet transforms," in *ICCV*, 2019.

[8] L. A. Gatys, A. S. Ecker, and M. Bethge, "Image style transfer using convolutional neural networks," in *CVPR*, 2016.

[9] J. Johnson, A. Alahi, e. B. Fei-Fei, Li", J. Matas, N. Sebe, and M. Welling, "Perceptual losses for real-time style transfer and super-resolution," in *ECCV*, 2016.

[10] Huang, Xun, and S. Belongie, "Arbitrary style transfer in real-time with adaptive instance normalization," in *ICCV*, 2017.

[11] Karras, Tero, Laine, Samuli, Aila, and Timo, "A style-based generator architecture for generative adversarial networks," in *CVPR*, 2019.

[12] D. Ulyanov, A. Vedaldi, and V. S. Lempitsky, "Instance normalization: The missing ingredient for fast stylization," *ArXiv*, 2016.

[13] W. Cho, S. Choi, D. K. Park, I. Shin, and J. Choo, "Image-to-image translation via group-wise deep whitening-and-coloring transformation," in *CVPR*, 2019.

[14] Y. Li, C. Fang, J. Yang, Z. Wang, X. Lu, and M.-H. Yang, "Universal style transfer via feature transforms," in *NIPS*, 2017.

[15] T. Chen and M. Schmidt, "Fast patch-based style transfer of arbitrary style," *arXiv*, vol. abs/1612.04337, 2016.

[16] Z. Zhong, T. Shen, Y. Yang, Z. Lin, and C. Zhang, "Joint sub-bands learning with clique structures for wavelet domain super-resolution," in *NIPS*, 2018.

[17] Y. Liu, Q. Li, and Z. Sun, "Attribute-aware face aging with wavelet-based generative adversarial networks," in *CVPR*, 2019.

[18] D. Silva, Vithanage, Fernando, and Piyatilake, "Multi-path learnable wavelet neural network for image classification," *arXiv*, vol. abs/1908.09775, 2019.

[19] S. Fujieda, K. Takayama, and T. Hachisuka, "Wavelet convolutional neural networks for texture classification," *arXiv*, vol. abs/1707.07394, 2017.

[20] H. Huang, R. He, Z. Sun, and T. Tan, "Wavelet-srnet: A wavelet-based cnn for multi-scale face super resolution," in *ICCV*, 2017.

[21] S. Zhang, R. He, Z. Sun, and T. Tan, "Demeshnet: Blind face inpainting for deep meshface verification," *IEEE Transactions on Information Forensics and Security*, vol. 13, no. 3, pp. 637–647, 2018.

[22] Z. Zhang, L. Yang, and Y. Zheng, "Translating and segmenting multimodal medical volumes with cycle- and shape-consistency generative adversarial network," in *CVPR*, 2018.

[23] R. Felix, V. B. G. Kumar, I. Reid, and G. Carneiro, "Multi-modal cycle-consistent generalized zero-shot learning," in *ECCV*, 2018.

[24] D. Engin, A. Genc, and H. Kemal Ekenel, "Cycle-dehaze: Enhanced cyclegan for single image dehazing," in *CVPR Workshops*, 2018.

[25] Y. Yao, J. Ren, X. Xie, W. Liu, Y.-J. Liu, and J. Wang, "Attention-aware multi-stroke style transfer," in *CVPR*, 2019.

[26]