ChipGAN: A Generative Adversarial Network for Chinese Ink Wash Painting Style Transfer

Bin He[1], Feng Gao[2], Daiqian Ma[1,3], Boxin Shi[1], Ling-Yu Duan[1*]
National Engineering Lab for Video Technology, Peking University, Beijing, China[1]
The Future Lab, Tsinghua University, Beijing, China[2]
SECE of Shenzhen Graduate School, Peking University, Shenzhen, China[3]
cs_binhe@outlook.com,gaofeng2018@mail.tsinghua.edu.cn,{madaiqian,shiboxin,lingyu}@pku.edu.cn

ABSTRACT

Style transfer has been successfully applied on photos to generate realistic western paintings. However, because of the inherently different painting techniques adopted by Chinese and western paintings, directly applying existing methods cannot generate satisfactory results for Chinese ink wash painting style transfer. This paper proposes ChipGAN, an end-to-end Generative Adversarial Network based architecture for photo to **Ch**inese **i**nk wash **p**ainting style transfer. The core modules of ChipGAN enforce three constraints – voids, brush strokes, and ink wash tone and diffusion – to address three key techniques commonly adopted in Chinese ink wash painting. We conduct stylization perceptual study to score the similarity of generated paintings to real paintings by consulting with professional artists based on the newly built Chinese ink wash photo and image dataset. The advantages in visual quality compared with state-of-the-art networks and high stylization perceptual study scores show the effectiveness of the proposed method.

KEYWORDS

Painting; style transfer; generative adversarial network

ACM Reference Format:
Bin He, Feng Gao, Daiqian Ma, Boxin Shi, Ling-Yu Duan. 2018. ChipGAN: A Generative Adversarial Network for Chinese Ink Wash Painting Style Transfer. In 2018 ACM Multimedia Conference (MM '18), October 22-26, 2018, Seoul, Republic of Korea. ACM, New York, NY, USA, 9 pages. https://doi.org/10.1145/3240508.3240655

1 INTRODUCTION

Any successful artist has his or her uniquely defined painting style. Studying such uniqueness in painting style is important in painting skill training. In addition to the traditional art theories training, computer vision and graphics techniques, such as style transfer [17, 37] and non photorealistic rendering [12, 36, 38], have been developed to help painting artists in systematically understanding how to apply an appropriate painting technique to present a type of unique style by observing a real scene or photo.

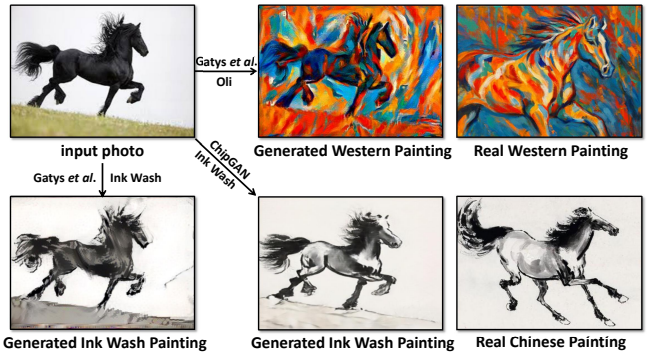

Figure 1: Given an input photo, existing style transfer technique (Gatys *et al.* [11]) is able to generate western painting with visually close style to the real painting (top row), but not for the Chinese ink wash painting (bottom left). The proposed ChipGAN with three constrains achieves realistic transfer result (bottom row).

Migrating the styles of paintings to images can be implemented through texture synthesis using low-level image features [8, 9, 26, 43], which ignores the semantic information of an image. To extract high-level semantic information from images for style transfer, Convolutional Neural Network (CNN) [25, 27] is utilized by [11, 21], which shows visually realistic results (Figure 1, photo to generated western painting according to the style of real western painting). However, directly applying existing style transfer techniques to Chinese ink wash paintings results in unrealistic results (in Figure 1, generated ink wash painting, note the chaotic lines and thick color). This is because there are several essential differences between western and Chinese ink wash painting techniques, as comparison between real paintings in the last column of Figure 1 shows: 1) In terms of the composition of a picture, western paintings are filled with colors over the whole image, while Chinese ink wash paintings contain certain areas of *voids*[1]; 2) In terms of expression skills, western paintings seldom use strong lines, while

*Ling-Yu Duan is the corresponding author.
Bin He and Feng Gao are joint first authors.

MM '18, October 22–26, 2018, Seoul, Republic of Korea

ACM ISBN 978-1-4503-5665-7/18/10…$15.00
https://doi.org/10.1145/3240508.3240655

[1]It refers to areas of the white paper, which Chinese ink wash painting artists purposely leave to inspire the viewers to imagine [5].

Chinese ink wash paintings adopt *brush strokes* with vigorous lines to emphasize the object in silhouette; 3) In terms of color richness, western paintings tend to use a great diversity of colors, while Chinese ink wash paintings mainly use ink with different gray levels that diffuses on a piece of rice paper (*ink wash tone and diffusion*).

To achieve style transfer for Chinese ink wash paintings, we propose a photo to **Ch**inese **i**nk wash **p**ainting style transfer solution based on Generative Adversarial Network (GAN) [14], named **Chip**GAN. We propose three special constrains according to the three techniques of Chinese ink wash painting: voids, brush strokes, and ink wash tone and diffusion. For voids, our constraint combines adversarial loss with cycle consistency loss [2, 50], since they aim to generate more realistic result by converting information to an imperceptible signal [4] thus leaves the white area. For brush strokes, we embed a pre-trained holistically-nested edge detector [44] and enforce a redesigned cross entropy loss [6] between edge maps of photo and fake painting to emphasize vigorous lines. For ink wash diffusion and tone, we use eroded and blurred images to mimic such painting properties and propose the ink wash discriminator to distinguish between processed real and fake paintings.

Existing painting datasets mainly contain artworks by western artists (*e.g.,* Van Gogh, Monet, *et al.*) [50], and there is no available dataset that consists of real photos and images of the corresponding Chinese ink wash paintings. For solving our problem, we present a Chinese ink wash painting dataset with **Ph**otos of real scene and **i**mages of paintings collected from the Internet and art studio, named "Chip**Phi**". Our dataset consists of Horse dataset containing 1630 photos of horses (with different colors and in various poses) and 912 images of paintings[2] by Xu Beihong and Landscape dataset with 1976 photos of landscapes (with famous landscapes around the world) and 1542 images of paintings by Huang Binhong.

In summary, the contributions of this paper are three-fold:

- We propose ChipGAN, the first[3] weakly supervised deep network architecture to perform photo to Chinese ink wash painting style transfer, with special considerations on three essential techniques of Chinese ink wash painting: void, brush stroke, and ink wash tone and diffusion.
- We introduce stylization perceptual study involving professional artists to evaluate the style consistency between generated and real paintings and analyze Chinese ink wash painters' techniques with the help of deep neural network.
- We build the first dataset with photos in real scenes and images of Chinese ink wash painting named ChipPhi to facilitate the training and testing of the proposed approach and benefit follow-up research on Chinese ink wash painting style transfer.

2 RELATED WORK

Image-level style transfer means migrating the style of a certain example image to the target one. Previous Image-level style transfer can be divided to texture synthesis and Convolutional Neural Network based approaches. Domain-level style transfer means rendering a given image (*e.g.,* photo) with style of a certain domain (*e.g.,* style of a certain painter). It is accomplished by approaches based on Generative Adversarial Network (GAN) [14, 19]. Besides, we also review some computational methods particularly designed for Chinese ink wash paintings.

Texture synthesis. There are some non-parametric algorithms [8, 9, 43] which can synthesize textures by resampling the given texture image. Efros and Freeman [8] propose a correspondence map which constrains the texture synthesis procedure according to image intensity of the target image. Ashikhmin[1] concentrates on transferring the high-frequency texture but preserves the scale of the target image. Hertzman *et al.* [16] apply image analogies to transfer style of a source image to the target one. However, since texture synthesis mainly depends on patches and low-level presentations, they fail to transfer semantic style of artistic works.

CNN based approaches. CNN based models target to extract semantic representations by pre-trained convolutional neural network. Gatys *et al.* [11] first use CNN to obtain the representations of images, and reproduce famous panting styles on the natural photos. Li *et al.* [30] find linear kernel is a good substitute for Maximum Mean Square. Yin [48] and Chen and Hsu [3] investigate content-aware neural style transfer and improve the results. Most of these approaches suffer from low speed and high computational cost, which can be accelerated by the methods in [21, 39]. Li and Wand [29] train a Markovian feed-forward network to solve the efficiency problem. Dumoulin *et al.* [7] propose to learn multiple styles at the same time. Although these methods have generated impressive stylized images for western painting, they fail to transfer Chinese ink wash style due to its essentially different properties.

GAN based approaches. When tackling the style transfer task from the perspective of GAN, some image-to-image translation approaches are reasonably effective. CoupledGAN [34] learns a joint distribution of multi-domain images by enforcing a weight-sharing constraint. However, this method can only take a noise vector as input to generate paired images. So it cannot be directly used as style transfer model. Liu *et al.* [33] combine CoupledGAN [34] with variational auto-encoder [24] and propose a framework named UNIT [33]. Zhu *et al.* introduce cycle consistency losses to reduce permutation of mappings and propose CycleGAN [50]. Based on architecture of CycleGAN [50], DistanceGAN [2] enforces the constraint where the distance of two samples in one domain should be preserved in the mapping to another domain. We also adopt cycle consistency losses in our model to overcome mode collapse [13], and combine it with adversarial loss to simulate voids. Though cycle consistency loss makes the model preserve some details in the original photo, it at the same time tends to remove some important brush strokes incorrectly, which motivates us to come up with additional constraints for modeling brush strokes of Chinese ink wash paintings.

Computational methods for Chinese ink wash paintings. Chinese ink wash paintings can be generated using different computational approaches. Yu *et al.* [49] combine the brush stroke texture from a real painting with color information of given landscape image to synthesize an ink wash painting. Xu *et al.* [45] decompose the brush strokes of a Chinese ink wash painting with a prepared

[2] All the paintings are cropped to remove the Chinese characters.

[3] Jing *et al.* [20] transfer the style of a Chinese ink wash painting to a given photo by directly using the method of Gatys *et al.* [11], without proposing a new approach specially for Chinese ink wash painting style transfer.

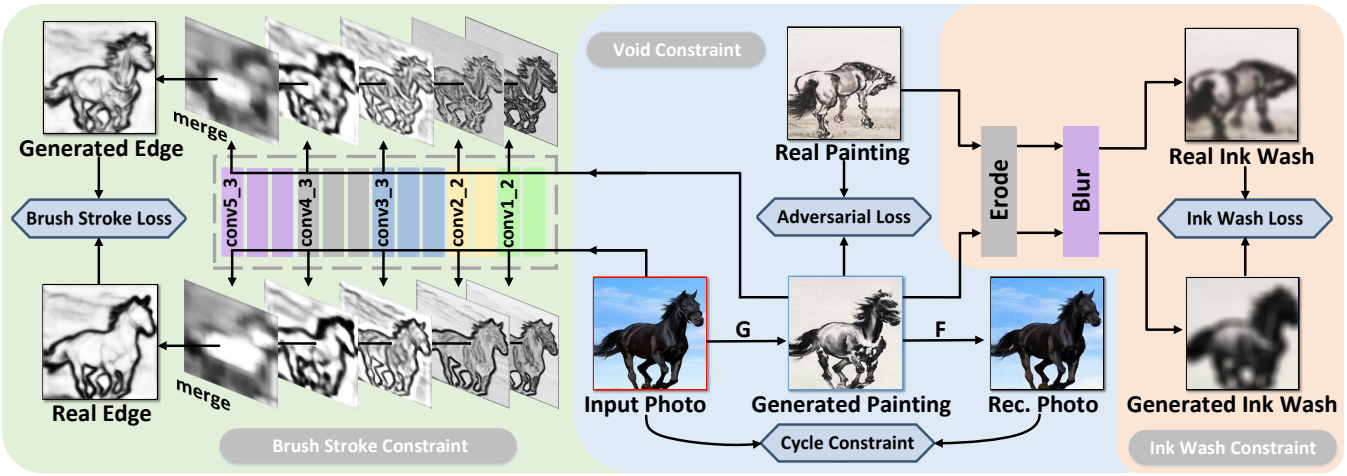

Figure 2: Pipeline of ChipGAN. We take an input photo (red box) of a horse to get a generated ink wash painting (blue box). The void constraint (blue part, middle; "Rec." for "Reconstructed"), brush stroke constraint (green part, left), and ink wash constraint (red part, right), are illustrated using this horse example.

brush strokes library to render animations. Yang and Xu [46] further refine the brush stroke decomposition method by providing automatic brush stroke trajectory estimation. Wang [41] propose an effective algorithm to simulate ink wash diffusion based on the Kubelka-Munk equation. Yeh *et al.* [47] and Way *et al.* [42] generate ink wash paintings based on the board lines strokes and interior shading of 3D models. Liang and Jin [31] generate ink wash painting from a given photo through image processing on edges, colors, and paper texture. Instead of relying on existing brush strokes simulation and low-level image features as prior, our method explores data-driven techniques to learn realistic Chinese ink wash painting feature representations.

3 PROPOSED METHOD

ChipGAN learns a mapping from the photo domain X (*e.g.,* defined by real-world photos of horses) to the painting domain Y (*e.g.,* defined by Chinese ink wash paintings of horses). We combine cycle consistency loss and adversarial loss as a constraint to deal with void technique in Section 3.1; we then propose brush stroke loss to remove unnecessary brush strokes while preserving essential ones in Section 3.2; we further introduce ink wash loss to ensure the correct tone of whole image and add the diffusion effect in Section 3.3. Our full objective and training details are provided in Section 3.4 and Section 3.5, respectively. The complete pipeline of ChipGAN is illustrated in Figure 2.

3.1 Void constraint

Intuitively speaking, applying voids means leaving blanks at proper places on the canvas [5]. Taking the horse as an example, appropriately applying voids requires the generated image completely ignores the sky and partly ignores the grass in photo while clearly keeping the horse silhouette, as shown in the middle part of Figure 2. The horse photo and a Chinese ink wash painting of horse have different entropies, because the photo has rich color and texture compared to the image of painting. Such different entropies between the source domain and target domain are utilized in image-to-image translation tasks [4] to effectively convert information about a source image into a nearly imperceptible signal, by combining the adversarial loss and cycle consistency loss. We therefore adopt the similar strategy to enforce the void constraint.

Adversarial loss. Given unpaired training sets which are regarded as two domains X and Y, our model includes two mappings: $G : X \rightarrow Y$ and $F : Y \rightarrow X$. For $G : X \rightarrow Y$ and its discriminator D_Y, the adversarial loss [19] is given by:

$$\begin{aligned}\mathcal{L}_{GAN}(G, D_Y, X, Y) =& \mathbb{E}_{y \sim p_{\text{data}}(y)}[\log D_Y(y)] \\ &+ \mathbb{E}_{x \sim p_{\text{data}}(x)}[\log(1 - D_Y(G(x)))],\end{aligned} \tag{1}$$

where G endeavors to generate samples that are similar to real ones from domain Y, while D_Y tries to discriminate between the fake and real samples. This objective is minimized over G and maximized over D_Y, *i.e.,* $\min_G \max_{D_Y} \mathcal{L}_{GAN}(G, D_Y, X, Y)$. For mapping $F : Y \rightarrow X$ and its discriminator D_X, there is a similar objective, *i.e.,* $\min_F \max_{D_X} \mathcal{L}_{GAN}(F, D_X, Y, X)$.

Cycle consistency loss. We add the cycle consistency constraint [50] by translating the given image x from domain X to target domain Y and then back to domain X, which should result in the same image, *i.e.,* $F(G(x)) \approx x$. Because the cycle consistency constraint requires recovery in both directions, for each image y in domain Y, there is also a cycle consistency constraint: $G(F(y)) \approx y$. Thus, the cycle consistency loss is defined as:

$$\begin{aligned}\mathcal{L}_{cycle}(G, F, X, Y) =& \mathbb{E}_{x \sim p_{\text{data}}(x)}[\|F(G(x)) - x\|_1] \\ &+ \mathbb{E}_{y \sim p_{\text{data}}(y)}[\|G(F(y)) - y\|_1].\end{aligned} \tag{2}$$

This constraint makes the generated image preserve some information of source domain, so that the generated one can be converted back to source domain.

3.2 Brush stroke constraint

Given the properly generated blank area, our next goal is to add brush strokes to clearly depict the silhouette of objects in Chinese ink wash painting style, *e.g.,* the head and body of the house should have vigorous silhouette. To model various types of brush strokes with different thicknesses in Chinese ink wash paintings [45] in a unified manner, we formulate our brush stroke constraint used to enforce the consistency between different levels of edge maps of real photos and generated paintings.

We adopt holistically nested edge detector [44] E to extract five levels of edges from the input image, to simulate five types of brush strokes of different thickness, as shown in the left part of Figure 2. We then merge edge maps generated from different stages of pre-trained VGG-16 feature extractor to obtain the final edge map. Different from regarding the edge detection task as a pixel-level binary classification problem, we train a multi-level edge detector from the perspective of regression to obtain smooth brush strokes with different thicknesses. Every pixel in training ground truth is labeled with a real number from 0 to 1 which indicates their probability to be a part of an edge [44]. By applying E, we obtain edge maps of real photo and generated painting $E(x)$ and $E(G(x))$. We then take $E(x)$ as ground truth and calculate balanced cross entropy loss to let G generate proper brush strokes as

$$\begin{aligned}\mathcal{L}_{brushstroke}(G,X) = \mathbb{E}_{x\sim p_{\text{data}}(x)}[-\frac{1}{N}\sum_{i=1}^{N}\mu E(x)_i \log E(G(x))_i \\ + (1-\mu)(1-E(x)_i)\log(1-E(G(x))_i)],\end{aligned} \tag{3}$$

where N is the total number of pixels in edge map of photo or fake painting and μ is a balancing weight. $\mu = N_-/N$ and $1-\mu = N_+/N$. N_- and N_+ are the sum of non-edge and edge probability of every pixel in $E(x)$, respectively.

3.3 Ink wash constraint

With the voids and brush strokes properly modeled, our final processing is to make the global tone (*e.g.,* the overall color temperature of the generated horse painting should be close to the real one) and diffusion effects (*e.g.,* the abdomen of the horse shows link diffuses to different gray levels on the rice paper) consistent between the real painting y and generated painting $G(x)$. Therefore, we further introduce the ink wash constraint.

The diffusion of ink wash on rice paper is approximately isotropic, so we simulate it with an erosion operation and followed by a Gaussian blur operation. With salient objects being blurred, such an operation suppresses explicit comparison of texture and content information [18], so that the model tends to focus more on the tone consistency, as illustrated in the right part of Figure 2.. Therefore, we add an adversarial discriminator D_I which is trained to distinguish between y_{eb} and $G(x)_{eb}$:

$$y_{eb}(i,j) = \sum_{k,l}(y \ominus B)_{i+k,j+l}\cdot G_{k,l}, \tag{4}$$

$$G(x)_{eb}(i,j) = \sum_{k,l}(G(x) \ominus B)_{i+k,j+l}\cdot G_{k,l}, \tag{5}$$

where y_{eb} is the real painting processed by erosion and blur, $G(x)_{eb}$ is the generated painting processed by erosion and blur, $\ominus$ is the erosion operator, B is an erosion kernel, and Gaussian blur kernel $G_{k,l} = \frac{1}{2\pi\sigma^2}\exp(-\frac{k^2+l^2}{2\sigma^2})$. Finally, the ink wash loss is defined as

$$\begin{aligned}\mathcal{L}_{inkwash}(G,D_I,X,Y) =& \mathbb{E}_{y\sim p_{\text{data}}(y)}[\log D_I(y_{eb})] \\ &+\mathbb{E}_{x\sim p_{\text{data}}(x)}[\log(1-D_I(G(x)_{eb}))].\end{aligned} \tag{6}$$

3.4 Full objective

Our full objective is a linear combination of the four types of losses introduced above:

$$\begin{aligned}\mathcal{L}(G,F,D_X,D_Y,D_{Ink}) =& \mathcal{L}_{GAN}(G,D_Y,X,Y) \\ &+\mathcal{L}_{GAN}(F,D_X,Y,X) + \lambda\mathcal{L}_{cycle}(G,F,X,Y) \\ &+\beta\mathcal{L}_{brushstroke}(G,X) + \gamma\mathcal{L}_{ink}(G,D_{Ink},X,Y),\end{aligned} \tag{7}$$

where hyper-parameters λ, β, and γ control the contributions of the individual objectives. We then aim to solve:

$$G^*,F^* = \arg\min_{G,F}\max_{D_X,D_Y,D_{Ink}}\mathcal{L}(G,F,D_X,D_Y,D_{Ink}). \tag{8}$$

In Section 5.2, we will analyze our method against ablation of full objective by removing $\mathcal{L}_{brushstroke}$ or $\mathcal{L}_{ink}$ or both of them to demonstrate that the losses specially designed for Chinese ink wash paintings are indispensable.

3.5 Architecture and training details

We build our generator networks with two stride-2 convolutions, 9 residual blocks [15] and two fractionally strided convolutions. Besides, we adopt instance normalization [40] in generator networks to generate stable and smooth images. The discriminator networks are constructed by 70×70 PatchGANs [19, 28, 29] which are designed to classify whether 70×70 overlapping image patches are real or fake. The pre-trained VGG-16 whose last pooling and all fully connected layers have been cut, is embedded into the edge extraction part. By adding fractionally strided convolutions into the modified VGG-16, the multi-level edge extraction part converts the feature maps from conv1_2, conv2_2, conv3_3, conv4_3 and conv5_3 to corresponding edge maps with the same size as input images. After that, all the edge maps are merged by a convolution to generate the final edge map.

During the training stage, all the input images are resized to 256×256. All networks are trained from scratch and weights are initialized from a Gaussian distribution with mean 0 and standard deviation 0.02. In all our experiments, the network is trained by Adam [23] solver with batch size of 1. The learning rate is initialized to 0.0002 for all generators and 0.0001 for all discriminators. We keep the same learning rate for the first 100 epochs and linearly decay the rate to zero over next 100 epochs. We set $\lambda = 10$, $\beta = 10$, and $\gamma = 0.05$ in Equation (7).

4 DATASET AND EVALUATION METHOD

4.1 The ChipPhi dataset

To the best of our knowledge, image dataset collected specially for Chinese ink wash paintings is not publicly available. We build the ChipPhi dataset containing photos of real scenes and images of Chinese ink wash paintings collected from the Internet to evaluate our method and hopefully to inspire the follow-up research. The

ChipPhi dataset consists two parts: Horse and Landscape, which are Chinese ink wash paintings of horses and landscapes drawn by Xu Beihong and Huang Binhong and photos of horses and landscapes. To ensure our dataset contains images with rich content diversity, we collect horse photos with various colors (*e.g.,* white, black, brown) and postures (*e.g.,* standing, running, part of the horse such as head), and the landscape photos covering famous hills from all over the world (*e.g.,* Mount Huangshan, Rocky Mountains, Great Smoky Mountains National Park).

To generate stylized images with high quality, we remove the photos in which the objects are unrecognizable or blocked by watermarks. For images of Chinese ink wash paintings, they usually contain some calligraphy to indicate the name of the artist, the year when the painting was created, or even some poems. We clean the painting images to get rid of Chinese characters by cropping.

Since we aim to learn the style of a painter, the ink wash paintings for a certain content (*e.g.,* horse) should be those drawn by the same artist (*e.g.,* Xu Beihong). Nonetheless, the total number of real paintings is rather limited. To compensate the deficiency, we augment our data by horizontal flip. Considering Horse, We first collect 456 ink wash paintings and 819 photos. For both the photo and painting domain, we divide them into training and testing set by a ratio of 9 to 1. After that, a horizontal flip is applied. We finally prepare 1478 photos and 822 paintings for training, 160 and 90 for testing. For Landscape, we collect 1774 photos and 1388 paintings for training, 202 and 154 for testing.

4.2 Stylization perceptual study

Since there is no ground truth to compare with, it is infeasible to quantitatively measure the style similarity of synthesized images to real paintings. We therefore design a stylization perceptual study [20], which asks ink wash painting artists to rank and rate scores about the style similarity to the real paintings from our generated paintings and other baselines.

We invite 60 artists who have studied Chinese ink wash paintings for eight years in average. Our stylization perceptual study is performed using the following steps:

(1) Artists are first told whose paintings styles we are going to generate using the given photos.
(2) Artists are asked to review and rate 40 groups of images. In each group, the leftmost image is the input photo randomly selected from the testing set, and other images, which are displayed in random order, are generated style-transferred paintings by our method and four baseline methods using the same input photo.
(3) Artists are asked to rank the generated paintings based on the criterion whether the void and brush strokes are applied properly, and whether the tone and ink wash diffusion looks natural. No time constraints are placed.
(4) The average score ϕ for a certain method is calculated from ranks as

$$\phi_k = \frac{1}{N_p} \sum_i \sum_j (N_m - \text{rank}_{i,j,k} + 1), \tag{9}$$

where N_m is the total number of evaluated methods in each group, N_p is the total number of participants, and i, j, k indicate the i-th participant, j-th group of images and k-th method, respectively.

5 EXPERIMENTS

We train and evaluate ChipGAN using ChipPhi dataset. We first introduce the baseline approaches adopted in our evaluation.

Gatys *et al.* [11] show that the content and style of a certain image are separable and synthesize a new image that simultaneously matches the content representation of photo and the style presentation of painting. The style representation by this method is calculated by Gram matrix, which depends on feature correlations.

Johnson *et al.* [21] train a feed-forward transformation network with perceptual loss of style and content to accelerate the process of style transfer. For our experiments, we train this style transfer network on Microsoft COCO dataset [32] based on the style of a painting which is randomly chosen from the painting set. Similar with Gatys *et al.*, this method also applies Gram matrix calculated from feature maps as style representation.

CycleGAN [50] learns a mapping $G : X \rightarrow Y$ to generate a new distribution $G(X)$ where the images are indistinguishable from the ones in domain Y. To further reduce the number of possible mappings, G is coupled with an inverse mapping $F : Y \rightarrow X$ and a cycle consistency constraint: $G(F(X)) \approx X$ is enforced. This method provides a solution to avoid mode collapse [13] and generate more realistic images in the target domain.

DistanceGAN [2] is based on CycleGAN [50] architecture. It further reduces the amount of mapping by enforcing the constraint that the distance of two samples in one domain should be preserved in the mapping to another domain.

5.1 Comparison with baselines

We compare the visual quality of generated paintings by our method against the baselines, and then use stylization perceptual study to evaluate the style similarity to real paintings of generated paintings from different approaches.

Visual quality comparison. As illustrated in the top row of Figure 3, for Horse , CNN based models (Gatys *et al.* [11] and Johnson *et al.* [21]) preserve the shapes of horses to some extent. But the generated paintings have thick strokes which are more similar to western oil paintings. Besides, these two methods fail to represent voids and show unexpected noise. In contrast, paintings generated by GAN based models (ChipGAN (ours), CycleGAN [50] and DistanceGAN [2]) look realistic, and they all have voids well represented. Compared against ChipGAN, CycleGAN [50] and DistanceGAN [2] lose some brush strokes, while adding some unnecessary ones. Among these methods, ChipGAN generates paintings with the most reasonable tone, thanks to the ink wash constraint.

The comparison of Landscape data is shown in the bottom row of Figure 3. Though CNN based models (Gatys *et al.* [11] and Johnson *et al.* [21]) can depict the contours of mountains, their results suffer from severe artifacts and cannot express the feeling of distance. As for GAN based models, the feeling of distance is well

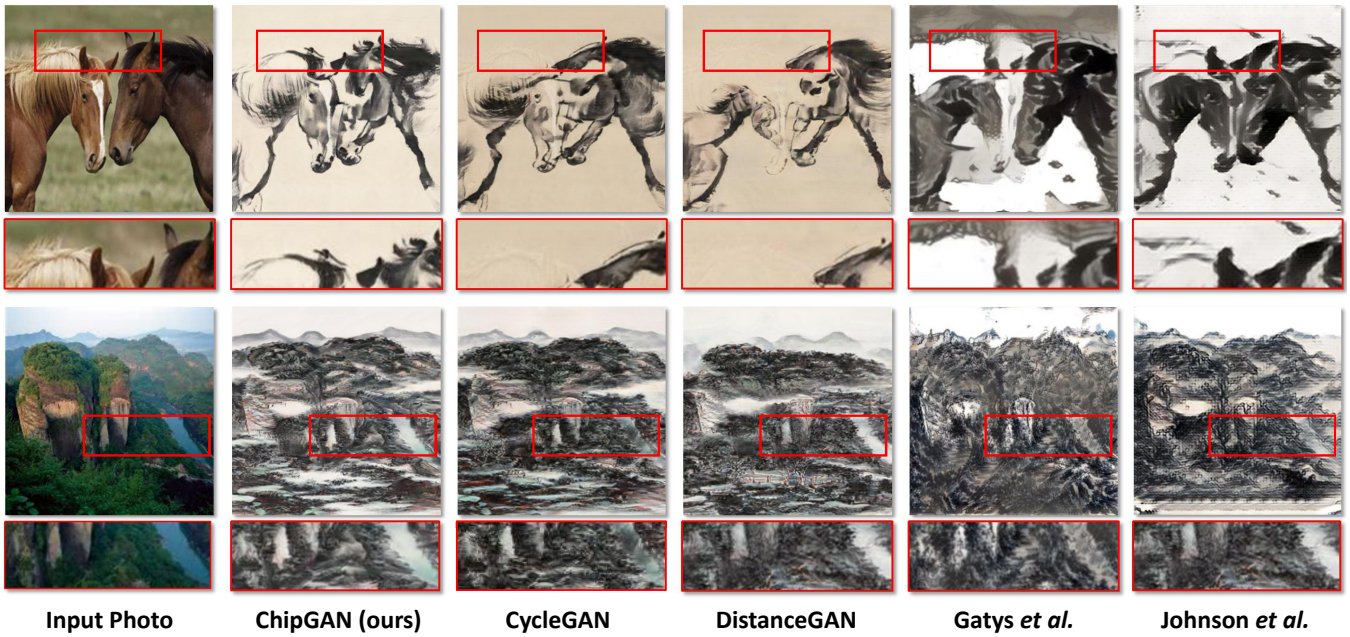

Figure 3: Visual quality comparison of different methods. From left to right: input, ChipGAN (ours), CycleGAN [50], DistanceGAN [2], Gatys *et al.* [11], and Johnson *et al.* [21]. The close-up views are provided in color boxes below the result and the detailed analysis is in Section 5.1.

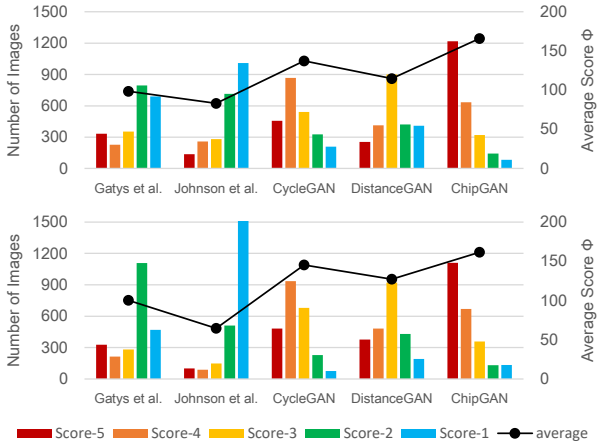

Figure 4: Stylization perceptual study result: The left vertical axis stands for the total number of images with different scores by different approaches on HORSE (top) and LANDSCAPE (bottom). The right vertical axis stands for the average score ϕ for each approach.

presented by leaving some voids. However, CycleGAN [50] and DistanceGAN [2] fail to correctly handle some smoothly textured region (*e.g.*, the close-up view of river in the second row), while ChipGAN represents it in a natural way. Similar to the results in HORSE , our method outperforms other baselines in correctly applying Chinese ink wash painting techniques.

Stylization perceptual study result. Figure 4 summarizes the scores of stylization perceptual study by different methods based on professional artists' evaluation. The test score reflects the style similarity of the generated horse paintings to the style of real paintings by Xu Beihong as well as the generated landscape paintings to the style of real paintings by Huang Binhong. It is obvious that ChipGAN has the highest score for the most numbers of images. For the average scores ϕ calculated by Equation (9), our approach outperforms the baselines in both datasets. Gatys *et al.* [11] and Johnson *et al.* [21] achieve similar scores on HORSE, but Johnson *et al.* [21] has lower score on LANDSCAPE due to the the lack of instance normalization [40] in feed-forward network which results in repetitive patterns which seldom appear in real paintings. The GAN based models have higher scores than CNN based models, since the well-presented voids areas look closer to the real paintings. Compared with CycleGAN [50], the distance preserving property of DistanceGAN [2] may result in the losing of brush strokes which is easily perceived in perceptual study and causing the lower scores. We use Kendall's W test to assess agreement among participants towards results generated by a certain model; Kendall's W [22] are 0.837 for HORSE and 0.825 for LANDSCAPE. Besides, the differences among methods are evaluated by Freidman test [10] (we observe a p-value $< \alpha$ (significance level set as 0.05) which indicates the differences are significant.).

5.2 Ablation study

Our proposed method includes three essential constraints to deal with voids, brush strokes, and ink wash tone and diffusion in Chinese ink wash paintings, respectively. Since the void constraint is

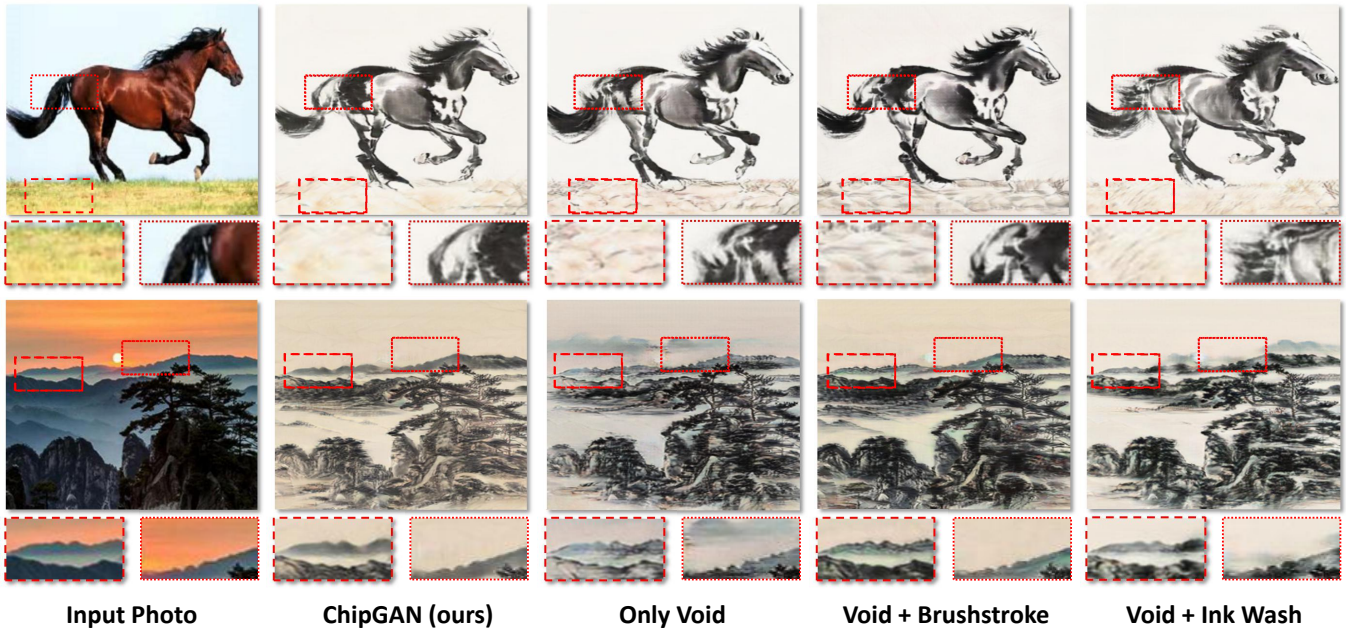

Figure 5: Visual quality comparison of different variants of our method. From left to right: input photo, ChipGAN (with full objective), with only void constraint, with void and brush stroke constraints, and with void and ink wash constraints. The dashed boxes represent differences in brush strokes and the dotted boxes represent the differences in ink wash diffusion and tone. The close-up views are provided below each image and the detailed analysis can be found in Section 5.2.

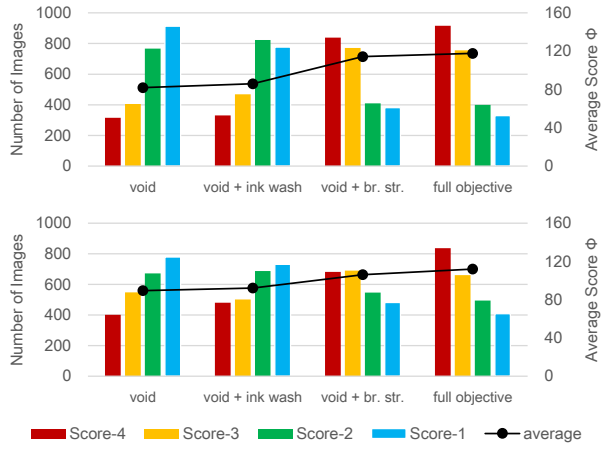

Figure 6: Stylization perceptual study result: The left vertical axis stands for the total number of images with different scores by different ablations and full objective on Horse **(top) and** Landscape **(bottom). The right vertical axis stands for the average score ϕ for each approach.**

the combination of generative adversarial loss and cycle consistency loss, which cannot be ablated from the complete network, we focus on evaluating the importance of other two constraints. We therefore train three variant networks, one for void constraint only, one for void and brush stroke constraints, the other one for void and ink wash constraints.

Visual quality comparison. We then show the results of ablation experiments in Figure 5. For Horse, method without brush stroke constraint loses essential brush strokes and adds inappropriate ones (*e.g.,* dotted boxes for the horse). And ink wash constraint helps to fade the unnecessary textures (*e.g.,* dashed boxes for the horse). As for Landscape, mountains are correctly depicted by adding brush stroke constraint (*e.g.,* dotted boxes for the landscape). Besides, methods with ink wash constraint can simulate the tone and diffusion effect of ink wash and represent a feeling of depth on the far away mountains (*e.g.,* the dashed boxes for the landscape).

Stylization perceptual study result. Figure 6 compares the complete ChipGAN against ablations of full objective in terms of style similarity to real paintings through perceptual study. Removing either brush stroke or ink wash constraint lowers the scores. When adding brush stroke constraint, the performance is largely improved. The influence of ink wash constraint is not obviously reflected in score, this difference is easy understand. The brush strokes depict essential parts of an object, so if they are applied inappropriately, the whole painting style is significantly biased. However, given the silhouette properly depicted, the ink wash constraint further refines the ink style and tone. Such improvements are not easily to perceive, but they are indispensable for the style of Chinese ink wash paintings. The Kendall's W test [22] is performed again, and we obtain Kendall's W 0.804 for Horse and 0.837 for Landscape, which indicates high agreement among participants. Similarly, the p-value $< \alpha$ (= 0.05) in the Freidman test [10] again indicates significant differences among different evaluated methods are observed.

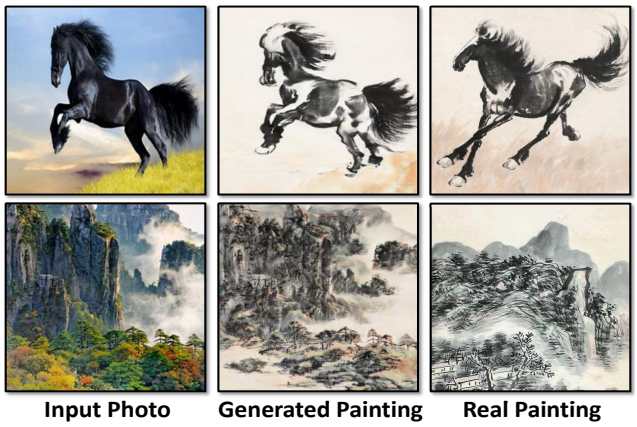

Figure 7: Comparison among input photos (left), generated paintings (middle), and real paintings (right, the horse by Xu Beihong and the landscape by Huang Binhong).

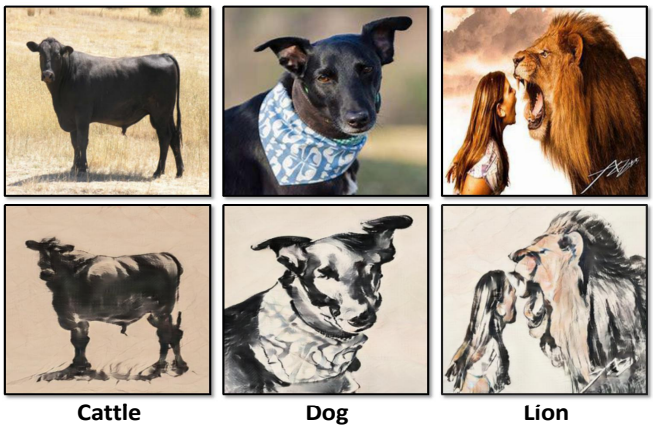

Figure 8: Chinese ink wash painting style transfer results of cattle, dog and lion with model trained on HORSE.

6 DISCUSSION

Computational methods for Chinese ink wash paintings have shown to assist ink wash painting artists and inspire their creations [45, 49]. It will be interesting to discuss how our generated paintings relate to the style of artists. Besides that, we will also discuss how our trained model generalizes to different subjects and how to combine our proposed constraints with other types of paintings.

Inspiration to painting artist. The first row of Figure 7 shows some typical techniques used to draw ink wash horses by Xu Beihong. In Xu's painting, the horsehair at tails is drawn in a fluttered style pointing to the sky to make the ink wash painting more vivid. The generated one is also depicted in a similar style, which is different from the drooping horsehair at tail in the original photo. Novice painters may be inspired by checking the subtle differences between the original photo and our generated painting to learn the expression spirit of an artist.

Different from Western realistic painters who attempt to present subject matter realistically, Chinese ink wash paintings apply voids to create artistic conceptions by omitting some unnecessary details. The second row of Figure 7 shows that in the real painting of Huang Binhong, the cloud and mist in a landscape scene is expressed by leaving voids. Since our generated painting learns the voids technique properly, by comparing the photo and painting generated by our model, the painters can learn to make a decision on what to preserve and what to omit when observing a real landscape scene.

Generalization of trained model. Because ChipGAN is designed to learn general painting techniques in Chinese ink wash paintings, the model trained on one dataset (*e.g.,* HORSE) can be adapted to other subjects. As shown in Figure 8, input photos of cattle, dog and lion with different poses are successfully transferred to ink wash painting style. The backgrounds are well handled by void constraint, and the subjects are depicted with proper brush strokes and correctly diffused ink wash.

Generalization to other types of painting. Since different types of painting share some common techniques, our constraints may be applied to other types of painting with slight modification. For example, though watercolor paintings have more abundant colors than Chinese ink wash paintings, they still require proper tones and pigment diffusions. By adjusting the erosion kernel size and deviation of Gaussian blur function, we may adapt the ink wash constraint to watercolor painting area. Another example could be woodcuts which consist of vigorous lines, we may generalize brush stroke constraint by changing its weight and adjusting the output layers of feature extractor.

Limitations. Because of the GPU memory limitation, we train our model on 256×256 images. When the resolution of input photos are high (*e.g.,* 1024×1024), the generated ink wash paintings contain chaotic lines, which is a common issue in the state-of-art methods based on fully-convolutional operation [35]. This problem can be partially solved by feeding the down sampled high-resolution images into generator and increasing the output resolution using pre-trained super resolution network [28]. An end-to-end high resolution solution for this task is our future work.

7 CONCLUSION

In this work, we propose an end-to-end weakly supervised network ChipGAN for photo to ink wash painting style transfer. This network is designed based on the three important techniques of Chinese ink wash painting: voids, brush strokes and ink wash. Experiments on the newly built "ChipPhi" dataset show effectiveness of our approach. Comparing photos, generated paintings with real paintings, we find our model is able to present techniques typically adopted by a certain artist. We hope our work can inspire computational and artistic study on Chinese ink wash painting.

ACKNOWLEDGMENTS

This work is supported by the National Natural Science Foundation of China (61661146005,U1611461), in part by the Key Research and Development Program of Beijing Municipal Science & Technology Commission (No. D171100003517002), and in part by the PKU-NTU Joint Research Institute (JRI) sponsored by a donation from the Ng Teng Fong Charitable Foundation.

REFERENCES

[1] N Ashikhmin. 2003. Fast texture transfer. *IEEE Computer Graphics and Applications* 23, 4 (2003), 38–43.

[2] Sagie Benaim and Lior Wolf. 2017. One-sided unsupervised domain mapping. In *Proc. Advances in Neural Information Processing Systems*. 752–762.

[3] Yi-Lei Chen and Chiou-Ting Hsu. 2016. Towards Deep Style Transfer: A Content-Aware Perspective.. In *Proc. British Machine Vision Conference*.

[4] Casey Chu, Andrey Zhmoginov, and Mark Sandler. 2017. CycleGAN: a Master of Steganography. *arXiv:1712.02950* (2017).

[5] Kwo Da-Wei. 1990. *Chinese Brushwork in Calligraphy and Painting: its history, aesthetics, and techniques*.

[6] Pieter-Tjerk De Boer, Dirk P Kroese, Shie Mannor, and Reuven Y Rubinstein. 2005. A tutorial on the cross-entropy method. *Annals of operations research* 134, 1 (2005), 19–67.

[7] Vincent Dumoulin, Jonathon Shlens, and Manjunath Kudlur. 2016. A learned representation for artistic style. *Computing Research Repository* 2, 4 (2016), 5.

[8] Alexei A Efros and William T Freeman. 2001. Image quilting for texture synthesis and transfer. In *Proc. Proceedings of the annual conference on Computer graphics and interactive techniques*. 341–346.

[9] Alexei A Efros and Thomas K Leung. 1999. Texture synthesis by non-parametric sampling. In *Proc. Computer Vision, 1999. The Proceedings of the Seventh IEEE International Conference*.

[10] Milton Friedman. 1937. The use of ranks to avoid the assumption of normality implicit in the analysis of variance. *Journal of the american statistical association* 32, 200 (1937), 675–701.

[11] Leon A Gatys, Alexander S Ecker, and Matthias Bethge. 2015. A neural algorithm of artistic style. *arXiv:1508.06576* (2015).

[12] Bruce Gooch and Amy Gooch. 2001. *Non-photorealistic rendering*.

[13] Ian Goodfellow. 2016. NIPS 2016 tutorial: Generative adversarial networks. *arXiv:1701.00160* (2016).

[14] Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. 2014. Generative adversarial nets. In *Proc. Advances in neural information processing systems*. 2672–2680.

[15] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. 2016. Deep residual learning for image recognition. In *Proc. Computer Vision and Pattern Recognition*. 770–778.

[16] Aaron Hertzmann, Charles E Jacobs, Nuria Oliver, Brian Curless, and David H Salesin. 2001. Image analogies. In *Proc. the annual conference on Computer graphics and interactive techniques*. 327–340.

[17] David Hockney and Charles M Falco. 2000. Optical insights into Renaissance art. *Optics and Photonics News* 11, 7 (2000), 52–59.

[18] Andrey Ignatov, Nikolay Kobyshev, Radu Timofte, Kenneth Vanhoey, and Luc Van Gool. 2017. WESPE: Weakly supervised photo enhancer for digital cameras. *arXiv:1709.01118* (2017).

[19] Phillip Isola, Jun-Yan Zhu, Tinghui Zhou, and Alexei A Efros. 2017. Image-to-image translation with conditional adversarial networks. *arXiv:1611.07004* (2017).

[20] Yongcheng Jing, Yezhou Yang, Zunlei Feng, Jingwen Ye, and Mingli Song. 2017. Neural style transfer: A review. *arXiv:1705.04058* (2017).

[21] Justin Johnson, Alexandre Alahi, and Li Fei-Fei. 2016. Perceptual losses for real-time style transfer and super-resolution. In *Proc. European Conference on Computer Vision*. Springer, 694–711.

[22] Maurice G Kendall and B Babington Smith. 1939. The problem of m rankings. *The annals of mathematical statistics* 10, 3 (1939), 275–287.

[23] Diederik P Kingma and Jimmy Ba. 2014. Adam: A method for stochastic optimization. *arXiv:1412.6980* (2014).

[24] Diederik P Kingma and Max Welling. 2013. Auto-encoding variational bayes. *arXiv:1312.6114* (2013).

[25] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton. 2012. Imagenet classification with deep convolutional neural networks. In *Proc. Advances in neural information processing systems*. 1097–1105.

[26] Vivek Kwatra, Arno Schödl, Irfan Essa, Greg Turk, and Aaron Bobick. 2003. Graphcut textures: image and video synthesis using graph cuts. In *Proc. ACM Transactions on Graphics*, Vol. 22. 277–286.

[27] Yann LeCun, Yoshua Bengio, and Geoffrey Hinton. 2015. Deep learning. *nature* 521, 7553 (2015), 436.

[28] Christian Ledig, Lucas Theis, Ferenc Huszár, Jose Caballero, Andrew Cunningham, Alejandro Acosta, Andrew Aitken, Alykhan Tejani, Johannes Totz, Zehan Wang, et al. 2016. Photo-realistic single image super-resolution using a generative adversarial network. *arXiv: 1609.04802* (2016).

[29] Chuan Li and Michael Wand. 2016. Precomputed real-time texture synthesis with markovian generative adversarial networks. In *Proc. European Conference on Computer Vision*. Springer, 702–716.

[30] Yanghao Li, Naiyan Wang, Jiaying Liu, and Xiaodi Hou. 2017. Demystifying neural style transfer. *arXiv preprint rX iv:1701.01036* (2017).

[31] Lingyu Liang and Lianwen Jin. 2013. Image-based rendering for ink painting. In *Proc. Systems, Man, and Cybernetics*.

[32] Tsung-Yi Lin, Michael Maire, Serge Belongie, James Hays, Pietro Perona, Deva Ramanan, Piotr Dollár, and C Lawrence Zitnick. 2014. Microsoft coco: Common objects in context. In *Proc. European Conference on Computer Vision*.

[33] Ming-Yu Liu, Thomas Breuel, and Jan Kautz. 2017. Unsupervised image-to-image translation networks. In *Proc. Advances in Neural Information Processing Systems*. 700–708.

[34] Ming-Yu Liu and Oncel Tuzel. 2016. Coupled generative adversarial networks. In *Proc. Advances in neural information processing systems*. 469–477.

[35] Jonathan Long, Evan Shelhamer, and Trevor Darrell. 2015. Fully convolutional networks for semantic segmentation. In *Proc. Computer Vision and Pattern Recognition*. 3431–3440.

[36] Paul Rosin and John Collomosse. 2012. *Image and Video-Based Artistic Stylisation*.

[37] David G Stork. [n. d.]. Computer vision and computer graphics analysis of paintings and drawings: An introduction to the literature. In *Proc. International Conference on Computer Analysis of Images and Patterns*.

[38] Thomas Strothotte and Stefan Schlechtweg. 2002. *Non-photorealistic computer graphics: modeling, rendering, and animation*.

[39] Dmitry Ulyanov, Vadim Lebedev, Andrea Vedaldi, and Victor S Lempitsky. 2016. Texture Networks: Feed-forward Synthesis of Textures and Stylized Images.. In *Proc. International Conference on Machine Learning*. 1349–1357.

[40] D. Ulyanov, A. Vedaldi, and V. Lempitsky. 2016. Instance Normalization: The Missing Ingredient for Fast Stylization. *ArXiv: 1607.08022* (2016).

[41] Ren-Jie Wang and Chung-Ming Wang. [n. d.]. Effective Color Ink Diffusion Synthesis. In *Proc. Intelligent Information Hiding and Multimedia Signal Processing*.

[42] Der-Lor Way, Yu-Ru Lin, Zen-Chung Shih, et al. 2002. The Synthesis of Trees in Chinese Landscape Painting Using Silhouette and Texture Strokes.. In *Proc. Inter.*

[43] Li-Yi Wei and Marc Levoy. 2000. Fast texture synthesis using tree-structured vector quantization. In *Proc. Proceedings of the annual conference on Computer graphics and interactive techniques*. 479–488.

[44] Saining Xie and Zhuowen Tu. 2015. Holistically-nested edge detection. In *Proc. International Conference on Computer Vision*. 1395–1403.

[45] Songhua Xu, Yingqing Xu, Sing Bing Kang, David H Salesin, Yunhe Pan, and Heung-Yeung Shum. 2006. Animating Chinese paintings through stroke-based decomposition. *ACM Transactions on Graphics* 25, 2 (2006), 239–267.

[46] LiJie Yang and TianChen Xu. 2013. Animating Chinese ink painting through generating reproducible brush strokes. *Science China Information Sciences* 56, 1 (2013), 1–13.

[47] Jun-Wei Yeh and Ming Ouhyoung. 2002. Non-photorealistic rendering in chinese painting of animals. *Journal of System Simulation* 1262 (2002).

[48] Rujie Yin. 2016. Content aware neural style transfer. *arXiv:1601.04568* (2016).

[49] JinHui Yu, GuoMing Luo, and QunSheng Peng. 2003. Image-based synthesis of Chinese landscape painting. *Journal of Computer Science and Technology* 18, 1 (2003), 22–28.

[50] Jun-Yan Zhu, Taesung Park, Phillip Isola, and Alexei A Efros. 2017. Unpaired image-to-image translation using cycle-consistent adversarial networks. *arXiv preprint arXiv:1703.10593* (2017).