Aesthetic-Aware Image Style Transfer

Zhiyuan Hu*
zy-hu16@tsinghua.org.cn
Department of Computer Science and Technology, Tsinghua University, Beijing 100084, China
Key Laboratory of Pervasive Computing, Ministry of Education
Beijing National Research Center for Information Science and Technology

Jia jia†
jjia@mail.tsinghua.edu.cn
Department of Computer Science and Technology, Tsinghua University, Beijing 100084, China
Key Laboratory of Pervasive Computing, Ministry of Education
Beijing National Research Center for Information Science and Technology

Bei Liu
Bei.Liu@microsoft.com
Microsoft Research

Yaohua Bu
boyh18@mails.tsinghua.edu.cn
Department of Computer Science and Technology, Tsinghua University, Beijing 100084, China
Key Laboratory of Pervasive Computing, Ministry of Education
Beijing National Research Center for Information Science and Technology

Jianlong Fu
jianf@microsoft.com
Microsoft Research

ABSTRACT

Style transfer aims to synthesize an image which inherits the content of one image while preserving a similar style of the other one. The "style" of an image usually refers to its unique feeling conveyed from visual features, which is highly related to the aesthetic effect of the image. Aesthetic effect can be mainly decomposed as two factors: *colour* and *texture*. Previous methods like Neural Style Transfer and Colour Transfer have shown strong abilities in transferring colour and texture features. However, such approaches neglect to further disentangle colour and texture, which makes some of unique aesthetic effects designed by human artists hard to express. In this paper, we propose a novel problem called Aesthetic-Aware Image Style Transfer task, which aims to transfer colour and texture separately and independently to manipulate the aesthetic effect of an image. We propose a novel *Aesthetic-Aware Model-Optimisation-Based Style Transfer (AAMOBST)* model to solve this problem. Specifically, AAMOBST is a multi-reference, two-path model. It uses different reference images to decide desired colour and texture features. It can segregate colour and texture into two distinct paths and transfer them independently. Qualitative and quantitative experiments show that our model can decide colour and texture features separately and is able to keep one of them fixed while changing the other one, which is not applicable for previous methods. Furthermore, on tasks that are applicable for previous methods (such as style transfer, colour-preserved transfer and colour-only transfer), our model shows comparable abilities with other baseline methods.

CCS CONCEPTS

• **Computing methodologies → Computational photography**; **Non-photorealistic rendering**.

KEYWORDS

style transfer, image aesthetics analysis

ACM Reference Format:
Zhiyuan Hu, Jia jia, Bei Liu, Yaohua Bu, and Jianlong Fu. 2020. Aesthetic-Aware Image Style Transfer. In *Proceedings of the 28th ACM International Conference on Multimedia (MM '20), October 12–16, 2020, Seattle, WA, USA.* ACM, New York, NY, USA, 10 pages. https://doi.org/10.1145/3394171.3413853

1 INTRODUCTION

An image can have its unique style that gives us a special feeling. Such style features are extremely obvious in art paintings: Vincent Willem van Gogh's paintings have delicate and colourful brush-strokes. The paintings of Impressionism blur brush-fire texture to create a sense of haze. Ink wash paintings use black as the only colour to create a feeling of quaintness. This kind of distinctive feeling can also be interpreted by the aesthetic effect of an image.

Finding the constituent elements of aesthetic effect and leveraging them to achieve style transfer have been discussed for a

*This work was performed when Zhiyuan Hu was visiting Microsoft Research as a research intern.
†Corresponding author

MM '20, October 12–16, 2020, Seattle, WA, USA

ACM ISBN 978-1-4503-7988-5/20/10…$15.00
https://doi.org/10.1145/3394171.3413853

(a) The Starry Night (b) Sunflowers (c) San Giorgio Maggiore by Twilight

Figure 1: Both colour and texture are indispensable components of aesthetic effect. They can influence aesthetic effect significantly and independently. [Best viewed in colour.]

long time. [20, 40, 42, 43] point out that the aesthetic effect conveyed by an image can be interpreted by its colour combination. Based on such observation, several Colour Transfer methods like [1, 3, 13, 26, 48] have been proposed to manipulate the aesthetic effect of an image by changing its colour patterns. However, to guarantee that the final changes on aesthetic effect are purely dependent on colour features, these methods fix the luminous channel of the original image to make its content and texture features unchanged. Gatys et al. [7] find that second-order statistics like Gram matrix are related to the colour and texture features of an image. By replacing such statistics [8], the aesthetic effect of the style reference image can be transferred to the content image. Style transfer models like [14, 25, 27, 47] explore different methods to better transfer second-order statistics. However, these methods usually use a single reference image to decide the colour and texture features at the same time. What's more, they represent colour and texture collaboratively as one single "style" feature. [9, 16, 45, 46] decompose style features into perceptual factors like spatial regions, colour patterns and brushstroke size. Although their methods can properly control texture features like brushstroke size, the significance of colour is neglected in their methods. In summary, the constituent elements of aesthetic effect proposed by these methods can be mainly classified into two independent categories: colour factors such as the colour theme, brightness, contrast ratio, etc. and texture factors which include the shape of lines, the size of brushstroke, the sharpness of contours, etc. For example, Figure 1(a) and Figure 1(b) both have curly and delicate brushstrokes. However, The Starry Night has dark, blue colour combinations and conveys a feeling of sadness and depression while Sunflowers looks much more joyous and positive, as it has a bright, yellow colour pattern. Figure 1(b) and Figure 1(c) both have a bright colour, but they convey very different feelings. The texture of the former is clear, and the latter has blurred texture thus perfectly convey the feeling of haze in San Giorgio Maggiore by Twilight. Inspired by these, we consider aesthetic effect from two aspects: *colour* and *texture*, which can both influence the aesthetic effect significantly and independently.

In this paper, we propose a novel problem called Aesthetic-Aware Image Style Transfer, which aims to transfer colour and texture separately and independently to manipulate the aesthetic effect of an image. A novel model called *Aesthetic-Aware Model-Optimisation-Based Style Transfer (AAMOBST)* [17] is proposed. Specifically, AAMOBST is a multi-reference, two-path model. It takes one original image and two reference images as input to decide the content, colour features and texture features of the target image separately. To guarantee the independence of colour and texture, we first exploit LAB transformation [18] to decompose content and reference images into two independent parts: L channel and AB channels. Colour and texture features are disentangled into these two distinct parts. Then we use two independent transfer modules in L-path and AB-path to transfer colour and texture features from reference images to the target image respectively.

We conduct extensive experiments to verify the effectiveness of our model: 1) We demonstrate that our model can completely control the aesthetic effect of an image by transferring colour features and texture features from different reference images, which is not applicable in previous models; 2) We conduct experiments like style transfer, colour-preserved transfer and colour-only transfer similar to previous works. We show that our model can produce comparable results with other baseline methods on these tasks.

We summarize the contributions of this work as follows:

- We introduce a novel Aesthetic-Aware Image Style Transfer task, which aims to control the aesthetic of an image completely. Such a task requires independent manipulation of colour features and texture features of an image. To our best knowledge, this is the first work that proposes such a problem and works on it.
- We use multiple reference images as input to decide the colour and texture features of the target image, which makes both colour and texture controllable for style transfer. Experiments show that our model can generate a target image that inherits colour and texture from different reference images.
- We design a two-path structure in our model, which can segregate colour features and texture features in two different domains and transfer them independently. Such results can not be produced by previous methods. Experiments also show that our model can fix one of these two features alone and only transfer the other one.

The rest of paper is organized as follows. Section 2 lists related works. Section 3 formulates the problem. Section 4 presents the methodologies. Section 5 introduces the dataset, training details and experiment results. Section 6 is the conclusion.

2 RELATED WORK

Neural Style Transfer. Neural Style Transfer (NST) takes a content image and a style reference image to synthesis an image that feels like the latter while preserving the content of the former. Gatys et al. [8] use co-variance matrix to represent style feature and propose a model which iteratively minimizes content and style loss. As such optimization based method is computationally expensive, feed-forward methods like [19, 23, 33, 41] are proposed to improve the efficiency problem. However, these per-style-per-model methods can hardly generalize to universal styles. Arbitrary-style-per-model methods like [4, 10, 14, 21, 24, 25, 30, 36] explore different second-order statistical transformations to conduct universal style transfer. AdaIN [14] transfer the mean and standard deviation from style feature map to content feature map. WCT [25] conduct whitening and colouring operation to ensure target feature map has the same co-variance matrix as style feature map. OST [30] formulates the

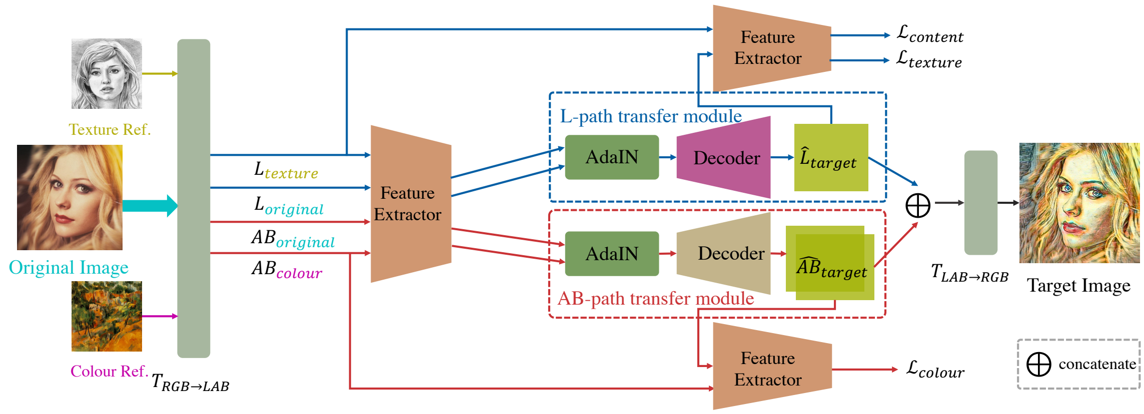

Figure 2: The architecture of Aesthetic-Aware Model-Optimisation-Based Style Transfer (AAMOBST) which consists of four parts: (1) LAB Transformation which transfers input image from RGB colour space to CIELAB colour space. (2) Feature Extractor which extracts content, colour and texture features of input image with CNN. (3) AB-path transfer module and (4) L-path transfer module which conduct colour and texture transfer separately to synthesis target images.[Best viewed in colour.]

universal style transfer problem and finds a mathematically optimal solution to the transformation.

Neural Style Transfer methods can modify the aesthetic effect of an image. But they neglect the significance of decomposing aesthetic effect into colour and texture features and controlling them independently. Thus reduce the diversity of aesthetic effects that can be generated.

Colour features controlling. To control the colour patterns of an image has also been widely discussed. Colour Transfer (or Image Recolouring) methods like [11, 31, 34, 49, 50] modify the colour pattern of an image based on given colour features. Most Colour Transfer models will first transfer original image into CIELAB colour space [18], fix the L channel to preserve the content unchanged and conduct colour transfer only on AB channels. Recently Palette-based Colour Transfer models which take a colour palette as input have become the mainstream. Cho et al. [3] propose a deep neural network which directly embed the colour palette into a U-net [35] framework. To avoid the network learning a trivial transformation, they also exploit the Conditional GAN [15] mechanism during training. Bahng et al. [1] propose a two stage model which generate colour palettes and then use the palettes generated to conduct palette based colourization. Their model takes natural language text as input and can generate diverse palettes correspond to text input. Li et al. [26] propose a palette extraction network and decompose original image into a colour space based on the palette extracted. Such decomposition can encode more elaborate details and can gives better guidance for the colourization process.

Texture features controlling. A serious of work have been proposed to control texture related features during style transfer. Gatys et al. [9] decompose style features into spatial regions, colour and brushstroke scales. By applying guidance channels, their model can control styles transferred in different regions. For colour controlling, their model mainly focuses on circumstances when the original colour of content image need to be preserved. They achieve this requirement by transforming content image's colour pattern to style reference image before style transfer or conducting style transfer on luminous channel only. As for stroke size control, their model transfer texture with desired stroke size to style reference image before style transfer to manipulate stroke scale in the target image. Among these three factors proposed by Gatys et al., stroke size has received much more attention from latter researchers. Jing et al. [16] propose a StrokePyramid module to extract different stroke size features and transfer them to continually control stroke size. Wang et al. [45] exploit a multimodal framework to apply coarse-to-fine stylisation procedure. Recently Yao et al. [46] apply self-attention mechanism to guide stroke transferring in different spatial positions.

Nevertheless, previous methods either entangle colour features and texture features together during transferring or are not able to control one of these two features. Such defect means that Aesthetic-Aware Image Style Transfer task can not be solved by these methods.

3 PROBLEM FORMULATION

To describe our problem more intuitively, we formulate our problem as follows: given a content image I_c, a colour reference image I_{rc} and a texture reference image I_{rt}, we aim to synthesis a target image I_t which : 1) remains the content of I_c, 2) has the same colour pattern as I_{rc} and 3) has the same texture feature as I_{rt}:

$$f : (I_c, I_{rc}, I_{rt}) \Rightarrow I_t. \quad (1)$$

4 METHODOLOGY

To transfer target aesthetic effect from reference images to content image, we propose the Aesthetic-Aware Model-Optimisation-Based Style Transfer (AAMOBST) model which mainly consists of four modules: LAB Transformation, Feature Extractor, AB-path transfer module and L-path transfer module. The overall architecture of our proposed approach is illustrated in Figure 2.

4.1 LAB Transformation

The main problem of previous style transfer models is that colour and texture features are represented together as the style features, which are extracted by an encoder from the original RGB image. To tackle this problem, we first build a LAB Transformation module $T_{RGB\to LAB}$ to transfer input images from RGB colour space into CIELAB [18] colour space:

$$AB_x, L_x = T_{RGB\to LAB}(I_x), x \in \{c, rc, rt\}, \quad (2)$$

where $I_x \in R^{3\times H\times W}$ is the input image in RGB space, $AB_x \in R^{2\times H\times W}$ and $L_x \in R^{1\times H\times W}$ are the A,B channels and L channel of the input image after transformation.

Such transformation is commonly used in Colour Transfer models. The input image is decomposed into AB channels which denote the colour distribution of the original image and L channel, which only contains information about content, contour and line features. As such, content, colour features and texture features are represented separately.

4.2 Feature Extractor

Similar to [14, 25], we use the first few layers (up to *relu4_1*) of a pre-trained and fixed VGG-19 [39] as our extractor E to encode features of AB_c, AB_{rc}, L_c and L_{rt}, :

$$f_{X_y} = E(X_y), (X, y) \in \{(AB, c), (AB, rc), (L, c), (L, rt)\}, \quad (3)$$

where f_{X_y} is the feature map of X_y.

Although VGG-19 is used to extract semantic information of an image, the shallow layers of it are believed to extract low level features such as lines, corners, etc. Colour pattern can also be persevered in the first few layers of VGG-19. Thus we directly apply the pre-trained VGG-19 on our AB channels and L channel to extract colour and texture features.

4.3 AB-path and L-path Transfer Modules

AB-path and L-path transfer modules are two independent modules that synthesize the AB channels and L channel of the target image. Both AB-path and L-path contain an AdaIN [14] module and a decoder D. AdaIN module replaces the channel-wise mean and standard deviation from one feature map to another:

$$\text{AdaIN}(x, y) = \sigma(y)(\frac{x - \mu(x)}{\sigma(x)}) + \mu(y), \quad (4)$$

where $\mu(x)$ calculates the mean of x and $\sigma(x)$ calculates the standard deviation of x.

As suggested in [7], second-order statistics like co-variance matrix of feature map can encode both colour and texture information. The standard deviation, which corresponds to terms in the main diagonal of co-variance matrix, can have a similar effect to co-variance matrix. Thus we use AdaIN in our model to transfer colour and texture features. Similar to [14, 25], we use two interpolation weights α and β to control the degree of colour and texture transferred to the target image:

$$\hat{f}_{AB_t} = (1 - \alpha)f_{AB_c} + \alpha\text{AdaIN}(f_{AB_c}, f_{AB_{rc}}), \quad (5)$$

$$\hat{f}_{L_t} = (1 - \beta)f_{L_c} + \beta\text{AdaIN}(f_{L_c}, f_{L_{rt}}), \quad (6)$$

where $\hat{f}_{AB_t}$ and $\hat{f}_{L_t}$ are predicted feature maps for the target image's AB channels and L channel.

Moreover, as L channel has ablated the effect of colour, AdaIN on AB-path is able to transfer colour feature alone to target image. While on L channel, only the texture features are transferred through AdaIN.

After AdaIN modules, we set the predicted features as the inputs of D_{AB} and D_L to generate the AB channels and L channel of the target image with two decoders:

$$\hat{AB}_t = D_{AB}(\hat{f}_{AB_t}), \quad (7)$$

$$\hat{L}_t = D_L(\hat{f}_{L_t}). \quad (8)$$

The decoders have similar structures to feature extractor except that we replace the pooling operation with up-sampling.

Finally, the AB channels and L channel of the target image are concatenated together and transferred back to RGB space:

$$\hat{I}_t = T_{LAB\to RGB}(\hat{AB}_t \oplus \hat{L}_t). \quad (9)$$

4.4 Objective Function

As our model factorizes image features into two paths, we need to carefully evaluate loss terms to reflect the model's performance in preserving content, transferring colour and texture. As colour features lie in AB-path and content and texture features lie in L-path, it is natural to evaluate colour loss in AB-path and content and texture loss in L-path.

In practice, we find that commonly used content loss that calculates the Euclidean distance between the feature map of the content image and the feature map of the target image can also work well in a luminous channel only situation. Thus similar to [8, 14], we define the content loss $\mathcal{L}_c$ as follows:

$$\mathcal{L}_{content} = ||E(L_c) - E(\hat{L}_t)||_2. \quad (10)$$

Feature map's second-order statistics like Gram matrix and standard deviation are frequently used in previous methods to evaluate style loss. Such loss will take colour and texture into consideration together. However, if the feature map itself does not contain any information about colour or texture, then the style loss can be used to evaluate colour or texture alone. For example, the colour features in L-path can be regarded as zero or the same for all images. Thus the colour features will contribute a constant term in style loss, which can be ignored. Such analysis can also be applied to the calculation of colour loss.

Thus we calculate the loss similar to [14] as the evaluation of colour and texture:

$$\mathcal{L}_{colour} = \Sigma_{i=1}^{N}||\mu(\phi_i(E(AB_{rc}))) - \mu(\phi_i(E(AB_t)))||_2 + \Sigma_{i=1}^{N}||\sigma(\phi_i(E(AB_{rc}))) - \sigma(\phi_i(E(AB_t)))||_2, \quad (11)$$

$$\mathcal{L}_{texture} = \Sigma_{i=1}^{N}||\mu(\phi_i(E(L_{rt}))) - \mu(\phi_i(E(L_t)))||_2 + \Sigma_{i=1}^{N}||\sigma(\phi_i(E(L_{rt}))) - \sigma(\phi_i(E(L_t)))||_2, \quad (12)$$

where each ϕ_i denotes a mid layer of VGG-19 and N is the total number of mid layers. We choose *relu1_1*, *relu2_1*, *relu3_1*, *relu4_1* as mid layers in our model.

The three loss terms are weighted and summed together as the total loss:

$$\mathcal{L} = \lambda_{content}\mathcal{L}_{content} + \lambda_{colour}\mathcal{L}_{colour} + \lambda_{texture}\mathcal{L}_{texture}. \quad (13)$$

5 EXPERIMENTS

In this section, we first introduce the datasets and training details of our model. Then we demonstrate three experiments to verify the effectiveness of our model.

5.1 Datasets

To conduct the training of our model, we need to collect three datasets as the input: content dataset, colour reference dataset and texture reference dataset. For the content dataset, we choose MS-COCO dataset [28] which contains 82,783 images and 80 different categories of objects. Such large-scale and diverse image dataset can help our model adapt to different domains.

For the colour reference dataset and texture reference dataset. It is unsuitable to choose human-photographed photo datasets like [5, 6, 22, 28]. Such photos are featureless in colour and texture. The colour features and texture features are always observed in art paintings. So we collect 8017 art paintings from WikiArt dataset [32], which contains masterpieces of lots of famous artists. We split these art paintings in half as our colour reference and texture reference datasets.

5.2 Training details

To fit the input format of VGG-19, we resize all input images to 256 × 256 before LAB transformation. As VGG-19 takes input image with 3-channel, we concatenate a channel with all zero terms to AB channels as the input of AB-path and duplicate L channel for 3 times for the input of L-path.

We train our model on one GeForce RTX 2080Ti GPU with 8 images in a mini batch. We set the initial learning rate to be 0.0001 and use a cosine scheduler [29] to adjust the learning rate during training. The weights for content loss is empirically set as 1.0. Texture loss and colour loss are both empirically set as 10.0. We train our model for 160,000 iterations. In each iteration, we randomly choose a mini batch of content, colour and texture images as inputs to train our model.

5.3 Experimental Setup

We conduct three experiments to verify the effectiveness of our model in Aesthetic-Aware Image Style Transfer task, traditional style transfer task and other traditional tasks like colour-preserved transfer and colour-only transfer.

5.3.1 Aesthetic-Aware Image Style Transfer. As is mentioned before, the aesthetic effect of an image can be decomposed into colour and texture. We argue that colour and texture must be both independent and controllable to conduct Aesthetic-Aware Image Style Transfer.

We verify our model's ability to control the aesthetic effect of an image in two aspects:

1) Transfer colour and texture together. We choose 3 colour reference images and 3 texture reference images. The free combination of colour and texture can generate 9 different aesthetic effects. We use our AAMOBST model to synthesize a target image with a specific aesthetic effect.

2) Content-Colour-Texture trade-off. Our model allows the trade-off between content, colour and texture by interpolating between feature maps of content, colour and texture. We gradually change α in Eq. 5 and β in Eq. 6 from 0 to 0.33, 0.67 and 1.0 separately and use our model to generate target images with different degrees of colour features and texture features.

5.3.2 Traditional Style Transfer. Traditional style transfer task transfers the colour and texture from the same style reference image, which is a special case of aesthetic effect controlling. As our model can manipulate colour and texture freely, it should also be able to conduct style transfer properly. To evaluate the effectiveness of our model, we test our model and other comparison methods in traditional style transfer task. We set the colour reference image and texture reference image in our model to the same one to transfer the overall style of this image.

We choose several classic and frequently used methods in style transfer tasks as comparison methods. **Gatys et al.** [8] propose the first optimization based, Neural Network model to conduct style transfer. **AdaIN** [14] is a Arbitrary-Style-Per-Model (ASPM), feed forward method. It exploits Adaptive Instance Normalization to transfer second-order statistic from reference image to content image. **WCT** [25] takes a similar framework as AdaIN, but it uses a Whitening and a Colouring process rather than Adaptive Instance Normalization to conduct transformation.

We conduct qualitative evaluation and user study to evaluate the performance of these methods.

1) Qualitative Experiments. We choose five content and style reference images to test our model and other baseline models. The style images we choose are from diverse art styles include Abstractionism, Sketch, Impressionism, Oil Painting, and Cubism.

2) User Study. As the evaluation of style transfer is highly subjective, we conduct a user study to evaluate the performance of 4 methods shown in Figure 5. We hire 24 volunteers for our user study. Their ages range from 21 to 49. 41.7% of them are females and 58.3% are males. Their majors include computer science, chemistry, machine engineering, mathematics, etc.

We randomly choose 12 content images and 20 style images to get 240 content-style pairs. Then we randomly choose 100 pairs from these 240 pairs as test pairs. We divide the 100 test pairs into 5 groups. We randomly choose a group for each subject to evaluate. We display the 4 stylized images by the compared methods in random order together with their style reference image. Each subject is asked to vote 3 results that are most similar with style image in terms of 1. colour, 2. texture and 3. overall feeling. The subject can choose the same result during these 3 votes. The average time for a volunteer to finish the questionnaire is 8.75 minutes. We collected the feedback from 24 subjects with totally 480 votes.

5.3.3 Colour-preserved transfer and colour-only transfer. How to preserve the original colour of the content image (colour-preserved style transfer) and how to transfer colour features only (photo-realistic style transfer) are two problems frequently discussed in the field of style transfer. Since colour and texture are independently controlled in our model, we can fix one of them and transfer the other one alone to conduct colour-preserved style transfer or colour-only style transfer.

1) Colour-preserved transfer. We fix the AB-channel of original content image and only use our L-path to transfer texture to preserve the original colour features. In this sense, our model is very similar to the first colour-preserving method mentioned in [9] to preserve colour patterns.

We select the method of **Gatys et al.**[9] and **AdaIN**[14] as the comparison methods.

2) Colour-only transfer. We fix the L-channel and conduct colour-only transfer by AB-path. We choose a series of paintings of *Monet*, who observed the light and shadow changes of the same water lily pool at different times, as our colour reference images.

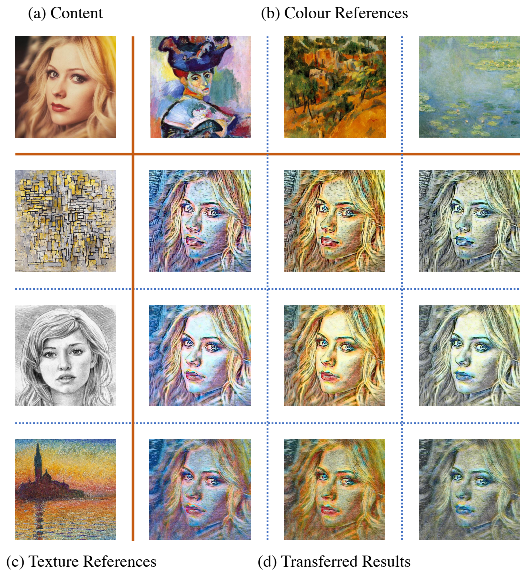

Figure 3: Result examples of transferring colour and texture together. The result in row i, column j of (d) inherits the texture of i-th image in (c) and the colour of j-th image in (b). [Best viewed in colour.]

We select the method of **Chang et al.**[2] and the method of **He et al.**[12] as the comparison methods. Note that, the method of **Chang et al.** is a palette based colour transfer method, the target palette is extracted from the colour reference images using the algorithm in their model.

5.4 Experiments Results and Analysis

5.4.1 Results of Aesthetic-Aware Image Style Transfer. The transferred results are shown in Figure 3. The images in the same row have similar texture features to the texture image on their left. Images in the same column have similar colour features to the colour image on top of them. Each of these 9 transferred results has its own art taste. For example, the image at the top-left has crack-like textures and cool colours which give a sense of tension. Image right in the middle has clear lines and fresh and bright colours. It looks much more cheerful than the former one. The image at right-bottom is blurred and has dark, green colour patterns. This image gives a sense of dark and depressing. All these images can not be generated by traditional style transfer methods since they can not fully control aesthetic effects.

Suppose that colour and texture are two independent axes that expand an Aesthetic Effect Space. Previous Style Transfer and Colour Transfer methods can not cover all this space as the two axes are not independent or one of them is not controllable. But each point in this space corresponds to a meaningful aesthetic effect. In our model, the independence and controllability of colour texture are properly reflected. The images generated from our model can cover the whole Aesthetic Effect Space and are more diverse than what previous methods can generate.

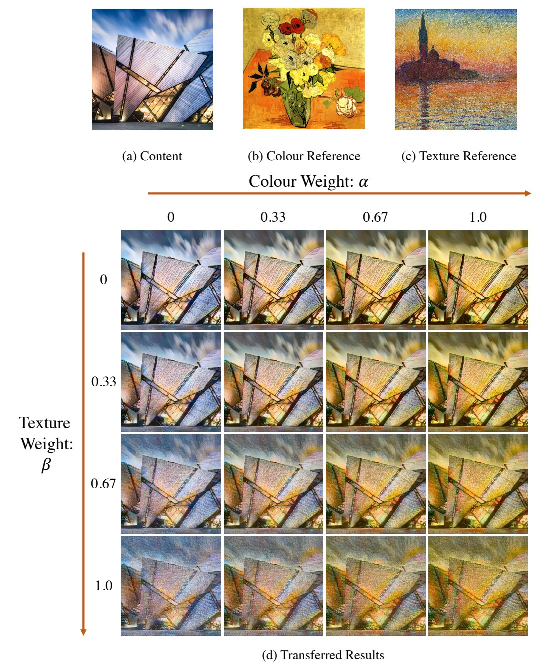

Figure 4: Result examples of Content-Colour-Texture trade-off. We control the hybrid degree of colour (α in Eq. 5) and texture (β in Eq. 6) in (d). We increase α from left to right and increase β from top to bottom. [Best viewed in colour.]

As shown in Figure 4, our model is able to control the texture similarity and colour similarity between content image and reference images smoothly. From left to right, the colour taste changes from cold to warm. From top to bottom, the texture changes from clear to blurred. Traditional style transfer models can only allow the trade-off between style and content and can only generate images on the diagonal of Figure 4. Our model has the capability to control the style (or aesthetic effect) in a finer manner. Colour and texture can be hybridized at any degree and generate more diverse texture-colour combinations.

Table 1: Results of user study. The user preferences of colour, texture and overall similarity among the 4 methods in Figure 5 are listed in this table. Our model can better transfer texture features and has comparable abilities with other methods in transferring colour and overall feeling.

Methods	Colour Preference/%	Texture Preference/%	Overall Preference/%
Gatys et al. [8]	5.71	5.71	6.79
AdaIN [14]	8.57	11.79	10.0
WCT [25]	**43.5**	38.86	40.62
AAMOBST (Ours)	42.22	**43.64**	**42.59**

5.4.2 Results of Traditional Style Transfer. The results of qualitative experiments in Figure 5 show that our model has comparable abilities in style transfer with other baseline methods. Our model

Figure 5: Results on Traditional Style Transfer. Column (a) and (b) are content and style images. Column (c) to (f) are transferred results of compared methods and our model.[Best viewed in colour.]

Figure 6: Zoom in results of the second test case in Figure 5. Our model can better preserve detailed features of content image. Our model can better preserve fine structures of the content image.

can faithfully transfer the style features of the reference image. Moreover, since our model transfers the luminous channel in a separate path, the luminous distribution of the content image is less influenced by colour features. Thus the fine-grained details of the content image, which is sensitive to luminous, can be better preserved. For example, we zoom in the third case in Figure 5 to get Figure 6, the wrinkles, thin hair and eyebrows are still clear in the result of our model, while for other models like WCT [25], since colour and texture are transferred together, such fine-grained structure is easy to be blurred. Such preservation of fine-grained structures can also be observed in other cases of Figure 5: the tiny tree branches are still clear in the forth. In the third and fifth cases, the lines of windows keep their straightness.

We show the percentage of the votes each method received in Table 1. The results show that our model can better transfer texture features. The ability to transfer colour and overall feeling of our model are also comparable with other methods.

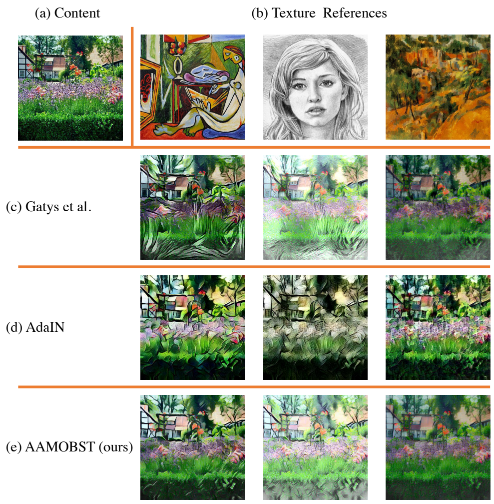

Figure 7: Results on colour-preserved transfer. Our model can preserve the colour distribution of content image. The results are comparable with other methods like Gatys et al.[9] and AdaIN[14]. [Best viewed in colour.]

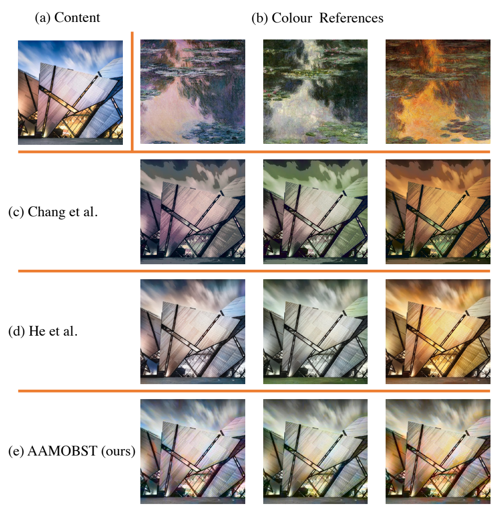

Figure 8: Results on colour-only transfer. Our model can faithfully transfer the colour features of reference images and the results are comparable with other baseline methods like Chang et al.[2] and He et al.[12]. [Best viewed in colour.]

The preferences of our model in colour is slightly lower than WCT. This is also reasonable since WCT tends to sacrifice more on the authenticity of content features and enhance the degree of stylization. But our model tend to be faithful to original content. It is worthwhile to take a small sacrifice in colour in exchange of much less distortion in content.

5.4.3 Results of colour-preserved transfer and colour-only transfer. As shown in Figure 7, our model is able to apply different textures to the content image, while keeping its colour distribution unchanged. Other methods like AdaIN[14] and the second colour preserving method in Gatys et al.[9] suggest to first transfer the colour patterns of the content image to the style image and then conduct normal style transfer. However, such methods can not fully ablate the colour features of style image because the colour transfer from content image to style image can not be perfect. While since our model transfer texture in the luminous channel only, it can fully remove the effect of the texture reference images.

Figure 8 shows that, similar to other classic Colour Transfer methods like Chang et al.[2] and He et al.[12], our model can keep the content totally unchanged while modifying its colour patterns. The transferred results keep the shape and lines of the content image but can reflect different impressions like the water lily pool paintings at different times drawn by Monet.

5.4.4 Error analysis. Our model has shown strong abilities in manipulating colour and texture to generate images with diverse aesthetic effects. However, as Aesthetic-Aware Image Style Transfer is a newly proposed problem. It has also brought some open problems:

1) Unlike RGB colour space in which the values of R, G and B are independent. The values of L, A, and B must meet some specific restrictions [18] to make the colour a valid one. Such restrictions are non-linear and difficult to express. How to apply such restrictions to ensure every predicted pixel has a valid colour remains a problem.

2) As is suggested by [37, 38, 44], colour and texture of an image must be suitable to each other to make the image looks harmony. How to refine the predicted colour and texture features to make them suitable to each other is a valuable question.

6 CONCLUSION

In this paper, we first propose a novel task called *Aesthetic-Aware Image Style Transfer*. We point out that colour and texture are two independent components of aesthetic effect. We clarify that to make the style transfer aesthetic-aware, colour and texture should be independent and controllable during transferring. Then we propose a novel Aesthetic-Aware Model-Optimisation-Based Style Transfer (AAMOBST) model, which can transfer colour and texture in two independent paths and evaluate colour and texture loss alone. It exploits LAB transformation to disentangle colour and texture features. Finally, we conduct extensive experiments including qualitative experiments and user study to show that, our model can freely control the aesthetic effect of an image by using different tastes of colour and texture or controlling the transformation degree of them. Such tasks are not applicable in previous style transfer methods. Compared with previous methods, our model is able to control the aesthetic effect of an image in a finer-grained manner and can generate more diverse images. On other tasks like style transfer, colour-preserved style transfer and colour-only style transfer, our model shows comparable capabilities with other baseline models and can better preserve fine structures of the original content image.

7 ACKNOWLEDGEMENT

This work is supported by the state key program of the National Natural Science Foundation of China (NSFC) (No.61831022),and the National Natural, and Science Foundation of China (61521002), Beijing Academy of Artificial Intelligence (BAAI).

REFERENCES

[1] Hyojin Bahng, Seungjoo Yoo, Wonwoong Cho, David Keetae Park, Ziming Wu, Xiaojuan Ma, and Jaegul Choo. 2018. Coloring with Words: Guiding Image Colorization Through Text-Based Palette Generation. In *Computer Vision - ECCV 2018 - 15th European Conference, Munich, Germany, September 8-14, 2018, Proceedings, Part XII (Lecture Notes in Computer Science)*, Vittorio Ferrari, Martial Hebert, Cristian Sminchisescu, and Yair Weiss (Eds.), Vol. 11216. Springer, 443–459.

[2] Huiwen Chang, Ohad Fried, Yiming Liu, Stephen DiVerdi, and Adam Finkelstein. 2015. Palette-based photo recoloring. *ACM Trans. Graph.* 34, 4 (2015), 139:1–139:11.

[3] Junho Cho, Sangdoo Yun, Kyoungmu Lee, and Jin Young Choi. 2017. PaletteNet: Image Recolorization with Given Color Palette. In *2017 IEEE Conference on Computer Vision and Pattern Recognition Workshops, CVPR Workshops 2017, Honolulu, HI, USA, July 21-26, 2017*. IEEE Computer Society, 1058–1066.

[4] Wonwoong Cho, Sungha Choi, David Keetae Park, Inkyu Shin, and Jaegul Choo. 2019. Image-To-Image Translation via Group-Wise Deep Whitening-And-Coloring Transformation. In *IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2019, Long Beach, CA, USA, June 16-20, 2019*. Computer Vision Foundation / IEEE, 10639–10647.

[5] Jia Deng, Wei Dong, Richard Socher, Li-Jia Li, Kai Li, and Fei-Fei Li. 2009. ImageNet: A large-scale hierarchical image database. In *2009 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR 2009), 20-25 June 2009, Miami, Florida, USA*. IEEE Computer Society, 248–255.

[6] Mark Everingham, Luc Van Gool, Christopher K. I. Williams, John M. Winn, and Andrew Zisserman. 2010. The Pascal Visual Object Classes (VOC) Challenge. *Int. J. Comput. Vis.* 88, 2 (2010), 303–338.

[7] Leon A. Gatys, Alexander S. Ecker, and Matthias Bethge. 2015. Texture Synthesis Using Convolutional Neural Networks. In *Advances in Neural Information Processing Systems 28: Annual Conference on Neural Information Processing Systems 2015, December 7-12, 2015, Montreal, Quebec, Canada*, Corinna Cortes, Neil D. Lawrence, Daniel D. Lee, Masashi Sugiyama, and Roman Garnett (Eds.). 262–270.

[8] Leon A Gatys, Alexander S Ecker, and Matthias Bethge. 2016. Image style transfer using convolutional neural networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2414–2423.

[9] Leon A. Gatys, Alexander S. Ecker, Matthias Bethge, Aaron Hertzmann, and Eli Shechtman. 2017. Controlling Perceptual Factors in Neural Style Transfer. In *2017 IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2017, Honolulu, HI, USA, July 21-26, 2017*. IEEE Computer Society, 3730–3738.

[10] Golnaz Ghiasi, Honglak Lee, Manjunath Kudlur, Vincent Dumoulin, and Jonathon Shlens. 2017. Exploring the structure of a real-time, arbitrary neural artistic stylization network. (2017).

[11] Sergio Guadarrama, Ryan Dahl, David Bieber, Jonathon Shlens, Mohammad Norouzi, and Kevin Murphy. 2017. PixColor: Pixel Recursive Colorization. (2017).

[12] Mingming He, Dongdong Chen, Jing Liao, Pedro V. Sander, and Lu Yuan. 2018. Deep exemplar-based colorization. *ACM Trans. Graph.* 37, 4 (2018), 47:1–47:16.

[13] Mingming He, Jing Liao, Dongdong Chen, Lu Yuan, and Pedro V. Sander. 2019. Progressive Color Transfer With Dense Semantic Correspondences. *ACM Trans. Graph.* 38, 2 (2019), 13:1–13:18.

[14] Xun Huang and Serge J. Belongie. 2017. Arbitrary Style Transfer in Real-Time with Adaptive Instance Normalization. In *IEEE International Conference on Computer Vision, ICCV 2017, Venice, Italy, October 22-29, 2017*. IEEE Computer Society, 1510–1519.

[15] Phillip Isola, Jun-Yan Zhu, Tinghui Zhou, and Alexei A. Efros. 2017. Image-to-Image Translation with Conditional Adversarial Networks. In *2017 IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2017, Honolulu, HI, USA, July 21-26, 2017*. IEEE Computer Society, 5967–5976.

[16] Yongcheng Jing, Yang Liu, Yezhou Yang, Zunlei Feng, Yizhou Yu, Dacheng Tao, and Mingli Song. 2018. Stroke Controllable Fast Style Transfer with Adaptive Receptive Fields. In *Computer Vision - ECCV 2018 - 15th European Conference, Munich, Germany, September 8-14, 2018, Proceedings, Part XIII (Lecture Notes in Computer Science)*, Vittorio Ferrari, Martial Hebert, Cristian Sminchisescu, and Yair Weiss (Eds.), Vol. 11217. Springer, 244–260.

[17] Yongcheng Jing, Yezhou Yang, Zunlei Feng, Jingwen Ye, and Mingli Song. 2017. Neural Style Transfer: A Review. *CoRR* abs/1705.04058 (2017). arXiv:1705.04058

[18] Garrett M Johnson and Mark D Fairchild. 2003. A top down description of S-CIELAB and CIEDE2000. *Color Research & Application: Endorsed by Inter-Society Color Council, The Colour Group (Great Britain), Canadian Society for Color, Color Science Association of Japan, Dutch Society for the Study of Color, The Swedish Colour Centre Foundation, Colour Society of Australia, Centre Français de la Couleur* 28, 6 (2003), 425–435.

[19] Justin Johnson, Alexandre Alahi, and Li Fei-Fei. 2016. Perceptual Losses for Real-Time Style Transfer and Super-Resolution. In *Computer Vision - ECCV 2016 - 14th European Conference, Amsterdam, The Netherlands, October 11-14, 2016, Proceedings, Part II (Lecture Notes in Computer Science)*, Bastian Leibe, Jiri Matas, Nicu Sebe, and Max Welling (Eds.), Vol. 9906. Springer, 694–711.

[20] Shigenobu Kobayashi. 1995. Art of color combinations. *Kosdansha International* (1995).

[21] Nicholas I. Kolkin, Jason Salavon, and Gregory Shakhnarovich. 2019. Style Transfer by Relaxed Optimal Transport and Self-Similarity. In *IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2019, Long Beach, CA, USA, June 16-20, 2019*. Computer Vision Foundation / IEEE, 10051–10060.

[22] Alex Krizhevsky, Geoffrey Hinton, et al. 2009. Learning multiple layers of features from tiny images. (2009).

[23] Chuan Li and Michael Wand. 2016. Precomputed Real-Time Texture Synthesis with Markovian Generative Adversarial Networks. In *Computer Vision - ECCV 2016 - 14th European Conference, Amsterdam, The Netherlands, October 11-14, 2016, Proceedings, Part III (Lecture Notes in Computer Science)*, Bastian Leibe, Jiri Matas, Nicu Sebe, and Max Welling (Eds.), Vol. 9907. Springer, 702–716.

[24] Xueting Li, Sifei Liu, Jan Kautz, and Ming-Hsuan Yang. 2018. Learning Linear Transformations for Fast Arbitrary Style Transfer. *CoRR* abs/1808.04537 (2018). arXiv:1808.04537

[25] Yijun Li, Chen Fang, Jimei Yang, Zhaowen Wang, Xin Lu, and Ming-Hsuan Yang. 2017. Universal Style Transfer via Feature Transforms. In *Advances in Neural Information Processing Systems 30: Annual Conference on Neural Information Processing Systems 2017, 4-9 December 2017, Long Beach, CA, USA*, Isabelle Guyon, Ulrike von Luxburg, Samy Bengio, Hanna M. Wallach, Rob Fergus, S. V. N. Vishwanathan, and Roman Garnett (Eds.). 386–396.

[26] Zhengqing Li, Zhengjun Zha, and Yang Cao. 2020. Deep Palette-Based Color Decomposition for Image Recoloring with Aesthetic Suggestion. In *MultiMedia Modeling - 26th International Conference, MMM 2020, Daejeon, South Korea, January 5-8, 2020, Proceedings, Part I (Lecture Notes in Computer Science)*, Yong Man Ro, Wen-Huang Cheng, Junmo Kim, Wei-Ta Chu, Peng Cui, Jung-Woo Choi, Min-Chun Hu, and Wesley De Neve (Eds.), Vol. 11961. Springer, 127–138.

[27] Jing Liao, Yuan Yao, Lu Yuan, Gang Hua, and Sing Bing Kang. 2017. Visual attribute transfer through deep image analogy. *ACM Trans. Graph.* 36, 4 (2017), 120:1–120:15.

[28] Tsung-Yi Lin, Michael Maire, Serge J. Belongie, James Hays, Pietro Perona, Deva Ramanan, Piotr Dollár, and C. Lawrence Zitnick. 2014. Microsoft COCO: Common Objects in Context. In *Computer Vision - ECCV 2014 - 13th European Conference, Zurich, Switzerland, September 6-12, 2014, Proceedings, Part V (Lecture Notes in Computer Science)*, David J. Fleet, Tomás Pajdla, Bernt Schiele, and Tinne Tuytelaars (Eds.), Vol. 8693. Springer, 740–755.

[29] Ilya Loshchilov and Frank Hutter. 2017. SGDR: Stochastic Gradient Descent with Warm Restarts. (2017).

[30] Ming Lu, Hao Zhao, Anbang Yao, Yurong Chen, Feng Xu, and Li Zhang. 2019. A Closed-Form Solution to Universal Style Transfer. In *2019 IEEE/CVF International Conference on Computer Vision, ICCV 2019, Seoul, Korea (South), October 27 - November 2, 2019*. IEEE, 5951–5960.

[31] Fujun Luan, Sylvain Paris, Eli Shechtman, and Kavita Bala. 2017. Deep Photo Style Transfer. In *2017 IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2017, Honolulu, HI, USA, July 21-26, 2017*. IEEE Computer Society, 6997–7005.

[32] Kiri Nichol. 2016. *Painter by numbers, wikiart.*

[33] Alec Radford, Luke Metz, and Soumith Chintala. 2015. Unsupervised representation learning with deep convolutional generative adversarial networks. *arXiv preprint arXiv:1511.06434* (2015).

[34] Erik Reinhard, Michael Ashikhmin, Bruce Gooch, and Peter Shirley. 2001. Color Transfer between Images. *IEEE Computer Graphics and Applications* 21, 5 (2001), 34–41.

[35] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. 2015. U-Net: Convolutional Networks for Biomedical Image Segmentation. In *Medical Image Computing and Computer-Assisted Intervention - MICCAI 2015 - 18th International Conference Munich, Germany, October 5 - 9, 2015, Proceedings, Part III (Lecture Notes in Computer Science)*, Nassir Navab, Joachim Hornegger, William M. Wells III, and Alejandro F. Frangi (Eds.), Vol. 9351. Springer, 234–241.

[36] Falong Shen, Shuicheng Yan, and Gang Zeng. 2018. Neural Style Transfer via Meta Networks. In *2018 IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2018, Salt Lake City, UT, USA, June 18-22, 2018*. IEEE Computer Society, 8061–8069.

[37] Guangyao Shen, Wenbing Huang, Chuang Gan, Mingkui Tan, Junzhou Huang, Wenwu Zhu, and Boqing Gong. 2019. Facial Image-to-Video Translation by a Hidden Affine Transformation. In *Proceedings of the 27th ACM International Conference on Multimedia*. 2505–2513.

[38] Guangyao Shen, Xin Wang, Xuguang Duan, Hongzhi Li, and Wenwu Zhu. 2020. MEmoR: A Dataset for Multimodal Emotion Reasoning in Videos. In *Proceedings of the 28th ACM International Conference on Multimedia.*

[39] Karen Simonyan and Andrew Zisserman. 2015. Very Deep Convolutional Networks for Large-Scale Image Recognition. (2015).

[40] Jianchao Tan, Jose I. Echevarria, and Yotam I. Gingold. 2018. Palette-based image decomposition, harmonization, and color transfer. *CoRR* abs/1804.01225 (2018). arXiv:1804.01225

[41] Dmitry Ulyanov, Vadim Lebedev, Andrea Vedaldi, and Victor S. Lempitsky. 2016. Texture Networks: Feed-forward Synthesis of Textures and Stylized Images. In *Proceedings of the 33nd International Conference on Machine Learning, ICML 2016, New York City, NY, USA, June 19-24, 2016 (JMLR Workshop and Conference Proceedings)*, Maria-Florina Balcan and Kilian Q. Weinberger (Eds.), Vol. 48. JMLR.org,

1349–1357.

[42] Xiaohui Wang, Jia Jia, Peiyun Hu, Sen Wu, Jie Tang, and Lianhong Cai. 2012. Understanding the emotional impact of images. In *Proceedings of the 20th ACM Multimedia Conference, MM '12, Nara, Japan, October 29 - November 02, 2012*, Noboru Babaguchi, Kiyoharu Aizawa, John R. Smith, Shin'ichi Satoh, Thomas Plagemann, Xian-Sheng Hua, and Rong Yan (Eds.). ACM, 1369–1370.

[43] Xiaohui Wang, Jia Jia, Hanyu Liao, and Lianhong Cai. 2012. Affective Image Colorization. *J. Comput. Sci. Technol.* 27, 6 (2012), 1119–1128.

[44] Xiaohui Wang, Jia Jia, Jie Tang, Boya Wu, Lianhong Cai, and Lexing Xie. 2015. Modeling Emotion Influence in Image Social Networks. *IEEE Trans. Affect. Comput.* 6, 3 (2015), 286–297.

[45] Xin Wang, Geoffrey Oxholm, Da Zhang, and Yuan-Fang Wang. 2017. Multimodal Transfer: A Hierarchical Deep Convolutional Neural Network for Fast Artistic Style Transfer. In *2017 IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2017, Honolulu, HI, USA, July 21-26, 2017*. IEEE Computer Society, 7178–7186.

[46] Yuan Yao, Jianqiang Ren, Xuansong Xie, Weidong Liu, Yong-Jin Liu, and Jun Wang. 2019. Attention-Aware Multi-Stroke Style Transfer. In *IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2019, Long Beach, CA, USA, June 16-20, 2019*. Computer Vision Foundation / IEEE, 1467–1475.

[47] Jaejun Yoo, Youngjung Uh, Sanghyuk Chun, Byeongkyu Kang, and Jung-Woo Ha. 2019. Photorealistic Style Transfer via Wavelet Transforms. In *2019 IEEE/CVF International Conference on Computer Vision, ICCV 2019, Seoul, Korea (South), October 27 - November 2, 2019*. IEEE, 9035–9044.

[48] Seungjoo Yoo, Hyojin Bahng, Sunghyo Chung, Junsoo Lee, Jaehyuk Chang, and Jaegul Choo. 2019. Coloring With Limited Data: Few-Shot Colorization via Memory Augmented Networks. In *IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2019, Long Beach, CA, USA, June 16-20, 2019*. Computer Vision Foundation / IEEE, 11283–11292.

[49] Richard Zhang, Phillip Isola, and Alexei A. Efros. 2016. Colorful Image Colorization. In *Computer Vision - ECCV 2016 - 14th European Conference, Amsterdam, The Netherlands, October 11-14, 2016, Proceedings, Part III (Lecture Notes in Computer Science)*, Bastian Leibe, Jiri Matas, Nicu Sebe, and Max Welling (Eds.), Vol. 9907. Springer, 649–666.

[50] Richard Zhang, Jun-Yan Zhu, Phillip Isola, Xinyang Geng, Angela S. Lin, Tianhe Yu, and Alexei A. Efros. 2017. Real-time user-guided image colorization with learned deep priors. *ACM Trans. Graph.* 36, 4 (2017), 119:1–119:11.