Anisotropic Stroke Control for Multiple Artists Style Transfer

Xuanhong Chen*, Xirui Yan*, Naiyuan Liu†, Ting Qiu and Bingbing Ni⊠
Shanghai Jiao Tong University, Shanghai, China
{chen19910528,xiruiYan,776398420,nibingbing}@sjtu.edu.cn,liunaiyuan27@gmail.com

ABSTRACT

Though significant progress has been made in artistic style transfer, semantic information is usually difficult to be preserved in a fine-grained locally consistent manner by most existing methods, especially when multiple artists styles are required to transfer within one single model. To circumvent this issue, we propose a *Stroke Control Multi-Artist Style Transfer* framework. On the one hand, we design an *Anisotropic Stroke Module (ASM)* which realizes the dynamic adjustment of style-stroke between the non-trivial and the trivial regions. *ASM* endows the network with the ability of adaptive semantic-consistency among various styles. On the other hand, we present an novel *Multi-Scale Projection Discriminator* to realize the texture-level conditional generation. In contrast to the single-scale conditional discriminator, our discriminator is able to capture multi-scale texture clue to effectively distinguish a wide range of artistic styles. Extensive experimental results well demonstrate the feasibility and effectiveness of our approach. Our framework can transform a photograph into different artistic style oil painting via only ONE single model. Furthermore, the results are with distinctive artistic style and retain the anisotropic semantic information. The code is already available on github: https://github.com/neuralchen/ASMAGAN.

CCS CONCEPTS

• **Computing methodologies → Computational photography**; *Computer vision representations.*

KEYWORDS

style transfer, non-real image rendering, image translation

ACM Reference Format:
Xuanhong Chen*, Xirui Yan*, Naiyuan Liu†, Ting Qiu and Bingbing Ni⊠. 2020. Anisotropic Stroke Control for Multiple Artists 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, 33 pages. https://doi.org/10.1145/3394171.3413770

*Equal contribution.
†Contributed to the work while he was a research assistant at Shanghai Jiao Tong University.
⊠Corresponding author: Bingbing Ni.

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

ACM ISBN 978-1-4503-7988-5/20/10…$15.00
https://doi.org/10.1145/3394171.3413770

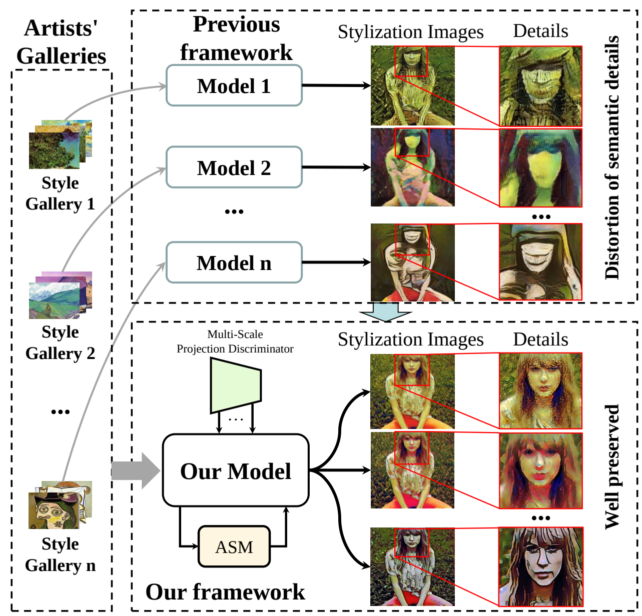


Figure 1: Previous method [32] needs separate models for each target style from artists' portfolios. Meanwhile, the semantic details are unrecognizable with excessive distortion. Our model is able to perform multi-artist style transfer within one model, while well preserving semantic information through anisotropic style-stroke controlling.

1 INTRODUCTION

Style transfer is a practical technique that transfers a natural photograph into an artistic painting. Recently, Convolutional Neural Network (CNN) based style transfer approaches [10, 32, 40] make significant progress in imitating style texture and tone. However, a convincing stylization is not just about the imitation of textures, it also needs to choose suitable stroke size according to different semantic regions, *e.g.,* face, background. Painters never use the same stroke size on the entire painting, where they use thicker strokes in trivial regions(e.g., sky, water surface), and use finer strokes in non-trivial regions(e.g., face, boats) to show more details. Therefore, anisotropic stroke control according to the content is a vital idea to make the style transfer closer to real painting. In addition, from a practical point of view, the style transfer model should have the ability of achieving multiple artistic stylization just via ONE single model.

Strokes refer to the fundamental element of the artistic paintings, artists tend to use different stroke size in various part of a painting. For example, to make the face more vivid, painters use fine brush strokes to outline facial details, while using thicker brush strokes to draw the background. Approaches such as [32] devote to learning style-stroke control in the style transfer. However, those frameworks can only adjust the overall brushstroke of the painting without distinction. Inspired by human perception, recent methods such as [25, 40, 42] incorporate attention mechanism to achieve the different generation granularity in non-trivial regions and trivial regions. Yao et al. [40] follow this scheme and introduce self-attention module which generates a salient map in hidden space, and then adjusts the style-stroke of different regions according to that map. However, this method suffers from the uncertainty of focus areas and poor effect of highlighting the pixel-wise salient part of semantic content.

Additionally, most contemporary style transfer methods [1, 7, 10, 14, 33, 35] focus on example guided stylization, which transfers the style characteristics of the example image onto a target content image. In this way, those approaches can only achieve the imitation of color and texture of a single painting rather than learning the overall artist style of an artist. Such learning strategy is completely different from human-artistic creation habits. The common way for humans to learn to paint the style of an artist is to delve into a set of works of the artist instead of a single piece of artwork. By analogy, the abstract yet comprehensive style-knowledge should be flexibly modeled from a quantity of artist paintings. Sanakoyeu et al. [32] train their model by a set of certain artist artworks and indeed achieve a substantial improvement in visual quality. However, this method faces strict limitations, only one single artistic stylization can be performed within one model. Such defects make this method have serious difficulties in deployment, *e.g.*, one thousand styles need a thousand models to deploy, which is unrealistic.

To address the two problems mentioned above, we propose a framework called *Anisotropic Stroke Multiple Artists GAN, ASMA-GAN*. To solve the anisotropic control of style-stroke problem, we present the *Anisotropic Stroke Module, ASM*. Worth mentioning, the size of the receptive field determines the size of the style-stroke [16]. In contrast to [40], our ASM does not explicitly generate the salient map to mark different style-stroke size regions, because this map is troublesome to learn without explicit supervised signals. In order to dynamically adjust the style-stroke, our AMS integrates features from different scales of receptive fields (equivalent to different strokes) according to the control signal. In detail, the control signal comes from the deepest hidden features, and the features have the largest receptive field, which will yield the thickest style-stroke. At the same time, these features have the most abundant semantic information, and this information can guide AMS to distinguish the non-trivial region from trivial region. Therefore, with the help of ASM, our framework is able to carry out dynamic and variable-grain style transfer. Unlike existing multi-domain translation issues [17], multi-artist style transfer requires the framework to have the ability to discriminate multi-scale texture textones. In view of this, a *Multi-Scale Projection Discriminator* is proposed to utilize multi-scale characteristic of style to integrally extract style information. In fact, many paintings contain plenty of micro-structures, which will be lost as the network deepens. Instead of single scale classification, our discriminator judges the authenticity of the painting by drawing the features of different receptive fields and using them comprehensively. Furthermore, our discriminator abandons the auxiliary classifier adopted by ACGAN, and uses the projection way [27] to embed the conditional information into the final output likelihood. With such a design, our discriminator can effectively encourage the generator to synthesize highly realistic stylized results. Extensive experimental results well demonstrate the effectiveness and high visual quality achieved by our framework.

2 RELATED WORK

Semantic focus. Motivated by the importance of attention in human perception, a lot of research efforts have been devoted to the semantic objects within an image [5, 6, 12, 36]. Methods in the field of image translation can be categorized as two-step and one-step. Two-step methods [3, 39] split models into two separated phases: 1) acquire semantic mask from a separated segmentation network; 2) process the semantic focal area using the main model. To augment the ability of adapting to variations of semantic context, Luan et al. [39] use DilatedNet [4] to generate image segmentation masks of the inputs and reference images for a set of common labels. Two-step methods require an extra pre-training network and millions of labels on semantic context. Therefore, they are time-consuming and also dilate the structures. By contrast, one-step models [24, 25] achieve semantic focus by incorporating the attention mechanism within the intact model. Ma et al. [24] decouple local textures from holistic shapes by attending to local objects of interest through square image regions, while it results in alteration of background during image translation. Mejjati et al. [25] explore an attention network to circumvent the problem.

Image Translation. Style transfer is a subfield of image translation where the goal is to learn the mapping between style and content images. The key issues of style translation are the presentation of style and the synthesis of image. Since the success of NST proposed by Gatys et al. [10], neural representation of image is widely applied in texture synthesis. To speed up the style transfer process, several algorithms have been proposed [1, 33, 35], which produce stylized results with a forward pass. To improve the flexibility, models incorporating multiple and arbitrary styles are proposed [7, 11, 14, 22, 38]. These works synthesize style texture by the representation of style captured from certain artwork rather than the style domain. Many works achieve domains mapping using generative adversarial networks (GANs) by unpaired images [18, 41, 46]. Sannakoyeu et al. [32] utilize related style images to train an adversarial discriminator and optimize the generator with content perceptual loss. AC-GAN [29] provides class information to generator and modifies the learning target of GANs by an auxiliary classifier. Instead of naively concatenating class information to the input, Projection Discriminator [28] proposes a specific form of the discriminator, motivated by a commonly occurring family of probabilistic models. However, it only utilizes the feature of the last layer, which would lose style information when transferring. Our method composes multi-scale style information.

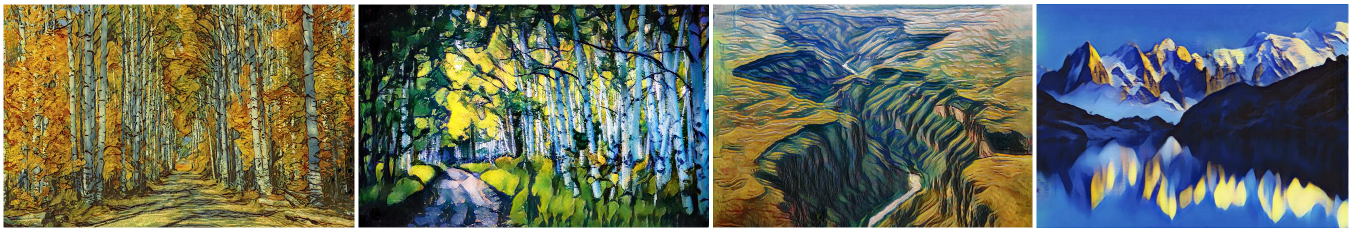


Figure 2: Some paintings generated by our ASMA-GAN. From left to right, the corresponding artists are Van Gogh, Samuel, Munch, Nicholas. These results are very similar to oil painting. More high resolution results can be found in suppl.

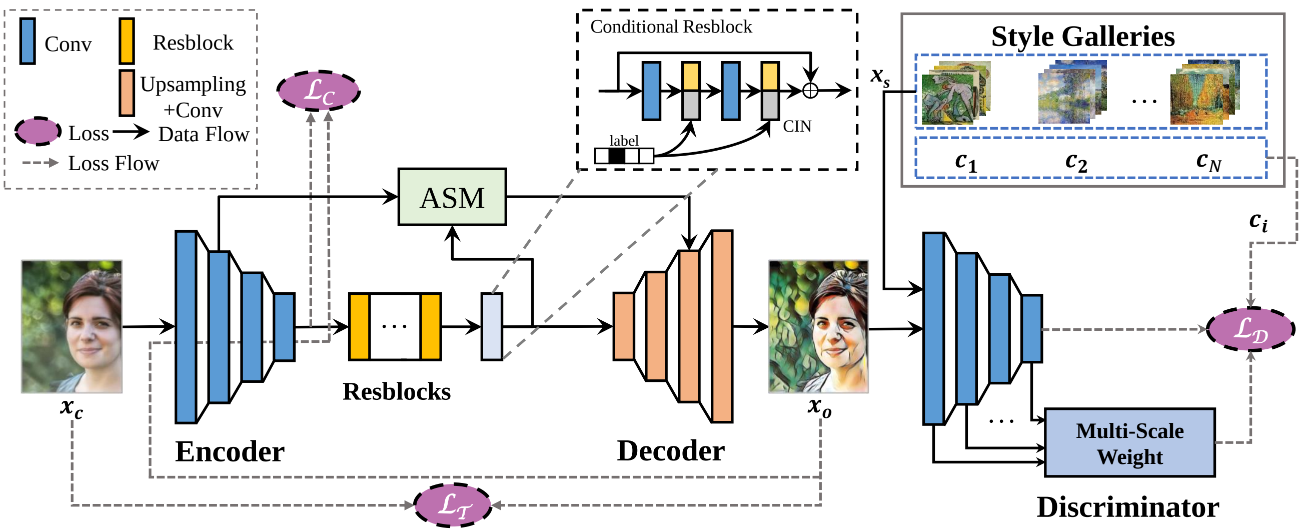


Figure 3: An overview of our framework. We train the ASM and multi-artist style transfer uniformly. ASM can dynamically adjust the stroke size of the corresponding region according to the semantic information. Our Multi-Scale Projection Discriminator can learn the multi-scale characteristics of oil painting to better guide the training of Generator. Each artist's style is trained by portfolio of certain artist instead of single painting.

3 METHODOLOGY

Our framework learns from artists' portfolios, instead of one single painting, for novel art creation, *i.e.*, multi-artist stylization with flexible style-stroke size. To this end, we propose a *ASMA-GAN* framework consisting of the following components: 1) a Conditional Generator $\mathcal{G}$ which efficiently leverages multi-artist style labels to synthesis corresponding stylized painting x_o; 2) a novel module called *Anisotropic Stroke Module* which endows the generator the capability to adjust style-stroke [16] size in different region according to the semantic information; and 3) a *Multi-Scale Projection Discriminator* $\mathcal{D}$, which encourages style consistency through the task of distinguishing artworks of different artists. Figure. 3 illustrates the full pipeline of our approach.

3.1 The Conditional Generator

From the perspective of art creation, style should be learned from artists' portfolios instead of a single painting [32]. Our model benefits from this conception: it is trained by artists' portfolios. Suppose $\mathbb{I}_i \in \{\mathbb{I}_{Monet}, \mathbb{I}_{Picasso}, ...\}$ denotes an artist's portfolio, $x_s \in \mathbb{I}_i$ denotes an artwork of portfolio $\mathbb{I}_i$. Given an input content image x_c and a target label l, the task is to generate a stylized result x_o using our Conditional Generator. Instead of unskillfully imitating a single painting, we manage to make use of more general characteristics of a certain artist. The Conditional Generator consists of four parts: the Encoder, the Resblocks, the conditional Resblock and the Decoder.

Style Information Injection. To generate multi-artist stylized images within a single model, efficient injection of style label information is necessary and crucial. Previous multi-domain translation method [8] directly concatenates one-hot label map with the input image or the feature map. This approach is only suitable for tasks with similar domains. Since there are a significant discrepancy between the content domain and the style domain in style transfer task. Therefore, it is invalid to inject the style information into the network through the direct concatenation. Another serious flaw is that directly concatenated one-hot vector label is invalid after reflection-padding. Validity of conditional input depends on

Figure 4: Different sizes of style-stroke regard to different down-sample rates. We put resblocks in places of different down-sample rates between the Encode and the Decoder to show the relationship between down-sample rate and stylization extent. As shown above, granularity of stylization increases as down-sample rate grows.

whether it would change data distribution in feature space [44]. Based on that idea, we design a conditional Resblock that uses Conditional Instance Normalization (CIN) [9] as the style information injection means. The structure of the conditional Resblock is shown in Fig. 3.

3.2 The Anisotropic Stroke Module

When watching an artwork, people are more sensitive to semantic content, such as people, faces and expect them to preserve details with less distortion. However, coarse granularity of stylization results in detail distortion. For example, fine granularity of stylization distorts subtle objects like edges, while coarse one causes distortion of large scale objects, like cars and human face. Actually, the granularity is closely related to receptive field of the network, and the larger the receptive field is, the coarser the granularity will be. In Fig. 4, it is shown that different down-sample (equivalent to receptive field size) results in stylization with varying granularities. Since this phenomenon extremely resembles drawing painting with different sizes of painting brushstrokes, it is named style-stroke [16]. We present our framework to employ the *Anisotropic Stroke Module* to dynamically adjust the style-stroke size according to the semantic information in various region. Dynamic style-stroke yield pleasing stylized results with meticulous strokes in rich semantics region and rough strokes in remaining region. The dynamic style-stroke make the stylized results maintain the legibility of the important content (*e.g.,* face, building parts and so on) in the photograph without being severely distorted and losing the meaning of the original picture.

The detailed structure of ASM is shown in Fig. 5. We re-design the reset and update gates with a spatial-wised attention mechanism [37] to be light weighted and still effective in information incorporation. Instead of removing the gates we choose to lighten them [30] because experiments show removing either of them

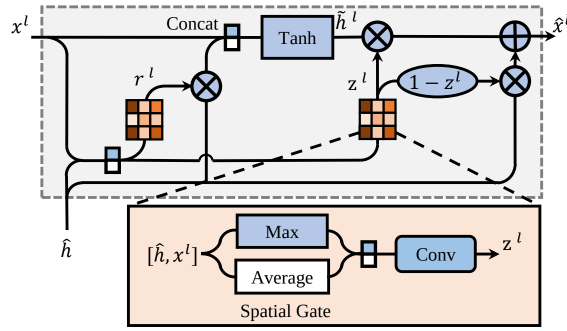


Figure 5: Structure of ASM. Our ASM is a variant of GRU with much lighterweight. The box below is our spatial gate structure. ASM can fuse features of two different scales to achieve dynamic adjustment of style-stroke.

would lead to a sharp decline in model performance. Without loss of generality, we choose the ASM located in the l-th layer as an analysis example. The l-th layer feature coming from the encoder side denotes as $\boldsymbol{x}^l$. $\boldsymbol{h}$ denotes the feature drawn from the last layer of the bottleneck. It contains rich semantic information to help ASM distinguish important and unimportant regions. $\boldsymbol{h}$ is firstly concatenated with style class information $\boldsymbol{c}$ to obtain up-sampled hidden state $\hat{\boldsymbol{h}}$. Then $\hat{\boldsymbol{h}}$ and $\boldsymbol{x}^l$ are combined to calculate the masks $\boldsymbol{r}^l, \boldsymbol{z}^l$ for the reset gate and update gate. W_T, W_r,W and W_z represent parameter matrix of transposed convolution, reset gate, merging and update gate. The further process is similar to GRU. $\hat{\boldsymbol{x}}^l$ combines two different receptive field features, in other words it blends the style-stroke of two scales. The equation of gates is shown below,

$$
\begin{aligned}
\hat{\boldsymbol{h}} &= \mathrm{W}_T * [\boldsymbol{h}, \boldsymbol{c}], \\
\boldsymbol{r}^l &= \mathrm{W}_r * \left[\hat{\boldsymbol{h}}, \boldsymbol{x}^l\right], \\
\tilde{\boldsymbol{h}}^l &= tanh\left(\mathrm{W} * \left[\boldsymbol{r}^l \circ \hat{\boldsymbol{h}}, \boldsymbol{x}^l\right]\right), \\
\boldsymbol{z}^l &= \mathrm{W}_z * \left[\hat{\boldsymbol{h}}, \boldsymbol{x}^l\right], \\
\hat{\boldsymbol{x}}^l &= \boldsymbol{z}^l \circ \tilde{\boldsymbol{h}}^l + \left(1 - \boldsymbol{z}^l\right) \circ \hat{\boldsymbol{h}}.
\end{aligned}
\tag{1}
$$

3.3 The Multi-Scale Projection Discriminator

The discriminator is the most important component of GANs, and it is trained in the game between the generator and itself. The Discriminator acts as an oil painting connoisseur in our framework, and its performance directly determines the visual quality of style transfer results. The existing translation frameworks achieve multi-domain discriminator in the following two ways: 1) Adding an auxiliary classifier similar to AC-GAN [29]; 2) Using multiple discriminators [43]. In the first method, the auxiliary classifier works well at low domain variance, but it is difficult to show good performance when the variance is high. In the second method, GANs are known for its notoriously difficult training, and multiple discriminators make training more unstable. Takeru Miyato et al. [28] propose a new conditional generation method for multi-class images synthesis. This method masterly projects the class label information

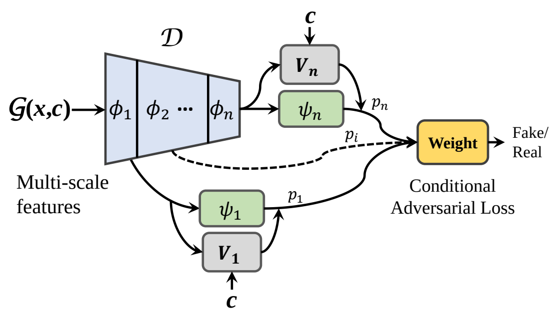


Figure 6: Unlike the Projection Discriminator [27], our model comprehensively uses the features of different scales, which greatly strengthens the discriminator to recognize the stroke textons of different scales in the painting.

into the likelihood and shows the state-of-the-art performance in the Imagenet [31]. They decompose the adversarial likelihood to the sum of two components:

$$\mathcal{D}(\boldsymbol{x_o}, \boldsymbol{c}) = \boldsymbol{c}^T \boldsymbol{V} \phi(\boldsymbol{x_o}) + \psi(\phi(\boldsymbol{x_o})), \tag{2}$$

where $\boldsymbol{V}$ denotes an embedding matrix. $\phi(\cdot)$ represents the input to the last layer of the convolution network part of $\mathcal{D}$. $\psi(\cdot)$ is the FC layers to scale the output.

Inspired by PatchGAN [15], we design a novel discriminator, called *Multi-Scale Projection Discriminator*, for extracting the multi-scale characteristics of oil paintings while achieving the multi-artist style transfer. Worth mentioning, the multi-scale character of artist style is exceedingly commonplace in oil paintings. For example, Van Gogh's paintings, his paintings have unique overall color and texture features, and are characterized by neatly arranged short strokes. The most obvious difference between our Multi-Scale Projection Discriminator and [15] is that we extend the discriminator into a multi-scale one that can fuse multiple scale features. It can capture the style characteristics of an artist's painting set from strokes to color schemes. The structure of our discriminator is shown in Fig. 6. Additionally, as the training process of GANs is extremely unstable, we apply the Spectral Normalization (SN) [26] in the Multi-Scale Projection Discriminator, which is able to force the weights in discriminator to regularize the Lipschitz constant yielding a stable training process. The mathematical expression of our Multi-Scale Projection Discriminator is shown below:

$$\mathcal{D}(\boldsymbol{x_o}, \boldsymbol{c}) = \sum_{i=0}^{N} w_i \cdot \left[\boldsymbol{c}^T V_i \phi_i(\boldsymbol{x_o}) + \psi_i(\phi_i(\boldsymbol{x_o})) \right], \tag{3}$$

where N indicates the total number of feature scales of the Multi-Scale Projection Discriminator, w_i denotes weighting factor of i-th output likelihood.

3.4 Objective Function

Perceptual Loss. Generator should try to ensure the semantic consistency of content while stylization. Most of the existing translation networks [1, 7, 16, 34, 40] use the pre-trained VGG model on Imagenet as the calculation function for perceptual loss [1]. However, when painting, the artist thinks about the content of the painting from an artistic point of view rather than the classification. Under the constraints of such perceptual loss, the generator can not realize the artistic reconstruction of the content details. Inspired by the style-aware content loss [32], we measure the similarity in content between input image $\boldsymbol{x_c}$ and stylized $\boldsymbol{x_o} = \mathcal{G}(\boldsymbol{x_c}, \boldsymbol{c})$ by a style-aware content loss. The loss directly uses the Encoder of the Generator instead of the pre-trained VGG model as the calculation function for perceptual loss, which makes the Encoder tend to retain the semantic region related to style. Therefore, our generator achieves better style transfer performance, but the content consistency is drastically reduced. This problem is exactly what our *ASM* had solved. Transform loss [32] $\mathcal{L}_T$ is introduced, as the extra signal, which initializes training and boosts the learning of the primary latent space:

$$\mathcal{L}_C = \mathbb{E}_{\boldsymbol{x_c}} \left[\|\mathcal{E}(\boldsymbol{x_c}) - \mathcal{E}(\mathcal{G}(\boldsymbol{x_c}, \boldsymbol{c}))\|_{\ell_1} \right], \tag{4}$$

$$\mathcal{L}_{\mathcal{T}} = \mathbb{E}_{\boldsymbol{x_c}} \left[\frac{1}{CHW} \|\mathcal{T}(\boldsymbol{x_c}) - \mathcal{T}(\mathcal{G}(\boldsymbol{x_c}, \boldsymbol{c}))\|_{\ell_1} \right], \tag{5}$$

where $\mathcal{E}$ is the Encoder of $\mathcal{G}$, $\mathcal{T}$ denotes a pooling layer, and C,H,W respectively represent channels, height, width of $\mathcal{T}(\cdot)$. $\|\cdot\|_{\ell_1}$ denotes $\mathcal{L}_1$ loss. Experiments show that compared to perceptual loss, training with sytle-aware loss can achieve better saturation in the stylized image.

Adversarial Loss. At the beginning of the training process, the stylization results are almost the same as the photographs. In order to speed up the discriminator to learn to distinguish between paintings and photographs, we also add photographs as the fake samples to the training of discriminator. Furthermore, we introduce the hinge loss [23] instead of WGAN loss [2] as the standard adversarial loss. Our standard adversarial loss is shown below:

$$\begin{aligned} \mathcal{L}_{\mathcal{D}} = \mathbb{E}_{\boldsymbol{c}} & \left[\mathbb{E}_{\boldsymbol{x_s}} \left[max(0, 1 - \mathcal{D}(\boldsymbol{x_s}, \boldsymbol{c})) \right] \right] \\ & + \mathbb{E}_{\boldsymbol{c}} \left[\mathbb{E}_{\boldsymbol{x_c}} \left[max(0, 1 + \mathcal{D}(\boldsymbol{x_c}, \boldsymbol{c})) \right] \right] \\ & \quad + \mathbb{E}_{\boldsymbol{c}} \left[\mathbb{E}_{\boldsymbol{x_c}} \left[max(0, 1 + \mathcal{D}(\hat{\mathcal{G}}(\boldsymbol{x_c}, \boldsymbol{c}), \boldsymbol{c})) \right] \right], \end{aligned} \tag{6}$$

$$\mathcal{L}_{\mathcal{G}} = -\mathbb{E}_{\boldsymbol{c}} \left[\mathbb{E}_{\boldsymbol{x_c}} \left[\hat{\mathcal{D}}(\mathcal{G}(\boldsymbol{x_c}, \boldsymbol{c}), \boldsymbol{c}) \right] \right], \tag{7}$$

where $\hat{\mathcal{G}}$ and $\hat{\mathcal{D}}$ indicates that the corresponding model parameters are fixed and no training.

Overall Loss. To summarize, the full objective of our model is:

$$\mathcal{L}(\mathcal{A}, \mathcal{G}, \mathcal{D}) = \mathcal{L}_{\mathcal{G}} + \mathcal{L}_{\mathcal{D}} + \lambda_C \mathcal{L}_C + \lambda_{\mathcal{T}} \mathcal{L}_{\mathcal{T}}, \tag{8}$$

where $\mathcal{A}$ indicates the ASM. The weight coefficients: λ_C, $\lambda_{\mathcal{T}}$ are mainly to balance the magnitude of different loss. We set $\lambda_C = 90$, $\lambda_{\mathcal{T}} = 100$ respectively.

4 EXPERIMENTS

4.1 Implementation details

Structure details. As mentioned above, the framework consists of the Conditional Generator, the Anisotropic Stroke Module and the Multi-Scale Projection Discriminator. The Conditional Generator contains three blocks: the Encoder, the Resblocks and the Decoder. The Encoder has 1 conv3-stride-1 and 4 conv3-stride-2, where each convolution layer is followed by an IN [34] and a LeakyRelu. 5 residual layers [13] and a conditional Resblock are connected in series

Figure 7: Results produced for different artist style from different style transfer methods, with a fixed content image. We evaluate the zoomed in cut-out of the same region in all results to compare the variation across different styles and details stylization effect. More high resolution results can be found in suppl.

to form the Resblocks. The Decoder is composed of 4 upsampling-conv3 layers and 1 conv3-stride-1 colorization layer. The backbone network of the Multi-Scale Projection Discriminator is a fully convolutional network with 6 conv5-stride2-SN-LeakyRelu blocks. More details can be found in suppl.

Training Data. The training data consists of two parts: the content images are sampled from Places365 [45] and the artistic style portfolios are collected from the Wiki Art dataset. Optimizer is Adam optimizer [19] and learning rate is set as 0.0001.

4.2 Stylization Assessment

In order to assess the quality of the stylization results of our framework, we propose two Quantitative metrics: *Semantic Retention Ratio (SRR)* and *Stylization Accuracy*. Actually, style is a relatively

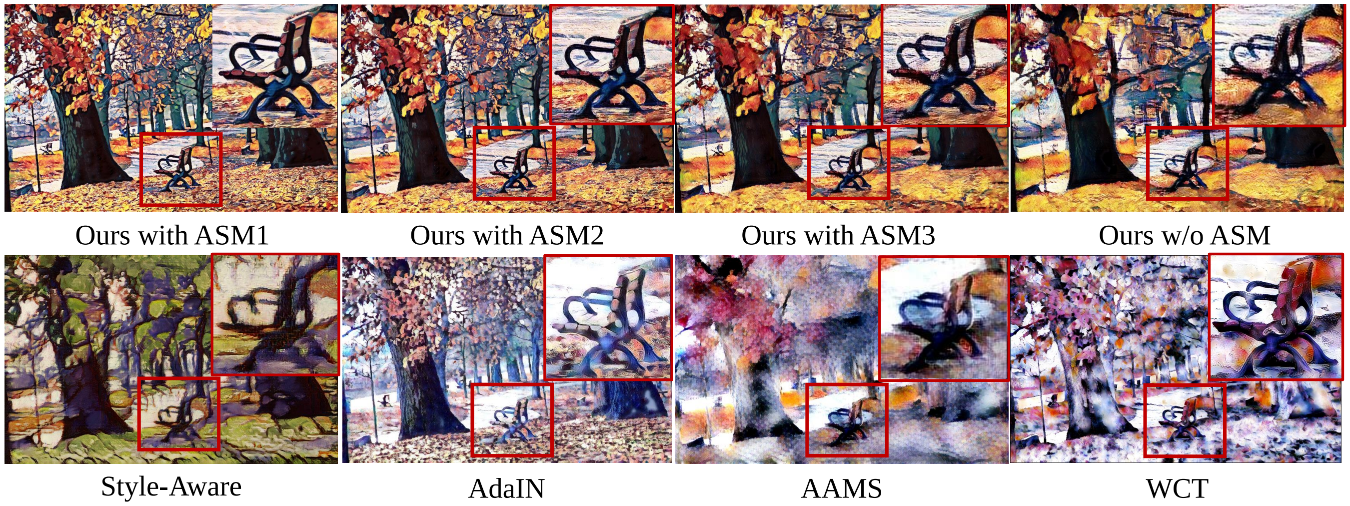


Figure 8: Comparison of anisotropic semantic preserving effect from Style-Aware, AAMS, AdaIN, WCT and ours. ASM1, ASM2, ASM3 indicate that ASM is placed in different layers of Generator.

Table 1: User studies scores (*mean*) of different methods, in terms of the style transfer effect and anisotropic semantic preserving effect. *Preference Score* is the final score, it is the average of the two scores.

Method	*Preference Score*	*Style Deception Score*	*Semantic Retention Score*
Ours	8.00	7.9	8.1
Style-Aware [32]	7.00	8.5	5.5
Style Swap [7]	5.15	2.7	7.6
WCT [22]	5.95	6.1	5.8
AdaIN [14]	6.15	6.4	5.9
AAMS [40]	6.25	4.6	7.9

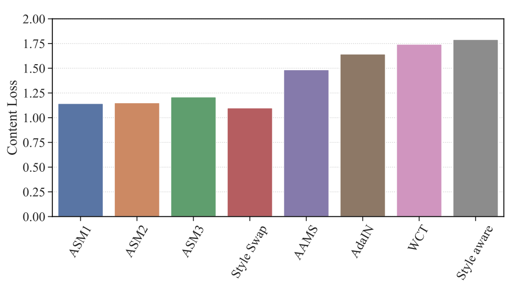


Figure 9: Semantic Retention Ratio histogram of different methods. The higher the score, the worse the semantic retention of the corresponding method.

abstract concept, it is difficult to use quantitative metrics for comprehensive measurement. Based on this fact, we introduced two types of user studies, *Style Deception Score, Semantic Retention Score*, with reference to [20, 21, 32] to perceptually evaluate the effectiveness of our algorithm.

Semantic Retention Ratio. This metric is designed to measure the degree of retention of the input image semantics after stylization. In fact, the main task of our ASM is to retain discriminative semantic information. Therefore, SRR can accurately and quantitatively evaluate the effectiveness of ASM. The mathematical form of our SRR is as follows:

$$SSR = \mathbb{E}_{\Delta x}\left[\frac{1}{n^2}\sum_{i,j}\left|\frac{D_{ij}^{\boldsymbol{x}_o}}{\sum_i D_{ij}^{\boldsymbol{x}_o}} - \frac{D_{ij}^{\boldsymbol{x}_c}}{\sum_i D_{ij}^{\boldsymbol{x}_c}}\right|\right], \tag{9}$$

where $D_{ij}^{\cdot}$ denotes the patch of corresponding image, n is the total number of patches.

Stylization Accuracy. We designed the Stylization Accuracy to evaluate the performance of our Multi artist style transfer. This metric is measured by an artistic style classifier that is isomorphic to our discriminator. To avoid overfitting, we collected 11 artists' paintings to generate more than 10^5 style patches to train our style classifier. Finally, the classifier achieves a 91.9% average accuracy on the paintings dataset.

4.3 Qualitative analysis

Style transfer results. We evaluate our approach with five state-of-the-art methods: AdaIN [14], Style-Aware [32], AAMS [40], Style Swap [7] and WCT [22]. Noting that style aware includes two subsequent works [20, 21] and their effects are not much different. For simplicity, we use style aware to represent this type of method. We pick the most representative paintings from training style portfolios as the style images to represent the portfolio. By comparing the zoomed in cut-outs in Fig. 7, our method shows much more stunning effect than other competitors. The details of the semantic contents (e.g the clock on the clock tower) are accurately preserved and the characteristic of target styles are vividly maintained. Although results of Style-Aware have the most prominent style

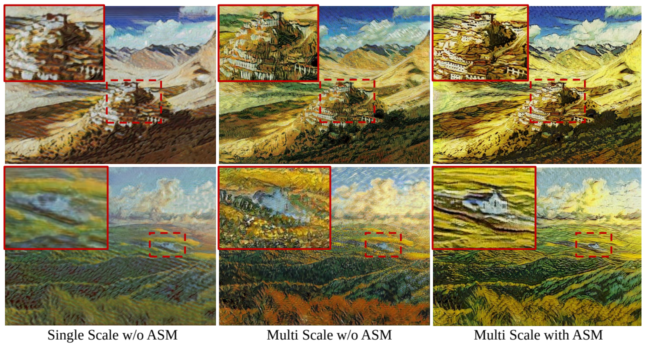


Figure 10: Comparison of the Multi-Scale Projection Discriminator and the Single-Scale Projection Discriminator.

characteristics, but details in the results are unrecognizable with excessive distortion, caused by the coarse granularity. It can be seen that the example based methods (WCT, AdaIN, Style Swap, AAMS) cannot effectively learn the characteristics of style. Their results only show strong tone changes and irregular details distortion. And these changes strongly depend on the example image.

4.4 Ablation Studies

The indispensable two components of our framework are the ASM and the Multi-Scale Projection Discriminator. We study the effectiveness of these two modules by individually removing them. First, by removing the ASM, our model degenerates into a multi-artist style transfer framework. The degraded model is trained using the same conditions as described in implementation details section, and the transfer results are shown in Fig. 8. Compared to other methods, our approach retains the most complete details. In addition, we place ASM on different layers in Generator, and use ASM1, ASM2, and ASM3 to represent $1/2$ downsampling, $1/4$ downsampling, and $1/8$ downsampling positions, respectively. It can be seen from the figure that the details of the bench gradually become blurred, which indicates that ASM can automatically integrate features from different scales so that our model can automatically use different stylized strokes in different areas.

Then, we remove the multi-scale style learning module from the Multi-Scale Projection Discriminator. Using this Single-Scale Projection Discriminator, we train the model on the portfolios mentioned above. The results are shown in Fig. 10. In the figure, the first column is the stylization result of the Single-Scale Discriminator. Although the results contain some characteristics of the corresponding style, while its details are over-smooth. It is because the discriminator lacks the use of shallow features, thus losing the ability to judge details. The results confirm that multi-scale style learning module is an indispensable part of our framework. The above two ablation experiments demonstrate the necessity and effectiveness of the two modules.

4.5 Quantitative analysis

User study. We use 200 groups of images, each consists of the input content image, the target style set and 5 results from [7, 14, 32, 40] and ours. The user studies include two parts in terms of the style transfer effect and anisotropic semantic preserving effect. In the first study, the participants are asked to score the results by the degree of style reduction, from 0 to 10 (10 is the best, 0 is the worst), i.e. *Style Deception Score*. In the second study, the participants score the results by the degree of detail retention of the semantic content, i.e. *Semantic Retention Score*. We calculate the mean value of two scores of each method over all participants, as shown in Tab. 1. The studies show that our multiple artists stylized results achieve approximate equivalent effect as the stylized results in [32], which is better than other methods. And our method outperforms others on the semantic detail retention.

Content Discrepancy. We carefully picked 200 pictures with abundant semantic information (*e.g.*, portraits, buildings, etc.) from the Place365 to form the benchmark. We estimate SSR based on this benchmark. It can be seen from Fig. 9 that style-swap has the highest score, but its stylization effect is too poor. Our method can achieve good semantic retention no matter where ASM is placed.

Style Accuracy. We generate 200 result images for each artist's style, and measure the Style Accuracy of the stylization by sending these result images to the style classifier. The higher the accuracy of the classification result, the closer the class is to the corresponding painting style. However, style is a perceptual concept, so this indicator can only be used as a reference. The classification results are shown in Fig. 11.

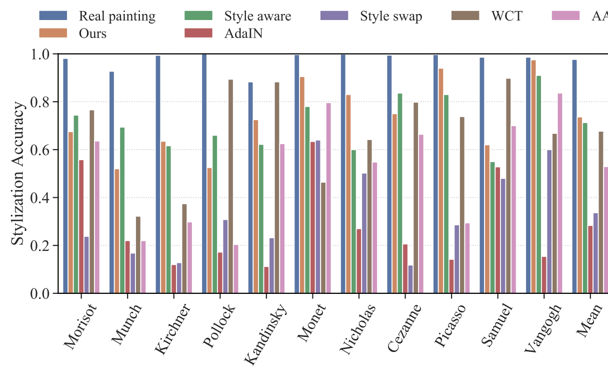


Figure 11: Our method achieves the best average classification accuracy.

5 CONCLUSION

In this paper, we propose a novel Multi-Scale Projection Discriminator, which overcomes the limitation of Single-Scale Projection Discriminator and gives our discriminator ability to deal with the multi-scale characteristics of style. Moreover, the ASM is able to dynamically adjust the strokes based on the semantic information of the picture. With the help of ASM, our model can retain the vital semantic information of the picture while transferring the style. Experimental results demonstrate the effectiveness and delicate visual performance of our method.

ACKNOWLEDGMENTS

This work was supported by National Science Foundation of China (61976137, U1611461, U19B2035) and STCSM(18DZ1112300).

REFERENCES

[1] Justin Johnson and· 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.* 694–711. https://doi.org/10.1007/978-3-319-46475-6_43

[2] Martín Arjovsky, Soumith Chintala, and Léon Bottou. 2017. Wasserstein GAN. *CoRR* abs/1701.07875 (2017). arXiv:1701.07875 http://arxiv.org/abs/1701.07875

[3] Alex J. Champandard. 2016. Semantic Style Transfer and Turning Two-Bit Doodles into Fine Artworks. *CoRR* abs/1603.01768 (2016). arXiv:1603.01768 http://arxiv.org/abs/1603.01768

[4] Liang-Chieh Chen, George Papandreou, Iasonas Kokkinos, Kevin Murphy, and Alan L. Yuille. 2018. DeepLab: Semantic Image Segmentation with Deep Convolutional Nets, Atrous Convolution, and Fully Connected CRFs. *IEEE Trans. Pattern Anal. Mach. Intell.* 40, 4 (2018), 834–848. https://doi.org/10.1109/TPAMI.2017.2699184

[5] Liang-Chieh Chen, Yi Yang, Jiang Wang, Wei Xu, and Alan L. Yuille. 2016. Attention to Scale: Scale-Aware Semantic Image Segmentation. In *2016 IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2016, Las Vegas, NV, USA, June 27-30, 2016.* 3640–3649. https://doi.org/10.1109/CVPR.2016.396

[6] Liang-Chieh Chen, Yukun Zhu, George Papandreou, Florian Schroff, and Hartwig Adam. 2018. Encoder-Decoder with Atrous Separable Convolution for Semantic Image Segmentation. In *Computer Vision - ECCV 2018 - 15th European Conference, Munich, Germany, September 8-14, 2018, Proceedings, Part VII.* 833–851. https://doi.org/10.1007/978-3-030-01234-2_49

[7] Tian Qi Chen and Mark Schmidt. 2016. Fast Patch-based Style Transfer of Arbitrary Style. *CoRR* abs/1612.04337 (2016). arXiv:1612.04337 http://arxiv.org/abs/1612.04337

[8] Yunjey Choi, Min-Je Choi, Munyoung Kim, Jung-Woo Ha, Sunghun Kim, and Jaegul Choo. 2017. StarGAN: Unified Generative Adversarial Networks for Multi-Domain Image-to-Image Translation. *CoRR* abs/1711.09020 (2017). arXiv:1711.09020 http://arxiv.org/abs/1711.09020

[9] Vincent Dumoulin, Jonathon Shlens, and Manjunath Kudlur. 2016. A Learned Representation For Artistic Style. *CoRR* abs/1610.07629 (2016). arXiv:1610.07629 http://arxiv.org/abs/1610.07629

[10] Leon A. Gatys, Alexander S. Ecker, and Matthias Bethge. 2015. A Neural Algorithm of Artistic Style. *CoRR* abs/1508.06576 (2015). arXiv:1508.06576 http://arxiv.org/abs/1508.06576

[11] Shuyang Gu, Congliang Chen, Jing Liao, and Lu Yuan. 2018. Arbitrary Style Transfer with Deep Feature Reshuffle. *CoRR* abs/1805.04103 (2018). arXiv:1805.04103 http://arxiv.org/abs/1805.04103

[12] Kaiming He, Georgia Gkioxari, Piotr Dollár, and Ross B. Girshick. 2017. Mask R-CNN. In *IEEE International Conference on Computer Vision, ICCV 2017, Venice, Italy, October 22-29, 2017.* 2980–2988. https://doi.org/10.1109/ICCV.2017.322

[13] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. 2016. Deep Residual Learning for Image Recognition. In *2016 IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2016, Las Vegas, NV, USA, June 27-30, 2016.* 770–778. https://doi.org/10.1109/CVPR.2016.90

[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.* 1510–1519. https://doi.org/10.1109/ICCV.2017.167

[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. https://doi.org/10.1109/CVPR.2017.632

[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.* 244–260. https://doi.org/10.1007/978-3-030-01261-8_15

[17] Hirokazu Kameoka, Takuhiro Kaneko, Kou Tanaka, and Nobukatsu Hojo. 2018. StarGAN-VC: Non-parallel many-to-many voice conversion with star generative adversarial networks. *CoRR* abs/1806.02169 (2018). arXiv:1806.02169 http://arxiv.org/abs/1806.02169

[18] Taeksoo Kim, Moonsu Cha, Hyunsoo Kim, Jung Kwon Lee, and Jiwon Kim. 2017. Learning to Discover Cross-Domain Relations with Generative Adversarial Networks. In *Proceedings of the 34th International Conference on Machine Learning, ICML 2017, Sydney, NSW, Australia, 6-11 August 2017.* 1857–1865. http://proceedings.mlr.press/v70/kim17a.html

[19] Diederik P. Kingma and Jimmy Ba. 2014. Adam: A Method for Stochastic Optimization. *CoRR* abs/1412.6980 (2014). arXiv:1412.6980 http://arxiv.org/abs/1412.6980

[20] Dmytro Kotovenko, Artsiom Sanakoyeu, Sabine Lang, and Björn Ommer. 2019. Content and Style Disentanglement for Artistic Style Transfer. In *2019 IEEE/CVF International Conference on Computer Vision, ICCV 2019, Seoul, Korea (South), October 27 - November 2, 2019.* IEEE, 4421–4430. https://doi.org/10.1109/ICCV.2019.00452

[21] Dmytro Kotovenko, Artsiom Sanakoyeu, Pingchuan Ma, Sabine Lang, and Björn Ommer. 2019. A Content Transformation Block for Image 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, 10032–10041. https://doi.org/10.1109/CVPR.2019.01027

[22] 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.* 385–395. http://papers.nips.cc/paper/6642-universal-style-transfer-via-feature-transforms

[23] Jae Hyun Lim and Jong Chul Ye. 2017. Geometric GAN. *CoRR* abs/1705.02894 (2017). arXiv:1705.02894 http://arxiv.org/abs/1705.02894

[24] Shuang Ma, Jianlong Fu, Chang Wen Chen, and Tao Mei. 2018. DA-GAN: Instance-Level Image Translation by Deep Attention Generative Adversarial Networks. In *2018 IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2018, Salt Lake City, UT, USA, June 18-22, 2018.* 5657–5666. https://doi.org/10.1109/CVPR.2018.00593

[25] Youssef Alami Mejjati, Christian Richardt, James Tompkin, Darren Cosker, and Kwang In Kim. 2018. Unsupervised Attention-guided Image-to-Image Translation. In *Advances in Neural Information Processing Systems 31: Annual Conference on Neural Information Processing Systems 2018, NeurIPS 2018, 3-8 December 2018, Montréal, Canada.* 3697–3707. http://papers.nips.cc/paper/7627-unsupervised-attention-guided-image-to-image-translation

[26] Takeru Miyato, Toshiki Kataoka, Masanori Koyama, and Yuichi Yoshida. 2018. Spectral Normalization for Generative Adversarial Networks. *CoRR* abs/1802.05957 (2018). arXiv:1802.05957 http://arxiv.org/abs/1802.05957

[27] Takeru Miyato and Masanori Koyama. 2018. cGANs with Projection Discriminator. In *6th International Conference on Learning Representations, ICLR 2018, Vancouver, BC, Canada, April 30 - May 3, 2018, Conference Track Proceedings.* https://openreview.net/forum?id=ByS1VpgRZ

[28] Takeru Miyato and Masanori Koyama. 2018. cGANs with Projection Discriminator. *CoRR* abs/1802.05637 (2018). arXiv:1802.05637 http://arxiv.org/abs/1802.05637

[29] Augustus Odena, Christopher Olah, and Jonathon Shlens. 2017. Conditional Image Synthesis with Auxiliary Classifier GANs. In *Proceedings of the 34th International Conference on Machine Learning, ICML 2017, Sydney, NSW, Australia, 6-11 August 2017.* 2642–2651. http://proceedings.mlr.press/v70/odena17a.html

[30] Mirco Ravanelli, Philemon Brakel, Maurizio Omologo, and Yoshua Bengio. 2018. Light Gated Recurrent Units for Speech Recognition. *IEEE Trans. Emerging Topics in Comput. Intellig.* 2, 2 (2018), 92–102. https://doi.org/10.1109/TETCI.2017.2762739

[31] Olga Russakovsky, Jia Deng, Hao Su, Jonathan Krause, Sanjeev Satheesh, Sean Ma, Zhiheng Huang, Andrej Karpathy, Aditya Khosla, Michael Bernstein, Alexander C. Berg, and Li Fei-Fei. 2015. ImageNet Large Scale Visual Recognition Challenge. *International Journal of Computer Vision (IJCV)* 115, 3 (2015), 211–252. https://doi.org/10.1007/s11263-015-0816-y

[32] Artsiom Sanakoyeu, Dmytro Kotovenko, Sabine Lang, and Björn Ommer. 2018. A Style-Aware Content Loss for Real-Time HD Style Transfer. In *Computer Vision - ECCV 2018 - 15th European Conference, Munich, Germany, September 8-14, 2018, Proceedings, Part VIII.* 715–731. https://doi.org/10.1007/978-3-030-01237-3_43

[33] 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.* 1349–1357. http://jmlr.org/proceedings/papers/v48/ulyanov16.html

[34] Dmitry Ulyanov, Andrea Vedaldi, and Victor S. Lempitsky. 2016. Instance Normalization: The Missing Ingredient for Fast Stylization. *CoRR* abs/1607.08022 (2016). arXiv:1607.08022 http://arxiv.org/abs/1607.08022

[35] Dmitry Ulyanov, Andrea Vedaldi, and Victor S. Lempitsky. 2017. Improved Texture Networks: Maximizing Quality and Diversity in Feed-Forward Stylization and Texture Synthesis. In *2017 IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2017, Honolulu, HI, USA, July 21-26, 2017.* 4105–4113. https://doi.org/10.1109/CVPR.2017.437

[36] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Lukasz Kaiser, and Illia Polosukhin. 2017. Attention is All you Need. In *Advances in Neural Information Processing Systems 30: Annual Conference on Neural Information Processing Systems 2017, 4-9 December 2017, Long Beach, CA, USA.* 6000–6010. http://papers.nips.cc/paper/7181-attention-is-all-you-need

[37] Sanghyun Woo, Jongchan Park, Joon-Young Lee, and In So Kweon. 2018. CBAM: Convolutional Block Attention Module. In *Computer Vision - ECCV 2018 - 15th European Conference, Munich, Germany, September 8-14, 2018, Proceedings, Part VII.* 3–19. https://doi.org/10.1007/978-3-030-01234-2_1

[38] Daan Wynen, Cordelia Schmid, and Julien Mairal. 2018. Unsupervised Learning of Artistic Styles with Archetypal Style Analysis. *CoRR* abs/1805.11155 (2018). arXiv:1805.11155 http://arxiv.org/abs/1805.11155

[39] Zhijiao Xiao, Xiaole Zhang, and Xiaoyan Zhang. 2018. Semantic Correspondence Guided Deep Photo Style Transfer. In *Advances in Multimedia Information Processing - PCM 2018 - 19th Pacific-Rim Conference on Multimedia, Hefei, China, September 21-22, 2018, Proceedings, Part I.* 81–93. https://doi.org/10.1007/978-3-

030-00776-8_8

[40] 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. https://doi.org/10.1109/CVPR.2019.00156

[41] Zili Yi, Hao (Richard) Zhang, Ping Tan, and Minglun Gong. 2017. DualGAN: Unsupervised Dual Learning for Image-to-Image Translation. In *IEEE International Conference on Computer Vision, ICCV 2017, Venice, Italy, October 22-29, 2017*. 2868–2876. https://doi.org/10.1109/ICCV.2017.310

[42] Weidong Yin, Ziwei Liu, and Chen Change Loy. 2018. Instance-level Facial Attributes Transfer with Geometry-Aware Flow. *CoRR* abs/1811.12670 (2018). arXiv:1811.12670 http://arxiv.org/abs/1811.12670

[43] Xiaoming Yu, Xing Cai, Zhenqiang Ying, Thomas H. Li, and Ge Li. 2018. SingleGAN: Image-to-Image Translation by a Single-Generator Network Using Multiple Generative Adversarial Learning. In *Computer Vision - ACCV 2018 - 14th Asian Conference on Computer Vision, Perth, Australia, December 2-6, 2018, Revised Selected Papers, Part V (Lecture Notes in Computer Science)*, C. V. Jawahar, Hongdong Li, Greg Mori, and Konrad Schindler (Eds.), Vol. 11365. Springer, 341–356. https://doi.org/10.1007/978-3-030-20873-8_22

[44] Xiaoming Yu, Zhenqiang Ying, and Ge Li. 2018. Multi-Mapping Image-to-Image Translation with Central Biasing Normalization. *CoRR* abs/1806.10050 (2018). arXiv:1806.10050 http://arxiv.org/abs/1806.10050

[45] Bolei Zhou, Àgata Lapedriza, Aditya Khosla, Aude Oliva, and Antonio Torralba. 2018. Places: A 10 Million Image Database for Scene Recognition. *IEEE Trans. Pattern Anal. Mach. Intell.* 40, 6 (2018), 1452–1464. https://doi.org/10.1109/TPAMI.2017.2723009

[46] Jun-Yan Zhu, Taesung Park, Phillip Isola, and Alexei A. Efros. 2017. Unpaired Image-to-Image Translation Using Cycle-Consistent Adversarial Networks. In *IEEE International Conference on Computer Vision, ICCV 2017, Venice, Italy, October 22-29, 2017*. 2242–2251. https://doi.org/10.1109/ICCV.2017.244

A CONTENT

In this supplementary material, we elaborate on the specific structures of the networks used and the training details. We also provide more generated paintings of different artists style to show the effectiveness of our method.

The contents are given in the following sequence:

(1) The network structure of ASMA-GAN. This section contains two sub-sections: 1) detailed network structure; 2) The training strategy of generator and discriminator.
(2) The paintings generated by the single network we proposed under the style of eleven painters. In this section, we provide the generation effect of more high-definition and large-size images.

Images are best viewed in color and zoomed in. Our source code will be made available soon in Github. And more generated pictures are in the compressed package.

B IMPLEMENT DETAILS OF ASMA-GAN

B.1 Detailed Network Structure

The structure of ASMA-GAN (mentioned in our paper) is given. The structure of the generator is shown in Fig. 12. The structure of the discriminator is shown in Fig. 13.

B.2 Training Strategy.

The three components of our framework are jointly trained. As the SNs make the discriminator training process become more difficult. We alternately train the Multi-Scale Projection Discriminator and the Conditional Generator, where 3 times for the Multi-Scale Projection Discriminator and once for the Conditionals Generator. Due to memory limitations, we start training the framework at the resolution of 256^2. After the 10^5 steps training, we increase

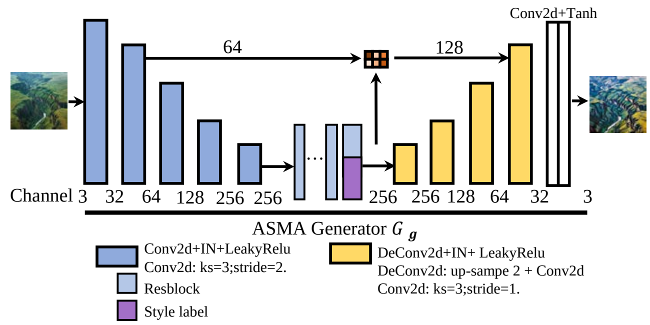


Figure 12: Network structure and parameter details of ASMA-GAN generator, where ks means the kernel size of convolution layers.

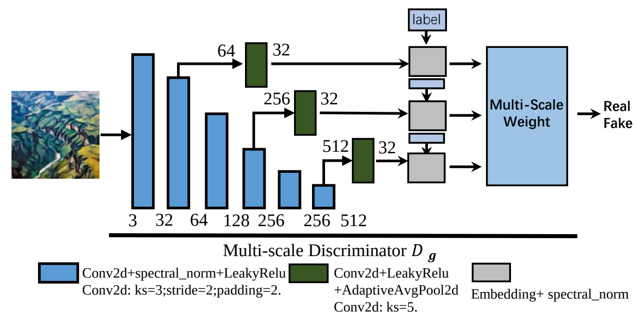


Figure 13: Network structure and parameter details of ASMA-GAN discriminator.

the resolution to 512^2 and continue training 10^5 steps. Finally, we increase the resolution to 768^2.

C RESULTS OF DIFFERENT STYLE TRANSFER METHODS

In the text, due to space limitations, we only show the details of the results of different style transfer methods. In Fig. 15, we provide the global result images of different methods on the same image of Munch's style.

D EXTRA RESULTS OF ASMA-GAN

We provide more generated HD paintings using our method under eleven artists styles. Each generated image has a minimal side size of 1600 pixels. In Fig. 15 to Fig. 35, we show the results of style transfer of eleven painter styles using ASMA-GAN. We select multiple types of pictures including portraits, woods, mountains, and buildings to show the generalization and flexibility of our method.

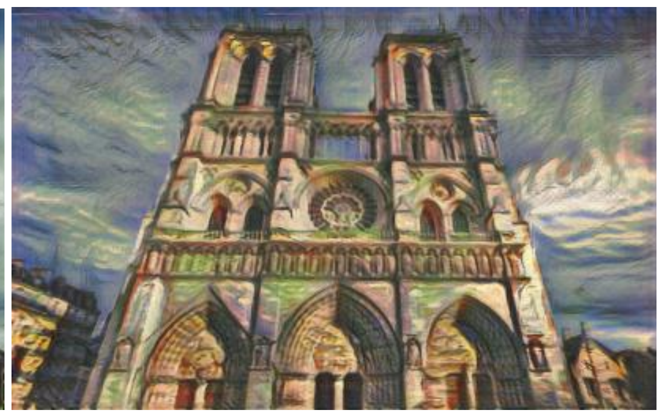
For each picture, the name of the picture is corresponding to the artist, the small picture in the upper left corner is the original picture, and the small picture in the lower left corner is one of the representative works of the artist.

From the results, we can see that the proposed ASMA-GAN realizes the style transfer while retaining the anisotropic semantic information. On one hand, the generated images realistically restore the original painter's style. By enlarging the picture, we can observe that the proposed method not only changes the color of

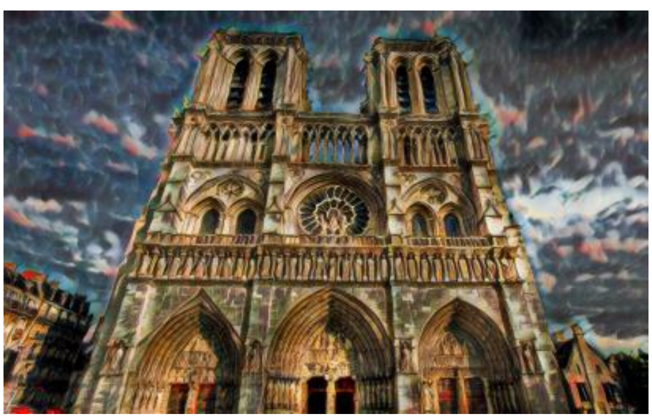
Munch

Original Picture

Ours



Style-Aware



AdaIN

AAMS

the picture, but also imitates the stroke of the artists, which is the really meaning of style. And the differences between the generated styles of different artists are big by using just single network we proposed. On the other hand, the semantic information of the original image is easily discarded because of the excessive transfer of style. Our proposed method achieves a compromise between style and content. The content retained in the generated picture is consistent with the artist's drawing habits.

Figure 15: Picasso

Figure 16: Samuel

Figure 17: Samuel

Figure 18: Morisot

Figure 19: Munch

Figure 20: Munch

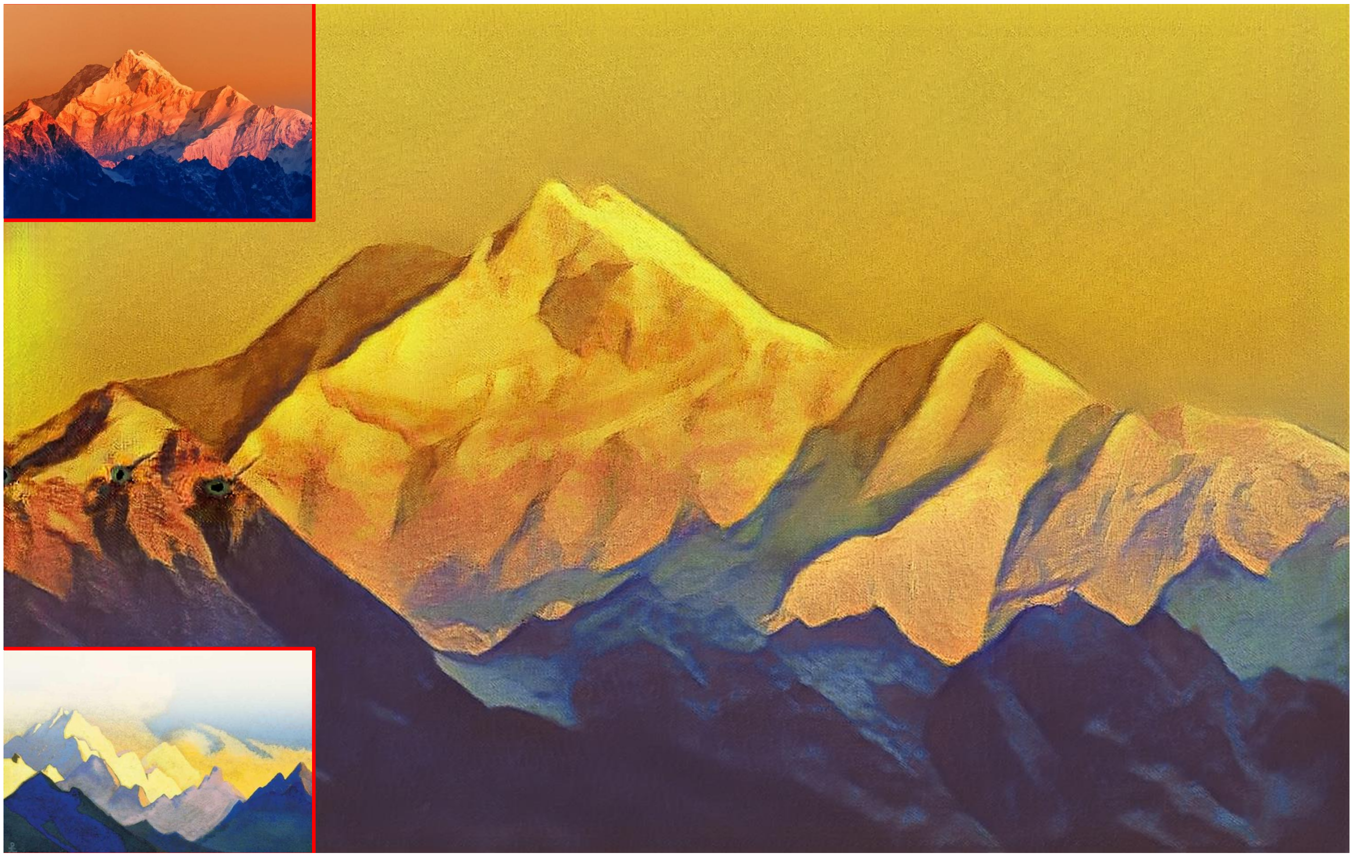


Figure 21: Nicholas

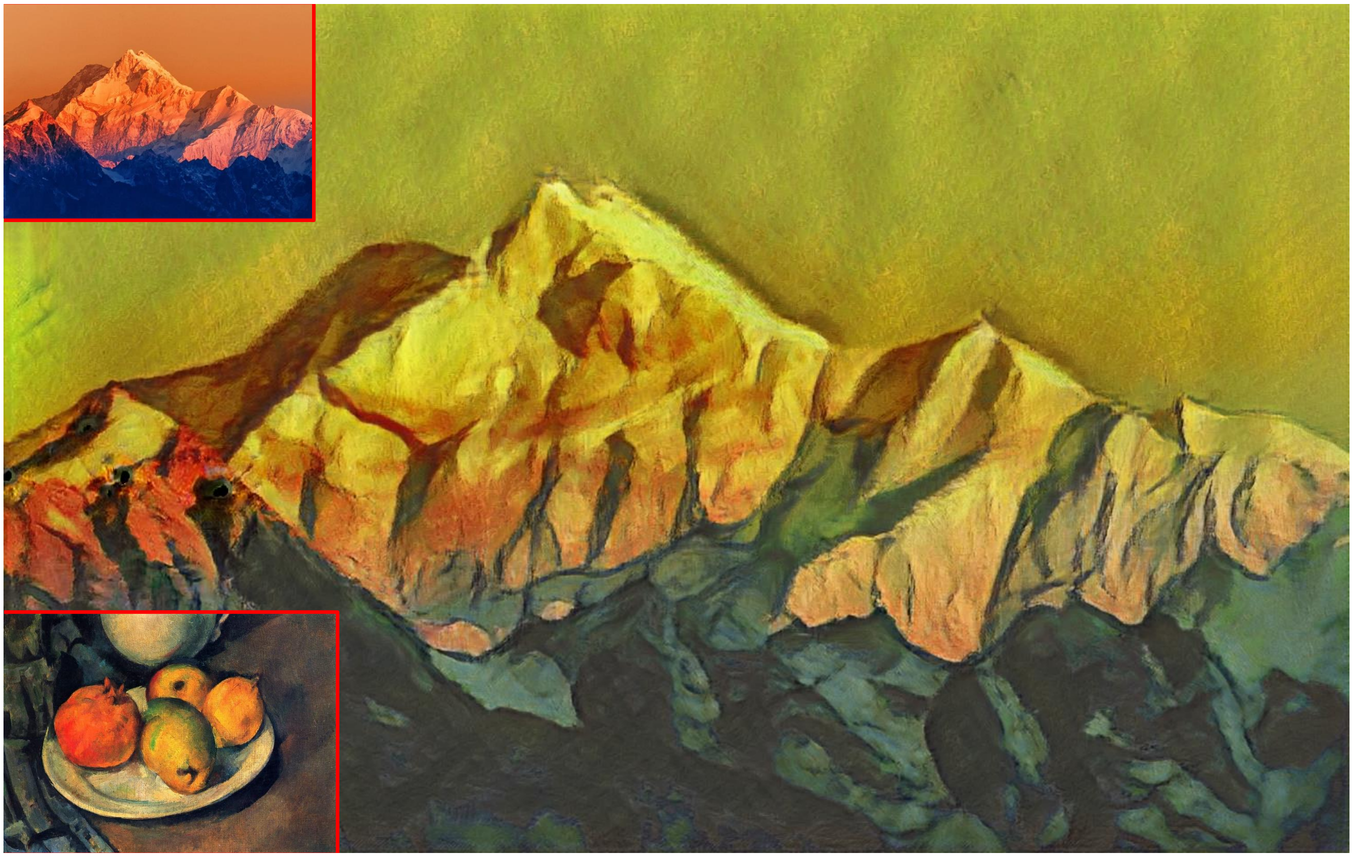


Figure 22: Cezanne

Figure 23: Samuel

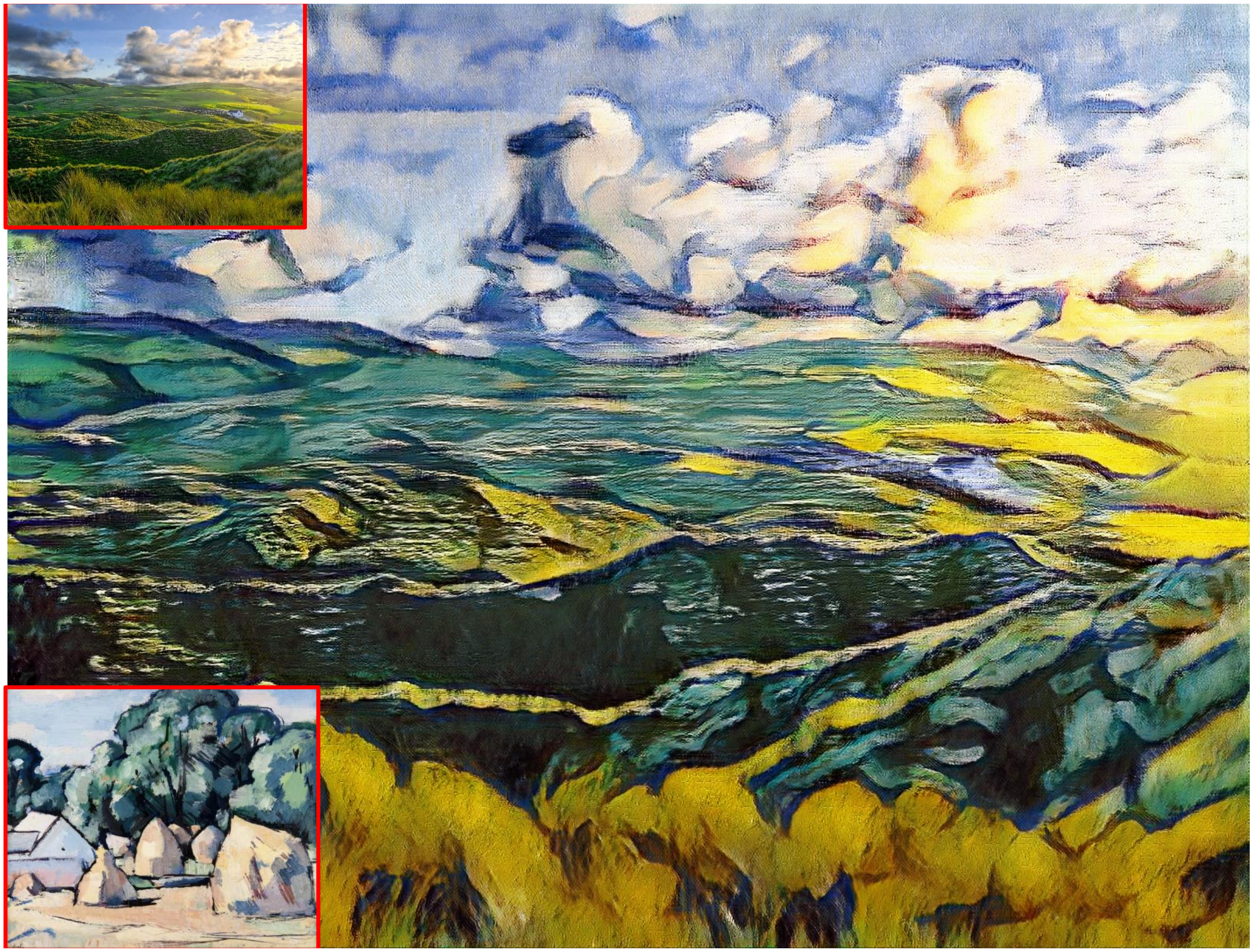


Figure 24: Samuel

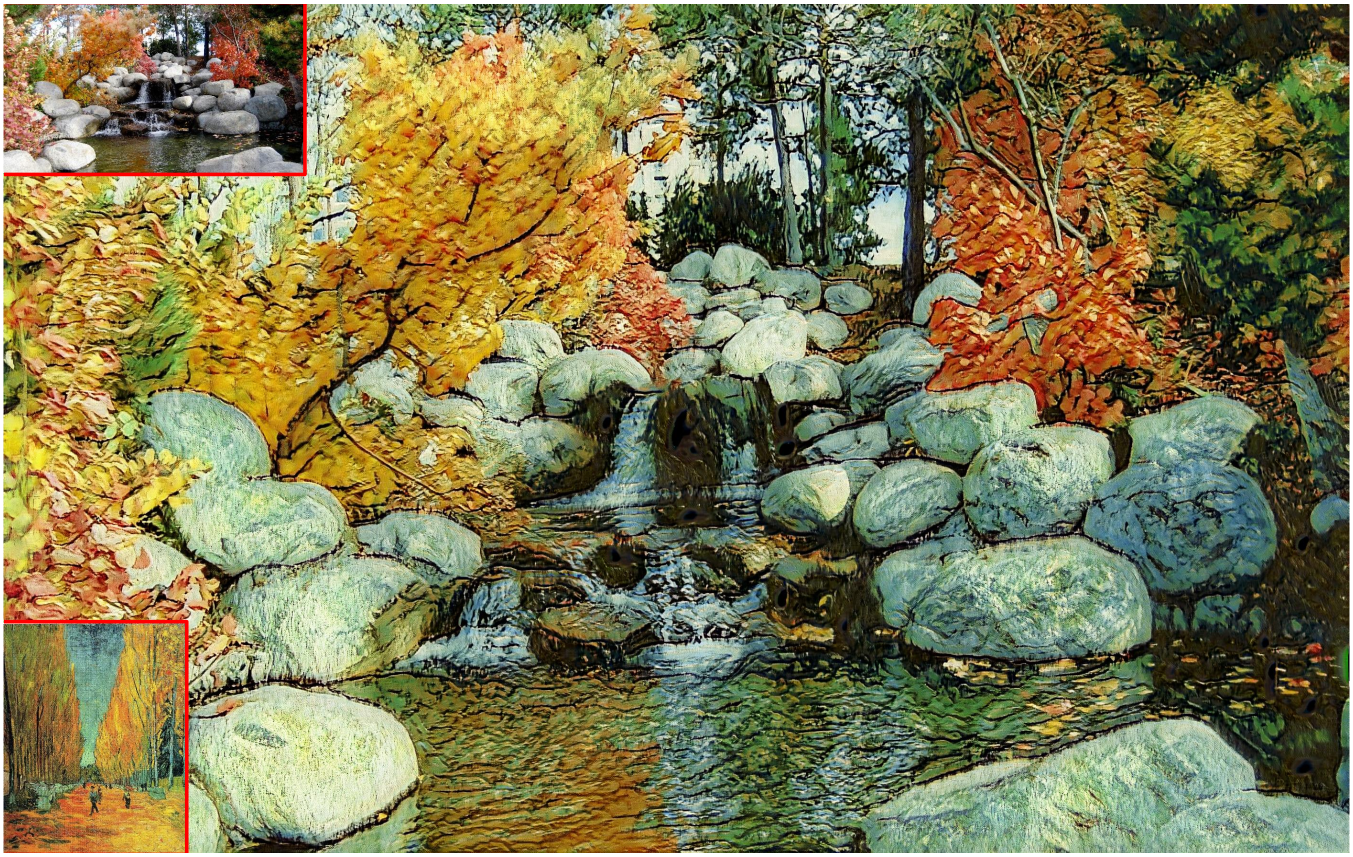


Figure 25: Vangogh

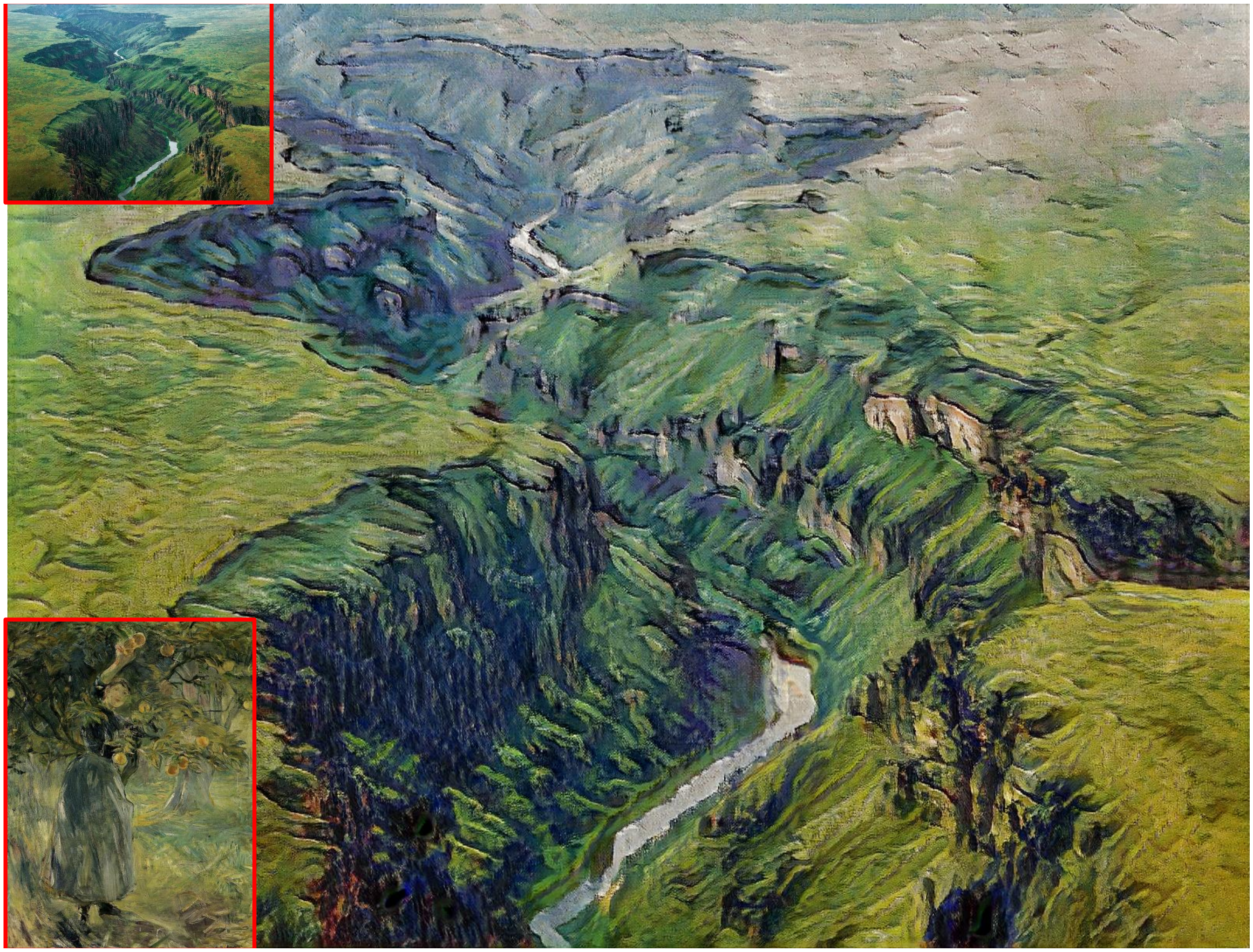


Figure 26: Morisot

Figure 27: Munch

Figure 28: Munch

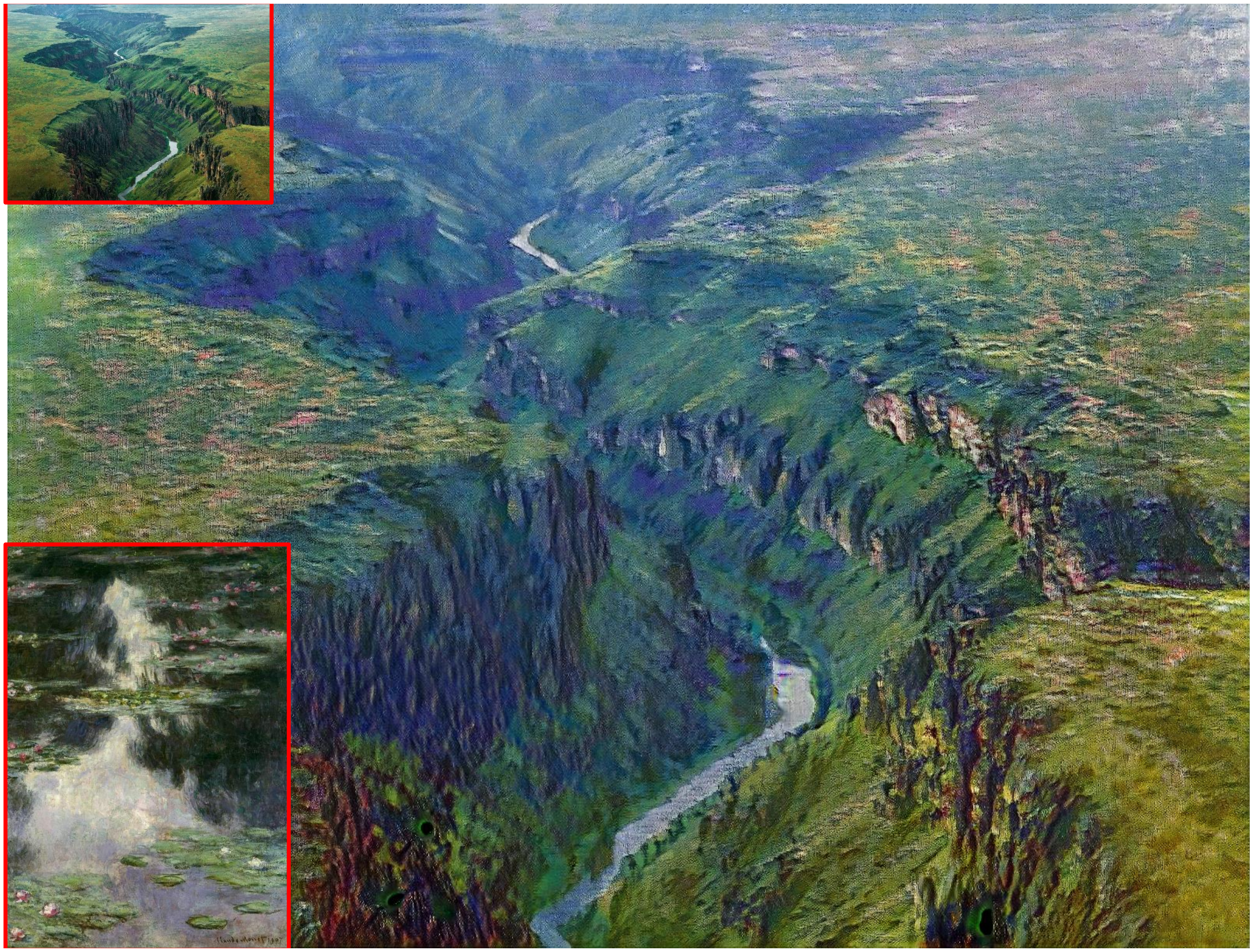


Figure 29: Monet

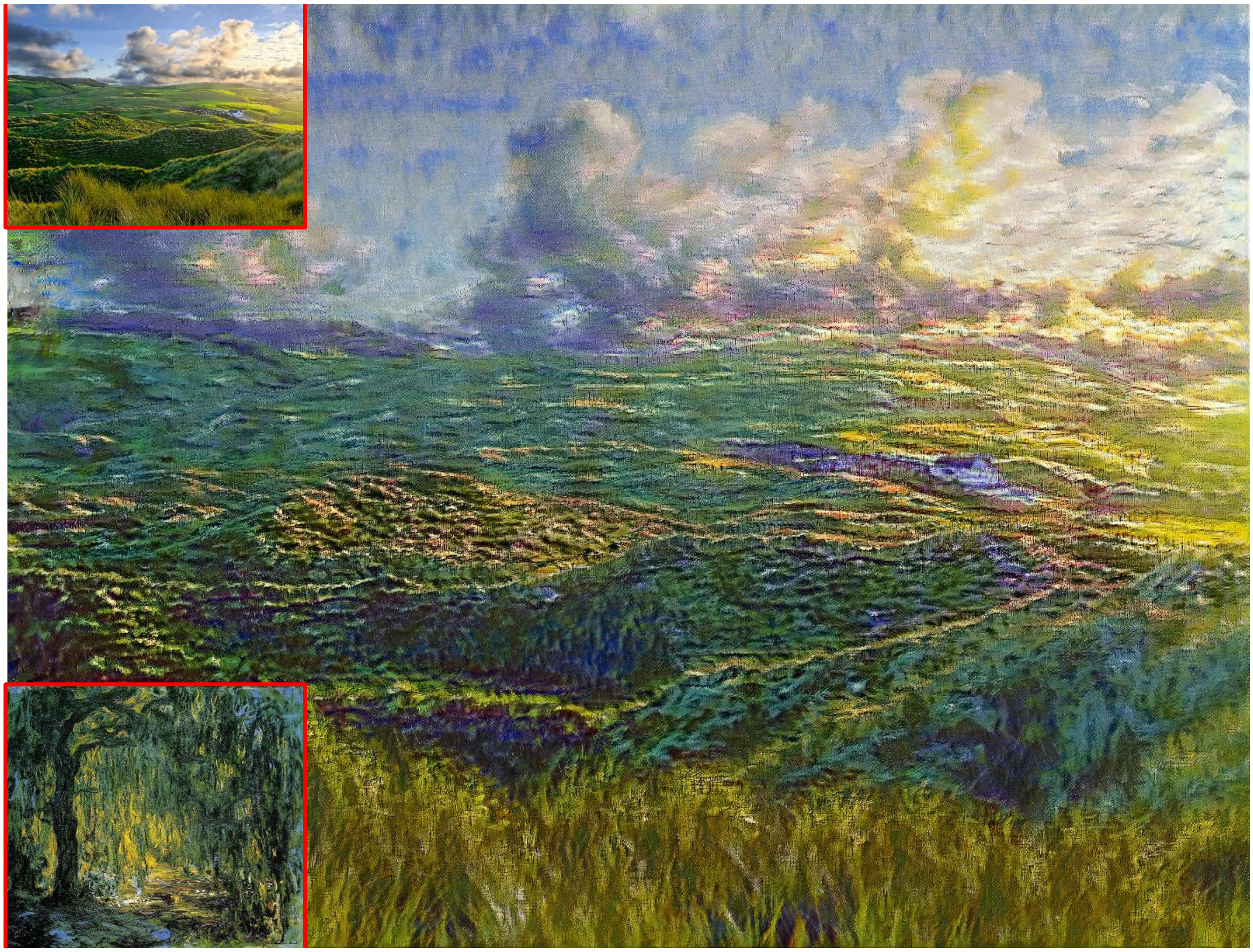


Figure 30: Monet

Figure 31: Samuel

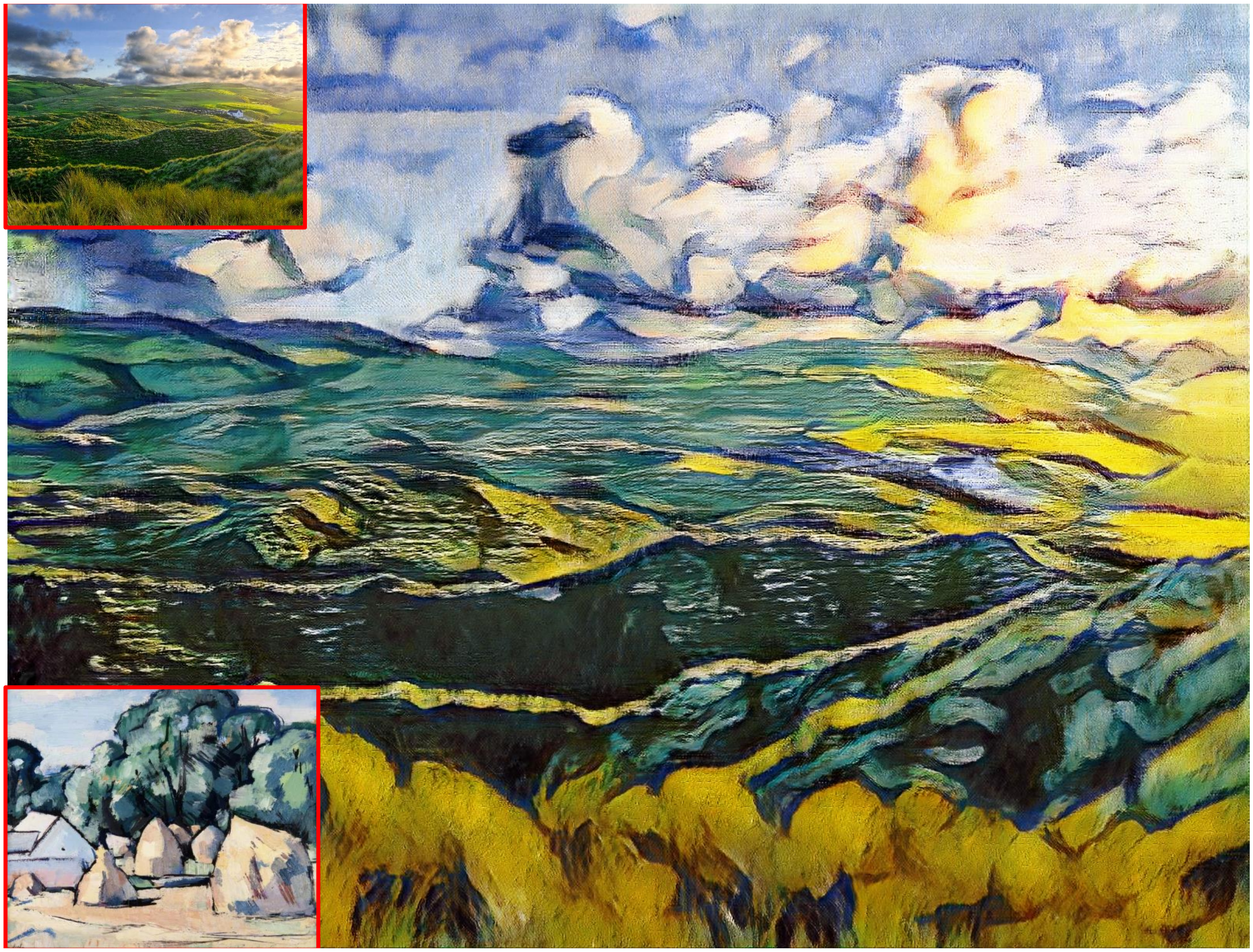


Figure 32: Samuel

Figure 33: Samuel

Figure 34: vangogh

Figure 35: Vangogh