Syntheses of Dual-Artistic Media Effects Using a Generative Model with Spatial Control

Heekyung Yang 1 and Kyungha Min 2,*

1 Division of SW Convergence, Sangmyung University, Seoul 03016, Korea; yanghk@smu.ac.kr
2 Department of Computer Science, Graduate School, Sangmyung University, Seoul 03016, Korea
* Correspondence: minkh@smu.ac.kr; Tel.: +82-2-2287-5377

Abstract: We present a generative model with spatial control to synthesize dual-artistic media effects. It generates different artistic media effects on the foreground and background of an image. In order to apply a distinct artistic media effect to a photograph, deep learning-based models require a training dataset composed of pairs of a photograph and its corresponding artwork images. To build the dataset, we apply some existing techniques that generate an artwork image including colored pencil, watercolor and abstraction from a photograph. In order to produce a dual artistic effect, we apply a semantic segmentation technique to separate the foreground and background of a photograph. Our model applies different artistic media effects on the foreground and background using space control module such as SPADE block.

Keywords: artistic style; media simulation; generative adversarial network; SPADE; SDAGAN

1. Introduction

Synthesizing artwork images is one of the most frequently challenged problems in computer graphics. The researchers in computer graphics community developed various computational models to mimic the artistic media effects to synthesize visually pleasing artistic styles expressed using real artistic media such as pencils or brushes [1–3]. Recently, the advancement of deep learning techniques has greatly accelerated artistic style synthesis techniques [4]. The style embedded in a sample artwork is transferred to a target photograph using deep convolutional neural networks without developing explicit computational models.

Using a texture-based approach, deep learning-based artistic style synthesis techniques transfer the styles extracted from a sample artwork image into a target photograph. Recently, pix2pix [5] successfully translated images in a domain to the images in a different domain by learning the styles inherent in the domain and applying them to images in the different domain. Pix2pix requires that the training set must consist of matched pairs. The content of the images that belong to different domains should be coincident. Consequently, pix2pix has hardly been applied to artistic style translation, since it is very difficult to collect a pair of image sets for photographs and their matched artwork images.

Zhu et al. [6] presented CycleGAN that resolves the limitation of a paired dataset of pix2pix. Instead of requiring a paired dataset, they present a cyclic generation of images in two distinct domains and devise a cycle consistency constraint for the original and generated images. CycleGAN successfully applies various artistic styles inherent in artwork domain to the images in the photograph domain. However, the style applied to photographs tends to be faint, since CycleGAN averages the styles extracted from many samples. Later, Park et al. [7] introduced GauGAN that includes a Spatially Adaptive Denormalization (SPADE) module, which regulates style transfer in locally segmented regions. They collect similar regions from a set of images and apply styles on the region-based normalization.
In some visual contents, the selected regions of a scene that the creator wants to emphasize are rendered in a different style. For example, in Schindler’s List, the famous black-and-white film, only the girl in a red coat was rendered in color, giving the audience more intense instillation of the girl’s emotions. The most widely-used method to implement different artistic styles is to build a mask that specifies the regions where different styles are applied. A stylized image with different artistic styles using a mask that segments an image into separate regions can be a groundtruth to the result of our method. With the recent development of deep learning-based methods that produce artistic styles, we aim to develop a technique that produces dual artistic styles to a scene in an end-to-end framework. Our model that produces both single and dual artistic styles in an identical end-to-end framework can reduce the efforts of synthesizing artistic styles using masks.

We present SDAGAN (Spatially-controlled Dual Artistic effect Generative Adversarial Network) to synthesize dual artistic effects from an input photograph. The first motivation of SDAGAN is to apply the pix2pix approach in style transfer in order to present a salient artistic media effect on an input photograph. To build a training dataset of matched pairs, we employ the existing techniques that mimic artistic media effects. We chose three most representative artistic media effects used in computer graphics and computer vision society: watercolor, color pencil drawing and abstraction with lines. We use Kang et al.’s work [8] for abstraction with lines, Bousseau et al.’s work [2] for watercolor and Yang et al.’s work [3] for color pencil drawing.

The second motivation is to apply different artistic media effects on an input photograph. For this purpose, we apply a semantic segmentation scheme to segment the photograph into foreground and background. We employ SPADE from Park et al.’s work [7] to apply dual artistic effects on a photograph. For example, the foreground of a photograph may be stylized with a watercolor effect, while the background may be stylized with a pencil drawing effect. Some researchers [9–12] have used this approach. However, these schemes have limitations in synthesizing artistic media effects.

We illustrate our framework in Figure 1. A single artistic effect is generated using our SDAGAN with deactivated SPADE blocks. We attach a control vector to select the artistic effect to generate. To generate dual artistic effects, we employ a segmentation module that segments the input photograph into foreground and background. This segmentation map is fed into the SPADE blocks to specify the region of applying the artistic effects. We assign two condition vectors for dual artistic effects. As illustrated in Figure 1b, we can generate various combination of dual artistic effects.

SDAGAN is trained in two different modes: single artistic media effect synthesis and dual artistic media synthesis. The single synthesis mode is trained like pix2pix [5]. For the training, pairs of input photographs and their matched stylized images are used. The SPADE block is deactivated in this mode. The pre-trained parameters from the single synthesis model are used in the dual synthesis mode. The segmentation information is also used for training in this mode. SDAGAN synthesizes dual artistic media effects on an input photograph after training with different effects applied to the foreground and background.

Our contribution can be summarized as follows:

- We build a paired dataset for photographs and their matched artwork images by employing the existing techniques that synthesize various artistic media effects. This approach enables us to generate artwork images with very realistic artistic media effects. Using this dataset, we can execute the pix2pix approach for translating a photograph to artwork images of various styles including abstraction, watercolor and pencil drawing.
- We develop a framework that synthesizes dual artistic media effects using region information extracted from a semantic segmentation scheme. The SPADE module introduced in GauGAN [7] is employed for this purpose. Our framework can successfully produce an artwork image whose foreground and background are depicted with different artistic media effects.
Figure 1. The overview of our framework: (a) single artistic effect generation using SDAGAN generator with deactivated SPADE blocks (The control vectors are used to specify user-desired effect), (b) dual artistic effect generation using SDAGAN generator with SPADE blocks (The segmentation map is employed in the SPADE blocks to apply different artistic media effects).

2. Related Work

2.1. Procedural Work

2.1.1. Abstraction and Line Drawing

Abstraction is a drawing technique that removes tiny textures and abstracts complex colors to present an object’s coarse shape. Many researchers [1,8,13–16] present abstraction techniques for photograph colors and tiny textures. They also presented various line drawing schemes for clearly presenting the distinguishing shape of objects [17–19].

2.1.2. Watercoloring

Simulating the physics of watercolor on paper surfaces was pioneered by Curtis et al. [20]. Later, Bousseau et al. [2,21] presented a texture-based approach to rendering video clips in watercolor style. Kang et al. [22] presented a computational model for watercolor and Laerhoven et al. [23] presented a physically-based model for watercolor.

2.1.3. Pencil Drawing

Sousa and Buchanan [24] analyzed the physics of pencils and paper surfaces to mimic pencil drawing. Later, Matsui et al. [25] and Murakami et al. [26] presented a pencil drawing stroke overlapping scheme. Kwon et al. [27] and Yang et al. [3] created a framework based on convolutions for pencil drawing on 3D meshes or photographs.

2.2. Deep Learning-Based Work

Gatys et al. [4] have recently presented a texture-based style transfer scheme that innovatively changed the artistic image synthesis framework. They define a Gram ma-
trix for extracting the styles from a texture scale sample image and present an iterative optimization process for applying the styles in the Gram matrix to a content image. This scheme prompts a series of follow-up studies on applying styles from samples to a target image in the absence of explicit procedures. These methods are known as image-based optimization stylization.

Model-based style transfer, another category of deep learning-based stylization, originates from Goodfellow et al.’s generative adversarial network (GAN) [28]. GAN comprises a couple of a generator and a discriminator. The generator creates an image to fool a discriminator and the discriminator’s goal is to detect the generator’s forgery. This antagonistic relationship increases the power of a generator. Following the training process, a generator can create any image recognized as a real photograph.

Radford et al. [29] enhanced the GAN structure by combining a convolutional network and a GAN to improve the image synthesis framework. Isola et al. [5] presented a pix2pix framework that learns styles from many samples and then applies the learned style to an input image. As a result, an abstract style, such as Gogh’s or Monet’s, can be applied to a photograph. This approach brings a limitation that the styles learned from many samples tend to get averaged. Later, Zhu et al. [6] presented CycleGAN, which converts images of matched pairs in crossway. An artwork image is converted to its paired photograph image, and vice versa.

Recently, some researchers focus on producing individual artistic effects. Chen and Tak [30] and Zhou et al. [31] applied GAN architecture to produce pencil drawing effects. Kim et al. [32] applied GAN with attention module to produce image abstraction effects from portraits. Platkevic et al. [33] and Sochorova and Jamriska [34] presented physically-based models on simulating artistic media effects such as watercolor and oilpainting.

2.3. Region-Based Work

Gatys et al. [35] presented a spatial control scheme for style transfer by applying a mask to select a series of styles from multiple style images. Champandard [9] proposed a semantic map scheme based on doodles for artistic image synthesis of spatial control. Rather than combining styles from various samples, he focuses on controlling the style of the semantic map to the doodles.

Huang and Belongie [10] presented an adaptive instance normalization (adaIN) scheme for artistic image synthesis. The adaIN replaces the batch normalization scheme, which normalizes input images on a batch scale, to improve the efficiency of learning and stylization. They also used a spatial control scheme based on a feed forward approach. Li et al. [11] also presented a mask-guided spatial control scheme that uses user-edited masks to apply a series of styles to an input image. Castillo et al. [12] presented an instance semantic segmentation-based spatial control scheme for artistic image synthesis. They blend different styles on the segmented regions using Markov random field.

Park et al. [7] recently presented a gauGAN that applies a doodle-based normalization (SPADE) for photographic image synthesis. The regions guided by a series of doodles are normalized for efficient image synthesis.

2.4. Dual Style Transfer

We briefly review several existing works that combine two different styles in a result image. Some of them focus on preserving both style and content and others consider two artistic styles.

Artists alter their styles over their lifetime. To follow this alternation of styles, Kotovenko et al. [36] presents a style transfer model that separates style and content for stylization. They aim to follow both a common style spread on overall samples and a fine-detailed style on a specific sample. For this purpose, their model involves two novel loss terms: fixpoint triple style loss and fixpoint disentanglement loss. These two loss terms enable a better of style distributions. Kotovenko et al. [37] presents a style transfer model that transfers both content and style simultaneously. They focus on the content transfer,
which has not been studied by the existing works. They claim that content details are altered according to the styles. Therefore, they devise a content transformation module located between an encoder and a decoder. This transformation module includes a local feature normalization layer, which is effective in reducing artifacts on stylized images. Recently, Svobada et al. [38] presented a stylization model that recombines style and content in a latent space using a two-stage peer-regularization layer. Since their model does not rely on the pre-trained network, their model allows a zero-shot style transfer. This approach preserves the content from content samples in a solid way, while transfers the style from style samples in a visually pleasing way.

Transferring styles samples from one or two images may produce unwanted biased results. Sanakoyeu et al. [39] resolved this limitation by devising style-aware content loss term that trains a style transfer network composed of an encoder and a decoder. This loss term enables real-time and high-resolution stylization. Chen et al. [40] presents a dual style transfer model that learns holistic artist style and specific artwork style simultaneously. The holistic artist style expresses the tone and specific artwork style depicts the detail of the stylization.

3. Our Generative Model with Spatial Control

3.1. Generator

We build the structure of our generative model for artistic style synthesis using CycleGAN structure [6]. CycleGAN’s original generator combines an encoder and a decoder block. However, the input of our approach includes bi-segmentation information for applying various artistic media effects to an image. We design our generator with residual blocks and SPADE [7] to properly use the bi-segmentation information. Our generator has four blocks: an encoder block, a residual block, a SPADE block and a decoder block. Figure 2 depicts our generator.

![Figure 2](image)

**Figure 2.** The architecture of our generator: An encoding block, a residual block and a decoder block is aligned. In the decoder block, we place three SPADE blocks to properly process the region information.

3.1.1. Encoder Block

In the encoder block, we downsample the input image by applying two convolutional layers. The input image whose dimension is $256 \times 256 \times 3$ is encoded to $128 \times 128 \times 128$ by conv3-64 and conv3-128 layers.
3.1.2. Residual Block
At the end of the encoder block, we devise a residual block that adds two feature maps in an element-by-element way. A conv3-256 layer produces the second feature map in the residual block. This residual block prevents the loss of information that may come from the convolution layers.

3.1.3. SPADE Block
In this SPADE block, the bi-segmentation information is convolved to be multiplied and added to the feature map element-wise. The feature map can be derived from the residual block or the decoder block’s transconvolutional layers. We locate the SPADE block for the feature maps of the transconvolutional layers because this bi-segmentation information separates the styles for the segmented regions. Figure 3 depicts the structure of our SPADE block.

![Figure 3. The structure of SPADE module in our generator.](image)

3.1.4. Decoder Block
Finally, our decoder block reconstructs the resolution of the input image by two transconvolution layers including transconv3-128 and transconv3-64. Each output of the transconvolution layers is post-processed through the SPADE block.

3.2. Discriminator
For our generative model’s discriminator, we use the patchGAN architecture for the CycleGAN discriminator [6]. Our discriminator is composed of six convolutional layers. At the end of the discriminator, a 14 × 14 one-dimensional map is produced. Each element of the map corresponds to the discrimination result of each patch. Figure 4 presents our discriminator’s architecture.

![Figure 4. The architecture of our discriminator.](image)
3.3. Loss Function

The loss functions in this work are presented in Equation (1), which is a weighted sum of GAN loss term in Equation (2) and cycle loss term in Equation (3). These equations were developed by Zhu et al. [6].

\[
L(G, F, D_X, D_Y) = L_{GAN}(G, D_Y, X, Y) + \lambda L_{cyc}(G, F),
\]

\[
L_{GAN}(G, D_Y, X, Y) = \mathbb{E}_{y \sim p_{data}(y)}[\log D_Y(y)] + \mathbb{E}_{x \sim p_{data}(x)}[\log(1 - D_Y(G(x)))],
\]

\[
L_{cyc}(G, F) = \mathbb{E}_{x \sim p_{data}(x)}[\|F(G(x)) - x\|_1] + \mathbb{E}_{y \sim p_{data}(y)}[\|G(F(y)) - y\|_1].
\]

Equation (4) was developed in Park et al.’s work [7].

\[
h_{n,c,y,x}^i = \gamma_{c,y,x}(m) \frac{h_{n,c,y,x}^i - \mu_{c}^i}{\sigma_{c}^i} + \beta_{c,y,x}^i(m),
\]

where

\[
\mu_{c}^i = \frac{1}{NHW} \sum_{n,y,x} h_{n,c,y,x}^i,
\]

\[
\sigma_{c}^i = \sqrt{\frac{1}{NHW} \sum_{n,y,x} (h_{n,c,y,x}^i - \mu_{c}^i)^2}.
\]

4. Training

4.1. Generation of the Training Dataset

We train our model by artistic images synthesized from existing non-photorealistic (NPR) studies. Kang et al.’s work [8] generates abstracted images by integrating color along smooth flow embedded in an image. This scheme can generate both abstracted images and abstracted images with lines. Bousseau et al.’s work [2] is used to create the watercolor style by synthesizing watercolor textures using Perlin noise. Yang et al.’s work [3] is designed using a line integral convolution (LIC) scheme to mimic various pencil styles including color pencil and monochrome pencil with various strokes. This scheme produces four variants of pencil drawing styles. Among the various styles that can be produced by these three techniques, we choose three styles: abstraction with lines, watercolor and color pencil drawing with thick stroke. We generate 1.5 K images for each style. Therefore, our dataset has 4.5 K images. The generated images using these schemes are suggested in Figure 5. The 3.3 K images of the dataset are used for training, 0.6 K images for validation and 0.6 K for test. Details of the dataset are suggested in Table 1.
Among the seven styles we can produce, we select three styles including abstraction with lines, watercolor and color pencil drawing with thick stroke for our dual style synthesis. The selected three styles are marked in red rectangles.
Table 1. Details of dataset.

|                     | Dataset                        | Size  | Resolution |
|---------------------|-------------------------------|-------|------------|
| training set        | abstraction with lines        | 1.1 K | 256 × 256  |
|                     | watercolor                    | 1.1 K | 3.3 K      |
|                     | color pencil with thick stroke| 1.1 K | 256 × 256  |
| validation set      | abstraction with lines        | 0.2 K | 256 × 256  |
|                     | watercolor                    | 0.2 K | 0.6 K      |
|                     | color pencil with thick stroke| 0.2 K | 256 × 256  |
| test set            | abstraction with lines        | 0.2 K |            |
|                     | watercolor                    | 0.2 K |            |
|                     | color pencil with thick stroke| 0.2 K |            |

4.2. Foreground/Background Segmentation

Many researchers have presented various deep learning-based schemes that segment a scene into foreground and background [41–43]. In these schemes, objects in a foreground of a scene are separated from their backgrounds through a series of convolutional neural nets. Bouwmans et al. presented a brief survey on these schemes [44]. In this study, we employ a BiSeNet [45] that effectively segments scenes into several regions. BiSeNet is very effective in segmenting regions with different depths. Since many landscape images are composed of closer regions and farther backgrounds, BiSeNet successfully segments foreground and background of a scene.

5. Implementation and Results

5.1. Implementation

We have implemented our framework on a PC with a Pentium i7 CPU and 16 GByte main memory. An nVidia TitanX GPU with 6 GByte memory accelerates our implementation. We used Pytorch version 1.2.0 and Cube version cu92 as our software environment. It takes 18 h to train our model for generating artistic media effects. In every training process, we set $\lambda$ of GAN loss as 10. We employed ADAM for our optimizer, and set the learning rate as 0.0002, epoch as 200, and batch size as 8. After the training, our model takes about 1.5 s to produce a result image.

5.2. Results

We apply SDAGAN to twenty images and produced a series of result images that include three images of single artistic media effects and six images of dual artistic media effects.

5.2.1. Synthesis of Single Artistic Media Effects

We generate images of single artistic styles by executing our SDAGAN with disabling SPADE module. Among the seven styles in Figure 5 that we can produce using existing methods, we focus on three styles: abstraction with lines, watercolor and color pencil drawing with thick stroke. After training SDAGAN using the dataset in Table 1, our model can produce images whose styles are similar to those of the training samples (see Figure 6).
5.2.2. Synthesis of Dual Artistic Media Effects

Artistic media effects produced in this study have their own benefits and shortcomings. Abstraction with lines, for example, can express the details of an object in the foreground, while sacrificing smooth tonal change in the background. The watercolor effect may lose the objects' detail because they express a scene with smooth tonal values. The pencil drawing effect can express objects in the foreground with details and tonal changes in the background, reducing a scene’s contrast.

Therefore, single artistic media effect’s shortcomings can be addressed by mixing two artistic media effects. We present dual-artistic media effects by synthesizing abstraction with watercolor effects in Figures 7–9. The original image of Figure 7 depicts mountains, an ocean and a blue sky. The foreground’s land appears dark, while the ocean and sky in the background appear bright. The abstraction effect depicts the details of the mountains, but loses the smooth tonal change of the blue sky and ocean. In contrast, the watercolor effect expresses the smooth tonal change in the background, but loses the details of the mountains. As a result, we separate the foreground and background and apply various artistic media effects. The foreground is expressed with abstraction and the background with a watercolor effect. This dual-artistic media effect can express details in the foreground with abstraction with lines and smooth tonal change in the background with watercolor effect. Figures 8 and 9 show similar results.

We present another dual-artistic media effect by synthesizing pencil drawing effects with watercolor effects in Figures 10–13. This synthesis aims for another effect as well.
as similar effects in the abstraction-watercolor synthesis. As shown in Figures 7–9, the pencil drawing effects brighten the results when compared to the original photographs. Figure 10 depicts the advantage of our approach. The pencil drawing effect applied to the original photograph boosts its brightness and restores the red color of the board player’s cloth. The watercolor effect used on the photograph preserves the image’s tone while depicting the mountain and sky with smooth tonal gradations. Combining these effects draws attention to the board player in the center of the image while preserving the overall tone. Figures 11–13 depict similar effects.

Figure 7. Dual artistic media effects with abstraction and watercolor using SDAGAN (1).

Figure 8. Dual artistic media effects with abstraction and watercolor using SDAGAN (2).
Figure 9. Dual artistic media effects with abstraction and watercolor using SDAGAN (3).

Figure 10. Dual artistic media effects with pencil drawing and watercolor using SDAGAN (1).
Figure 11. Dual artistic media effects with pencil drawing and watercolor using SDAGAN (2).

Figure 12. Dual artistic media effects with pencil drawing and watercolor using SDAGAN (3).
5.3. Comparison to Existing Studies

Champandard [9] proposed a 2-bit doodle-based approach that segments an image into four regions and applies different styles on the regions (see Figure 14a). Objects in the scene are rendered with the appropriate styles because they apply styles based on the segmented regions. They did not, however, experiment with dual-artistic styles for their results. Huang and Belongie [10] proposed an adaptive instance normalization scheme to address existing style transfer issues like slow optimization and texture smearing. They calculate the mean and variance of style features and apply them to content features. Using this strategy, they segment an image into two regions and apply different style textures to create a dual-styled result. However, this scheme does not consider the effects of artistic media on their styles (see Figure 14b). Castillo et al. [12] proposed an instance-aware style transfer scheme based on MRF transfer. They segment an input image and choose a region to apply their style (see Figure 14c). They do not think about applying different styles to different regions and combining them into single final result. Li et al. [11] proposed a universal style transfer scheme that addresses existing works’ limitations, such as poor generation to unseen styles or visual quality. They separate the stylization process into whitening and coloring transform. They apply this strategy to a multi-segmented image and produce a multi-styled result (see Figure 14d).
Figure 14. The similar results from existing works: (a) doodle-based style transfer [9], (b) style transfer on bi-segmented image [10], (c) stylization on segmented region [12], (d) universal style transfer on multi-segmented image [11].

5.4. Evaluation

In many existing style transfer studies, applying different styles on the different regions of a single image is hardly observed. Therefore, the evaluation of our scheme with similar existing studies is not feasible. Instead of comparing our study with the existing works, we compare our results with the images stylized by a single medium. We present two evaluation tests: the first evaluation is using FID and the second one is user survey.

5.4.1. FID Evaluation

We evaluate FID (Frechet Inception Distance) for the three stylized images: a stylized image for the foreground, a stylized image for the background and our result that combines the foreground and background stylized images. We compare the images in Figure 15 and suggest the FID values for the three stylized images in the left column of Table 2, where our result shows smallest FID values for nine cases from ten cases. Ten images in Figure 15 are grouped into five subsets, each of which contains one input image.

5.4.2. User Survey

For a user survey, we build two questions in order to evaluate our results. We prepare 10 sets of an input image and three stylization images: stylized with the foreground effect, stylized with the background effect, and our result. The 10 sets of the input images are grouped into five subsets, each of which share the input image. Each input image is stylized into two different result images. We aim to estimate the effects of stylization on the identical input image. We present two questions: one for the degree of stylization and the other for the preservation of content.

- **q1.** Evaluate the stylization for each stylized image. Mark the degree of the stylization in five-point metric: 1 for least stylized, 2 for less stylized, 3 for medium, 4 for more stylized, 5 for maximally stylized.

- **q2.** Evaluate how much information of the input image is preserved for each stylized image. The information includes details of objects in the scene as well as tone and
color. Mark the degree of preservation of the information in five-point metric: 1 for least preserved, 2 for less preserved, 3 for medium, 4 for somewhat preserved, 5 for maximally preserved.

We hire twenty persons including nine female and eleven male. Twelve of them are in their twenties and eight of them are in thirties. The results of user survey are summarized in the right column of Table 2. The values in Table 2 are averaged over twenty participants.

Table 2. FID values and user survey scores (F/G denotes foreground, and B/G denotes background).

| Case | Evaluation 1: FID | Evaluation 2: User survey |
|------|------------------|---------------------------|
|      | F/G  | B/G  | Ours | F/G  | B/G  | Ours | F/G  | B/G  | Ours |
| 1    | 381.1 | 333.6 | 236.3 | 3.4  | 3.3  | 3.9  | 2.6  | 3.4  | 4.2  |
| 2    | 314.0 | 333.6 | 280.1 | 3.5  | 3.3  | 4.1  | 3.1  | 3.4  | 4.1  |
| 3    | 487.6 | 381.5 | 236.2 | 3.6  | 3.5  | 4.2  | 2.4  | 3.3  | 4.1  |
| 4    | 489.4 | 381.5 | 247.1 | 3.5  | 3.5  | 4.3  | 3.0  | 3.3  | 3.9  |
| 5    | 275.1 | 324.4 | 262.3 | 3.4  | 3.1  | 4.0  | 4.2  | 3.1  | 3.9  |
| 6    | 275.1 | 315.2 | 251.4 | 3.4  | 3.4  | 4.3  | 4.2  | 3.0  | 3.8  |
| 7    | 241.8 | 225.2 | 150.3 | 2.9  | 3.3  | 3.8  | 2.8  | 3.4  | 4.2  |
| 8    | 108.5 | 225.2 | 118.9 | 3.4  | 3.3  | 4.4  | 2.7  | 3.4  | 4.4  |
| 9    | 242.5 | 201.7 | 174.5 | 3.1  | 3.5  | 3.3  | 2.8  | 3.5  | 4.4  |
| 10   | 233.8 | 201.7 | 179.8 | 3.2  | 3.5  | 3.8  | 3.2  | 3.5  | 4.6  |
| avg  | 312.09 | 292.36 | 213.69 | 3.34 | 3.37 | 4.01 | 3.10 | 3.33 | 4.16 |
| std  | 106.97 | 71.73 | 53.77 | 0.21 | 0.13 | 0.33 | 0.63 | 0.16 | 0.25 |

5.5. Analysis

We execute two analysis on the FID values and user survey scores: t-test and effect size.

5.5.1. t-Test

We execute t-test on two pairs of the stylized images: (i) foreground and ours and (ii) background and ours. For the FID values, the p values record 0.022 and 0.012, respectively. This denotes that the differences are significant for $p < 0.05$. For the user survey scores, $p$ values score $3.7 \times 10^{-5}$ and $9.7 \times 10^{-5}$ for the stylization and $0.00033$ and $7.7 \times 10^{-8}$ for the preservation of content. These denote that the differences are significant for $p < 0.01$. These values are presented in the upper row of Table 3.

5.5.2. Effect Size

We estimate effect size on two pairs of the stylized images: (i) foreground and ours and (ii) background and ours. For the FID values, the Cohen's d values record 1.64 and 1.76, respectively. This denotes that the effect sizes between two groups are large. For the user survey scores, Cohen's d values are 3.43 and 3.61 for the stylization and 3.14 and 5.48 for the preservation of content. From these values, we conclude that the effect sizes between two groups are large.
Figure 15. The images used in FID estimation and user study. Images in green box are rendered in pencil style, images in blue are in abstraction, and images in red are in watercolor.
Table 3. Analysis of the evaluation results.

| Evaluation 1: FID | Evaluation 2: User Survey |
|------------------|---------------------------|
|                  | FID | Q1: Stylization | Q2: Preservation of Content |
| p value          |     |                |                             |
| F/G and ours     | 0.022 | 3.7 \times 10^{-5} | 0.00033                     |
| B/G and ours     | 0.012 | 9.7 \times 10^{-5} | 7.7 \times 10^{-8}          |
| Cohen's d        |     |                |                             |
| F/G and ours     | 1.64 | 3.43           | 3.14                        |
| B/G and ours     | 1.76 | 3.61           | 5.48                        |

5.6. Discussion

The existing stylization studies concentrated on applying styles while preserving content or synthesizing various styles captured from various samples. On the other hand, our technique differs from the existing research in that it can segment the image into foreground and background and apply different artistic styles to the segmented regions. This approach, which has not been tried by the existing studies, can be applied to various images as a way to express the images more dramatically by applying different artistic styles on them. This approach can be extended to develop a method that applies different styles on various regions segmented from an image.

6. Conclusions and Future Work

This study proposes a framework for applying two different art effects to a photograph. Most of deep learning-based models require a paired dataset to generate distinct art effects. To resolve this requirement, we applied existing artistic media simulation techniques that produce color pencils, watercolors, and abstraction effects to build the dataset composed of a photograph and its artistic images. By training our model with this dataset, our model successfully produces visual pleasing artistic media effects. Furthermore, in order to apply the two different artistic effects, an input photograph is segmented into foreground and background using a semantic segmentation model. By applying a series of SPADE blocks that apply region information to style transfer, our model successfully produce different artistic media effects in the foreground and background of a photograph.

In future work, we plan to improve the generative model to express more diverse artistic media effects. Furthermore, we will enrich our styles by segmenting an image into more detailed regions and applying proper styles to the regions. Finally, we will extend our dataset of artistic styles by applying more existing artistic style synthesis schemes.

Author Contributions: Conceptualization, K.M.; methodology, H.Y.; software, H.Y.; validation, H.Y. and K.M.; formal analysis, K.M.; investigation, H.Y.; resources, H.Y.; data curation, H.Y.; writing—original draft preparation, H.Y.; writing—review and editing, K.M.; visualization, K.M.; supervision, K.M.; project administration, K.M.; funding acquisition, H.Y. All authors have read and agreed to the published version of the manuscript.

Funding: This research was supported by Sangmyung University Research Fund 2020.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Kyprianidis, J.; Kang, H. Image and video abstraction by coherence-enhancing filtering. *Comput. Graph. Forum* 2011, 30, 593–602.[CrossRef]
2. Bousseau, A.; Neyret, F.; Thollot, J.; Salesin, D. Video watercolorization using bidirectional texture advection. *ACM Trans. Graph.* 2007, 26, 104:1–104:10.[CrossRef]
3. Platkevic, A.; Curtis, C.; Sykora, D. Fluidymation: Stylizing Animations Using Natural Dynamics of Artistic Media. *Comput. Graph. Forum* 2022, 31, 1471–1480. [CrossRef]

4. Gatys, L.; Ecker, A.; Bethge, M. Image style transfer using convolutional neural networks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition 2016, Las Vegas, NV, USA, 30–27 June 2016; pp. 2411–2423.

5. Isola, P.; Zhu, J.; Zhou, T.; Efros, A. Image-to-image translation with conditional adversarial networks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition 2017, Honolulu, HI, USA, 21–26 June 2017; pp. 1125–1134.

6. Zhu, J.; Park, T.; Isola, P.; Efros, A. Unpaired image-to-image translation using cycle-consistent adversarial networks. In Proceedings of the IEEE International Conference on Computer Vision 2017, Honolulu, HI, USA, 21–26 June 2017; pp. 2223–2232.

7. Park, T.; Liu, M.; Wang, T.; Zhu, J. Semantic image synthesis with spatially-adaptive normalization. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition 2019, Long Beach, CA, USA, 15–20 June 2019; pp. 2337–2346.

8. Kang, H.; Lee, S.; Chui, C. Flow-based image abstraction. *IEEE Trans. Vis. Comput. Graph.* 2009, 15, 62–76. [CrossRef] [PubMed]

9. Champandard, A. Semantic style transfer and turning two-bit doodles into fine artworks. *arXiv 2016*, arXiv:1603.01768

10. Huang, X.; Belongie, S. Arbitrary style transfer in real-time with adaptive instance normalization. In Proceedings of the IEEE International Conference on Computer Vision 2017, Honolulu, HI, USA, 21–26 June 2017; pp. 1501–1510.

11. Li, Y.; Fang, C.; Yang, J.; Wang, Z.; Lu, X.; Yang, M. Universal style transfer via feature transforms. In Proceedings of the Neural Information Processing Systems 2017, Long Beach, CA, USA, 4–9 December 2017; pp. 385–395.

12. Castillo, C.; De, S.; Han, X.; Singh, B.; Yadav, A.; Goldstein, T. Son of zorn’s lemma: Targeted style transfer using instance-aware semantic segmentation. In Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing 2017, New Orleans, LA, USA, 5–9 March 2017; pp. 1348–1352.

13. DeCarlo, D.; Santella, A. Stylization and abstraction of photographs. In Proceedings of the ACM Computer Graphics and Interactive Techniques 2002, San Antonio, TX, USA, 21–26 June 2002; pp. 769–776.

14. Winnemoller, H.; Olsen, S.; Goock, B. Real-time video abstraction. In Proceedings of the ACM Computer Graphics and Interactive Techniques 2006, Boston, MA, USA, 30 July–3 August 2006; pp. 1221–1226.

15. Kang, H.; Lee, S. Shape-simplifying image abstraction. *Comput. Graph. Forum* 2008, 27, 1773–1780. [CrossRef]

16. Kyprianidis, J.; Kang, H.; Dollner, J. Image and video abstraction by anisotropic Kuwahara filtering. *Comput. Graph. Forum* 2009, 28, 1955–1963. [CrossRef]

17. DeCarlo, D.; Finkelstein, A.; Rusinkiewicz, S.; Santella, A. Suggestive contours for conveying shape. *ACM Trans. Graph.* 2003, 22, 848–855. [CrossRef]

18. Kang, H.; Lee, S.; Chui, C. Coherent line drawing. In Proceedings of the 5th Non-Photorealistic Animation and Rendering Symposium, San Diego, CA, USA, 4–5 August 2007; pp. 43–50.

19. Winnemoller, H.; Kyprianidis, J.; Olsen, S. XDoG: An eXtended difference-of-Gaussians compendium including advanced image stylization. *Comput. Graph.* 2012, 36, 740–753. [CrossRef]

20. Curtis, C.; Anderson, S.; Seims, J.; Fleischer, K.; Salesin, D. Computer-generated watercolor. In Proceedings of the ACM Computer Graphics and Interactive Techniques 1997, Los Angeles, CA, USA, 3–8 August 1997; pp. 421–430.

21. Bousseau, A.; Kaplan, M.; Thollot, J.; Sillion, F. Interactive watercolor rendering with temporal coherence and abstraction. In Proceedings of the Non-Photorealistic Animation and Rendering Symposium, Annecy, France, 5–7 June 2006; pp. 141–149.

22. Kang, H.; Chui, C.; Chakraborty, U. A unified scheme for adaptive stroke-based rendering. *Vis. Comput.* 2006, 22, 814–824. [CrossRef]

23. van Laerhoven, T.; Lisenborgs, J.; van Reeth, F. Real-time watercolor painting on a distributed paper model. In Proceedings of the Computer Graphics International 2004, Crete, Greece, 16–19 June 2004; pp. 640–643.

24. Sousa, M.; Buchanan, J. Computer-generated graphite pencil rendering of 3D polygonal models. *Comput. Graph. Forum* 1999, 18, 195–208. [CrossRef]

25. Matsui, H.; Joha, H.; Nishita, T. Creating colored pencil style images by drawing strokes based on boundaries of regions. In Proceedings of the Computer Graphics International 2005, Stony Brook, NY, USA, 22–24 June 2005; pp. 148–155.

26. Murakami, K.; Tsuruno, R.; Genda, E. Multiple illuminated paper textures for drawing strok. In Proceedings of the Computer Graphics International 2005, Stony Brook, NY, USA, 22–24 June 2005; pp. 156–161.

27. Kwon, Y.; Yang, H.; Min, K. Pencil rendering on 3D meshes using convolution. *Comput. Graph.* 2012, 36, 930–944. [CrossRef]

28. Goodfellow, I.; Pouget-Abadie, J.; Mirza, M.; Xu, B.; Warde-Farley, D.; Ozair, S.; Bengio, Y. Generative adversarial nets. In Proceedings of the Advances in Neural Information Processing Systems 2014, Montreal, QC, Canada, 8–13 December 2014; pp. 2672–2680.

29. Radford, A.; Metz, L.; Chintala, S. Unsupervised representation learning with deep convolutional generative adversarial networks. In Proceedings of the International Conference on Learning Representations 2016, San Diego, CA, USA, 2–4 May 2016; pp. 1–16.

30. Chen, H.; Tak, U. Image Colored-Pencil-Style Rendering Based on Generative Adversarial Network. In Proceedings of the International Conference on Wavelet Analysis and Pattern Recognition, Adelaide, Australia, 2 December 2020; pp. 90–95.

31. Zhou, H.; Zhou, C.; Wang, X. Pencil Drawing Generation Algorithm Based on GMD. *IEEE Access* 2021, 9, 41275–41282. [CrossRef]

32. Kim, J.; Kim, M.; Kang, H.; Lee, K. U-gat-it: Unsupervised generative attentional networks with adaptive layer-instance normalization for image-to-image translation. In Proceedings of the ICLR 2020, Online, 26 April–1 May 2020.

33. Platkevic, A.; Curtis, C.; Sykora, D. Fluidymation: Stylizing Animations Using Natural Dynamics of Artistic Media. *Comput. Graph. Forum* 2021, 40, 21–32. [CrossRef]
34. Sochorova, S.; Jamriska, O. Practical pigment mixing for digital painting. *ACM Trans. Graph.* 2021, 40, 234. [CrossRef]
35. Gatys, L.; Ecker, A.; Bethge, M.; Hertzmann, A.; Shechtman, E. Controlling perceptual factors in neural style transfer. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition 2017, Honolulu, HI, USA, 21–26 July 2017; pp. 3985–3993.
36. Kotovenko, D.; Sanakoyeu, A.; Lang, S.; Ommer, B. Content and Style Disentanglement for Artistic Style Transfer. In Proceedings of the ICCV 2019, Seoul, Korea, 27 October–2 November 2019; pp. 4422–4431.
37. Kotovenko, D.; Sanakoyeu, A.; Ma, P.; Lang, S.; Ommer, B. A Content Transformation Block For Image Style Transfer. In Proceedings of the CVPR 2019, Long Beach, CA, USA, 16–20 June 2019; pp. 10032–10041.
38. Svozil, J.; Annosheh, A.; Osendorf, C.; Malis, J. Two-Stage Peer-Regularized Feature Recombination for Arbitrary Image Style Transfer. In Proceedings of the CVPR 2020, Online, 14–19 June 2020; pp. 13816–13825.
39. Sanakoyeu, A.; Kotovenko, D.; Lang, S.; Ommer, B. A Style-Aware Content Loss for Real-time HD Style Transfer. In Proceedings of the ECCV 2018, Munich, Germany, 8–14 September 2018; pp. 698–714.
40. Chen, H.; Zhao, L.; Wang, Z.; Zhang, H.; Zuo, Z.; Li, A.; Xing, W.; Lu, D. DualAst: Dual Style-Learning Networks for Artistic Style Transfer. In Proceedings of the CVPR 2021, Online, 19–25 June 2021; pp. 872–881.
41. Lim, L.; Keles, H. Foreground segmentation using a triplet convolutional neural network for multiscale feature encoding. *Pattern Recognit. Lett.* 2018, 112, 256–262. [CrossRef]
42. Wang, X.; Juang, J.; Chan, S. Automatic foreground extraction from imperfect backgrounds using multi-agent consensus equilibrium. *J. Vis. Commun. Image Represent.* 2020, 72, 102907. [CrossRef]
43. Tezcan, M.; Ishwar, P.; Konrad, J. BSUV-Net: A Fully-convolutional neural network for background subtraction of unseen videos. In Proceedings of the WACV 2020, Snowmass Village, CO, USA, 2–5 March 2020; pp. 2774–2783.
44. Bouwmans, T.; Javed, S.; Sultana, M.; Jung, S. Deep neural network concepts for background subtraction: A systematic review and comparative evaluation. *Neural Netw.* 2019, 117, 8–66. [CrossRef] [PubMed]
45. Yu, C.; Wang, J.; Peng, C.; Gao, C.; Yu, G.; Sang, N. BiSeNet: Bilateral segmentation network for real-time semantic segmentation. In Proceedings of the ECCV 2018, Munich, Germany, 8–14 September 2018; pp. 325–341.