Learning-based Artificial Intelligence Artwork: Methodology Taxonomy and Quality Evaluation

QIAN WANG, Guangdong University of Technology, Guangzhou, China
HONG-NINGH DAI, Hong Kong Baptist University, Hong Kong, Hong Kong
JINGHUA YANG, Southwest Jiaotong University, Chengdu, China
CAI GUO, Hanshan Normal University, Chaozhou, China
PETER CHILDS, Imperial College London, London, United Kingdom of Great Britain and Northern Ireland
MAAIKE KLEINSMANN, Delft University of Technology, Delft, Netherlands
YIKE GUO*, The Hong Kong University of Science and Technology, Hong Kong, Hong Kong
PAN WANG*, Delft University of Technology, Delft, Netherlands

With the development of the theory and technology of computer science, machine or computer painting is increasingly being explored in the creation of art. Machine-made works are referred to as artificial intelligence (AI) artworks. Early methods of AI artwork generation have been classified as non-photorealistic rendering (NPR) and, latterly, neural-style transfer methods have also been investigated. As technology advances, the variety of machine-generated artworks and the methods used to create them have proliferated. However, there is no unified and comprehensive system to classify and evaluate these works. To date, no work has generalised methods of creating AI artwork including learning-based methods for painting or drawing. Moreover, the taxonomy, evaluation and development of AI artwork methods face many challenges. This paper is motivated by these considerations. We first investigate current learning-based methods for making AI artworks and classify the methods according to art styles. Furthermore, we propose a consistent evaluation system for AI artworks and conduct a user study to evaluate the proposed system on different AI artworks. This evaluation system uses six criteria: beauty, color, texture, content detail, line and style. The user study demonstrates that the six-dimensional evaluation index is effective for different types of AI artworks.

CCS Concepts: • **General and reference** → **Surveys and overviews**.

Additional Key Words and Phrases: AI Art, Artwork, Style Transform, Painting, Methodology Taxonomy, Quality Evaluation

1 Introduction

In the late 19th century, the emergence of photographic technology stimulated artistic diversity. In the early 1990s, the successes of photorealistic computer graphics encouraged alternative techniques for non-photorealistic styles of rendering [81, 82, 124, 142]. Recently, creation of computer artworks has become popular along with

*Both authors are the Corresponding Authors.

Authors' Contact Information: Qian Wang, Guangdong University of Technology, Guangzhou, Guangdong, China; e-mail: wangqian@gdut.edu.cn; Hong-NingH Dai, Hong Kong Baptist University, Hong Kong, Hong Kong; e-mail: hndai@ieee.org; Jinghua Yang, Southwest Jiaotong University, Chengdu, Sichuan, China; e-mail: yangjinghua110@126.com; Cai Guo, Hanshan Normal University, Chaozhou, Guangdong, China; e-mail: c.guo@hstc.edu.cn; Peter Childs, Imperial College London, London, United Kingdom of Great Britain and Northern Ireland; e-mail: p.childs@imperial.ac.uk; Maaike Kleinsmann, Delft University of Technology, Delft, Zuid-Holland, Netherlands; e-mail: m.s.kleinsmann@tudelft.nl; Yike Guo, The Hong Kong University of Science and Technology, Hong Kong, Hong Kong; e-mail: yikeguo@ust.hk; Pan Wang, Delft University of Technology, Delft, Zuid-Holland, Netherlands; e-mail: p.wang-2@tudelft.nl.

ACM 1557-7341/2024/10-ART
https://doi.org/10.1145/3698105

related research studies, and new advances in machine learning and deep learning have led to an acceleration in the development of AI artworks [12]. In this review, we consider state-of-the-art methods in AI artworks, i.e. non-photorealistic creative drawings or paintings generated by AI models.

Many artists and computer researchers have used technologies and methodologies for automatically transforming images into synthetic artworks. Since the 1990s, stroke-based rendering (SBR) methods first proposed by [48] have become popular in computer-generated artwork. In 2003, Hertzmann reviewed SBR algorithms and art styles of machine paintings [54]. Although diverse SBR methods offer many types of art style for synthesised artworks, these methods require significant use of computer memory and are time-consuming. With the development of machine learning and reinforcement learning, methods and technologies addressing AI artworks optimise these issues. In 2013, the authors of [81] reviewed technologies and methods of non-photorealistic rendering (NPR) that transferred input photographic images or videos into non-photorealistic stylised results. Latterly, the authors of [70] investigated neural-style transfer (NST) methods that belong to the field of NPR. Their work extended the review of NPR based on the work of [81]. However, to date, no work has generalised the methods of creating AI artwork including learning-based methods for painting or drawing. Moreover, the evaluation of AI artwork methods is not systematic. Researchers have tended to use their own evaluation methods to compare their own work with prior works. However, a reasonable and consistent evaluation system is important for fair comparison of the differing methods of generating AI artworks. Although [70] summarised the current approaches to evaluating NPR artworks, most evaluation approaches are not suited to different algorithms. It is necessary to develop a consistent evaluation system for diverse styles of AI artwork.

To solve the above problems, we investigate current learning-based methods for AI artworks and classify these methods according to different art styles. Furthermore, inspired by art vocabulary [134] and the representation of art paintings [22], we propose a consistent evaluation system for AI artworks and conduct a user study to evaluate the adaptability of the evaluation system. The proposed evaluation system contains six criteria: beauty, color, texture, content detail, line and style. In particular, since beauty [107] is a dominant factor in the judgement of artwork by humans, we set a weighting of 50% of the score for beauty, and the other five aspects account for 10% each, respectively. The results of the user study indicate that the proposed evaluation system is effective for different types of artworks, and the score distribution also demonstrates that the percentage setting is reasonable. Based on the analysis of the current methods and experiments on the evaluation system, we propose and analyse challenges and opportunities for AI artworks as well as areas of possible development.

We summarise the contributions of this survey as follows:

- We investigate recent works on existing AI artworks and classified these according to different art types to produce a clear taxonomy and consistent evaluation.
- We propose a unified evaluation system for different AI artworks to ensure fair comparison of different AI models.
- We analyse challenges and opportunities for the development of AI artworks.

The paper takes into consideration methods, art styles, and the evaluation system. To ensure the comprehensiveness and reliability of the literature review, we collected relevant literature from multiple databases, including Google Scholar, IEEE Xplore, ACM Digital Library, and ArXiv. Our keywords included"artificial intelligence art", "deep learning", "generative adversarial networks (GAN)", "diffusion model", "computer vision", "creative generation", "line drawing", "oil painting", and "stroke". The search range was limited to publications from 2015 to 2024 to provide bounding of information but also ensure the timeliness and relevance of the literature. The initial search yielded about 2500 papers, and an additional 50 papers were identified from other sources. After removing duplicate entries, we screened 600 papers. By reading the titles and abstracts, we excluded 300 less relevant papers, leaving 300 papers. We then conducted a full-text review of these remaining papers and excluded

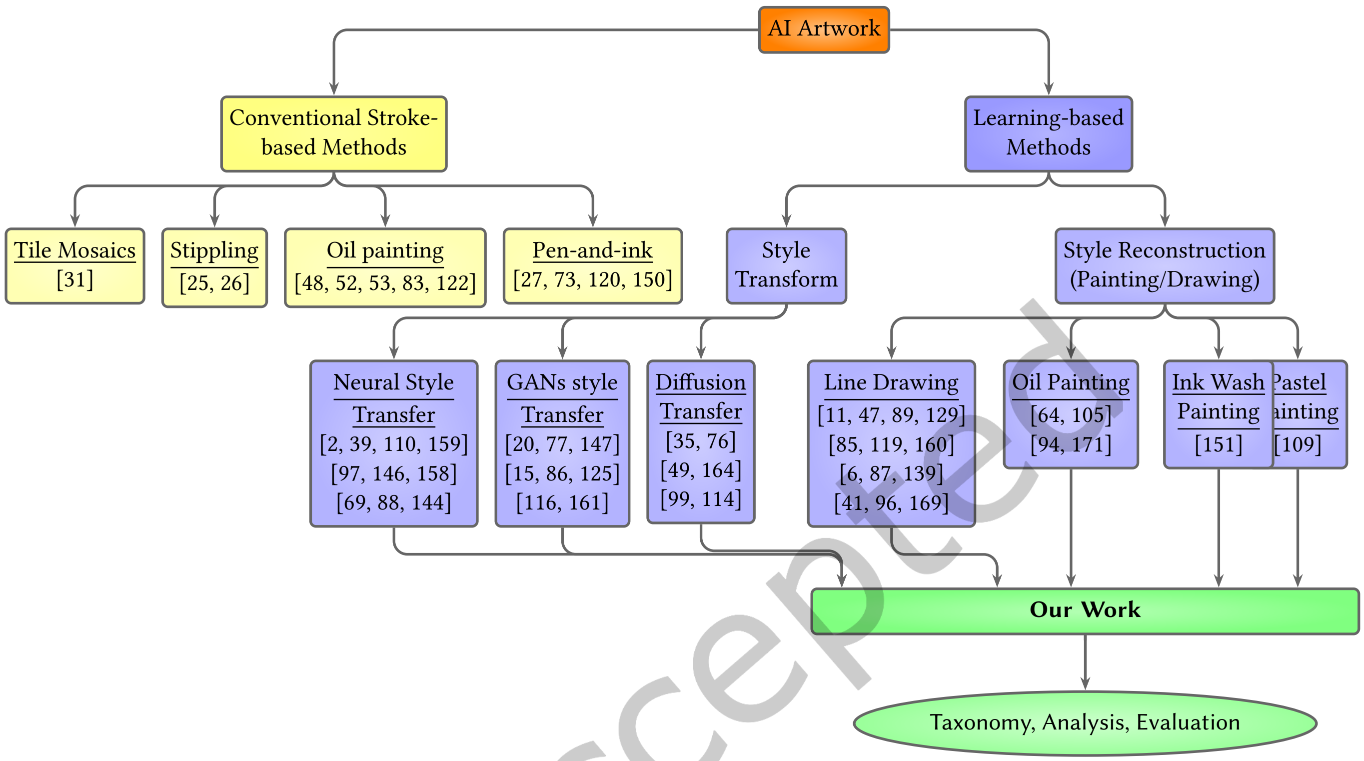

Fig. 1. Taxonomy of AI artwork based on methods and art styles.

100 that did not meet the inclusion criteria. Ultimately, we selected 200 highly-relevant papers as the basis for this study.

As Fig. 1 shows, AI artworks are classified into two preliminary categories based on the method used: conventional stroke-based methods and learning-based methods. Since conventional stroke-based methods have been extensively investigated and we mainly focus on learning-based methods, we only discuss conventional stroke-based methods briefly, in Section 2. We further categorise learning-based methods into style transformation and style reconstruction (painting/drawing) based on the way the style is produced. In each category, the number of references is extensive. Due to space constraints, we have selected only a subset to represent each category. Section 3 introduces the concepts and related methodologies of learning-based AI artworks. As stated in Section 3, we categorise and analyse current research on AI artwork based on neural networks in Section 4. Section 5 presents the resultant evaluation system for AI artworks and the experimental results to test the system on different methods. We aim to build a standardised, comprehensive evaluation system in follow-up studies. This evaluation system is able to evaluate various types of AI artworks adaptively. In Section 6, we analyse the opportunities and challenges of AI artworks while pointing out possible ways to address them in the not-too-distant future. Finally, we present the conclusions of this paper in Section 7 and proposed several worthy issues for future research. For a further discussion, we make a supplementary to discuses the application of AI Art and the Ethics and Artistic Integrity for AI Art.

2 Conventional stroke-based AI Artworks

Conventional SBR methods mainly reconstruct images into non-photorealistic imagery with stroke-based models. Researchers have proposed many SBR methods adapted to different types of artwork, e.g. paintings [48, 52, 53, 83, 122], pen-and-ink drawings [27, 36, 148, 150] and stippling drawings [25, 26]. The work [48] introduced a semi-automatic painting method based on a greedy algorithm commonly used for SBR. This work shows that different stroke shapes and stroke sizes can be used to draw paintings with different styles; however, this method needs substantial human intervention to control the stroke shapes and select the stroke location. The authors of [52] also proposed a style design for their painting method by using spline-brush strokes to *draw* the image. They used a set of parameters to define the style of the brush strokes. The painting effects can be changed when the parameters are altered by the designer (user). Thus, this method requires users to have a high level of drawing skill. The work of [83] proposed a method to segment an image into areas with similar levels of salience to control the brushstrokes. The detail level of brushstrokes in the salient area can be increased to improve the realism of painterly rendering, though users are also required to control the number of levels. Other researchers also proposed pen-and-ink drawing and stippling drawing methods [25–27, 36, 148, 150] to improve the drawing effect. Most of these methods decompose strokes utilizing a greedy algorithm [54] into steps and require substantial human intervention.

Most SBR methods are relatively slow, so their usability is limited, especially in interactive applications [54]. It is also difficult for inexperienced or unskilled users to choose key parameters in SBR methods to produce satisfying paintings. Moreover, SBR methods can generate a limited number of styles, making them inflexible.

3 Learning-based AI Artworks

Learning-based AI artworks are non-photorealistic images reconstructed by deep neural networks. We classify learning-based AI artworks into two categories: end-to-end image reconstruction by style-transform models and drawing/painting with digital strokes by art-style-reconstruction models.

3.1 Style-transform AI Artworks

Style-transform methods mainly focus on reconstructing an image into another visual style according to a reference style image or a style image dataset. Image neural style transfer (NST) methods take a content image and a style image as the input and then output a stylised result containing the content features of the content image; the visual representation of this stylised result looks like the style image. Most GAN-based methods transform the input image into another style image according to the style of the training dataset. The output image contains its own content and presents the visual style in the same style as the dataset.

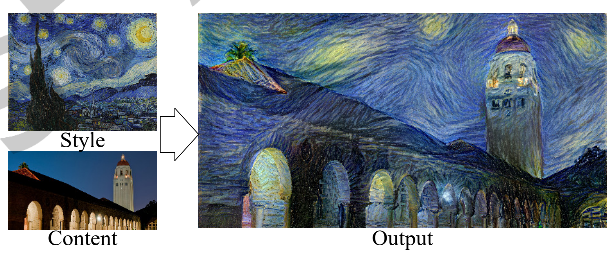

Fig. 2. Sample of results generated by the neural style transfer method [38].

3.1.1 Neural Style Transfer. Neural Style Transfer (NST) is a prototypical style-transform AI artwork method. Fig. 2 shows an NST result generated by [38]. NST works in an image-to-image manner, extracting texture features from a style image and content features from a content image, then fusing them to synthesize a new image. Modeling the style image and extracting its texture features is crucial. The goal is to reconstruct an image with the style textures from the style image while preserving the content of the content image.

The NST method, introduced in [38], uses CNNs to transfer style texture to a target image while resolving its content. The Gram matrix models the style image's representation, and the pre-trained VGG network's high-level features represent the content image. By minimizing content and style losses, the method synthesizes an image with both input images' content and style. However, this style representation focuses on texture rather than global arrangement, resulting in unsatisfactory results for long-range symmetric structures. The work [5] improved this by imposing a Markov structure on high-level features. [69]'s StrokePyramid module considers receptive field and scale, producing variant stroke sizes.

NST-generated images often have hard style features, making them appear unnatural. Careful selection of input-style images is essential to avoid unattractive results.

3.1.2 GAN-based Style Transfer. GANs, introduced by [42], have been widely applied in various research fields. GANs consist of a generator and discriminator, trained in an adversarial manner. The generator learns to produce realistic images, while the discriminator aims to distinguish between real and generated images. This minimax optimization process ends at a saddle point, balancing the two networks. GANs generate visually compelling fake images, blending authenticity with novelty.

GAN-based methods have revolutionized AI art, with notable applications like CycleGAN [170], AttentionGAN [132], and Gated-GAN [14]. These models learn the style features from datasets, transforming real photos into artistic styles without harsh style features. However, GAN-based methods have their drawbacks: the difficulty of training, large model size, sometimes poor detailed representation, and even mistakes.

3.1.3 Diffusion-model Style Transfer. Diffusion model style transfer represents a major breakthrough in Artificial Intelligence Generated Content (AIGC). It harnesses the power of Diffusion Models, which transform random noise into novel data samples through a unique stochastic diffusion process. This technology has fueled the rise of AI drawing platforms like OpenAI's DALL·E 2 [84, 111] and Google's Imagen [118], showcasing their remarkable image generation capabilities. In style transfer, diffusion models apply their generative prowess to imagery, enabling the seamless transformation of any input image into a specified artistic style. Their working mechanism seamlessly integrates noising and denoising processes, gradually degrading and then reconstructing the image with the desired style while preserving its original content.

This approach not only offers exceptional controllability, allowing users to fine-tune generated images with precision, but also guarantees diversity and flexibility. It effortlessly accommodates a wide spectrum of style requirements and reference images, yielding results ranging from photorealistic fakes [8, 49, 113, 118] to artistic interpretations [35, 49, 76, 99, 114, 164]. Furthermore, diffusion models exhibit remarkable stability and robustness, consistently producing high-quality stylized images even under noisy or varying input conditions. This reliability has sparked interest in research exploring partial image re-editing [51, 80], further underscoring the versatility of this technology.

3.2 Art-style-reconstruction AI Artworks

In this paper, we refer to art-style-reconstruction AI artworks as those images that are generated via simulated strokes. Note that the art style is neither transferred from the style image nor learned from the dataset: it is determined by the elements rendered onto the canvas. Therefore, when the models use different strokes to render the canvas, the generated image presents a different style. We first propose the concept of art-style-reconstruction AI artworks for these methods. It is important to recognise the difference between style-transform methods and style-reconstruction methods for AI artworks. Style-transform methods do not consider the generating process of the result, while style-reconstruction methods with simulated strokes pay significant attention to the generating process, since the result is built by strokes. For fairness, methods in these different categories should be evaluated by different evaluation metrics. According to the types of style, we classify art-style-reconstruction AI artworks into line drawings, oil paintings and watercolor paintings, and ink wash paintings.

3.2.1 Line Drawing. Line-drawing artworks such as sketches [6, 11, 47, 85, 89, 119, 129, 160], pencil drawings [87], portraits [96, 139] and doodles [105, 169] are created by line strokes. Significant research has been undertaken on line-drawing methods. Many studies have concerned the generation of line-drawing artworks by reconstructing input photos into line drawings. Compared with the input photos, generated line drawings lose much detailed content but retain the key contour of the object. Photo-sketch methods are mainly focused on the approach for capturing the contour information of an object in a photo, then mimicking the human sketching process to present the object. We usually consider photo-to-sketch synthesis as a cross-domain reconstruction issue. For example, the work of [129] constructed a generative sequence model with a recurrent neural network (RNN) acting as a neural sketcher. Their neural sketcher reconstructed a photo into a synthesis sketch by learning the noisy photo-sketch pairs dataset. Many methods for reconstructing photos into line drawings have been proposed. Line-drawing methods emphasise extracting the edge features of the object but not paying attention to the image's color information. In particular, when comparing methods of line drawings, the key point is the line stroke or the shade drawn by line strokes. Portraits and pencil drawings (except with colored pencils) similar to sketches usually have black-and-white color characteristics.

3.2.2 Oil Painting and Watercolor Painting. Painting is an important form of visual art. Oil painting and watercolor painting, distinct from line drawings, emphasise color and tone. The essence of painting is color, which is made up of hue, saturation, and value, dispersed over a surface. In generating oil paintings and watercolor paintings, mimicing the color and stroke texture of paintings is a main task for the reconstruction of image-to-painting. With deep learning coming into widespread use, researchers have conducted studies on training machines to learn to paint like human artists. In particular, the work [105] proposed a neural network SPIRAL++ to doodle human portraits. The style of the generated image is close to that of an oil painting, although the results lose detailed content. The work by [68] proposed a self-supervised learning algorithm to achieve painting stroke by stroke, and the results outperformed SPIRAL++ on the presentation of details, although the detailed contents were still not sharp. The authors of [64] designed a painting model based on reinforcement learning (RL) to mimic the painting process of a human artist. The color strokes rendered onto the digital canvas in a certain order made their generated images similar to oil paintings, although the texture of the strokes was different from human artists' strokes. The work [171] proposed an automatic image-to-painting model that generates oil paintings with controllable brushstrokes. The authors re-framed the stroke prediction as a parameter searching process so that it mimicked the human painting process. The authors of [123] also proposed a model using content masked loss to generate paintings stroke by stroke, although they lost some detailed contents of the image. For the stroke-based methods, the key point is how to present the detailed contents of the input image when reconstructing it to the painting stroke by stroke. The problem is that retaining as many details as possible will produce to a close-to-photo result instead of a painting.

3.2.3 Ink Wash Painting. Ink wash painting is a type of Chinese ink brush painting that uses black or colored ink in different concentrations. The stroke texture and character of ink wash painting are so different from that of oil painting and watercolor painting that teaching a machine or computer to do ink wash painting is difficult. Research has been conducted on methods to simulate the special stroke of ink wash painting; for example, in a conventional stroke-based method in [34], the authors used B-spline curves to simulate the trajectory of the Chinese brush. This method inspired later researchers to improve the simulation of Chinese brushstrokes for deep neural networks. The authors of [151] first modelled the tip of the Chinese brush and then utilised a reinforcement learning algorithm to formulate the automatic stroke generator.

3.2.4 Robtic Painting. Robotic painting, an intersection of art and robotics, has seen significant advancements. Researchers and interdisciplinary artists have employed various painting techniques and human-machine collaboration models to create visual media on canvas. While robot paintings differ from the AI artworks discussed in this paper, they share some similarities. Robotic painting requires the use of physical robotic arms or robots to complete stroke-by-stroke painting, ultimately resulting in physical paintings. However, the AI paintings discussed in this article are almost exclusively electronic versions, and do not require the use of robotic arms or robots. Their similarity lies in the stroke-by-stroke painting algorithm, as most AI models for stroke-by-stroke painting, after processing, can be applied to robotic painting. Nevertheless, since the focus of this paper is not an in-depth exploration of algorithms, in section 4.4.5, we conduct a more comprehensive analysis and discussion on robotic painting.

4 Methods comparison

For different types of AI artworks, we have classified existing research into several categories based on artistic types. Correspondingly, we propose an algorithm taxonomy according to the different types of AI artwork. We first classify AI artworks into two categories according to the generating process mentioned in Section 3. This section explains the algorithms of different methods for different types of AI artwork.

4.1 NST Method

DeepDream [1] first synthesized artistic pictures by reversing CNNs' representations with image-style fusion through online image reconstruction techniques. This method aimed to improve the interpretability of deep CNNs by visualizing patterns that maximize neuron activation. Although producing a psychedelic and unrealistic style, it became popular for digital art. Subsequent methods [38–40, 46, 62, 63, 71, 88, 100, 101, 117] optimized digital art by combining visual-texture-modeling techniques with style transfer, inspiring the proposal of Neural Style Transfer (NST). The basic idea is to model and extract style and content features from input style and content images, respectively, then recombine them into a target image through iterative reconstruction to produce a stylized result with features of both images.

Generally, image-style fusion NST algorithms share the same image reconstruction theory but differ in techniques to model the visual style. For example, some methods [97, 146, 154, 157] adjust parameters to tune the style or content ratio, while others [9, 69, 79, 142, 158, 159] control stroke size to represent the stylized results. A common limitation is their computation-intensive nature due to the iterative image optimization procedure.

The classical NST algorithm by Gatys et al. [38] reconstructs representations from intermediate layers of the VGG-19 network, showing that CNN-extracted content and style representations are separable. The algorithm combines these features to synthesize a new image displaying both the style and content of the original images. The detailed algorithm is as follows:

Given a pair of images, the content image (I_c) and the style image (I_s), the algorithm of [38] synthesizes a target image (I_t) by minimizing the following function:

$$\widetilde{I} = \arg\min_{I_t} \alpha \mathcal{L}_c(I_c, I_t) + \beta \mathcal{L}_s(I_s, I_t), \tag{1}$$

where $\mathcal{L}_c$ is the content loss between the content image and the generated target image, and $\mathcal{L}_s$ is the style loss between the style image and the synthesized target image. The parameters α and β tune the ratio of content and style in the target image. While tuning α and β changes the visual expression of the result, it does not allow for detailed style texture adjustments.

Further methods proposed controlling model parameters to achieve different stylization outcomes. The authors of [142] introduced intuitive guidance and artistic control on style-transfer models by adjusting pattern density and stroke strength. Based on the style transfer concept of [38], this method also minimizes content loss and style loss, as shown in Eq. 1, but with a different style loss definition in Eq. 2c. In particular, Eq. 2a defines the centered Gram matrix, Eq. 2b is the style representation by Eq. 2a, and δ_l controls the importance of each network layer. X denotes the input, and $\varphi_{(l)}(X)$ denotes the feature activation from the VGG-19 network.

$$\underset{c}{Gram}(X) = \mathbb{E}[(X - \mathbb{E}[X])(X - \mathbb{E}[X])^T], \tag{2a}$$

$$f_s(X, l) = \underset{c}{Gram}\left(\varphi_{(l)}(X)\right), \tag{2b}$$

$$\mathcal{L}_s = \sum_l \delta_l ||f_s(I_t, l) - f_s(I_s, l)||_2^2. \tag{2c}$$

To control the visual effect of the stylized results, research has proposed using *stroke size*, *style scale*, or *pattern density* to control the artistic style in the synthesized image. These methods adjust the graininess of style feature representation to change the visual art effect. [142]'s *pattern density* controls stroke sizes, frequency, and graininess overall for the entire image through style resolution changes and variance-aware adaptive weighting. Pattern density is inversely proportional to image resolution size, and variance-aware adaptive weighting prioritizes dense pattern features to affect style representation. Additionally, [142] used *pattern density* and *stroke strength* together to control the art style, defining stroke strength as the salience of texture edges to tune without affecting other features.

While pattern density and stroke strength can adjust the visual performance of the stylized image, such as sharpening or lightening edge details, or zooming in or out on the style pattern grain, they cannot change the percentage of style or content features in the results. This highlights the need for more flexible methods that allow detailed adjustments of both style and content features.

4.2 GAN Method

4.2.1 Per-model-per-style. GAN is a min-max game between two neural networks with different objectives. One network, the generator (G), aims to trick the other, the discriminator (D), by generating images that resemble the dataset from a random latent vector z. The objective of G is to create images closer to the dataset, while D tries to distinguish between real and generated images. Both networks optimize their tasks according to their objective functions. The dataset image is denoted as x, and $D(x)$ represents the probability that x is from the dataset. $G(z)$ denotes the image generated by the generator, and the cost for G is $\log(1 - D(G(z)))$. The overall loss function is:

$$\mathcal{L}_{\text{GAN}} = V(\underset{\max}{D}, \underset{\min}{G}) = \mathbb{E}_{x \sim p_{data}(x)}[-\log(D(x))] + E_{z \sim p_z(z)}[\log(1 - D(G(z)))]. \tag{3}$$

The discriminator aims to maximize its ability to distinguish between real training data images and those generated by the generator. In the loss function 3, minimizing $-\log(D(x))$ equates to maximizing the discriminator's probability. The generator, on the other hand, minimizes $\log(1 - D(G(z)))$ to generate images that can trick the discriminator. Training a GAN, being a two-player adversarial game, is complex and challenging.

When Goodfellow et al. first proposed GANs, they were not capable of generating stylized images. As shown in Eq.(3), the generator aims to minimize its cost to produce images similar to the real data. Building on the GAN framework, researchers developed image-to-image translation methods[66, 130, 170] to achieve style transfer. CycleGAN, proposed by Zhu et al. [170], transforms photos into paintings that closely resemble the styles of various artists using unpaired data. This method maps a source image data domain S to a target image domain T, learning the mapping $G : S \to T$. It employs an adversarial loss to distinguish between the data distribution of T and the distribution of images generated by $G(S)$.

Since the mapping $G : S \to T$ lacks constraints, another generator $\widetilde{G}$ is introduced for the reverse mapping $\widetilde{G} : T \to S$ to ensure consistent results. Cycle consistency loss is added to enforce $\widetilde{G}(G(S)) \approx S$. When G translates an image from S to T, $\widetilde{G}$ should be able to translate it back to S, ensuring the reconstructed image $\widetilde{G}(G(S))$ closely matches the original image S. Similarly, for each image from T, the reverse should hold. For the mapping

$G : S \rightarrow T$ and its discriminator D_T, the objective function is:

$$\mathcal{L}_{\text{GAN}}(G, D_T, S, T) = \mathbb{E}_{t \sim p_{data}(t)}[\log D_T(t)] + \mathbb{E}_{s \sim p_{data}(s)}[\log(1 - D_T(G(s))]. \tag{4}$$

For each image s from the source image domain S, the image reconstruction cycle should be able to bring s back the original image, i.e. $s \rightarrow G(s) \rightarrow \widetilde{G}(G(s)) \approx s$. This gives the forward cycle consistency. On the other hand, for each image t from the target image domain T, G and $\widetilde{G}$ should also finish backward cycle consistency: $t \rightarrow \widetilde{G}(t) \rightarrow G(\widetilde{G}(t)) \approx t$. Therefore, we get the cycle consistency loss function written as follows:

$$\mathcal{L}_{\text{cyc}}(G, \widetilde{G}) = \mathbb{E}_{s \sim p_{data}(s)}[\|\widetilde{G}(G(s)) - s\|_1] + \mathbb{E}_{t \sim p_{data}(t)}[\|G(\widetilde{G}(t)) - t\|_1]. \tag{5}$$

The whole loss function of CycleGAN is:

$$\mathcal{L}(G, \widetilde{G}, D_S, D_T) = \mathcal{L}_{\text{GAN}}(G, D_T, S, T) + \mathcal{L}_{\text{GAN}}(\widetilde{G}, D_S, T, S) + \gamma \mathcal{L}_{\text{cyc}}(G, \widetilde{G}), \tag{6}$$

CycleGAN allows the generation of stylised images that contain both the content of input images and the style of the training dataset, controlled by γ. It enriches diverse art styles for unpaired image datasets, enabling reconstructions like transforming a modern photo into a Monet or Van Gogh painting. As shown in Fig. 3, CycleGAN's stylised results exhibit harmonious stylised characteristics, closely resembling Monet's style, compared to neural style transfer methods like AAMS [159], ASTSAN [110], and URUST [144], which contain varied features not truly reflective of Monet's style.

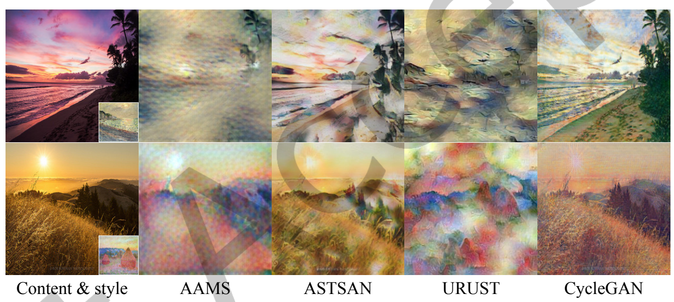

Fig. 3. Comparison of results: The first column displays content and style images. The last column shows CycleGAN's output, while the others present results from various neural style transfer methods.

CycleGAN has drawbacks, such as unclear detailed contents. To improve image quality, AttentionGAN [132] incorporates the attention mechanism [140] into CycleGAN. AttentionGAN redesigns the second generator $\widetilde{G}$ to generate content and attention masks, fusing them with the generated image $G(s)$ to restore the source image s. This process is formulated as: $\widetilde{G}(G(s)) = C_s * A_s + G(s) * (1 - A_s)$. The term of $\widetilde{G}$ consists of an encoder $\widetilde{G}_E$, an attention mask module $\widetilde{G}_A$, and a content mask module $\widetilde{G}_C$. $\widetilde{G}_C$ generates content masks, while $\widetilde{G}_A$ generates attention masks for both background and foreground. These masks are fused with $G(s)$ to restore s, formulated as: $\widetilde{G}(G(s)) = \sum_{f=1}^{n-1}(C_s^f * A_s^f) + G(s) * A_s^b$, where the reconstructed image $\widetilde{G}(G(s))$ should closely match the input source image s. Similarly, for a target image t, the cycle is formulated, and the reconstructed image should closely match t.

Fig. 4 compares CycleGAN and AttentionGAN. The first row shows real photos (small images) and subsequent rows display style-reconstructed results. AttentionGAN generates images with more detailed content than CycleGAN, especially in photo-to-Monet transformations, due to its attention mask mechanism. Different datasets yield distinct styles, enabling diverse AI artwork. For instance, training CycleGAN with a photo-to-anime dataset transforms real photos into anime images. CartoonGAN [15] and MS-CartoonGAN [125] focus on

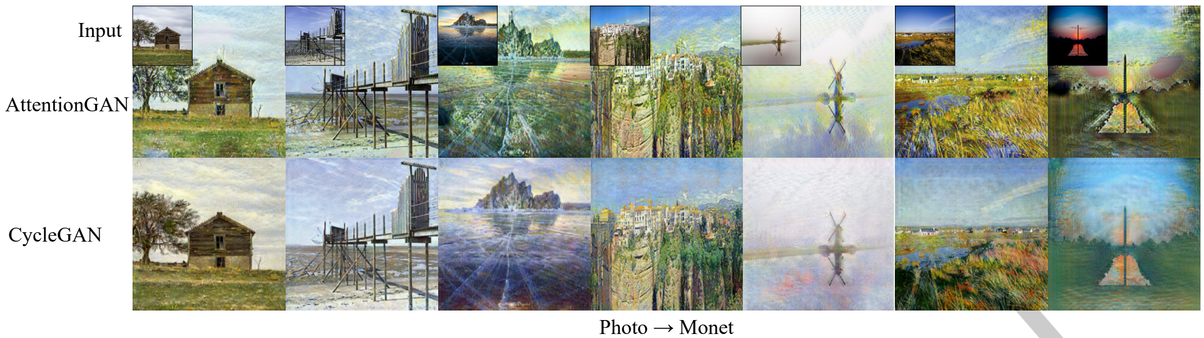

Fig. 4. Visual comparison between CycleGAN [170] and AttentionGAN [132].

reconstructing photos to anime, emphasizing sharp edges, smooth shading, and abstract textures. CartoonGAN's edge-promoting adversarial loss is given by:

$$\mathcal{L}_{\text{adv}}(G, D) = \mathbb{E}_{c_r \sim S_{\text{data}}(c_r)}\Big[\log D(c_r)\Big] + \mathbb{E}_{c_e \sim S_{\text{data}}(c_e)}\Big[\log\big(1 - D(c_e)\big)\Big] + \mathbb{E}_{P_I \sim S_{\text{data}}(P_I)}\Big[\log\Big(1 - D\big(G(P_I)\big)\Big)\Big]. \quad (7)$$

The discriminator D maximizes the probability of distinguishing the generated image $G(P_I)$, cartoon images without sharp edges, and real cartoon images. CartoonGAN also introduces a content loss for smooth shading:

$$\mathcal{L}_{\text{con}}(G, D) = \mathbb{E}_{P \sim S_{\text{data}(\text{P}_\text{I})}}[||VGG_l(G(P_I)) - VGG_l(P_I)||_1], \quad (8)$$

where l denotes a specific layer of VGG [126] for feature extraction. This loss uses ℓ_1 sparse regularization for better representation and regional characteristic preservation. While mimicking real art styles is crucial for AI artworks, diversity is also important. CycleGAN-based methods contribute to vivid art styles but generate only one style per model, which is inconvenient for diverse art style applications.

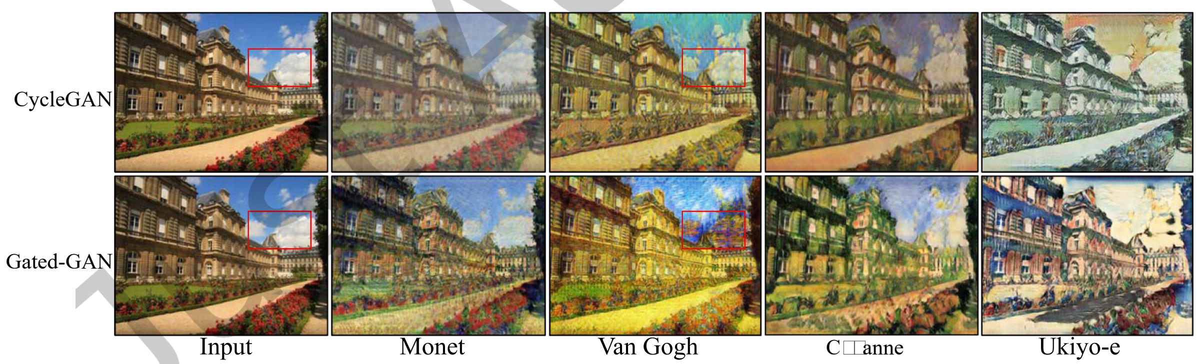

Fig. 5. Examples generated by Gated-GAN [14].

4.2.2 Per-model multi-style. Gated-GAN, proposed by [14], enables the generation of multiple styles within a single framework. It uses an adversarial gated network, known as the gated transformer, for multi-collection style transfer. The model includes a switching trigger to select the desired style for the output. The gated transformer processes a set of photos $\{p_i\}_{i=1}^{N} \in P$ and multiple painting collections $Q = \{Q_1, Q_2, ..., Q_K\}$, where K is the number of collections, each containing N_c images $\{q_i\}_{i=1}^{N_c}$. The network generates multiple styles $G(p, c)$ by applying the style of collection c to the input photo: $G(p, c) = Dec(T(Enc(p), c))$. Here, $T(.)$ is a transformer built with residual networks, and $Enc(p)$ denotes the encoded feature space. Each style-specific branch in the transformer module contains additional parameters, minimizing the overall model complexity. Inspired by LabelGAN [135], Gated-GAN incorporates an auxiliary classifier to handle multiple style categories, optimizing the entropy to improve classification confidence. This design enables the model to generate diverse styles within a unified framework.

Despite its ability to produce multiple styles, Gated-GAN has limitations, such as occasionally lacking detailed content. Fig. 5 shows examples generated by Gated-GAN, highlighting issues like the unnatural color block in the cloud region of the Van Gogh-styled image.

Gated-GAN's per-model multi-style approach contrasts with per-model-per-style methods like CycleGAN and CartoonGAN. While CycleGAN and CartoonGAN generate one style per model, Gated-GAN supports multiple styles, enhancing versatility. However, models like AttentionGAN, which builds on CycleGAN, tend to produce higher-quality images with more detailed content. Gated-GAN's strength lies in its ability to manage multiple styles efficiently, but it sometimes sacrifices detail. Combining the advantages of these approaches could lead to models that handle multiple styles and maintain high-quality, detailed outputs.

4.3 Diffusion Model Method

Early research on Diffusion Models (DM) began with deep unsupervised learning using nonequilibrium thermodynamics [128] in 2015. However, the key breakthrough came with Denoising Diffusion Probabilistic Models (DDPM) [58]. Unlike other models, DMs generate images by gradually "sampling" from Gaussian noise, forming images through a series of steps.

DMs consist of two processes: the forward (diffusion) process and the reverse (denoising) process, both parameterized as Markov chains. The forward process adds Gaussian noise to the input image I_0 over T steps, transforming it into pure Gaussian noise Y_T. The reverse process denoises this to generate realistic images.

For real data $\mathbf{y}_0 \sim q(\mathbf{y}_0)$, the forward process is: $q(\mathbf{y}_t|\mathbf{y}_{t-1}) = \mathcal{N}(\mathbf{y}_t; \sqrt{1-\beta_t}\mathbf{y}_{t-1}, \beta_t\mathbf{I})$, where β_t is the variance at each step. The reverse process generates data using parameterized Gaussian distributions:

$$\begin{cases} p_\theta(\mathbf{y}_0 : T) = p(\mathbf{y}_T)\prod_{t=1}^{T} p_\theta(\mathbf{y}_{t-1}|\mathbf{y}_t), \\ p_\theta(\mathbf{y}_{t-1}|\mathbf{y}_t) = \mathcal{N}(\mathbf{y}_{t-1}; \psi_\theta(\mathbf{y}_t, t), \pi_\theta(\mathbf{y}_t, t)), \end{cases} \tag{9}$$

where $p(\mathbf{y}_T) = \mathcal{N}(\mathbf{y}_T, \mathbf{0}, \mathbf{I})$, and $p_\theta(\mathbf{y}_{t-1}|\mathbf{y}_t)$ is the parameterised Gaussian distribution. The trained networks of $\psi_\theta(\mathbf{y}_t, t)$ and $\pi_\theta(\mathbf{y}_t, t)$ give the means and variances. The diffusion model is to obtain the trained networks for the final-generation model. The objective function of Denoising Score Matching, integrating Score Matching [65] and denoising principles [141], is: $\mathbb{E}_{y\sim p(y)}\mathbb{E}_{\tilde{y}\sim q(\tilde{y}|y)}\left[\|s_\theta(\tilde{y}) - \Delta_{\tilde{y}} \log q(\tilde{y}|y)\|_2^2\right]$, where s_θ is (Stein Score) is the real noisy data. For Gaussian noise, this simplifies to:

$$\sum_{\epsilon\in B} \lambda(\epsilon)\mathbb{E}_{y\sim p(y)}\mathbb{E}_{\tilde{y}\sim\mathcal{N}(y,\epsilon)}\left[\left\|s_\theta(\tilde{y},\epsilon) - \frac{\tilde{y}-y}{\epsilon^2}\right\|\right], \tag{10}$$

where B is the set of standard deviations and $\lambda(\epsilon)$ is a coefficient function. Using Langevin dynamics principles, the iterative update is: $\mathbf{y}_k \leftarrow \mathbf{y}_{k-1} + \varphi\Delta_{\mathrm{y}} \log p(\mathbf{y}_{k-1}) + \sqrt{2\varphi}\mathbf{z}_k, 1 \leq k \leq K$. This method allows the gradual transformation of noise into the desired data. The work [58] proposed an objective function for optimization based on variational bounds, leading to: $\mathbb{E}_{t,\xi}\left[C\|\xi - \xi_\theta(\sqrt{\delta_t}\mathbf{y}_0 + \sqrt{1-\delta_t}\xi, t)\|_2^2\right]$, where C is a definite constant, ξ is the noise generated randomly from a standard Gaussian distribution and δ is also a constant changing with t. Let $\beta_t \sim \mathcal{N}(0, 1), \delta = 1 - \beta_t, \delta_t = \Pi_{i=1}^{t}\delta_i$, where we can set $\beta_t = 0.5$.

Compared to GANs, DMs offer significant advantages in stability and simplicity. While GANs require training both a generator and discriminator, DMs focus solely on the generator with a straightforward Gaussian-based loss, avoiding the adversarial nature that often causes instability in GANs. The work [28] demonstrated that

DMs outperform GANs in image quality, achieving lower Fréchet Inception Distance (FID) scores across multiple resolutions on ImageNet. This indicates superior fidelity and diversity in generated samples.

DMs benefit from simpler training processes and avoid issues like mode collapse common in GANs. Additionally, classifier guidance in DMs effectively balances diversity and fidelity, further enhancing image quality. These features make DMs more computationally efficient and easier to optimize, marking a significant advance in generative modeling and image synthesis.

In summary, DMs streamline the training process, reduce computational complexity, and achieve superior performance compared to GANs. The success of DMs lies in their ability to mimic a straightforward reverse process, fitting simple Gaussian distributions, which significantly enhances optimization and performance.

4.4 Art-style-reconstruction Algorithm

For comparison fairness, we classify the AI artworks into style transfer and style reconstruction. Meanwhile, we take the methodology and the art style to consider. This section analyses different method algorithms under one art style.

4.4.1 Line Drawings. As neural style transfer methods achieve sketching directly from images (e.g., APDrawing-GAN [161], Synthesizing human-like sketches [74]), we analyze line drawing methods focusing on the drawing process.

The work [47] proposed sketch-rnn, an RNN, capable of generating stroke-based drawings. A sketch is defined as a point list, where each point is a vector with five elements: (Δx, Δy, st_1, st_2, st_3). The sketch-rnn model employs a sequence-to-sequence VAE architecture, similar to those in [78, 121]. It encodes a sketch image into a latent vector and decodes it stroke-by-stroke, guided by the encoded states.

The encoding process involves two RNNs processing the sketch sequence and its reverse, resulting in final hidden states $\underrightarrow{h}$ and $\underleftarrow{h}$, combined into h_s. The process can be written as follows:

$$\underrightarrow{h} = \underrightarrow{\text{encode}}(Sq), \ \underleftarrow{h} = \underleftarrow{\text{encode}}(Sq_{reverse}), h_s = [\underrightarrow{h}; \underleftarrow{h}]. \tag{11}$$

The sketch-rnn encoder processes the concatenated hidden states h_s into δ and $\hat{\eta}$ of size V_z. $\hat{\eta}$ is transformed into the non-negative standard deviation η via exponentiation. Using δ, η, $\mathcal{N}(0, 1)$, and a vector of 2-D Gaussian variables, a random latent vector $z \in \mathbb{R}^{V_z}$ is constructed, akin to the VAE approach in [78]. z is conditioned on the input sketch, differing from deterministic outputs.

The auto-regressive RNN decoder of sketch-rnn sequentially predicts strokes using the last point, previous sketch sequence Sq_{di-1}, and latent vector z. It iterates through drawing steps to generate simple object sketches and can produce ablation sketches by adjusting the Kullback-Leibler loss weight. However, sketch-rnn struggles with complex images and supports limited sketch styles, allowing human participation only in predicting unfinished sketches.

The Creative Sketch Generation method [41] introduces DoodlerGAN, which leverages styleGAN2 [41] to sequentially generate sketch parts guided by human observations. Its part selector facilitates a human-in-the-loop sketching process but is currently limited to birds and creative creatures.

An alternative approach [169] uses reinforcement learning (Deep Q-learning) in Doodle-SDQ to train an agent to draw strokes on a virtual canvas, aiming to reconstruct a reference image stroke-by-stroke. The similarity metric $\mathbb{S}_k$ evaluates the canvas's closeness to the input image: $\mathbb{S}_k = \frac{\sum_{i=1}^{L}\sum_{j=1}^{L}\left(P_{ij}^{k} - P_{ij}^{\text{ref}}\right)}{L^2}$, where P_{ij}^{k} and P_{ij}^{ref} are pixel values at position (i, j) on the canvas and input image, respectively, at step k. The pixel reward $R_{\text{P}} = \mathbb{S}_k - \mathbb{S}_{k+1}$ optimizes the executing action at each step.

Doodle-SDQ's line-stroke sketching penalizes slow movements (P_{s} for <5 pixels/step or pen lift) and incorrect color choices (P_{c} with β adjusted for grayscale/color input). The final reward $R_k = R_{\text{P}} + P_{\text{s}} + \beta P_{\text{c}}$ combines pixel similarity and penalties. While Doodle-SDQ reproduces reference sketches well, it cannot sketch from real photos and lacks artistic creativity. In [169], strokes are simulated by a virtual ‘pen’, with reinforcement learning mapping actions to strokes. This inspires the development of diverse stroke types, potentially mimicking oil paintings and ink-wash paintings.

4.4.2 Oil Painting. The method in [64] utilizes a model-based deep deterministic policy gradient (DDPG) [91] algorithm to simulate a stroke-by-stroke oil-painting process. Bézier curves mimic brushstroke paths, and a circle represents the brush tip. The control points of the Bézier curves serve as actions, enabling action-to-stroke mapping. Given an input photo P_I and an initial canvas C_0, the model generates an action sequence $(b_0; b_1, ..., b_{n-1})$ to sequentially render strokes onto the canvas, producing the final painting C_N. This task is formulated as a Markov decision process with a state space $\mathfrak{S}$, action space $\mathfrak{B}$, transition function trans(s_n, b_n), and reward function $R(s_n, b_n)$ designed to minimize the distance between the input image and the canvas at each step: $R(s_n, b_n) = L_n - L_{n+1}$, where L_n and L_{n+1} represent the losses between P_I and the current/next canvases, respectively. The model aims to maximize the accumulated discounted future reward $R_n = \sum_{i=n}^{T} \epsilon^{(i-n)} R(s_i, b_i)$ with a discount factor $\epsilon \in (0, 1)$.

The original DDPG algorithm comprises an actor network $\Phi(s)$ that maps state s_n to actions b_n, and a critic network $\Psi(s, b)$ that estimates reward to guide the actor. Both networks are trained using the Bellman equation 12, with an experienced replay buffer storing the latest 800 episodes to enhance data usage:

$$\Psi(s_n, b_n) = R(s_n, b_n) + \epsilon \Psi(s_{n+1}, \Phi(s_{n+1})). \tag{12}$$

The MDRL Painter method in [64] improves upon line drawing approaches by simulating oil-painting brushstrokes using Bézier curves and circles. This method is improved from the line drawing of method [169] by designing the brushstroke. Although it can create paintings from various input images, the details are coarse, and the simulated stroke textures lack realism compared to human-made oil paintings.

The work [7] (ASRP) aimed to mimic human artist styles by generating brushstroke samples with similar textures. It uses Bézier curves to simulate strokes without tuning transparency, ensuring realism. VAEs were trained to capture artist brushstroke features, resulting in stroke textures close to human artists’, but the final paintings lacked content detail.

The work [123] improved painting quality by proposing Content Masked Loss (CML), a reinforcement learning model based on [64]. CML emphasizes salient regions using VGG-16 features and ℓ_2 distance, mimicking the human painting process. However, while the model captures the painting process well, it loses detailed content and stroke texture clarity.

Another AI painting model [171] SNP for AI oil painting contributes to stroke modelling by generating strokes with realistic oil-painting textures. A dual-pathway neural network independently generates stroke colors and textures. The model predicts and renders strokes step-by-step to optimize the final canvas C_N to resemble the input image I_r: $C_N = \phi_{n=1 \sim N}(\tilde{s}) \approx I_r$, where $\phi_{n=1 \sim N}(.)$ maps stroke parameters to canvas states. The model optimizes stroke parameters $\tilde{s} = [s_1, ..., s_N]$ using gradient descent to minimize the visual similarity loss $\mathcal{L}(C_N, I_r)$: $\tilde{s} \leftarrow \tilde{s} - \theta \frac{\partial \mathcal{L}(C_N, I_r)}{\partial \tilde{s}}$, where θ is the learning rate.

The SNP method [171] produces paintings with more details and realistic oil-painting stroke textures compared to [7, 64, 123], as shown in Fig. 6. While ASRP [7] and SNP exhibit clear oil-painting textures, SNP’s output size is fixed, requiring input images with the same aspect ratio. This can distort non-conforming images, and some input details may become blurry. Additionally, SNP requires more computation time than MDRLP.

4.4.3 Ink Wash Painting. Ink wash painting seems difficult to achieve with learning-based methods, and there are only a few research studies [151, 152]. For example, the texture of Chinese hair brush is difficult to mimic, although conventional SBR methods make contributions [131] to stroke modelling. The method of [151] proposed using the Markov decision process (MDP) to imitate drawing a stroke. The authors first used a tip V and a circle

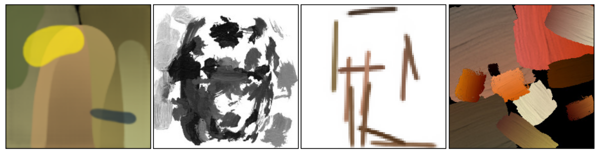

Fig. 6. Stroke comparison. The images, from left to right, are generated by MDRLP of [64], ASRP [7], CML [123] and SNP [171], respectively.

Fig. 7. The pastel-like stroke samples and the painting result generated by the method of neural painter [109].

with centre C_o and radius r_o to model the brush agent. MDP consists of a tuple $(\hat{\mathcal{S}}, \hat{\mathcal{A}}, P_d, P_T, \phi)$, where $\hat{\mathcal{S}}$ is a set of continuous states of the canvas, $\hat{\mathcal{A}}$ is a set of continuous actions, P_d is the probability-density of the initial state. $P_T(\hat{s}'|\hat{s}, \hat{a})$ is the transition of the probability density from the current state of the canvas $\hat{s} \in \hat{\mathcal{S}}$ to the next state $\hat{s}' \in \hat{\mathcal{S}}$ when taking action $\hat{a} \in \hat{\mathcal{A}}$. The term $\phi(\hat{s}, \hat{a}, \hat{s}')$ denotes the immediate reward for the transition from $\hat{s}$ to $\hat{s}'$. Let $\mathcal{T} = (\hat{s}_1, \hat{a}_1, \hat{s}_L, \hat{a}_L, \hat{s}_{L+1})$ be a trajectory of length L. Then, the return (i.e. the sum of the accumulating discounted future rewards) along $\mathcal{T}$ is written as: $\phi(\mathcal{T}) = \sum_{l=1}^{L} \sigma^{L-1} \phi(\hat{s}_l, \hat{a}_l, \hat{s}_{l+1})$, where $\sigma \in [0, 1)$ is the discount value for the future reward. Meanwhile, the authors designed four actions to move the brush agent, and in the reinforcement learning (RL) model, the brush agent was trained to generate hair brushstrokes.

Since the algorithm achieves high fidelity of hair brushstroke textures, the RL model is, at last, able to use the brush agent to generate ink wash paintings or Chinese paintings. Although the painting results contain textures of hair brushstrokes and characteristics of ink wash paintings, the method does not provide the painting process. Therefore, we do not know what happened during the painting procedure. We are not sure if the paintings are painted stroke-by-stroke. Moreover, the method description does not explain how the painting agent processes the input reference images and how the agent decomposes the images into strokes.

4.4.4 Pastel-like Painting. The method of neural painters (NP) in [109] uses GAN-based model and VAE-based model to simulate an intrinsic style-transform painting. Since the stroke textures are close to the pastel-painting style, we have called this form of painting pastel-like painting. However, the finished paintings express few characteristics of pastel paintings. The GAN-based and VAE-based models in the method were used to generate pastel-like strokes by training the models on the stroke dataset provided by the MyPaint program. When training the GAN and VAE-based models, the author labelled the dataset for the action space mapping a single action to a single brushstroke. The entire model (a neural painter) then used the GAN or VAE-based model to generate pastel-like strokes rendering on the canvas. By dividing the canvas into grids with the same size as the stroke image generated by the GAN or VAE-base model, the neural painter was able to recreate a pastel-like painting based on the given image. However, the paintings generated by neural painters lost much detailed content and the pastel-painting stroke textures were not clear. As Fig. 7 shows, with images from [109], the stroke samples contained characteristics of pastel-painting stroke textures, but the painting not only lost too much detailed content but also had few pastel-painting characteristics.

4.4.5 Robotic Painting. Robotic painting has long captivated both artists and robotics experts. Most artistic painting robots use acrylic paints [75], which are nearly as versatile as oil paints but are water-soluble, eliminating the need for harsh or toxic thinners and solvents. An example of an acrylic painting robot is the e-David robot [43, 92, 93], developed by Dessoin, Lindmeier and colleagues. This system comprises an industrial robot equipped with a paintbrush and a visual feedback system, utilizing a set of premixed colors. Additional color mixing is achieved by applying translucent brushstrokes to the canvas, considering the Kubelka-Munk paint film theory. The e-David robot can also learn to replicate brushstrokes through trial and error. The work of LETI painting robot [75] introduces a new type of robot capable of precisely metering and mixing acrylic paints, demonstrating high-quality painting results. The robotic system's capabilities are showcased through four artworks: replicas of landscapes by Claude Monet and Arkhip Kuindzhi, and synthetic images generated by StyleGAN2 and Midjourney neural networks. These results can be applied to computer-generated creativity, art replication and restoration, and color 3D printing.

The work by [7] presents a new approach that integrates artistic style into the process of robotic painting through collaboration with human artists. The method involves collecting brushstroke samples from artists, training a generative model to imitate the artist's style, and then fine-tuning the brushstroke rendering model to adapt it to robotic painting. Their user studies have shown that this method can effectively apply the artist's style to robotic painting. The use of a Visual Measurement System (VMS) and a Robotic Painting System (RPS) to simulate brushstrokes is presented by [44]. The specific method involves using VMS to capture the interaction trajectories and environmental state information during the artist's painting process. Then, RPS mimics human painting actions based on this information, utilizing real-time visual feedback to adjust the robot's movements, thus achieving precise brushstroke simulation. Through these methods, the proposed ShadowPainter system can simulate brushstroke effects that are close to human levels.

Work by [106] explores whether AI-driven robots can be regarded as artists and create real works of art. Two experiments were conduction to investigate people's perception of the artistic quality of robot paintings and their acceptance of the identity of robot artists. Experimental results show that although people generally believe that robot paintings are not much different from human works in terms of artistic quality, they have reservations about identifying robots as artists.

In conclusion, robotic painting has become a fascinating field that bridges art and technology. Various systems and methods have been developed to mimic and even surpass human artistic abilities. From using acrylic paints to precise metering and mixing techniques, these robots have demonstrated extraordinary painting capability. The integration of artistic styles through human-machine collaboration further enhances the creative possibilities of robotic painting. As technology advances, we can expect more innovative and captivating artworks to emerge from this exciting field, breaking the boundaries of traditional art forms and opening new avenues for artistic expression. However, the debate over whether AI-driven robots can truly be considered artists remains unresolved. Despite the increasing technical proficiency and artistic quality approaching human standards, societal acceptance of robots as genuine creators of art continues to lag. Future research and development in this field may focus on bridging this gap, enhancing the creative capabilities of robots, and addressing the ethical and philosophical issues surrounding AI and art.

5 Evaluation

From the SBR methods of the early nineties to increasingly learning-based methods of drawing/painting and generating for image processing, research into AI painting has reached a new pinnacle. We have analysed recent methods based on the taxonomy of generation methods and art styles. Different models and algorithms have been proposed to achieve diverse kinds of creative artwork. Although these methods are rich in AI artworks, their drawbacks are still obvious as well as their advantages. The discussion about the evaluation of aesthetics and usability catches much attention of researchers in both industry and academia.

We propose that AI artworks should be compared within the same field or category. However, for existing evaluations of methods and the artworks generated by these methods, there are no uniform standards. Some evaluation aspects do not fit certain methods or artworks. For example, we should not take the details of the content in artwork into account only when comparing the method and its outputs. We are comparing artworks instead of the high resolution of an image: we should be taking the art elements into account.

5.1 Evaluation Metrics

Currently, there are four principal representative metrics widely used for image quality evaluation, namely Inception Score (IS), Fréchet Inception Distance (FID), Contrastive Language-Image Pre-training (CLIP), and Generated Image Quality Assessment (GIQA) [143]. IS evaluates the effectiveness of generative models, mainly measuring the quality and diversity of generated images. It assesses the classification effectiveness of generated images based on the image classifier Inception v3. FID evaluates the effectiveness of generative models, measuring the distance between the distribution of generated images and the distribution of real images. FID calculates the difference between these two distributions based on the Inception network. CLIP is an artificial intelligence model developed by OpenAI that can simultaneously understand text and images. It is not just an evaluation metric but also a bridge connecting language and visual information. GIQA evaluates the quality of generated images, defining "quality" as the similarity between the distribution of generated images and real datasets. This metric can score individual-generated images, which is a capability that some previous generative model evaluation metrics lacked.

These four metrics cannot be directly compared due to their different calculation methods and result ranges. Moreover, none of these evaluation metrics target elements related to artistic aesthetics. When image evaluation is needed from the perspective of the image or artwork itself, these evaluation metrics are not very applicable. To this end, we propose a six-dimensional evaluation index to focus on evaluating images from an artistic aesthetic perspective, which perfectly fills this gap.

We have referred to some elements used for evaluation from the artistic field. Art vocabulary [134] describes the elements of art and principle of design as:

- The elements of art: form, line, shape, space, texture, color. Color is light reflected off objects. There are three main characteristics: hue (the name of the color: red, green, blue, etc.), value (how light or dark it is) and intensity (how bright or dull it is).
- The principles of design: balance, movement, emphasis, repetition, proportion, pattern, rhythm, unity, variety.

When evaluating AI-generated images, we cannot only consider the quality of the generated images, namely just using the four evaluation metrics mentioned above. From an artistic perspective, we should evaluate the artistic characteristics of the works. Thus, we design several items of the evaluation for AI artworks inspired by the AI criticism [37], Exploring the Representativity of Art Paintings [22], Beauty in abstract paintings [102], Aesthetic-Aware Image Style Transfer [61], and Aesthetics-Guided Graph Clustering [165]. We mainly design the items on two aspects, the **Beauty** of the entire painting and the **art elements**. In particular, the beauty of the painting takes 50% of the score, and the elements, too. The art elements are **Line Smooth, Stroke Texture, Colors, Contents, and Art Style** recognisability. As Table 1 indicates, the beauty of the entire artwork is the core characteristic of artwork so the item of beauty takes 50% of an artwork. Each of the other elements takes 10% of an artwork. We ask the participants to score the paintings on the beauty of the entire artwork and all the elements according to a five-point Likert scale [90] (the points being: strongly good (5′), good (4′), neither good nor bad (3′), bad(2′), strongly bad(1′)). The questions are as follows:

Table 1. Evaluation items used in the user study.

Item	Explanation
Beauty	The aesthetic evaluation of the entire artwork
Line	The expression and smoothness of the lines in the artwork
Texture	The stroke texture expressed in the artwork
Color	The treatment of light and shade in the artwork
Contents	The features of the whole artwork, including the details
Style	The art style of the artwork, for example, oil-painting style

- How beautiful is this artwork?
- How well are lines expressed in this artwork?
- How well are stroke textures expressed in this artwork?
- How well is the light and shade of the color treated in this artwork?
- How detailed are the contents contained in this artwork?
- How easy is it to recognise the art style of this artwork?

5.2 Experiments and Analysis

Experiments were conducted using the the methods described on the same platform with the authors providing codes and pre-trained models. We then choose the best results of the compared methods as the test images for visual comparison and user study.

5.2.1 Visual Comparison. We first compare the results generated by the methods of image-style-transfer. In particular, the stylised images are synthesised by the content image and the style image. Fig. 8 shows the sample results generated by methods of AAMS [159], ASTSAN [110], and URUST [144]. The first column contains the content images and style images (small). Rest columns from left to right are the generated images of AAMS [159], ASTSAN [110], and URUST [144], respectively. All of the results present the style features well.

However, as can be seen from the top row (Fig. 8), the style image is a pencil drawing in the top row. However, the image generated by ASTSAN [110] still retains the original color features of the content image, indicating incomplete style transfer. Although the image generated by URUST [144] exhibits pencil drawing features, the content of the bird is blurred, indicating imperfect content expression. The image generated by AAMS [159] presents clear content of the target image, and the style features are also harmoniously synthesized into the target image. From a visual aesthetic perspective, considering overall aesthetic "Beauty", "Lines", "Colors", "Stroke Texture", "Content" details, and "Style", the image generated by AAMS [159] appears more aesthetically pleasing than the others. Therefore, we conclude that the results of image style transfer should contain detailed content of the target image, and the features of the style image should not overshadow the content image.

Fig. 9 shows the visual results of new style transfer methods. The visual effects of the images generated by AesPA-Net [60], EFDM [167], AdaIN [63], CAST [168], StyTR2 [23] and AdaAttN [95] are quite impressive. They maintain high clarity and content detail, with good color reproduction and contrast. The stroke and line textures are also well-presented. The cat's image is vivid, and the background environments have their own characteristics, showcasing different artistic styles. However, in terms of style transfer, they do not fully embody the features of the style image, so they are not the best in this aspect.

The images generated by MAST [24] and SID [21] are slightly inferior in content detail. Although they basically capture the cat's image and background environment, they are slightly lacking in clarity, color reproduction, and contrast. Some details may be blurry, and the colors may be somewhat distorted, affecting the overall visual effect. The line sense and stroke texture are not very obvious. The content detail expression in images generated

Fig. 8. Visual comparison of existing neural style transfer methods. The first column shows the content and style images (the small images). The second to the fourth columns contain the results of AAMS [159], ASTSAN [110], URUST [144], respectively.

Fig. 9. Visual comparison of existing style transfer methods. The first is the style image and the first image in the top row is the content image. The compared images refer to the work of Style Injection in Diffusion (SID) [21]. The methods are DiffuseIT [80], MAST [24], AesPA-Net [60], EFDM [167], SID [21], AdaIN [63], InST [166], CAST [168], StyTR2 [23], DiffStyle [67], AdaAttN [95].

by DiffuseIT [80], InST [166], and DiffStyle [67] is very poor. For InST [166] and DiffStyle [67], the cat's image is almost indistinguishable. On the contrary, InST [166] expresses more content from the style image. Although it is hard to recognize the content of the image generated by DiffStyle [67], its overall color expression creates a fresh and 'cute' effect.

In summary, the evaluation of style transfer results across various models highlights several key features necessary for generating high-quality, new-style artistic images. From the perspective of **Beauty**, an ideal artistic image should exhibit a balanced composition of visually pleasing elements, including harmonious color schemes and well-composed subjects. Regarding **Lines**, clarity and sharpness are crucial for defining objects and subjects, contributing to the overall structural readability of the image. In terms of **Colors**, accurate color reproduction and contrast are essential for enhancing visual appeal and reflecting the desired mood and atmosphere. **Stroke texture** plays a vital role in conveying the sense of artistic technique and traditional medium, providing a tactile experience for the viewer. **Content** details are important for maintaining the recognizability and realism of the main subject, ensuring that key elements are neither lost nor distorted during the transformation process. Finally, the **Style** itself must be faithfully reproduced, capturing the unique characteristics and nuances of the reference style image. Balancing these elements ensures that the generated artistic image not only adheres to the desired style but also stands out as a cohesive and aesthetically engaging piece of art.

Fig. 10. Visual comparison of existing GAN-based methods for photo-to-cartoon. The first column shows the input images, and the remaining columns from left to right are generated images by methods of GANs N' Roses [20], U-GAT-IT [77], and WBC [147], respectively.

Fig. 10 shows the results generated by style transfer methods. Note that the style of the generated images is learned from the training dataset, not synthesised from a style image. The first column shows the input images, and the rest of the columns, from left to right, are generated images by GANs N' Roses [20], U-GAT-IT [77] and WBC [147], respectively. The first row input image is from the dataset provided by U-GAT-IT [77], and the last input image is from the sample image test provided by WBC [147]. When comparing the first three rows of images, we observe that images generated by WBC [147] retain more realistic contents of the input images than the others. The images generated by GANs N' Roses [20] and U-GAT-IT [77] present more non-realistic cartoon features than WBC [147]. However, when comparing the bottom row images, we observe that the image generated by U-GAT-IT [77] has few cartoon features but blurred contents. Based on the analysis, we conclude that U-GAT-IT [77] has a low generalization.

Fig. 11 shows the results generated by line drawings methods. The top row shows the input reference images (small images), and the rest of the rows, from top to bottom, show the results generated by photo-sketching [85] and APDrawingGAN [161], respectively. The images generated by photo-sketching [85] lose so much content that it is difficult to recognise the object in the image. Although results generated by APDrawingGAN [161] contain sufficient image content, the expression of the girl's hair is not satisfactory.

Fig. 12 shows another line drawing results generated by DoodlerGAN [41]. The images are created by the online demo provided by the authors. The model only creates birds or bird-like creatures. The images are generated

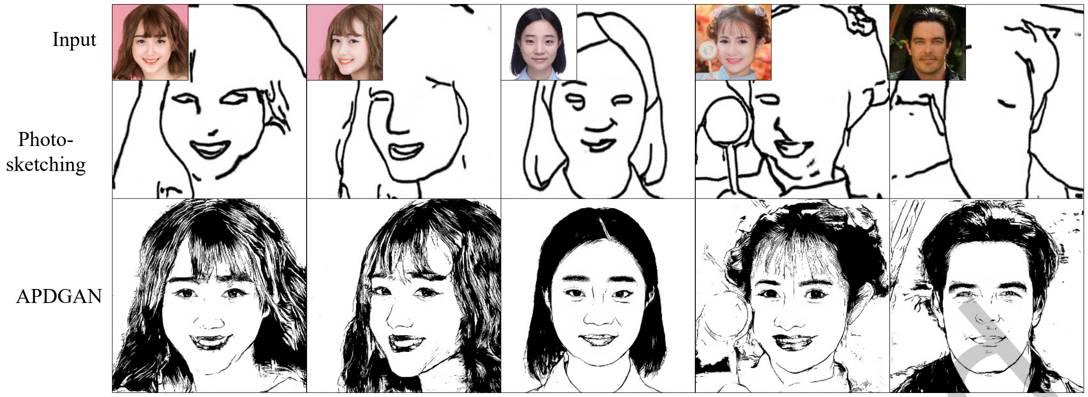

Fig. 11. Visual comparison for photo-to-sketch. The top row shows the reference images. The middle row shows the results of the photo-sketching method [85], and the last row shows the results of APDrawingGAN method [161].

step by step. The whole image consists of several components of a bird or bird-like creatures. The human or the computer draws a final step in the process to finish a component. Images (a) and (c) are finished by the cooperation of a human and a computer. Images (b) and (d) are generated by the computer only. We observe that all the images are like birds but not real birds.

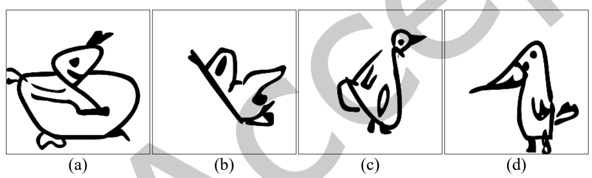

Fig. 12. Line drawings generated by DoodlerGAN [41].

Fig. 13 shows the results generated by methods of painting. The results are created stroke by stroke. The left column shows the input images, and the remaining columns from left to right are the results generated by methods of MDRLP [64], SNP [171], Stroke-GAN Painter [145] and NP [109], respectively. The images in the three middle columns have colors closer to the input images than the right-column images. Images generated by SNP [171] present clearer stroke textures than others. Images generated by MDRLP [64] contain more details than others. Images generated by MDRLP [64], Stroke-GAN Painter [145] and SNP [171] look like oil-painting, especially the brushstroke texture of SNP [171]. The style of images generated by NP [109] is difficult to recognize since the stroke texture is more like pastel-painting than oil-painting, but the art style is close to watercolor painting.

5.2.2 User Study. To make an objective evaluation of the generated images, we undertake a two-step user study. For a fair comparison, we conduct a blind-trial test among the participants. The participants know neither the authors of the methods used for generating comparison paintings nor the experimenter. The participants are chosen from various backgrounds (69.2% in the art field, and 85.1% known about AI Art), age groups (18–60), and genders (74 females and 127 males).

We designed the user study as a two-step test for the six-dimensional evaluation index analysis to find suitable items for a certain art style. We design the two-step user study inspired by the work of [136]. For the first step, we mix all the painting results in the same questionnaire and then ask the participants to score all the paintings

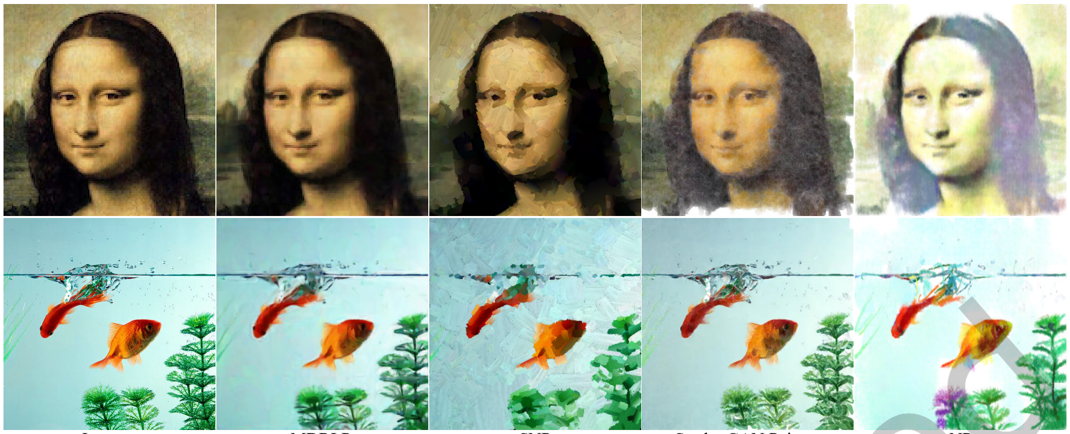

Fig. 13. Visual comparison for paintings. The left column contains the input reference images. The other columns are the painting results of different methods. The three middle-column methods use oil-painting strokes to create paintings. The right column uses pastel-like strokes to generate paintings.

according to the six evaluation items. In the second step, we classify the paintings into two categories: style-transform paintings and style-reconstruction paintings (stroke-by-stroke paintings). The style-reconstruction paintings contain the painting process images, and the paintings with the same style are put in the same group. We then ask the participants to score the paintings based on a five-point Likert scale [90]. The participants finish the user study's Step 1 and Step 2.

Table 2. Intra-class Correlation Coefficient Results of the Step 1 test.

Two-way Mixed/Random Consistency	ICC	95% CI
Single Measure ICC (C,1)	0.437	0.373 ∼ 0.513
Average Measure ICC (C,K)	0.985	0.980 ∼ 0.989

Table 3. Intra-class Correlation Coefficient Results of the Step 2 test.

Two-way Mixed/Random Consistency	ICC	95% CI
Single Measure ICC (C,1)	0.498	0.432 ∼0.5740.437
Average Measure ICC (C,K)	0.988	0.985 ∼0.991

Table 2 and Table 3 show the Intraclass Correlation Coefficient (ICC) Results of the two-step user study. In analyzing two sets of ICC data, we observed similar trends regarding the reliability of single and average measurements. In both datasets, the single measure ICC(C,1) values, 0.437 and 0.498 respectively, indicate a certain to moderate degree of correlation in single measurements, but not particularly strong. The 95% confidence intervals for these single measures show a range of fluctuation, suggesting room for improvement and reflecting the potential impact of random errors or individual differences. However, the average measure ICC(C,K) values exhibit extremely high reliability in both sets, reaching 0.985 and 0.988. The narrow confidence intervals further confirm that averaging multiple measurements significantly enhances measurement accuracy and consistency. These findings underscore the importance of repeated measurements in improving data quality and reliability. In the subsequent data analysis, we mainly took the average score of each question for further analysis.

Table 4 shows the experimental results of Step 1, and Table 5 shows the results of Step 2. Scores in the two tables are marked with different colors for observation. Red indicates the highest scores, blue indicates the lowest scores, and orange represents scores lower than 3 except blue ones.

Table 4 shows the six-dimensional evaluation index scores on mixed artworks. In the beauty column of Table 4, we observe that the results generated by the method of photo-sketching [85] give the lowest scores (2.849). Compared with other paintings, the sketches generated by photo-sketching [85] have little content from the input image, and we cannot readily recognise what the sketches express in some cases (as Fig. 11 shows). The score is 2.849, which means that most participants judged the sketches to be poor in terms of beauty. The sketches generated by DoodlerGAN [41] also obtain a lower score (3.000) compared with other paintings. However, when comparing the line smoothness of the paintings, we observe that paintings generated by the method of APDrawingGAN [161] gained higher scores than most. Paintings generated by DiffStyle [67], ASTSAN [110], DiffuseIT [80], and H-SRC [72] obtained scores lower than 3. This means these paintings have poor line expressions. The texture column compares the stroke texture of the test artworks. MAST [24], H-SRC [72] obtain scores lower than 3; however, AesPA-Net [60], APDrawingGAN [161], StyTR2 [23], CAST [168] and PST [98] obtain scores higher than 3.6. That means these methods express stroke texture well. Methods obtaining high scores, especially PST [98] (3.823), present clear stroke textures in their paintings. In the color column, most of the methods score higher than 3 except Photo-Sketching [85], MAST [24], DiffStyle [67], H-SRC [72] and DoodlerGAN [41]. For the content comparison, only H-SRC [72], Photo-Sketching [85] and DoodlerGAN [41] obtain a score lower than 3. Scanning Fig. 11, the images generated by Photo-Sketching [85] lose too much content. Thus, the line drawings or sketches, when compared with other paintings with rich contents, only gain lower scores. When compared in terms of art style recognisability, only the paintings generated by H-SRC [72] obtained low scores (2.940). That is, most of the participants cannot recognize the art style of the paintings created by H-SRC [72]. Table 5 shows the

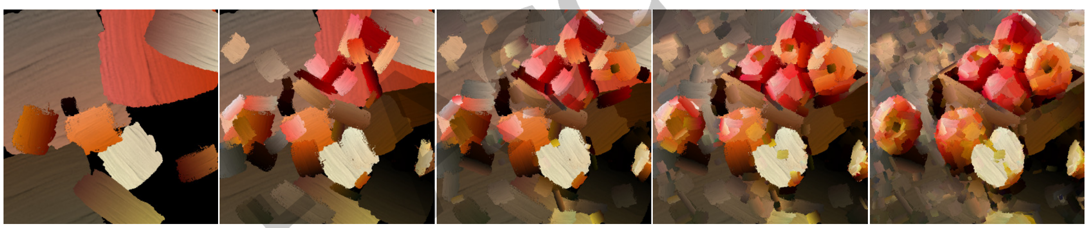

Fig. 14. Example of the painting process.

scores of the six-dimensional evaluation index on the classified artworks. The artworks are divided into four groups: style transfer/transform, photo-to-cartoon, line drawing, and stroke-by-stroke painting. Some of the artworks created stroke-by-stroke also exhibit the painting process images (Fig. 14). In the user study step 2, the scores were significantly higher than those of step 1, especially since the number of scores below 3 was much fewer. The reason is that in the second test, users were informed of the style type and the image generation method so that users had a fuller understanding of the object they were evaluating. Therefore, users would be more tolerant and accepting of some less distinguishable options, thus giving higher scores. In the beauty column of Table 5, results of PST [98], AAMS [159], Im2Oil [137], APDrawingGAN [161], Intelli-paint [127] obtained higher scores than most others. Especially, in the style column, the lowest score is higher than 3, which means when users are informed of the styles and generation methods, their scores for artworks will be more accurate in the style confirmation item. In addition, it is in line with the principle of fairness to evaluate paintings by classifying them according to their styles and generation methods.

To conduct a more detailed analysis of the user study, we have sorted and classified the scores of the users based on their backgrounds. Fig. 15 shows the scores of all artworks by five backgrounds of users: all users, users

Table 4. Scores on evaluation items in the user study, Step 1. All the painting results are put in the same questionnaire.

Methods	Beauty (50%)	Line (10%)	Texture (10%)	Color (10%)	Content (10%)	Style (10%)	Mixed Total
AAMS [159]	3.756	3.532	3.582	3.677	3.587	3.613	3.677
ASTSAN [110]	3.095	2.935	3.069	3.069	3.000	3.185	3.073
URUST [144]	3.164	3.000	3.224	3.086	3.125	3.267	3.152
SID [21]	3.741	3.444	3.504	3.478	3.483	3.586	3.620
AesPA-Net [60]	3.836	3.612	3.716	3.556	3.746	3.716	3.753
CAST [168]	3.625	3.444	3.608	3.526	3.483	3.539	3.572
StyTR2 [23]	3.884	3.591	3.711	3.591	3.716	3.651	3.768
EFDM [167]	3.595	3.323	3.341	3.418	3.487	3.448	3.499
MAST [24]	3.108	3.004	2.918	2.996	3.116	3.065	3.064
AdaAttN [95]	3.582	3.358	3.371	3.293	3.379	3.362	3.467
AdaIN [63]	3.685	3.405	3.565	3.466	3.440	3.539	3.584
DiffuseIT [80]	3.233	2.978	3.185	3.082	3.065	3.151	3.163
InST [166]	3.496	3.216	3.353	3.233	3.341	3.388	3.401
DiffStyle [67]	3.246	2.892	3.125	2.978	3.121	3.043	3.139
CycleGAN [170]	3.543	3.188	3.338	3.297	3.358	3.345	3.424
Gated-GAN [14]	3.853	3.491	3.591	3.690	3.634	3.763	3.744
StarGAN [18]	3.353	3.168	3.250	3.134	3.297	3.254	3.287
StarGANv2 [19]	3.366	3.134	3.190	3.095	3.233	3.216	3.270
H-SRC [72]	2.961	2.845	2.901	2.884	**2.836**	2.940	2.921
MSC [10]	3.522	3.203	3.280	3.306	3.315	3.224	3.394
U-GAT-IT [77]	3.670	3.391	3.460	3.432	3.485	3.460	3.558
WBC [147]	3.432	3.263	3.319	3.235	3.310	3.262	3.355
CartoonGAN [15]	3.358	3.172	3.315	3.284	3.263	3.280	3.310
MSCartoonGAN [125]	3.457	3.272	3.379	3.241	3.366	3.379	3.392
GANs N'Roses [20]	3.865	3.553	3.585	3.586	3.658	3.726	3.743
LGLD [13]	3.862	3.625	3.595	3.366	3.603	3.828	3.733
APDrawingGAN++ [162]	3.565	3.504	3.582	3.220	3.526	3.608	3.526
APDrawingGAN [161]	3.875	3.694	3.642	3.302	3.612	3.741	3.728
Photo-Sketching [85]	2.849	**2.784**	2.845	2.828	2.853	3.194	2.875
DoodlerGAN [41]	3.000	3.022	2.970	2.918	2.927	3.263	3.010
NP [109]	3.427	3.190	3.310	3.241	3.379	3.397	3.365
MDRLP [64]	3.534	3.310	3.418	3.448	3.418	3.474	3.474
SNP [171]	3.659	3.392	3.491	3.547	3.445	3.582	3.576
Stroke-GAN Painter [145]	3.613	3.430	3.516	3.521	3.456	3.453	3.544
PaintTransformer [94]	3.621	3.512	3.447	3.342	3.452	3.567	3.543
Intelli-paint [127]	3.653	3.521	3.522	3.601	3.485	3.587	3.598
Im2Oil [137]	3.732	3.311	3.554	3.663	3.512	3.601	3.630
RST [79]	3.712	3.344	3.558	3.628	3.523	3.612	3.623
PST [98]	**4.112**	3.603	3.823	3.892	3.884	3.974	3.983
Average	3.529	3.299	3.389	3.337	3.383	3.443	3.450

Table 5. Scores on evaluation items in the user study, Step 2. All the painting results are classified into categories according to the generating procedure and art styles.

Category	Methods	Beauty (50%)	Line (10%)	Texture (10%)	Color (10%)	Content (10%)	Style (10%)	Categorised Total
Style Transfer/ Transform New Style	AAMS [159]	3.910	3.637	3.672	3.706	3.682	3.881	3.813
	ASTSAN [110]	3.378	3.328	3.308	3.318	3.338	3.373	3.356
	URUST [144]	3.244	3.104	3.234	3.164	3.209	3.239	3.217
	SID [21]	3.602	3.318	3.423	3.323	3.498	3.473	3.504
	AesPA-Net [60]	3.861	3.448	3.622	3.493	3.537	3.552	3.696
	CAST [168]	3.741	3.433	3.562	3.488	3.512	3.562	3.626
	StyTR2 [23]	3.811	3.532	3.602	3.582	3.562	3.642	3.698
	EFDM [167]	3.692	3.353	3.567	3.443	3.522	3.493	3.584
	MAST [24]	3.478	3.119	3.174	3.219	3.164	3.343	3.341
	AdaAttN [95]	3.736	3.343	3.438	3.403	3.398	3.463	3.573
	AdaIN [63]	3.746	3.373	3.537	3.502	3.488	3.612	3.624
	DiffuseIT [80]	3.388	3.139	3.279	3.159	3.184	3.214	3.292
	InST [166]	3.493	3.229	3.323	3.279	3.289	3.428	3.401
	DiffStyle [67]	3.458	3.065	3.323	3.119	3.164	3.149	3.311
	CycleGAN [170]	3.674	3.378	3.376	3.453	3.398	3.425	3.540
	Gated-GAN [14]	3.881	3.532	3.597	3.542	3.542	3.776	3.739
	StarGAN [18]	3.537	3.164	3.363	3.358	3.333	3.249	3.415
	StarGANv2 [19]	3.493	3.204	3.333	3.224	3.289	3.388	3.390
	H-SRC [72]	3.224	2.945	3.085	3.025	3.070	3.055	3.130
	MSC [10]	3.562	3.249	3.483	3.284	3.378	3.423	3.463
Photo to cartoon	GANs N' Roses [20]	3.826	3.458	3.653	3.522	3.595	3.784	3.714
	U-GAT-IT [77]	3.690	3.378	3.530	3.439	3.479	3.464	3.574
	WBC [147]	3.578	3.362	3.453	3.374	3.408	3.311	3.480
	CartoonGAN [15]	3.577	3.179	3.507	3.338	3.224	3.373	3.451
	MSCartoonGAN [125]	3.552	3.299	3.393	3.343	3.328	3.358	3.448
Line drawing	LGLD [13]	3.831	3.532	3.577	3.368	3.662	3.697	3.699
	APDrawingGAN++ [162]	3.682	3.353	3.612	3.348	3.468	3.597	3.579
	APDrawingGAN [161]	3.905	3.537	3.617	3.418	3.572	3.796	3.747
	Photo-Sketching [85]	3.109	2.900	2.960	**2.771**	2.950	3.279	3.041
	DoodlerGAN [41]	3.308	3.144	3.134	2.905	3.119	3.279	3.212
Stroke by Stroke Painting	NP [109]	3.776	3.338	3.527	3.433	3.473	3.408	3.606
	MDRLP [64]	3.627	3.318	3.393	3.363	3.423	3.498	3.513
	SNP [171]	3.697	3.343	3.488	3.403	3.463	3.602	3.578
	Stroke-GAN Painter [145]	3.893	3.433	3.513	3.423	3.664	3.725	3.722
	PaintTransformer [94]	3.653	3.375	3.443	3.378	3.491	3.564	3.552
	Intelli-paint [127]	3.985	3.226	3.586	3.441	3.786	3.786	3.775
	Im2Oil [137]	3.901	3.315	3.688	3.412	3.878	3.823	3.762
	RST [79]	3.866	3.443	3.557	3.389	3.927	3.886	3.753
	PST [98]	3.987	3.586	3.732	3.443	**3.998**	3.923	3.862
Average		3.650	3.318	3.453	3.349	3.448	3.510	3.533

with artistic backgrounds who understand AI art, users with artistic backgrounds but do not understand AI art, users without artistic backgrounds but understand AI art, and users without artistic backgrounds who also do not understand AI art.

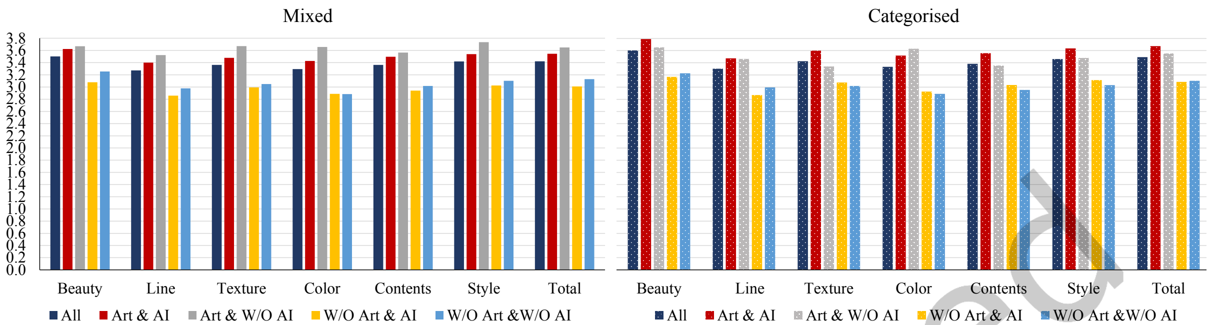

Fig. 15. The average scores of different background users in the mixed test and categorised test.

The analysis identified that the average scores of users with artistic backgrounds are higher than those of other users, whether in artworks-mixed or artworks-categorised tests. In the artworks-mixed test, users with an artistic background but no knowledge of AI art gave the highest scores, followed by users with an artistic background and knowledge of AI art. In the artworks-categorised tests, users with an artistic background and knowledge of AI art gave the highest scores except for the color item, followed by users with an artistic background but no knowledge of AI art. Especially in the color item, the latter group gave the highest scores. Interestingly, in the two-step user study, the average scores given by users with an artistic background were higher than the average scores given by all users. Among users without an artistic background, in the artworks-mixed test, the scores given by users who understand AI art are lower than those who do not understand AI art in every category. In the categorized test, only the Beauty and Line items have lower scores from users who understand AI art compared to those who do not. Overall, in both tests, users who understand AI art gave lower scores than those who do not.

6 Challenges and Opportunities

AI technologies have been applied in many fields, including industry, art and education, and have attracted significant attention. Methods for creating digital art are diverse, and the performance of these is steadily rising. However, there are still many challenges as well as opportunities. 1) When converting a photo to an artwork, the balance of fidelity and creativity is still an ill-posed issue. 2) For painting/drawing methods, the creation order of generating an artwork is still a machine order and very different from the human order. 3) For most learning-based methods, the framework almost generates one art style instead of multiple styles. 4) It is difficult to generate artworks without reference images; in other words, existing methods have to refer to an input image to finish the painting process. 5) The existing evaluations for AI artworks (conducting user studies) are still subjective. However, there are still many opportunities for AI artworks in areas such as science and technology big-bang society [4]. There are requirements and opportunities for AI artworks in many fields, such as social community, education, art and commerce.

6.1 Challenges

6.1.1 Fidelity VS. Creativity. Creativity has a profound impact on society [16, 163], especially in art. No matter whether we are considering style-transform AI artworks or art-style-reconstruction artworks, existing methods can 'almost' turn a photo into an artwork. Therefore, it is worth discussing the fidelity and creativity [50] of the results. Unfortunately, most painting/drawing methods have difficulty in achieving high fidelity because of the art style representation. For example, methods such as [7, 94, 123, 171], although presenting the stroke texture of oil painting well, produce results that lose much detailed content owing to the invariant stroke shape or type. The method in [64] also mimics the oil painting process and can generate high fidelity results when giving a large number of strokes, but the high fidelity result is almost a photo rather than an oil painting because the strokes lack oil-painting stroke textures. In summary, turning a photo into a painting is a creative task requiring the result to not be the same as the photo itself, but the fidelity requiring the preservation of as many details as possible is still a difficult challenge, and we have yet to deliver pleasing results consistently.

6.1.2 Creation Order. Most painting/drawing methods claim that they can mimic the human painting/drawing process. In reality, they model stroke generation to render a large number of strokes onto the canvas to finish the creation of an artwork. However, the generation process is so different from the human painting process that they ignore the creation order that humans follow. In particular, when human artists create artwork, e.g. an oil painting, they tend to draft the main objects by lines first and then paint the background and the objects progressively. It is worthwhile to teach machines to really mimic the human painting process so as to reveal the mysterious veil of art creation, even though it is difficult to achieve this task. If we make a step to achieve the real human painting process, we make the machine painting more intelligent and closer to the human artist; if we endow the machine or computer with inspiration and motivation for its creation (as pointed out by Hertzmann [56, 57]), then we may claim that the machine or computer can create art.

6.1.3 Abstract Art. Existing methods for creating AI artworks usually refer to the input image to recreate the artwork. However, a human artist can create artwork without real referent objects thanks to their human inspiration and imagination. Consequently, teaching a machine or computer to create artworks without reference images is a very challenging task. Although the work [155] achieved the generation of images from fine-grained text, the result was photorealistic and could not really be called artwork. The work [32] generated abstract artworks with their creative adversarial networks, but the model itself could not name the artwork according to its creation. In other words, this model just generates abstract images but does not know what the image is or what meaning the image represents. However, researchers can obtain inspiration from these two works, since the combination of text-to-image and abstract artworks can prompt areas of consideration and development for future AI art creation.

6.1.4 Multi-style. The work [14] managed to generate multiple styles of results within an unified framework for image-style transfer. It is popular to design a model to address multiple tasks; however, it is difficult to design a model that paints with multiple art styles. Although [64, 94, 171] could change the visual representation of the results by replacing different stroke styles, the art style stayed the same, almost close to oil paintings. Can machines or computers create different art styles of artworks within the unified framework? Similar to a human artist who can create a watercolor painting, a pastel painting and an oil painting, seemingly by changing their painting tools, can a painting system create different art styles of paintings by changing its stroke style? It is an interesting and challenging issue for both artists and computer scientists.

6.1.5 Aesthetic Evaluation. Aesthetic evaluation is a critical issue for AI artworks. The works [33, 45, 55, 61, 103, 108, 133] argued that aesthetic evaluation is important to develop methods for AI artworks. Especially for such diverse types of AI artworks as mentioned in [115], a fair and scientific evaluation system is very important. In this paper, we propose an evaluation system to cover several types of AI artworks so as to unify the diverse evaluating methods as well as make the evaluation fair when facing different types of AI artworks. However, even the proposed evaluation system is still based on user studies. Can we evaluate AI artwork and its methods via computing indexes? The proposed six-dimensional evaluation index may give some ideas and inspirations for the following research. For the development of AI artworks, fair, objective and scientific evaluation is still an important and challenging area to be addressed.

6.2 Technological Advancement

To address the aforementioned challenges, the following technological advancements need to be achieved: Firstly, the development of advanced image synthesis techniques and creative algorithms is necessary to enhance the fidelity of paintings and exhibit greater creativity. This can be accomplished by improving technological or algorithmic models such as Convolutional Neural Networks (CNNs), Generative Adversarial Networks (GANs), transformers, and diffusion models. Secondly, sequential modeling and reinforcement learning techniques should be utilized to enable AI to mimic the creative sequence of humans, from composition to detail refinement. For instance, by simulating the painting process of artists through deep learning techniques, a system can be developed that adjusts based on feedback during the creative process, allowing robots to more intelligently imitate the artistic creation sequence of humans. Thirdly, exploring unreferenced generation techniques and inspiration and imagination modules is crucial to enable AI to create abstract artworks without specific input. This can be achieved by advancing unsupervised learning and generative model-related technologies, while introducing a Natural Language Processing (NLP)-based inspiration and imagination generation module. Additionally, through multi-task learning and style transfer modules, AI can process multiple artistic styles within a single framework and dynamically change brushstroke styles, resulting in works of various styles. Finally, the introduction of computational aesthetics evaluation metrics and the proposed six-dimensional evaluation system is essential for objective, fair, and scientific evaluation of AI artworks. This can be accomplished through Image Quality Assessment (IQA) algorithms and visual aesthetic feature extraction techniques.

All these technological advancements rely on powerful computing capabilities and sufficient data support. Therefore, it is necessary to continuously enhance computing power and collect more diversified art datasets for model learning and training. By achieving these technological advancements, significant breakthroughs can be made in improving the quality, creativity, and diversity of AI artworks, while promoting the further development of human-machine collaborative creation.

6.3 Opportunities

6.3.1 Social Media Requirements. The application of AI artworks in the social media community is very popular. In an era of ever-higher aesthetic aspirations and requirements, self-actualisation and self-creation are areas of increasing attention and demand resources accordingly. Current techniques and algorithms cannot meet the demand of interaction and creation for everyone. Whether via social application software or on social websites, people are enthusiastic about making their own virtual characters or turning photos they have taken into artworks. However, it is difficult to make high technology and applications accessible universally for all people. First, the operation of creating an artwork based on a photo should be convenient and easy. Second, the method itself should have a small model size and a short inference time. Last but not least, the aesthetic quality should be acceptable to a relevant proportion of people.

6.3.2 Education Requirements. If the virtual artworks are visible but untouchable, that reduces subjective feelings: real artworks give a more direct sensory experience. When talking about direct sensory experience, painting artwork by oneself must be the act that gives the most comprehensive sensory experience. However, learning to paint from scratch is so difficult that most people do not know how to start. Not everyone who likes to paint needs or wants to go to school to learn how. Learning to paint by referring to videos or websites is popular; even so, it is not convenient for people who want to paint a certain artwork. Imagining that an application in your mobile phone can generate any artwork process according to your input: is this not more convenient or interesting? Such AI-aided art education can enrich individualised art education [156], which will bring more opportunities and possibilities for art education.

6.3.3 Art Diversity. AI technologies bring diversity and possibility for all kinds of art. GAN-based methods in particular have made a visual feast of style transfer or feature texture fusion. In traditional art history, it is always humans that create and present art. In this AI era, can computers really create art and diversify the presentation of art, differentiating from human art? As Hertzmann gave a viewpoint, computers cannot make art [57] because they have no creation, motivation or emotion, but people do. In addressing the motivation and emotion of computers, we may have a long way to go, and it is not only the issue of AI artworks. Is it impossible for AI to create enriching forms of art and occupy a place in art history? The answer is no! We can, at least, make efforts to apply collaborative intelligence to the creation of digital art. As mentioned in [149], humans should collaborate with AI so that, when creating a new artwork, we have a clear motivation and emotion, and even create an amazing artwork out of our imagination. Meanwhile, Cécile Paris pointed out that collaborative intelligence is the next scientific frontier of digital transformation [153]. It must be an interesting task to achieve the collaboration of AI and human artists to create a new form of art, and collaborative intelligence must do something wonderful in this task [3].

6.3.4 Commercial Values. Since AI artworks can be used in many scenarios, it is necessary to discuss the value of AI artworks. The work [12] proposed that the novelty of AI art should be taken into account when we talk about the values of this type of artwork in the context of art history. This type of art as *generative art* [30] has been extensively theoretically and practically explored in the last few decades [29]. Recently, Chohan [17] noted that there is a category of blockchain-based virtual assets known as non-fungible tokens (NFTs), attracting an incredible amount of interest from investors in a very recent and short period. Digital artworks can be added to the growing list of uses for the blockchain technology that is now becoming a part of modern life in application such as accounting and auditing, agriculture, AI, business supply chains, and creative and artistic endeavours [138]. Researchers also investigated the price value of machine-made artworks compared with man-made artworks by user studies [59]. The work found that man-made and machine-made artworks are not judged equivalent in their artistic value. The authors pointed out when the participants are told that the artworks are made by machines, then the evaluation is not influenced compared with participants not knowing. We can predict that AI artworks can be traded online and offline in the future, and people have a stable evaluation of artworks. Of course, we should take into account that the sale and subsequent reaction to the work resurrect venerable questions regarding autonomy, authorship, authenticity and intention in computer-generated art [104].

6.3.5 AI Evaluation for AI Artworks. Inspired by [22, 70, 112], we focus on making a unified evaluation system for AI artworks. Note that the unified system contains several items (color, content, stroke texture, style and beauty), and for a certain type of artwork, certain items should be chosen. For example, line drawings without color design should choose content, stroke texture, style and beauty without the color item. We conduct a comparable experiment to find out the relation of the six items and different types of artworks. We first design the user study with all the artworks put together, composing the questionnaire. We then compose the second questionnaire by classifying the artworks according to art types. In these two questionnaires, the evaluation items are the same. From the analysis of Section 5, we determine that the six evaluation items are reasonable, and for different types of artworks, certain items gain very low scores, demonstrating that they are inappropriate for that type. We propose a unified evaluation system for AI artworks, where the items are flexible and are to be chosen for a certain type of artwork. This six-dimensional evaluation index is able to cover many types of AI artworks as well as assign the abstract aesthetic evaluation into several concrete dimensions. However, it is still not enough to cover all kinds of AI artworks, and it needs to be developed into a more objective evaluation system based on computational aesthetics in the future.

7 Conclusion

We have investigated current learning-based methods for AI artworks and classified the methods according to art styles. In particular, we first classified the methods into style-transform methods and art-style-reconstruction methods according to the artwork generation process. For the style-transform field, we further classified the methods as neural-style transfer, GAN-based, and diffusion-model-based. For art-style-reconstruction methods, we classified the methods according to the traditional artistic art style of the generated results, such as line drawing, oil painting, ink wash painting, pastel painting, and the more specialized robot paintings. Furthermore, we proposed a consistent evaluation (based on previous works) for AI artworks and conducted a user study to evaluate the proposed AI artwork evaluation system. This evaluation system contains six items: beauty, color, texture, content detail, line, and style. The user study demonstrates that this evaluation system is suitable for different styles of artwork. This consistent evaluation system containing six items is sufficiently flexible to enable the selection of certain items when evaluating different styles of artwork. There are many more art styles than those considered in this paper, and we hope that, in the future, further art styles will be generated and more methods can be evaluated by a unified evaluation system.

References

[1] Mordvintsev Alexander, Olah Christopher, and Tyka Mike. 2015. Inceptionism: Going Deeper into Neural Networks.

[2] Jie An, Siyu Huang, Yibing Song, Dejing Dou, Wei Liu, and Jiebo Luo. 2021. ArtFlow: Unbiased Image Style Transfer via Reversible Neural Flows. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. Virtual, 862–871.

[3] Ivan V. Bajić, Weisi Lin, and Yonghong Tian. 2021. Collaborative Intelligence: Challenges and Opportunities. In *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. 8493–8497.

[4] Dominik Balazka and Dario Rodighiero. 2020. Big data and the little big bang: an epistemological (R) evolution. *Frontiers in big Data* 3 (2020), 31.

[5] Guillaume Berger and R. Memisevic. 2017. Incorporating long-range consistency in CNN-based texture generation. In *International Conference on Learning Representations*. 1–10.

[6] Ayan Kumar Bhunia, Pinaki Nath Chowdhury, Yongxin Yang, Timothy M. Hospedales, Tao Xiang, and Yi-Zhe Song. 2021. Vectorization and Rasterization: Self-Supervised Learning for Sketch and Handwriting. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. Virtual, 5672–5681.

[7] Ardavan Bidgoli, Manuel Ladron De Guevara, Cinnie Hsiung, Jean Oh, and Eunsu Kang. 2020. Artistic Style in Robotic Painting; a Machine Learning Approach to Learning Brushstroke from Human Artists. In *IEEE International Conference on Robot and Human Interactive Communication*. 412–418.

[8] Andreas Blattmann, Robin Rombach, Kaan Oktay, Jonas Müller, and Björn Ommer. [n. d.]. Semi-Parametric Neural Image Synthesis. In *Thirty-sixth Conference on Neural Information Processing Systems*.

[9] Benito Buchheim, Max Reimann, Sebastian Pasewaldt, Jürgen Döllner, and Matthias Trapp. 2021. StyleTune: Interactive Style Transfer Enhancement on Mobile Devices. In *ACM SIGGRAPH 2021 Appy Hour* (Virtual Event, USA) *(SIGGRAPH '21)*. Association for Computing Machinery, New York, NY, USA, Article 8, 2 pages.

[10] Jianlu Cai, Frederick WB Li, Fangzhe Nan, and Bailin Yang. 2024. Multi-style cartoonization: Leveraging multiple datasets with generative adversarial networks. *Computer Animation and Virtual Worlds* 35, 3 (2024), e2269.

[11] Nan Cao, Xin Yan, Yang Shi, and Chaoran Chen. 2019. AI-Sketcher : A Deep Generative Model for Producing High-Quality Sketches. *Proceedings of the AAAI Conference on Artificial Intelligence* 33, 01 (Jul. 2019), 2564–2571.

[12] Eva Cetinic and James She. 2022. Understanding and Creating Art with AI: Review and Outlook. *ACM Trans. Multimedia Comput. Commun. Appl.* 18, 2, Article 66 (feb 2022), 22 pages.

[13] Caroline Chan, Frédo Durand, and Phillip Isola. 2022. Learning to generate line drawings that convey geometry and semantics. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 7915–7925.

[14] Xinyuan Chen, Chang Xu, Xiaokang Yang, Li Song, and Dacheng Tao. 2019. Gated-GAN: Adversarial Gated Networks for Multi-Collection Style Transfer. *IEEE Transactions on Image Processing* 28, 2 (2019), 546–560.

[15] Yang Chen, Yu-Kun Lai, and Yong-Jin Liu. 2018. CartoonGAN: Generative Adversarial Networks for Photo Cartoonization. In *Conference on Computer Vision and Pattern Recognition*. Los Alamitos, CA, USA, 9465–9474.

[16] Peter Childs, Ji Han, Liuqing Chen, Pingfei Jiang, Pan Wang, Dongmyung Park, Yuan Yin, Elena Dieckmann, and Ignacio Vilanova. 2022. The creativity diamond—a framework to aid creativity. *Journal of Intelligence* 10, 4 (2022), 73.

[17] Usman W. Chohan. 2021. Non-Fungible Tokens: Blockchains, Scarcity, and Value. *Critical Blockchain Research Initiative (CBRI) Working Papers* (March 2021), 1–14.
[18] Yunjey Choi, Minje Choi, Munyoung Kim, Jung-Woo Ha, Sunghun Kim, and Jaegul Choo. 2018. Stargan: Unified generative adversarial networks for multi-domain image-to-image translation. In *Proceedings of the IEEE conference on computer vision and pattern recognition*. 8789–8797.
[19] Yunjey Choi, Youngjung Uh, Jaejun Yoo, and Jung-Woo Ha. 2020. Stargan v2: Diverse image synthesis for multiple domains. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. 8188–8197.
[20] Min Jin Chong and David Forsyth. 2021. GANs N' Roses: Stable, Controllable, Diverse Image to Image Translation (works for videos too!). arXiv:2106.06561 [cs.CV]
[21] Jiwoo Chung, Sangeek Hyun, and Jae-Pil Heo. 2024. Style Injection in Diffusion: A Training-free Approach for Adapting Large-scale Diffusion Models for Style Transfer. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 8795–8805.
[22] Y. Deng, F. Tang, W. Dong, C. Ma, F. Huang, O. Deussen, and C. Xu. 2020. Exploring the Representativity of Art Paintings. *IEEE Transactions on Multimedia* (2020), 1–12.
[23] Yingying Deng, Fan Tang, Weiming Dong, Chongyang Ma, Xingjia Pan, Lei Wang, and Changsheng Xu. 2022. StyTr2: Image Style Transfer With Transformers. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. 11326–11336.
[24] Yingying Deng, Fan Tang, Weiming Dong, Wen Sun, Feiyue Huang, and Changsheng Xu. 2020. Arbitrary Style Transfer via Multi-Adaptation Network *(MM '20)*. Association for Computing Machinery, New York, NY, USA, 2719–2727.
[25] Oliver Deussen, Stefan Hiller, Cornelius Van Overveld, and Thomas Strothotte. 2000. Floating points: A method for computing stipple drawings. *Computer Graphics Forum* 19 (2000).
[26] O. Deussen and Tobias Isenberg. 2013. Halftoning and Stippling. In *Image and Video-Based Artistic Stylisation*, Vol. 42. 45–61.
[27] Oliver Deussen and Thomas Strothotte. 2000. Computer-Generated Pen-and-Ink Illustration of Trees. In *Proceedings of the 27th Annual Conference on Computer Graphics and Interactive Techniques (SIGGRAPH '00)*. ACM Press/Addison-Wesley Publishing Co., USA, 13–18.
[28] Prafulla Dhariwal and Alexander Nichol. 2021. Diffusion models beat gans on image synthesis. *Advances in Neural Information Processing Systems* 34 (2021), 8780–8794.
[29] Alan Dorin. 2013. Chance and Complexity: Stochastic and Generative Processes in Art and Creativity. In *Proceedings of the Virtual Reality International Conference: Laval Virtual* (Laval, France) *(VRIC '13)*. Association for Computing Machinery, New York, NY, USA, Article 19, 8 pages.
[30] Alan Dorin, Jonathan McCabe, Jon McCormack, Gordon Monro, and Mitchell Whitelaw. 2012. A framework for understanding generative art. *Digital Creativity* 23, 3-4 (2012), 239–259.
[31] Gershon Elber and George Wolberg. 2003. Rendering traditional mosaics. *The Visual Computer* 19, 1 (2003), 67–78.
[32] Ahmed M. Elgammal, Bingchen Liu, Mohamed Elhoseiny, and Marian Mazzone. 2017. CAN: Creative Adversarial Networks, Generating "Art" by Learning About Styles and Deviating from Style Norms, In ICCC. *IEEE Intelligent Systems*.
[33] Chia-Hui Feng, Yu-Chun Lin, Yu-Hsiu Hung, Chao-Kuang Yang, Liang-Chi Chen, Shih-Wei Yeh, and Shih-Hao Lin. 2020. Research on Aesthetic Perception of Artificial Intelligence Style Transfer. In *HCI International 2020 - Posters*, Constantine Stephanidis and Margherita Antona (Eds.). Springer International Publishing, Cham, 641–649.
[34] Fenghui Yao and Guifeng Shao. 2005. Painting brush control techniques in Chinese painting robot. In *IEEE International Workshop on Robot and Human Interactive Communication*. 462–467.
[35] Tsu-Jui Fu, Xin Eric Wang, and William Yang Wang. 2022. Language-Driven Artistic Style Transfer. In *Computer Vision – ECCV 2022*, Shai Avidan, Gabriel Brostow, Moustapha Cissé, Giovanni Maria Farinella, and Tal Hassner (Eds.). Springer Nature Switzerland, Cham, 717–734.
[36] G. Winkenbach and D. Salesin. 1996. Rendering parametric surfaces in pen and ink. In *ACM SIGGRAPH*. 469–476.
[37] Shunryu Colin Garvey. 2021. The "General Problem Solver" Does Not Exist: Mortimer Taube and the Art of AI Criticism. *IEEE Annals of the History of Computing* 43, 1 (2021), 60–73.
[38] Leon A. Gatys, Alexander S. Ecker, and Matthias Bethge. 2015. A Neural Algorithm of Artistic Style. *ArXiv* abs/1508.06576 (2015).
[39] Leon A. Gatys, Alexander S. Ecker, and Matthias Bethge. 2016. Image Style Transfer Using Convolutional Neural Networks. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. LAS VEGAS, 2414–2423.
[40] Leon A. Gatys, Alexander S. Ecker, Matthias Bethge, Aaron Hertzmann, and Eli Shechtman. 2017. Controlling Perceptual Factors in Neural Style Transfer. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. Honolulu, HI, USA, 3985–3993.
[41] Songwei Ge, Vedanuj Goswami, Larry Zitnick, and Devi Parikh. 2021. Creative Sketch Generation. In *International Conference on Learning Representations*. OpenReview.net, Vienna, Austria, 1–26.
[42] Ian J. Goodfellow, Jean Pouget-Abadie, M. Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron C. Courville, and Yoshua Bengio. 2014. Generative Adversarial Nets. In *Conference on Neural Information Processing Systems*.
[43] Jörg Marvin Gülzow, Liat Grayver, and Oliver Deussen. 2018. Self-Improving Robotic Brushstroke Replication. *Arts* 7, 4 (2018).

[44] Chao Guo, Tianxiang Bai, Xiao Wang, Xiangyu Zhang, Yue Lu, Xingyuan Dai, and Fei-Yue Wang. 2022. ShadowPainter: Active learning enabled robotic painting through visual measurement and reproduction of the artistic creation process. *Journal of Intelligent & Robotic Systems* 105, 3 (2022), 61.

[45] Xiaoying Guo, Yuhua Qian, Liang Li, and Akira Asano. 2018. Assessment model for perceived visual complexity of painting images. *Knowledge-Based Systems* 159 (2018), 110–119.

[46] Agrim Gupta, Justin Johnson, Alexandre Alahi, and Li Fei-Fei. 2017. Characterizing and Improving Stability in Neural Style Transfer. In *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*. Venice, Italy, 4067–4076.

[47] David Ha and Douglas Eck. 2018. A Neural Representation of Sketch Drawings. In *International Conference on Learning Representations*. OpenReview.net, Vancouver, BC, Canada, 1–16.

[48] Paul Haeberli. 1990. Paint by Numbers: Abstract Image Representations. *ACM SIGGRAPH Computer Graphics* 24, 4 (1990), 207–214.

[49] Jun Hao Liew, Hanshu Yan, Daquan Zhou, and Jiashi Feng. 2022. MagicMix: Semantic Mixing with Diffusion Models. *arXiv e-prints* (2022), arXiv–2210.

[50] Kamyar Hazeri, Peter Childs, David Cropley, et al. 2017. Proposing a new product creativity assessment tool and a novel methodology to investigate the effects of different types of product functionality on the underlying structure of factor analysis. In *DS 87-8 Proceedings of the 21st International Conference on Engineering Design (ICED 17) Vol 8: Human Behaviour in Design, Vancouver, Canada, 21-25.08. 2017*. 579–588.

[51] Amir Hertz, Ron Mokady, Jay Tenenbaum, Kfir Aberman, Yael Pritch, and Daniel Cohen-Or. 2022. Prompt-to-prompt image editing with cross attention control. *arXiv preprint arXiv:2208.01626* (2022).

[52] Aaron Hertzmann. 1998. Painterly Rendering with Curved Brush Strokes of Multiple Sizes. In *ACM SIGGRAPH*. 453–460.

[53] Aaron Hertzmann. 2002. Fast Paint Texture. In *International Symposium on Non-Photorealistic Animation and Rendering*. 91–97.

[54] Aaron Hertzmann. 2003. A survey of stroke-based rendering. *IEEE Computer Graphics and Applications* 23 (2003), 70–81.

[55] Aaron Hertzmann. 2010. Non-Photorealistic Rendering and the Science of Art. In *Proceedings of the 8th International Symposium on Non-Photorealistic Animation and Rendering* (Annecy, France) *(NPAR '10)*. Association for Computing Machinery, New York, NY, USA, 147–157.

[56] Aaron Hertzmann. 2018. Can Computers Create Art? *Arts* 7, 2 (2018).

[57] Aaron Hertzmann. 2020. Computers Do Not Make Art, People Do. *Commun. ACM* 63, 5 (April 2020), 45–48.

[58] Jonathan Ho, Ajay Jain, and Pieter Abbeel. 2020. Denoising diffusion probabilistic models. *Advances in Neural Information Processing Systems* 33 (2020), 6840–6851.

[59] Joo-Wha Hong and Nathaniel Ming Curran. 2019. Artificial Intelligence, Artists, and Art: Attitudes Toward Artwork Produced by Humans vs. Artificial Intelligence. *ACM Trans. Multimedia Comput. Commun. Appl.* 15, 2s, Article 58 (July 2019), 16 pages.

[60] Kibeom Hong, Seogkyu Jeon, Junsoo Lee, Namhyuk Ahn, Kunhee Kim, Pilhyeon Lee, Daesik Kim, Youngjung Uh, and Hyeran Byun. 2023. AesPA-Net: Aesthetic Pattern-Aware Style Transfer Networks. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*. 22758–22767.

[61] Zhiyuan Hu, Jia Jia, Bei Liu, Yaohua Bu, and Jianlong Fu. 2020. Aesthetic-Aware Image Style Transfer. In *Proceedings of the 28th ACM International Conference on Multimedia* (Seattle, WA, USA) *(MM '20)*. Association for Computing Machinery, New York, NY, USA, 3320–3329.

[62] Haozhi Huang, Hao Wang, Wenhan Luo, Lin Ma, Wenhao Jiang, Xiaolong Zhu, Zhifeng Li, and Wei Liu. 2017. Real-Time Neural Style Transfer for Videos. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. Honolulu, HI, USA, 783–791.

[63] Xun Huang and Serge Belongie. 2017. Arbitrary Style Transfer in Real-Time With Adaptive Instance Normalization. In *Proceedings of the IEEE International Conference on Computer Vision*.

[64] Zhewei Huang, Wen Heng, and Shuchang Zhou. 2019. Learning to Paint With Model-Based Deep Reinforcement Learning. *International Conference on Computer Vision* (2019), 8708–8717.

[65] Aapo Hyvärinen and Peter Dayan. 2005. Estimation of non-normalized statistical models by score matching. *Journal of Machine Learning Research* 6, 4 (2005).

[66] Phillip Isola, Jun-Yan Zhu, Tinghui Zhou, and Alexei A. Efros. 2017. Image-To-Image Translation With Conditional Adversarial Networks. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. Honolulu, HI, USA, 1125–1134.

[67] Jaeseok Jeong, Mingi Kwon, and Youngjung Uh. 2024. Training-Free Content Injection Using H-Space in Diffusion Models. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*. 5151–5161.

[68] Biao Jia, Jonathan Brandt, Radomír Mech, Byungmoon Kim, and Dinesh Manocha. 2019. LPaintB: Learning to Paint from Self-Supervision.. In *Pacific Graphics*. 1–7.

[69] 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 *European Conference on Computer Vision*. 1–17.

[70] Y. Jing, Y. Yang, Z. Feng, J. Ye, Y. Yu, and M. Song. 2020. Neural Style Transfer: A Review. *IEEE Transactions on Visualization and Computer Graphics* 26, 11 (2020), 3365–3385.

[71] Justin Johnson, Alexandrel Alahi, and Li Fei-Fei. 2016. Perceptual Losses for Real-Time Style Transfer and Super-Resolution. In *Computer Vision – ECCV 2016*, Bastian Leibe, Jiri Matas, Nicu Sebe, and Max Welling (Eds.). Springer International Publishing, Cham, 694–711.

[72] Chanyong Jung, Gihyun Kwon, and Jong Chul Ye. 2022. Exploring Patch-Wise Semantic Relation for Contrastive Learning in Image-to-Image Translation Tasks. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 18260–18269.

[73] Evangelos Kalogerakis, Derek Nowrouzezahrai, Simon Breslav, and Aaron Hertzmann. 2012. Learning Hatching for Pen-and-Ink Illustration of Surfaces. *ACM Trans. Graph.* 31, 1, Article 1 (Feb. 2012), 17 pages.

[74] Moritz Kampelmuhler and Axel Pinz. 2020. Synthesizing human-like sketches from natural images using a conditional convolutional decoder. In *Winter Conference on Applications of Computer Vision (WACV)*. The Westin Snowmass Resort in Snowmass village, Colorado, 3203–3211.

[75] Artur Karimov, Ekaterina Kopets, Sergey Leonov, Lorenzo Scalera, and Denis Butusov. 2023. A robot for artistic painting in authentic colors. *Journal of Intelligent & Robotic Systems* 107, 3 (2023), 34.

[76] Gwanghyun Kim, Taesung Kwon, and Jong Chul Ye. 2022. DiffusionCLIP: Text-Guided Diffusion Models for Robust Image Manipulation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 2426–2435.

[77] Junho Kim, Minjae Kim, Hyeonwoo Kang, and Kwanghee Lee. 2020. U-GAT-IT: Unsupervised Generative Attentional Networks with Adaptive Layer-Instance Normalization for Image-to-Image Translation. In *Conference on Computer Vision and Pattern Recognition*. OpenReview.net, Virtual, 1–19.

[78] Diederik P Kingma and Max Welling. 2014. Auto-Encoding Variational Bayes. arXiv:1312.6114 [stat.ML]

[79] Dmytro Kotovenko, Matthias Wright, Arthur Heimbrecht, and Bjorn Ommer. 2021. Rethinking Style Transfer: From Pixels to Parameterized Brushstrokes. In *Conference on Computer Vision and Pattern Recognition*. Virtual, 12196–12205.

[80] Gihyun Kwon and Jong Chul Ye. 2023. Diffusion-based Image Translation using disentangled style and content representation. In *The Eleventh International Conference on Learning Representations*. 1–22.

[81] Jan Eric Kyprianidis, John Collomosse, Tinghuai Wang, and Tobias Isenberg. 2013. State of the "Art": A Taxonomy of Artistic Stylization Techniques for Images and Video. *IEEE Transactions on Visualization and Computer Graphics* 19, 5 (2013), 866–885.

[82] Yu-Chi Lai, Bo-An Chen, Kuo-Wei Chen, Wei-Lin Si, Chih-Yuan Yao, and Eugene Zhang. 2017. Data-Driven NPR Illustrations of Natural Flows in Chinese Painting. *IEEE Transactions on Visualization and Computer Graphics* 23, 12 (2017), 2535–2549.

[83] Hochang Lee, Sanghyun Seo, Seungtaek Ryoo, Keejoo Ahn, and Kyunghyun Yoon. 2013. A multi-level depiction method for painterly rendering based on visual perception cue. *Multimedia Tools and Applications* 64, 2 (2013), 277–292.

[84] Sangyun Lee. 2022. DALLE-2. (2022).

[85] M. Li, Z. Lin, R. Mech, E. Yumer, and D. Ramanan. 2019. Photo-Sketching: Inferring Contour Drawings From Images. In *IEEE Winter Conference on Applications of Computer Vision*. 1403–1412.

[86] Y. Li and G. Baciu. 2021. SG-GAN: Adversarial Self-Attention GCN for Point Cloud Topological Parts Generation. *IEEE Transactions on Visualization and Computer Graphics* (2021), 1–14.

[87] Y. Li, C. Fang, A. Hertzmann, E. Shechtman, and M. Yang. 2019. Im2Pencil: Controllable Pencil Illustration From Photographs. In *IEEE Conference on Computer Vision and Pattern Recognition*. Long Beach, California, 1525–1534.

[88] Yijun Li, Chen Fang, Jimei Yang, Zhaowen Wang, Xin Lu, and Ming-Hsuan Yang. 2017. Universal Style Transfer via Feature Transforms. In *Conference on Neural Information Processing Systems*. Long Beach Convention, 385–395.

[89] Yi Li, Yi-Zhe Song, Timothy M. Hospedales, and Shaogang Gong. 2017. Free-Hand Sketch Synthesis with Deformable Stroke Models. *International Journal of Computer Vision* 122, 1 (2017), 169–190.

[90] Torrin M. Liddell and John K. Kruschke. 2018. Analyzing ordinal data with metric models: What could possibly go wrong? *Journal of Experimental Social Psychology* 79 (2018), 328–348.

[91] Timothy P. Lillicrap, Jonathan J. Hunt, Alexander Pritzel, Nicolas Heess, Tom Erez, Yuval Tassa, David Silver, and Daan Wierstra. 2015. Continuous control with deep reinforcement learning. *CoRR* arXiv:1509.02971 (2015). arXiv:1509.02971

[92] T. Lindemeier, J. M. Gülzow, and O. Deussen. 2018. Painterly rendering using limited paint color palettes. In *Proceedings of the Conference on Vision, Modeling, and Visualization* (Stuttgart, Germany) *(EG VMV '18)*. Eurographics Association, Goslar, DEU, 135–145.

[93] Thomas Lindemeier, Jens Metzner, Lena Pollak, and Oliver Deussen. 2015. Hardware-Based Non-Photorealistic Rendering Using a Painting Robot. *Comput. Graph. Forum* 34, 2 (may 2015), 311–323.

[94] Songhua Liu, Tianwei Lin, Dongliang He, Fu Li, Ruifeng Deng, Xin Li, Errui Ding, and Hao Wang. 2021. Paint Transformer: Feed Forward Neural Painting with Stroke Prediction. In *IEEE International Conference on Computer Vision*. 6598–6607.

[95] Songhua Liu, Tianwei Lin, Dongliang He, Fu Li, Meiling Wang, Xin Li, Zhengxing Sun, Qian Li, and Errui Ding. 2021. AdaAttN: Revisit Attention Mechanism in Arbitrary Neural Style Transfer. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*. 6649–6658.

[96] Shao Liu, Jiaqi Yang, Sos S. Agaian, and Changhe Yuan. 2021. Novel features for art movement classification of portrait paintings. *Image and Vision Computing* 108 (2021), 104–121.

[97] S. Liu and T. Zhu. 2021. Structure-Guided Arbitrary Style Transfer for Artistic Image and Video. *IEEE Transactions on Multimedia* (2021), 1–14.

[98] Xiao-Chang Liu, Yu-Chen Wu, and Peter Hall. 2023. Painterly Style Transfer With Learned Brush Strokes. *IEEE Transactions on Visualization and Computer Graphics* (2023), 1–12.

[99] Zhi-Song Liu, Li-Wen Wang, Wan-Chi Siu, and Vicky Kalogeiton. 2022. Name Your Style: An Arbitrary Artist-aware Image Style Transfer. *arXiv preprint arXiv:2202.13562* (2022).

[100] Fujun Luan, Sylvain Paris, Eli Shechtman, and Kavita Bala. 2017. Deep Photo Style Transfer. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. Honolulu, HI, USA, 4990–4998.

[101] Kumar M. P. Pavan, B. Poornima, H. S. Nagendraswamy, and C. Manjunath. 2019. A comprehensive survey on non-photorealistic rendering and benchmark developments for image abstraction and stylization. *Iran Journal of Computer Science* 2 (May 2019), 131–165.

[102] Birgit Mallon, Christoph Redies, and Gregor Hayn-Leichsenring. 2014. Beauty in abstract paintings: perceptual contrast and statistical properties. *Frontiers in Human Neuroscience* 8 (2014), 161.

[103] Regan L. Mandryk, David Mould, and Hua Li. 2011. Evaluation of Emotional Response to Non-Photorealistic Images. In *Proceedings of the ACM SIGGRAPH/Eurographics Symposium on Non-Photorealistic Animation and Rendering* (Vancouver, British Columbia, Canada) *(NPAR '11)*. Association for Computing Machinery, New York, NY, USA, 7–16.

[104] Jon McCormack, Toby Gifford, and Patrick Hutchings. 2019. Autonomy, Authenticity, Authorship and Intention in Computer Generated Art. In *Computational Intelligence in Music, Sound, Art and Design*, Anikó Ekárt, Antonios Liapis, and María Luz Castro Pena (Eds.). Springer International Publishing, Cham, 35–50.

[105] John F. J. Mellor, Eunbyung Park, Yaroslav Ganin, I. Babuschkin, T. Kulkarni, Dan Rosenbaum, Andy Ballard, T. Weber, Oriol Vinyals, and S. Eslami. 2019. Unsupervised Doodling and Painting with Improved SPIRAL. In *Neural Information Processing Systems Workshops*.

[106] Elzundefined Sigutundefined Mikalonytundefined and Markus Kneer. 2022. Can Artificial Intelligence Make Art?: Folk Intuitions as to whether AI-driven Robots Can Be Viewed as Artists and Produce Art. *J. Hum.-Robot Interact.* 11, 4, Article 43 (sep 2022), 19 pages.

[107] Alan Moore. 2018. *Do Design: Why beauty is key to everything*. Do Books.

[108] David Mould. 2014. Authorial Subjective Evaluation of Non-Photorealistic Images. In *Proceedings of the Workshop on Non-Photorealistic Animation and Rendering* (Vancouver, British Columbia, Canada) *(NPAR '14)*. Association for Computing Machinery, New York, NY, USA, 49–56.

[109] Reiichiro Nakano. 2019. Neural Painters: A learned differentiable constraint for generating brushstroke paintings. In *Neural Information Processing Systems Workshops*.

[110] Dae Young Park and Kwang Hee Lee. 2019. Arbitrary Style Transfer With Style-Attentional Networks. In *Conference on Computer Vision and Pattern Recognition*. Long Beach, California, 5880–5888.

[111] Aditya Ramesh, Mikhail Pavlov, Gabriel Goh, Scott Gray, Chelsea Voss, Alec Radford, Mark Chen, and Ilya Sutskever. 2021. Zero-shot text-to-image generation. In *International Conference on Machine Learning*. PMLR, 8821–8831.

[112] Jian Ren, Xiaohui Shen, Zhe Lin, Radomir Mech, and David J. Foran. 2017. Personalized Image Aesthetics. In *Proceedings of the IEEE International Conference on Computer Vision*. Venice, Italy, 638–647.

[113] Robin Rombach, Andreas Blattmann, Dominik Lorenz, Patrick Esser, and Björn Ommer. 2022. High-Resolution Image Synthesis With Latent Diffusion Models. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. 10684–10695.

[114] Robin Rombach, Andreas Blattmann, and Björn Ommer. 2022. Text-Guided Synthesis of Artistic Images with Retrieval-Augmented Diffusion Models. *arXiv preprint arXiv:2207.13038* (2022).

[115] Paul Rosin and John Collomosse. 2012. *Image and video-based artistic stylisation*. Vol. 42. Springer Science & Business Media.

[116] Li Ru, Wu Chi-Hao, Liu Shuaicheng, Wang Jue, Wang Guangfu, Liu Guanghui, and Zeng Bing. 2021. SDP-GAN: Saliency Detail Preservation Generative Adversarial Networks for High Perceptual Quality Style Transfer. *IEEE Transactions on Image Processing* 30 (2021), 374–385.

[117] Manuel Ruder, Alexey Dosovitskiy, and Thomas Brox. 2016. Artistic Style Transfer for Videos. In *Pattern Recognition*, Bodo Rosenhahn and Bjoern Andres (Eds.). Springer International Publishing, Cham, 26–36.

[118] Chitwan Saharia, William Chan, Saurabh Saxena, Lala Li, Jay Whang, Emily Denton, Seyed Kamyar Seyed Ghasemipour, Burcu Karagol Ayan, S Sara Mahdavi, Rapha Gontijo Lopes, et al. 2022. Photorealistic Text-to-Image Diffusion Models with Deep Language Understanding. *arXiv preprint arXiv:2205.11487* (2022).

[119] Aneeshan Sain, Ayan Kumar Bhunia, Yongxin Yang, Tao Xiang, and Yi-Zhe Song. 2021. StyleMeUp: Towards Style-Agnostic Sketch-Based Image Retrieval. In *Conference on Computer Vision and Pattern Recognition*. Virtual, 8504–8513.

[120] Michael P. Salisbury, Sean E. Anderson, Ronen Barzel, and David H. Salesin. 1994. Interactive Pen-and-Ink Illustration. In *Proceedings of the 21st Annual Conference on Computer Graphics and Interactive Techniques (SIGGRAPH '94)*. Association for Computing Machinery, New York, NY, USA, 101–108.

[121] R. Bowman Samuel and Vilnis Luke. 2016. Generating Sentences from a Continuous Space. In *Conference on Computational Natural Language Learning*. Berlin, Germany, 10–21.

[122] Anthony Santella and D. DeCarlo. 2002. Abstracted painterly renderings using eye-tracking data. In *International Symposium on Non-Photorealistic Animation and Rendering*. 75–83.

[123] Peter Schaldenbrand and Jean Oh. 2021. Content Masked Loss: Human-Like Brush Stroke Planning in a Reinforcement Learning Painting Agent. In *Proceedings of the AAAI Conference on Artificial Intelligence*, Vol. 35. 505–512.

[124] Bin Sheng, Ping Li, Chenhao Gao, and Kwan-Liu Ma. 2019. Deep Neural Representation Guided Face Sketch Synthesis. *IEEE Transactions on Visualization and Computer Graphics* 25, 12 (2019), 3216–3230.

[125] Yezhi Shu, Ran Yi, Mengfei Xia, Zipeng Ye, Wang Zhao, Yang Chen, Yu-Kun Lai, and Yong-Jin Liu. 2021. GAN-based Multi-Style Photo Cartoonization. *IEEE Transactions on Visualization and Computer Graphics* (2021), 1–14.

[126] Karen Simonyan and Andrew Zisserman. 2015. Very Deep Convolutional Networks for Large-Scale Image Recognition. In *International Conference on Learning Representations*. OpenReview.net, The Hilton San Diego Resort, 1–14.

[127] Jaskirat Singh, Cameron Smith, Jose Echevarria, and Liang Zheng. 2022. Intelli-Paint: Towards developing more human-intelligible painting agents. In *European Conference on Computer Vision*. Springer, 685–701.

[128] Jascha Sohl-Dickstein, Eric Weiss, Niru Maheswaranathan, and Surya Ganguli. 2015. Deep unsupervised learning using nonequilibrium thermodynamics. In *International Conference on Machine Learning*. PMLR, 2256–2265.

[129] Jifei Song, Kaiyue Pang, Yi-Zhe Song, Tao Xiang, and Timothy M. Hospedales. 2018. Learning to Sketch With Shortcut Cycle Consistency. In *Computer Vision and Pattern Recognition*. Los Alamitos, CA, USA, 801–810.

[130] Yaniv Taigman, Adam Polyak, and Lior Wolf. 2016. Unsupervised Cross-Domain Image Generation. *CoRR* abs/1611.02200 (2016). arXiv:1611.02200

[131] Fan Tang, Weiming Dong, Yiping Meng, Xing Guo Mei, Feiyue Huang, Xiaopeng Zhang, and Oliver Deussen. 2018. Animated Construction of Chinese Brush Paintings. *IEEE Transactions on Visualization and Computer Graphics* 24 (2018), 3019–3031.

[132] Hao Tang, Hong Liu, Dan Xu, Philip H. S. Torr, and Nicu Sebe. 2021. AttentionGAN: Unpaired Image-to-Image Translation Using Attention-Guided Generative Adversarial Networks. *IEEE Transactions on Neural Networks and Learning Systems* (2021), 1–16.

[133] Zineng Tang. 2019. Adaptive Aesthetic Photo Filter learning. In *Proceedings of the 2019 3rd International Conference on Virtual and Augmented Reality Simulations* (Perth, WN, Australia) *(ICVARS '19)*. Association for Computing Machinery, New York, NY, USA, 67–72.

[134] The J. Paul Getty Museum/ Education. 2021. Art Vocabulary Words: Elements of Art/ Principles of Design. [Accessed 16-August-2021].

[135] Salimans Tim, Goodfellow Ian, Zaremba Wojciech, and Cheung Vicki. 2016. Improved Techniques for Training GANs. In *Conference on Neural Information Processing Systems*. Barcelona, Spain, 1–9.

[136] Zhengyan Tong, Xuanhong Chen, Bingbing Ni, and Xiaohang Wang. 2021. Sketch Generation with Drawing Process Guided by Vector Flow and Grayscale. In *AAAI Conference on Artificial Intelligence*, Vol. 53. 609–616.

[137] Zhengyan Tong, Xiaohang Wang, Shengchao Yuan, Xuanhong Chen, Junjie Wang, and Xiangzhong Fang. 2022. Im2Oil: Stroke-BasintingPainting Rendering with Linearly Controllable Fineness Via Adaptive Sampling. In *Proceedings of the 30th ACM International Conference on Multimedia* (Lisboa, Portugal) *(MM '22)*. Association for Computing Machinery, New York, NY, USA, 1035–1046.

[138] Lawrence J. Trautman. 2021. Virtual Art and Non-fungible Tokens. *50 Hofstra Law Review, Forthcoming* (April 2021), 1–76.

[139] Patrick Tresset and Frederic Fol Leymarie. 2013. Portrait drawing by Paul the robot. *Computers and Graphics* 37, 5 (2013), 348 – 363.

[140] Ashish Vaswani et al. 2017. Attention is All you Need. In *Advances in Neural Information Processing Systems*. Curran Associates, Inc.

[141] Pascal Vincent. 2011. A Connection Between Score Matching and Denoising Autoencoders. *Neural Computation* 23, 7 (2011), 1661–1674.

[142] J. J. Virtusio, D. S. Tan, W. Cheng, M. Tanveer, and K. Hua. 2020. Enabling Artistic Control Over Pattern Density and Stroke Strength. *IEEE Transactions on Multimedia* (2020), 1–13.

[143] Boheng Wang, Yunhuai Zhu, Liuqing Chen, Jingcheng Liu, Lingyun Sun, and Peter Childs. 2023. A study of the evaluation metrics for generative images containing combinational creativity. *Artificial Intelligence for Engineering Design, Analysis and Manufacturing* 37 (2023), e11.

[144] Huan Wang, Yijun Li, Yuehai Wang, Haoji Hu, and Ming-Hsuan Yang. 2020. Collaborative Distillation for Ultra-Resolution Universal Style Transfer. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 1860–1869.

[145] Qian Wang, Cai Guo, Hong-Ning Dai, and Ping Li. 2023. Stroke-GAN Painter: Learning to paint artworks using stroke-style generative adversarial networks. *Computational Visual Media* 9, 4 (2023), 787–806.

[146] Wenjing Wang, Shuai Yang, Jizheng Xu, and Jiaying Liu. 2020. Consistent Video Style Transfer via Relaxation and Regularization. *IEEE Transactions on Image Processing* 29 (2020), 9125–9139.

[147] Xinrui Wang and Jinze Yu. 2020. Learning to Cartoonize Using White-Box Cartoon Representations. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition*. Virtual, 8090–8099.

[148] B. Wilson and K. Ma. 2004. Rendering complexity in computer-generated pen-and-ink illustrations. In *International Symposium on Non-Photorealistic Animation and Rendering*. 129–137.

[149] H James Wilson and Paul R Daugherty. 2018. Collaborative intelligence: humans and AI are joining forces. *Harvard Business Review* 96, 4 (2018), 114–123.

[150] G. Winkenbach and D. Salesin. 1994. Computer-generated pen-and-ink illustration. In *ACM SIGGRAPH*. 91–100.

[151] Ning Xie, Hirotaka Hachiya, and Masashi Sugiyama. 2013. Artist Agent: A Reinforcement Learning Approach to Automatic Stroke Generation in Oriental Ink Painting. *IEICE Transactions on Information and Systems* E96.D, 5 (2013), 1134–1144.
[152] Ning Xie, Yang Yang, Heng Tao Shen, and Ting Ting Zhao. 2018. Stroke-based stylization by learning sequential drawing examples. *Journal of Visual Communication and Image Representation* 51 (2018), 29 – 39.
[153] China.org Xinhua. 2021. Australian scientists establish platform to combine human, machine intelligence. *China news* (2021).
[154] Kai Xu, Longyin Wen, Guorong Li, Honggang Qi, Liefeng Bo, and Qingming Huang. 2021. Learning Self-Supervised Space-Time CNN for Fast Video Style Transfer. *IEEE Transactions on Image Processing* 30 (2021), 2501–2512.
[155] Tao Xu, Pengchuan Zhang, Qiuyuan Huang, Han Zhang, Zhe Gan, Xiaolei Huang, and Xiaodong He. 2018. AttnGAN: Fine-Grained Text to Image Generation With Attentional Generative Adversarial Networks. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*.
[156] Yunqing Xu, Yi Ji, Peng Tan, Qiaoling Zhong, and Ming Ma. 2021. Intelligent Painting Education Mode Based on Individualized Learning Under the Internet Vision. In *Intelligent Human Systems Integration 2021*, Dario Russo, Tareq Ahram, Waldemar Karwowski, Giuseppe Di Bucchianico, and Redha Taiar (Eds.). Springer International Publishing, Cham, 253–259.
[157] Y. Zhang Y. Zhang and W. Cai. 2020. A Unified Framework for Generalizable Style Transfer: Style and Content Separation. *IEEE Transactions on Image Processing* 29 (2020), 4085–4098.
[158] Lingchen Yang, Lumin Yang, M. Zhao, and Youyi Zheng. 2018. Controlling Stroke Size in Fast Style Transfer with Recurrent Convolutional Neural Network. *Comput. Graph. Forum* 37 (2018), 97–107.
[159] Yuan Yao, Jianqiang Ren, Xuansong Xie, Weidong Liu, Yong-Jin Liu, and Jun Wang. 2019. Attention-Aware Multi-Stroke Style Transfer. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition*. Long Beach, California, 1467–1475.
[160] Meijuan Ye, Shizhe Zhou, and Hongbo Fu. 2019. DeepShapeSketch : Generating hand drawing sketches from 3D objects. *International Joint Conference on Neural Networks* (2019), 1–8.
[161] Ran Yi, Yong-Jin Liu, Yu-Kun Lai, and Paul L. Rosin. 2019. APDrawingGAN: Generating Artistic Portrait Drawings From Face Photos With Hierarchical GANs. In *Conference on Computer Vision and Pattern Recognition*. Long Beach, California, 10743–10752.
[162] Ran Yi, Mengfei Xia, Yong-Jin Liu, Yu-Kun Lai, and Paul L. Rosin. 2021. Line Drawings for Face Portraits From Photos Using Global and Local Structure Based GANs. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 43, 10 (2021), 3462–3475.
[163] Yuan Yin, Kamyar Hazeri, Shafina Vohra, Haoyu Zuo, Shu Huang, Bowen Zhan, Peter Childs, et al. 2022. Using Creativity Levels as a Criterion for Rater Selection in Creativity Assessment. *DS 118: Proceedings of NordDesign 2022, Copenhagen, Denmark, 16th-18th August 2022* (2022), 1–10.
[164] Chiyu Zhang, Jun Yang, Lei Wang, and Zaiyan Dai. 2022. S2WAT: Image Style Transfer via Hierarchical Vision Transformer using Strips Window Attention. *arXiv preprint arXiv:2210.12381* (2022).
[165] Luming Zhang, Yiyang Yao, Zhenguang Lu, and Ling Shao. 2019. Aesthetics-Guided Graph Clustering With Absent Modalities Imputation. *IEEE Transactions on Image Processing* 28, 7 (2019), 3462–3476.
[166] Yuxin Zhang, Nisha Huang, Fan Tang, Haibin Huang, Chongyang Ma, Weiming Dong, and Changsheng Xu. 2023. Inversion-Based Style Transfer With Diffusion Models. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 10146–10156.
[167] Yabin Zhang, Minghan Li, Ruihuang Li, Kui Jia, and Lei Zhang. 2022. Exact Feature Distribution Matching for Arbitrary Style Transfer and Domain Generalization. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 8035–8045.
[168] Yuxin Zhang, Fan Tang, Weiming Dong, Haibin Huang, Chongyang Ma, Tong-Yee Lee, and Changsheng Xu. 2022. Domain Enhanced Arbitrary Image Style Transfer via Contrastive Learning. In *ACM SIGGRAPH 2022 Conference Proceedings* (Vancouver, BC, Canada) *(SIGGRAPH '22)*. Association for Computing Machinery, New York, NY, USA, Article 12, 8 pages.
[169] Tao Zhou, Chen Fang, Zhaowen Wang, Jimei Yang, Byungmoon Kim, Zhili Chen, Jonathan Brandt, and Demetri Terzopoulos. 2018. Learning to Doodle with Deep Q-Networks and Demonstrated Strokes. In *British Machine Vision Conference*. Newcastle Upon Tyne, 1–13.
[170] 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)*. 2223–2232.
[171] Zhengxia Zou, Tianyang Shi, Shuang Qiu, Yi Yuan, and Zhenwei Shi. 2021. Stylized Neural Painting. In *IEEE Conference on Computer Vision and Pattern Recognition*. 15689–15698.

Received 9 August 2023; revised 30 July 2024; accepted 12 September 2024