sustainability

MDPI

Article

Artificial Intelligence-Empowered Art Education: A Cycle-Consistency Network-Based Model for Creating the Fusion Works of Tibetan Painting Styles

Yijing Chen [1], Luqing Wang [2], Xingquan Liu [1,*] and Hongjun Wang [2,*]

[1] School of Art, Southwest Minzu University, Chengdu 610041, China
[2] School of Computing and Artificial Intelligence, Southwest Jiaotong University, Chengdu 611756, China
* liuxingquanswmu@gmail.com (X.L.); wanghongjun@swjtu.edu.cn (H.W.)

Abstract: The integration of Tibetan Thangka and other ethnic painting styles is an important topic of Chinese ethnic art. Its purpose is to explore, supplement, and continue Chinese traditional culture. Restricted by Buddhism and the economy, the traditional Thangka presents the problem of a single style, and drawing a Thangka is time-consuming and labor-intensive. In response to these problems, we propose a Tibetan painting style fusion (TPSF) model based on neural networks that can automatically and quickly integrate the painting styles of the two ethnicities. First, we set up Thangka and Chinese painting datasets as experimental data. Second, we use the training data to train the generator and the discriminator. Then, the TPSF model maps the style of the input image to the target image to fuse the two ethnicities painting styles of Tibetan and Chinese. Finally, to demonstrate the advancement of the proposed method, we add four comparison models to our experiments. At the same time, the Frechet Inception Distance (FID) metric and the questionnaire method were used to evaluate the quality and visual appeal of the generated images, respectively. The experimental results show that the fusion images have excellent quality and great visual appeal.

Keywords: Tibetan Thangka; TPSF model; neural networks; ethnic painting style fusion

Citation: Chen, Y.; Wang, L.; Liu, X.; Wang, H. Artificial Intelligence-Empowered Art Education: A Cycle-Consistency Network-Based Model for Creating the Fusion Works of Tibetan Painting Styles. *Sustainability* **2023**, *15*, 6692. https://doi.org/10.3390/su15086692

Academic Editor: Aras Bozkurt

Received: 29 January 2023
Revised: 7 April 2023
Accepted: 13 April 2023
Published: 15 April 2023

1. Introduction

Thangka is a popular research topic in Chinese ethnic painting, and it usually represents the typical painting style of Tibet and other Tibetan-related areas. In the complex Tibetan Buddhist culture, Thangka stands out in the field of Chinese painting with its long history of development [1,2]. For centuries, other ethnic groups have had close contacts with Tibetans, and Thangkas have constantly merged and drawn on the painting styles of other ethnic groups.

"Thangka" is also called Tangga, which means canvas, and mainly refers to religious scroll paintings mounted and hung in colorful satin. Thangka was introduced from India along with Buddhism in the seventh century. During the Han, Tang, Song and Yuan dynasties, the communication between the Tubo and Han people became closer, which also contributed to the fusion of earlier Thangka and Tubo flag painting [3]. Under the continuous nourishment of the Tibetan cultural background for thousands of years, Thangka presents a unique style, which has been widely inherited and developed. So, it is also known as the "Encyclopedia of Tibetan Culture" [4]. However, with the development of the society and economy, learners have more and more diverse needs for Thangka styles, and the single style of traditional Thangka can no longer meet the needs of the public aesthetic. Meanwhile, it usually takes a long time for learners to learn Thangka style. Compared with other ethnic groups and other forms of painting, Thangka is also difficult to be selected as teaching content in the classroom. As a result, this makes it difficult for many learners to access Thangka, which is extremely detrimental to the inheritance and development of Thangka.

Although most Thangka creators still use traditional forms, more people have become interested in artificial intelligence painting over the past few decades [5]. The trend of using artificial intelligence technology to generate images, fuse image styles, and help people learn to paint, etc., is irreversible [6]. In recent years, deep neural networks have continuously entered the public's field of vision and are widely used in image feature recognition [7], image style fusion [8], image generation [9], etc. Early non-realistic rendering [10] and texture migration [11] are the main traditional style migration methods. Mainstream style transfer models include Generative Adversarial Networks (GAN) [12], Cycle-consistent Generative Adversarial Networks (CycleGAN) [13], Star Generative Adversarial Networks (StarGAN) [14], Star Generative Adversarial Networks v2 (StarGAN V2) [15], Style Generative Adversarial Networks (StyleGAN) [16], Anime Generative Adversarial Networks [17], Conditional Generative Adversarial Networks [18] , Cari Generative Adversarial Networks [19], Adversarial consistence loss-Generative Adversarial Networks [20] and other style migration models.

At present, a large number of scholars use artificial intelligence technology to assist learners in painting creation. For example, deep dream generator assists learners in style transfer, Dall-E2 helps learners generate images through text descriptions, Nvidia Canvas helps learners convert images with abstract strokes into images with realistic photographic effects, AI Gahaku helps learners convert real portraits for abstract painting style effects, and Deoldify AI can assist learners to colorize black and white videos or photos. However, few scholars have studied the fusion of Thangka styles. To fill this gap, we propose a Thangka Painting Style Fusion (TPSF) model based on CycleGAN. The TPSF model can automatically and efficiently integrate the lines, colors, and other forms and content of Tibetan and Chinese paintings, effectively solving the problem of a single Thangka style. At the same time, the TPSF model is easy to operate, which is beneficial to assist learners to understand the style of Thangka, and to effectively solve the problem of the difficulty of teaching Thangka.

How the model is constructed determines the characteristics of the model, and an analysis showed that the TPSF model has the following characteristics. First, the model is stable. As a highly robust network model, the quality of the paintings generated by the TPSF model is stable. Second, the training process of the model is unsupervised learning. The TPSF model does not require a labeling process for the samples, and it can directly be trained and modeled on the data. These characteristics of the TPSF model can facilitate the fusion of Tibetan and Chinese painting styles.

The contributions of this paper are as follows.

- The TPSF model is proposed to solve problems regarding the limited content and the similar styles of Thangka. Additionally, a digital approach to the fusion of Tibetan-Chinese painting styles is provided.
- We propose that the use of a TPSF model in art learning empowers art education and provides learners with a new model of interactive learning.
- Comparison experiments were performed on real data sets. Firstly, the converged objective function proved the feasibility of the model. Secondly, the TPSF model outperformed the other four comparison models according to the Frechet Inception Distance (FID) metric, which proves how advanced it is. Finally, the questionnaire method was used to evaluate the visual appeal of the generated images.

Section 2 of this article describes the relevant work. Section 3 describes the framework design, whereby learners use the model process and the objective function of the TPSF model. At the same time, four groups of comparative experiments, FID metrics and the questionnaire method were used to confirm the progressiveness of the TPSF model and the attractive visual effects of the fusion works.

2. Related Work

The TPSF model is based on artificial intelligence methods to help learners incorporate Thangka styles into other ethnic paintings. Therefore, the related work covers Artificial intelligence in education (AIED) and the study of intelligent methods of dealing with Thangka characteristics and styles.

2.1. Artificial Intelligence Education

In the area of education, scholars have conducted a lot of research on AIED and made remarkable achievements. Timms et al. [21] have used AI technology to enable teachers to work with assistant robots to assist students in teaching, which can help students improve their learning ability. With the continuous promotion and development of computers, computer technology began to permeate into different educational fields from the middle of the 20th century. Specifically, more and more disciplines are beginning to bring computer-assisted teaching and learning methods into classroom interaction [22]. At this stage, the field of AIED is relatively mature, and the field of AI-enabled education has had a huge impact, including the improvement of learning efficiency, personalized learning, and smarter content [21]. Hwang et al. [23] and others studied the definition and function of AIED, proposed the AIED framework and displayed it in different learning and teaching environments, so as to help guide researchers with computer and educational backgrounds to conduct AIED research. At the same time, Darayseh et al. [24] confirmed that artificial intelligence technology has a high degree of acceptance in teachers' teaching, and believed that teachers' use of artificial intelligence can improve their positive attitudes and effectively exert their self-efficacy.

During this period, countless researchers have confirmed the feasibility of artificial intelligence in the field of education. Chen et al. [25] accelerated an intelligent bibliometrics-driven literature analysis by leveraging deep learning for automatic literature screening. By using artificial intelligence to empower education and obtain a new learning model, Chiu et al. [26] launched a deep-learning-based art learning system to help increase students' art appreciation and creativity. Lin et al. [27] developed learners' complex professional skills by using a virtual reality inversion learning method. This interactive simulation technology enables learners to demonstrate higher learning motivation and self-efficacy. In addition, Zhu et al. [28] proposed a high-resolution detail-recovering image-deraining network to effectively increase the quality of deraining images. The above research showed the feasibility of using AI to empower education, and it inspired us to use artificial intelligence to intervene in the fusion of Thangka styles and empower art education.

The fusion of Thangka styles is a specific form of style fusion. Realistic picture effects can only be produced by the accurate semantic analysis of Thangka style characteristics. The authors of several previous works have addressed the challenge of extracting and representing Thangka style features. Ma et al. [29] constructed a small dataset called Chomo Yarlung Tibet version 1, which consisted of images of Tibet Thangkas. Additionally, the dataset was semantically annotated using a deep learning model. Zhang et al. [30] proposed the use of parametric modeling to generate mandala Thangka patterns, which solved the time-consuming and labor-intensive problem of drawing mandalas. Qian et al. [31] proposed a method based on least squares curve fitting to express image contour features. This method solved the problem of extracting and expressing the features of the Thangka Buddha headdress. Liu L et al. [32] proposed a method to repair the damaged images by using similar blocks adjacent to the damaged area of the image. Hu et al. [33] proposed a reference-free quality assessment method, which addresses the quality assessment problem of rendered Thangka images. The authors of the above work focused on studying the characteristics of a single style, which laid the foundation for the fusion of multiple styles.

Many scholars have conducted in-depth research on style fusion. The earliest style fusion is style fusion under the premise of fixed style and content. The idea is simple: treat images as trainable variables and optimize them to reduce differences in content and style between images. After repeated iterations of the model training, the images generated by the generator will tend to be consistent. Gatys et al. [34] studied the art style problem of the fusion of images, and proposed a solution by using the Gram matrix statistics of the optimization method, whereby the deep features are matched. Johnson et al. [35] proposed a feedforward network method to attain the fusion goal, and used this method to approximate the solution to the optimization problem. Risser et al. [36] increased the quality and stability of the image texture by applying a histogram loss and a stronger constraint and larger dispersion of the texture statistical library. Thus, the image texture and style fusion increased. Moreover, Li et al. [37] proposed a Laplace loss, which addresses the loss of image detail in style migration and the "artificiality" of image styles. Li et al. [38] proposed a new explanation for neural style fusions. They theoretically proved that Gram's proof matching is equal to the specified maximum average difference process. The style information in the neural style fusion is essentially represented by the activation distribution in the convolutional neural network, and a style fusion can be achieved through a distribution alignment. Meanwhile, Li et al. think that the fusion of neural styles can be regarded as a domain adaptation problem.

2.2. Image Domain Adaption

Image domain adaptation is a technique that allows the model to behave close to the original domain in the target image domain. It reduces the gap between the two domains in the feature space, which causes the model to be more generalizable and domain-invariant. In recent years, domain adaptation has been explored to fuse image styles. Since Goodfellow et al. [39] proposed the generative confrontation network model in 2014, the model has made remarkable progress in many fields, such as image generation and video prediction. At the same time, it also directly promoted the style fusion milestone. The main inspiration for the GAN came from the idea of a zero-sum game in game theory. The model achieves the goal of Nash equilibrium [40] by training the generation network and the discriminant network, by optimizing the mutual minimax game that occurs between the two. In turn, the generators learn about the distribution of the input data. If the model is used for image generation, after the training is completed, the generators can generate a realistic image from a random number. However, GAN also has the following problems. For example, GAN needs paired samples, and problems regarding unstable training, gradient disappearance, and mode collapse also exist. GAN achieves Nash equilibrium via gradient descent, but this is unstable.

Before Zhu et al. [41] proposed the CycleGan, the use of the Markov random field in image processing, which was proposed by Li et al. [42], segmented the image into interconnected blocks instead of enacting a pixel-to-pixel correspondence. Castillo et al. [43] enhanced the smoothing effect of the boundary between the target object and the background after local fusion by increasing the loss of the Markov field. Based on Markov random field image segmentation, Champandard [44] used the patch algorithms to achieve the style migration of the dissimilar parts of an image and manually labeled the semantic segmentation of the parts. Chen et al. [45] used the partial output to constrain the spatial correspondence, which improves the accuracy of style fusion in the specified region and avoids the influence of non-fused target information and background on the image as much as possible. Lu et al. [46] increased the speed of the style fusion method in the original image domain, which was based on the selection of the feature space. Additionally, the semantic style transformation method with context loss for free segmentation was proposed by Mechrez et al. [47].

In 2017, Zhu et al. [41] and others proposed CycleGAN. Inspired by the "*pix2pix*" idea, the model adjusts the structured idea by using the idea of transferability and cycle-consistent supervised training to obtain a CycleGAN model with dual generators. This model structure solves the limitations of the GAN model's instability and paired samples. The purpose of CycleGAN is to achieve domain matching of the data, which enhances performance by learning a certain number of mappings between non-matching data domains. This satisfies the image generation needs of multistyle fusion. So far, CycleGAN has been widely used in the field of computer vision to enhance image quality, transform image styles, transform objects, etc. With further applications and the increased practice of more scholars, CycleGAN has started become involved in various aspects such as character reidentification [48], multidomain fusion, and game-graphics modeling. Especially in image style fusion, it is more stable and superior. Cycle consistency has also been applied to structured data, such as 3D model matching [49], motion structure [50], three-cone shape matching [51], cosegmentation [52], dense semantic alignment [53], and depth estimation [54] models.

The related work shows that very few methods related to the adaption fusion of Thangka styles exist. Thus, the TPSF model was proposed for Tangka style fusion, which solves the problems caused by the limited content and similar styles of Thangkas. The TPSF model can perfectly integrate the Tibetan painting style with the Chinese painting style, making the experimental work both artistic and scientific. The TPSF model promotes the development of Tibetan intangible cultural heritage with the Thangka and recognizes more ethnic painting expressions. At the same time, it also promotes the fusion of different ethnic cultures.

3. TPSF Model Design

The TPSF model consists of dual generators and discriminators. At the same time, we use a cycle-consistency loss, similar to that used by Zhou et al. [55] and Godar et al. [54], to drive the dual generators G and F in the TPSF model. The TPSF generators are mainly composed of three parts: style encoder, residual block, and style decoder. The style decoder mainly uses convolution, in regularization, leaky rectified linear units (ReLU), as well as image augmentation methods to improve the resolution of the fused image. In addition, the residual block module is also used to enhance the data effect between the style decoder and encoder. In the encoder part, activation functions such as transpose convolution, IN normalization, and the ReLU are used to recover the amount of data used. Then, the image resolution is increased by ReflectionPad2d, and the image is convolved again to return to its original size. This method can solve the problem of processing the edge information of objects. In the objective function part, the cycle-consistency loss function used in the TPSF model is mainly used to limit the image generated by the generator, which can maintain the characteristics of the original image domain. The expected goal of our experiment is to realize the spatial mapping of the real input data of the Tibetan painting image domain X to the real data of the Chinese painting image domain Y and to fuse the styles of different ethnic groups.

We assume that two different image domains are provided in the model, which are the real Thangka style image domain $X = \{x_1, x_2, ...x_N\}$, $x_i \in R$, and the real Chinese painting style image domain $Y = \{y_1, y_2, ...y_N\}$, $y_i \in R$, where R represents the real dataset. And we constructed two ethnic-style-fusion generators G and F for X and Y, respectively, and two additional ethnic style fusion discriminators D_X and D_Y.

3.1. TPSF Generators

The specific operation is shown in Figure 1, and specific information about the TPSF generators is shown in Table 1.

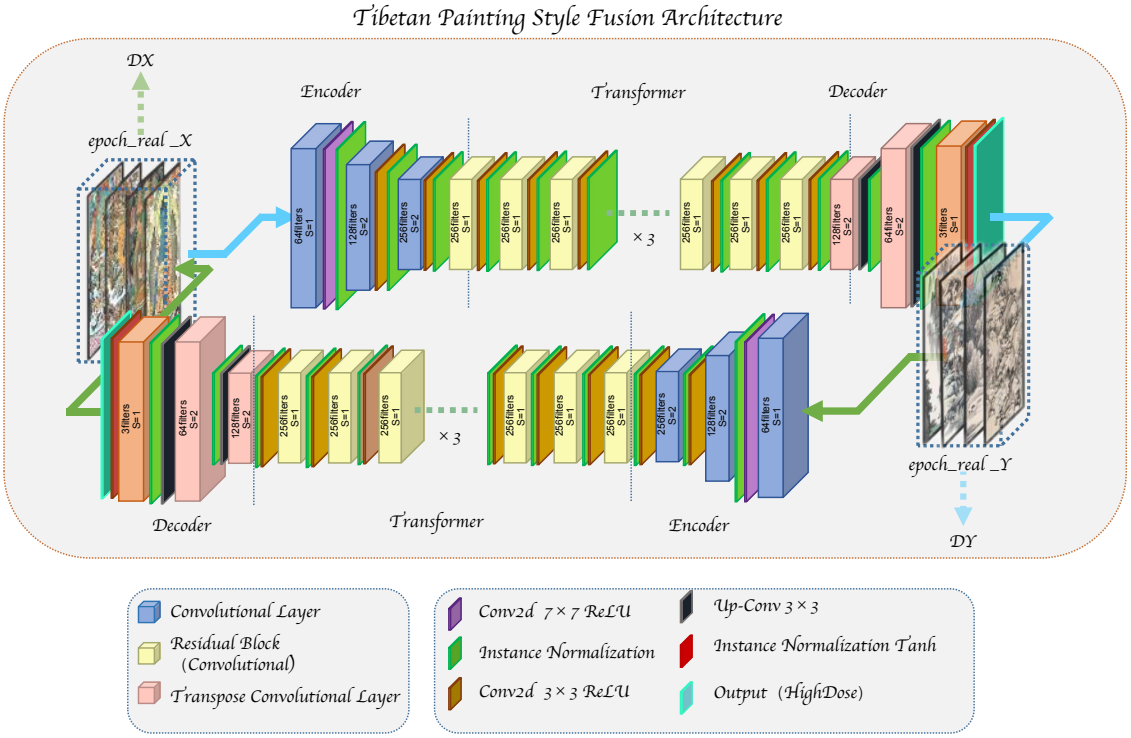

Figure 1. TPSF generators structure diagram: *epoch_real_X* in the image denotes the real input Tangka image set, and *epoch_real_Y* denotes the real input Chinese painting image set.

Table 1. The information table of TPSF generators.

Convolution Number	Kernel	Strides	Padding	Norm	Activation
Conv	(7,7,64)	1	1	InstanceNorm	ReLU
Conv	(3,3,128)	2	1	InstanceNorm	ReLU
Conv	(3,3,256)	2	1	InstanceNorm	ReLU
Resblock	(3,3,256)	1	1	InstanceNorm	ReLU
Transposed Conv	(3,3,128)	2	1	InstanceNorm	ReLU
Transposed Conv	(3,3,64)	2	1	InstanceNorm	ReLU
Transposed Conv	(3,3,64)	1	1	InstanceNorm	ReLU

The following is the detailed design of the TPSF generators' operation.

- The input is two real three-channel image sets that are 256 × 256 pixels and are named *epoch_real_X* and *epoch_real_Y*.
- This image set enters the decoder for undersampling, and the first layer uses a convolution kernel with the number 64 and a size of 7 × 7. The sliding step length is one and the fill size is three. Then, the instance normalization occurs, and the ReLU is finally implemented.
- The second and third layers use 128 and 256 convolutional kernels of size 3 × 3, respectively, and both the second and third layers slide two steps. Additionally, they have a padding size of one, undergo instance normalization, and finally implement ReLU activation.
- The last layers use the nine residual block model. Nine convolutional kernels are present in the residual module, each with 256 3 × 3 convolutional kernels. They slide one step, undergo instance normalization, and finally implement ReLU activation.

3.2. TPSF Discriminators

The specific operation of the TPSF discriminators is shown in Figure 2, the specific information is shown in Table 2.

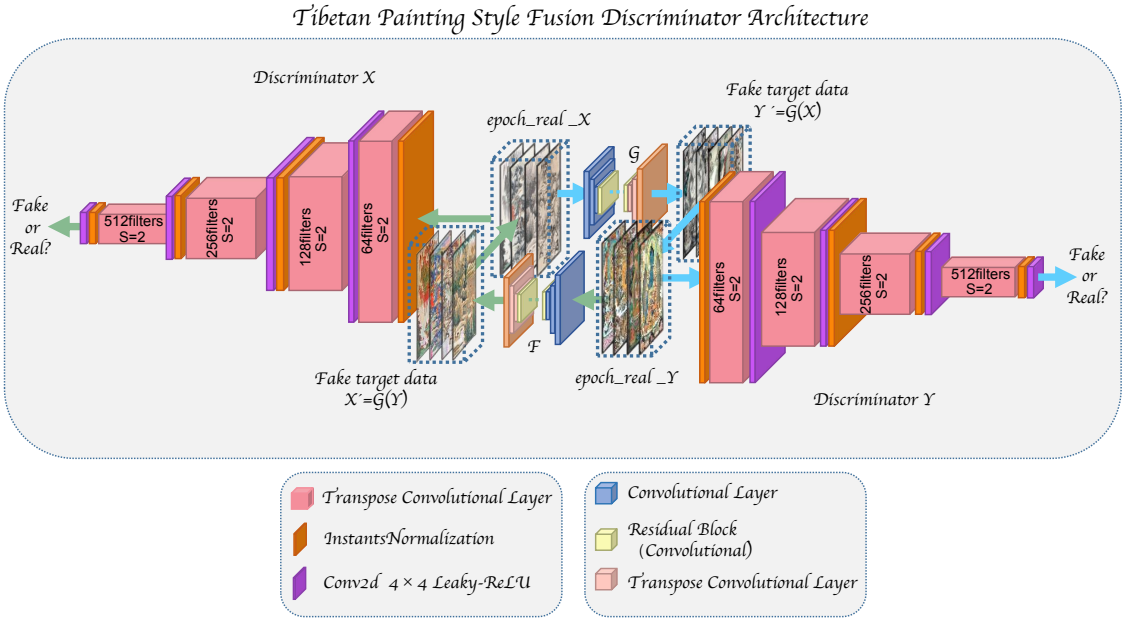

Figure 2. The structure diagram of TPSF discriminator: *epoch_real_X* in the image denotes the real input Tangka image set, and *epoch_real_Y* denotes the real input Chinese painting image set.

Table 2. The information table of the TPSF discriminators.

Convolution Number	Kernel	Strides	Padding
Conv	(4,4,64)	2	1
Conv	(4,4,128)	2	1
Conv	(4,4,256)	2	1
Conv	(4,4,512)	2	1

- The input is a real three-channel image set, each channel of which is 256 × 256 pixels, and are named *epoch_real_X* and *epoch_real_Y*. The output is the fake target data $X' = G(Y)$ and $Y' = G(X)$.
- The first layer has 64 four × four convolutional kernels with a sliding step of two and a fill size of one. Additionally, it undergoes an instance normalization process and, finally, a ReLU activation process.
- The reason for the ReLu activation is that the second to fifth layers all use four × four convolutional kernels, which are 128, 256. That has 521 in number with a sliding step size of two, and padding of one, and they are subjected to average pooling.

3.3. Interactive Learning Process of TPSF Model

The process of learners using the TPSF model to create fusion works is shown in Figure 3. Before the experiment, the learners chose an ethnic painting style that they liked, and we collected such ethnic paintings to create an experimental dataset. Then, we input the selected ethnic painting style dataset and Thangka style dataset into the TPSF model for training. In the experiment, the Thangka and target ethnic painting styles were fused by the dual generator and discriminator of the TPSF model. The fusion works generated automatically, quickly and efficiently through the TPSF model are rich in content and novel in style. After the experiment, the learners selected their favorite images from the fusion works produced by the TPSF model and created their own unique collection of

Thangka style fusion works. The datasets collected from the learners can be shared with other users, which provides more possibilities for more scholars to create Thangka style fusions. Additionally, learners can efficiently generate fusion works with the TPSF model and thereby avoid the influence of low aesthetic experience and creative ability on their works. At the same time, the model can also enable learners to participate in the creation of Thangka painting styles and arouse learners' interest in learning about art.

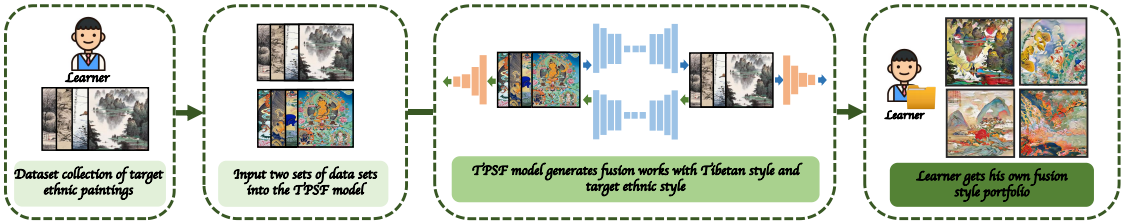

Figure 3. Flow chart of TPSF model interactive learning.

3.4. Objective Function of TPSF Model

The TPSF model consists of three main parts: two sets of adversarial loss, one set of cycle-consistency loss, and an identity loss function. The generative adversarial loss function mainly consists of a GAN, recognition network, and objective function loss, in which the least squares loss is used to replace the original negative log-likelihood loss to ensure the robustness of the objective function and to obtain more experimental results. The specific function is to map two sets of real Tibetan style images to generate and convert the images in the Tibetan style and cause the generated images to be closer to the target images in terms of distribution. The adversarial loss is expressed as

$$L_{GAN}(G, D_Y, X, Y) = E_{y \sim pdata(y)}[\log D_Y(y)] + E_{x \sim pdata(x)}[\log(1 - D_y(G(x)))], \quad (1)$$

where the style fusion generator G tries to generate an image set $G(x)$ similar to the target painting style image set Y. By contrast, the style fusion discriminator D_Y aims to distinguish the fake image Y' from the real image Y. G aims to minimize this objective, whereas the opponent D tries to maximize this objective, which is denoted as Equation (2). A similar adversarial loss will be used for the mapping function F:Y→ X and its discriminator D_Y, which is denoted as Equation (3).

$$L_{GAN}(G, D_Y, X, Y) = E_{y \sim pdata(y)}[(D_Y(y) - 1)^2] + E_{x \sim padata(x)}[D_Y(G(x))^2], \quad (2)$$

$$L_{GAN}(G, D_X, Y, X) = E_{x \sim pdata(x)}[(D_X(x) - 1)^2] + E_{y \sim padata(y)}[D_X(F(y))^2]. \quad (3)$$

Identity loss is used to ensure the continuity of the style fusion into an image and is denoted as Equation (4). When X_i passes through one of the generators, identity loss can cause the generated image $G(x)$ to be as close to the original image as possible, which prevents generators G and F from changing the hue of the input image.

$$Lidentity(G, F) = E_{y \sim pdata}(y)[\| G_y - y \|_1] + E_{x \sim pdata}(x)[\| F_x - x \|_1]. \quad (4)$$

In order to solve the non-matching data training problem of the adversarial network, so as to achieve the style transfer between Thangka and Chinese painting image sets, the cycle-consistency loss is necessary, which prevents the generators G and F from contradicting each other while increasing the mapping Image realism. That is, the adversarial loss causes the generated Chinese painting image set G_x to conform to the distribution of the input Thangka image domain y and thereby preserves multiple mapping relationships, but it does not cause the generated Chinese painting image domain G_x to retain its content from the real Chinese painting dataset X training process. $Epoch\_real\_x \rightarrow G_{(x)} \rightarrow F(G_{(x)})$ for forward cycle consistency. Similarly, for the generation process of the input image domain y,

G and F should also satisfy the backward cycle consistency, i.e., $y \to F_{(y)} \to G_{(F_{(y)})} \approx y$. For each image element X_i of the X, the cycle process of image fusion can bring back the stylistic features of X_i to the input X. Thus, the cycle-consistency loss should be expressed as

$$L_{cyc}(G, F) = E_{x \sim pdata(x)}[\| F(G(x)) - x \|_1] + E_{y \sim pdata}(y)[\| G(F(y)) - y \|]. \quad (5)$$

The full jobs objective is as follows. The λ controls the relative importance of the two objectives.

$$L(G, F, D_X, D_Y) = L_{GAN}(G, D_Y, X, Y) + L_{GAN}(F, D_X, Y, X) + \lambda L_{cyc}(G, F), \quad (6)$$

and the TPSF model aims to solve

$$G^*, F^* = \arg\min_{G,F} \max_{D_x, D_Y} L(G, F, D_X, D_Y). \quad (7)$$

In terms of training details, the least squares loss [56] is used to make the model more robust. For example, in Equation (2), the goal of the discriminator D is minimized to $E_{y \sim pdata(y)}[(D_Y(y) - 1)^2] + E_{x \sim padata(x)}[D_Y(G(x))^2]$, which means that $D_Y(y)$ is as close as possible to 1 and $D_Y(G(x))$ is as close as possible to 0. Among them, the goal of the generator G is to try to generate an image $G(X)$ similar to the image domain Y image, which means that the goal of G is to minimize the opponent D, which tries to maximize the output; that is, the generator is expected to minimize $D_Y(G(x)$, so $D_Y(G(x))$ needs to be as close to 0 as possible. The goal of the discriminator D_Y is to determine the difference between the generated image $G(x)$ and the real input sample Y, which means that the goal of D_Y is to maximize a $G(x)$ that tries to minimize the difference, that is, the discriminator is expected to maximize $D_Y(y)$; therefore, $D_Y(y)$ needs to be as close to *one* as possible.

4. Experiment and Results

4.1. Setup of Experiment

The computer configuration required for the TPSF model includes an AMD Ryzen 7 5800X processor, Window10×64 operating system, and TiTan XP×2 graphics card. The TPSF model code was mainly written in Python, and the framework was implemented with Pytorch. The small-squares loss was applied instead of the original maximum likelihood function, the weight of *lambda* was set to 10.0, the batch size was set to 1, and the learning rate of the Adam optimization parameter was set to 0.001 for the actual optimization after repeated iterations.

4.2. Results and Analysis

We collected a Tibetan and Chinese painting style dataset and established a Tibetan painting style fusion model. The entire dataset contained 400 works in total, which consisted of 200 Tibetan Thangkas and 200 Han Chinese paintings. Additionally, we applied 10-fold cross-validation. Thus, the dataset was divided into ten parts; nine of them were used as training data and one was used as test data.

Figure 4 shows the TPSF model loss of the training process, which is composed of GAN loss, cycle-consistency loss, and identity loss. According to the loss function graph, the cycle-consistency and identity loss in the graph greatly fluctuated, but the two showed an overall downward trend, which showed that the training results of the fusion of the Tibetan painting styles reached the expected goal. Although the GAN loss of the generator was on the rise during the operation of the generator, it indicated that the closer the reconstructed image was to the original image, the more realistic the generated image was.

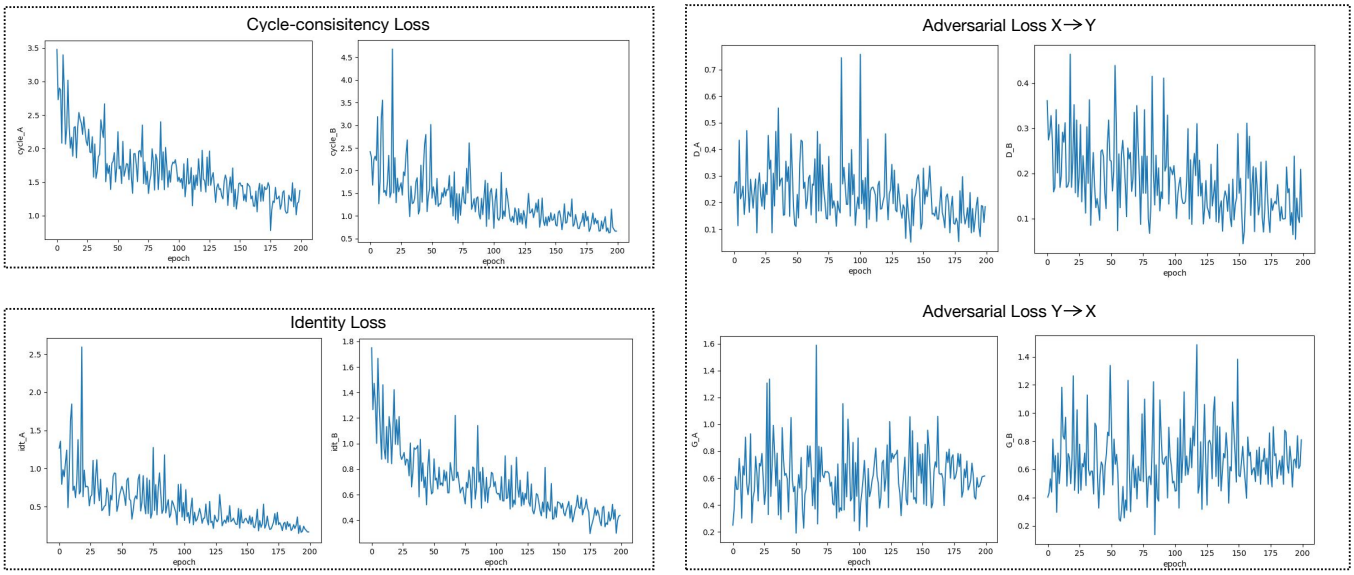

Figure 4. Three kinds loss of TPSF model, including two adversarial losses, a cycle-consistency loss, and a identity loss.

To objectively evaluate the experimental results, four sets of comparative models and the Frechet Inception Distance (FID) metric were used in the comparative experiments, and the experimental results were visualized.

When evaluating the veracity and variety of generated images, FID is a reliable and thorough evaluation metric that is more similar to human vision. The closer the data distribution is to the actual data distribution, the more accurate the picture generation will be. And the smaller the FID score, the closer the created data distribution is to the actual data distribution. Thus, calculating the FID score of the target image and the fused image allows one to assess the TPSF model's quality.

The average FID of the 10 training results of the TPSF model is shown in Figure 5. The model has a recursive network structure. The $G(X)$ score of the input Thangka image set was 236.82, and the $G(Y)$ score of the input Chinese painting image set was 166.06. The four comparison models added were all one-way networks, so under the FID score, the score of the Thangka image set $G(X)$ input of StarGAN V2 was 337.56; the score of the Thangka image set $G(X)$ input of StyleGAN was 321.89; the score of the Thangka image set $G(X)$ input of GAN was 368.03; and the score of the Thangka image set $G(X)$ input of StarGAN was 371.45. By arranging the average FID score of each model in descending order, we found that the FID score of the TPSF model was the smallest, which proved that the fusion works produced by this model were of higher quality.

To demonstrate the objectivity of the experiment, 50 professional reviewers (professors and students from relevant disciplines) and 100 general reviewers were invited to evaluate 10 randomly selected fused images in four aspects: attractive color, attractive visual, help study and ease operate. Details are shown in Figure 6. The questionnaire showed that most of the reviewers found the images generated by the model visually appealing and the model could help them learn the Thangka. Of the 50 professional judges, 81% found the randomly selected fusion work very attractive and 74% found the TPSF model easy to operate and effective in helping them learn the Thangka. Of the 100 general jurors, 84% found the TPSF model effective in helping them learn the Thangka and 86% found it easy to operate.

Model	Average value of FID ↓	
	Thangka	Chinese paintin
CycleGan	236.82	166.06
StarGAN V2	321.89	\
StyleGAN	337.56	\
GAN	368.03	\
StarGAN	371.45	\

Figure 5. Average value of FID: sort in descending order according to the FID score of the model.

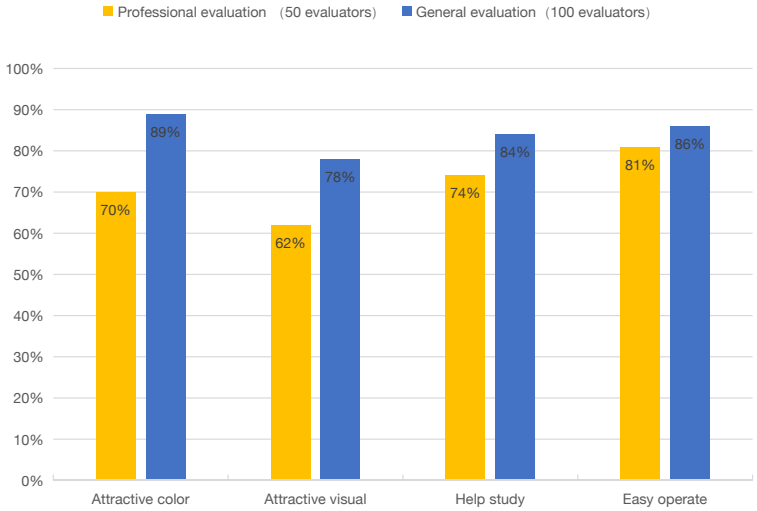

Figure 6. The questionnaire for fusion images. The numbers in the figure are the proportion of votes that were considered the best in each evaluation of the fusion work.

As shown in Figure 7, by comparing the real, fake, and idt images, we found that the fused works produced by the TPSF model had the style characteristics of both Tibetan and Chinese paintings. And they also had a strong visual appeal. In addition, some experimental results are shown in Figure 8.

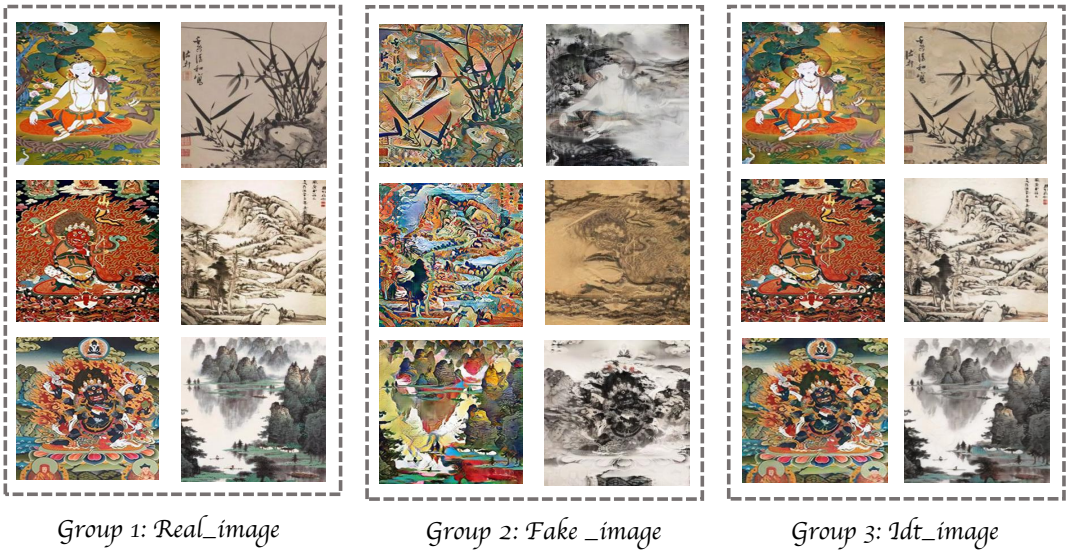

Figure 7. Some example results of TPSF model for qualitative evaluation. Group one is the selected three groups of real Tibetan painting experimental data and three groups of real Han painting test data. Group two is the fusion data of three groups of ethnic painting styles corresponding to group one. Group three is the equivalent data generated by the three corresponding to group one. The style images of group one are in the public domain.

Figure 8. Some Tibetan-Han style examples which were generated by the TPSF model.

5. Conclusions

We propose a Tibetan painting style fusion model based on CycleGAN. First, we collected a Tibetan Thangka and Chinese painting dataset to use as the experimental training and test data. We used the training dataset to train a generator and two discriminators. Meanwhile, we added cycle-consistency loss to the model so that the output works had Tibetan and Chinese painting style characteristics, rich picture content, and real picture effects. To more accurately verify the effectiveness of the integration of the Tibetan painting styles, four groups of contrast models were added to the experiment and were used for experimental evaluation. The TPSF model outperformed the other models and could quickly generate visually appealing fused image compositions. The advantages of the TPSF model to automatically and quickly integrate the painting styles of the two ethnicities can help more learners participate in the creation of the ethnic painting and stimulate more learners to have a strong interest in the study of ethnic painting.

Author Contributions: Methodology, H.W.; Software, L.W.; Formal analysis, H.W.; Investigation, Y.C. and L.W.; Data curation, Y.C.; Writing—original draft, Y.C.; Writing—review & editing, Y.C., L.W., X.L. and H.W.; Visualization, Y.C.; Supervision, X.L.; Funding acquisition, H.W. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the National Natural Science Foundation of China under grant number 62276216.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: https://github.com/90ii/TPSF-data.git (accessed on 28 January 2023).

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

References

1. Béguin, G.; Colinart, S. *Les Peintures du Bouddhisme Tibétain*; Réunion des Musées Nationaux: Paris, France, 1995; p. 258.
2. Jackson, D.; Jackson, J. *Tibetan Thangka Painting: Methods and Materials;* Serindia Publications: London, UK, 1984; p. 10.
3. Elgar, J. Tibetan thang kas: An overview. *Pap. Conserv.* **2006**, *30*, 99–114. [CrossRef]
4. Beer, R. *The Encyclopedia of Tibetan Symbols and Motifs*; Serindia Publications: London, UK, 2004; p. 373.
5. Cetinic, E.; She, J. Understanding and creating art with AI: Review and outlook. *ACM Trans. Multimed. Comput. Commun. Appl. (TOMM)* **2022**, *18*, 1–22. [CrossRef]
6. Hao, K. China has started a grand experiment in AI education. It could reshape how the world learns. *MIT Technol. Rev.* **2019**, *123*, 1.
7. Song, J.; Li, P.; Fang, Q.; Xia, H.; Guo, R. Data Augmentation by an Additional Self-Supervised CycleGAN-Based for Shadowed Pavement Detection. *Sustainability* **2022**, *14*, 14304. [CrossRef]

8. Ramesh, A.; Dhariwal, P.; Nichol, A.; Chu, C.; Chen, M. Hierarchical text-conditional image generation with clip latents. *arXiv* **2022**, arXiv:2204.06125.
9. Gregor, K.; Danihelka, I.; Graves, A.; Rezende, D.; Wierstra, D. Draw: A recurrent neural network for image generation. In Proceedings of the International Conference on Machine Learning, PMLR, Lille, France, 6–11 July 2015; pp. 1462–1471.
10. Hertzmann, A. Non-photorealistic rendering and the science of art. In Proceedings of the 8th International Symposium on Non-Photorealistic Animation and Rendering, Annecy, France, 7–10 June 2010; pp. 147–157.
11. Park, J.; Kim, D.H.; Kim, H.N.; Wang, C.J.; Kwak, M.K.; Hur, E.; Suh, K.Y.; An, S.S.; Levchenko, A. Directed migration of cancer cells guided by the graded texture of the underlying matrix. *Nat. Mater.* **2016**, *15*, 792–801. [CrossRef]
12. AlAmir, M.; AlGhamdi, M. The Role of generative adversarial network in medical image analysis: An in-depth survey. *ACM Comput. Surv.* **2022**, *55*, 1–36. [CrossRef]
13. Mo, Y.; Li, C.; Zheng, Y.; Wu, X. DCA-CycleGAN: Unsupervised single image dehazing using Dark Channel Attention optimized CycleGAN. *J. Vis. Commun. Image Represent.* **2022**, *82*, 103431. [CrossRef]
14. Choi, Y.; Choi, M.; Kim, M.; Ha, J.W.; Kim, S.; Choo, J. Stargan: Unified generative adversarial networks for multi-domain image-to-image translation. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–22 June 2018; pp. 8789–8797.
15. Liu, Y.; Sangineto, E.; Chen, Y.; Bao, L.; Zhang, H.; Sebe, N.; Lepri, B.; Wang, W.; De Nadai, M. Smoothing the disentangled latent style space for unsupervised image-to-image translation. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Nashville, TN, USA, 20–25 June 2021; pp. 10785–10794.
16. Karras, T.; Laine, S.; Aittala, M.; Hellsten, J.; Lehtinen, J.; Aila, T. Analyzing and improving the image quality of stylegan. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Seattle, WA, USA, 13–19 June 2020; pp. 8110–8119.
17. Chen, J.; Liu, G.; Chen, X. AnimeGAN: A novel lightweight GAN for photo animation. In Proceedings of the International Symposium on Intelligence Computation and Applications, Guangzhou, China, 16–17 November 2019; Springer: New York, NY, USA, 2019; pp. 242–256.
18. Mirza, M.; Osindero, S. Conditional generative adversarial nets. *arXiv* **2014**, arXiv:1411.1784.
19. Cao, K.; Liao, J.; Yuan, L. Carigans: Unpaired photo-to-caricature translation. *arXiv* **2018**, arXiv:1811.00222.
20. Zhao, Y.; Wu, R.; Dong, H. Unpaired image-to-image translation using adversarial consistency loss. In Proceedings of the European Conference on Computer Vision, Glasgow, UK, 23–28 August 2020; Springer: New York, NY, USA, 2020; pp. 800–815.
21. Timms, M.J. Letting artificial intelligence in education out of the box: Educational cobots and smart classrooms. *Int. J. Artif. Intell. Educ.* **2016**, *26*, 701–712. [CrossRef]
22. Cairns, L.; Malloch, M. Computers in education: The impact on schools and classrooms. *Life in Schools and Classrooms: Past, Present and Future*; Springer: Berlin, Germany, 2017; pp. 603–617.
23. Hwang, G.J.; Xie, H.; Wah, B.W.; Gašević, D. Vision, challenges, roles and research issues of Artificial Intelligence in Education. In *Computers and Education: Artificial Intelligence;* Elsevier: Amsterdam, The Netherlands, **2020**, *1*, 100001.
24. Al Darayseh, A. Acceptance of artificial intelligence in teaching science: Science teachers' perspective. *Comput. Educ. Artif. Intell.* **2023**, *4*, 100132. [CrossRef]
25. Chen, X.; Xie, H.; Li, Z.; Zhang, D.; Cheng, G.; Wang, F.L.; Dai, H.N.; Li, Q. Leveraging deep learning for automatic literature screening in intelligent bibliometrics. *Int. J. Mach. Learn. Cybern.* **2022**, *14*, 1483–1525. [CrossRef]
26. Chiu, M.C.; Hwang, G.J.; Hsia, L.H.; Shyu, F.M. Artificial intelligence-supported art education: A deep learning-based system for promoting university students' artwork appreciation and painting outcomes. *Interact. Learn. Environ.* **2022**, 1–19. [CrossRef]
27. Lin, H.C.; Hwang, G.J.; Chou, K.R.; Tsai, C.K. Fostering complex professional skills with interactive simulation technology: A virtual reality-based flipped learning approach. *Br. J. Educ. Technol.* **2023**, *54*, 622–641. [CrossRef]
28. Zhu, D.; Deng, S.; Wang, W.; Cheng, G.; Wei, M.; Wang, F.L.; Xie, H. HDRD-Net: High-resolution detail-recovering image deraining network. *Multimed. Tools Appl.* **2022**, *81*, 42889–42906. [CrossRef]
29. Ma, Y.; Liu, Y.; Xie, Q.; Xiong, S.; Bai, L.; Hu, A. A Tibetan Thangka data set and relative tasks. *Image Vis. Comput.* **2021**, *108*, 104125. [CrossRef]
30. Zhang, J.; Zhang, K.; Peng, R.; Yu, J. Parametric modeling and generation of mandala thangka patterns. *J. Comput. Lang.* **2020**, *58*, 100968. [CrossRef]
31. Qian, J.; Wang, W. Main feature extraction and expression for religious portrait Thangka image. In Proceedings of the 2008 the 9th International Conference for Young Computer Scientists, Hunan, China, 18–21 November 2008; pp. 803–807.
32. Liu, H.; Wang, W.; Xie, H. Thangka image inpainting using adjacent information of broken area. In Proceedings of the International MultiConference of Engineers and Computer Scientists, Hong Kong, China, 19–21 March 2008; Volume 1.
33. Hu, W.; Ye, Y.; Zeng, F.; Meng, J. A new method of Thangka image inpainting quality assessment. *J. Vis. Commun. Image Represent.* **2019**, *59*, 292–299. [CrossRef]
34. Gatys, L.A.; Ecker, A.S.; Bethge, M. Image style transfer using convolutional neural networks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27–30 June 2016; pp. 2414–2423.
35. Johnson, J.; Alahi, A.; Fei-Fei, L. Perceptual losses for real-time style transfer and super-resolution. In Proceedings of the European Conference on Computer Vision, Amsterdam, The Netherlands, 11–14 October 2016; Springer: New York, NY, USA, 2016; pp. 694–711.

36. Risser, E.; Wilmot, P.; Barnes, C. Stable and controllable neural texture synthesis and style transfer using histogram losses. *arXiv* **2017**, arXiv:1701.08893.
37. Li, S.; Xu, X.; Nie, L.; Chua, T.S. Laplacian-steered neural style transfer. In Proceedings of the 25th ACM International Conference on Multimedia, Mountain View, CA, USA, 23–27 October 2017; pp. 1716–1724.
38. Li, Y.; Wang, N.; Liu, J.; Hou, X. Demystifying neural style transfer. *arXiv* **2017**, arXiv:1701.01036.
39. Goodfellow, I.; Pouget-Abadie, J.; Mirza, M.; Xu, B.; Warde-Farley, D.; Ozair, S.; Courville, A.; Bengio, Y. Generative adversarial nets. *arXiv* **2014**, arXiv:1406.2661.
40. Ratliff, L.J.; Burden, S.A.; Sastry, S.S. Characterization and computation of local Nash equilibria in continuous games. In Proceedings of the 2013 51st Annual Allerton Conference on Communication, Control, and Computing (Allerton), Monticello, IL, USA, 2–4 October 2013; pp. 917–924.
41. Zhu, J.Y.; Park, T.; Isola, P.; Efros, A.A. Unpaired image-to-image translation using cycle-consistent adversarial networks. In Proceedings of the IEEE International Conference on Computer Vision, Venice, Italy, 22–29 October 2017; pp. 2223–2232.
42. Li, S.Z. Markov random field models in computer vision. In Proceedings of the European Conference on Computer Vision, Stockholm, Sweden, 2–6 May 1994; Springer: New York, NY, USA, 1994; pp. 361–370.
43. Castillo, L.; Seo, J.; Hangan, H.; Gunnar Johansson, T. Smooth and rough turbulent boundary layers at high Reynolds number. *Exp. Fluids* **2004**, *36*, 759–774. [CrossRef]
44. Champandard, A.J. Semantic style transfer and turning two-bit doodles into fine artworks. *arXiv* **2016**, arXiv:1603.01768.
45. Chen, Y.L.; Hsu, C.T. Towards Deep Style Transfer: A Content-Aware Perspective. In Proceedings of the BMVC, York, UK, 19–22 September 2016; pp. 8.1–8.11.
46. Lu, X.; Zheng, X.; Yuan, Y. Remote sensing scene classification by unsupervised representation learning. *IEEE Trans. Geosci. Remote Sens.* **2017**, *55*, 5148–5157. [CrossRef]
47. Mechrez, R.; Talmi, I.; Zelnik-Manor, L. The contextual loss for image transformation with non-aligned data. In Proceedings of the European Conference on Computer Vision (ECCV), Munich, Germany, 8–14 September 2018; pp. 768–783.
48. Liu, J.; Zha, Z.J.; Chen, D.; Hong, R.; Wang, M. Adaptive transfer network for cross-domain person re-identification. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Long Beach, CA, USA, 15–20 June 2019; pp. 7202–7211.
49. Chen, J.; Li, S.; Liu, D.; Lu, W. Indoor camera pose estimation via style-transfer 3D models. *Comput.-Aided Civ. Infrastruct. Eng.* **2022**, *37*, 335–353. [CrossRef]
50. Zach, C.; Klopschitz, M.; Pollefeys, M. Disambiguating visual relations using loop constraints. In Proceedings of the 2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, San Francisco, CA, USA, 13–18 June 2010; pp. 1426–1433.
51. Huang, Q.X.; Guibas, L. Consistent shape maps via semidefinite programming. In Proceedings of the Computer Graphics Forum, Guangzhou, China, 16–18 November 2013; Wiley Online Library: Hoboken, NJ, USA, 2013; Volume 32, pp. 177–186.
52. Wang, F.; Huang, Q.; Guibas, L.J. Image co-segmentation via consistent functional maps. In Proceedings of the IEEE International Conference on Computer Vision, Sydney, Australia, 1–8 December 2013; pp. 849–856.
53. Zhou, T.; Jae Lee, Y.; Yu, S.X.; Efros, A.A. Flowweb: Joint image set alignment by weaving consistent, pixel-wise correspondences. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Boston, MA, USA, 7–12 June 2015; pp. 1191–1200.
54. Godard, C.; Mac Aodha, O.; Brostow, G.J. Unsupervised monocular depth estimation with left-right consistency. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017; pp. 270–279.
55. Zhou, T.; Krahenbuhl, P.; Aubry, M.; Huang, Q.; Efros, A.A. Learning dense correspondence via 3d-guided cycle consistency. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27–30 June 2016; pp. 117–126.
56. Mao, X.; Li, Q.; Xie, H.; Lau, R.Y.; Wang, Z.; Paul Smolley, S. Least squares generative adversarial networks. In Proceedings of the IEEE International Conference on Computer Vision, Venice, Italy, 22–29 October 2017; pp. 2794–2802.