Article

Icon Generation Based on Generative Adversarial Networks

Hongyi Yang [1], Chengqi Xue [1], Xiaoying Yang [1] and Han Yang [2,*]

[1] School of Mechanical Engeering, Southeast University, Nanjing 211189, China; 220194514@seu.edu.cn (H.Y.); ipd_xcq@seu.edu.cn (C.X.); 220194543@seu.edu.cn (X.Y.)

[2] School of Aeronautic Science and Engineering, Beihang University, Beijing 100191, China

* Correspondence: yang_han@buaa.edu.cn

Abstract: Icon design is an important part of UI design, and a design task that designers often encounter. During the design process, it is important to highlight the function of icons themselves and avoid excessive similarity with similar icons, i.e., to have a certain degree of innovation and uniqueness. With the rapid development of deep learning technology, generative adversarial networks (GANs) can be used to assist designers in designing and updating icons. In this paper, we construct an icon dataset consisting of 8 icon categories, and introduce state-of-the-art research and training techniques including attention mechanism and spectral normalization based on the original StyleGAN. The results show that our model can effectively generate high-quality icons. In addition, based on the user study, we demonstrate that our generated icons can be useful to designers as design aids. Finally, we discuss the potential impacts and consider the prospects for future related research.

Keywords: user interface; icon design; deep learning; generating adversarial network

1. Introduction

Icons are an international language that can be observed ubiquitously in everyday life and are one of the most critical parts of a user interface (UI). It is a carrier between the user and the graphical interface, a mutual integration of function and aesthetics. A high-quality icon should not only enable users to quickly recognize and understand the meaning of the function it represents, but should also bring out the commercial character of the brand to some extent [1,2].

The design team often needs to design different types of icons according to different application scenarios and customer needs, while the design process often requires significant communication and coordination between designers and customers. The design team needs to make the icon language-independent and provide quick navigation for users in different countries in an interactive interface of limited size, while it has to fit the brand concept and reflect the unique ethos of the brand's product. The drafting phase, where an icon is conceived and created, is considered to be the most time-consuming part of the entire design process, and this phase often requires the design team to invest extra time and a wealth of inspiration. The designer must ensure that the icon created expresses both the functionality of the application and its own uniqueness. In particular, the large number of new applications created every day and the graphic structure of the icon that fits the psychological metaphor of most people brings about the phenomenon of similarity in design (Figure 1), thus requiring designers to constantly adjust design details and strive for innovation in the icon design process to avoid too much similarity with icons of similar products.

The use of artificial intelligence technology to assist in design and art creation has always been an interesting and meaningful research problem. With its superior computational and memory capabilities, artificial intelligence technology is able to provide the best design solutions through continuous accumulation of empirical knowledge and continuous optimization. Previous studies have proved that this technology has been successfully applied in the fields of apparel design [3], layout design [4], and circuit design [5].
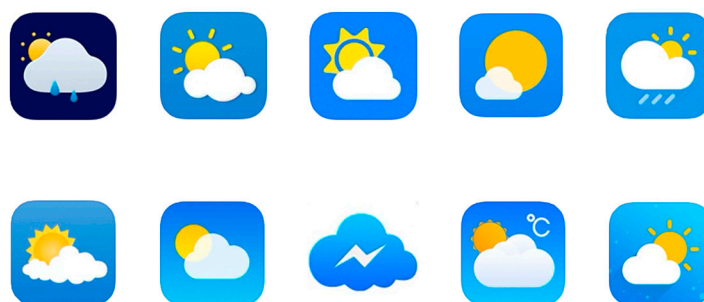
Figure 1. When searching for weather apps in the Apple Store, you can see a strong similarity in the icon design of these apps.

With the success of Generative Adversarial Networks (GANs) [6] in image generation related tasks [7–10], researchers have started to apply this technique to design, where GANs can learn from a large number of past design cases to find the regularities and eventually generate high-fidelity and diverse design solutions. For example, Liu et al. proposed a GAN-based model to assist designers in chair design, which consists of an image synthesis module and a super-resolution module to generate high-quality chair design solutions [11]; Li et al. proposed a LayoutGAN that enables automatic page layout, and for a random set of input parameters, the model can output an improved set of parameters in a wireframe layout [12]; and HouseGAN++ [13] proposed by Nauata.N et al. enables the task of automatic design of house floor plans.

GANs has also applied to icon design. The earliest related research could be traced back to Sage et al. in 2017, where they first constructed a large-scale logo dataset containing 60w corporate logos, and gave the icons different labels by clustering methods to avoid the mode collapse that often occurs in GANs, and finally achieved the logo generation task [14]. Mino et al. combined the ACGAN [15] with the WGAN-GP [16] to construct the LoGAN, which is based on the LLD-logo datasets [17] and eventually generated 12 types of logos with color as the label [18]. In 2019 this team constructed the LoGANv2, thus realizing the task of high-resolution icon generation, and the study was conducted on the StyleGAN [19] architecture. It experimented with conditional extensions and demonstrated that adding high-quality conditions to the unconditional model can effectively control the output of the synthetic network [20].

Since then, inspired by research combining logo design with GANs, researchers have started to focus their attention on the icon generation task. In graphic design, both logos and icons are visual symbols, but a logo is more to carry a company's values, and its purpose is to spread the value connotation of the brand; while an icon has stronger functionality and unity, it carries more single information, and its existence value makes it easier for users to identify and understand when using or browsing a product. Thus, unlike the logo generation task, the generated icon needs to have more functional characteristics and not just be an abstract visual symbol. Sun et al. constructed a bi-conditional GAN with color and shape for the first time in 2019, which could automatically color the icons desired by designers [21]. Takuro Karamatsu et al. used the CycleGAN [22] for the domain transformation task from natural images to icons [23].

However, there are two key challenges in applying GANs directly to the research of icon design: (1) Icon styles differ (linear, faceted, etc.), and the icons applicable to different scenarios have little in common (medical treatment, emotion, traffic, etc.). Therefore, for deep learning models such as GANs, it is challenging to capture the local correlation in the icon. (2) GANs require a large quantity of high-quality data; no such relevant icon datasets currently exist, and this could hinder research progress.

In view of the two aforementioned challenges, in this study, we constructed a large-scale icon dataset comprising 8 different categories. This study is the first icon generation study based on a large dataset (about 20,000 samples), but in addition, in contrast to previous studies, which used color and shape as the classification conditions, we classified the

icons according to application scenarios. Inspired by the study of Cedric Oeldorf et al. [20], in this study, we introduced conditional information on the StyleGAN to enable it to generate icons by category. Besides this, we added a self-attention mechanism and a spectral normalization operation to the original model to improve the quality and diversity of the model. We also adopted the idea of "relative accuracy" in the RGAN [24] and reconstructed the loss function of the model, which gives RGAN better training stability and generation effects with a limited number of samples.

The main contributions of this study are as follows.

1. It is an effective attempt to introduce deep learning techniques into the field of icon design, and for the first time the task of icon generation is implemented based on GAN, expanding the field of this research.
2. The first construction of an icon dataset consisting of 8 types of icons with different styles.
3. Applying conditional constraints in the StyleGAN and introducing new methods such as the self-attention mechanism to make it have better icon generation results.
4. Making some suggestions for future related research.

2. Materials and Methods

GANs have achieved good results in image generation since they were proposed, but due to the training challenges of the original GANs such as training instability, mode collapse, and difficulty in judging convergence [25], researchers have therefore successively proposed various improvements to GANs, such as WGAN [26], WGAN-GP [16] and other models for penalty functions, StackGAN [27], ProGAN [28], etc. in terms of structural improvements. In this section, we give a brief introduction to GAN, StyleGAN and some improvements introduced in the original model.

2.1. Generative Adversarial Network, GAN

GAN is a deep learning model which consists of two modules, the generator and the discriminator (Figure 2a), and the samples generated by the generator obey the real data distribution by adversarial training [6]. Among them, the discriminator aims to determine whether the input samples are real samples or fake samples generated by the generator; while the generator tries to generate samples that the discriminator cannot distinguish, and the two modules have opposite objectives and are trained alternately in an attempt to reach Nash equilibrium in this "MiniMax" game [29].
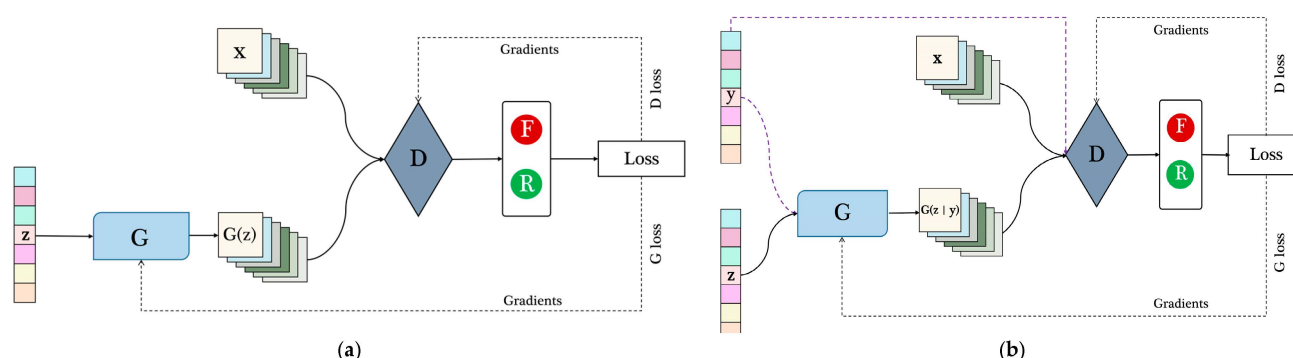

Figure 2. (**a**) The basic framework of GAN. Where G is the generator; D is the discriminator; z is generally random noise sampled from a Gaussian distribution, z generates a false sample G(z) after passing through the G; x is the true sample, and the discriminator makes a true or false judgment on the input sample. (**b**) The structure of CGAN. The biggest difference to traditional GAN is that the label of the sample will be used as input to both the generator and the discriminator.

The mutual game process between generator and discriminator can be expressed by a value function, and the generation problem of the model can also be transformed

into a minimax problem solving $V(G, D)$, and the objective function of the GAN can be expressed as:

$$\min_{G} \max_{D} V(D,G) = E_{x \sim p_{data}(x)}[log D(x)] + E_{z \sim p_z(z)}[log(1 - D(G(z)))], \quad (1)$$

where x is a real sample and z is random noise. The goal of the generator is to minimize the upper expression, and the goal of the discriminator is to maximize the upper expression.

2.2. Conditional Generative Adversarial Network, CGAN

The conditional GAN (CGAN) carries out conditional constraints on the basis of the original GAN and introduces the condition variable into the generator and discriminator to change the model from an unsupervised network to a supervised network [30], thus guiding the data generation process (Figure 2b).

Based on the original GAN, the objective function of the CGAN can be expressed in the following form:

$$\min_{G} \max_{D} V(D,G) = E_{x \sim p_{data}(x)}[log D(x|y)] + E_{z \sim p_z(z)}[log(1 - D(G(z|y)))] \quad (2)$$

The input of the generator in the model is noise, z, a conditional variable, y, and the conditional data that can guide the training process of the generator.

2.3. Style-Based Generator, StyleGAN

The traditional GAN samples noise z from a Gaussian distribution as input, which essentially learns the mapping function from the Gaussian distribution to the real image distribution through a neural network. StyleGAN [19], on the other hand, uses a style parameter to determine the output of the generator, which essentially learns the mapping from the style space to the real image distribution, and thus can artificially control the style of the generated image.

StyleGAN borrows the progressive structure of ProGAN [28], and the authors of the original paper designed a nonlinear Mapping Network consisting of 8 sets of fully connected networks to process the transformation $Z \rightarrow W$ ($w \in W$) in the potential space. A learnable affine transformation is then introduced to transform w into $y = (y_s, y_b)$, and y undergoes an adaptive instance normalization (AdaIN) after passing through each set of convolution layers in the Synthesis Network. Each set of convolution layers in the Synthesis Network is followed by an AdaIN operation [20].

$$AdaIN(x_i, y) = y_{s, i} \frac{x_i - \mu(x_i)}{\sigma(x_i)} + y_{b, i}, \quad (3)$$

where x_i is the feature map after performing normalization and then combining with y for scaling and biasing.

In addition, the model authors add a Gaussian noise layer after each convolution, thus providing the generator with a way to generate random details. The generator and discriminator structure follows the progressive structure proposed by ProGAN, which has been successfully able to significantly generate higher resolution images [19].

2.4. Self-Attention Mechanism

Since the receptive field size of GANs based on CNNs is limited, the convolutional kernel can only cover a small area around the central pixel, and thus there is a shift in the key locations in the generated image [31]. By contrast, the self-attention mechanism could compute the relationship between any 2 pixels in the image, and thus learn the global, long-range dependencies for generating images. Mirza, M. noted that the self-attention mechanism in the GAN could significantly improve the diversity of the generated images and, simultaneously, improve the quality of the generated images [32].

2.5. Spectral Normalization

Unlike previous models such as WGAN [26] or WGAN-GP [16] that make the discriminator satisfy Lipschitz continuity [33] by adding a constraint term to the objective function, spectral normalization directly constrains the Lipschitz constants of the discriminator by constraining the weight parameters of each layer of the GAN discriminative network, thus improving the stability of the GAN during training [34]. This operation is computationally simple and can be applied to GANs with different structures. Also, the original authors compared it with the existing normalization operation [35,36] and demonstrated that the spectral normalization operation has a more significant improvement on the diversity and quality of the generated images.

Specifically, when each parameter is updated during the training process, spectral normalization decomposes the weights of each layer of the model with singular values and normalizes their will to be limited to 1.

2.6. Relativistic Generative Adversarial Network, RGAN

The discriminator of a traditional GAN is designed to measure the probability that the generated data is true, while the discriminator of RGAN measures the probability that the generated data is "more real" than the real data. The researchers used relative distance measurements to measure this relative truth [23]. In other words, the traditional GAN attempts to increase the probability that the generated sample looks true, while the RGAN simultaneously attempts to reduce the probability that the real data looks true, which leads to faster convergence. Meanwhile, compared with GAN, Spectral GAN [24], and WGAN-GP [25], RGAN can generate higher quality samples, even when sample size is limited (approximately 2000 samples).

Specifically, the essence of the GANs is to make the generated distribution approximate to the unknown true distribution. The loss function in the GAN measures the distance between the two distributions and attempts to minimize it as much as possible. RGAN's loss function applied to the original GAN is stated as:

$$
\begin{aligned}
L_G^{GAN} &= E_{x_r \sim P}[g_1(C(x_r))] + E_{z \sim P_z}\left[g_2\left(C\left(x_f\right)\right)\right]; \\
L_D^{GAN} &= E_{x_r \sim P}[f_1(C(x_r))] + E_{z \sim P_z}\left[f_2\left(C\left(x_f\right)\right)\right]
\end{aligned}
\tag{4}
$$

RGAN modifies the loss function of traditional GAN by replacing $C(x_r)$ and $C(x_g)$ with $C(x_r) - C(x_g)$ to measure the relative authenticity of the measured input samples of the discriminant. The authors of RGAN have also successfully demonstrated that the relative discriminator is highly versatile, can be trained in combination with any type of GAN, and could obtain better performance together with techniques such as spectral normalization and gradient penalty.

$$
\begin{aligned}
L_G^{RGAN} &= E_{(x_r,\,x_f) \sim (\,P,\,Q)}\left[g_1\left(C(x_r) - C\left(x_f\right)\right)\right] + E_{(x_r,\,x_f) \sim (P,\,Q)}\left[g_2\left(C(x_r) - C\left(x_f\right)\right)\right] \\
L_D^{RGAN} &= E_{(x_r,\,x_f) \sim (\,P,\,Q)}\left[f_1\left(C(x_r) - C\left(x_f\right)\right)\right] + E_{(x_r,\,x_f) \sim (P,\,Q)}\left[f_2\left(C(x_r) - C\left(x_f\right)\right)\right]
\end{aligned}
\tag{5}
$$

3. Experiments and Results

The study is based on StyleGAN to enable it to generate corresponding icons by class according to the category of icons, and we hope to demonstrate that the generated icons can assist the design to some extent from a subjective and objective point of view. This section consists of four main parts: collection and processing of the dataset, model improvement and training details, generation results, and evaluation of subjective subjects.

3.1. Data

In our study, we used icons from www.iconfont.cn (accessed on 19 April 2021) [37] and www.icons8.com (accessed on 19 April 2021) [38], which contain a large number of

unwatermarked high-definition icons. We selected the top 30,000 flat style icons according to the number of downloads and collections. Flat style icons were selected as they are more available than other styles, such as linear style or 2.5D style icons.

In addition, the number of samples was confirmed and the icons were divided into 8 major categories and application scenarios, such as weather, emotion, and human, and the remainder of the icons were deleted. In the 8 categories selected, we also attempted to ensure the diversity of samples in each category. For example, there are trees, leaves, and flowers in the plant category icon, and planes, cars, and motorcycles in the transportation category icon. Simultaneously, we deleted some samples that were notably repetitive in style, excessively low in resolution, or rarely appeared in the same category (such as ties and socks in the clothing category), or where the background of the icons was uniformly processed into white, and set the resolution to 256×256 in PNG files (Figure 3). The final sample set consisted of 21,000 icon images, with a sample size exceeding 2000 in each category.

Figure 3. T-SNE [39] was used to visualize part of the dataset (Perplexity = 30, learning rate = 150, iterations = 1500). We selected the 8 categories of icons with the largest sample size from the collected samples: emotion, weather, human, plant, avatar, clothes, house, transport. It should be noted that human icons are mainly images of people with different movements and postures, while avatar icons are mostly busts of people with different professions.

3.2. Model Structure

The original StyleGAN is an unsupervised model. In this study, we change the label of the samples into the mapping network, style generator and discriminator, thus changing it into a supervised model conditional on the icon category. Meanwhile, both the generator and discriminator of the original StyleGAN are 9-layer progressive structure, and the generator starts from a feature map of size 4×4 to gradually generate images of size 1024×1024. Considering the size of our samples, the progressive structure is changed to 7 layers in this study, and the final image generation size is 256×256 (Figure 4).
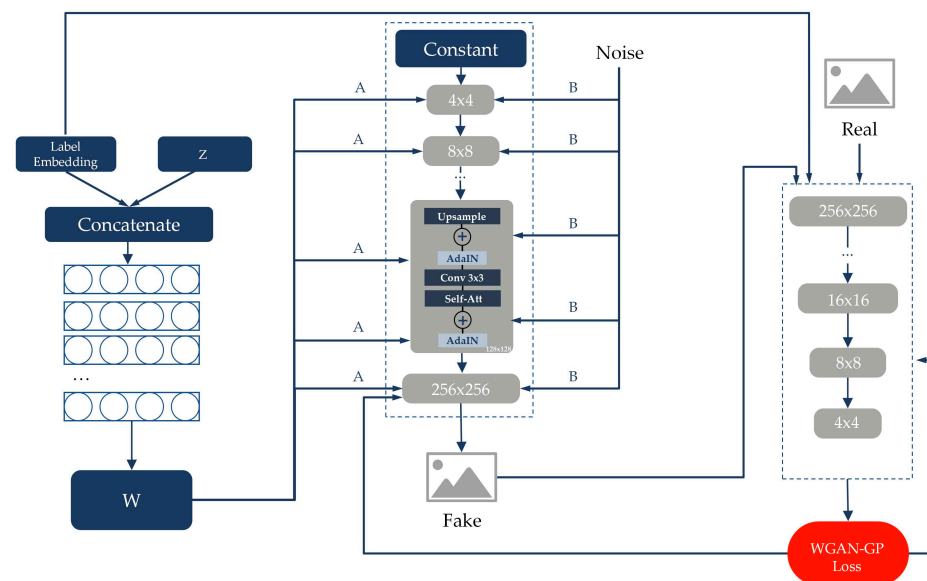
Figure 4. The improved model based on StyleGAN in the study. We input the label embedding of the samples into the mapping network with discriminator; we added the self-attention mechanism in the generator and discriminator 64×64, 128×128 and 256×256 feature maps.

Introducing the self-attention mechanism in GANs and imposing it on large feature maps could more efficiently improve the multi-class learning capability of GANs and generate more diverse images [31]. Thus, we add the self-attention mechanism to the 64×64, 128×128 and 256×256 sets of feature maps in the generator and discriminator. In addition, the spectral normalization not only improves the quality and diversity of the generated images, but also could be applied to any GANs [32,34]; thus, we applied spectral normalization in each convolution layer and linear layer of the generator and discriminator.

3.3. Train Detail

In the experiment, two NVIDIA Tesla P100-16GB units were selected as the GPU. The two-timescale update rule [40] was adopted to set the learning rate, the generator learning rate was set at 0.0002, and the discriminator learning rate was set at 0.0004.

Other hyperparameters in the model training include: Both the generator and discriminator optimizer used the Adam optimizer [41], the parameters of the optimizer used are 0.5 and 0.999; the hidden vector length is set to 256, the batch size is set to 8, and 50 epochs are trained.

Wang, Z. et al. pointed out that the form of RGAN can be introduced into any GAN model (IPM-GANs) whose objective function employs integrated probabilities, so in our study we chose the WGAN-GP loss function belonging to the form of IPM-GANs [32]. WGAN-GP uses the Wasserstein distance to calculate the distribution of the generated samples and the real samples of the loss function and applies a gradient penalty to the loss function [16].

In addition, we use one-sided label smoothing (Change the label 1 of the real sample to 0.9, thus guiding the discriminator to make some smoother predictions [25]) and feature matching (the comparison of the output value of the last layer in the original objective function is replaced by the comparison of the output of the penultimate layer [25]) to improve the model's performance.

3.4. Generate Results

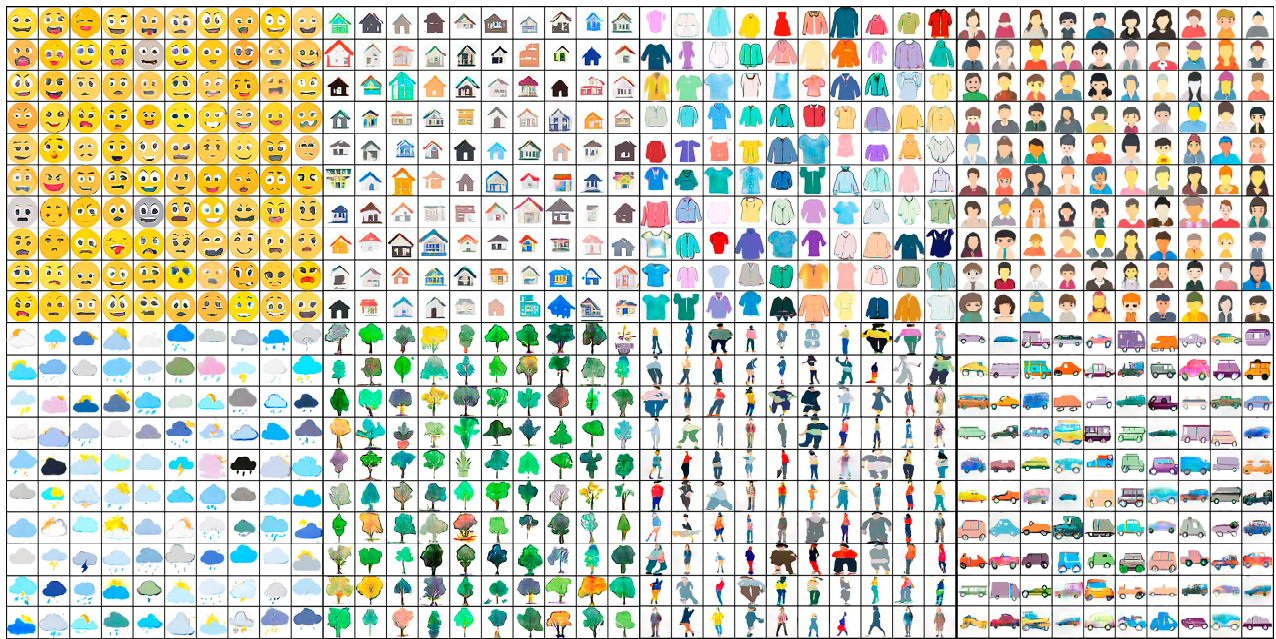The results generated in the study are shown in Figure 5.

Figure 5. The generation effect of the model at the 50th epoch, and the StyleGAN with the conditional constraints is able to perform the icon generation task by category.

We conducted a comparison experiment using the StyleGAN as the baseline, under the same experimental environment and training parameters, and the generation results are shown in Table 1, where the data are randomly collected from the icons generated at the 50th epoch. Intuitively, the icons generated by the improved model have clearer contours and less noise, and we will further compare the results of the two models in Sections 3.5 and 3.7.

Table 1. Comparison between StyleGAN and ours model generated results.

	StyleGAN	Ours
Weather		
Emotion		
Cloths		
Plants		
House		
Avatar		
Human		
Transport		

3.5. Quantitative Evaluation

We used the Inception Score (IS) [42] and the Frechet Inception Distance (FID) [40] to compare multiple sets of experiments, which are effective metrics for measuring the quality and diversity of the generated images. Where IS is a method to quantitatively evaluate

the results of the generative model by calculating the KL divergence with the help of the Inception-v3 network [43]:

$$IS = exp(E_{x \sim p_g(x)} KL(p(y|x)||p(y))),$$ (6)

where $p(y|x)$ is the probability of a category for a generated image x, after feeding it to the pre-training network Inception-v3 classification network, and $p(y)$ represents the expectation of the probability of the category output by this pre-training network for all generated images. For high-quality generated images, the classification network will determine the image as a certain category with a high confidence level, when $p(y|x)$ has a small entropy value. In addition, when $p(y)$ has a large entropy value, this indicates that the generated images are diverse. Therefore, when $p(y|x)$ has a lower entropy value and $p(y)$ has a higher entropy value, the KL divergence between the two is greater at that time, and so the IS value is higher, which means that the generated samples have higher quality and diversity.

FID, by contrast, compares the real samples with the generated samples, again by feeding the generated samples into the Inception-v3 classification network and calculating them by computing the distance between the real and generated sample feature maps. A lower value indicates that the generated sample is closer to the statistics of the real sample, i.e., the generated image has higher quality and diversity.

$$FID = \mu_{data} - \mu_g{}^2 + Tr\left(\sum\nolimits_{data} + \sum\nolimits_g - 2\left(\sum\nolimits_{data} \sum\nolimits_g\right)^{1/2}\right),$$ (7)

where $(\mu_{data}, \sum_{data})$ and $\left(\mu_g, \sum_g\right)$ denote the mean and covariance matrices of the real and generated samples, respectively, and Tr denotes the trace of the matrix.

We conducted several sets of comparison experiments, and the quantitative comparison of IS, and FID showed that our model has a more significant improvement relative to the original StyleGAN. Among them, the self-attentive mechanism has the greatest effect on the model to improve IS, and the relative discriminator has the greatest improvement on FID (Table 2).

Table 2. Quantitative evaluation according to IS and FID.

Model	IS	FID
StyleGAN	5.29	157.67
StyleGAN + AM	**12.23**	134.35
StyleGAN + SN	7.87	144.59
StyleGAN + RD	8.80	**112.83**
Ours	16.54	101.52

3.6. Visualization of the Generation Process

We visualized the generation process (Figure 6). In general, the model learns both the outline and the color combination of icons. After 1 epoch of training, the model generates abstract images with various color combinations; after 5 epochs, the model can gradually learn some general outlines of the icons and the composition of the body color; after 10 epochs, the model can already generate clearer icons of people's postures, but other categories of icons have not yet been able to generate obvious shape features; after 15 epochs, more categories of icons are gradually generated and the outline is clear, such as weather, clothes, avatar, etc., while we find that the model has not been able to generate house and plant icons with clear outline; after 25 epochs, some details of icons start to be reflected; and after 40 epochs, the model has been able to generate icons with clear outlines and less noise, and the generated icons have stronger diversity.
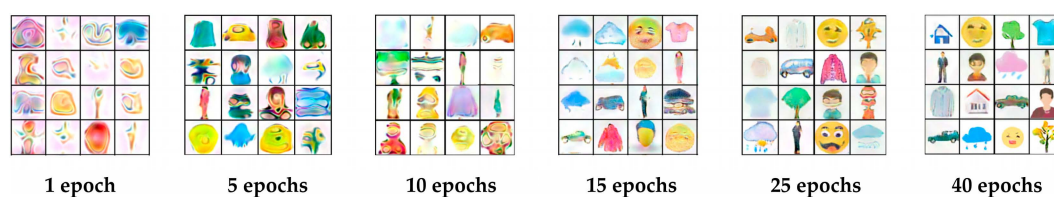
Figure 6. Visualization of the generation process.

3.7. User Study

We conducted a subjective evaluation experiment in the form of a visual Turing test with a questionnaire. A total of 105 subjects aged between 22 and 35 years old participated in our study, consisting of 60 males and 45 females. Among them were 25 graduate students in design, 15 graduate students in artificial intelligence, 35 UI designers or product designers from the Internet and finance industries, and 30 practitioners from other industries (specifically those not related to the design industry) (Figure 7). Twenty of the other industry practitioners were found through the Credamo online platform [44], which is a professional online research platform and allows for restrictions on participants' age, occupation, and other information. Other than that, all other subjects conducted their research offline.
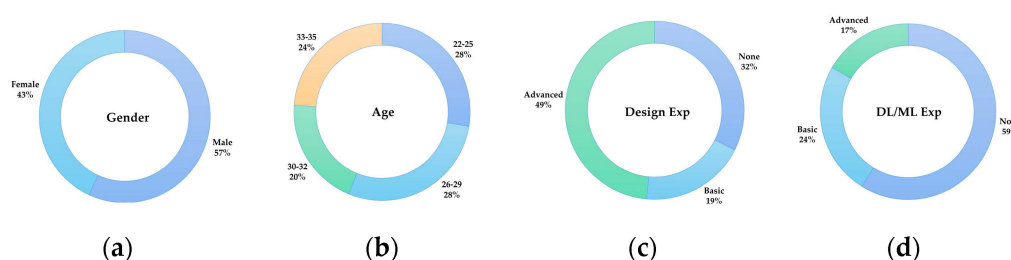

Figure 7. Profile distribution of the participants of our user study; (**a**) Age, (**b**) Gender, (**c**) UI design experience, (**d**) Deep Learning/Machine Learning experience.

3.7.1. Visual Turing Test

We randomly selected 15 samples from each of the 8 categories of icons for the visual Turing test, and subjects were asked to determine whether the given samples were from real icons or generated icons. Each class of test icons was equally divided into those from real icons and those generated by StyleGAN and ours, respectively.

We show 5 random icons in the test page (Figure 8), which may come from real icons, StyleGAN-generated icons, or icons generated by our improved model, and we want to get the most direct judgment from the subjects, so we ask them to complete each set of judgment tasks within 8 s, and each subject needs to complete 24 sets of judgment experiments. The results are shown in Table 3.
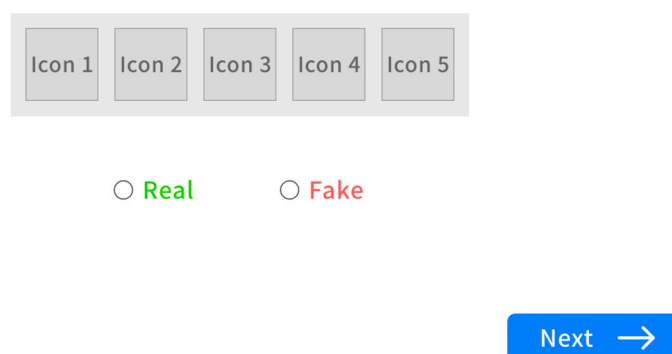

Figure 8. A sample of survey.

Table 3. The visual Turing test accuracy of 8 types of icons.

	Weather	Emotion	Clothes	Plants	House	Avatar	Human	Transport	Total
StyleGAN	86.3%	83.5%	97.4%	88.9%	83.4%	99.2%	94.5%	100%	91.6%
Ours	68.4%	77.0%	87.5%	76.2%	71.7%	75.6%	89.1%	97.1%	80.3%

By comparing the judgment accuracy in the visual Turing test, the icons generated by the improved model are harder to recognize as generated (lower judgment accuracy), although it should also be noted that the improvement in the correct judgment rate is also very limited (the overall value decreases by 5.8%), indicating that there is still a large visual moment of difference between most of the generated icons of the improved model and the real icons. The visual effect of Avatar icons generated by the improved model is the most improved, from 99.2% correct judgment rate to 75.6%. The weather, house and expression icons have lower judgment accuracy in both models, i.e., the most "realistic" effect, while the traffic and character gesture icons have the highest judgment accuracy, which also indicates that the visual effect of these two categories is poor. Another interesting phenomenon in the study is that subjects with more design experience did not have higher accuracy rates (73.6% for those with no design experience; 75.4% for those with design experience), contrary to our assumptions.

3.7.2. Questionnaire Survey

We selected 24 design graduate students and 20 UI designers to conduct a questionnaire study on the quality of the generated icons. The survey was conducted offline, and all participants had experience in designing visual symbols such as icons or logos. The UI designers were mainly from the UX Lab of Samsung Electronics China Research & Development Center in Nanjing.

We combined the icon design requirements and suggestions for developers from Apple's official website [2], and selected 5 subjective evaluation criteria for icon quality: legibility, simplicity, uniqueness, artistry and attraction. We asked the subjects to evaluate the generated 8 categories of icons (10 generated icons of each category were randomly selected) based on the aforementioned five criteria (Table 4). In addition, we also set a question at the end of the questionnaire: 'How helpful do you think the generated icons in this study are to your daily icon design work?'. All the aforementioned questions were asked using 5-point Richter scale points. The 5-point scale is expressed as '1 = very low, 2 = little low, 3 = moderate, 4 = little high, 5 = very high.'

Table 4. Quantitative metrics of our model from the user perceptual study.

	Legibility		Simplicity		Uniqueness		Artistry		Attraction	
	Mean	Std.	Mean	Std.	Mean	Std.	Mean	Std.	Mean	Std.
Weather	4.31	1.06	3.93	0.86	3.29	1.51	3.04	1.29	3.23	1.39
Emotion	3.41.	1.09	3.56	1.15	2.75	0.95	2.40	1.02	2.60	1.23
Cloths	3.52	0.81	2.36	1.04	2.95	0.85	2.20	1.11	2.27	1.14
Plants	3.54	1.05	3.11	0.75	2.50	1.19	2.90	1.53	2.65	1.36
House	3.80	1.09	3.18	0.83	2.27	1.23	2.54	1.19	2.38	1.15
Avatar	3.97	0.78	3.59	0.49	3.00	0.43	2.25	1.13	2.40	1.19
Human	3.75	1.22	3.02	0.83	3.09	0.70	1.95	0.82	2.90	1.32
Transport	2.90	0.79	2.43	1.07	2.23	1.22	1.72	0.78	1.97	0.72
Average	**3.78**	1.05	3.12	0.88	2.76	1.13	**2.38**	1.21	2.55	1.22

Generated icons scored higher on average on the recognizability and simplicity metrics than on others, with legibility being rated highest and artistry being lowest. As with the results of the visual Turing test, the traffic and human gesture icons scored relatively lower than the other categories of icons. We believe that the reason for this phenomenon is that there are fewer commonalities in these two icon categories-specifically, the traffic category

contains more categories, such as trucks, motorcycles, cars, etc., while the human gesture category is more challenging to generate due to the rich diversity of human actions.

We conducted the aforementioned research on the results generated by StyleGAN in the same environment, and the comparison between the results and our final model is shown in Figure 9a: we can see that our model has the most obvious improvement for uniqueness and legibility, while the improvement effect of artistry and attractiveness is weaker, so overall the improved model has better generation results. The results of the survey on the extent to which the generated icons help design work are shown in Figure 9b: more designers chose moderate (17) and little high (13), indicating that the generated icons in the study can assist designers' icon design work to a certain extent. Furthermore, surprisingly, through the communication with all the participating designers at the end of the questionnaire, it was found that the figure pose icon was given the most favorable comments, and they thought that although the visual effect of this type of generated icon was poor, as a more abstract icon, it could give designers more inspiration in creating human movements.
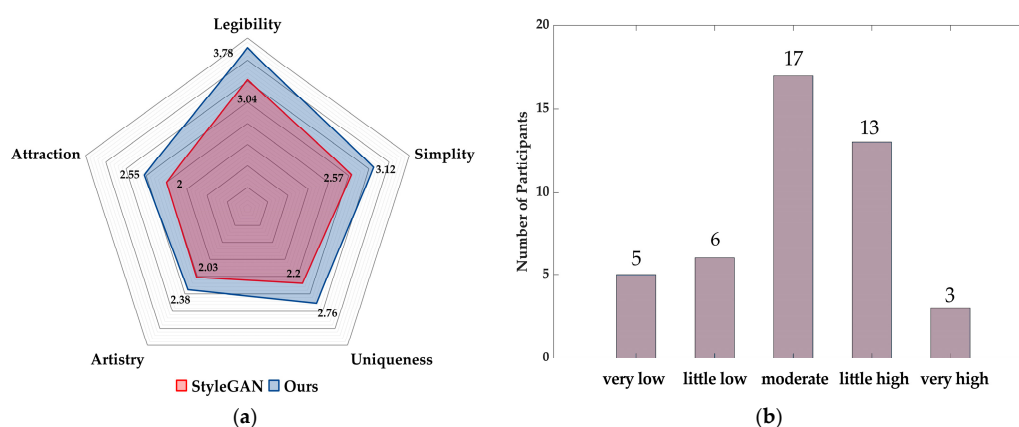

Figure 9. (**a**) Comparison of StyleGAN and ours evaluations; (**b**) Results of the survey on the usefulness of generating icons for design work.

4. Discussion

This study is an effective attempt at the intersection of icon design and deep learning. A large dataset consisting of 8 categories of icons is constructed, and StyleGAN is slightly improved and conditioned to generate icons by category. By comparing the IS and FID metrics with 105 visual Turing tests, the self-attention, spectral normalization, and relative discriminative distance operations added to the original model can improve the quality and diversity of the generated icons. Among them, the self-attention mechanism operation improves the IS most significantly, while the relative discriminative distance improves the FID most. We also visualize the icon generation process, and we can see that the model learns both color combinations and contour shapes. Notably, the model generates the first clear silhouette for the pose icons, which are considered to be of poor quality (about 5 epochs), while the house icons, which are of good quality, start to generate a clearer silhouette at the 28th epoch.

In addition, five evaluation indexes were defined according to Apple's icon design requirements and suggestions: legibility, simplicity, uniqueness, artistry and attraction, and 44 subjects with previous icon design experience were asked to evaluate the generated icons through a questionnaire. In the end, the three categories of expressions, weather and houses received higher scores, while the traffic and people's gestures categories received the lowest scores, which is consistent with the results of the visual Turing test. At the same time, the recognizability and simplicity of the generated icons were more recognized by the designers, and most of the UI designers who participated in the study thought that the generated icons could assist them in their daily icon design work. One interesting point in the study is that, compared with the weather and plant icons, which gave the subjects

higher visual perception and evaluation, the quality of the generated icons was poorer than that of the character pose icons, but the participating designers generally thought that they could get more inspiration from the generated icons for character movement, thus improving their design efficiency and quality. Therefore, more abstract visual symbol generation can be studied in future research.

Our study also has limitations: firstly, the data set we constructed in our study is only a small class of icons, and the quantity and quality of the training data are crucial for generation models like GAN. We believe that a major focus of future icon generation research is still on the construction of datasets, and it would be interesting to try to study other categories or styles of icons in addition to the 8 categories of face icons covered in our study. However, notably, in the process of icon collection, we found that the number of icons of other styles (linear, 2.5D icons) was significantly smaller compared to face icons, which would make it challenging for related generation studies. It is also shown in our study that when there is less commonality in the sample, it affects the quality of generated icons. Of course, with the rapid development of GANs on few-shot learning [44], it is believed that the generation of high-quality images on a small number of samples will gradually become mainstream in the future, which will undoubtedly give great help to icon generation studies. In addition, future research should be expanded on the applications related to the combination of icons and even interface design with GANs to make it more application-worthy, so that it could better assist designers in their work.

5. Conclusions

How to rely on artificial intelligence techniques to aid design is an interesting research direction, and with the success of generative adversarial networks in deep learning, new research ideas and methods have been provided for this crossover area. The factors limiting the intelligent generation of icons are the scarcity of data sets and the necessity of the generated icons to reflect the corresponding functions, both of which challenge the training and generation effectiveness of the models. The research in this paper is an effective attempt in this field and can be used as a basis for related research. We constructed a dataset of 20,000 icons and realized the generation of icons by class based on StyleGAN, and combined it with user research methods to confirm that the generated icons are functional and minimalistic, and can assist designers in their design work to a certain extent. However, it should be pointed out that the generated icons can meet the designers' initial draft requirements in terms of functionality and simplicity, but designers still need to add artificial design elements to make them more artistic and attractive.

With the rapid development of GANs, greater progress has been made in the fields of image super-resolution reconstruction, image style conversion, etc. Thus, how such models could be introduced into the field of icon creation and even graphic design is a problem worthy of further research. Finally, it would be an interesting research direction to understand the design process from this perspective by better visualizing the generation process of these GANs to understand this abstract learning process.

The importance of high-quality data for GANs-based design assistance is self-evident. However, when the data we collect has certain universality, the generated samples also have certain universality, that is, it is difficult to generate more unique samples according to different application scenarios. Therefore, how to make better use of the computing power, learning ability and memory ability of AI technology in the future, so that it can more reasonably model and analyze different design problems or application scenarios, and generate optimal design solutionswill be worth more in-depth research and discussion.

Author Contributions: Conceptualization, H.Y. (Hongyi Yang), X.Y., C.X. and H.Y. (Han Yang); methodology, H.Y. (Hongyi Yang); software, H.Y. (Hongyi Yang); validation, X.Y.; formal analysis, H.Y. (Hongyi Yang); investigation, H.Y. (Hongyi Yang) and X.Y.; resources, H.Y. (Hongyi Yang); data curation, H.Y. (Hongyi Yang) and X.Y.; writing—original draft preparation, H.Y. (Hongyi Yang); writing—review and editing, H.Y. (Han Yang); visualization, X.Y.; supervision, C.X., H.Y. (Han Yang);

References

1. Apple. Available online: https://developer.apple.com/design/human-interface-guidelines/ios/icons-and-images/app-icon/ (accessed on 10 March 2021).
2. Bonnie Kate Wolf. A Complete Guide to Iconography. Available online: https://www.designsystems.com/iconography-guide/ (accessed on 10 March 2021).
3. Dubey, A.; Bhardwaj, N.; Abhinav, K.; Mani, K.S.; Jain, S.; Arora, V. AI Assisted Apparel Design. *arXiv* **2020**, arXiv:2007.04950.
4. Yang, X.; Mei, T.; Xu, Y.Q.; Rui, Y.; Li, S. Automatic Generation of Visual-Textual Presentation Layout. *ACM Trans. Multimed. Comput. Commun. Appl.* **2016**, *12*, 1–22. [CrossRef]
5. Wang, H.R.; Wang, K.A.; Yang, J.C.; Shen, L.X.; Sun, N.; Lee, H.S.; Han, S. GCN-RL Circuit Designer: Transferable Transistor Sizing with Graph Neural Networks and Reinforcement Learning. In Proceedings of the 57th Design Automation Conference, San Francisco, CA, USA, 20–24 July 2020.
6. Goodfellow, I.J.; Pouget-Abadie, J.; Mirza, M.; Xu, B.; Warde-Farley, D.; Ozair, S.; Courville, A.; Bengio, Y. Generative adversarial nets. In Proceedings of the 27th International Conference on Neural Information Processing Systems, Cambridge, MA, USA, 8–13 December 2014; pp. 2672–2680.
7. Bodla, N.; Hua, G.; Chellappa, R. Semi-supervised fused GAN for conditional image generation. In Proceedings of the 15th European Conference on Computer Vision, Munich, Germany, 8–14 September 2018.
8. Ci, Y.; Ma, X.; Wang, Z.; Li, H.; Luo, Z. User-guided deep anime line art colorization with conditional adversarial networks. In Proceedings of the 9th ACM International Conference on Multimedia Systems, Seoul, Korea, 22–26 October 2018.
9. Yeh, R.A.; Chen, C.; Lim, T.Y. Semantic Image Inpainting with Deep Generative Models. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27–30 June 2016.
10. Iizuka, S.; Simo-Serra, E.; Ishikawa, H. Globally and Locally Consistent Image Completion. *ACM Trans. Graph.* **2017**, *36*, 1–14. [CrossRef]
11. Liu, Z.B.; Gao, F.; Wang, Y.Z. A Generative Adversarial Network for AI-Aided Chair Design. In Proceedings of the 2019 IEEE Conference on Multimedia Information Processing and Retrieval (MIPR), San Jose, CA, USA, 28–30 March 2019.
12. Li, J.; Yang, J.; Hertzmann, A.; Zhang, J.; Xu, T. LayoutGAN: Generating Graphic Layouts with Wireframe Discriminators. *arXiv* **2019**, arXiv:1901.06767.
13. Nauata, N.; Chang, K.H.; Cheng, C.Y.; Mori, G.; Furukawa, Y. House-GAN: Relational Generative Adversarial Networks for Graph-constrained House Layout Generation. In *European Conference on Computer Vision*; Springer: Cham, Switzerland, 2020.
14. Sage, A.; Agustsson, E.; Timofte, R.; van Gool, L. Logo synthesis and manipulation with clustered generative adversarial networks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–23 June 2018.
15. Odena, A.; Olah, C.; Shlens, J. Conditional image synthesis with auxiliary classifier GANs. In Proceedings of the International Conference on Machine Learning, New York, NY, USA, 19–24 June 2016.
16. Gulrajani, I.; Ahmed, F.; Arjovsky, M.; Dumoulin, V.; Courville, A. Improved training of Wasserstein GANs. In Proceedings of the 31st International Conference on Neural Information Processing Systems, Long Beach, CA, USA, 4–9 December 2017; pp. 5767–5777.
17. Sage, A.; Agustsson, E.; Timofte, R.; Van Gool, L. Lld-Large Logodataset—Version 0.1. 2017. Available online: https://data.vision.ee.ethz.ch/cvl/lld (accessed on 21 March 2021).
18. Mino, A.; Spanakis, G. LoGAN: Generating Logos with a Generative Adversarial Neural Network Conditioned on color. In Proceedings of the 17th IEEE International Conference on Machine Learning and Applications (ICMLA), Orlando, FL, USA, 17–20 December 2018.
19. Karras, T.; Laine, S.; Aila, T. A style-based generator architecture for generative adversarial networks. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–22 June 2018.
20. Oeldorf, C.; Spanakis, G. LoGANv2: Conditional Style-Based Logo Generation with Generative Adversarial Networks. In Proceedings of the 18th IEEE International Conference on Machine Learning and Applications (ICMLA), Boca Raton, FL, USA, 16–19 December 2019.

21. Sun, T.-H.; Lai, C.-H.; Wong, S.-K.; Wang, Y.-S. Adversarial colorization of icons based on structure and color conditions. In Proceedings of the 27th ACM International Conference on Multimedia, Orlando, FL, USA, 3–7 November 2019.

22. Zhu, J.Y.; Park, T.; Isola, P.; Efros, A.A. Unpaired image-to-image translation using cycle-consistent adversarial networks. In Proceedings of the International Conference on Computer Vision, Venice, Italy, 22–29 October 2017.

23. Karamatsu, T.; Benitez-Garcia, G.; Yanai, K.; Uchida, S. Iconify: Converting photographs into icons. In Proceedings of the 2020 Joint Workshop on Multimedia Artworks Analysis and Attractiveness Computing in Multimedia, Dublin, Ireland, 8 June 2020.

24. Martineau, A.J. The relativistic discriminator: A key element missing from standard GAN. *arXiv* **2018**, arXiv:1807.00734.

25. Salimans, T.; Goodfellow, I.; Zaremba, W.; Cheung, V.; Radford, A.; Chen, X. Improved techniques for training GANs. In Proceedings of the Advances in Neural Information Processing Systems, Barcelona, Spain, 5–10 December 2016; pp. 2234–2242.

26. Arjovsky, M.; Chintala, S.; Bottou, L. Wasserstein GAN. *arXiv* **2017**, arXiv:1701.07875.

27. Zhang, H.; Xu, T.; Li, H.; Zhang, S.; Wang, X.; Huang, X.; Metaxas, D. StackGAN: Text to photo-realistic image synthesis with stacked generative adversarial networks. In Proceedings of the IEEE International Conference on Computer Vision, Venice, Italy, 22–29 October 2017.

28. Karras, T.; Aila, T.; Laine, S.; Lehtinen, J. Progressive growing of gans for improved quality, stability, and variation. *arXiv* **2017**, arXiv:1710.10196.

29. Nash, J.F. Equilibrium points in n-person games. *Proc. Natl. Acad. Sci. USA* **1950**, *36*, 48–49. [CrossRef] [PubMed]

30. Mirza, M.; Osindero, S. Conditional generative adversarial nets. *arXiv* **2014**, arXiv:1411.1784.

31. Zhang, H.; Goodfellow, I.; Metaxas, D.; Odena, A. Self-attention generative adversarial networks. In Proceedings of the International Conference on Machine Learning, Long Beach, CA, USA, 9–15 June 2019.

32. Wang, Z.W.; She, Q.; Tomas, E.W. Generative Adversarial Networks in Computer Vision: A Survey and Taxonomy. *ACM Comput. Surv.* **2020**, *54*, 1–38.

33. Armijo, L. Minimization of functions having lipschitz continuous first partial derivatives. *Pac. J. Math.* **1966**, *16*, 1–3. [CrossRef]

34. Miyato, T.; Kataoka, T.; Koyama, M.; Yoshida, Y. Spectral normalization for generative adversarial networks. In Proceedings of the 6th International Conference on Learning Representations, Vancouver, BC, Canada, 30 April–3 May 2018.

35. Ioffe, S.; Szegedy, C. Batch normalization: Accelerating deep network training by reducing internal covariate shift. In Proceedings of the International Conference on Machine Learning, Lille, France, 6–11 July 2015.

36. Salimans, T.; Kingma, D.P. Weight normalization: A simple reparameterization to accelerate training of deep neural networks. In Proceedings of the Advances in Neural Information Processing Systems, Barcelona, Spain, 5–10 December 2016.

37. Iconfont. Available online: www.iconfont.cn (accessed on 19 April 2021).

38. Icons8. Available online: www.icons8.cn (accessed on 19 April 2021).

39. Van der Maaten, L.; Hinton, G. Visualizing Data using t-SNE. *J. Mach. Learn. Res.* **2008**, *9*, 2579–2605.

40. Heusel, M.; Ramsauer, H.; Unterthiner, T.; Nessler, B.; Hochreiter, S. GANs trained by a two time-scale update rule converge to a local Nash equilibrium. In Proceedings of the 31st International Conference on Neural Information Processing Systems, Long Beach, CA, USA, 4–9 December 2017.

41. Kingma, D.P.; Ba, J. Adam: A method for stochastic optimization. *arXiv* **2015**, arXiv:1412.6980.

42. Barratt, S.; Sharma, R.A. Note on the Inception Score. In Proceedings of the ICML 2018 Workshop on Theoretical Foundations and Applications of Deep Generative Models, Stockholm, Sweden, 10–15 July 2018.

43. Szegedy, C.; Vanhoucke, V.; Ioffe, S.; Shlens, J. Rethinking the Inception Architecture for Computer Vision. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27–30 June 2016; pp. 2818–2826.

44. Credamo. Available online: www.credamo.com (accessed on 1 July 2021).