Footwear Sketches Colorization Method based on Generative Adversarial Network

Xin Li[1]*, Yihang Zhang[2]
School of Art and Design, Guangdong University of Technology, Guangzhou, China[1]
College of Mathematics and Informatics, South China Agricultural University, Guangzhou, China[2]

***Abstract*—The coloring of sketches has a constant market demand in the area of research. The difficulty of the coloring sketch outline is its lack of texture and color. Take footwear design as an example, it is difficult for designers to complete a colorful sketch in a limited time, so an artificial intelligence technology for coloring shoes is required. Though we do not build a new GAN, which is based on pix2pix. We try to integrate the existing model in four ways, including generator, discriminator, loss function and comparison. In this paper, given a set of edges-to-shoes that have 50,025 shoe images, our approach produces an image with vivid shoes images. Unlike the recent research, our approach is not based on a unique adversarial training. We show that shoe sketches can be synthesized from simple lines by a GAN into a high-resolution picture. In particular, we offer a new model to synthesize high-resolution photo-realistic images of shoes, and apply a multi-discriminator to train and distinguish the generated images. Our model enables the shoe designer to benefit from the colorization design.**

Keywords—Footwear sketch; generative adversarial network; image to image translation; colorization

I. Introduction

In the footwear industry, color sketch is a time-consuming step in design period for designers. Creating a concept map of colorful shoes requires a professional color, composition and valid use of shade and texture. This process demands experienced drawing expertise and a good sense of design aesthetics. Even the experts would spend much time coloring the sketches.

In the design process, designer may decrease the needless hours if they overcome the colorization issues. As a result, a standalone coloring system can be a suitable solution for the footwear design industry. With the help of this system, the newcomer can be inspired, while professors save more time on color compositions of product sketches. This idea may totally change the fundamental structure of product development, which do optimize the industrial structure.

However, a lot of challenges remain to achieve this process. At first it is hard for the machine to understand the sketches of footwear which have innumerable drawing styles. As well, footwear sketches have a limited expression. What's more, there is no guidance for a machine to make colorizing decisions.

While deep learning in IT vision research is becoming a hot topic, alternative learning-based methods have been developed. For example, Mathias Eitz presented an interactive pattern recognition system that can identify a human sketch object [1]. Christopher Hesse has operated an online system that can generate cat images from the edges [2]. Although these systems can successfully turn a user sketch into a colorful object, such an application still cannot meet the designer's expectation. In addition, it is hard to shape high-resolution images and images with details and texture. Due to the limitation of an experimental topic, there are still has many opportunities to make progress.

In this article, we propose a method that allows users to enrich their design with hand-made shoes, color based on the Generative Adversarial Networks (GAN). GANs can classify the real or fake images, while forming a model that can minimize the loss. We form our network on an open access edge-to-shoes dataset using a new approach that creates high-resolution images. Different from the previous results with little detail and realistic textures, we explore a new, robust adversarial learning method with multi-scale generator and discriminator framework. This framework can produce a result with better visual quality. In this process, we receive the result with adversarial training rather than any loss by hand-made or pre-trained networks. This method shows that it is possible to improve the addition of perception losses from pre-trained networks. What is more, a multi-discriminator allows a better performance in training and alternative quantitative comparison such as PSNR, SSIM and MSE can have a thorough analysis of the coloring footwear results. Our contributions may be summarized as follows:

1) We introduce a new method called local amplifier to synthesize high-resolution photo-realistic images of shoes.

2) We consume the discriminator by extending the GAN into a multiple framework while improving the adversarial loss.

3) A quantitative comparison using the PSNR, SSIM and MSE is included.

II. Related Work

A. Generative Adversarial Network

In recent years, deep learning has unleashed another wave of artificial intelligence. In particular, in the areas of image recognition, the method based on deep learning [3, 4] has much improved over traditional methods. Their accuracy rate is near to or even greater than that of the manual identification.

The most typical task in unsupervised learning is an image generation [5, 6], and the first image generation model based

*Corresponding Author.

on deep learning is Autoencoder [7]. However, Autoencoder do not have a specific link to measure the error between the reconstructed sample and the real sample. The upgrade model VAE, simply makes the generated images more similar to the database images, instead of learning the generative paradigm for obtaining new images.

A new architecture called the generative adversarial network [8] attempts to resolve this problem. The standard GAN model is made up of two parts, one generator and one discriminator. It no longer updates the generator solely by measuring the similarity between the generated image and the actual image. However, it implements adversarial training through a discriminator, so that the generator can learn then latent picture mode. Meanwhile the distribution is nearer to the distribution of reality.

Today, GAN has become the main model for picture generation [9, 10, 11] and even unsupervised learning. It not only occupies the mainstream in academia, but also has made great achievements in fashion, advertising, audio and video industries, etc.

B. Image-to-Image Translation

The conditional generative adversarial network [12] highlights, automatic Image-to-Image translation, which teaches input mapping to output images has been applied to various tasks. For instance, generating photographs from sketches [13] or attributes, semantic layouts [14]. Concretely, the generative model of Image-to-Image translation has two parameters or variants that are not parametric and distributes with special algorithms. In an Image-to-Image translation task, a generative model can utilize the distribution of the target domain by generating perfect “fake” data which named translated images to derive from the target domain’s distribution [15]. Popular Image-to-Image translation methods include two-domain and multi-domain [16]. First, two-domain can solve problems like computer vision, image processing by using image style transfer in photo editors to benefit from autonomous driving and image colorization [17]. Secondly, multi-domains focus on creating multiple outputs made up different semantic contents or style textures.

In the future, Image-to-Image translation will increase in resolution and generate variable outputs. And the researchers will generalize the Image-to-Image translation methods within the image field to other aspects such as text, language, speech and also multimodal translation tasks.

C. Colorization

Colorization is a computer assisted method of adding color to an image or film [18]. In 1987, Markle invents a colorization process that paints at least one reference frame. It solved the image problem of visual fatigue, by segmenting images into regions and filling in colors.

In the past, coloring was a time-consuming task that required a professional ability to paint. In order to better reduce processing time of sketching as shown in Fig. 1, several interactive technologies have been offered. Levin uses strokes to both indicate the colors of positive pixels and colorized images with optimization where similar pixels have similar colors [19]. Colorization technologies embrace that image division can be well defined in a different space. Take interactive manga colorization technology as an example, several groups gather in a mensurable cluster before colorization. Luan presents an interactive system through color labelling and color mapping for greyscale images, in the absence of complex texture grouping techniques that further improve the system usability [20]. Cheng has a high-quality comprehensive colorization method, but it needs to train on a huge set of reference image data that contains all of the objects [21]. Gupta introduced a sample-based solution to colorize a grey image, adopting a quick cascade feature mapping scheme to voluntarily match matches between reference and target images. While the process still needs to prepare a color example, which is semantically similar to the grey image [22]. In a word, automatic segmentation often failed to identify the complex drawing color, especially the alternative colorized selection of different experimental subjects. They have multiple sketch lines and gray image feature. For instance, a difference between fashion and sporty shoes. To solve this problem, we are going to invent more and more techniques of colorization.

Nowadays, we are making advances in colorization. The machine can produce a colorful sketch with little mistakes which is impossible in the previous time presented in Fig. 2. Subversive colorization techniques such as CIC, LTBC, Pix2pix and DC used CNNs appear [23]. Generative coloring designs extend the unconditional image generation to a better vision. These optimization-based methods and learning-based methods have put forward to successfully colorize line sketches or grayscale photos. They can create dramatic color images without any hand-made help.

In the coming years, especially in the field of coloring sketches, increasingly efficient method will be taken. And the object will be a variable that can inspire further investigation in this direction in the generative modeling of the sketch.

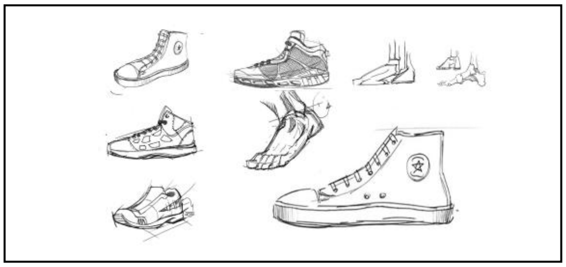

Fig. 1. The footwear handmade sketches.

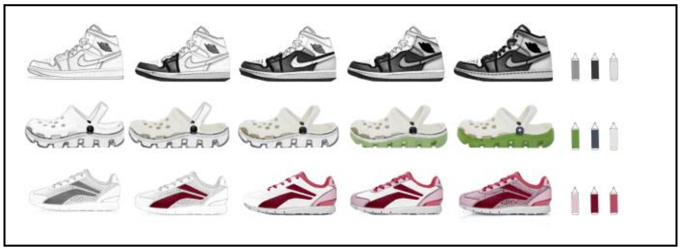

Fig. 2. Sketches manual colorization.

III. Method

To optimize outputs, we construct a conditional adversarial frame to generate a high-resolution shoe image from line sketches. Apart from generation, we use multi-discriminator variants which increase the difficulty of the generator screening sequence. Initially, a basic model of pix2pix is proven. Second, we explain how to accelerate the image reality and resolution of the picture with a GAN frame design. In addition, the adversarial loss is also considered, which can stabilize the training process.

A. The pix2pix Baseline

Pix2pix method is one of the extensions included in the conditional GAN framework for image-to-image translation [24]. The design framework composed of a generator G and a discriminator D. In our research, the aim of our generator G is to colorize the footwear line sketches to complete concept images, while the discriminator D calculated with the purpose of distinguish real images from the forger. The operation consists of a supervised setting. To sum up, the training dataset is provided as a pair of comparative images {(s_i, x_i)}, where s_i is a purely footwear shape line sketch and x_i is a corresponding real photo. The purpose of conditional GANs is modeling the assumed distribution of real shoe images by enriching the input line sketches via the underlined minimax operation:

$$\min_{G}\max_{D}\mathcal{L}_{\text{GAN}}(G,D) \tag{1}$$

And the objection function L_{GAN} (G, D) is given by:

$$\mathcal{L}_{\text{GAN}}(G,D)=\mathbb{E}_{(\mathbf{s},\mathbf{x})}[\log D(\mathbf{s},\mathbf{x})]+\mathbb{E}_{\mathbf{s}}[\log(1-D(\mathbf{s},G(\mathbf{s}))] \tag{2}$$

The pix2pix method applies U-Net as the generator [25], simultaneously applies fully convolutional patch-base network as discriminator [26]. We put a series of sketches of footwear line and comparative image to the discriminator as our input.

B. Coarse-to-fine Generator

As the resolution of the training images is unstable and the lesser quality, we further improve our pix2pix frame based on the baseline above, the figure presented in Fig. 3. In the beginning, we split the generator into two parts, named G_1 and G_2. Whereas G_1 is a global network of generators, G_2 has a local network of multipliers. Below is an overview of G_1 and G_2. The overall generator network's resolution is 256 x 256, and the local multiplier network is calculated with an output twice the size of the latter. To integrate a better product image resolution, we can join local excess multiplier networks. For example, when using G_1 and G_2 as the generator sections, the resolution of the final output image is G= {G_1, G_2} is 512 x 512. However, the resolution grew when G= {G_1, G_2, G_3} is 1024 x 1024.

As a vital example of global generator, Johnson et al. [27] constructed a neural network framework for 512 x 512 image style transfer. The framework is composed of three parts including $G_1^{(F)}$, $G_1^{(R)}$ and $G_1^{(B)}$. $G_1^{(F)}$ represented a convolutional front-end, $G_1^{(R)}$ represented as a set of residual blocks and $G_1^{(B)}$ represented a transposed convolutional back-end. The input of the generator is a 256 x 256 image, and output of our process operating three parts is an image of 256 x 256 in resolution.

Apart from global generator, the local enhancer network is composed of three sections as well. And the components, respectively named $G_2^{(F)}$, $G_2^{(R)}$, $G_2^{(B)}$. Firstly, $G_2^{(F)}$ is called a convolutional front-end. Secondly, $G_2^{(R)}$ is called a group of residual blocks, $G_2^{(B)}$ is called a transposed convolutional back-end. We use an input line sketch image to G_2 resolution of which is 512 x 512. Comparing to the global generator, the input of residual block $G_2^{(R)}$ is made up of two feature schedules, the output of $G_2^{(F)}$ and the $G_1^{(B)}$. This manipulation assists to synthesize the entire image information from G_1 to G_2.

In our training course, we operate the global generator before the local enhancer at a certain resolution. After adjusting all the element, the global and local material for shoe line sketches is assembled for image compound. It is obvious that this image generates framework with two-scale is efficient [28]. Other architectures with common usage can be found, such as unconditional GANs or image generator.

C. Multi-scale Discriminators

It's difficult for the GAN discriminator to generate high-resolution images. In order to seperate the high-resolution real and forger images, the discriminator should have a comprehensive production program [29]. Aim to accelerate the framework's ability and decrease the overfitting, our process commands an embedded network and larger convolutional kernels. Our memory space usage, which has a high-resolution image generate function should be large enough for training.

As a maintain problem, we address multi-scale discriminators to cope with it in Fig. 4. We apply multiple discriminators which have a definite network framework but operate multiple image scales. It's obvious that a multi-discriminator which approximate max D V (D, G) can be a good transfer to the generator.

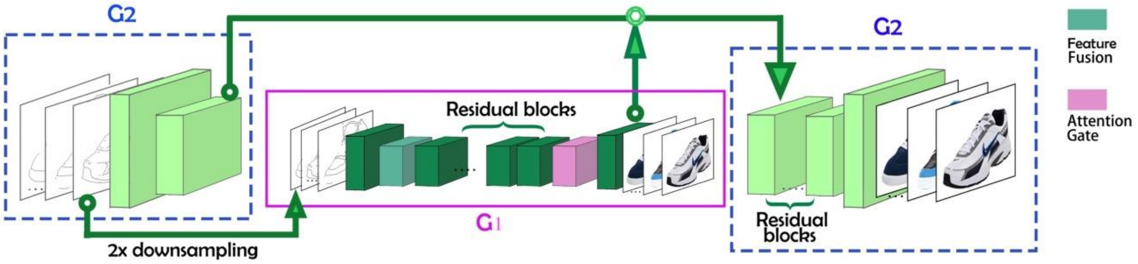

Fig. 3. Two sub-network generators, G_1 for low resolution images, G_2 for high resolution images appending to G_1 input and utilizing the last map from G_1.

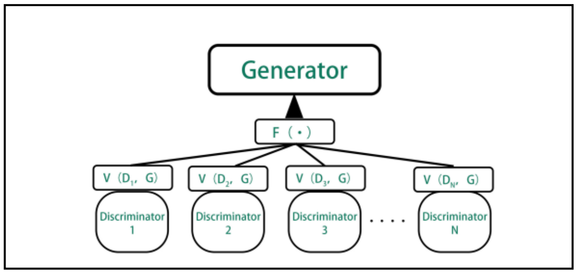

Fig. 4. Multiple discriminators.

To provide a better source for generator, it's necessary to maximize G(V$_i$). The discriminators are randomly presented, gathering all the optimized V such as the stochastic gradient ascent. $\max_{i\in}$ {1. N} V (D$_i$, G) is the loss of the generator which copes with the problem rendered by the non-convexity of complexes V. In order to minimize the max forces G, we must monitor all of the N discriminators. When we are operating, $\max_{D_i \in D}$ V (D$_i$, G) doesn't have countless elements, so the above framework is complicated. However, the quantity of the discriminator is still an effective factor.

For example, we construct three discriminators called D$_1$, D$_2$, D$_3$. In particular, we collect the actual images and compound at high-resolution images by several parts to create a pyramid of image of three scales. All of the three discriminators D$_1$, D$_2$, D$_3$ trained to seperate the real and forger at three different scales. To analyze one of the three discriminators, the discriminator that used the roughest scale with the entire field, generating corresponding overall images. Moreover, there is a discriminator with the role of enhancing the finest scale that can produce more specific. This operation allows the calculation to produce higher resolution images that can be an optimized method. By comparing with the previous equation, the multi-scale discriminators equation is:

$$\min_{G} \max_{D_1,D_2,D_3} \sum_{k=1,2,3} \mathcal{L}_{\text{GAN}}(G, D_k) \tag{3}$$

It has been shown that this is a useful method for using multiple GAN discriminators on the same scale. Many researchers have applied the multi-discriminator model to synthesize a better image output [30]. And our purpose is the production of high-resolution images.

D. Improved Adversarial Loss

Our former equation of the objection function LGAN (G, D) has already improved the GAN loss. The steady operation of this loss function helps to produce dramatic results at multiple scales. The process identifies various characteristics of the alternative layers of the discriminators in order to match uncertain representation images. To increase completeness, the following calculation formula is used:

$$\mathcal{L}_{FM}(G, D_k) = \mathbb{E}_{(s,x)} \sum_{i=1}^{T} \frac{1}{N_i} \left[\left\| D_k^{(i)}(s,x) - D_k^{(i)}(s, G(s)) \right\|_1 \right] \tag{4}$$

where T in the equation means the total quantity of layers and Ni represents the sum of elements in layers. We use the discriminator related to perceptual loss, and we discuss how to make more progress in operational performance. With the combination of GAN loss and characteristic mismatch loss, the equation is:

$$\min_{G} \left(\left(\max_{D_1,D_2,D_3} \sum_{k=1,2,3} \mathcal{L}_{\text{GAN}}(G, D_k) \right) + \lambda \sum_{k=1,2,3} \mathcal{L}_{FM}(G, D_k) \right) \tag{5}$$

In this equation, λ is the critical point of two terms, explaining that for loss LFM, Dk is a main extractor.

E. Quantitative Comparison Metrics

To better measure the quality of our experiment, we conduct some evaluation on the synthesized images and true label with these metrics below.

Firstly, SSIM is also a famous quality measurement for testing difference between two image samples [31]. Wang developed the SSIM for calculation related to the image quality of the human visual system [32]. The difference of SSIM is its applications of three factors apart from considering the addition of all image error. The luminance comparison function calculates the similarity of two samples' luminance ($\mu_f\ \mu_g$). The contrast comparison function calculates the closeness of two images' contrast ($\sigma_f \sigma_g$). And the structure comparison function measures the correlation coefficients of image *f* and image *g* noting that σ_{fg} is the covariance. The SSIM index positive values are in [0, 1].

The SSIM can be presented as:

$$SSIM(f,g) = l(f,g)c(f,g)s(f,g) \tag{6}$$

Moreover, the *l (f, g)*, *c (f, g)* and *s (f, g)* It can be described as follow:

$$l(f,g) = \frac{2m_f m_g + c_1}{m_f^2 + m_g^2 + c_1} \tag{7}$$

$$c(f,g) = \frac{2s_f s_g + c_2}{s_f^2 + s_g^2 + c_2} \tag{8}$$

$$s(f,g) = \frac{s_{fg} + c_3}{s_f s_g + c_3} \tag{9}$$

Secondly, MSE (Mean Square Error) is one of the most popular applications which utilized the integrated sample which measured by square intensity differences of various images pixels [33]. In addition, MSE considers the whole reference metric and the final score are better if it values closed to 0. The equation can be defined as:

$$MSE = \frac{1}{MN} \sum_{n=0}^{M} \sum_{m=1}^{N} \left[\hat{g}(n,m) - g(n,m) \right]^2 \tag{10}$$

Lastly, PSNR (Peak signal-to-noise ratio) is a common coherence quality measurement for system operation especially in the video or image sample. It is a simple metric that researchers utilize the method to evaluate and develop the visual sample quality. Huynh-Thu et al. show that PSNR could be a useful indicator when the content and codec have analogue. However, the situation changed when the different samples mix together [34]. It is said that the higher PSNR value gains a higher visual sample quality, low quality results from image otherness.

Sometimes noisy samples may affect our evaluation results, and studies have shown that PSNR mixed with MSE have the best performance of evaluating the quality of noisy samples. For example, when applying image degradation PSNR value presented between 30-50 dB for 8-bit data or 16-bit data. The equation can be defined as (11) below, peak value (which we denote by *peakval* in the equation below) means the maximum of the image sample data. For an unsigned 8-bit integer data type, the *peakval* is 255.

$$PSNR = 10\log_{10}(peakval^2) \,/\, MSE \tag{11}$$

IV. Experiment

A. Datasets

We perform our method on the Edges2shoes data set, which has been obtained from the official Pix2Pix Datasets directory from Pix2Pix Datasets [35]. The total data set is 50,025, one data pair contains an art line and the color image corresponding to the art line. The task is to map the line art to the color picture, such as the colorization task. To form a model with enhanced generalizability, we randomly sampled 49,825 pairs of images from the original dataset as a training set. After completing the model formation process, we need the test set to evaluate quality, so that the remaining 200 data pairs are used as the test set. Fig. 5 shows edge2shoes sample training information:

B. Implement Details

Our model is improved from pix2pix [15], to check the quality of the model generation, we compared with some models in the colorization task. All of these methods are performed using public codes and the Edges2shoes data set for training and testing in the same experimental environment for comparison purposes. We are experimenting on the computer with an RTX 3090 with 24 GB GPU memory and the PyTorch frame. The model parameters are optimized by the Adam optimizer [36], whose hyper-parameters β_1 and β_2 is set to 0.5 and 0.999, respectively. The training periods are 100 to ensure the convergence of model weights can converge.

Our loss model includes multi-scale GAN loss and feature mapping loss and the latter one plays a significant role in making the training process more stable. We carry out a hyper-parameter search on λ in order to obtain the most effective value for the model. We limit the lambda value between 1 and 10 since λ will bring a blurred image. Fig. 6 shows that all the metrics SSIM, MSE, PSNR achieve best at λ = 5. Consequently, we define λ = 5 in the training process.

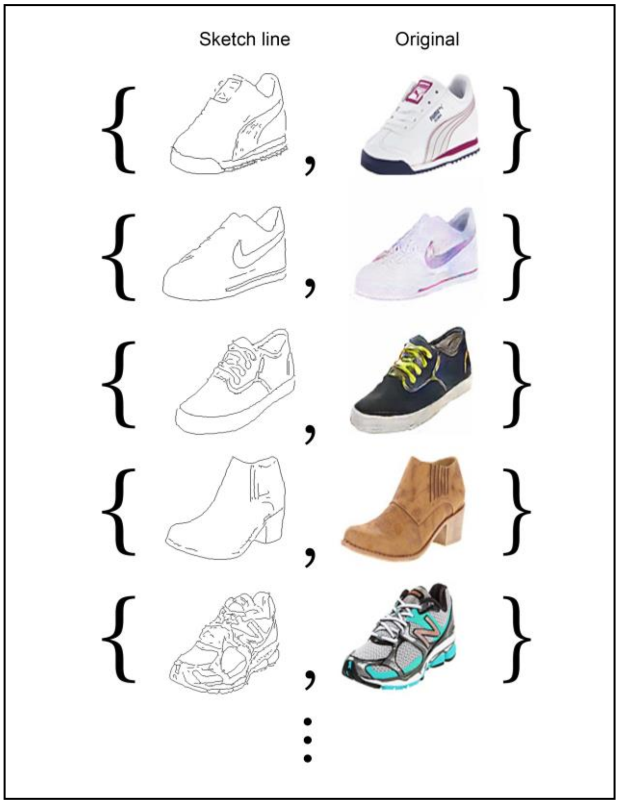

Fig. 5. Samples of edge2shoes in the training set.

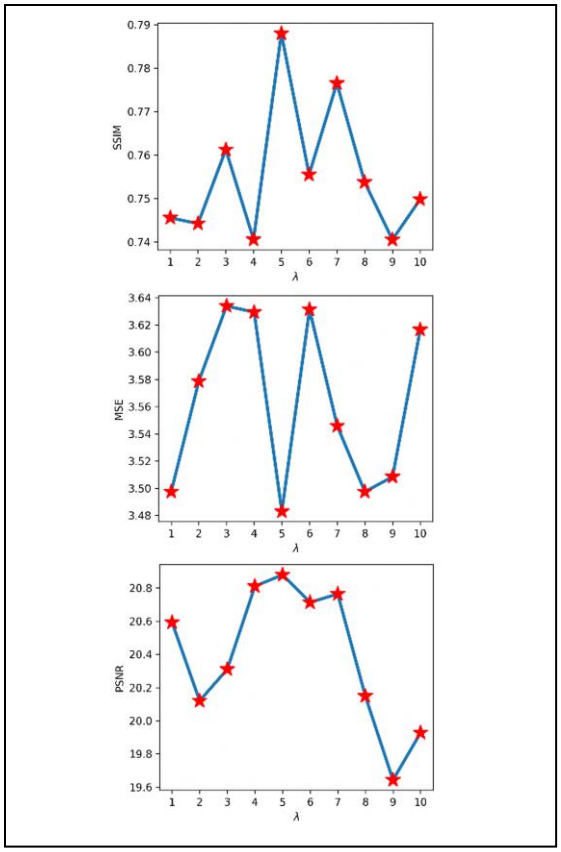

Fig. 6. SSIM, MSE, PSNR with different λ selection.

C. Comparison with state-of-the-arts

From the Table I, we can obtain that our ATTGAN method achieved a significant improvement in SSIM with the SOTA models. Regarding the performance of the PSNR metric for image quality measurement, our method may outperform the other models. In terms of MSE, our ATTGAN method can also achieve the lowest score, which shows that the difference between the image generated and the real image is smaller in pixel. The visualization results can be shown in Fig. 7. It is not difficult to see that our model can get different shaded areas according to precise colors. But they also have local details.

D. Ablation Study

We conduct ablation studies on the Edges2shoes data set to find out how the effectiveness of our upgrades are. From Table II, we can see that all assessment measures are improved, proving that all our improvements are effective. Among them, the metric PSNR has the largest optimization effect.

TABLE I. PERFORMANCE WITH DIFFERENT METHODS OF TRANSLATION ON EDGES2SHOES

Method	**Evaluation Metrics**		
	SSIM	***MSE***	***PSNR***
CycleGAN	0.5812	4.7832	16.1214
Pix2Pix	0.7192	3.6011	20.1621
NiceGAN	0.6374	4.5172	17.8964
ATTGAN	**0.7881**	**3.4835**	**20.8526**

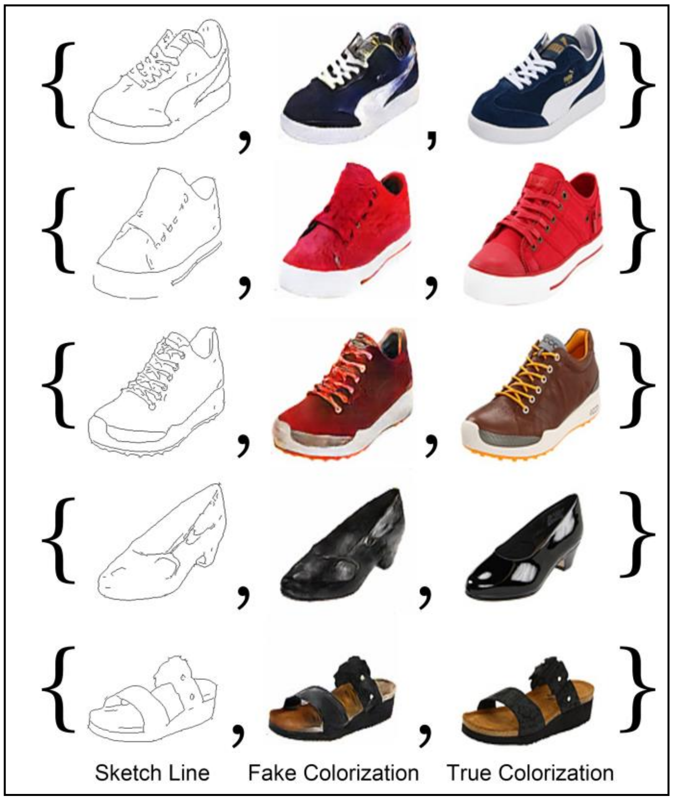

Fig. 7. Examples of the Edges2shoes colorization by our model.

TABLE II. PERFORMANCE WITH DIFFERENT COMPONENTS OF TRANSLATION ON EDGE2SHOES DATASET

Components			**Evaluation Metrics**		
Attention	Feature Fusion	Multi-scale D	PSNR	SSIM	MSE
×	×	×	20.1621	0.7192	3.6011
×	×	✓	20.2112	0.7231	3.5813
✓	×	×	20.1922	0.7205	3.5926
×	✓	×	20.5721	0.7649	3.5387
✓	✓	✓	**20.8526**	**0.7881**	**3.4835**

V. DISCUSSION

Here, we try to analyze the influence and importance of several aspects in our experiment.

First of all, the results of the experiment prove that in conditional GANs are able to generate high-resolution images of real footwear photos without any manual control or preparation. The combination of perception loss can improve the image generation results [37]. This approach is effective for studies where the aim is to produce high-resolution image results. At the same time, the researcher has benefited from this method in which pre-training for the exam is prohibited.

What's more, we used multiple discriminators in our GAN network and developed discriminator features exchanging an immersive opponent into effective filters. This GAN network introduces the N-discriminator extension, called Generative Multi-Adversary Networks, which make automatically progress in performance.

Third, the loss matching function fixes the formation where the generator is required to produce image statistics at various scales. This process improves the operational mechanism of the loss function.

Finally, we examine the evaluation measure by PSNR, SSIM, MSE. Treating the pix2pix method as a baseline, we evaluate that ATTGAN has the best performance compared with CycleGAN, Pix2pix and NiceGAN. By three translation methods on Edges2shoes, λ = 5 in the training procedure gain the best consequence to all. There is no doubt that all measures of evaluation have the highest ratings when it comes to drawing attention, presenting a characteristic and choosing a multi-scale discriminator. While others may have less effort and just below the best.

We believe that the automatic footwear colorization gives designers a chance to minimize the coloring time of the finish drawing line. To implement this automatic colorization footwear design, we are improving the resolution based on an earlier study in another area. At the same time, we make certain multi-discrimination to help the quality of the generation image. We may conclude that the approach for footwear colorization helps to accelerate the effectiveness of the coloring process. However, there are still limitations of the experiment. When the sketch line presented too complex or messy, it would have unavoidable identification errors throughout period. Nevertheless, the color slightly changed which caused by computing constraint.

In the future, designers may have more measures for eliminating colorization time. The development of automatic colorizing solves the problems for selecting color styles and finishing handmade draft. Furthermore, it becomes possible for customers to do DIY based on their individual sketching creation. With the help of artificial intelligence, the distance between manufacturer and consumer brings closeness. It might totally change the current situation of the footwear industry, especially in the post-pandemic era. Consumers may realize their own ideas by interaction media, which even reduce the working-hours of designers. It is no doubt that co-creation of products will be a trend of the times.

VI. Conclusion

In this research, we introduce a conditional GAN network for coloring sketches of shoes. The whole process divided into generator and discriminator, which help to perform better for coloring shoes. These methods decrease the difficulties of comparing evaluation measures. The results of our experiments are consistently visual. We believe that our method for coloring shoe sketches can have a good inspiration in the area of research to generate color sketch image samples.

References

[1] Eitz M, Hays J, Alexa M. How do humans sketch objects? [J]. ACM Transactions on graphics (TOG), 2012, 31(4): 1-10.

[2] Hesse C. Image-to-image demo: Interactive image translation with pix2pix-tensorflow[J]. Affinelayer. com, February, 2017, 19.

[3] Krizhevsky A, Sutskever I, Hinton G E. Imagenet classification with deep convolutional neural networks[J]. Communications of the ACM, 2017, 60(6): 84-90.

[4] He K, Zhang X, Ren S, et al. Deep residual learning for image recognition[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2016: 770-778.

[5] Schmarje L, Santarossa M, Schröder S M, et al. A survey on semi-, self- and unsupervised learning for image classification[J]. IEEE Access, 2021, 9: 82146-82168.

[6] Shorten C, Khoshgoftaar T M. A survey on image data augmentation for deep learning[J]. Journal of big data, 2019, 6(1): 1-48.

[7] LeCun Y, Bengio Y, Hinton G. Deep learning[J]. nature, 2015, 521(7553): 436-444.

[8] Goodfellow I, Pouget-Abadie J, Mirza M, et al. Generative adversarial networks[J]. Communications of the ACM, 2020, 63(11): 139-144.

[9] Radford A, Metz L, Chintala S. Unsupervised representation learning with deep convolutional generative adversarial networks[J]. arXiv preprint arXiv:1511.06434, 2015.

[10] Wang X, Liu H. Data supplement for a soft sensor using a new generative model based on a variational autoencoder and Wasserstein GAN[J]. Journal of Process Control, 2020, 85: 91-99.

[11] Karras T, Laine S, Aila T. A style-based generator architecture for generative adversarial networks[C]//Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2019: 4401-4410.

[12] Mirza M, Osindero S. Conditional generative adversarial nets[J]. arXiv preprint arXiv:1411.1784, 2014.

[13] Sangkloy P, Lu J, Fang C, et al. Scribbler: Controlling deep image synthesis with sketch and color[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2017: 5400-5409.

[14] Karacan L, Akata Z, Erdem A, et al. Learning to generate images of outdoor scenes from attributes and semantic layouts[J]. arXiv preprint arXiv:1612.00215, 2016.

[15] Pang Y, Lin J, Qin T, et al. Image-to-image translation: Methods and applications[J]. IEEE Transactions on Multimedia, 2021.

[16] Lee H Y, Tseng H Y, Mao Q, et al. Drit++: Diverse image-to-image translation via disentangled representations[J]. International Journal of Computer Vision, 2020, 128(10): 2402-2417.

[17] Liu M Y, Breuel T, Kautz J. Unsupervised image-to-image translation networks[J]. Advances in neural information processing systems, 2017, 30.

[18] Zhang R, Isola P, Efros A A. Colorful image colorization[C]//European conference on computer vision. Springer, Cham, 2016: 649-666.

[19] Levin A, Lischinski D, Weiss Y. Colorization using optimization[M]//ACM SIGGRAPH 2004 Papers. 2004: 689-694.

[20] Liu X, Luan R, Huang C. A novel encoding method for visual two-dimensional barcode using pattern substitution[C]//2010 International Conference on Machine Vision and Human-machine Interface. IEEE, 2010: 662-665.

[21] Cheng Z, Yang Q, Sheng B. Deep colorization[C]//Proceedings of the IEEE international conference on computer vision. 2015: 415-423.

[22] Gupta R K, Chia A Y S, Rajan D, et al. Image colorization using similar images[C]//Proceedings of the 20th ACM international conference on Multimedia. 2012: 369-378.

[23] Kumar M, Weissenborn D, Kalchbrenner N. Colorization transformer[J]. arXiv preprint arXiv:2102.04432, 2021.

[24] Wang T C, Liu M Y, Zhu J Y, et al. High-resolution image synthesis and semantic manipulation with conditional gans[C]//Proceedings of the

[25] Ronneberger O, Fischer P, Brox T. U-net: Convolutional networks for biomedical image segmentation[C]//International Conference on Medical image computing and computer-assisted intervention. Springer, Cham, 2015: 234-241.

[26] Long J, Shelhamer E, Darrell T. Fully convolutional networks for semantic segmentation[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2015: 3431-3440.

[27] Johnson J, Alahi A, Fei-Fei L. Perceptual losses for real-time style transfer and super-resolution[C]//European conference on computer vision. Springer, Cham, 2016: 694-711.

[28] Burt P J, Adelson E H. The Laplacian pyramid as a compact image code[M]//Readings in computer vision. Morgan Kaufmann, 1987: 671-679.

[29] Yildirim G, Jetchev N, Vollgraf R, et al. Generating high-resolution fashion model images wearing custom outfits[C]//Proceedings of the IEEE/CVF International Conference on Computer Vision Workshops. 2019: 0-0.

[30] Durugkar I, Gemp I, Mahadevan S. Generative multi-adversarial networks[J]. arXiv preprint arXiv:1611.01673, 2016.

[31] Hore A, Ziou D. Image quality metrics: PSNR vs. SSIM[C]//2010 20th international conference on pattern recognition. IEEE, 2010: 2366-2369.

[32] Wang Z, Bovik A C, Sheikh H R, et al. Image quality assessment: from error visibility to structural similarity[J]. IEEE transactions on image processing, 2004, 13(4): 600-612.

[33] Sara U, Akter M, Uddin M S. Image quality assessment through FSIM, SSIM, MSE and PSNR—a comparative study[J]. Journal of Computer and Communications, 2019, 7(3): 8-18.

[34] Huynh-Thu Q, Ghanbari M. Scope of validity of PSNR in image/video quality assessment[J]. Electronics letters, 2008, 44(13): 800-801.

[35] Qu Y, Chen Y, Huang J, et al. Enhanced pix2pix dehazing network[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2019: 8160-8168.

[36] Zhang Z. Improved adam optimizer for deep neural networks[C]//2018 IEEE/ACM 26th International Symposium on Quality of Service (IWQoS). Ieee, 2018: 1-2.

[37] Johnson J, Alahi A, Fei-Fei L. Perceptual losses for real-time style transfer and super-resolution[C]//European conference on computer vision. Springer, Cham, 2016: 694-711.