ArtPDGAN: Creating Artistic Pencil Drawing With Key Map Using Generative Adversarial Networks

SuChang Li[1], Kan Li[*1], Ilyes Kacher[2], Yuichiro Taira[3], Bungo Yanatori[3], and Imari Sato[4]

[1] Beijing Institute of Technology, Beijing, China
[2] Qwant Research
[3] Tokyo University of the Arts, Tokyo, Japan
[4] National Institute of Informatics, Tokyo, Japan

Abstract. A lot of researches focus on image transfer using deep learning, especially with generative adversarial networks (GANs). However, no existing methods can produce high quality artistic pencil drawings. First, artists do not convert all the details of the photos into the drawings. Instead, artists tend to use strategies to magnify some special parts of the items and cut others down. Second, the elements in artistic drawings may not be located precisely. What's more, the lines may not relate to the features of the items strictly. To address above challenges, we propose ArtPDGAN, a novel GAN based framework that combines an image-to-image network to generate key map. And then, we use the key map as an important part of input to generate artistic pencil drawings. The key map can show the key parts of the items to guide the generator. We use a paired and unaligned artistic drawing dataset containing high-resolution photos of items and corresponding professional artistic pencil drawings to train ArtPDGAN. Results of our experiments show that the proposed framework performs excellently against existing methods in terms of similarity to artist's work and user evaluations.

Keywords: Generative Adversarial Networks · Deep learning · Artistic pencil drawing

1 Introduction

Pencil drawing is one of the most appreciated technique in quick sketching or finely-worked depiction. Researchers are quite interested in pencil drawings because it is a combination of observation, analysis and experience of the authors. So study pencil drawing can be help to the progress of artificial intelligence. And it usually takes several hours to finish a fine drawing (Fig 1), even for an experienced artist with professional training, which attracts people work on the pencil drawing generation algorithms. In former methods, pencil drawing generation

---

[*] Corresponding author: likan@bit.edu.cn

was split into two components, the structure map that define region boundaries, and the tone map that reflects differences in the amount of light falling on a region as well as its intensity or tone and even texture[20]. However, we learned from artists that artistic pencil drawing should be able to capture the characteristics of the items and emphasize them. We give the images which key part have been labeled a name called key map. The key maps labeled by artists are also shown in Fig 1.

Designing an algorithm or a framework which can study from artistic drawings and automatically transform an input photo into high-quality artistic drawings is highly desired. It can be used in many areas such as animation and advertisement. In particular, the development of deep learning which uses networks to perform image style transfer was also proposed[5]. Recently, generative adversarial network (GAN)[8] based style transfer methods (e.g. [1, 2, 11, 30]) with datasets of (paired or unpaired) photos and stylized images have achieved abundant good results.

Based the knowledge of artists, generating artistic pencil drawings are quite different with pencil styles studied in previous work[17, 20]. The differences can be summarized into three aspects. First, the artists will not convert all the details of the photos directly into their drawings, they will find the most important regions to magnify and simplify other parts at the same time. Second, artists will not locate the elements in pencil drawings precisely, which makes it a challenge for the methods based on similarity or correspondence (e.g. Pix2Pix[11]). Finally, artists put lines in pencil drawings that are not directly related to the basic vision features in the view or photograph of the items. Therefore, even state-of-the-art image style transfer algorithms (e.g. [5, 11, 16, 18, 22, 30]) often fail to produce vivid and realistic artistic pencil drawings.

To address the above challenges, we propose ArtPDGAN, a novel GAN based architecture which combines with an image-to-image network for transforming photos to high-quality artistic pencil drawings. ArtPDGAN can generate key maps for the original photos and use the key maps to synthesis the artistic pencil drawings. To learn key region for different object shapes effectively, our architecture involves a specialized image-to-image network to capture key map.

The main contributions of this work are summarized as follows:

- We propose a GAN based framework for artistic pencil drawing generation, which combines with a specialized image-to-image network to generate high-quality and expressive artistic drawings from real photos.
- In order to imitate artists better, we also propose a key map dedicated to the artists' emphasizing parts. This make our model more imitation of the artist than previous works. To our knowledge, it is the first one to apply the key map in artistic style transfer, which is an idea based on the knowledge of artists.
- The experiments demonstrate the ability of our model to synthesize artistic pencil drawings which are more close to the artists' drawings than the state-of-the-art methods.
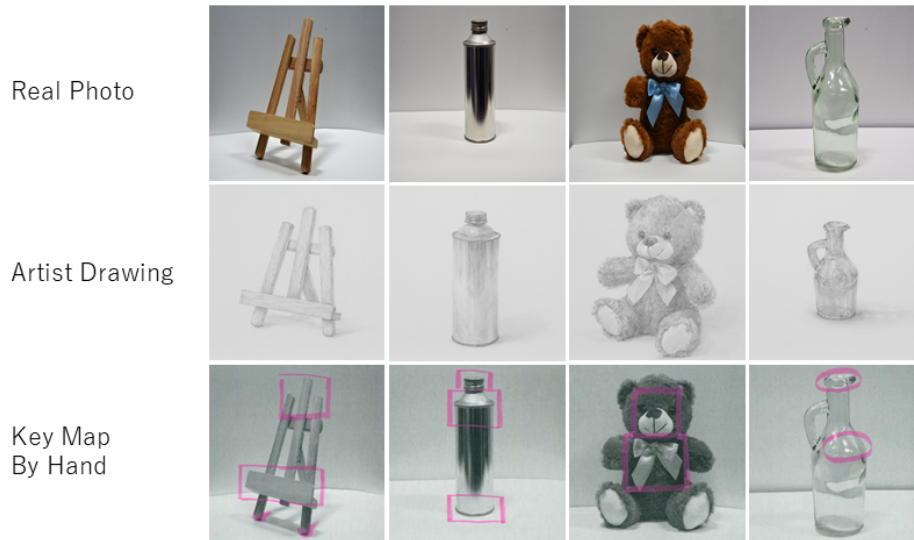
Fig. 1. Examples of artistic pencil drawing and key map

2 Related Work

Pencil drawing generation has been widely studied in sketch extraction and deep learning style transfer. In this section, we will summarize related work in these two aspects respectively.

2.1 Sketch Extraction

Traditional edge extraction methods like[7, 14, 3] usually deal with the edge extraction problem with fuzzy mathematics as well as other algebraic algorithms. However, the edges are not as natural as human-made ones even though they are easy to calculate because of their discontinuity. Works such as [29, 26] are also based on neural network. However, their results are still quite different with artistic drawings.

2.2 Style transfer using neural networks

Large numbers of approaches have been proposed to learn style transfer from examples, because it is too hard to describe the styles semantically. The Image Analogy approach[9] requires the input and the output are strictly aligned because it is designed for pairs training. Liao et al.[18] proposed a method called Deep Image Analogy which also requires the example images and target image to have similar content, if they are not aligned with each other. By finding semantically meaningful correspondences between two input images, they first compute correspondences between feature maps extracted by a network, and

then finish visual attribute transfer. Deep Image Analogy works well on photo-to-style transfer problem, but when being applied to our artistic pencil drawing style, the subjects in the generated images look fuzzy due to the light texture. Due to the difficulty of gathering aligned and paired data, advanced neural style transfer methods (e.g. [11, 25, 31]) can be hardly used in common style transfer applications.

There are also methods (e.g. [19, 13, 15, 10, 30]) which can learn mappings from unpaired data. Most of them are designed with cycle-consistency theorem. However, these methods do not work well on capturing the tone texture.

Gatys et al. [5] is a milestone for neural style transfer research. They use the Gram matrices[4] in a VGG network to capture the content and style representations of images. The stylization process is achieved by minimizing the distances between the style features of the content image and the style image. This method works well on oil painting style transfer and get good results for famous artists' styles like Van Gogh. But it is not suitable for pencil drawing style transfer because it takes style as a kind of texture while style in pencil drawing is too little to be easily captured. The perceptual loss which based on high-level features was proposed by Johnson et al.[12], and became one of the best loss in image style transfer problem. For the same reason as [6], their texture-based loss function is not suitable for our style. In addition to aforementioned limitations for artistic pencil drawing style transfer, most existing methods require the style image to be close to the content image. And they do not apply the key map to improve the generated results.

Our models are different with the works above. We use an image-to-image network to generate key map, then combine it with structure map and tone map as the input to our generator. Our ArtPDGAN is able to use the features from these three kinds of maps and generate artistic pencil drawings.

3   Model

Our approach is based on the knowledge learned from artists that pencil drawings can be separated into three main components: structure map, tone map and key map. Each of them delineate different parts of the pencil drawings. The structure map is used to detect object boundaries as well as other boundaries in the scene. The cross hatching and other tonal technique for lighting, texture, and materials should be shown in tone map. The key map which our method replied on can help to find the key parts of the items. Our method combines all the maps above to do the style transformation.

With the dataset provided by artists, our model is trained to learn to translate from input maps to artistic pencil drawings. Our training process is based on the idea of paired image-to-image translation frameworks, though our model does not need the aligned data. Since the artists will abstract and transform the photos, the aligned training data are not helpful to artistic characteristics. We use two GAN-based models in our framework, one is used for generating the artistic pencil drawings and the other is for the key map generation. These two

models are trained together because we want the whole framework can learn how to imitate the artists. Fig 2 shows the main modules of the proposed framework. We generate the training maps by using abstraction filters and functions on pencil drawings to estimates edges and tones. These filters can produce similar abstract maps from pencil drawings as from real photographs. Hence, at test time, we use the same abstraction filters on the input photograph, to produce input maps which are in the same domain as the training inputs.
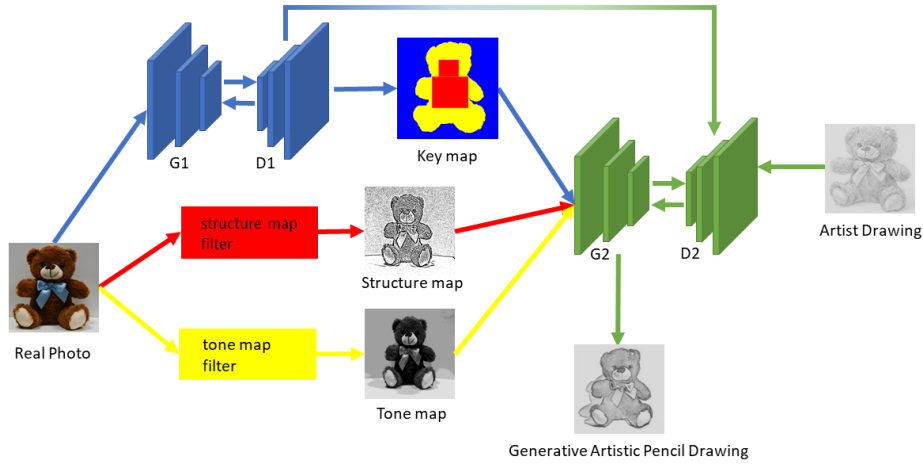

Fig. 2. The structure of the ArtPDGAN

3.1 Structure map and tone map

For generating structure map, we use the adaptive filter[28], which can mark material edges of images, even the highly-textured ones. To extract the tone map, we apply a simple mapping on the pixels of photos to generate a smoothing output as the tone extraction. The mapping can be formulated as

$$Tone\ map = \lfloor original\ image/9 \rfloor \tag{1}$$

This formula is based on the experience of the artists. The artists only use a few channels of tone compared with real photos. So we directly compress the values of the pixels. Examples of structure map and tone map are shown in Fig 3. The tone maps are actually invisible and we managed to show them as images.

3.2 Key Map

The goal of our key map is to help the generator network learn the artistic characteristics. Our model mainly based on the fact that artists will use their
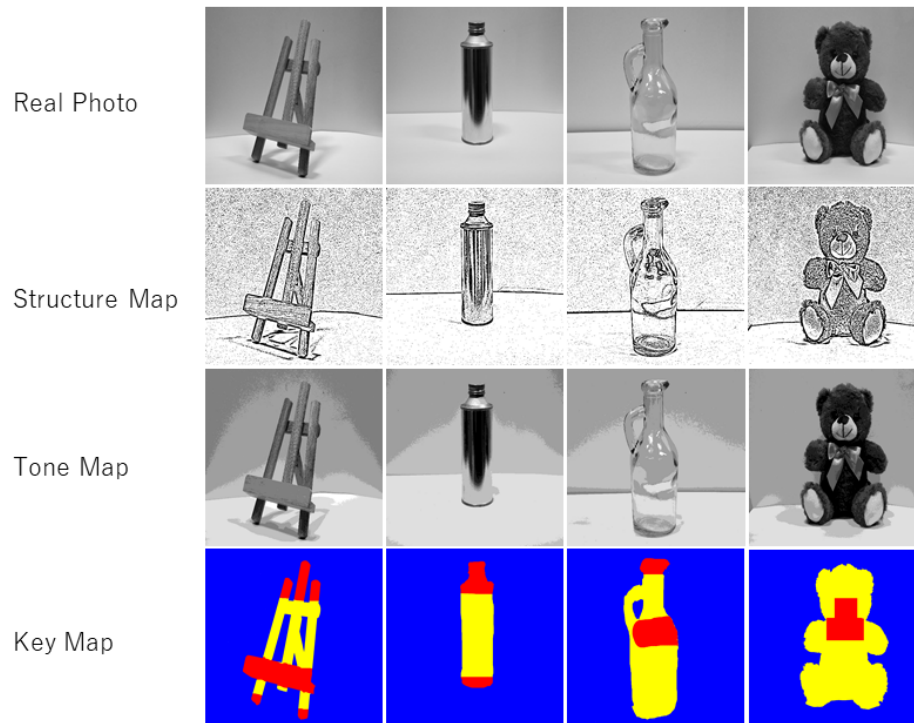
Fig. 3. Examples of structure map, tone map and key map

experience to find the key parts and emphasize them. Our model can learn from the labeled key map and find the key parts by itself as the artists do. Examples of key map used by our model are shown in Fig 3. The key map network aims at extracting key regions from the photos. The important parts in the key maps are marked in red. As it's impossible to learn the key map for every photo at test time, we apply a specialized image-to-image network based GAN model to preprocess inputs and generate the key map at test time. Our training key map pair is labeled by artist. And the 300 training pairs contains almost all common shapes. The image-to-image network contains the Generator G_1 and the Discriminator D_1, and use the loss as [11]

$$L = arg \min_{G_1} \max_{D_1} L_{cGAN}(G_1, D_1) + \lambda L_{L1}(G_1) \qquad (2)$$

where

$$L_{cGAN} = E_{x,y}[log\, D_1(x,y)] + E_{x,z}[log(1 - D_1(G_1(x,z)))] \qquad (3)$$

During the training stage, the model G_1 and D_1 will work together to learn how to find the key map by using aligned and paired data created from key map labeled by artists and real photos. Then the whole framework learns to apply the emphasizing according to the key maps. At test-time, the model will generate the key map using real photos only.

3.3 Pencil Drawing Network

The G_2 and D_2 are the Generator and Discriminator of pencil drawing network(Fig 2). We first use the key map network to generate the key maps. Then we encode key map and the real photos as feature maps and concatenated with the features of the structure map and the tone map. The pencil drawing network will use the features to generate artistic pencil drawings. Due to the lines of artistic pencil drawings are not related to the real photo precisely, there always some artifacts around the lines and make the results look fuzzy. To address this issue, we apply feature matching based on the idea of [21]. We extract the features from the Discriminator D_2 and regulate them to match the features from the Discriminator D_1 of the key map network.

3.4 Loss function

As tried in the previous works[17], loss functions do not perform well alone, our loss function is combined by three loss functions. However, we do not use the classical pixel-based reconstruction loss L_{rec}. The reason is described above, as the artists will not directly change every detail of the real photos into drawings, the L_{rec} will lead the model to a less artistic results. Our loss function can be described as

$$L_{all} = \alpha * L_{adv} + \beta * L_{fea} + \gamma * L_{per} \qquad (4)$$

Adversarial loss As the traditional conditional GAN, we use the discriminator network D_2 to discriminate the real samples from the pencil drawings and

generated results. And the goal of the generator G_2 works on opposite, trying to generate images which cannot be judged from the real ones by the discriminator D_2. This can be achieved by using an adversarial loss:

$$L_{adv} = \min_{G_2} \max_{D_2} P_Y[logD_2(y)] + P_X[log(1 - D_2(G_2(x)))] \tag{5}$$

In which P_Y and P_X represent the distributions of pencil drawing samples y and their generated samples x.

Feature match loss To match features extract from the Discriminator of image-to-image network D_1 and the Discriminator of pencil drawing network D_2, we use formula introduced by [21].

$$L_{fea} = \|f(D_1(x)) - f(D_2(x))\|_2^2 \tag{6}$$

In which $f(x)$ denote activations on an intermediate layer of the discriminator.

Perceptual loss The perceptual loss[12] was performed well in minimizing the feature differences. It makes the results sharper than traditional reconstruction loss L_{rec}. So we also apply it to help the generated samples look more close to the artistic pencil drawing.

$$L_{per} = \sum_{i=1}^{4} \|\Phi_i(G_2(x)) - \Phi_i(y)\|_2^2 \tag{7}$$

where x, y are the input and the pencil drawing from artists, G is the translation model, and Φ stands for the VGG-19[23] network up to the $ReLU\_i\_1$ layer.

4   Experiments

We implemented ArtPDGAN using PyTorch[24] and execute experiments on a computer with an NVIDIA Titan X GPU. Our model only need 200 epochs' training, and the training time is about 6 hours. The generator G takes color photos as input and output the gray drawings whose size is 512*512. So the numbers of input and output channels are 3 and 1, respectively. In all our experiments, the parameters in Eq.4 are fixed at $\alpha = 1.0, \beta = 0.5, \gamma = 0.5$. In order to guarantee the fairness, all the evaluation results showed in this section are based on the generated results of test data, and all the images are resized to 256*256.

The generated results are shown in the Fig 4. We use 100 images in user study 1 and 80 images in user study 2. We finally collect the feedback from 108 users of totally 9920 scores.

4.1   Ablation study in ArtPDGAN

We perform an ablation study on our unique factor, the key map. As it shown in Fig 5, the user study between ArtPDGAN and ArtPDGAN without key map show that the key map is critical to our ArtPDGAN and help to produce high-quality results of artistic pencil drawings.
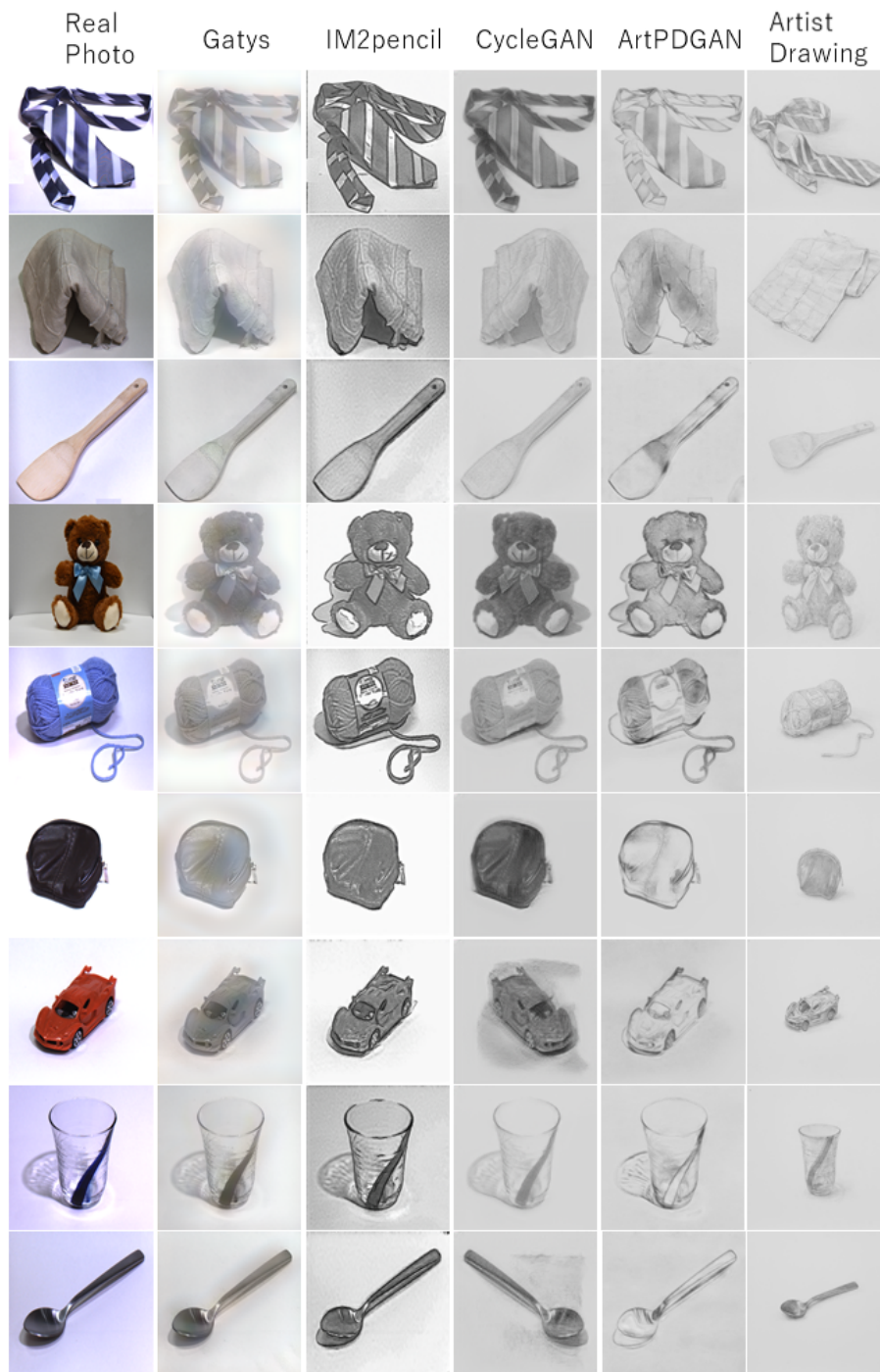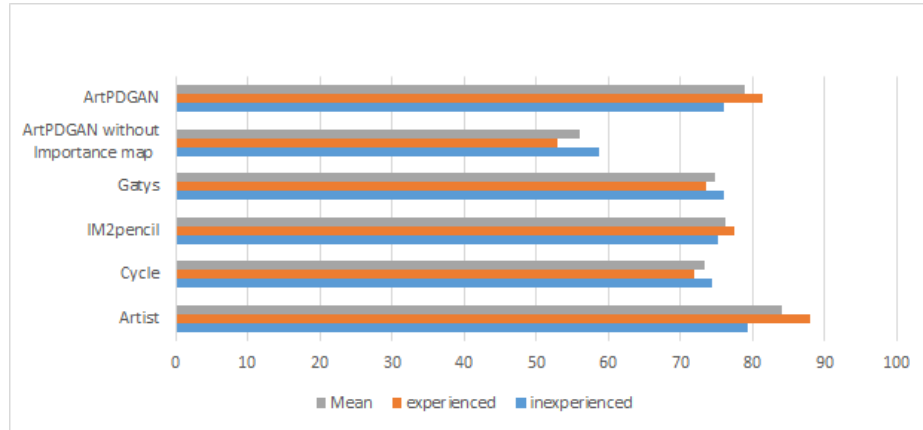
Fig. 4. The generated results

Fig. 5. The results of user study 1

4.2   Comparison with state-of-the-art

We compare ArtPDGAN with three state-of-the-art style transfer methods: Gatys, CycleGAN and Im2Pencil.
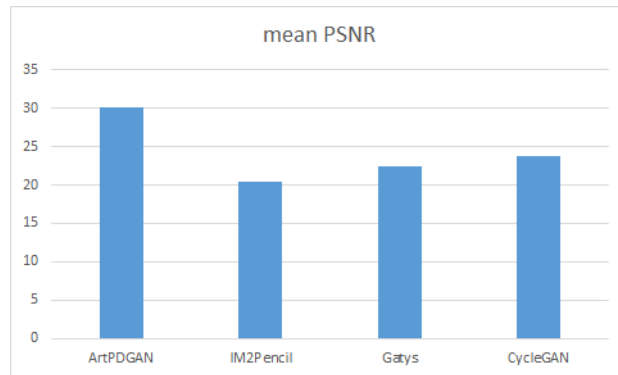

Fig. 6. The results of PSNR

Gatys, Pix2Pix and CycleGAN are classical and famous style transfer models of different types of data. Im2Pencil is accepted by CVPR 2019, which stands for the latest approach. Our dataset is made up of paired and unaligned data, which do not satisfy the conditions of the Pix2Pix model. So we choose CycleGAN as the comparative method because it can be applied in the unpaired and unaligned dataset. Qualitative results of comparison with Gatys, CycleGAN and Im2Pencil are shown in Fig 5, Fig 6, Fig 7, Fig 8 and Table 1. Fig 8 shows the results of one of the user studies, which rates the similarity between the results of different
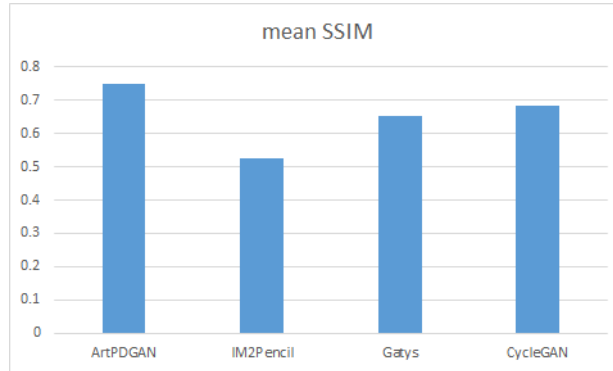
Fig. 7. The results of SSIM

algorithms and the artists' works. In the user study, we divided the users into two groups, an inexperienced user group and an experienced user group, based on their drawing experience. The reason is that users with different drawing experience may have different focuses while looking at images. Experienced users my easily to find out the key map like the artist, while inexperienced users may pay more attention on realistic details of the object in the drawings.

Table 1. Experiment results

Model	mean PSNR	max PSNR	mean SSIM	max SSIM
ArtPDGAN	30.0945	34.5141	0.7502	0.8368
Im2Pencil	18.5237	24.1215	0.5256	0.7295
Gatys	22.3423	26.6630	0.6521	0.7226
CycleGAN	23.7886	31.7173	0.6838	0.7801

Gatys' method takes one content image and one style image as input by default, so we use the exact drawing in the training set as the style and content image to model the target style for a fair comparison. Im2pencil provides many styles to choose from, we just use the fundamental style in all the experiment. As Fig 5 shows, Gatys' method generates good results for artistic pencil stylization. The CycleGAN gets good results among PSNR, SSIM and user study, while the Im2pencil only gets very good manual score. And our ArtPDGAN is always the best one among these methods.

We can find that Gatys's results lose some features and have two or more kinds of styles of regions. And the results contain many artifacts. The reason behind is that the method regards style as texture information in the Gram matrix, which cannot capture our artistic pencil style which only has little texture. And the artistic deformation caused the imprecise content loss. CycleGAN
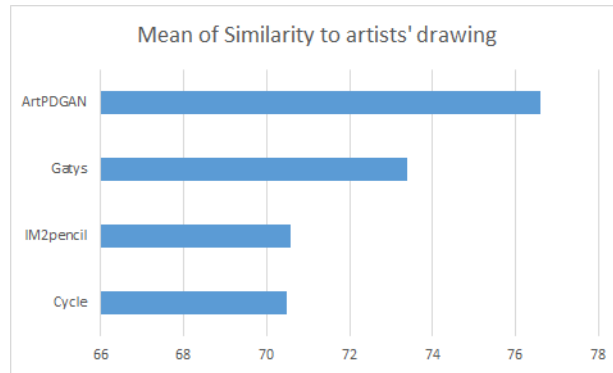
Fig. 8. The results of user study 2

also cannot imitate the artistic drawings well. As shown in Fig 8, CycleGAN's results do not look like an artist's drawing. CycleGAN is unable to preserve important features because it uses the cycle-consistency to constrain network. The cycle-consistency which uses only unsupervised information, is less accurate than a supervised method and leads to problems that it's hard for the loss to accurately recover all the details of domains. Im2pencil generates results that preserve some aspect of artistic drawings, but they also have many artifacts. The structural lines, making the stylized result unlike the artistic drawings, and the xDOG filter[27] makes the tone maps look too close to black-and-white photos. The main reason of these problems is that they did not use the real pictures drawing by artists.

In comparison, our method captures the key regions accurately and generates high-quality results with artist's drawing style. Moreover, our results are very close to the drawings drawn by the artists than other methods. For quantitative evaluation, we compare our ArtPDGAN with artists' works, Gatys, CycleGAN and Im2Pencil using the user study, which are a widely used in GAN evaluation. We also measure the similarity between generated artistic pencil drawings and real artistic pencil drawings use the PSNR, SSIM and the user study at the same time. The comparison results are presented in Table 1. The scores show that our method has a higher PSNR value and SSIM value, which means that our drawings is closer to the artistic pencil drawings than Im2Pencil, CycleGAN and Gtays. The same as the results of user investigation. In other words, these results indicate ArtPDGAN captures better artistic pencil drawings distribution than comparative methods.

5   Conclusion and Future Work

In this paper, we propose ArtPDGAN, a framework which can transform photos into artistic pencil drawings. We use different filters to produce structure maps and tone maps of the pencil drawings, as well as the key maps which generated

from the original photos to train the network to transfer photos into artistic drawings. To imitate the artists' skills as much as possible and avoid the time consuming of aligned data collection for future application, our model uses unaligned data for the training. Experiment results and user study both show that our method can successfully complete artistic style transfer and outperform the state-of-the-art methods. In the future, we will try to use our key maps combined with the technology of instance segmentation to deal with more complex photos.

References

1. Azadi, S., Fisher, M., Kim, V.G., Wang, Z., Shechtman, E., Darrell, T.: Multi-content gan for few-shot font style transfer. arXiv: Computer Vision and Pattern Recognition (2017)
2. Chen, Y., Lai, Y., Liu, Y.: Cartoongan: Generative adversarial networks for photo cartoonization pp. 9465–9474 (2018)
3. Gao, W., Zhang, X., Yang, L., Liu, H.: An improved sobel edge detection **5**, 67–71 (2010)
4. Gatys, L.A., Ecker, A.S., Bethge, M.: Texture synthesis using convolutional neural networks. arXiv: Computer Vision and Pattern Recognition (2015)
5. Gatys, L.A., Ecker, A.S., Bethge, M.: Image style transfer using convolutional neural networks pp. 2414–2423 (2016)
6. Gatys, L.A., Ecker, A.S., Bethge, M., Hertzmann, A., Shechtman, E.: Controlling perceptual factors in neural style transfer pp. 3730–3738 (2017)
7. Gonzalez, C.I., Melin, P., Castro, J.R., Mendoza, O., Castillo, O.: An improved sobel edge detection method based on generalized type-2 fuzzy logic. soft computing **20**(2), 773–784 (2016)
8. Goodfellow, I., Pougetabadie, J., Mirza, M., Xu, B., Wardefarley, D., Ozair, S., Courville, A., Bengio, Y.: Generative adversarial nets pp. 2672–2680 (2014)
9. Hertzmann, A., Jacobs, C.E., Oliver, N., Curless, B., Salesin, D.: Image analogies pp. 327–340 (2001)
10. Huang, X., Liu, M., Belongie, S., Kautz, J.: Multimodal unsupervised image-to-image translation. arXiv: Computer Vision and Pattern Recognition (2018)
11. Isola, P., Zhu, J., Zhou, T., Efros, A.A.: Image-to-image translation with conditional adversarial networks. arXiv: Computer Vision and Pattern Recognition (2016)
12. Johnson, J., Alahi, A., Feifei, L.: Perceptual losses for real-time style transfer and super-resolution. arXiv: Computer Vision and Pattern Recognition (2016)
13. Kim, T., Cha, M., Kim, H., Lee, J.K., Kim, J.: Learning to discover cross-domain relations with generative adversarial networks pp. 1857–1865 (2017)
14. Kumar, S., Saxena, R., Singh, K.: Fractional fourier transform and fractional-order calculus-based image edge detection. Circuits Systems and Signal Processing **36**(4), 1493–1513 (2017)

15. Lee, H., Tseng, H., Huang, J., Singh, M., Yang, M.: Diverse image-to-image translation via disentangled representations pp. 36–52 (2018)
16. Li, C., Wand, M.: Combining markov random fields and convolutional neural networks for image synthesis. arXiv: Computer Vision and Pattern Recognition (2016)
17. Li, Y., Fang, C., Hertzmann, A., Shechtman, E., Yang, M.: Im2pencil: Controllable pencil illustration from photographs pp. 1525–1534 (2019)
18. Liao, J., Yao, Y., Yuan, L., Hua, G., Kang, S.B.: Visual attribute transfer through deep image analogy. ACM Transactions on Graphics **36**(4), 120 (2017)
19. Liu, M., Breuel, T.M., Kautz, J.: Unsupervised image-to-image translation networks. arXiv: Computer Vision and Pattern Recognition (2017)
20. Lu, C., Xu, L., Jia, J.: Combining sketch and tone for pencil drawing production pp. 65–73 (2012)
21. Salimans, T., Goodfellow, I., Zaremba, W., Cheung, V., Radford, A., Chen, X.: Improved techniques for training gans. arXiv: Learning (2016)
22. Shih, Y., Paris, S., Barnes, C., Freeman, W.T., Durand, F.: Style transfer for headshot portraits. international conference on computer graphics and interactive techniques **33**(4), 148 (2014)
23. Simonyan, K., Zisserman, A.: Very deep convolutional networks for large-scale image recognition. arXiv: Computer Vision and Pattern Recognition (2014)
24. Steiner, B., Devito, Z., Chintala, S., Gross, S., Paszke, A., Massa, F., Lerer, A., Chanan, G., Lin, Z., Yang, E., et al.: Pytorch: An imperative style, high-performance deep learning library (2019)
25. Wang, T., Liu, M., Zhu, J., Tao, A., Kautz, J., Catanzaro, B.: High-resolution image synthesis and semantic manipulation with conditional gans pp. 8798–8807 (2018)
26. Wei, Y., Wang, Z., Xu, M.: Road structure refined cnn for road extraction in aerial image. IEEE Geoscience and Remote Sensing Letters **14**(5), 709–713 (2017)
27. Winnemoller, H.: Xdog: advanced image stylization with extended difference-of-gaussians pp. 147–156 (2011)
28. Wittenmark, B.: Adaptive filter theory. Automatica **29**(2), 567–568 (1993)
29. Xie, S., Tu, Z.: Holistically-nested edge detection. arXiv: Computer Vision and Pattern Recognition (2015)
30. Zhu, J., Park, T., Isola, P., Efros, A.A.: Unpaired image-to-image translation using cycle-consistent adversarial networks pp. 2242–2251 (2017)
31. Zhu, J., Zhang, R., Pathak, D., Darrell, T., Efros, A.A., Wang, O., Shechtman, E.: Toward multimodal image-to-image translation. arXiv: Computer Vision and Pattern Recognition (2017)