Final Report

Wang Ao
Tsinghua University
2017010395
wa17@mails.tsinghua.edu.cn

Sun Ziping
Tsinghua University
2015013249
sunziping2016@yeah.net

Cui Yanfei
Tsinghua University
2017012326
929881841@qq.com

Abstract

We study to generate anime avatars using GAN. In the past few years, the field of computer vision has achieved rapid development, and the Generative Adverarial Network [5] has attracted wide attention since it was proposed in 2014. It is considered to be one of the biggest breakthrough in deep learning in recent years. The number of papers surrounding GANs has also increased rapidly. In many projects around GAN, generating anime avatars is an interesting and practical task. On the one hand, there have been many articles to achieve the task of generating second-dimensional images, such as dcGAN [12] and CartoonGAN [1]. On the other hand, this project also has many problems to be solved. For example, it is difficult to have a good evaluation standard in the quality evaluation of generated images. We compare the different results between CartoonGAN [1] and dcGAN [12], and try to give a new loss function to perform better.

1. Introduction

Generating anime avatars is a very interesting task. Since the introduction of GAN [5], there have been many networks using GAN [5] to generate anime avatars. There are currently two main ways to achieve this task: the first is to directly generate a anime avatar and the input is random noise; the second belongs to the category of picture style transfer, that is generating corresponding two-dimensional animation Style avatar with given real character avatar.

For the first approach, the current dcGAN [12] has achieved very good results. The character's head of the output picture is very clear, the lines are very coherent, and there is no serious deformation. But it may be because the input is random noise and the data set is not large enough, the homogeneity of the generated content is more serious, and the characters always look like the same. At the same time, it is difficult for us to have a better evaluation index to judge the effectiveness of the model. In addition, the training of GAN is very unstable, which leads to the need for a large amount of hyperparameter tuning work. Continue adding more information about direct generating...

As for the second approach, CartoonGAN [1] is able to successfully convert landscape photos into anime style. When we apply CartoonGAN [1] to the task of character head style transfer, we need a new dataset, a new feature extraction module, and an adaptation to the face. In the feature extraction module, we use the classic VGG19 [17] as the feature extraction module of the picture to extract the animation elements. For the human face adaptation, we added a landmark loss function of the human face to mark the key points of the human face and help the discriminator to better identify whether the generated image is a human face. In addition, we have also added style loss function, image content loss function, grayscale loss function and color loss function, making it less difficult to judge the quality of image generation.

2. Related Work

2.1. Non-photorealistic rendering (NPR)

In order to mimic specific artistic styles (including animation [14]), many automatic and semi-automatic NPR algorithms have been developed. Most of the works are animated by using a simple shadow rendering method [16]. One technique, called *cel shading*, is widely used in the creation of games, animations, and movies, saving artists a lot of time [11].

To mimic the cartoon style, people have developed various methods to create images with flat shadows. These methods either use image filtering [18] or use specific transformation formulas [19] in optimization problems. However, it is difficult to capture rich artistic styles using only simple mathematical formulas. In particular, filtering or optimizing the entire image does not create the high-level abstractions that artists typically require. For portraits, people also have special algorithms [20, 15], in which semantic segmentation can be automatically derived by detecting facial components. However, these methods cannot handle general images.

2.2. Convolutional neural networks

With the great success of convolutional neural networks [8, 9] in the field of computer vision, people are looking forward to their performance in the field of image style transfer and image generation. Compared with the traditional complex NPR algorithm [16, 11], CNN is indeed more convenient and more applicable. For example, in the task of image style transfer, the VGG network [17] has a good ability to extract picture features.

For the style and content of the image, Gatys et al. [4] first proposed a CNN-based neural style transfer (NST) method that can transfer the style of a picture from one picture to another. They use a feature map of a pre-trained VGG network to extract picture content and optimize the resulting image, so that it can match the corresponding texture information described by the global Gram matrix [3] while retaining the original content of the image. However, such operations caused a serious loss of the edge information and shadow information of the picture.

2.3. Image synthesis with GANs

The other genre using GANs [5] has achieved great imporvement. It has achieved the state of the art results in the fields of text-to-image translation [13], image inpainting [21], and image super-resolution [10], etc. The key idea of the GAN model is to train two networks (generator and discriminator). Iteratively, the adversarial loss provided by the discriminator transforms the generated image into a target manifold [21]. However, GANs are very unstable to train and often make the generator produce meaningless output.

Some literatures [2, 6, 7] provided solutions using GANs for pixel-level image synthesis problems. However, these methods require paired data sets during the training process, but such high-quality data sets are difficult to obtain and therefore cannot be used for our training.

On the one hand, CartoonGAN effectively solved the above problems by using a GAN model to learn the mapping between photos using unpaired training data and cartoon manifolds. They formulate the process of learning to transform realworld photos into cartoon images as a mapping function which maps the photo manifold to the cartoon manifold [1]. Additionally, they proposed edge-promoting adversarial loss in order to generate sharper edges just like anime works. However, their work places high demands on the quality of the training set, the input figure must have a high color saturation to get anime-like outputs. At the same time, it doesn't provide a good method to guarantee that the image is a portrait.

On the other hand, dcGAN [12] made some great improvements on the basis of traditional GAN. It set a series of restrictions for the network topology of CNN to make it stable training, and used the obtained feature representations for image classification so that it get better results to verify the expression ability of the generated image feature representations. Continue adding more information about dcGAN's advantages and limitation...

3. Approach

4. Experiment

5. Conclusion

References

[1] Y. Chen, Y.-K. Lai, and Y.-J. Liu. Cartoongan: Generative adversarial networks for photo cartoonization. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 9465–9474, 2018.

[2] V. Dumoulin, I. Belghazi, B. Poole, O. Mastropietro, A. Lamb, M. Arjovsky, and A. Courville. Adversarially learned inference. *arXiv preprint arXiv:1606.00704*, 2016.

[3] L. A. Gatys, A. S. Ecker, and M. Bethge. Texture synthesis and the controlled generation of natural stimuli using convolutional neural networks. *arXiv preprint arXiv:1505.07376*, 12:4, 2015.

[4] L. A. Gatys, A. S. Ecker, and M. Bethge. Image style transfer using convolutional neural networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 2414–2423, 2016.

[5] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio. Generative adversarial nets. In *Advances in neural information processing systems*, pages 2672–2680, 2014.

[6] P. Isola, J.-Y. Zhu, T. Zhou, and A. A. Efros. Image-to-image translation with conditional adversarial networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1125–1134, 2017.

[7] L. Karacan, Z. Akata, A. Erdem, and E. Erdem. Learning to generate images of outdoor scenes from attributes and semantic layouts. *arXiv preprint arXiv:1612.00215*, 2016.

[8] A. Krizhevsky, I. Sutskever, and G. E. Hinton. Imagenet classification with deep convolutional neural networks. In *Advances in neural information processing systems*, pages 1097–1105, 2012.

[9] S. Lawrence, C. L. Giles, A. C. Tsoi, and A. D. Back. Face recognition: A convolutional neural-network approach. *IEEE transactions on neural networks*, 8(1):98–113, 1997.

[10] C. Ledig, L. Theis, F. Huszár, J. Caballero, A. Cunningham, A. Acosta, A. Aitken, A. Tejani, J. Totz, Z. Wang, et al. Photo-realistic single image super-resolution using a generative adversarial network. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 4681–4690, 2017.

[11] R. R. Luque. The cel shading technique. Technical report, Citeseer, 2012.

[12] A. Radford, L. Metz, and S. Chintala. Unsupervised representation learning with deep convolutional generative adversarial networks. In *International Conference on Learning Representations*, 2016.

[13] S. Reed, Z. Akata, X. Yan, L. Logeswaran, B. Schiele, and H. Lee. Generative adversarial text to image synthesis. *arXiv preprint arXiv:1605.05396*, 2016.

[14] P. Rosin and J. Collomosse. *Image and video-based artistic stylisation*, volume 42. Springer Science & Business Media, 2012.

[15] P. L. Rosin and Y.-K. Lai. Non-photorealistic rendering of portraits. In *Proceedings of the workshop on Computational Aesthetics*, pages 159–170. Eurographics Association, 2015.

[16] T. Saito and T. Takahashi. Comprehensible rendering of 3-d shapes. In *ACM SIGGRAPH Computer Graphics*, volume 24, pages 197–206. ACM, 1990.

[17] K. Simonyan and A. Zisserman. Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*, 2014.

[18] H. Winnemöller, S. C. Olsen, and B. Gooch. Real-time video abstraction. In *ACM Transactions On Graphics (TOG)*, volume 25, pages 1221–1226. ACM, 2006.

[19] L. Xu, C. Lu, Y. Xu, and J. Jia. Image smoothing via l 0 gradient minimization. In *ACM Transactions on Graphics (TOG)*, volume 30, page 174. ACM, 2011.

[20] M. Yang, S. Lin, P. Luo, L. Lin, and H. Chao. Semantics-driven portrait cartoon stylization. In *2010 IEEE International Conference on Image Processing*, pages 1805–1808. IEEE, 2010.

[21] R. Yeh, C. Chen, T. Y. Lim, M. Hasegawa-Johnson, and M. N. Do. Semantic image inpainting with perceptual and contextual losses. *arXiv preprint arXiv:1607.07539*, 2:3, 2016.