Literature Review

Guorui Lu

2018.8.15

1 History and Significance

Synthesizing realistic photo from sketches drawn by human has been a challenging and hot problem for a long time. The research of this technique can be traced back to 2009 [3]. However, the technique of image generation and image-to-image transfer did not rise up until 2014, when the Generative Adversarial Nets(GAN) [8] was proposed. Since then, variety of methods have been proposed to generate a more realistic photo for its broad application prospect. For example, police can use it to catch suspects through synthesised photos based on the drawn sketches, which requires that our photo should be as accurate as possible. As for the entertainment industry, we may need the model is able to generate images of multiple style, which is also a popular but difficult task.

2 Mainstream Method

I will introduce the main method of this project from 2015 to the present because the actual reasearch on sketch-photo synthesis did't qppear until 2014, which I think is largely due to the fact that GAN [8] was present in 2014. Before it, sketches were mainly used to do retrieval work [1, 2, 7, 12–14, 17, 23].

2.1 Generative Adversarial Nets(GAN)

Generative Adversarial Nets and its thoughts from game theory is the most popular and useful method in image generation. Here, I would clarify the GANs' development and show some recent work which can reflect the use of GAN in our project.

2.1.1 Generative Adversarial Nets

The original GAN brought us a completely new idea. It is corresponding a minimax two-player game, which is simple and efficient. But it also has an obvious shortcoming that the constraint is too week so that we can't control what it will generate.

2.1.2 cGAN and InfoGAN

Soon after the GAN's appearance, conditional GAN [19] was proposed, and then the InfoGAN [5]. These two models are able to partly control the outputs by adding extra codes. However, the pictures they synthesize are still blurry and low-resolution.

2.1.3 Pix2pix and Pix2pixHD

Pix2pix [16] and pix2pixHD [22] are two of the improved versions of the GANs. They took a big step forward to addressing the resolution issues. What they didn't resolve is that only when we have lots of paired images can we train the models.

2.1.4 CycleGAN

Inspired by NLP, Jun-Yan Zhu and Taesung Park et al. came up with a new model named CycleGan [26]. Although it has some drawbacks such as the high computing cost, CycleGAN still works very well for many problems. So I think I can apply it to my project.

2.1.5 *Rescent Work*

Here I would show you some rescent work of sketch-photo in 2017 and 2018.

- **SketchyGAN: Towards Diverse and Realistic Sketch to Image Synthesis** [4]:

  A new network structure proposed for generative task. It works better than retrieval model but the result is still blurry and low-resolution. Moreover, it is too faithful to the badly drawn pictures to keep the realism.

- **Facial Attributes Guided Deep Sketch-to-Photo Synthesis** [11]:

  This paper puts forward a method that add an auxiliary attribute discriminator to find the false attributes in the output of the generator. However, editing an attribute can cause unwanted structural edition of the image in some area, which is also the weakness of the previous models.

- **TextureGAN: Controlling Deep Image Synthesis with Texture Patches** [24]:

  Just as its name implies, TextureGAN adds a texture patch on a sketch at arbitrary locations and scales, which allows it to synthesize high-quality photos. And its novel losses for training deep image synthesis can encourage the generative network to handle new textures never seen on existing objects.

- **Scribbler: Controlling Deep Image Synthesis with Sketch and Color** [21]:

  In this research, Pastsorn Sangkloy et al. use sketched with sparse color "scribbles" descided by human to generate a higher resolution and more diverse images. However, if users specifies uncommon or even wrong colors and shapes, this model would deem them correct. In addition, sometimes the boundaries between object parts or regions of defferent colors are likely to become vague.

- **Image Generation from Sketch Constraint Using Contextual GAN** [18]:

  This is the paper I appreciate most for it proposed a creative idea that we can see the generation of image based on sketches as a inpainting process so that the model can avoid being too strictly corresponding to the input paintings.

2.2 Deep Convolutional Neural Networks($DNNs$)

Although GANs has been a remarkable success in image generation field, deep convolutional neural network is still also an effective tool. However, the use of this method is obviously less than GANs because it is difficult for a user to control what the network produce, which limits the application in image generation field.

- **Convolutional Sketch Inversion** [10]:

  Unlike GANs, this model can synthesize realistic photos from sketches without extra constraint. But the output is still low-resolution.

- **Attribute-controlled face photo synthesis from simple line drawing** [9]:

  This model can generate photo with desired attribute even though the simple line drawing is not complete. However, the style is random and a style photo is needed if we want to control the style of synthsized photo.

3 Other Related Work

Apart from the above work that relate to sketch-photo synthesis directly, there are some other researches which are associated with the project.

time	title	contribution	limitation
2016	Sketch me that shoe [20]	Solve the problem of fine-grained instance-level SBIR using free-hand sketches and contribute two new fine-grained SBIR datasets with extensive ground truth annotations	
2016	Colorful Image Colorization [25]	Introduce a novel framework for testing colorization algorithms, potentially applicable to other image synthesis tasks.	
2016	Let there be color!: joint end-to-end learning of global and local image priors for automatic image colorization with simultaneous classification [15]	The proposed architecture can process images of any resolution.	The main limitation lies in the fact that it is data-driven and thus will only be able to colorize images that share common properties with those in the training set. The model will use the color it have learned while training despite outher colors may also be suitable.
2017	Real-Time User-Guided Image Colorization with Learned Deep Priors [6]	It can colorize an image well with limitated time. And it can even generate unusual colorizations despite being trained on natural images.	It can be over-optimistic and produce undesired non-local effects

4 *Focus Problems and My Ideas*

While reading papers, I found the main controversies constrate on the flollowing aspects:

1. **Lack of sketches, especially the paired images and sketches**

   We can use CycleGAN to avoid this problem because CycleGAN doesn't need paired images.

2. **Uncontrollable results**

   It seems difficult for models to learn the latent information that controls the results directly, so the main idea of solving this problem is that use extra restricts, such as colors and textures, to specify what we want to get from the sketch, which requires the restricts given by people must be real and correct.

3. **Low-resolution**

   Just like the previous problems, adding constraints makes it easier for us to improve the resolution. Experiments have shown that using sketches with color specified can synthsize more realistic and higher resolution photos, so I suspect that we can apply another model which is used to colorize the picture first, and then use the colorized sketch to synthesize the photo.

4. **Badly drawn pictures**

   We can pose the image generation problem as an image completion problem by using sketch as a weak constextual constraint. By doing so, the generated image may exhibit different poses and shapes beyond the input sketches which may not strictly corresponding to photographic objects.

5. **Models can't fit the objects that don't appear during the training procedure**

   I haven't found any solution from papers above. Maybe zero-shot learning can be helpful.

5 Status and Future Direction

Sketch-to-photo synthesis has reached a quite realistic level, but some intractable problems still exit, which I have refered in *Focus Problems and My Ideas*. To my surprise, image colorization has also reached an amazing effect, which gives me the inspiration that we can combine the colorization and generation to get better results.

So this is my thoughts: we can select a model which can colorize the sketch, based on which we use the generative model to synthesize the realistic and high-resolution photo. But there is a final problem still remained: how can the model deal with the situation tha doesn't appear during the training procedure? I'll continue to study in depth.

References

[1] Yang Cao, Changhu Wang, Liqing Zhang, and Lei Zhang. Edgel index for large-scale sketch-based image search. In *Computer Vision and Pattern Recognition*, pages 761–768, 2011. 1

[2] Yang Cao, Hai Wang, Changhu Wang, Zhiwei Li, Liqing Zhang, and Lei Zhang. Mindfinder:interactive sketch-based image search on millions of images. In *International Conference on Multimedea 2010, Firenze, Italy, October*, pages 1605–1608, 2010. 1

[3] Tao Chen, Ming Ming Cheng, Ping Tan, Ariel Shamir, and Shi Min Hu. Sketch2photo. *Acm Transactions on Graphics*, 28(5), 2009. 1

[4] Wengling Chen and James Hays. Sketchygan: Towards diverse and realistic sketch to image synthesis. 2018. 1

[5] Xi Chen, Yan Duan, Rein Houthooft, John Schulman, Ilya Sutskever, and Pieter Abbeel. Infogan: Interpretable representation learning by information maximizing generative adversarial nets. 2016. 1

[6] Alexei A. Efros, Alexei A. Efros, Alexei A. Efros, Alexei A. Efros, Alexei A. Efros, Alexei A. Efros, and Alexei A. Efros. Real-time user-guided image colorization with learned deep priors. *Acm Transactions on Graphics*, 36(4):119, 2017. 2

[7] Mathias Eitz, Kristian Hildebrand, Tamy Boubekeur, and Marc Alexa. An evaluation of descriptors for large-scale image retrieval from sketched feature lines. *Computers Graphics*, 34(5):482–498, 2010. 1

[8] Ian J. Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. Generative adversarial nets. In *International Conference on Neural Information Processing Systems*, pages 2672–2680, 2014. 1

[9] Qi Guo, Ce Zhu, Zhiqiang Xia, Zhengtao Wang, and Yipeng Liu. Attribute-controlled face photo synthesis from simple line drawing. pages 2946–2950, 2017. 2

[10] Yamur Gltrk, Umut Gl, Rob Van Lier, and Marcel A. J. Van Gerven. Convolutional sketch inversion. 9913:810–824, 2016. 2

[11] A Dabouei S Soleymani NM Nasrabadi H Kazemi, M Iranmanesh. Facial attributes guided deep sketch-to-photo synthesis. *IEEE Winter Applications of Computer Vision Workshops (WACVW)*. 2

[12] Rui Hu, Mark Barnard, and John Collomosse. Gradient field descriptor for sketch based retrieval and localization. In *IEEE International Conference on Image Processing*, pages 1025–1028, 2010. 1

[13] Rui Hu and John Collomosse. A performance evaluation of gradient field hog descriptor for sketch based image retrieval . *Computer Vision Image Understanding Cviu*, 117(7):790–806, 2013. 1

[14] Rui Hu, Tinghuai Wang, and John Collomosse. A bag-of-regions approach to sketch-based image retrieval. In *IEEE International Conference on Image Processing*, pages 3661–3664, 2011. 1

[15] Satoshi Iizuka, Edgar Simoserra, and Hiroshi Ishikawa. Let there be color!: joint end-to-end learning of global and local image priors for automatic image colorization with simultaneous classification. *Acm Transactions on Graphics*, 35(4):1–11, 2016. 2

[16] Phillip Isola, Jun Yan Zhu, Tinghui Zhou, and Alexei A. Efros. Image-to-image translation with conditional adversarial networks. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 5967–5976, 2017. 1

[17] Yen Liang Lin, Cheng Yu Huang, Hao Jeng Wang, and Wei Chou Hsu. 3d sub-query expansion for improving sketch-based multi-view image retrieval. pages 3495–3502, 2014. 1

[18] Yongyi Lu, Shangzhe Wu, Yu Wing Tai, and Chi Keung Tang. Image generation from sketch constraint using contextual gan. 2017. 2

[19] Mehdi Mirza and Simon Osindero. Conditional generative adversarial nets. *Computer Science*, pages 2672–2680, 2014. 1

[20] Y.-Z. Song T. Xiang T. M. Hospedales C.-C. Loy Q. Yu, F. Liu. Sketch me that shoe. *The IEEE Conference on Computer Vision and Pattern Recognition(CVPR)*, june 2016. 2

[21] Patsorn Sangkloy, Jingwan Lu, Chen Fang, Fisher Yu, and James Hays. Scribbler: Controlling deep image synthesis with sketch and color. pages 6836–6845, 2016. 2

[22] Jun-Yan Zhu Andrew Tao Jan Kautz-Bryan Catanzaro NVIDIA Corporation UC Berkeley Ting-Chun Wang, Ming-Yu Liu. High-resolution image synthesis and semantic manipulation with conditional gans. In *Computer Vision and Pattern Recognition*, 2017. 1

[23] Changhu Wang, Zhiwei Li, and Lei Zhang. Mindfinder:image search by interactive sketching and tagging. In *International Conference on World Wide Web*, pages 1309–1312, 2010. 1

[24] Wenqi Xian, Patsorn Sangkloy, Varun Agrawal, Amit Raj, Jingwan Lu, Chen Fang, Fisher Yu, and James Hays. Texturegan: Controlling deep image synthesis with texture patches. 2017. 2

[25] Richard Zhang, Phillip Isola, and Alexei A. Efros. Colorful image colorization. *European Conference on Computer Vision(ECCV)*, pages 649–666, 2016. 2

[26] Jun Yan Zhu, Taesung Park, Phillip Isola, and Alexei A Efros. Unpaired image-to-image translation using cycle-consistent adversarial networks. pages 2242–2251, 2017. 1