Emerging Applications of Generative Adversarial Networks

Yu Xinyu
Electronic Information Science and Technology(Photoelectron), Physical Science and Technology, Southwest Jiaotong University, Chengdu, China

Abstract. With the idea of generative adversarial network (GAN) being brought up since 2014, a serial of research emerged in this field. While bring many exciting opportunities, these applications also raise the awareness of the risk of fake images which may cause huge damage. In this paper, we revisit the history of GAN’s development and summarize some important models and applications. Additionally, we discuss the worries which GAN would bring to people and possible solutions.

1. Introduction
Before the rise of deep learning [1], the computer vision field is dominant by shadow machine learning models, e.g., support vector machine and random forests, etc. With the backpropagation algorithm and the acceleration of computation by graphics processing unit (GPU), deep learning models become dominant again and achieve a better performance than human-beings on competitions including the famous ImageNet [2]. While having been successfully applied in many fields, e.g., face recognition [3], object detection [4], traffic forecasting [5], and stock prediction [6], deep learning has been stuck in a dilemma that deeper neural networks become harder to train and achieve little improvement.

To break the dilemma, the idea of generative adversarial network (GAN) was proposed by Ian GoodFellow in 2014 [7], who claimed GAN as a rich and hierarchical model of deep learning. He and his group trained two models: a generative model G which contains data distribution. A discriminative model called D which estimates the probability that a sample comes from the training data which based on the result that the model G given. The training procedure for the generative model would maximize the discriminative model D’s probability of making an error. It is a huge progress of deep learning models, because Markov chains or unrolled approximate inference networks is unnecessary in this way.

Afterwards, GAN has been improved and further studied by many researchers. In Mehdi Mirza and Simon Osindero’s work [8], they have discovered a solution to direct the data progress by conditioning the model on additional information (CGAN). The original GANs generate data from random noise. If we use more than one species of images to train this model, it would be confused and generate a vague image just like all of these training species mixed. CGAN solve this problem. Specifically, CGAN links a one-hot vector $y$ with random noise vector $z$, so that we could use the same GAN to produce more than one species of images.

In 2016, Alec Radford, Luke Metz and Soumith Chintala introduced a group of convolutional networks which called deep convolutional generation adversarial networks (DCGANs) [9] which is a strong candidate for unsupervised learning. Before this idea was brought up, GANs were limited to traditional fully connected neural network structures without convolutional operations. DCGANs used a method named transposed convolution, which could help us transform low resolution images into...
In 2018, the idea of BigGAN [10] achieved a huge progress of GAN on ImageNet. With dynamic GAN training, the scalability of images improved. The generated image is nearly true as captured in the real world. In BigGAN, the authors trained their models with two to four times as many parameters and eight times the original size compared to previous models. This solution obviously increased the resolution of an image and improved stability.

The progress of GAN is twofold. On one hand, the development of GANs has create many useful applications, such as ACGANs, which is known as a powerful instrument to produce an anime character which is determined by your own. On the other hand, the unrestrained use of GANs would also bring many troubles, someone may use this technique to make a fake image to misguide others.

In this paper, we introduce some important models in Section 2, and some important applications in Section 3. We point out the possible future directions and threatens in Section 4. We conclude this paper in Section 5.

2. Models
In this section, we introduce some important variants of GANs.

2.1 Wgan [11]
In Wasserstein GAN or WGAN, the researchers propose a new kind of function which could improve the old GAN minimax formula. Virtually we need to make a hypothesis of a noise to make the maximum likelihood approach work, but the researchers define a random variable $Z$ with a fixed distribution $p(z)$ and they pass $Z$ through a parametric function that directly generates samples following a certain distribution which called $P_{\theta}$. So that WGANs could solve main training problems of traditional GANs. Specifically, training WGANs do not need to find a balance between the discriminator and the generator, neither do the careful design of the network architecture.

2.2 Cogans [12]
Coupled generative adversarial networks (CoGANs) train two GAN in the same time. For example, there are two group in a competition, and these two groups are consisted by two members. They compose one image in two different field to blur the model. The discriminator needs to distinguish the image from two fields. The cooperation of a group depends on the weights they use. Due to the shared weights, a CoGAN needs fewer parameters, compared with two individual GANs.

2.3 Sagan [13]
Self-attention generative adversarial network (SAGAN) generates details by using cues from all feature locations, and it is an evolutorial progress. Traditional convolutional GANs could only visit the nearest likelihood approach work, but the researchers define a random variable $Z$ with a fixed distribution $p(z)$ and they pass $Z$ through a parametric function that directly generates samples following a certain distribution which called $P_{\theta}$. So that WGANs could solve main training problems of traditional GANs. Specifically, training WGANs do not need to find a balance between the discriminator and the generator, neither do the careful design of the network architecture.

2.4 CycleGAN [14]
Cycle-consistent adversarial networks (CycleGAN) generates a picture with both features from two different animals, by solving a problem named image-to-image translation. CycleGAN is not a new GAN architecture, but rather a clear way of applying GANs. CycleGAN translates unpaired images and learns the mapping from one domain to another domain. For example, CycleGAN can extract the style from one specific image and translate it into another image.

2.5 Progan [15]
Progressive growing of GANs (ProGAN) tries to strengthen the stability during the training process of GANs. The instability of training process is caused by the influence between the generator and discriminator. Sometimes, this instability would cause the generation of very scary images. ProGAN is
a technique of training GNAs by gradually increasing the image resolution. At first, ProGAN trains a 4x4 generator and a 4x4 discriminator. The generated 4x4 images are further mapped into 16x16, 32x32, until 1024x1024. This procedure helps to mitigate the training instability problem.

2.6 Stylegan [16]
Style-based generator (StyleGAN) doesn’t focus on creating the real image, but improves the GAN’s fine control ability of the generated image. To achieve this goal, StyleGAN uses the state-of-art techniques including mapping network, style mixing, stochastic variation, etc.

2.7 Lapgan [17]
Laplacian pyramid of adversarial networks (LAPGAN) uses plenty of convolutional networks with a Laplacian framework in order to generate images in a coarse-to-fine fashion. The uniqueness of the generator is that it combines the image from a higher-level network with the noise to generate a lower-level image, which is equivalent to a conditional generative adversarial network. The cascade structure lowers the difficulty of the learning content for each GAN, thus improving the overall learning ability.

2.8 Aae [18]
Adversarial autoencoder (AAE) has two objectives, one is traditional reconstruction mistake criterion, the other is adversarial training criterion. It suits the aggregated posterior distribution of the potential representation of the automatic encoder with any prior distribution. A standard autoencoder is firstly used to reconstruct the image from the potential code and a next network is trained to predict if the sample is from the hidden code of the autoencoder or a user-specified sample distribution.

2.9 Infogan [19]
Information maximizing generative adversarial nets (InfoGAN) tries to get interpretable feature representations via unsupervised learning, using GAN and the mutual information between the generated image and input encoding. The benefits of this approach is the avoid of using supervised learning and the massive computing to get the interpretable features.

3. Applications
In this section, we introduce some important applications of GANs.

3.1 Image-to-Image Translation [20]
These networks are creative. They do not only learn the mapping from the input picture resource to the output picture resource. Additionally, they learn a waste function to train the mapping procedure, which allows them to apply the same generic approach to the problem which is hard to solve in traditional way. Significantly, this approach is helpful and useful at synthesizing photos from label maps, reconstructing vague objects from edge maps, and colorizing images. Many twitter users have posted their products using this system.

3.2 Image Super-Resolution [21]
A generative adversarial network can be used for image super-resolution (SR). It is the first tool able to infer photo-realistic natural images for 4×upscaling factors framework, and the factors recover the finer texture details. The adversarial loss in a perceptual loss function pushes their result to the original image manifold by using a discriminator network which is trained to distinguish the super-resolved images and natural realistic photos.

3.3 Text to Image [22]
In this work, the researchers raise a novel deep architecture and generative adversarial networks formulation to strongly bridge an advance in text and image modeling, translating visual ideas from
characters to simple pixels. Based on the neural network architecture which is developed to learn discriminative text feature representations, this work can generate plausible images from detailed text descriptions.

3.4 Image Editing [23, 24]
Bad weather conditions, for example, heavy rain or blizzard would affect the visual quality when taking images. In this work, the researchers attempt to adjust generative modeling capabilities of conditional generative adversarial networks by adding an additional constraint which could remove the raindrop that affect the image quality [23]. And the de-rained image could be indistinguishable from its corresponding ground truth original clean image.

Another group of researchers have developed an invertible conditional GANSs for image editing [24]. They evaluate encoders to inverse the mapping of a conditional generative adversarial networks to re-generate real images with deterministic complex modifications.

3.5 Story Visualization [25]
StoryGAN is proposed to conduct the task of translating the input story to images, which could be used to explain the story to readers. This model is developed from the conditional generative adversarial network. The image generation for each sentence is enriched with contextual information from the Context Encoder. Two discriminators at different levels guide the generation process.

3.6 Architecture Design [26]
ArchigAN is an application of pix2pix in the task of architecture design. This task is divided into the workflow: footprint massing, program repartition and furniture layout, with three models separately. ArchigAN learns the topology and space room from the architecture design pictures.

3.7 Music Synthesis [27]
GANsynth is the first successful application in the synthesized, high-quality musical audio using GANS with fast generation. Compared with Wavenet, GANsynth has a smoother interpolations and a faster speed by 50,000 times. Instead of generating the music with the series sequentially, GANsynth generate the whole time series in parallel, which contributes its fast speed.

4. Future Directions
As we know, generative adversarial networks are popular nowadays, and we have already developed many important applications through this progressive invention. In this section, we introduce the possible future directions of GANs and how to prevent the threatens brought by this technique.

4.1 Possible Applications
In the future, we could make more progress on restoring old photos which may help people remember their ancestors, so that people could get a clearer photo of their old family members. Through this application, we could also make a good research condition of archeology, because there are many photos which are not clear enough to figure out some important information.

Other than the applications, there are many possible theoretical research problems. GANs require plenty of input images for training. How to design a few-shot GAN is a possible research direction. Another direction is to solve the generator collapsing problem that bothers the generator.

4.2 Possible Threatens
GANs have the possibility of being used for illegal or evil purposes. The fingerprints can be counterfeited by GANs and used to unlock the mobile devices and access control systems, which brings a high risk of privacy leakage or burglary [28]. By replacing the faces in a video with those of famous people, e.g., DeepFake, the faked videos would not only bring troubles to the star who is fooled, but also bring the possible illegal business by selling these videos [29]. With automatic audio
generation, even the official speech including the UN’s announcement could be faked. This kind of speech could be used to misguide the government, the companies, and the individuals to take wrong actions [30].

It is still a difficult task for humans to tell the distinguishable points between the original and the after-generated images. And it is becoming more and more difficult with the development of GANs. Some researchers propose to use a GAN to fight with another GAN. Google recently release a public dataset with videos generated by DeepFake and encourage the researchers to propose AI models to distinguish the real and faked ones.

5. Conclusion

In this paper, we introduce the state-of-the-art progress of the models and applications of generative adversarial networks. As a revolutionary idea, GANs have been widely researched and applied in different research problems. But they are still facing many challenges and require future investigations. Our discussions in this paper would inspire the future research directions.

References

[1] LeCun Y, Bengio Y, Hinton G. Deep learning[J]. nature, 2015, 521(7553): 436.
[2] Deng J, Dong W, Socher R, et al. Imagenet: A large-scale hierarchical image database[C]//2009 IEEE conference on computer vision and pattern recognition. Ieee, 2009: 248-255.
[3] Sun Y, Chen Y, Wang X, et al. Deep learning face representation by joint identification-verification[C]//Advances in neural information processing systems. 2014: 1988-1996.
[4] Liu W, Anguelov D, Erhan D, et al. Ssd: Single shot multibox detector[C]//European conference on computer vision. Springer, Cham, 2016: 21-37.
[5] Jiang W, Zhang L. Geospatial data to images: A deep-learning framework for traffic forecasting[J]. Tsinghua Science and Technology, 2018, 24(1): 52-64.
[6] Zhou X, Pan Z, Hu G, et al. Stock market prediction on high-frequency data using generative adversarial nets[J]. Mathematical Problems in Engineering, 2018, 2018.
[7] Goodfellow I, Pouget-Abadie J, Mirza M, et al. Generative adversarial nets[C]//Advances in neural information processing systems. 2014: 2672-2680.
[8] Mirza M, Osindero S. Conditional generative adversarial nets[J]. arXiv preprint arXiv:1411.1784, 2014.
[9] Radford A, Metz L, Chintala S. Unsupervised representation learning with deep convolutional generative adversarial networks[J]. arXiv preprint arXiv:1511.06434, 2015.
[10] Brock A, Donahue J, Simonyan K. Large scale gan training for high fidelity natural image synthesis[J]. arXiv preprint arXiv:1809.11096, 2018.
[11] Arjovsky M, Chintala S, Bottou L. Wasserstein gan[J]. arXiv preprint arXiv:1701.07875, 2017.
[12] Liu M Y, Tuzel O. Coupled generative adversarial networks[C]//Advances in neural information processing systems. 2016: 469-477.
[13] Zhang H, Goodfellow I, Metaxas D, et al. Self-attention generative adversarial networks[J]. arXiv preprint arXiv:1805.08318, 2018.
[14] Zhu J Y, Park T, Isola P, et al. Unpaired image-to-image translation using cycle-consistent adversarial networks[C]//Proceedings of the IEEE international conference on computer vision. 2017: 2223-2232.
[15] Karras T, Aila T, Laine S, et al. Progressive growing of gans for improved quality, stability, and variation[J]. arXiv preprint arXiv:1710.10196, 2017.
[16] Karras T, Laine S, Aila T. A style-based generator architecture for generative adversarial networks[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2019: 4401-4410.
[17] Denton E L, Chintala S, Fergus R. Deep generative image models using adaptive laplacian pyramid of adversarial networks[C]//Advances in neural information processing systems. 2015:
1486-1494.

[18] Makhzani A, Shlens J, Jaitly N, et al. Adversarial autoencoders[J]. arXiv preprint arXiv:1511.05644, 2015.

[19] Chen X, Duan Y, Houthooft R, et al. Infogan: Interpretable representation learning by information maximizing generative adversarial nets[C]//Advances in neural information processing systems. 2016: 2172-2180.

[20] Isola P, Zhu J Y, Zhou T, et al. Image-to-image translation with conditional adversarial networks[C]//Proceedings of the IEEE conference on computer vision and pattern recognition, 2017: 1125-1134.

[21] Ledig C, Theis L, Huszár F, et al. Photo-realistic single image super-resolution using a generative adversarial network[C]//Proceedings of the IEEE conference on computer vision and pattern recognition, 2017: 4681-4690.

[22] Reed S, Akata Z, Yan X, et al. Generative adversarial text to image synthesis[J]. arXiv preprint arXiv:1605.05396, 2016.

[23] Zhang H, Sindagi V, Patel V M. Image de-raining using a conditional generative adversarial network[J]. IEEE Transactions on Circuits and Systems for Video Technology, 2019.

[24] Perarnau G, Van De Weijer J, Raducanu B, et al. Invertible conditional gans for image editing[J]. arXiv preprint arXiv:1611.06355, 2016.

[25] Li Y, Gan Z, Shen Y, et al. StoryGAN: A Sequential Conditional GAN for Story Visualization[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2019: 6329-6338.

[26] Stanislas C. ArchiGAN: A Generative Stack for Apartment Building Design[J]. 2019.

[27] Engel J, Agrawal K K, Chen S, et al. Gansynth: Adversarial neural audio synthesis[J]. arXiv preprint arXiv:1902.08710, 2019.

[28] Bontrager P, Roy A, Togelius J, et al. DeepMasterPrints: Generating masterprints for dictionary attacks via latent variable evolution[C]//2018 IEEE 9th International Conference on Biometrics Theory, Applications and Systems (BTAS). IEEE, 2018: 1-9.

[29] Karras T, Aila T, Laine S, et al. Progressive growing of gans for improved quality, stability, and variation[J]. arXiv preprint arXiv:1710.10196, 2017.

[30] Merity S, Keskar N S, Socher R. Regularizing and optimizing LSTM language models[J]. arXiv preprint arXiv:1708.02182, 2017.