Critical Analysis

Generative Adversarial Nets (Goodfellow et al NIPS'14)

Main Trends [1]:

- Generative algorithms have wide practical applications.
- In the adversarial idea, the recent event that AlphaGo defeats world's top human player engages public interest in artificial intelligence. The intermediate version of AlphaGo utilizes two networks competing against each other.
- Adversarial examples have the adversarial idea, too. Adversarial examples are those examples which are very different from the real examples, but are classified into a real category very confidently, or those that are slightly different than the real examples, but are classified into a wrong category. This is a very hot research topic recently. To be against adversarial attacks, references, utilize GANs to conduct the right defense.
- Adversarial machine learning is a minimax problem. The defender, who builds the classifier that we want to work correctly, is searching over the parameter space to find the parameters that reduce the cost of the classifier as much as possible. Simultaneously, the attacker is searching over the inputs of the model to maximize the cost.
- The adversarial idea exists in adversarial networks, adversarial learning, and adversarial examples. However, they have different objectives.

There are a large number of papers related to GANs according to Google scholar. For example, there are about 11,800 papers related to GANs in 2018. That is to say, there are about 32 papers everyday and more than one paper every hour related to GANs in 2018.

Key Ideas since the publication:

- [3] cGAN: Conditional Generative Adversarial Network. Use of information in addition to the image as input both to the generator and the discriminator models.
- [4] InfoGAN: Information Maximizing Generative Adversarial Network. Attempts to structure the input or latent space for the generator. Add specific semantic meaning to the variables in the latent space.
- [5] AC-GAN: Auxiliary Classifier Generative Adversarial Network. Changes the generator to be class conditional as with the cGAN, and adds an additional or auxiliary model to the discriminator that is trained to reconstruct the class label. The discriminator both predicts the likelihood of the image given the class label and the class label given the image.
- [6] StackGAN: Stacked Generative Adversarial Network. Generate images from text using a hierarchical stack of conditional GAN models. The proposed method achieves significant improvements on generating photo-realistic images conditioned on text descriptions.
- [9] WGAN: Wasserstein Generative Adversarial Network. Changes the training procedure to update the discriminator model (called a critic) many more times than the generator model for each iteration. The critic and generator models are both trained using "*Wasserstein loss*," designed to provide linear gradients that are useful for updating the model.

- [10] CycleGAN: Cycle-Consistent Generative Adversarial Network. image-to-image translation without paired image data. Their approach seeks "*cycle consistency*" such that image translation from one domain to another is reversible.
- [11] Progressive GAN: Progressive Growing Generative Adversarial Network. Involves progressively increasing the model depth during the training process keeping the generator and discriminator symmetric in-depth during training and adding layers stepwise.
- [12] StyleGAN: Style-Based Generative Adversarial Network. Extension of the generator that allows the latent code to be used as input at different points of the model to control features of the generated image. Leads to demonstrably better interpolation properties, and also better disentangles the latent factors of variation.
- [13] BigGAN: Big Generative Adversarial Network. High-quality output images can be created by scaling up existing class-conditional GAN models. A "*truncation trick*" is used where points are sampled from a truncated Gaussian latent space at generation time that is different from the untruncated distribution at training time.

Main Problems:

- [1] Evaluation metric: How to select a good evaluation metric for GANs is still a hard problem. An appropriate evaluation metric ought to differentiate true samples from fake ones, verify mode drop, mode collapse, and detect overfitting.
- [1] Mode collapse: distributions learned by GANs suffer from mode collapse. Each iteration of generator over-optimizes for a particular discriminator, and the discriminator never manages to learn its way out of the trap. As a result, the generators rotate through a small set of output types. This form of GAN failure is called mode collapse.
- [1] Divergence/Distance: Training of GAN may not have good generalization properties, e.g., training may look successful, but the generated distribution may be far from real data distribution in standard metrics.
- [1] Inverse mapping: GANs cannot learn the inverse mapping - projecting data back into the latent space.
- [1] Memorization: As for "memorization of GANs", making the generator "learn to memorize" the training data is a more difficult task than making it "learn to output realistic but unseen data".
- [16] Vanishing Gradients: if your discriminator is too good, then generator training can fail due to vanishing gradients. In effect, an optimal discriminator doesn't provide enough information for the generator to make progress.
- [16] Failure to Converge: GANs frequently fail to converge. Researchers have tried to use various forms of regularization to improve GAN convergence.

Remaining Problems:

- [1] New Divergences: This deserves further study.
- [1] Estimation uncertainty: Generally speaking, as we have more data, uncertainty estimation reduces. GANs do not give the distribution that generated the training examples and GANs aim to generate new samples that come from the same distribution of the training examples. Therefore, GANs have neither a likelihood nor a well-defined posterior.

- [1] Generalization: Useful theory should enable choice of model class, capacity, and architectures. This is an interesting issue to be investigated in future work.

Most interesting unsolved problem:

- [1] GANs for discrete data: GANs rely on the generated samples being completely differentiable with respect to the generative parameters. Therefore, GANs cannot produce discrete data directly, such as hashing code and one-hot word. Solving this problem is very important since it could unlock the potential of GANs for NLP and hashing. There are other methods towards this research direction. More work needs to be done in this interesting area.

References:

[1] Jie Gui, Zhenan Sun, Yonggang Wen, Dacheng Tao, Jieping Ye, "A Review on Generative Adversarial Networks: Algorithms, Theory, and Applications" [Online]. Available: https://arxiv.org/abs/2001.06937 (2020)

[2] Alec Radford, Luke Metz, Soumith Chintala, "Unsupervised Representation Learning with Deep Convolutional Generative Adversarial Networks" [Online]. Available: https://arxiv.org/abs/1511.06434 (2016)

[3] Mehdi Mirza, Simon Osindero, "Conditional Generative Adversarial Nets" [Online]. Available: https://arxiv.org/abs/1411.1784 (2014)

[4] Xi Chen, Yan Duan, Rein Houthooft, John Schulman, Ilya Sutskever, Pieter Abbeel, "InfoGAN: Interpretable Representation Learning by Information Maximizing Generative Adversarial Nets" [Online]. Available: https://arxiv.org/abs/1606.03657 (2016)

[5] Augustus Odena, Christopher Olah, Jonathon Shlens, "Conditional Image Synthesis With Auxiliary Classifier GANs" [Online]. Available: https://arxiv.org/abs/1610.09585 (2017)

[6] Han Zhang, Tao Xu, Hongsheng Li, Shaoting Zhang, Xiaogang Wang, Xiaolei Huang, Dimitris Metaxas, "StackGAN: Text to Photo-realistic Image Synthesis with Stacked Generative Adversarial Networks" [Online]. Available: https://arxiv.org/abs/1612.03242 (2017)

[8] Phillip Isola, Jun-Yan Zhu, Tinghui Zhou, Alexei A. Efros, "Image-to-Image Translation with Conditional Adversarial Networks" [Online]. Available: https://arxiv.org/abs/1611.07004 (2018)

[8] Martin Arjovsky, Soumith Chintala, Léon Bottou, "Wasserstein GAN" [Online]. Available: https://arxiv.org/abs/1701.07875 (2017)

[10] Jun-Yan Zhu, Taesung Park, Phillip Isola, Alexei A. Efros. "Unpaired Image-to-Image Translation using Cycle-Consistent Adversarial Networks" [Online]. Available: https://arxiv.org/abs/1703.10593 (2020)

[11] Tero Karras, Timo Aila, Samuli Laine, Jaakko Lehtinen, "Progressive Growing of GANs for Improved Quality, Stability, and Variation" [Online]. Available: https://arxiv.org/abs/1710.10196 (2018)

[12] Tero Karras, Samuli Laine, Timo Aila, "A Style-Based Generator Architecture for Generative Adversarial Networks" [Online]. Available: https://arxiv.org/abs/1812.04948 (2019)

[13] Andrew Brock, Jeff Donahue, Karen Simonyan, "Large Scale GAN Training for High Fidelity Natural Image Synthesis" [Online]. Available: https://arxiv.org/abs/1809.11096 (2019)

[14] Yongjun Hong, Uiwon Hwang, Jaeyoon Yoo, Sungroh Yoon, "How Generative Adversarial Networks and Their Variants Work: An Overview" [Online]. Available: https://arxiv.org/abs/1711.05914 (2019)

[15] Martin Arjovsky, Soumith Chintala, Léon Bottou, "Wasserstein GAN" [Online] Available. https://arxiv.org/abs/1701.07875 (2017)

[16] Martin Arjovsky, Léon Bottou , Towards Principled Methods for Training Generative Adversarial Networks [Online]. Available. https://arxiv.org/abs/1701.04862 (2017)