Unsupervised learning by GAN

Animesh Devendra Chourey – 210765551

Queen Mary University of London – March 2022

Generative Adversarial Networks

Generative Adversarial Networks (GANs) [6] is a framework proposed for estimating the generative models via an adversarial process. Discriminative models have had striking success in deep learning so far especially whenever mapping a high-dimensional, rich sensory input to a label is needed[9]. However, the same cannot be said for deep generative models because of difficulties with intractable probabilistic computations which arise in maximum likelihood estimation and also due to difficulty of leveraging the benefits of piecewise linear units in the generative context. In the proposed framework, a new generative model estimation procedure sidesteps these above difficulties, as here the generative model is pitted against an adversary i.e. an discriminative model. Adversarial Nets has generative model generating samples by passing random noise through a multilayered perceptron and a discrminative model which is also a multilayered perceptron. Both these models can be trained using only highly successful backpropogation and dropout algorithms[10] and sampling from the generative models by using only forward propagation. No approximate inference or Markov chains are necessary. As adversarial networks do not require feedback loops during generation, they are able to leverage piecewise linear units [11, 5, 7] better, because of which the performance of backpropogation is improved. However, adversarial nets do face problems with unbounded activation when used in a feedback loop.

In these two networks, one network takes noise as input and generates the data samples called the generator(G) network, and the other network called discriminator (D) receives data from both the generator and the training data, which distinguishes between the two data coming from these two sources. G is associated with a prior on input noise variables p_z(z) from which G draws sample and z, and creates fake samples G(z;θ_g). D(x; θ_d) outputs a scalar value representing the probability that x came from real data rather than from the genertor's distribution(p_g). D is trained in such a way that it maximizes the probability of assigning correct label to both training examples and samples from G. Simultaneously G is trained such that it minimizes log(1 - D(G(z))) i.e. G is able to fool the discriminator into thinking data is coming from the real source and not the fake one.

Main trends on the topic since the publication of the paper:-

Ever since it's introductions, GANs have been implemented in various fields. Their most prominent and successful use has been in the field of computer vision. It's application include image generation, image super resolution, image translation, video generation and many more. Carl Vondrick, et al [18] uses GANs for video prediction specifically predicting up to a second of video frames with success.In other domains, GANs have been extended to language processing, audio and music processing. text-2-image [16] made significant progress in generating meaningful images based on explicit textual description. Not all GANs produce images. They are also being used to produce synthesized speech from text input[20]. SRGAN[14] generates output images with higher pixel resolution. This has given healthcare industry much needed boost, as there is a strong need to decrease the effect of radiation on patients. Poor quality pictures can be enhanced with minimal risks on patients. In entertainment industry also GANs[6] are being implemented. Yanghua Jin et al [12] demonstrates the training and use of a GANs[6] for generating faces of anime characters.

Main problems solved or improvements over the original work:-

Various methods have been proposed to overcome the stability issues related to GANs.

Deep Convolutional Generative Adversarial Networks (DCGAN) [15] were the first ever model introduced having major improvement over the original GANs[6] architecture for image generation. *ReLU* activation function is used in all the layers of generator except the in the output layer which uses *tanh* function. In the discriminator *LeakyReLU* is used.There are no fully connected layers and all the max pooling is replaced with convolutional stride. Also, batch normalization is used in except in the input layer of the discriminator and the output layer of the generator. These modifications to the original GANs[6] result in success of DCGANs.

Wasserstein Generative Adversarial Networks (WGAN) [1] is an extension to the GANs[6] that both improves the stability when training the model and also provides a loss function that correlates to the quality of the images generated. Instead of using discriminator that predicts the probability of generated images being fake or real, WGAN[1] changes discriminator model with a critic who scores the magnitude of the image being real or fake. This change prevents *mode collapse* without us worrying about vanishing gradients.

Wasserstein Generative Adversarial Networks with Gradient Penalty (WGAN-GP) [8] slightly tweaks the architecture of WGANs[1] as WGANs[1] generates low quality samples and also often fails to converge due to the use of weight clipping which arises due to Lipschitz constraint in the discriminator. WGAN-GP[8] uses gradient penalty instead of the weight clipping to enforce the Lipschitz constraint. WGAN-GP[8] stables the training process and improves the training speed when compared to WGANs[1].

Remaining problems from the published works so far:-

None of the problems mentioned below have been completely resolved, and they are still in the area of active research.

One of the main question regarding GANs has been how to scale them beyond image synthesis. SeqGAN[21] and MaskGAN [4] have been proposed in the text domain but the discrete nature of text has made it difficult to apply GANs on them. There still exists areas for development in NLP (Natural Language Processing) and IR (Information Retrieval). Audio is the domain in which GANs are closest in achieving similar success they've had with images. WaveGAN[3] is the first serious attempt at applying GANs to unsupervised audio synthesis. So the question, *"does scaling GANs to other domains require new training techniques or does it simply require better implicit priors for each domain"* will require hard thinking about what makes sense and whether it is also computationally feasible in a given domain. These question may require fundamental research progress.

Another issue is does training of GANs[6] scale with batch size. There has been some evidence that increasing minibatch size does improves the quantitative results and reduces training time [2]. If this phenomenon is robust then it would suggest that gradient noise is a dominating factor. However, it hasn't been systematically studied and the question remains open. The instability in the training process is still a challenge that needs to be addressed. GANs need to reach Nash equilibrium during the training, but is proved difficult to achieve [1].

An unsolved problem on the topic most interesting to you to solve and why:-

When it comes to evaluating GANs, there are many proposals but a little consensus. MS-SSIM[19], Precision and Recall assessing generative models[17], Inception Score and FID, Geometery Score[13] are just a smal fraction of the proposed GAN evaluation schemes. Even though Inception score and FID are relatively popular, from what it seems evaluation metric for GAN is clearly an unsettled issue. Therefore I believe getting a novel solution that unifies all the marking and evaluation criteria approved by all would be the most promising issue to solve which will definitely guide all the research in *Generative Adversarial Networks* to even greater heights.

References

[1] Martin Arjovsky, Soumith Chintala, and Léon Bottou. Wasserstein generative adversarial networks. In *International conference on machine learning*, pages 214–223. PMLR, 2017.

[2] Andrew Brock, Jeff Donahue, and Karen Simonyan. Large scale gan training for high fidelity natural image synthesis. *arXiv preprint arXiv:1809.11096*, 2018.

[3] Chris Donahue, Julian McAuley, and Miller Puckette. Adversarial audio synthesis. *arXiv preprint arXiv:1802.04208*, 2018.

[4] William Fedus, Ian Goodfellow, and Andrew M Dai. Maskgan: better text generation via filling in the_. *arXiv preprint arXiv:1801.07736*, 2018.

[5] Xavier Glorot, Antoine Bordes, and Yoshua Bengio. Deep sparse rectifier neural networks. In *Proceedings of the fourteenth international conference on artificial intelligence and statistics*, pages 315–323. JMLR Workshop and Conference Proceedings, 2011.

[6] Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. Generative adversarial nets. *Advances in neural information processing systems*, 27, 2014.

[7] Ian Goodfellow, David Warde-Farley, Mehdi Mirza, Aaron Courville, and Yoshua Bengio. Maxout networks. In *International conference on machine learning*, pages 1319–1327. PMLR, 2013.

[8] Ishaan Gulrajani, Faruk Ahmed, Martin Arjovsky, Vincent Dumoulin, and Aaron C Courville. Improved training of wasserstein gans. *Advances in neural information processing systems*, 30, 2017.

[9] Geoffrey Hinton, Li Deng, Dong Yu, George E Dahl, Abdel-rahman Mohamed, Navdeep Jaitly, Andrew Senior, Vincent Vanhoucke, Patrick Nguyen, Tara N Sainath, et al. Deep neural networks for acoustic modeling in speech recognition: The shared views of four research groups. *IEEE Signal processing magazine*, 29(6):82–97, 2012.

[10] Geoffrey E Hinton, Nitish Srivastava, Alex Krizhevsky, Ilya Sutskever, and Ruslan R Salakhutdinov. Improving neural networks by preventing co-adaptation of feature detectors. *arXiv preprint arXiv:1207.0580*, 2012.

[11] Kevin Jarrett, Koray Kavukcuoglu, Marc'Aurelio Ranzato, and Yann LeCun. What is the best multi-stage architecture for object recognition? In *2009 IEEE 12th international conference on computer vision*, pages 2146–2153. IEEE, 2009.

[12] Yanghua Jin, Jiakai Zhang, Minjun Li, Yingtao Tian, Huachun Zhu, and Zhihao Fang. Towards the automatic anime characters creation with generative adversarial networks. *arXiv preprint arXiv:1708.05509*, 2017.

[13] Valentin Khrulkov and Ivan Oseledets. Geometry score: A method for comparing generative adversarial networks. In *International Conference on Machine Learning*, pages 2621–2629. PMLR, 2018.

[14] Christian Ledig, Lucas Theis, Ferenc Huszár, Jose Caballero, Andrew Cunningham, Alejandro Acosta, Andrew Aitken, Alykhan Tejani, Johannes Totz, Zehan 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.

[15] Alec Radford, Luke Metz, and Soumith Chintala. Unsupervised representation learning with deep convolutional generative adversarial networks. *arXiv preprint arXiv:1511.06434*, 2015.

[16] Scott Reed, Zeynep Akata, Xinchen Yan, Lajanugen Logeswaran, Bernt Schiele, and Honglak Lee. Generative adversarial text to image synthesis. In *International conference on machine learning*, pages 1060–1069. PMLR, 2016.

[17] Mehdi SM Sajjadi, Olivier Bachem, Mario Lucic, Olivier Bousquet, and Sylvain Gelly. Assessing generative models via precision and recall. *Advances in Neural Information Processing Systems*, 31, 2018.

[18] Carl Vondrick, Hamed Pirsiavash, and Antonio Torralba. Generating videos with scene dynamics. *Advances in neural information processing systems*, 29, 2016.

[19] Zhou Wang, Eero P Simoncelli, and Alan C Bovik. Multiscale structural similarity for image quality assessment. In *The Thrity-Seventh Asilomar Conference on Signals, Systems & Computers, 2003*, volume 2, pages 1398–1402. Ieee, 2003.

[20] Shan Yang, Lei Xie, Xiao Chen, Xiaoyan Lou, Xuan Zhu, Dongyan Huang, and Haizhou Li. Statistical parametric speech synthesis using generative adversarial networks under a multi-task learning framework. In *2017 IEEE Automatic Speech Recognition and Understanding Workshop (ASRU)*, pages 685–691. IEEE, 2017.

[21] Lantao Yu, Weinan Zhang, Jun Wang, and Yong Yu. Seqgan: Sequence generative adversarial nets with policy gradient. In *Proceedings of the AAAI conference on artificial intelligence*, volume 31, 2017.