17 March 2022

ECS795P DEEP LEARNING AND COMPUTER VISION, 2022

COURSE WORK 2: GENERATIVE ADVERSARIAL NETS

1. **What are the main trends on the topic since the publication of the paper discussed?**

   One of the primary goals of GANs is to produce high-quality realistic images. Variants have been proposed to overcome the vanilla GAN architecture's limited capacity. Few of the different types of architectures are as mentioned below:

   i. **DC GAN** – It is a GAN with deep convolutions. It is one of the most widely used, powerful, and successful GAN architectures. It is implemented using ConvNets instead of a Multi-layered Perceptron. ConvNets employ a convolutional stride and are constructed without the use of max pooling, and the layers in this network are not completely connected.[3]

   ii. **Conditional GAN (CGAN)** – Conditional GAN is a deep learning neural network with some extra parameters. Labels are also added to Discriminator inputs to assist the discriminator in correctly classifying inputs that are not easily filled by the generator.[3]

   iii. **SRGAN** – A Super Resolution GAN (SRGAN) is one such ML method capable of upscale images to extremely high resolutions. An SRGAN learns how to generate upscaled images by combining the adversarial nature of GANs with deep neural networks (up to four times the resolution of the original). The resulting super resolution images are more accurate and have high mean opinion scores (MOS).[4]

   iv. **Cycle GAN** – It was released in 2017 and performs Image Translation. Assume we trained it on a dataset of horse images and can translate it into zebra images. [3]

   v. **Info GAN** – A more advanced version of GAN that can learn to disentangle representations using an unsupervised learning approach. [3]

   GANs, and more specifically their discriminators and generators, can be architected in a variety of ways to solve a wide range of image processing problems.

2. **What are the key ideas of related published works since the original publication?**

   i. **Generative Model:** A generative model is a statistical model of the joint probability distribution P(X,Y) on given observable variable X and target variable Y. The term "generative model" can also refer to models that generate instances of output variables in ways that have no obvious relationship to probability distributions over potential samples of input variables. This class of generative models includes generative adversarial networks, which are judged primarily by the similarity of specific outputs to potential inputs. These are not classifiers.

   ii. **Discriminative Model:** A discriminative model is a model of the conditional probability P(Y|X=x) of the target Y, given an observation x. In contrast to generative modelling, which investigates the joint probability P(x,y), discriminative modelling investigates the P(y|x) or maps the given unobserved variable (target) 'x' to a class label 'y' dependent on the observed variables (training samples). In object recognition, for example, 'x' is most likely a vector of raw pixels (or features extracted from the raw pixels of the image). This is accomplished within a probabilistic framework by modelling the conditional probability distribution P(y|x), which can be used to predict 'y' from 'x'.

3. **What are the main problems solved or improvements over the original work?**

   i. The original paper mentions several issues of the ***Vanilla variant of GANs***, the most important of which was that it was not State of the Art at the time and did not outperform most models available at the time. Because of the architecture's limited capacity, the original GAN was only applied to MNIST, the

Toronto face dataset, and CIFAR-10This problem has been helped by the deconvolution and up-sampling processes.

ii. Another issue that has been raised in relation to GANs is convergence during training. As the generator improves with training, the discriminator's performance deteriorates because the discriminator cannot easily distinguish between real and fake samples. If the generator succeeds, the discriminator will eventually have a 50% accuracy. This progression is problematic because the discriminator feedback becomes less meaningful over time, and the generator relies on this random feedback, resulting in poor quality output. There are two ways to solve this: add random noise to the discriminator input and use regularisation for the discriminator's network weights.

iii. In another case, if the discriminator is overly good, generator training may fail due to vanishing gradients because the discriminator is unable to provide enough information for the generator to progress. To deal with vanishing gradients, the introductory paper on GANs[1] proposed a modification to the minimax loss. Also proposed is the Wasserstein loss[2], but these loss functions vary depending on the architecture, making it an architecture-specific loss that cannot be generalised to other architectures.

4. What are the remaining problems from the published works so far?

i. One of the goals of training GANs is to increase the diversity of generated images. When the generator finds one or a few samples regardless of the input, and the discriminator cannot tell the difference between a real input and the generator's output, the generator may learn to produce only that output.

ii. This appears to be an ideal training progression, but it may result in a failure known as Mode collapse. This problem can be alleviated somewhat by using a more diverse set of samples or by employing the Wasserstein loss, which prevents the discriminator from becoming stuck in a local optima.

5. What is an unsolved problem on the topic most interesting to you to solve and why?

i. GAN evaluation is primarily divided into two types: qualitative, which refers to the visual quality of generated images from a human perspective, and quantitative, which evaluates the model based on criteria such as overfitting, low diversity of generated samples, and mode dropping for GANs.

ii. Despite the existence of a plethora of metrics, it remains difficult to compare the performance of models and evaluate the quality of images generated by GAN on par with human judgement. The performance of GANs that are conditioned on labels, such as CGAN and its successors, is dependent on a well-labeled dataset, which may pose difficulties in some real-world applications.

iii. There are variations of GANs that cater to specific tasks, and thus there is no general version of GAN that can be applied to a variety of tasks. GANs have primarily been used in the field of computer vision. It will be interesting to work toward the use of GANs for multi-modal data in order to cater to a broader set of Natural Language applications for Virtual Agents and Machine Translation.

References:

1. *Goodfellow Ian,et al.“Generative Adversarial Nets”. Advances in neural information processing systems. 2016.*
2. *M. Arjovsky, S. Chintala, and L. Bottou. Wasserstein GAN. ArXiv e-prints, Jan. 2017.*
3. https://www.analyticsvidhya.com/blog/2021/10/an-end-to-end-introduction-to-generative-adversarial-networksgans/#h2_7
4. https://towardsdatascience.com/five-gans-for-better-image-processing-fabab88b370b
5. https://en.wikipedia.org/wiki/Generative_model#Generative_models
6. https://en.wikipedia.org/wiki/Discriminative_model
7. https://developers.google.com/machine-learning/gan/generative#:~:text=A%20generative%20model%20includes%20the,to%20a%20sequence%20of%20words.