Deep Learning and Computer Vision

P2 : Unsupervised learning by GAN

Vishisht Sharma

1 Introduction

A generator network and a discriminator network are the two networks that make up a generative adversarial network, or GAN [1]. These two networks could be neural networks, such as autoencoders, recurrent neural networks, or convolutional neural networks. In this configuration, two networks are concurrently competing with one another and striving to outdo one another while assisting one another with their respective jobs. A random noise vector is transformed into a sample from a real data set by the generator network from a latent space (note that not all GANs sample from a latent space). The technique of training a GAN is quite intuitive. Both networks are simultaneously trained, and as time passes, they both get better. There are many practical applications for GANs, including the creation of images, works of art, music, and videos. Also, they may improve the quality of your photographs, stylize or colourize them, create faces, and carry out a variety of other fascinating jobs.

2 What are the main trends on the topic since the publication of the paper discussed?

There have been a number of notable advancements and trends in the field of GANs since Ian Goodfellow's "Generative Adversarial Networks" study was published in 2014, improved convergence and stability, The instability and difficulty in training GANs was one of its major drawbacks. In order to overcome this, a number of methods have been developed, such as batch normalisation, weight initialization, and alternate goal functions, leading to more stable training and improved convergence.

3 What are the key ideas of related published works since the original publication?

With DCGAN [2] Convolutional neural networks were applied within GANs for the first time, and the results were impressive. It featured significant architectural changes to address issues including training instability, mode collapse, and internal covariate shift, which made it a significant turning point in the research on GANs. Since then, a large number of GAN designs built on the DCGAN architecture have been released. BigGAN [3] improved with the use of large-scale architecture and higher quality data to generate high-resolution and diverse images with more control over image features. BigGAN also introduced a truncation trick to reduce the variation in generated images, and class-conditional self-attention to improve the quality of specific image classes. This resulted in state-of-the-art performance on various image generation benchmarks. StyleGANs [4] improved by the ability to control specific features of generated images, such as facial attributes and pose, by manipulating learned style vectors. This is achieved through a progressive training process and the use of adaptive instance normalization (AdaIN) layers, which allow for more fine-grained control over the style of each feature map in the generator. As a result, StyleGAN generated more realistic and diverse images compared to the older counterparts. The capacity to produce high-quality photos with exact control over image properties is Pix2pix's [5] key advancement over GANs. Pix2pix employs a conditional GAN architecture to map input photos to output images with desired qualities, in contrast to typical GANs, which learn to generate images from random noise.

4 What are the main problems solved or improvements over the original work?

DCGAN introduced a set of architectural guidelines for GANs, which greatly simplified the process of training stable and high-quality GAN models. It also showed that convolutional neural networks (CNNs) are effective in generating high-resolution images. BigGAN solved the problem of generating high-quality images at scale by introducing novel techniques such as large batch training, multi-resolution training, and hierarchical latent spaces. It achieved state-of-the-art performance on multiple image generation benchmarks. StyleGAN addressed the problem of limited control over generated images by introducing a new generator architecture that allows for more fine-grained control over specific features of generated images. It also introduced the concept of "style mixing" to generate images with a combination of features from different input images. Pix2pix solved the problem of image-to-image translation by introducing a conditional GAN that can learn a mapping between two image domains. It achieved state-of-the-art performance on various image translation tasks such as style transfer, semantic segmentation, and edge-to-photo translation.

5 What are the remaining problems from the published works so far?

(In context of this paper as i haven't discussed any models other than discussed in the report) While DCGAN introduced a set of architectural guidelines that greatly simplified the process of training GANs, it still requires significant hyperparameter tuning and is prone to mode collapse and instability during training. BigGAN achieved state-of-the-art performance on various image generation benchmarks, it requires a large amount of compute and is not accessible to most researchers, StyleGAN allows for more fine-grained control over generated images, it can still suffer from mode collapse and is computationally expensive. It also requires a large amount of labeled data for training. Pix2pix achieved state-of-the-art performance on various image translation tasks, it can still suffer from overfitting and is limited by the availability of paired training data.

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

We often see images from the cameras in extreme environments such as in space or small cameras such as the ones in medical surgical applications that they are often blurry and very low resolution. One way of tackling this problem is using better cameras but there is a limit to the amount of light a camera sensor can sense. A much better way in my opinion is to focus on new methods using GANs.Existing methods for video generation are often limited by short video sequences, low resolution, or lack of control over the content. Developing more effective methods for video generation that can scale to longer and higher-resolution videos while maintaining temporal consistency and controllability would have significant implications in fields such as entertainment, advertising, and surveillance.

References

[1] Generative Adversarial Networks https://arxiv.org/abs/1406.2661

[2] Unsupervised Representation Learning with Deep Convolutional Generative Adversarial Networks, https://arxiv.org/abs/1511.06434

[3] [Large Scale GAN Training for High Fidelity Natural Image Synthesis, https://arxiv.org/abs/1511.04587

[4] A Style-Based Generator Architecture for Generative Adversarial Networks, https://arxiv.org/abs/1812.04948

[5] Image-to-Image Translation with Conditional Adversarial Networks, https://arxiv.org/abs/1611.07004