GAN

Nicholas Klardie
University of New Hampshire

Introduction

A generative adversarial network (GAN) is a relatively new class of machine learning system which seeks the generation of new content rather than the classification of existing content. The network itself is made up of two networks, a generator and a classifier, operating in competition, with the generator trying to create output which the classifier misclassifies as genuine. Unlike the models we have learned so far, the goal of a GAN is not to learn classifications from its input data, but to learn to mimic the data accurately enough to fool the classifier, and hopefully a human. (Goodfellow, et al, 2014)

These networks, along with those with similar capabilities, have found several uses, ranging from image upscaling, (Ledig, et al, 2016) to the production of photorealistic images. (Karras, Laine and Aila, 2018) In addition, the dual-network architecture has been exploited in order to train more accurate classifier networks in situations where the amount of input data available is limited. (Ahsan, Sun and Essa, 2018) The photorealistic images produced by GANs and similar deep learning architectures have become so accurate as to be targeted by lawmakers. (H.R. 3230, 2019)

Milestones and Exemplary Work

The original proof of concept for the GAN came from Goodfellow et al, in a paper in which a network was demonstrated which is both able to generate accurate enough images and avoid the issues inherent in other comparable techniques. Unfortunately, the GAN is susceptible to what Goodfellow termed the “Helvetica problem”, where the networks are trained out of sync, and as a result the whole process fails. Proof of concept images were generated from the MNIST, TFD and CIFAR-10 datasets, and shows acceptable realism. The simple model, two three layer deep perceptrons shows the amount of promise this technique had, given the generated images. The door was left open for further improvement, and several areas of focus were suggested. (Goodfellow, 2014)

Goodfellow’s GAN would soon be improved to consume labels, and therefore be able to filter the output. This new GAN is the Conditional GAN, or CGAN. The primary difference between the GAN and the CGAN is the presence of labels, without which resulting is seemingly random output. The CGAN was also tested on MNIST data, and was able to produce satisfactory output, but it was also tested on data from the YFCC100 dataset, and was successfully able to predict human-looking tags for a given image. (Mirsa and Osindero, 2014)

Because two models are trained, the overall goal of the system does not have to be the generation of images. Rather, at the end of the training process, the discriminator can be used as a pre-trained model for further tuning. This is particularly useful in situations where input data is scarce. When used in this way, the GAN’s training process is used as a semi-supervised process. There has been success using GANs in this was using a variant known as the Deep Convolutional GAN, or DCGAN. (Ahsan, Sun and Essa, 2018) DCGANs were originally introduced with this purpose in mind, with the goal of finding a network which is as flexible and accurate as a convolutional neural network, or CNN, and as flexible as a GAN. (Radford, Metz and Chintala, 2015) DCGANs have been shown to produce suitable models for further refinement. (Ahsan, Sun and Essa, 2018)

In 2018, researchers at NVIDIA released a state-of-the-art architecture known as StyleGAN which extends existing GANs by making large changes to the overall architecture, and borrowing ideas from other models. While the generator in a typical GAN is a single relatively simple network, the StyleGAN splits it further into a mapping network, which preprocesses the input data, and a synthesis network, which does the actual generation. The mapping network is relatively simple, being a regular multi-layer perceptron, however the synthesizer adds complexity by adding style data from the mapping network at several stages, along with random noise. The addition of noise at several stages in the training process allows for fine details in the output image which make it look more realistic compared to previous models. The “style” data is an adaptation of previous work to apply an artistic style to real photographs in order to make them look as though they were either made without photography, or as if they were retouched in photo-editing software, known as “style transfer”. (Karras, Laine and Aila, 2018) (Huang and Belongie, 2017) The network is further augmented with the application of a technique named “style mixing”, where a limited subset of style data from two different generated images is combined, in effect augmenting the input data as the model is being

trained. While the actual amount of data mixed is relatively small, the differences are significant enough that the generated images are not recognizable between eachother. StyleGAN is described as being superior to traditional GANs in every way. (Karras, Laine and Aila, 2018)

Applications

As the state-of-the-art for GANs and similar networks, such as autoencoders, has grown, they have found extensive use in several areas, which previously were either not possible, computationally expensive, or intensely flawed. Image upscaling on many classes of raster images using state-of-the-art traditional non deep learning algorithms yields a version of the original image which is satisfactory, but degraded. By comparison, methods using CNNs exhibit a great improvement of quality over traditional algorithms, and advanced GANs have been shown to produce results comparable to or better than CNNs. Particularly with the proliferation smartphones as the typical consumer's personal camera, image upscaling is an important tool to increase the quality of images which were taken either at a far distance from their subject, or at a low optical zoom level. Accordingly, there is a high demand for accurate and realistic digital zooming techniques. (Guan, Pan, Li and Yu, 2019)

Since the publication of the paper on the original paper on DCGANs, there has been a very large number of papers published on the application of them to semi-supervised learning to tackle the problem of small datasets. These are used on a wide variety of applications, including but not limited to wildfire detection, (Aslan, Güdükbay, Töreyin and Çetin, 2019) ichthyoplankton classification, (Aljaafari, 2018) pedestrian detection, (Huang and Ramanan, 2017) and cancer research. (Rubin, et al, 2018) These are all situations where coming across a dataset of the size required to properly train a traditional learning system is either difficult, expensive, or both, and generally take similar approaches.

Clearly there are no end to the uses of GANs for toy uses, utilities and for the betterment of science and mankind, but only because of the accuracy of these models and the utility of faked data. The same techniques can be used as tools for criminals and other bad actors. While there are surely applications for the unscrupulous in check and wire fraud, phone scams, and other criminal activity that depends on subterfuge and is countered by recognizing the fake situation, and those are likely being exploited by some enterprising criminals, most of the attention is currently focused on generated depictions of real people. While these depictions have drawn the eye of lawmakers, with bills introduced in state legislatures and in Congress, some successfully, (H.R. 3230, 2019) (CA A.B. 602, 2019) whether or not they can survive a First Amendment test has yet to be seen. (Dold, 2019)

Bibliography

Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., & Bengio, Y. (2014). Generative Adversarial Networks.

Ledig, C., Theis, L., Huszar, F., Caballero, J., Cunningham, A., Acosta, A., Aitken, A., Tejani, A., Totz, J., Wang, Z., & Shi, W. (2016). Photo-Realistic Single Image Super-Resolution Using a Generative Adversarial Network.

Karras, T., Laine, S., & Aila, T. (2018). A Style-Based Generator Architecture for Generative Adversarial Networks.

Defending Each and Every Person from False Appearances by Keeping Exploitation Subject to Accountability Act of 2019, H.R. 3230, 116th Cong. (2019).

Mirza, M., & Osindero, S. (2014). Conditional Generative Adversarial Nets.

Ahsan, U., Sun, C., & Essa, I.. (2018). DiscrimNet: Semi-Supervised Action Recognition from Videos using Generative Adversarial Networks.

Radford, A., Metz, L., & Chintala, S.. (2015). Unsupervised Representation Learning with Deep Convolutional Generative Adversarial Networks.

Huang, X., & Belongie, S.. (2017). Arbitrary Style Transfer in Real-time with Adaptive Instance Normalization.

Guan, J., Pan, C., Li, S., & Yu, D.. (2019). SRDGAN: learning the noise prior for Super Resolution with Dual Generative Adversarial Networks.

Aslan, S., Güdükbay, U., Töreyin, B.U., & Çetin, A.E. (2019). Early Wildfire Smoke Detection Based on Motion-based Geometric Image Transformation and Deep Convolutional Generative Adversarial Networks. *ICASSP 2019 - 2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 8315-8319.

Aljaafari, N. (2018). Ichthyoplankton Classification Tool using Generative Adversarial Networks and Transfer Learning.

Huang, S., & Ramanan, D. (2017). Recognition in-the-Tail: Training Detectors for Unusual Pedestrians with Synthetic ImpostersArXiv, abs/1703.06283.

Rubin, M., Stein, O., Turko, N., Nygate, Y., Roitshtain, D., Karako, L., Barnea, I., Giryes, R., & Shaked, N.. (2018). TOP-GAN: Label-Free Cancer Cell Classification Using Deep Learning with a Small Training Set.

Depiction of individual using digital or electronic technology: sexually explicit material: cause of action, A.B. 602, Calif. State Legis. 2019-2020 session. (2019).

Dold, K. (2019). Face-Swapping Porn: How a Creepy Internet Trend Could Threaten Democracy. *The Rolling Stone.*