Queen Mary, University of London
School of Electronic Engineering and Computer Science

Unsupervised Learning by Generative Adversarial Nets

ECS795P - Deep Learning and Computer Vision

Mughees Asif | 180288337

Due: 17 March 2022

Generative Adversarial Networks (GANs) are an unsupervised modelling method that use two conflicting models to establish uniformities or patterns in input data that can be applied to unseen examples to produce a neoteric output. The technical architecture of GANs involves a combination of two different supervised networks, where a generator network (neural network with gradient descent) is designed to produce new outputs, while a discriminator model (classifier with gradient ascent) facilitates the classification of the output as either real or fake. The architecture reduces the overall computational complexity by implementing backpropagation to reduce the total loss and eliminating the usage of Markov chains to generate precise data distributions.

The computer vision domain has experienced various applications of GANs that have been implemented as possible solutions to challenges such as plausible image generation, image-to-image translation, and facial attribute manipulation [1]. The main trends in current research involve the development/modification of different architectures and loss functions for GANs that allow the generation of high-quality images, diversity of image generation, and model stability whilst training [1]. The latest research is diverting away from the mainstream GAN implementations that required a fully connected feed forward generator neural network to work synchronously with a discriminator classifier towards more complex architectures and custom-defined loss functions.

A possible configuration for the architecture involved the use of deep neural networks with convolutional layers for unsupervised representation learning called Deep Convolutional Generative Adversarial Networks (DCGANs) [2]. The authors introduced three key improvements over previous attempts: (1) use of strided convolutions, instead of the vanilla deterministic spatial pooling functions, enabled autonomous learning of spatial up/downsampling for unbalanced data; (2) directly connecting the highest convolutional features to the input and output layers improved learning stability; (3) employment of batch normalization to offset training problems associated with poor initialization and congested gradient flow in deep models [2]. Karras et al. developed a custom GAN with a progressive neural network (PROGAN), where both the competing models' (generator and discriminator) trainable parameters expand symbiotically during training, thus improving the quality of the media progressively until a benchmark is achieved [3]. Consequently, Karras et al. also suggested an alternative generator architecture that utilises the application of a non-linear mapping of the input to an intermediate latent space, instead of entering latent code via the input layer only. This methodology yields state-of-the-art distribution quality metrics that led to demonstrable improvements in the interpolation properties [4], thereby, becoming an industry-standard application.

In a corresponding study, Wang et al. proposed a custom loss function for image-to-image transformations with GANs involving two feed forward neural networks that combined perceptual loss with generative adversarial loss. This enabled the model to nullify discrepancies autonomously and continuously between the ground-truth images and the correlated transformed images [5]. In a complimentary study, Mirza et al. proposed the notion of a conditional GAN, where both the discriminator and generator models are instilled with pertinent data that aligns both sub-models towards the desired output [6].

As an extension to vanilla feature extraction for GANs where features are only captured in the local/sparse spatial regions, Zhang et al. designed a self-attention GAN (SAGAN) that uses high-resolution details to generate cues that can discern features from all locations in the state space [7]. SAGAN uses a self-attention module that leverages joint input from separated spatial regions to improve generation in multi-class image categories, thereby allowing the GAN to "... efficiently find global, long-range dependencies within internal representations of images" [7]. This work was extended by Brock et al. who demonstrated tangible improvements in SAGAN by scaling up the model with two to three times more trainable parameters and allocating a bigger batch size [8]. Additionally, the work also covered adding orthogonal regularization to the generator, therefore rendering fine control over the trade-off between reliability and diversity of high-resolution image synthesis.

Arjovsky et al. worked on offsetting mode collapse, improve learning stability by using a continuous loss function, and work on the vanishing gradient problem through the usage of the Wasserstein distance (WGAN) as a heuristic to model the distance between two probability distributions. This research helped to develop a framework for measuring the similarity between two probability distributions which was a perennial GAN issue [9]. Mao et al. proposed a GAN with the least square loss function (LSGAN) for the discriminator as opposed to a sigmoid cross entropy loss function to improve learning stability, in addition to minimizing the objective function to reduce Pearson divergence [10]. Zhu et al. worked on resolving the image translation conundrum ("... translate an image from a source domain X to a target domain Y in the absence of paired examples" [11]) by using a cycle consistency loss function such that the model learns "... a mapping $G : X \longmapsto Y$ such that the distribution of images from $G(X)$ is indistinguishable from the distribution Y" [11].

GANs are inherently hard to implement, train and evaluate due to a plethora of problems including achieving perfect synchronization between the generator and discriminator, learning instability due to conflicting protagonistic networks, and non-convergence due to the oscillation of the hyper-parameters during training. Although, research is being conducted in the mentioned domains for computer vision problems, more work needs to be done in extrapolating GAN results from images to other types of media such as voice and video.

References

[1] Zhengwei Wang, Qi She, and Tomás E. Ward. "Generative Adversarial Networks in Computer Vision: A Survey and Taxonomy". In: *ACM Comput. Surv.* 54.2 (Feb. 2021). ISSN: 0360-0300. DOI: 10.1145/3439723. URL: https://doi.org/10.1145/3439723 (page 1).

[2] Alec Radford, Luke Metz, and Soumith Chintala. "Unsupervised Representation Learning with Deep Convolutional Generative Adversarial Networks". In: *CoRR* abs/1511.06434 (2016) (page 1).

[3] Tero Karras, Timo Aila, Samuli Laine, and Jaakko Lehtinen. "Progressive Growing of GANs for Improved Quality, Stability, and Variation". In: *ArXiv* abs/1710.10196 (2018) (page 1).

[4] Tero Karras, Samuli Laine, and Timo Aila. "A style-based generator architecture for generative adversarial networks". In: *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. 2019, pp. 4401–4410 (page 1).

[5] Chaoyue Wang, Chang Xu, Chaohui Wang, and Dacheng Tao. "Perceptual Adversarial Networks for Image-to-Image Transformation". In: *IEEE Transactions on Image Processing* 27.8 (2018), pp. 4066–4079. DOI: 10.1109/TIP.2018.2836316 (page 1).

[6] Mehdi Mirza and Simon Osindero. "Conditional generative adversarial nets". In: *arXiv preprint arXiv:1411.1784* (2014) (page 1).

[7] Han Zhang, Ian Goodfellow, Dimitris Metaxas, and Augustus Odena. "Self-attention generative adversarial networks". In: *International conference on machine learning*. PMLR. 2019, pp. 7354–7363 (pages 1, 2).

[8] Andrew Brock, Jeff Donahue, and Karen Simonyan. "Large scale GAN training for high fidelity natural image synthesis". In: *arXiv preprint arXiv:1809.11096* (2018) (page 2).

[9] Martin Arjovsky, Soumith Chintala, and Léon Bottou. "Wasserstein generative adversarial networks". In: *International conference on machine learning*. PMLR. 2017, pp. 214–223 (page 2).

[10] Xudong Mao, Qing Li, Haoran Xie, Raymond YK Lau, Zhen Wang, and Stephen Paul Smolley. "Least squares generative adversarial networks". In: *Proceedings of the IEEE international conference on computer vision*. 2017, pp. 2794–2802 (page 2).

[11] Jun-Yan Zhu, Taesung Park, Phillip Isola, and Alexei A Efros. "Unpaired image-to-image translation using cycle-consistent adversarial networks". In: *Proceedings of the IEEE international conference on computer vision*. 2017, pp. 2223–2232 (page 2).