A Study on Different Types of Generative Adversarial Networks

Arnab Sinha

Abstract—The abstract goes here.

Index Terms—Computer Society, IEEE, IEEEtran, journal, LATEX, paper, template.

1 Introduction

THIS demo file is intended to serve as a "starter file" for IEEE Computer Society journal papers produced under LATEX using IEEEtran.cls version 1.8b and later. I wish you the best of success.

mds
August 26, 2015

1.1 Subsection Heading Here

Subsection text here.

1.1.1 Subsubsection Heading Here

Subsubsection text here.

/describe all of them (a), (b), etc., within the main caption.

2 Method

During the course of this internship, I learnt about and explored three types of GAN variations: Vanilla GANs , Super Resolution GANs and Conditional GANs. The following section shall focus upon the mathematics underlying these novel models.

2.1 Vanilla GANs

The preliminary version of generative adversarial networks is pretty straightforward and easy to understand. The fact that it was implemented with simple multilayer perceptrons made it easier to generate an intuition behind it. It focuses on creating a mapping, $G(z; \theta_g) \rightarrow y$, from an input vector z, belonging to a prior distribution p_z, to an output image y. G is a differentiable function represented by the multilayer perceptron with θ_g as parameters of neural network. The final theoretical aim is to make $p_{data} = p_g$, where p_{data} corresponds the probability distribution of the true data whereas p_g corresponds to the probability distribution of the data generated by the Generator G. A second multilayer perceptron, the opposition network, called the Discriminator, $D(x; \theta_d)$, is defined, where, for $\mathbf{x} \sim p_{data}$, $D(\mathbf{x})$ is the probability of $\mathbf{x}$ belonging to p_{data} rather than p_g. We train D to maximize this probability by training the actual data and generated data with their respective correct labels, 1 for real images and 0 for fake images. We simultaneously train G to minimize the probability of $\log(1 - D(G(\mathbf{z})))$. The entire equation goes as follows:

$$\underset{G}{min}\ \underset{D}{max}\ \mathrm{V}(G, D)\ \mathrm{E}_{\mathbf{x}\sim p_{data}(\mathbf{x})} \log[D(\mathbf{x})] + \mathrm{E}_{\mathbf{z}\sim p_z(\mathbf{z})} \log[1 - D(G(\mathbf{z}))] \quad (1)$$

The process is an iterative and numerical one, as mentioned by Ian Goodfellow *et al.* in [1], for every k iterations of discriminator training, 1 iteration of generator training is implemented. This ensures optimality of the discriminator and slow training of the generator. However, it has been observed and recommended to maximize $\log(D(G(\mathbf{z})))$ rather than minimizing $\log(1 - D(G(\mathbf{z})))$ proves beneficial as the latter does not produce sufficient gradients in the intial stages of training in order to train G well.

2.2 Super Resolution GANs

GANs finds its applications in image super resolution as well. Convolutional neural networks have proven to give supreme results [2]–[4] in the past with image related tasks and it is no different here as well. Ledig et al. in [5] have described both the generator and discriminator to be convolutional networks, with the core of the ResNet architecture [6]. The two networks, combined, attempt to create a mapping, $G(\mathbf{x}^{\mathbf{LR}}; \theta_g) \rightarrow \mathbf{x}^{\mathbf{HR}}$, where $\mathbf{x}^{\mathrm{LR}} \sim p^{\mathrm{LR}}$, the probability distribution of low resolution images, $\mathbf{x}^{\mathrm{HR}} \sim p^{\mathrm{HR}}$, the probability distribution of actual high resolution images, and θ_g defines the parameters of G. In totality, a minimax game is played, which can be mathematically presented as follows:

$$\underset{G}{min}\ \underset{D}{max}\ \mathrm{V}(G, D)\ \mathrm{E}_{\mathbf{x}^{HR}\sim p^{HR}} \log[D(\mathbf{x}^{HR})] + \mathrm{E}_{\mathbf{x}^{LR}\sim p^{LR}} \log[1 - D(G(\mathbf{x}^{LR}))] \quad (2)$$

The discriminator D is trained so as to reolve the conflict between super resolved images and real high resolution images. The generator, on the other hand, has a slightly modified loss function. While the adversarial loss stays the same as in Vanilla GANs, an additional perceptual loss is added. The perceptual loss is derived from pixel-wise MSE

- *Arnab Sinha is a student at The LNM Institute of Information Technology, Jaipur, India, 302031.*
E-mail: 17ucs034@lnmiit.ac.in

loss. The latter uses absolute values of the pixel intensities, whereas the perceptual loss uses values from the ReLU activation layers of a pre-trained VGG19 network as described in [4]. The loss is defined as the euclidean distance between the feature map values of the super-resolved image $G(\mathbf{x}^{LR})$ and the real image $\mathbf{x}^{HR}$. The formula, adapted from [5] goes as follows:

$$l^{SR}_{VGG/i,j} = \frac{1}{W_{i,j}H_{i,j}} \sum_{x=1}^{W_{i,j}} \sum_{y=1}^{H_{i,j}} (\phi_{i,j}(I^{HR})_{x,y} - \phi_{i,j}(G_{\theta_G}(I^{LR}))_{x,y})^2 \quad (3)$$

3 Conclusion

The conclusion goes here.

Appendix A
Proof of the First Zonklar Equation

Appendix one text goes here.

Appendix B

Appendix two text goes here.

Acknowledgments

The authors would like to thank...

References

[1] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, "Generative adversarial nets," in *Advances in neural information processing systems*, 2014, pp. 2672–2680.

[2] Y. LeCun *et al.*, "Lenet-5, convolutional neural networks," *URL: http://yann. lecun. com/exdb/lenet*, vol. 20, no. 5, p. 14, 2015.

[3] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," in *Advances in neural information processing systems*, 2012, pp. 1097–1105.

[4] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," *arXiv preprint arXiv:1409.1556*, 2014.

[5] C. Ledig, L. Theis, F. Huszár, J. Caballero, A. Cunningham, A. Acosta, A. Aitken, A. Tejani, J. Totz, Z. 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*, 2017, pp. 4681–4690.

[6] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 770–778.

John Doe Biography text here.

Jane Doe Biography text here.