Deep Generative Models: Practical Comparison Between Variational Autoencoders and Generative Adversarial Networks

Mohamed EL–KADDOURY[1], Abdelhak MAHMOUDI[2], and Mohammed Majid HIMMI[1]

[1] *LIMIARF, Faculty of Sciences, Mohammed V University, Rabat, Morocco, {mh.kadouri,himmi.fsr}@gmail.com*

[2] *LIMIARF, Ecole Normale Superieure, Mohammed V University, Rabat, Morocco, abdelhak.mahmoudi@um5.ac.ma*

Introduction

Deep Learning models can achieve impressive performance in supervised learning but not for unsupervised one. In image generation problem for example, we have no concrete target vector. Generative models have been proven useful for solving this kind of issues. In this work, we will compare two types of generative models: Generative Adversarial Networks (GANs) and Variational Autoencoders (VAEs). We apply those methods to different data sets to point out their differences and see their capabilities and limits as well. We find that, while VAEs are easier and faster to train, their results are in general more blurry than the images generated by GANs. These last are more realistic but noisy.

Methods

- Variational auto-encoders[1] are a class of deep generative models based on variational methods.

  The objective function is:

$$\tilde{\mathcal{L}}(\phi,\theta,\lambda) = \mathbb{E}_{x\sim q(x)}[\frac{1}{N}\sum_{l=1}^{L}(\ln p_\theta(x|z_\phi^{(l)})$$

$$+\ln p_\lambda(z_\phi^{(l)}) - \ln q_\phi(z_\phi^{(l)}|x))],$$

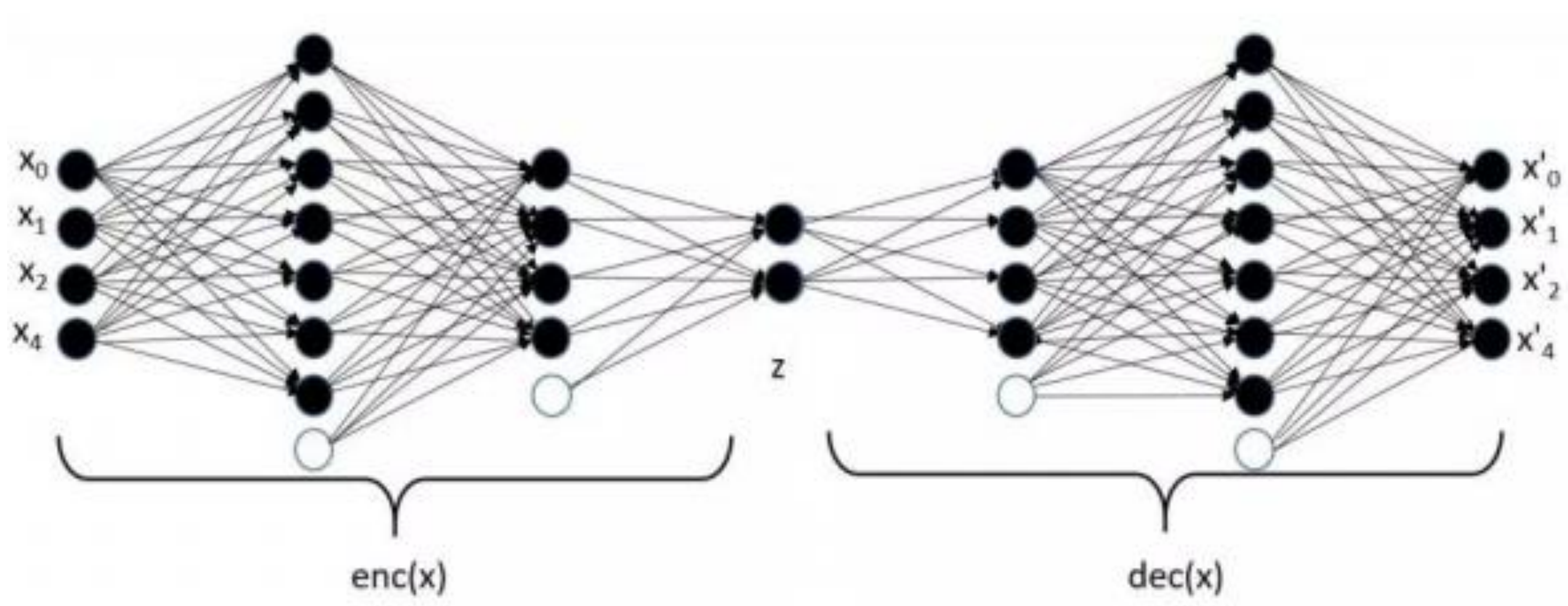


Fig. 1. VAE Architecture.

- The GANs [2] framework is composed of a generator G(z) and a discriminator D(x), where z is random noise. The generator G(z) tries to generate more and more likely data to fools the discriminator D(x), while the discriminator D(x) aims to tell apart the fake data from the real data. The value function of this adversarial process is as follows:

$$min_G max_D V(D,G) =$$

$$\mathbb{E}_{x\sim p_{data}(x)}[logD(x)] + \mathbb{E}_{z\sim p_z(z)}[log(1-D(G(z)))]$$

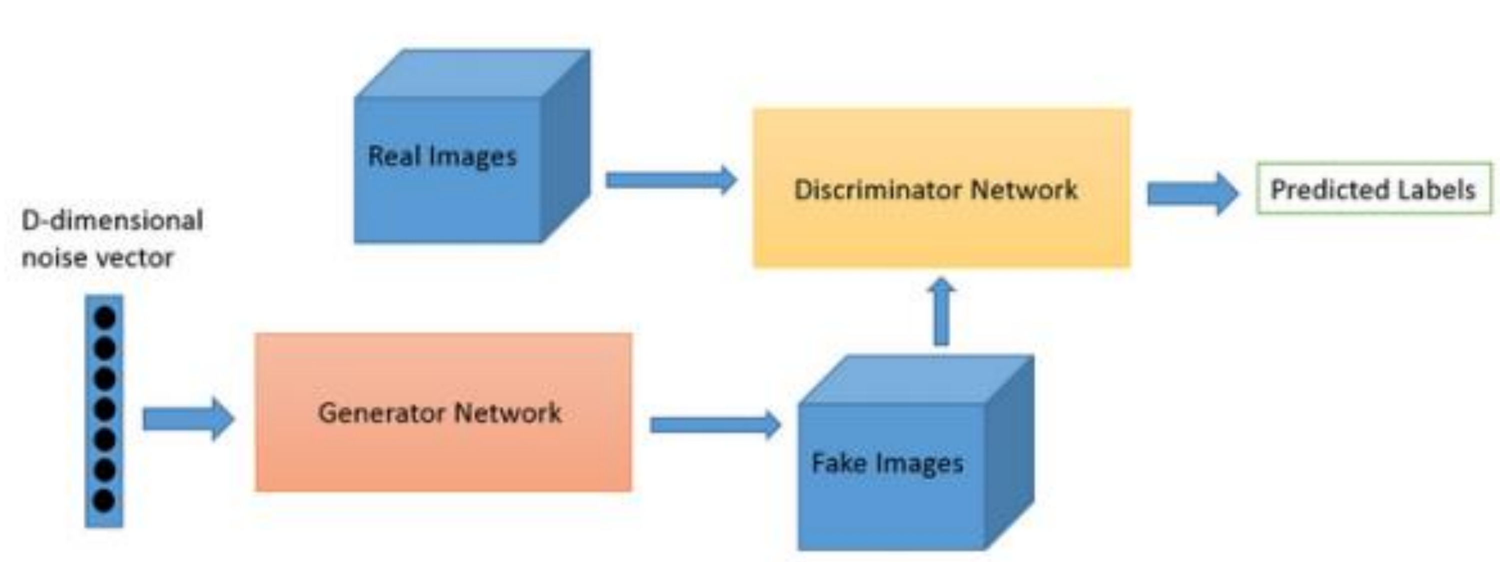
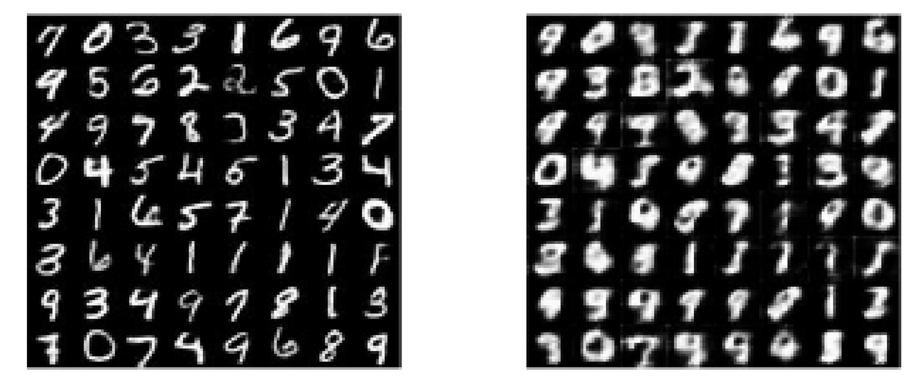


Fig. 2. GAN Architecture.

Results



a. Samples generated by GAN b. Samples generated by VAE

Fig. 3. Comparison of sampled images of the two models based on the MNIST dataset.

- Both models generate images which can easily be recognized as digits. While the GANs generates sharper images, the VAE tends to smooth the edges of the digits.

a. Samples generated by GAN b. Samples generated by VAE

Fig. 4. Comparison of sampled images of the two models based on the CIFAR10 dataset.

- The images of the VAE are once again blurry and no realistic objects can be recognized. The GAN generates images with sharper edges

a. Samples generated by GAN b. Samples generated by VAE

Fig. 5. Comparison of sampled images of the two models based on the CelebA dataset.

- The GAN produces again much sharper images than the VAE. Nevertheless, the faces produced by the VAE own a more natural appearance.

Dataset	Method	Inception score
MNIST	VAEs	3.32
	GANs	**9.05**
CelebA	VAEs	2.78
	GANs	**7.02**
CIFAR-10	VAEs	3.0
	GANs	**6.8**

Table 1. Comparison of Inception scores of the two models based on a different dataset.

Conclusion

The main difference between VAEs and GANs is their learning process. VAEs are minimizing a loss reproducing a certain image, and can, therefore, be considered as solving a semi-supervised learning problem. GANs, on the other hand, are solving an unsupervised learning problem. The most important difference found in this work was the training time for the two methods. GANs took longer time to train. Therefore the use of GANs was considered and proved a lot more stable. With GANs this does not necessarily occur. Eventually, we conclude, that for low-diversity datasets like MNIST, both methods give sufficiently realistic images. Finally, using VAEs one can achieve results in less time, but with decreased image quality compared to results of GANs.

References

1. D. P. Kingma and M. Welling. “Auto-encoding variational Bayes”. arXiv:1312.6114, 2013.
2. 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, pages 26722680, 2014.