Learning latent space representations (VAE)

Mohammed Jassem Hlel, Yassine Abou Hadid, Ismail El Hadrami

Data Science Project
Nov 8, 2022

Introduction

Problem : Generate images from the dataset MNIST

Methods explored :

- Auto-Encoder (AE)
- Variational Auto-Encoder (VAE)
- Vector Quantized Variational Autoencoder (VQ-VAE)
- Deep Convolutional Generative Adversarial Network (DeepConvGAN)

CVAE (2)

- The VAE is composed of two components: the probabilistic decoder which is naturally defined by p(x/z) (likelihood) whereas the probabilistic encoder is defined by p(z/x)(posterior). (1)
- CVAE is a Convolutional Neural Networks based model that is more adequate for image reconstructions.

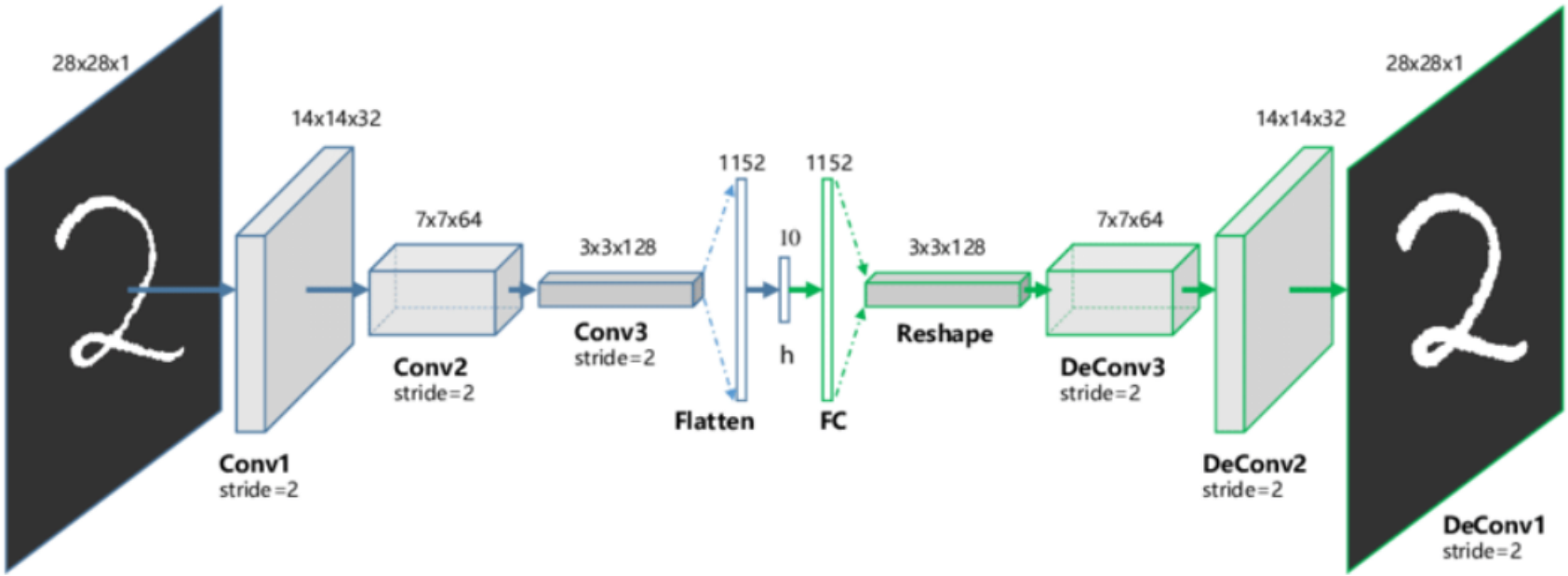


Figure: CVAE Network

VQ-VAE (3)

- VAE learns a continuous latent representation, whereas VQ-VAE learns a discrete latent representation.

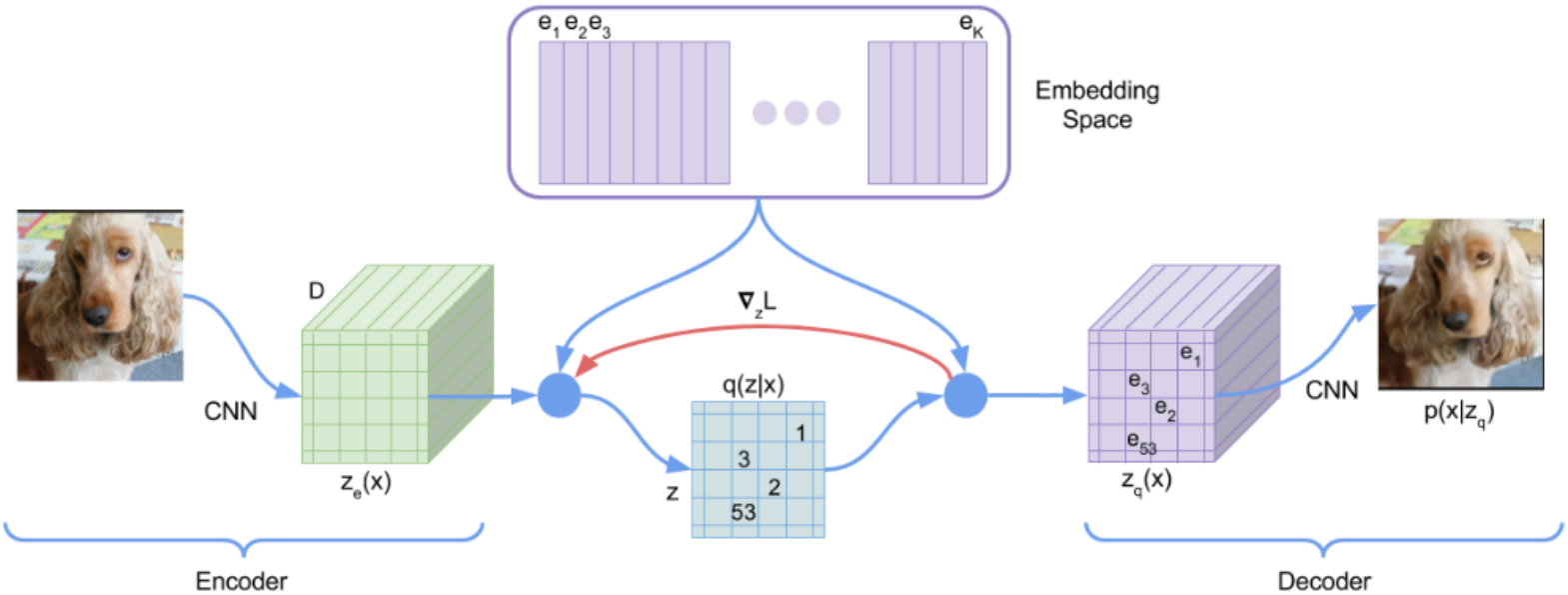


Figure: VQ-VAE architecture

$$L = \log p\left(x \mid z_q(x)\right) + \left\|\mathrm{sg}\left[z_e(x)\right] - e\right\|_2^2 + \beta \left\|z_e(x) - \mathrm{sg}[e]\right\|_2^2 \qquad (1)$$

Deep Convolutional GAN (4)

- The GAN is a class of generative models formed by 2 neural networks: Generator and Discriminator
- DeepConvGAN uses convolutional and convolutional-transpose layers in the discriminator and the generator, respectively.

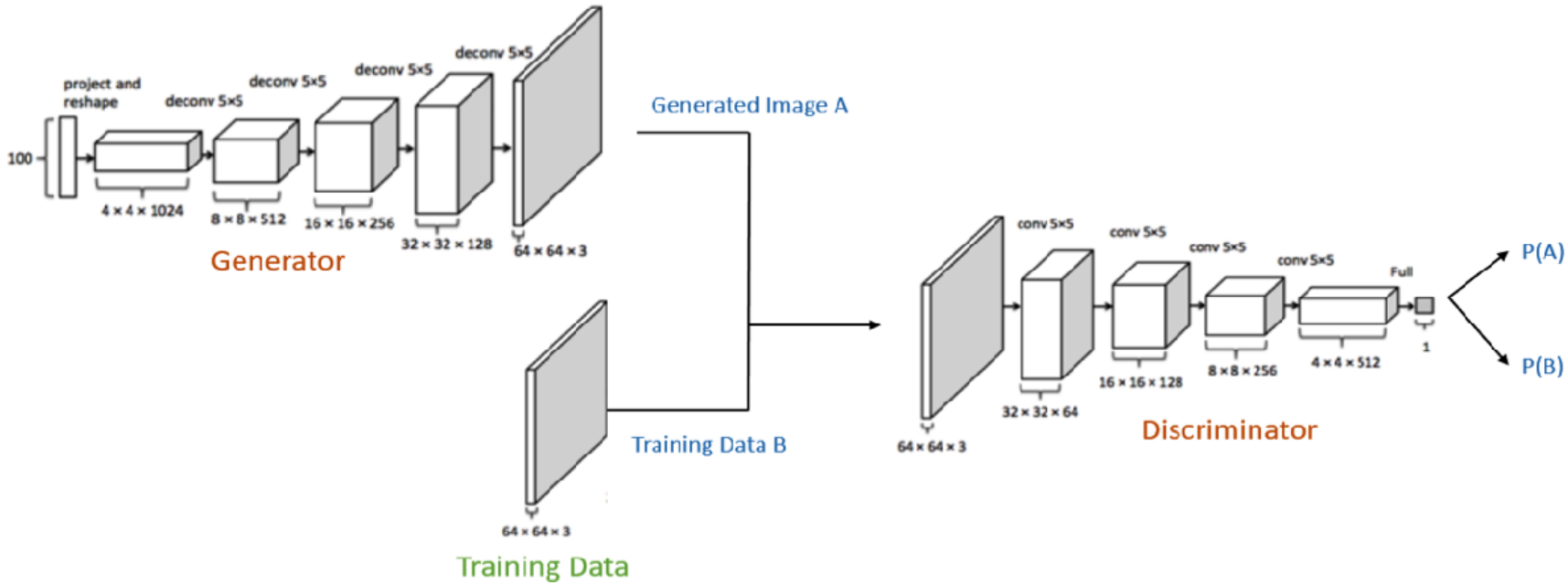
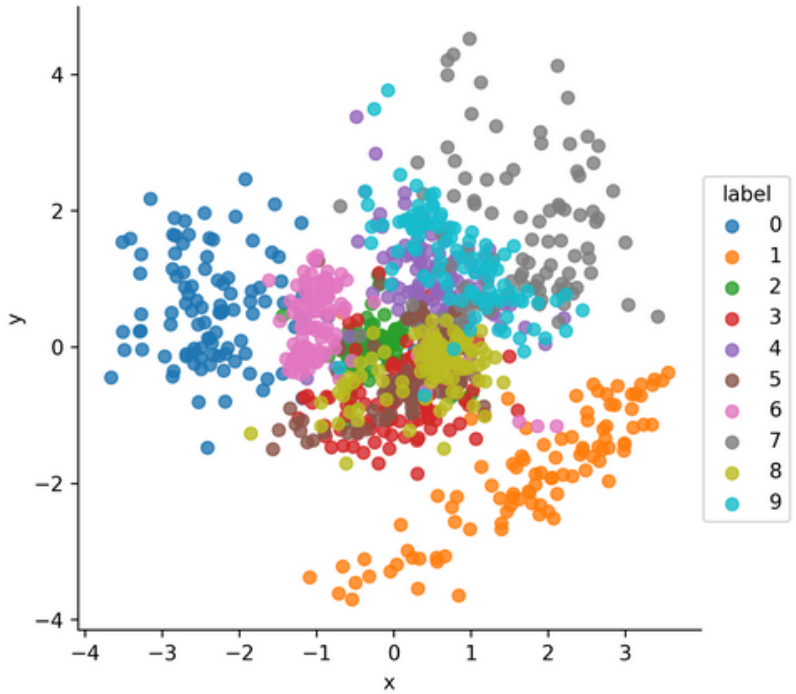
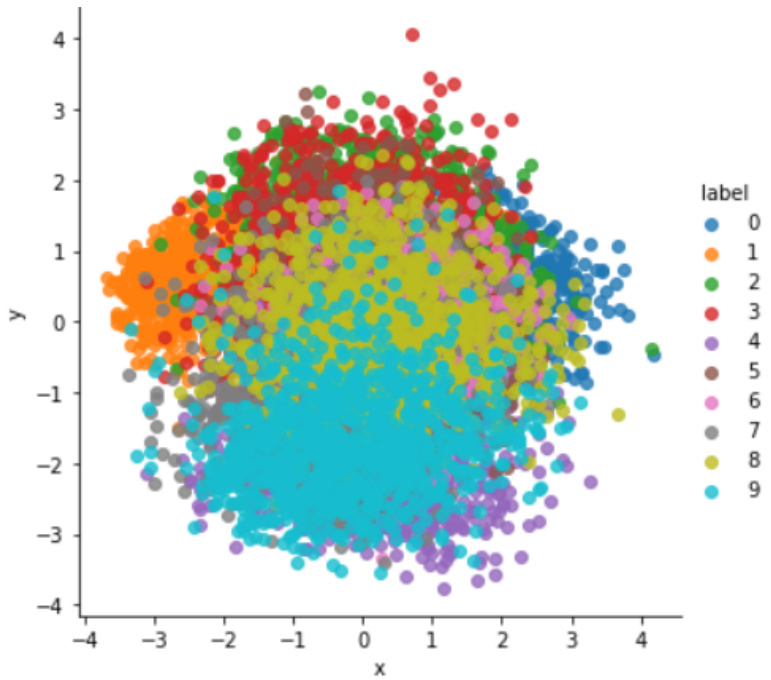


Figure: GAN Network

Results



(a) VAE



(b) CVAE

Figure: Representation of the latent space

Results

Figure: Real Data

(a) VQ-VAE Reconstruction

(b) CVAE Reconstruction

Model	FID
Real Data	0.005
CVAE	0.603
VQ-VAE	12.855
DeepConv GAN	36.186
Random Gaussian noise	83.064

Future work

- Re-run the tests with similar number of parameters for the models to draw a more accurate conclusion
- Try to improve the implemented models to work better with noise as input
- Explore other architectures and see if any is more adapted for our problem
- Create the generate.py file

References

[1] D. P. Kingma and M. Welling, "Auto-encoding variational bayes," *arXiv preprint arXiv:1312.6114*, 2013.

[2] K. Sohn, X. Yan, and H. Lee, "Learning structured output representation using deep conditional generative models," in *Proceedings of the 28th International Conference on Neural Information Processing Systems - Volume 2*, NIPS'15, (Cambridge, MA, USA), p. 3483–3491, MIT Press, 2015.

[3] A. v. d. Oord, O. Vinyals, and K. Kavukcuoglu, "Neural discrete representation learning," 2017.

[4] A. Radford, L. Metz, and S. Chintala, "Unsupervised representation learning with deep convolutional generative adversarial networks," 2015.

Thank you!

Annexe

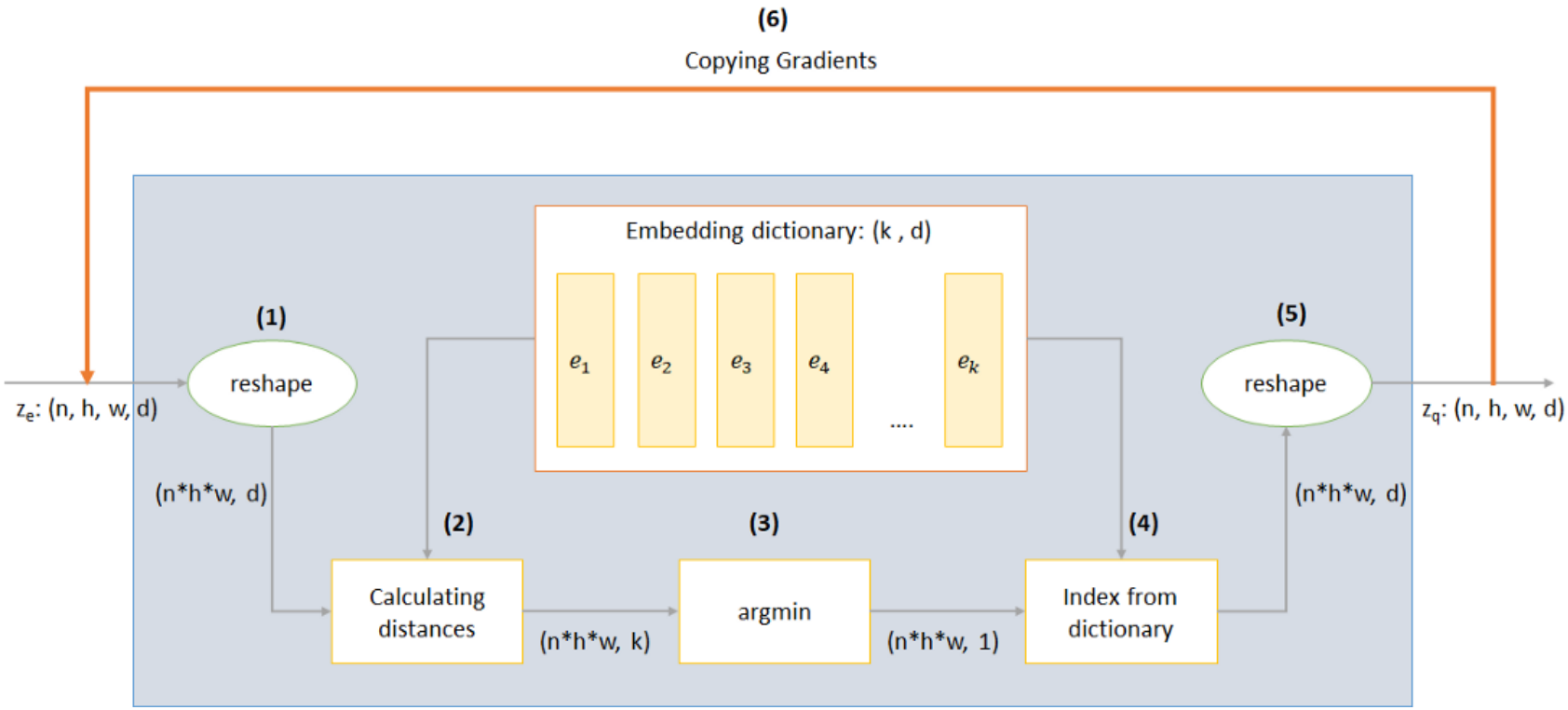


Figure: Vector Quantization Layer

Figure is from this **article**