Latent Bernoulli Autoencoder
Supplementary Material
ICML 2020

Jiri Fajtl [1] **Vasileios Argyriou** [1] **Dorothy Monekosso** [2] **Paolo Remagnino** [1]

Appendix A Model Architecture

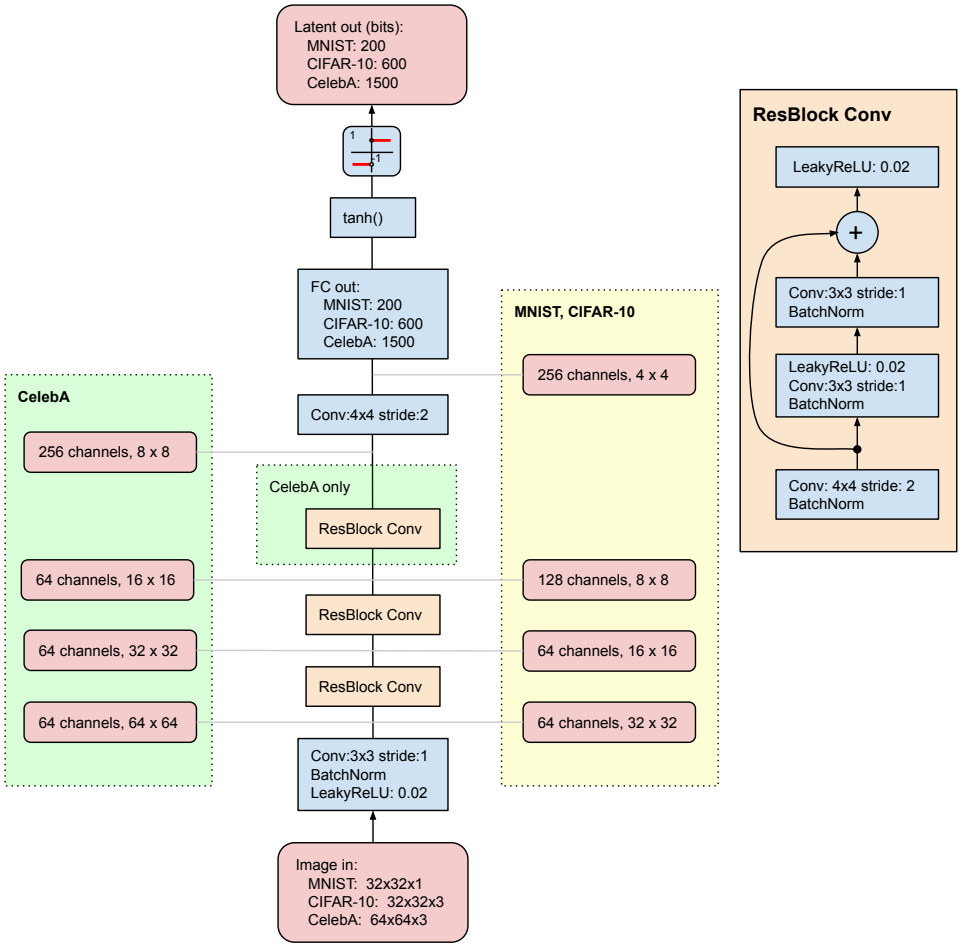


Figure 1. LBAE Encoder

[1]Kingston University, London, UK [2]Leeds Beckett University, Leeds, UK. Correspondence to: Jiri Fajtl <j.fajtl@kingston.ac.uk>.

Proceedings of the 37^{th} International Conference on Machine Learning, Online, PMLR 119, 2020.

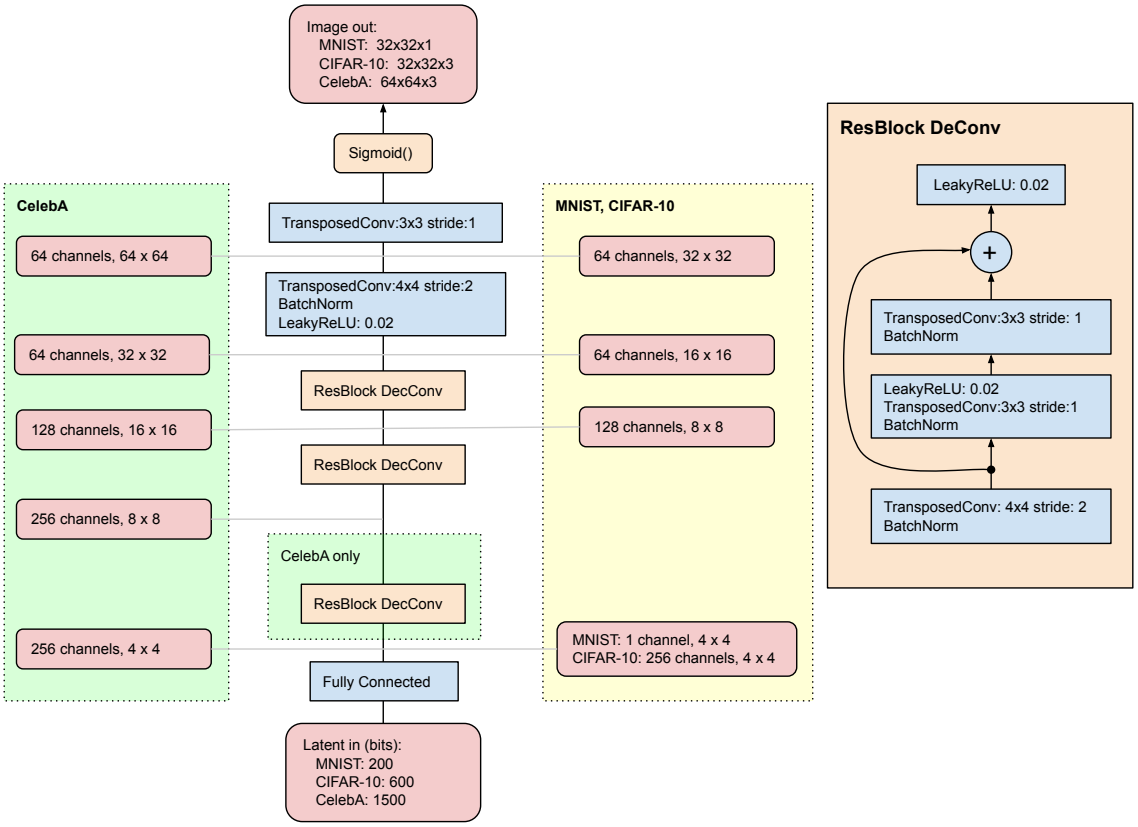


Figure 2. LBAE Decoder

Appendix B Compression

While not a comprehensive evaluation, the Table 1 shows compression performance of our method LBAE compared with VAE and, only on the CIFAR-10 dataset, to the VQ-VAE (van den Oord et al., 2017). Compression ratio on the ImageNet(Russakovsky et al., 2015) images with resolution $128 \times 128 \times 3$, as presented in the VQ-VAE work, has not been yet explored with our method.

The compression is reported as the input sample (image) size over the compressed latent representation, both in bits, similar to the method in VQ-VAE publication. Additionally, we relate the compression ratio to the reconstruction quality reported as FID. On the CIFAR-10, the VQ-VAE discrete latent code indexes $8 \times 8 \times 10$ embeddings in a dictionary with 512 entries. Therefore, each index requires 9 bits and together the code consumes $8 \times 8 \times 10 \times 9 = 5760$ bits. Size of the real-valued VAE latents is estimated as 32 bits (32 bits floating-point variables) per dimension. Arguably, the latents do not saturate all 32 bits at each dimension, thus the reported values are just informative. More thorough evaluation in this direction is a subject of upcoming research work.

In Table 1, we can observe that LBAE shows significantly higher compression compared to VAE as well as higher quality in FID. LBAE offers also higher compression than the VQ-VAE, although we could not compare the reconstruction quality, thus this result can not be considered conclusive.

Appendix C Qualitative results

Appendix C.1 CIFAR-10

Table 1. Comparison of input/latent size compression ratio and corresponding FIDs on the test dataset reconstruction. VQ-VAE compression is based on data from the van den Oord et al. 2017 publication, available only for CIFAR-10.

	MNIST Image size: 1024 bits (32x32x1)			CIFAR-10 Image size: 24576 bits (32x32x3x8)			CelebA Image size: 98304 bits (64x64x3x8)		
Method	Latent size(bits)	Compression ratio	FID	Latent size(bits)	Compression ratio	FID	Latent size(bits)	Compression ratio	FID
LBAE	200	5.12	8.11	600	40.96	19.37	1500	65.54	7.71
VAE (OURS)	512	2	8.77	4096	6	37.9	2048	48	34.96
VQ-VAE				5760	4.27				

Figure 3. Reconstruction on the test dataset.

Figure 4. Random samples with LBAE method.

Figure 5. Interpolation on the test dataset.

Appendix C.2 MNIST

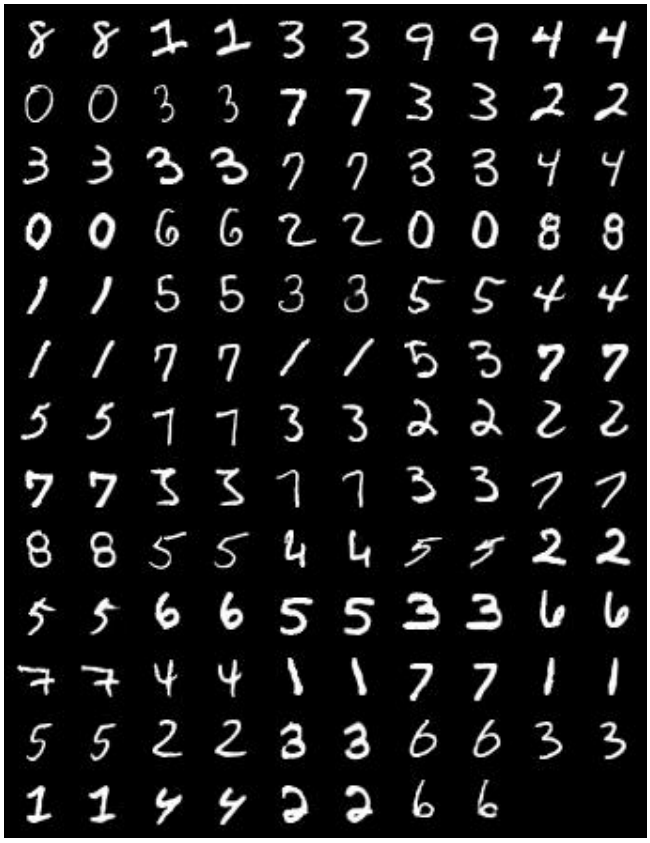


Figure 6. Reconstruction on the test dataset.

Figure 7. Random samples with LBAE method.

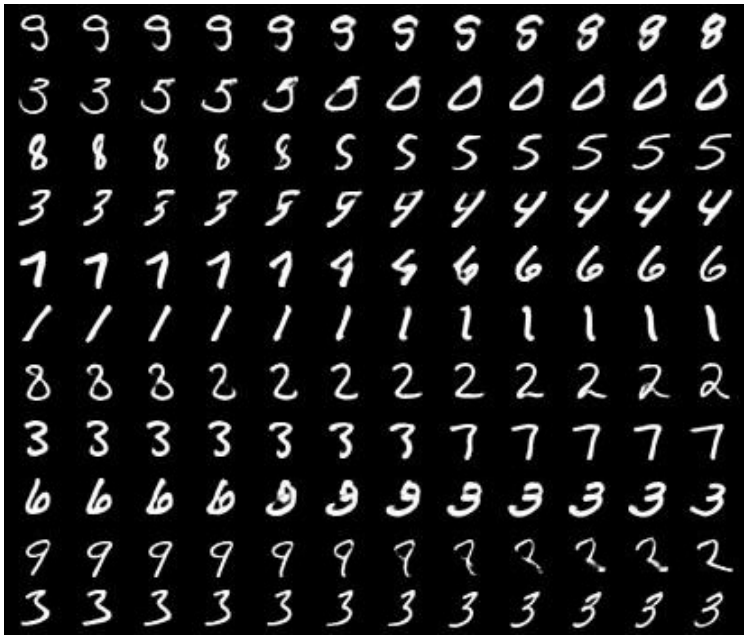


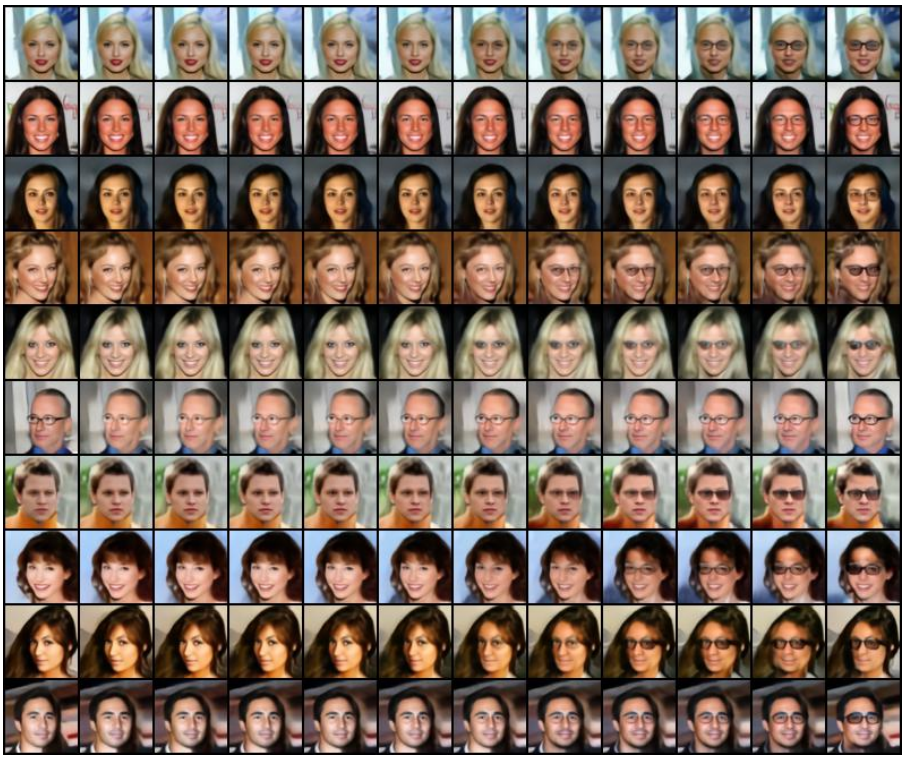
Figure 8. Interpolation on the test dataset.

Appendix C.3 CelebA

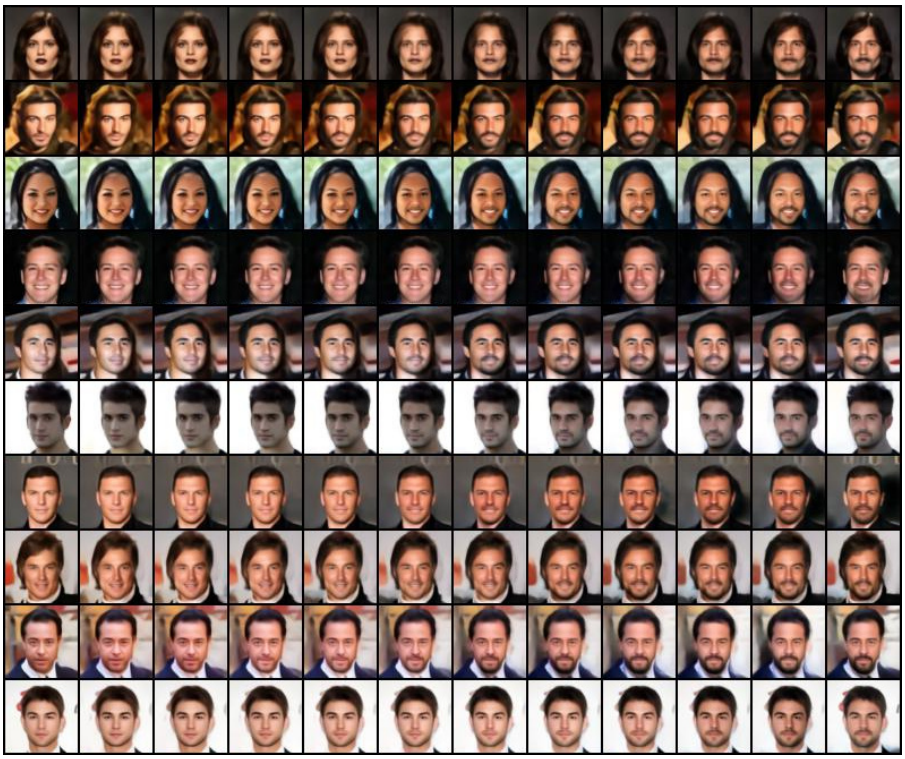
Figure 9. Reconstruction on the test dataset.

Figure 10. Random samples with LBAE method.

Figure 11. Interpolation on the test dataset.



(a) Setting eyeglasses attribute



(b) Setting goatee attribute

Figure 12. Interpolation between test images (left) and the same images (right) with modified attributes.

References

Russakovsky, O., Deng, J., Su, H., Krause, J., Satheesh, S., Ma, S., Huang, Z., Karpathy, A., Khosla, A., Bernstein, M., and Others. Imagenet large scale visual recognition challenge. *International Journal of Computer Vision*, 115(3):211–252, 2015.

van den Oord, A., Vinyals, O., et al. Neural discrete representation learning. In *Advances in Neural Information Processing Systems*, pp. 6306–6315, 2017.