Supplementary material for Batch norm with entropic regularization turns deterministic autoencoders into generative models

Amur Ghose and **Abdullah Rashwan** and **Pascal Poupart**
University of Waterloo and Vector Institute, Ontario, Canada
$\{$a3ghose,arashwan,ppoupart$\}$@uwaterloo.ca

On the next two pages, we present additional qualitative results.

We add in this page general notes on the training of EAEs, as requested by reviewers. In particular, across extensive experiments, we noted the following empirical trends and heuristics which we choose to pass on for the sake of ease of implementation.

- Results for all EAEs that use the Kozachenko-Leonenko or similar KNN based entropy estimators can be improved in general by using ≥ 2 neighbours per minibatch. However, we do not recommend this. Performance gains from this are slight, and for most applications, using 1 suffices.
- We recommend using at least 3 conv-deconv layers in the encoder and decoder for any autoencoding pair for all three datasets for best FID scores.
- For both CIFAR and CelebA, we recommend a minimum latent size of 32.
- For CIFAR-10 in particular, learning rate decay is critical when using the ADAM optimizer. We use an exponential decay with a decay rate ≥ 0.98. It should be noted that (Ghosh et al., 2019) use a more complex schedule that involves looking at the validation loss. We did not require such.
- Good samples emerge early - samples for all three datasets generated by epoch 10 as evaluated by a human eye are highly predictive of eventual best performance in terms of FID. As such, it is recommended to periodically generate samples and visually inspect them.

Preprocessing datasets

Here, we detail the pre-processing of datasets common to our methods and the methods we benchmark against. We carry out no pre-processing for CIFAR-10. For MNIST, we pad with zeros to reach 32×32 as the shape. For CelebA, pre-processing is important and can vastly change FID scores. We perform a center-crop to 140×140 before resizing to 64×64.

Details of following material

In Table 1 below, we present a larger version of the FID results from Table 1 in the main paper. In Figures 1 and 2 below, we also present qualitative results including reconstruction and interpolations on the latent space that serve to show that the latent spaces obtained by EAEs are meaningful. These experiments on latent spaces mirror (Ghosh et al., 2019).

	CIFAR-10		CelebA	
Architectures(Isotropic)	FID	Reconstruction	FID	Reconstruction
VAE	106.37	57.94	48.12	39.12
CV-VAE	94.75	37.74	48.87	40.41
WAE	117.44	35.97	53.67	**34.81**
2SVAE	109.77	62.54	49.70	42.04
EAE	**85.26(84.53)**	29.77	**44.63**	40.26
Architectures(MVG)	FID	Reconstruction	FID	Reconstruction
RAE	83.87	29.05	48.20	40.18
RAE-L2	80.80	32.24	51.13	43.52
RAE-GP	83.05	32.17	116.30	39.71
RAE-SN	84.25	**27.61**	44.74	36.01
AE	84.74	30.52	127.85	40.79
AE-L2	247.48	34.35	346.29	44.72
EAE	**80.07**	29.77	**42.92**	40.26
Architectures(GMM)	FID	Reconstruction	FID	Reconstruction
VAE	103.78	57.94	45.52	39.12
CV-VAE	86.64	37.74	49.30	40.41
WAE	93.53	35.97	42.73	**34.81**
2SVAE	N/A	62.54	N/A	42.04
RAE	76.28	29.05	44.68	40.18
RAE-L2	74.16	32.24	47.97	43.52
RAE-GP	76.33	32.17	45.63	39.71
RAE-SN	75.30	**27.61**	40.95	36.01
AE	76.47	30.52	45.10	40.79
AE-L2	75.40	34.35	48.42	44.72
EAE	**73.12**	29.77	**39.76**	40.26

Table 1: FID scores for relevant VAEs and VAE-like architectures. Scores within parentheses for EAE denote a regularization on a linear map. Isotropic denotes samples drawn from a latent space of $\mathcal{N}(0, I)$. GMM denotes sampling from a mixture of 10 Gaussians of full covariance. These evaluations correspond to analogous benchmarking for RAEs (Ghosh et al., 2019). Alongside FID values appearing in Table 1 of the main paper, we add results obtained when a Multivariate Gaussian (MVG) i.e. $\mathcal{N}(\mu, \Sigma)$ of full covariance is used for ex-post density estimation. Note that values for reconstruction are not changed by change of density estimators.

Figure 1: Qualitative comparisons to RAE variants and other standard benchmarks on CIFAR-10. On the left, we have reconstructions (top row being ground truth GT), the middle has generated samples, the right has interpolations. From top to bottom ignoring GT: VAE, CV-VAE, WAE, 2SVAE, RAE-GP, RAE-L2, RAE-SN, RAE, AE, EAE. Non-EAE figures reproduced from (Ghosh et al., 2019)

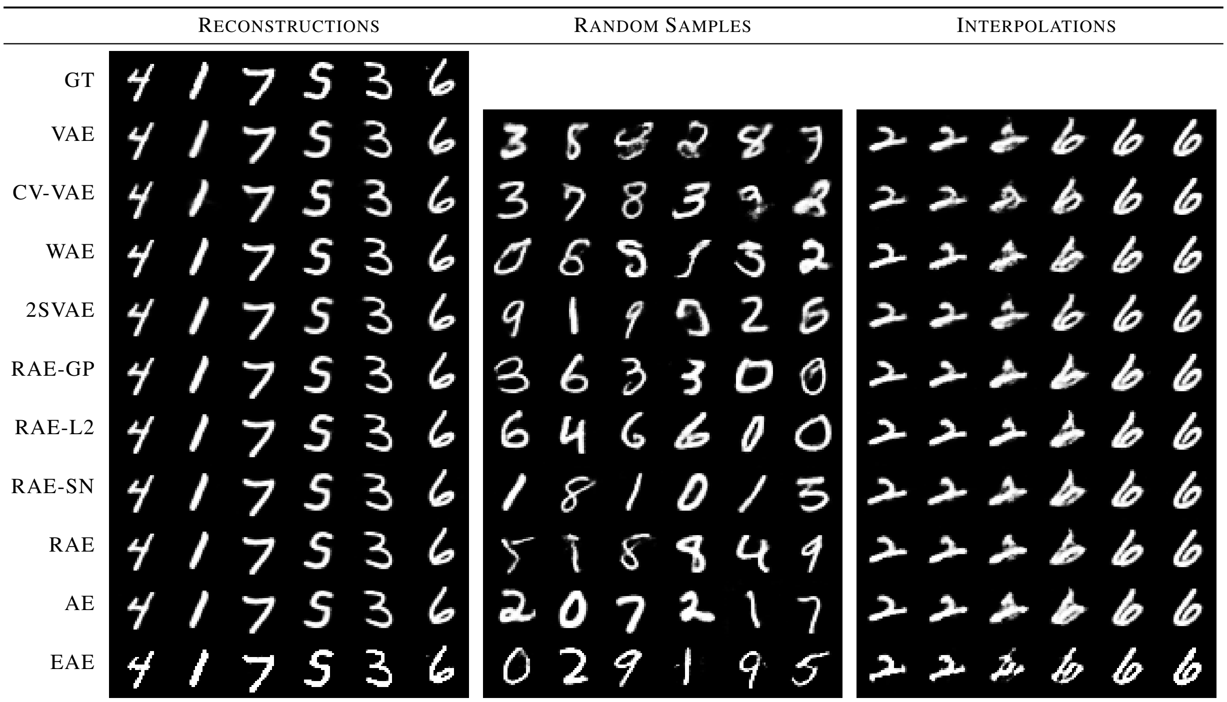

Figure 2: Qualitative comparisons to RAE variants and other standard benchmarks on MNIST. On the left, we have reconstructions (top row being ground truth GT) , the middle has generated samples, the right has interpolations. From top to bottom ignoring GT: VAE, CV-VAE, WAE, 2SVAE, RAE-GP, RAE-L2, RAE-SN, RAE, AE, EAE. Non-EAE figures reproduced from (Ghosh et al., 2019)

References

P Ghosh, M SM Sajjadi, A Vergari, M Black, and B Schölkopf. From variational to deterministic autoencoders. *arXiv:1903.12436*, 2019.