Y-Autoencoders: disentangling latent representations via sequential-encoding

Massimiliano Patacchiola*
University of Edinburgh
mpatacch@ed.ac.uk

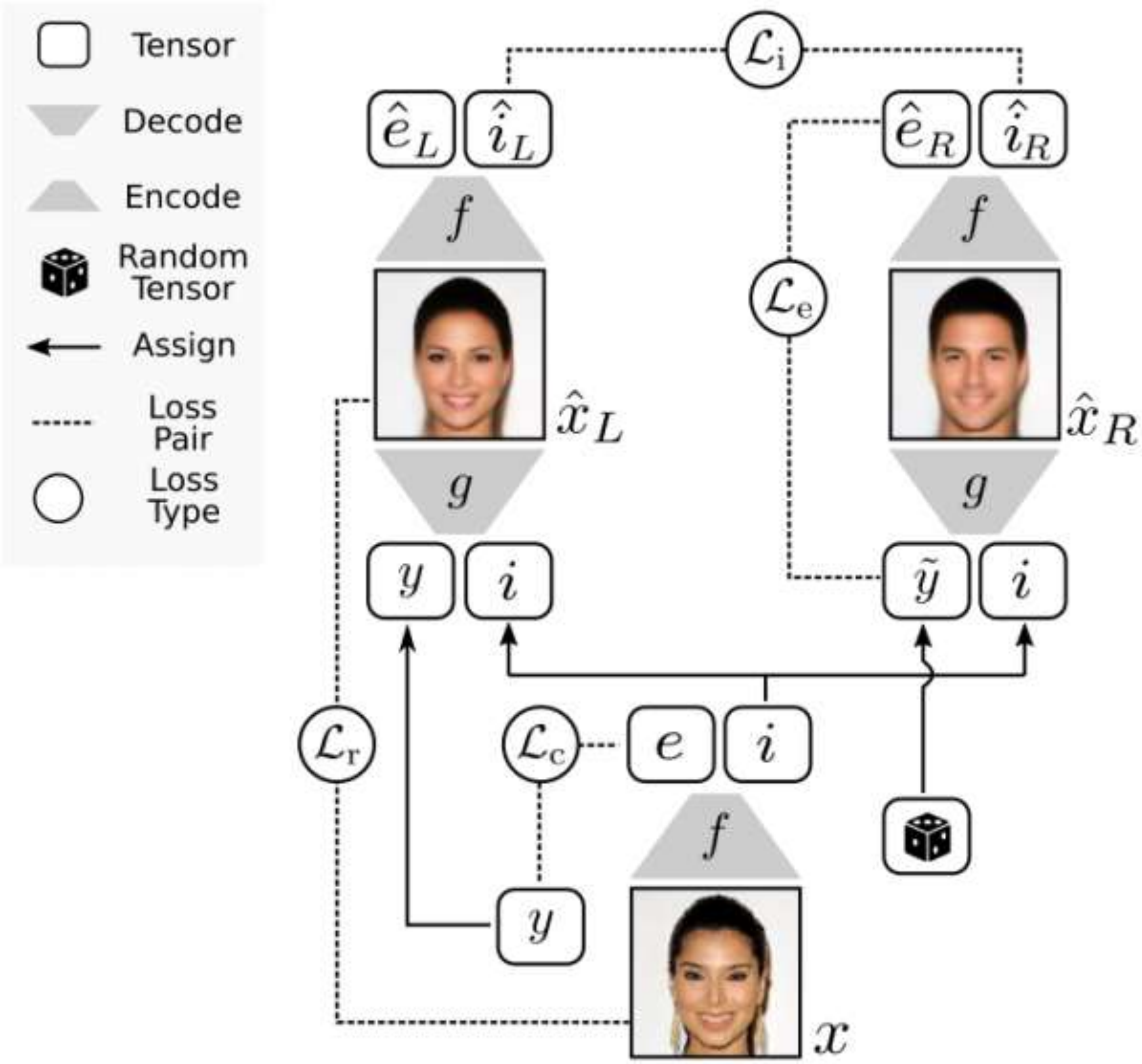
Patrick Fox-Roberts
Snap Inc.
pfoxroberts@snap.com

Edward Rosten
Snap Inc.
erosten@snap.com

arxiv

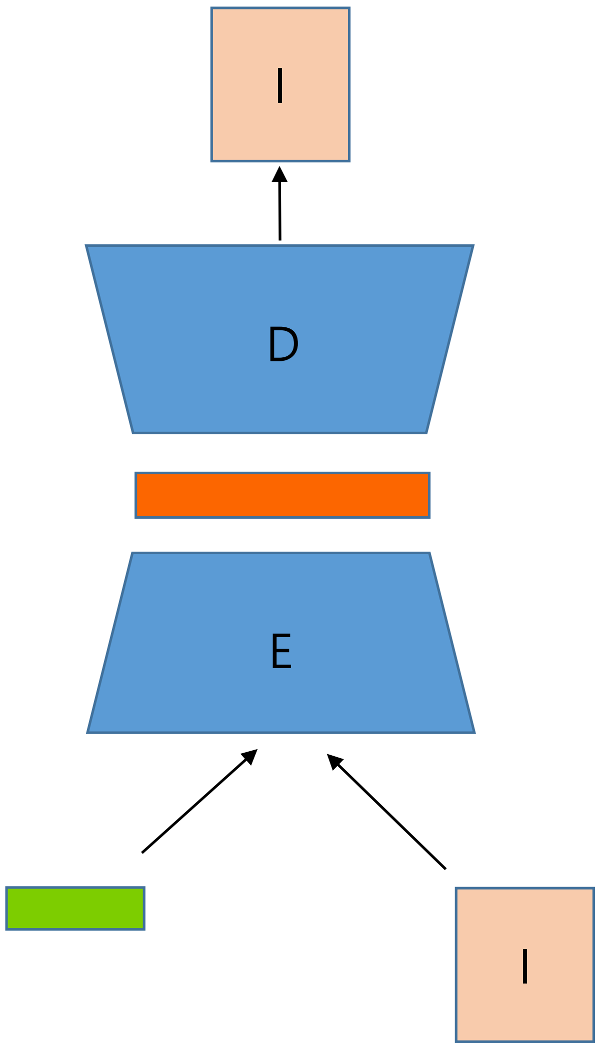
19/08/08, Yonggyu Kim

Y-AutoEncoder



Y-AEs generally represent **explicit information** via discrete latent units, and **implicit information** vias continuous units.

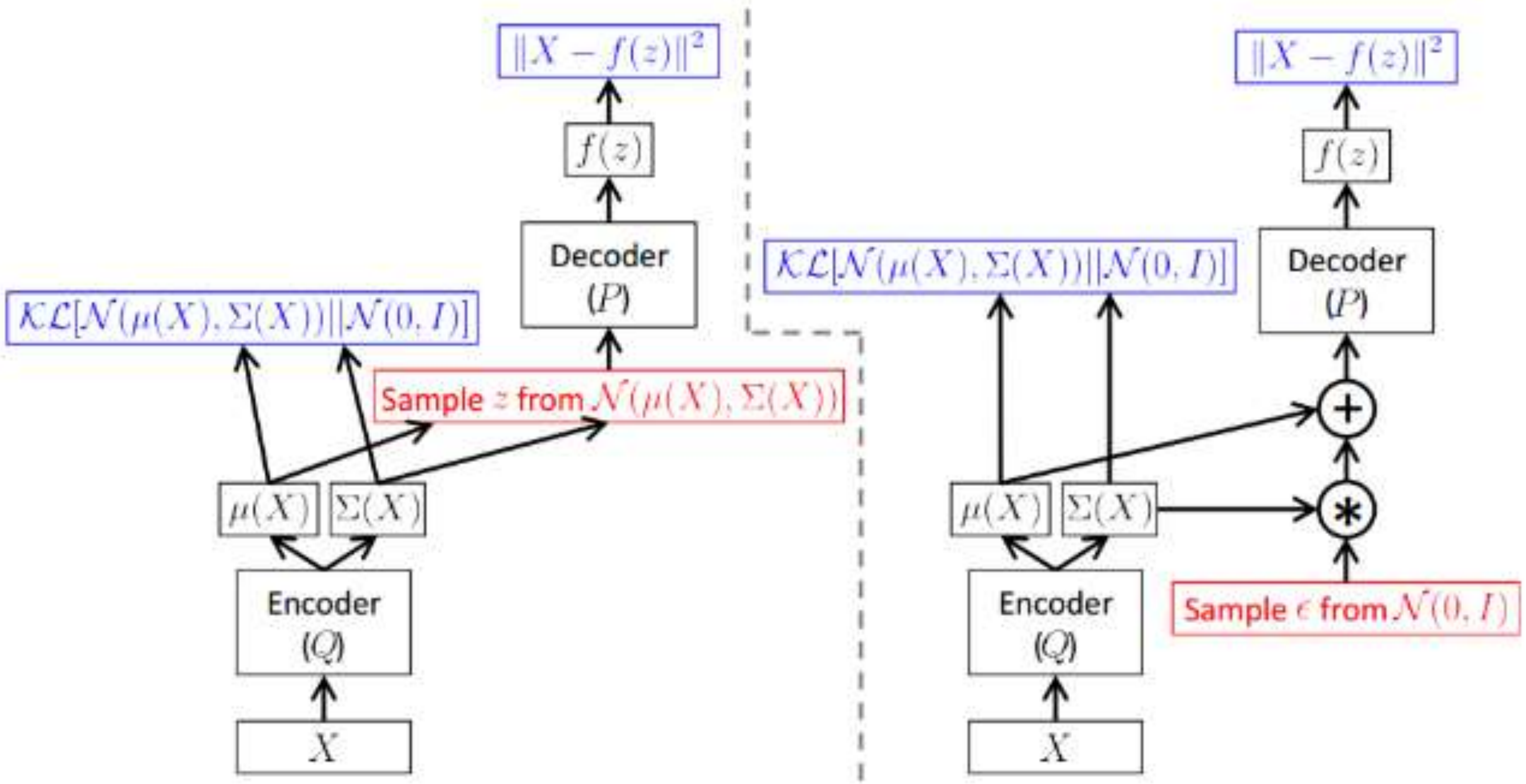
Conditional AutoEncoder



cAEs often struggle disentangling the latent representation,
Because there is no effective regularization to enforce an effect.

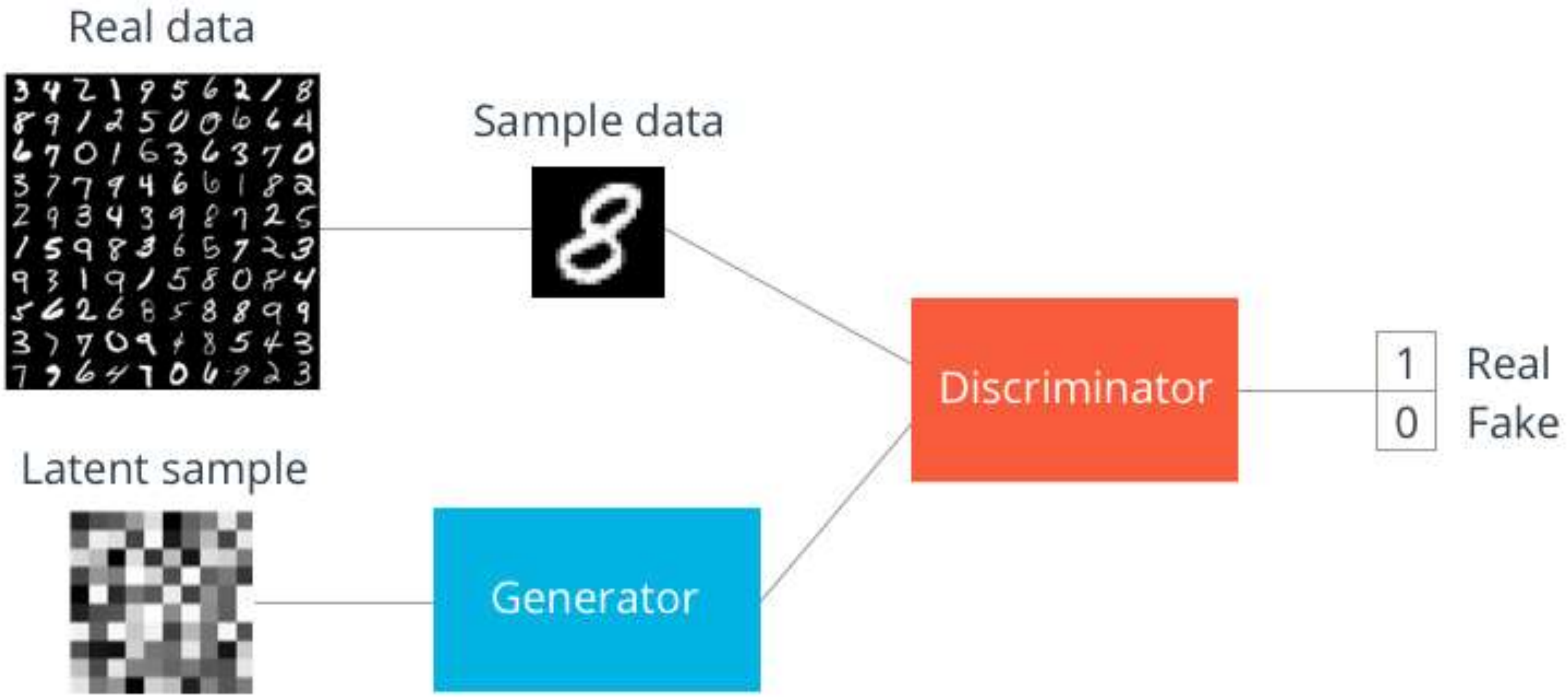
VAEs

VAEs rarely include discrete units due to the inability to apply backpropagation through those layers.

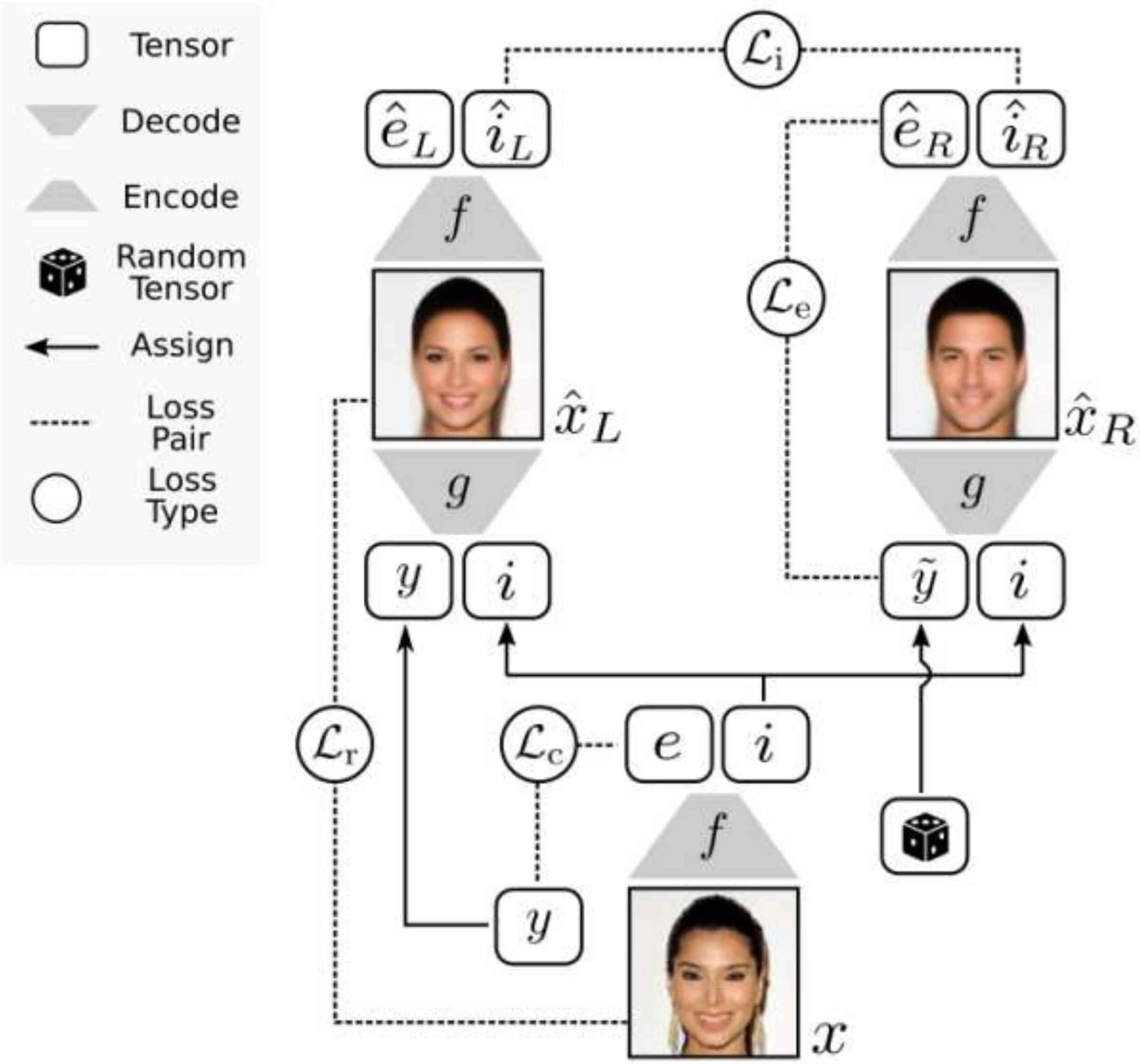


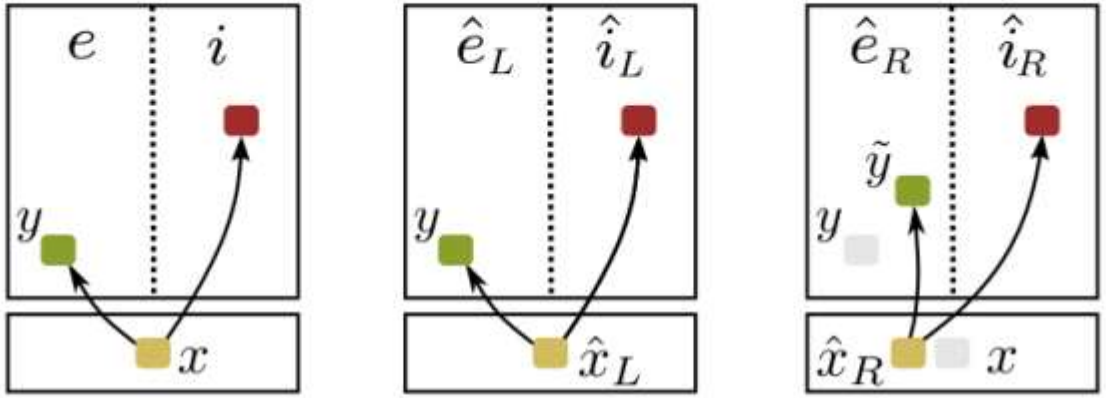
GANs & VAEs

GANs are notoriously difficult to train, and may suffer of mode collapse when the state space is implicitly multimodal.



Model



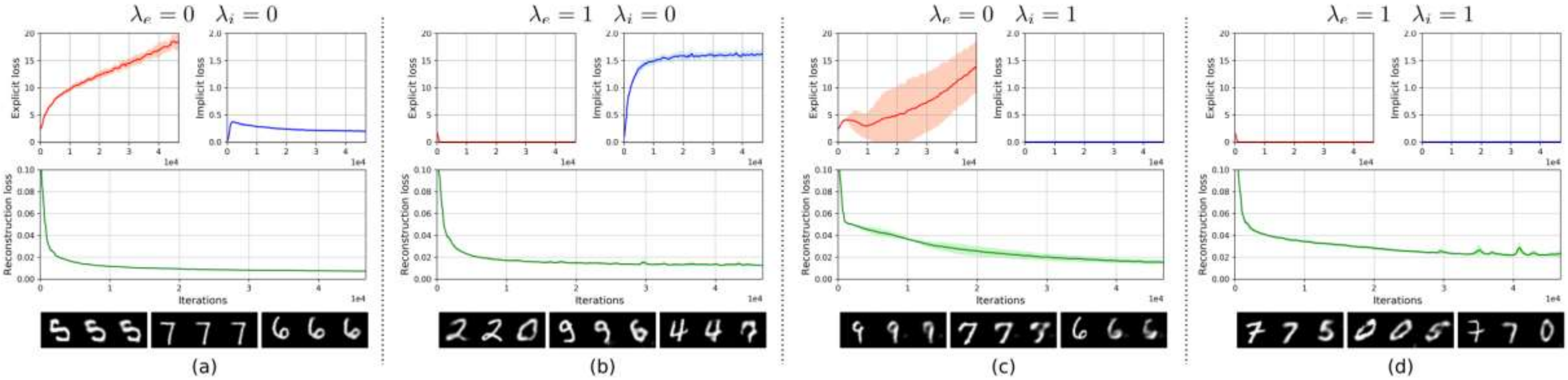


(a) 1st encoding (b) 2nd encoding (left) (c) 2nd encoding (right)

$$\mathcal{L} = \mathcal{L}_\mathrm{r} + \mathcal{L}_c + \lambda_e \mathcal{L}_e + \lambda_i \mathcal{L}_i,$$

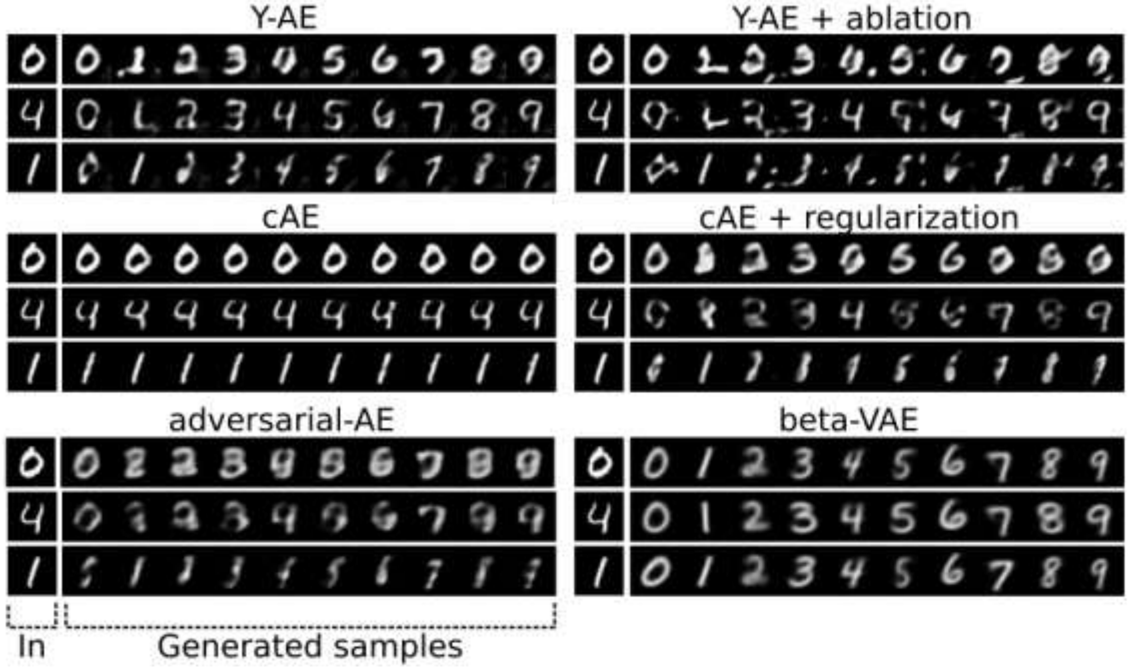
Experiments

Ablation Study

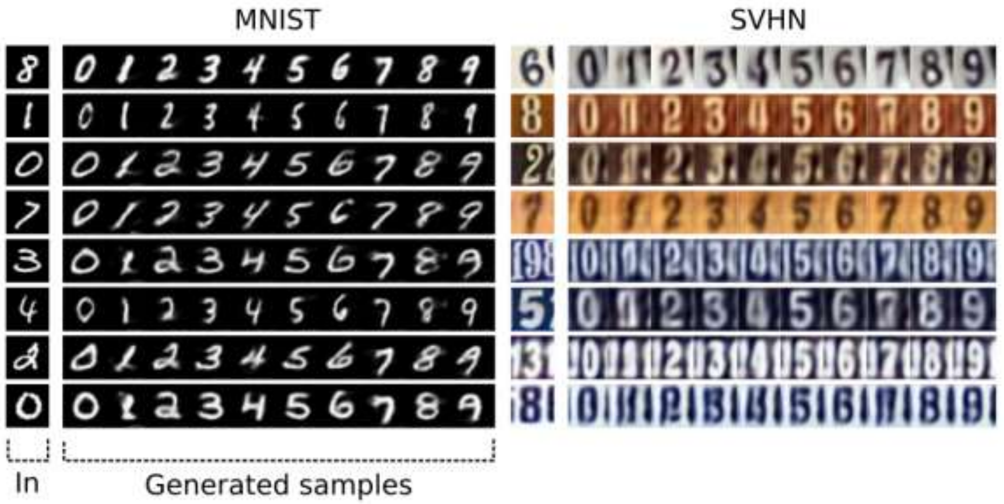


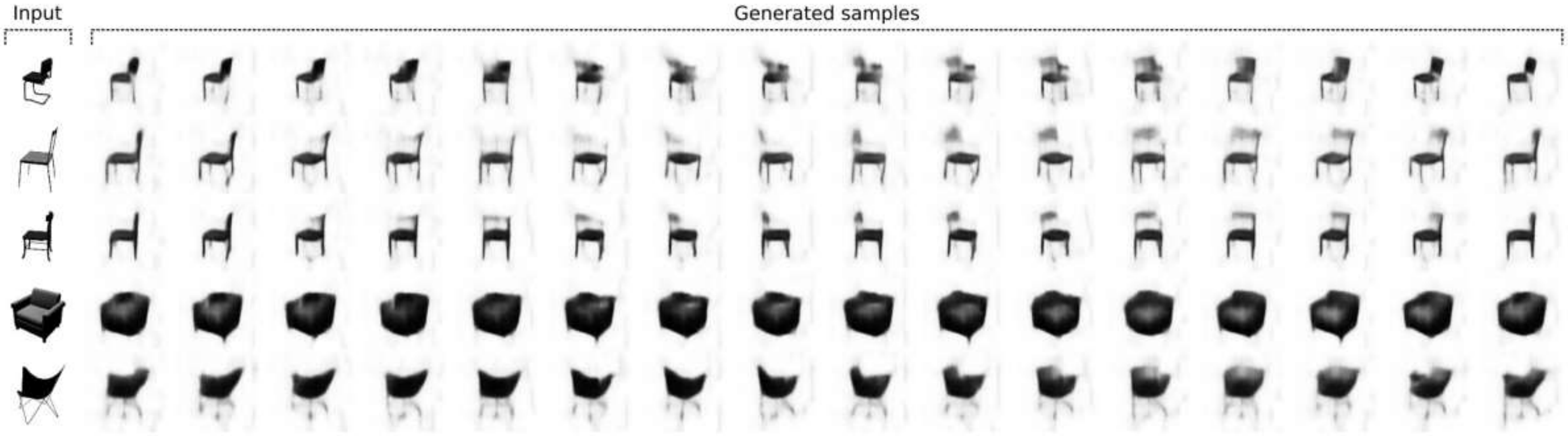
Experiments

Method	Accuracy (%)	SSIM	MSE
cAE	10.6 ± 0.1	0.87	17.52
cAE + regularizer	66.9 ± 17.5	0.55	26.43
adversarial-AE [21]	43.4 ± 10.5	0.57	27.4
cVAE [15]	96.7 ± 1.6	0.50	27.05
beta-VAE [12]	$\mathbf{99.7 \pm 0.1}$	0.42	30.43
Y-AE + ablation [our]	90.5 ± 2.9	0.59	27.38
Y-AE [our]	99.5 ± 0.1	**0.37**	**42.99**



Experiments





Experiments