Adversarial Autoencoders

Marcin Sidorowicz, Wojciech Wojnar

January 22, 2020

Table of contents

1. VAE - Variational Autoecnoders
2. GAN - Generative Adversarial Networks
3. AAE - Adversarial Autoencoder
4. AAE applications
   - Supervised learning
   - Semi-supervised learning
   - Unsupervised Clustering
   - Dimensionality Reduction
5. 3d-AAE applications
   - Generating by interpolations in latent space
   - Generating by algebraic operations in latent space
   - Clustering using 3d-AAE

VAE - Variational Autoecnoders

$$\mathbb{E}_{\mathbf{x}\sim p_d(\mathbf{x})}[-\log p(\mathbf{x})] < \mathbb{E}_{\mathbf{x}}[\mathbb{E}_{z\sim q(\mathbf{z}|\mathbf{x})}[-\log p(\mathbf{x}|\mathbf{z})]] + \mathbb{E}_{\mathbf{x}}\left[\mathbf{KL}\left(q(\mathbf{z}|\mathbf{x})\|p(\mathbf{z})\right)\right] = \text{reconstruction} + \text{regularization} \quad (1)$$

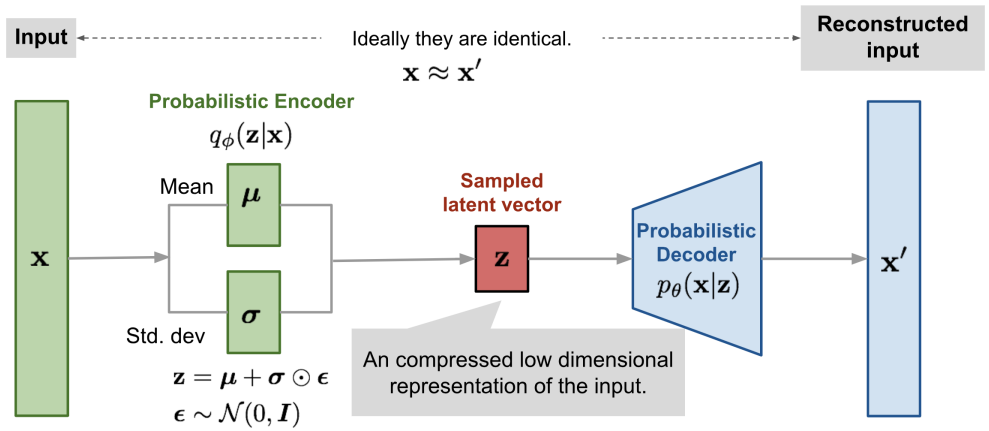

Figure 1: VAE architecture with multivariate normal prior

GANs

$$\min_G \max_D \mathrm{E}_{\mathrm{x}\sim p_{\mathrm{data}}}[\log D(\mathbf{x})] + \mathrm{E}_{\mathbf{z}\sim p(\mathbf{z})}[\log(1 - D(G(\mathbf{z}))] \quad (2)$$

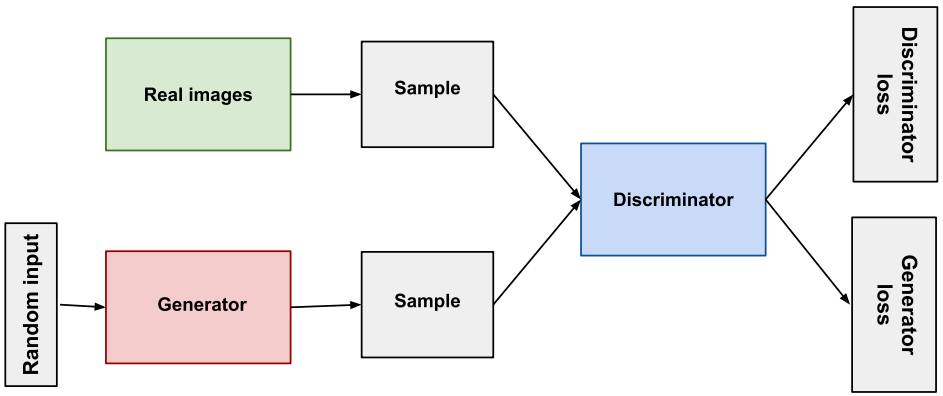

Figure 2: GAN architecture

AAE - Adversarial Autoencoder

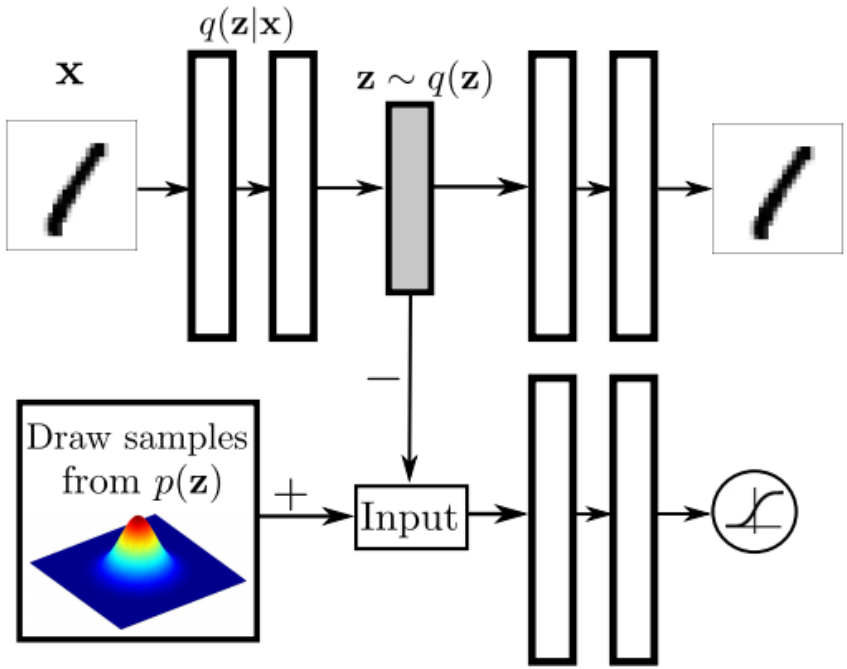

Figure 3: Basic AAE architecture

AAE - Adversarial Autoencoder

$$\min_E \max_D V(E, D) = \mathbb{E}_{\mathbf{z} \sim p(\mathbf{z})}[\log D(\mathbf{z})] + \mathbb{E}_{\mathbf{x} \sim p_d(\mathbf{x})}[\log (1 - D(E(\mathbf{x})))] \quad (3)$$

$$L_D = -V(E, D)$$

$$L_{EG} = \textit{reconstruction} + V(E, D)$$

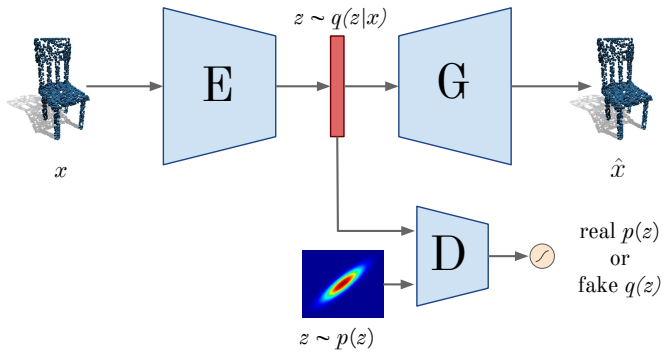

Figure 4: AAE architecture

Imposing a prior distribution - examples

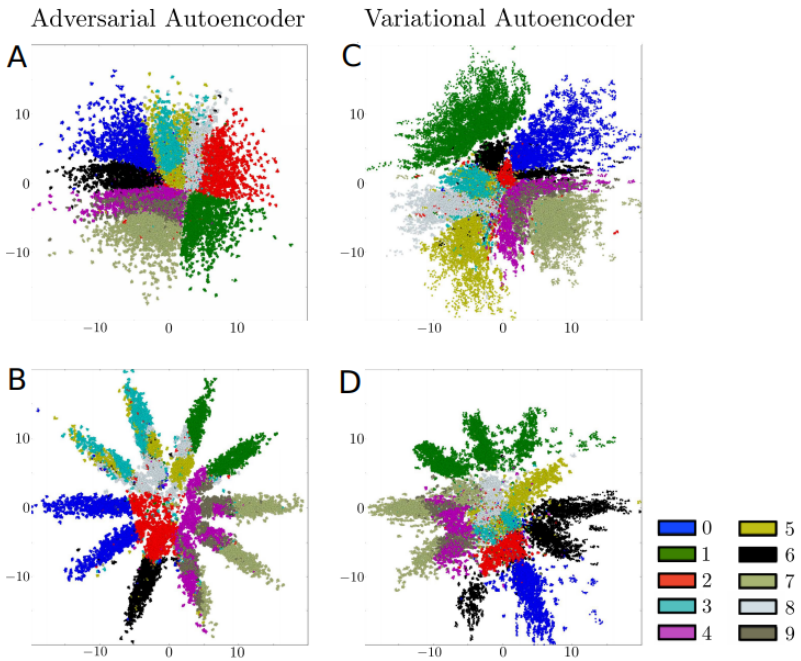

Figure 5: Posterior distributions of the latent vectors - 2D Gaussian and a mixture of 10 2D Gaussians

Imposing a prior distribution - ctd.

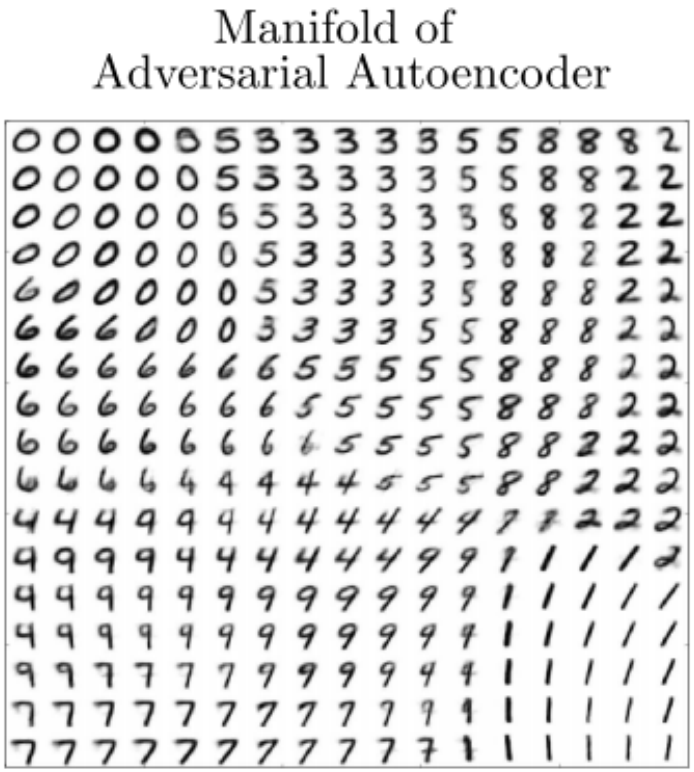

Figure 6: Manifold of the posterior distribution - images generated by taking samples of the 2D Gaussian from (a)

Incorporating label information

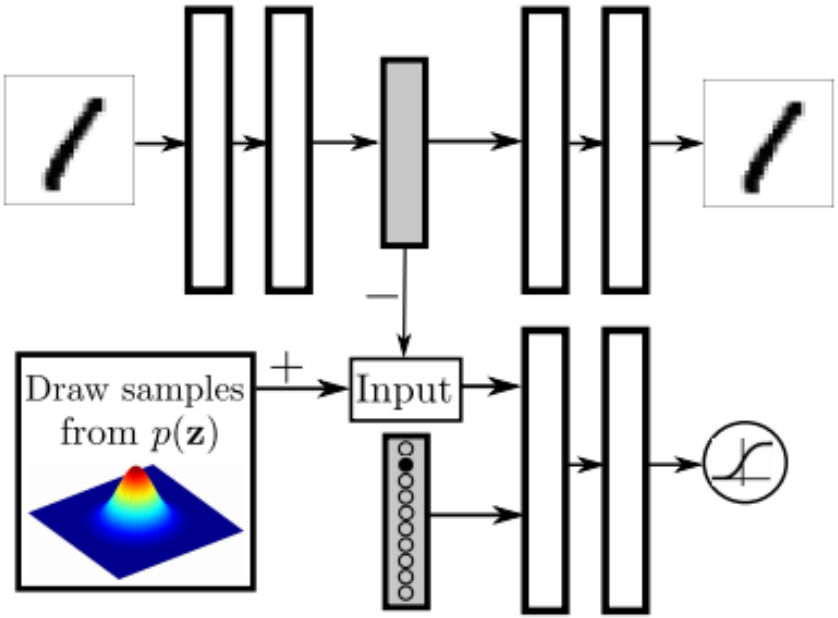

Figure 7: Providing label information to the discriminative network

Incorporating label information - ctd.

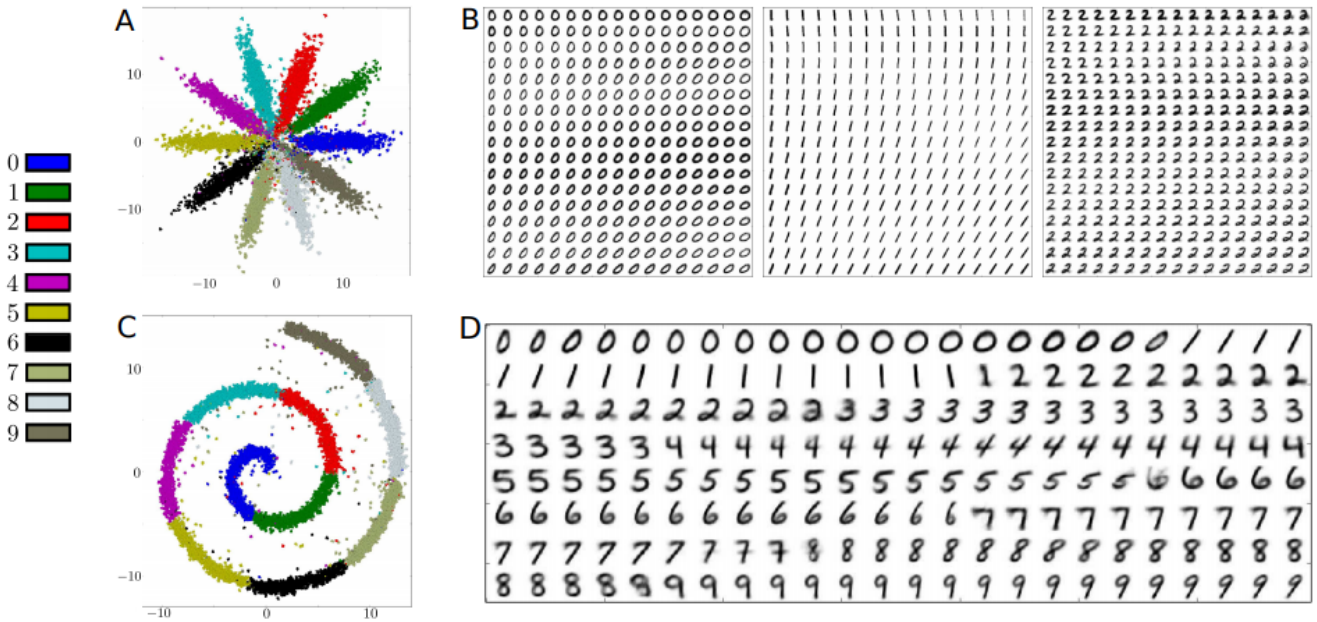

Figure 8: Imposing a distribution with forced labels

Supervised AAE

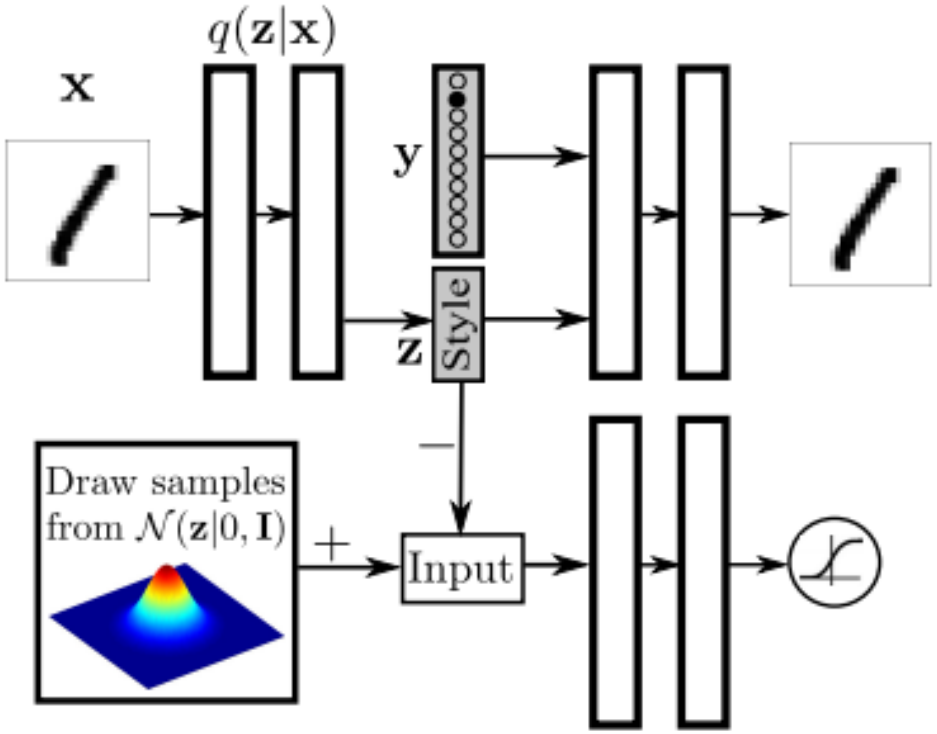

Figure 9: Architecture of a supervised AAE - providing a one-hot vector to the generative model

Supervised learning

Figure 10: Disentangling content and style (15-D Gaussian) on MNIST and SVHN datasets.

Semi-supervised AAE

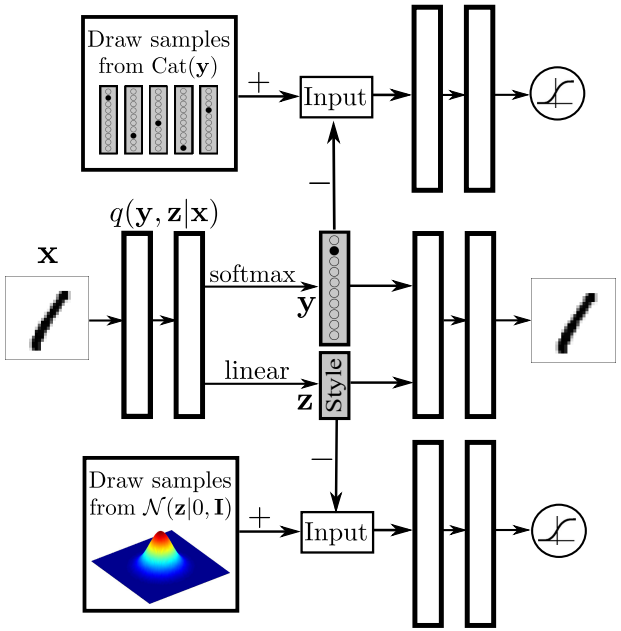

Figure 11: Semi-Supervised AAE: the top adversarial network imposes a Categorical distribution on the label representation and the bottom adversarial network imposes a Gaussian distribution on the style representation. $q(\mathbf{y}|\mathbf{x})$ is trained on the labeled data in the semi-supervised settings.

Semi-Supervised learning comparison

	MNIST (100)	MNIST (1000)	MNIST (All)	SVHN (1000)
NN Baseline	25.80	8.73	1.25	47.50
VAE (M1) + TSVM	11.82 ($\pm$0.25)	4.24 ($\pm$0.07)	-	55.33 ($\pm$0.11)
VAE (M2)	11.97 ($\pm$1.71)	3.60 ($\pm$0.56)	-	-
VAE (M1 + M2)	3.33 ($\pm$0.14)	2.40 ($\pm$0.02)	0.96	36.02 ($\pm$0.10)
VAT	2.33	1.36	0.64 ($\pm$0.04)	24.63
CatGAN	1.91 ($\pm$0.1)	1.73 ($\pm$0.18)	0.91	-
Ladder Networks	1.06 ($\pm$0.37)	0.84 ($\pm$0.08)	0.57 ($\pm$0.02)	-
ADGM	0.96 ($\pm$0.02)	-	-	16.61 ($\pm$0.24)
Adversarial Autoencoders	1.90 ($\pm$0.10)	1.60 ($\pm$0.08)	0.85 ($\pm$0.02)	17.70 ($\pm$0.30)

Table 1: Semi-supervised classification performance (error-rate) on MNIST and SVHN.

Unsupervised Clustering

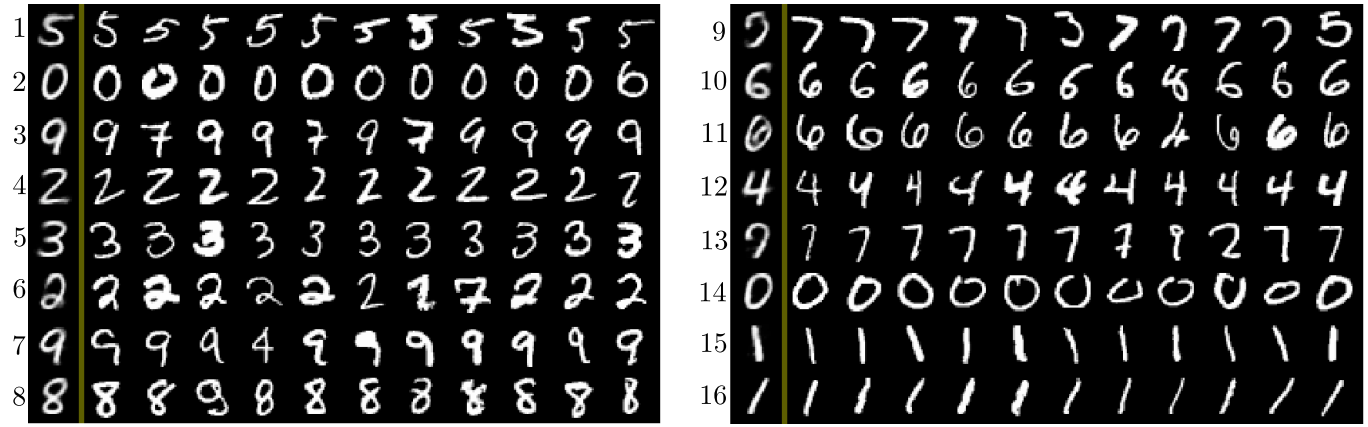

Figure 12: Unsupervised clustering of MNIST using the AAE with 16 clusters. Each row corresponds to one cluster with the first image being the cluster head.

Dimensionality Reduction

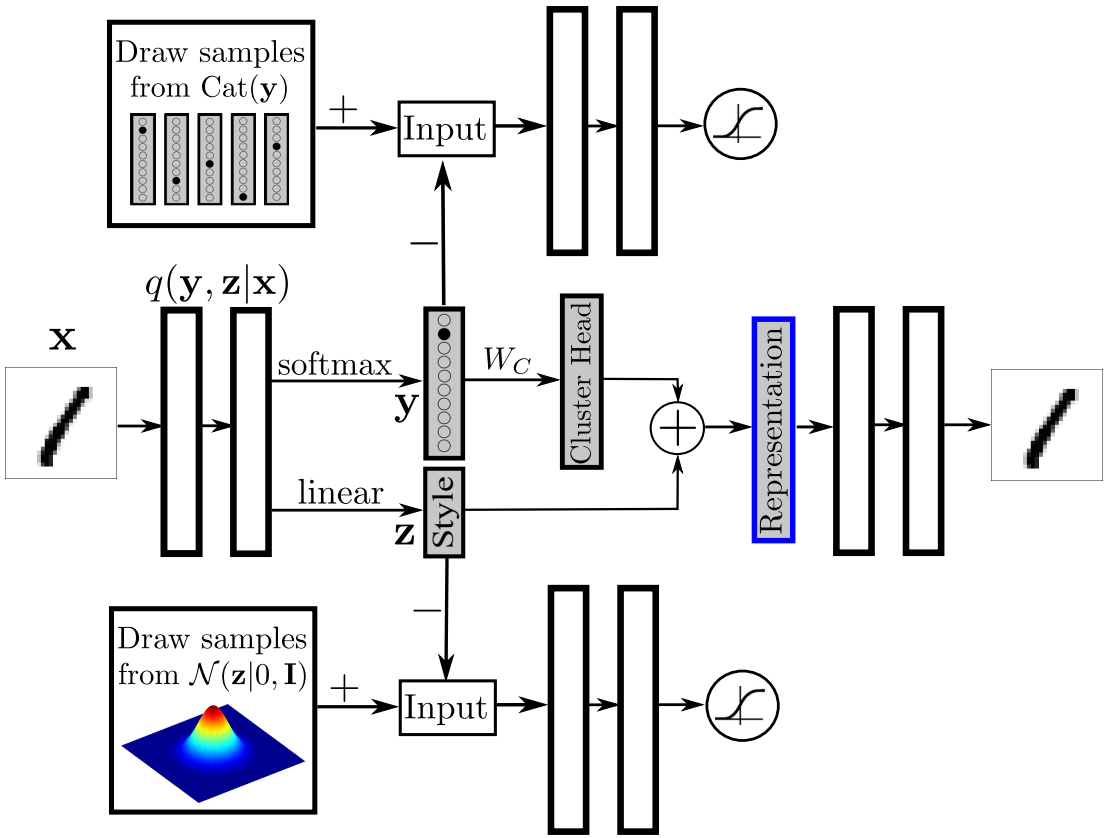

Figure 13: AAE architecture for Dimensionality Reduction

Dimensionality Reduction on MNIST

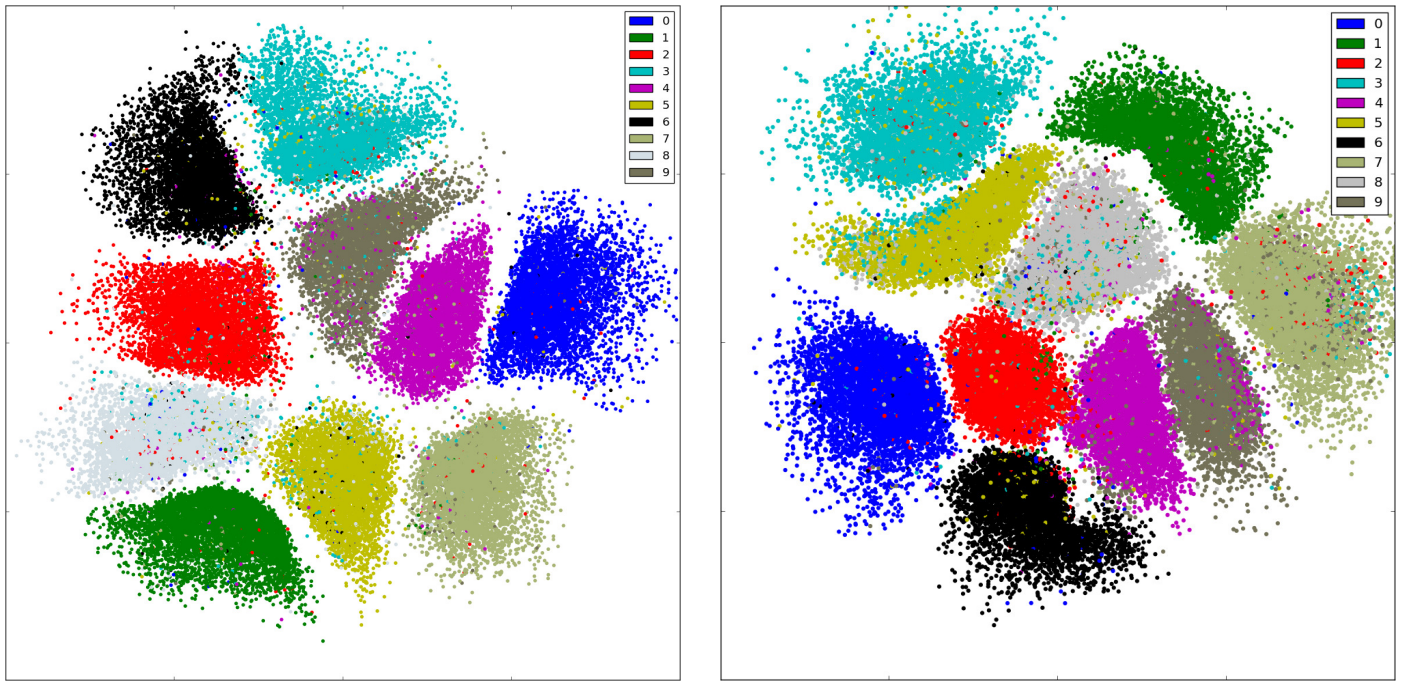

Figure 14: Semi-Supervised Dimensionality Reduction with AAE on MNIST.

Dimensionality Reduction on MNIST

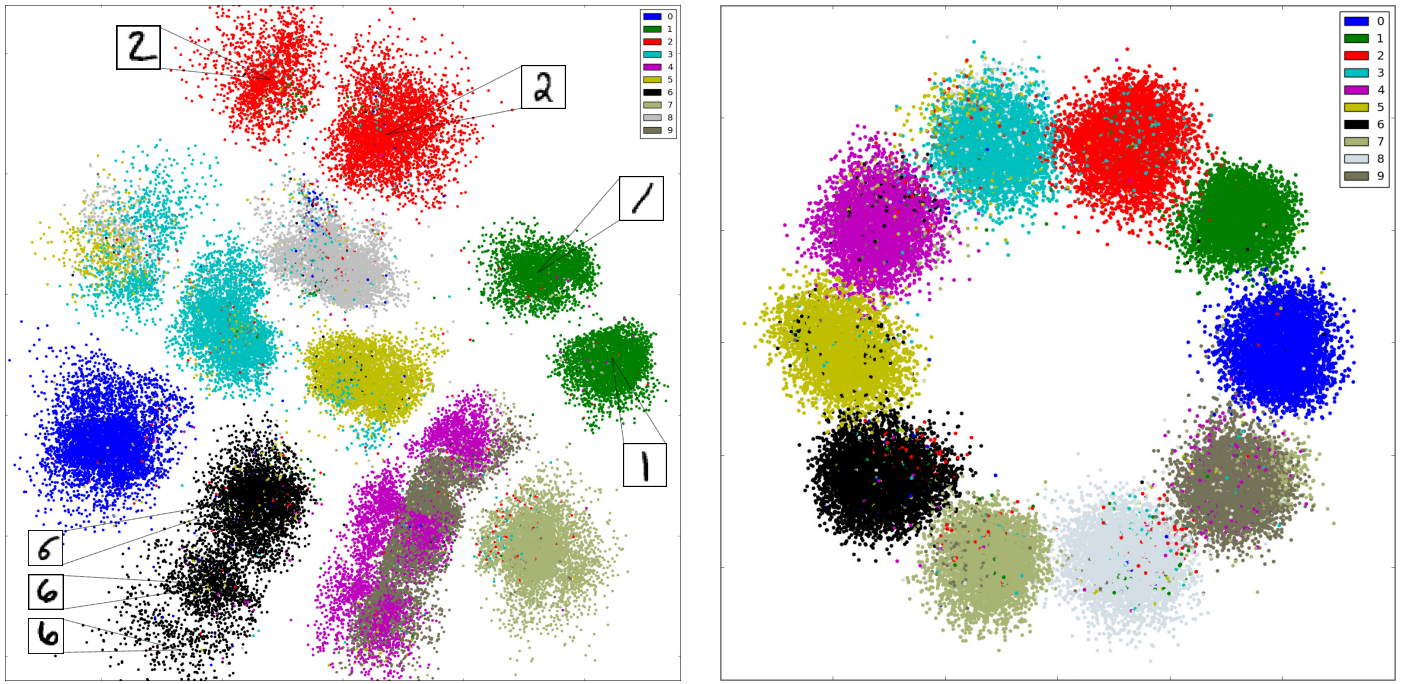

Figure 15: Semi-Supervised and Unsupervised Dimensionality Reduction with AAE on MNIST.

Interpolations in latent space

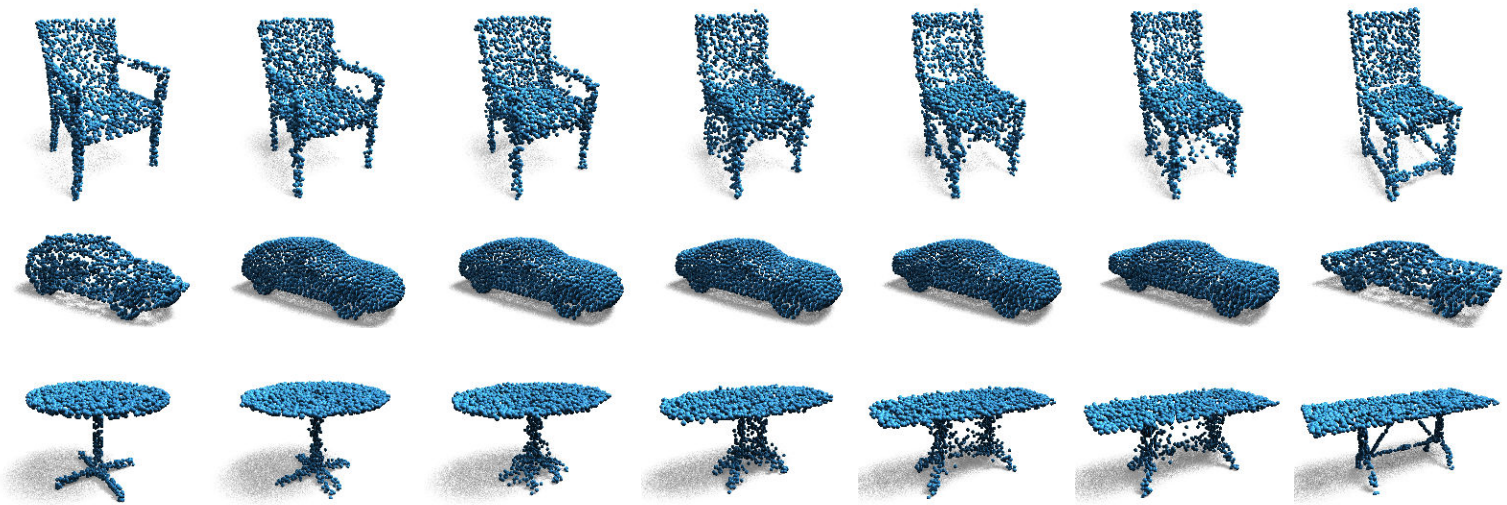

Figure 16: Interpolations in latent space

Interpolations in latent space

Figure 17: Porsche Macan S '14

Interpolations in latent space

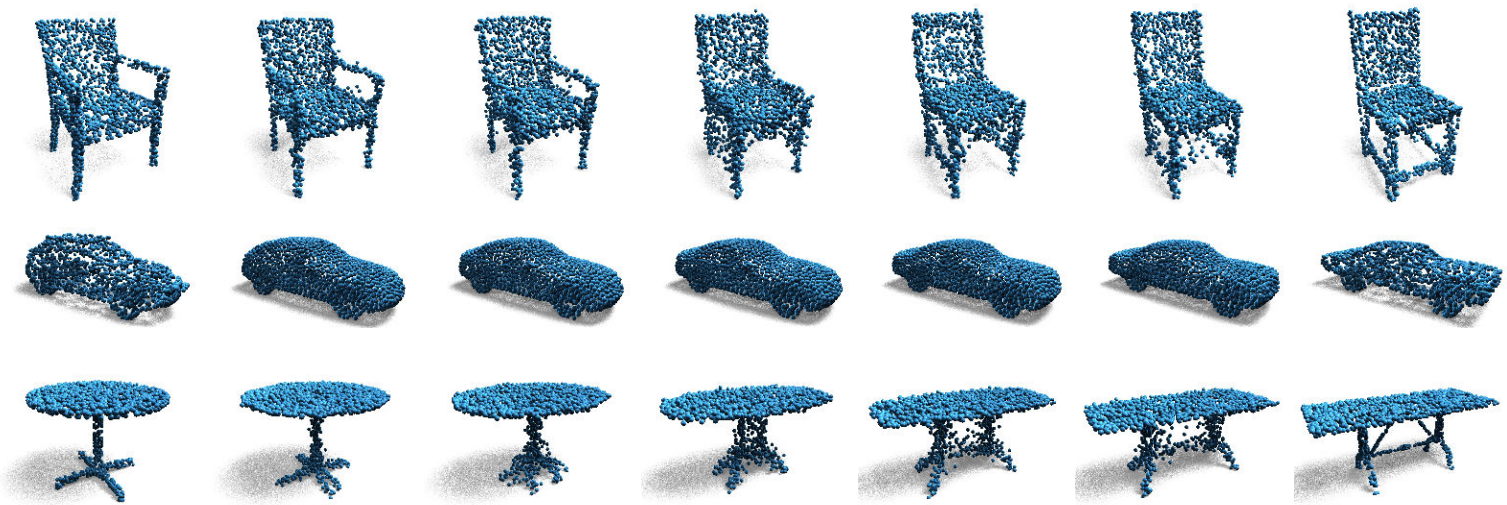

Figure 18: Interpolations in latent space

Interpolations in latent space

Figure 19: Porsche 911 '97

Interpolations in latent space

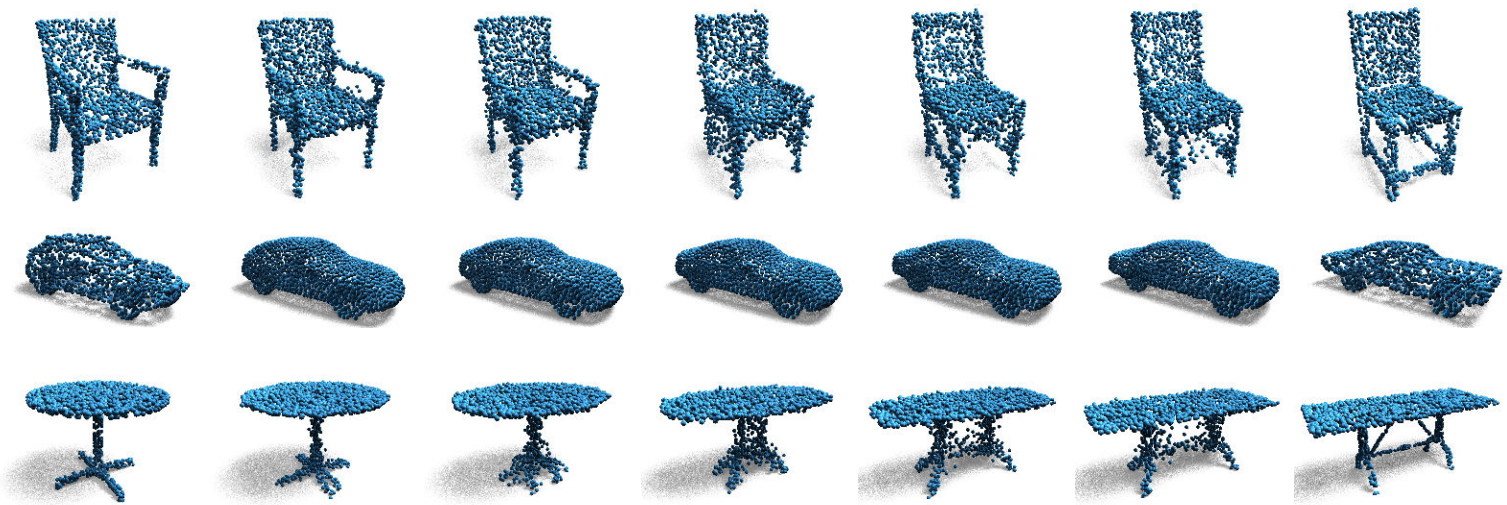

Figure 20: Interpolations in latent space

Interpolations in latent space

Figure 21: Ford Mustang '67

Algebraic operations in latent space

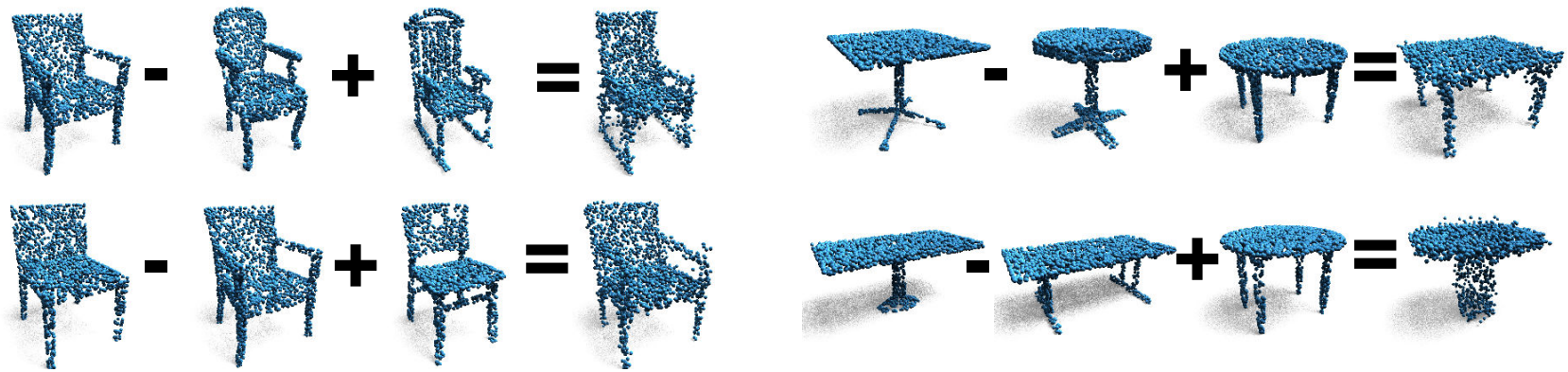

Figure 22: Algebra on latent space

Clustering

Figure 23: Clustering by 3d-AAE

References


Alireza Makhzani et al., *Adversarial Autoencoders*, https://arxiv.org/abs/1511.05644, 2015


Maciej Zamorski et al., *Adversarial Autoencoders for Compact Representations of 3D Point Clouds* https://arxiv.org/pdf/1811.07605.pdf, 2019.

Pictures sources

- Porsche 911
- Ford Mustang
- Macans
- GAN's structure
- VAE's structure
- AAE's structure