Implicit Autoencoders

Alireza Makhzani

Vector Institute for Artificial Intelligence
University of Toronto

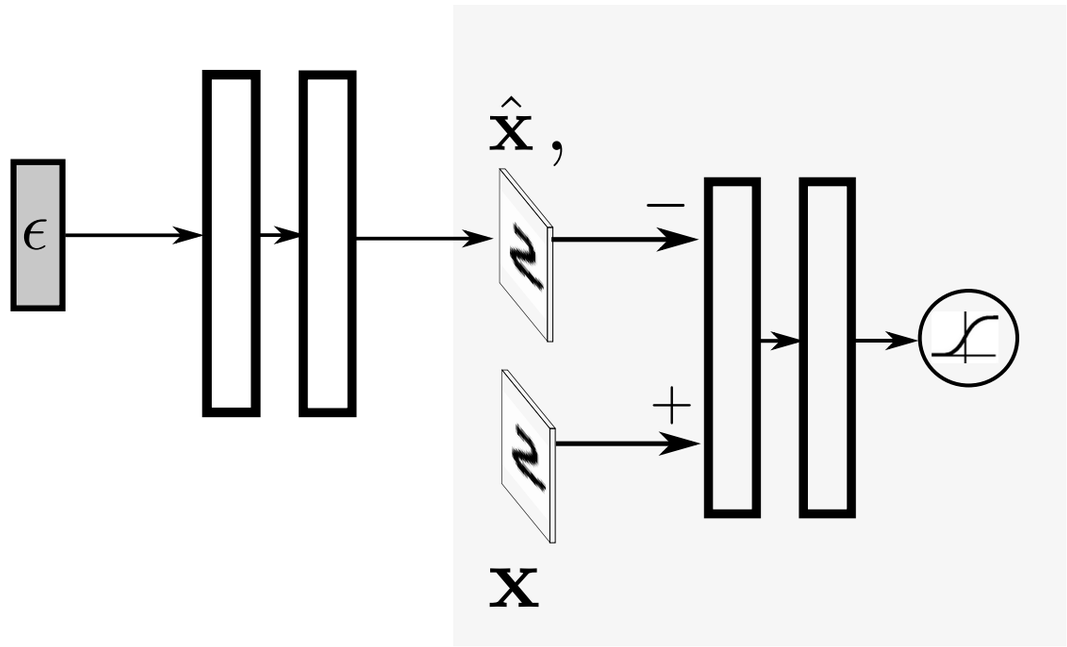
August 7th, 2018

Outline

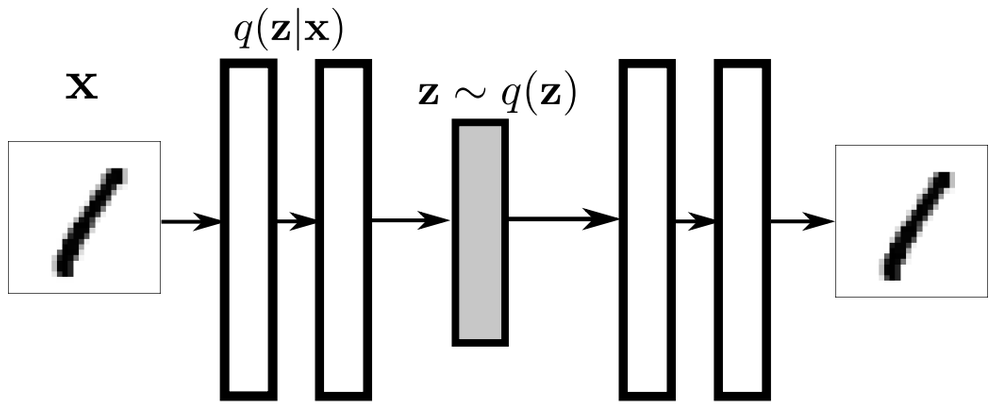
- **Background**
- **Adversarial Autoencoders**
- **Implicit Autoencoders**

Implicit Distributions

- Implicit distributions: densities obtained by passing a noise vector through a neural network.
- Densities are not tractable; but can be easily sampled from.
- Implicit distributions were initially only being used in generative modelling with GANs, but recently have found many more applications such as variational inference.



Variational Autoencoders



$$\mathbb{E}_{\mathbf{x}\sim p_{\text{data}}(\mathbf{x})}[\log p(\mathbf{x})] \geq -\underbrace{\mathbb{E}_{\mathbf{x}\sim p_{\text{data}}(\mathbf{x})}\Big[\mathbb{E}_{q(\mathbf{z}|\mathbf{x})}[-\log p(\mathbf{x}|\mathbf{z})]\Big]}_{\text{VAE Reconstruction}} - \underbrace{\mathbb{E}_{\mathbf{x}\sim p_{\text{data}}(\mathbf{x})}\Big[\mathrm{KL}(q(\mathbf{z}|\mathbf{x})\|p(\mathbf{z}))\Big]}_{\text{VAE Regularization}}$$

1. **The posterior is a factorized Gaussian.**
   - Normalizing flows.
   - MCMC-based methods.
   - Implicit Distributions: **Adversarial Autoencoders**
2. **The conditional likelihood is a factorized distribution.**
   - Autoregressive Decoders: PixelVAE, VLAE, PixelGAN Autoencoders
   - Implicit Distributions: **Implicit Autoencoders**

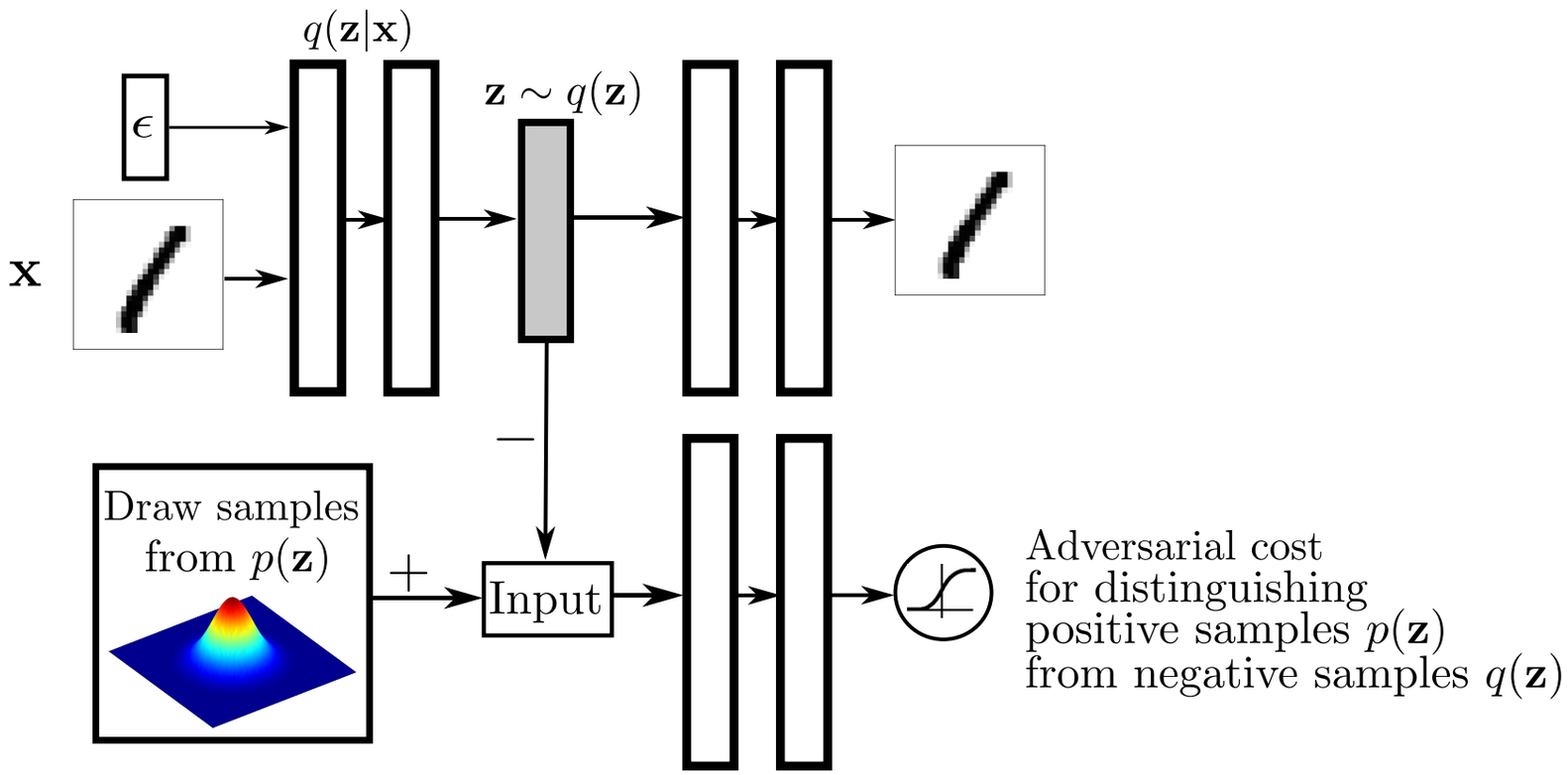
Kingma et al., 2013

Outline

- Background
- **Adversarial Autoencoders**
- Implicit Autoencoders

Adversarial Autoencoders

$$q(\mathbf{z}) = \int_{\mathbf{x}} q(\mathbf{z}|\mathbf{x}) p_{\text{data}}(\mathbf{x}) d\mathbf{x}$$



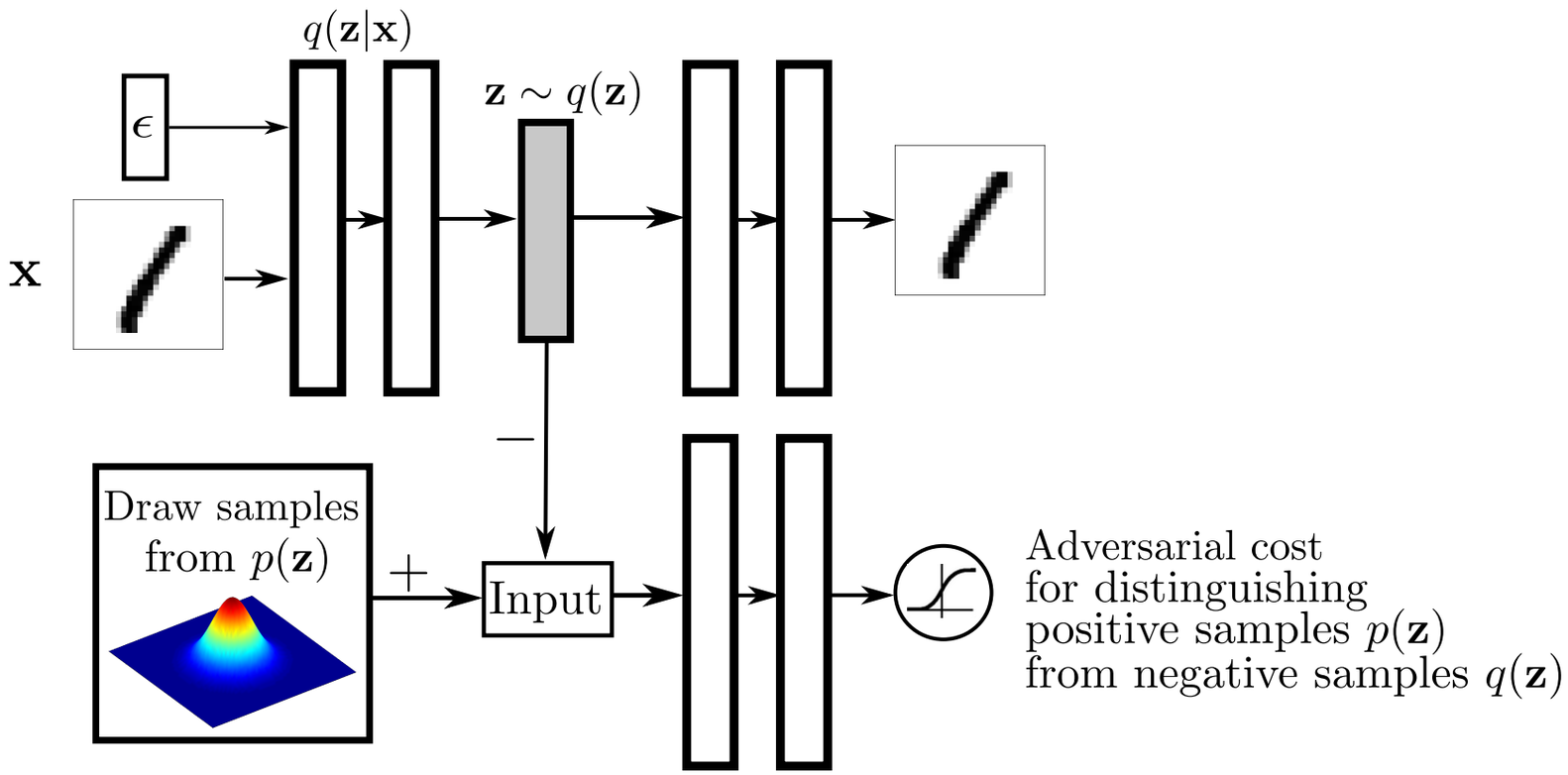
$$\mathbb{E}_{\mathbf{x}\sim p_{\text{data}}(\mathbf{x})}[\log p(\mathbf{x})] \geq -\underbrace{\mathbb{E}_{\mathbf{x}\sim p_{\text{data}}(\mathbf{x})}\Big[\mathbb{E}_{q(\mathbf{z}|\mathbf{x})}[-\log p(\mathbf{x}|\mathbf{z})]\Big]}_{\text{VAE Reconstruction}} - \underbrace{\mathbb{E}_{\mathbf{x}\sim p_{\text{data}}(\mathbf{x})}\Big[\text{KL}(q(\mathbf{z}|\mathbf{x})\|p(\mathbf{z}))\Big]}_{\text{VAE Regularization}} \quad (4)$$

$$= -\underbrace{\mathbb{E}_{\mathbf{x}\sim p_{\text{data}}(\mathbf{x})}\Big[\mathbb{E}_{q(\mathbf{z}|\mathbf{x})}[-\log p(\mathbf{x}|\mathbf{z})]\Big]}_{\text{AAE Reconstruction}} - \underbrace{\text{KL}(q(\mathbf{z})\|p(\mathbf{z}))}_{\text{AAE Regularization}} - \underbrace{\mathcal{I}(\mathbf{z};\mathbf{x})}_{\text{Mutual Info.}} \quad (5)$$

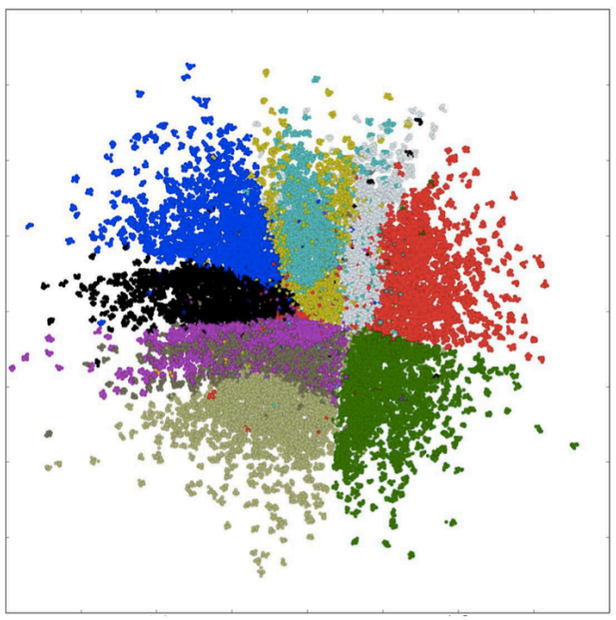
Makhzani et al., 2016

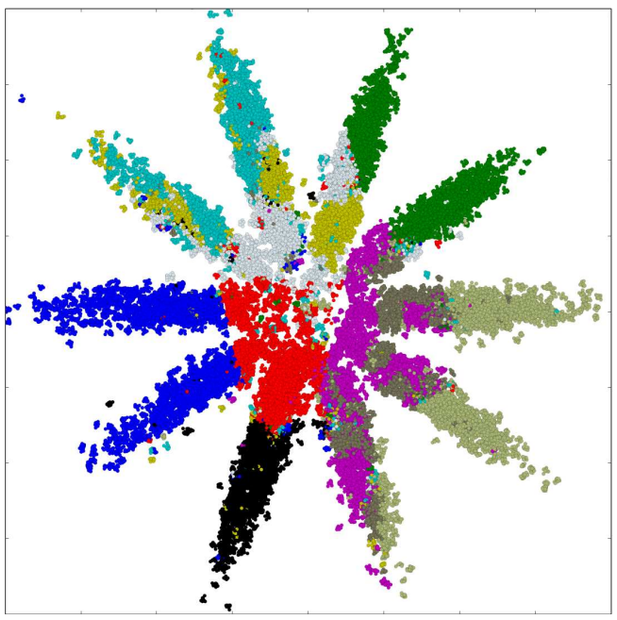
Adversarial Autoencoders

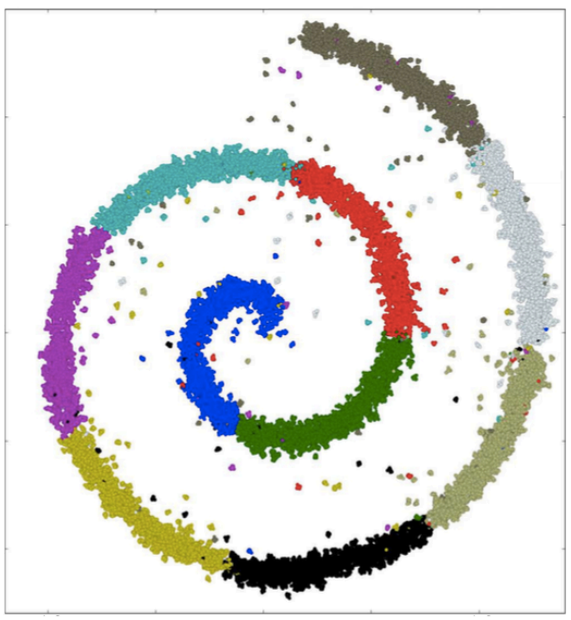
$$q(\mathbf{z}) = \int_{\mathbf{x}} q(\mathbf{z}|\mathbf{x}) p_{\text{data}}(\mathbf{x}) d\mathbf{x}$$



Code Space of MNIST:







Prior: Gaussian | Gaussian Mixture | Swiss Roll

Adversarial Variational Bayes

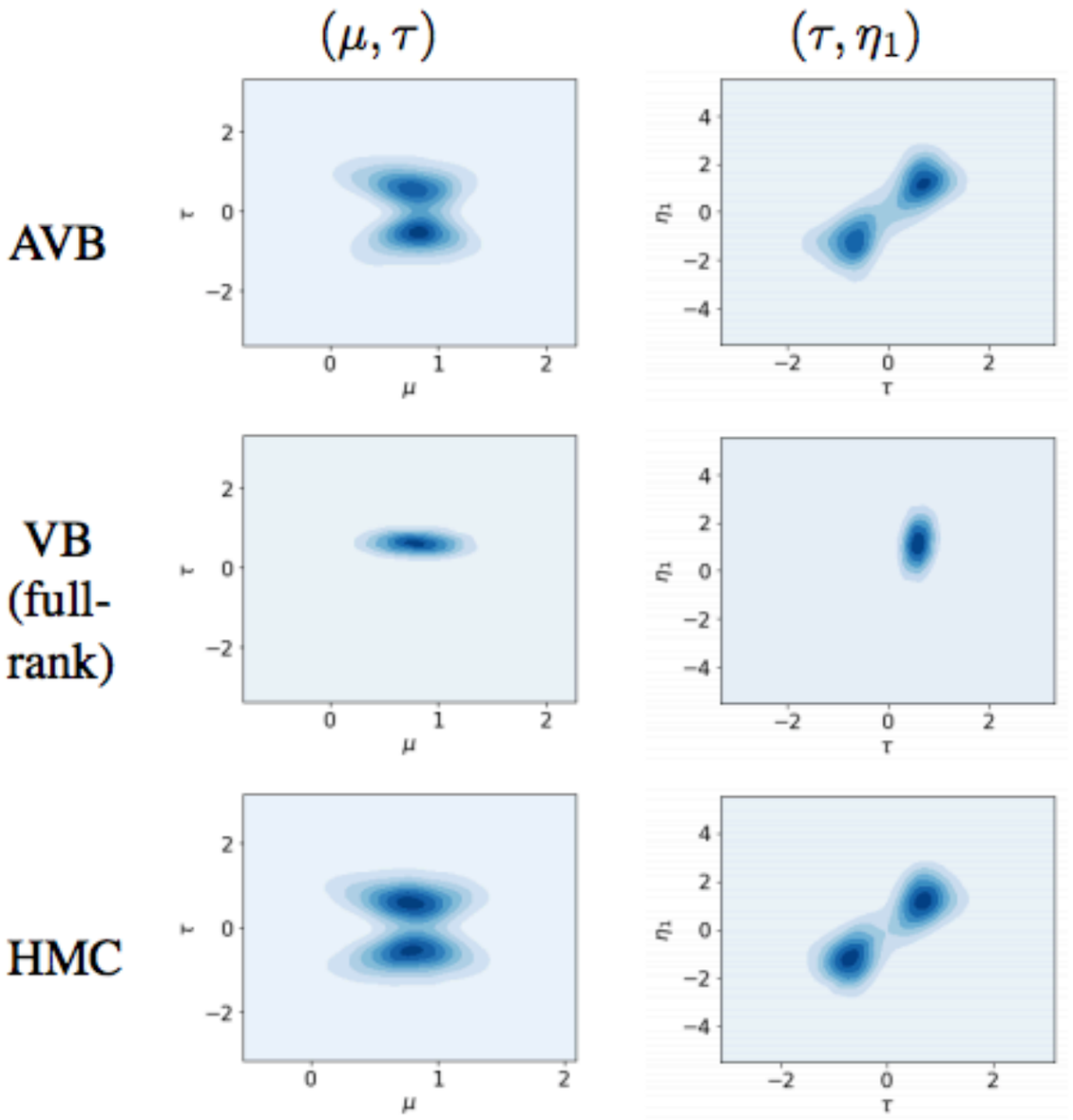
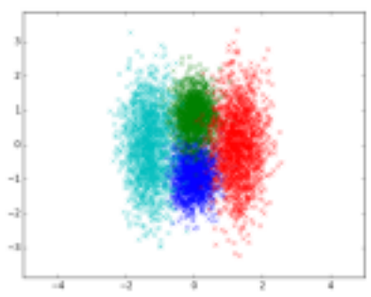
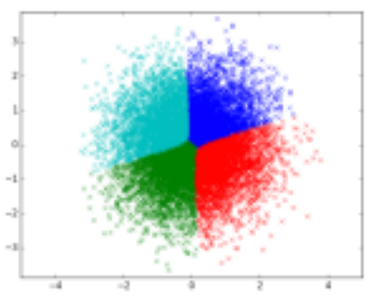


Figure 5. Training examples in the synthetic dataset.



(a) VAE

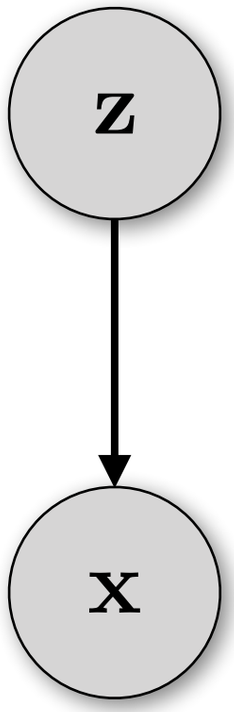


(b) AVB

Mescheder et al., 2017

Limitations of Adversarial Autoencoders

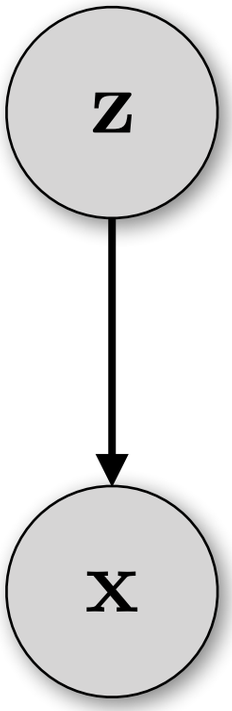
- AAE does not optimize a principled objective.
- All the image statistics are captured by the single latent vector.



	AAE/VAE	
$p(\mathbf{z})$	**Latent Variable**	label, style global and local
$p(\mathbf{x}\|\mathbf{z})$	**Deterministic** **(factorized Gaussians)**	None

Implicit Autoencoders (IAE)

- The image statistics are captured jointly by the latent vector and the implicit decoder distribution.
- IAE optimizes the ELBO.

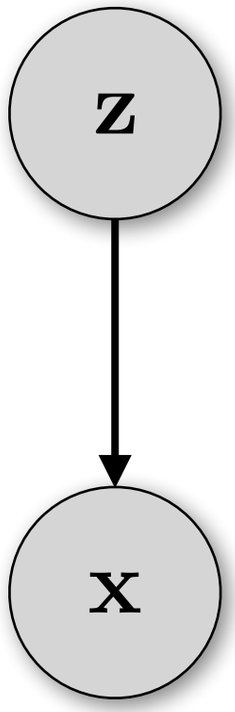


$p(\mathbf{z})$ **Latent Variable**

$p(\mathbf{x}|\mathbf{z})$ **Implicit Likelihood**

Implicit Autoencoders (IAE)

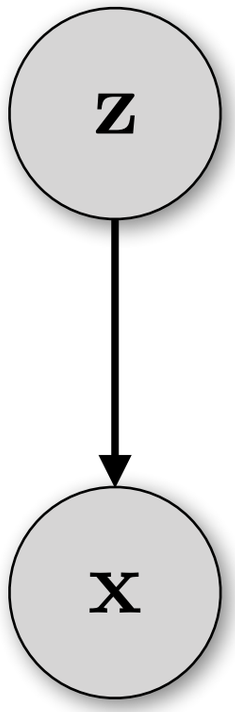
- The image statistics are captured jointly by the latent vector and the implicit decoder distribution.
- IAE optimizes the ELBO.



		IAE (Gaussian)	IAE (Categorical)
$p(\mathbf{z})$	**Latent Variable**	Global (high-level)	Discrete (label)
$p(\mathbf{x}\|\mathbf{z})$	**Implicit Likelihood**	Local (low-level)	Continuous (Style)

Implicit Autoencoders (IAE)

- The image statistics are captured jointly by the latent vector and the implicit decoder distribution.
- IAE optimizes the ELBO.



$p(\mathbf{z})$ **Latent Variable**

$p(\mathbf{x}|\mathbf{z})$ **Implicit Likelihood**

	IAE (Gaussian)	**IAE (Categorical)**
	Global (high-level)	Discrete (label)
	Local (low-level)	Continuous (Style)

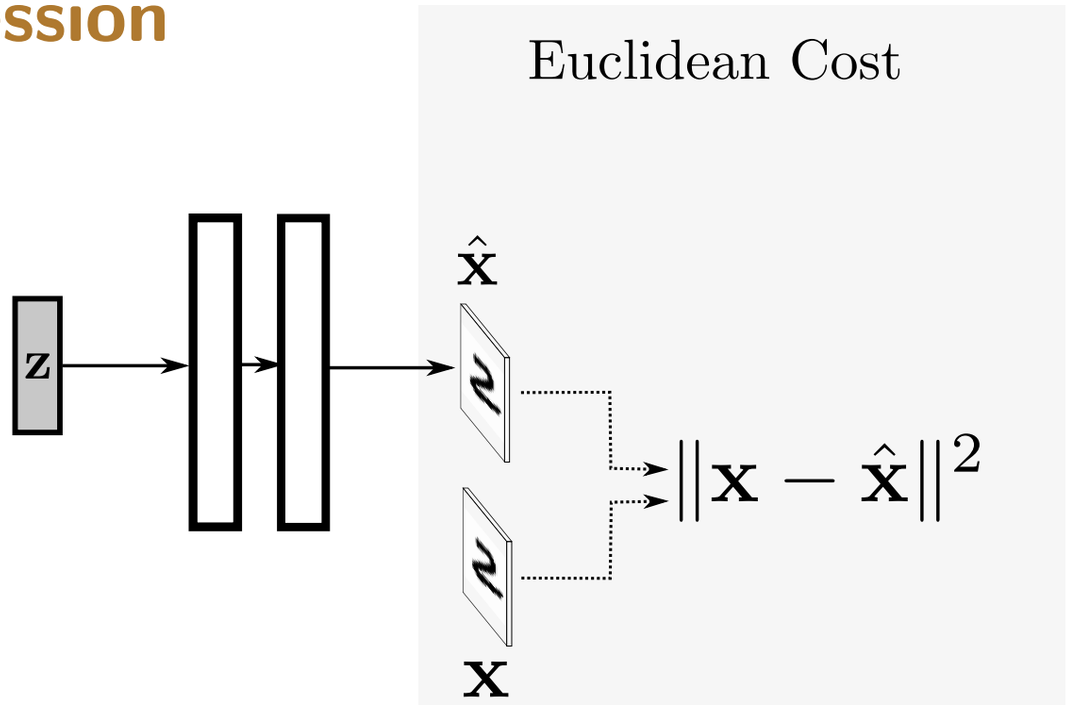
Semi-supervised Learning

Outline

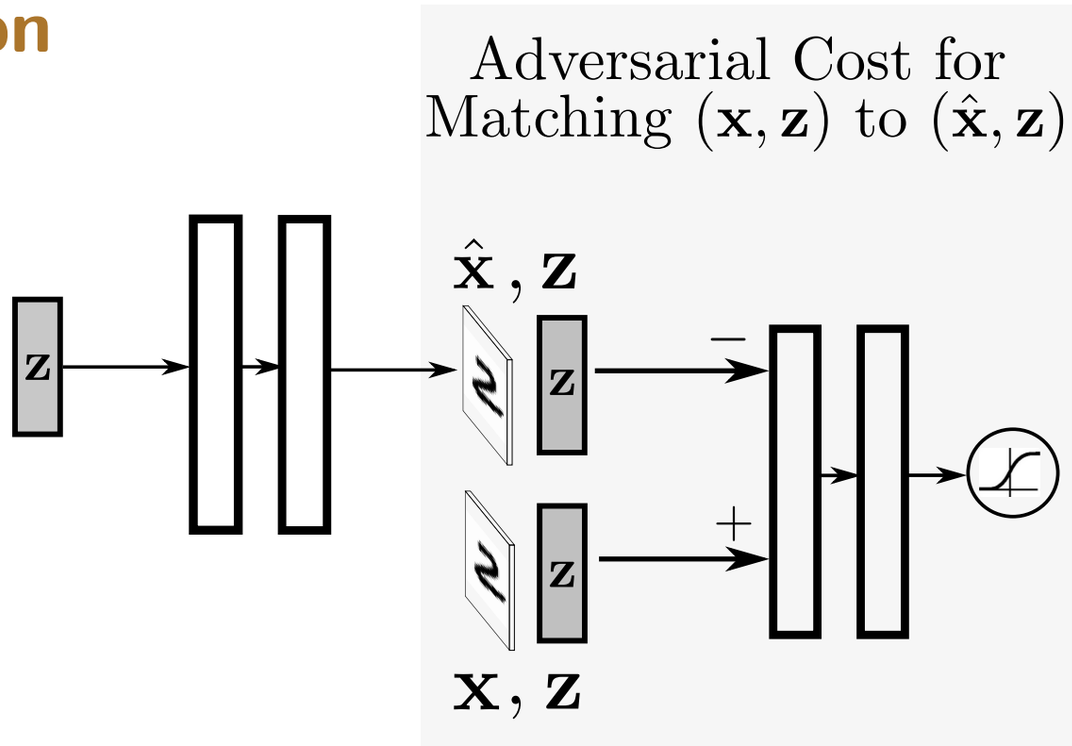
- Background
- Adversarial Autoencoders
- **Implicit Autoencoders**

Regression

- **Least Squares Regression**

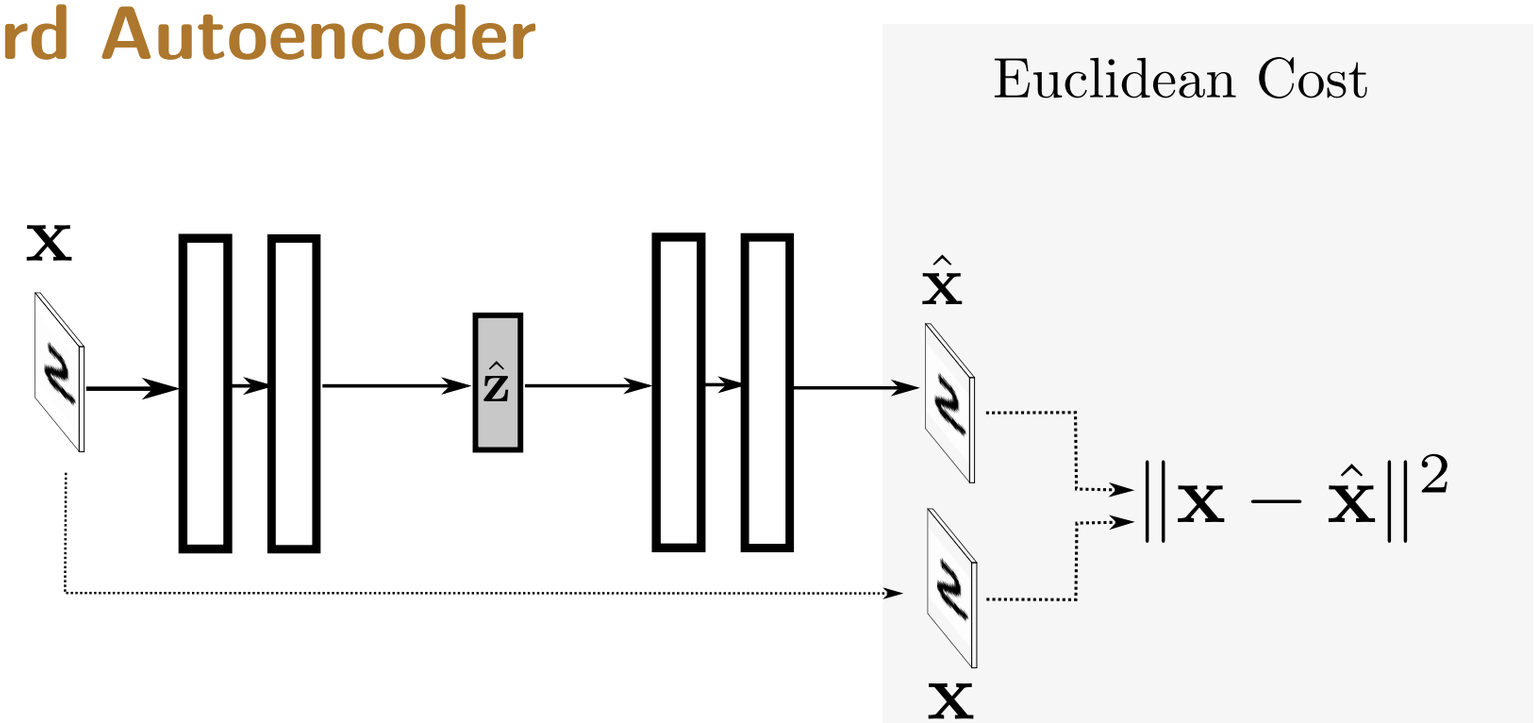


- **Adversarial Regression**

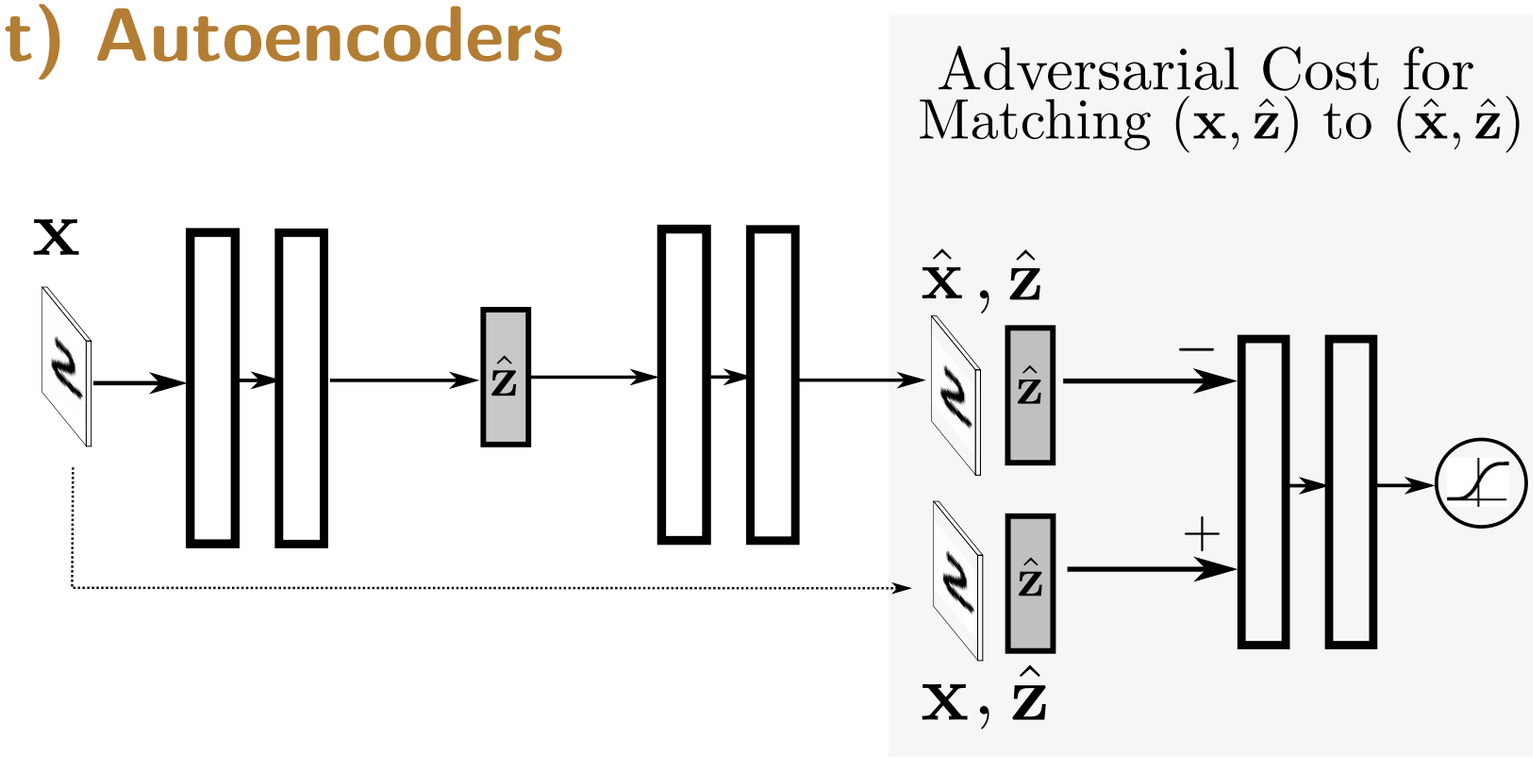


Autoencoders

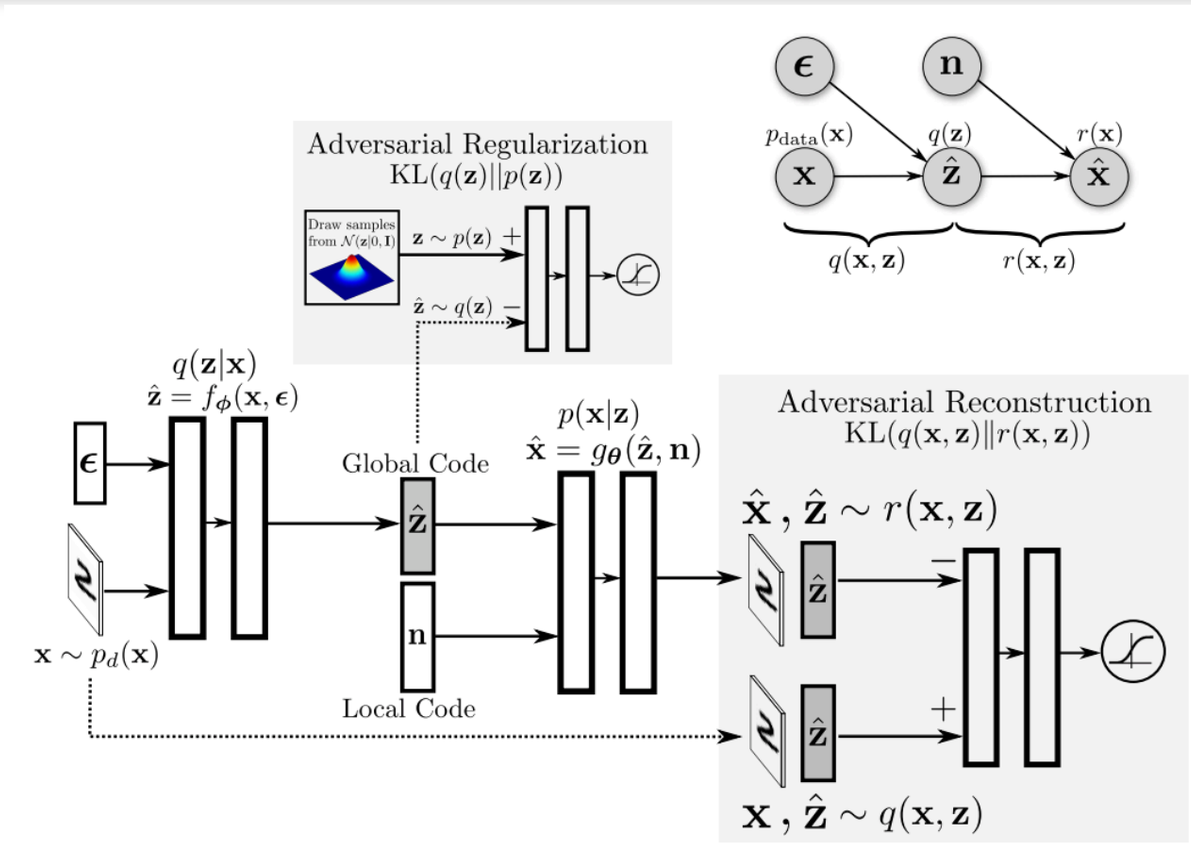
- **Standard Autoencoder**



- **(Implicit) Autoencoders**



Implicit Autoencoders



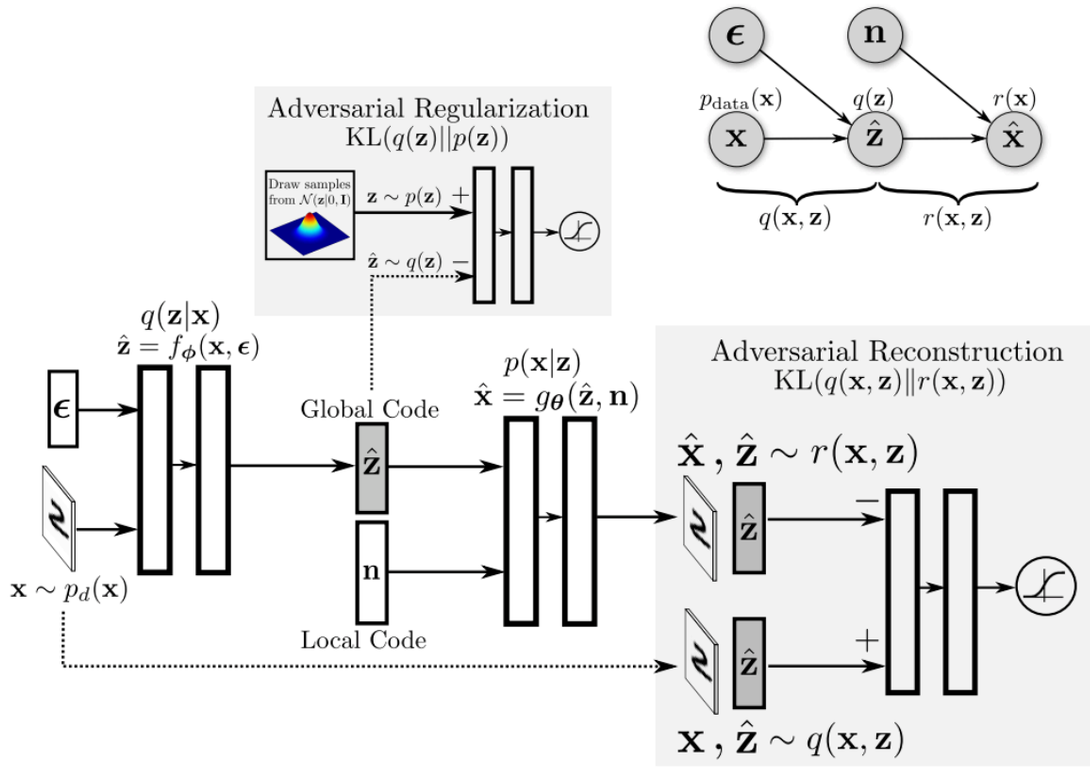
Distributions over z: $p(z)$ prior — $q(z) = \int p_d(x)q(z\,|\,x)dx$ aggregated posterior

Distributions over x: $p_d(x)$ data dist. — $p(x) = \int p(z)p(x\,|\,z)dz$ model dist. — $r(x) = \int q(z)p(x\,|\,z)dz$ agg. reconstruction dist.

Distributions over (x,z): $p(x,z) = p(z)p(x\,|\,z)$ joint model dist. — $q(x,z) = q(z\,|\,x)p_d(x)$ joint data dist. — $r(x,z) = q(z)p(x\,|\,z)$ joint reconst. dist.

Conditionals of (x,z): $p(z\,|\,x)$ true post. — $q(z\,|\,x)$ variational post. — $p(x\,|\,z)$ cond. likelihood — $q(x\,|\,z)$ inverse post.

Implicit Autoencoders



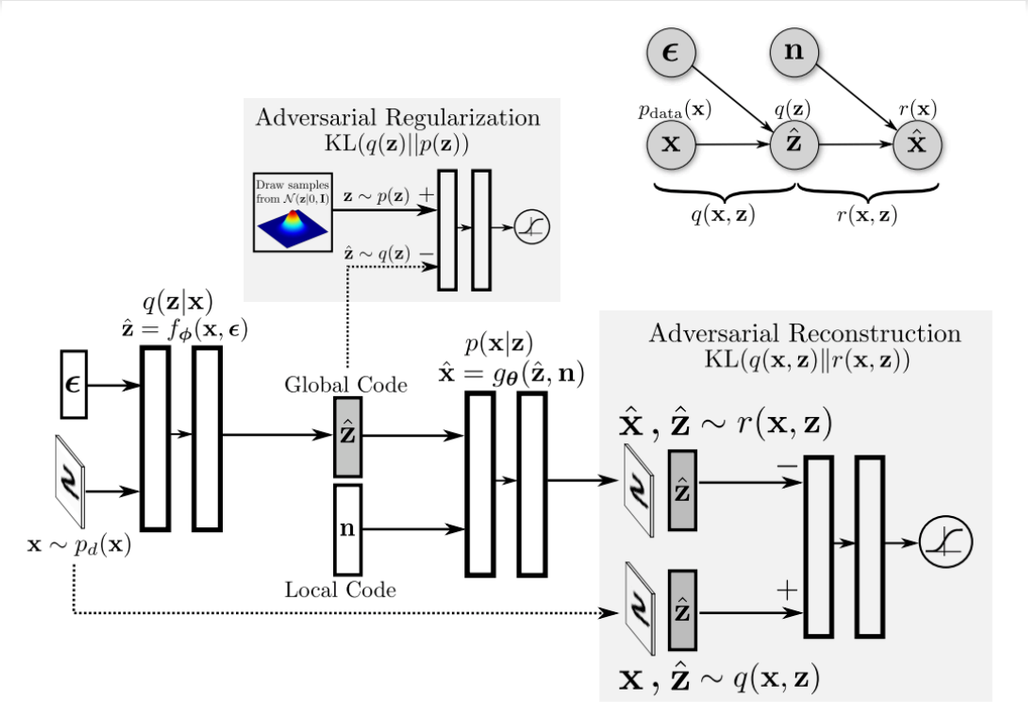
$$\mathbb{E}_{\mathbf{x}\sim p_{\text{data}}(\mathbf{x})}[\log p(\mathbf{x})] \geq \underbrace{-\mathbb{E}_{\mathbf{x}\sim p_{\text{data}}(\mathbf{x})}\Big[\mathbb{E}_{q(\mathbf{z}|\mathbf{x})}[-\log p(\mathbf{x}|\mathbf{z})]\Big]}_{\text{VAE Reconstruction}} \underbrace{-\mathbb{E}_{\mathbf{x}\sim p_{\text{data}}(\mathbf{x})}\Big[\mathrm{KL}(q(\mathbf{z}|\mathbf{x})\|p(\mathbf{z}))\Big]}_{\text{VAE Regularization}}$$

$$= \underbrace{-\mathbb{E}_{\mathbf{x}\sim p_{\text{data}}(\mathbf{x})}\Big[\mathbb{E}_{q(\mathbf{z}|\mathbf{x})}[-\log p(\mathbf{x}|\mathbf{z})]\Big]}_{\text{AAE Reconstruction}} \underbrace{-\mathrm{KL}(q(\mathbf{z})\|p(\mathbf{z}))}_{\text{AAE Regularization}} \underbrace{-\mathcal{I}(\mathbf{z};\mathbf{x})}_{\text{Mutual Info.}}$$

$$= \underbrace{-\mathbb{E}_{\mathbf{z}\sim q(\mathbf{z})}\Big[\mathrm{KL}(q(\mathbf{x}|\mathbf{z})\|p(\mathbf{x}|\mathbf{z}))\Big]}_{\text{IAE Reconstruction}} \underbrace{-\mathrm{KL}(q(\mathbf{z})\|p(\mathbf{z}))}_{\text{IAE Regularization}} \underbrace{-\mathcal{H}_{\text{data}}(\mathbf{x})}_{\text{Entropy of Data}}$$

$$= \underbrace{-\mathrm{KL}(q(\mathbf{x},\mathbf{z})\|r(\mathbf{x},\mathbf{z}))}_{\text{IAE Reconstruction}} \underbrace{-\mathrm{KL}(q(\mathbf{z})\|p(\mathbf{z}))}_{\text{IAE Regularization}} \underbrace{-\mathcal{H}_{\text{data}}(\mathbf{x})}_{\text{Entropy of Data}}$$

More on Adversarial Reconstruction



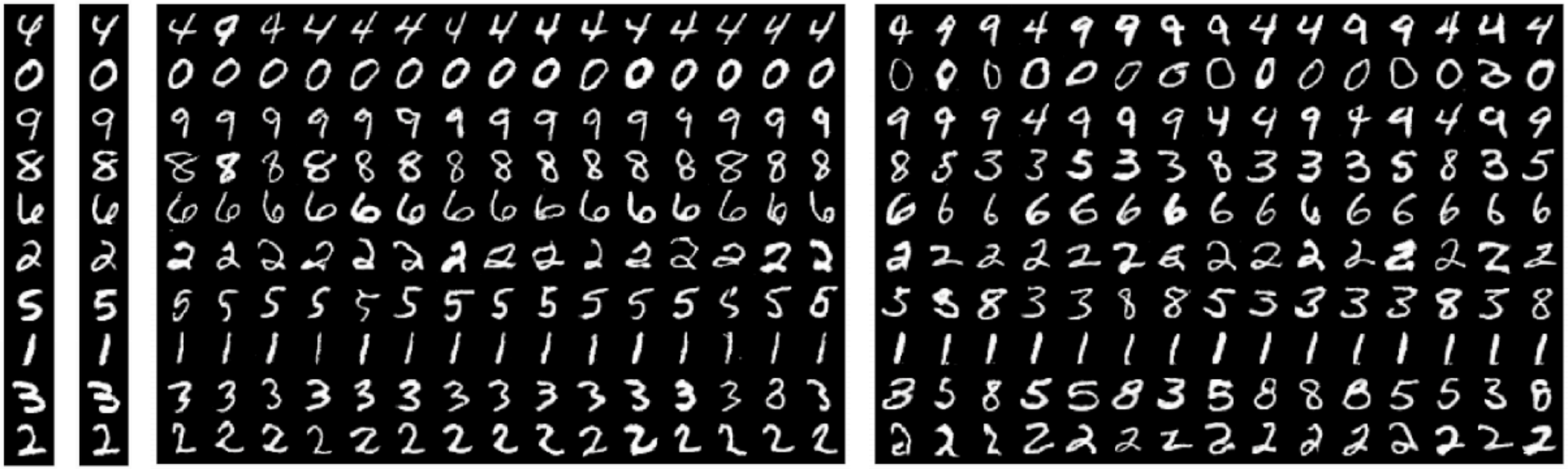
- IAE objective $= \underbrace{\mathbb{E}_{\mathbf{z}\sim q(\mathbf{z})}\Big[\mathrm{KL}(q(\mathbf{x}|\mathbf{z})\|p(\mathbf{x}|\mathbf{z}))\Big]}_{\text{IAE Reconstruction}} + \underbrace{\mathrm{KL}(q(\mathbf{z})\|p(\mathbf{z}))}_{\text{IAE Regularization}}$

$$= \underbrace{\mathrm{KL}(q(\mathbf{x},\mathbf{z})\|r(\mathbf{x},\mathbf{z}))}_{\text{IAE Reconstruction}} + \underbrace{\mathrm{KL}(q(\mathbf{z})\|p(\mathbf{z}))}_{\text{IAE Regularization}}$$

- If the encoder/decoder is deterministic: **Deterministic Reconstructions**
- If the encoder/decoder is stochastic: **Stochastic Reconstructions**
- Adversarial Reconstruction vs. Euclidean Reconstruction:

$$\underbrace{\mathbb{E}_{\mathbf{x}\sim p_{\text{data}}(\mathbf{x})}\Big[\mathbb{E}_{q(\mathbf{z}|\mathbf{x})}[-\log p(\mathbf{x}|\mathbf{z})]\Big]}_{\text{AE Reconstruction}} = \underbrace{\mathbb{E}_{\mathbf{z}\sim q(\mathbf{z})}\Big[\mathrm{KL}(q(\mathbf{x}|\mathbf{z})\|p(\mathbf{x}|\mathbf{z}))\Big]}_{\text{IAE Reconstruction}} + \underbrace{\mathcal{H}(\mathbf{x}|\mathbf{z})}_{\text{Cond. Entropy}}$$

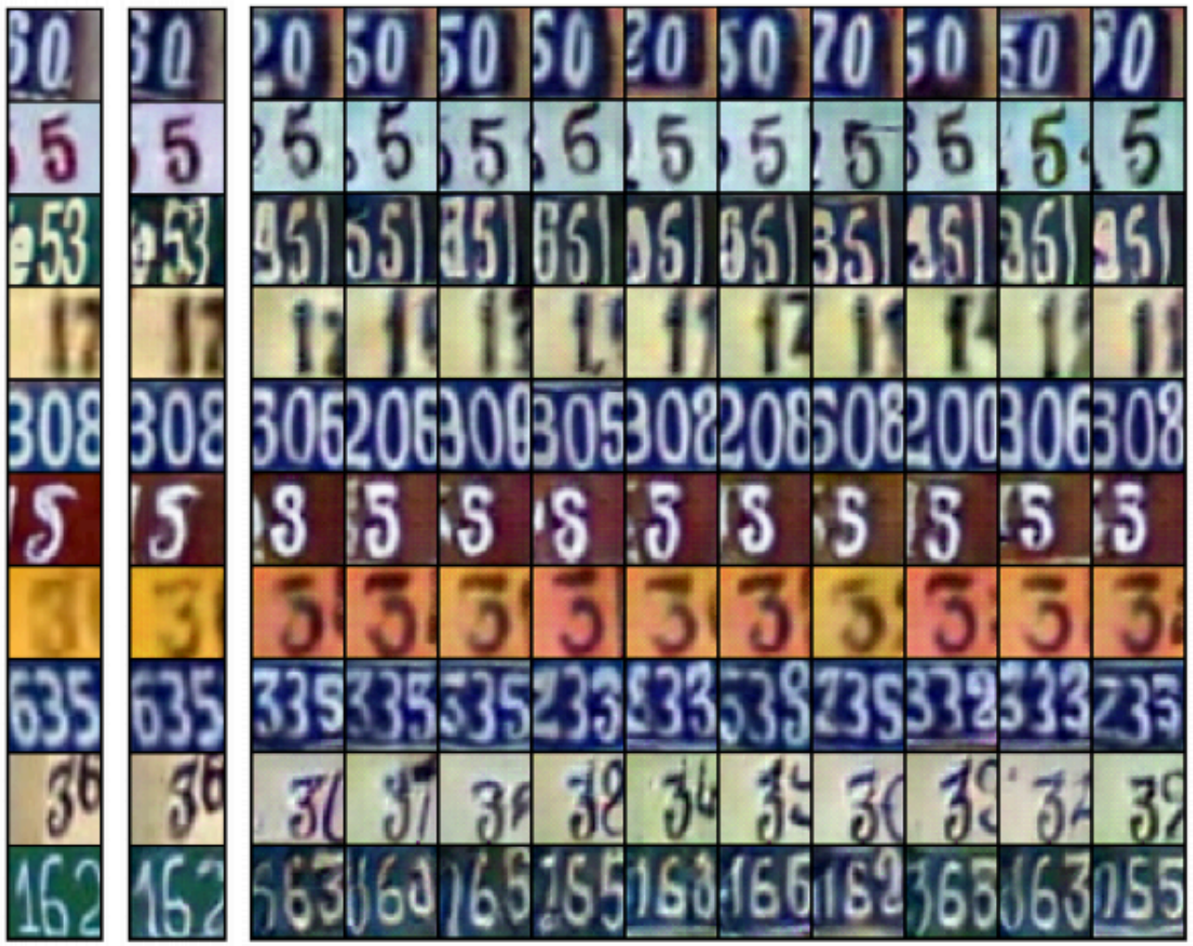
Implicit Autoencoders: MNIST



(a) (b) (c) (d)

Figure 2: MNIST dataset. (a) Original images. (b) Deterministic reconstructions with 20D global vector. (c) Stochastic reconstructions with 10D global and 100D local vector. (d) Stochastic reconstructions with 5D global and 100D local vector.

Implicit Autoencoders: SVHN and CelebA



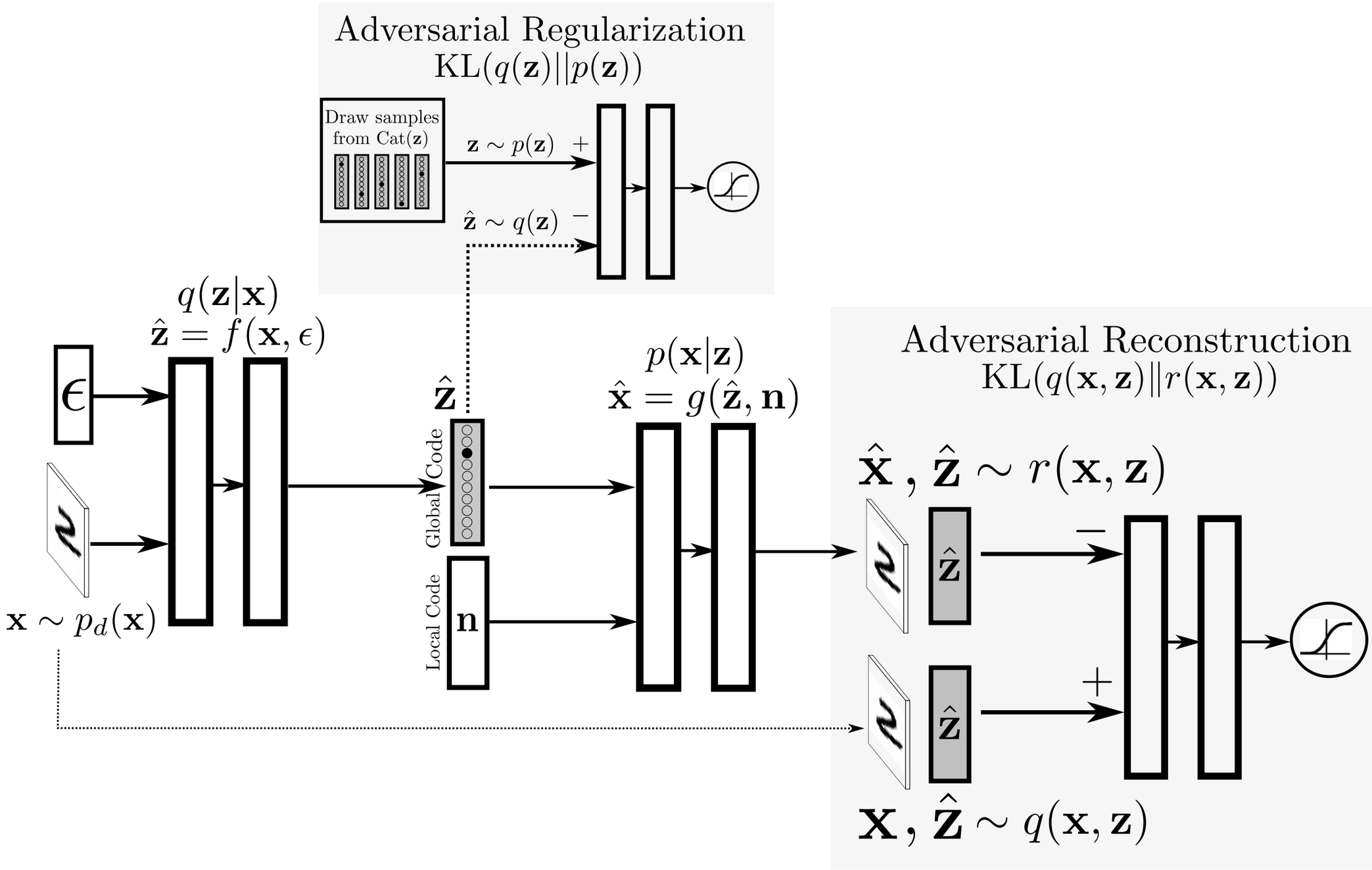
(a) (b) (c)

Figure 3: SVHN dataset. (a) Original images. (b) Deterministic reconstructions with 150D global vector. (c) Stochastic reconstructions with 75D global and 1000D local vector.

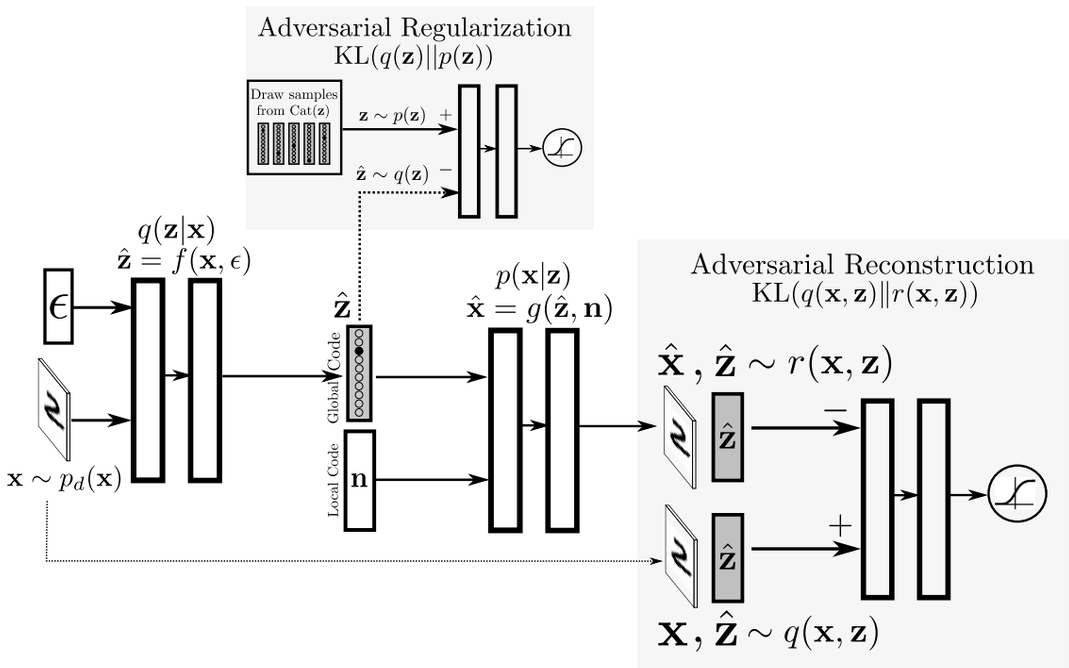
(a) (b) (c)

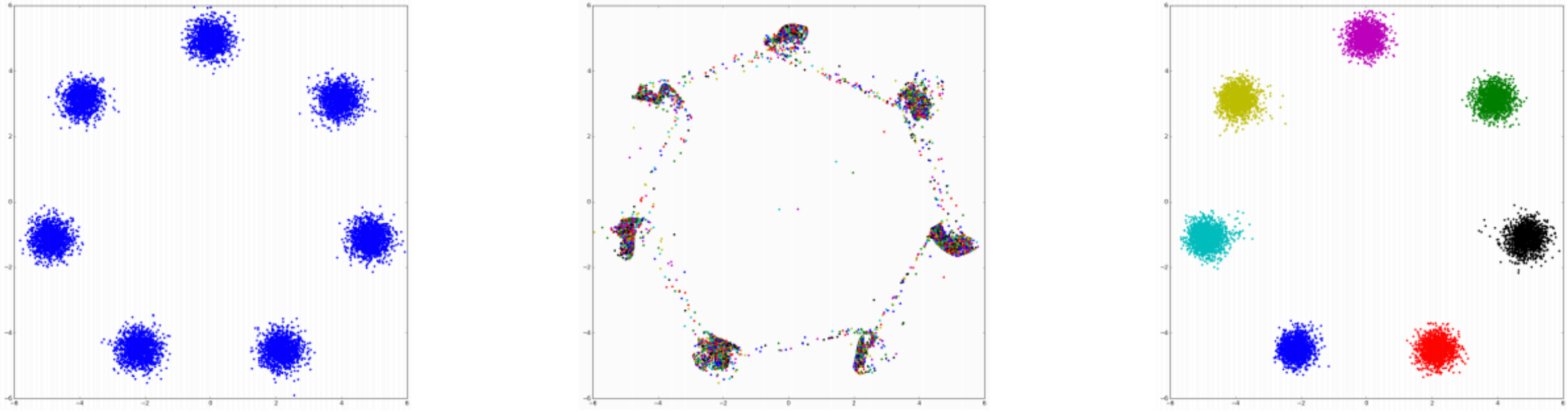
Figure 4: CelebA dataset. (a) Original images. (b) Deterministic reconstructions with 150D global vector. (c) Stochastic reconstructions with 50D global and 1000D local vector.

Clustering with IAEs



Clustering with IAEs





(a) Original data (b) GAN (c) IAE

Figure 5: Learning the mixture of Gaussian distribution by the standard GAN and the IAE.

Clustering with IAEs

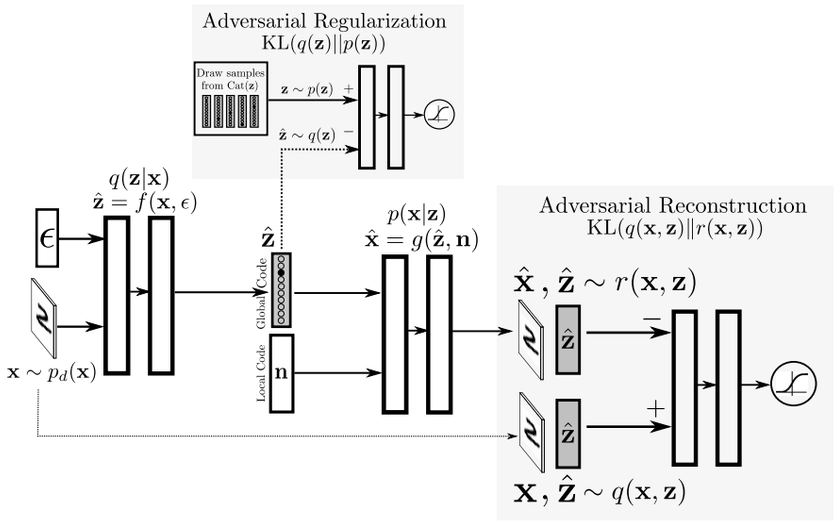
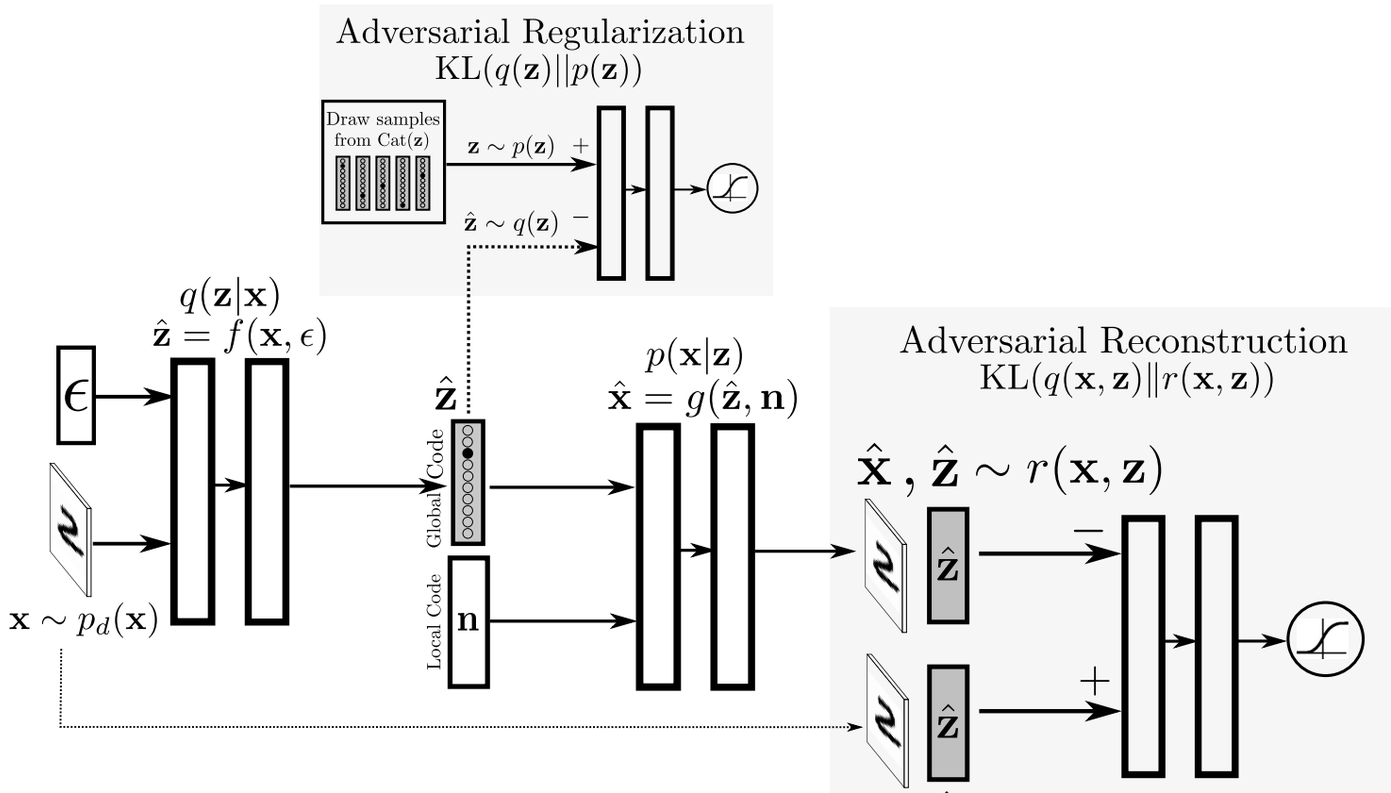


Figure 6: Disentangling the content and style of the MNIST digits in an unsupervised fashion with implicit autoencoders. Each column shows samples of the model from one of the learnt clusters. The style (local noise vector) is drawn from a Gaussian distribution and held fixed across each row.

Semi-Supervised Classification with IAEs



	MNIST (100 Labels)	SVHN (1000 Labels)
Adversarial Autoencoders	1.90%	17.70%
Improved-GAN	0.93%	8.11%
Implicit Autoencoders	1.40%	9.80%

Aggregated-ELBO

$$\mathbb{E}_{\mathbf{x}\sim p_{\text{data}}(\mathbf{x})}[\log p(\mathbf{x})] \geq \underbrace{-\mathbb{E}_{\mathbf{x}\sim p_{\text{data}}(\mathbf{x})}\Big[\mathbb{E}_{q(\mathbf{z}|\mathbf{x})}[-\log p(\mathbf{x}|\mathbf{z})]\Big]}_{\text{VAE Reconstruction}} \underbrace{-\mathbb{E}_{\mathbf{x}\sim p_{\text{data}}(\mathbf{x})}\Big[\mathrm{KL}(q(\mathbf{z}|\mathbf{x})\|p(\mathbf{z}))\Big]}_{\text{VAE Regularization}}$$

$$= \underbrace{-\mathbb{E}_{\mathbf{x}\sim p_{\text{data}}(\mathbf{x})}\Big[\mathbb{E}_{q(\mathbf{z}|\mathbf{x})}[-\log p(\mathbf{x}|\mathbf{z})]\Big]}_{\text{AVB Reconstruction}} \underbrace{-\mathrm{KL}(q(\mathbf{z},\mathbf{x})\|p(\mathbf{z})p_{\text{data}}(\mathbf{x}))}_{\text{AVB Regularization}}$$

$$= \underbrace{-\mathbb{E}_{\mathbf{x}\sim p_{\text{data}}(\mathbf{x})}\Big[\mathbb{E}_{q(\mathbf{z}|\mathbf{x})}[-\log p(\mathbf{x}|\mathbf{z})]\Big]}_{\text{AAE Reconstruction}} \underbrace{-\mathrm{KL}(q(\mathbf{z})\|p(\mathbf{z}))}_{\text{AAE Regularization}} \underbrace{-\mathcal{I}(\mathbf{z};\mathbf{x})}_{\text{Mutual Info.}}$$

$$= \underbrace{-\mathbb{E}_{\mathbf{z}\sim q(\mathbf{z})}\Big[\mathrm{KL}(q(\mathbf{x}|\mathbf{z})\|p(\mathbf{x}|\mathbf{z}))\Big]}_{\text{IAE Reconstruction}} \underbrace{-\mathrm{KL}(q(\mathbf{z})\|p(\mathbf{z}))}_{\text{IAE Regularization}} \underbrace{-\mathcal{H}_{\text{data}}(\mathbf{x})}_{\text{Entropy of Data}}$$

$$= \underbrace{-\mathrm{KL}(q(\mathbf{x},\mathbf{z})\|r(\mathbf{x},\mathbf{z}))}_{\text{IAE Reconstruction}} \underbrace{-\mathrm{KL}(q(\mathbf{z})\|p(\mathbf{z}))}_{\text{IAE Regularization}} \underbrace{-\mathcal{H}_{\text{data}}(\mathbf{x})}_{\text{Entropy of Data}}$$

$$= \underbrace{-\mathrm{KL}(q(\mathbf{x},\mathbf{z})\|p(\mathbf{x},\mathbf{z}))}_{\text{ALI/BiGAN Cost}} \underbrace{-\mathcal{H}_{\text{data}}(\mathbf{x})}_{\text{Entropy of Data}}$$

Thank you!