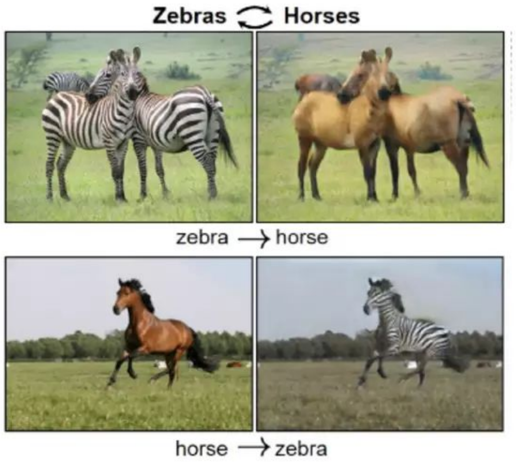
Variational Autoencoder

Do you recognize these people?

Applications

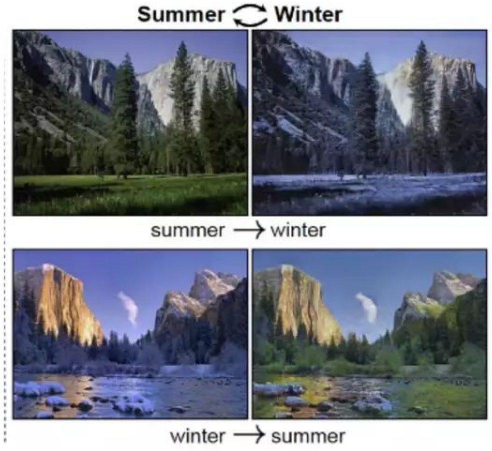
- Image editing



From “Unpaired Image-to-Image Translation using Cycle-Consistent Adversarial Networks,” https://arxiv.org/abs/1703.10593

Applications

- Style transfer



From “Unpaired Image-to-Image Translation using Cycle-Consistent Adversarial Networks,” https://arxiv.org/abs/1703.10593

Applications

- Pose interpolation

From “Representation Learning by Rotating Your Faces,”

https://arxiv.org/abs/1705.11136

Applications

- Pose interpolation

Input Generated

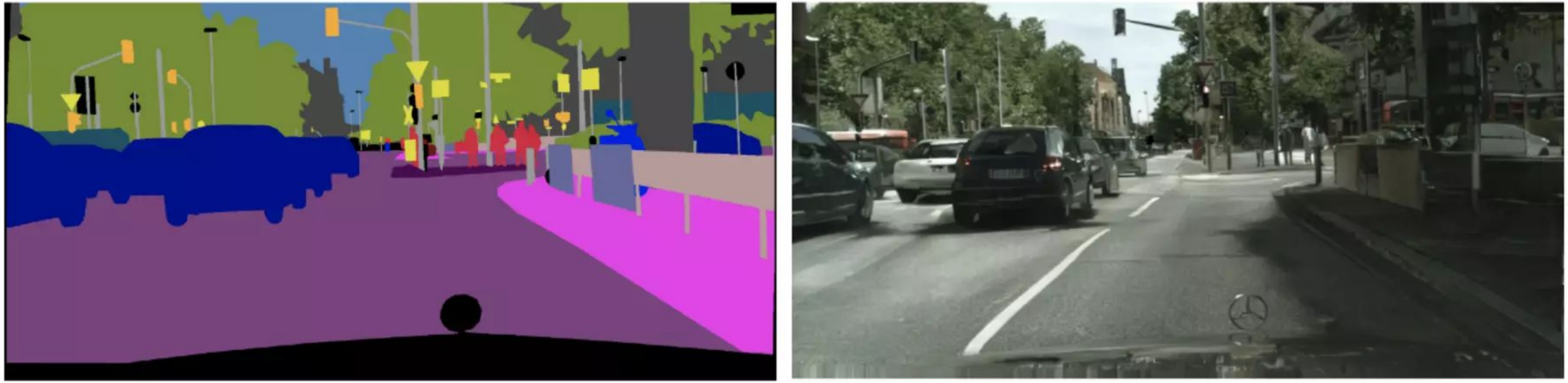
From “Representation Learning by Rotating Your Faces,”

https://arxiv.org/abs/1705.11136

Applications

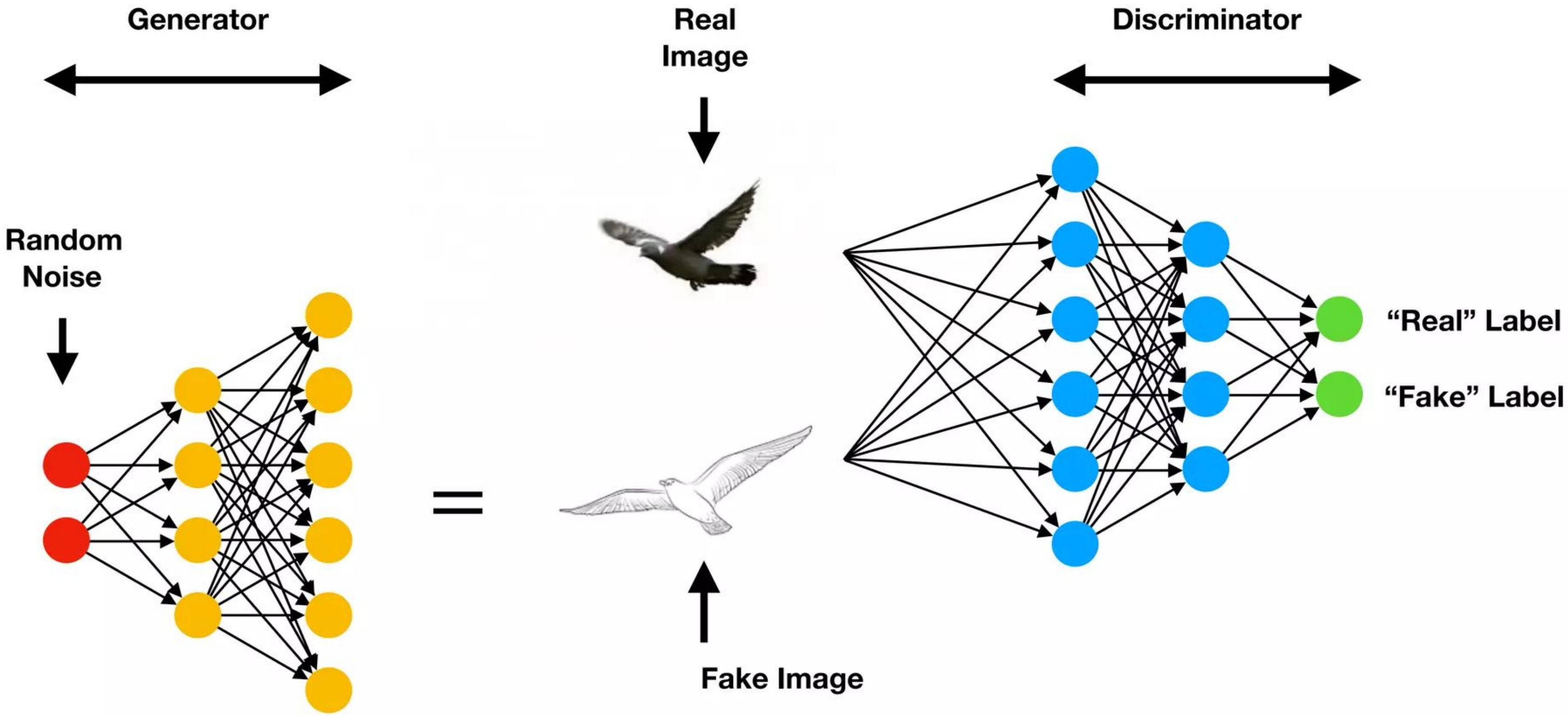
- Image synthesis

Input labels Synthesized image



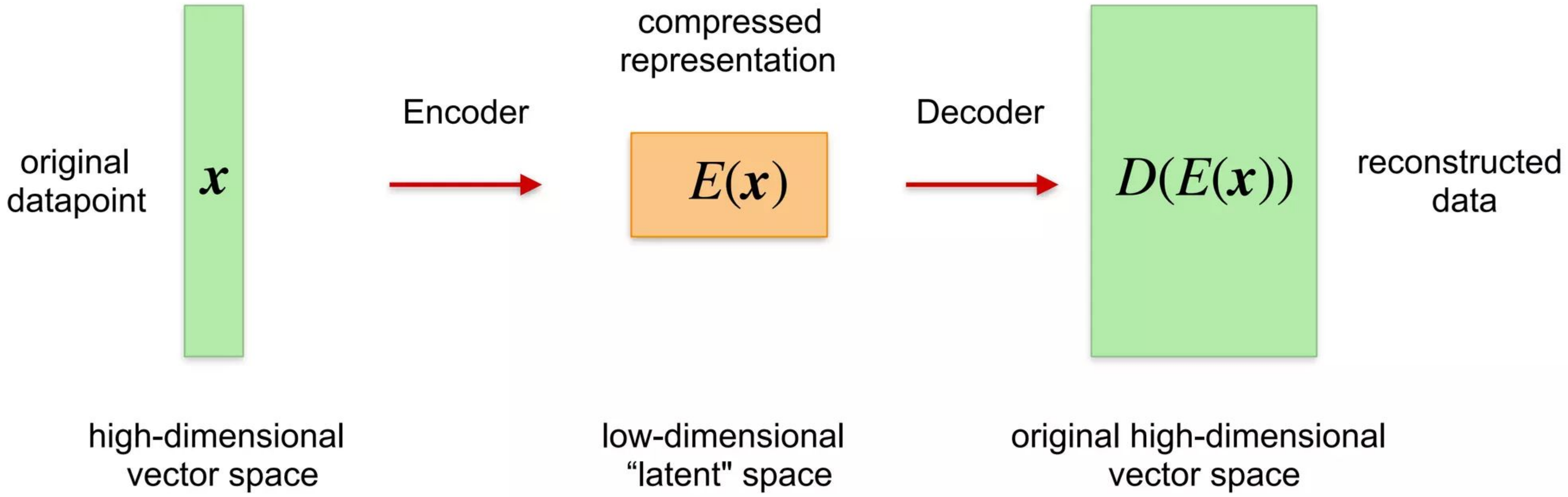
From “High-Resolution Image Synthesis and Semantic Manipulation with Conditional GANs,” https://arxiv.org/abs/1711.11585

Generative Adversarial Network (GAN)



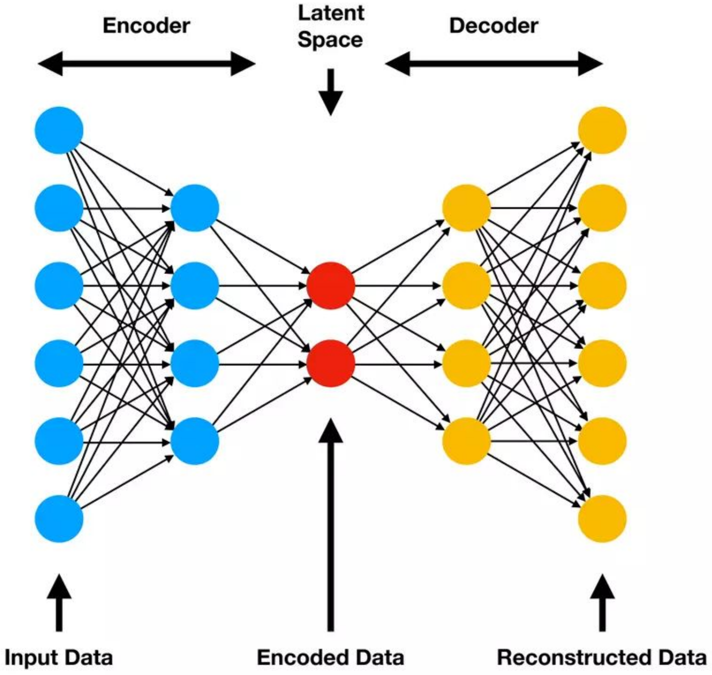
The autoencoder (AE)

- Learn a low-dimensional representation of high-dimensional data.
- Macro-architecture comprises an encoder followed by a decoder.



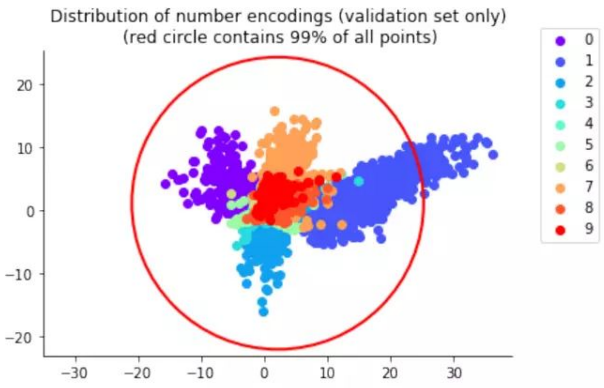
AE microarchitecture

- The encoder and decoder are usually neural networks:
- Layers are often fully connected or convolutional (for image data):

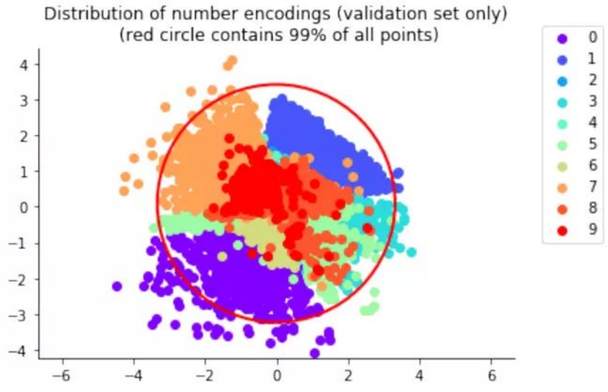


AE limitations

- Encoded representations optimize for data reconstruction, not generation.
- Encoding clusters have irregular shape, which make them hard to sample.
- As a result, random generation of good imitation data is hard to do.



The encodings we got.

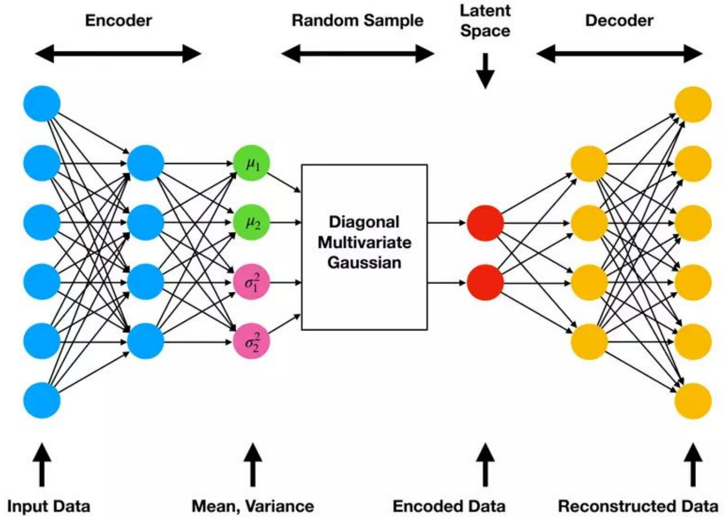


The encodings we want.

Variational autoencoders (VAE)

A variational autoencoder (VAE) is an AE with two adaptations.

- Encoder maps datapoints to probability distributions.

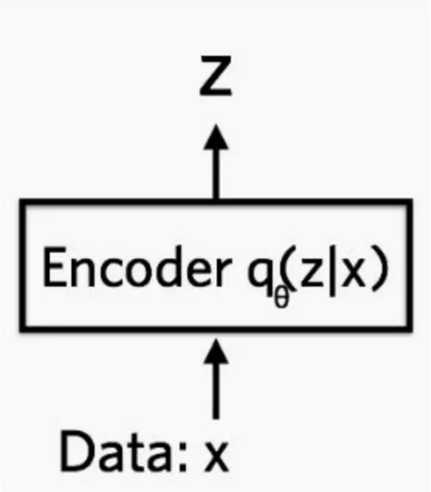


- Add a new “KL-divergence” term to the loss function during training.

Variational Auto-Encoder

Neural network prespective

The approximated function starts to shape up as a neural encoder, going from training datapoints **x** to the likely **z** points following $Q(z|X)$, which in turn is similar to the real $P(z|X)$.



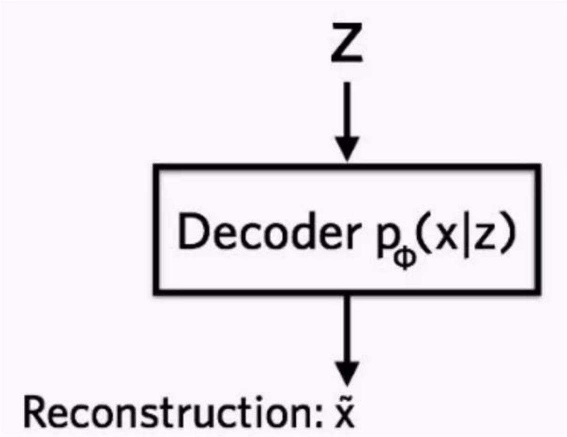
Credit: Altosaar

[What is a variational autoncoder? (Altosaar 2017)](#)

Variational Auto-Encoder

Neural network prespective

The (latent→ data) mapping starts to shape up as a neural decoder, where we go from our sampled **z** to the reconstruction, which can have a very complex distribution.



Credit: Altosaar

VAE learning objective

- KL-divergence measures “distance” between two probability distributions.

$$L = \frac{1}{|T|} \sum_{x \in T} d(\boldsymbol{x}, D(E(\boldsymbol{x}))) + \mathrm{KL}(\mathcal{N}(\boldsymbol{\mu}(\boldsymbol{x}), \boldsymbol{\sigma}(\boldsymbol{x})^2) \;||\; \mathcal{N}(0,1))$$

The KL term in L encourages datapoints to map near to unit-Gaussians.

- Update weights in the encoder E and decoder D via gradient descent.