Deep Unsupervised Learning

Naeemullah Khan
naeemullah.khan@kaust.edu.sa

جامعة الملك عبدالله
للعلوم والتقنية
King Abdullah University of
Science and Technology

LMH
Lady Margaret Hall

KAUST Academy
King Abdullah University of Science and Technology

Autoencoders

- Family of neural networks for which the input is the same as the output. They work by compressing the input into a latent-space representation, and then reconstructing the output from this representation.
- The idea is to project the input into a latent space and then reconstruct the input from that latent space representation
- Consist of two parts: Encoder and decode.
  - Encoder projects the input to a latent space Z.
  - Decoder takes the encoded embedding vector and reconstructs the input from it.
  - We also use altered versions of input as output which can be even more interesting.

Autoencoders (cont.)

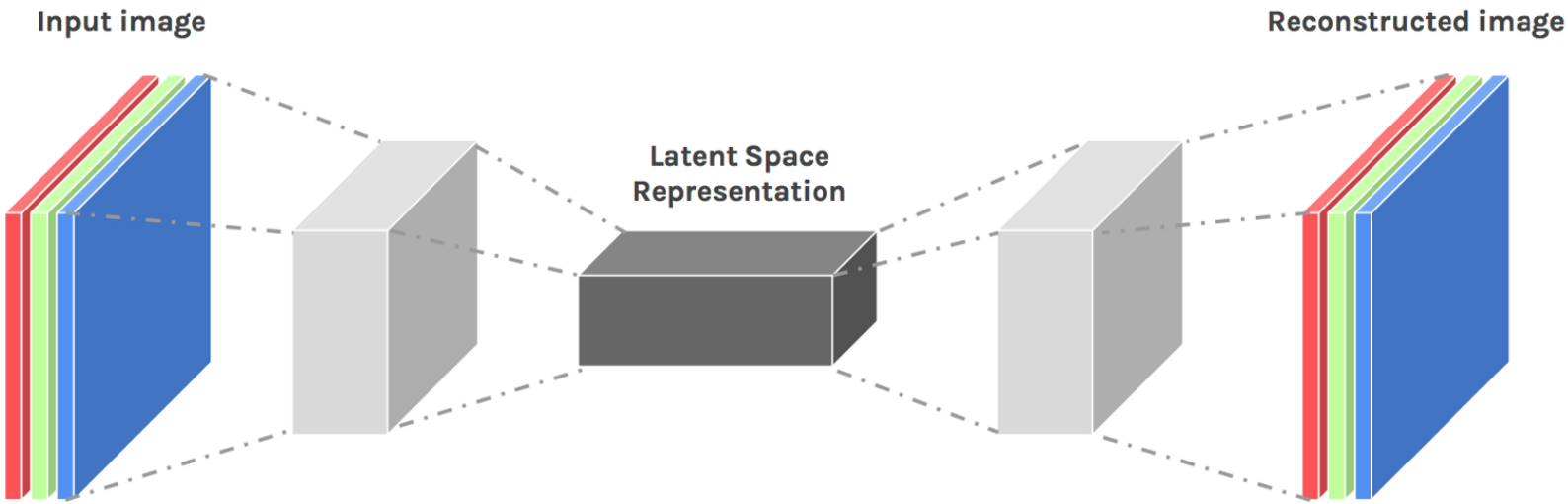

Figure 2: Autoencoder architecture

Autoencoders (cont.)

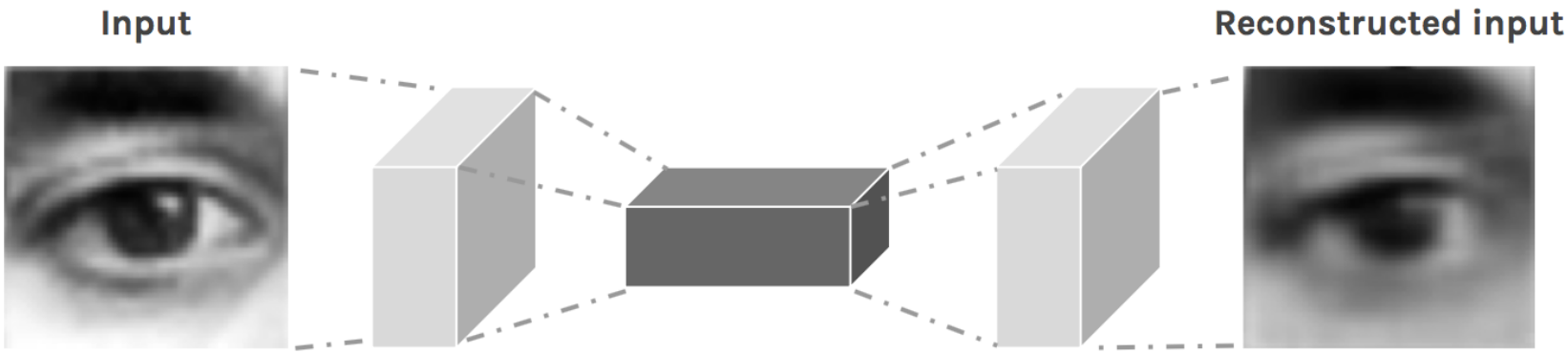

Figure 3: Sample Autoencoder

Autoencoders - Interactive Demo

https://douglasduhaime.com/posts/visualizing-latent-spaces.html

Autoencoders as generative models

- Autoencoders project data into a latent space Z.
- What if we sample a new embedding vector from Z and then have the decoder reconstruct the image from it?
- **Does not work**. Autoencoders just learn a function that maps input to output. The learned latent space is too discontinous to work as a generative model.

Autoencoders as generative models (cont.)

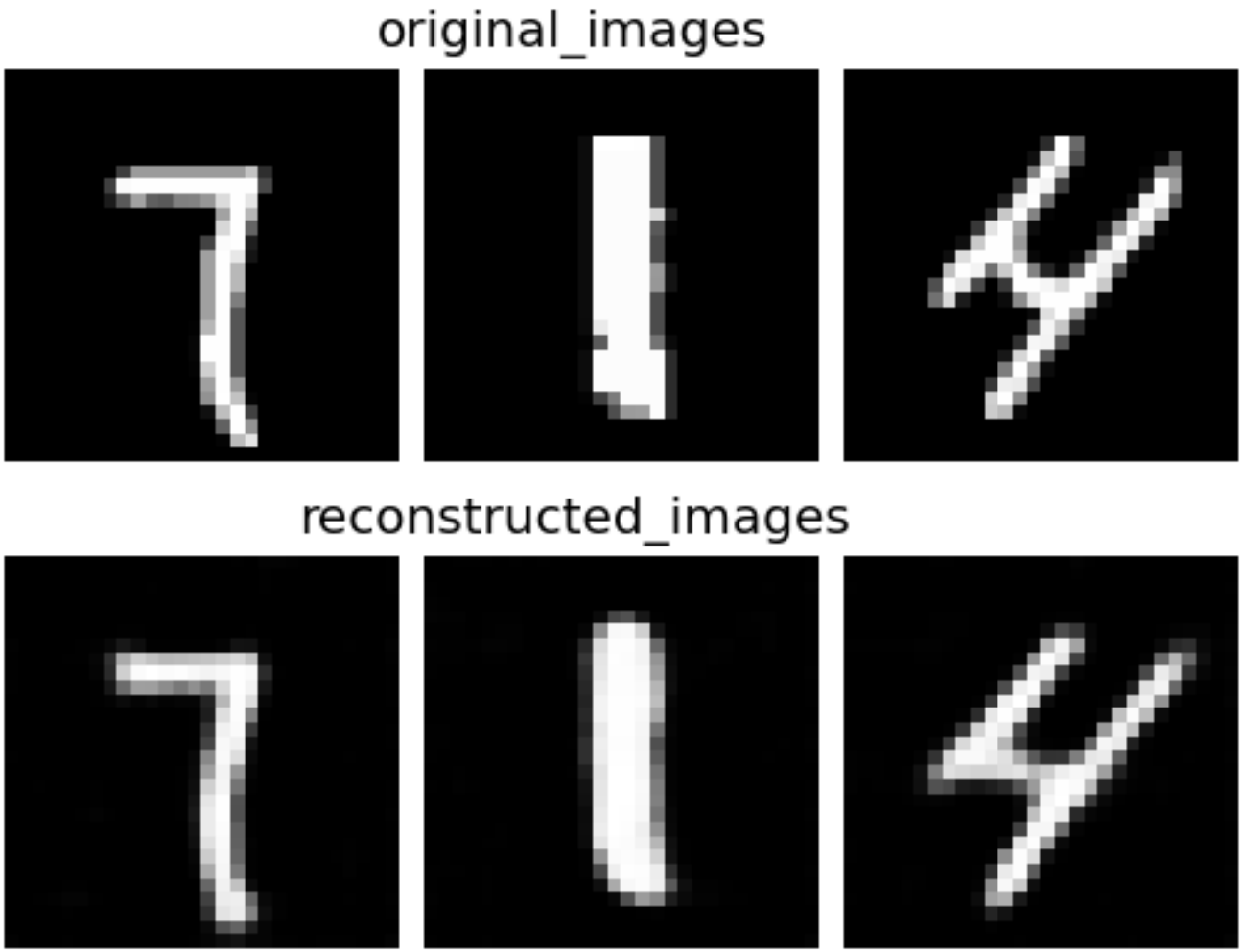

Figure 4: Image reconstruction with autoencoder trained on MNIST digits

Autoencoders as generative models (cont.)

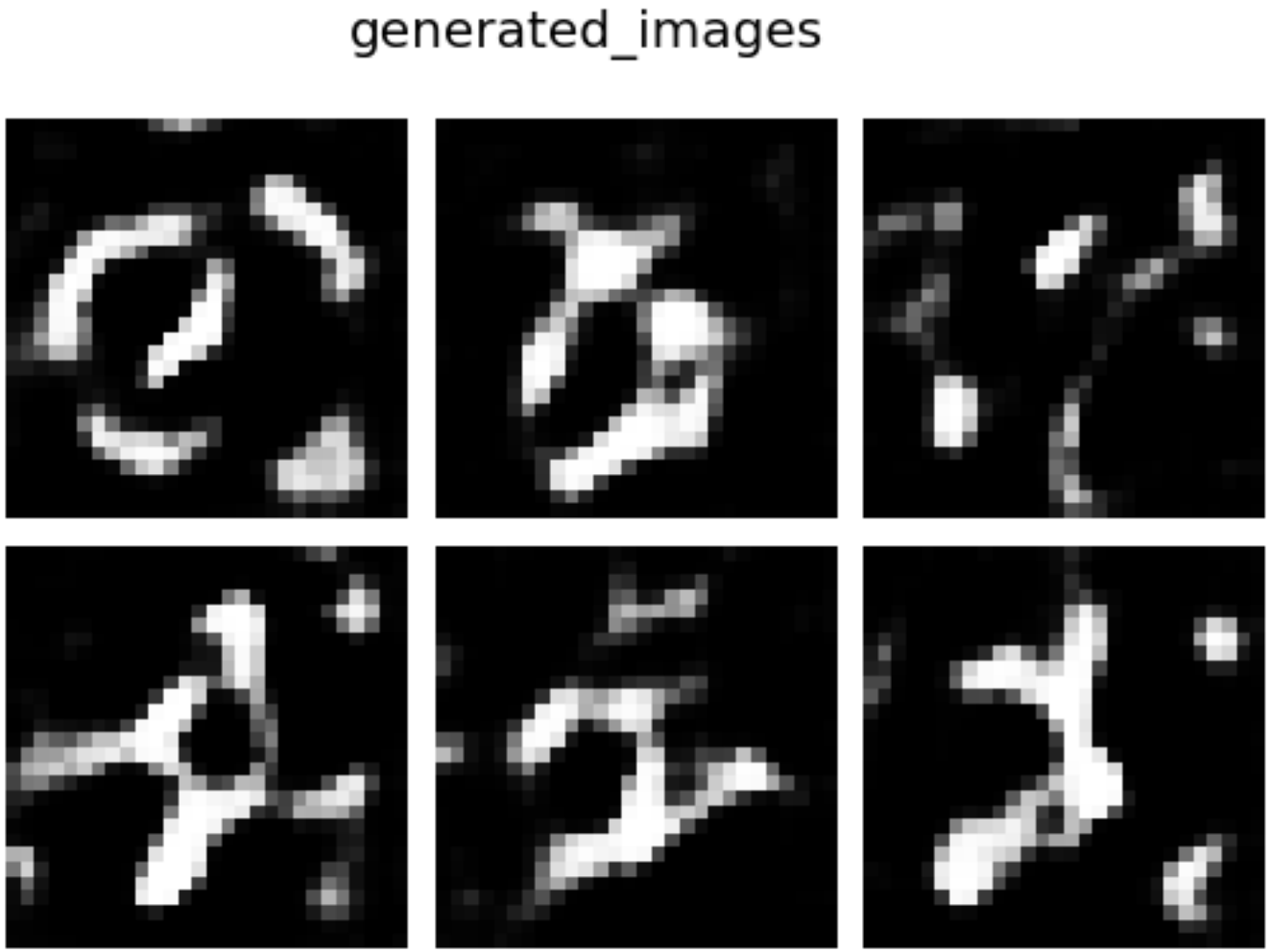

Figure 5: Image generation with autoencoder trained on MNIST digits. Encoding vector sampled from latent space Z and the passed to decoder.

Autoencoders - Applications

- While autoencoders themselves have very low generative power, we will soon talk about a type of autoencoders called **Variational Autoencoders** which are specifically designed for generative modeling.
- Other use cases of Autoencoders include:
  - Data encoding and dimensionality reduction
  - Image denoising and super-resolution
  - Image completion
  - Image colorization

Autoencoders - Applications (cont.)

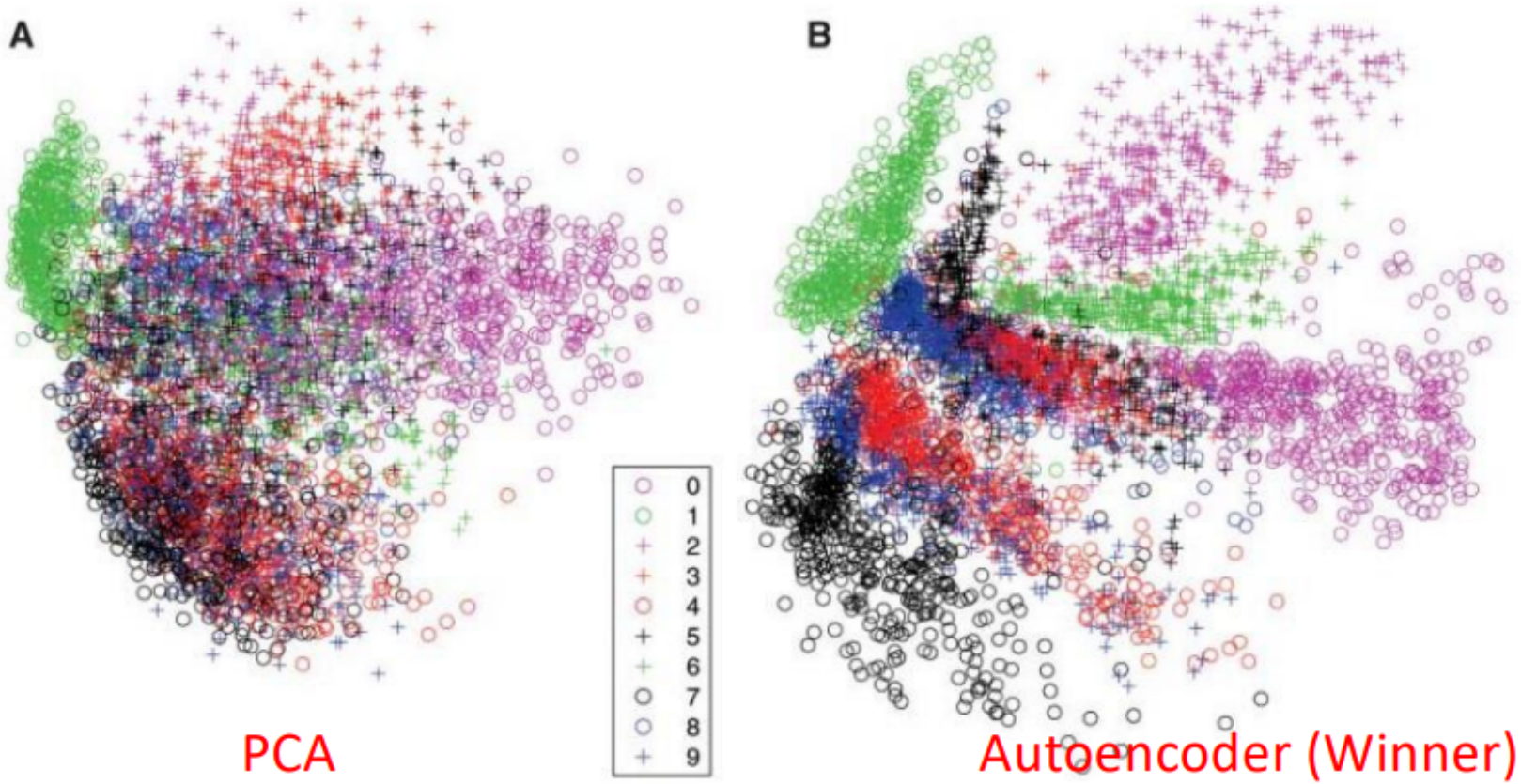

Figure 6: t-SNE visualization on MNIST digits dataset. PCA vs. Autoencoders. The image vector is projected into $\mathbb{R}^2$.

Autoencoders - Applications (cont.)

Figure 7: Image super-resolution using Autoencoders

Autoencoders - Applications (cont.)

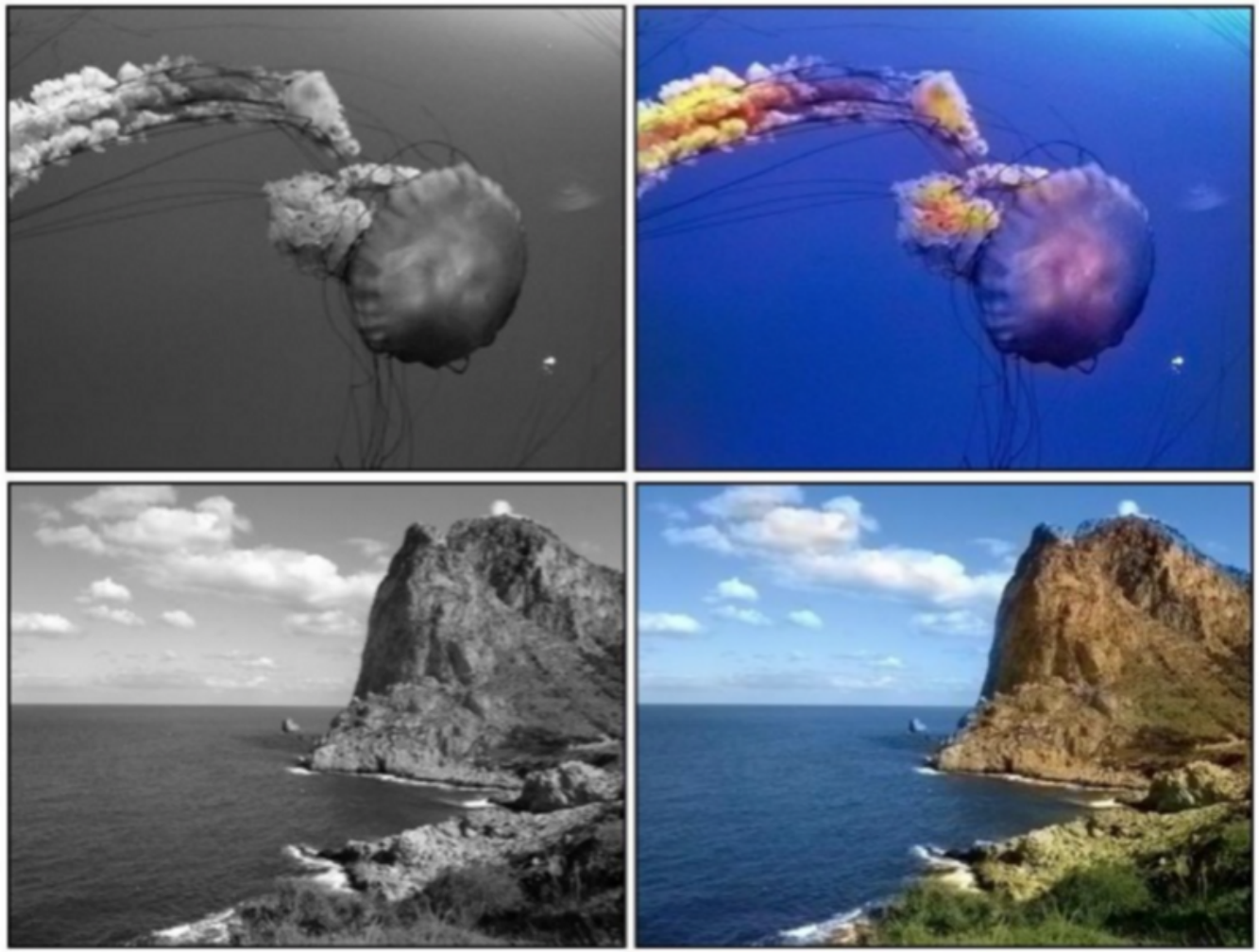

Figure 8: Image colorization using Autoencoders

References

Reference Slides

- Fei-Fei Li "Generative Deep Learning" CS231
- Murtaza Taj "Deep Learning" CS437