EPFL

Alexander Rusnak

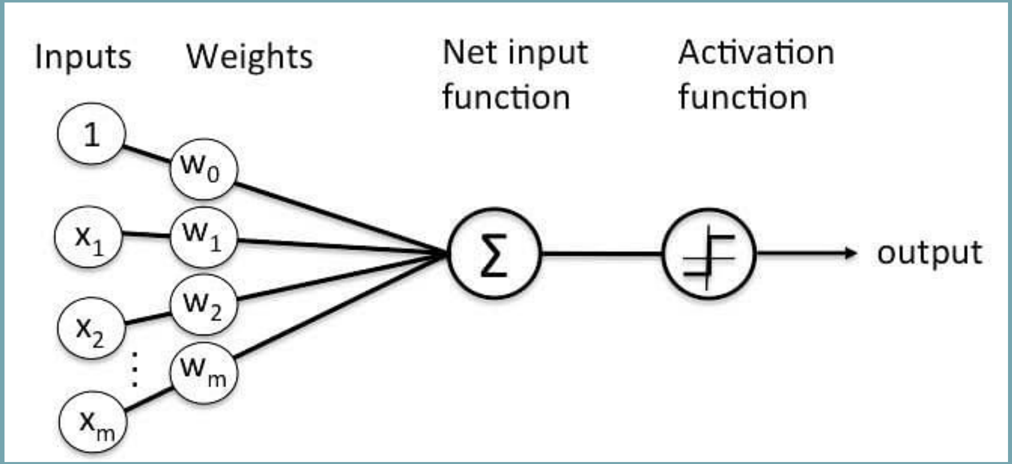
Conditional Generative Modeling with Diffusion

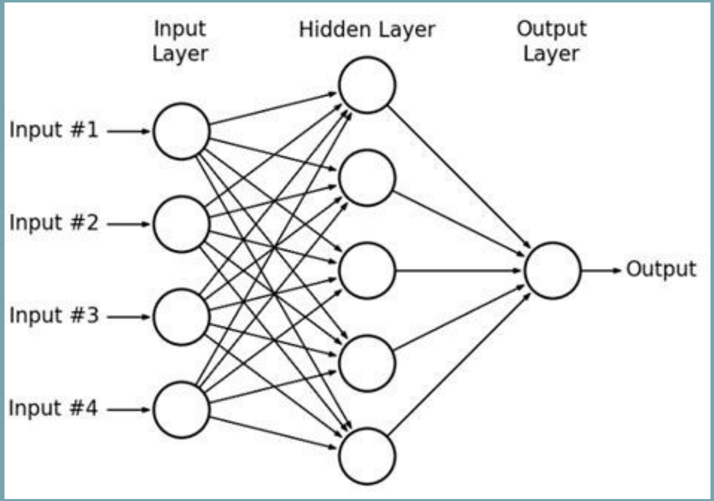
Introduction

- Motivation: increase in excitement around generative imaging and proliferation of approaches.
- Object of presentation is to understand the fundamentals of diffusion models rather than learn to use highly abstracted foundational tools:
  - But by understanding the architectures of the foundational models, maybe you will understand how to use them better.
  - Or you can work on your own architectures!
- Generating samples from small datasets has huge advantages in many ML problems - not just for generating pretty pictures.

Neural Networks

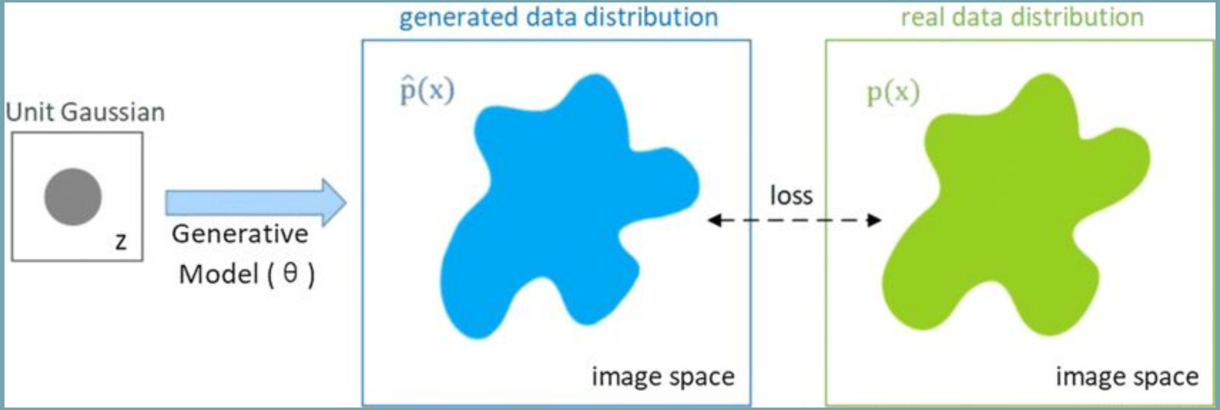
- Assumed as background knowledge of this lecture





Generative Modeling

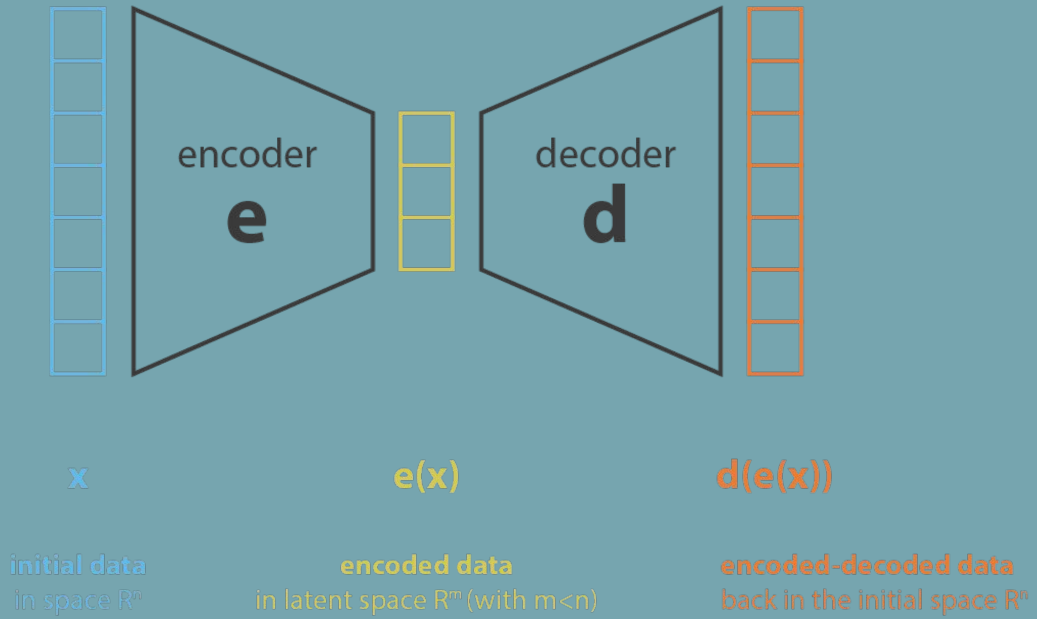
- Learning to create novel samples that conform to distribution of target data P(x).
- Either by explicitly seeking to define P(x) or to approximate a transformation from another distribution to a sample conforming to P(x).



Generative Modeling (cont.)

- Deep learning is often the backbone for learning this transformation.
- Self supervised learning -> create a sample and loss function determines how realistic it is in comparison with real samples or desired output.
- Other popular SSL deep generative modeling techniques: Transformers, GANs, Autoencoders.

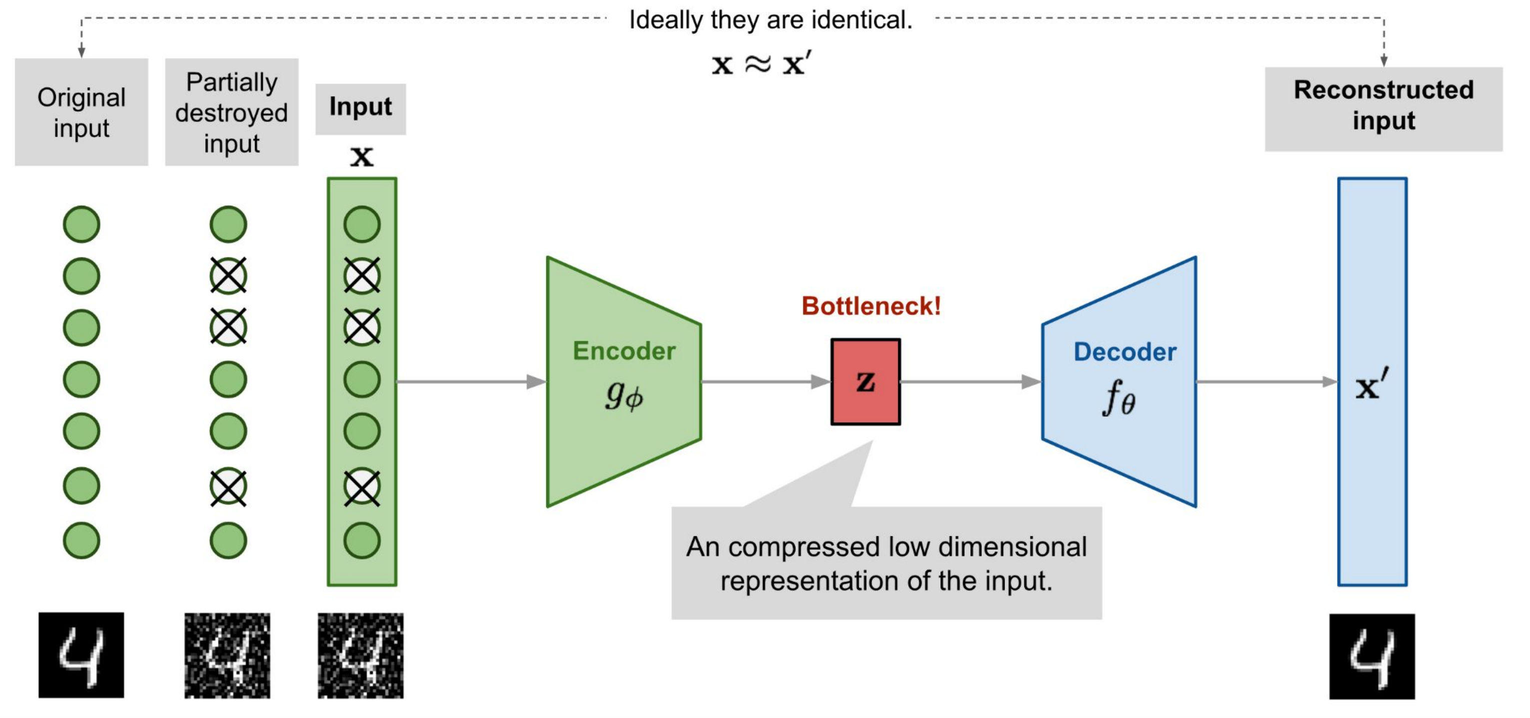
Autoencoders



$x = d(e(x))$ → **lossless encoding**
no information is lost when reducing the number of dimensions

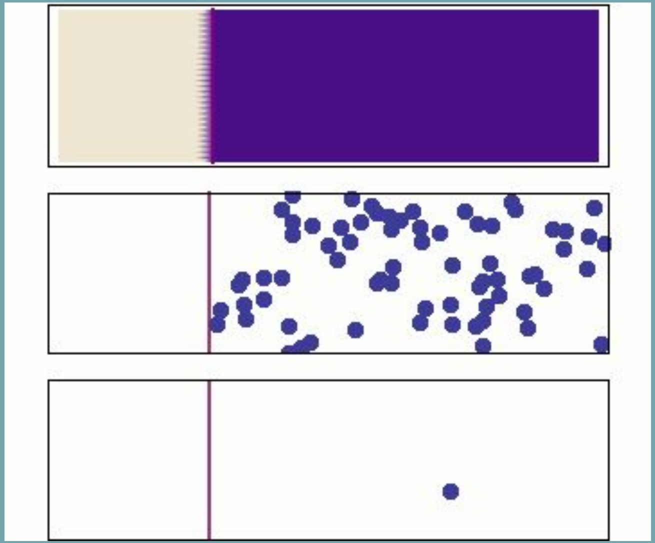
$x \neq d(e(x))$ → **lossy encoding**
some information is lost when reducing the number of dimensions and can't be recovered later

Denoising Autoencoders



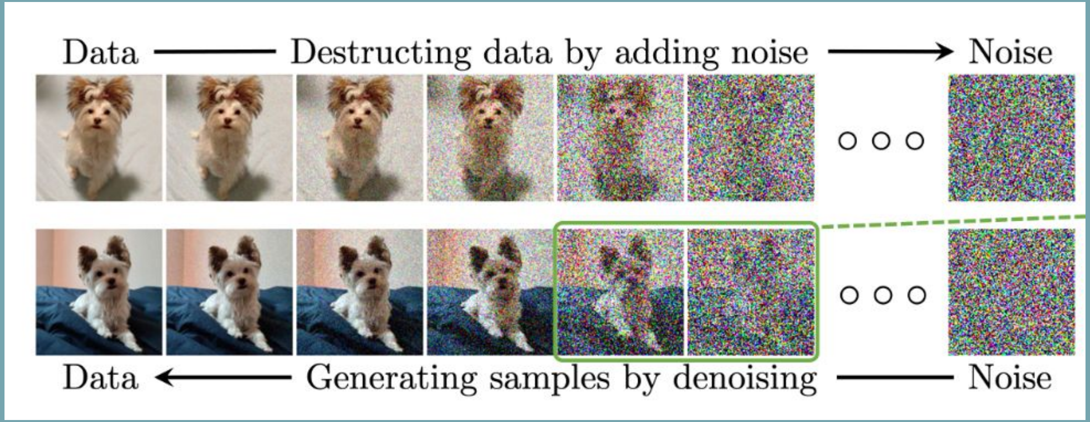
Physics Origins of Diffusion Models

- Diffusion: The process whereby particles of liquids, gases, or solids intermingle as the result of their spontaneous movement caused by thermal agitation and in dissolved substances move from a region of higher to one of lower concentration.
- This is a highly difficult process to model directly / in one shot because of the high degree of randomness.
- Observation: Movements of individual particles at any step can be modeled by gaussian process.
- Semi intractable and complex transformation reduced to smaller steps of tiny additions/subtractions of gaussian noise.



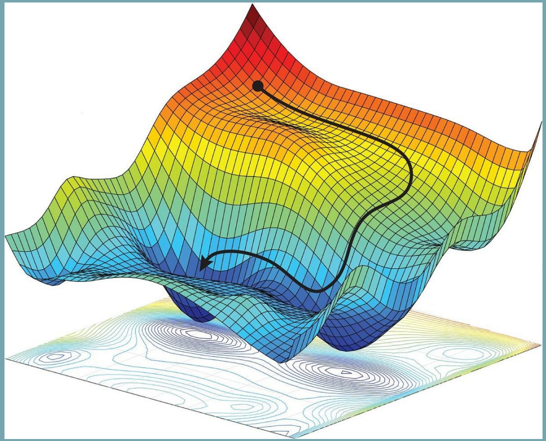
Denoising Diffusion Models

- Predicts what is noise at each step rather than what is signal.
- Reduces each transformation to a small gaussian which is far easier to model.
- Step by step reconstruction from noise rather than one shot or autoregressive at inference.



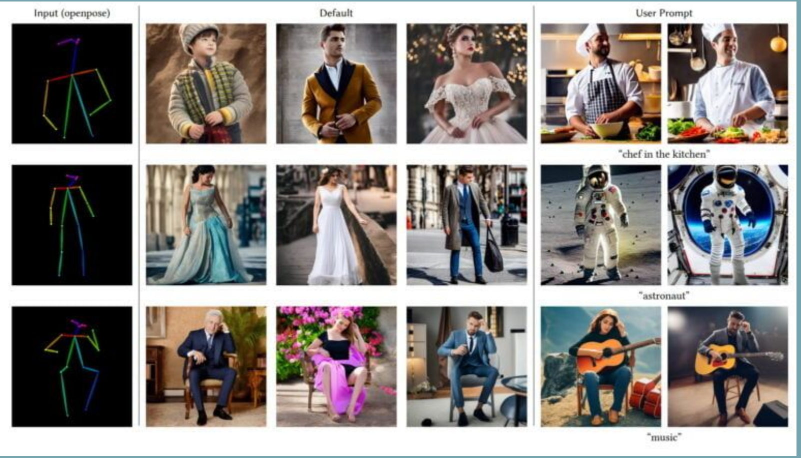
Language Vectors and CLIP

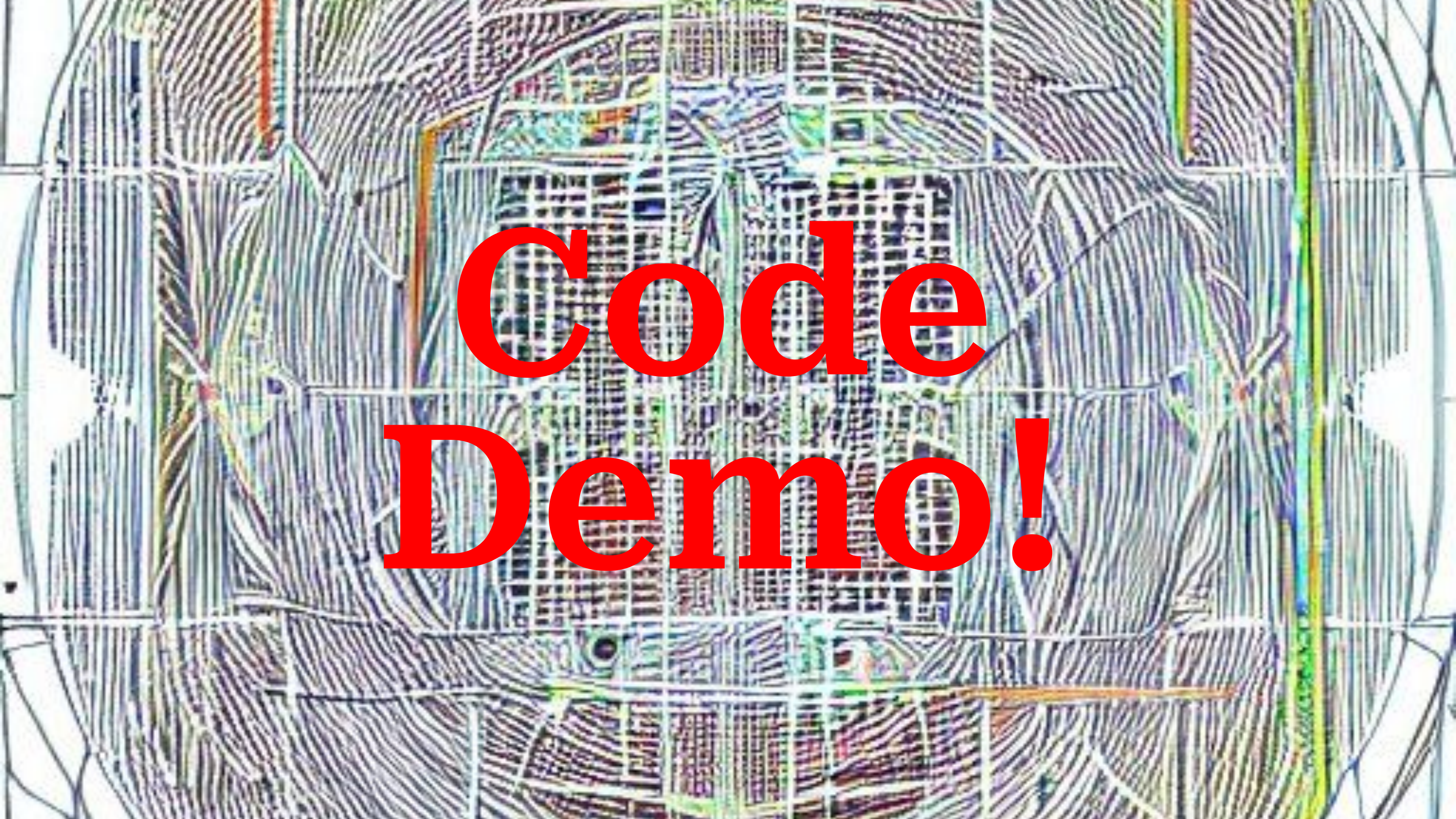
- Contrastive Language-Image Pre-training.
- Creates joint embedding space of images and natural language.
- Can be used to guide diffusion process:
  - Get embedding of text prompt.
  - At each generative step, find embedding of image or subsections of image.
  - Loss function -> distance between prompt and image embedding.
  - Slowly move image in the direction of prompt as you denoise.



Related Models

- Latent Diffusion Models
  - Stable Diffusion
  - Midjourney
- ControlNet
- Dall-E (not diffusion but related to CLIP) (GPT-4)




Code
Demo!