Auto-encoders

CPSC 383: Explorations in Artificial Intelligence and Machine Learning
Fall 2025

Jonathan Hudson, Ph.D.
Associate Professor (Teaching)
Department of Computer Science
University of Calgary

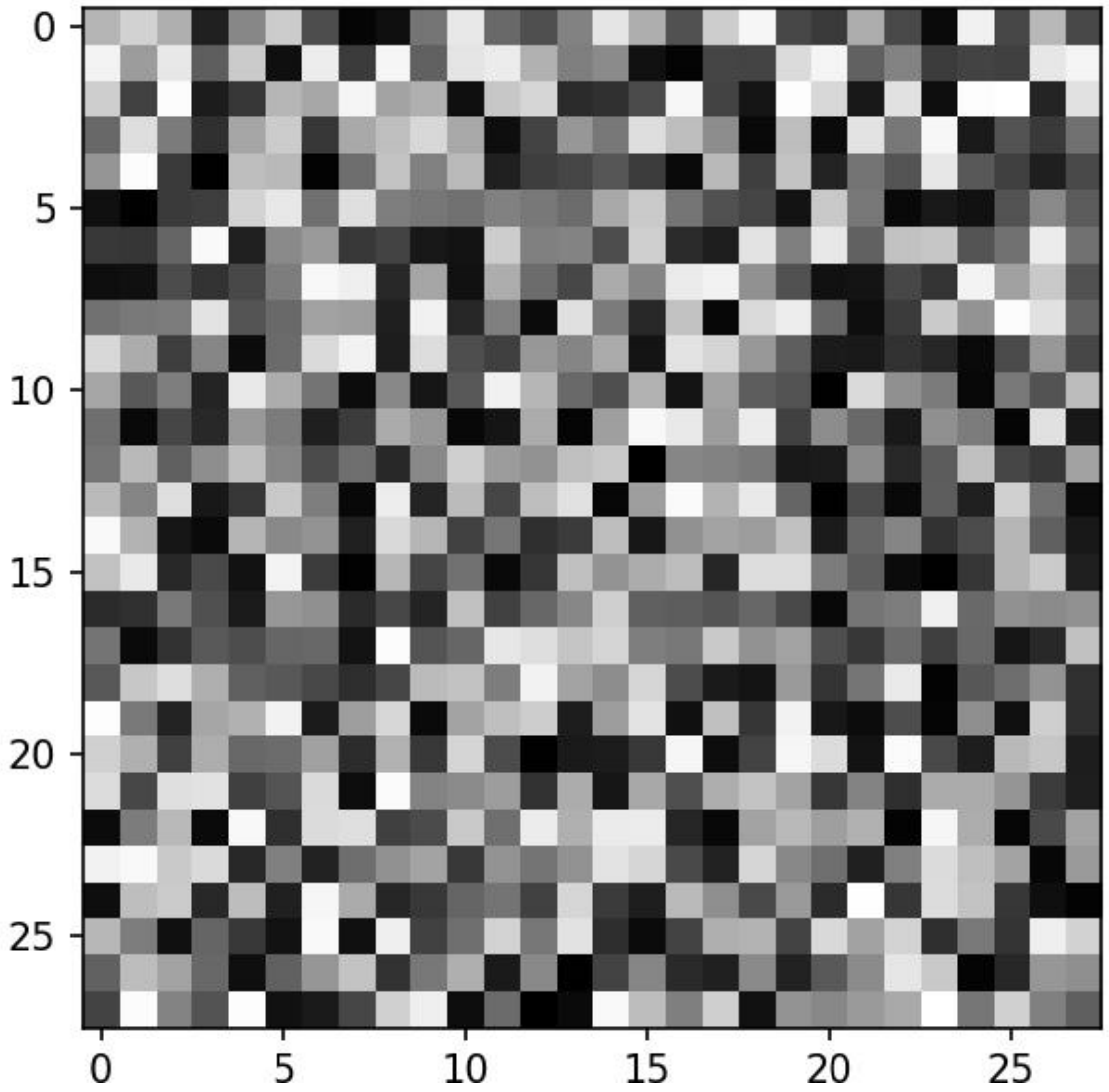
August 27, 2025

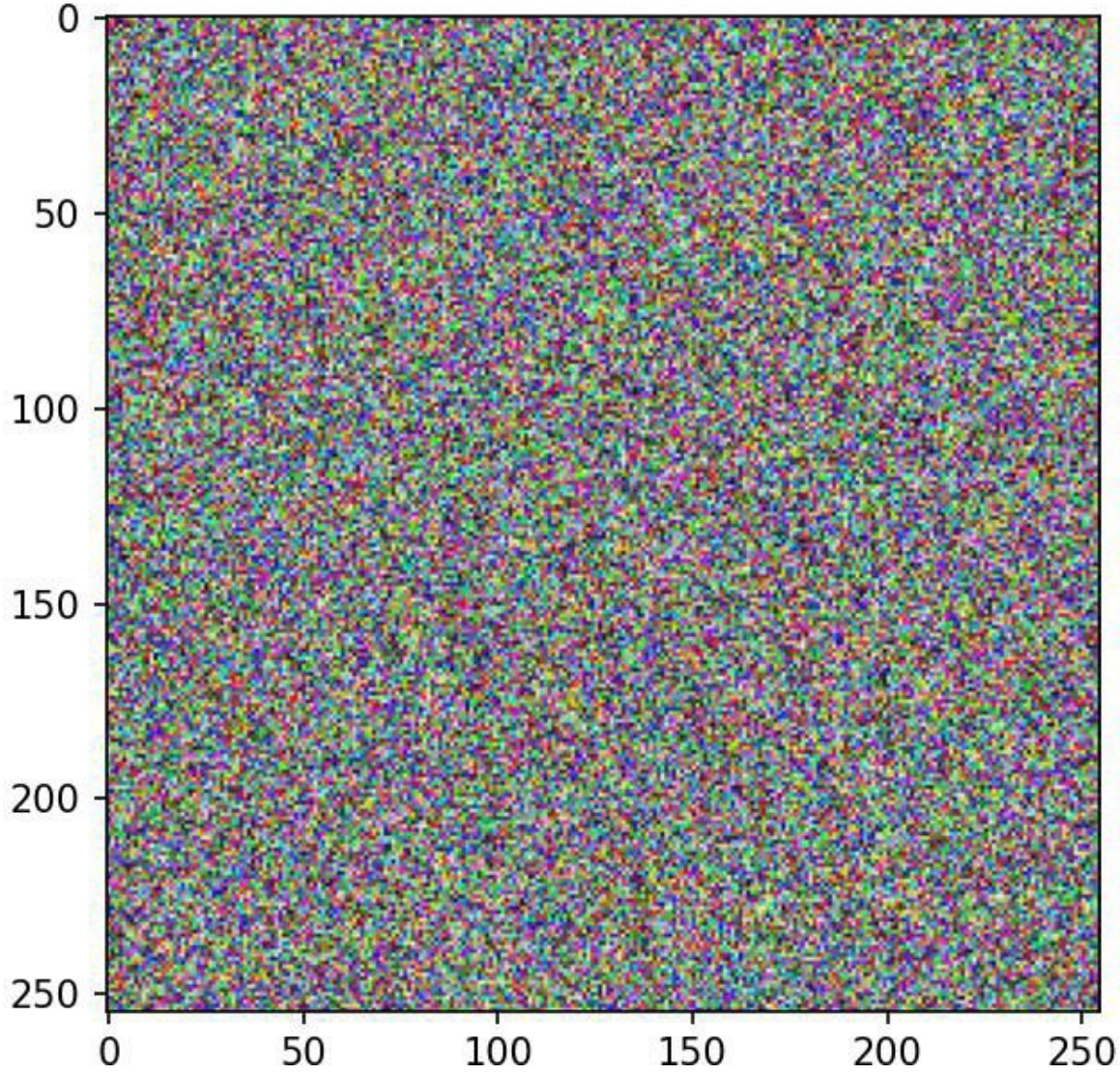
UNIVERSITY OF CALGARY

Idea

There is an inherent structure in “normal” data that we expect to see.

For example, most possible arrangements of pixels just look like random noise or won’t make sense to us.


0
5
10
15
20
25
0
5
10
15
20
25


0
50
100
150
200
250
0
50
100
150
200
250

Idea

Recall that ideally, we want similar data values to have similar encodings.

We want to take advantage of the structure of “natural” images to more efficiently encode them.

Data encoding

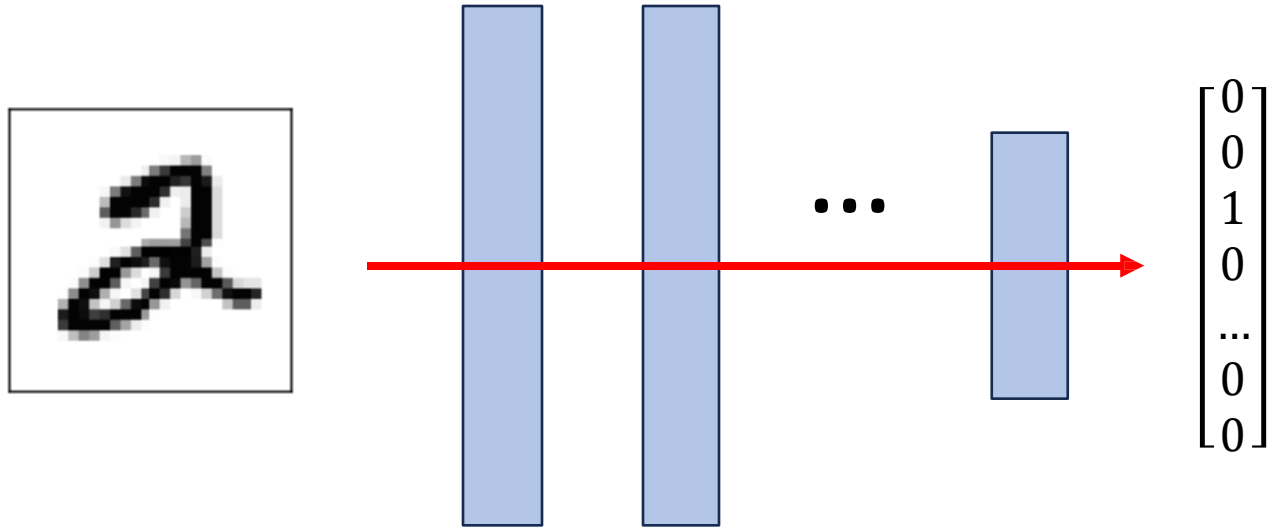
We can’t always describe well what this pattern looks like…

..but we can train a model to try and figure it out!

Example

This is one way you can think about what a neural net for MNIST is doing.

- i.e. compressing 256^{196608} possible input values to 10 possible values representing the aspect of the image that we care about.



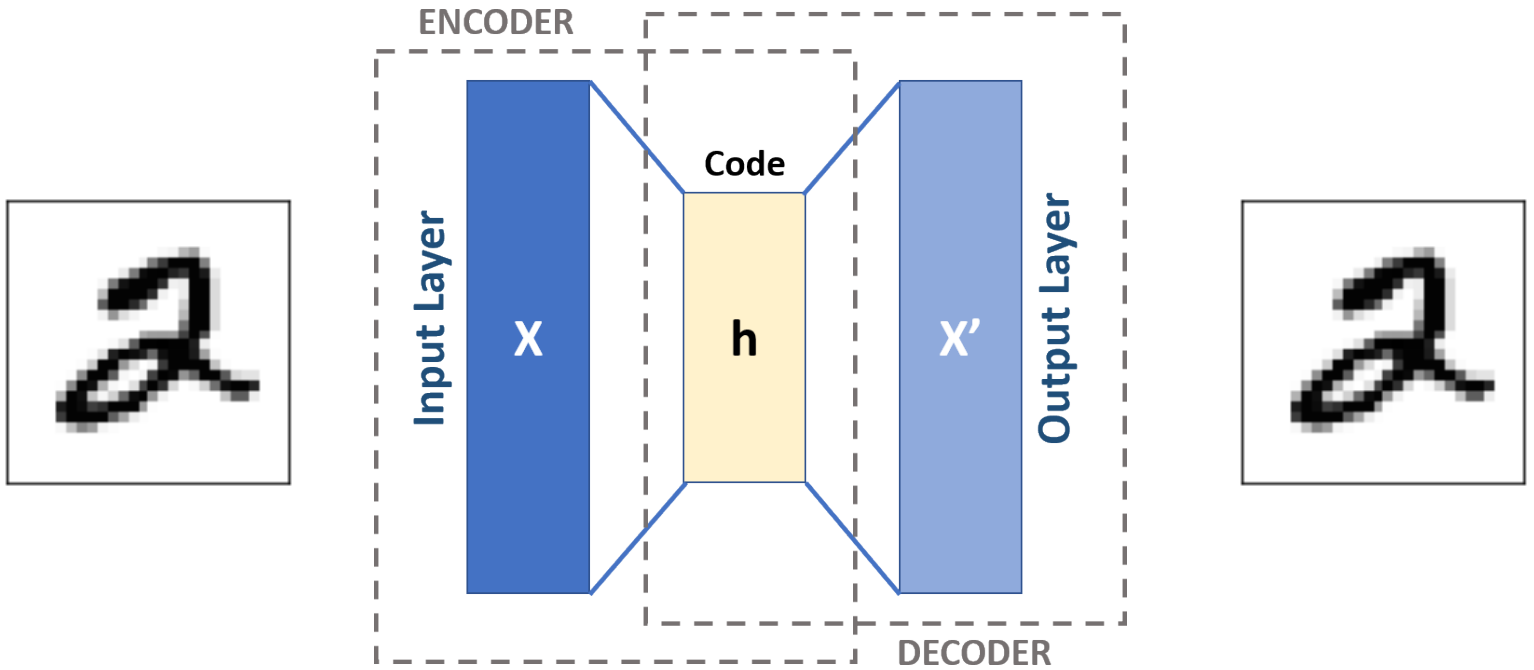
Autoencoding

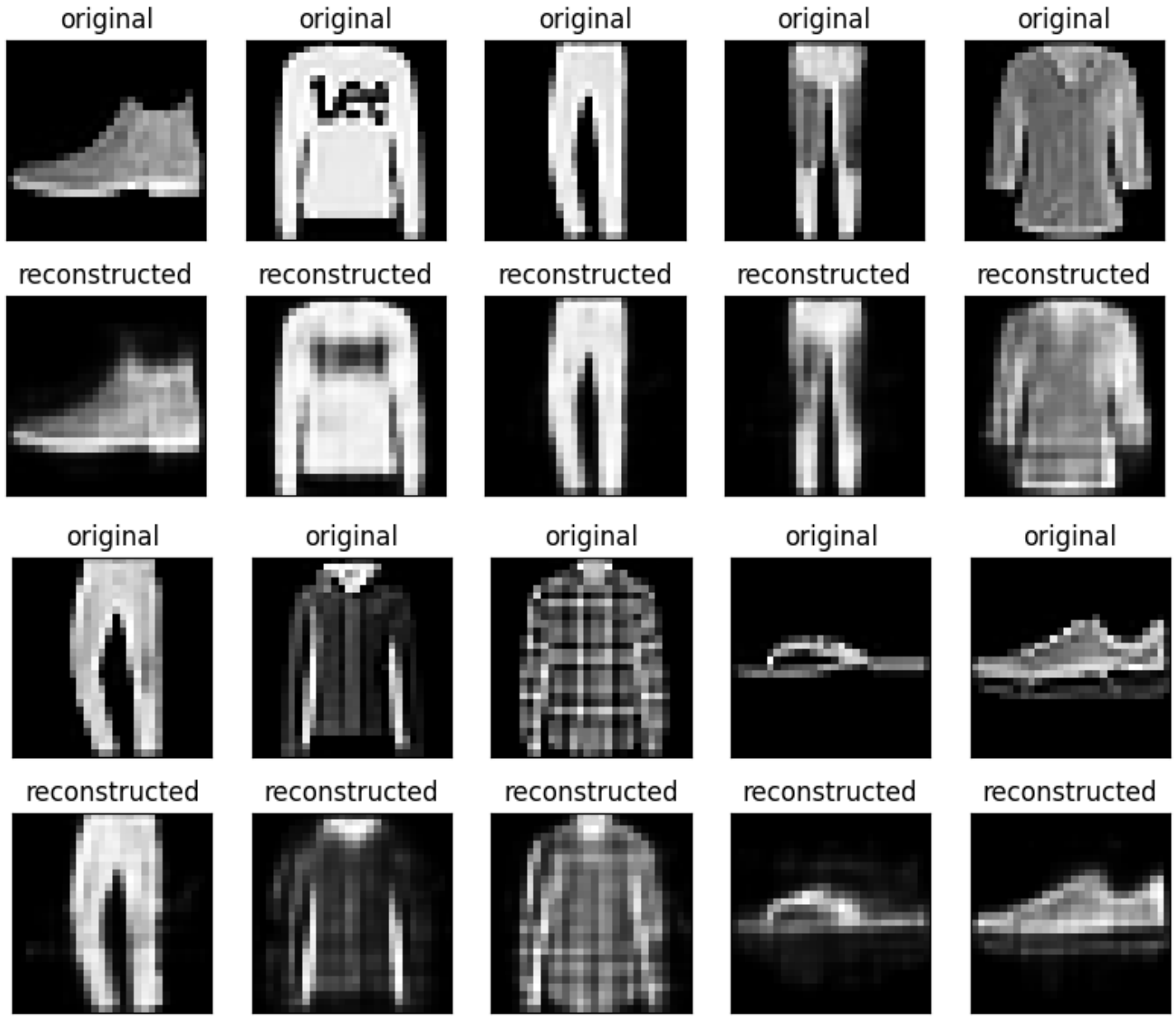
Autoencoding is a type of unsupervised learning used to find feature-preserving representations of unlabeled data in a smaller **latent space**.

It is a type of dimensionality reduction, where the reduction is learned by a model through training.

Autoencoding

- An autoencoder consists of two parts: an **encoder** and a **decoder**.
- The system is trained to produce the identity function, passing data through a small intermediate layer.

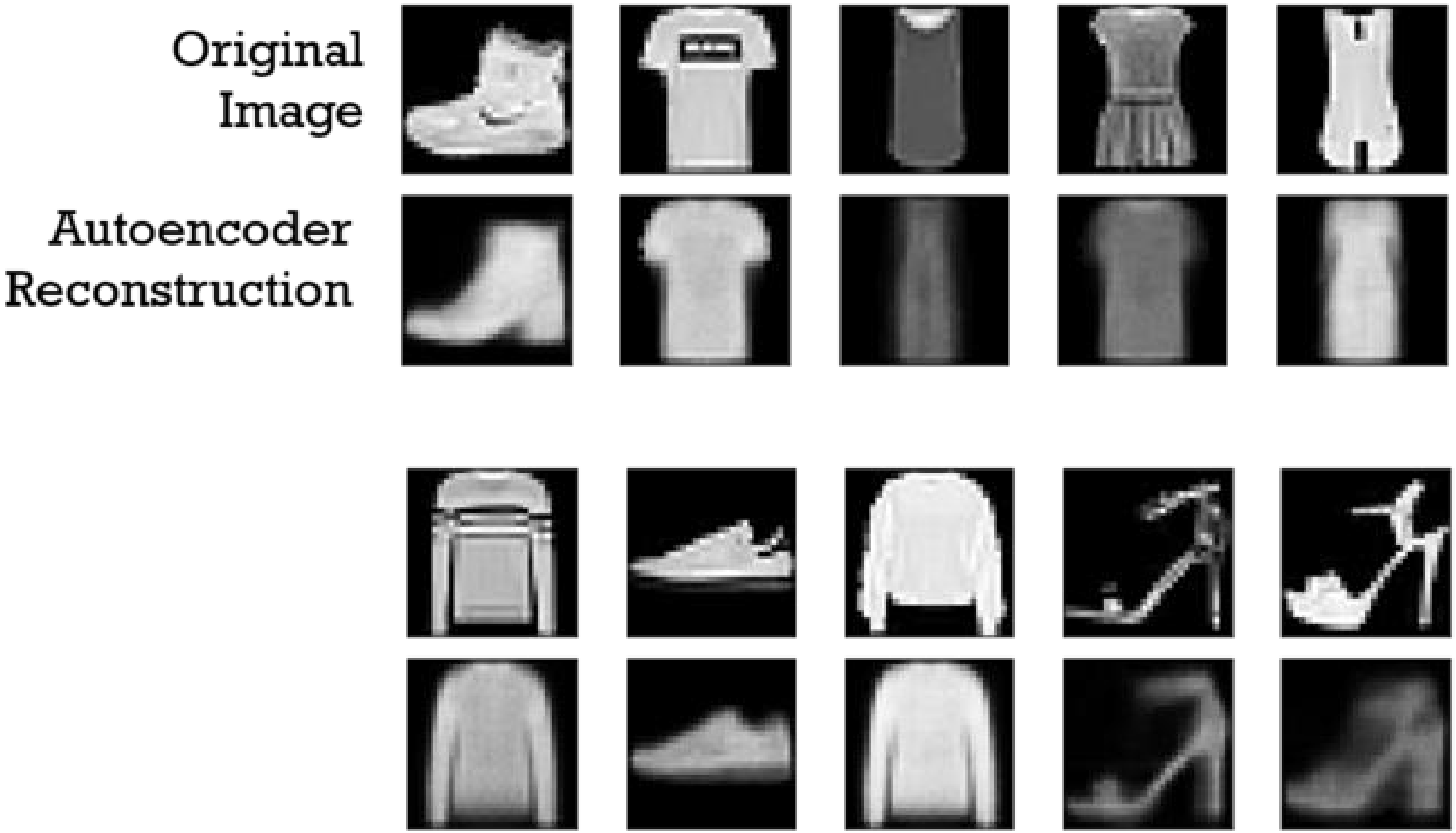




FashionMNIST

0: T-shirt/top
1: Trouser
2: Pullover
3: Dress
4: Coat
5: Sandal
6: Shirt
7: Sneaker
8: Bag
9: Ankle boot

https://www.tensorflow.org/tutorials/generative/autoencoder



https://en.wikipedia.org/wiki/Autoencoder#/media/File:Reconstruction_autoencoders_vs_PCA.png

Autoencoding

We can think of any neural networks as learning to encode the data they see in training.

- This encoding depends on the nature of the training set

The decoder then takes this encoding and makes a final decision based on the encoded data it sees.

Representations

There is the concept in psychology of a “grandmother” neuron.

This is one type of analogy for a **sparse** encoding, where different concepts are divided between different neurons.

In practice, AI (and the brain) seem to use **distributed** encodings.

Applications

Once we have a useful encoding, we can also use it to preprocess data for other applications.

- Can help with learning in subsequent layers
- Also used for generative AI, detecting outliers, data synthesis, etc.
- Used as pre-processing for other ML

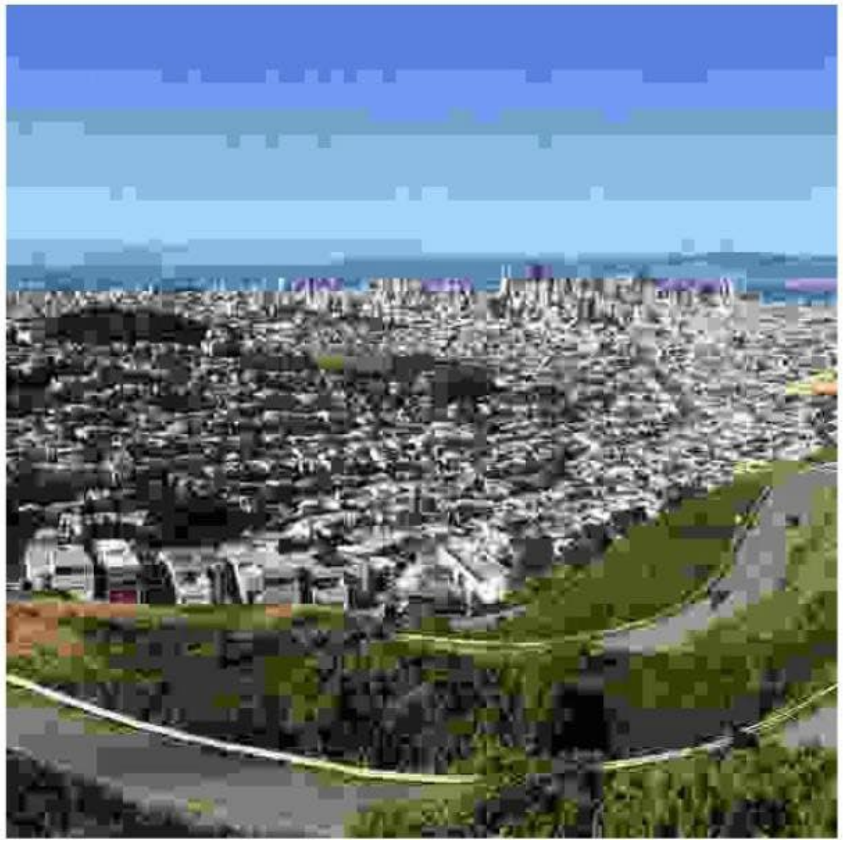
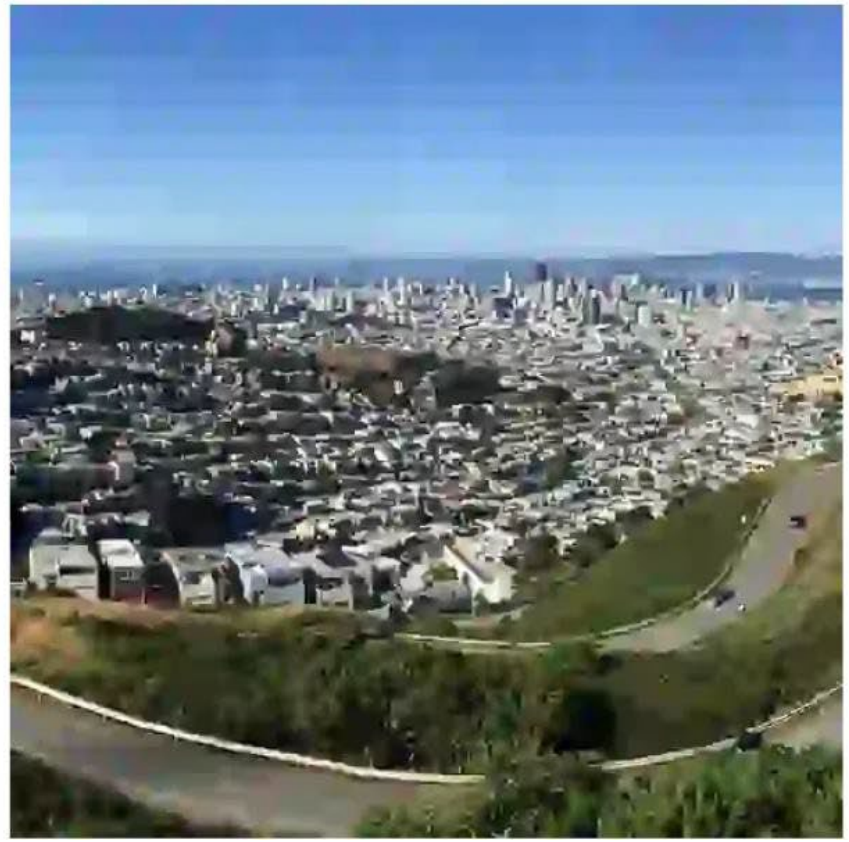
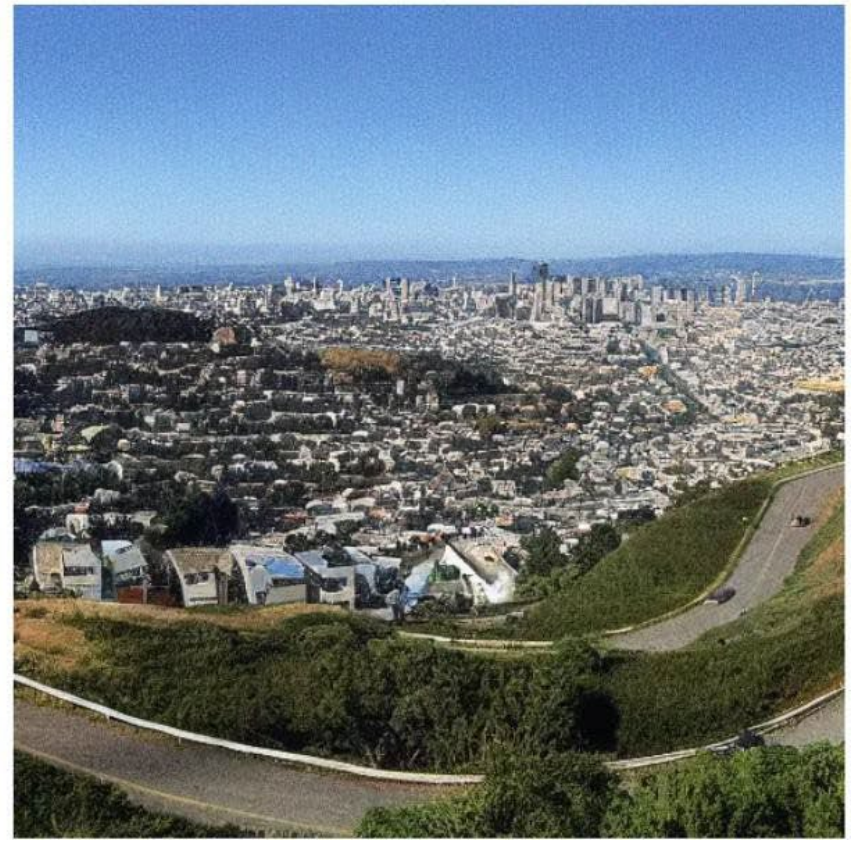
Better than JPEG? Researcher discovers that Stable Diffusion can compress images

Lossy compression bypasses text-to-image portions of Stable Diffusion with interesting results.

BENJ EDWARDS – SEP 27, 2022 2:59 P.M. | 143

https://arstechnica.com/information-technology/2022/09/better-than-jpeg-researcher-discovers-that-stable-diffusion-can-compress-images/

VAE roundtrip, from left to right: 512×512@24bpp ground truth, 64×64@128bpp latent space representation, 512×512@24bpp decoded image

https://pub.towardsai.net/stable-diffusion-based-image-compresssion-6f1f0a399202



https://pub.towardsai.net/stable-diffusion-based-image-compresssion-6f1f0a399202

https://pub.towardsai.net/stable-diffusion-based-image-compresssion-6f1f0a399202

https://pub.towardsai.net/stable-diffusion-based-image-compresssion-6f1f0a399202

https://pub.towardsai.net/stable-diffusion-based-image-compresssion-6f1f0a399202

Reading

https://www.newyorker.com/tech/annals-of-technology/chatgpt-is-a-blurry-jpeg-of-the-web

VAE

A **variational autoencoder** (**VAE**) is a type of autoencoder.

VAEs are used as a component in image generators like DALL-E and Stable Diffusion.

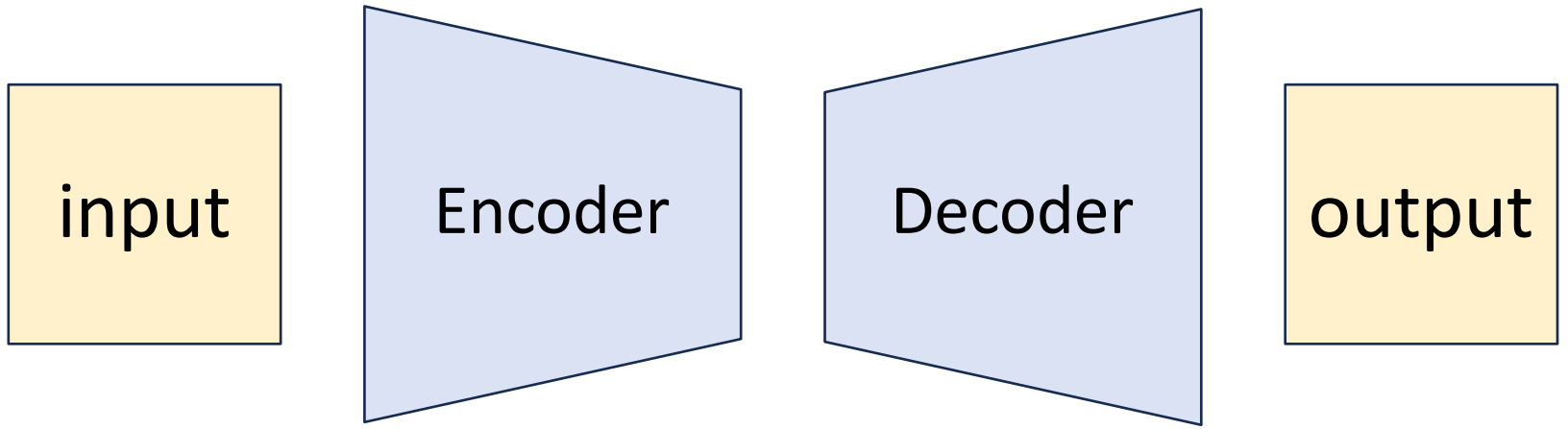
These VAEs are trained on a huge corpus of images pulled from the internet

- e.g. LAION-5B, which was trained on 5 billion scraped images

Image generation

Image generators like DALL-E and Stable Diffusion work by applying a **diffusion** process to images in an encoded space.

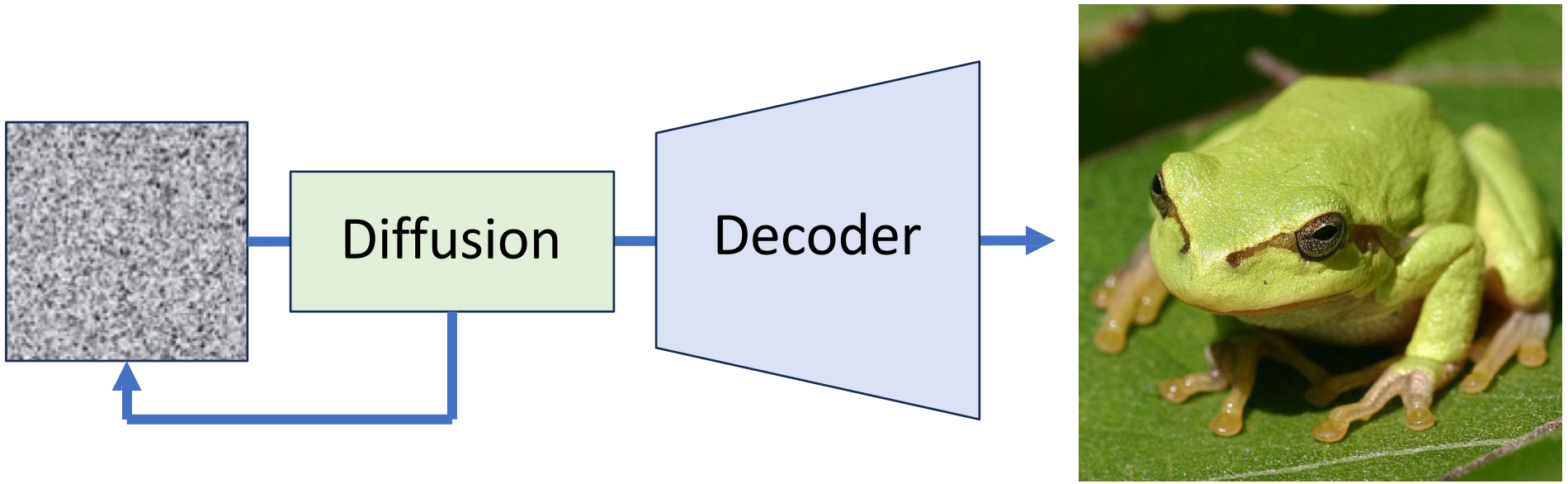
- A VAE is responsible for the encoding/decoding steps

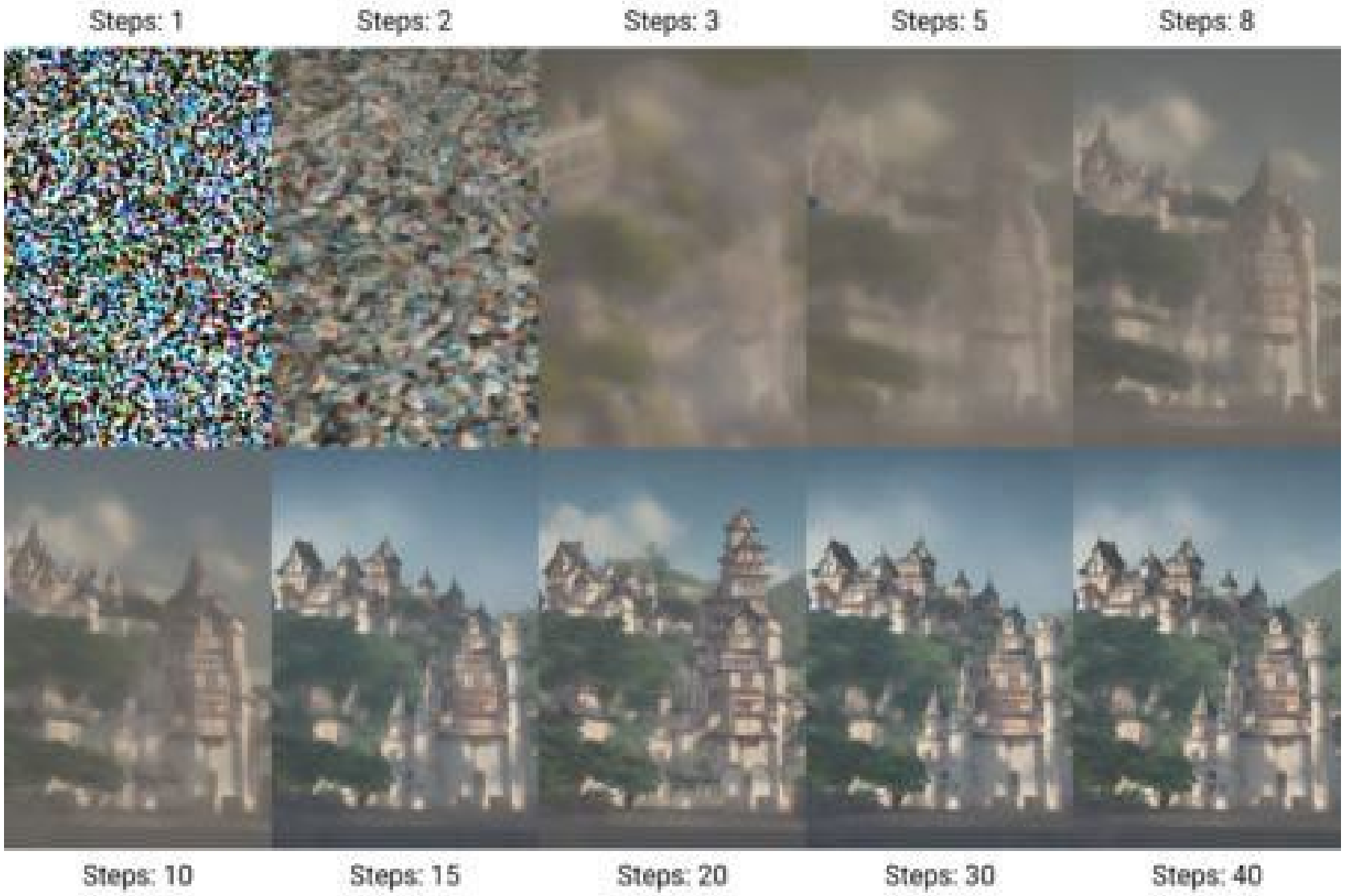


Diffusion

A **diffusion** network is trained to remove noise from images.

We can use this to generate images by applying it repeatedly to a starting image of random noise.





https://en.wikipedia.org/wiki/Stable_Diffusion#/media/File:X-Y_plot_of_algorithmically-generated_AI_art_of_European-style_castle_in_Japan_demonstrating_DDIM_diffusion_steps.png

Image generator

The final diffusion-based image generator looks like the following:

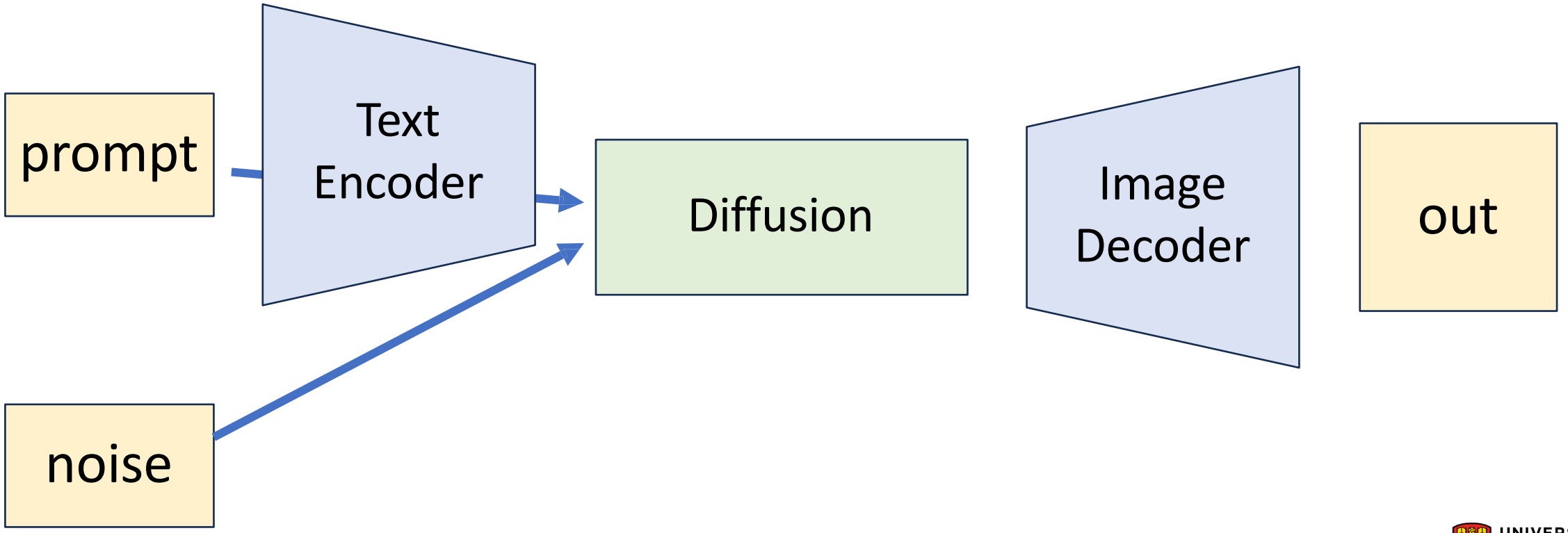
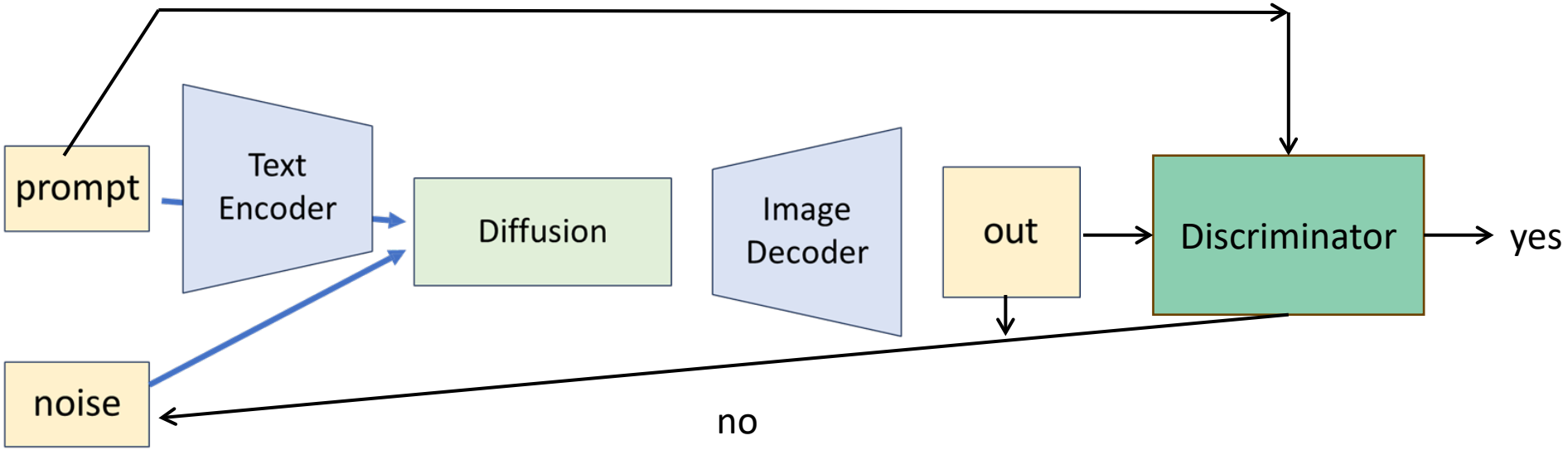


Image generator

The final diffusion-based image generator looks like the following:



Show me a random image

Here are two random, surreal landscapes—let me know if you want to see more or explore a different theme!

https://chatgpt.com/

show me a picture of random pixels

Here are two abstract images of randomly arranged pixels—let me know if you want to explore more patterns or styles!

https://chatgpt.com/

Can you show me actual random pixels?

Generating truly random pixels—like digital noise where each pixel's color is completely independent and random—falls outside the scope of my image generation abilities. However, I can create abstract representations that resemble random pixel noise.

Would you like me to try a denser, more chaotic version, or are you looking for something specific like static from an old TV screen?

https://chatgpt.com/

Next…language processing

Jonathan Hudson, Ph.D.
jwhudson@ucalgary.ca
https://cspages.ucalgary.ca/~jwhudson/

UNIVERSITY OF CALGARY