Multimodal Generative Models

Think About…

- Think about the image generator we used in lab. How might its training data be structured? Where would we get it?

- How might we set up a model so its *output* is an *image*? Think about how you could combine the building blocks we've used (convolutional nets, Transformers, etc.)

Which was real?

- First slide: FFHQ dataset (https://github.com/NVlabs/ffhq-dataset)
- Second slide: mid-training snapshot of fine-tuning StyleGAN2

This Artwork Does Not Exist

Using generative adversarial networks (GAN), we can learn how to create realistic-looking fake versions of almost anything, as shown by this collection of sites that have sprung up in the past month. Learn how it works.

This Person Does Not Exist

The site that started it all, with the name that says it all. Created using a style-based generative adversarial network (StyleGAN), this website had the tech community buzzing with excitement and intrigue and inspired many more sites.

Created by Phillip Wang.

This Cat Does Not Exist

These purr-fect GAN-made cats will freshen your feeline-gs and make you wish you could reach through your screen and cuddle them. Once in a while the cats have visual deformities due to imperfections in the model – beware, they can cause nightmares.

Created by Ryan Hoover.

This Rental Does Not Exist

Why bother trying to look for the perfect home when you can create one instead? Just find a listing you like, buy some land, build it, and then enjoy the rest of your life.

Created by Christopher Schmidt.

https://thisxdoesnotexist.com/

How might an AI *generate* an image?

Goal: **Why** can generative models work?

- For details on *how*, see the readings for this week
- What about the world makes this work?

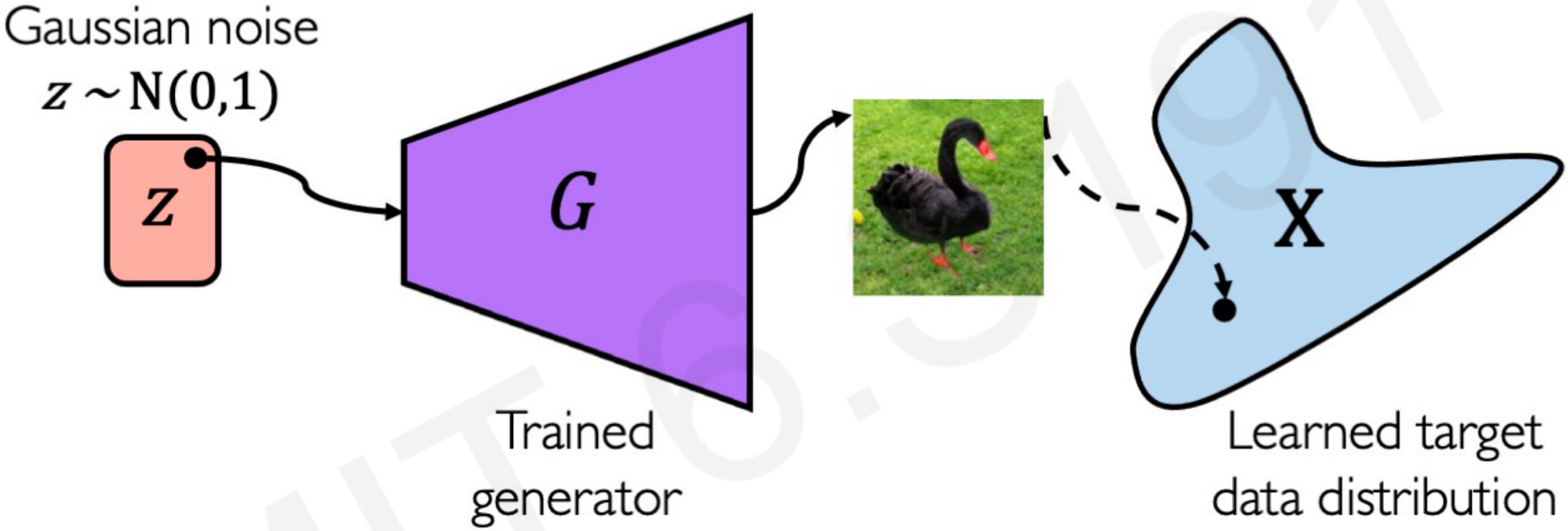
Main Ideas Today

- Real things occupy a tiny portion of the space of all possible things.
  - Images, sentences, sounds, etc.
  - We can describe that tiny space as a *distribution*
- We can compose learned functions in various ways to model that distribution

Real things occupy a tiny portion of the space of all possible things.

- Natural images are a small subset of all possible images

Generator Network



Demo: https://observablehq.com/@stwind/latent-flowers-ganden

http://introtodeeplearning.com/slides/6S191_MIT_DeepLearning_L4.pdf

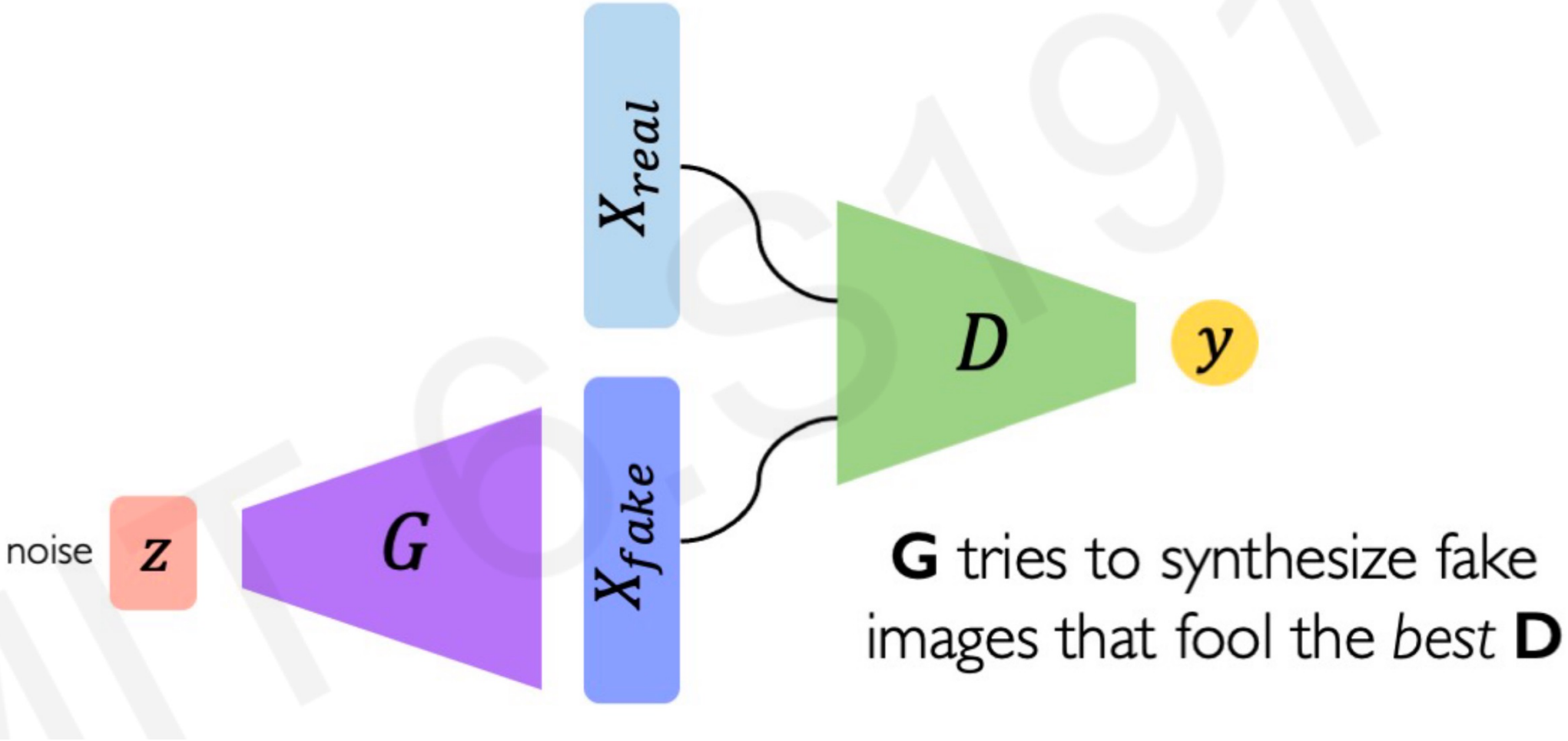
How to compose functions to model distributions?

And how do we train those functions??

Ways of modeling distributions

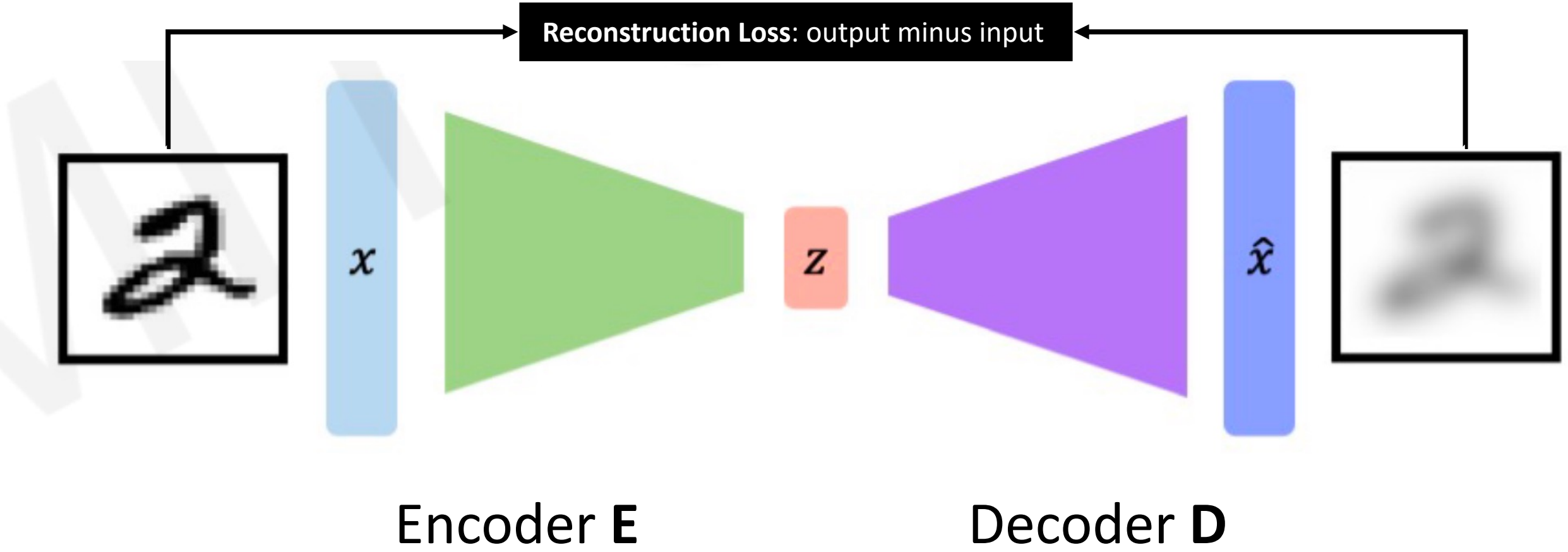
- Single decisions (softmax classifier)
- Sequential decisions (autoregressive)
- Latent variables
- Diffusion-based

Generative Adversarial Network



Try it at [GANLab]

Autoencoder

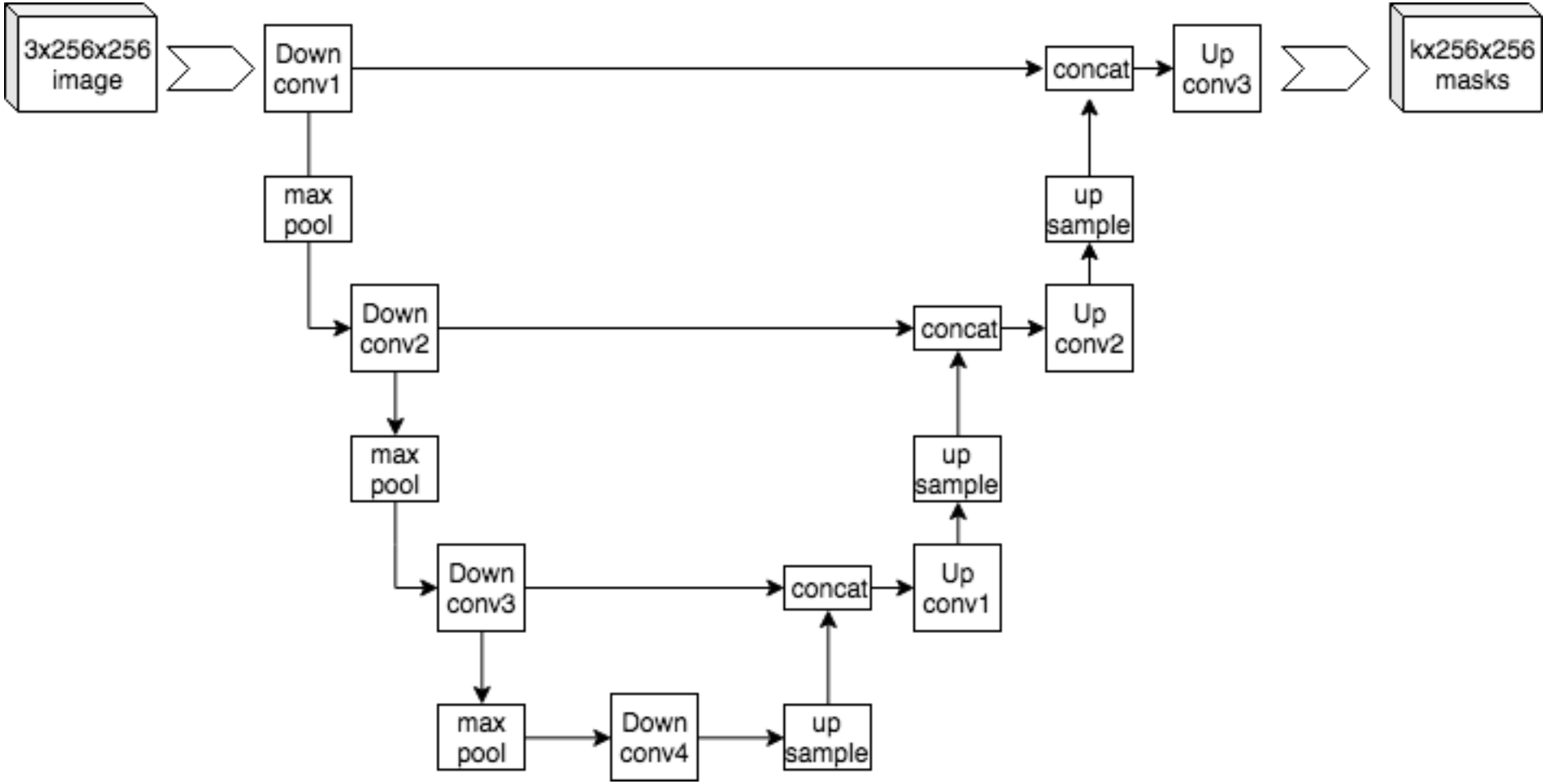


Encoder **E**

Decoder **D**

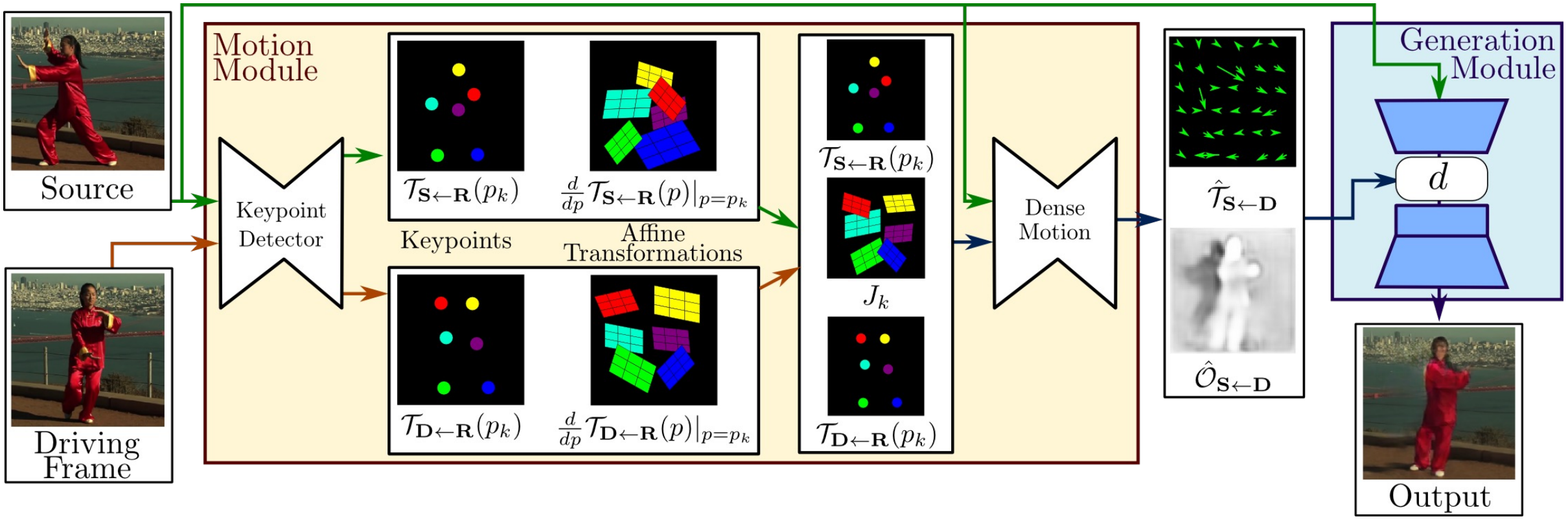
http://introtodeeplearning.com/slides/6S191_MIT_DeepLearning_L4.pdf

U-Net



https://commons.wikimedia.org/wiki/File:Example_architecture_of_U-Net_for_producing_k_256-by-256_image_masks_for_a_256-by-256_RGB_image.png

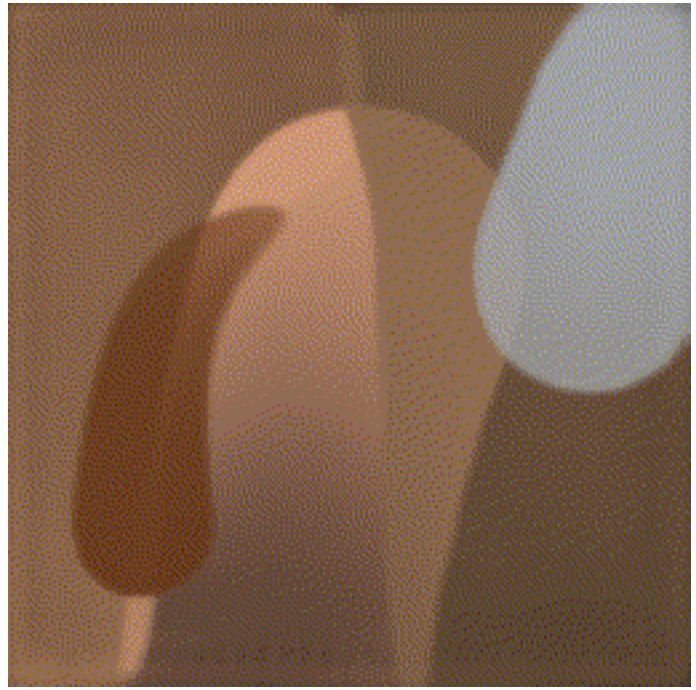
You don’t need all this for a simple deepfake



https://aliaksandrsiarohin.github.io/first-order-model-website/

Demo: https://github.com/eyaler/avatars4all

Many other creative approaches



- Learning to draw an image stroke-by-stroke
  - PaintBot / https://arxiv.org/abs/1904.02201
  - Model-based RL: https://arxiv.org/abs/1903.04411
- Interactive image editing (GANPaint)
- Controls (e.g., GANSpace)
- Use a sequence model (Transformers)
  - https://compvis.github.io/taming-transformers/
- Neural Rendering (NeRF)
  - https://github.com/Kai-46/nerfplusplus
  - https://github.com/yenchenlin/awesome-NeRF

Applications

- Video conferencing
  - Background segmentation (e.g., virtual backgrounds)
  - regenerate the face on the client side (e.g., NVIDIA Maxine)
- Sim2real: train robots in simulations that perform well in real world
- Super-resolution, image/sound restoration
- Inpainting: remove distracting objects (now in Photoshop)
- Anomaly detection (e.g., tumor detection)
- Causal inference

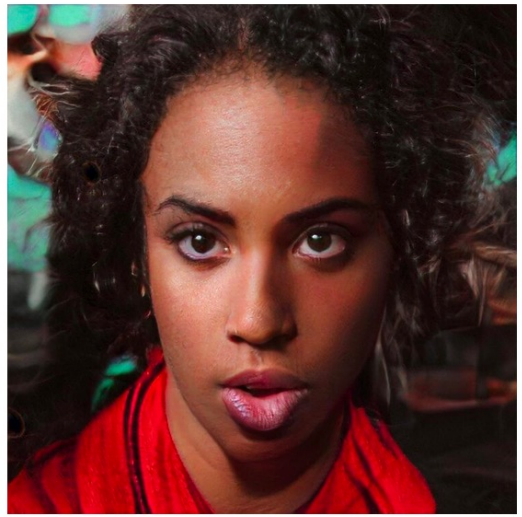
Practical tips

- Use pre-trained models
- Use battle-tested code
- Start with already-running code and tweak it

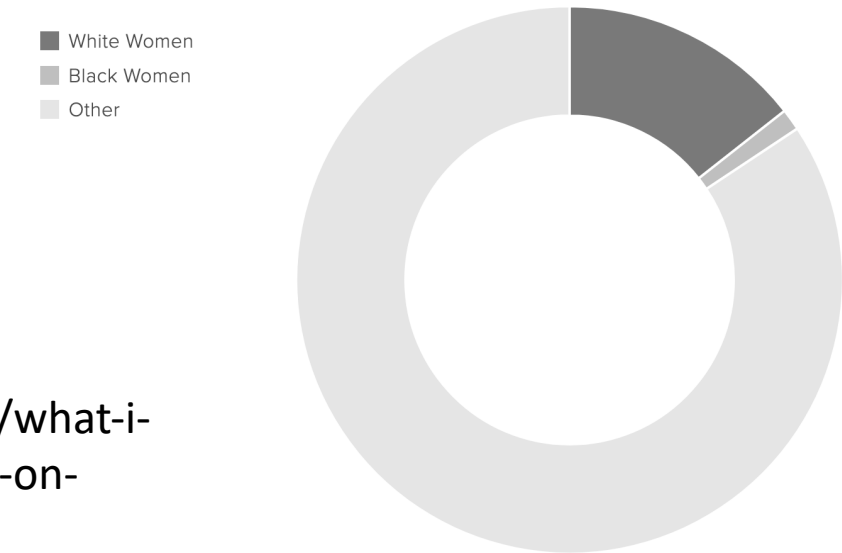
Perspectives and Ethical Considerations

This Black Woman Does Not Exist

Below you see the first three “black” women generated in that session by StyleGAN, and one randomly selected “white” woman from the same session for comparison. The first image is less detailed, the following images is highly distorted; the final image (and notably a small variation of the other) is somewhat convincing, but unflattering.

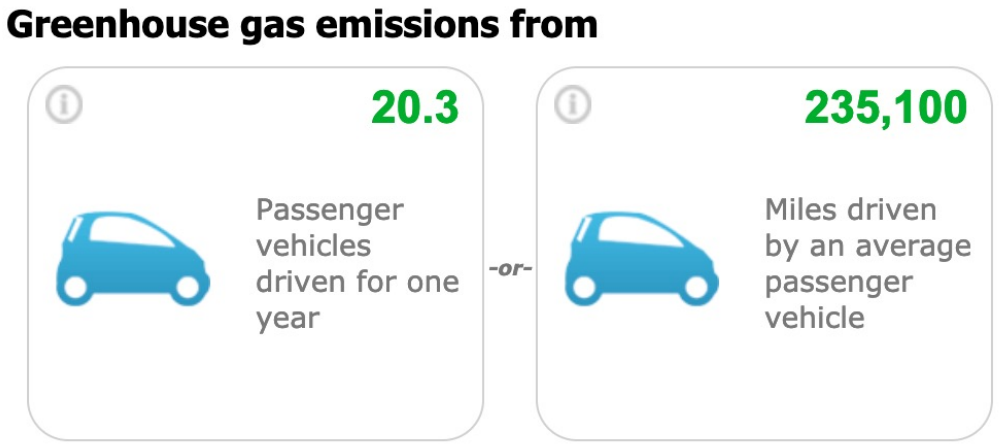



https://www.intersectsy.com/intersectsy/2019/10/3/what-i-learned-about-bias-from-the-first-ai-generated-faces-on-wikimedia-commons



StyleGAN2's Energy Cost

The entire project, including all exploration, consumed 132 MWh of electricity, of which 0.68 MWh went into training the final FFHQ model. In total, we used about 51 single-GPU years of computation (Volta class GPU). A more detailed discussion is available in Appendix F.

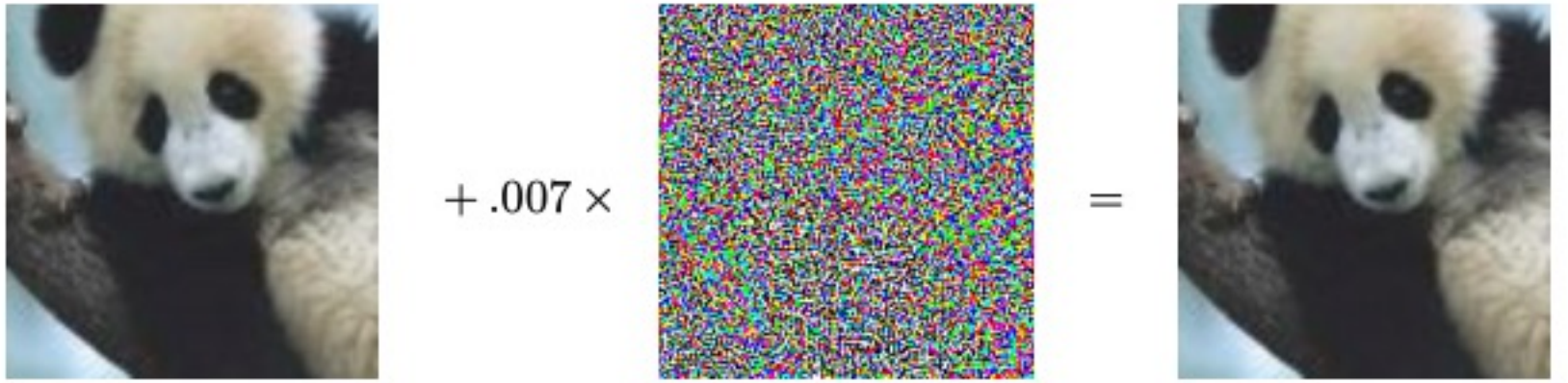


https://arxiv.org/abs/1912.04958

https://www.epa.gov/energy/greenhouse-gas-equivalencies-calculator

Adversarial *Examples*

- Generating an image to fool a classifier



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

Who owns the image? Who owns the model?

Other questions?