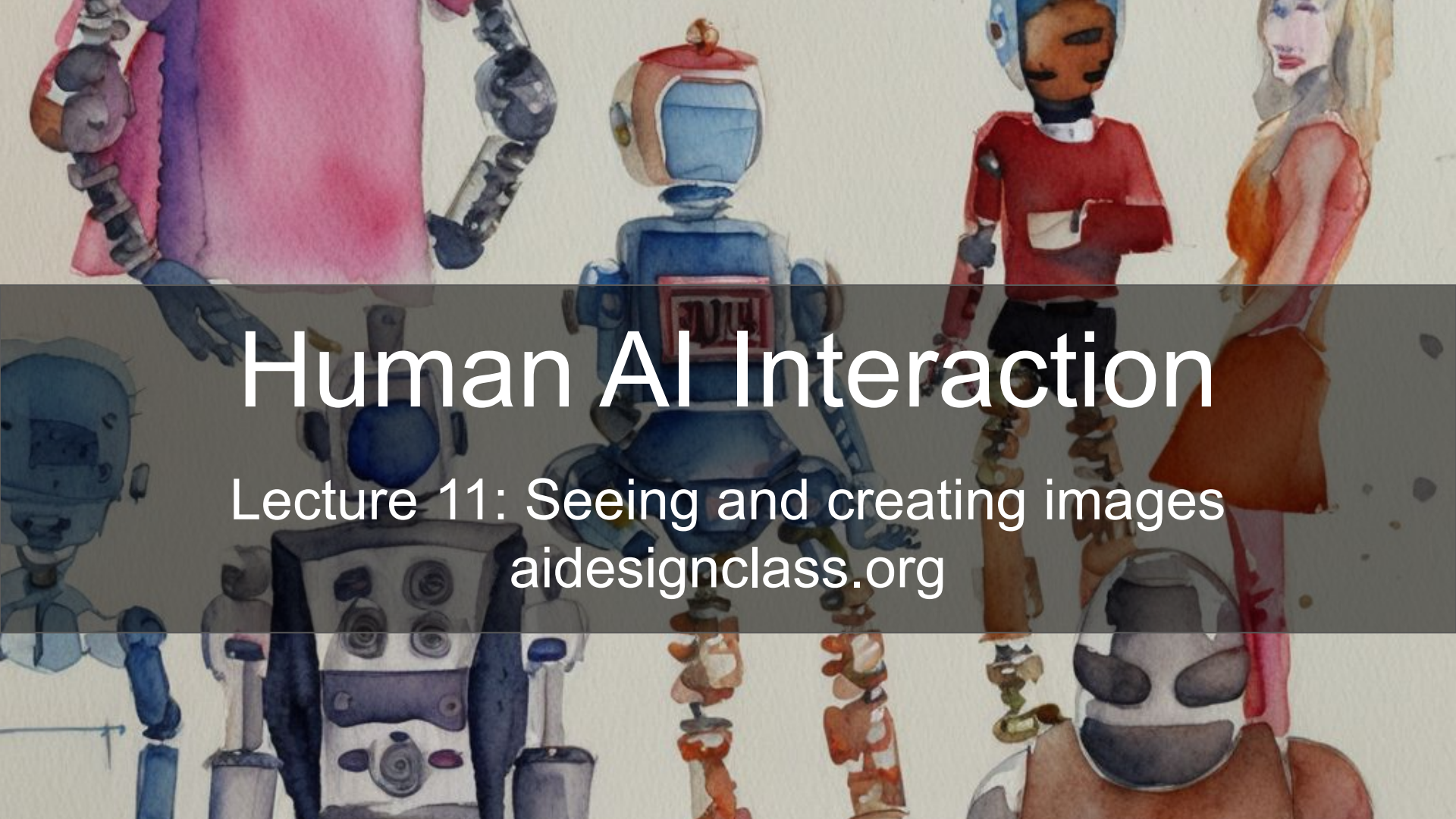


Human AI Interaction

Lecture 11: Seeing and creating images

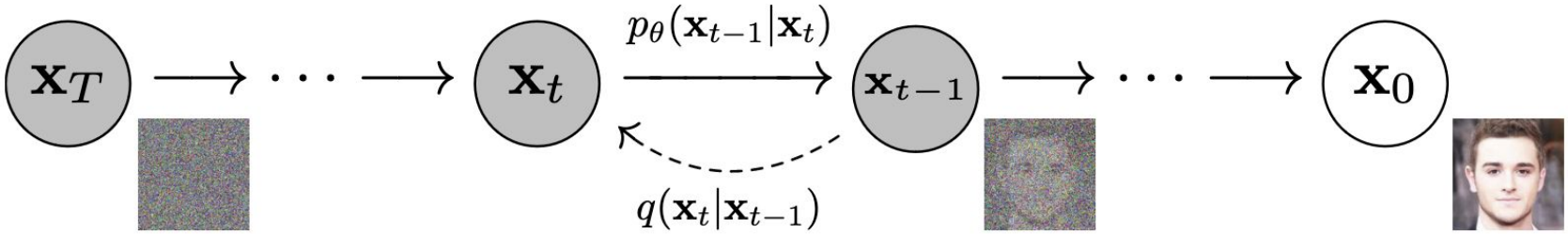
aidesignclass.org

Today

- Image generation
  - What is possible and why (gentle ML intro)
  - User goals with image generation - and tying it back to ML tasks
  - Images and culture
  - Questions of credit and copyright
- Computer vision
  - What is possible and why
  - User goals with vision - and tying it back to ML tasks
  - Vision and bias
  - Privacy

Basic idea behind diffusion

Learn how to “denoise” an image



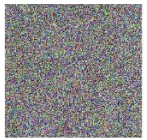
From Denoising Diffusion Probabilistic Models https://arxiv.org/pdf/2006.11239.pdf

Some fun outcomes of diffusion models

Which features are preserved?

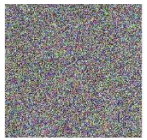
Some fun outcomes of diffusion models

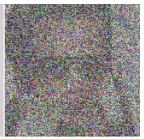
How does this face look like? Are they wearing sunglasses? Color of hair? Possible gender?



Some fun outcomes of diffusion models

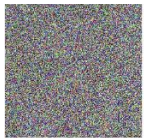
How does this face look like? Are they wearing sunglasses? Color of hair? Possible gender?





Some fun outcomes of diffusion models

How does this face look like? Are they wearing sunglasses? Color of hair? Possible gender?



Some fun outcomes of diffusion models

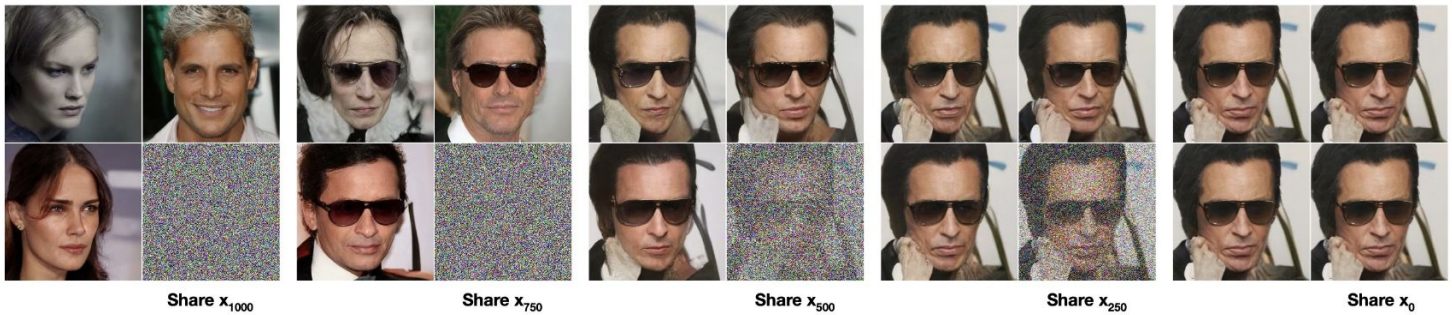
How does this face look like? Are they wearing sunglasses? Color of hair? Possible gender?

Some fun outcomes of diffusion models

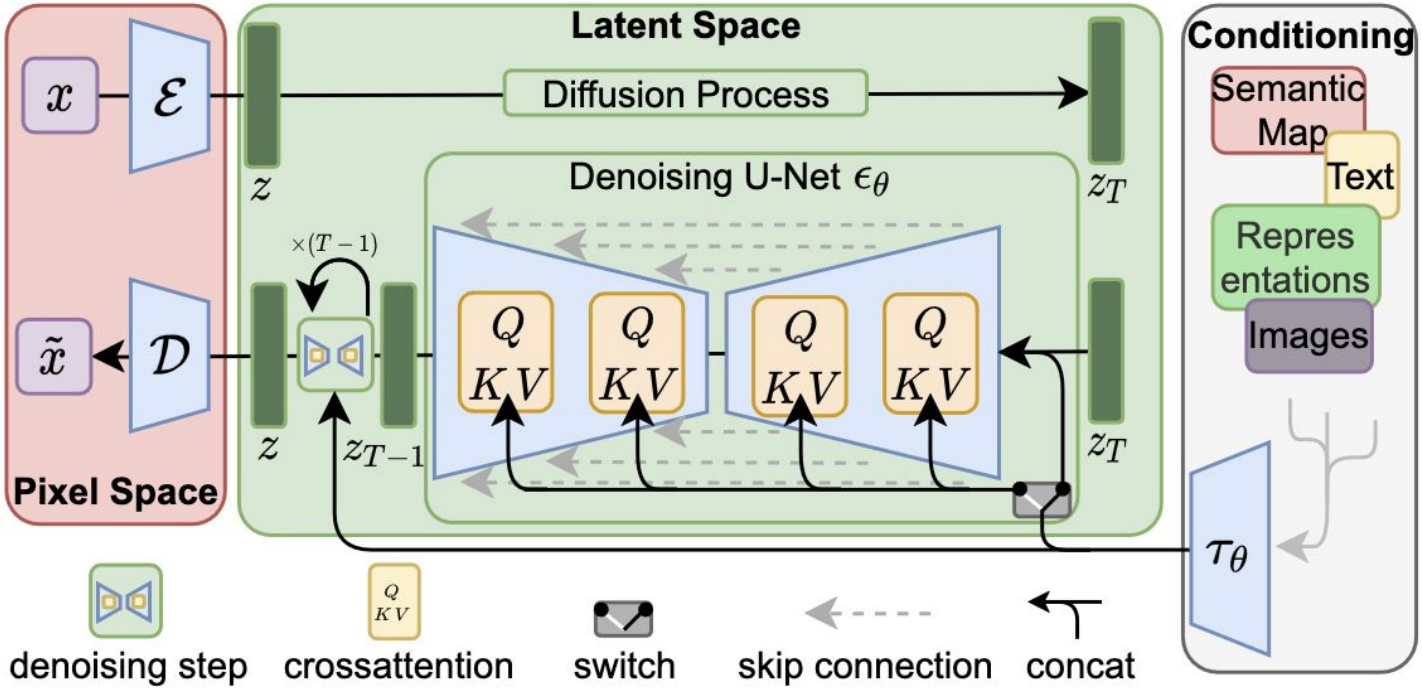
What variants emerge from the same noise?

Some fun outcomes of diffusion models

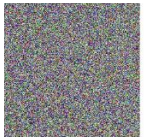
What variants emerge from the same noise?



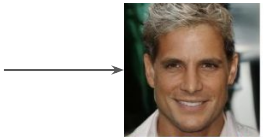
Guiding diffusion with text

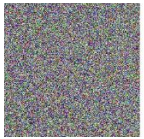


At a conceptual level



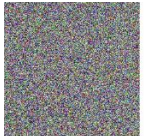
\+ “There is a guy in this image, wearing a white shirt. He has graying hair”





\+ “There is a woman in this image, with dark hair”

What can we do with this?

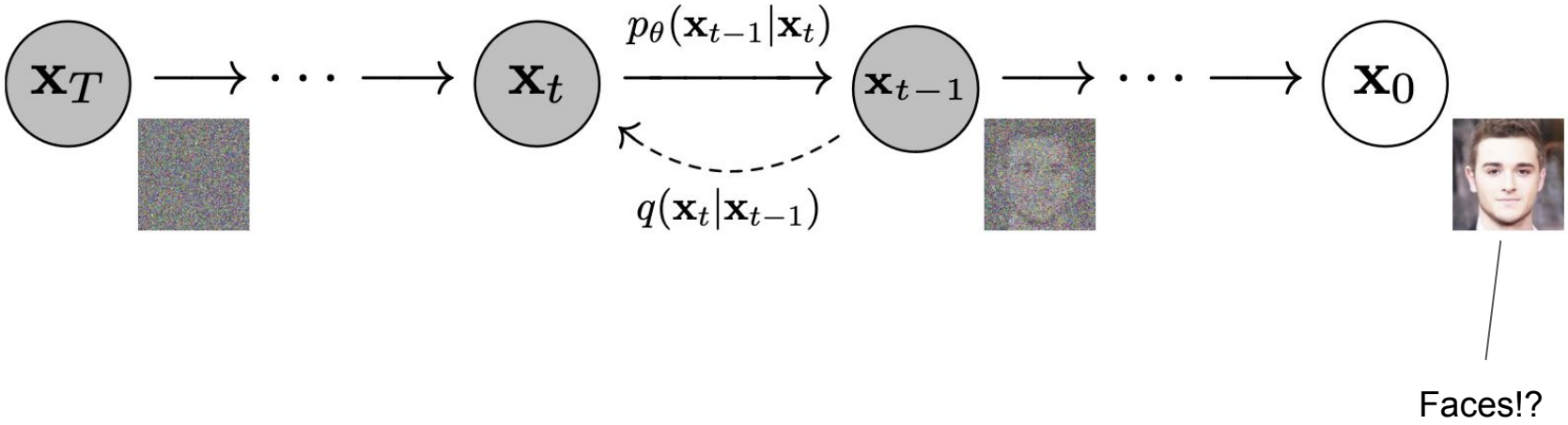
- Generate an image given a text description: start with text + 

- Extend an image (start with image, add noise to edges, help with text)

- Replace an object in the image - how?
- Change “style” – e.g. photo to painting – how?

Brainstorm a few other ideas!

What data should we train on?



Some options

- Photographs of Faces
- Good photos you have taken
- All the photos you have taken
- Just all the photos ever
- Paintings of historic (no longer alive) artists
- Paintings by live artists
- Your own paintings
- All the images in the world
- …

What makes an image a “good” image?

Image from:
https://ai.meta.com/research/publications/emu-enhancing-image-generation-models-using-photogenic-needles-in-a-haystack/

Figure 1. With quality-tuning, Emu generates *highly* aesthetic images. Prompts: (top) a glass of orange juice; a woman in an apron works at a local bar; a coffee mug; (bottom) an egg and a bird made of wheat bread; a corgi; a shake is next to a cake.

Meta’s choice of “good”

“In the first stage, we train generalist annotators to down select the image pool to 20K images. Our primary goal during this stage is to optimize recall, ensuring the exclusion of medium and low quality that may have passed through the automatic filtering. In the second stage, we engage specialist annotators who have a good understanding of a set of photography principles. Their task is to filter and select images of the highest aesthetic quality (see Figure). During this stage, we focus on optimizing precision, meaning we aim to select only **the very best images.”** (emphasis mine)

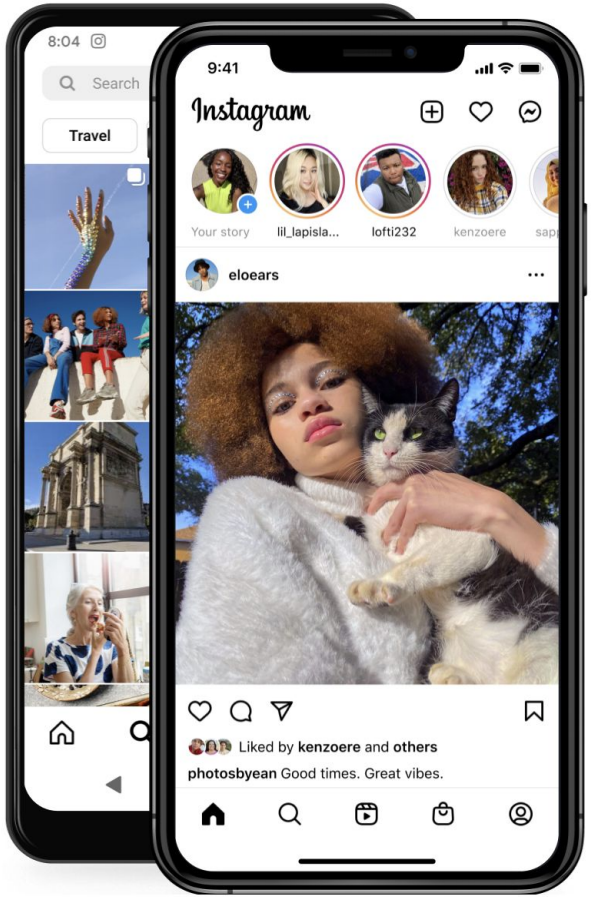
Ended up with 2000 “finetuning” images

Figure 4. **Visual Appealing Data**. Examples of visually appealing data that can meet our human filtering criterion.

What else is possible for good images?

Image from Instagram on right

How would you choose “good” images if you wanted to use text-to-image for Instagram?



What is a good image?

Native American chiefs, 1865

From https://medium.com/@socialcreature/ai-and-the-american-smile-76d23a0fbfaf

What is a good image?

How people depict themselves in a photo is culturally informed

- Do you smile?
- What kind of smile?

What we choose as a good image is our choice about how we think the world looks. Perhaps we shouldn't be so arrogant about our choices.

Native American chiefs, 1865

From https://medium.com/@socialcreature/ai-and-the-american-smile-76d23a0fbfaf

Recognizing a cat

The original idea is to not teach a computer exactly what to look for… instead let it discover high level features from pixels.

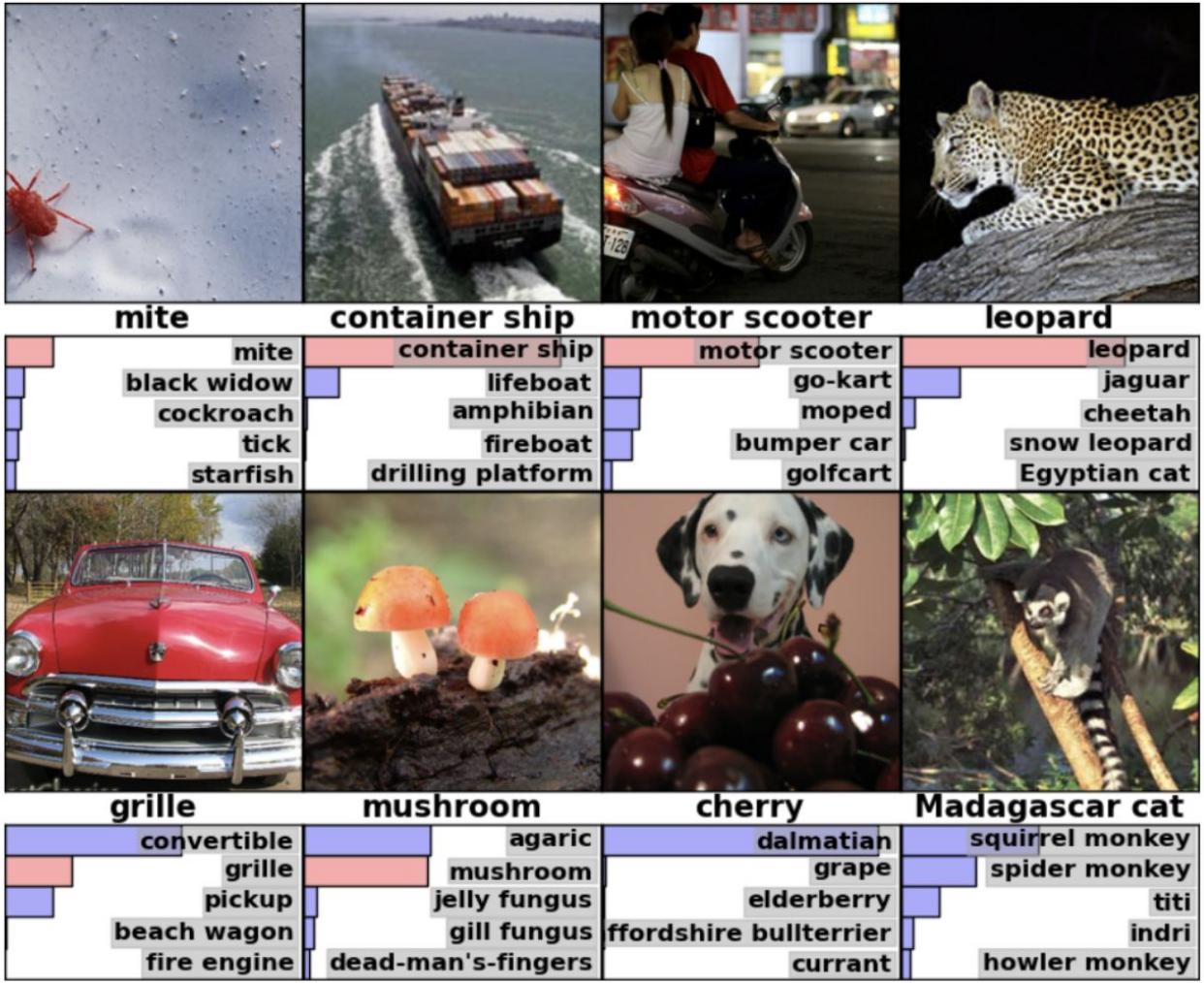
The paper below used frames from 10 million YouTube videos.

From Building High-level Features Using Large Scale Unsupervised

Recognize -> classify

To learn a lot of categories, you need to know the name and how they look

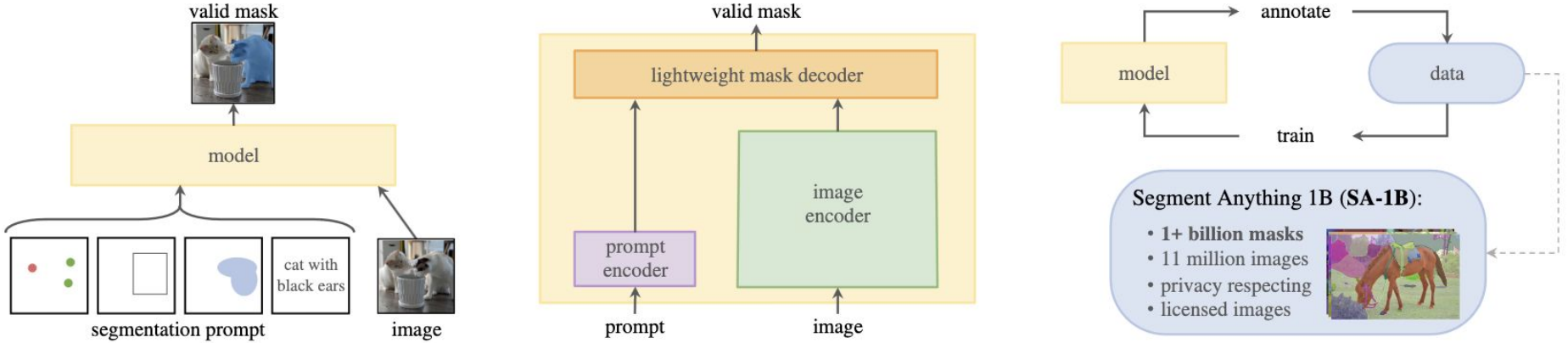
ImageNet has ~22k categories



From [ImageNet Classification with Deep Convolutional Neural Networks]

Segmentation

How it works



(a) **Task**: promptable segmentation (b) **Model**: Segment Anything Model (**SAM**) (c) **Data**: data engine (top) & dataset (bottom)

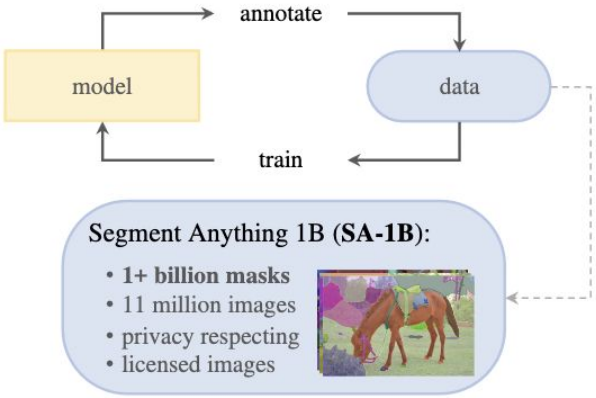
Figure 1: We aim to build a foundation model for segmentation by introducing three interconnected components: a promptable segmentation *task*, a segmentation *model* (SAM) that powers data annotation and enables zero-shot transfer to a range of tasks via prompt engineering, and a *data* engine for collecting SA-1B, our dataset of over 1 billion masks.

Remember the prompt encoder from diffusion models?

The most important advancement is a data flywheel

A “data-engine” or flywheel is a setup that helps you best allocate human effort.

“we first automatically detected confident masks. Then we presented annotators with images prefilled with these masks and asked them to annotate any additional unannotated objects” (from their paper)



(c) **Data**: data engine (top) & dataset (bottom)

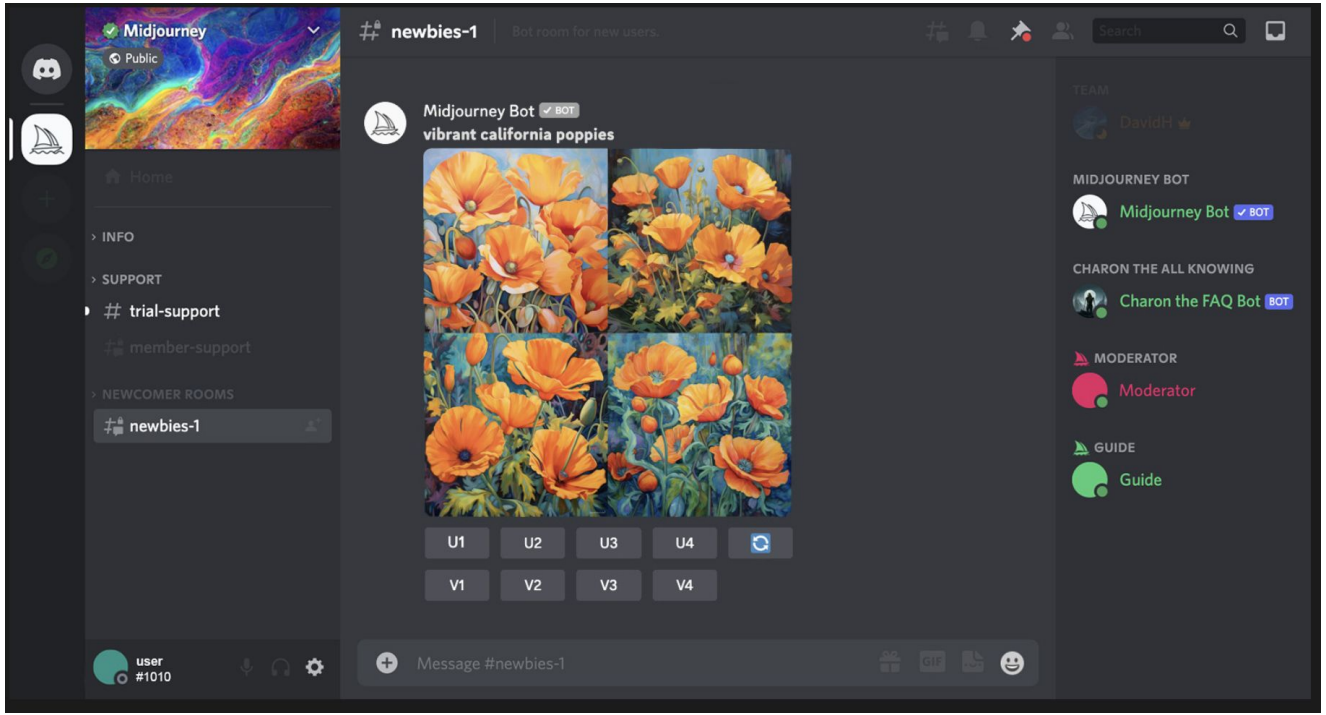
ee interconnected components: a prompt-
and enables zero-shot transfer to a range
et of over 1 billion masks.

Interactive applications need both image understanding and creation

“Replace an object” = label + segment + “in-fill” painting

What does the interaction model for image applications look like?

Image from Midjourney



What is good about this interface?

What is not?

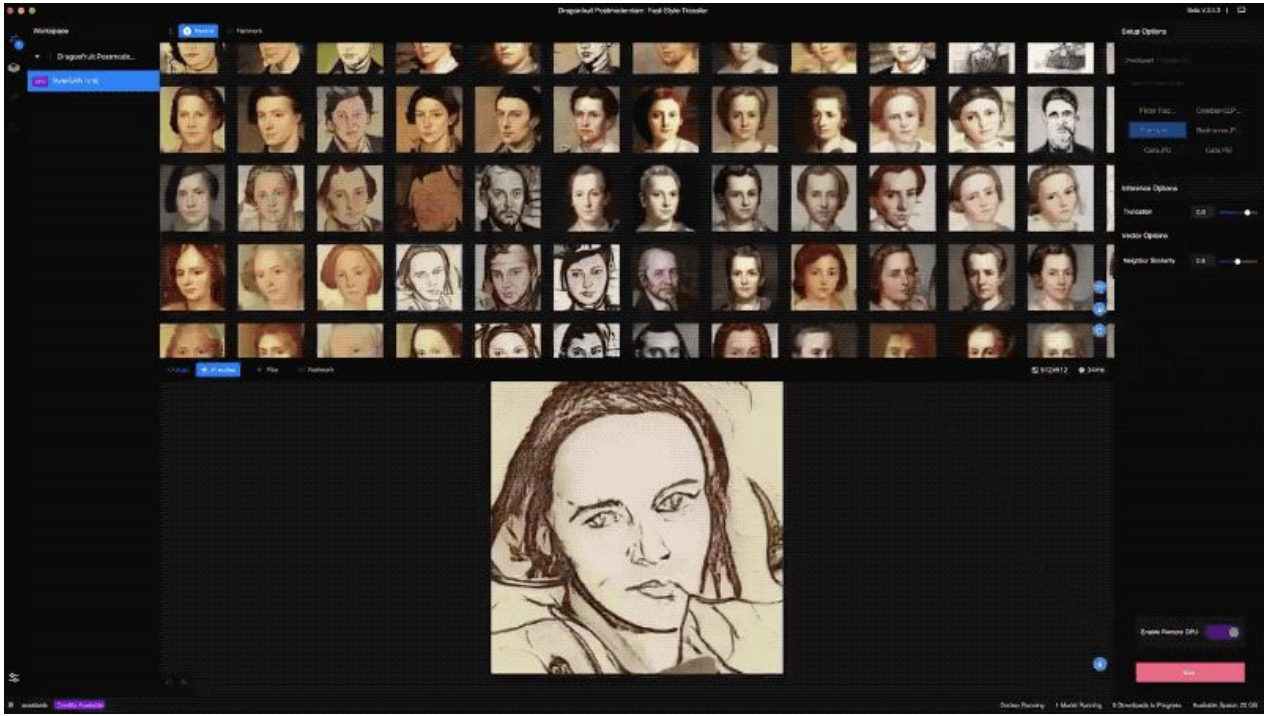
UX for images

RunwayML

Image from Verge

What is good about this UX?

What isn't?



Key takeaways

- Datasets matter – what is aesthetic is not a trivial question
- Joint understanding (text + image, etc) can allow for a variety of new tasks
  - With creative tasks, examples of work are a powerful learning mechanism
- What UX you choose depends on how you want users to learn the creative power of the medium