

# Text-to-Image Generation



Compute vision:

- Image classification
- Object detection
- Segmentation
- Image generation

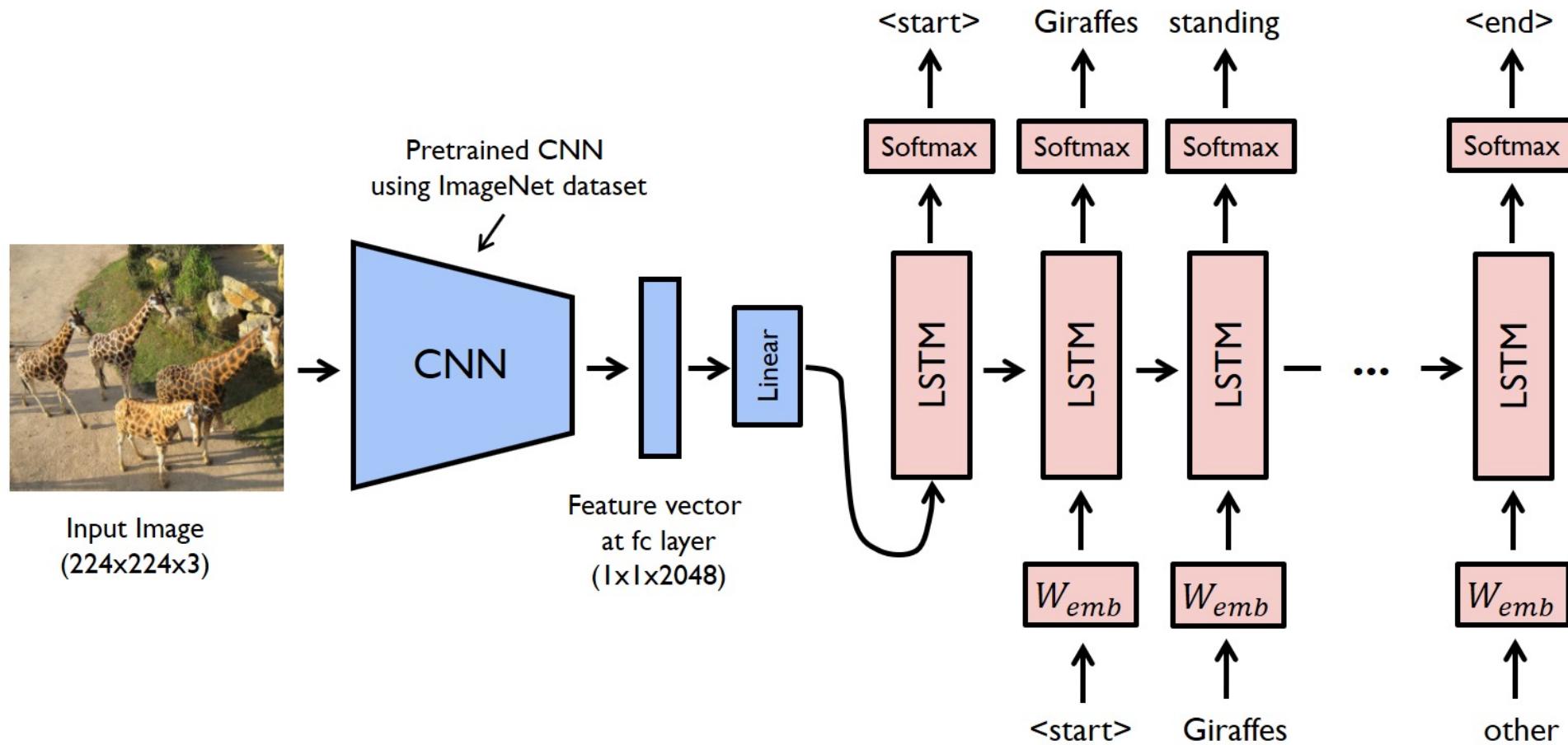
CNN's are widely used

Natural Language Programming:

- Text classification
- Text prediction/ generation
- Translation

Transformers are widely used

# Idea - Mix CV + NLP



# VQ-GAN + CLIP

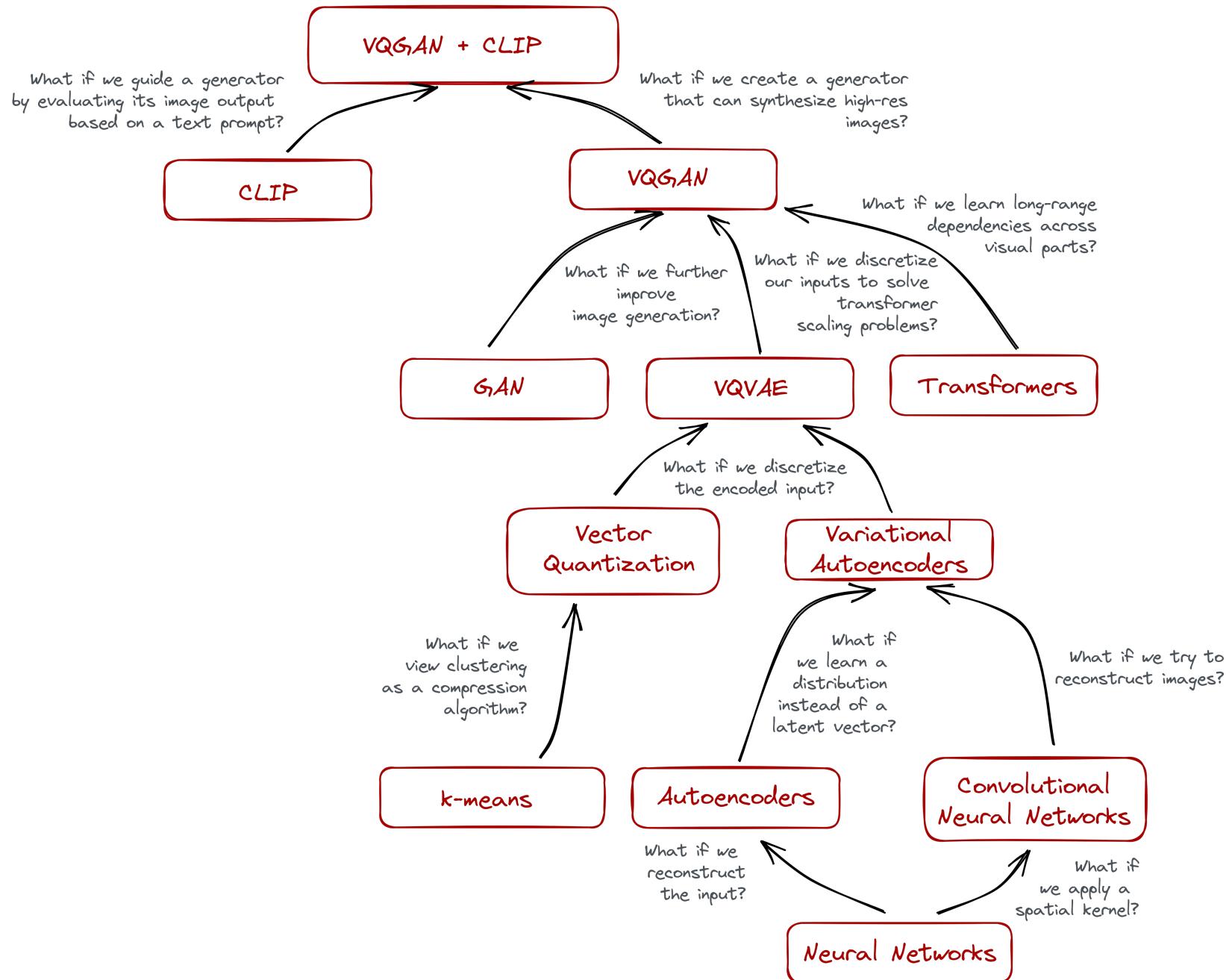
1) VQGAN = **V**ector **Q**uantized **G**enerative **A**dversarial **N**etwork

Takes in a noise vector, and outputs a (realistic) image.

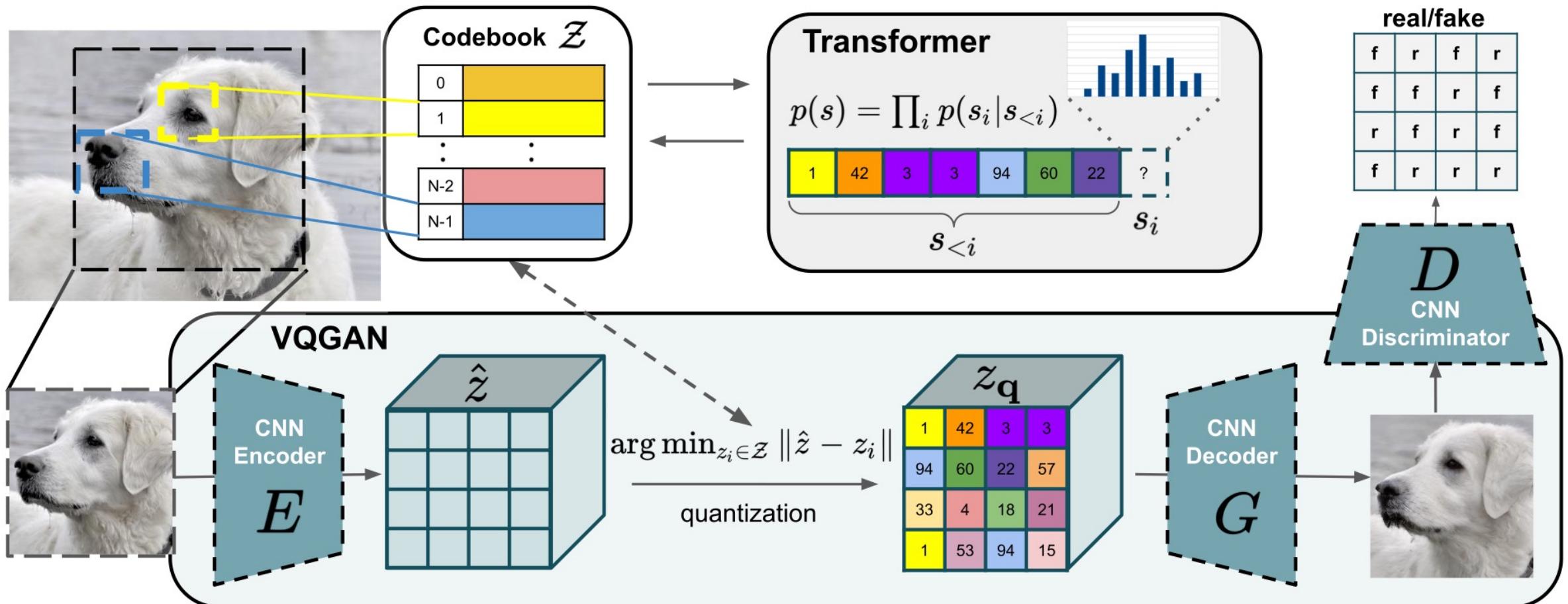
2) CLIP = **C**ontrastive **L**anguage–**I**mage **P**re-training

Takes in:

- (a) an image, and outputs the image features; or
- (b) a text, and outputs text features.

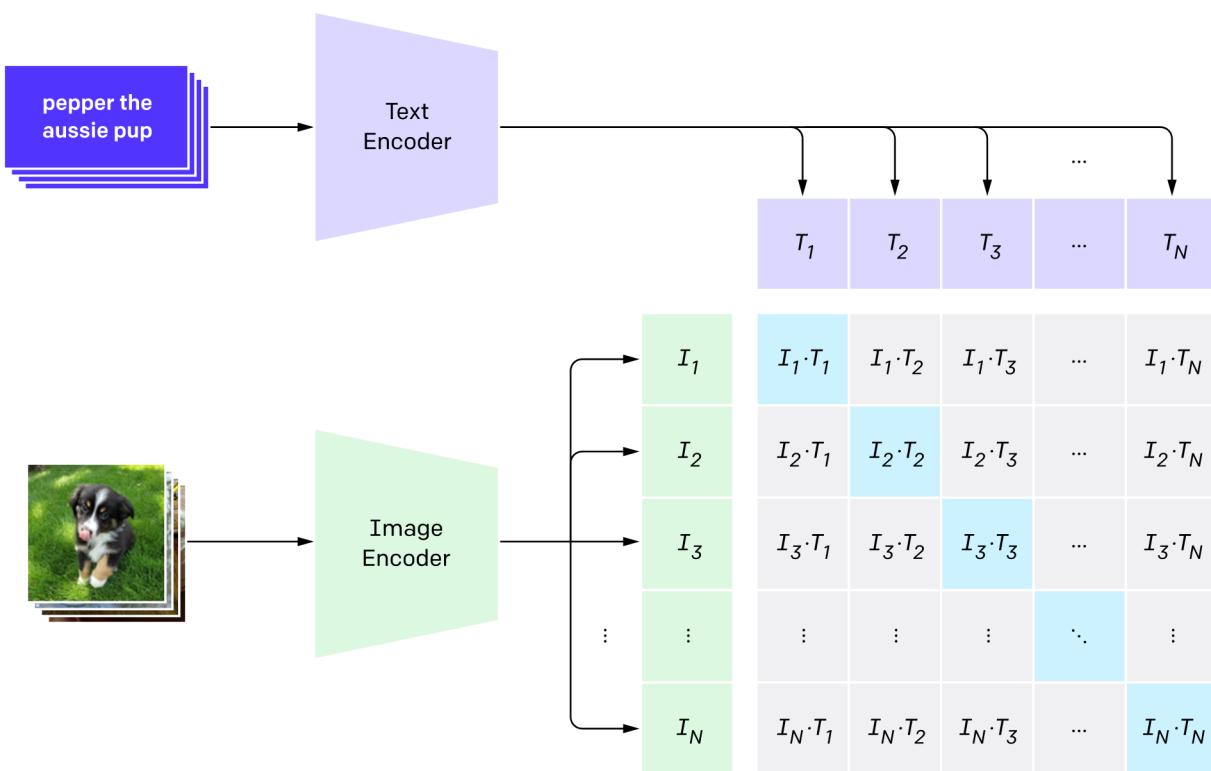


# VQ-GAN

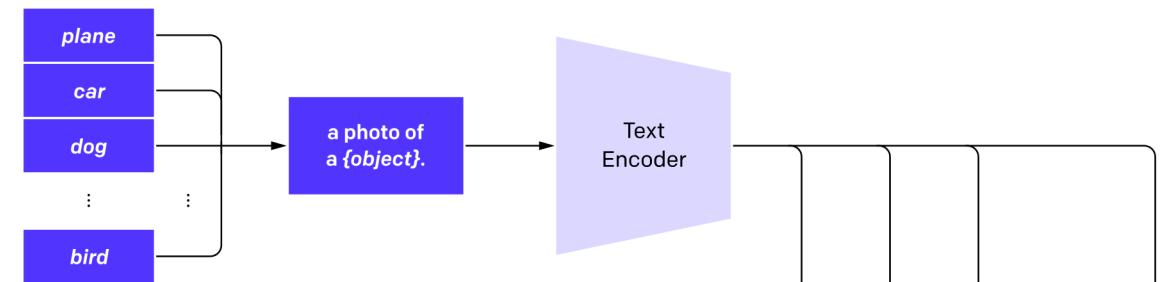


# CLIP

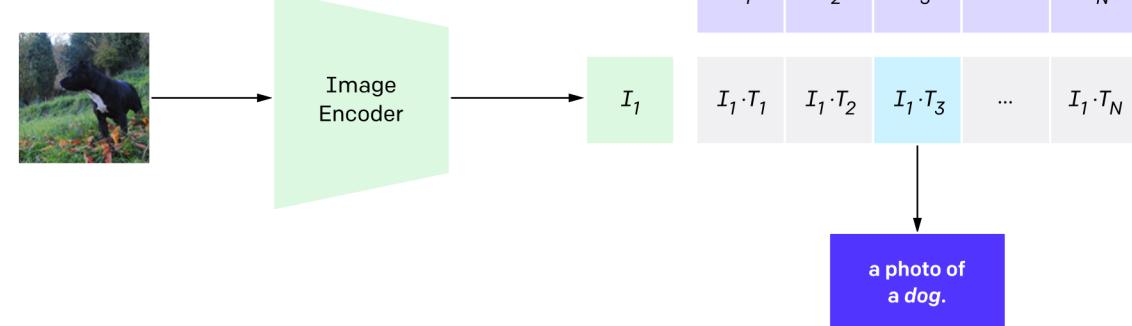
## 1. Contrastive pre-training

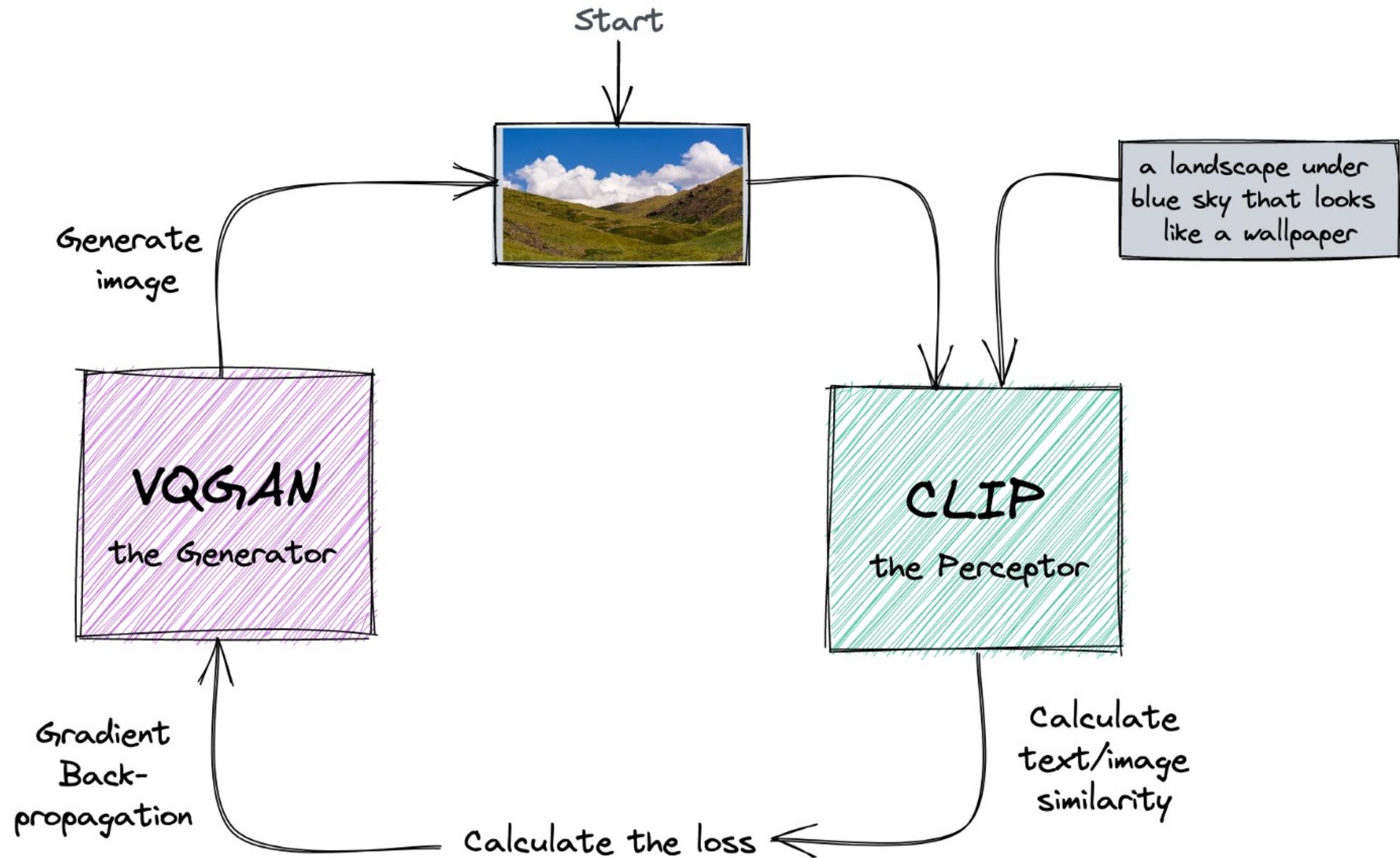


## 2. Create dataset classifier from label text



## 3. Use for zero-shot prediction



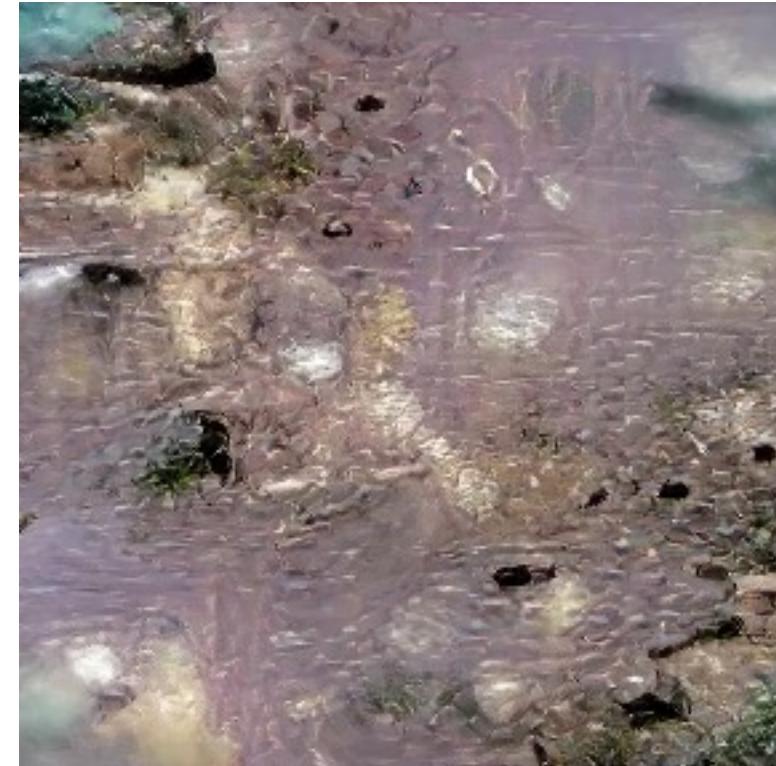


# Hands-on: VQ-GAN + CLIP

yellow flower pigeons flying under water



Frida Kahlo hat cat flowers red clouds



DALL-E and DALL-E2