Image generation foundation models

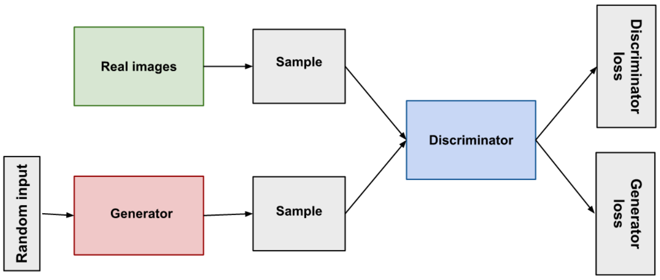
GAN (Generative Adversarial Network)

Generator:
- Generates images from noise
- Tries to trick Discriminator with generated Images

Discriminator
- Tries to predict if image is generated or real
- Used for training the Generator

Use cases:
- Image upscaling / removing noise/artifacts
- Image to image translation (changing style)
- Generating training data for other models

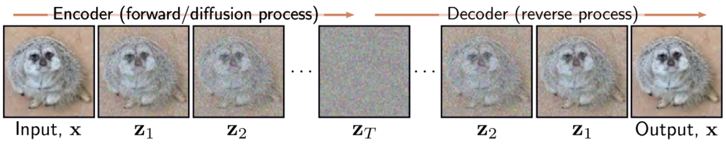
Key models:
- CycleGAN (2017, Jun-Yan Zhu et al.)
- StyleGAN (2018, NVIDIA)



Diffusion Model

Forward process:
- Random noise is added to an images at multiple steps
- Used to train the model

Reverse process/generation:
- Model gets noisy image as input and tries to predict what noise was added at specific step (model learns noise at each time step)
- The predicted noise gets removed from the image
- Cleaned up image gets used as input for next step



Architecture:
- Fist models operated on pixel space (DDMP) -> Slow training and generation time
- Modern models work with Variational Auto Encoder to encode pixels in latent space (LDM). Noise gets predicted in latent space -> Fewer parameters and faster training and generation. Also known as Stable Diffusion

Key models:
- Open Source:
  - Qwen Image (Alibaba Cloud)
  - Stable Diffusion 1.5 / Stable Diffusion XL (Open Source Project)
  - Playground v2.5 (Playground AI)
- Closed Models:
  - DALL-E 3 (OpenAI)
  - Google Imagen 3 (Google)
  - Midjourney V5.2 (Midjourney)

Flow Matching Model

Forward process:
- Random noise is added to an images at multiple steps
- Used to train the model

Reverse process/generation:
- Model gets noisy image as input and tries to flow/velocity of the noise
- With predicted flow the added noise can be removed with fewer bigger steps because the direction of the noise is known -> faster than diffusion because the steps are much bigger
- Cleaned up image gets used as input for next step

Key models:
- Open Source
  - FLUX.1 [schnell] (Black Forest Labs)
  - Stable Diffusion 3.5 / Stable Diffusion XL (Open Source Project)
- Closed Source Models
  - Google Imagen 3.5 (Google)
  - Midjourney V7 (Midjourney)

Image sources

- https://developers.google.com/machine-learning/gan/gan_structure?hl=de
- https://github.com/udlbook/udlbook/releases/download/v5.0.3/UnderstandingDeepLearning_02_09_26_C.pdf

Other Sources

- https://github.com/udlbook/udlbook/releases/download/v5.0.3/UnderstandingDeepLearning_02_09_26_C.pdf
- https://hojonathanho.github.io/diffusion/
- https://github.com/CompVis/latent-diffusion
- https://github.com/junyanz/cyclegan
- https://github.com/nvlabs/stylegan
- https://developers.google.com/machine-learning/gan/gan_structure?hl=de
- https://www.preprints.org/manuscript/202501.1502
- https://developer.nvidia.com/blog/generative-ai-research-spotlight-demystifying-diffusion-based-models/
- https://www.youtube.com/watch?v=iv-5mZ_9CPY
- https://www.youtube.com/watch?v=DDq_plfHqLs
- https://www.youtube.com/watch?v=firXjwZ_6KI