Plan

Recap

Logistics

Diffusion

Architecture

Latent Space

Text Conditioning

Recap

Contrastive Learning

(x, x') positive $\rightarrow$ $f_\theta(x)^T f_\theta(x')$ large

(x, x') negative $\rightarrow$ $f_\theta(x)^T f_\theta(x')$ small

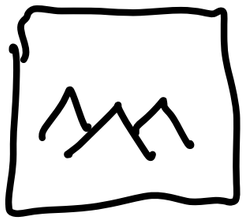
$$\mathcal{L}(\theta) = -2 \underset{x, x^+ \sim pos}{\mathbb{E}} \left[f_\theta(x)^T f_\theta(x^+) \right] + \underset{x, x^- \sim neg}{\mathbb{E}} \left[\left[f_\theta(x)^T f_\theta(x^-) \right]^2 \right]$$

$$\underset{\theta}{\operatorname{argmin}} \mathcal{L}(\theta) = \underset{\theta}{\operatorname{argmin}} \| \bar{A} - F_\theta F_\theta^T \|_F^2$$

θ: Idk

$F_\theta F_\theta^T$: well understood

CLIP



beautiful snowy mountains

$\in \mathbb{R}^k$ $\in \mathbb{R}^k$

Logistics

Form: last two are today and tomorrow

Grades:

↳ self grade due 5pm tonight

↳ come talk with concerns

↳ address issues now

↳ focus <u>on</u> project (30 points)

Homework:

↳ 5pm Wednesday

↳ 5pm Friday self grade

Wednesday

↳ no class/demo

↳ I'm available

↳ board games at 4pm

Thursday

↳ presentations 10-noon

↳ no demo

Project

↳ deliverable: code, talk, report

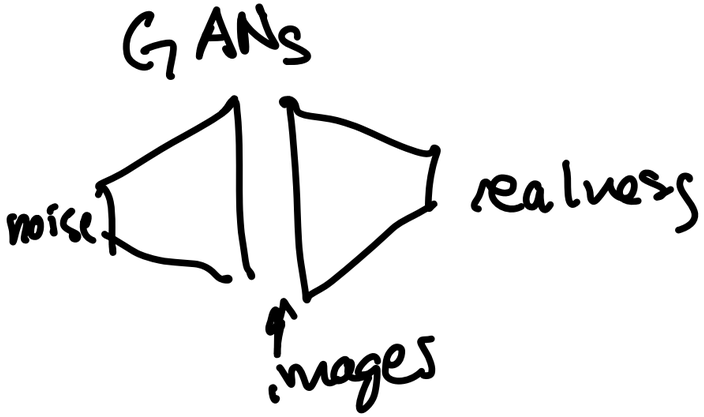
↳ working example

↳ get <u>unstuck</u>

Questions for Projects

↳ I can give <u>high</u> level ideas

↳ post on canvas



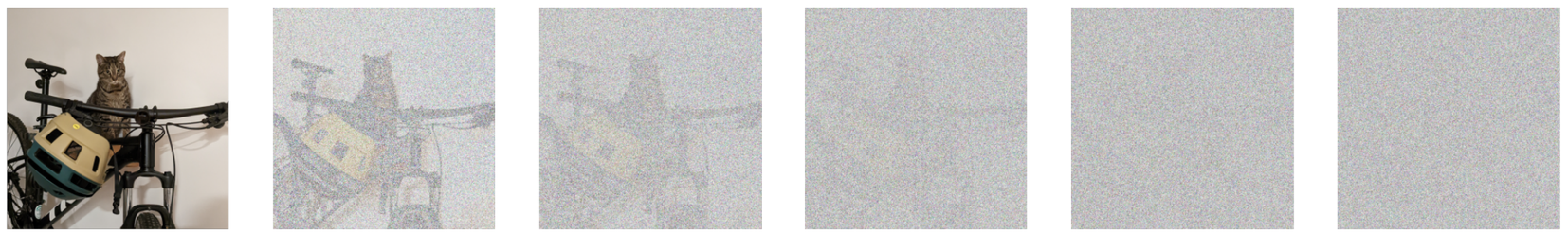
Mode Collapse

↳ WGAN

↳ regularization

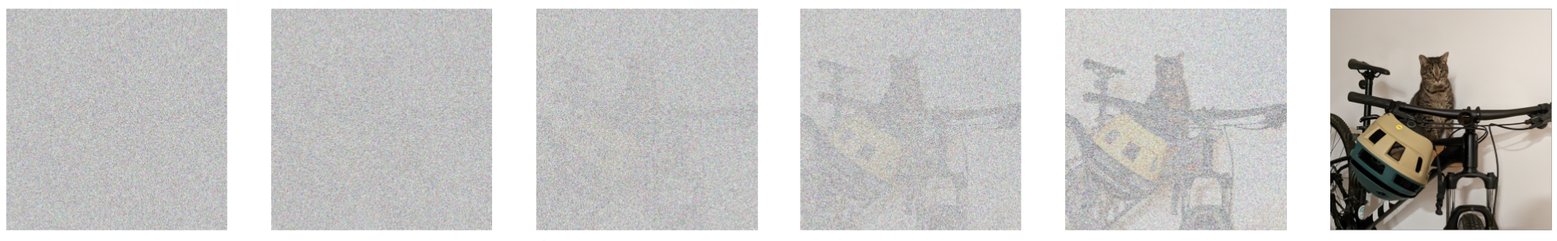
Diffusion

Build image from noise (very hard)

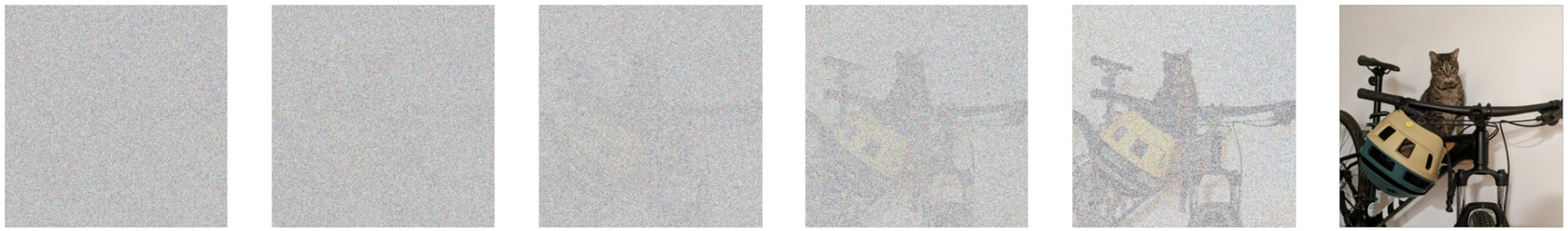


$$x_t = x_{t-1} + \epsilon_t \quad \text{where} \quad \epsilon_t \sim \mathcal{N}(0, \sigma^2 I)$$

ϵ_t: epsilon



$$f_\theta(x_t, t) \approx x_{t-1} \qquad \mathcal{L}(\theta) = \| f_\theta(x_t, t) - x_{t-1} \|_2^2$$



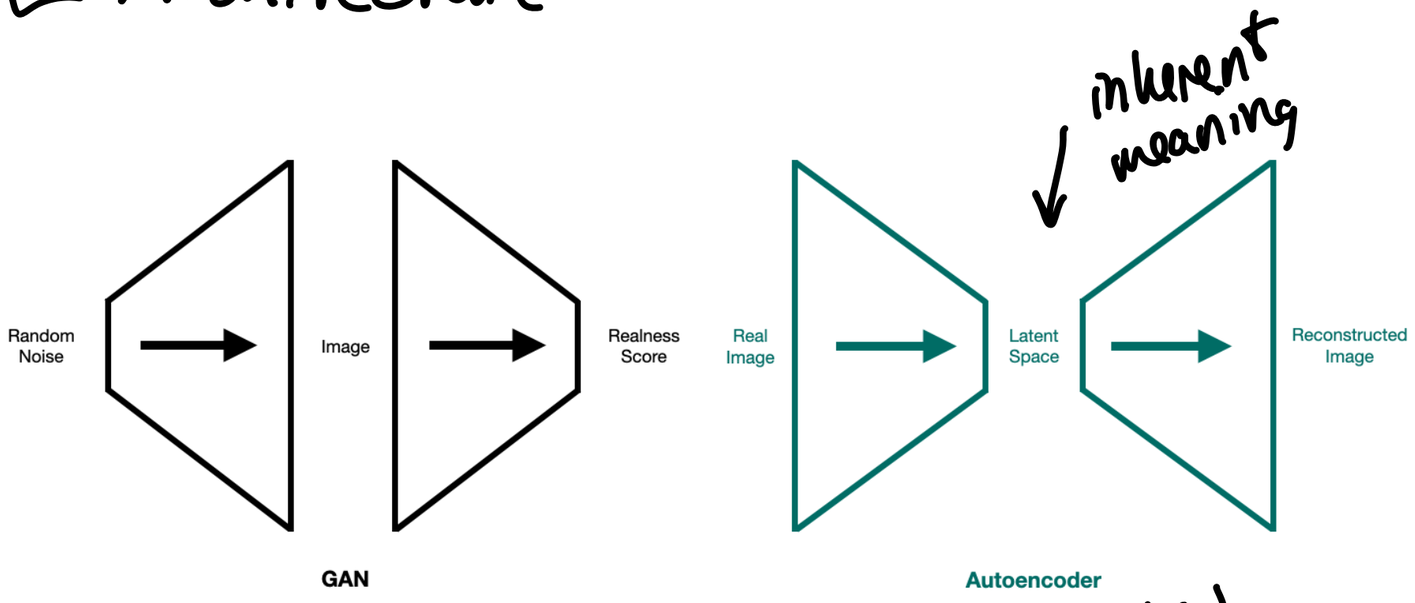
$$f_\theta(x_t, t) \approx \epsilon_t$$

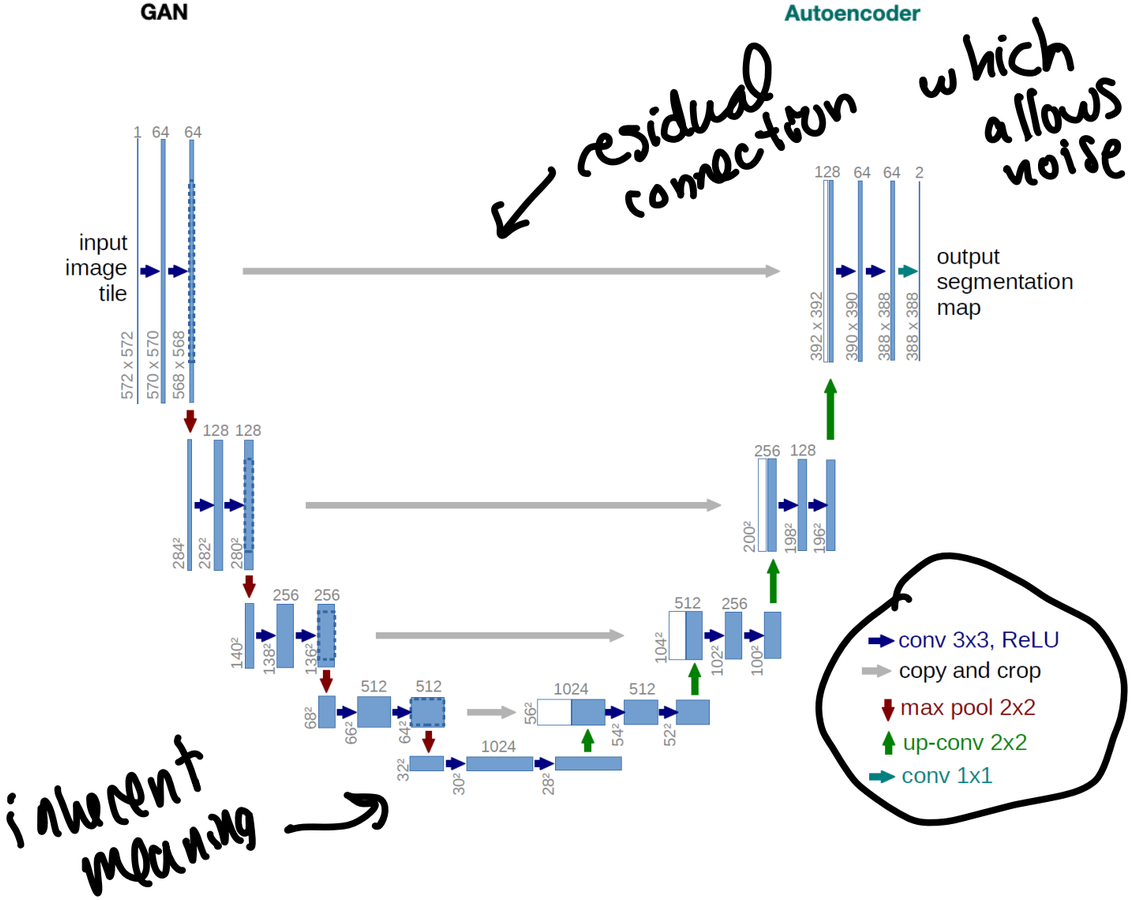
$$x_{t-1} = x_t - \epsilon_t$$

$$\mathcal{L}(\theta) = \mathbb{E}_{x_t, \epsilon_t}\left[\| f_\theta(x_t, t) - \epsilon_t \|_2^2\right]$$

- ☑ Loss
- ☑ Optimizer

☑ Architecture





Diffusion in Latent Space

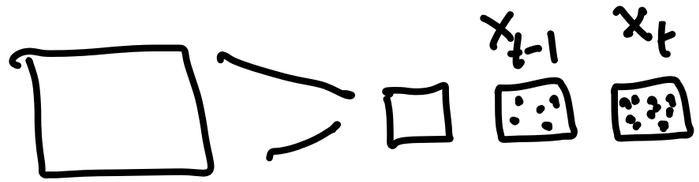
Problems

↳ really big architecture ✓

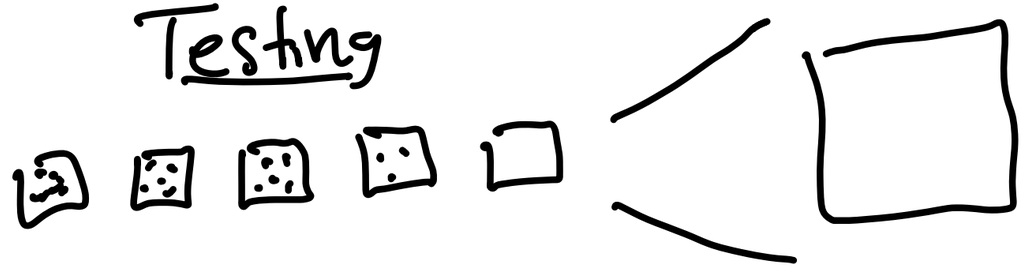
↳ rigid because dimensions ✓ are fixed

Training

$\epsilon_t = f_\theta(x_t, t)$



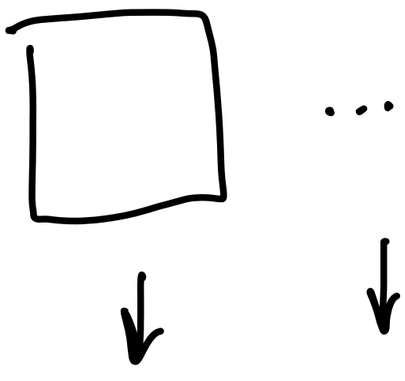
Testing



Key insight of stable diffusion:
model doesn't "see" like we do

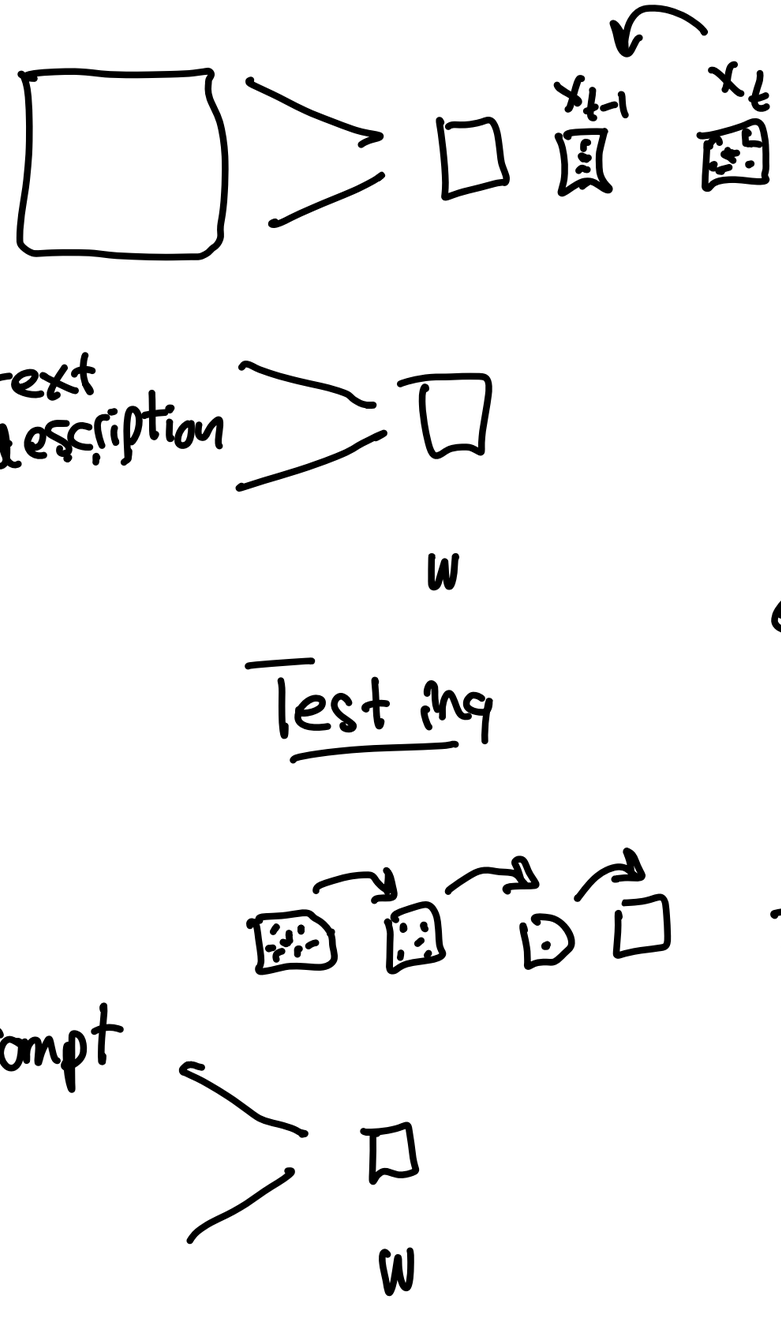
CLIP + Diffusion

Text prompt
↳ guide through
image generation



Training

$$\epsilon_t \approx f_\theta(x_t, t, w)$$



Testing

$$\epsilon_t \approx f_\theta(x_t, t, w)$$