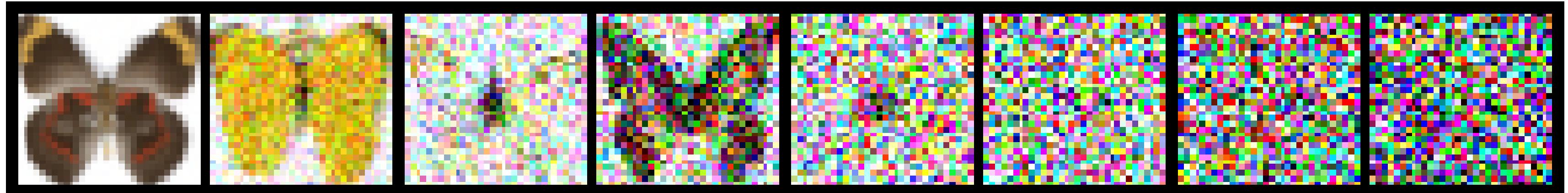


# Intro to Diffusion Models & Guidance/Conditioning

Kerem Aslan

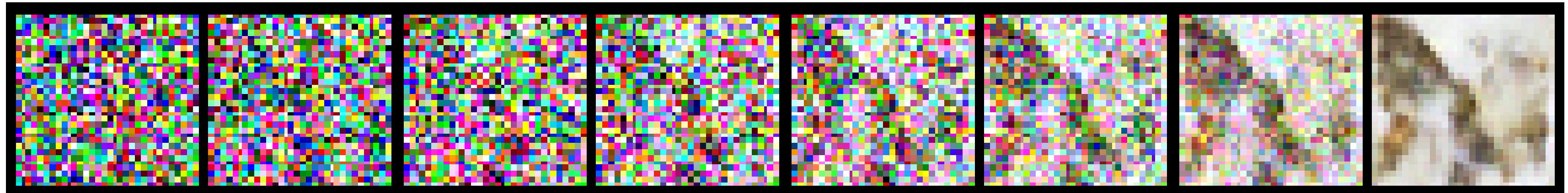
# Idea of Diffusion:

Diffusion



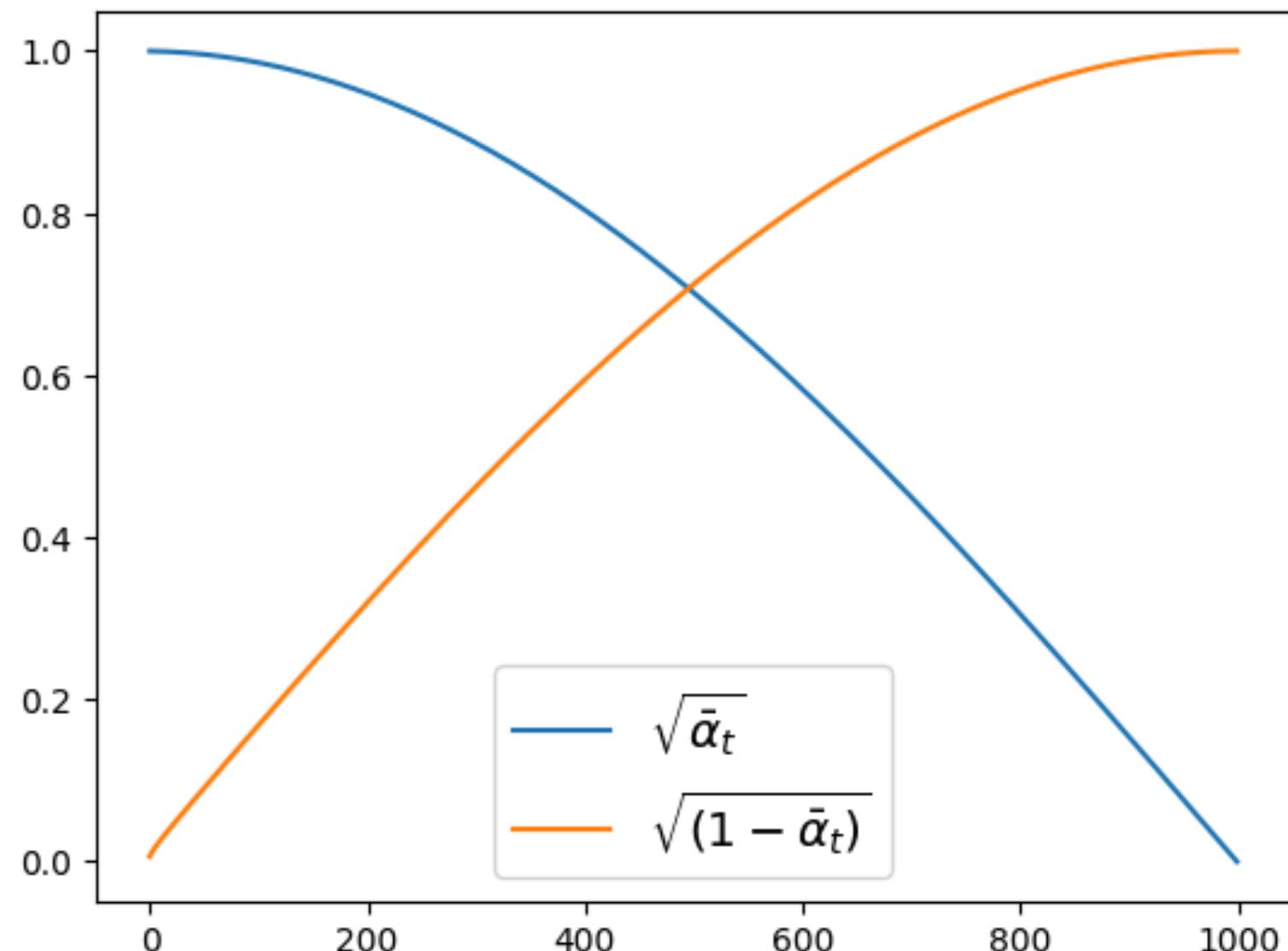
*Image from: huggan smithsonian\_butterflies\_subset*

*Reverse Diffusion*

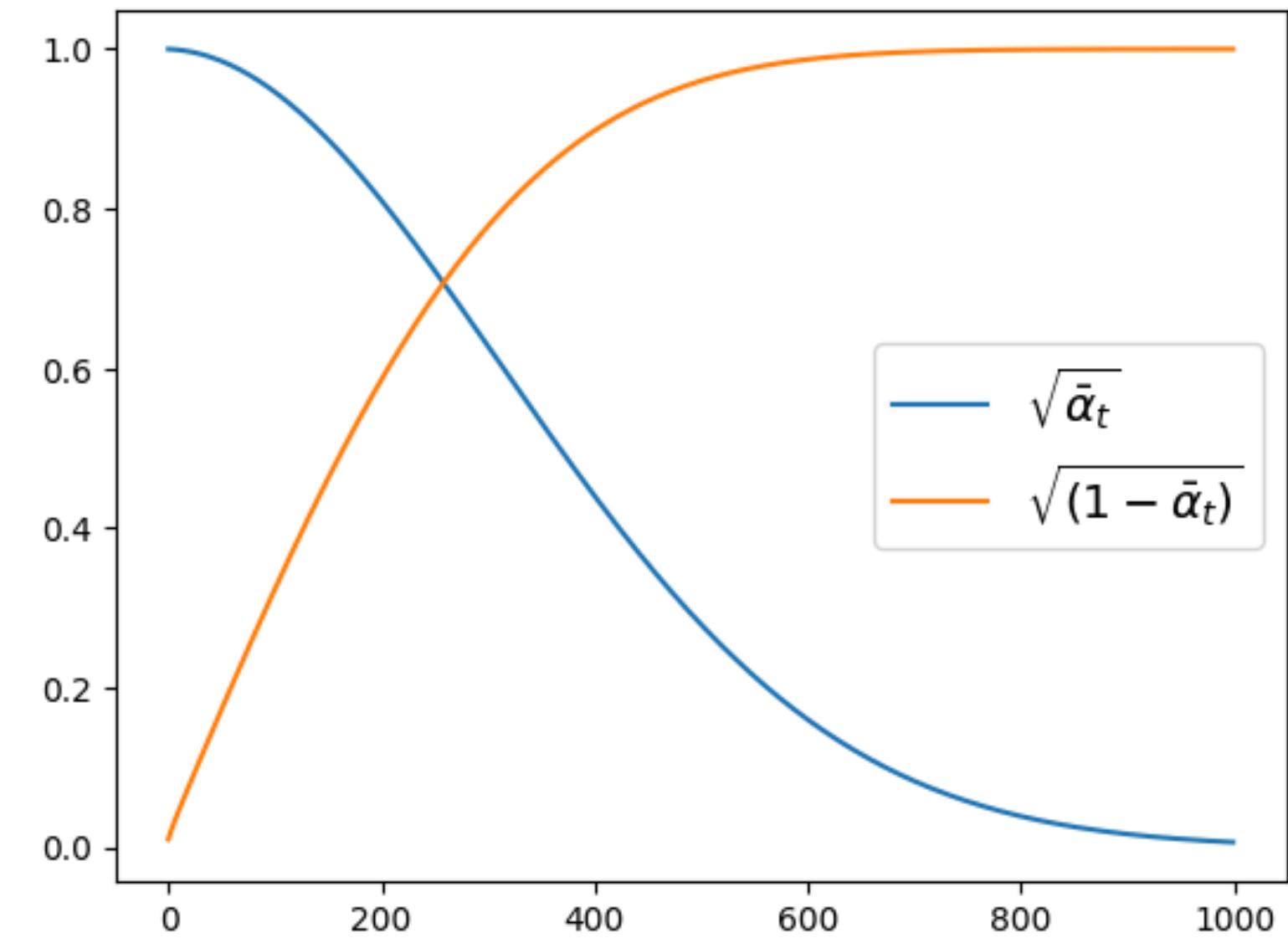


*Used Pipeline: keremmaslan/butterflies-sd-32*

# Diffusion Scheduler

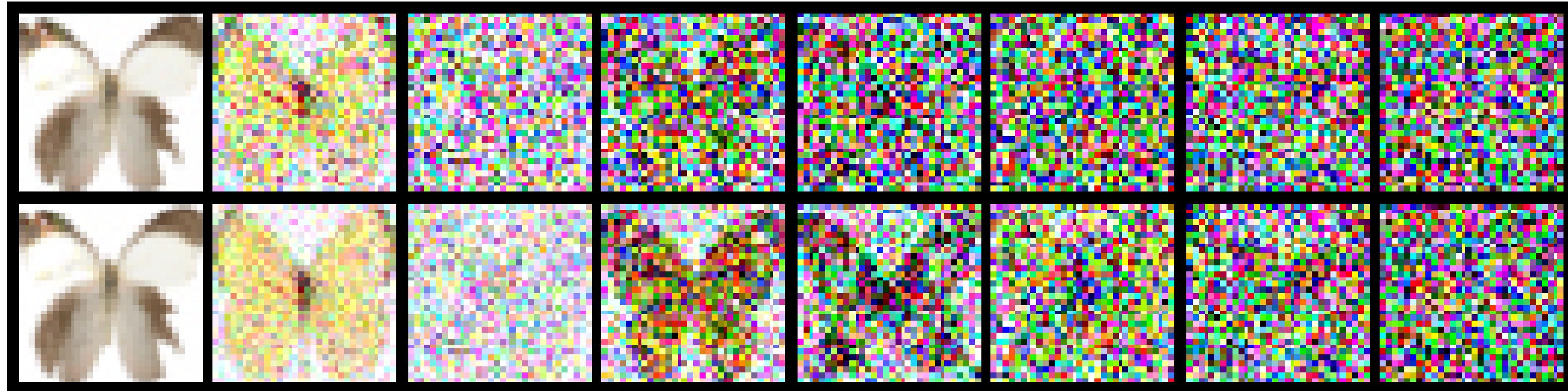


`beta_schedule="squaredcos_cap_v2"`



`beta_schedule="linear"`

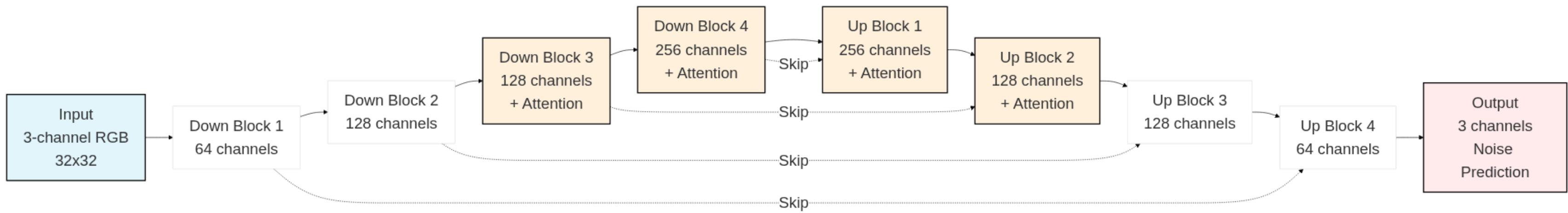
# Diffusion Scheduler



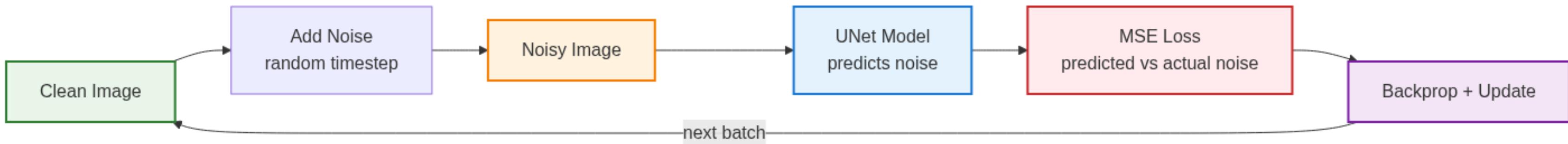
Top: `beta_schedule="linear"`

Bottom: `beta_schedule="squaredcos_cap_v2"`

# UNet Model Architecture



# Training Loop



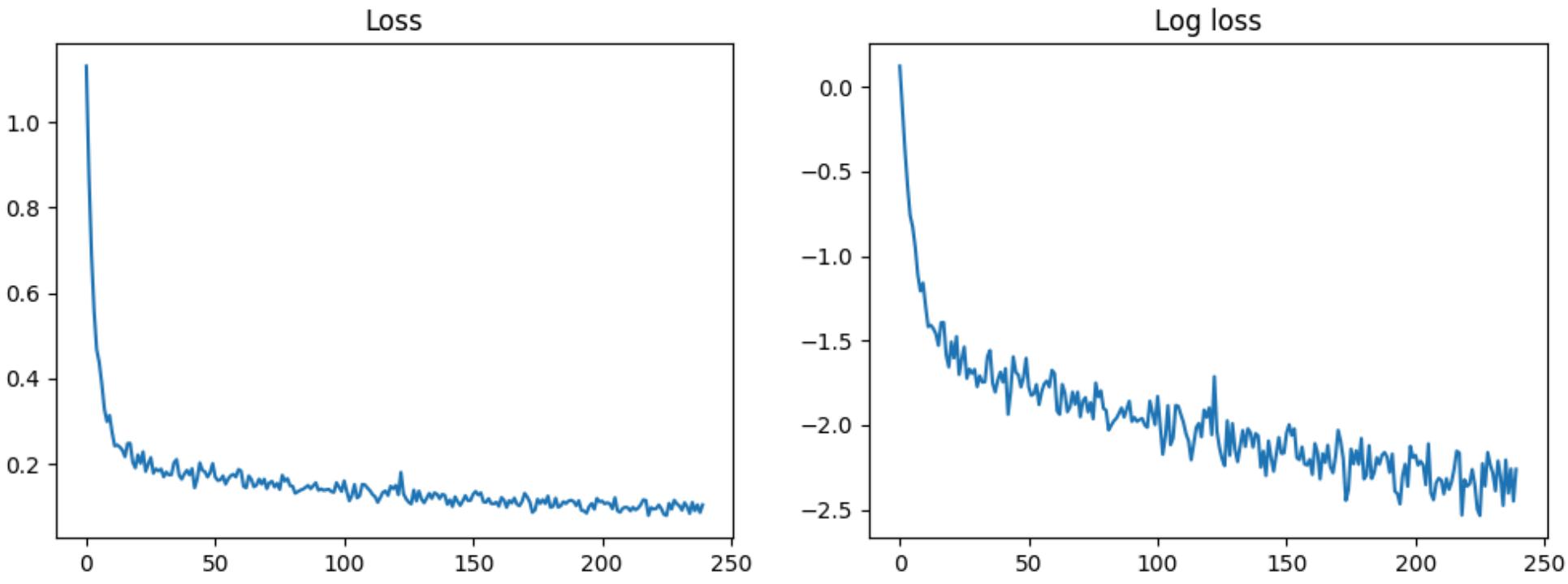
## Key Settings:

Scheduler: DDPM, 1000 timesteps, squaredcos\_cap\_v2

Optimizer: AdamW, lr=4e-4

Epochs: 30

```
Epoch:5, loss: 0.18275297060608864
Epoch:10, loss: 0.1560902539640665
Epoch:15, loss: 0.13007053080946207
Epoch:20, loss: 0.11560267396271229
Epoch:25, loss: 0.09990465361624956
Epoch:30, loss: 0.09857593476772308
```

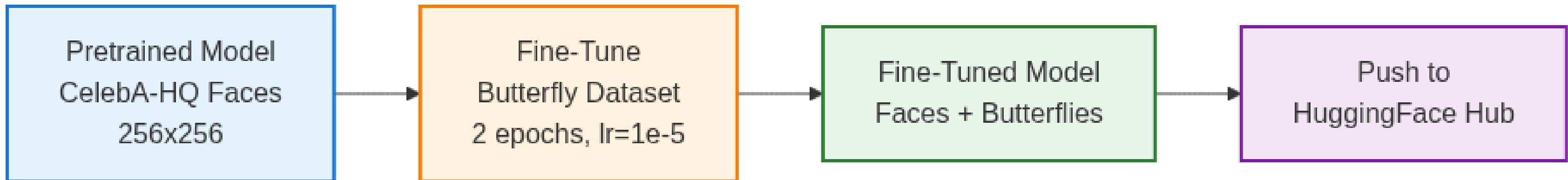


# Generated Images

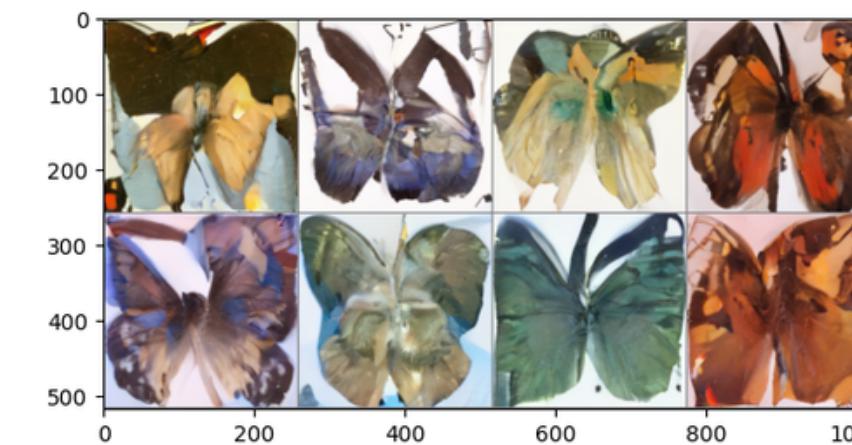
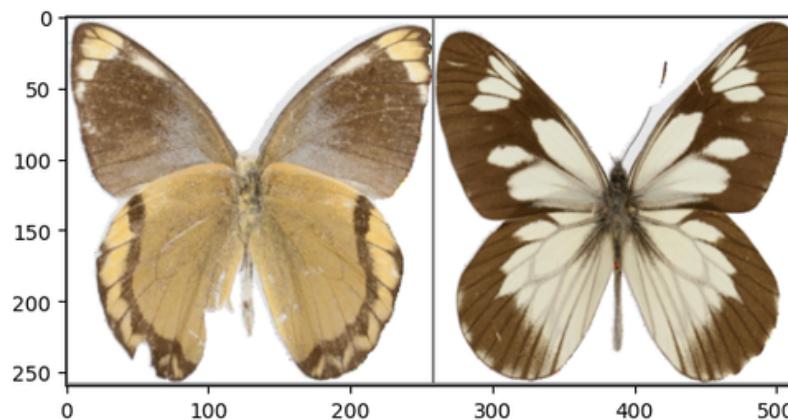


*Pipeline could be found @ keremaslan/butterflies-sd-32*

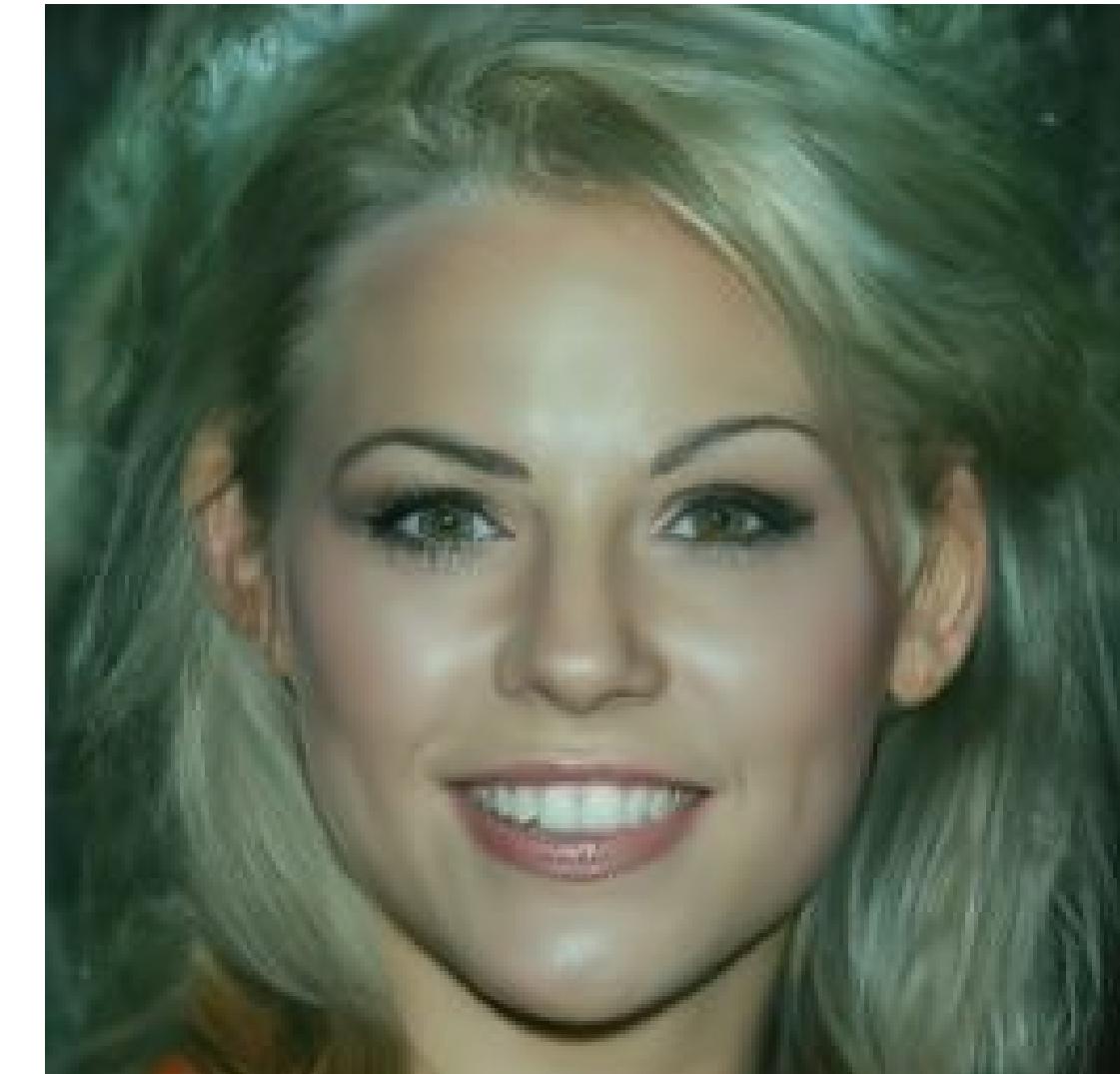
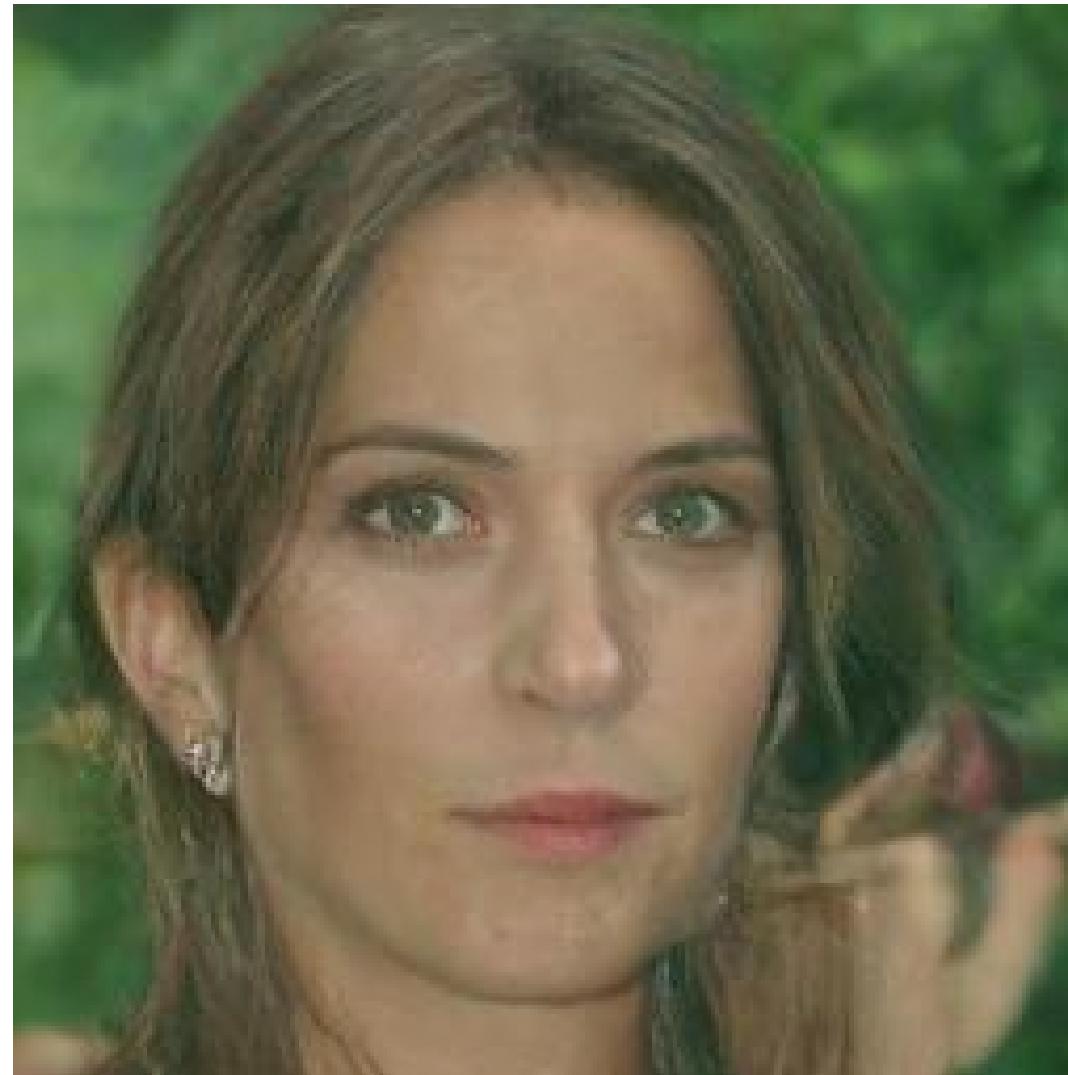
# Fine Tuning



Used Dataset:  
*huggan smithsonian\_butterflies\_subset*



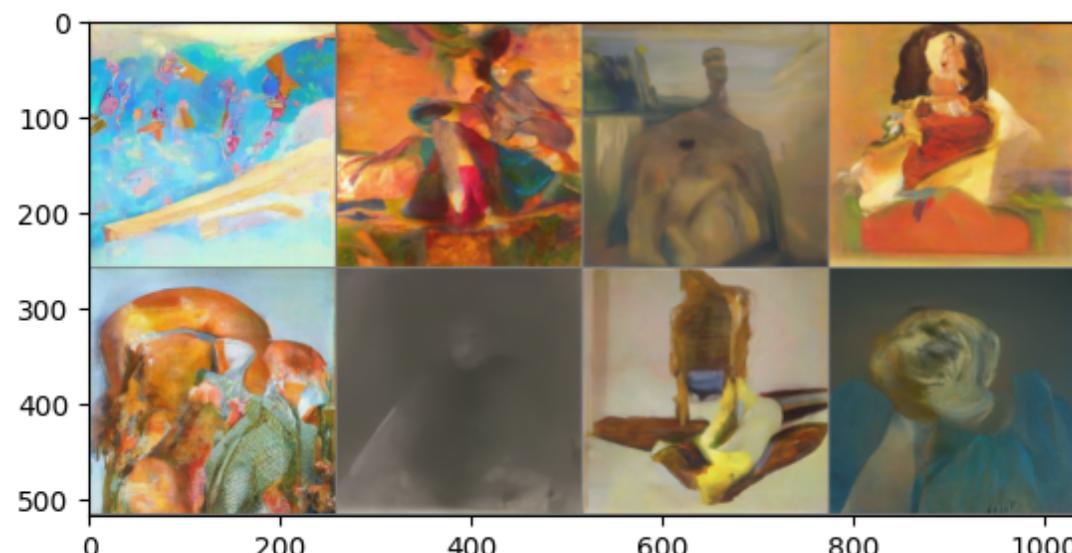
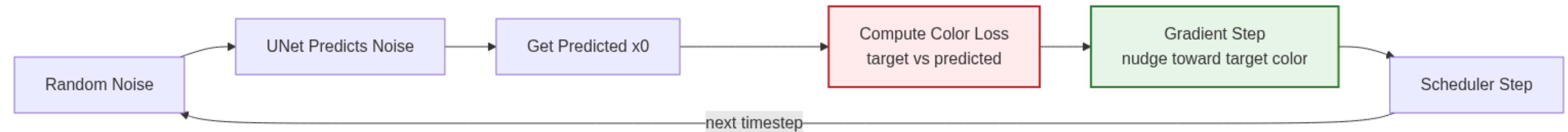
# Speed: DDPM vs DDIM Scheduler



DDPM 1000 steps -> 1m 21s    DDIM 40 steps → 3.9s

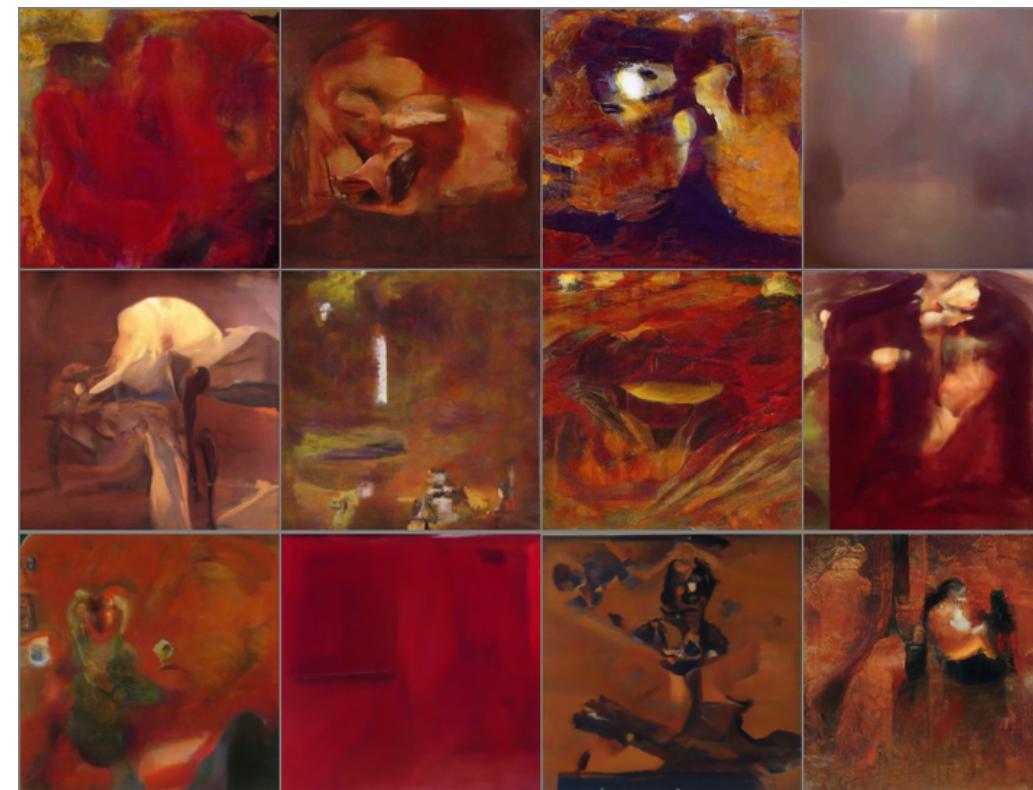
DDIM 100 steps → 8.3s

# Guidance

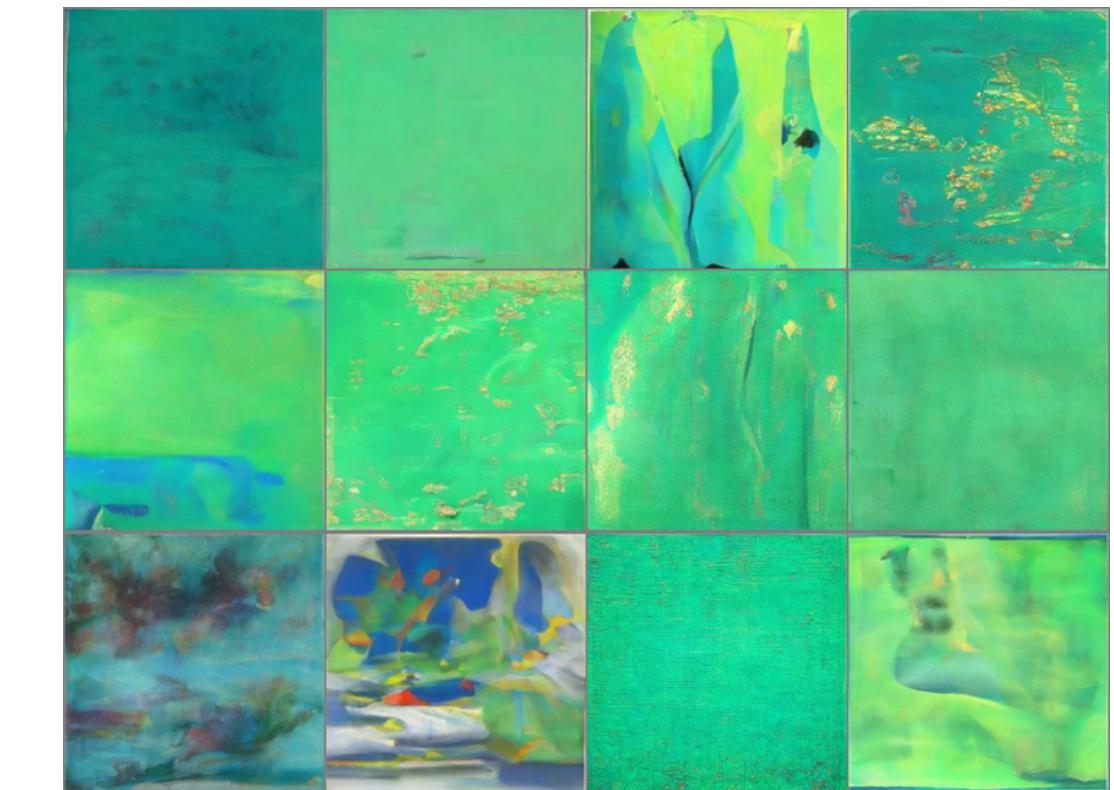


Images generated without  
guidance

Used pipeline: [johnowhitaker/sd-class-wikiart-from-bedrooms](#)



Images generated with guidance  
(color)



# Variant 1 vs. Variant 2

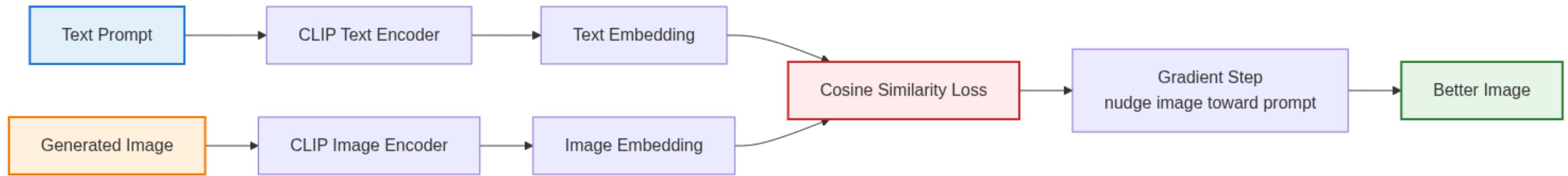
- **Variant 1:** Gradient only through scheduler, less VRAM used
- **Variant 2:** Gradient through UNet + scheduler, more accurate, more VRAM used



Images generated using Variant 1

Images generated using Variant 2

# CLIP Guidance

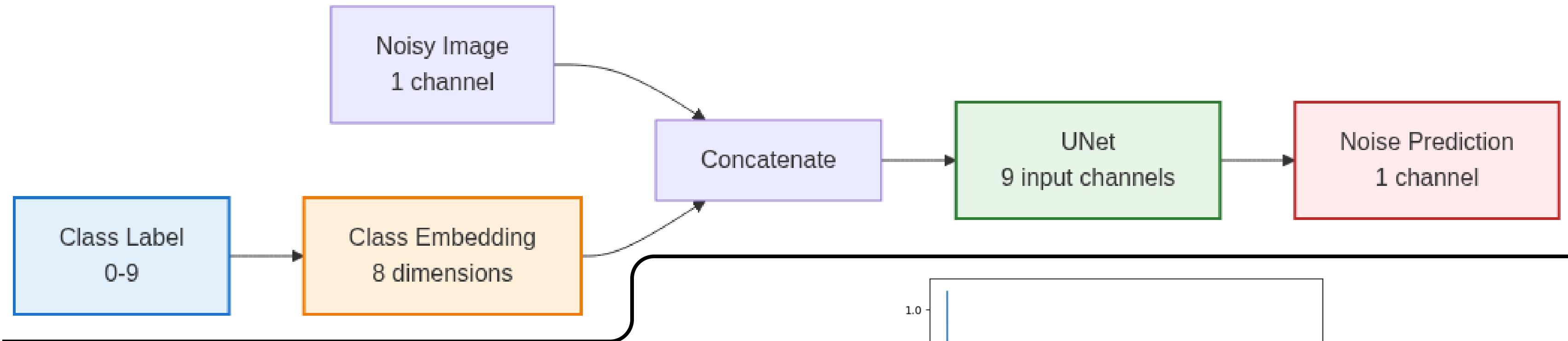


*Used model: ViT-B-32*

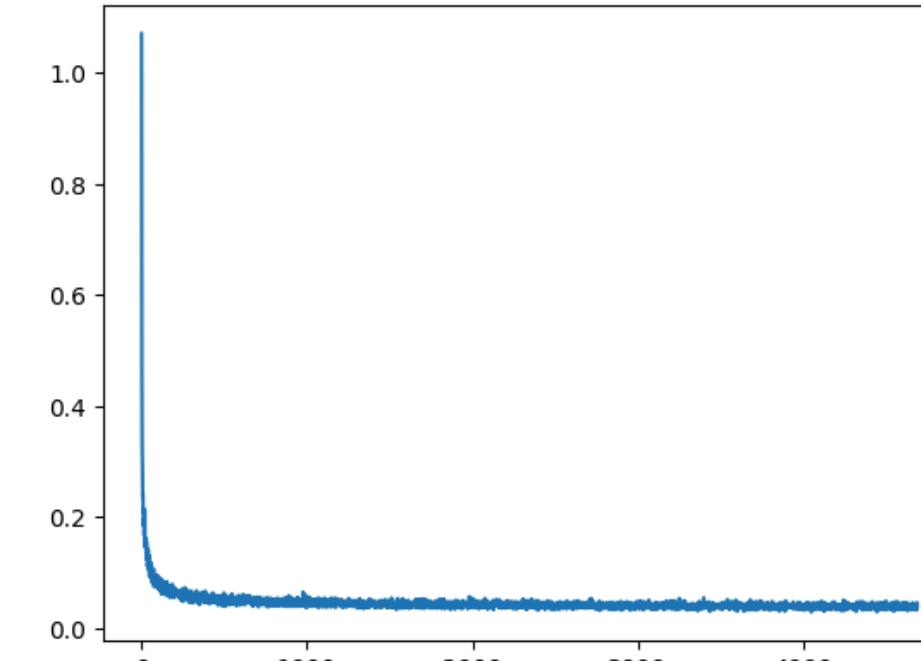


*prompt = "Blue orchid, blue flower painting"*

# Class-Conditioned Diffusion Model



*Used Dataset: MNIST 28x28*



*Loss curve*

# Class-Conditioned Diffusion Model



*Same noise, different class labels →  
different digit outputs*

Thank you for listening.