

FreeU: Free Lunch in Diffusion U-Net



Chenyang Si



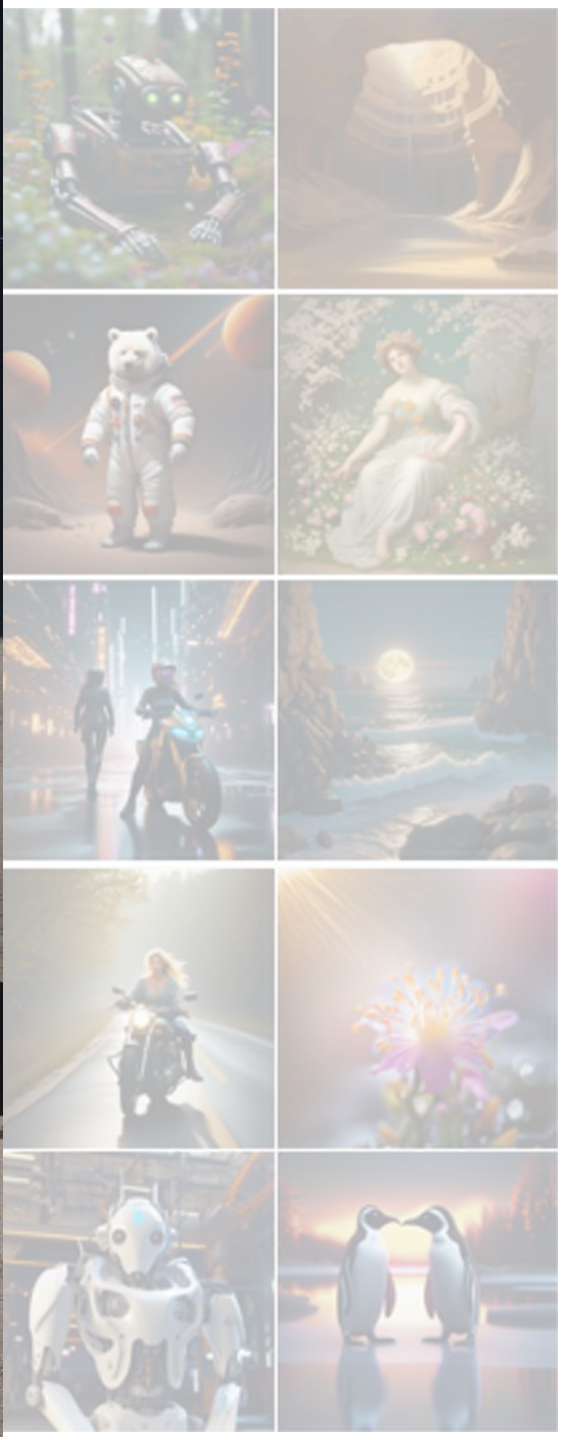
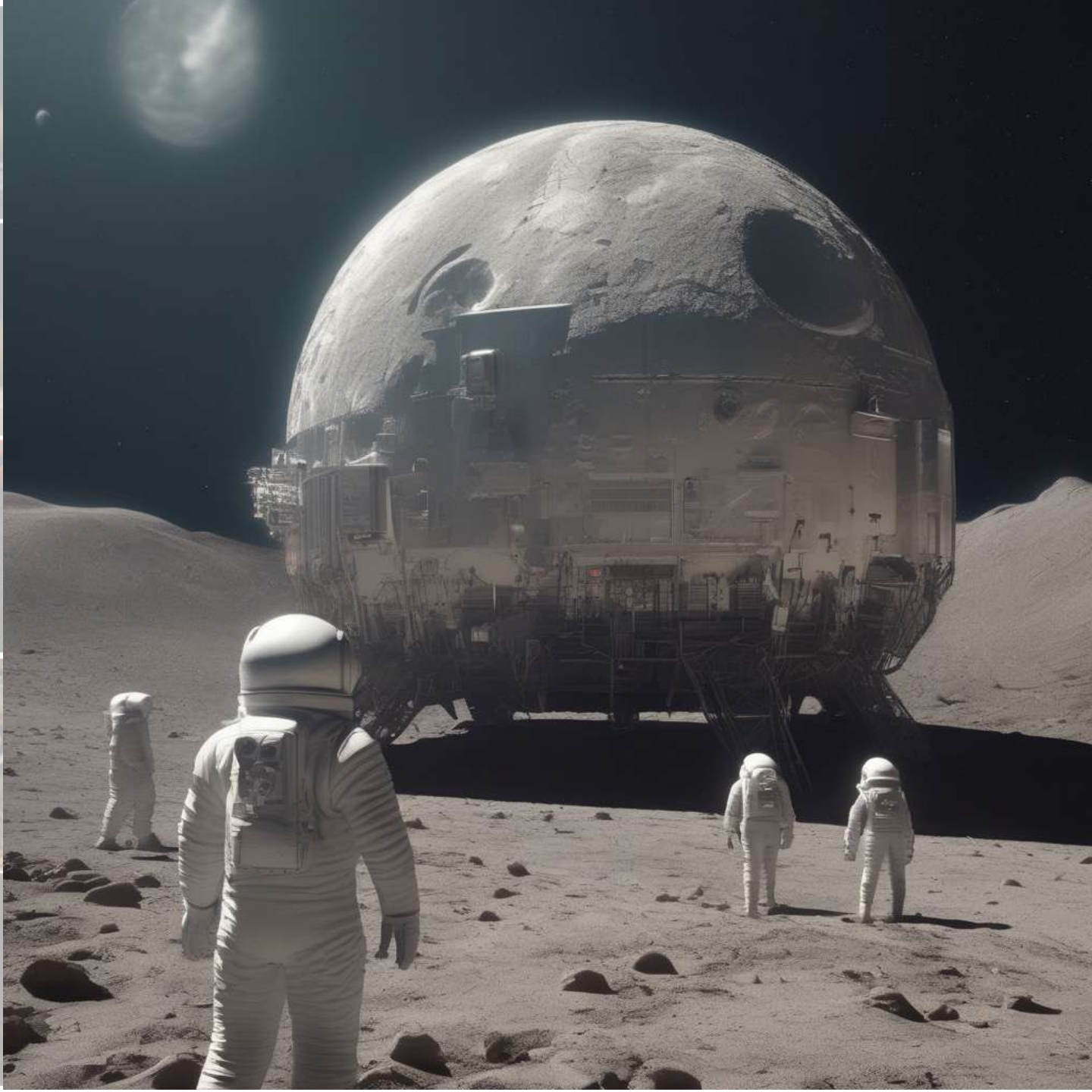
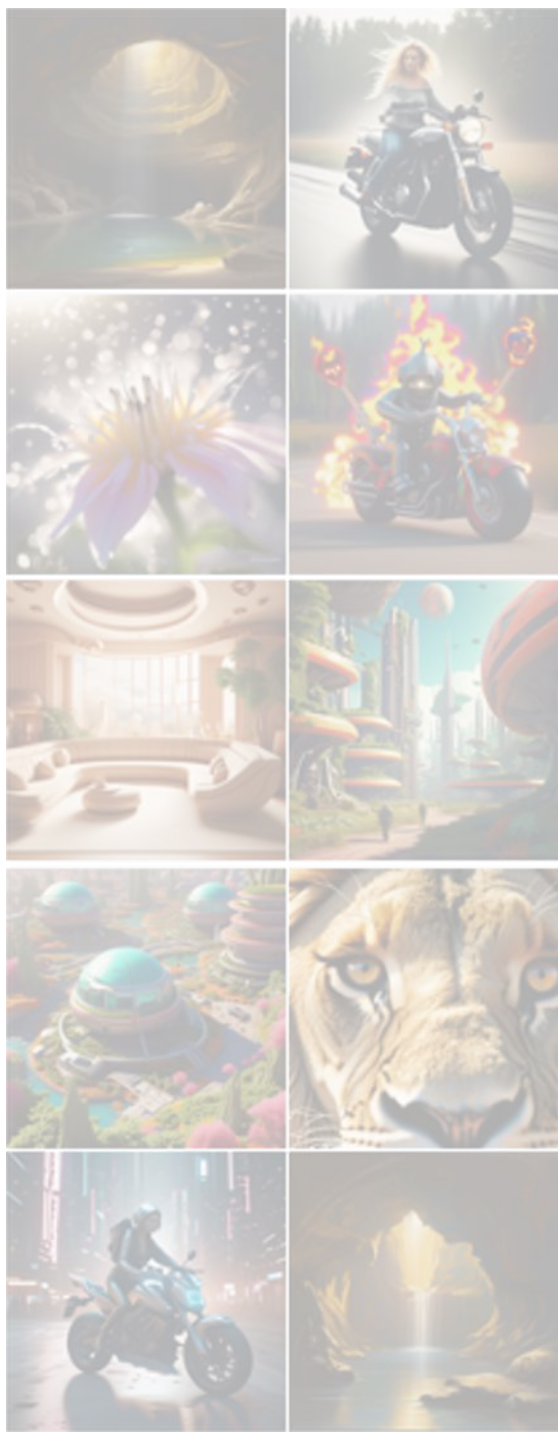
Ziqi Huang

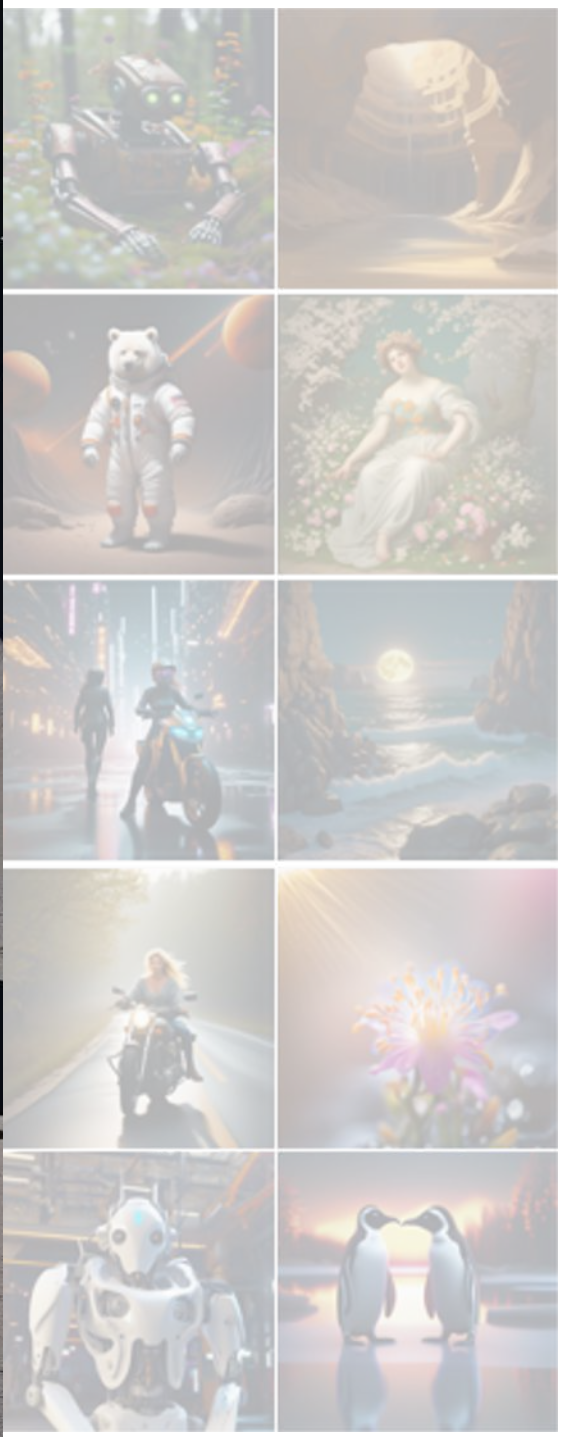
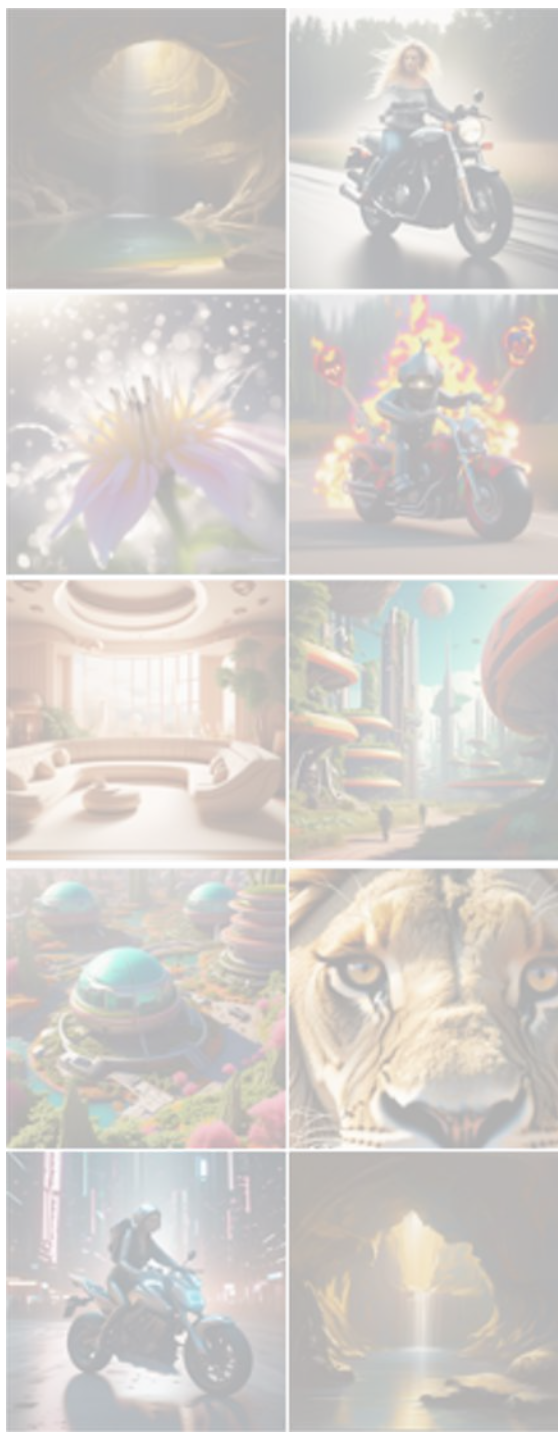


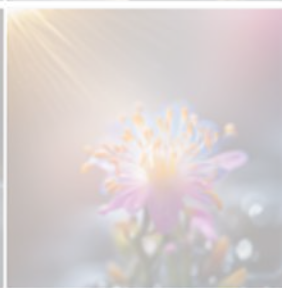
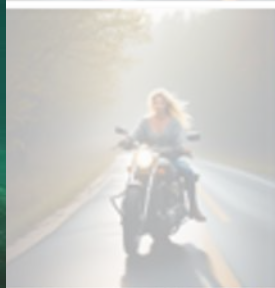
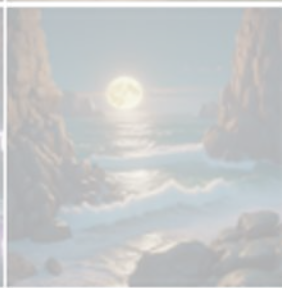
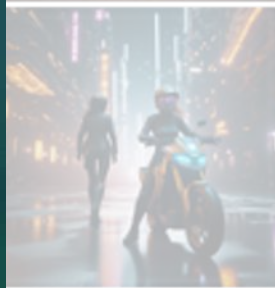
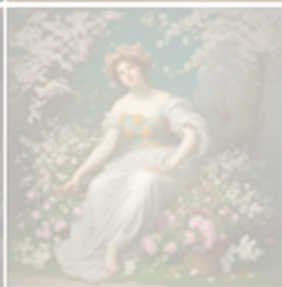
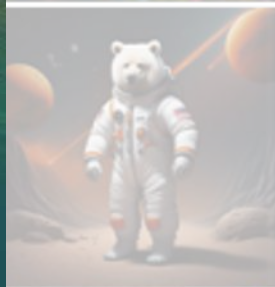
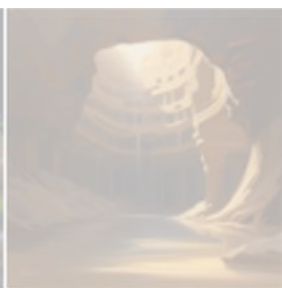
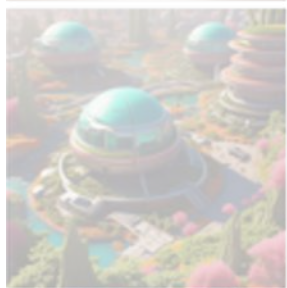
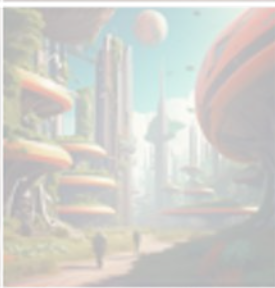
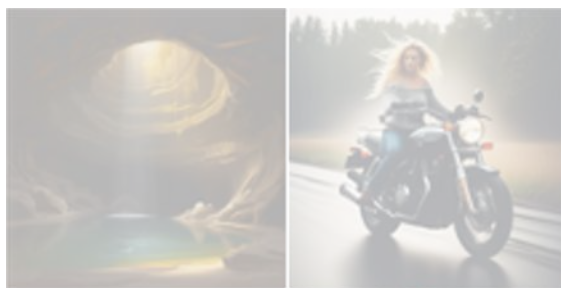
Yuming Jiang

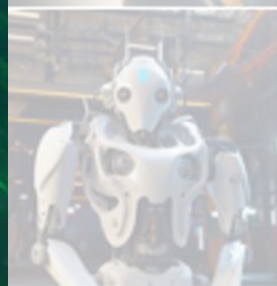
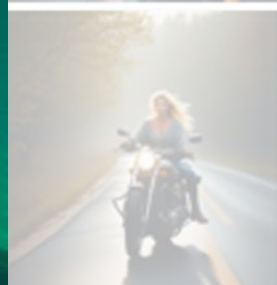
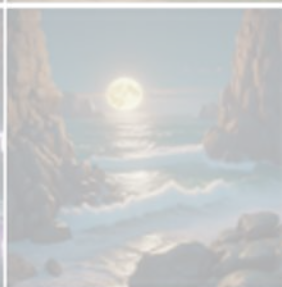
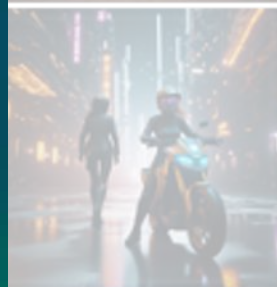
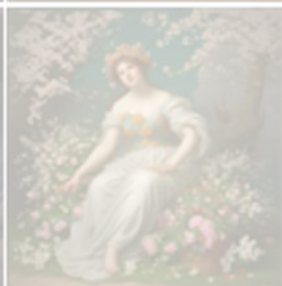
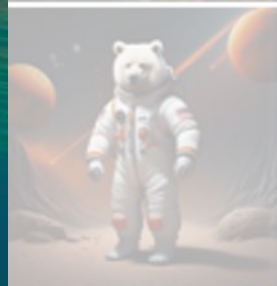
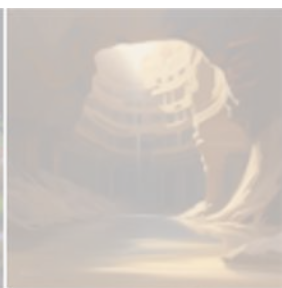
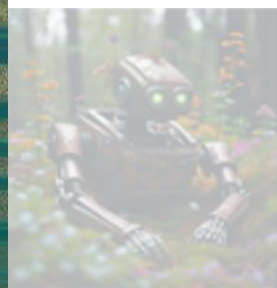
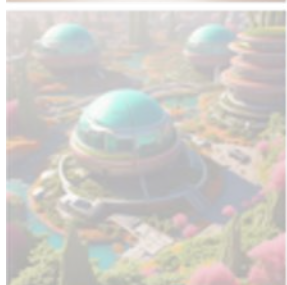
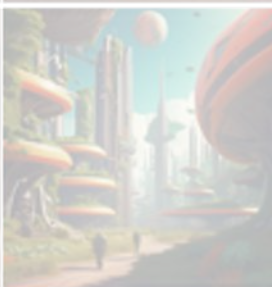
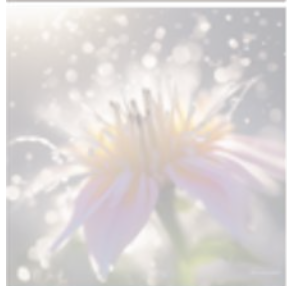
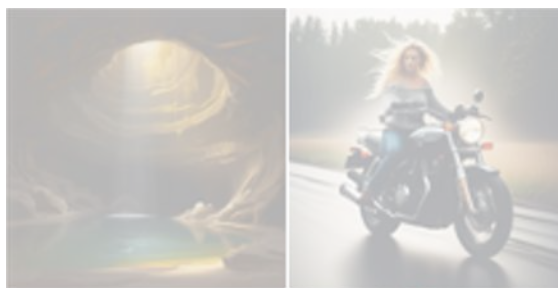


Ziwei Liu









Pre-trained Diffusion Models

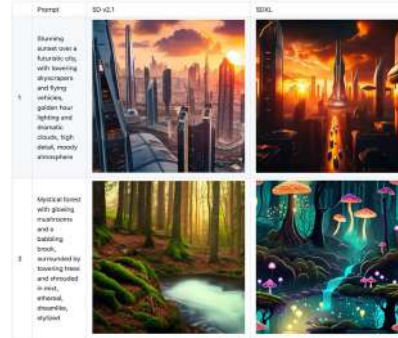
image generation



ADM



LDM



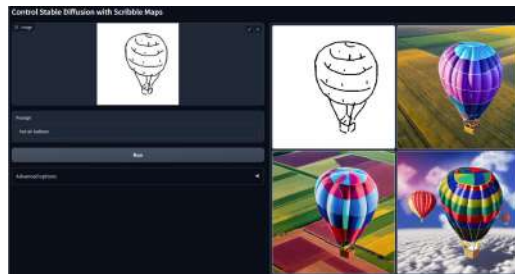
SDXL

video generation



VideoCrafter

controllable generation, customization, editing



ControlNet



Input images

in the Acropolis

DreamBooth



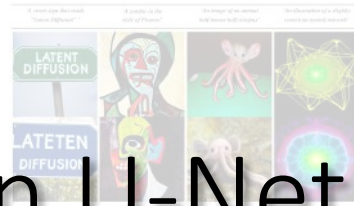
InstructPix2Pix

Pre-trained Diffusion U-Nets

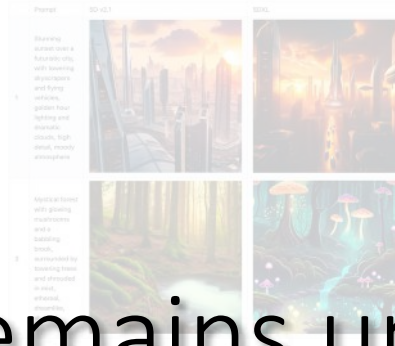
image generation



ADM



LDM



SDXL

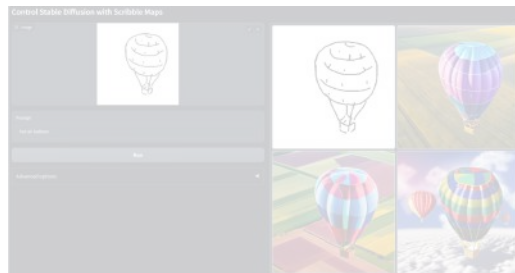
Diffusion U-Net remains under-explored

video generation



VideoCrafter

controllable generation, editing, customization



ControlNet



Input images

in the Acropolis

DreamBooth



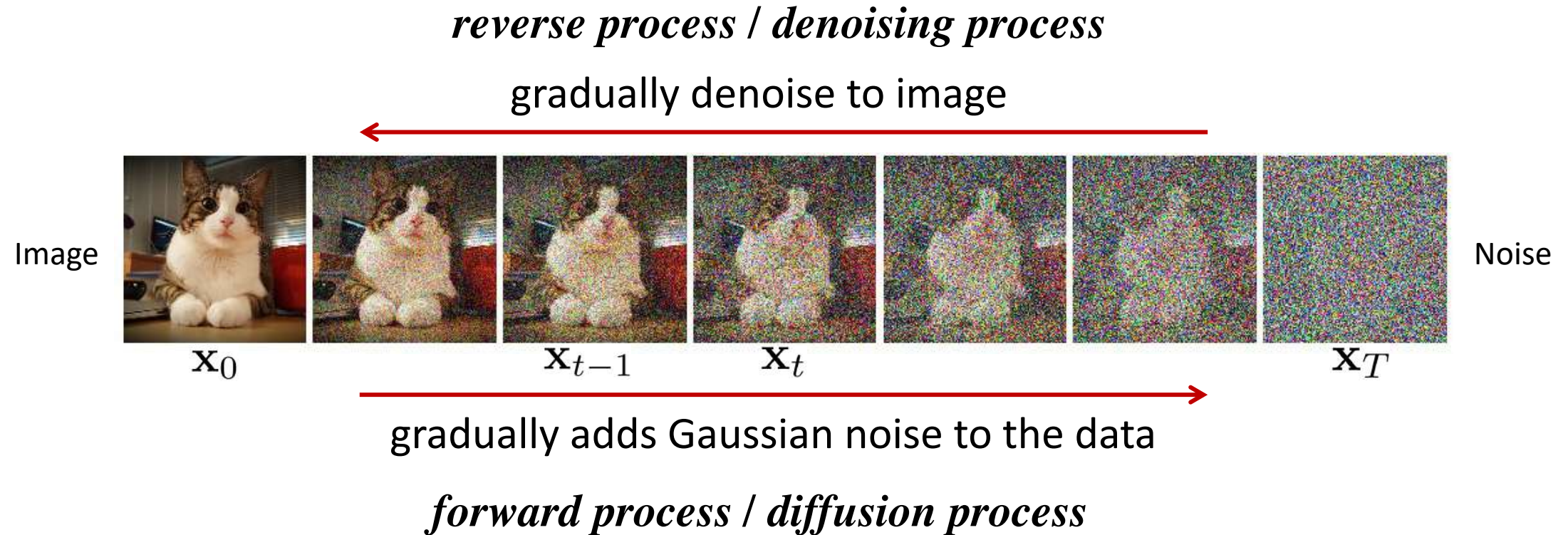
InstructPix2Pix

Motivation

- Downstream applications
 - directly utilizing pre-trained diffusion U-Nets
 - internal properties of diffusion U-Net features remain under-explored
- Train better foundation models
 - expensive (*e.g.*, SDXL)
 - besides scaling up (*e.g.*, data scale, model size), what else can we do?
- Why not exploit pre-trained diffusion models?
 - Let's take a closer look at *diffusion U-Net* and the *denoising process*



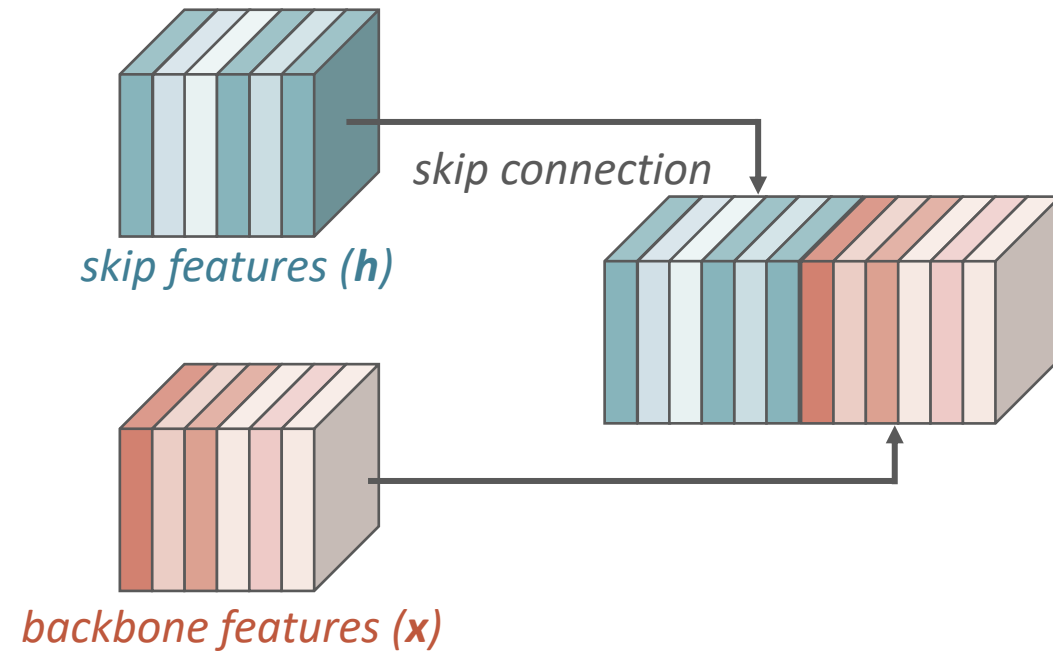
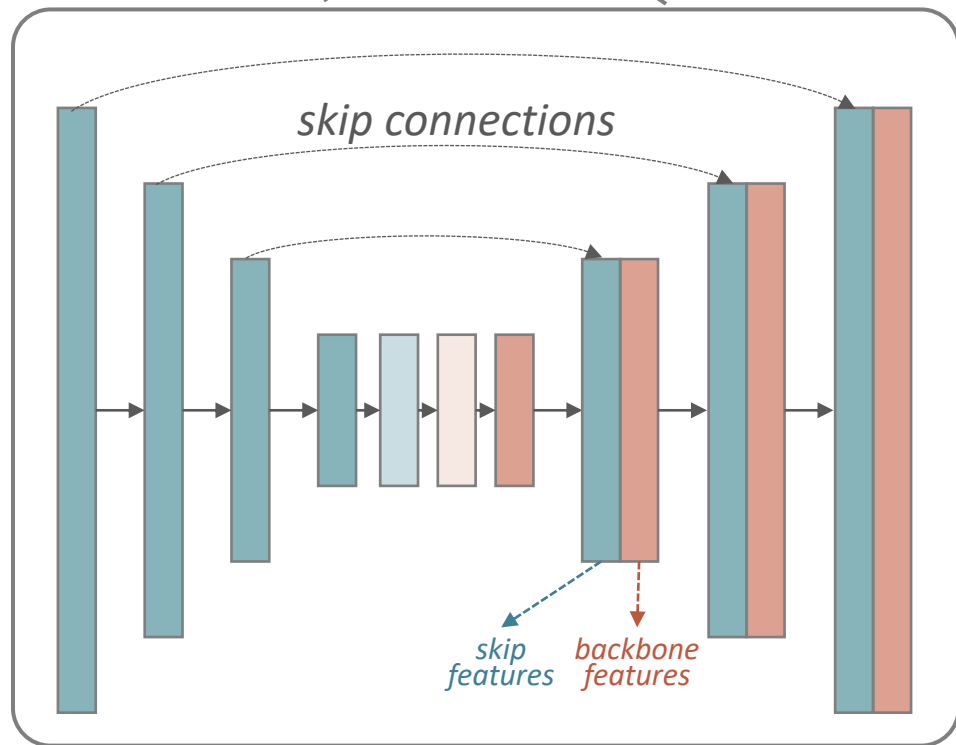
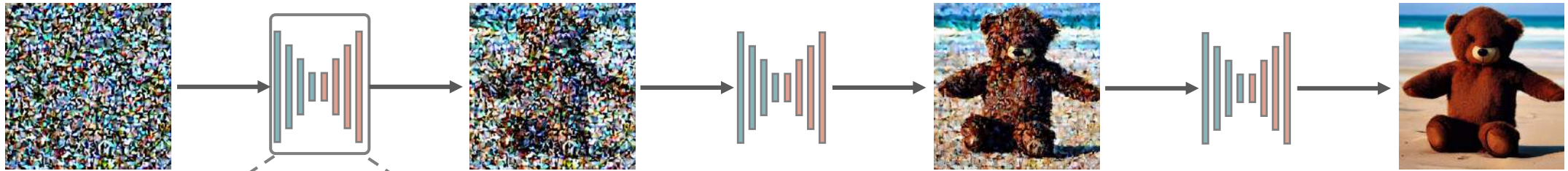
Recap: Diffusion Models



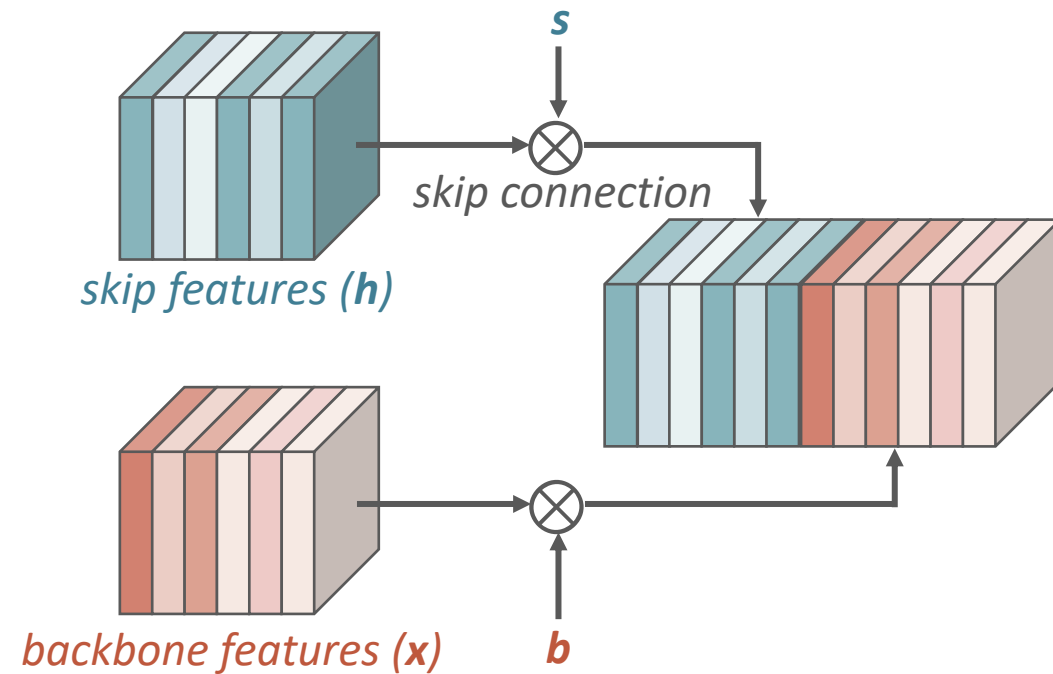
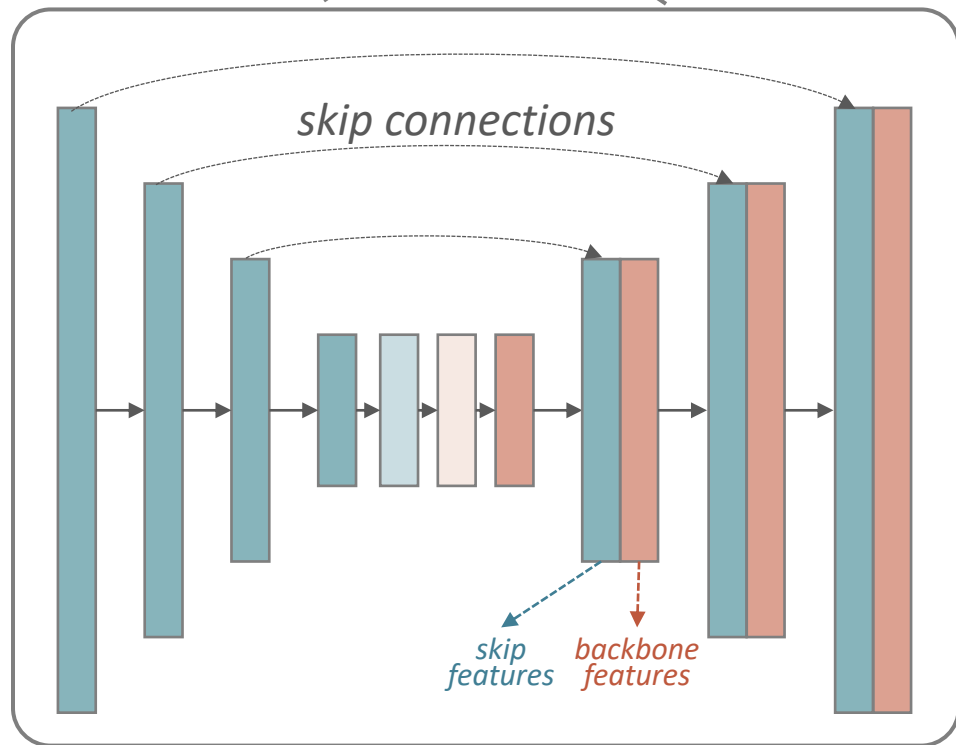
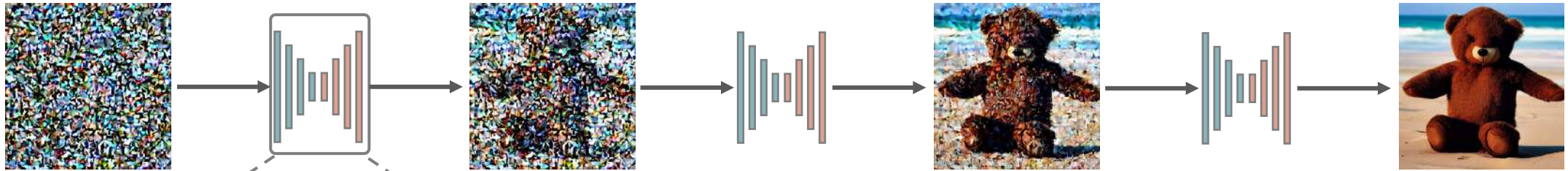
How does diffusion U-Net perform denoising?



Denoising Process: U-Net

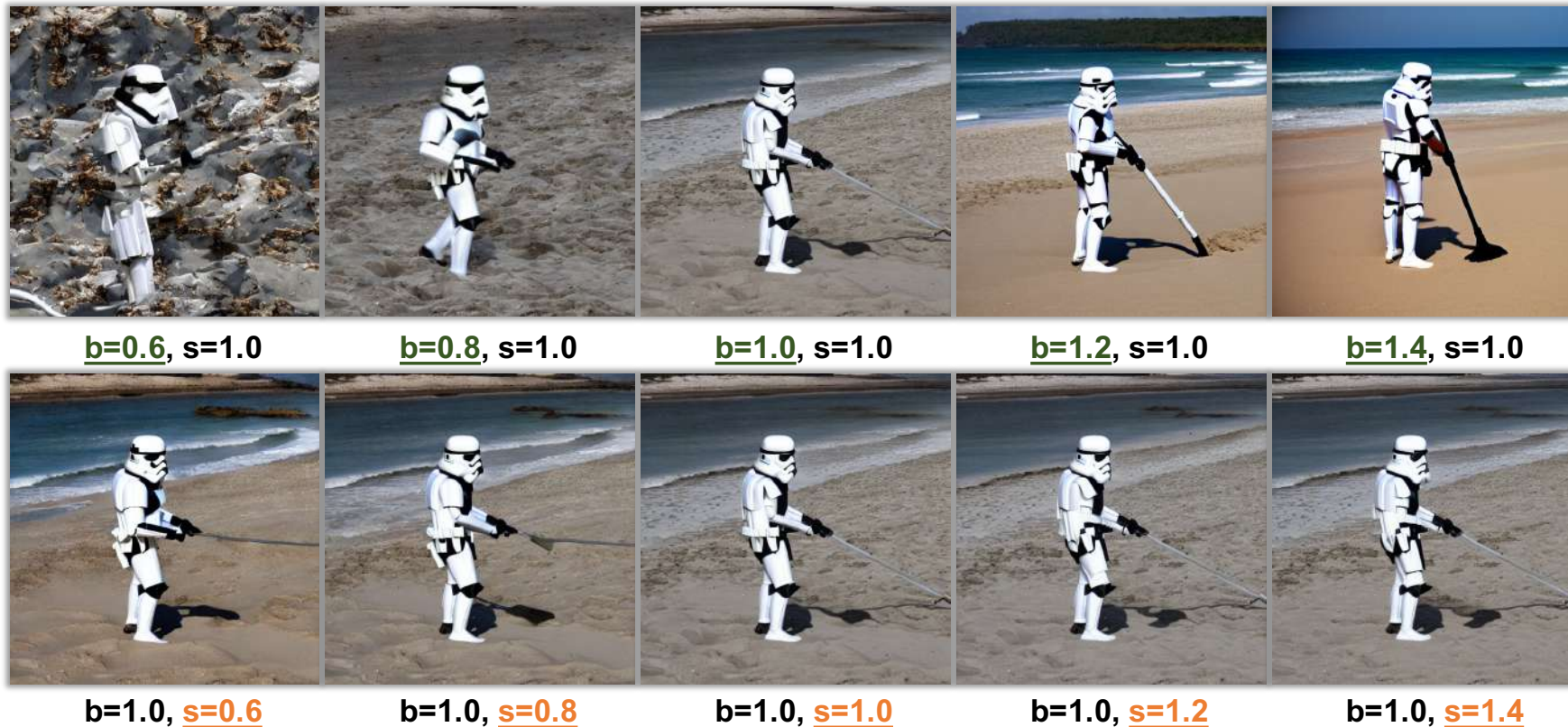


Denoising Process: U-Net



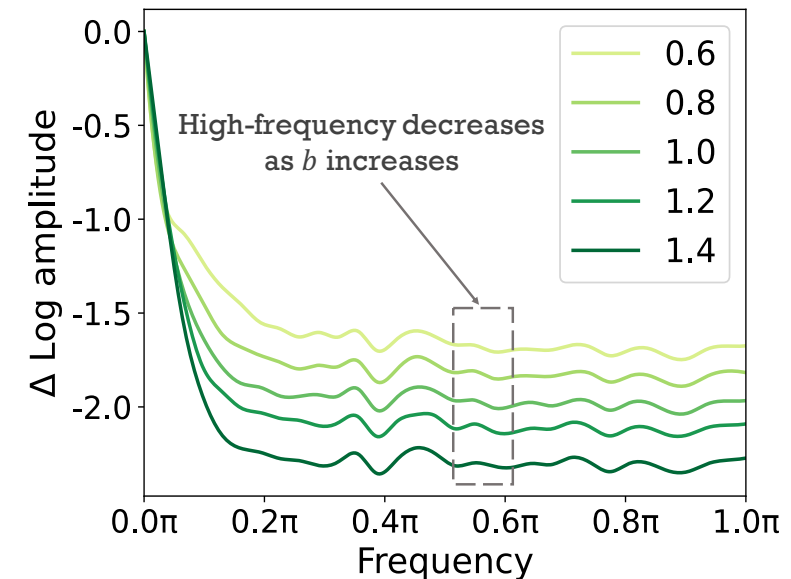
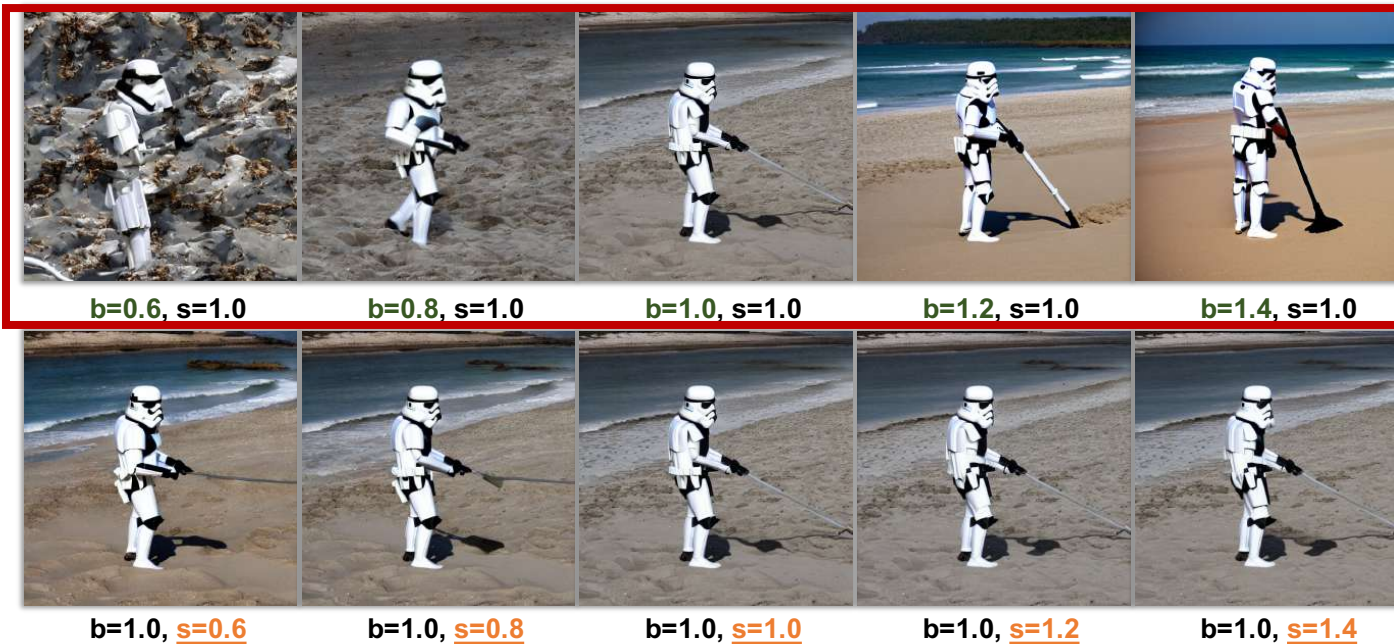
Role of Backbone and Skip Features

- Backbone: denoising
- Skip: limited impact during inference



How Diffusion U-Net Perform Denoising?

- **Backbone features**: primarily contributes to denoising
 - Consistent with visualization on the next page

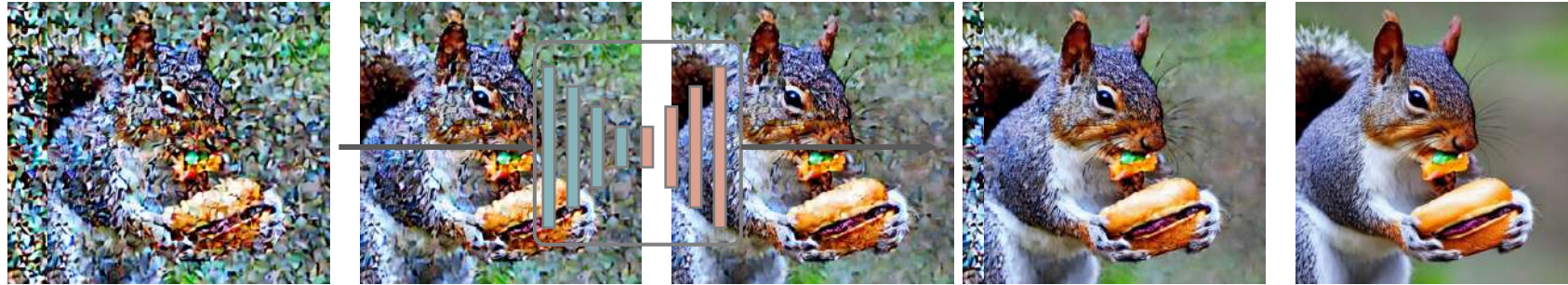


Fourier relative log amplitudes of variations of b



Denoising Process

Input: A squirrel eating a burger



Denoising

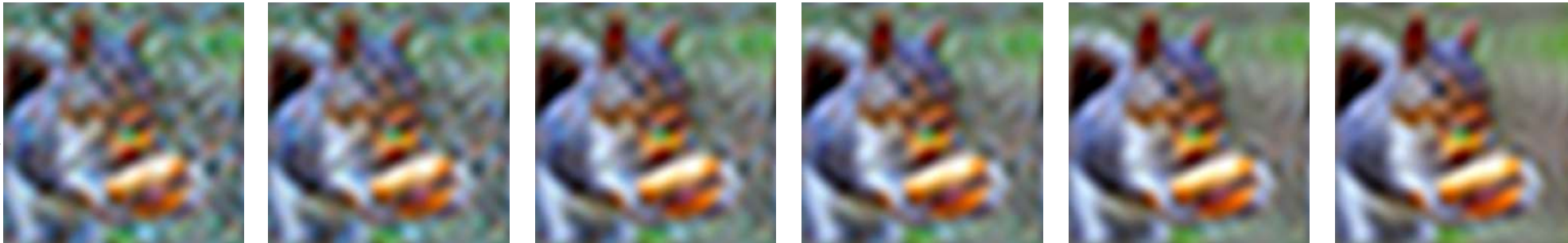


Denoising Process

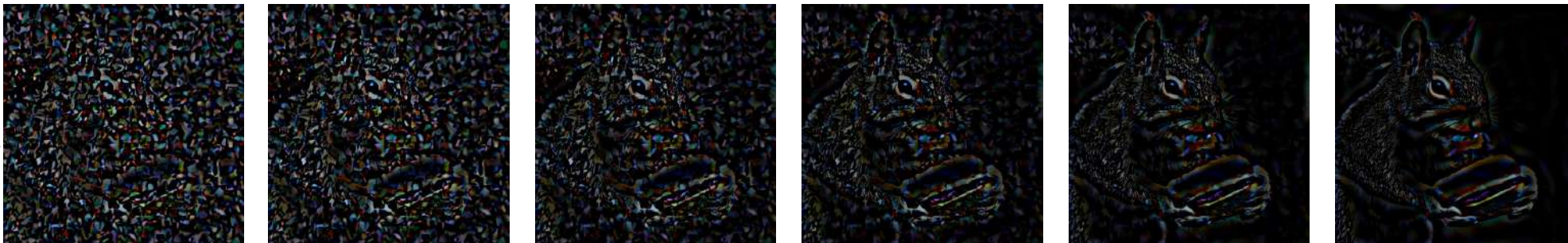
Input: A squirrel eating a burger



*Low
frequency*

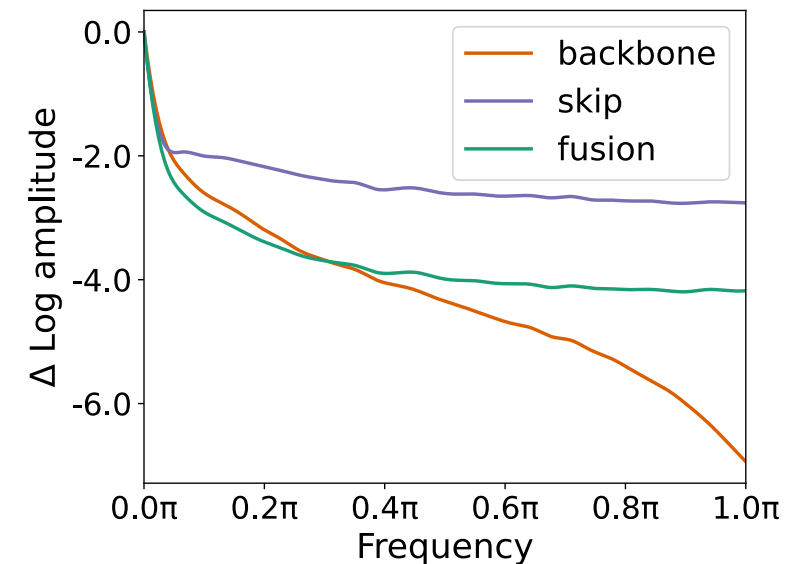
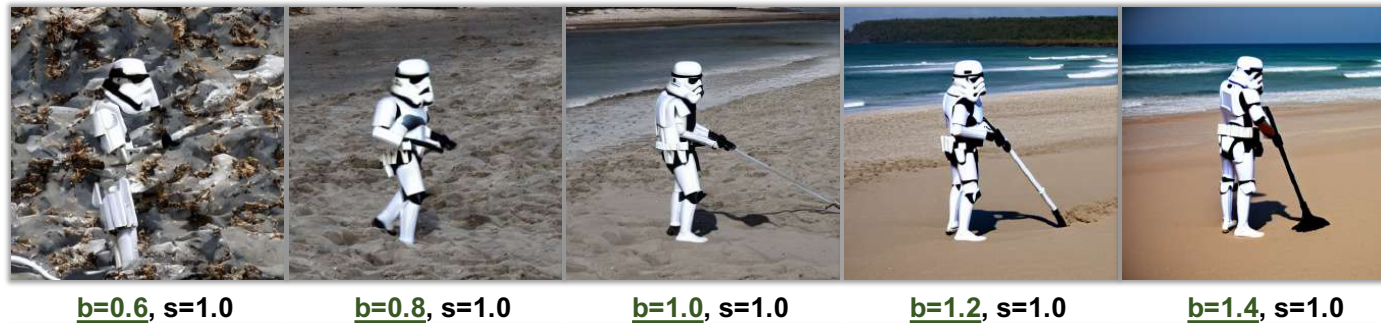


*High
frequency*



How Diffusion U-Net Perform Denoising?

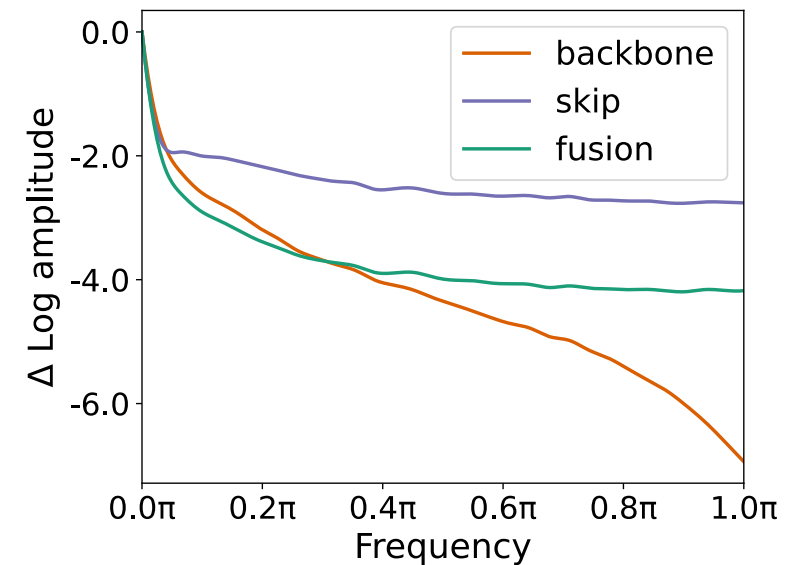
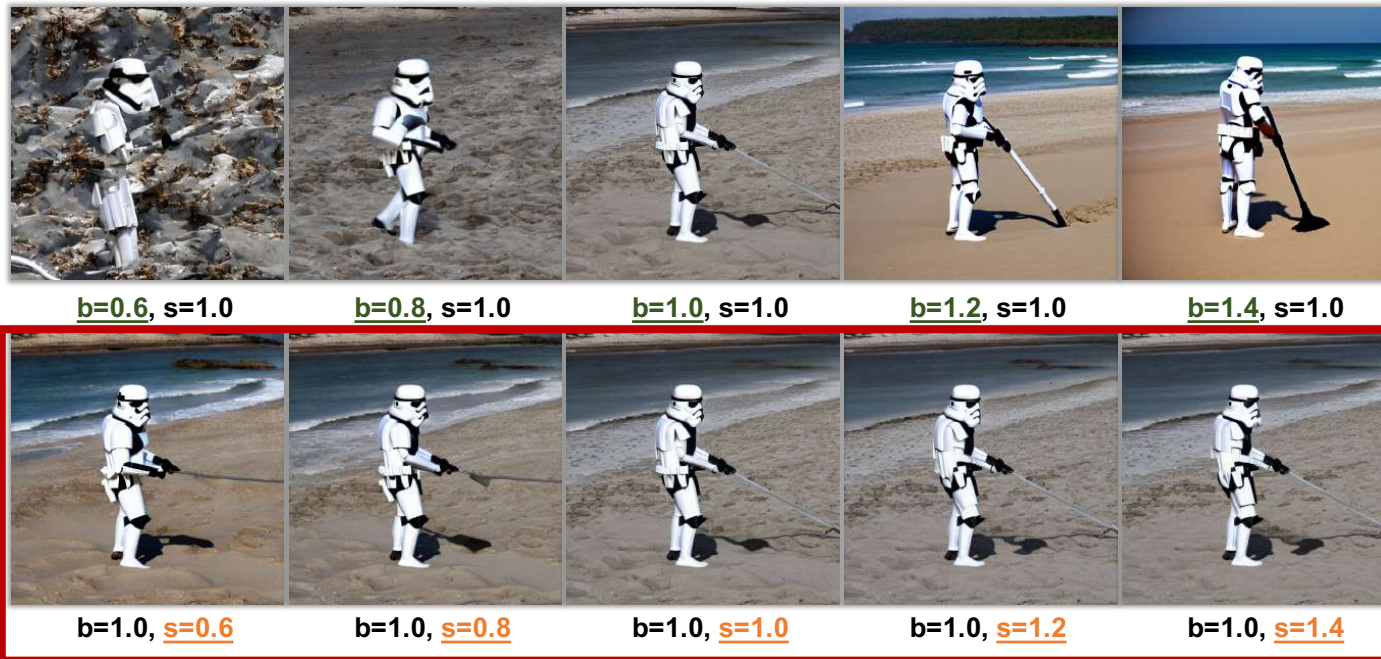
- **Backbone**: primarily contributes to denoising
- **Skip**: introduce high-frequency features into the decoder module



Fourier relative log amplitudes of backbone, skip, and their fused feature maps

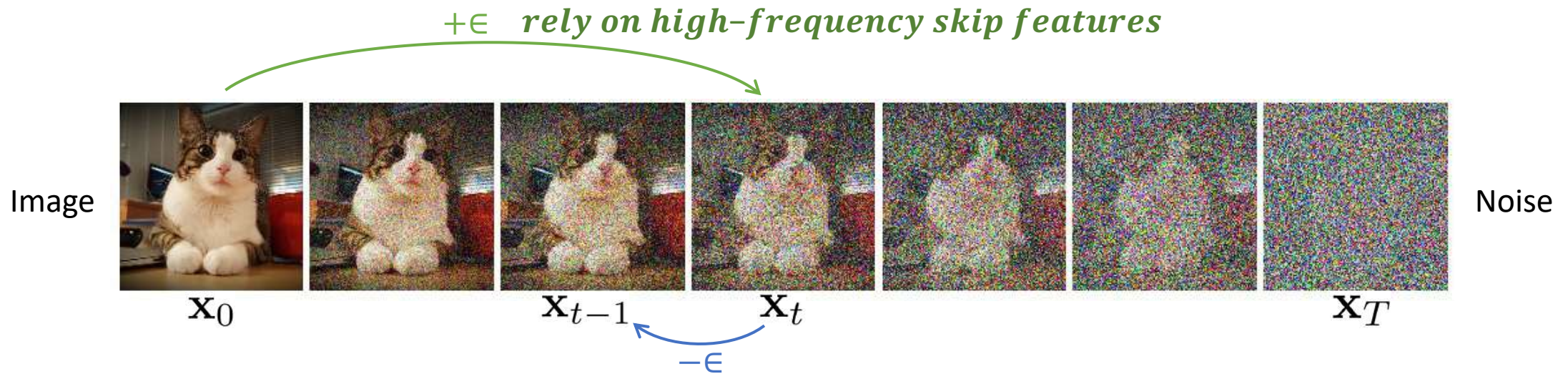
How Diffusion U-Net Perform Denoising?

- Gap between training and sampling



Fourier relative log amplitudes of backbone, skip, and their fused feature maps

Training & Sampling



Algorithm 1 Training

- 1: **repeat**
- 2: $\mathbf{x}_0 \sim q(\mathbf{x}_0)$
- 3: $t \sim \text{Uniform}(\{1, \dots, T\})$
- 4: $\epsilon \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$
- 5: Take gradient descent step on

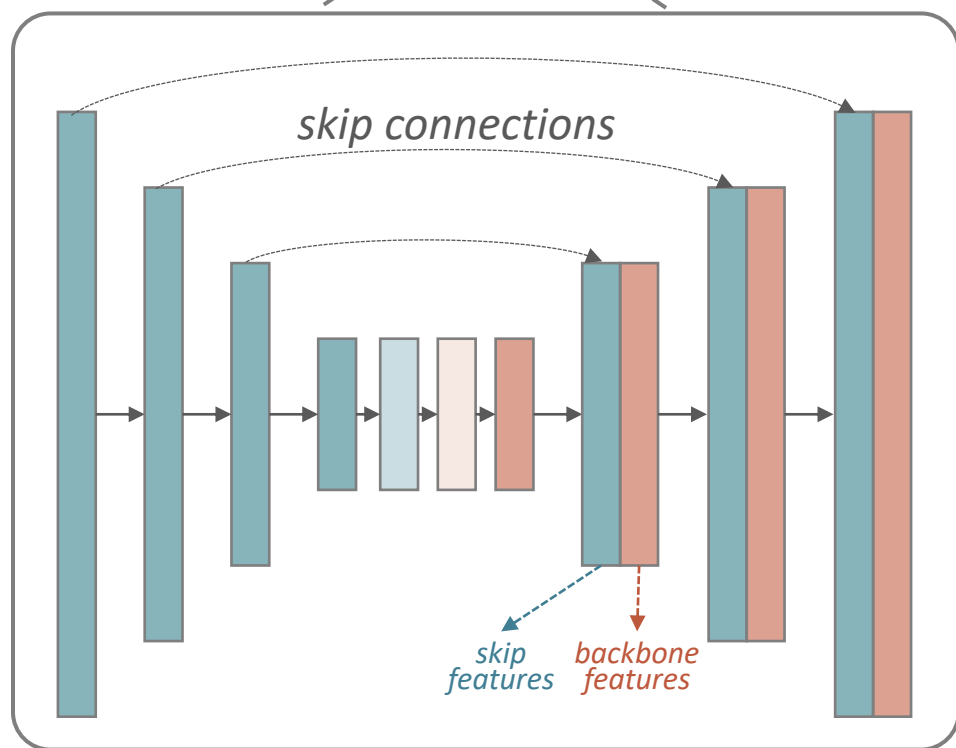
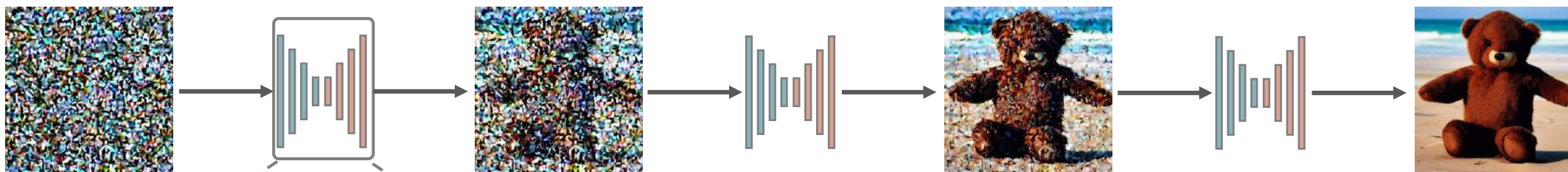
$$\nabla_{\theta} \left\| \epsilon - \epsilon_{\theta} \left(\sqrt{\bar{\alpha}_t} \mathbf{x}_0 + \sqrt{1 - \bar{\alpha}_t} \epsilon, t \right) \right\|^2$$
- 6: **until** converged

Algorithm 2 Sampling

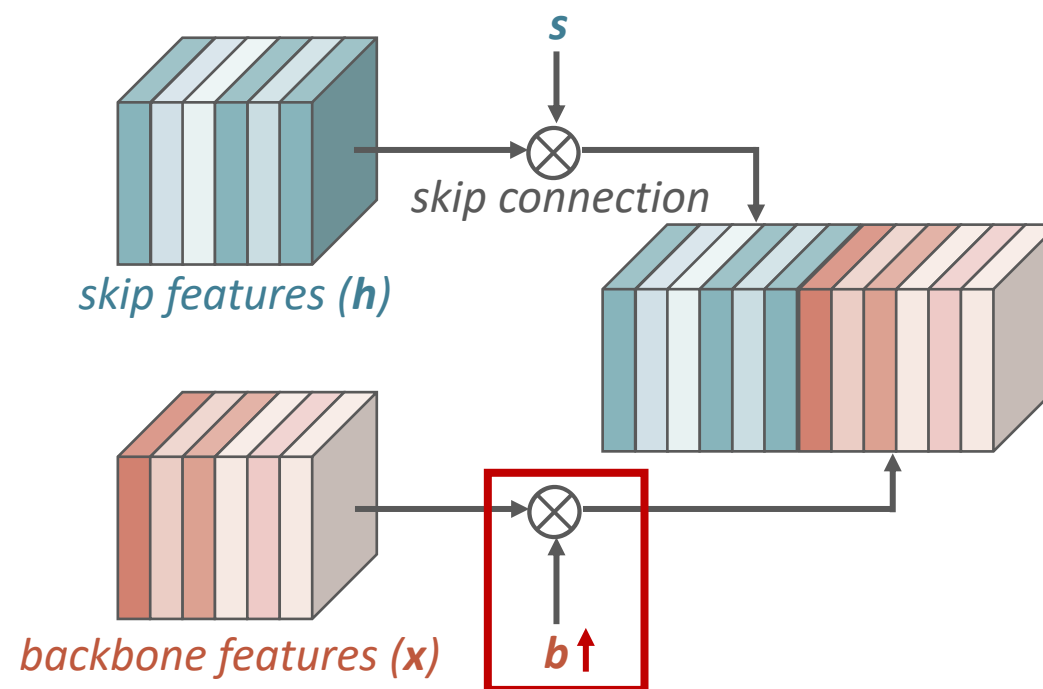
- 1: $\mathbf{x}_T \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$
- 2: **for** $t = T, \dots, 1$ **do**
- 3: $\mathbf{z} \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$ if $t > 1$, else $\mathbf{z} = \mathbf{0}$
- 4: $\mathbf{x}_{t-1} = \frac{1}{\sqrt{\alpha_t}} \left(\mathbf{x}_t - \frac{1 - \alpha_t}{\sqrt{1 - \bar{\alpha}_t}} \epsilon_{\theta}(\mathbf{x}_t, t) \right) + \sigma_t \mathbf{z}$
- 5: **end for**
- 6: **return** \mathbf{x}_0

FreeU Method

(1) enhance backbone features



(a) UNet Architecture



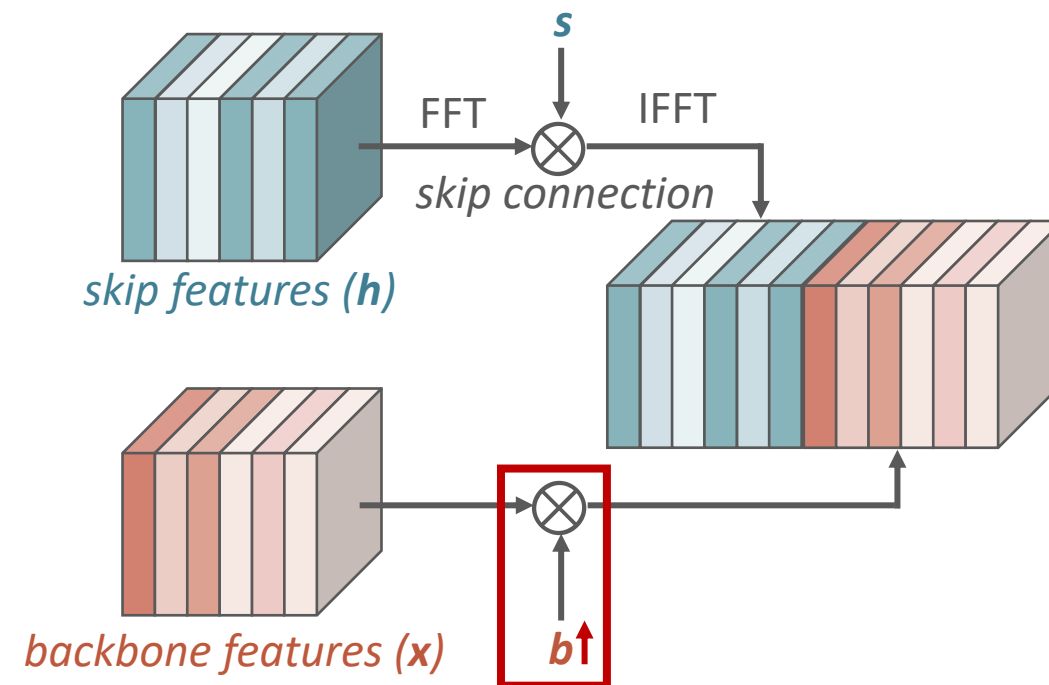
(b) FreeU Operations



FreeU Method

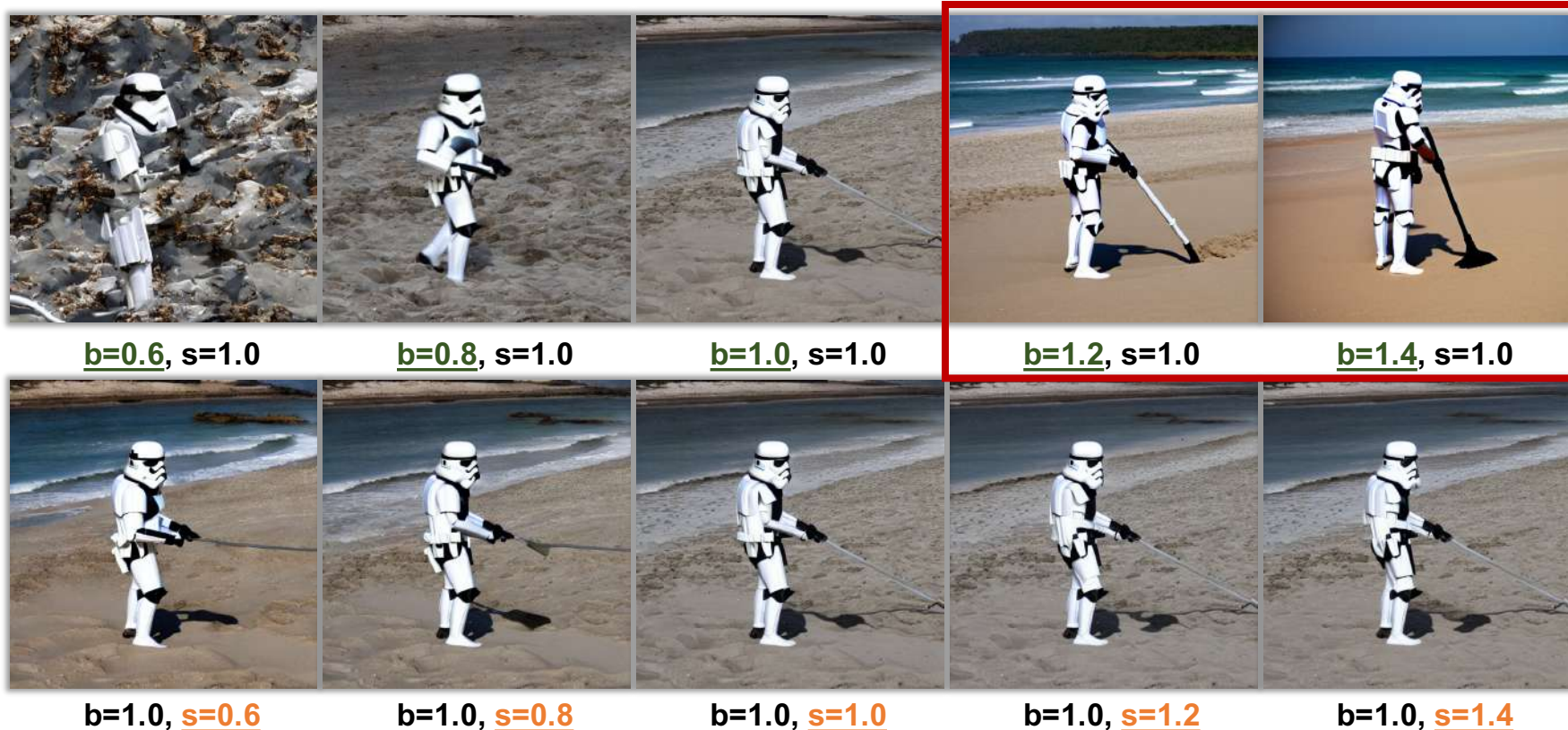
(1) enhance backbone features

Scale backbone features up
by a factor of b (e.g., $b=1.4$)



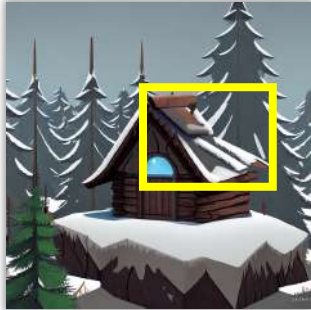
Ablation: Backbone Scaling Factor

- Enhancing backbone features can improve image quality



Ablation: Backbone Scaling Factor

$b = 1.0$



$b = 1.2$



$b = 1.4$



$b = 1.6$



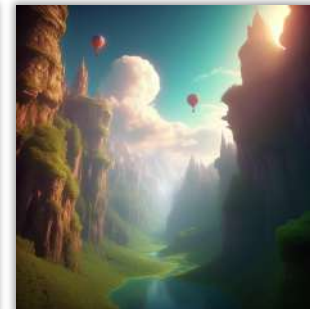
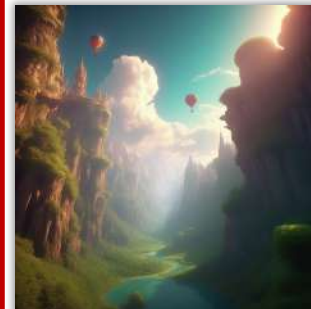
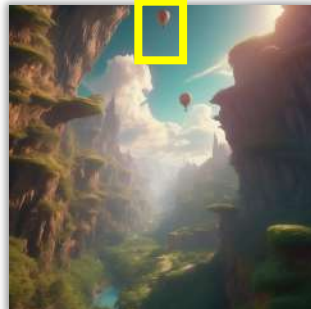
$b = 1.8$



A small cabin on top of a snowy mountain in the style of Disney, artstation



A drone view of celebration with Christmas tree and fireworks, starry sky - background.

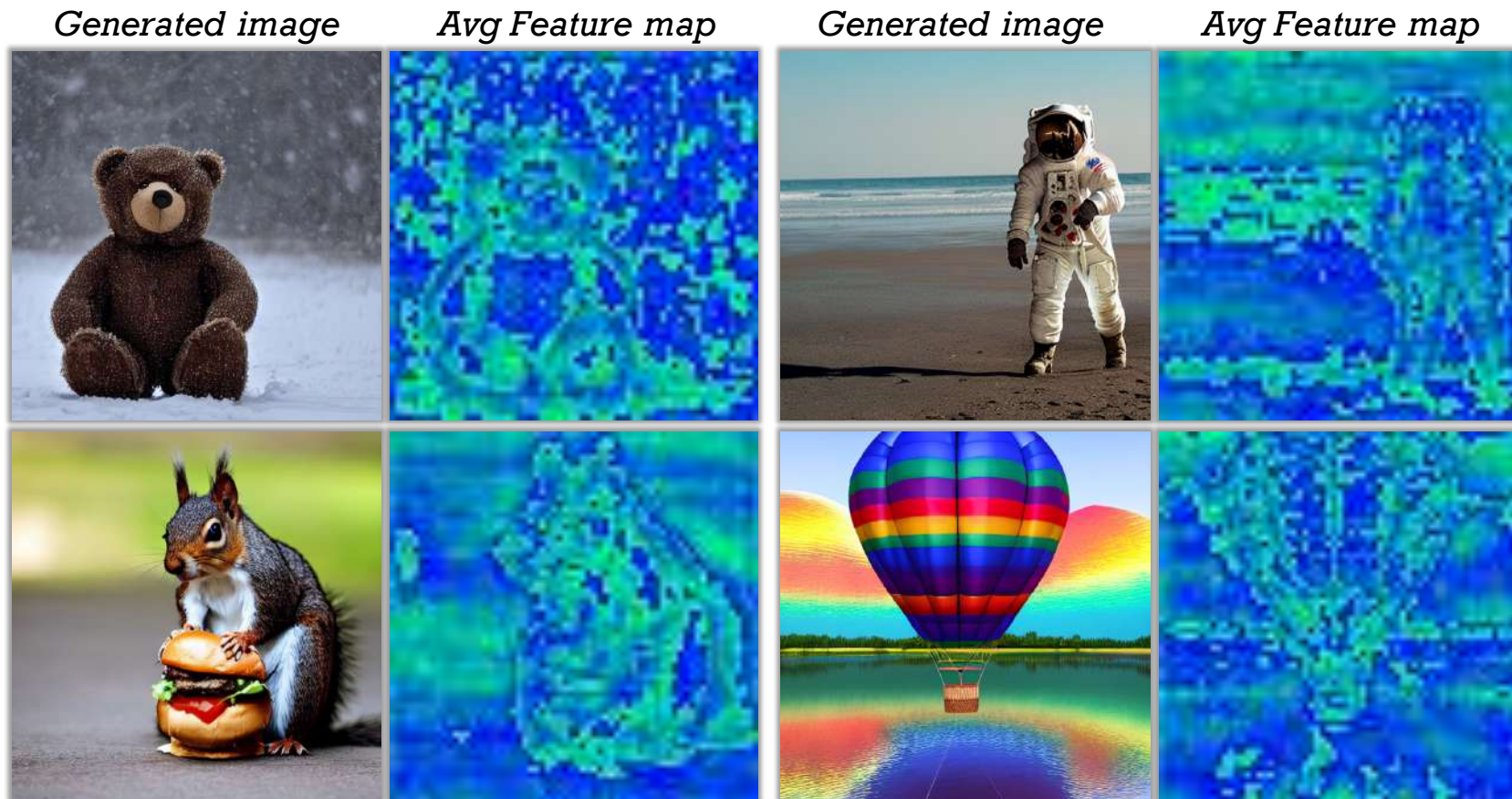


Flying through fantasy landscapes, 4k, high resolution.



Average Backbone Feature Maps

- Now: same backbone scaling everywhere.
- Is there a better way?



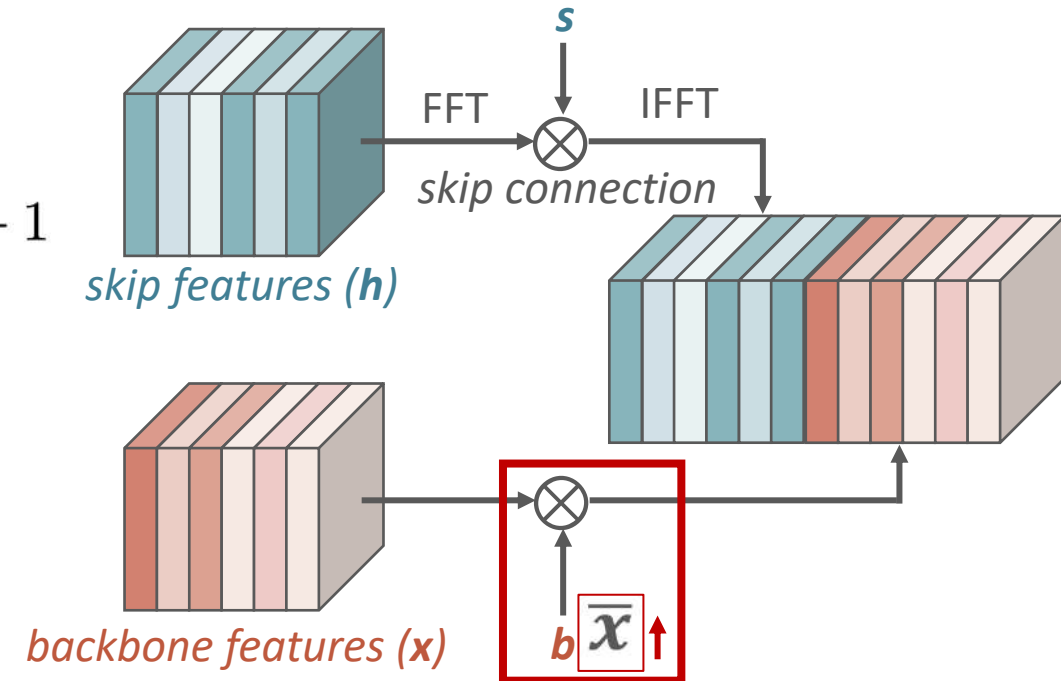
FreeU Method

(1) enhance backbone features

(2) content-aware backbone enhancement

$$\bar{x}_l = \frac{1}{C} \sum_{i=1}^C x_{l,i} \quad \alpha_l = (b_l - 1) \cdot \frac{\bar{x}_l - \text{Min}(\bar{x}_l)}{\text{Max}(\bar{x}_l) - \text{Min}(\bar{x}_l)} + 1$$

- spatially adaptive
- instance specific



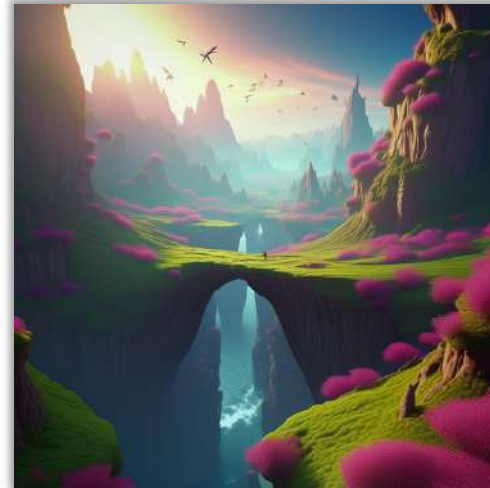
Content-Aware Backbone Scaling

Without FreeU



(a)

Constant
Backbone Scaling



(b)

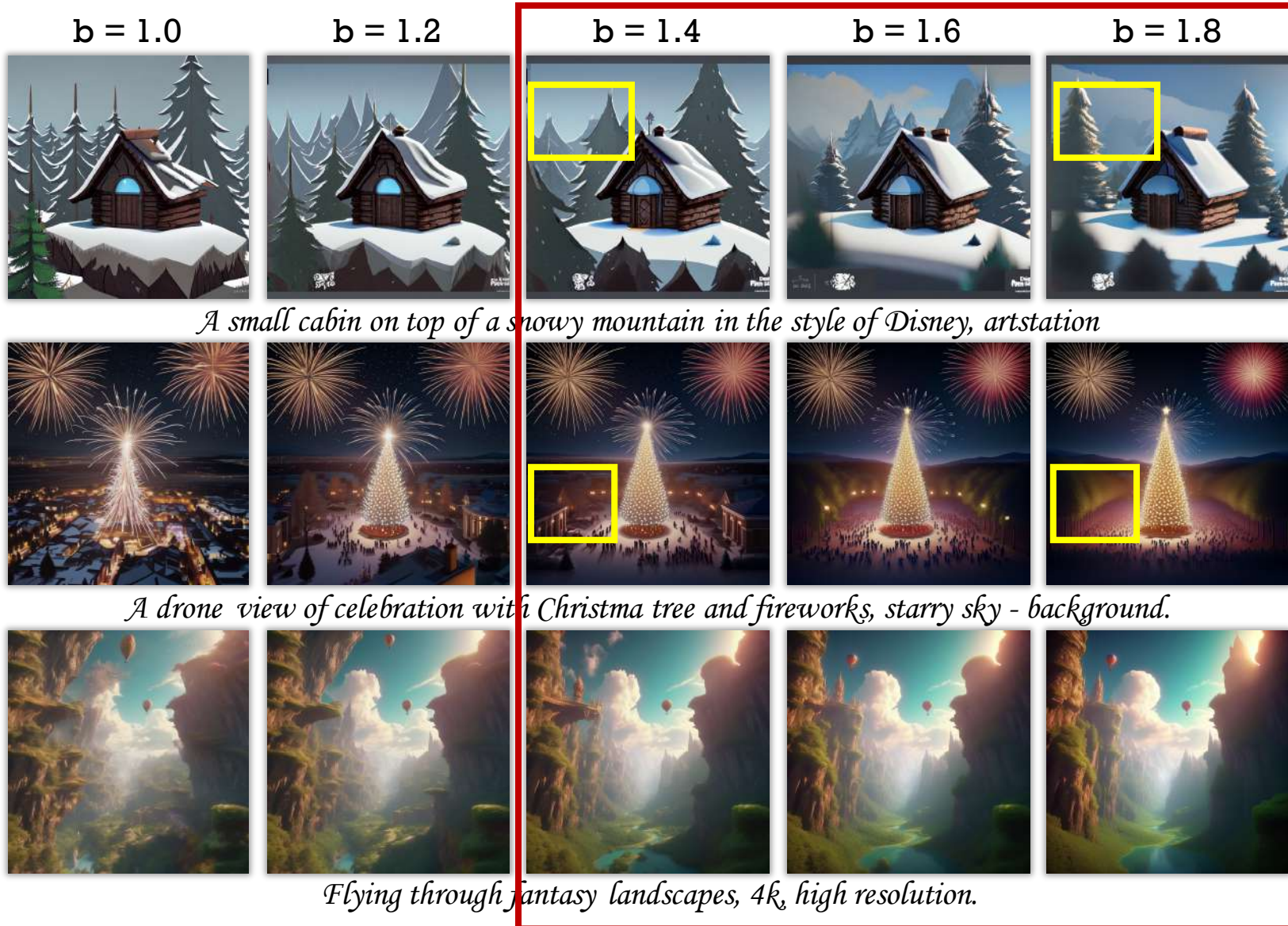
Content-Aware
Backbone Scaling



(c)



Ablation: Backbone Scaling Factor



with increased backbone scaling, image can be oversmoothed



FreeU Method

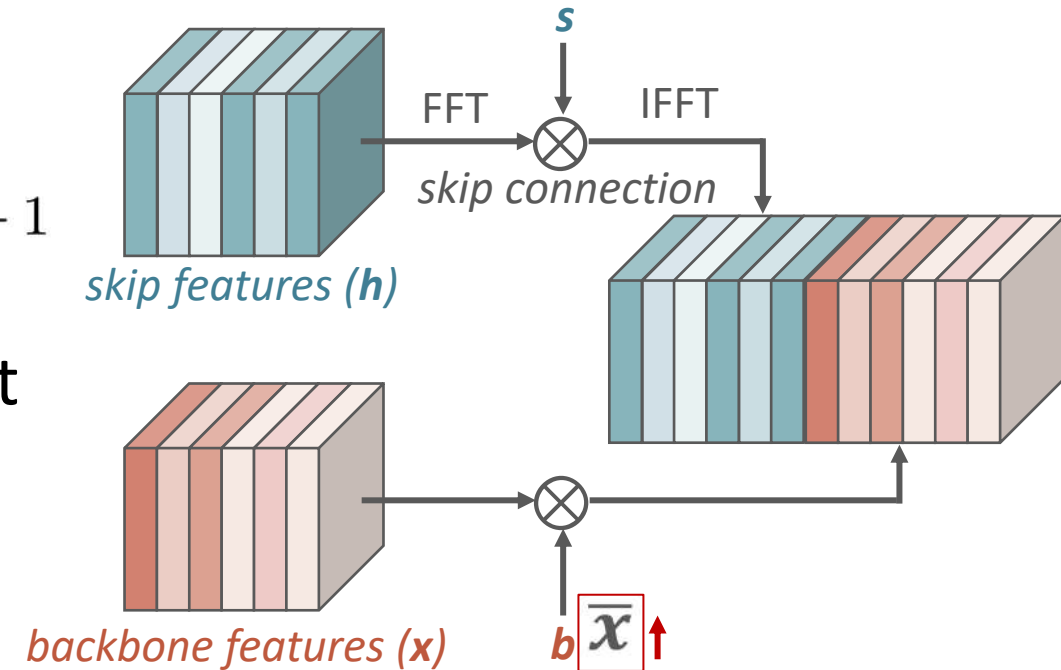
(1) enhance backbone features

(2) content-aware backbone enhancement

$$\bar{x}_l = \frac{1}{C} \sum_{i=1}^C x_{l,i} \quad \alpha_l = (b_l - 1) \cdot \frac{\bar{x}_l - \text{Min}(\bar{x}_l)}{\text{Max}(\bar{x}_l) - \text{Min}(\bar{x}_l)} + 1$$

(3) channel-selective backbone enhancement

$$x'_{l,i} = \begin{cases} x_{l,i} \odot \alpha_l, & \text{if } i < C/2 \\ x_{l,i}, & \text{otherwise} \end{cases}$$



Channel Selection of Backbone Scaling

No Scaling

Scale All

Select
First Half

Select
Second Half

Uniform
Selection



A drone view of celebration with Christmas tree and fireworks, starry sky - background.



Flying through fantasy landscapes, 4k, high resolution.



A fat rabbit wearing a purple robe walking through a fantasy landscape.



FreeU Method

(1) enhance backbone features

(2) content-aware backbone enhancement

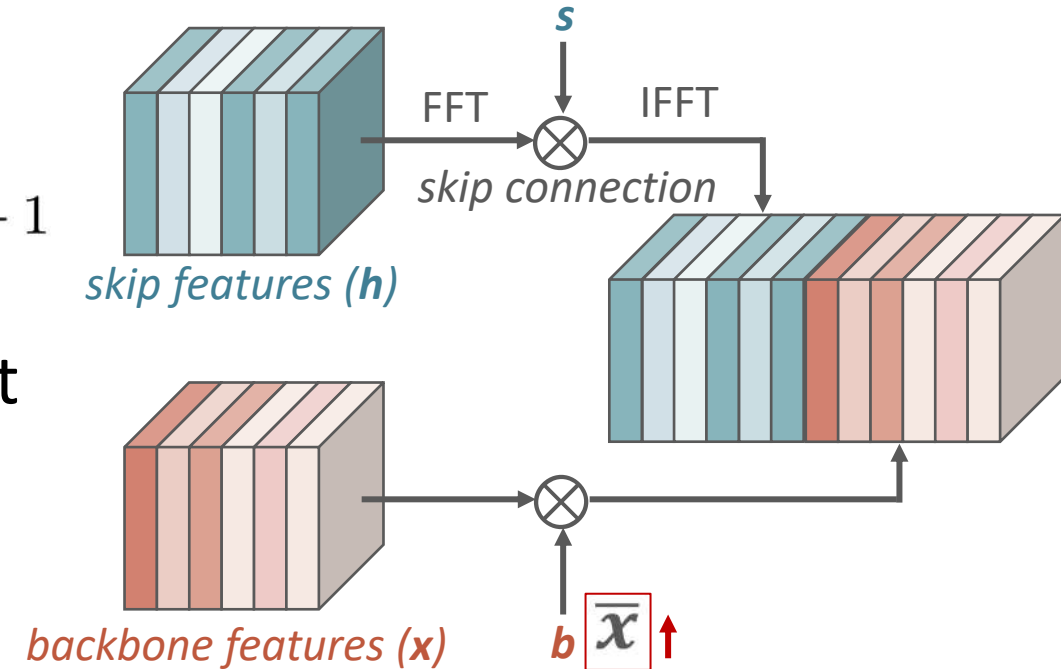
$$\bar{x}_l = \frac{1}{C} \sum_{i=1}^C x_{l,i} \quad \alpha_l = (b_l - 1) \cdot \frac{\bar{x}_l - \text{Min}(\bar{x}_l)}{\text{Max}(\bar{x}_l) - \text{Min}(\bar{x}_l)} + 1$$

(3) channel-selective backbone enhancement

$$x'_{l,i} = \begin{cases} x_{l,i} \odot \alpha_l, & \text{if } i < C/2 \\ x_{l,i}, & \text{otherwise} \end{cases}$$

(4) suppress low-frequency in skip features

$$\beta_{l,i}(r) = \begin{cases} s_l & \text{if } r < r_{\text{thresh}}, \\ 1 & \text{otherwise.} \end{cases} \quad \begin{aligned} \mathcal{F}(h_{l,i}) &= \text{FFT}(h_{l,i}) \\ \mathcal{F}'(h_{l,i}) &= \mathcal{F}(h_{l,i}) \odot \beta_{l,i} \\ h'_{l,i} &= \text{IFFT}(\mathcal{F}'(h_{l,i})) \end{aligned}$$



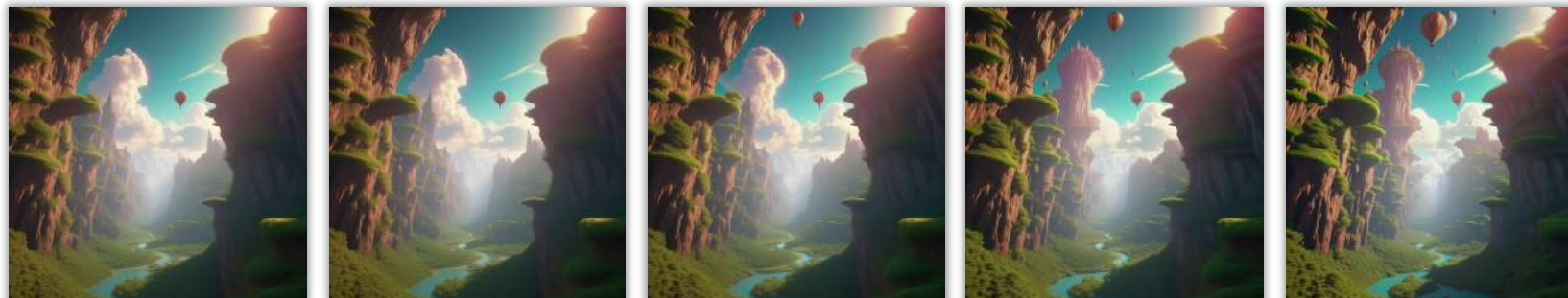
Ablation: Skip Scaling Factor



A small cabin on top of a snowy mountain in the style of Disney, artstation



A drone view of celebration with Christmas tree and fireworks, starry sky - background.



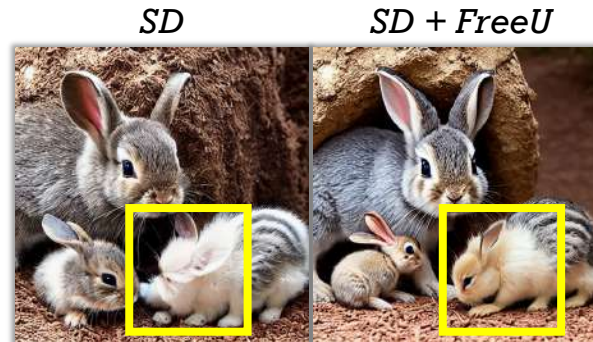
Flying through fantasy landscapes, 4k, high resolution.



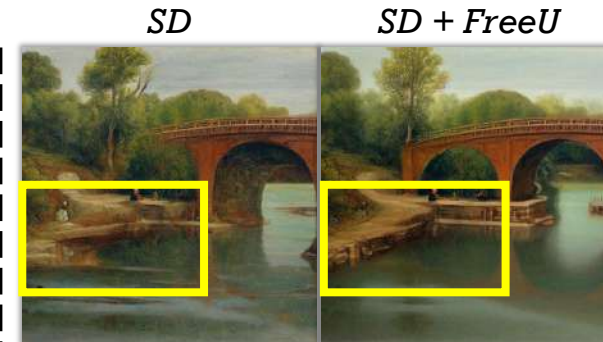
Visual Results: Text-to-Image



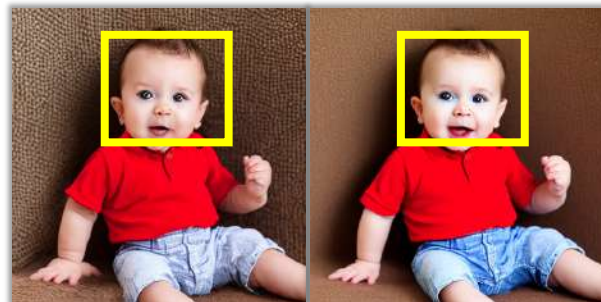
a blue car is being filmed



Mother rabbit is raising baby rabbits



A bridge is depicted in the water



a baby in a red shirt



a attacks an upset cat and is then chased off



A teddy bear walking in the snowstorm



A cat riding a motorcycle.



A panda standing on a surfboard in the ocean



A boy is playing pokémon



Visual Results: Text-to-Video

ModelScope



ModelScope + FreeU



Pacific coast, carmel by the sea ocean and waves.

ModelScope



ModelScope + FreeU



Michelangelo's sculpture of David wearing headphones djing.

ModelScope



ModelScope + FreeU



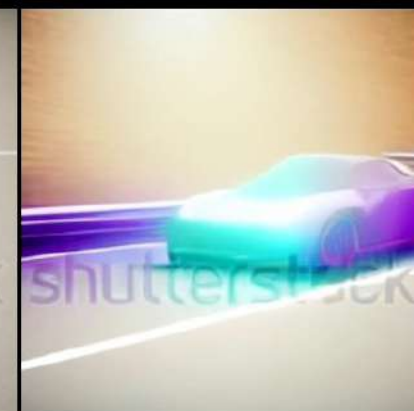
Milk dripping into a cup of coffee



An astronaut flying in space



Fireworks



synthwave sports car

Visual Results: Text-to-Video

ModelScope



ModelScope + FreeU



Fireworks

ModelScope



ModelScope + FreeU



A galloping horse

ModelScope



ModelScope + FreeU



A horse galloping on the ocean



Picturesque autumn scene of Altausseer See lake.

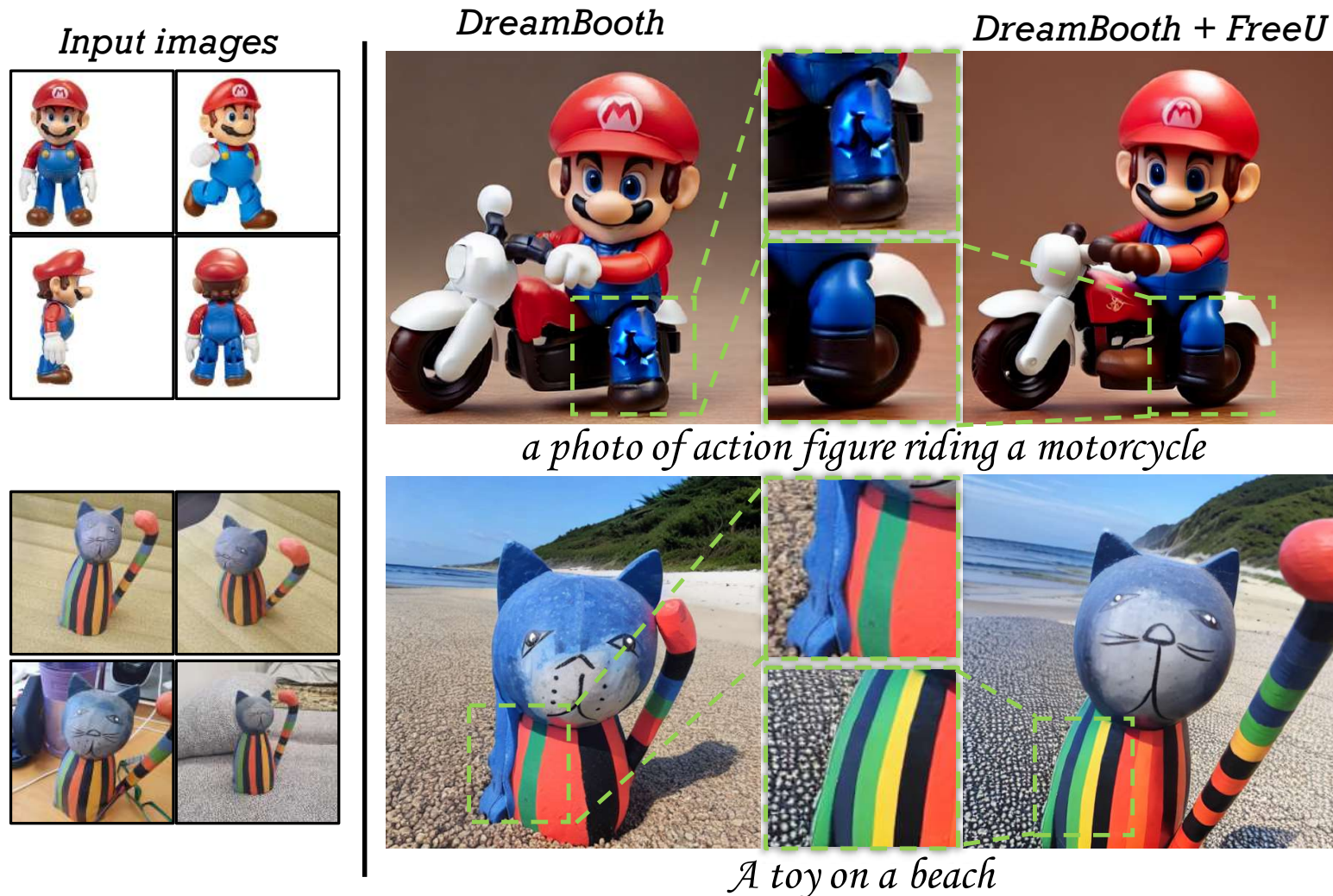


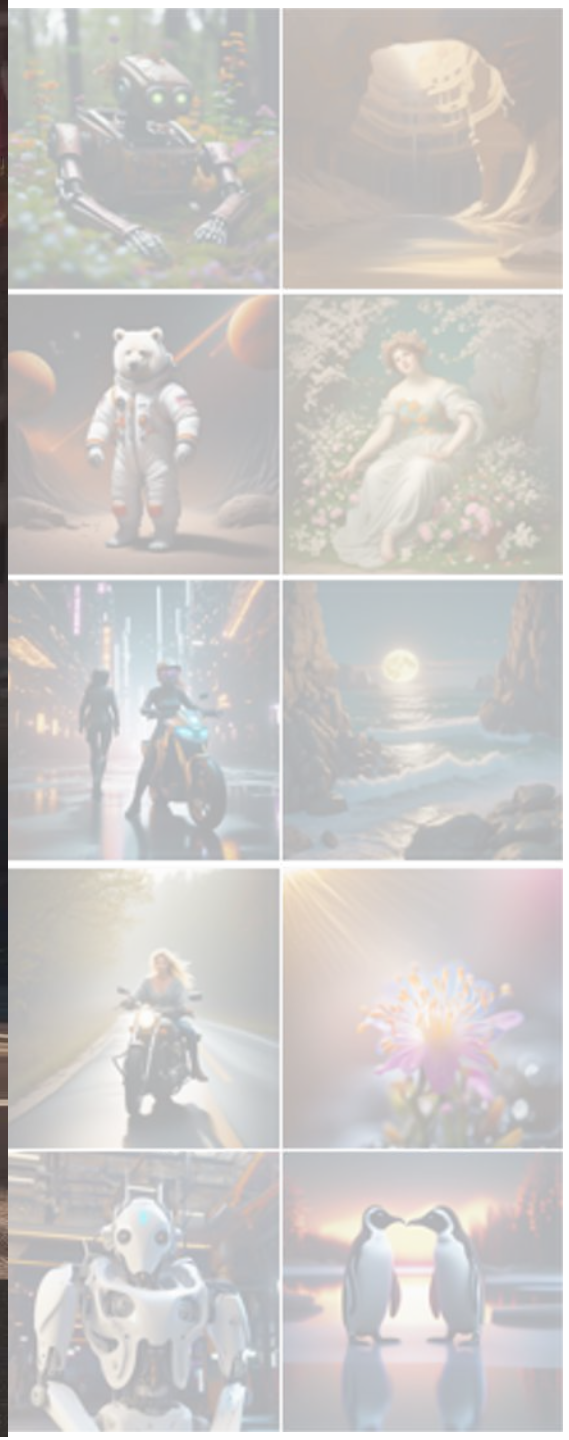
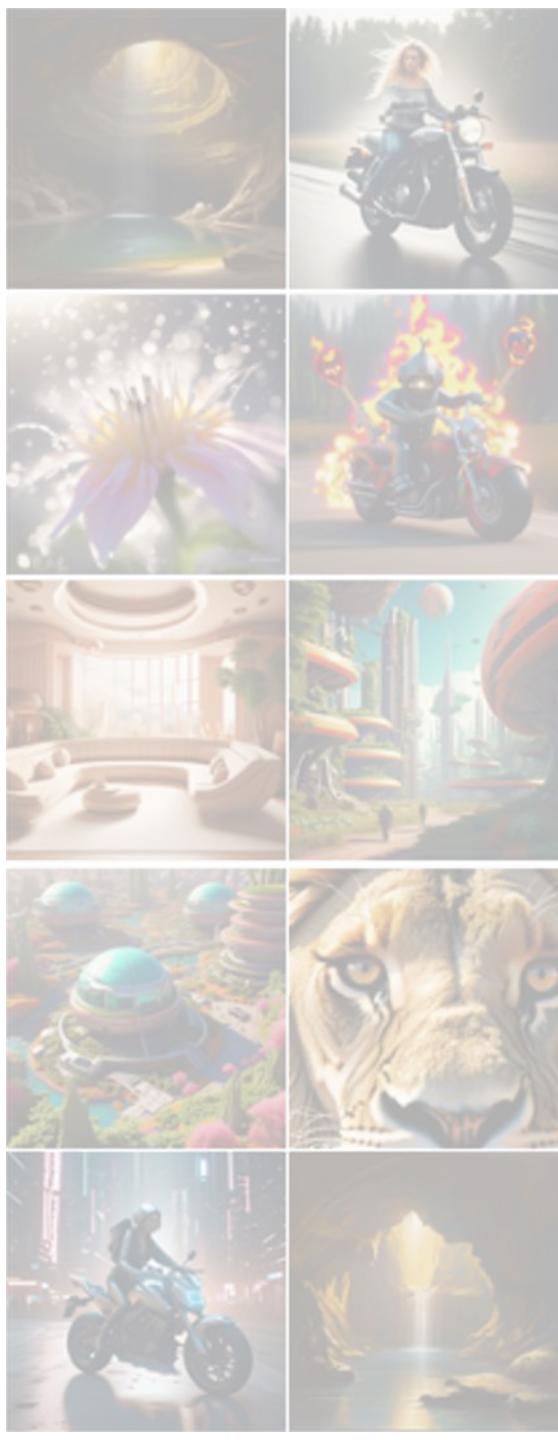
Sunset time lapse at the beach with moving clouds and colors in the sky

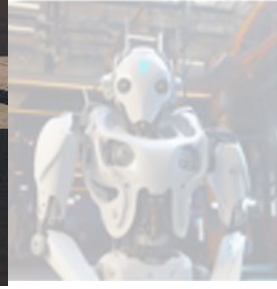
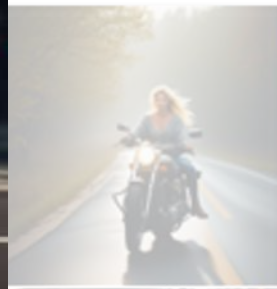
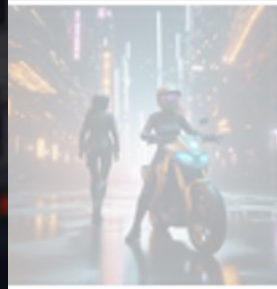
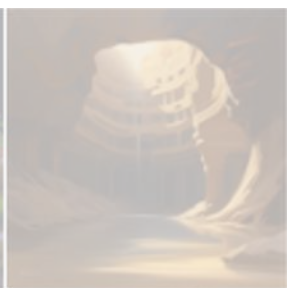
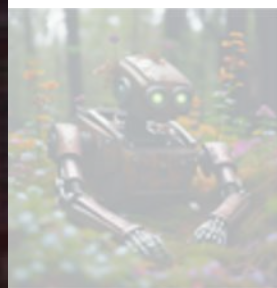
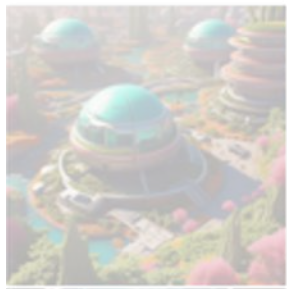
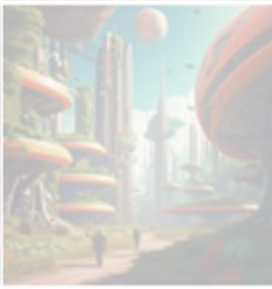
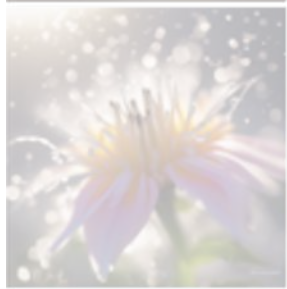
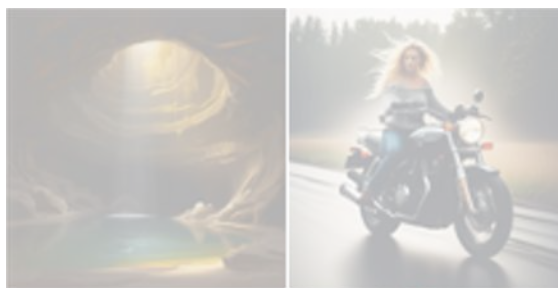


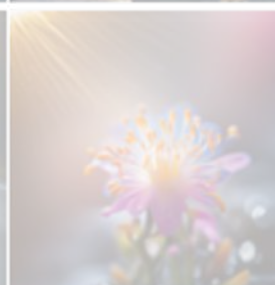
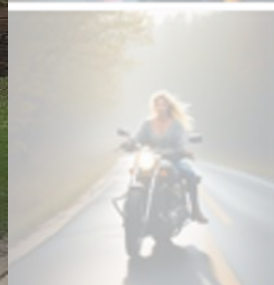
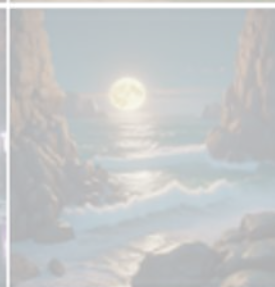
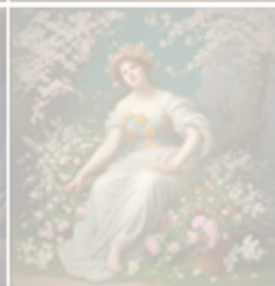
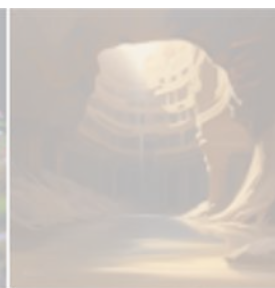
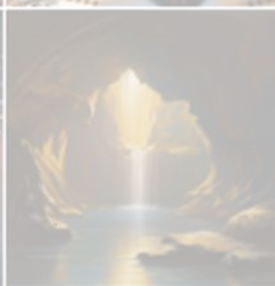
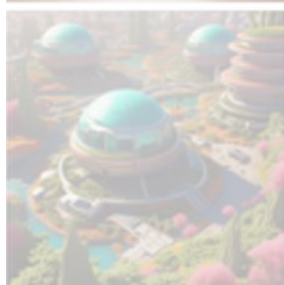
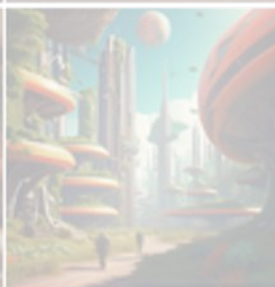
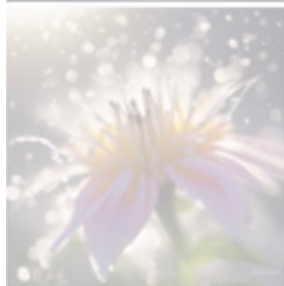
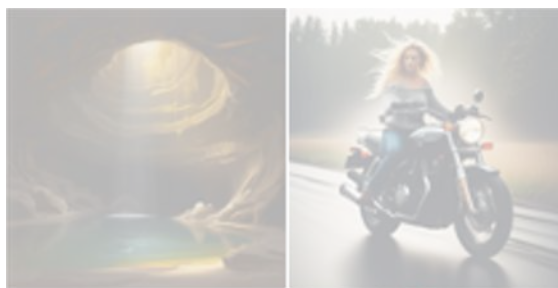
a shark is swimming in the ocean.

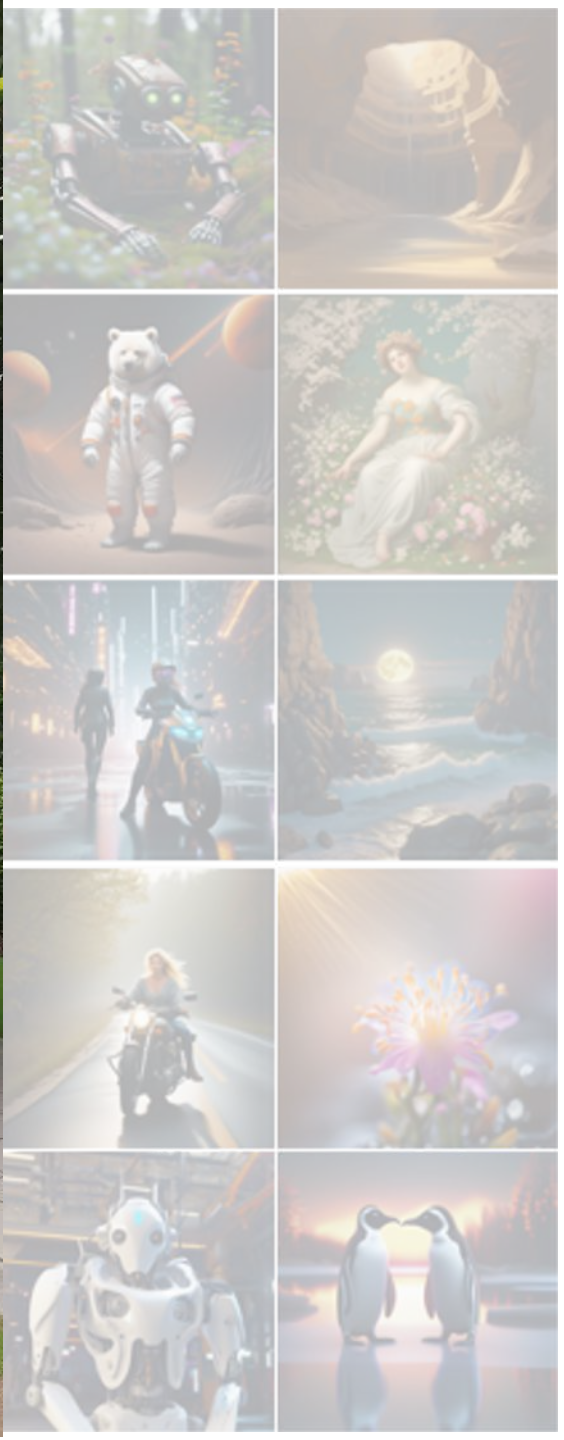
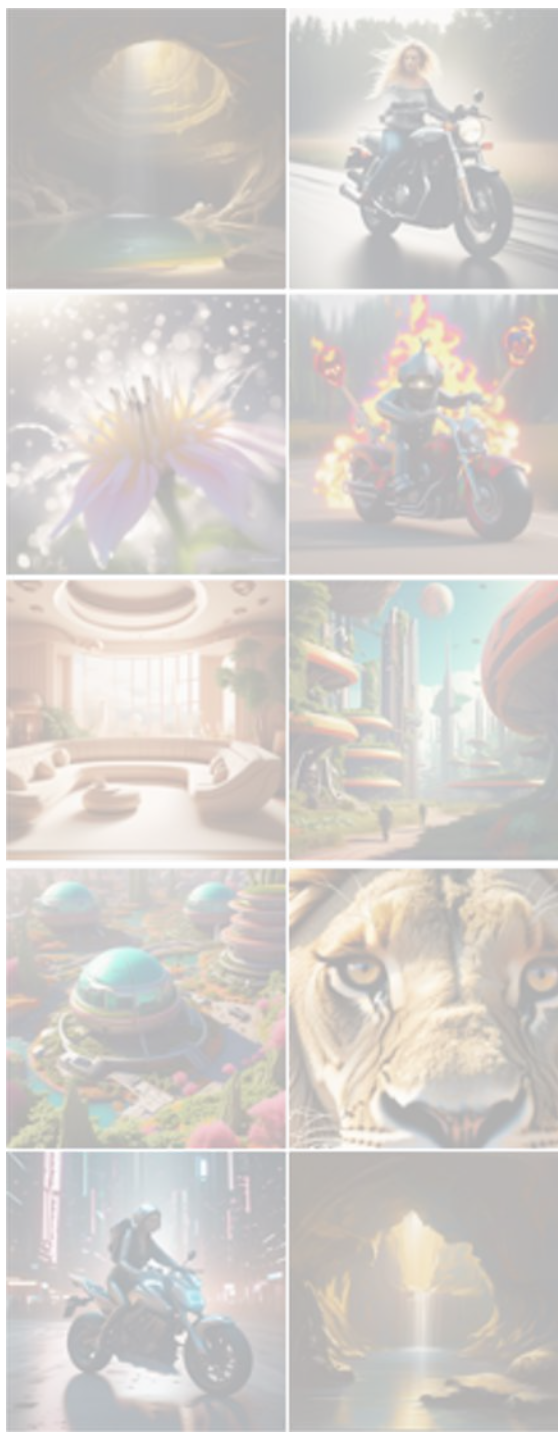
Visual Results: Personalized Text-to-Image

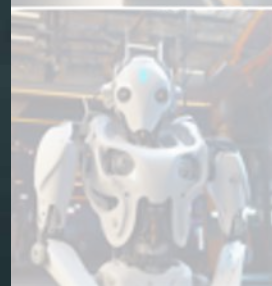
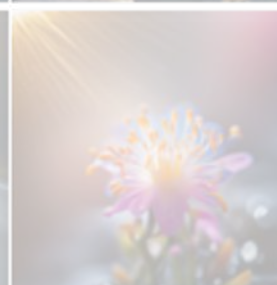
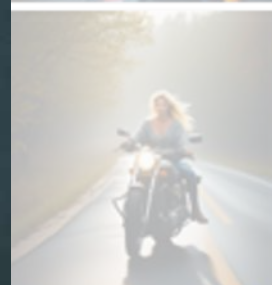
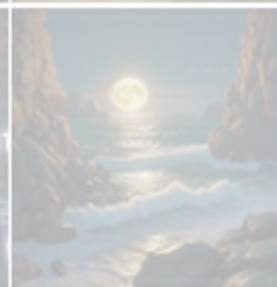
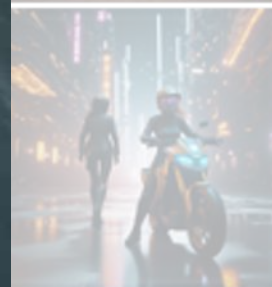
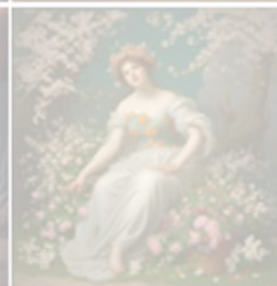
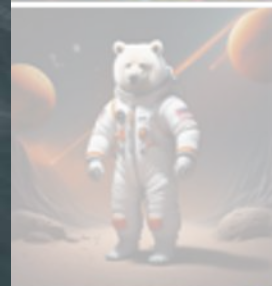
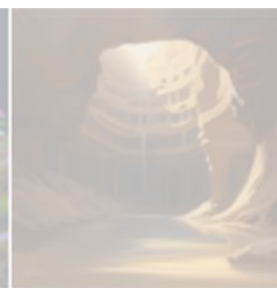
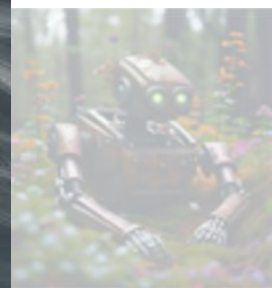
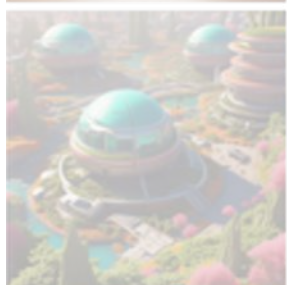
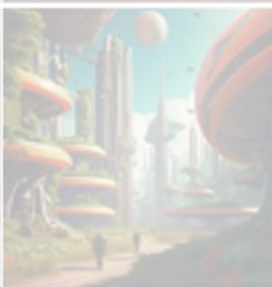
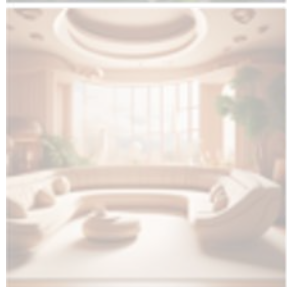
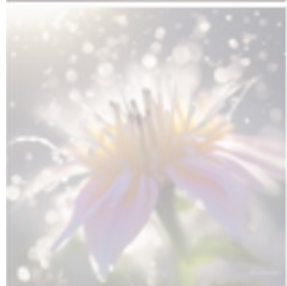
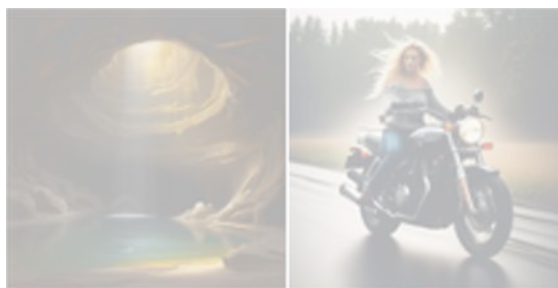


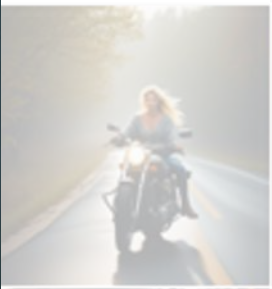
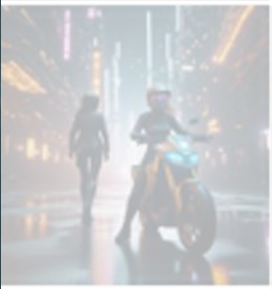
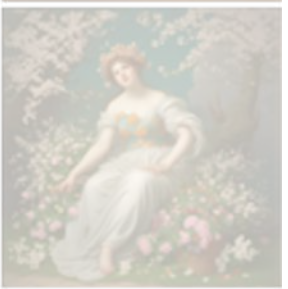
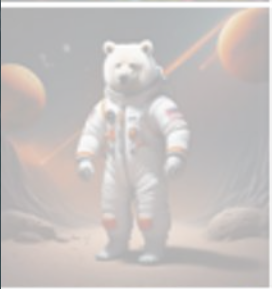
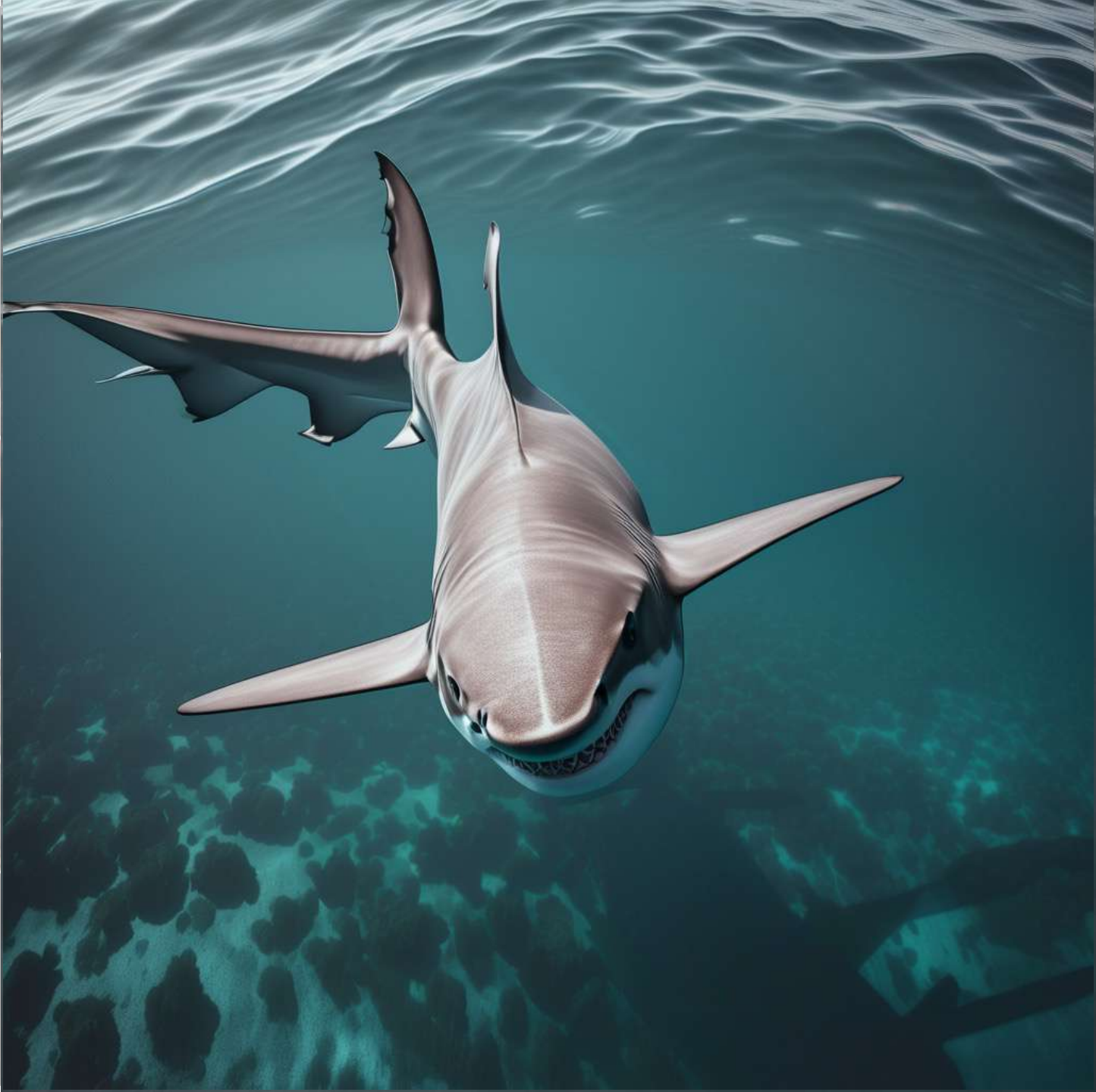
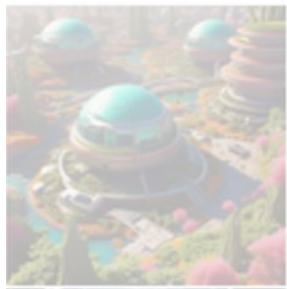
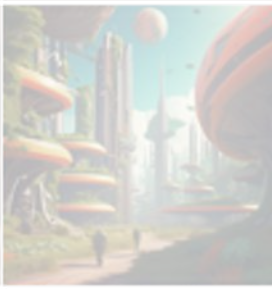
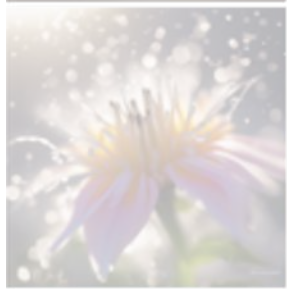
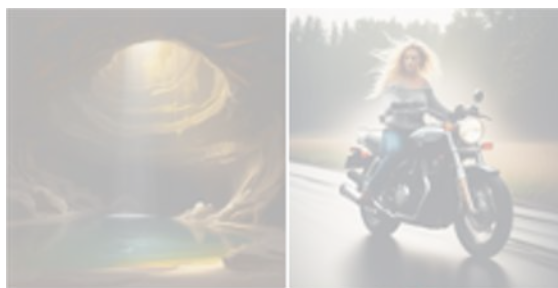


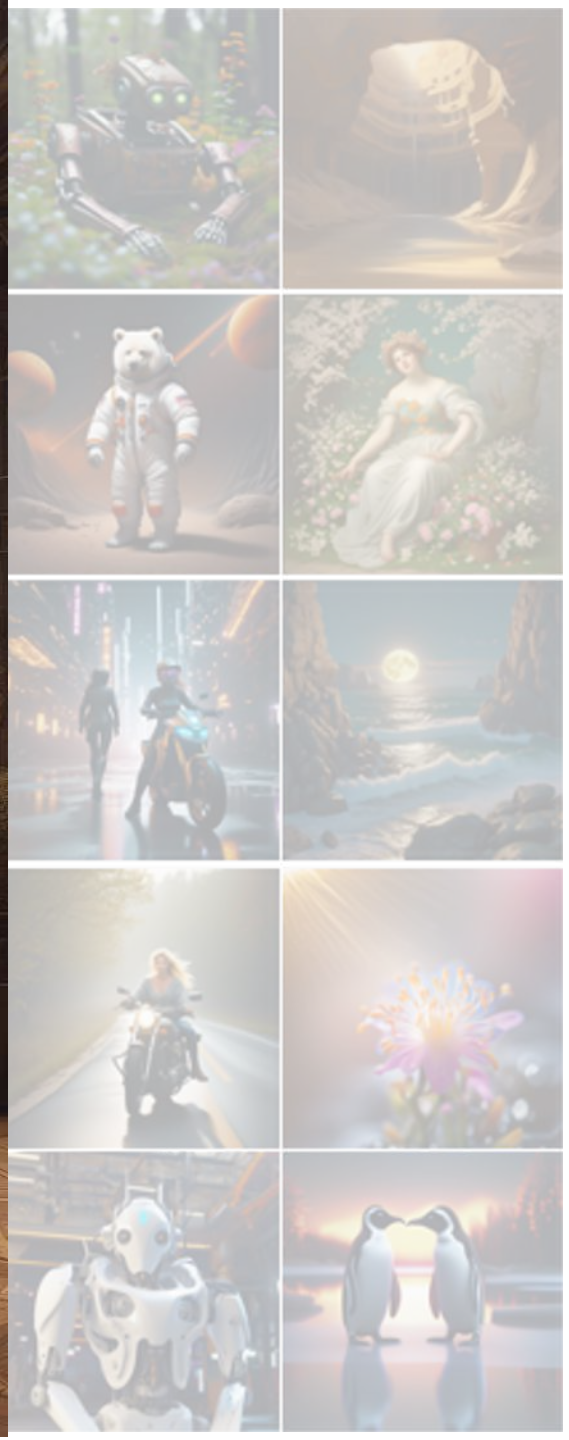
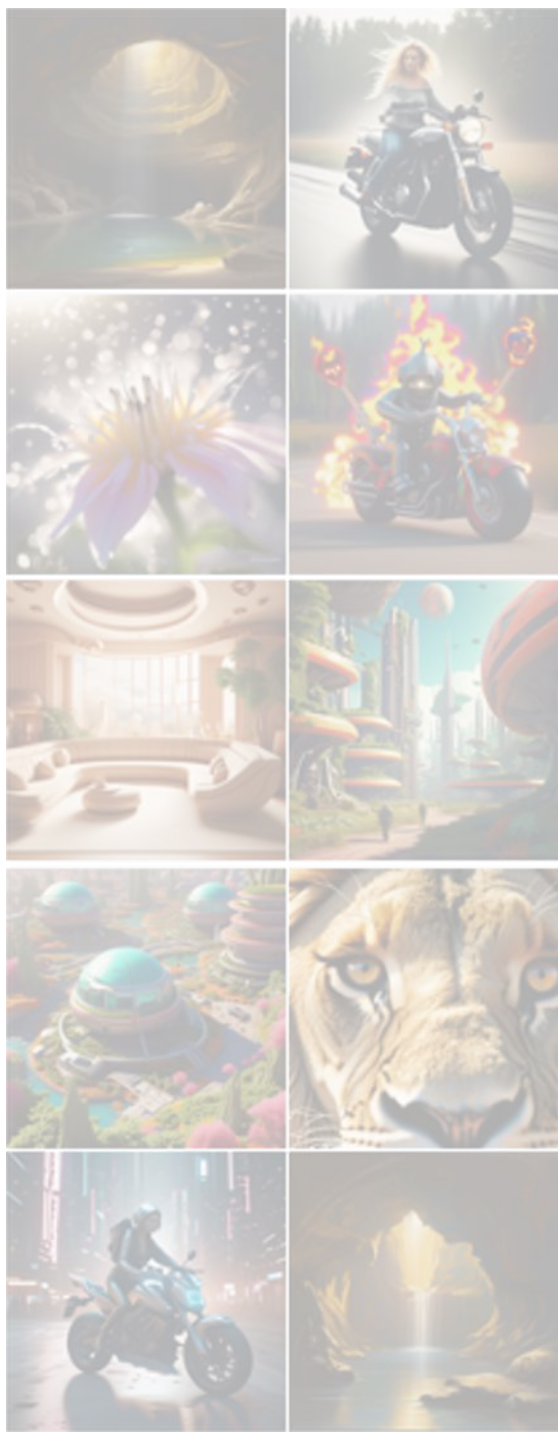


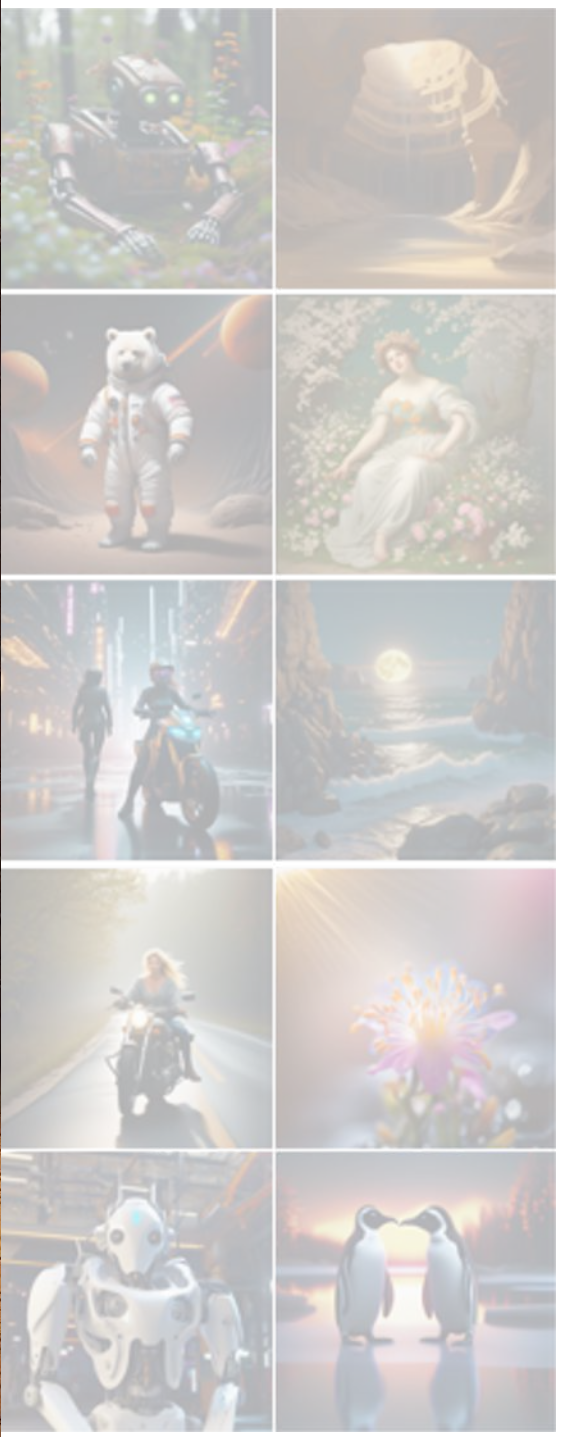
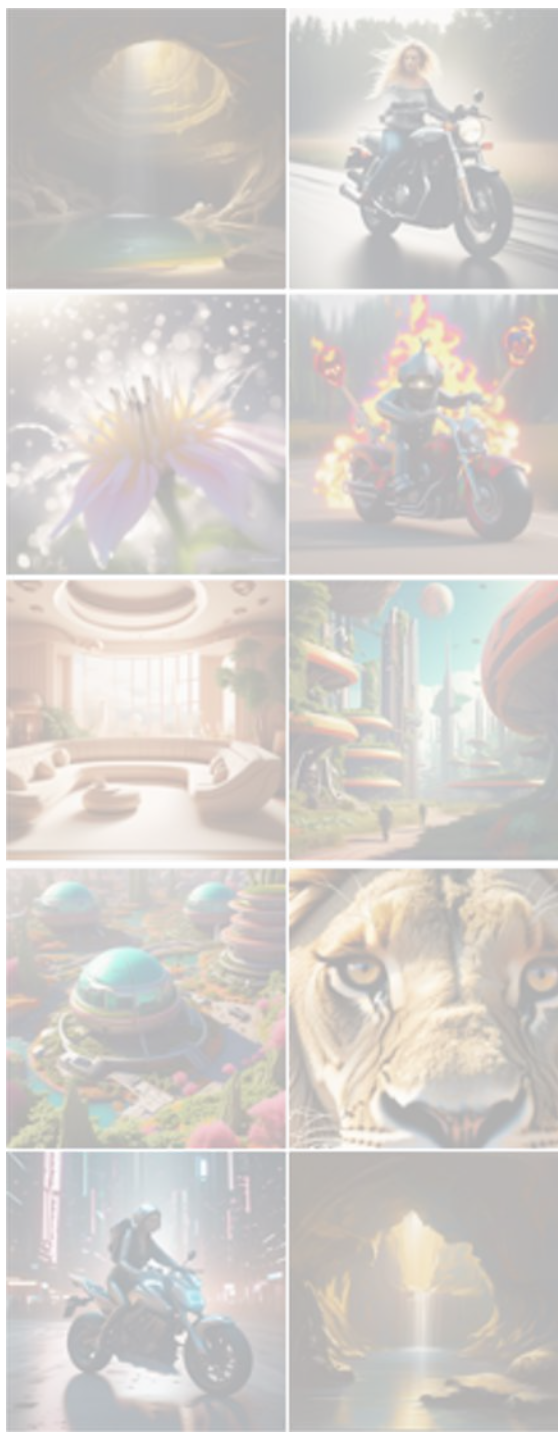


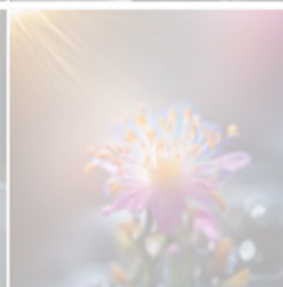
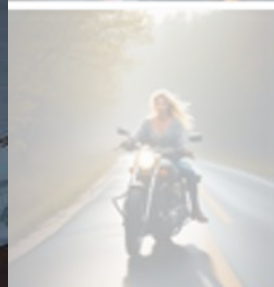
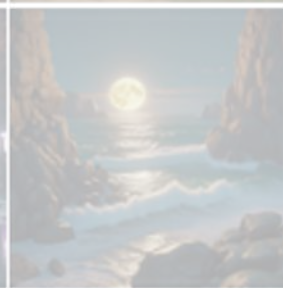
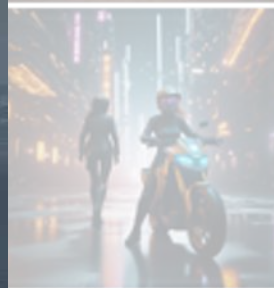
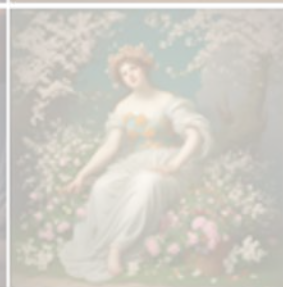
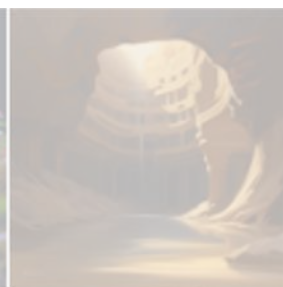
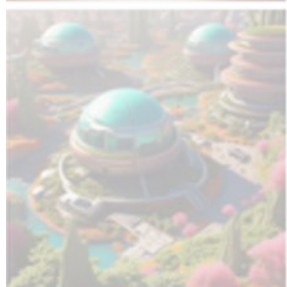
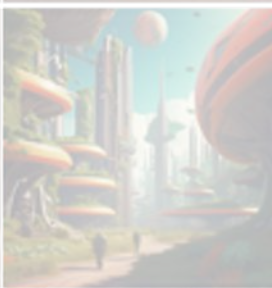
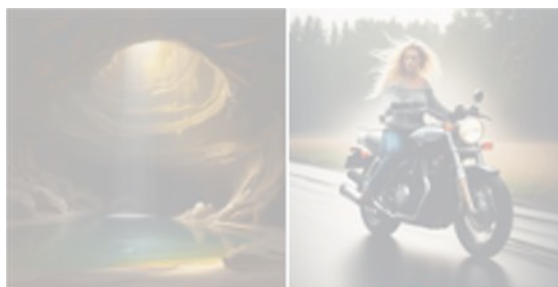


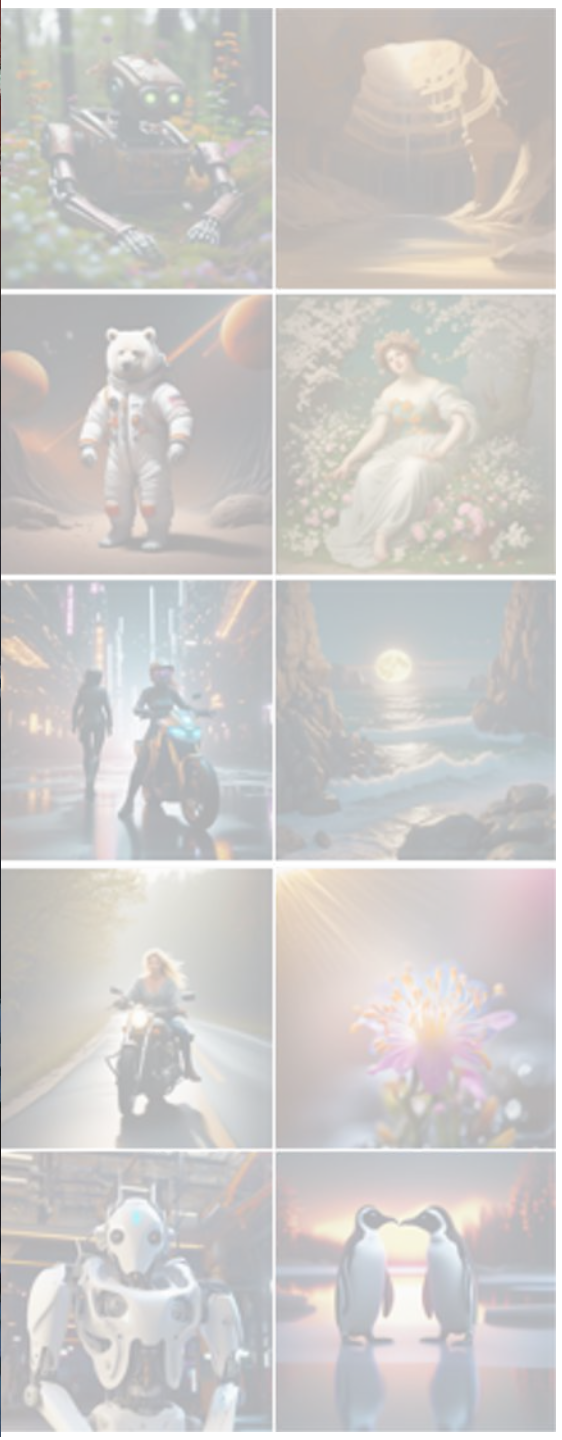
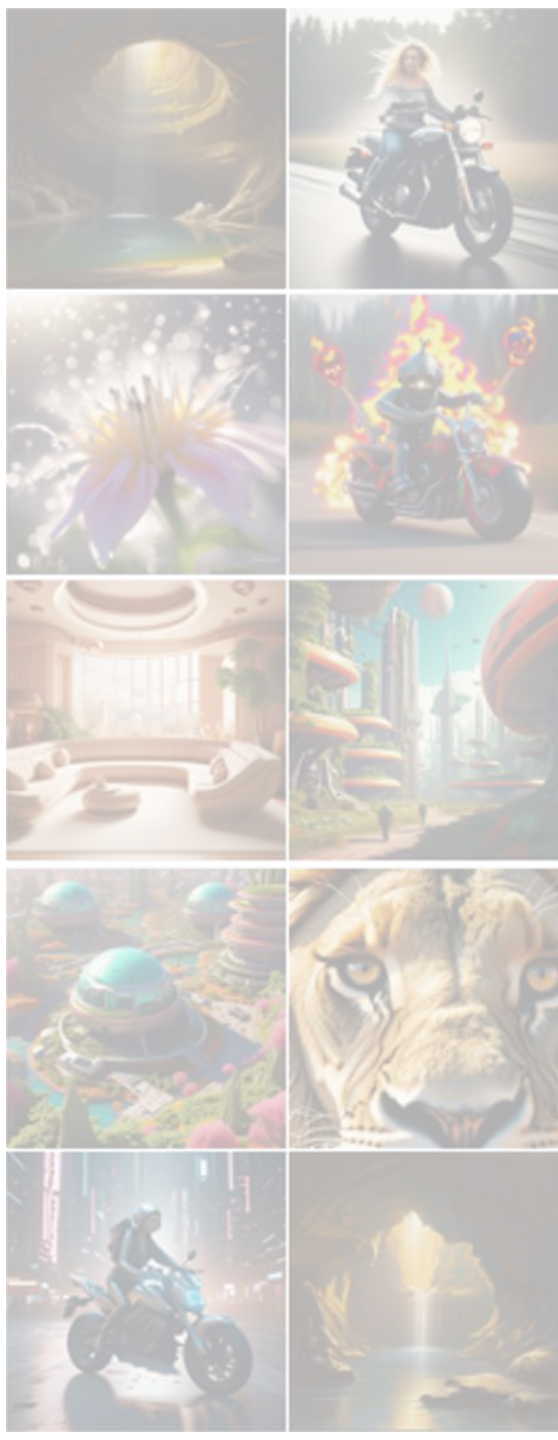


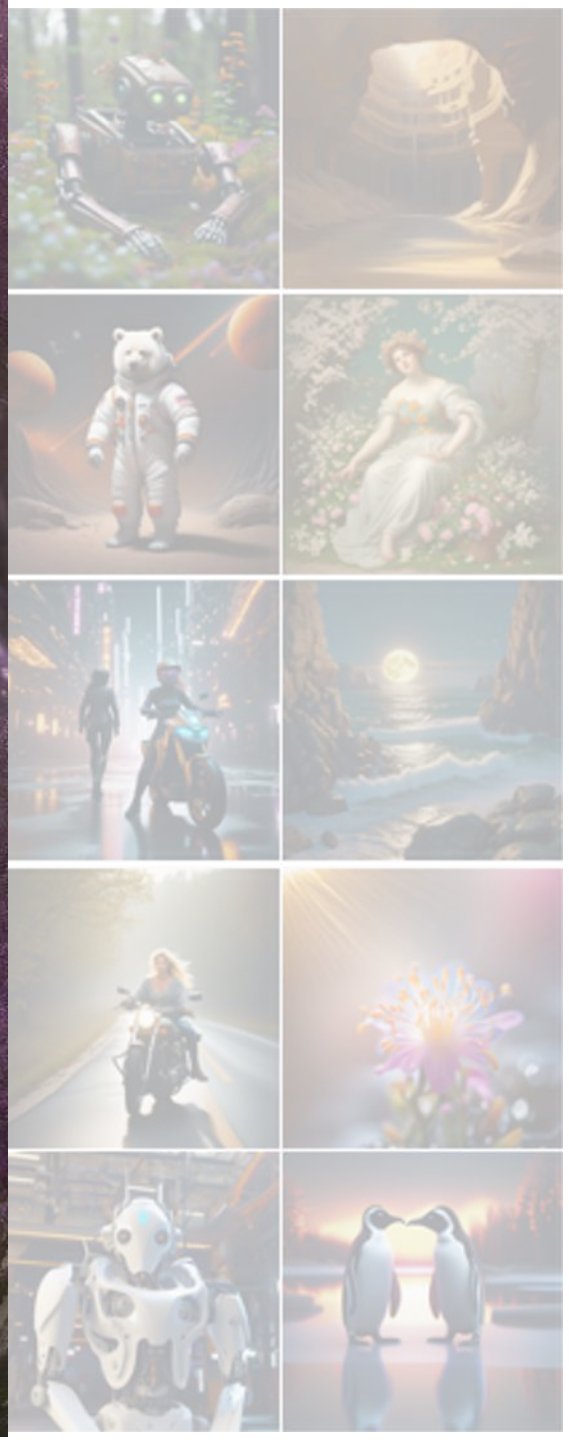
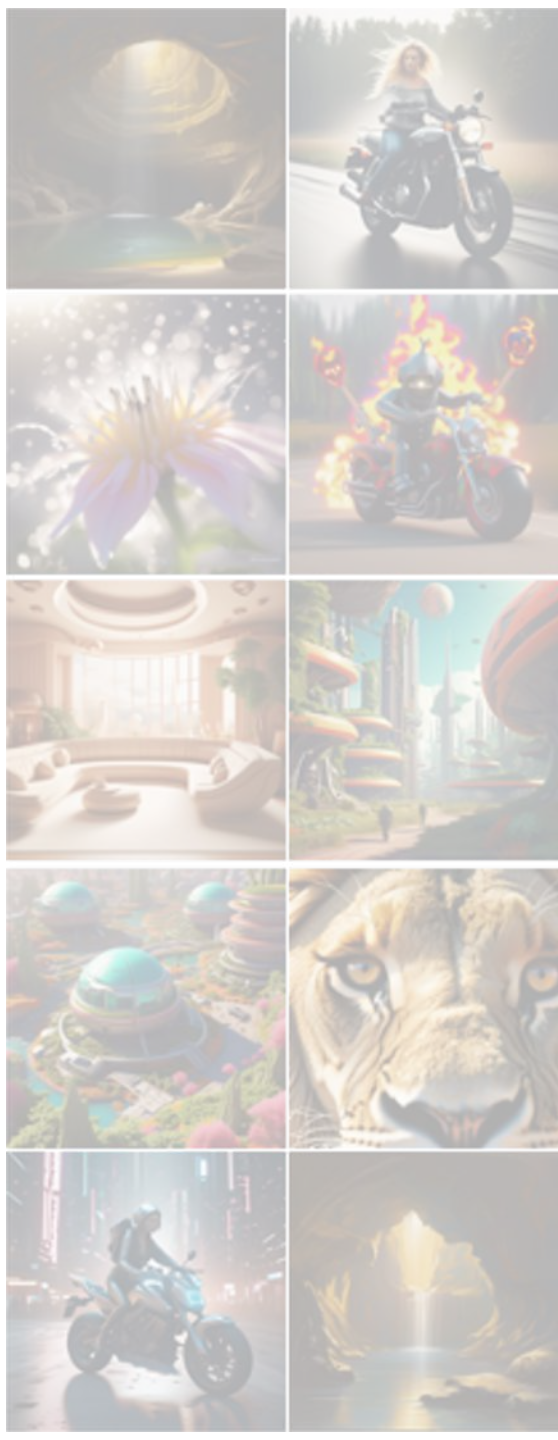


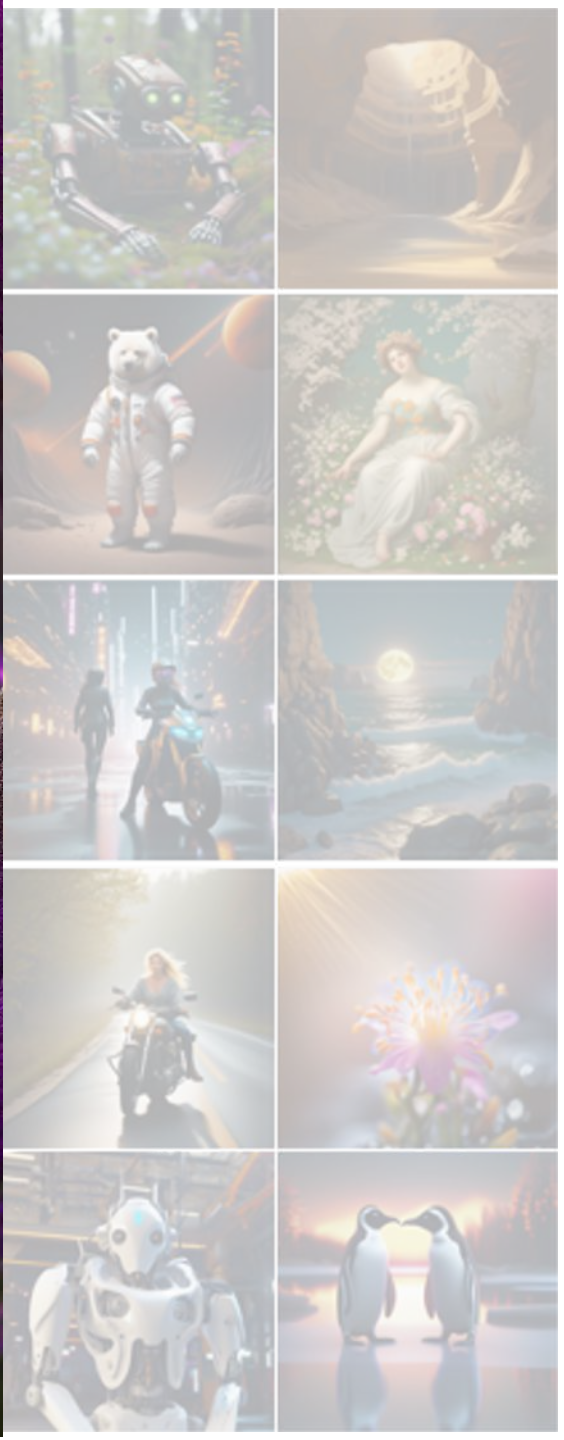
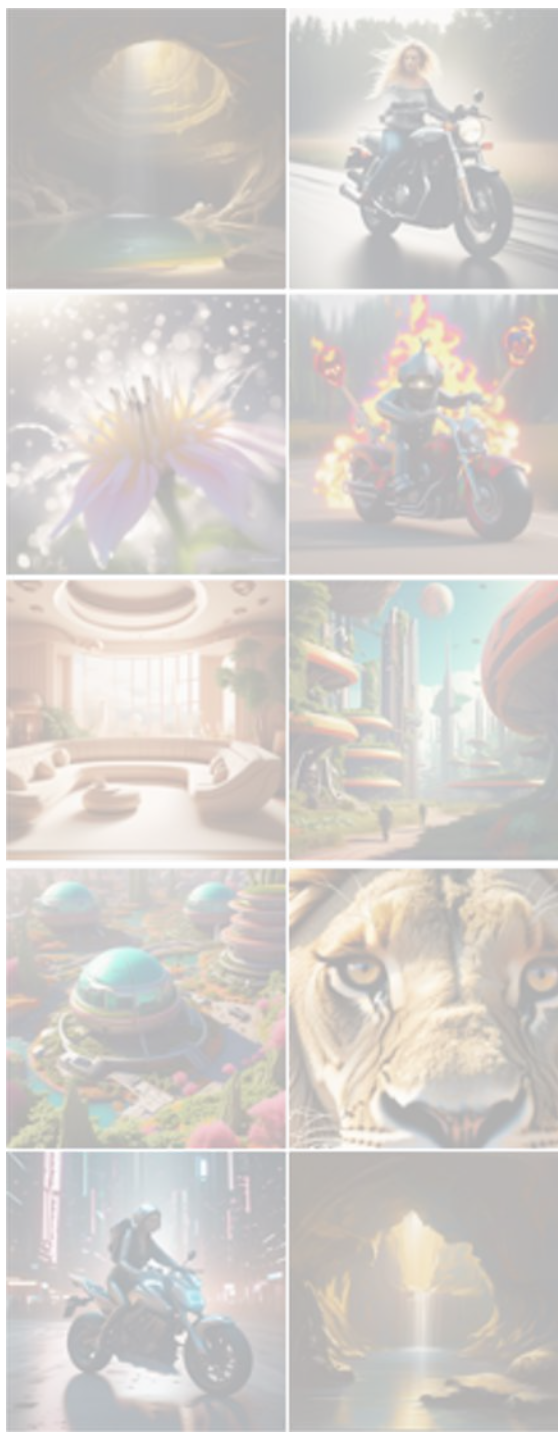


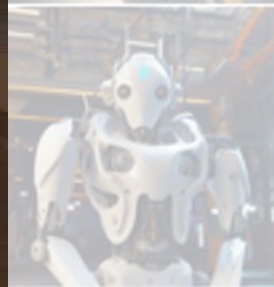
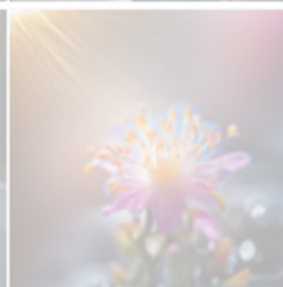
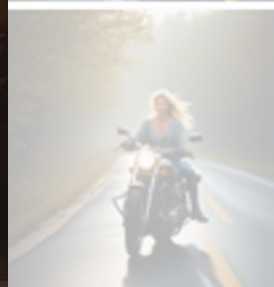
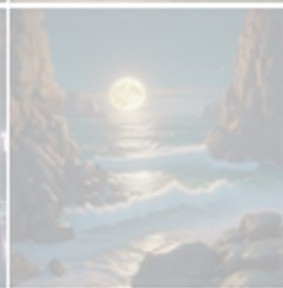
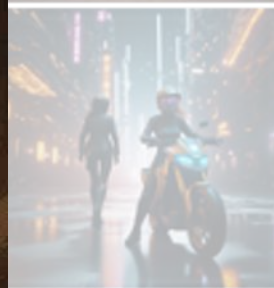
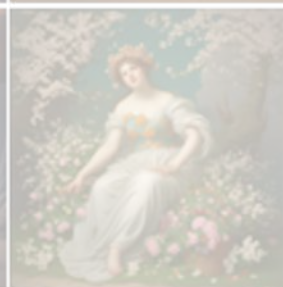
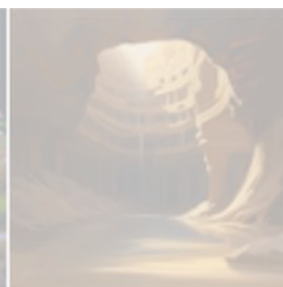
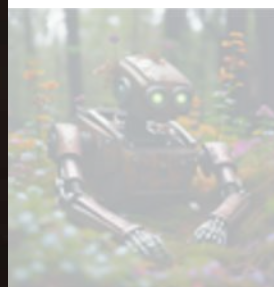
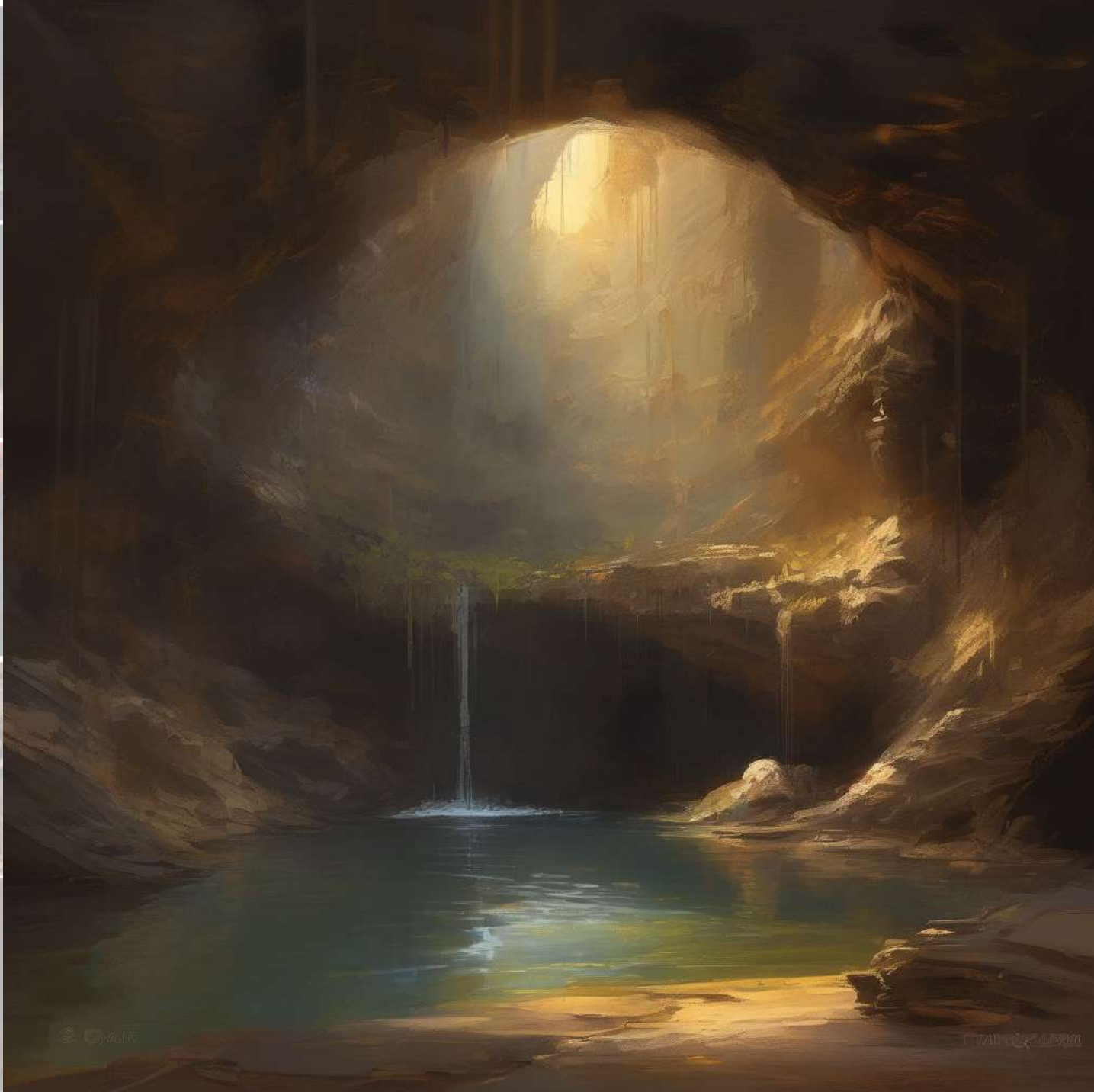
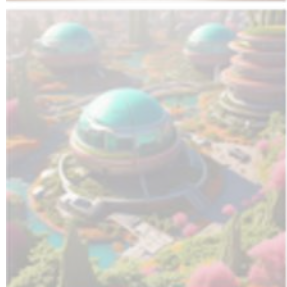
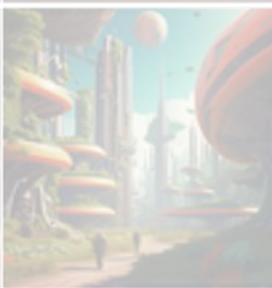
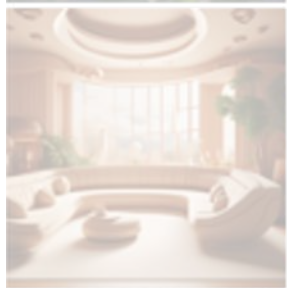
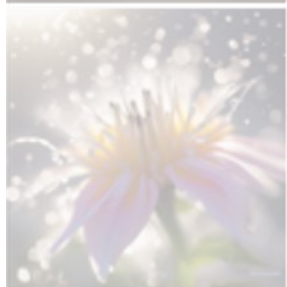
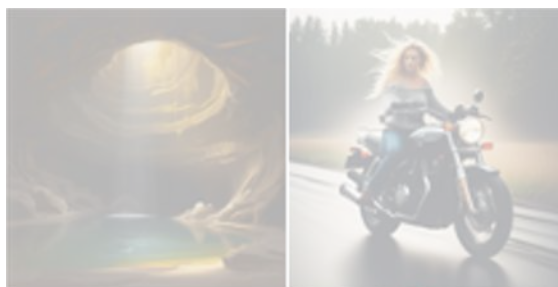


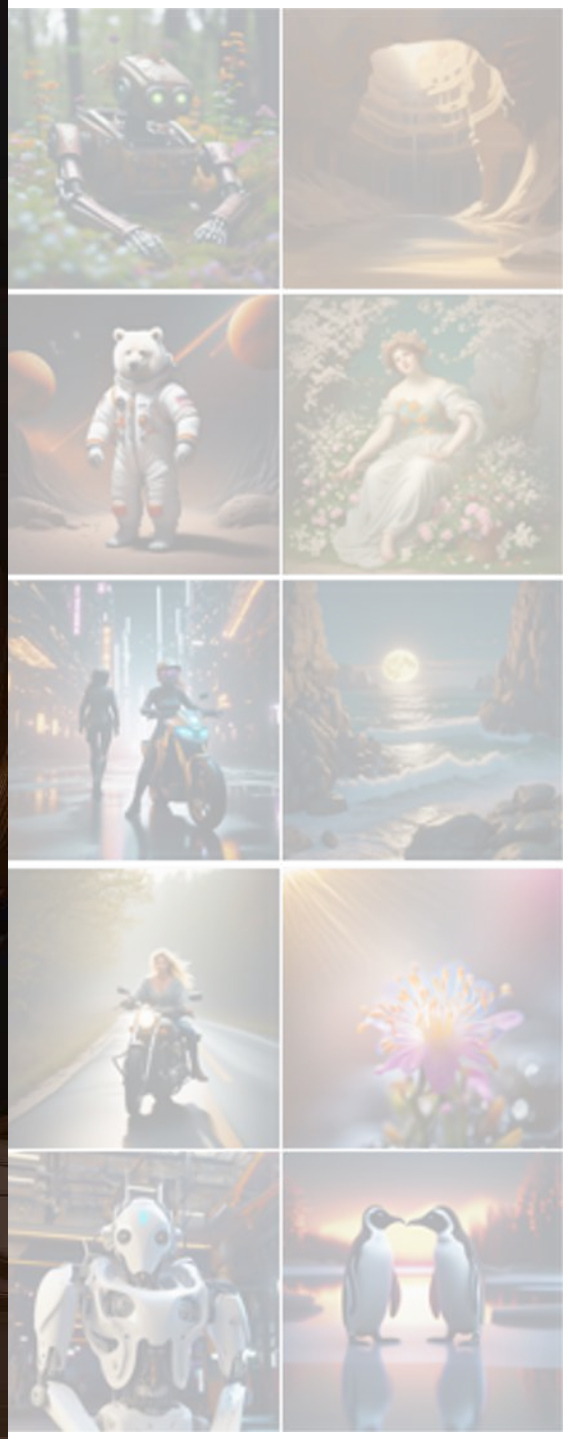
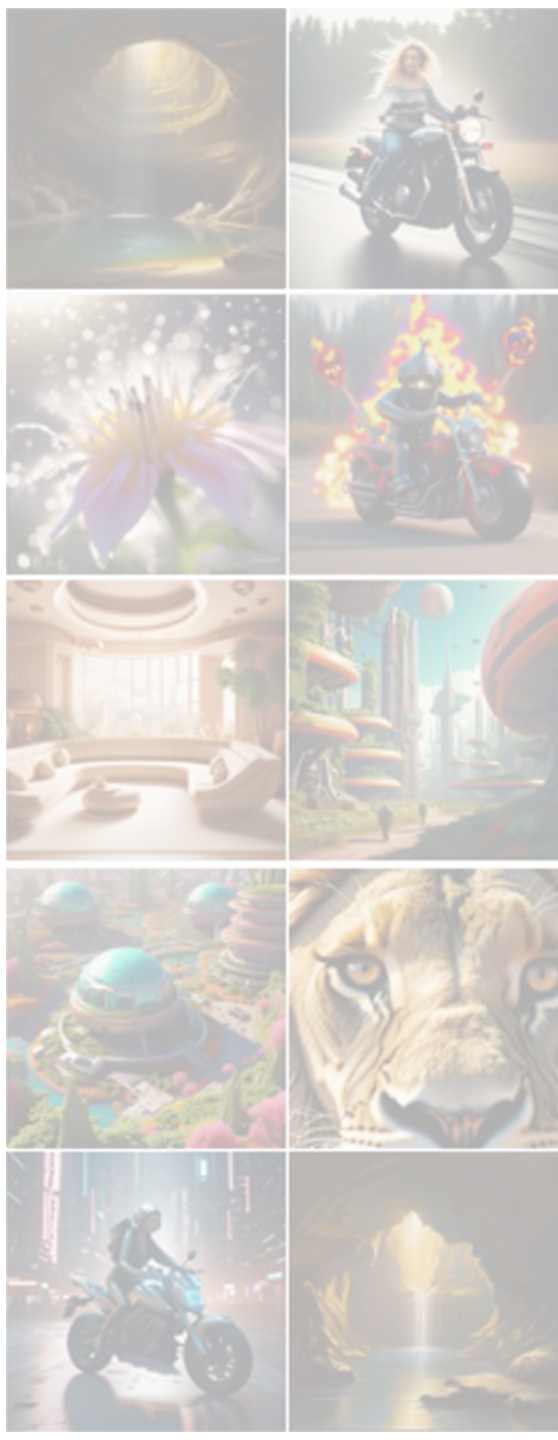






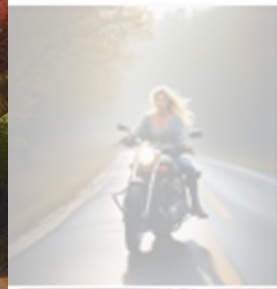
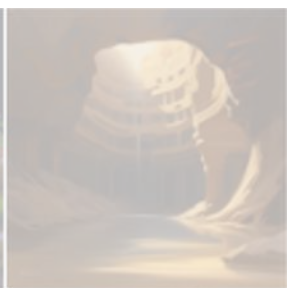
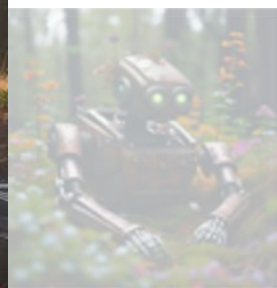
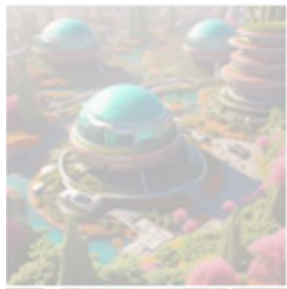
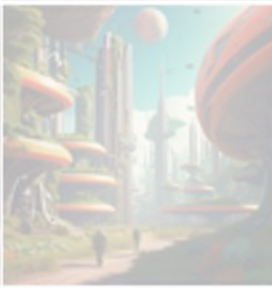
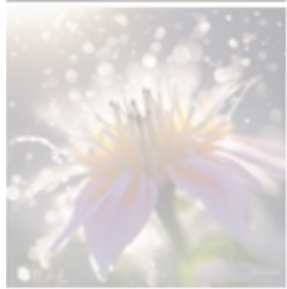
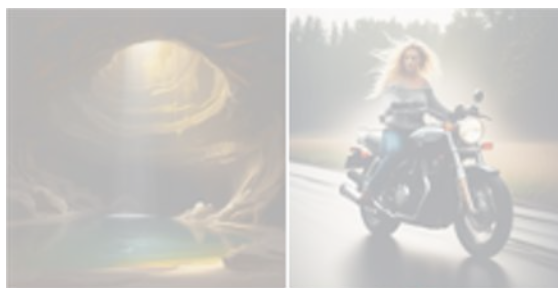


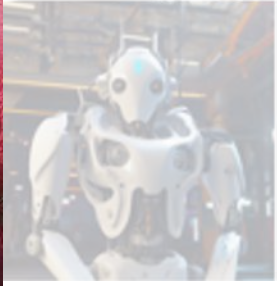
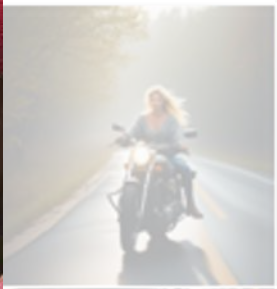
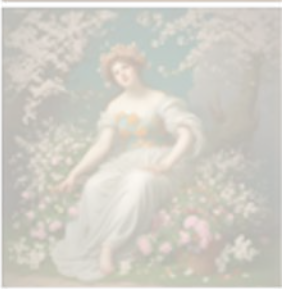
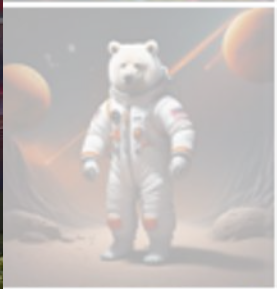
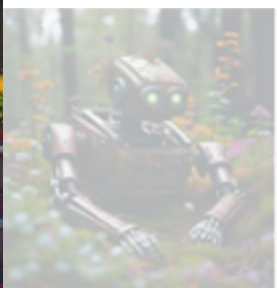
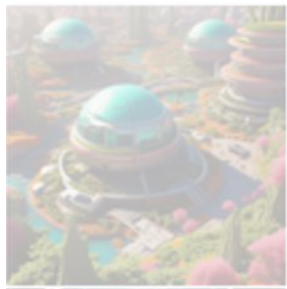
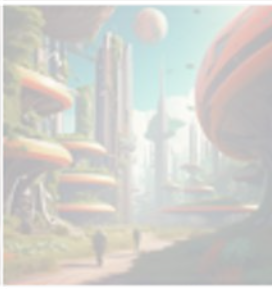
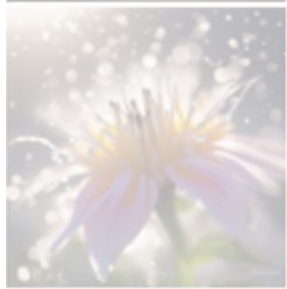
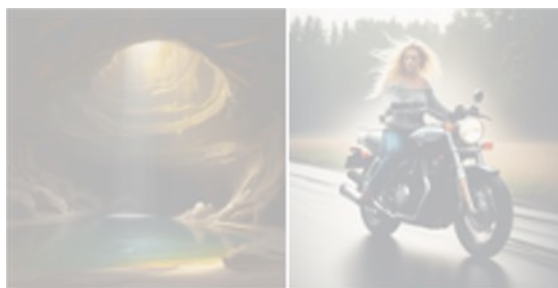


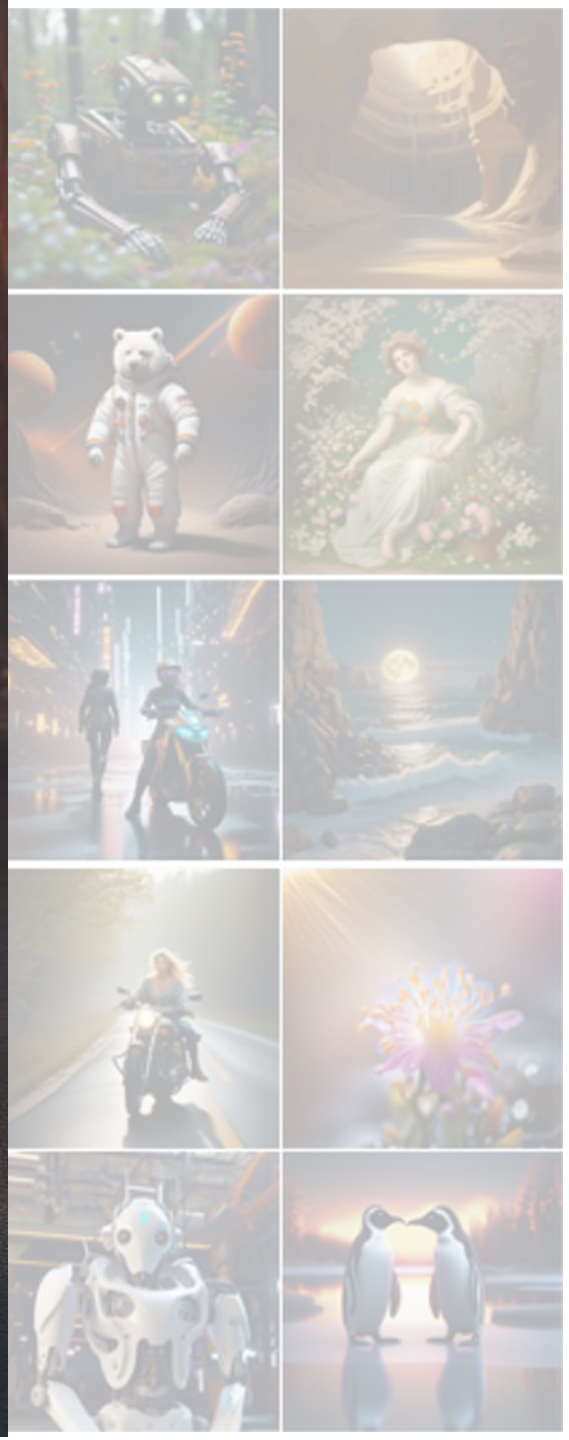
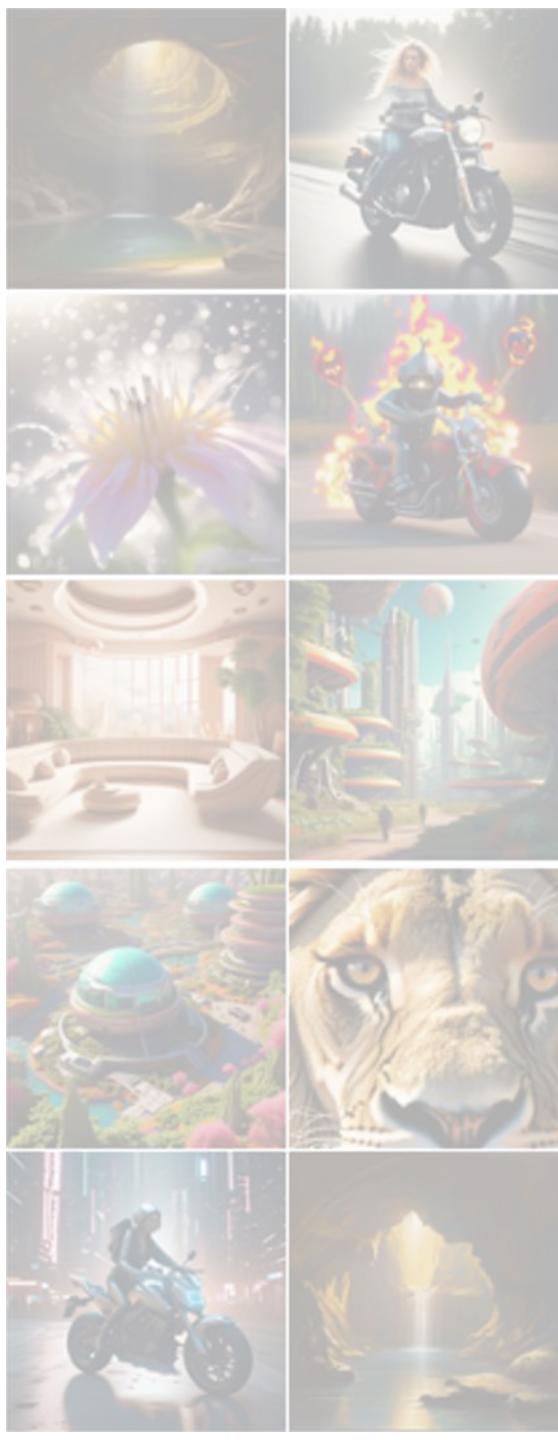


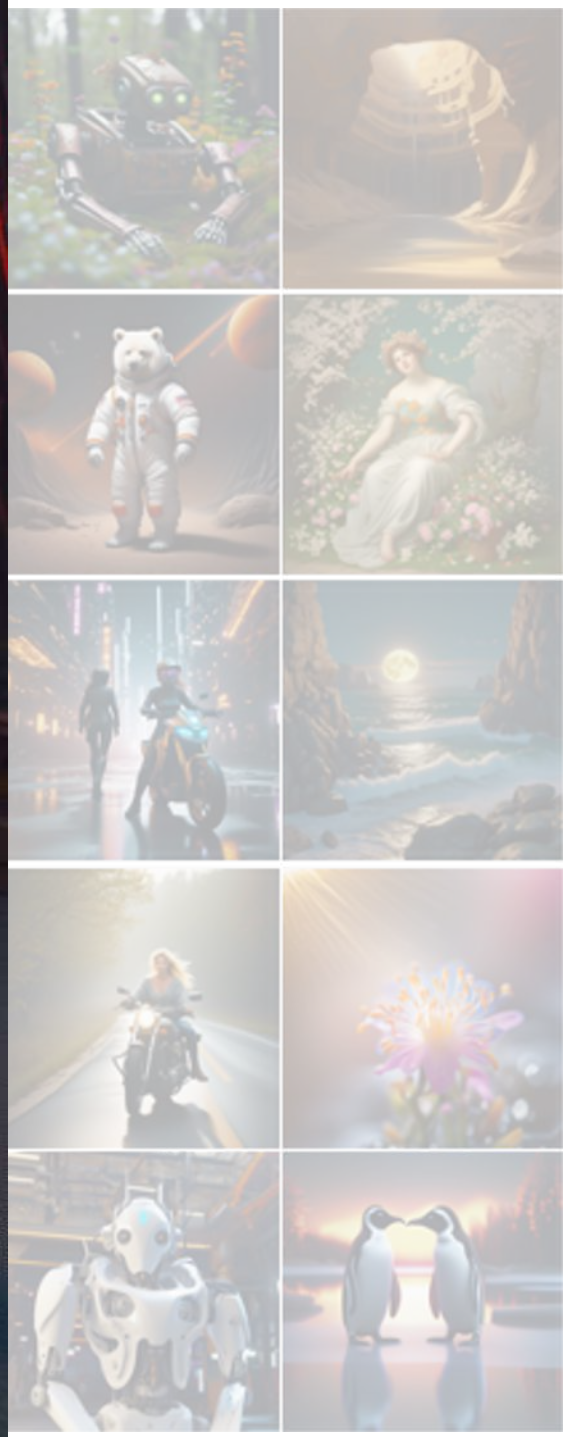
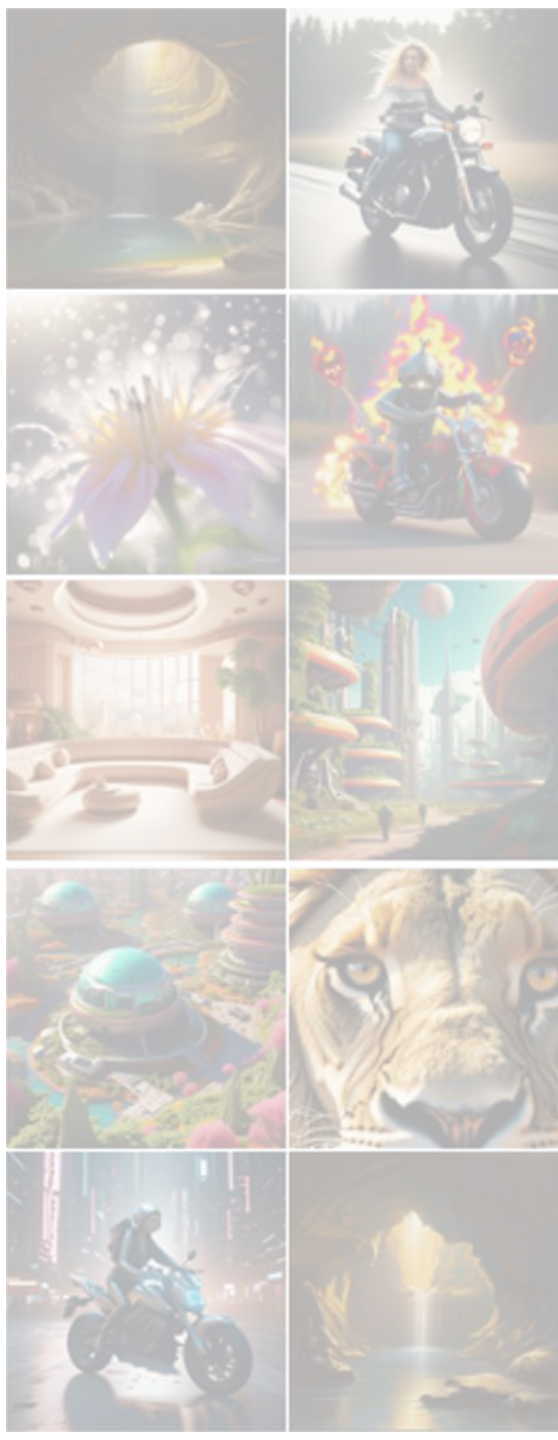
© 2023

ИЗРАИЛЬ









Community Contributions

Sebastian
@seb_cawai

Spent a few hours experimenting with FreeU and I'm very pleased with the results! It's remarkable how it boosts the detail levels of SDXL without any impact on process time. I'm definitely keeping this in my workflow! 🥰

github.com/ChenyangSi/FreeU



10:12 PM · Sep 24, 2023 · 18.3K Views

2 16 85 58



FreeU

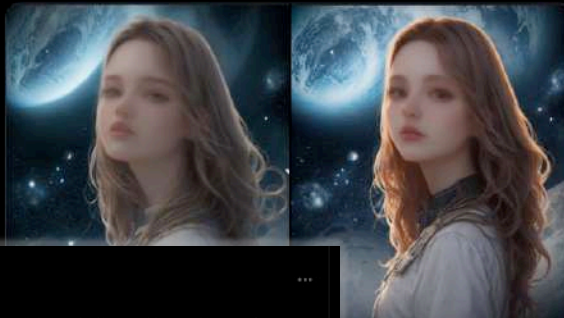
Peps
@Peps 61

exp 01) LCM, 4-steps, freeU (Y/N)

With proper hyperparameters, freeU gives better quality even with LCM.

seed=1024
"photo of a beautiful girl in the space, universe, earth in the background"
pipe.unet.enable_freeu(s1=0.2, s2=0.2, b1=0.8, b2=1.4)


#LCM #huggingface #diffusers



GM I've just uploaded the SD freeU ComfyUI workflow – give it a try and share your thoughts with me! Cheers! huggingface.co/bramvera/comfyui-stablediffusion-freeu #stablediffusion #comfyui #AIArtCommunity #aigirls #AIArtwork cc @scy894



11:17 AM · Oct 21, 2023 · 1,987 Views



11:55 PM · Sep 27, 2023 · 1,007 Views

2 3 8 1



Q&A

Poster Session

- Today 5 PM
- Arch 4A-E Poster #153
- Welcome any questions & discussions



S-LAB
FOR ADVANCED
INTELLIGENCE



NANYANG
TECHNOLOGICAL
UNIVERSITY
SINGAPORE

FreeU: Free Lunch in Diffusion U-Net



Chenyang Si Ziqi Huang Yuming Jiang Ziwei Liu
MMLab@NTU | S-Lab, Nanyang Technological University



Code



Project Page