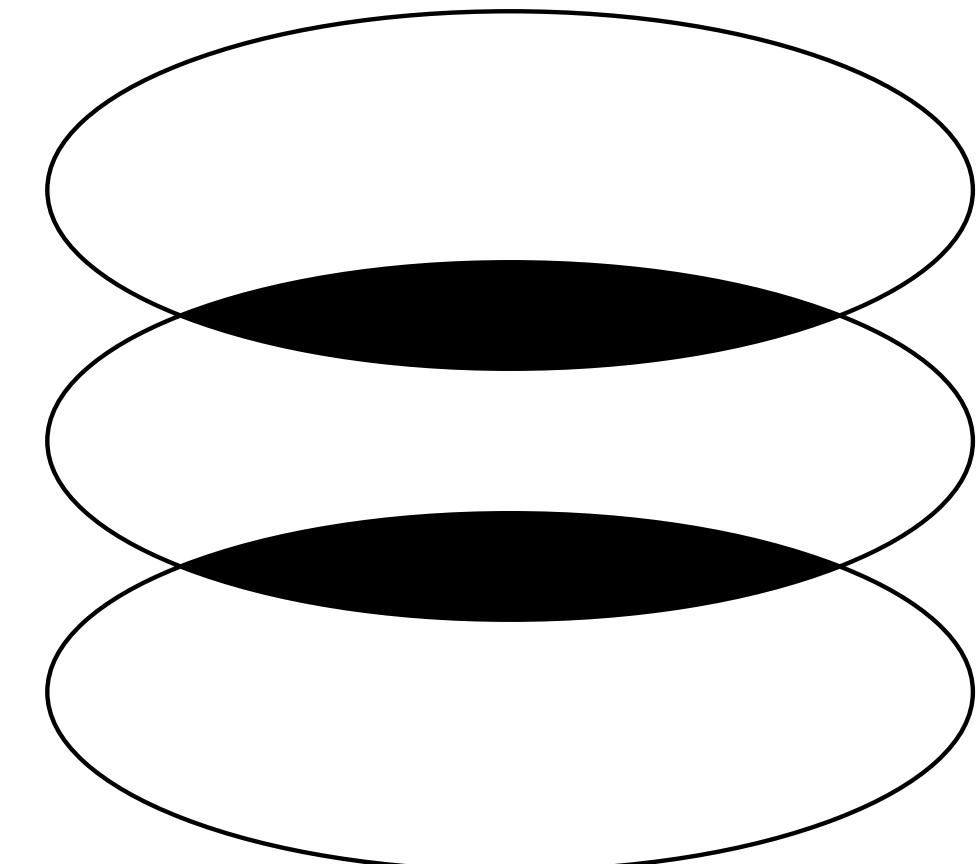


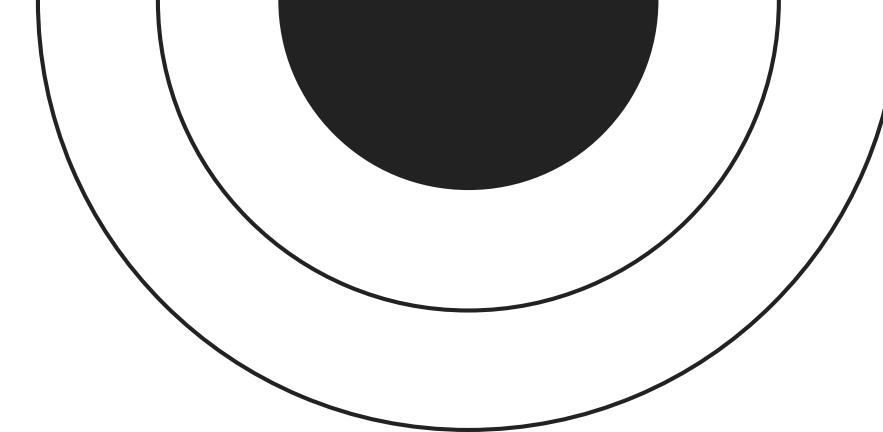
How to *generate* new versions of an original character?

An Application of LoRA and Dream Booth Fine-Tuning

Milestone #3

Maëlys Boudier, Natalia Beltrán,
Arianna Michelangelo





Goal

Conclusion of Feedback : work towards improving the model to increase quality ratio

Problem Overview

Training Data

Implement LoRA only code

Compare Outputs per Model

Compare Model Advantages

Tuning Best Model

Next Steps

Problem Overview

How to generate new versions of an original character?

Situation: data scarcity, if an artist only has 6 images of a character they invented.

Generate: Image generation through **fine-tuning** by ‘teaching’ concept (the character) to pre-trained model, get a new image from a text prompt naming the character and a position.

New Versions: the same character (it looks the same) but in **new positions not in train data**, i.e., lifting an arm, closing eyes.

Importance: Limit use of large models because of **limited computational resources** but also to limit CO₂ emissions in the environment.

Original Character: only **one character** in our paper, but in the future this could be integrated alongside other characters, or backgrounds.



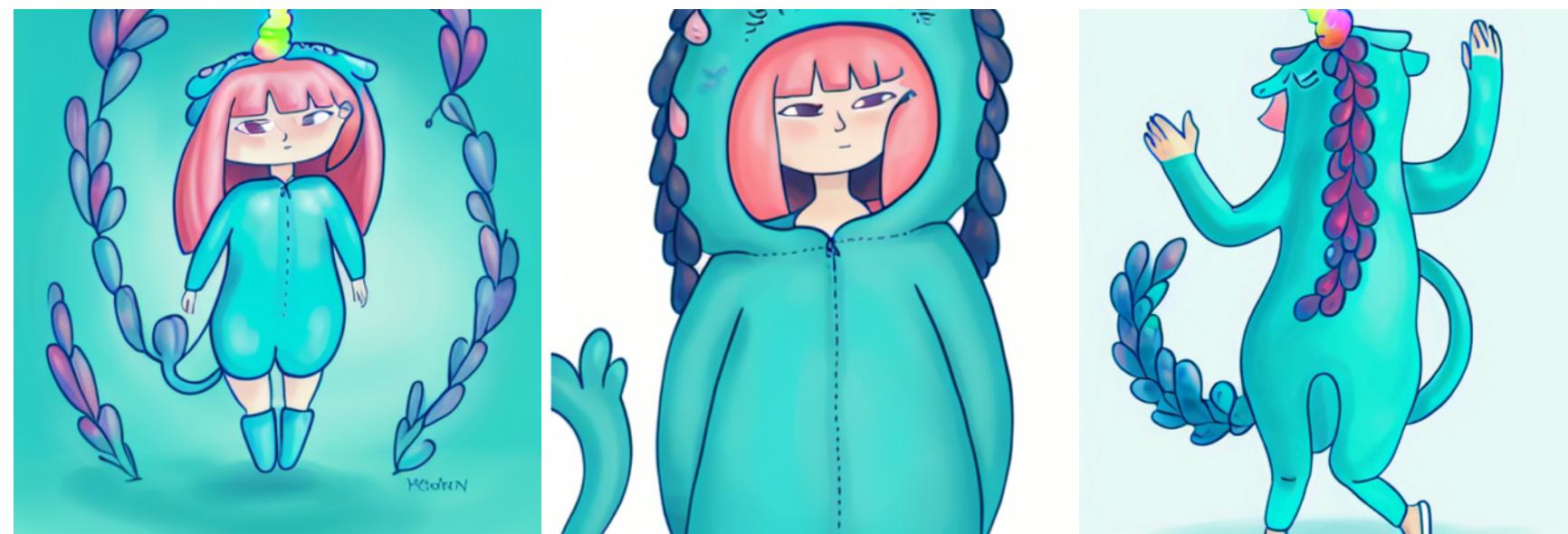
Training Data



Designed by Jordane Meignaud

Results: Dream Booth

Dream Booth to retrain full Stable Diffusion (v1.5) Gradient to learn our concept



Training:

1000 Steps

Inference:

50 Steps

Concept: "A picture of UnicornGirl"

Analysis: Dream Booth

- Long to run ~ 2h
- Heavy files to store: ~4 GB
- Quality is low, uncanny valley:
 - extra unicorn hair
 - some wrong colors
 - different hair in some versions



Results: LoRA

LoRA to retrain full Stable Diffusion (vXL) Gradient to learn our concept



Training:

200 Steps

Inference:

50 Steps

Concept: “A picture of UnicornGirl”

Analysis: LoRA

- too long to run (over 12 hours) so runtime error on Kaggle
- Extra light files to store: ~ 3.23 MB
- Additional Step: manual labeling of images
- Quality is lowest, uncanny valley:
 - extra horns
 - extra tails
 - wrong colors
 - wrong style
 - human figures



Results: DreamBooth + LoRA

Mix of Dream Booth and Lora applied to Stable Diffusion (vXL) Gradient to learn our concept



Training:

1000 Steps

Inference:

50 Steps

Concept: "A picture of UnicornGirl"

Analysis: Dream Booth + LoRA

- Fastest to run (under 2 hours)
- Light files to store: ~25 MB
- Quality is high:
 - sometimes extra features or some imperfections but overall very similar

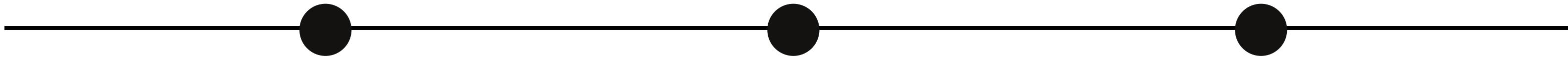


Comparison

| Model Name | Weight Size | Run Time | Visual Quality |
|-------------|-------------|-----------|----------------|
| Dream Booth | 4 GB | 2 hours | low |
| LoRA | 3.23 MB | 12 hour + | low |
| Both | 25 MB | 1 hour | high |

Comparison

| Model Name | Weight Size | Run Time | Visual Quality |
|-------------|-------------|-----------|----------------|
| Dream Booth | 4 GB | 2 hours | low |
| LoRA | 3.23 MB | 12 hour + | low |
| Both | 25 MB | 1 hour | high |



STEPS

Train Steps:

- 100 • 1100
- 200 • 1200
- 300 • 1300
- 400 • 1400
- 500 • 1500
- 600 • 1600
- 700 • 1700
- 800 • 1800
- 900 • 1900
- 1000 • 2000

LR

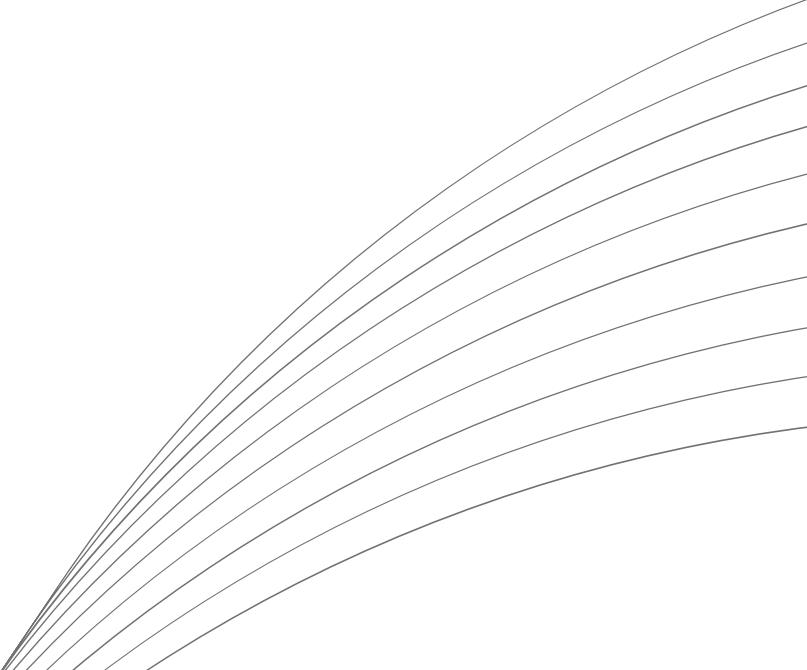
Learning Rate:

- 1e-1
- 1e-2
- 1e-3
- **1e-4**
- 5e-5
- 1e-5
- 1e-6

INF STEPS

Inference Steps:

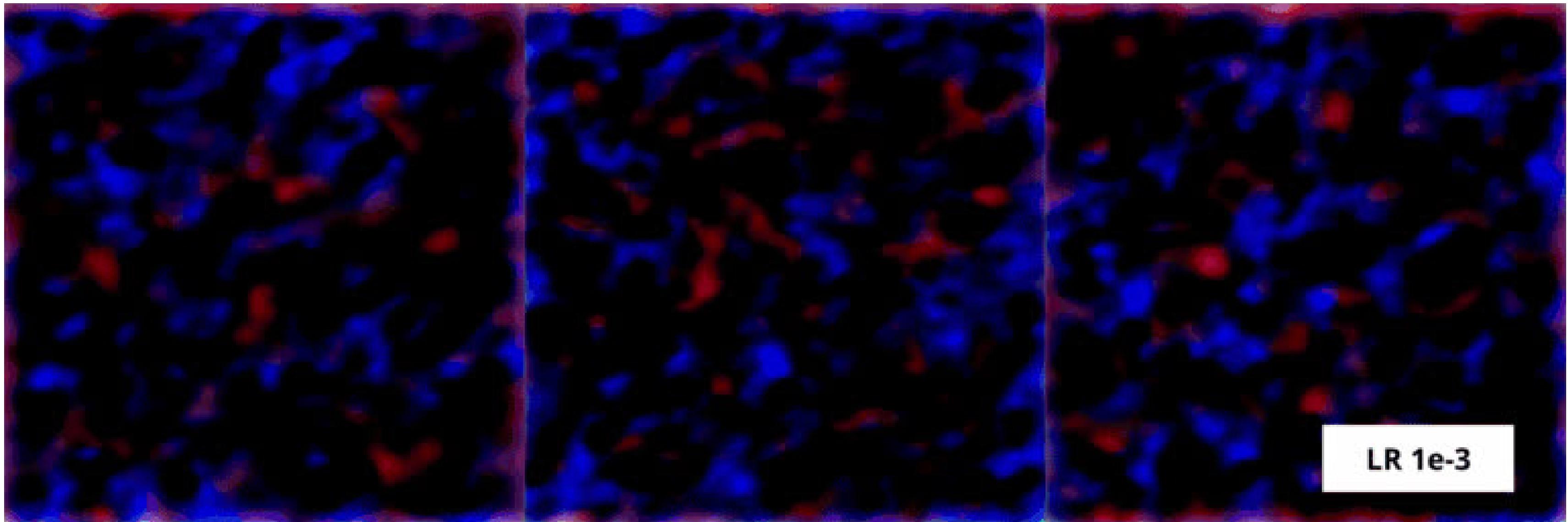
- 10
- 20
- 30
- 40
- **50**
- **60**
- **70**
- **80**
- **90**
- **100**



Tuning: Train Steps

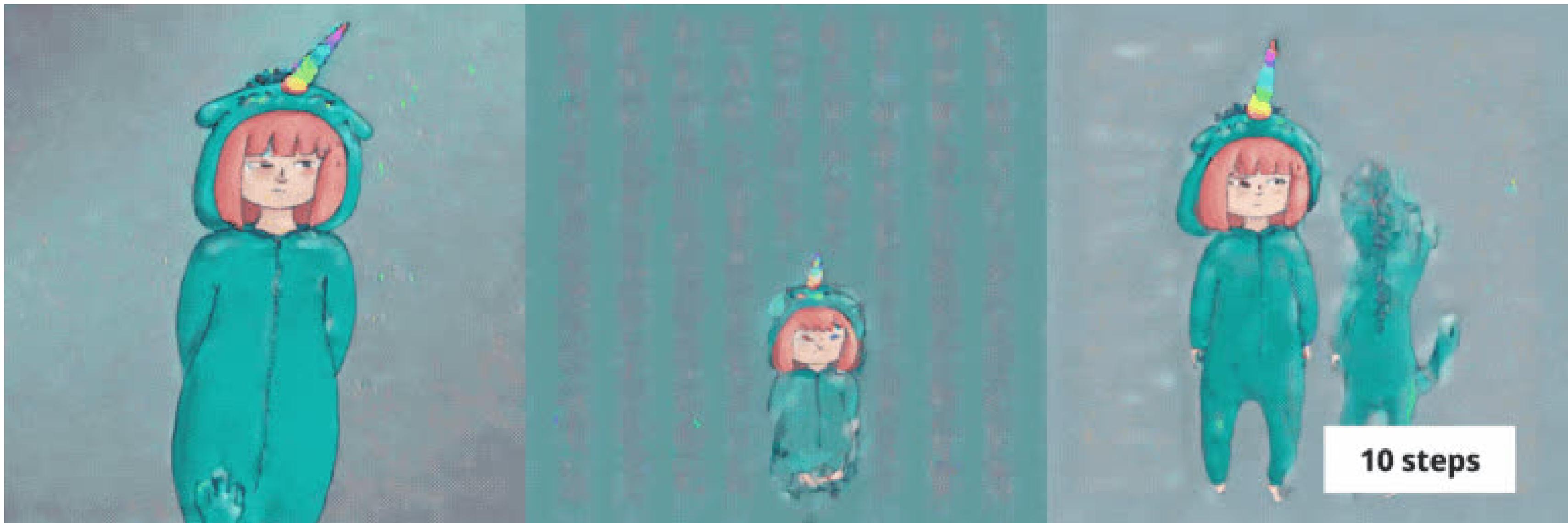


Tuning: Learning Rate



LR 1e-3

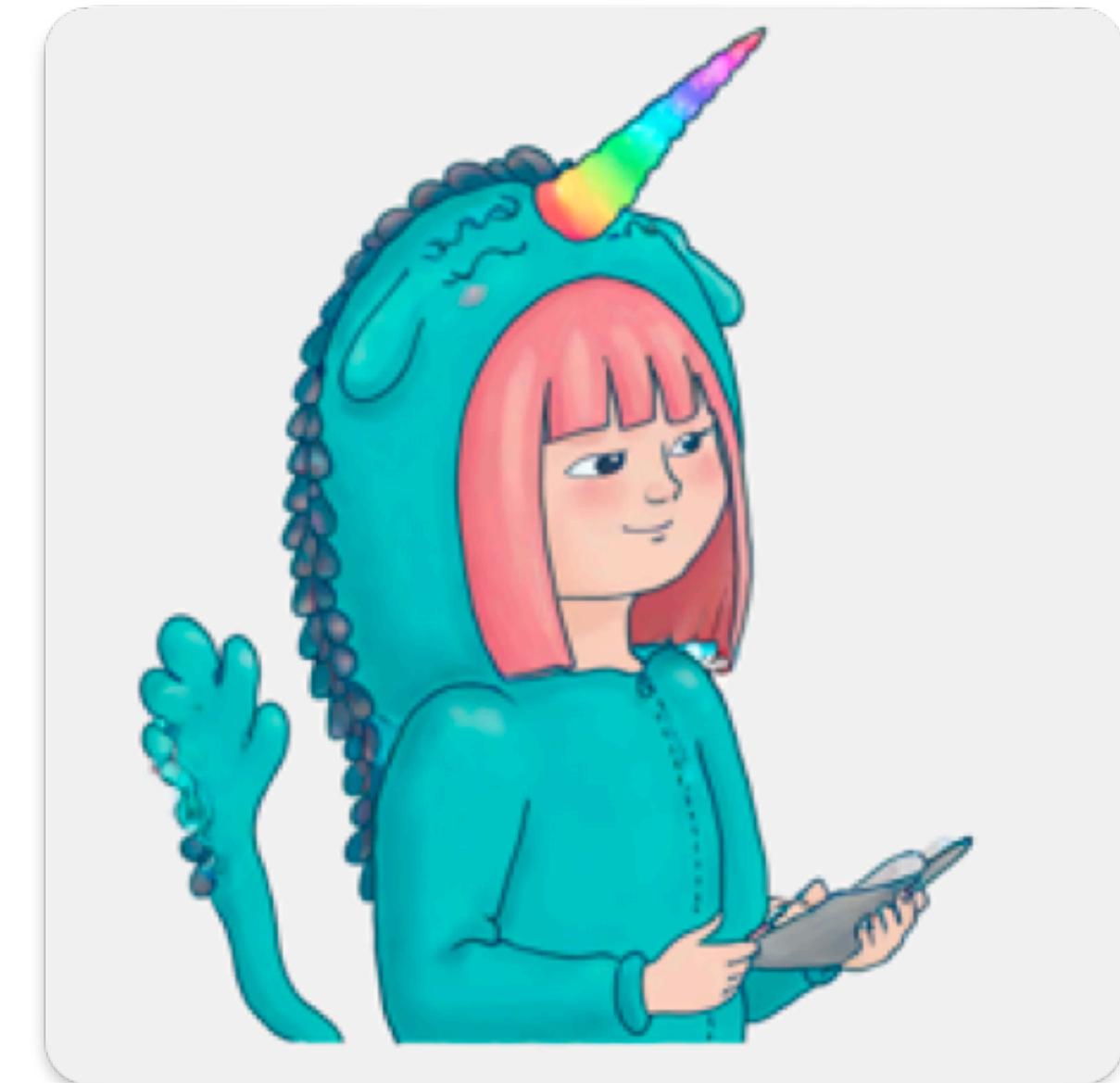
Tuning: Inference Steps



Next Steps

2 weeks until submission:

- Attempt to get the loss per step plots
- Generate a story prototype
- Write paper (30 pages)



Appendix

Model Overview:

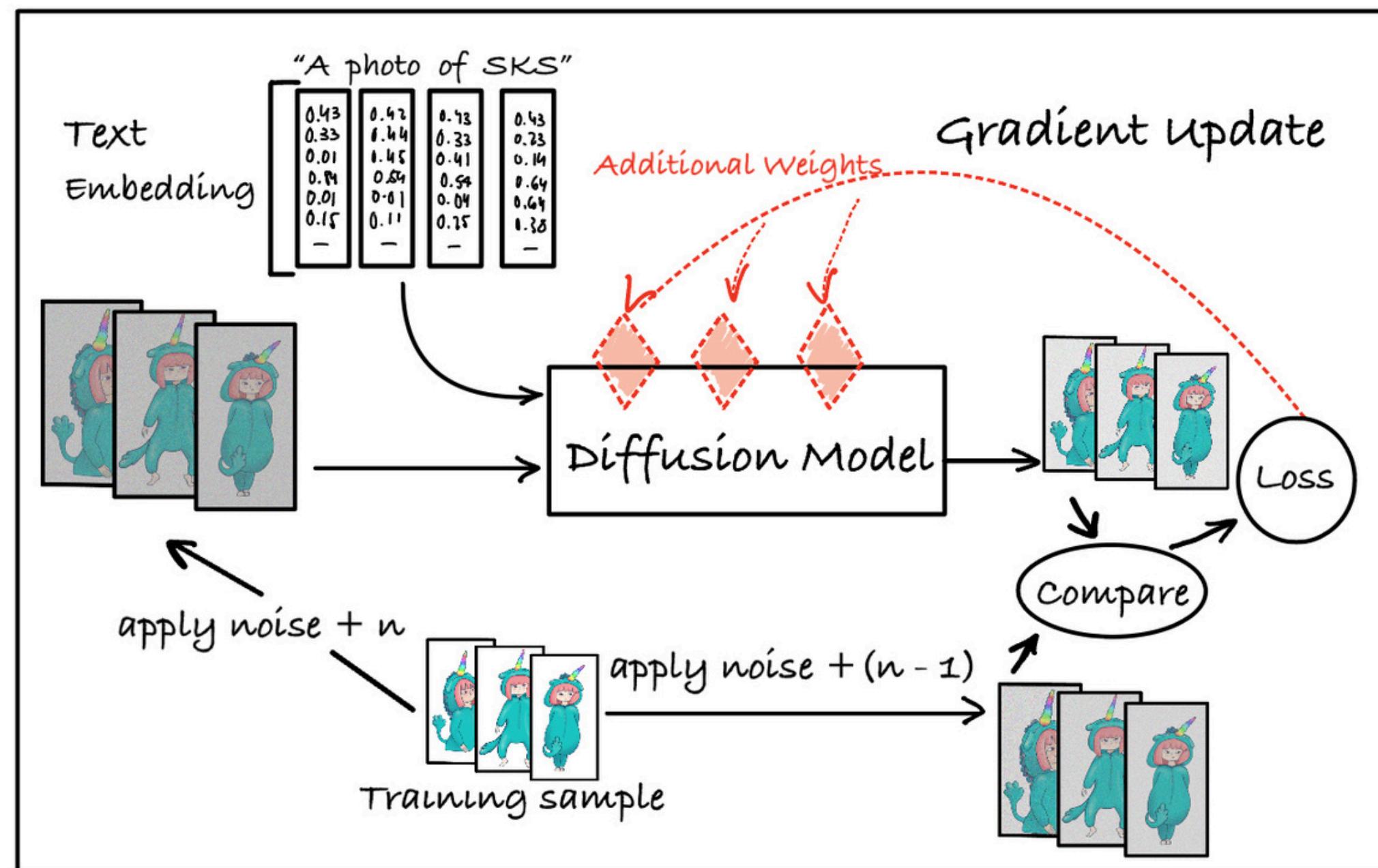
- LoRA
- Dream Booth
- GAN Baseline
- Diffusion Baseline



The End

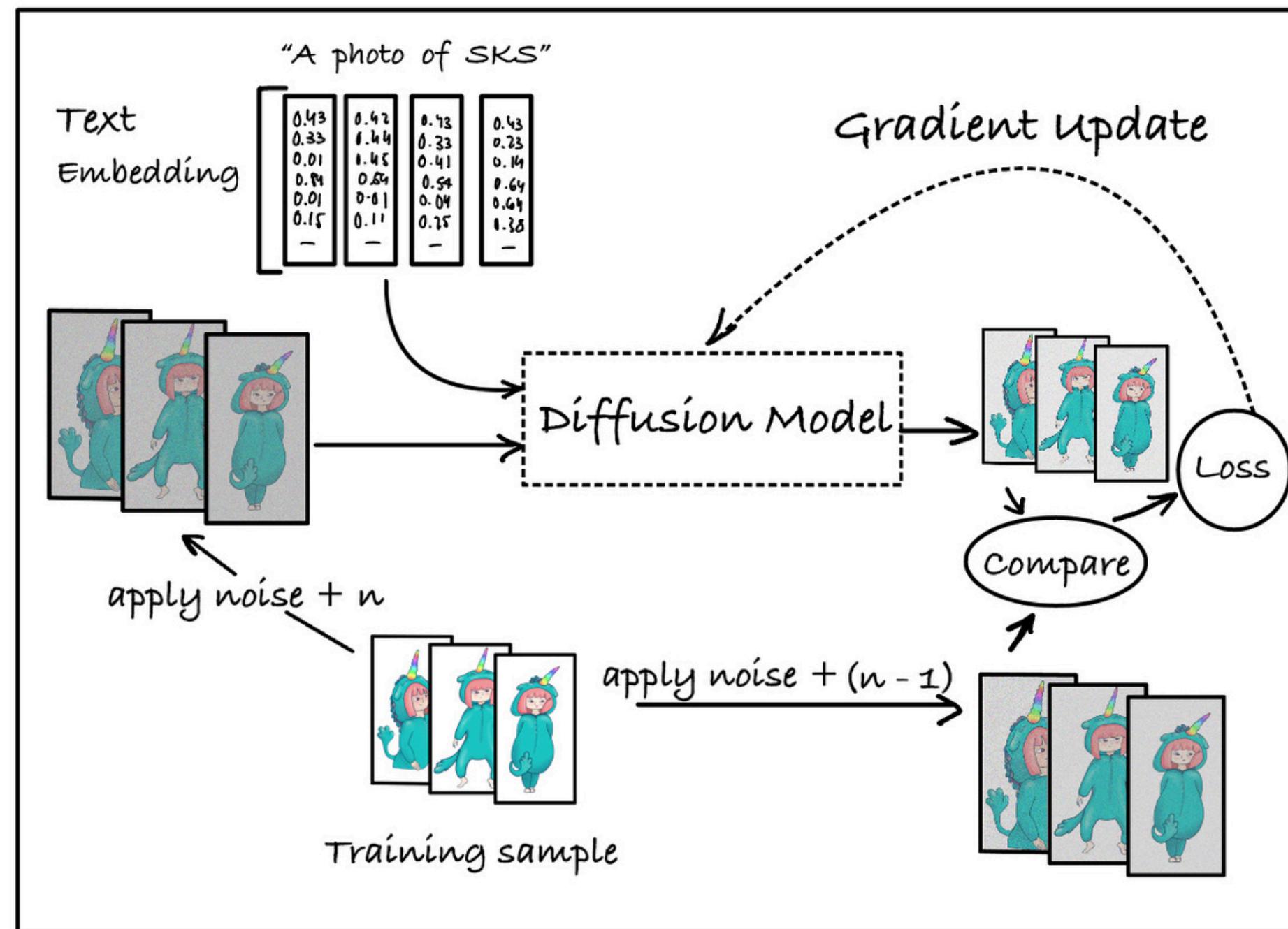
LoRA

Objective: Teach the model a new concept without creating a whole new model. Add a number of layers to the diffusion model and then train the model. The weights (of the new layers) are modified as the model understands the concept.

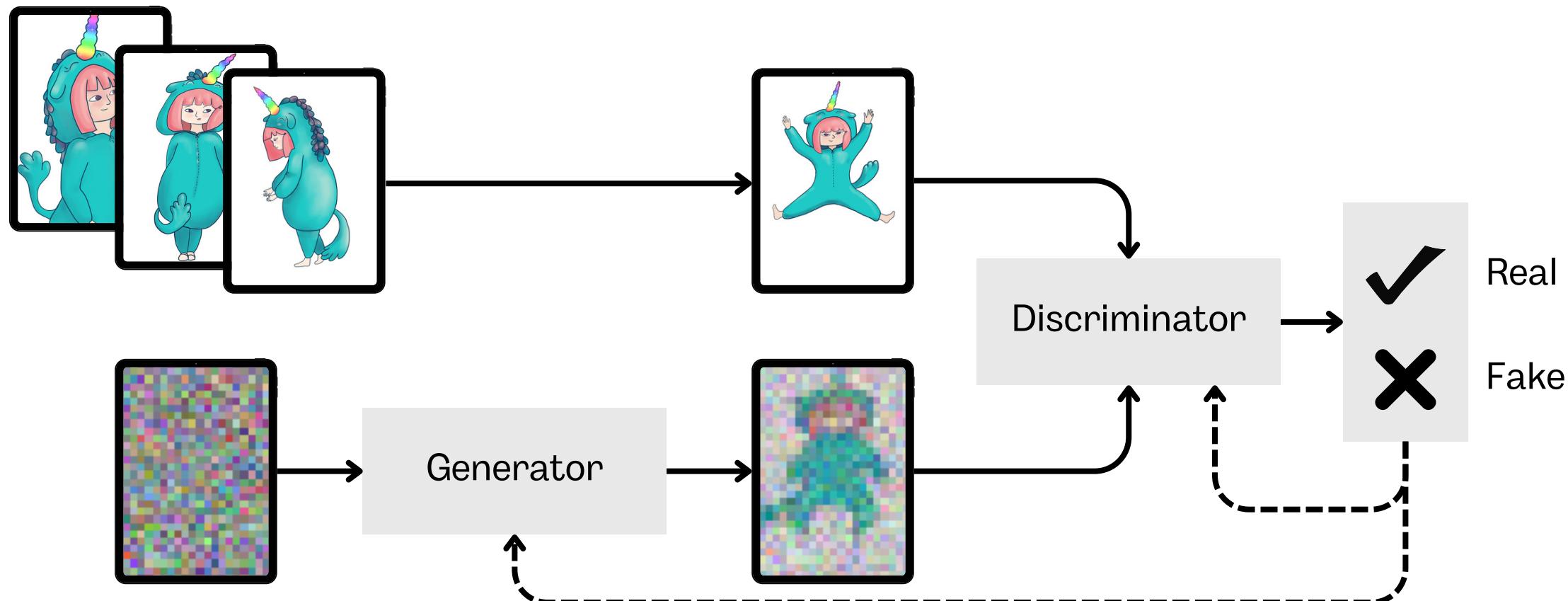


Dream Booth

Objective: Teach the model to associate a unique identifier to a specific concept. Fine-tunes the diffusion model itself until it understands the new concepts.



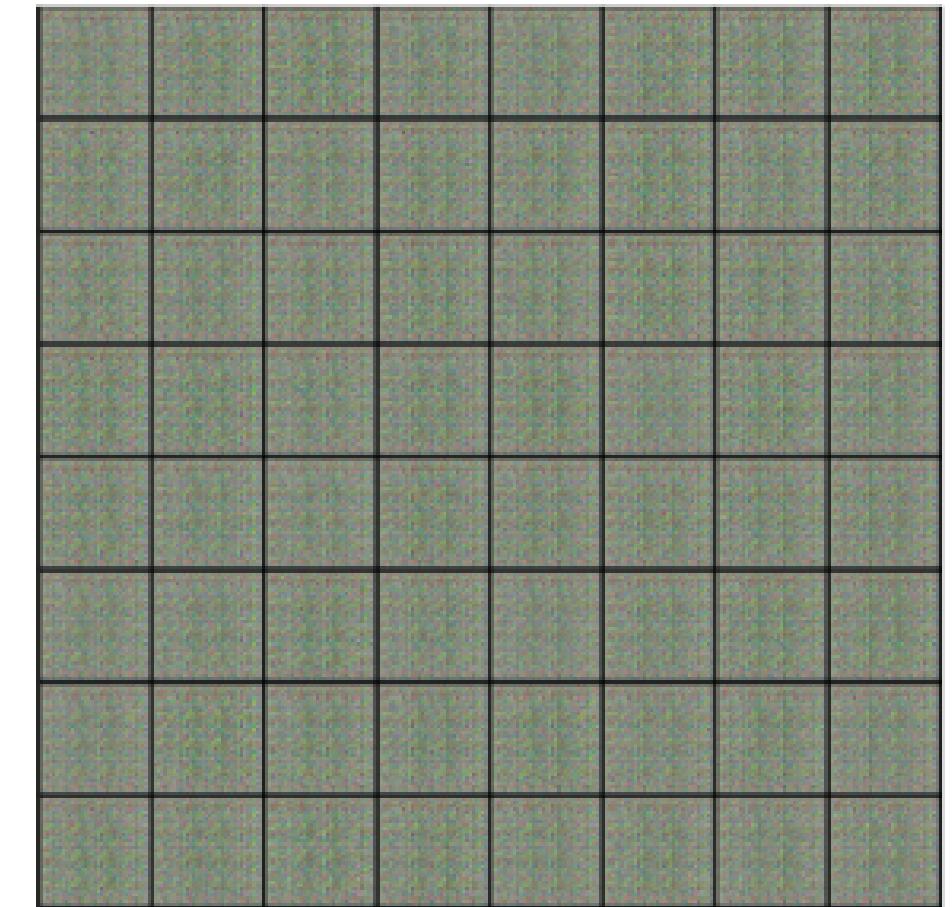
GAN Baseline



Conclusion:

- GANs work well with very high amount of data (100k+)
- unable to perform well on the generator side from 6 images

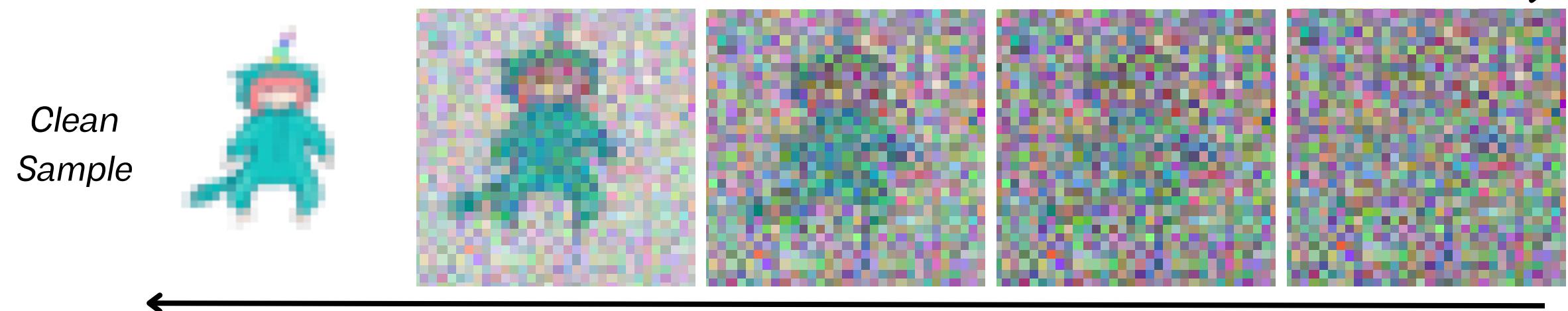
Results:



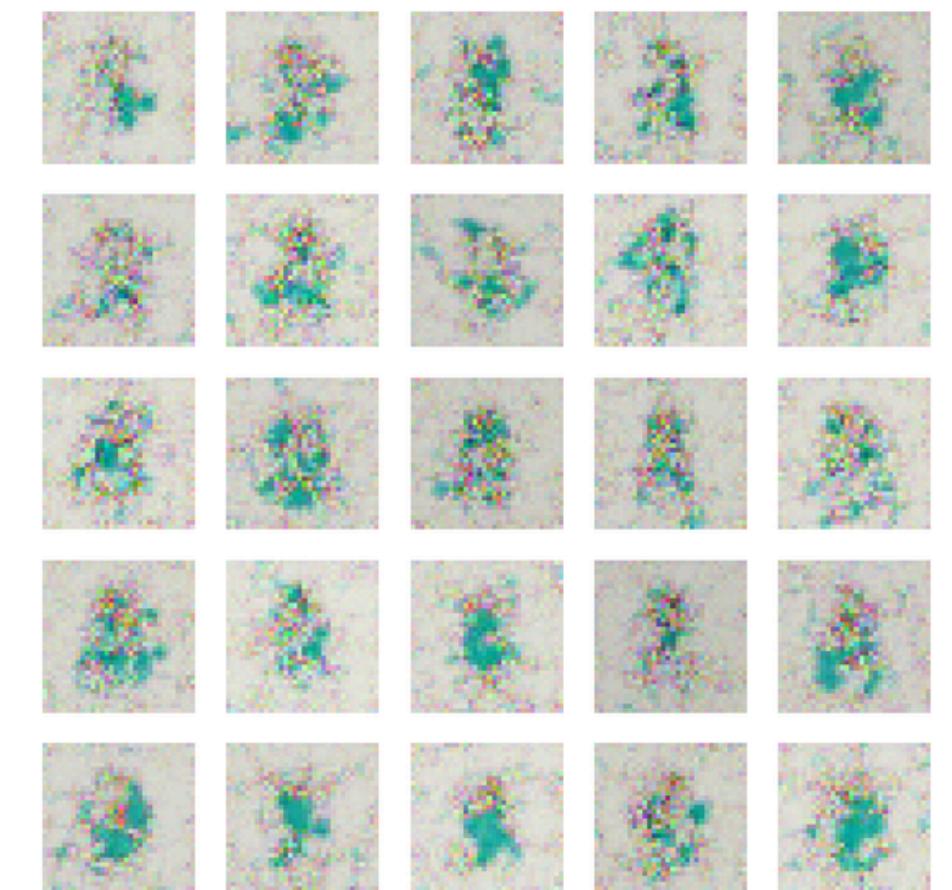
Diffusion Baseline

Forward / Noising Process

Sample data -> turn into noise



Results:



Reverse / Denoising Process

Sample noise -> turn into data

Conclusion:

- Learns colors better and faster than GAN
- Needs more data (Stable Diffusion trained with 2.3 Billion images)