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

An Application of LoRA and Dream Booth Fine-Tuning

Milestone #2

Maëlys Boudier, Natalia Beltrán,
Arianna Michelangelo

Overview

Problem Overview

Data Overview

GAN Baseline

Stable Diffusion Baseline

Prelim Results: Dream Booth

Prelim Results: Dream Booth + LoRA

Next Steps

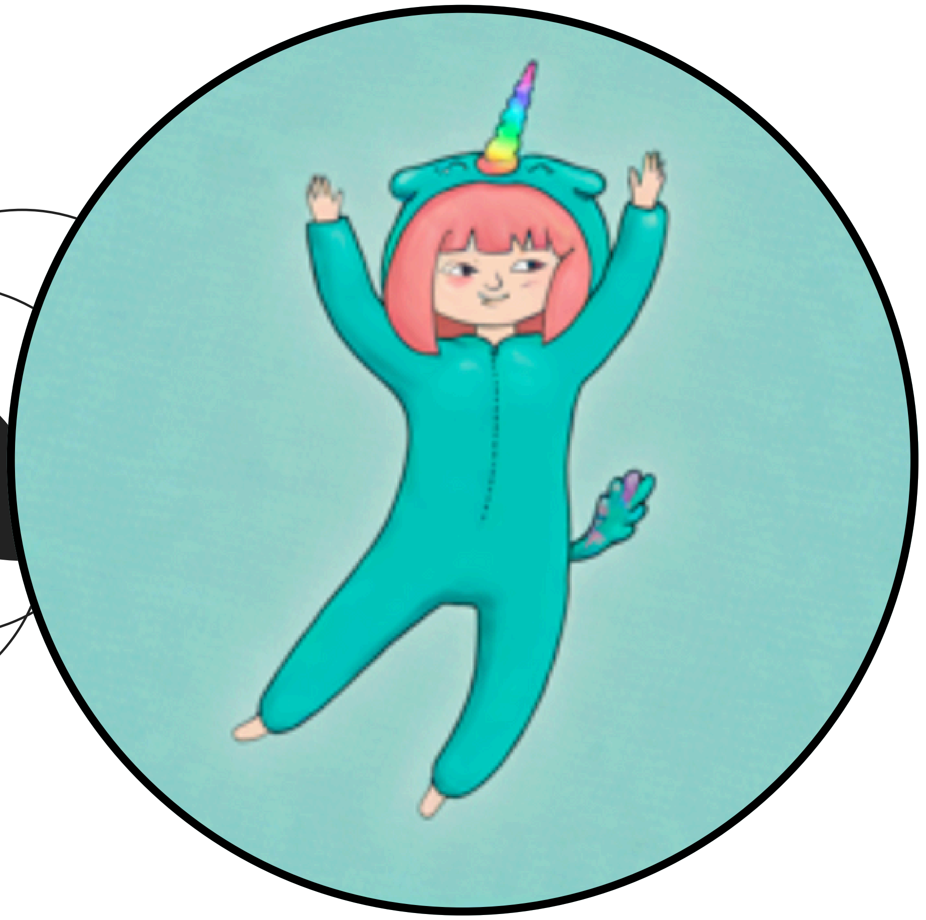
Problem Overview

Project Goal:

Given limited images of a character as input, train model to generate new positions of this same character based on text prompts. Tackle data scarcity with model fine-tuning.

Project Importance:

Provide a tool for artists in the industry to save time and resources in initial mock-up designs. Favor models limiting CO2 emissions for the environment.



In: 6 Images

Model

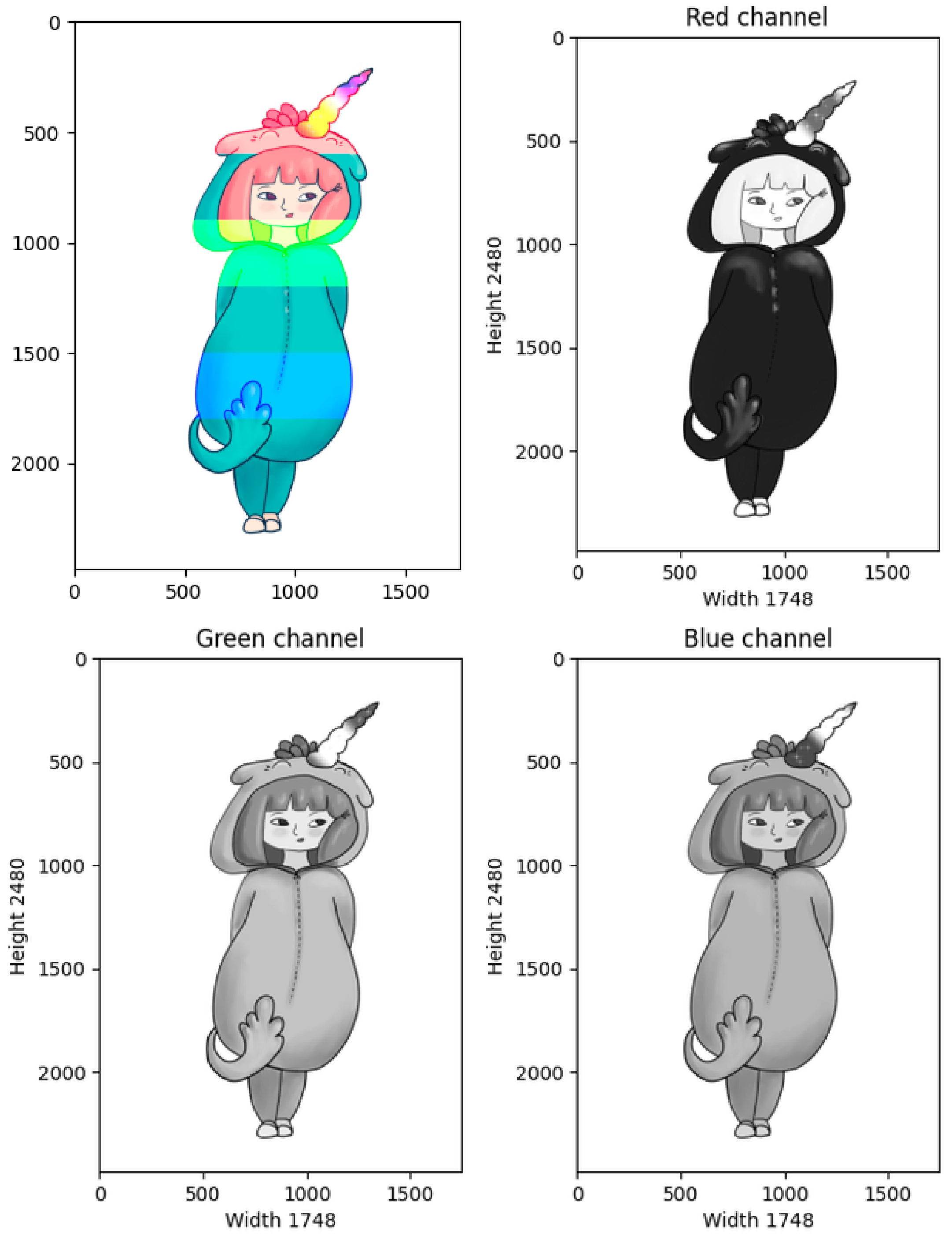
Out: New Positions

Original, never-seen-before images of a character

Fine-tune Stable Diffusion with LoRA / Dream Booth

Same character in new positions not in train data

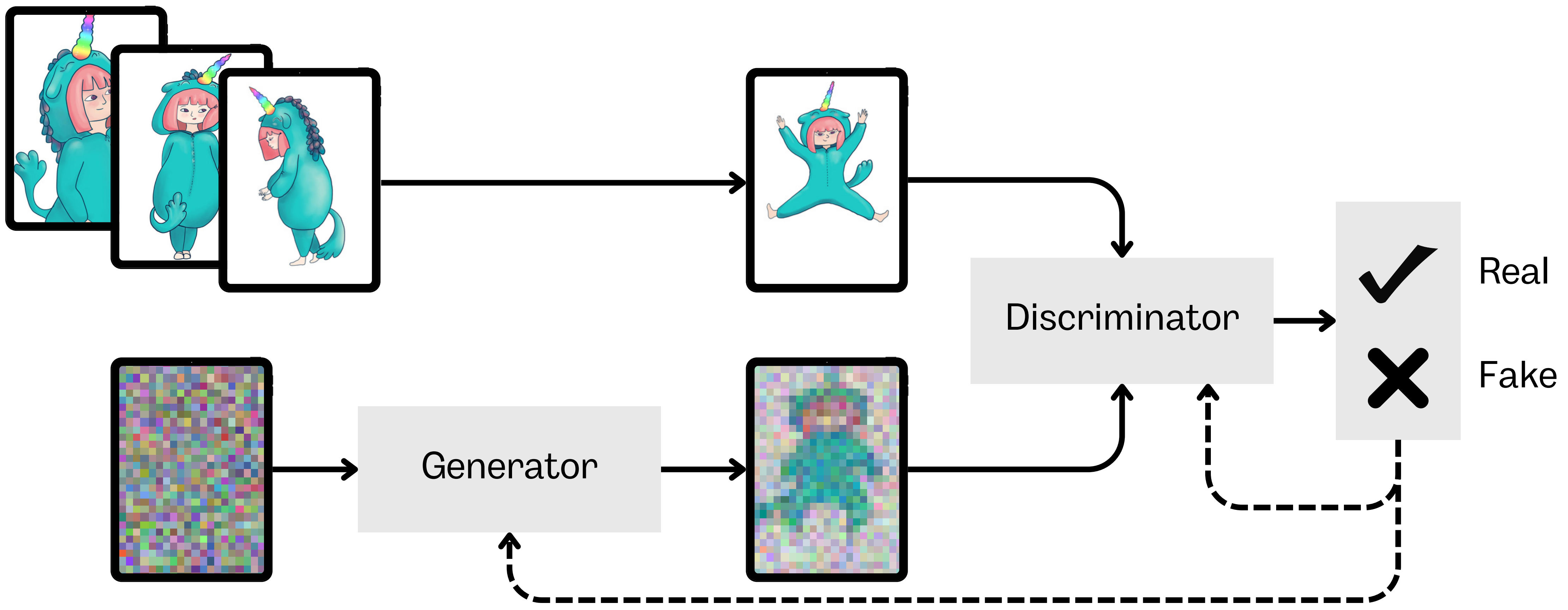
Original Data



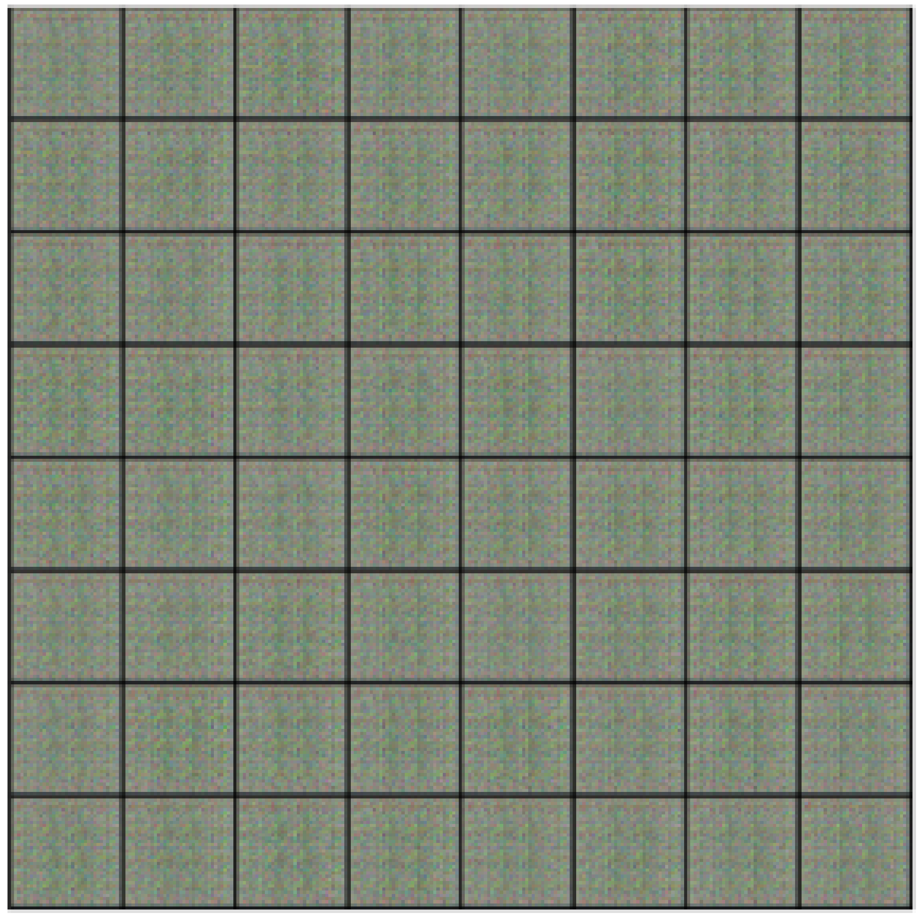
Designed by Jordane Meignaud

	width	height	size (bytes)	mean red	mean green	mean blue
min	1748	2480	407787	198	228	229
mean	1748	2480	484481	210	235	235
max	1748	2480	586804	219	240	240

GAN Baseline



Results:



Conclusion:

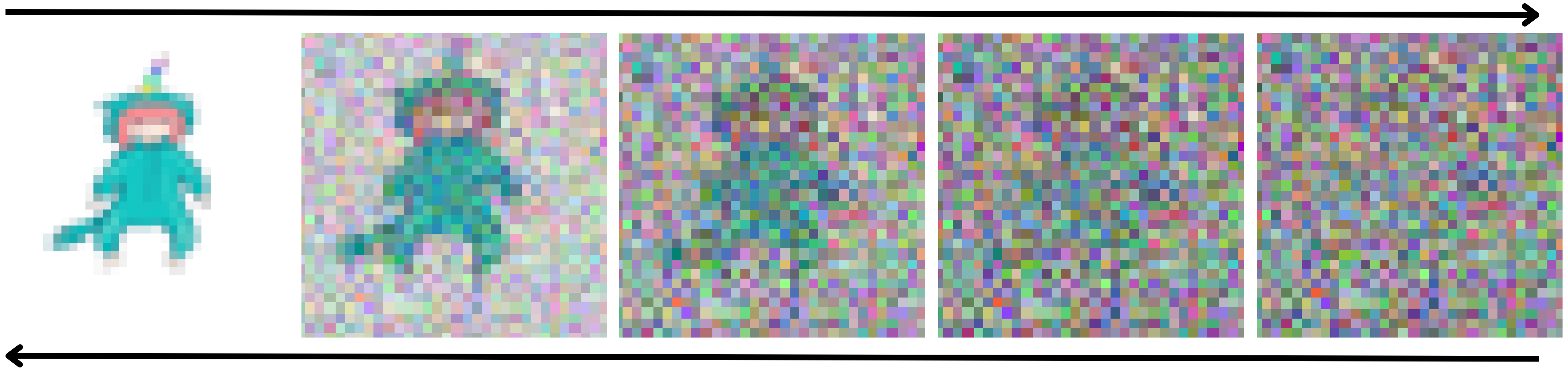
- GANs work well with very high amount of data (100k+)
- unable to perform well on the generator side from 6 images

Diffusion Baseline

Forward / Noising Process

Sample data -> turn into noise

Clean Sample

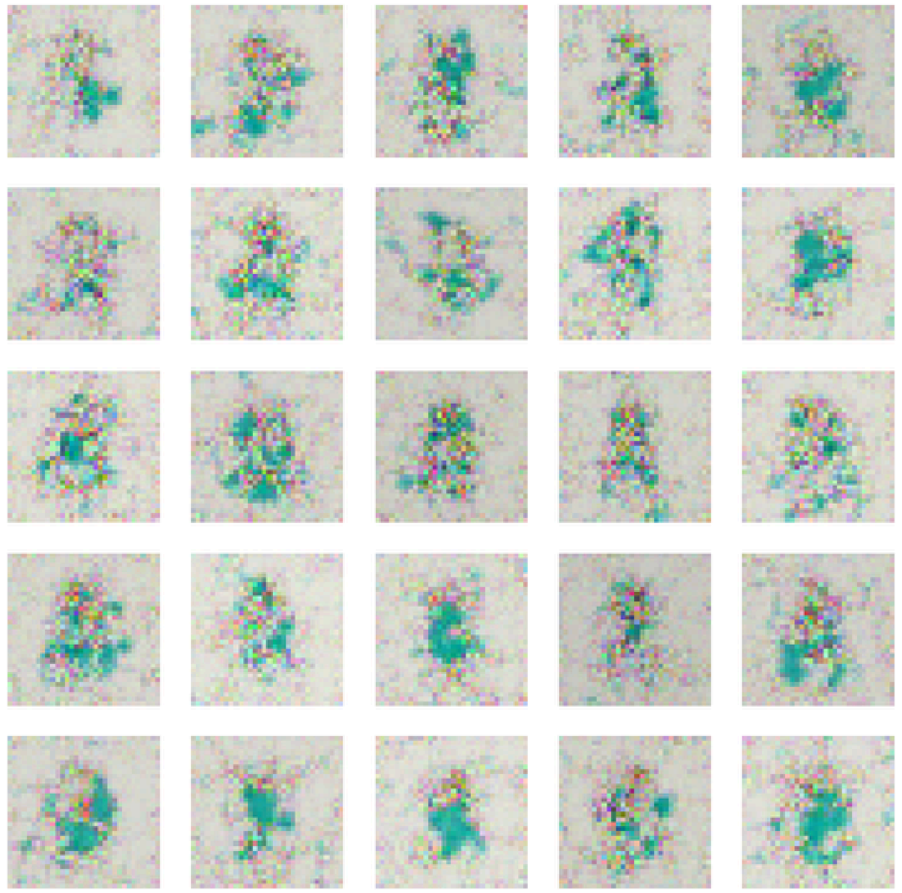


Pure Noise

Reverse / Denoising Process

Sample noise -> turn into data

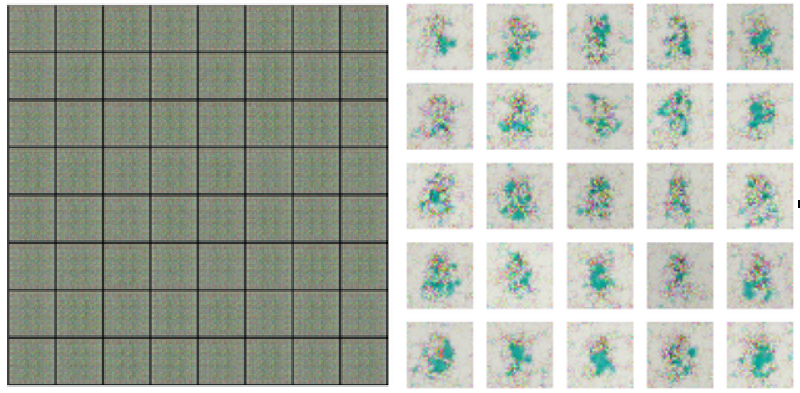
Results:



Conclusion:

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

Baseline

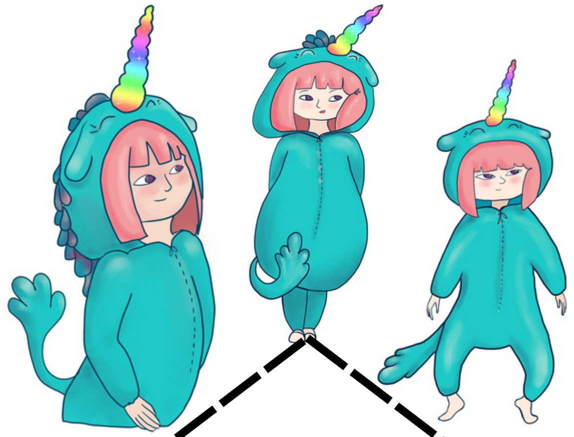


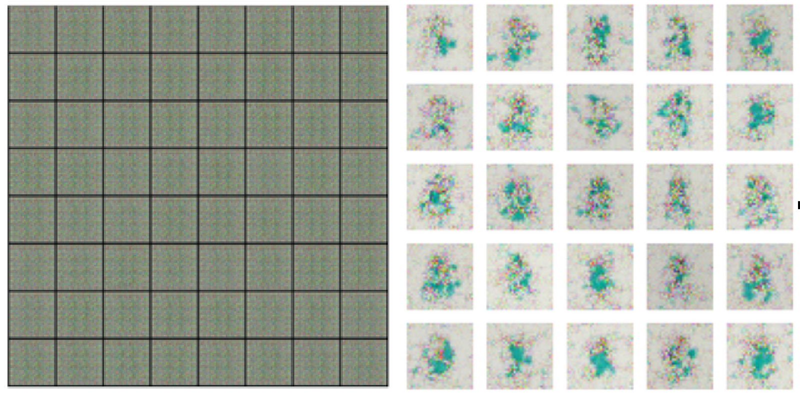
Full GAN / Diffusion

Open Source SD (vXL) Model

Baseline

Fine Tuned Model





Full GAN / Diffusion

Open Source SD (vXL) Model

Results: *Dream Booth*

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

Training:

600 Steps

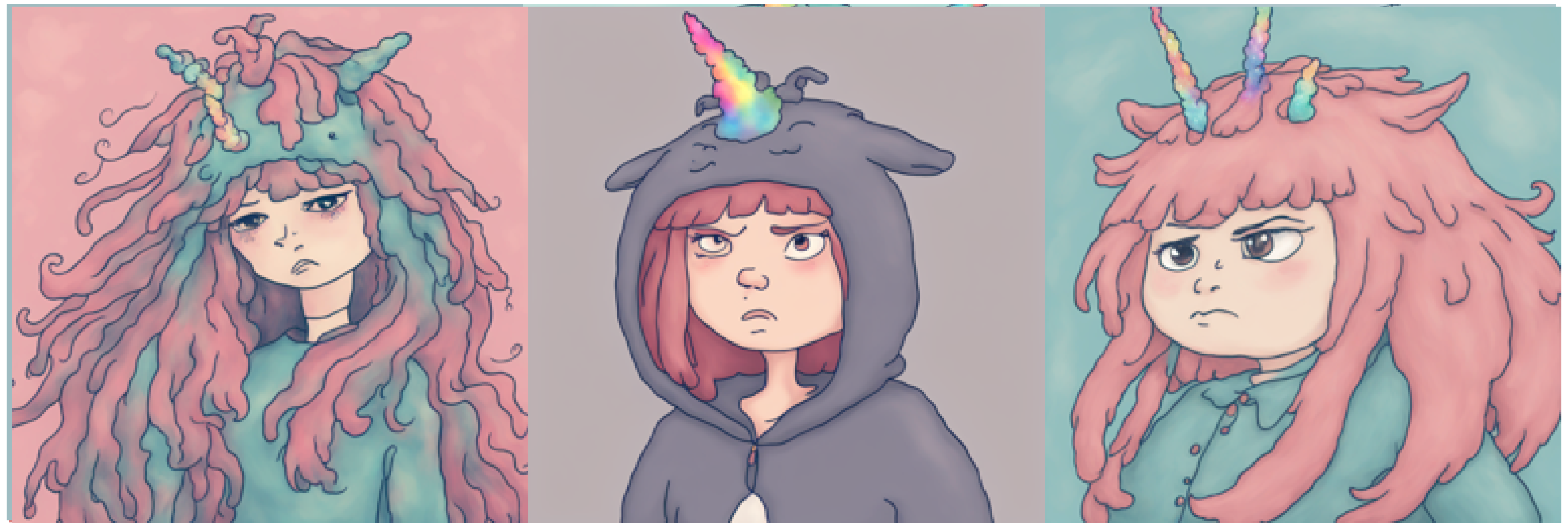
Inference:

150 Steps

Concept: "A picture of UnicornGirl"

Results: Dream Booth + LoRA

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



Training: 100 Steps

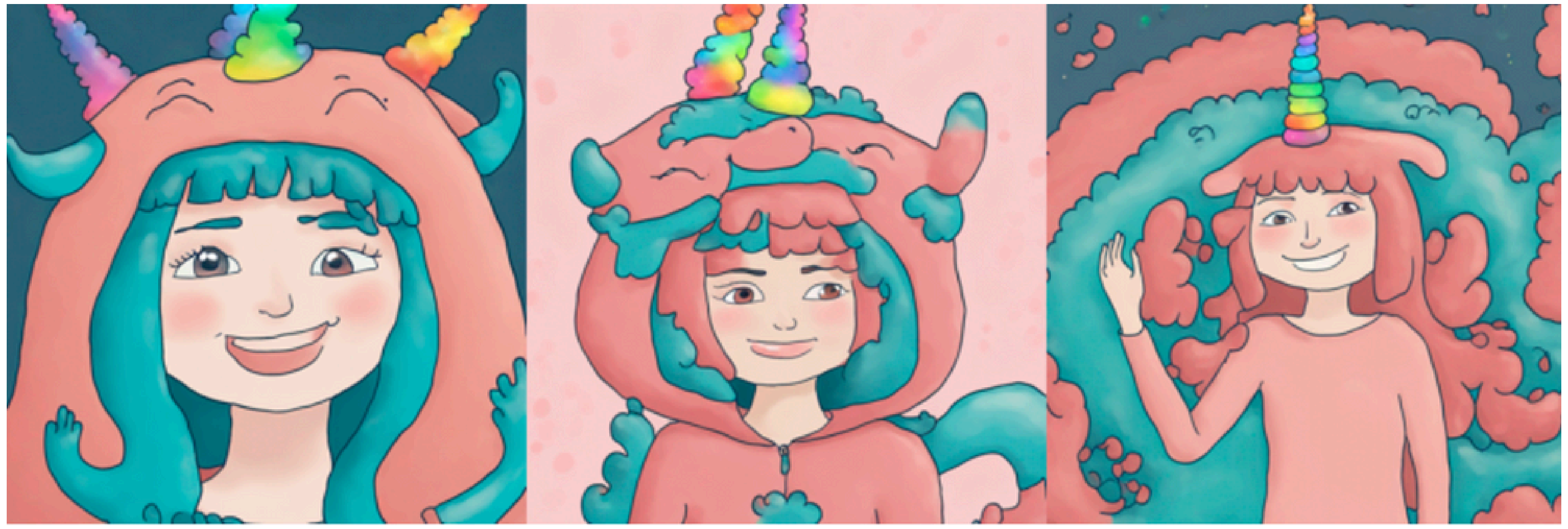
Inference: 25 Steps

Concept: "A picture of UnicornGirl"

+ angry

Results: *Dream Booth + LoRA*

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



Training:

200 Steps

Inference:

25 Steps

Concept: "A picture of UnicornGirl"

+ smiling

Results: *Dream Booth + 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”

Prompts

using the transferred learning text embeddings from SDXL

Training:

1000 Steps

Inference:

50 Steps

Concept: "A picture of UnicornGirl"

+ hands on face

+ angry

+ eating ice cream

+ eating sandwich

+ praying

Key Statistics

Training Time: Dream Booth & LoRA

- 1000 steps without GPU: 800 mins
- 1000 steps with GPU: 60 mins

Inference Time: Dream Booth & LoRA

- 50 steps without GPU: 60 mins
- 50 steps with GPU: 5 mins

Weights (Safetensors File):

- Dream Booth & LoRA: 25 MB
- Dream Booth: 4GB

"A picture of UnicornGirl sitting on a chair"

Challenges

long runtimes, iterative prompt engineering, absurd results, quality ratio: 1/5

Training: 1000 Steps

Inference: 50 Steps

Concept: “A picture of UnicornGirl”

Next Steps

General Tasks:

- Develop LoRA-only code
- Tune the models
- Discuss benchmark to compare (~5 positions)
- Get the loss per step

Tuning:

- Learning Rate
- Number of Input Images
- Model Steps
- Inference Steps

Appendix

Model Overview:

- LoRA
- Dream Booth

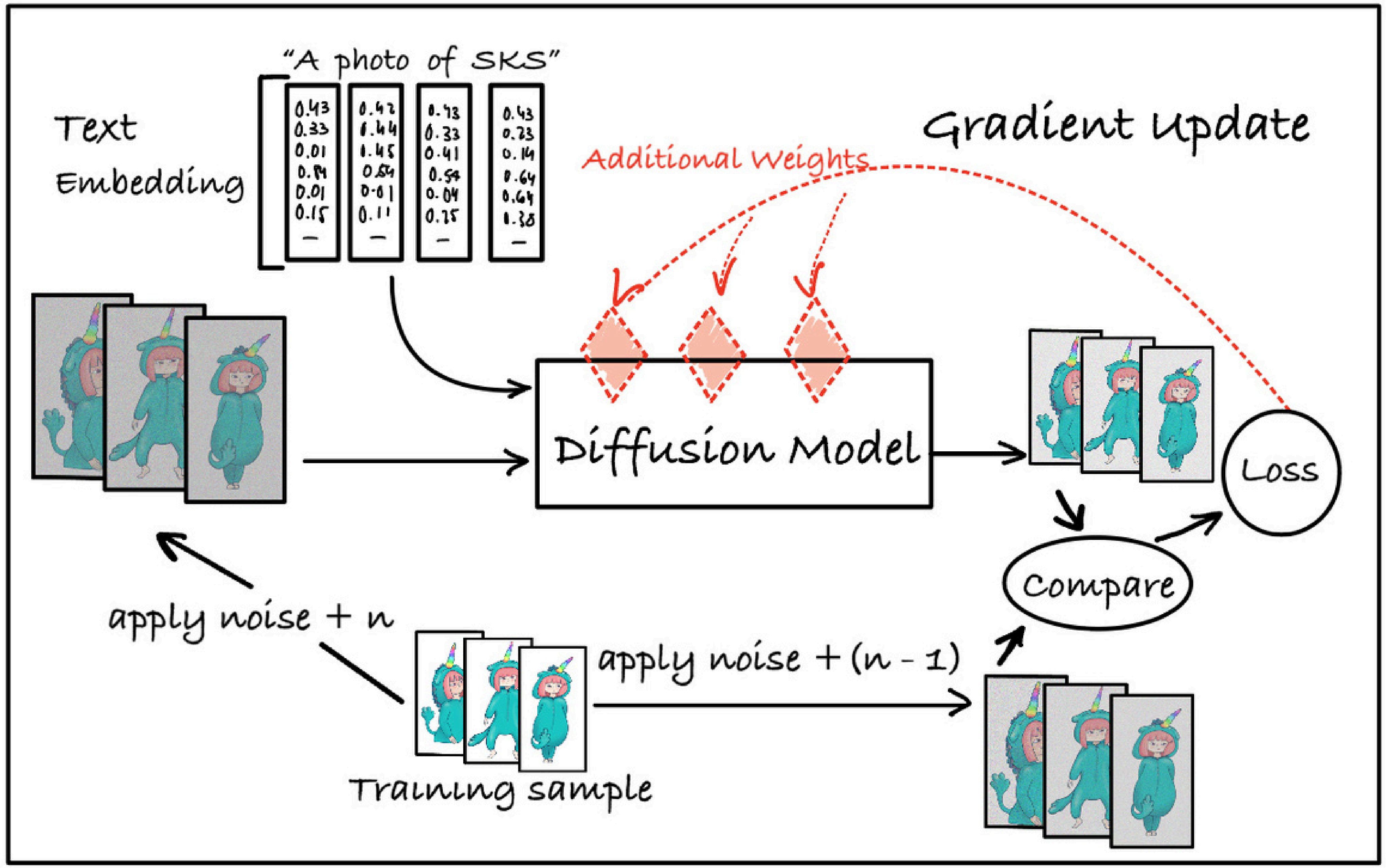
Results:

- Prompt Fail
- 1000 steps (no GPU)

The End

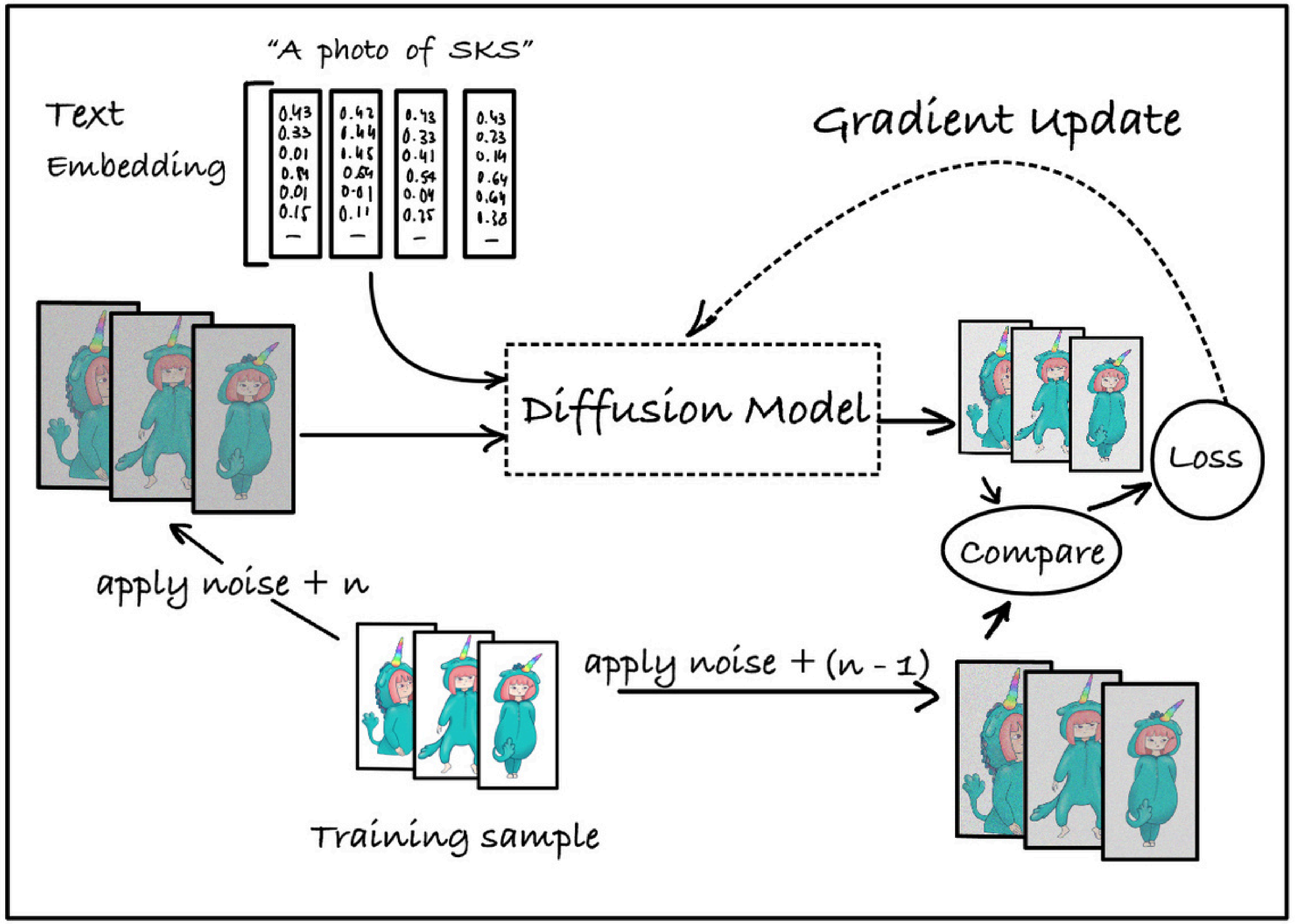
LoRA

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.



Prompt Fail

Concept: "A picture of UnicornGirl"

+ petting a dog

Results: *Dream Booth + LoRA*

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

Training:

1000 Steps - no GPU

Inference:

25 Steps

Concept: “A picture of UnicornGirl”