

# Personalization of Diffusion Models

Using 🧨 Diffusers



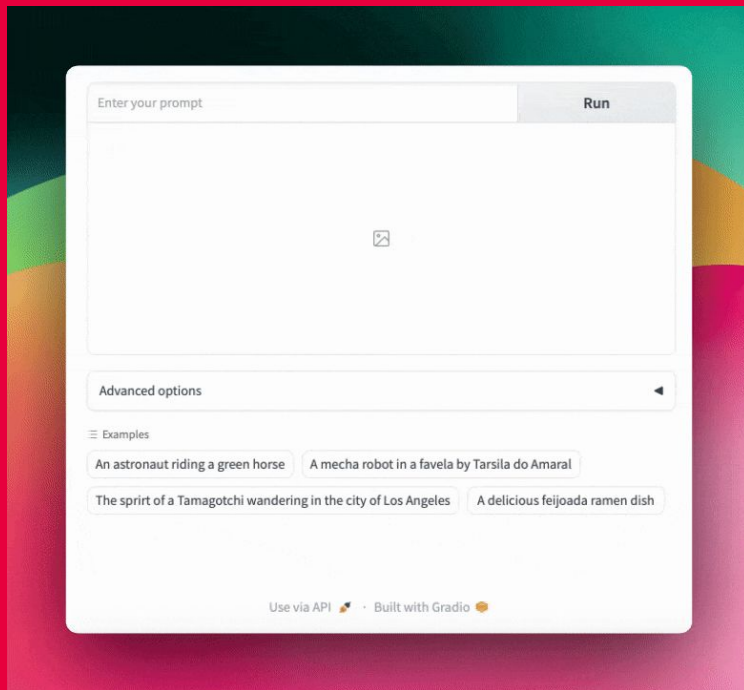
X



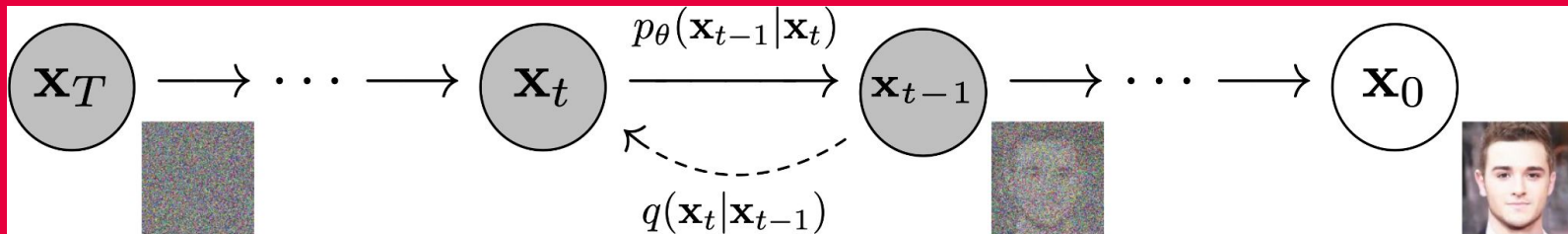
**Linoy Tsaban**  
ML Engineer  
Hugging Face

February 29th, 2024

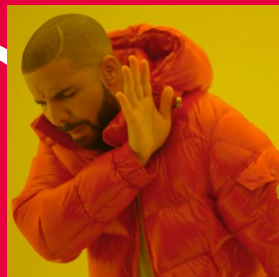
# Diffusion Models



# Diffusion Models



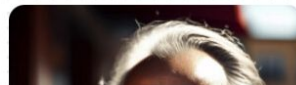
# Stable Diffusion



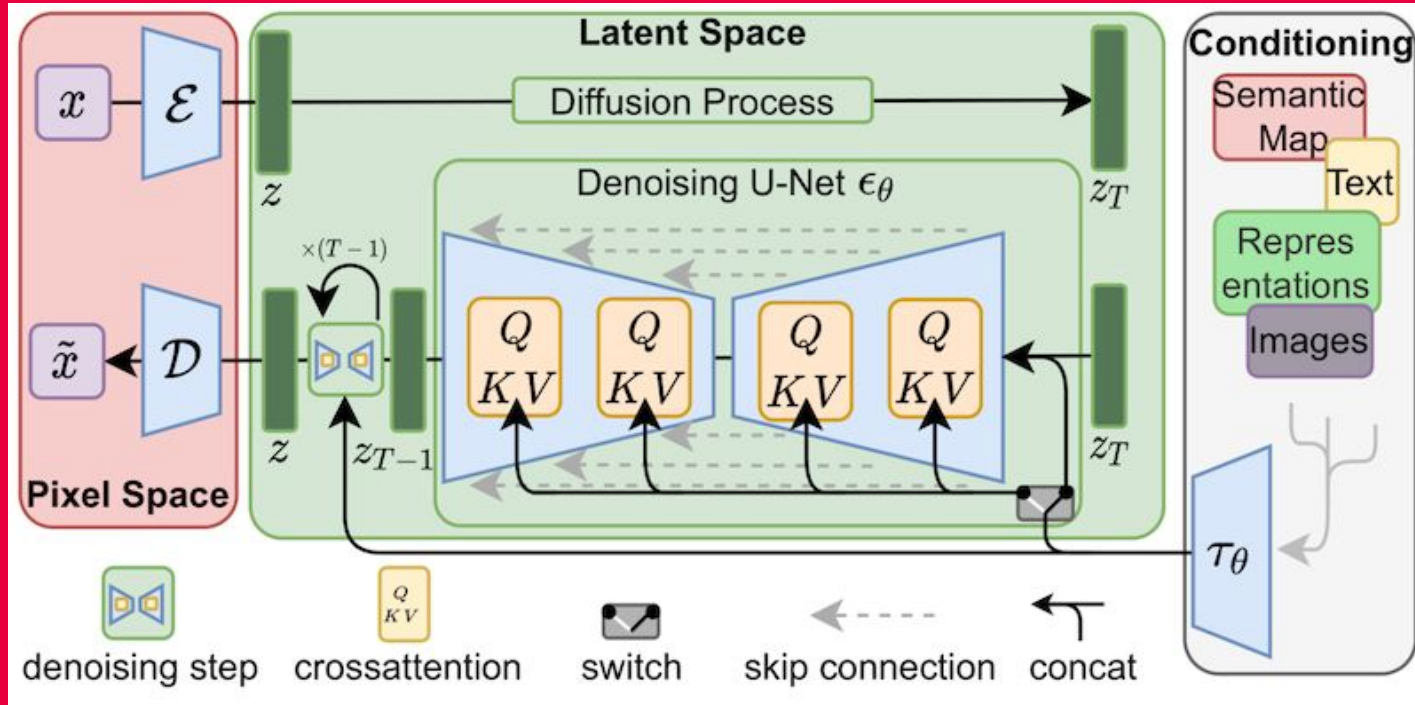
Pixel  
space



Latent  
space



# Stable Diffusion



# Latest Release - Stable Cascade



Click [here](#) to try it out!

# What is Personalization?

*Usually\**, refers to the process of teaching new concepts to a diffusion model by using specialized forms of fine-tuning.

E.g. Dreambooth, Textual Inversion, LoRA

\*





# Textual Inversion

## Textual Inversion

*(Fine Tuning Technique)*

When we teach a new concept to Stable Diffusion by finding new text embeddings that represent the concept. These embeddings are then linked to new pseudo-words, which can be incorporated into new sentences like any other word.



Input samples



"The streets of Paris  
in the style of  $S_*$ "



"Adorable corgi  
in the style of  $S_*$ "



"Painting of a black hole  
in the style of  $S_*$ "

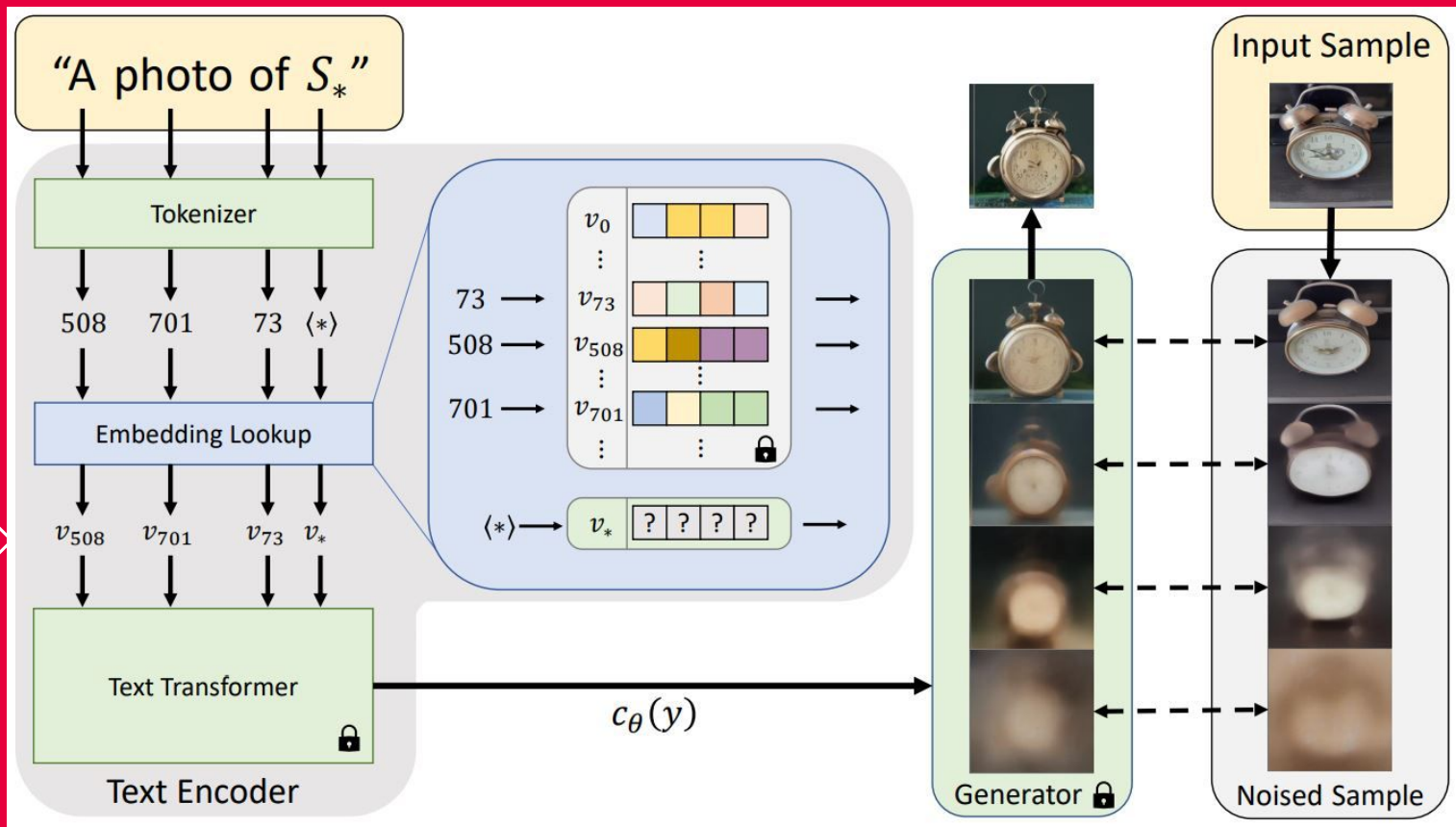


"Times square  
in the style of  $S_*$ "

Gal, Rinon, et al. "An image is worth one word: Personalizing text-to-image generation using textual inversion."



# Textual Inversion



# Dreambooth

## Dreambooth

*(Fine Tuning Technique)*

When we teach a new concept to Stable Diffusion by fine tuning on 3-5 input images paired with a text prompt containing a unique identifier and the name of the class the subject belongs to (e.g., "A photo of a [T] dog")



Input images



[V] dog in the Acropolis



swimming



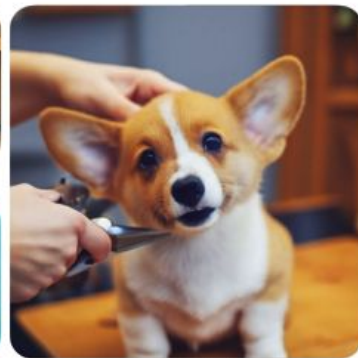
sleeping



in a doghouse

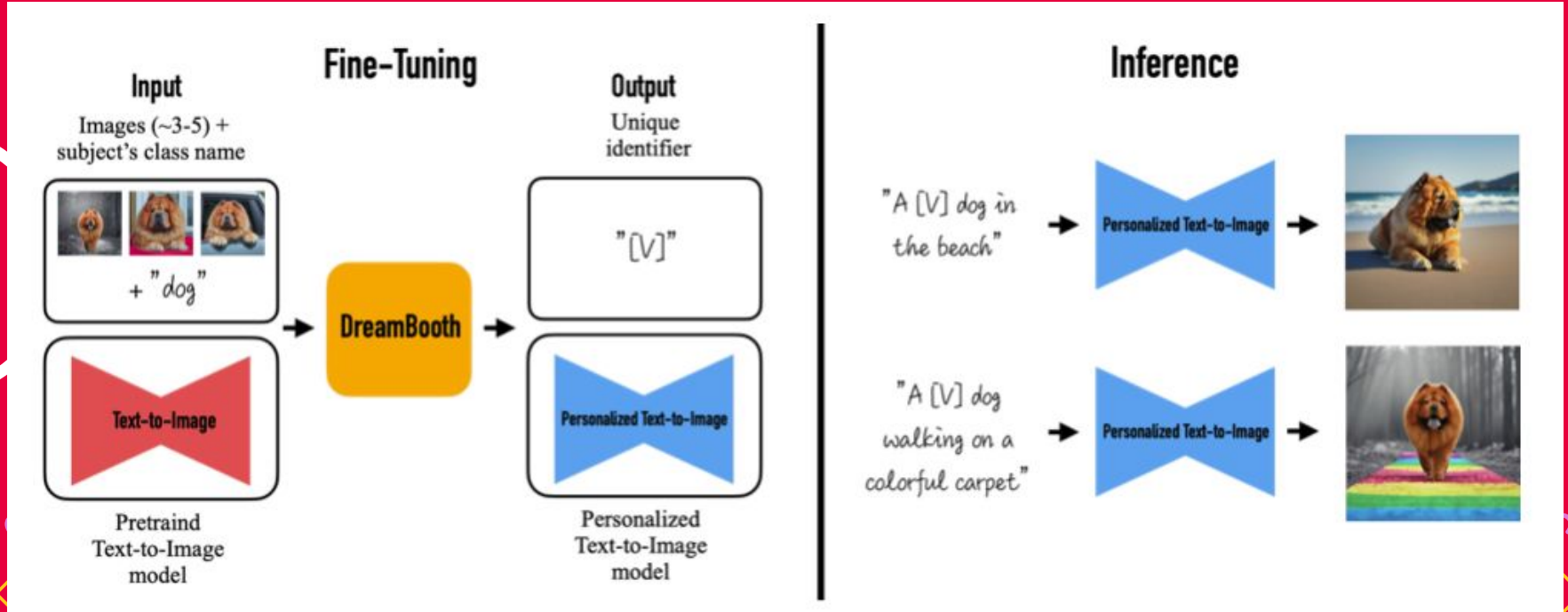


in a bucket



getting a haircut

# Dreambooth



# LoRA

Hu, Edward J., et al. "Lora: Low-rank adaptation of large language models."

1. First introduced for LLMs, and later proposed by Simo Ryu for diffusion models
2. Proposes to freeze pre-trained model weights and inject trainable layers in transformer blocks
3. Quality on par with full fine-tuning while being faster & needing less compute



<https://github.com/cloneofsimo/lora>

# LoRA

## LoRA: Low-Rank Adaptation of Large Language Models

For a pre-trained weight matrix  $W_0$ , LoRA decomposes its update into  $W_0 + \Delta W$  and represents  $\Delta W$  as a product of **two low-rank matrices, B and A**.

Both B and A have ranks significantly smaller than the dimensions of  $W_0$ , dramatically reducing the number of trainable parameters.

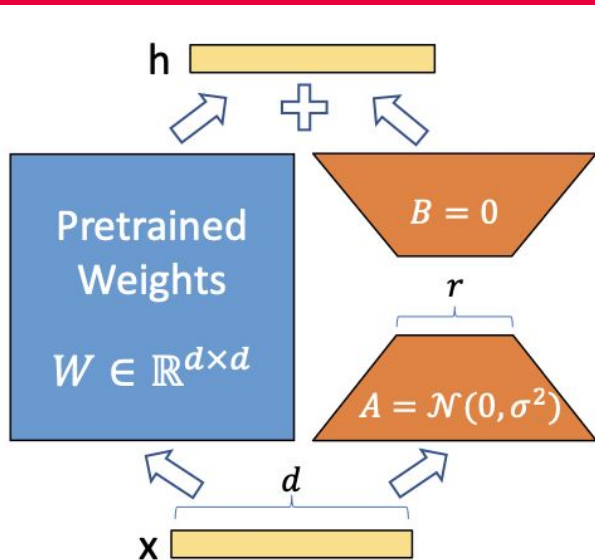


Figure 1: Our reparametrization. We only train  $A$  and  $B$ .

# Recap

	Output
Dreambooth	Fine tuned model
LoRA	Injectable layers
Textual Inversion	Text embeddings

# Let's Train

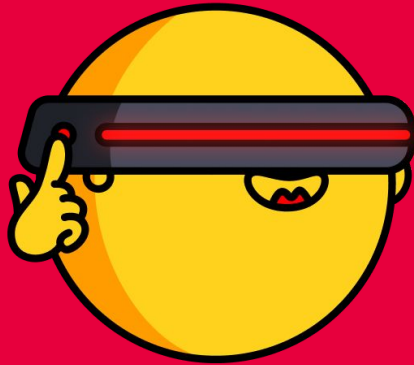




What if  
we combine  
**Dreambooth**  
with **LoRA**?



# Notebook #1



10101  
1110101001  
1010100101010110  
101011110010101010011  
101011010101101010001010  
111100010010101010111  
101010100000001011  
1010101010101110  
101110101  
101010




# Recap

	Output
Dreambooth	Fine tuned model
LoRA	Injectable layers
Textual Inversion	Text embeddings
<b>Dreambooth LoRA</b>	<b>Injectable layers</b>

What if #2:  
We add in  
**Textual  
inversion?**



# Share your models with the community!


 linoyts/2000\_ads like 2

Text-to-Image Diffusers stable-diffusion-xl stable-diffusion-xl-diffusers lora template:sd-lora License: openrail++


Model card Files and versions Community Settings Deploy Use in Diffusers

Edit model card


### SDXL LoRA DreamBooth - LinoyTsan/2000\_ads




Prompt  
<s0><s1> ad of a llama wearing headphones




Prompt  
<s0><s1> ad for banana flavored toothpaste




Prompt  
<s0><s1> ad of the Mona Lisa eating ramen



Prompt  
<s0><s1> ad for speakers




Prompt  
<s0><s1> ad of Santa clause



Prompt  
<s0><s1> ad of the an astronaut

Downloads last month  
232




#### Inference API

Text-to-Image Examples

<s0><s1> ad for speakers Compute

This model can be loaded on the Inference API on-demand.



# Pivotal Tuning

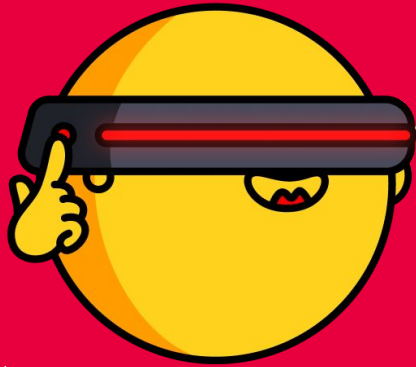
## Pivotal Tuning

---

*(Fine Tuning Technique)*

combines Textual Inversion with regular diffusion fine-tuning by inserting new tokens into the text encoders of the model, instead of reusing existing ones and optimize the associated embeddings. We continue to fine-tune the LoRA layers of the unet with the new textual embeddings.

# Notebook #2





# Recap

	Output
Dreambooth	Fine tuned model
LoRA	Injectable layers
Textual Inversion	Text embeddings
<b>Dreambooth LoRA</b>	<b>Injectable layers</b>
<b>Dreambooth LoRA with Pivotal Tuning</b>	<b>Injectable layers &amp; text embeddings</b>

# Learn More 🤗

🎬 **Introduction to Diffusion Models:**  
Hugging Face [Diffusion Models Course](#)

💌 **Blog Posts:**  
[Blog post about Dreambooth](#)  
[Blog post about LoRAs](#)  
[Blog post about current SOTA practices for Dreambooth](#)  
[LoRA](#)

📁 **Useful Docs:**  
[Dreambooth with Diffusers](#)  
[Textual Inversion](#) with Diffusers  
[LoRA](#) with Diffusers

# Personalization Playground

Demos worth trying 🔥

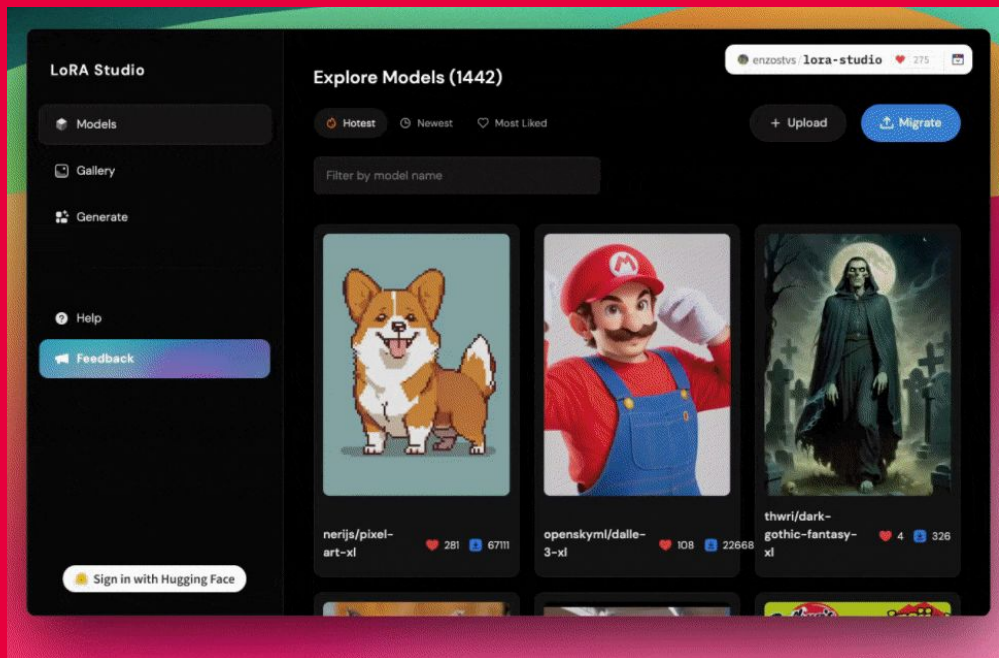
► [LoRA Studio](#)

► [LoRA the explorer](#)

► [LoRA roulette](#)

► [IP Adapter](#)

► [Instant ID](#)





**Thank You!**