JUNE 18-22, 2023
CVPR
VANCOUVER, CANADA

DreamBooth: Fine Tuning Text-to-Image Diffusion Models for Subject-Driven Generation

Poster Session: THU-PM-181


Nataniel Ruiz


Yuanzhen Li


Varun Jampani


Yael Pritch


Michael Rubinstein

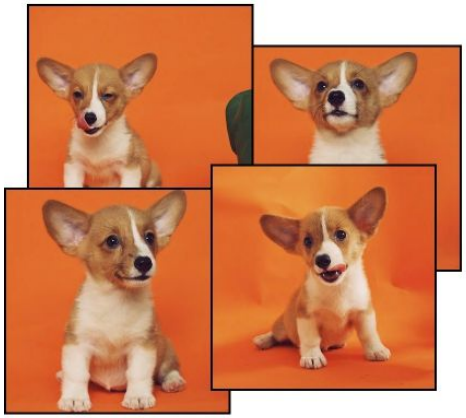

Kfir Aberman

Google Research

BOSTON UNIVERSITY

DreamBooth

Fine Tuning Text-to-Image Diffusion Models for Subject-Driven Generation



Input images

in the Acropolis

swimming

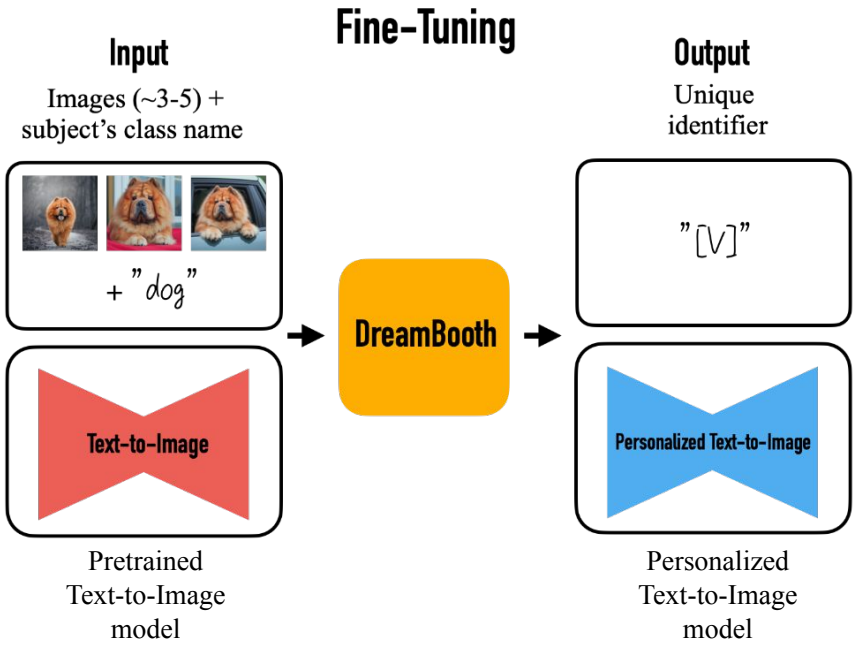
sleeping

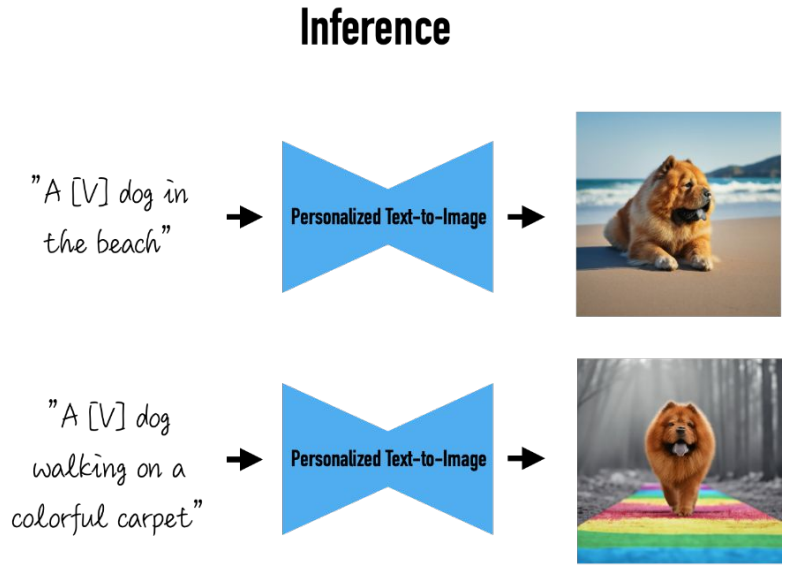
in a doghouse

in a bucket

getting a haircut

Method Summary





Subject-Driven Generation

Subject-Driven Generation

✓ New contexts

✗ Subject Fidelity

"A yellow alarm clock with a large yellow number 3 on the right side of the clock face, in the jungle."

Detailed Text Prompt?

Saharia et al. "Photorealistic text-to-image diffusion models with deep language understanding." NeurIPS 2022

Subject-Driven Generation

❌ New contexts

❌ Subject Fidelity

Condition on CLIP Embedding from reference images

Ramesh, Aditya, et al. "Hierarchical text-conditional image generation with clip latents." 2022

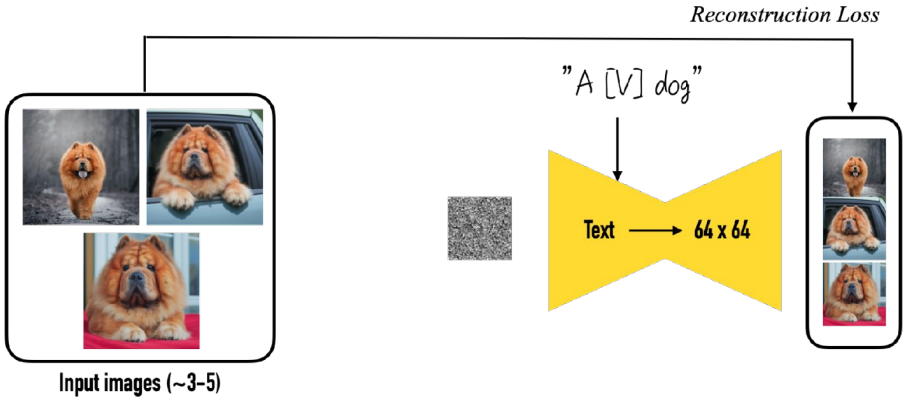
Subject-Driven Generation

✓ New contexts

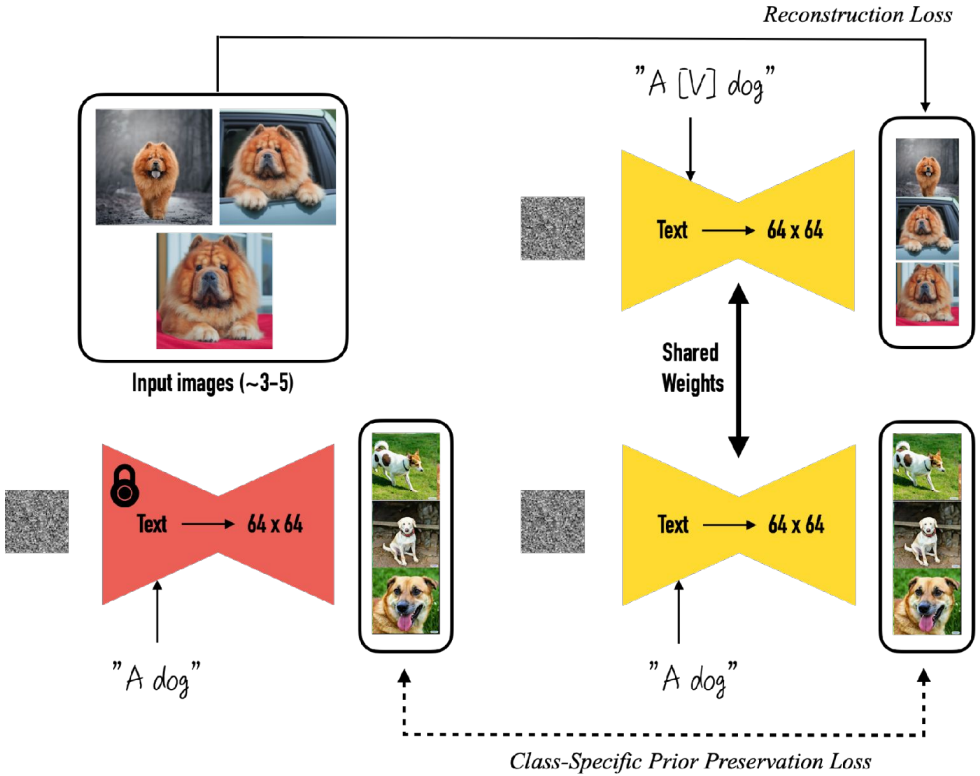
✓ Subject Fidelity

DreamBooth

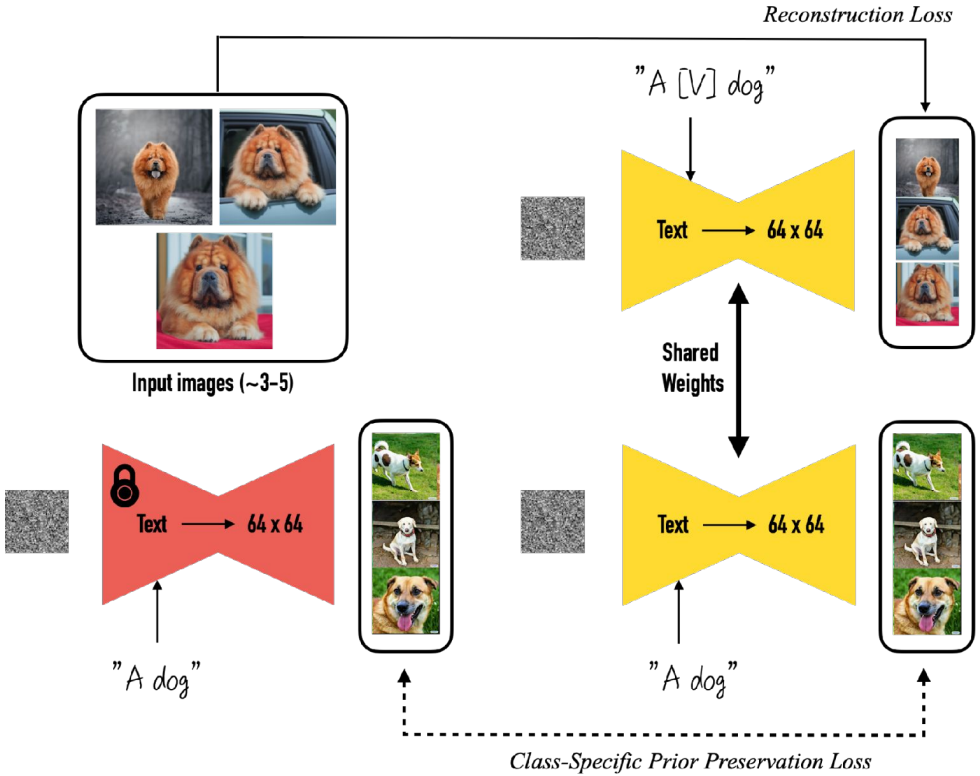
Full Approach

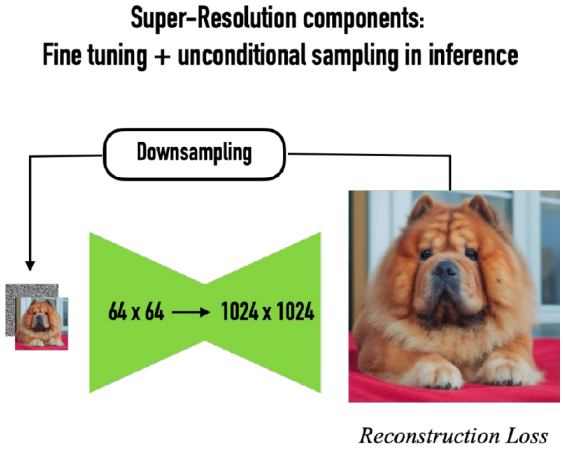


Full Approach



Full Approach





Recontextualization

Input images

Input images

A [V] backpack in the Grand Canyon

A [V] backpack with the night sky

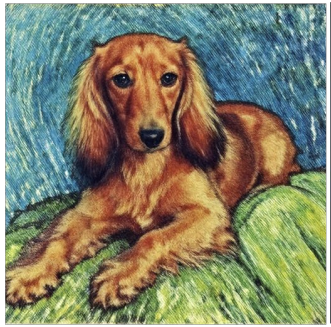
A [V] backpack in Boston

A [V] backpack in the city of Versailles

A wet [V] backpack in water

Artistic Renditions

Input images



Vincent Van Gogh

Michelangelo

Rembrandt

Johannes Vermeer

Pierre-Auguste Renoir

Leonardo da Vinci

Property Modification

Input

Hybrids (*"A cross of a [V] dog and a [target species]"*)

Bear

Panda

Koala

Lion

Hippo

Accessorization

Input images

Chef Outfit

Witch Outfit

Ironman Outfit

Nurse Outfit

Purple Wizard Outfit

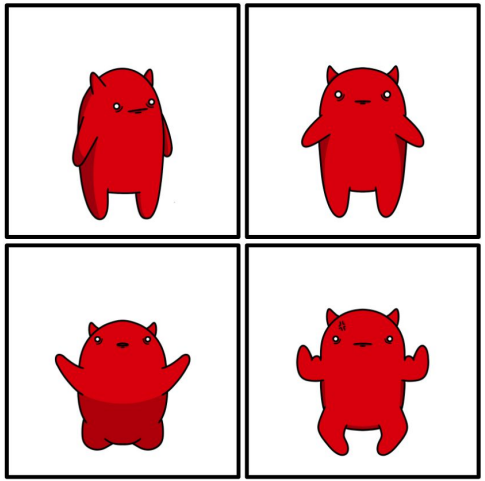
Superman Outfit

Police Outfit

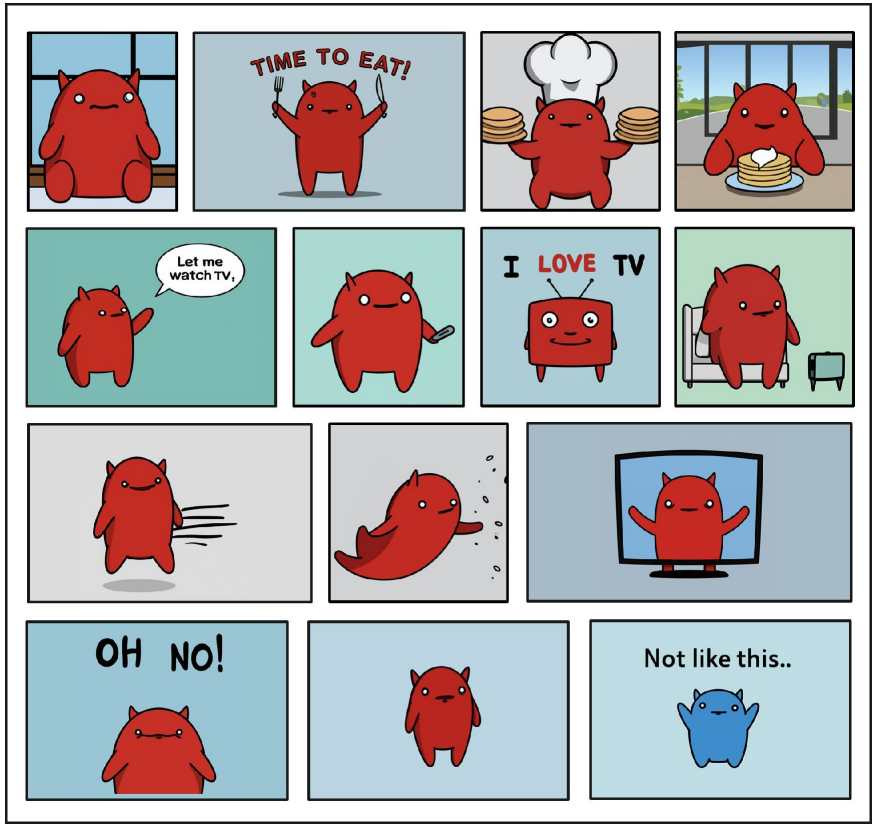
Angel Wings

Comics

Real



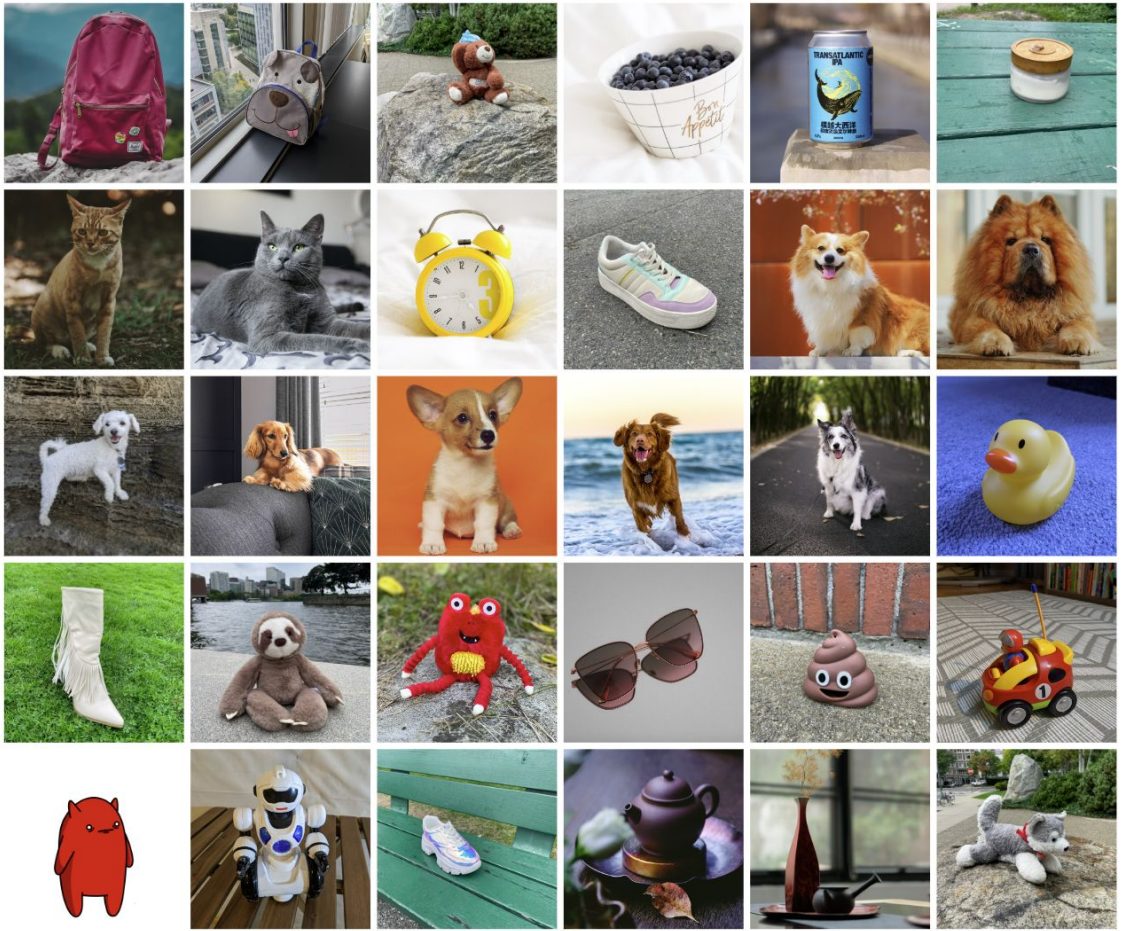
Generated



How to Evaluate?

Dataset

- 30 Subjects
- 3-6 Images per Subject
- 21 inanimate objects
- 9 pets
- 25 prompts



https://github.com/google/dreambooth

Dataset Prompts

a [V] [class] floating in an ocean of milk

a wet [V] [class]

a red [V] [class]

a [V] [class] in the jungle

a [V] [class] with a blue house in the background

a [V] [class] on top of a purple rug

a [V] [class] wearing a santa hat

a [V] [class] with a tree and autumn leaves in the background

a [V] [class] wearing a black top hat and a monocle

DreamBooth and Textual Inversion

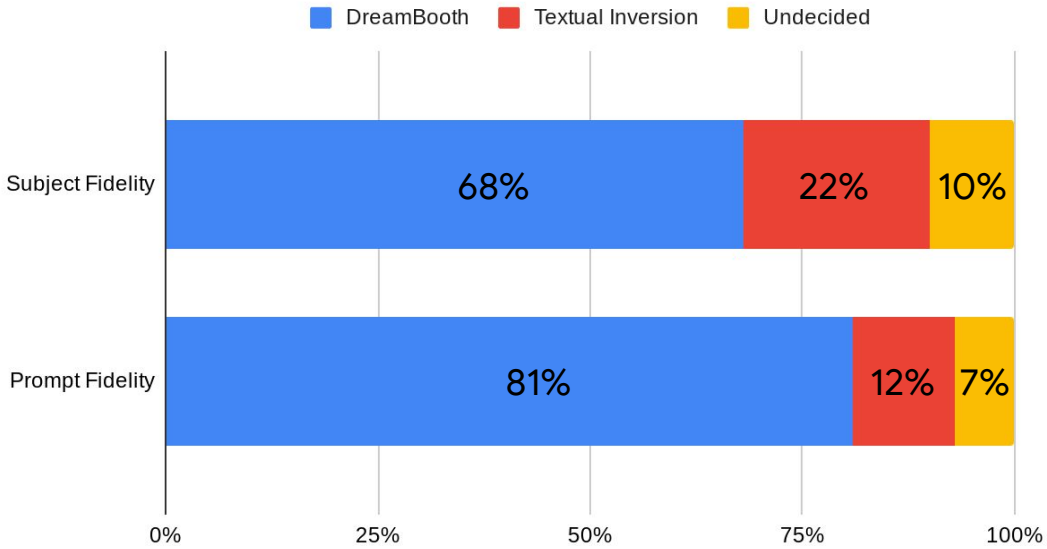
DreamBooth

Textual Inversion

Gal et al. "An image is worth one word: Personalizing text-to-image generation using textual inversion." ICLR 2022

User Study

- 72 users
- 25 item questionnaires
- 3 users / questionnaire
- Total of 1800 answers



Subject Fidelity Metrics

Reference Real Sample

Real Sample

0.783

Synthetic Samples

0.792

CLIP-Image Scores 

Subject Fidelity Metrics

Reference Real Sample

Real Sample

0.772

Synthetic Samples

0.678

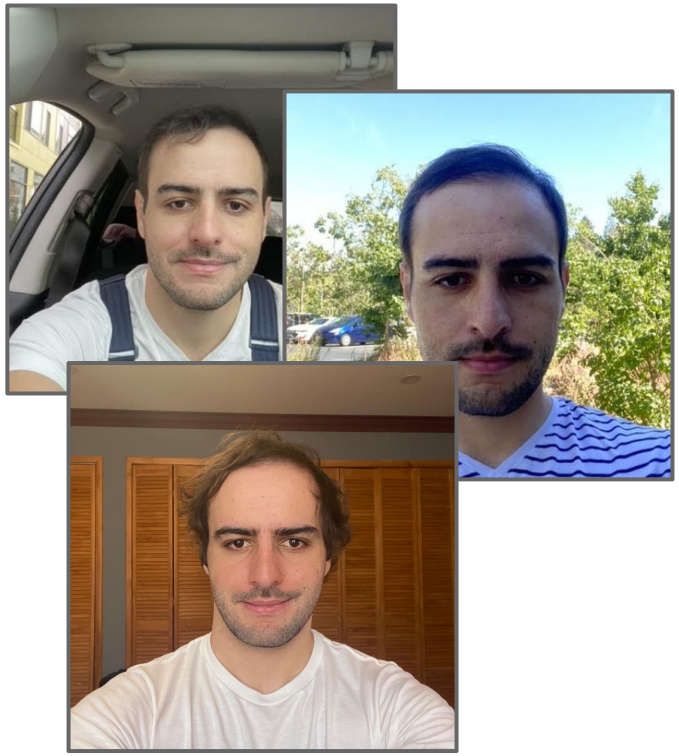
DINO ✓

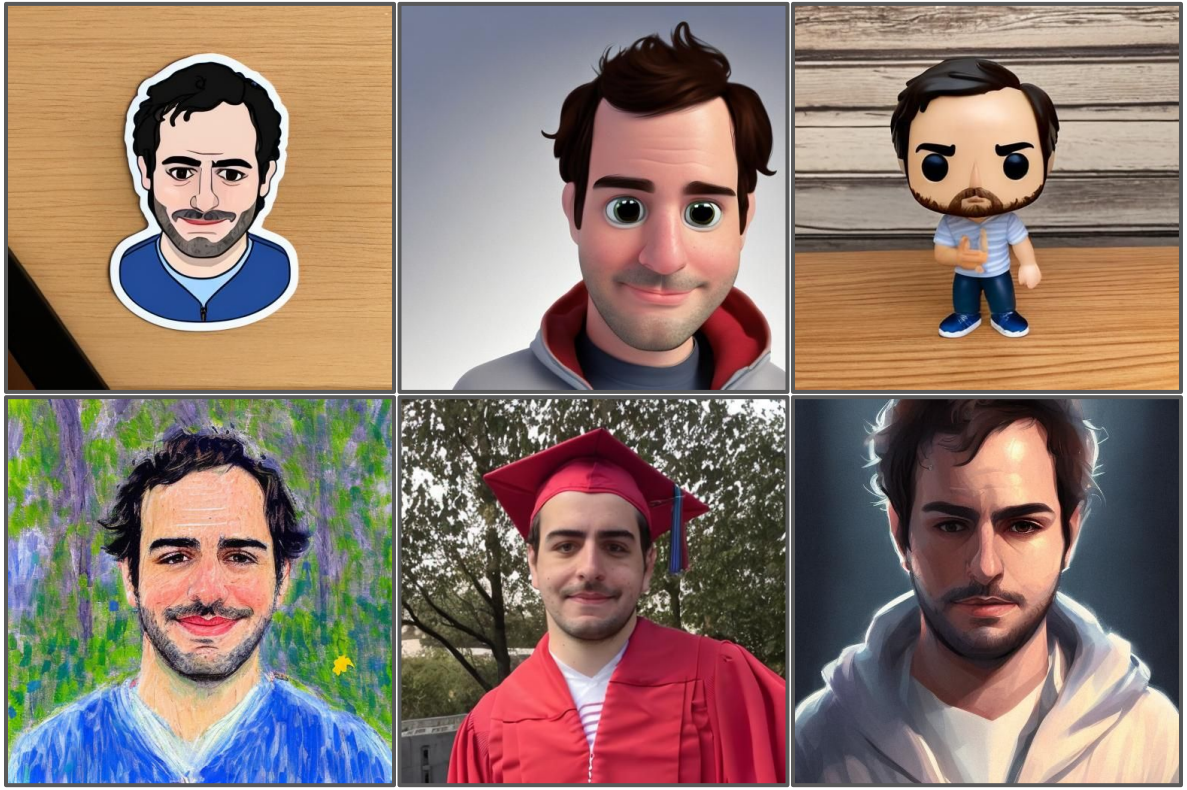
Quantitative Metrics

Method	DINO ↑	CLIP-I ↑	CLIP-T ↑
Real Images	0.774	0.885	N/A
DreamBooth (Imagen)	**0.696**	**0.812**	**0.306**
DreamBooth (Stable Diffusion)	0.668	0.803	0.305
Textual Inversion (Stable Diffusion)	0.569	0.780	0.255

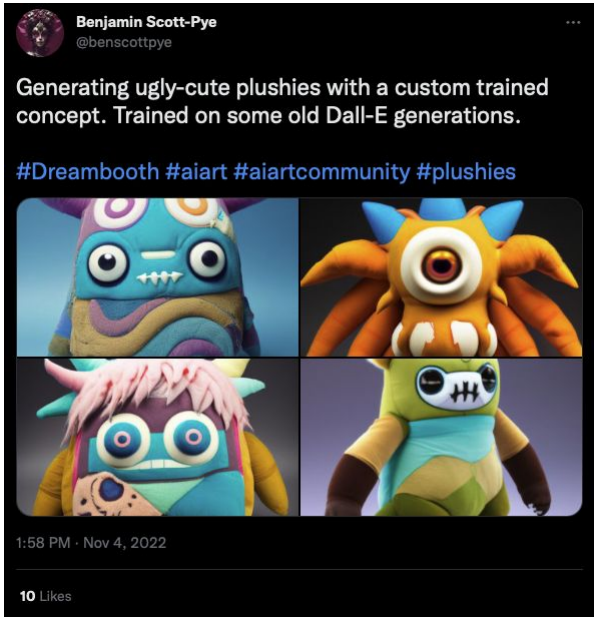
Application-Driven Exploration

AI Selfies





Thematic Plushie Creation

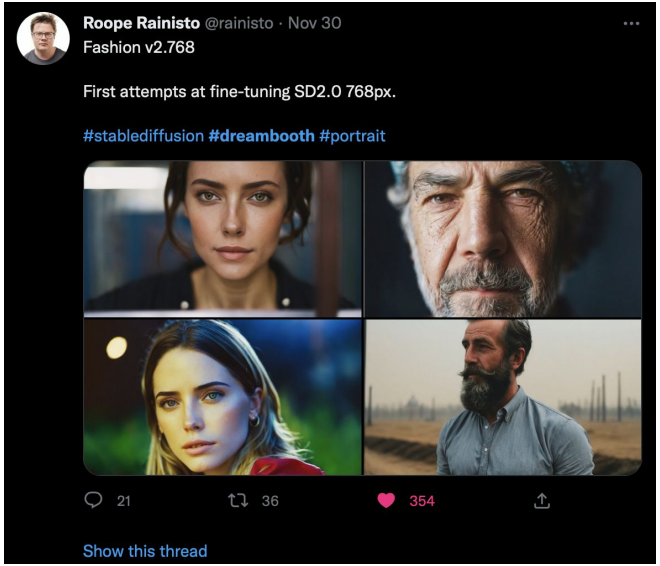


Twitter: @benscottpye

Video Game Asset Generation

Twitter: @emmanuel_2m

Artistic Images



Credit: @rainisto

JUNE 18-22, 2023
CVPR
VANCOUVER, CANADA

DreamBooth: Fine Tuning Text-to-Image Diffusion Models for Subject-Driven Generation

Poster Session: THU-PM-181


Nataniel Ruiz


Yuanzhen Li


Varun Jampani


Yael Pritch


Michael Rubinstein


Kfir Aberman

Google Research

BOSTON UNIVERSITY