Outline (Cont. from part 1 covered in Lec#17)

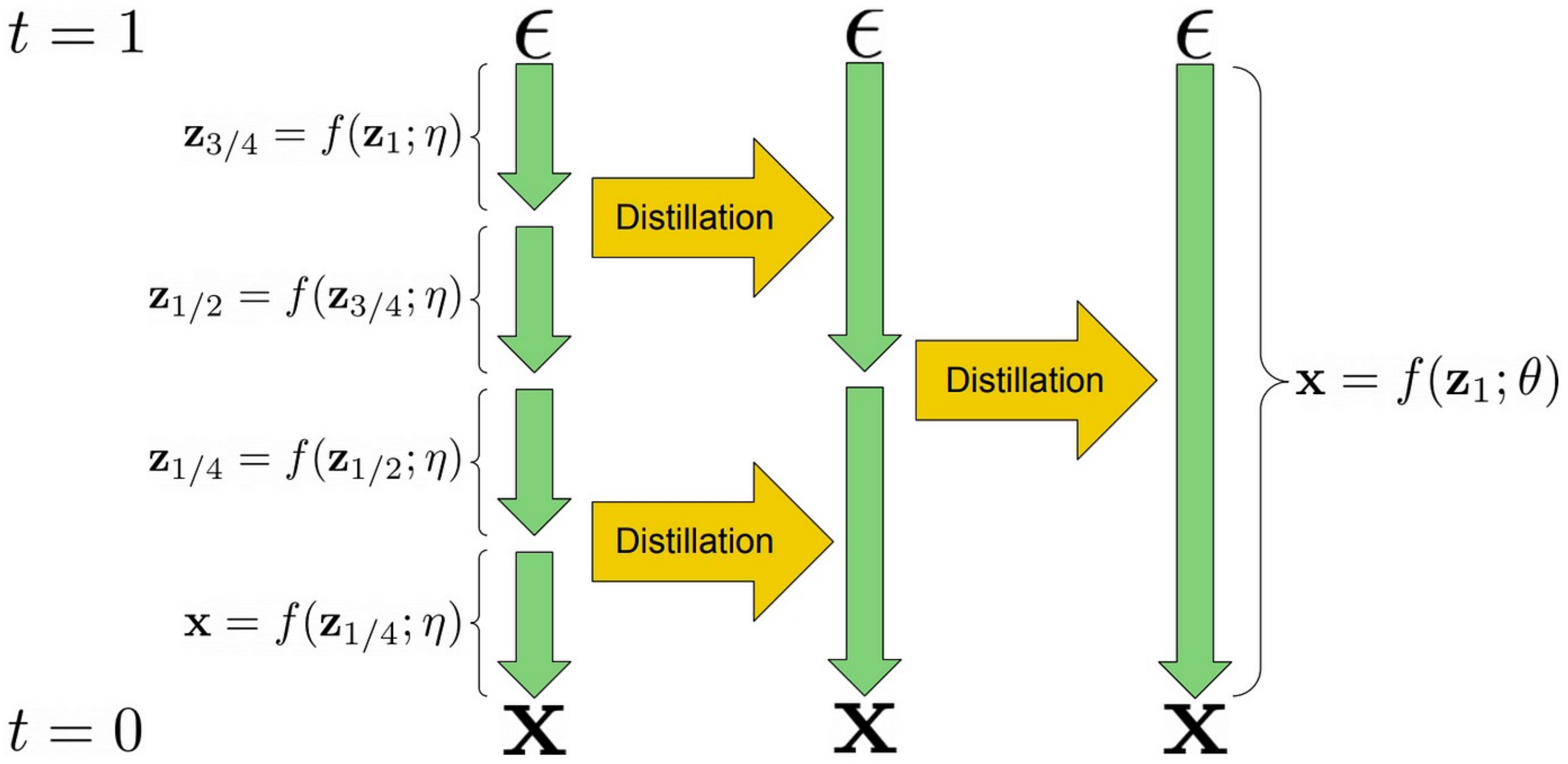
Part 1: Basics

- Denoising diffusion probabilistic models (DDPMs)
- Conditional diffusion models
- Large-scale models: DALL-E 2, Stable Diffusion, Imagen

Part 2: Recent Advances

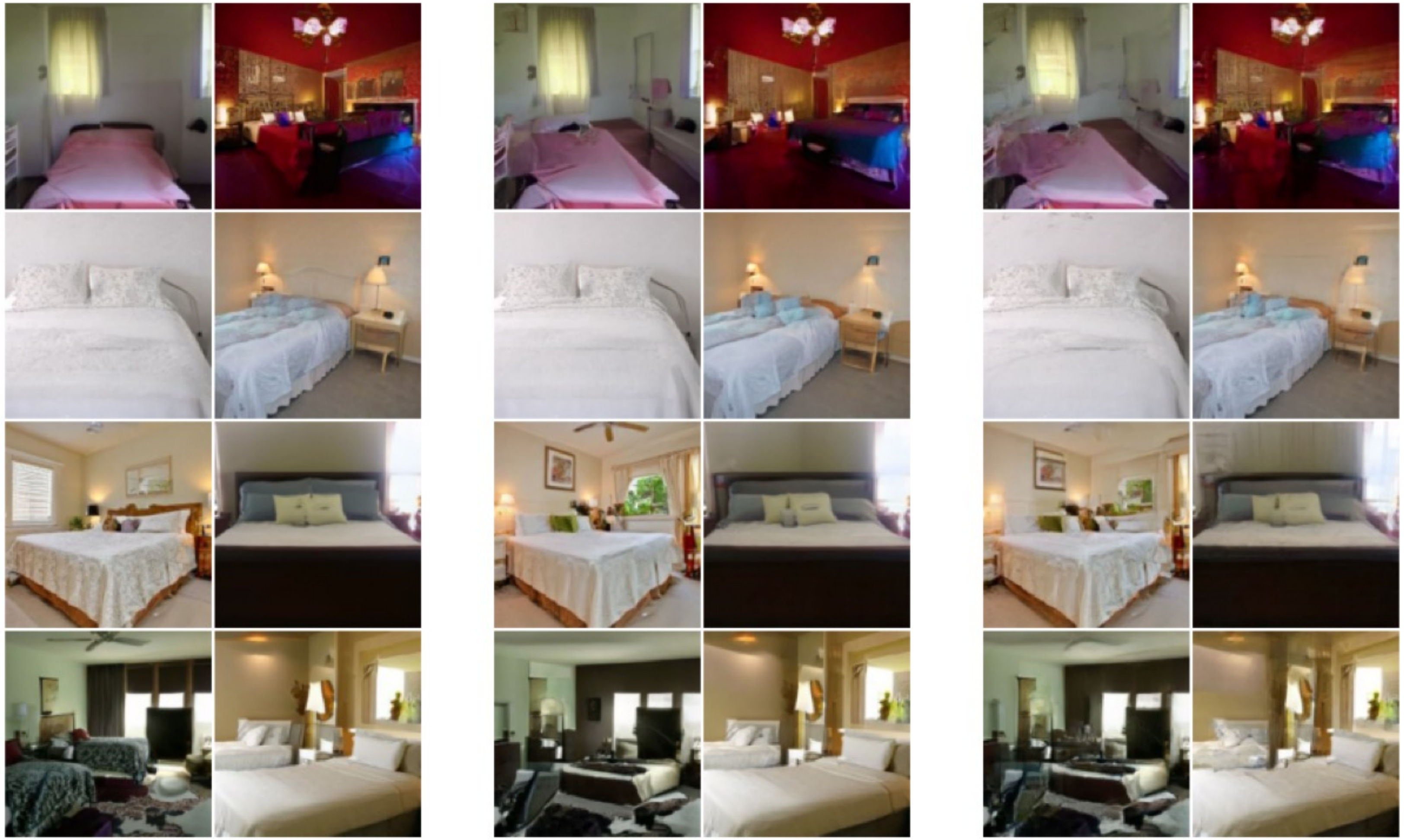
- Denoising diffusion implicit models (DDIMs)
- Stable Diffusion XL, Stable Diffusion 3
- Progressive Distillation

Progressive Distillation



[Salimans et al, Progressive distillation for fast sampling of diffusion models]

Progressive Distillation: Results



(a) 256 sampling steps (b) 4 sampling steps (c) 1 sampling step

Figure 10: Random samples from our distilled LSUN bedrooms models, for fixed random seed and for varying number of sampling steps.

Outline

Part 1: Basics

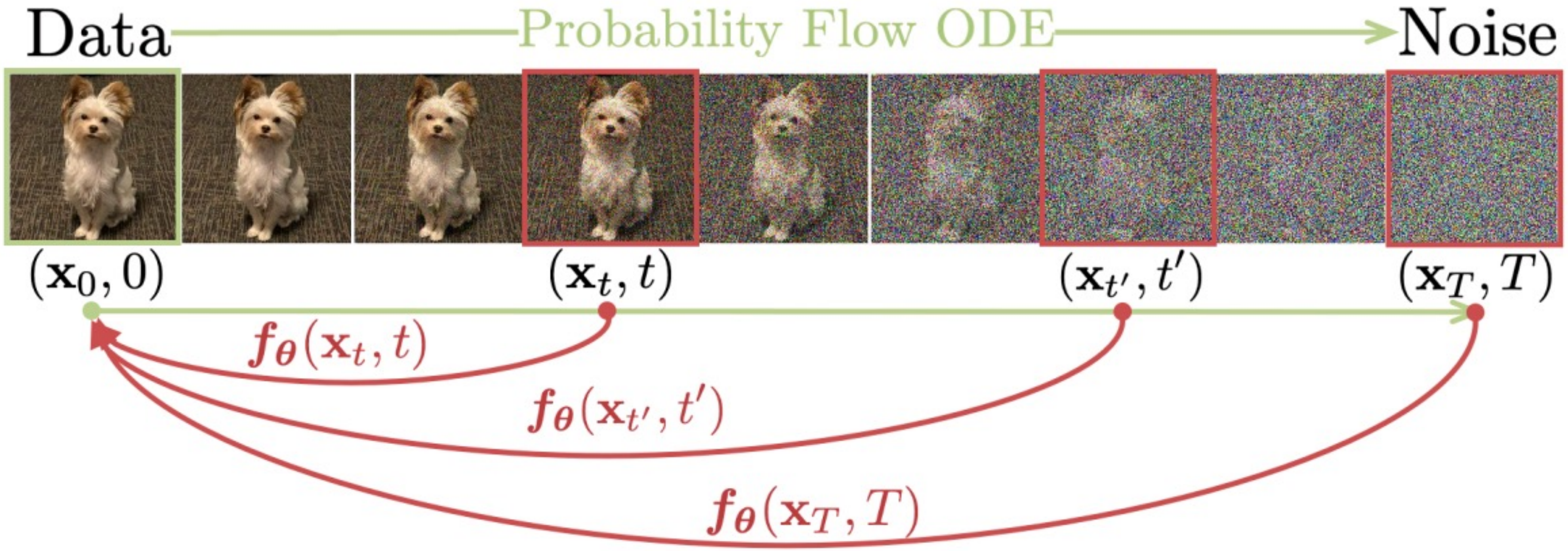
- Denoising diffusion probabilistic models (DDPMs)
- Conditional diffusion models
- Large-scale models: DALL-E 2, Stable Diffusion, Imagen

Part 2: Recent Advances

- Denoising diffusion implicit models (DDIMs)
- Stable Diffusion XL, Stable Diffusion 3
- Model Distillation
- Latent Consistency Models (LCM)

Latent Consistency Models

Consistency Models

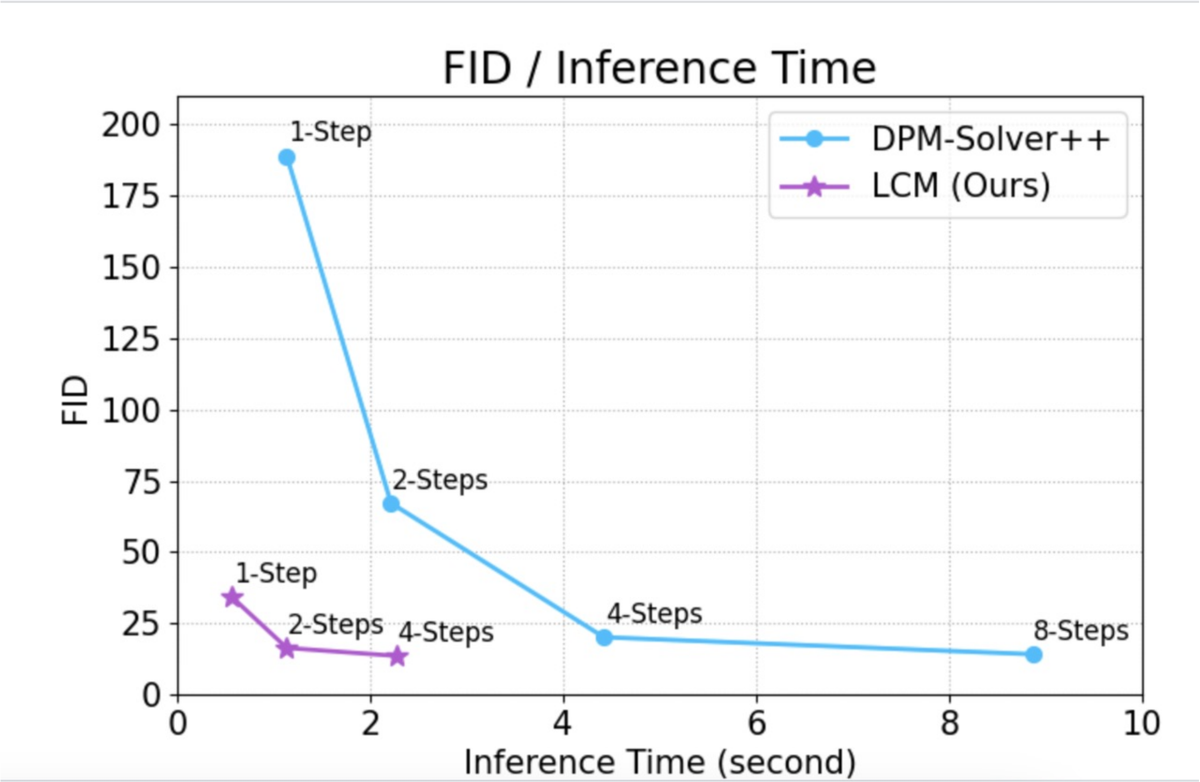


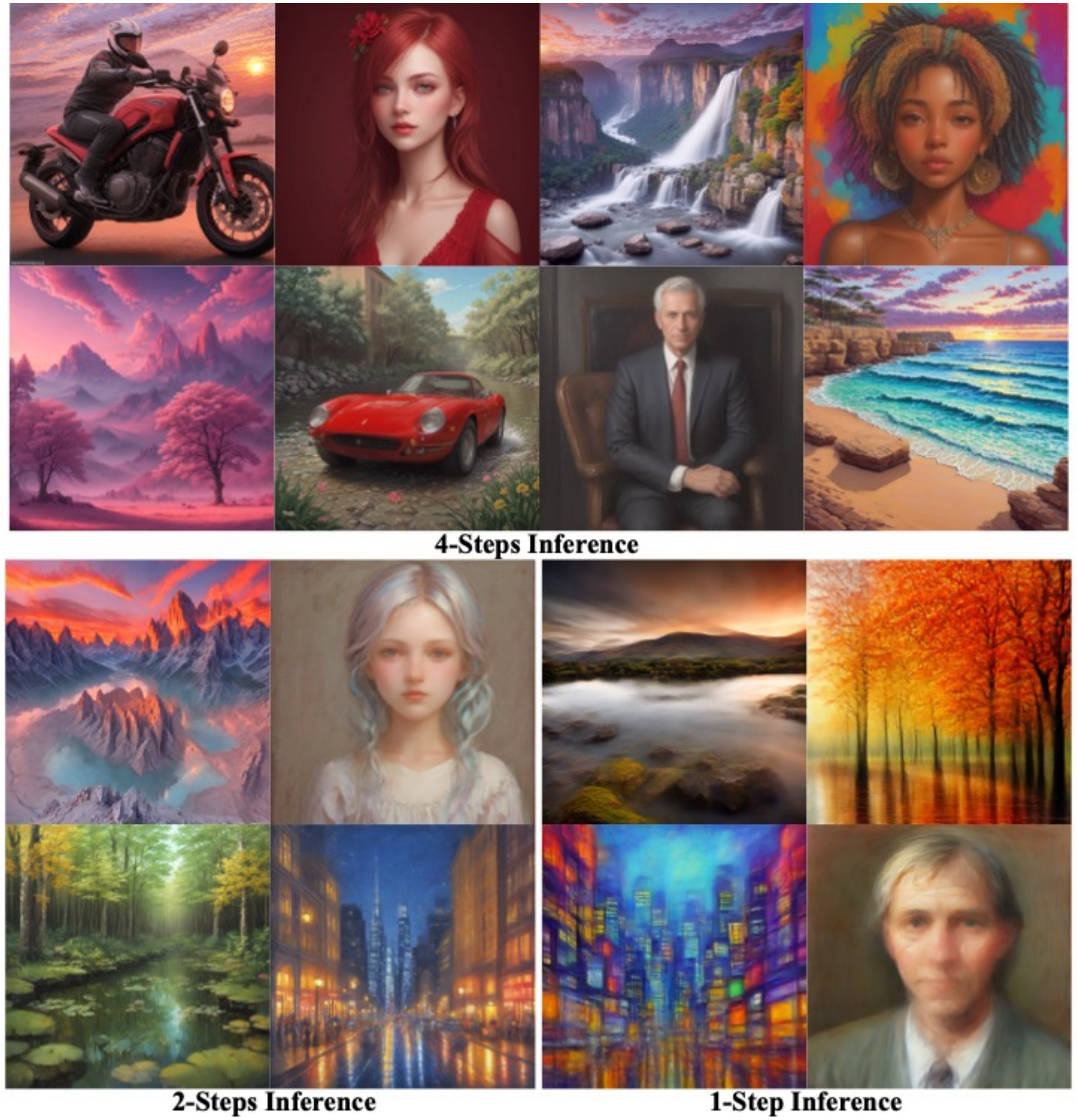
Latent Consistency Models: combine the above idea with Latent Diffusion Models

[Song et al., Consistency Models]

[Luo et al, Latent Consistency Models: Synthesizing High-Resolution Images with Few-step Inference]

Latent Consistency Models: Results





Outline

Part 1: Basics

- Denoising diffusion probabilistic models (DDPMs)
- Conditional diffusion models
- Large-scale models: DALL-E 2, Stable Diffusion, Imagen

Part 2: Recent Advances

- Denoising diffusion implicit models (DDIMs)
- Stable Diffusion XL, Stable Diffusion 3
- Progressive Distillation
- Latent Consistency Models (LCM)

(next class) Part 3: Applications and Implementation, Ethical issues

Outline

Part 3: Applications and Implementation; Ethical Issues

- Customizing Diffusion Models
  - Textual Inversion
  - DreamBooth
  - Low Rank Approximation (LoRA)
  - ZipLoRA
- ControlNet
- Prompt-to-Prompt
- InstructPix2Pix
- DreamFusion
- Working with Diffusion Models: Implementation aspects
- Societal, ethical, and legal issues

Outline

Part 3: Applications and Implementation; Ethical Issues

- Customizing Diffusion Models
  - Textual Inversion
  - DreamBooth
  - Low Rank Approximation (LoRA)
  - ZipLoRA

Customizing DMs: Textual inversion

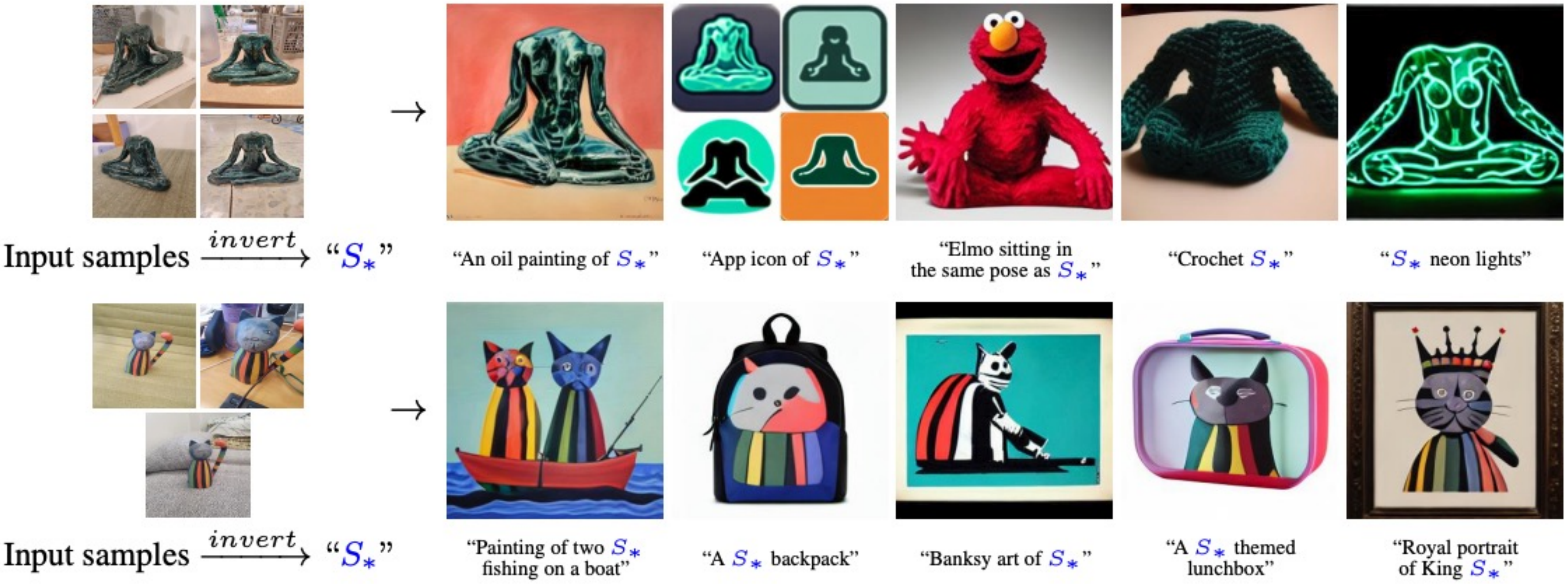


Figure 1: (left) We find new pseudo-words in the embedding space of pre-trained text-to-image models which describe specific concepts. (right) These pseudo-words are composed into new sentences, placing our targets in new scenes, changing their style or ingraining them into new products.

R. Gal et al. [An Image is Worth One Word: Personalizing Text-to-Image Generation using Textual Inversion](). ICLR 2023

Customizing DMs: Textual inversion

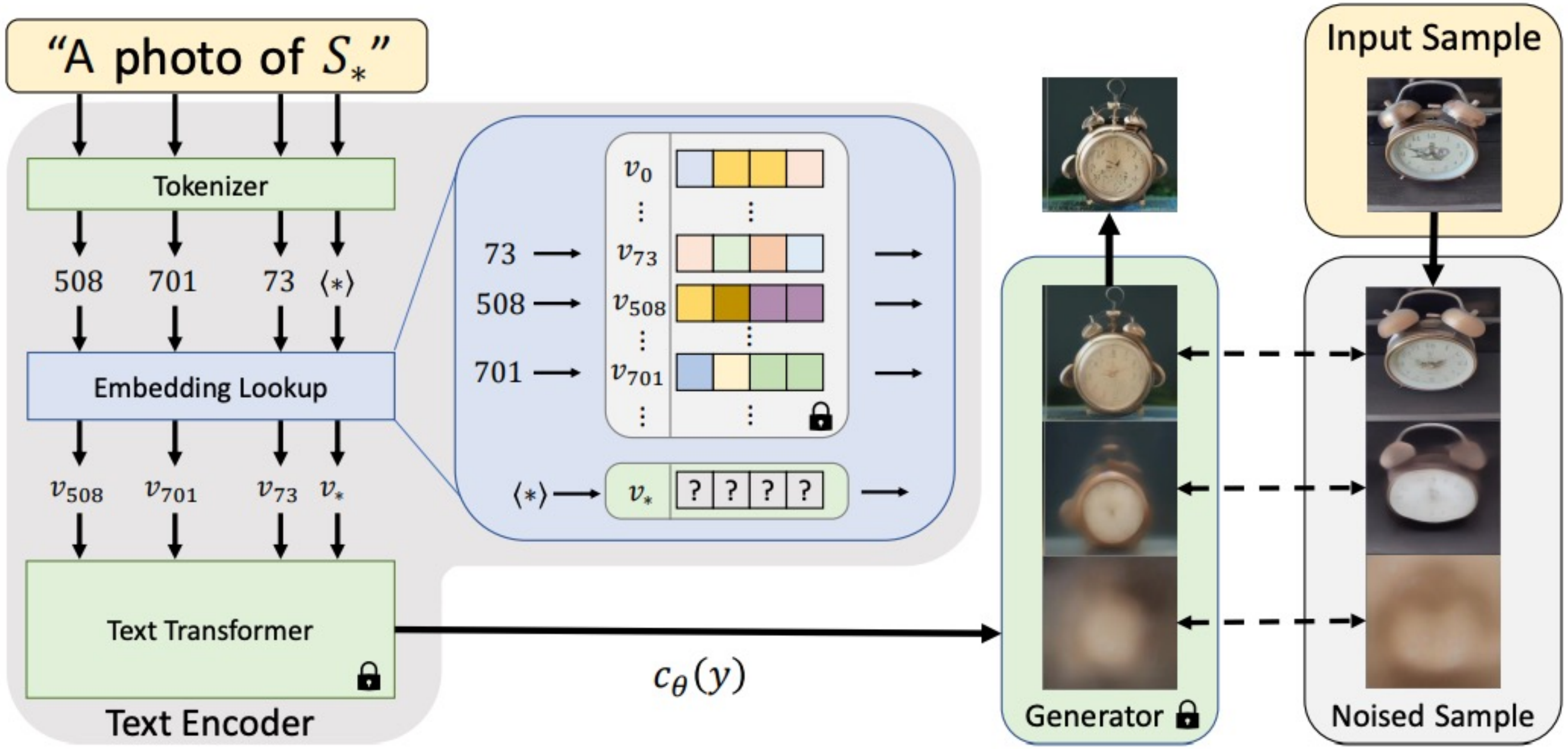


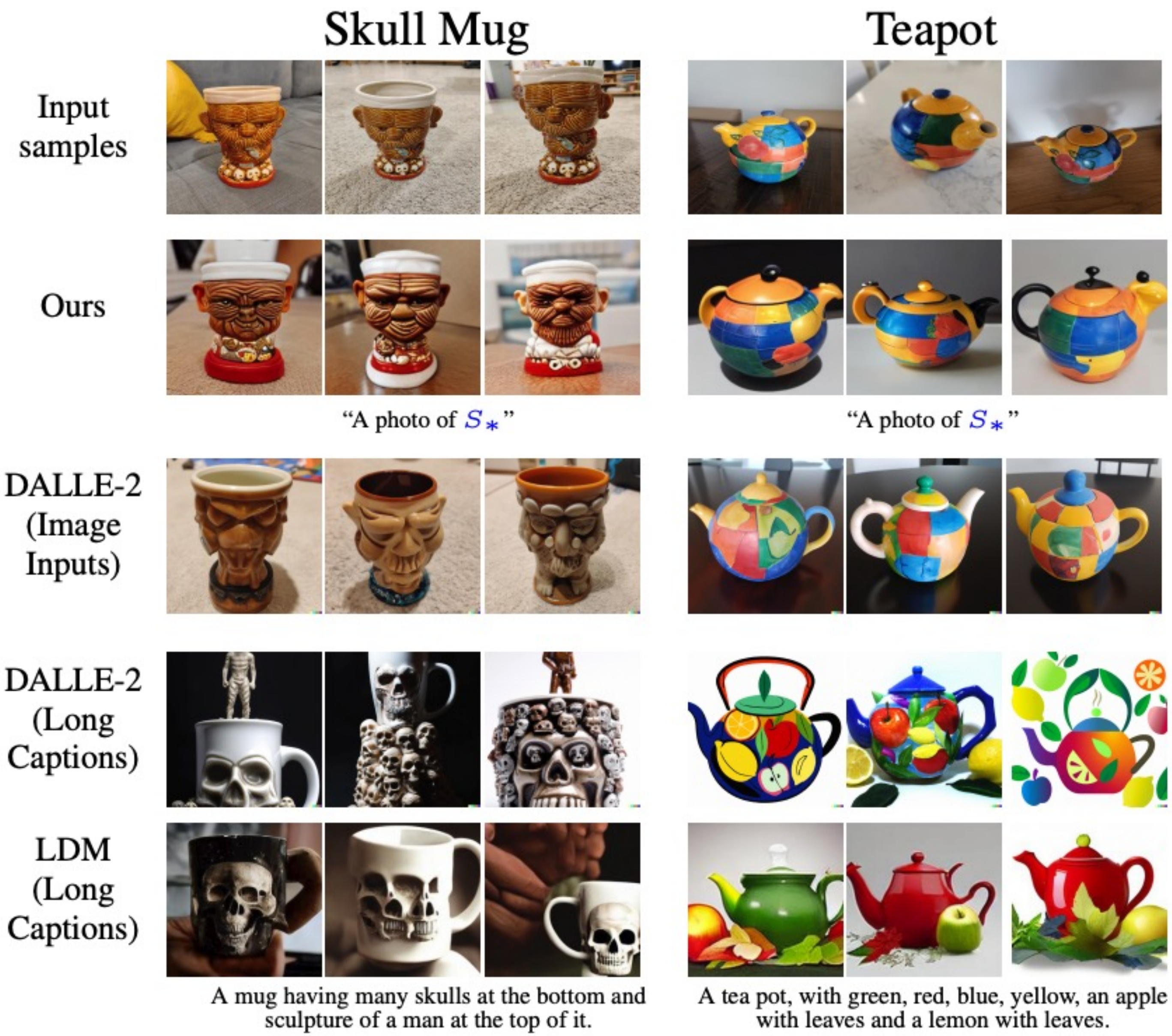
Figure 2: Outline of the text-embedding and inversion process. A string containing our placeholder word is first converted into tokens (*i.e.* word or sub-word indices in a dictionary). These tokens are converted to continuous vector representations (the "embeddings", v). Finally, the embedding vectors are transformed into a conditioning code $c_\theta(y)$ that guides the generation. We optimize the embedding vector v_* associated with our pseudo-word S_*, using a reconstruction objective.

R. Gal et al. An Image is Worth One Word: Personalizing Text-to-Image Generation using Textual Inversion. ICLR 2023

Textual inversion: Results

Textual inversion: Results

Textual inversion: Comparisons



Customizing DMs: DreamBooth

N. Ruiz et al. [DreamBooth: Fine Tuning Text-to-Image Diffusion Models for Subject-Driven Generation](). arXiv 2022

Customizing DMs: DreamBooth

Input Images

Image-guided, DALL-E2

Text-guided, Imagen

Ours

Prompt: "retro style yellow alarm clock with a white clock face and a yellow number three on the right part of the clock face in the jungle"

N. Ruiz et al. [DreamBooth: Fine Tuning Text-to-Image Diffusion Models for Subject-Driven Generation](). arXiv 2022

Customizing DMs: DreamBooth

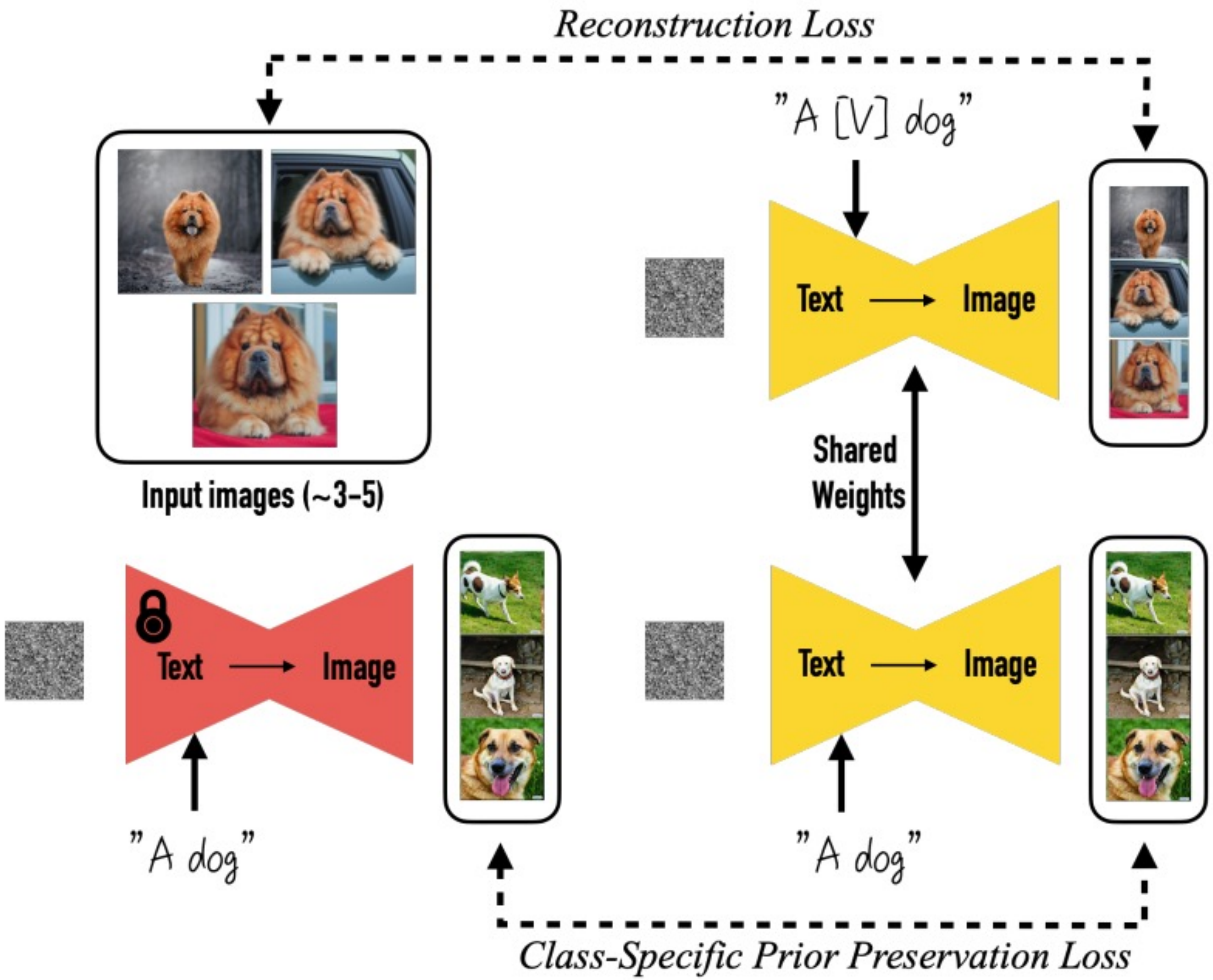
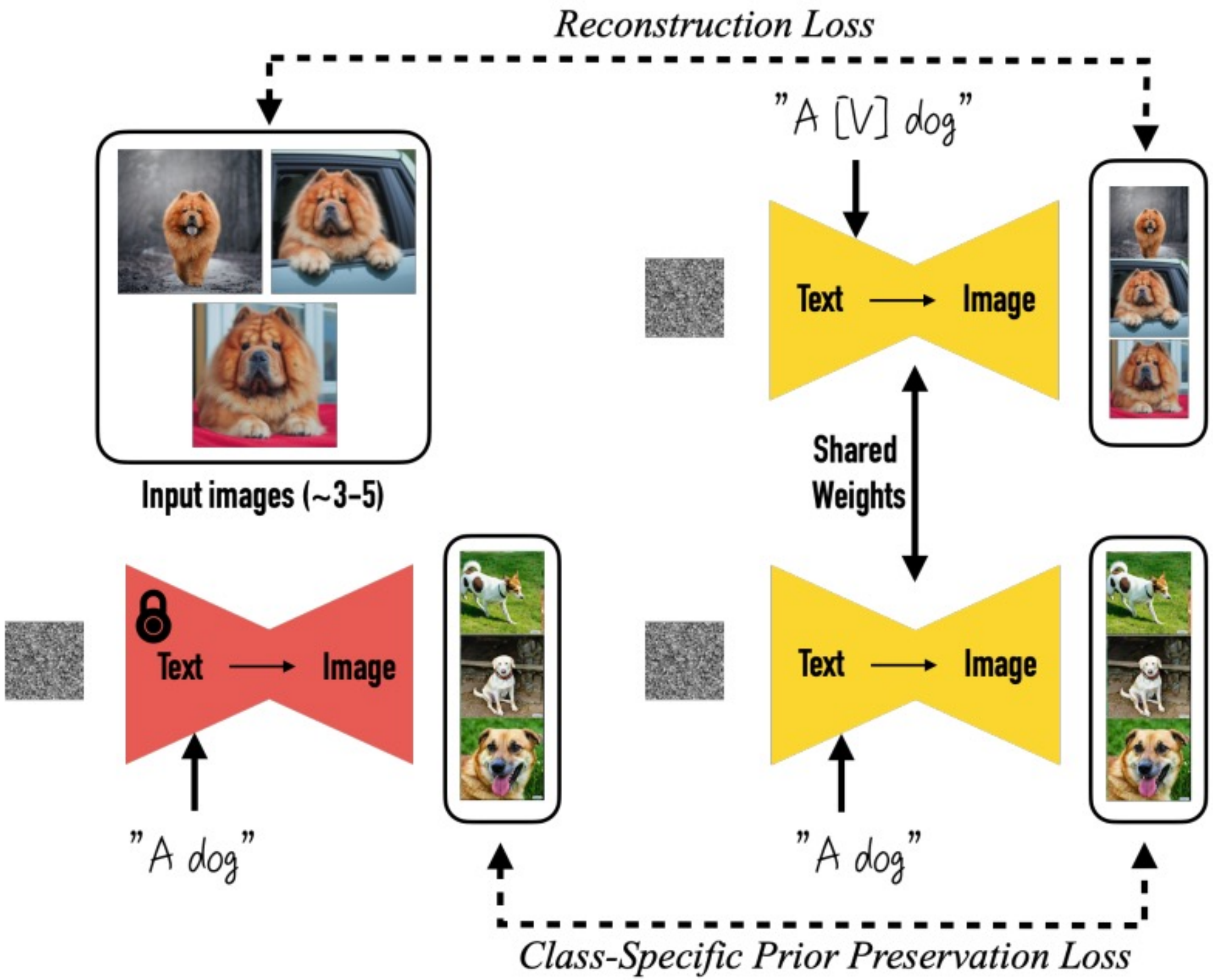


Figure 3. **Fine-tuning.** Given $\sim 3-5$ images of a subject we fine-tune a text-to-image diffusion model with the input images paired with a text prompt containing a unique identifier and the name of the class the subject belongs to (e.g., "A [V] dog"), in parallel, we apply a class-specific prior preservation loss, which leverages the semantic prior that the model has on the class and encourages it to generate diverse instances belong to the subject's class using the class name in a text prompt (e.g., "A dog").

N. Ruiz et al. [DreamBooth: Fine Tuning Text-to-Image Diffusion Models for Subject-Driven Generation](). arXiv 2022

Customizing DMs: DreamBooth



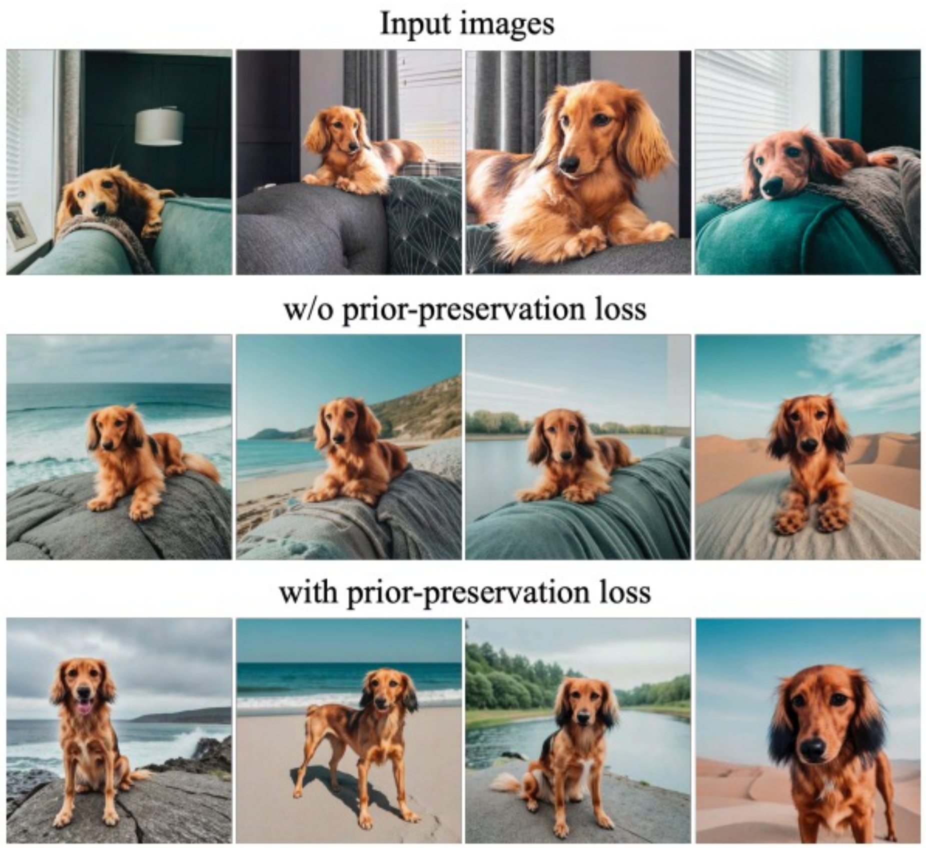
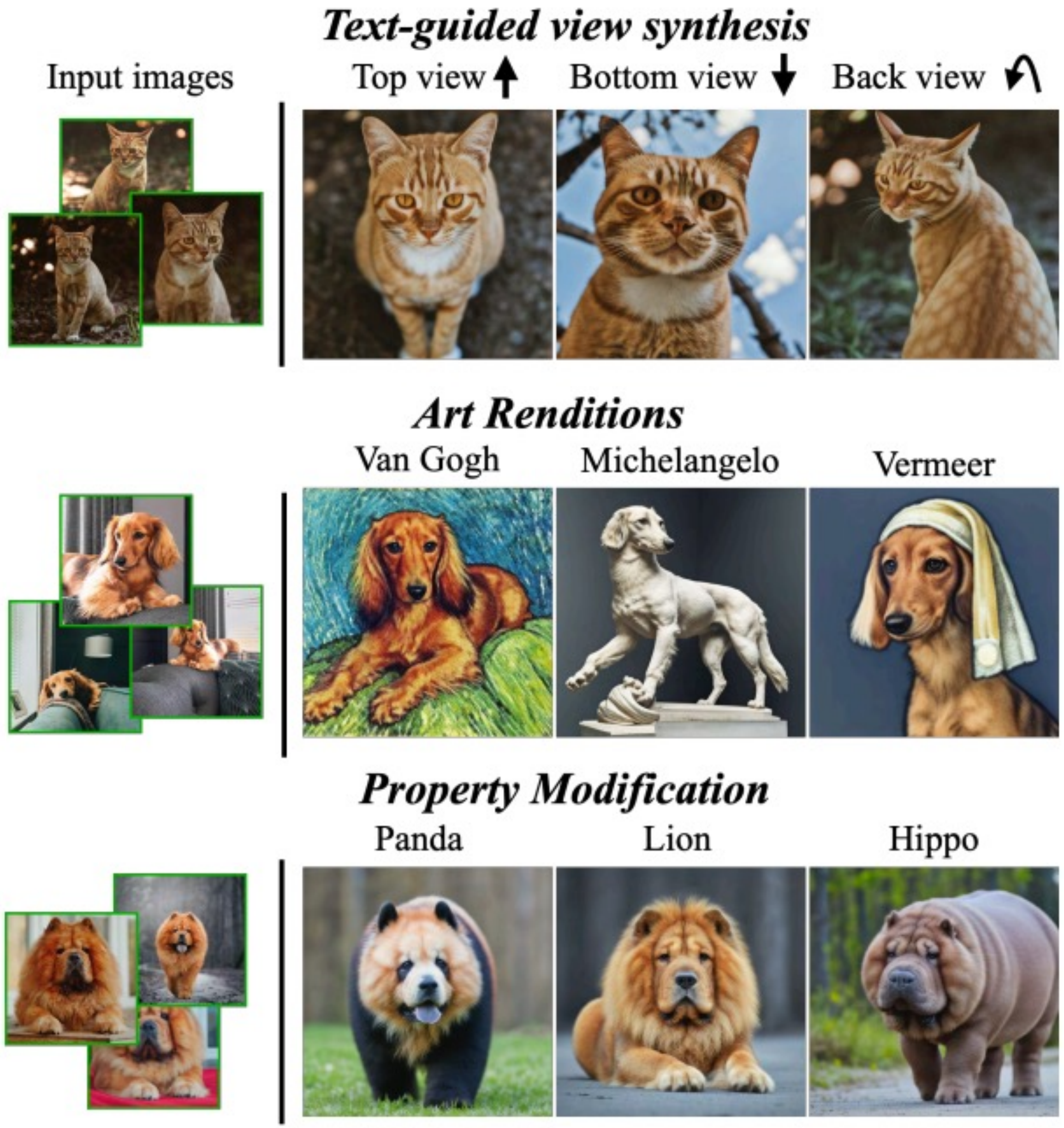


Figure 6. **Encouraging diversity with prior-preservation loss.** Naive fine-tuning can result in overfitting to input image context and subject appearance (e.g. pose). PPL acts as a regularizer that alleviates overfitting and encourages diversity, allowing for more pose variability and appearance diversity.

N. Ruiz et al. [DreamBooth: Fine Tuning Text-to-Image Diffusion Models for Subject-Driven Generation](). arXiv 2022

DreamBooth: Results

DreamBooth: Results



“a [painting/sculpture] of a [V] [class noun] in the style of [famous artist]”

“a cross of a [V] dog and a [target species]”

DreamBooth: Comparison with textual inversion

"a [V] vase in the snow" "a [V] vase on the beach" "a [V] vase in the jungle" "a [V] vase with Eiffel Tower in the background

Method	Subject Fidelity ↑	Prompt Fidelity ↑
DreamBooth (Stable Diffusion)	**68%**	**81%**
Textual Inversion (Stable Diffusion)	22%	12%
Undecided	10%	7%

DreamBooth: Limitations

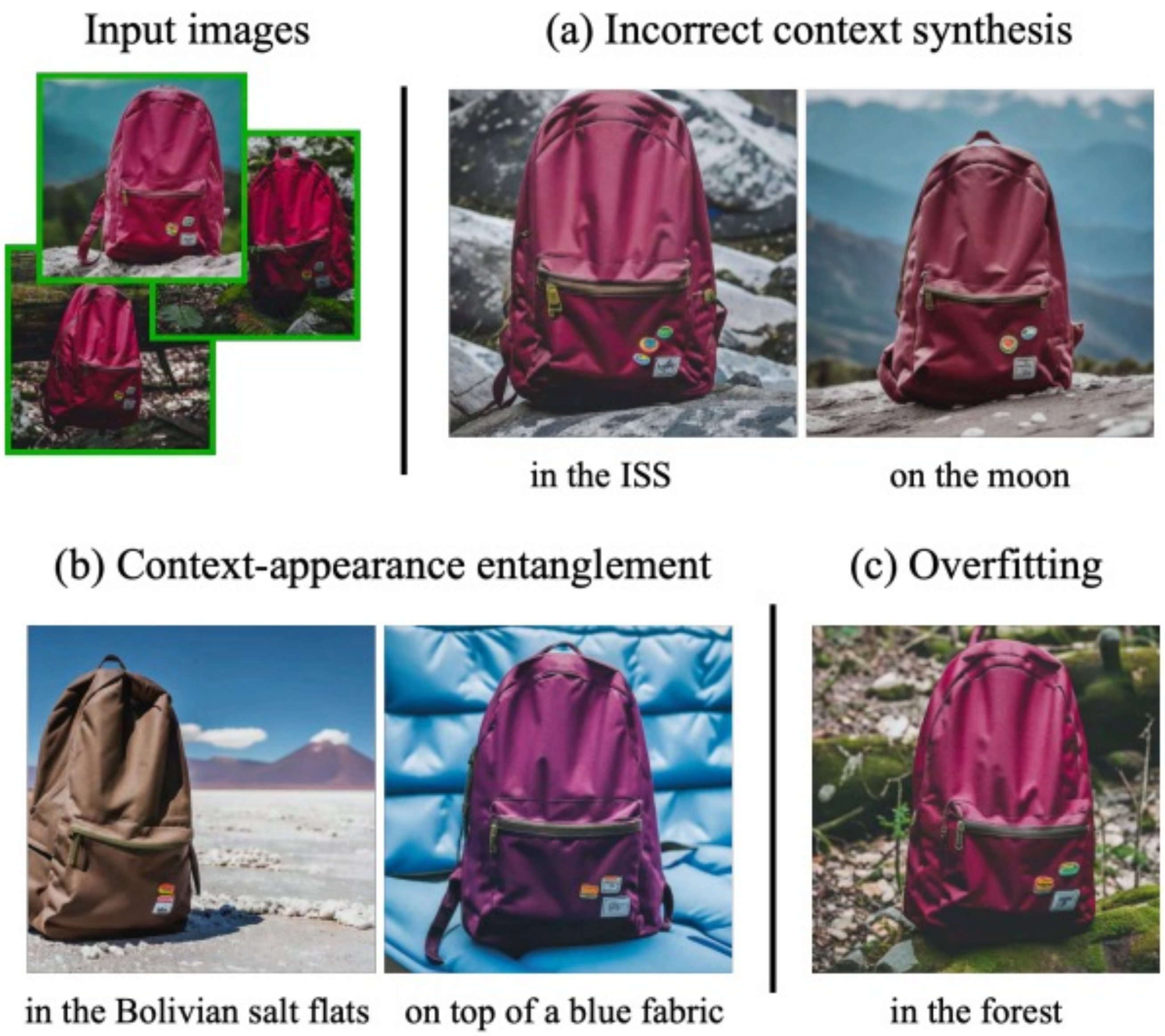
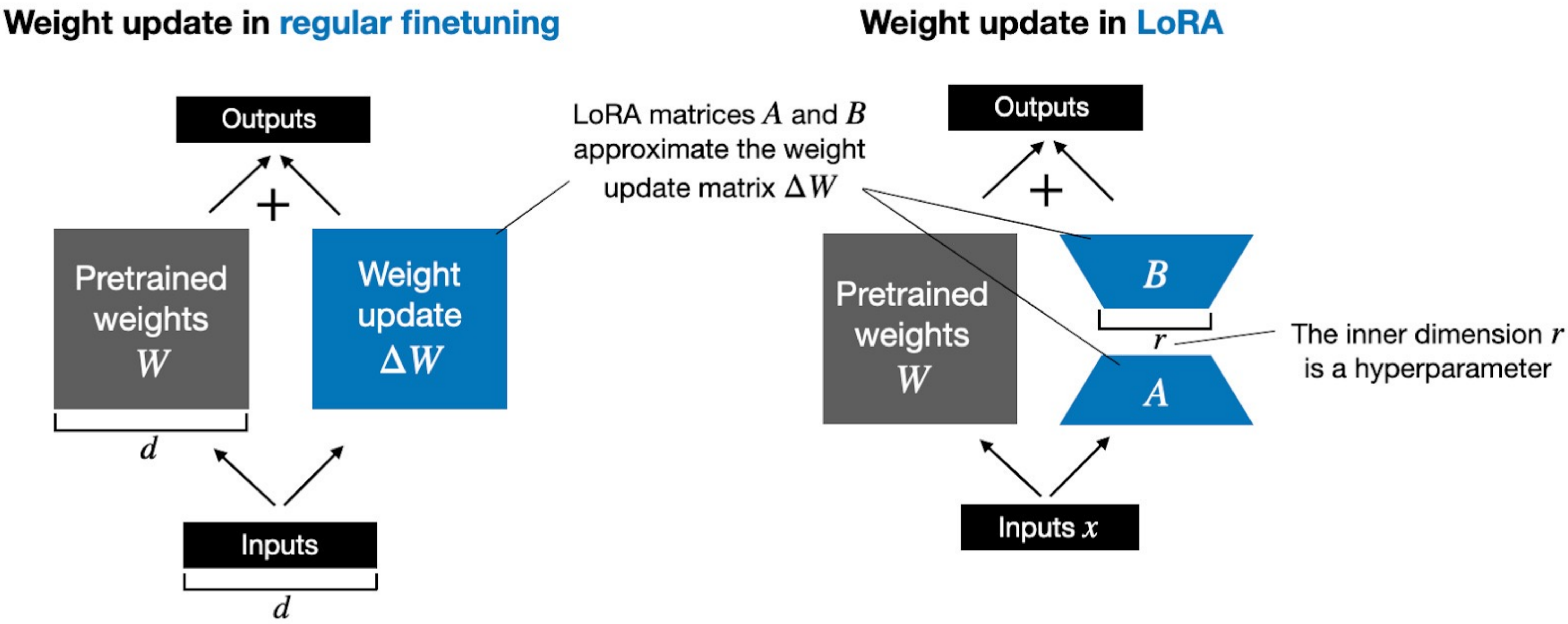


Figure 9. **Failure modes.** Given a rare prompted context the model might fail at generating the correct environment (a). It is possible for context and subject appearance to become entangled (b). Finally, it is possible for the model to overfit and generate images similar to the training set, especially if prompts reflect the original environment of the training set (c).

Efficient Customization: DreamBooth with LoRA



$\boldsymbol{A}$ is initialized with standard normal; $\boldsymbol{B}$ is initialized with zeros

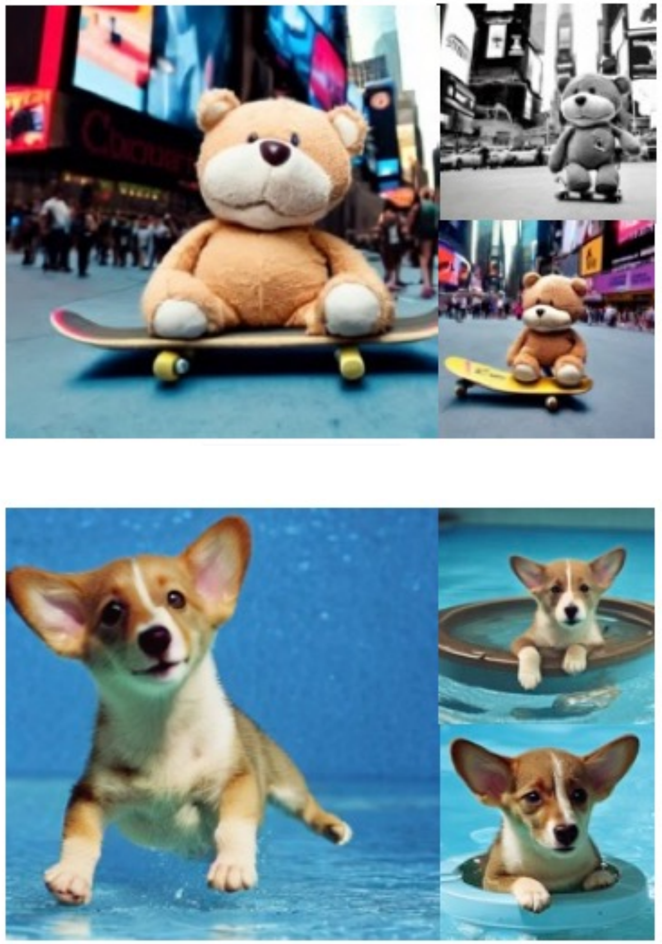
Hu et al., [LoRA: Low-Rank Adaptation of Large Language Models](), ICLR'22; https://magazine.sebastianraschka.com/p/practical-tips-for-finetuning-llms

LoRA DreamBooth: Results

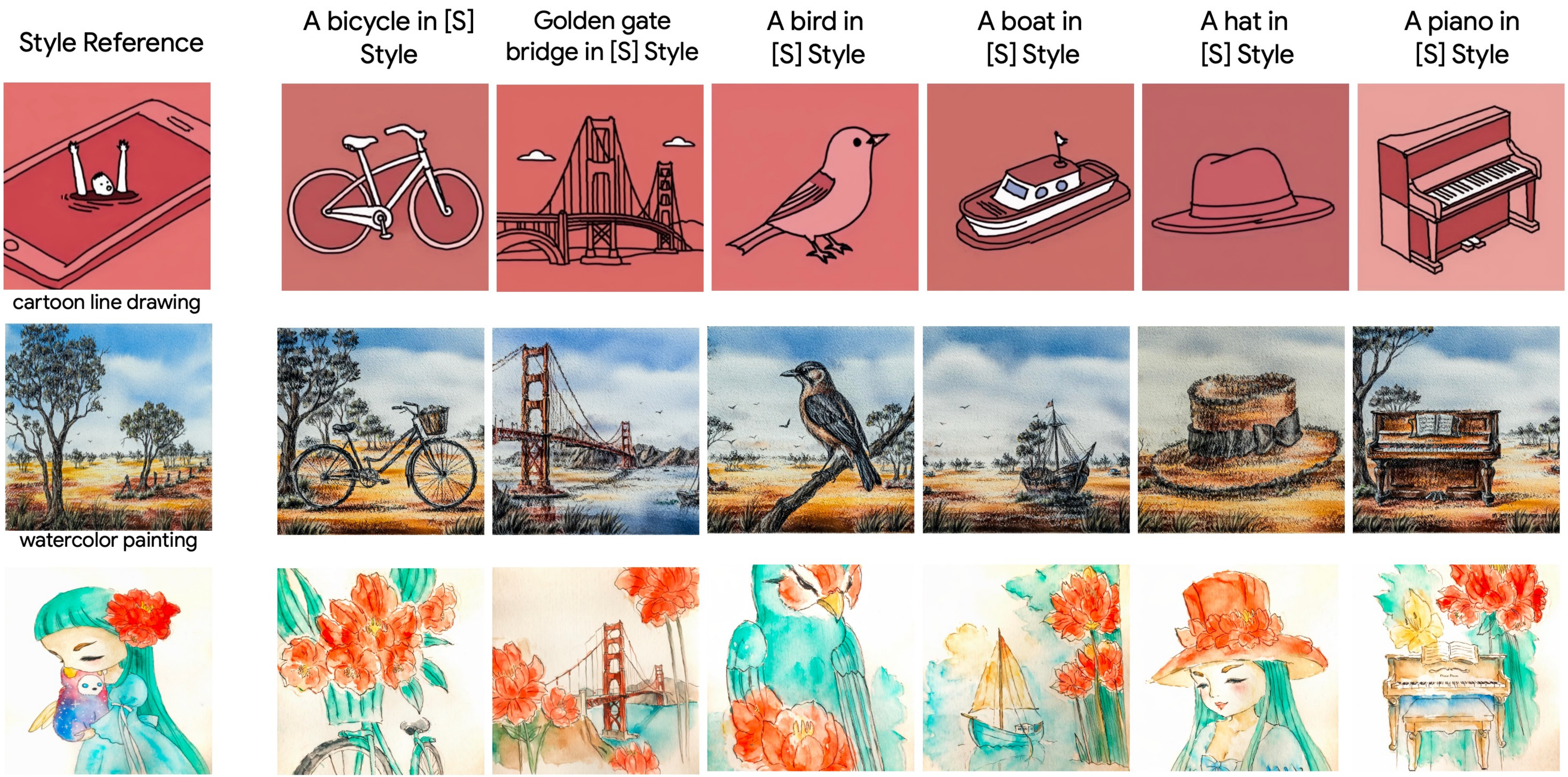
Input Images

LoRA DreamBooth (r=4)

DreamBooth



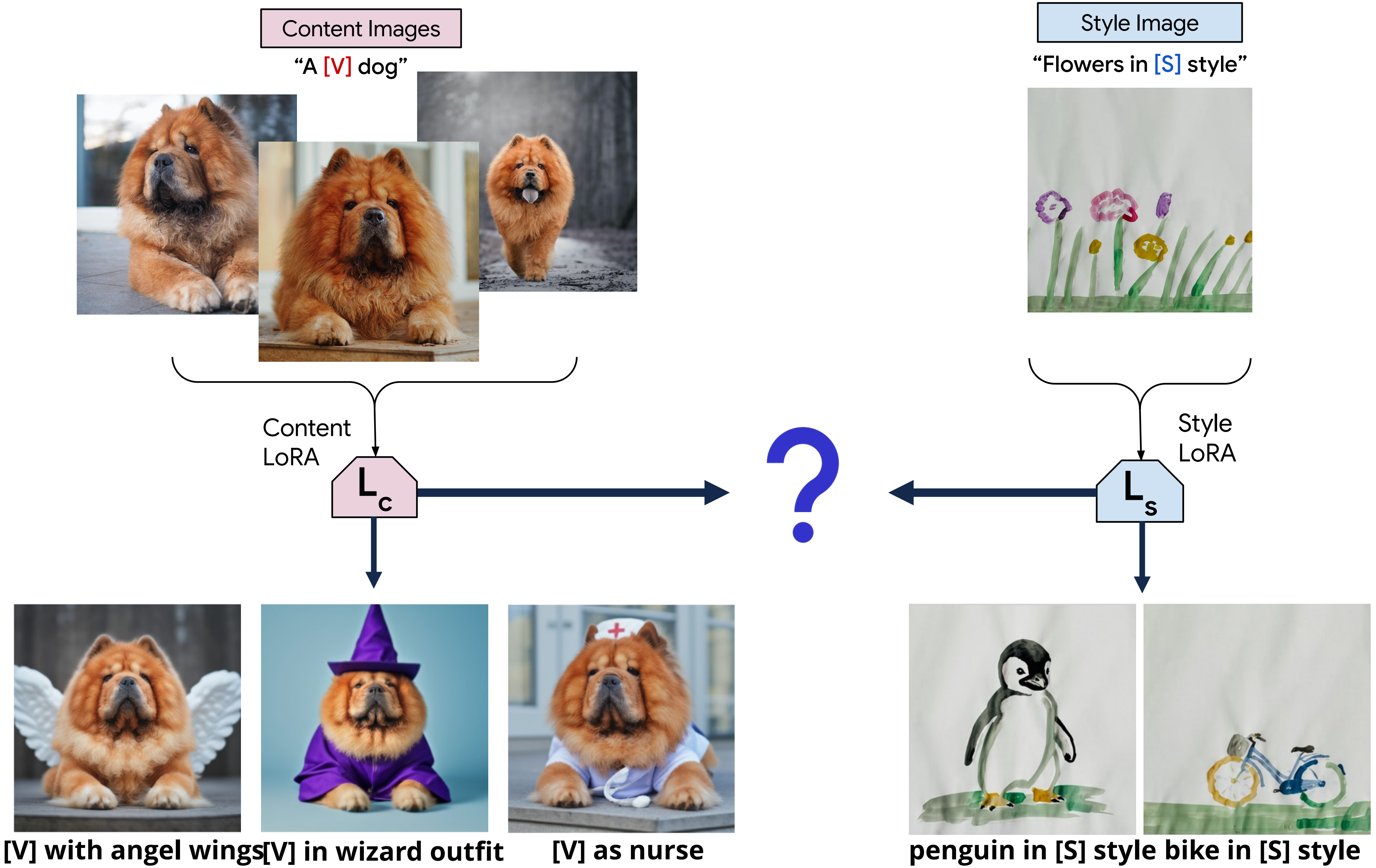
LoRA DreamBooth for Stylizations (on SDXL)



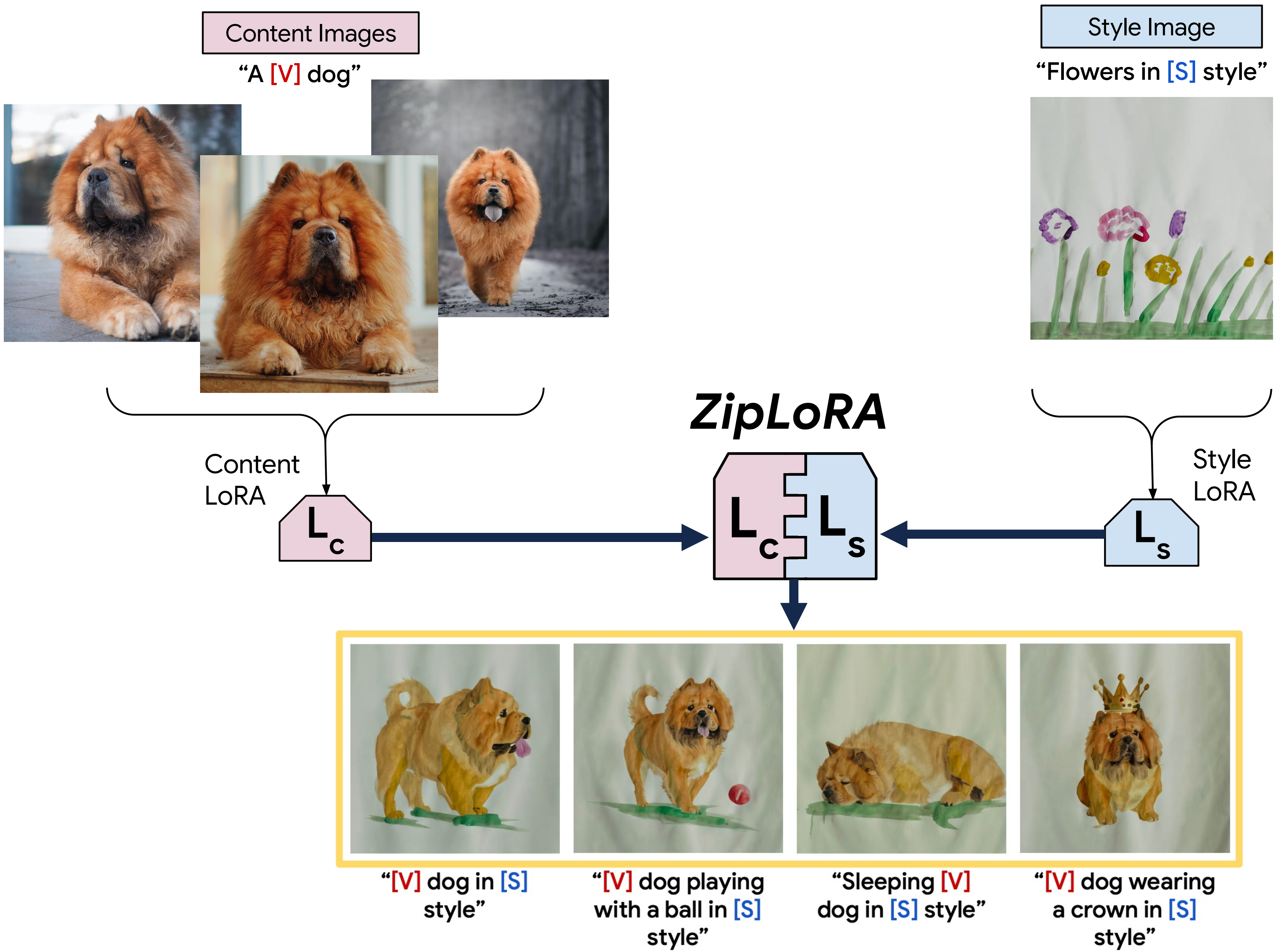
Stylizations obtained using DreamBooth on SDXL with LoRA

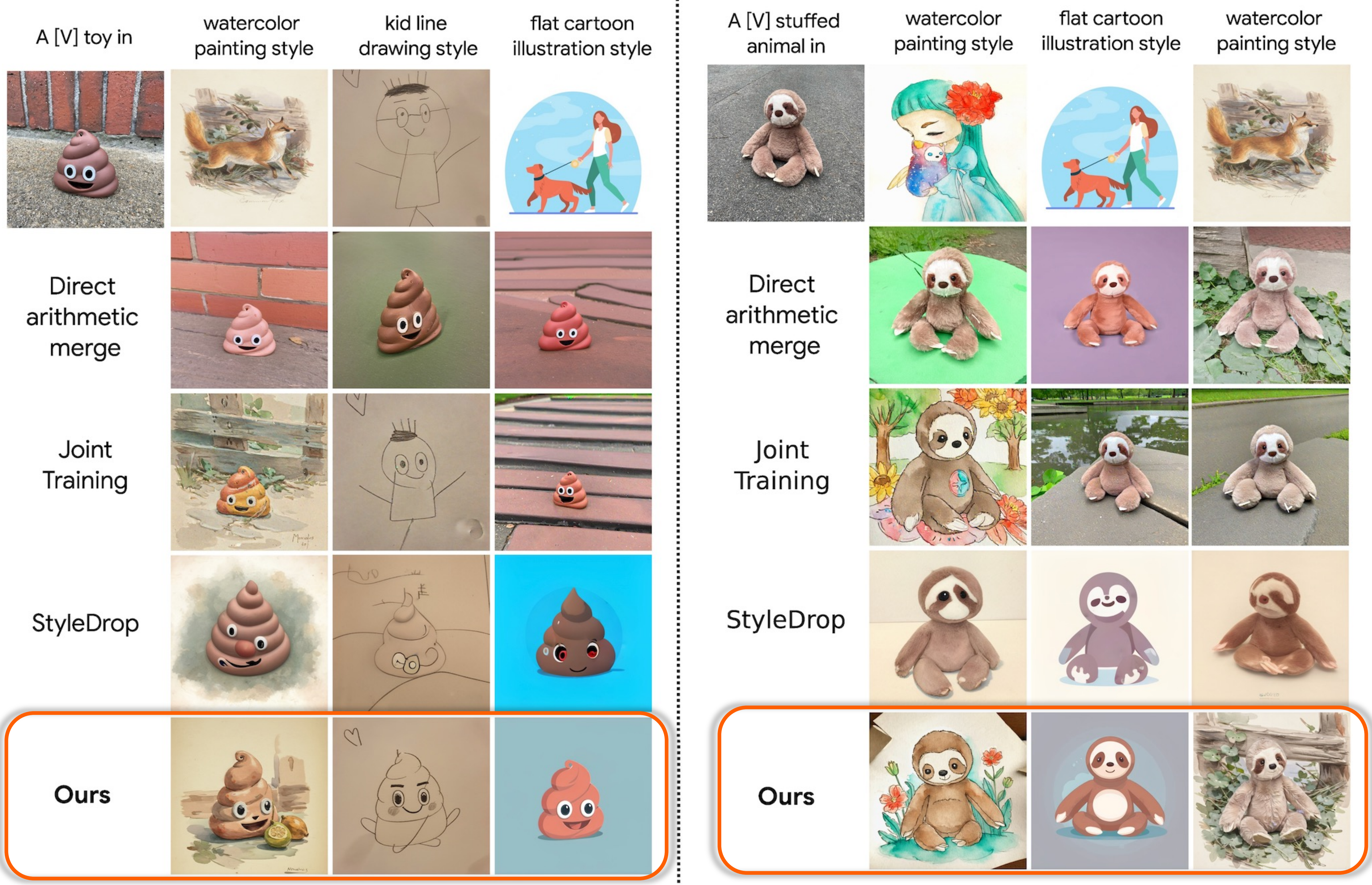
Shah et al., [ZipLoRA: Any Subject in Any Style by Effectively Merging LoRAs](), arxiv'23

Can we Merge Content and Style LoRAs?



Can we Merge Content and Style LoRAs?




A [V] toy in
watercolor painting style
kid line drawing style
flat cartoon illustration style
Direct arithmetic merge
Joint Training
StyleDrop
Ours
A [V] stuffed animal in
watercolor painting style
flat cartoon illustration style
watercolor painting style
Direct arithmetic merge
Joint Training
StyleDrop
Ours

ZipLoRA produces successful recontextualizations

Anti-DreamBooth

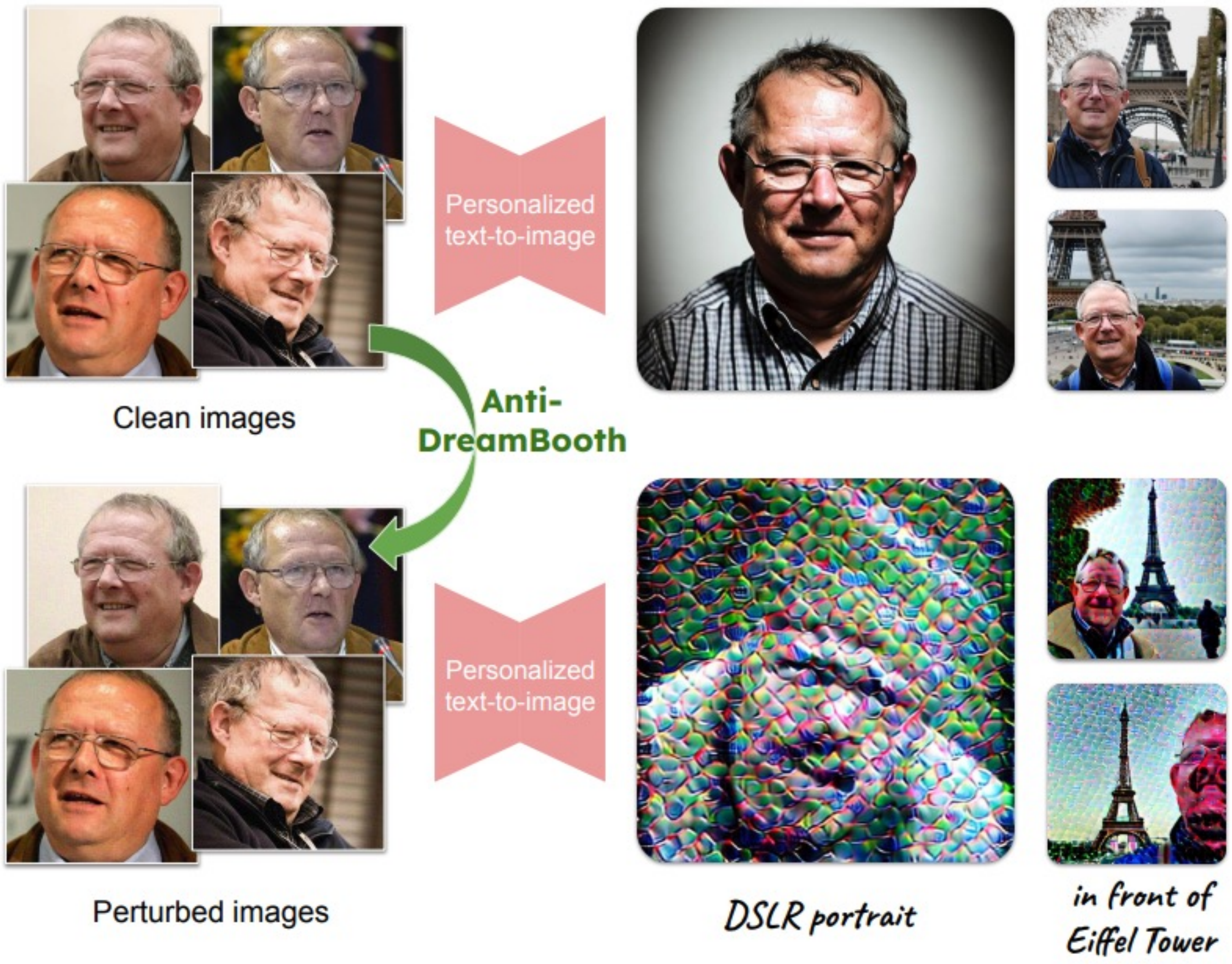


Figure 1: A malicious attacker can collect a user's images to train a personalized text-to-image generator for malicious purposes. Our system, called Anti-DreamBooth, applies imperceptible perturbations to the user's images before releasing, making any personalized generator trained on these images fail to produce usable images, protecting the user from that threat.

T. V. Le et al. [Anti-DreamBooth: Protecting users from personalized text-to-image synthesis](). arXiv 2023

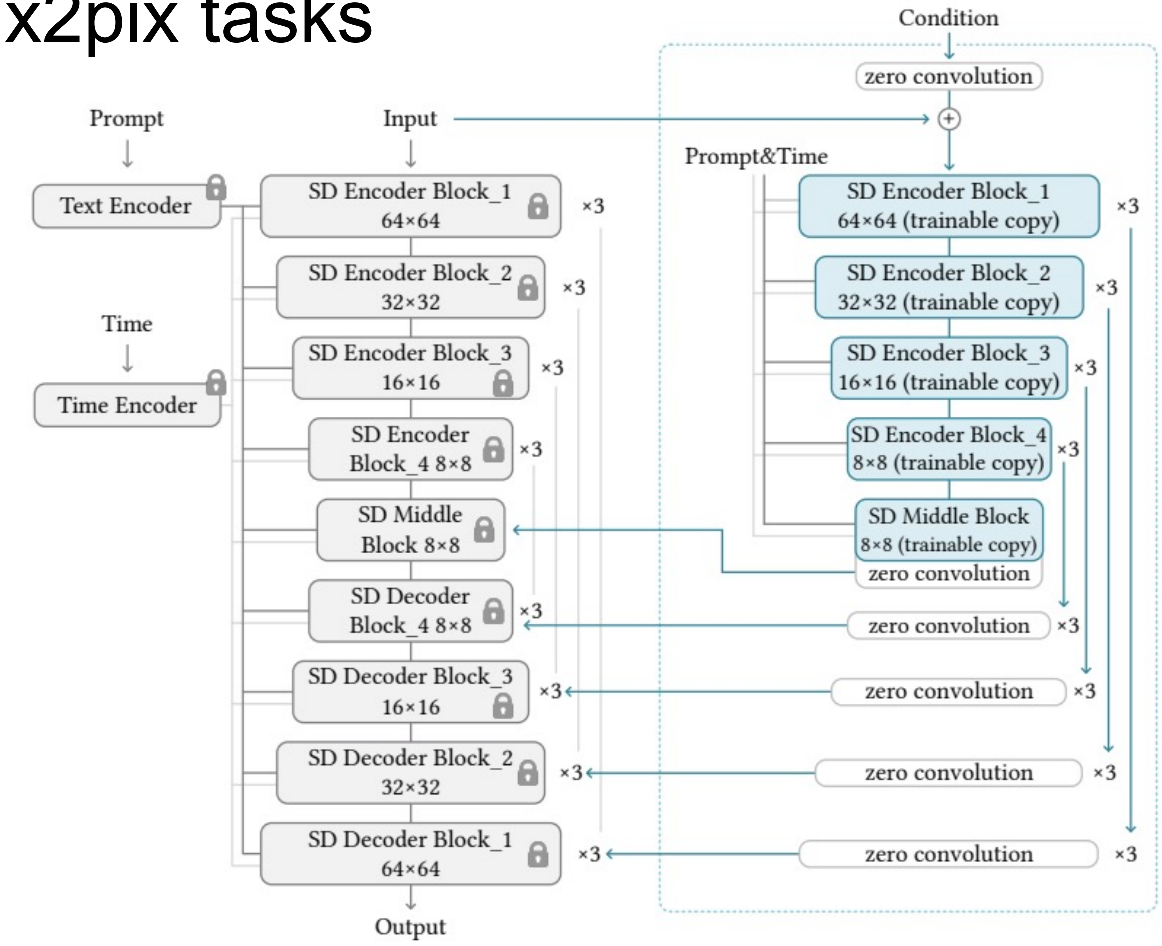
Outline

Part 3: Applications and Implementation; Ethical Issues

- Customizing Diffusion Models
  - Textual Inversion
  - DreamBooth
  - Low Rank Approximation (LoRA)
  - ZipLoRA
- ControlNet

ControlNet

- Add a trainable “wrapper” around a pre-trained DM to fine-tune it for pix2pix tasks



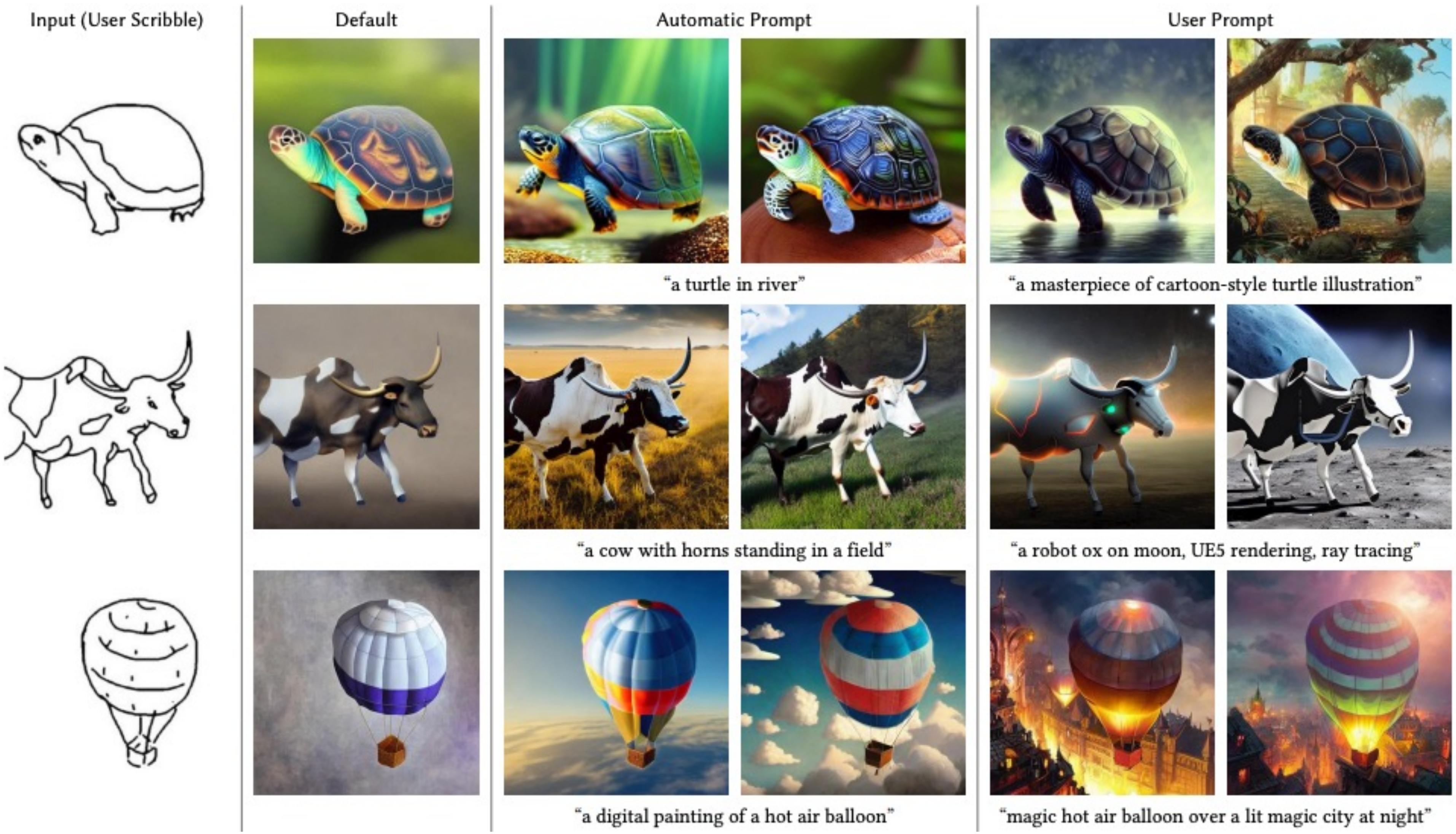
(a) Stable Diffusion (b) ControlNet

L. Zhang and M. Agrawala. [Adding Conditional Control to Text-to-Image Diffusion Models](). arXiv 2023

ControlNet

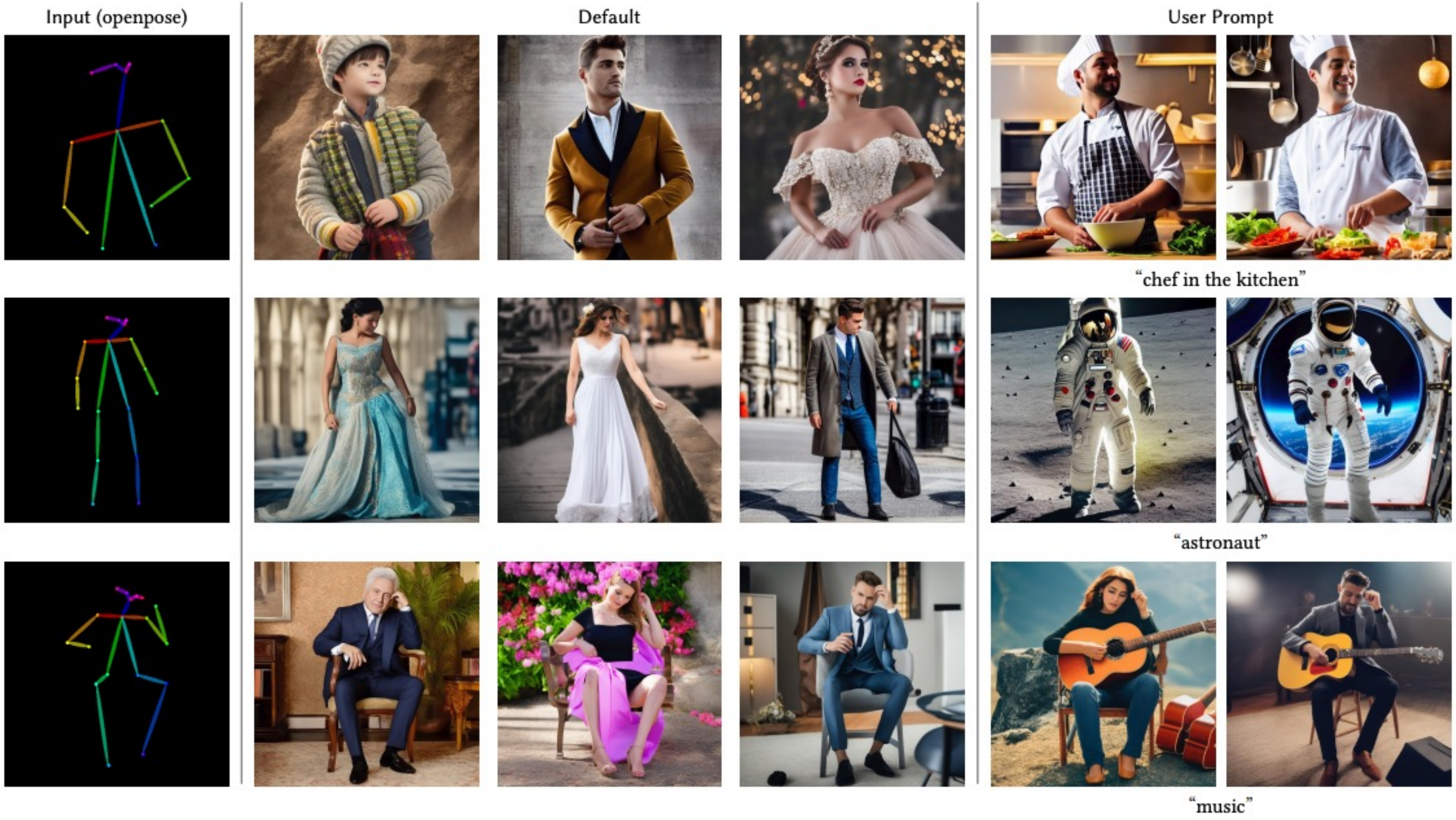
L. Zhang and M. Agrawala. [Adding Conditional Control to Text-to-Image Diffusion Models](). arXiv 2023

ControlNet



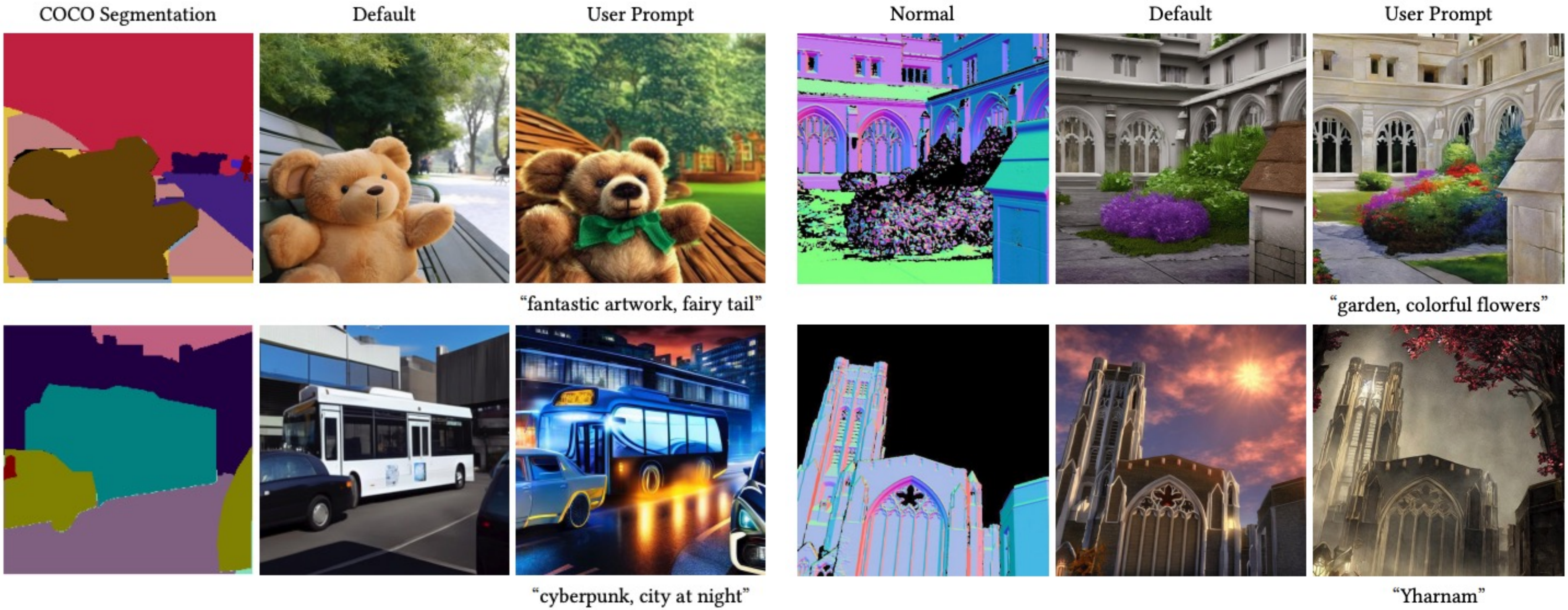
L. Zhang and M. Agrawala. [Adding Conditional Control to Text-to-Image Diffusion Models](). arXiv 2023

ControlNet



L. Zhang and M. Agrawala. Adding Conditional Control to Text-to-Image Diffusion Models. arXiv 2023

ControlNet



L. Zhang and M. Agrawala. [Adding Conditional Control to Text-to-Image Diffusion Models](). arXiv 2023

Outline

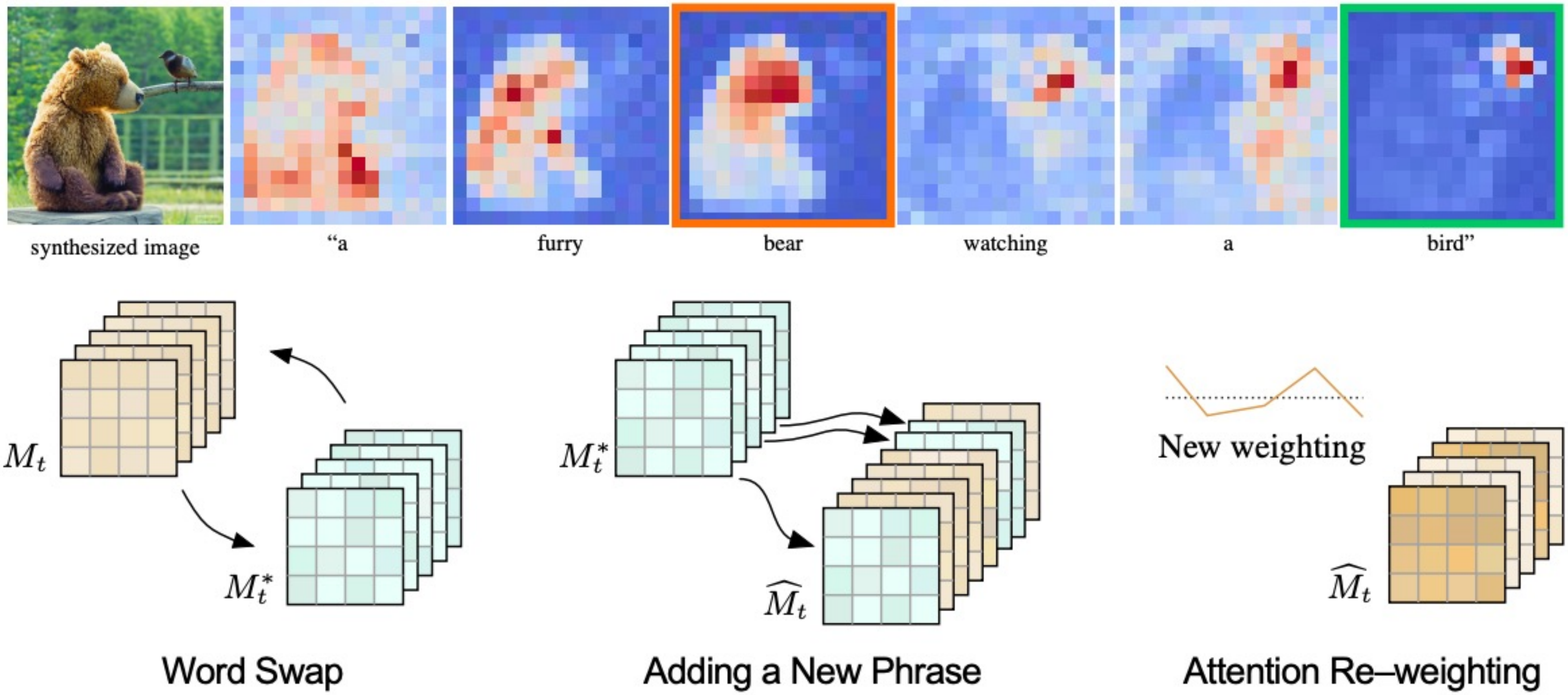
Part 3: Applications and Implementation; Ethical Issues

- Customizing Diffusion Models
  - Textual Inversion
  - DreamBooth
  - Low Rank Approximation (LoRA)
  - ZipLoRA
- ControlNet
- Prompt-to-Prompt and InstructPix2Pix

Prompt-to-Prompt Image Editing

A. Hertz et al. [Prompt-to-Prompt Image Editing with Cross Attention Control](). arXiv 2022

Prompt-to-Prompt Image Editing



Prompt-to-Prompt Image Editing

InstructPix2Pix

Figure 1. Given **an image** and **an instruction** for how to edit that image, our model performs the appropriate edit. Our model does not require full descriptions for the input or output image, and edits images in the forward pass without per-example inversion or fine-tuning.

T. Brooks, A. Holynski, and A. Efros. [InstructPix2Pix: Learning to Follow Image Editing Instructions](). CVPR 2023

InstructPix2Pix

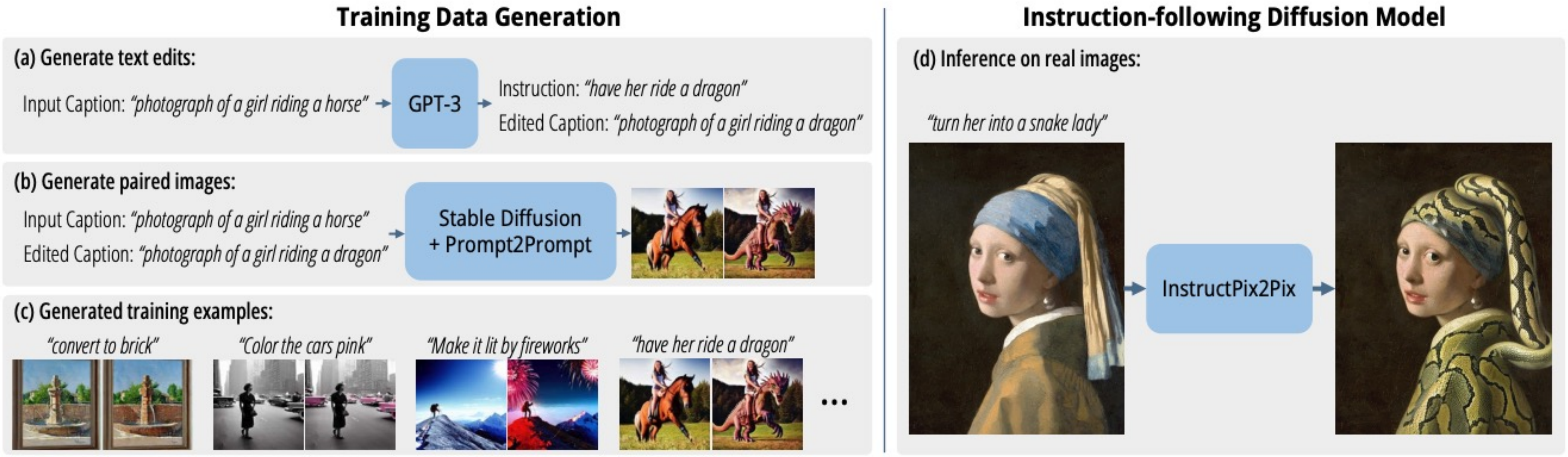


Figure 2. Our method consists of two parts: generating an image editing dataset, and training a diffusion model on that dataset. (a) We first use a finetuned GPT-3 to generate instructions and edited captions. (b) We then use StableDiffusion [52] in combination with Prompt-to-Prompt [17] to generate pairs of images from pairs of captions. We use this procedure to create a dataset (c) of over 450,000 training examples. (d) Finally, our InstructPix2Pix diffusion model is trained on our generated data to edit images from instructions. At inference time, our model generalizes to edit real images from human-written instructions.

T. Brooks, A. Holynski, and A. Efros. [InstructPix2Pix: Learning to Follow Image Editing Instructions](). CVPR 2023

InstructPix2Pix

- Fine-tuning GPT-3:

	Input LAION caption	Edit instruction	Edited caption
Human-written (700 edits)	*Yefim Volkov, Misty Morning*	*make it afternoon*	*Yefim Volkov, Misty Afternoon*
	girl with horse at sunset	*change the background to a city*	*girl with horse at sunset in front of city*
	painting-of-forest-and-pond	*Without the water.*	*painting-of-forest*
	...	...	...
GPT-3 generated (>450,000 edits)	*Alex Hill, Original oil painting on canvas, Moonlight Bay*	*in the style of a coloring book*	*Alex Hill, Original coloring book illustration, Moonlight Bay*
	The great elf city of Rivendell, sitting atop a waterfall as cascades of water spill around it	*Add a giant red dragon*	*The great elf city of Rivendell, sitting atop a waterfall as cascades of water spill around it with a giant red dragon flying overhead*
	Kate Hudson arriving at the Golden Globes 2015	*make her look like a zombie*	*Zombie Kate Hudson arriving at the Golden Globes 2015*
	...	...	...

Table 1. We label a small text dataset, finetune GPT-3, and use that finetuned model to generate a large dataset of text triplets. As the input caption for both the labeled and generated examples, we use real image captions from LAION. Highlighted text is generated by GPT-3.

InstructPix2Pix

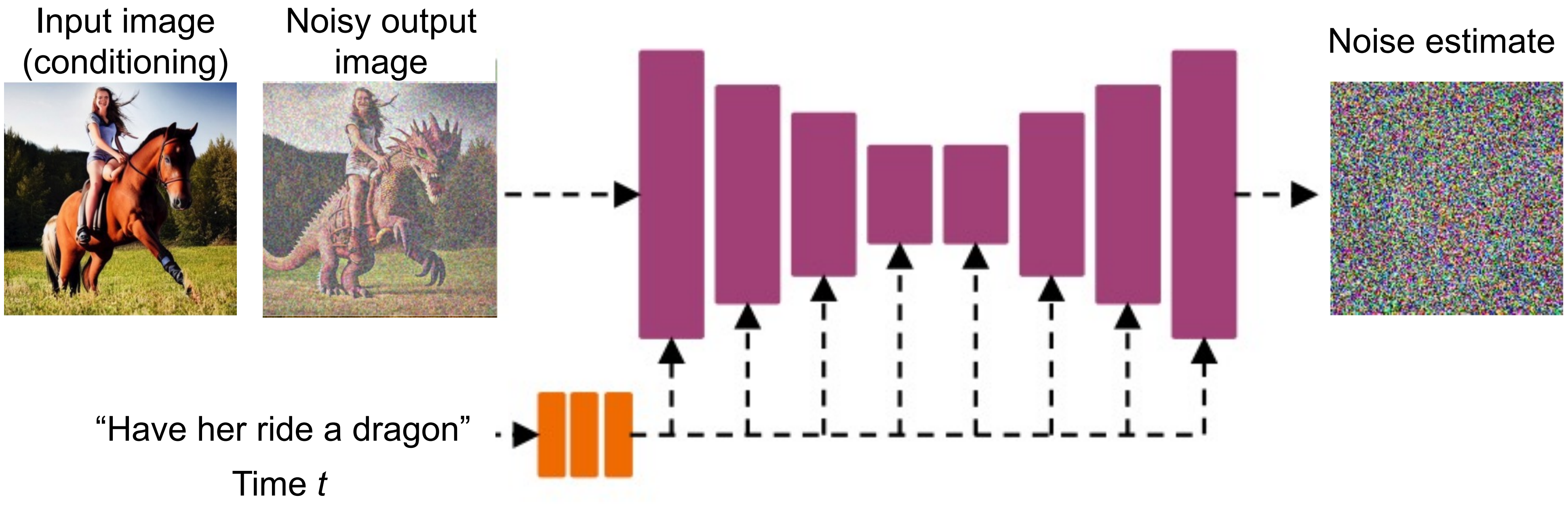
- Generating input-output image pairs:

(a) Without Prompt-to-Prompt. (b) With Prompt-to-Prompt.

Figure 3. Pair of images generated using StableDiffusion [52] with and without Prompt-to-Prompt [17]. For both, the corresponding captions are *"photograph of a girl riding a horse"* and *"photograph of a girl riding a dragon"*.

InstructPix2Pix

- Fine-tuning a DM for image-to-image translation:



InstructPix2Pix: Results

Input | "Make it a Modigliani painting" | "Make it a Miro painting" | "Make it an Egyptian sculpture" | "Make it a marble roman sculpture"

Figure 5. *Mona Lisa* transformed into various artistic mediums.

Input | "Put them in outer space" | "Turn the humans into robots"

Figure 6. *The Creation of Adam* with new context and subjects (generated at 768 resolution).

InstructPix2Pix: Results

"What if she were in an anime?" *"Make her terrifying"* *"Make her more sad"* *"Make her James Bond"* *"Turn her into Dwayne The Rock Johnson"*

InstructPix2Pix: Results

"Make it Paris" "Make it Hong Kong" "Make it Manhattan" "Make it Prague"

"Make it evening" "Put them on roller skates" "Turn this into 1900s" "Make it underwater"

"Make it Minecraft" "Turn this into the space age" "Make them into Alexander Calder sculptures" "Make it a Claymation"

InstructPix2Pix: Failure cases

"Zoom into the image" *"Move it to Mars"* *"Color the tie blue"* *"Have the people swap places"*

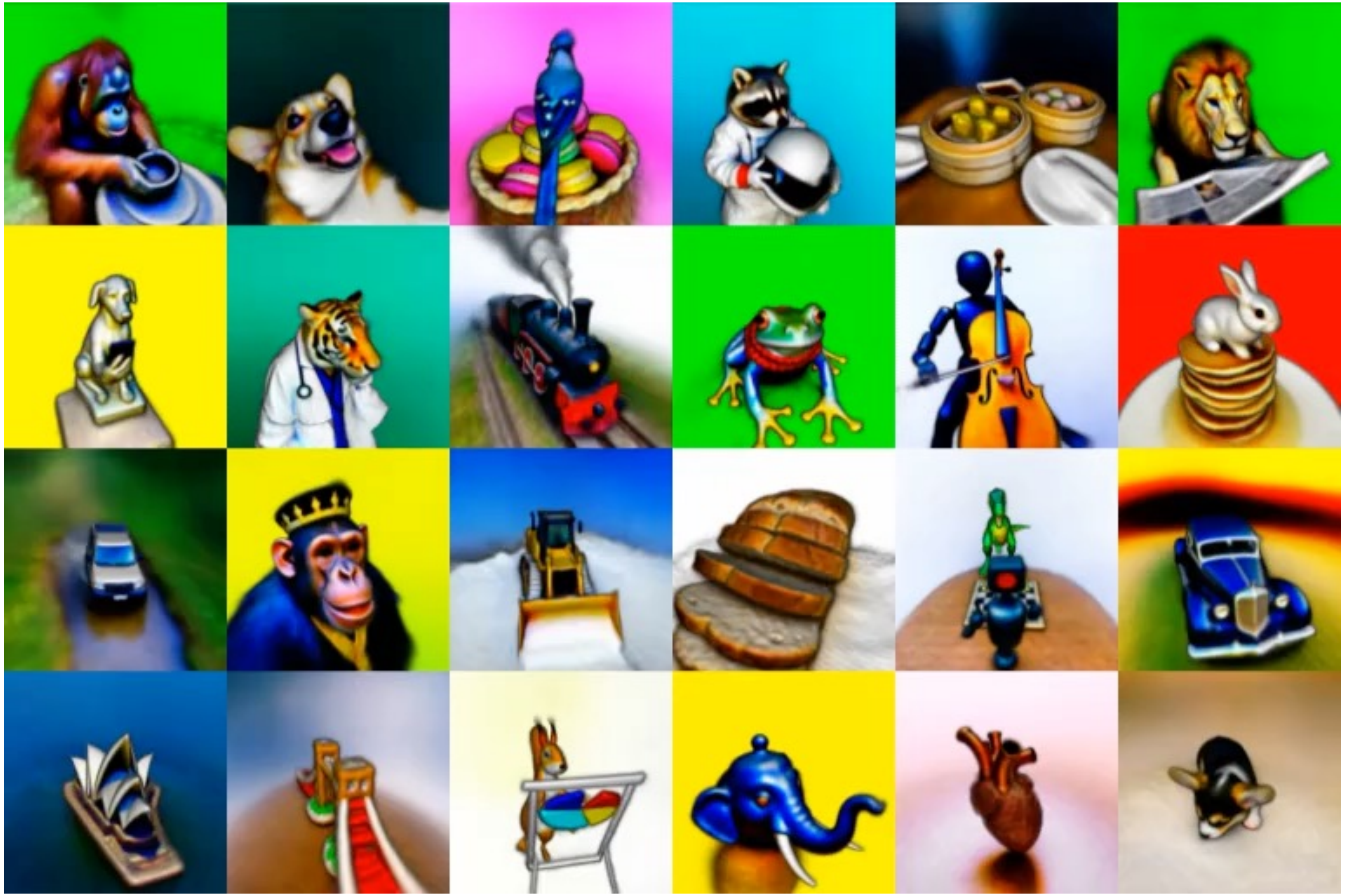
Figure 13. Failure cases. Left to right: our model is not capable of performing viewpoint changes, can make undesired excessive changes to the image, can sometimes fail to isolate the specified object, and has difficulty reorganizing or swapping objects with each other.

Outline

Part 3: Applications and Implementation; Ethical Issues

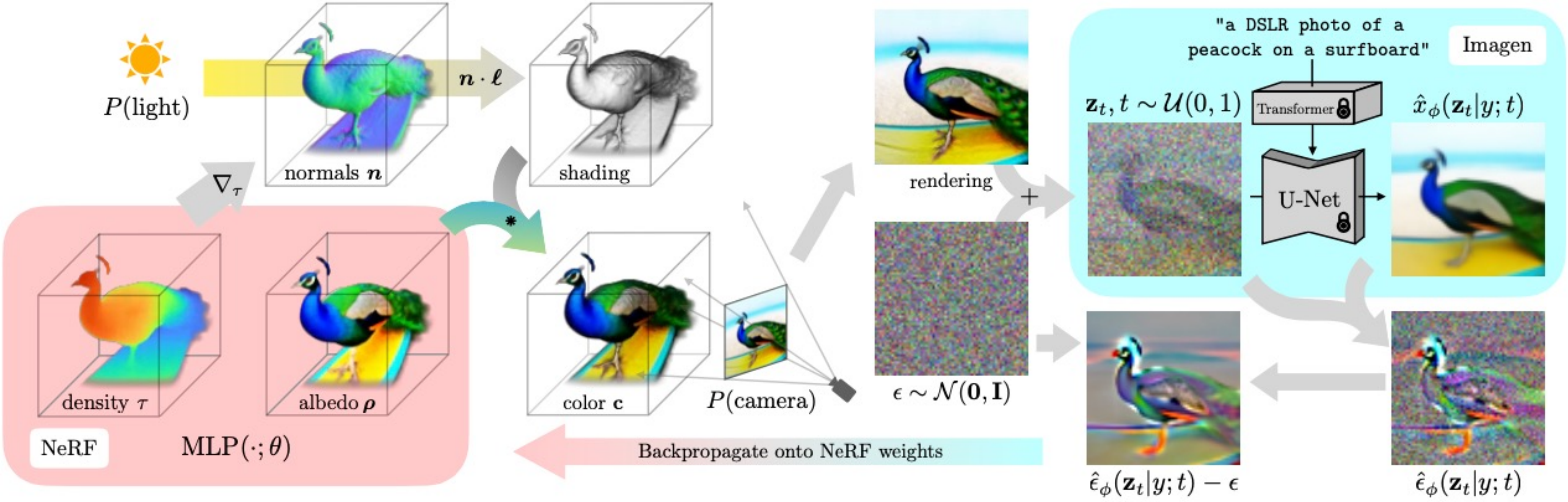
- Customizing Diffusion Models
  - Textual Inversion
  - DreamBooth
  - Low Rank Approximation (LoRA)
  - ZipLoRA
- ControlNet
- Prompt-to-Prompt
- InstructPix2Pix
- DreamFusion

Connecting 2D to 3D: DreamFusion



B. Poole, A. Jain, J. Barron, B. Mildenhall. [DreamFusion: Text-to-3D using 2D Diffusion](). arXiv 2022

Connecting 2D to 3D: DreamFusion



B. Poole, A. Jain, J. Barron, B. Mildenhall. [DreamFusion: Text-to-3D using 2D Diffusion](). arXiv 2022

Outline

Part 3: Applications and Implementation; Ethical Issues

- Customizing Diffusion Models
  - Textual Inversion
  - DreamBooth
  - Low Rank Approximation (LoRA)
  - ZipLoRA
- ControlNet
- Prompt-to-Prompt
- InstructPix2Pix
- DreamFusion
- Working with Diffusion Models: Implementation aspects

Working with Diffusion Models

- Fast moving area with new research coming in everyday
- Variety of pre-trained, open-source models available
- Working directly with the implementations and codebases helps!
- Starting point: diffusers library by huggingface
- Popular open-source models:
  - Stable Diffusion
  - SDXL
  - SD3 (recently announced)
  - Deep-Floyd IF (open-source alternative of Imagen)
- Some useful websites
  - Huggingface, Reddit threads on Stable Diffusion; Civit.ai; publicprompts.art

Outline

Part 3: Applications and Implementation; Ethical Issues

- Customizing Diffusion Models
  - Textual Inversion
  - DreamBooth
  - Low Rank Approximation (LoRA)
  - ZipLoRA
- ControlNet
- Prompt-to-Prompt
- InstructPix2Pix
- DreamFusion
- Working with Diffusion Models: Implementation aspects
- Societal, ethical, and legal issues

Societal, ethical, and legal issues

- Closed or open?
- Safe or unsafe?
- Potential for generating DeepFakes and misinformation
- Dataset image rights
- Artists’ rights
- The nature of creativity

In the news

ARTIFICIAL INTELLIGENCE / TECH / LAW

Getty Images is suing the creators of AI art tool Stable Diffusion for scraping its content

An image created by Stable Diffusion showing a recreation of Getty Images' watermark. Image: The Verge / Stable Diffusion

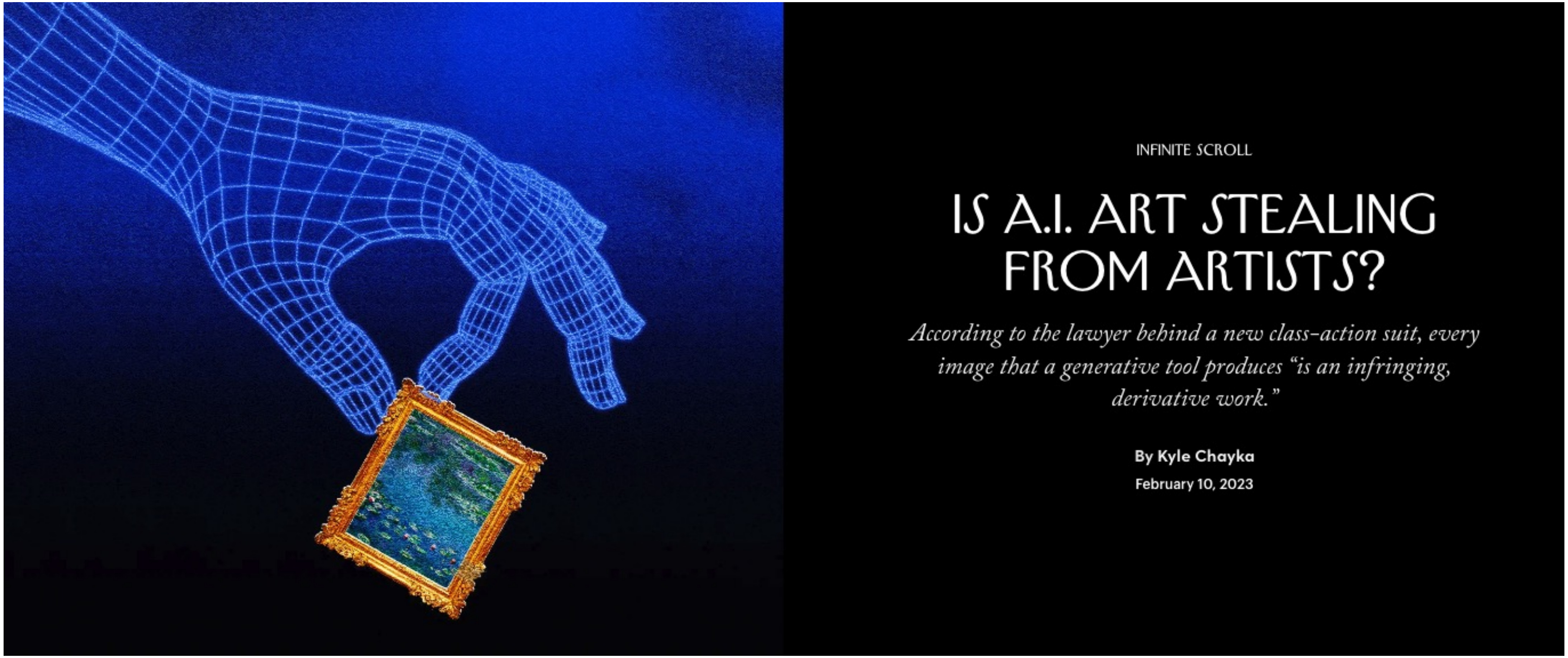
/ Getty Images claims Stability AI 'unlawfully' scraped millions of images from its site. It's a significant escalation in the developing legal battles between generative AI firms and content creators.

By JAMES VINCENT

Jan 17, 2023, 4:30 AM CST | 18 Comments / 18 New

https://www.theverge.com/2023/1/17/23558516/ai-art-copyright-stable-diffusion-getty-images-lawsuit

In the news



https://www.newyorker.com/culture/infinite-scroll/is-ai-art-stealing-from-artists

In the news

Fake Trump arrest photos: How to spot an AI-generated image

This image looks realistic, but take a closer look at Trump's right arm and neck

https://www.bbc.com/news/world-us-canada-65069316

In the news

Midjourney Bans AI Images of Chinese President Xi Jinping

APR 03, 2023 MATT GROWCOOT

https://petapixel.com/2023/04/03/midjourney-bans-ai-images-of-chinese-president-xi-jinping/