

Editing in text2img models

Drobyshevskiy Ilya

Outline:

1. What is editing in text2img model?
2. Options for solving this problem
3. Prompt2Prompt
4. DreamBooth

What is editing in text2img models?

We have:

I – original image

\mathcal{P} – prompt for current image

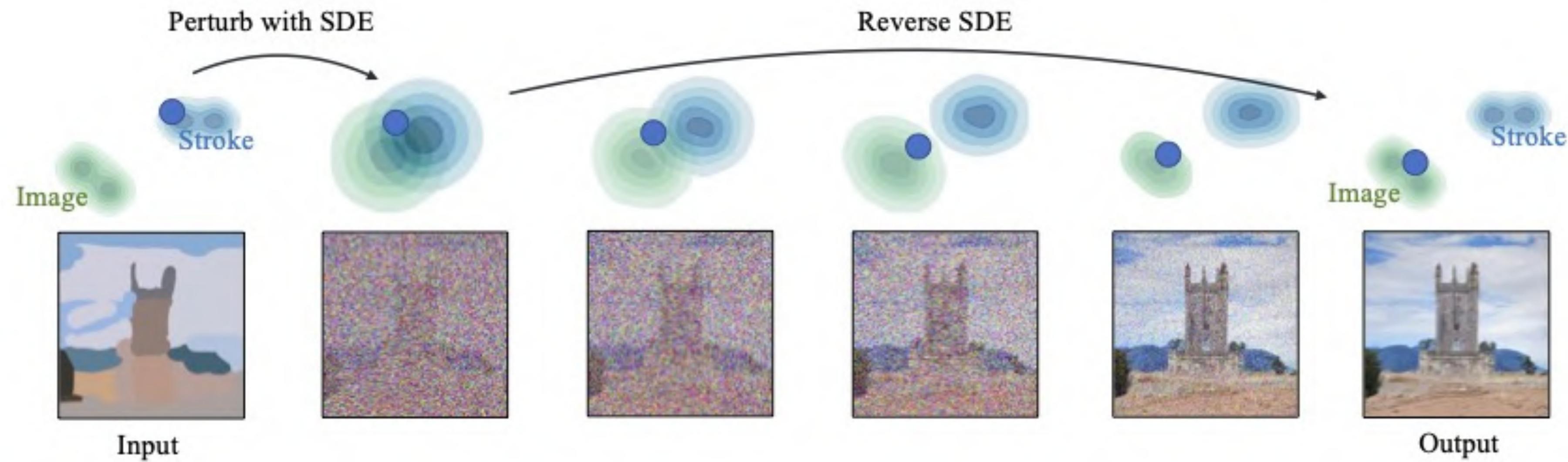
\mathcal{P}^* – new prompt, s.t. $\mathcal{P} \cap \mathcal{P}^* \neq \emptyset$

Our aim is to generate I^* – new image based on \mathcal{P}^* and similar with original image I .

Some options

- Add some noise and denoise with a new prompt
- Same as above but with a mask
- Cross-attention Control
- Fine-tune

Add some noise and denoise with a new prompt

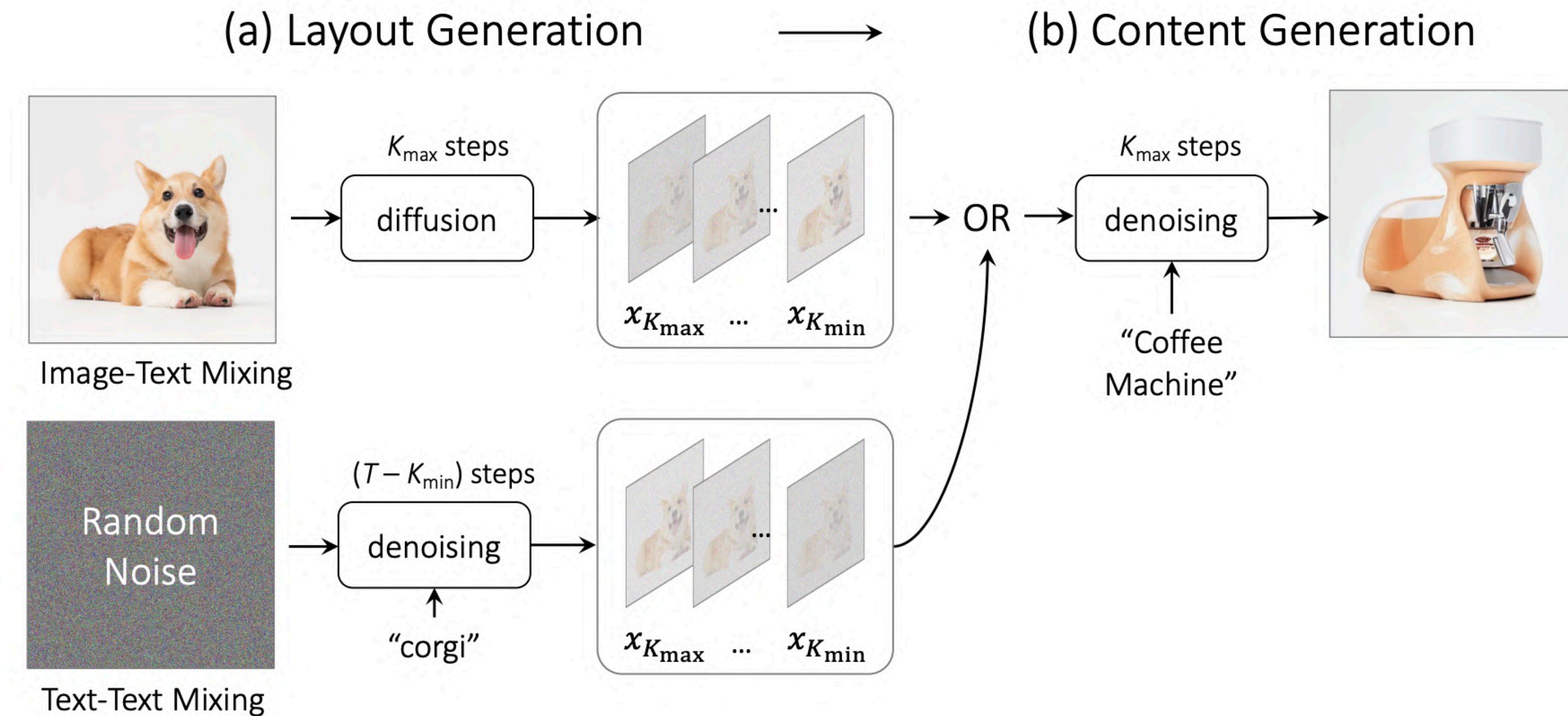


<https://sde-image-editing.github.io>

Add some noise and denoise with a new prompt

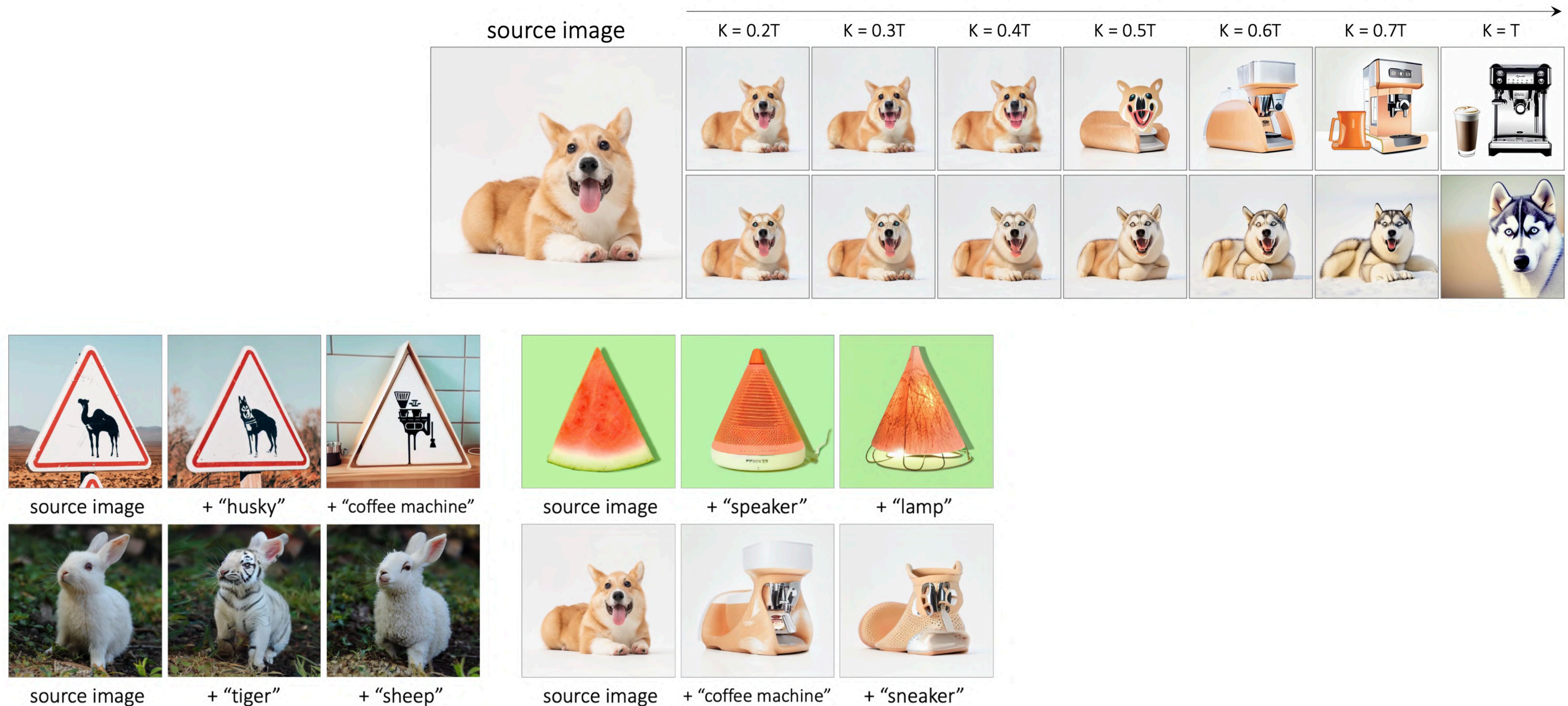


Add some noise and denoise with a new prompt



<https://magicmix.github.io>

Add some noise and denoise with a new prompt



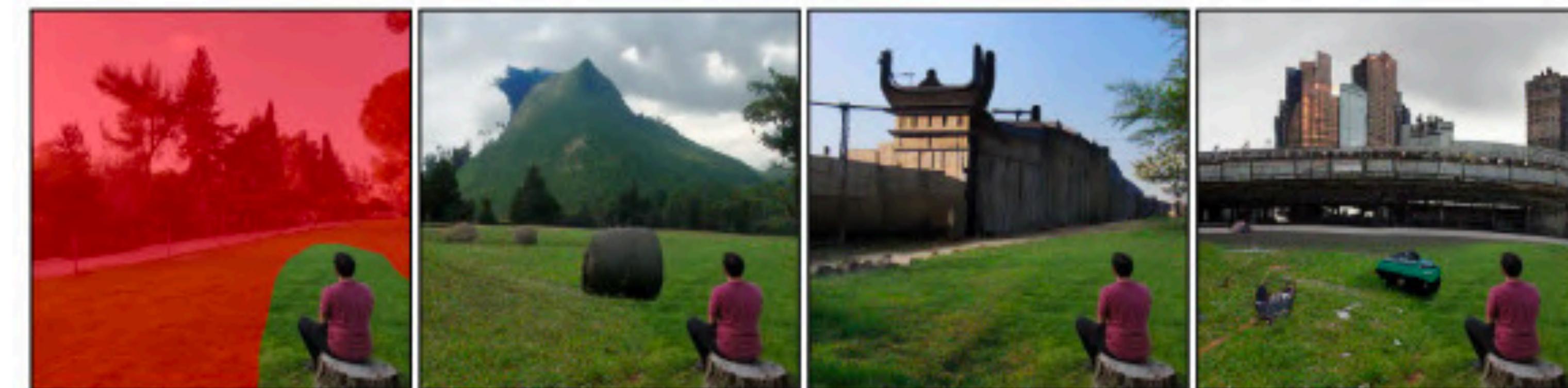
Same but with a mask



input+mask

no prompt

“white ball” “bowl of water”



input+mask

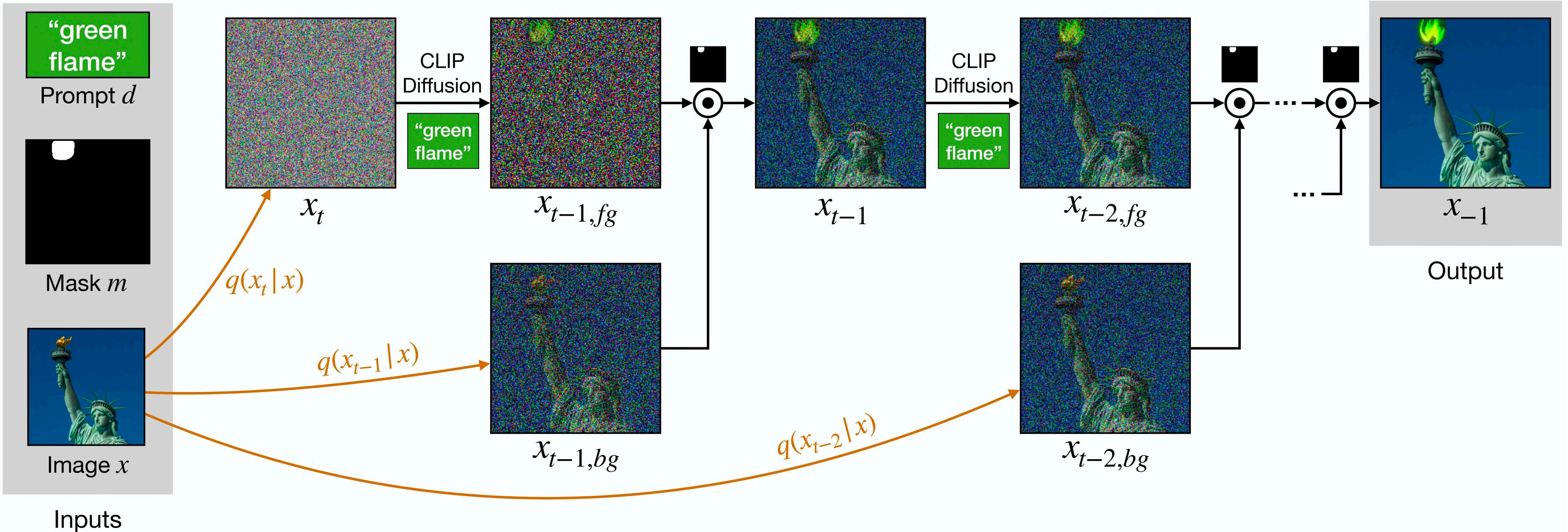
“big mountain”

“big wall”

“New York City”

<https://omriavrahami.com/blended-diffusion-page>

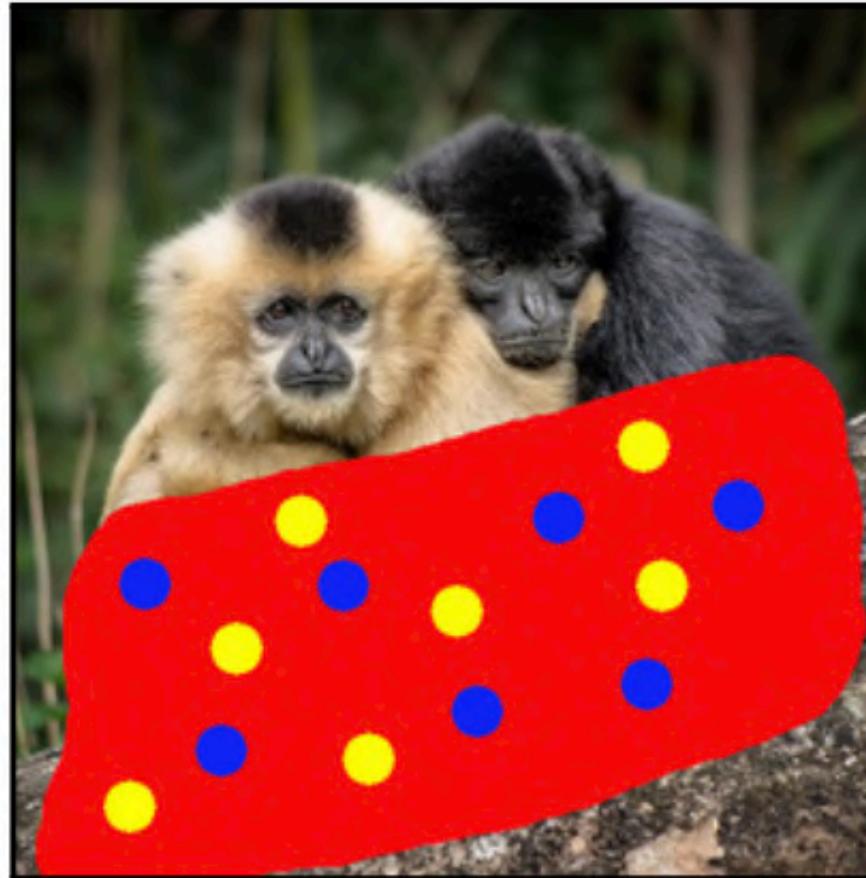
Same but with a mask



Same but with a mask



Original image



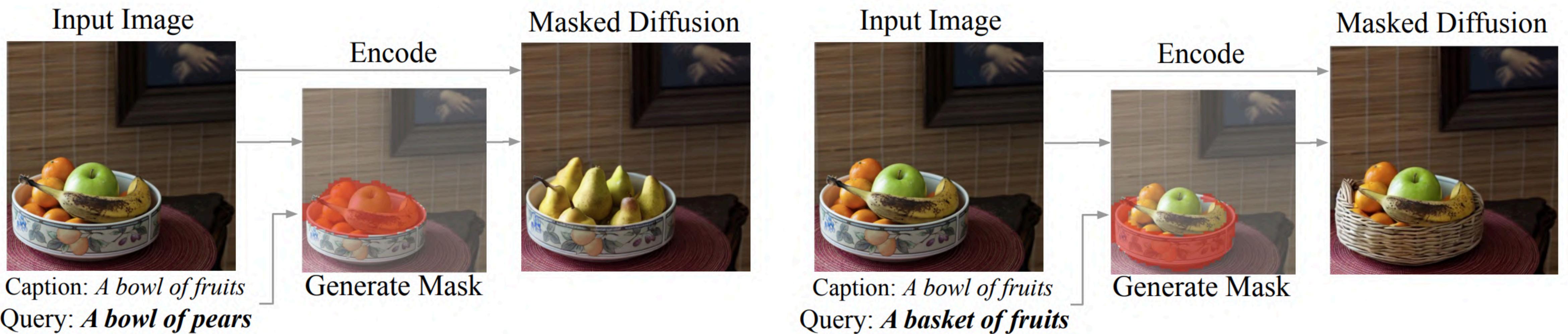
Input scribble



Result 1

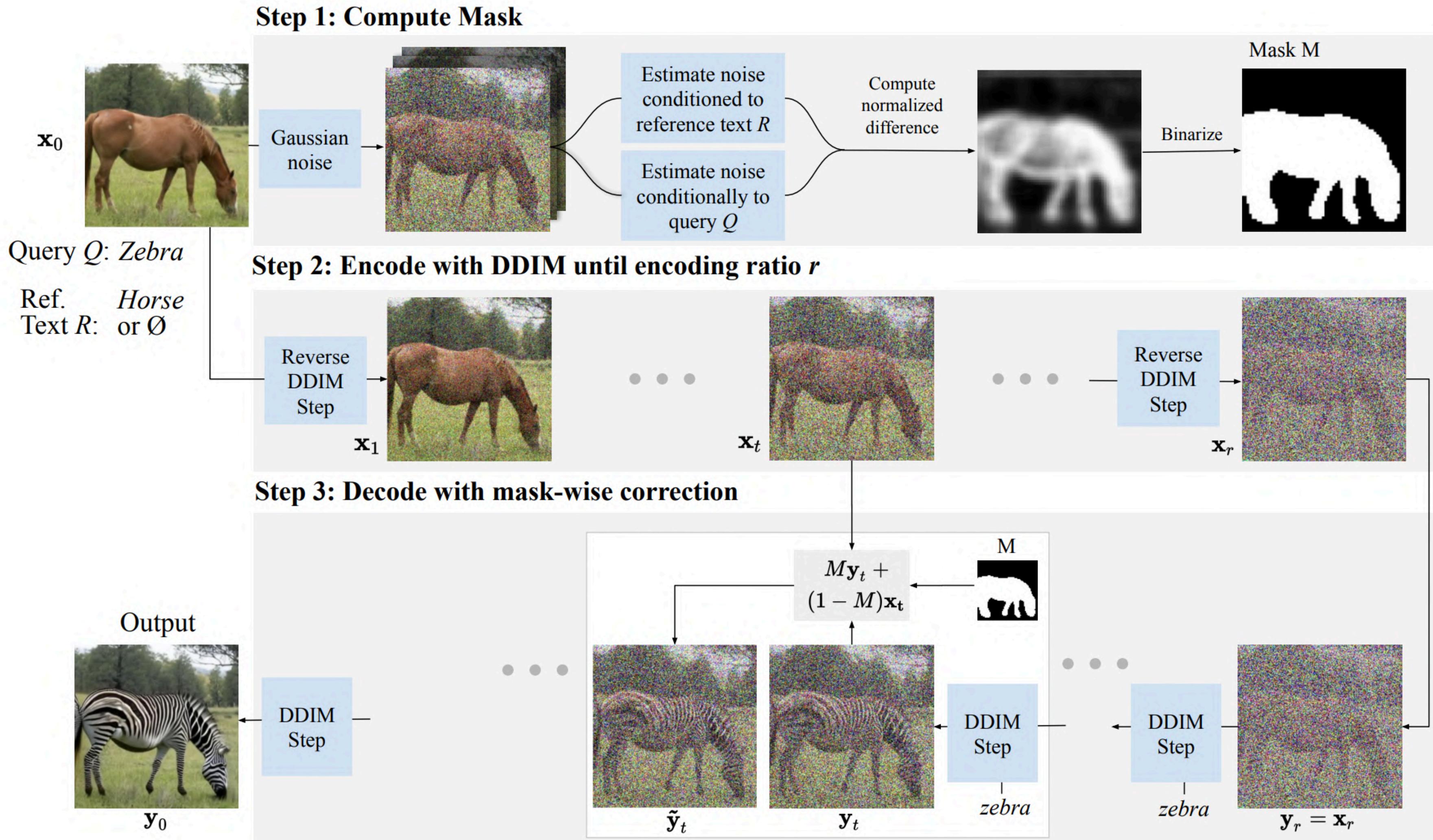


Same but with a mask



<https://arxiv.org/abs/2210.11427>

Same but with a mask



Prompt2Prompt

Original Prompt:

A photo of a teddy bear wearing sunglasses on and having a drink

Edited Prompt by Adding a Word:

A photo of a teddy bear wearing sunglasses on and having a **red** drink

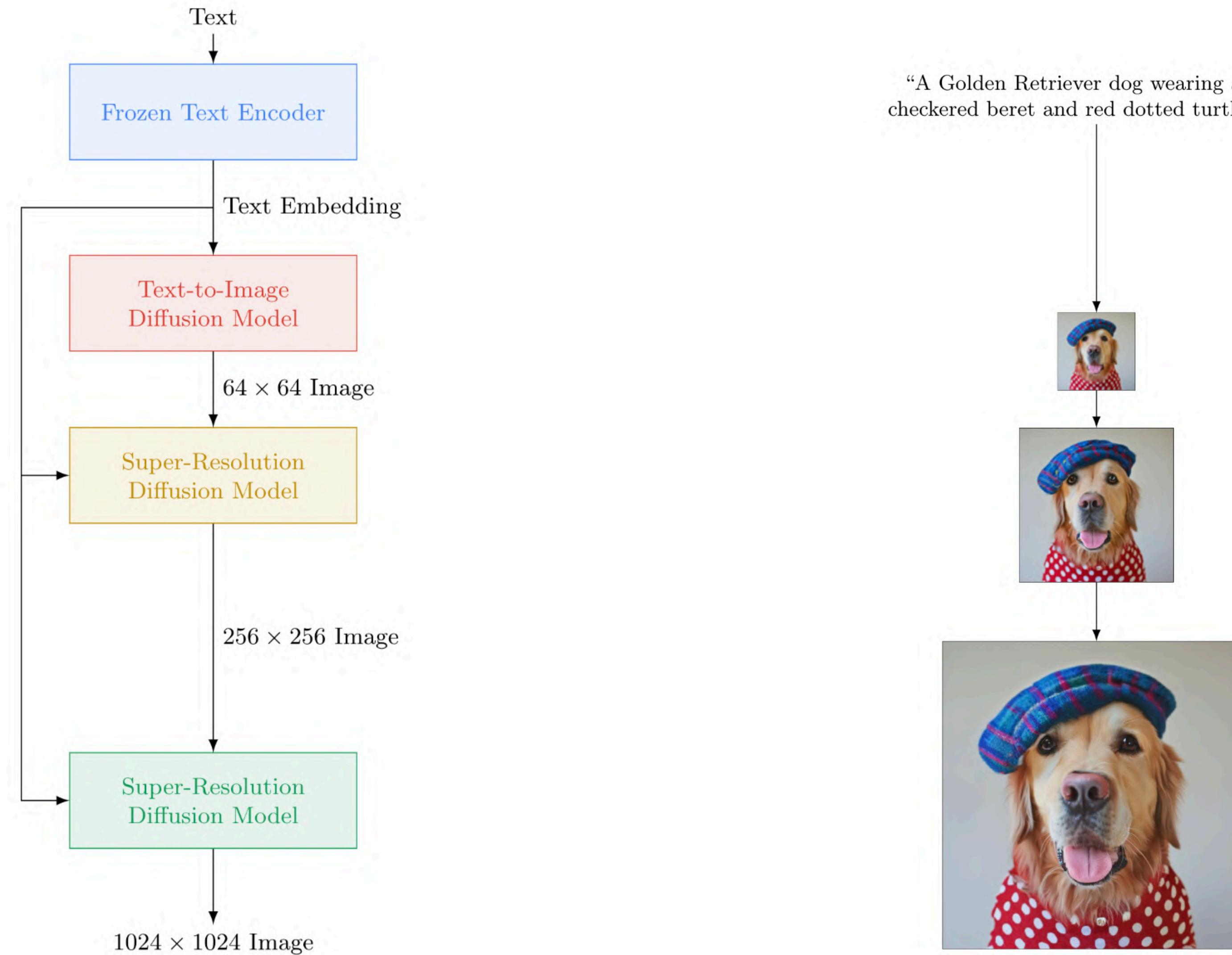
Generated Image without Promp Edit



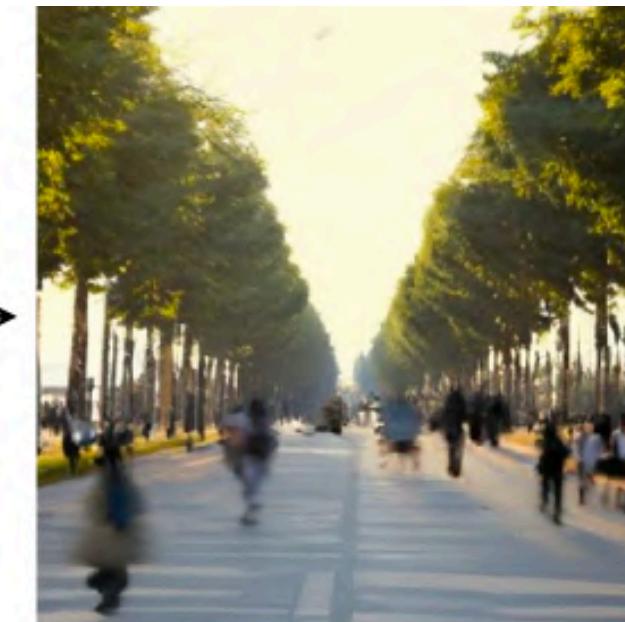
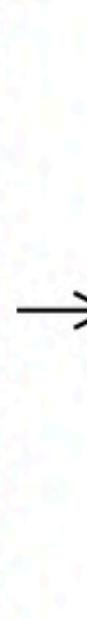
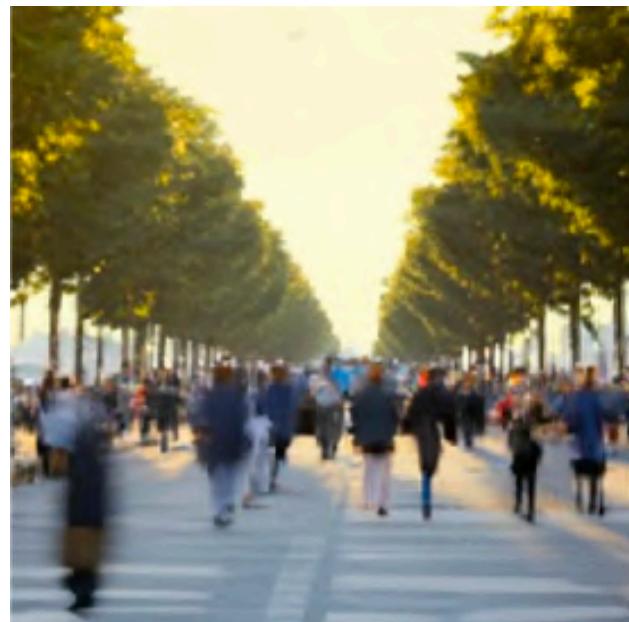
Generated Image with Promp Edit



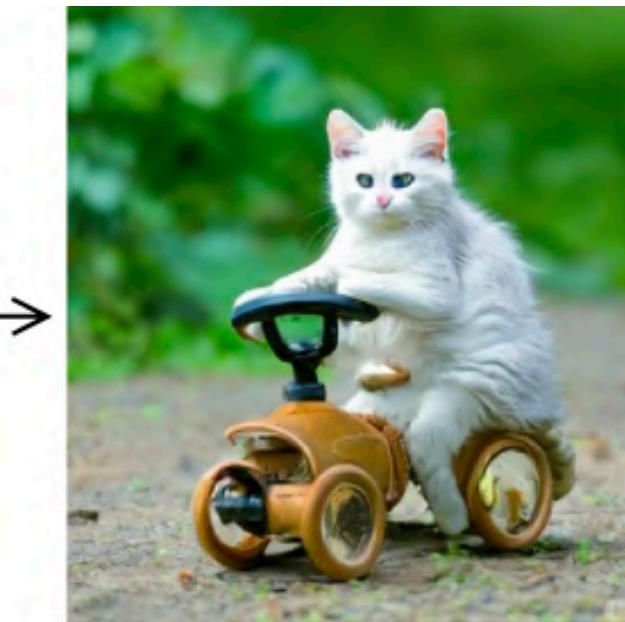
Imagen Architecture



Prompt2Prompt



“The boulevards are crowded today.”
↓↓↓↓↓



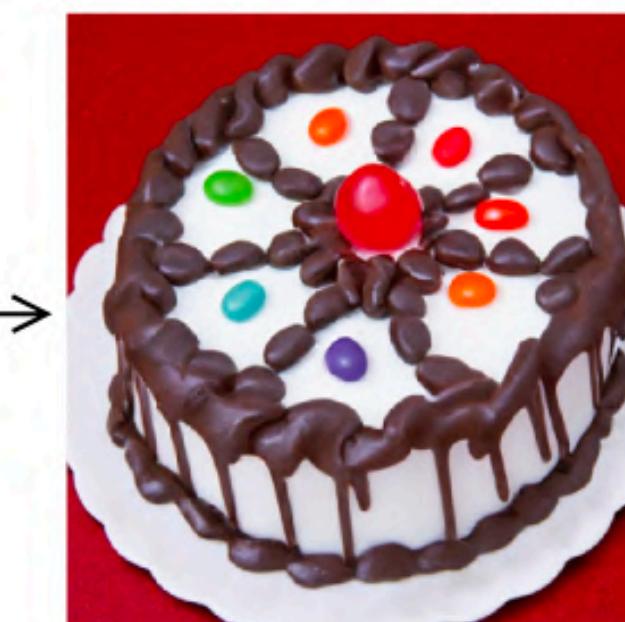
“Photo of a cat riding on a bicycle.”
car



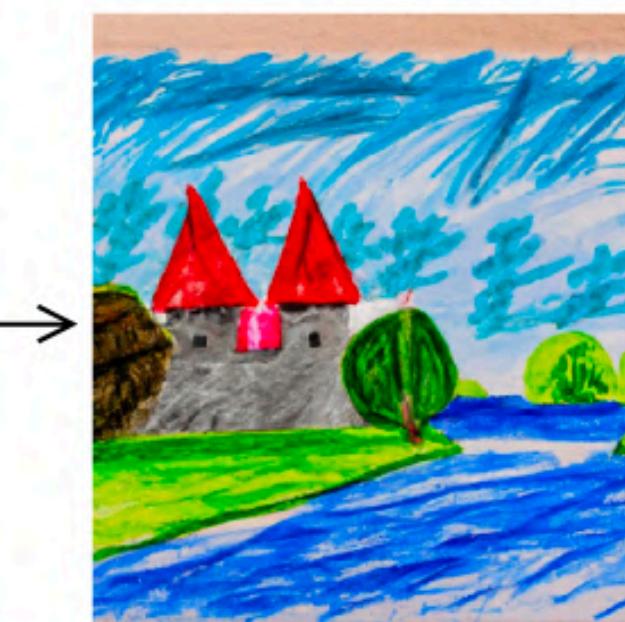
“Landscape with a house near a river
and a rainbow in the background.”



“My fluffy bunny doll.”
↑↑↑↑↑

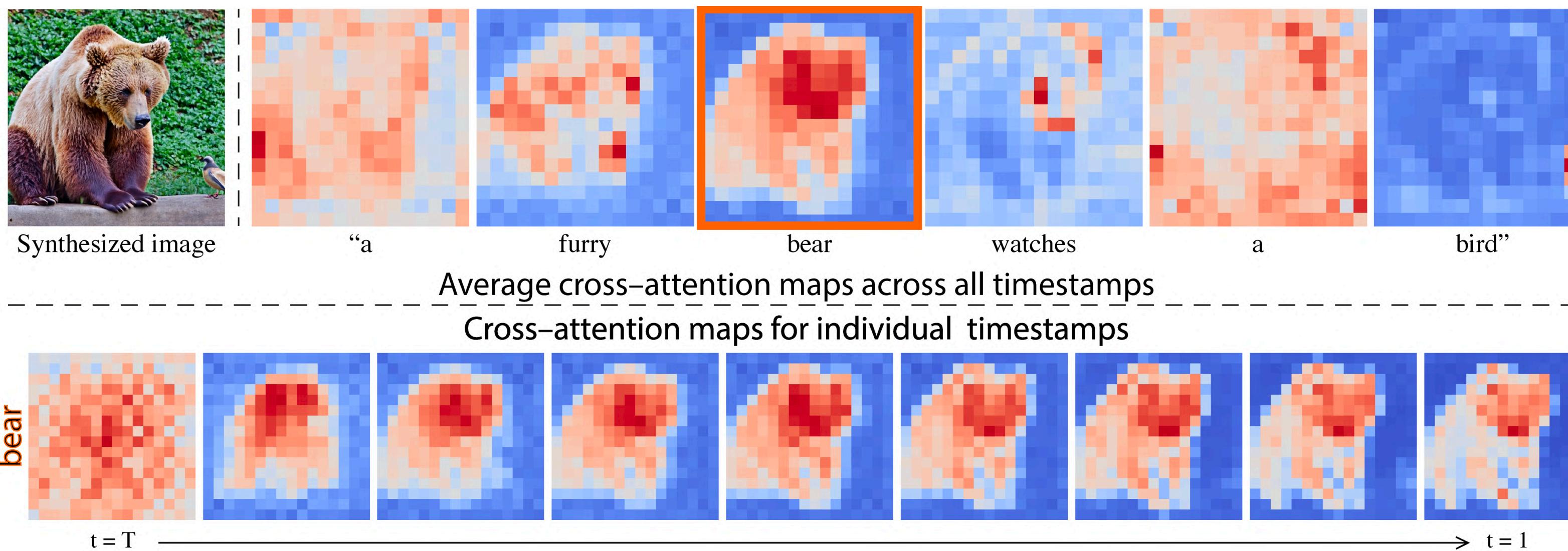
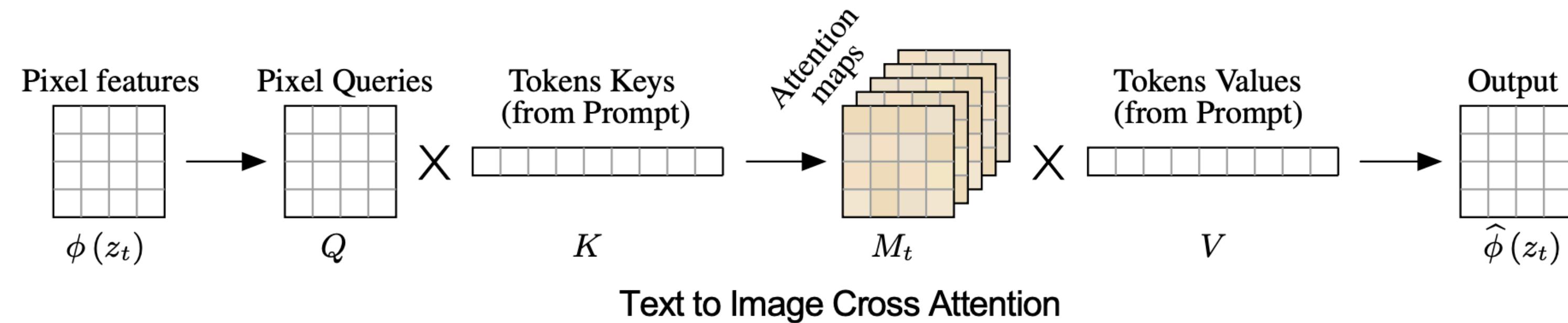


“a cake with decorations.”
jelly beans



“Children drawing of a castle next to a river.”

Prompt2Prompt



Prompt2Prompt

Source image and prompt:

“photo of a cat riding on a bicycle.”



bicycle → motorcycle



bicycle → car



bicycle → airplane



bicycle → train



W.O. attention injection

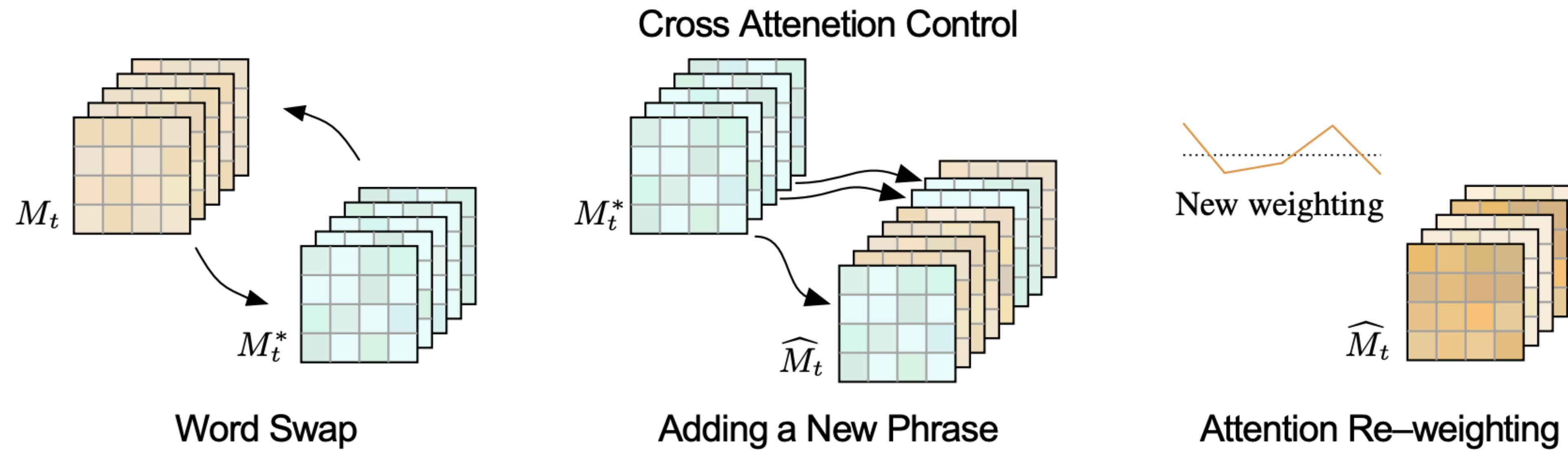
Full attention injection

Prompt2Prompt

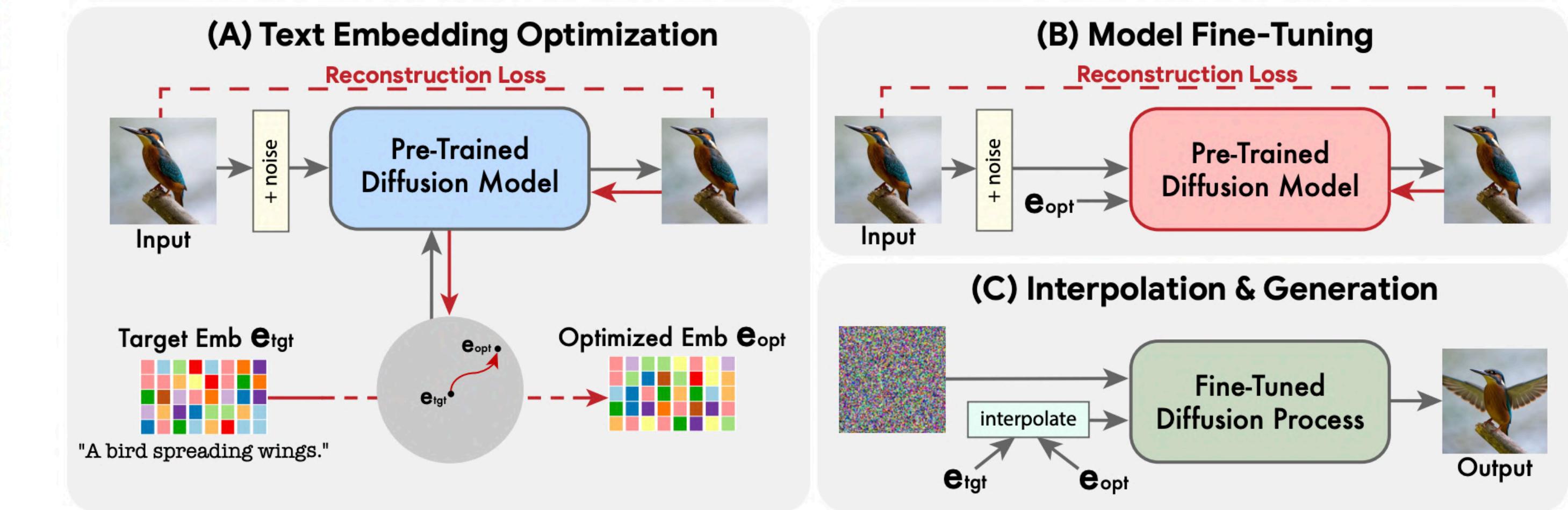
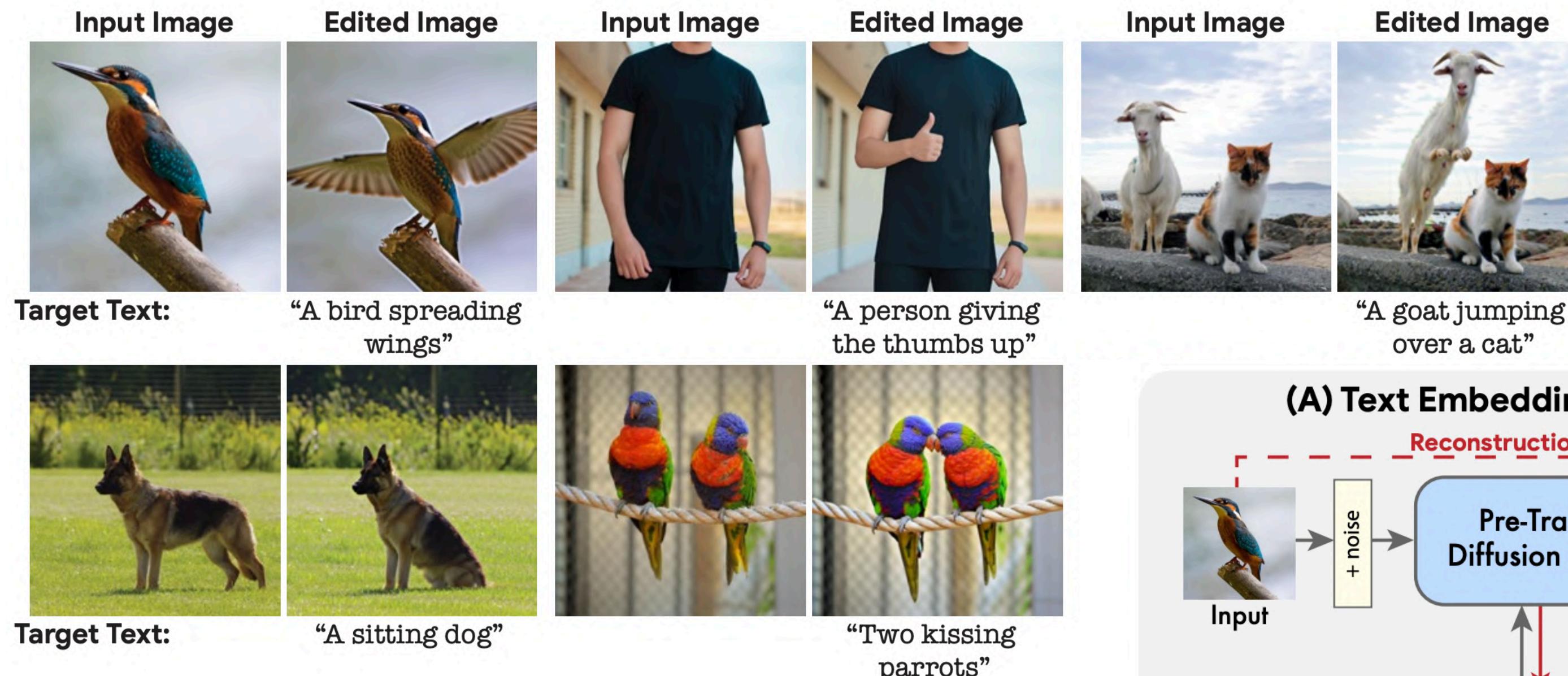
Algorithm 1: Prompt-to-Prompt image editing

- 1 **Input:** A source prompt \mathcal{P} , a target prompt \mathcal{P}^* , and a random seed s .
- 2 **Output:** A source image x_{src} and an edited image x_{dst} .
- 3 $z_T \sim N(0, I)$ a unit Gaussian random variable with random seed s ;
- 4 $z_T^* \leftarrow z_T$;
- 5 **for** $t = T, T - 1, \dots, 1$ **do**
- 6 $z_{t-1}, M_t \leftarrow DM(z_t, \mathcal{P}, t, s)$;
- 7 $M_t^* \leftarrow DM(z_t^*, \mathcal{P}^*, t, s)$;
- 8 $\widehat{M}_t \leftarrow Edit(M_t, M_t^*, t)$;
- 9 $z_{t-1}^* \leftarrow DM(z_t^*, \mathcal{P}^*, t, s_t) \{M \leftarrow \widehat{M}_t\}$;
- 10 **end**
- 11 **Return** (z_0, z_0^*)

Prompt2Prompt



Fine-tune



<https://arxiv.org/pdf/2210.09276.pdf>

DreamBooth



Input images



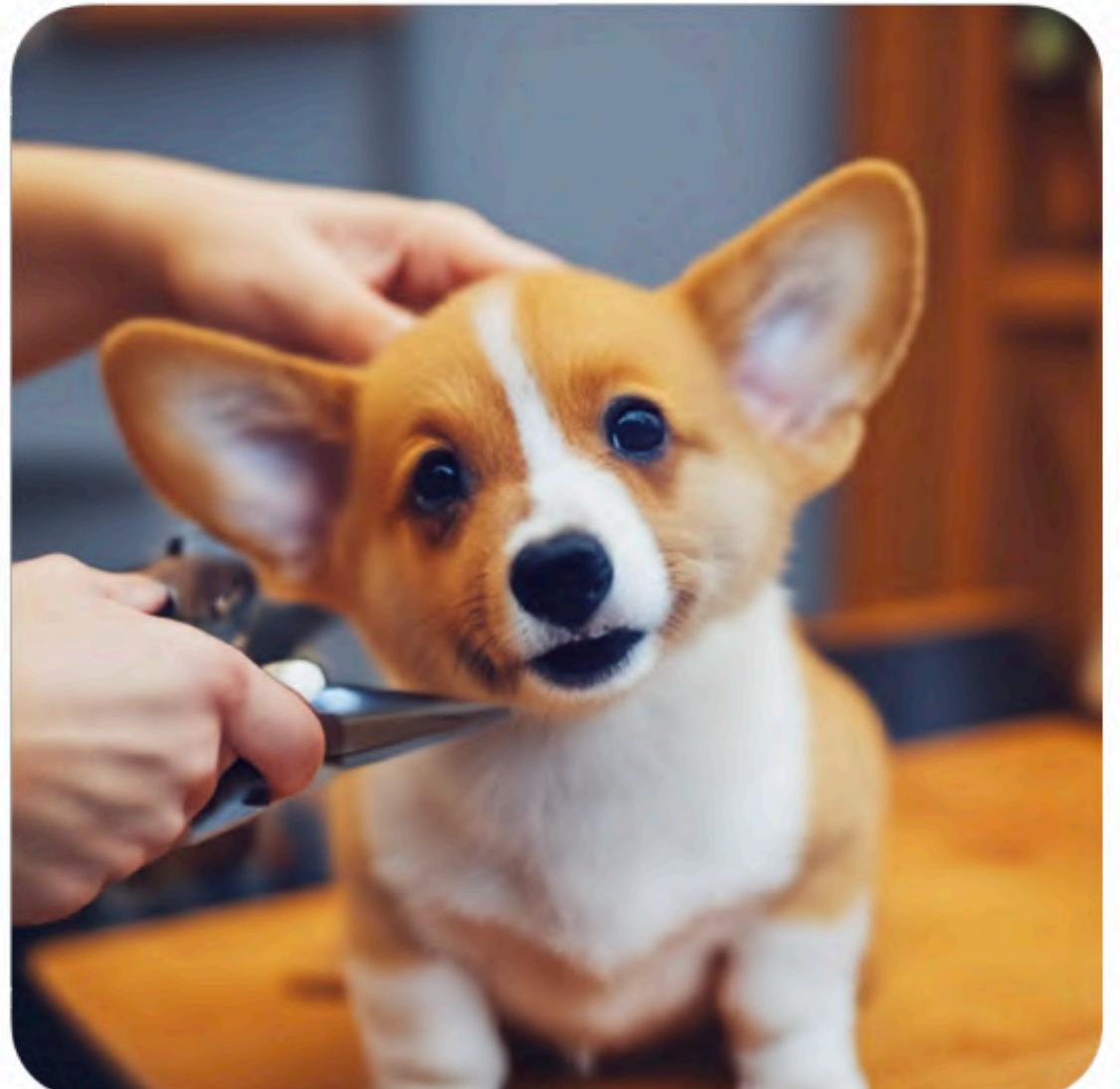
in the Acropolis



swimming

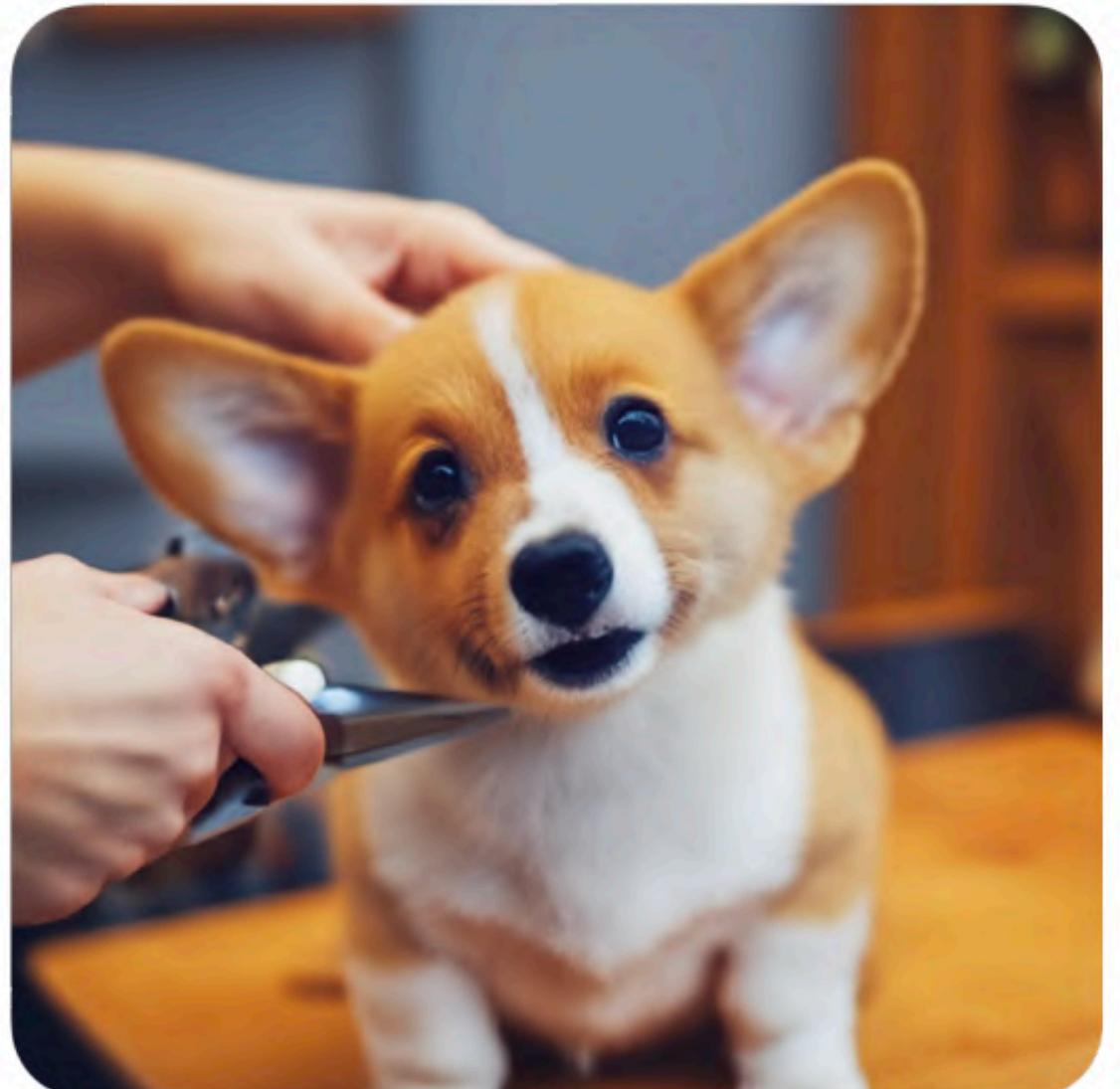


sleeping



in a doghouse

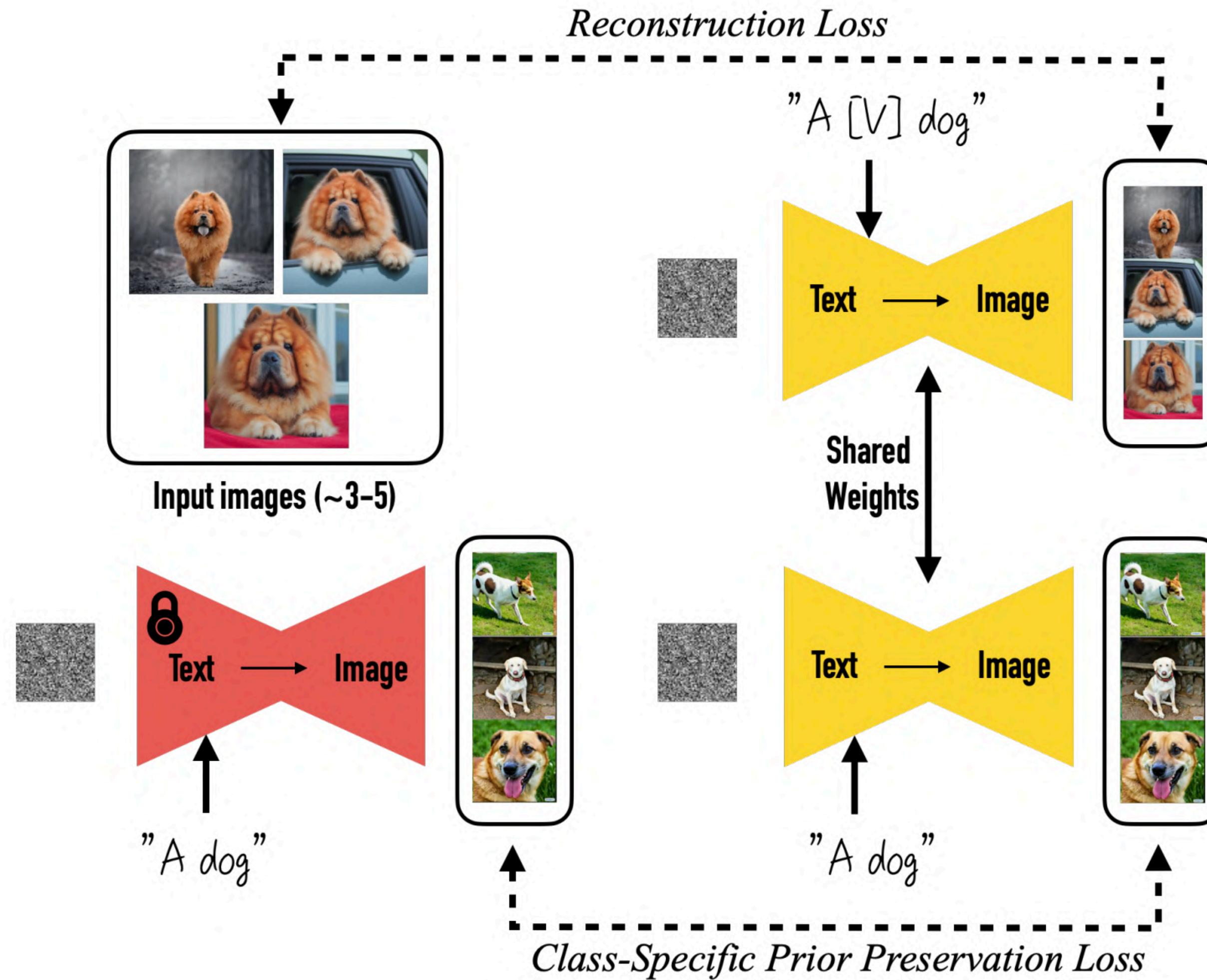
in a bucket



getting a haircut

<https://arxiv.org/pdf/2208.12242.pdf>

DreamBooth



Problems of “stupid” fine-tuning

1. *Language drift* – model that is pre-trained on a large text corpus and later fine-tuned for a specific task progressively loses syntactic and semantic knowledge of the language.
2. *Reduced output diversity* – text2img models naturally posses high amounts of output diversity, so when fine-tuning on a small set of images we would like to be able to generate the subject in novel viewpoints, poses and articulations. Yet, there is a risk of reducing the amount of variability in the output poses and views of the subject.

DreamBooth

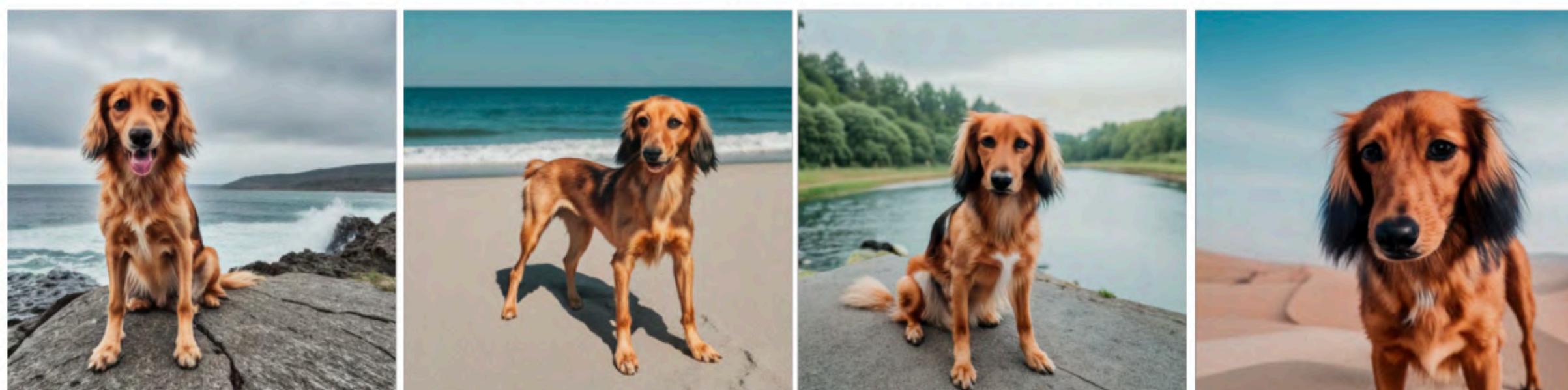
Input images



w/o prior-preservation loss

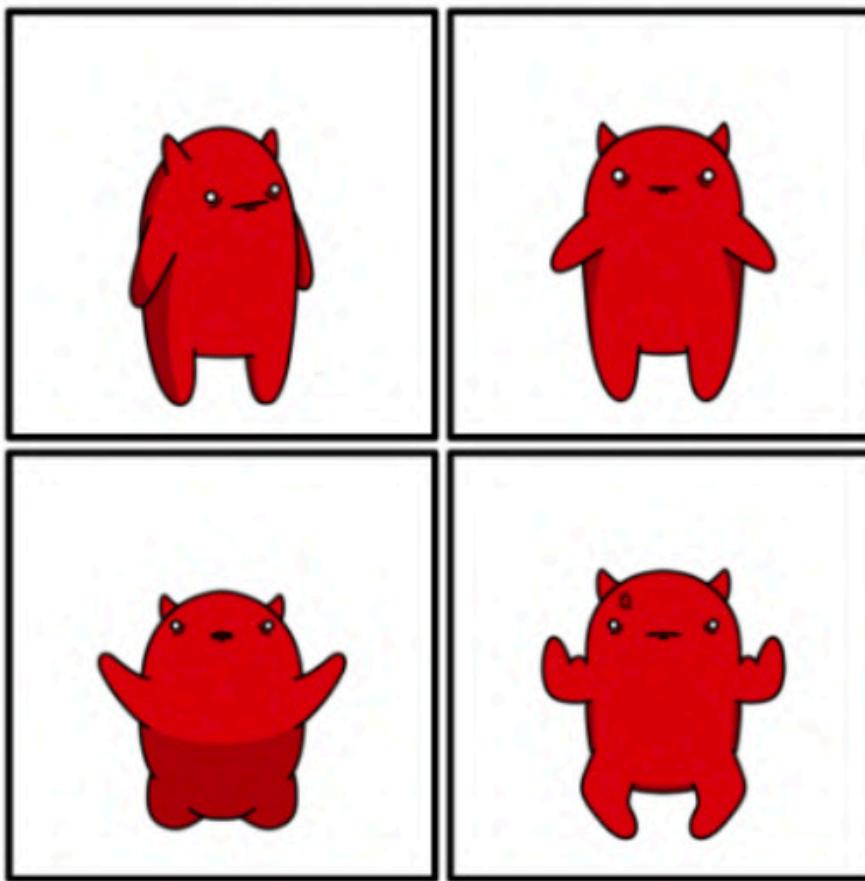


with prior-preservation loss

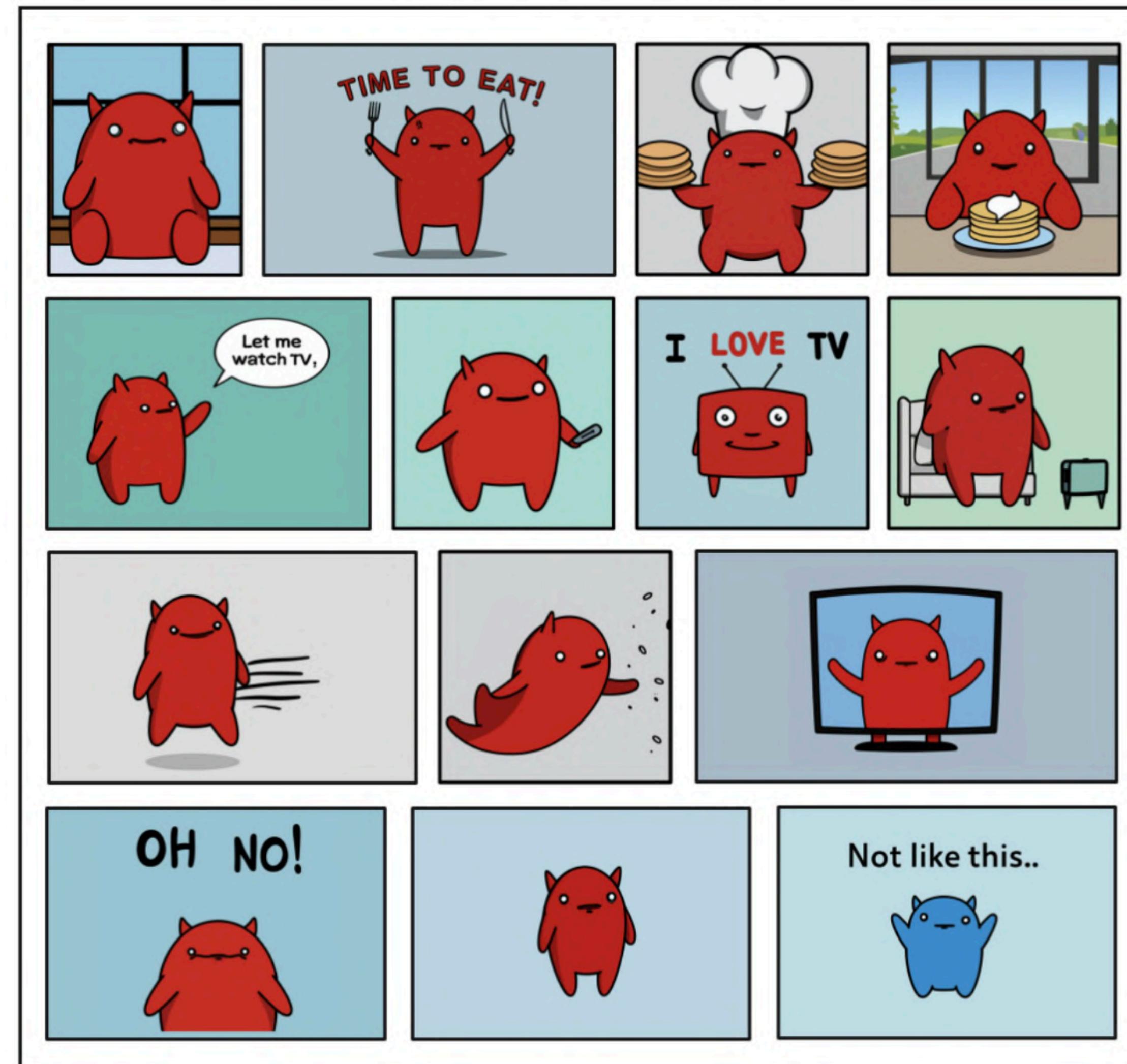


DreamBooth

Input Images



Generated Comic



References

1. Lit review of different approaches
2. Intro to text2img and image editing