NULL-text Inversion for Editing Real Images using Guided Diffusion Models

Ron Mokady*, Amir Hertz*, Kfir Aberman, Yael Pritch, Daniel Cohen-Or

** equal contribution*

Research at Google

TUE-PM-183

JUNE 18-22, 2023 CVPR VANCOUVER, CANADA

Text-Guided Diffusion Models

DALL E 2

Imagen

Stable Diffusion

Prompt-to-Prompt Image Editing with Cross Attention Control, Hertz et al.

Generated Image

S[illegible]ttan

Prompt-to-Prompt Image Editing with Cross Attention Control, Hertz et al.

Generated Image

Prompt-to-Prompt over a **real** image?

NULL-Text Inversion

Input caption: “A baby wearing a blue shirt lying on the sofa.”

Input Image

“... blond baby...”

“... floral shirt...”

“... golden shirt...”

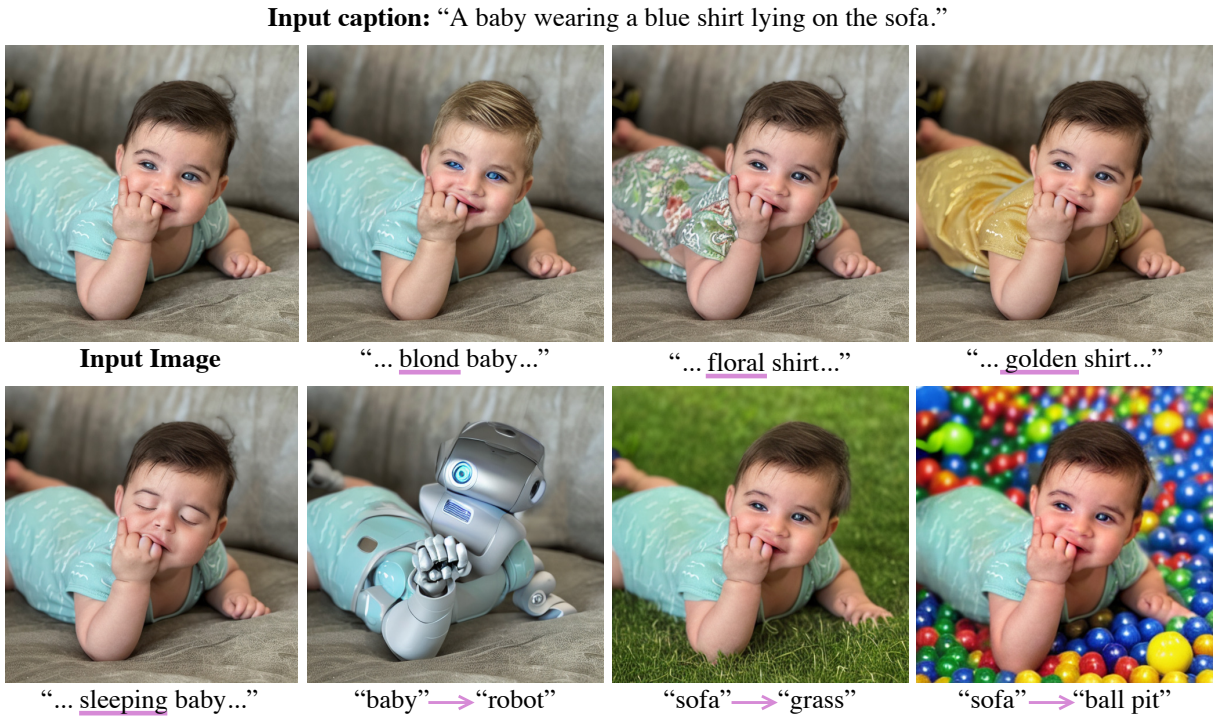
“... sleeping baby...”

“baby” → “robot”

“sofa” → “grass”

“sofa” → “ball pit”

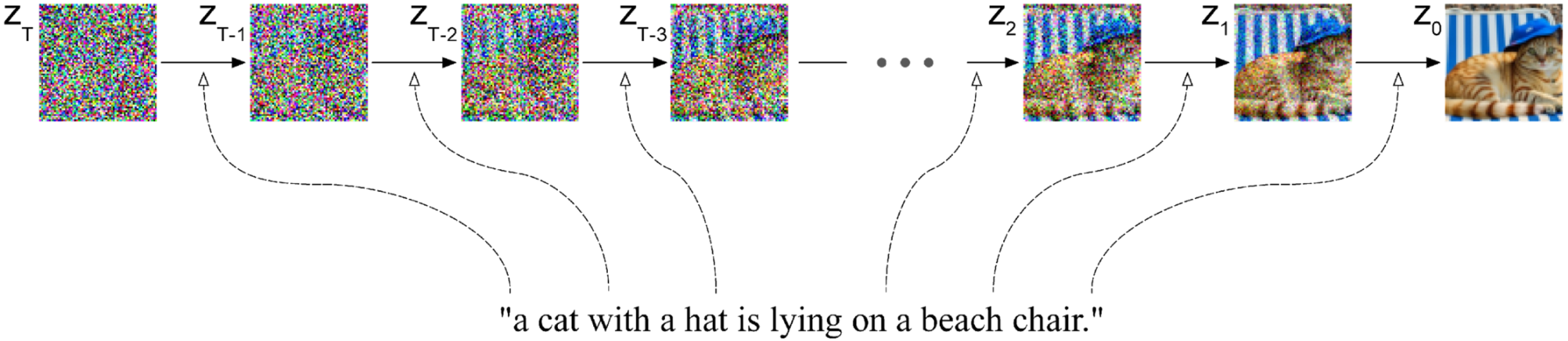
NULL-text Inversion



1. Pivotal Inversion for Diffusion

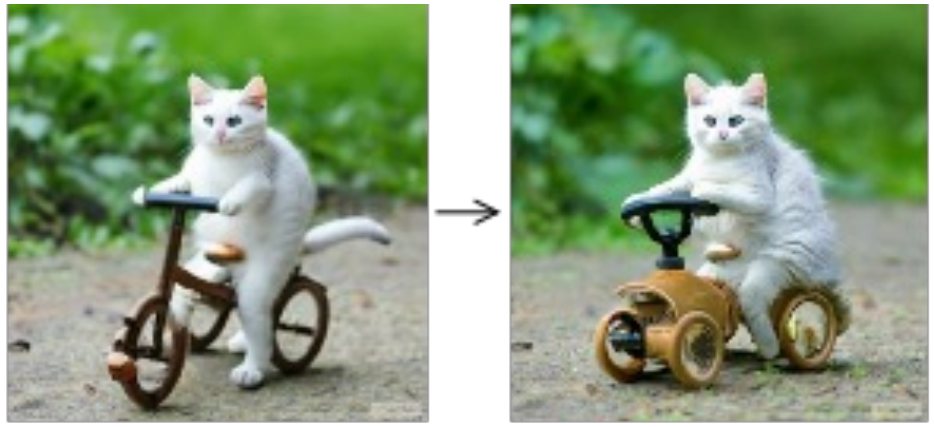
2. Null-Text Optimization

Text-to-Image Generation with Diffusion Models



Prompt-to-Prompt Image Editing with Cross Attention Control

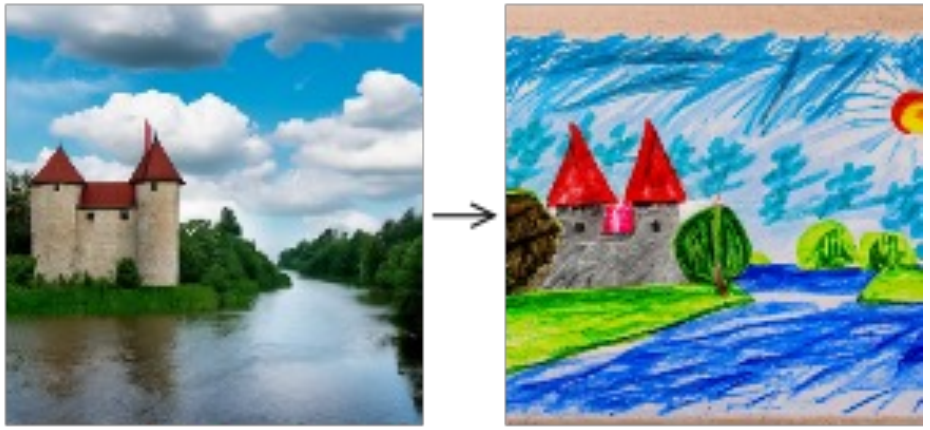
Hertz et al.



"Photo of a cat riding on a ~~bicycle~~ cat."

"Landscape with a secluded house near a river and a rainbow in the background."

"a cake with decorations." jelly beans



"Children drawing of a castle next to a river."

Prompt-to-Prompt over a **real** image?

Inversion

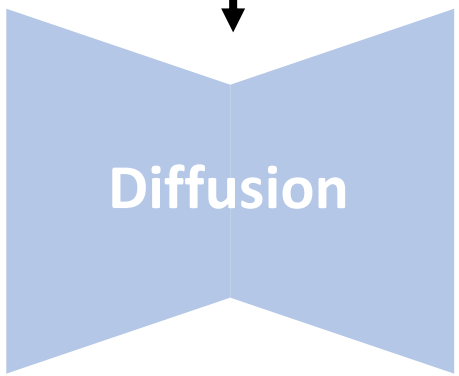
Input

A baby on a sofa

A baby on a sofa

Output:

z_t



Inversion

Output:

z_t

A baby on a sofa

Diffusion

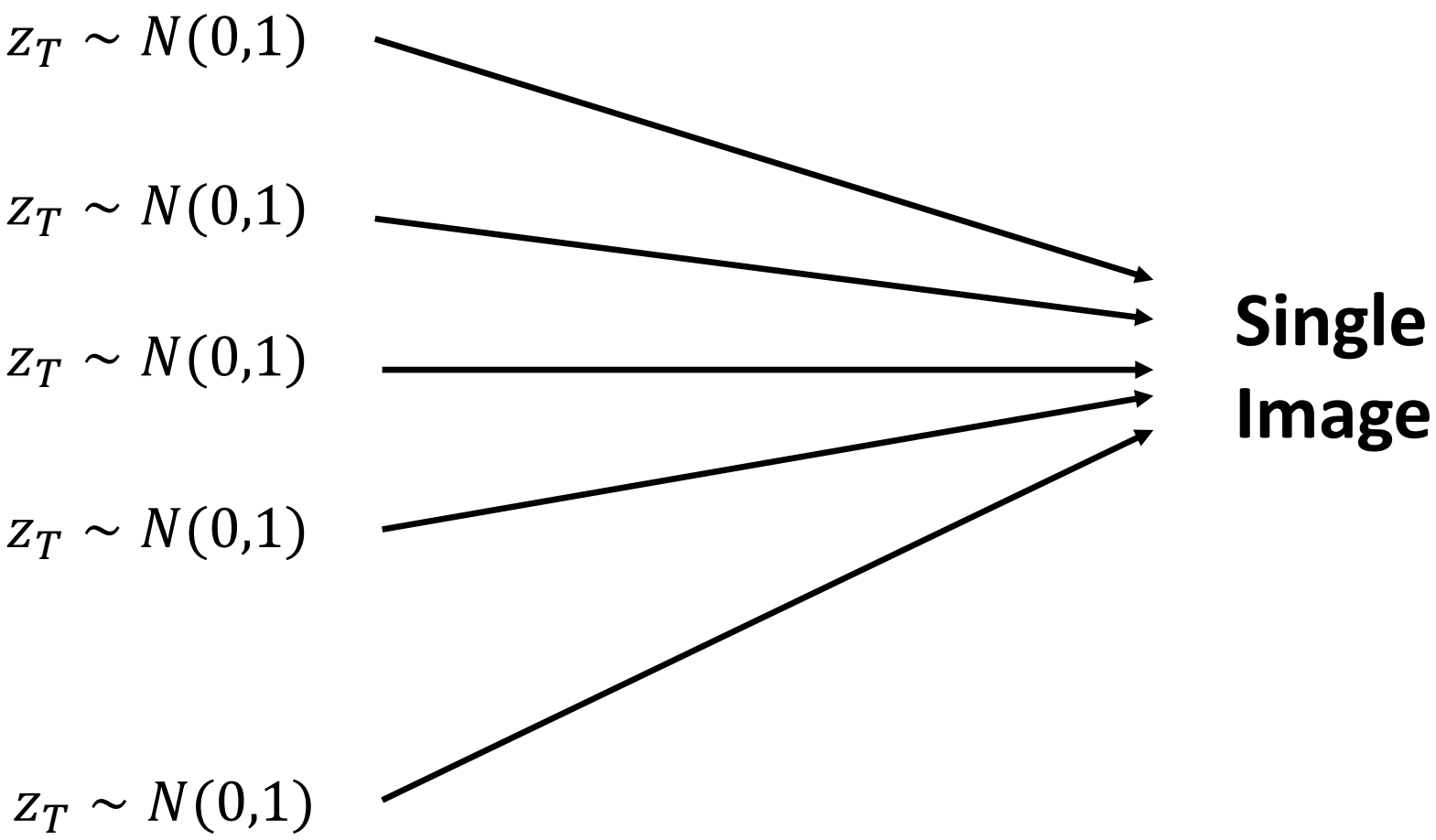
Inference

A baby on **a grass**

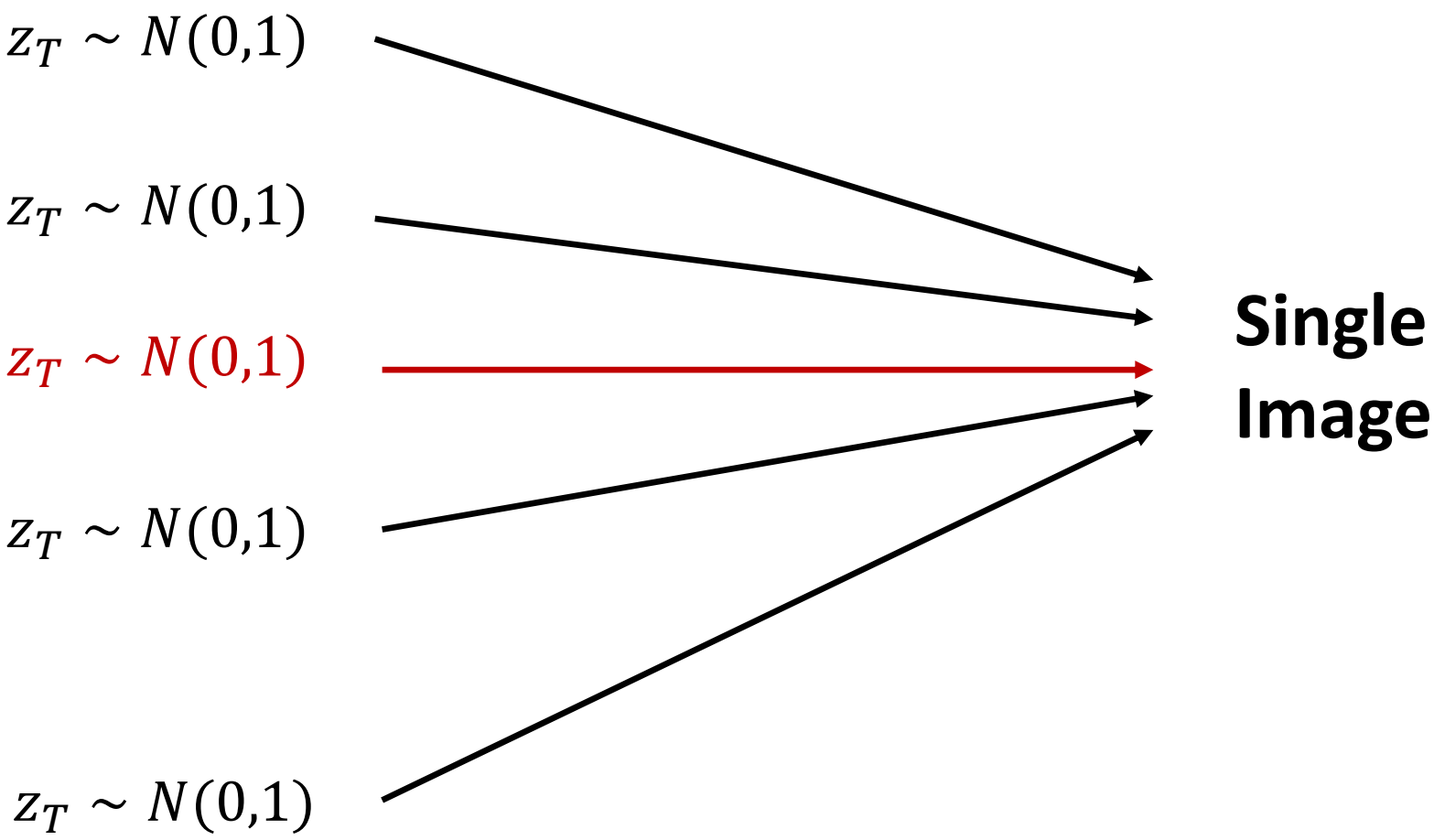
Diffusion

1. Pivotal Inversion for Diffusion

Training of Other Approches



Inference of Other Approaches



DDIM Inversion

DENOISING DIFFUSION IMPLICIT MODELS, *Song et al.*

Diffusion Models Beat GANs on Image Synthesis, *Dhariwal et al.*

Without classifier-free guidance.

Original

Inversion

Not editable!

DDIM Inversion

Using classifier-free guidance during **inversion**.

And **using** classifier-free guidance during **inference**.

Original

Inversion

Without classifier-free guidance during **inversion**.

And **using** classifier-free guidance during **inference.**

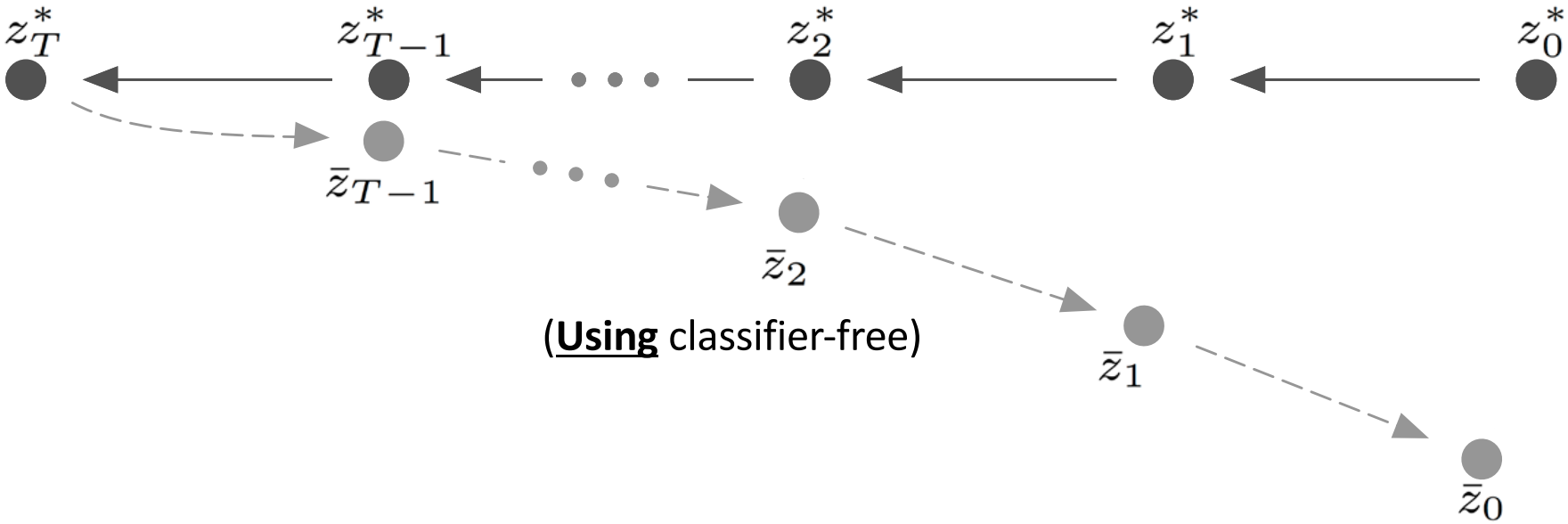
Original

Inversion

Good Starting Point!

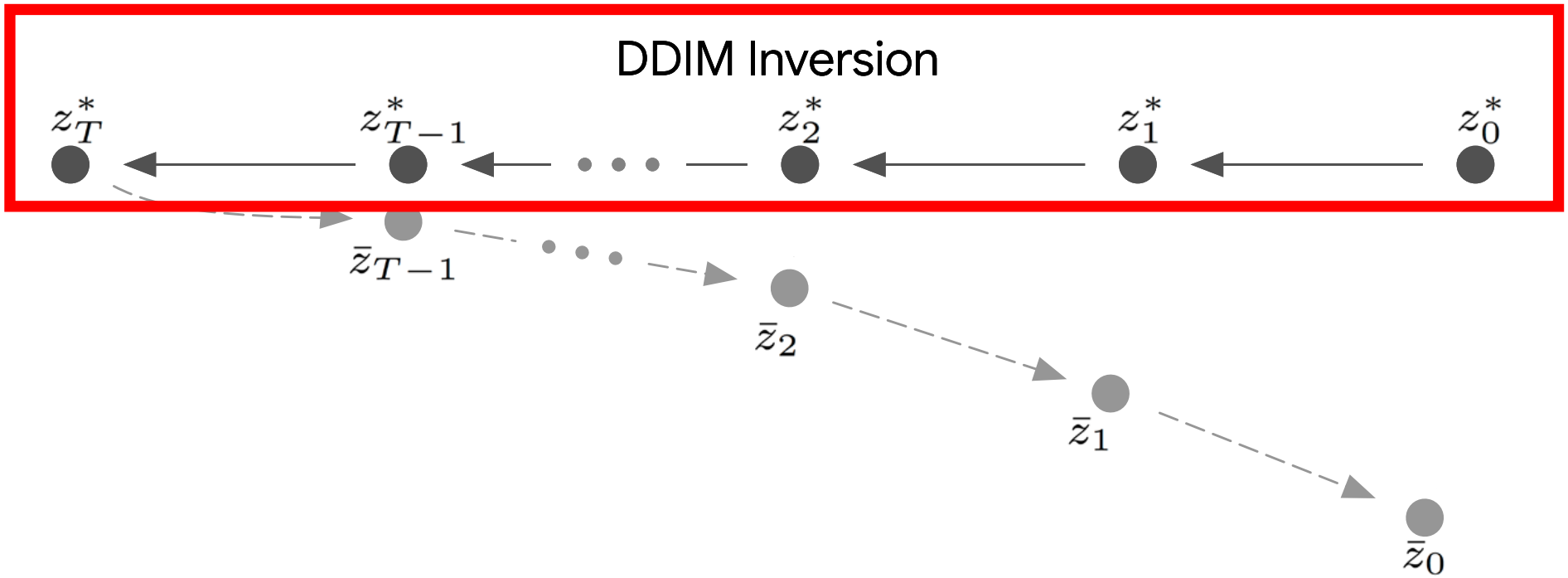
Input Image

DDIM Inversion (**Without** classifier-free)



(**Using** classifier-free)

Pivot:



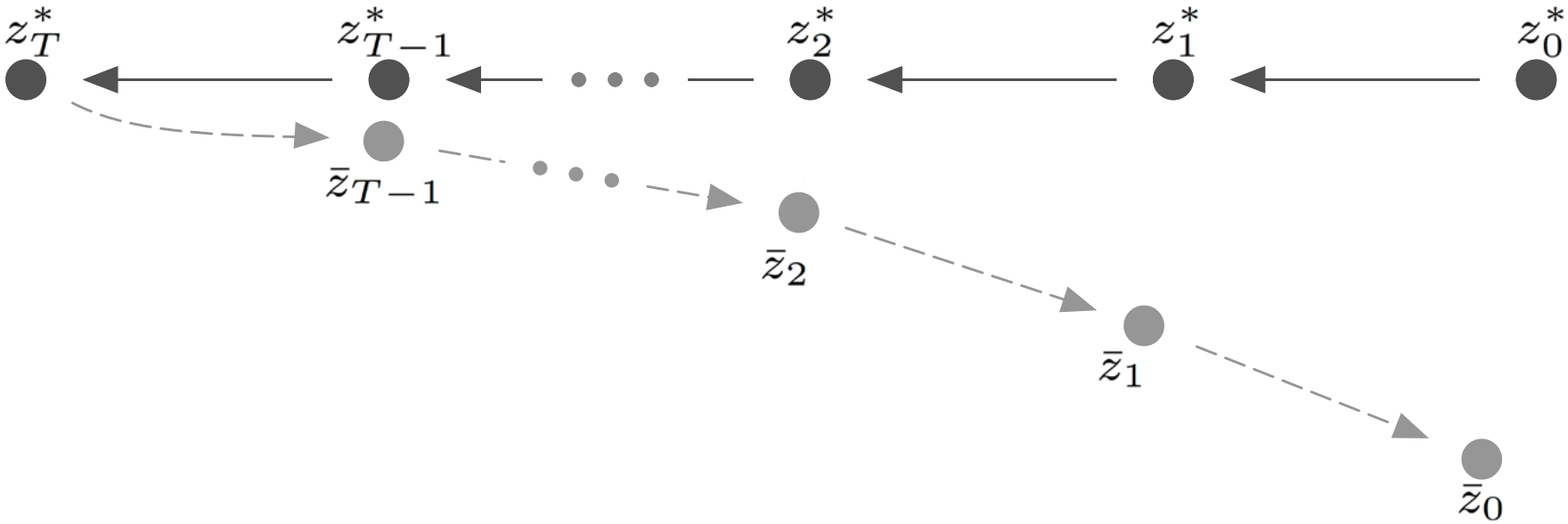
Input Image

Initial Inversion

Pivotal Tuning for Latent-based Editing of Real Images

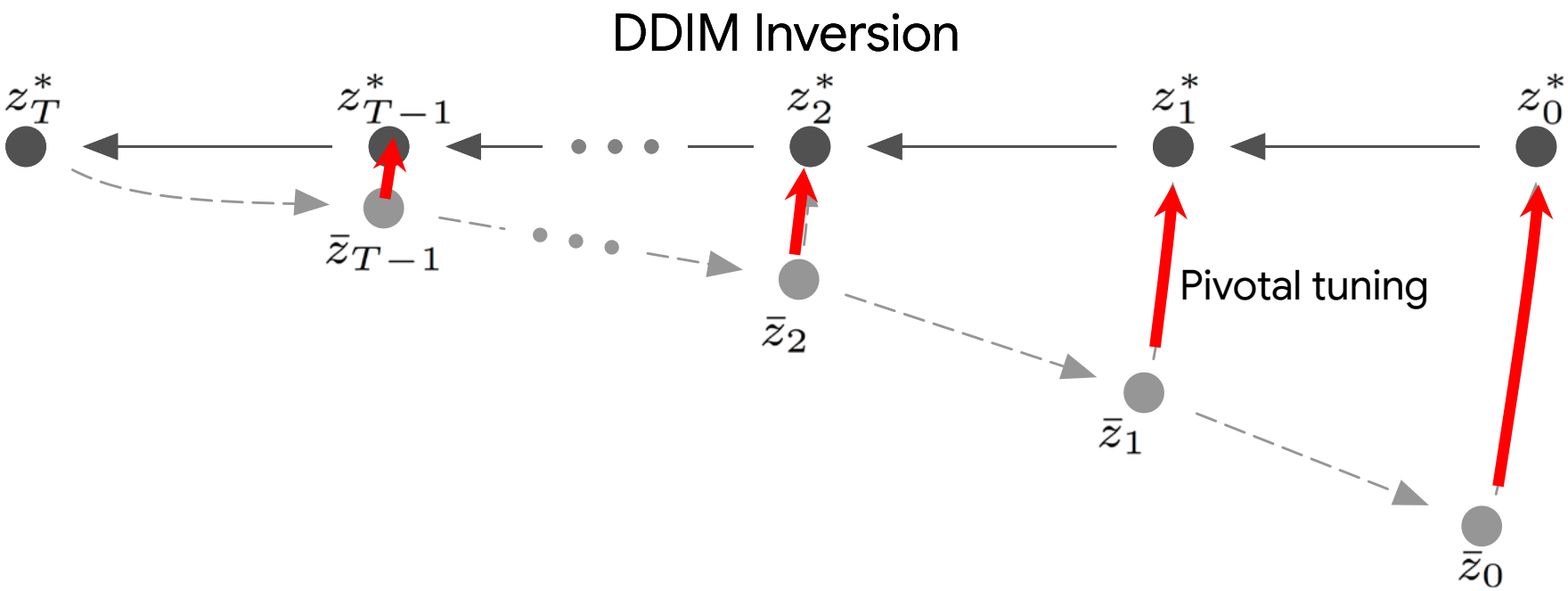
Daniel Roich, Ron Mokady, Amit H. Bermano, Daniel Cohen-Or

DDIM Inversion



Input Image

Initial Inversion



Input Image

Initial Inversion

Final Inversi

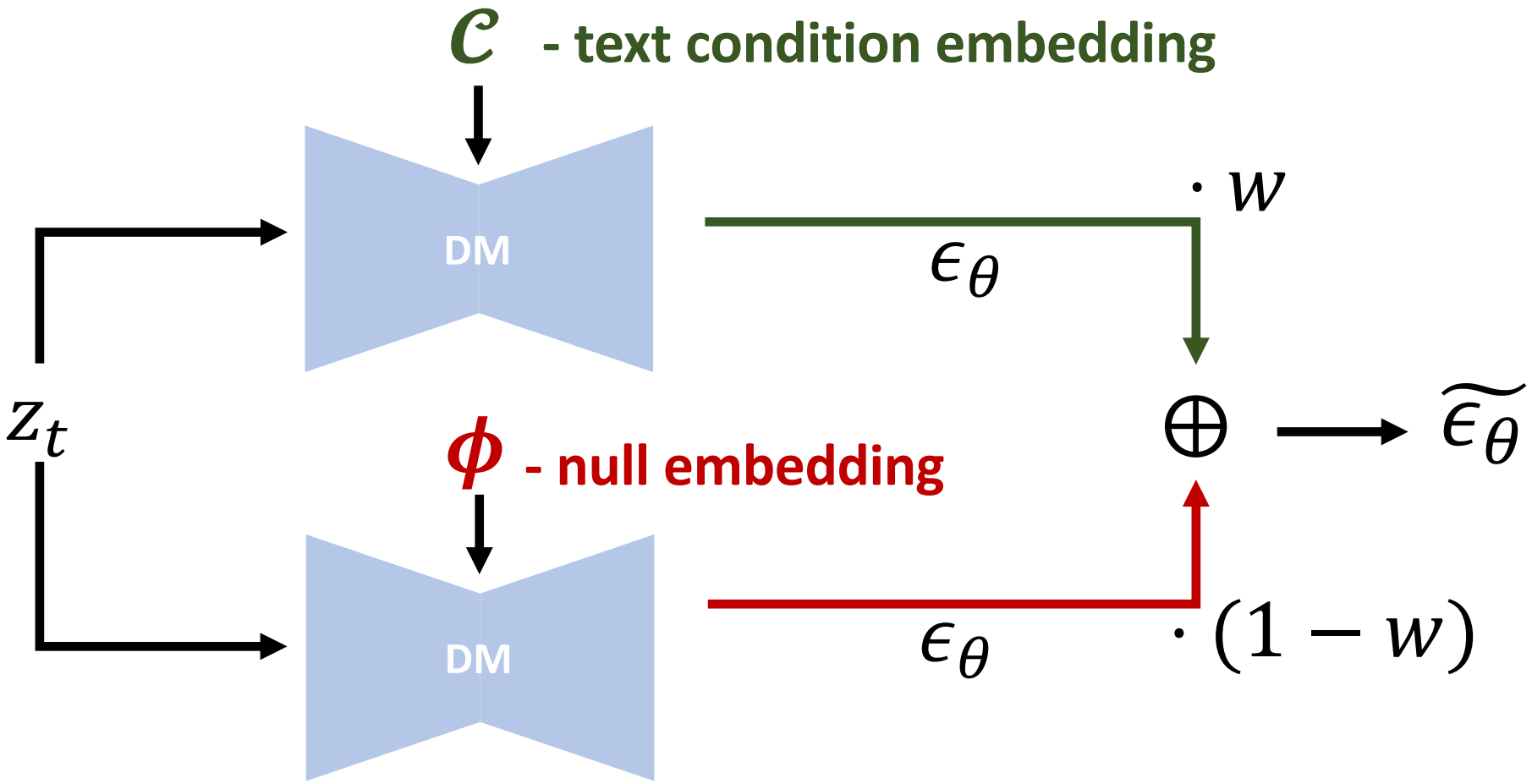
How to optimize?

Can we avoid fine-tuning the model?

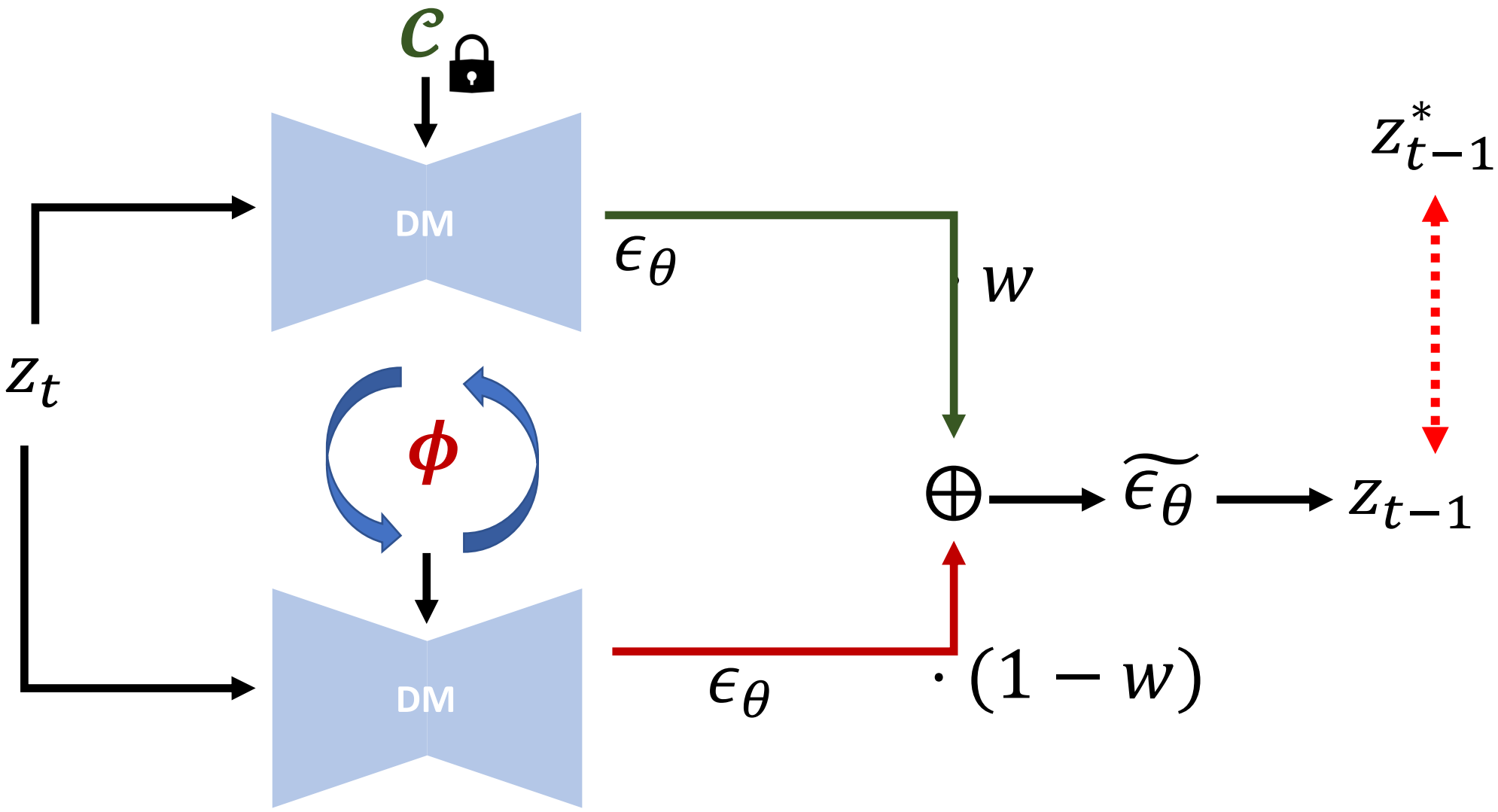
2. Null-Text Optimization

Classifier-free guidance

$$\tilde{\epsilon}_\theta(z_t, \mathcal{C}, \phi) = w \cdot \epsilon_\theta(z_t, \mathcal{C}) + (1 - w) \cdot \epsilon_\theta(z_t, \phi)$$



Null-Text Optimization



Input Image

DDIM Inversion

z_T^* z_{T-1}^* z_2^* z_1^* z_0^*

$\bar{z}_{T-1}$ $\bar{z}_2$ $\bar{z}_1$ $\bar{z}_0$

Pivotal tuning

ϕ_{T-1} ϕ_2 ϕ_1 ϕ_0

Input caption: “A man in glasses eating a doughnut in the park.”

Input Image

“... red-haired man...”

“glasses” → “sunglasses”

“angry man...”

“doughnut” → “pizza”

“glasses” → “Joker mask”

“...the park at sunset.”

“park” → “desert”

Input caption: “Two birds sitting on a branch.”

Input Image | branch → rainbow | branch → metal pole | branch → electric cable | “...Lego birds” | “...crochet birds” | “...origami birds” | “...jello birds”

Input caption: “Cake on a table.”

Input Image | "Chocolate cake..." | "Strawberry cake..." | “Spinach moss cake...” | "Macaroni cake..." | "Purple neon cake..." | “Monster cake...” | "Pepperoni cake...

Thank You!

Code:

github.com/google/prompt-to-prompt/#null-text-inversion-for-editing-real-images