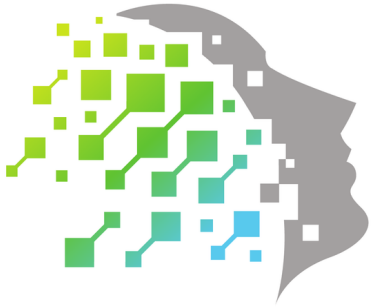
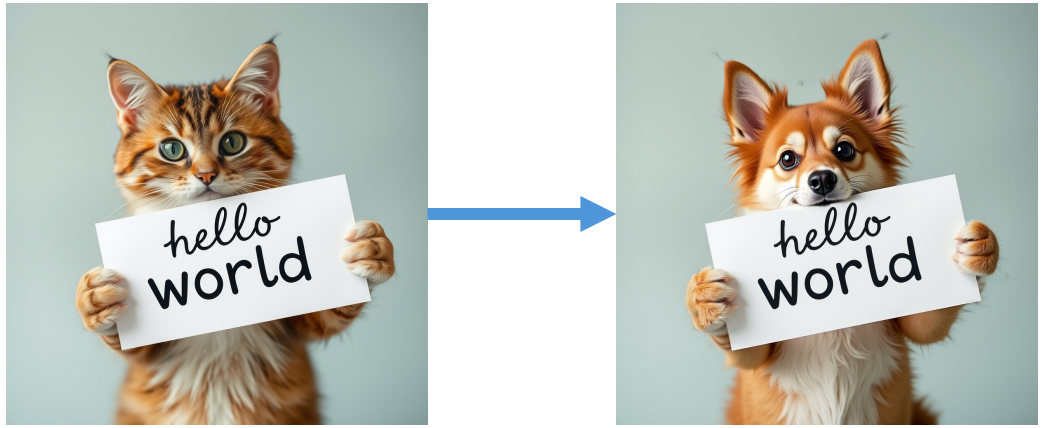


ICLR 2025

Lightning-Fast Image Inversion and Editing for Text-to-Image Diffusion Models

Dvir Samuel, Barak Meiri, Haggai Maron, Yoad Tewel, Nir Darshan, Shai Avidan,
Gal Chechik, Rami Ben-Ari

Image Editing from Text

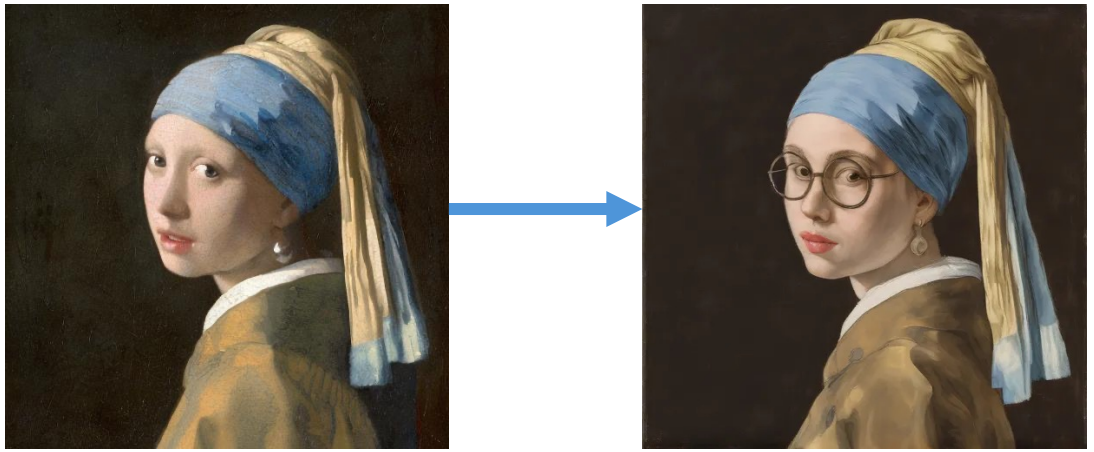


A ~~cat~~ holding up a “Hello world” sign

dog

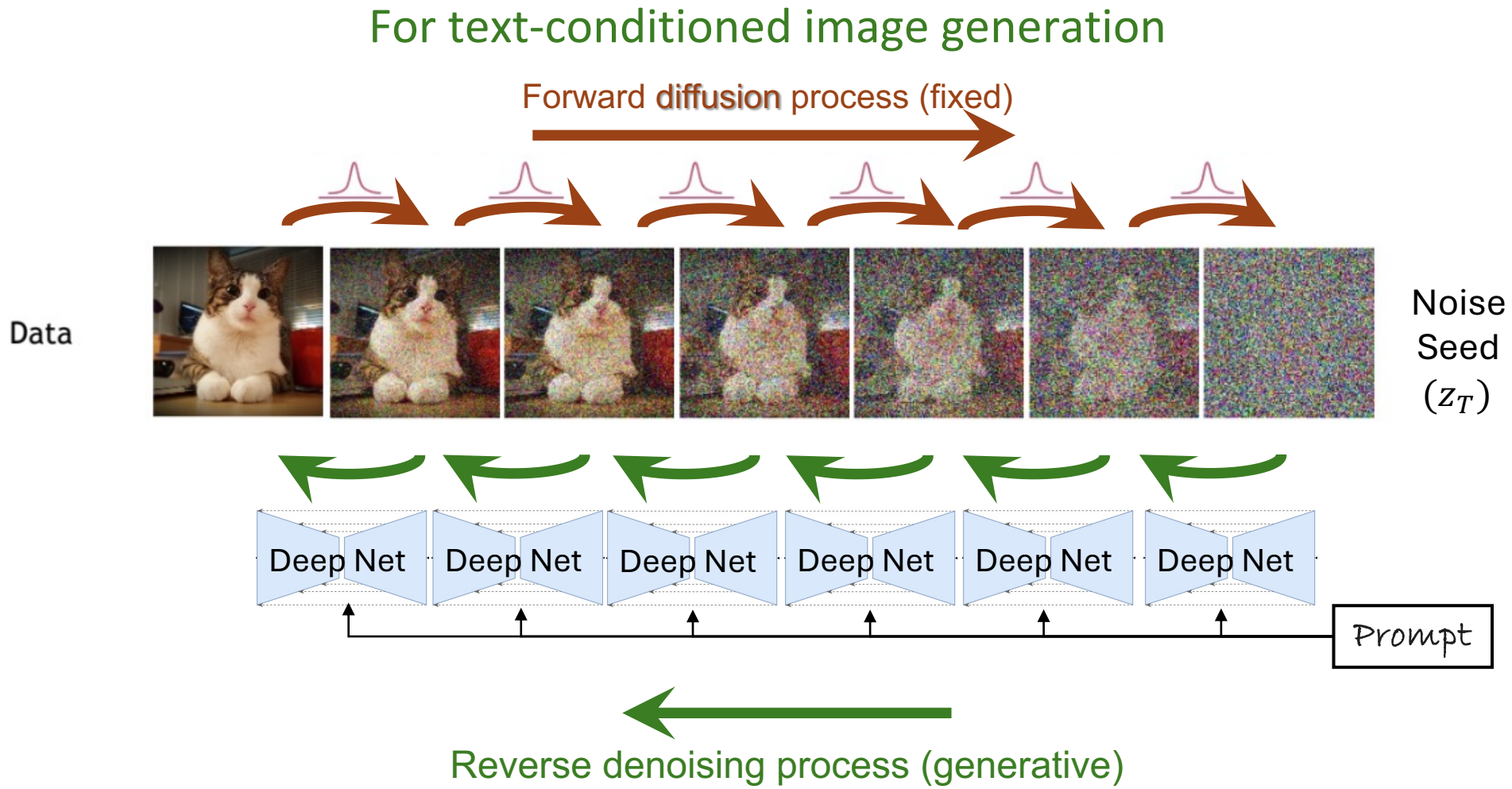
A table with wine, glasses and ~~pizza~~

bread



An oil painting of a woman with glasses

Diffusion Models



Diffusion Models

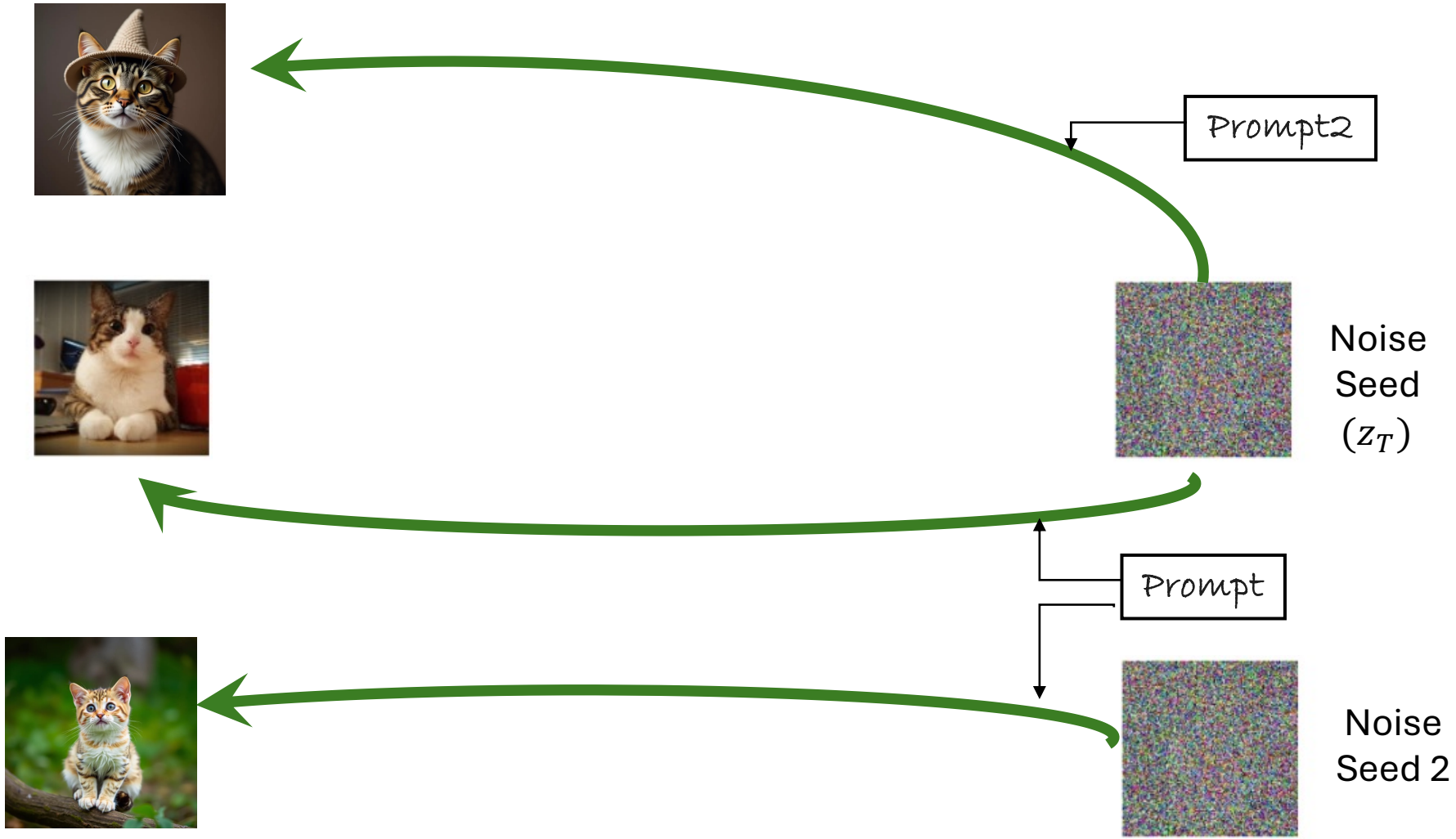
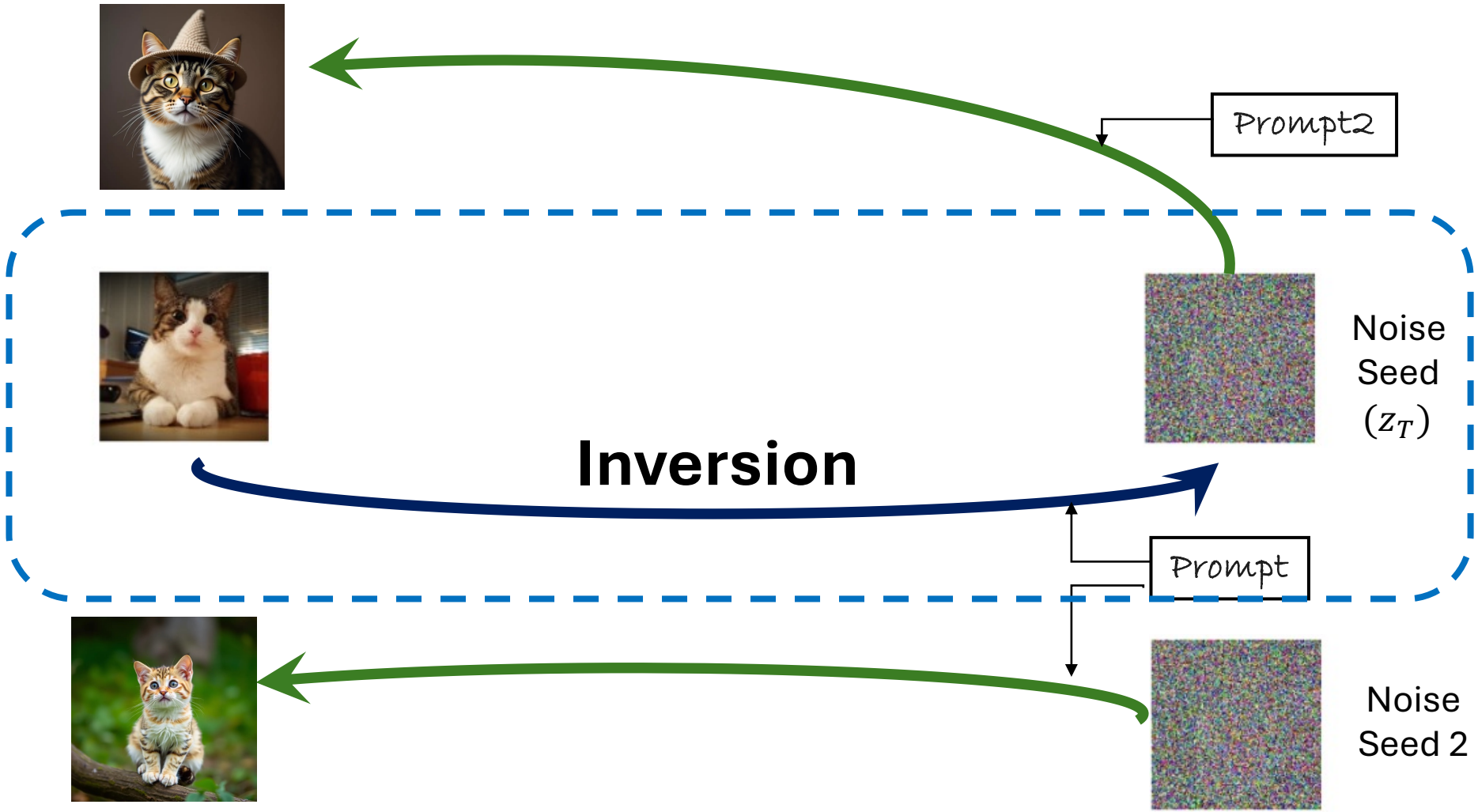
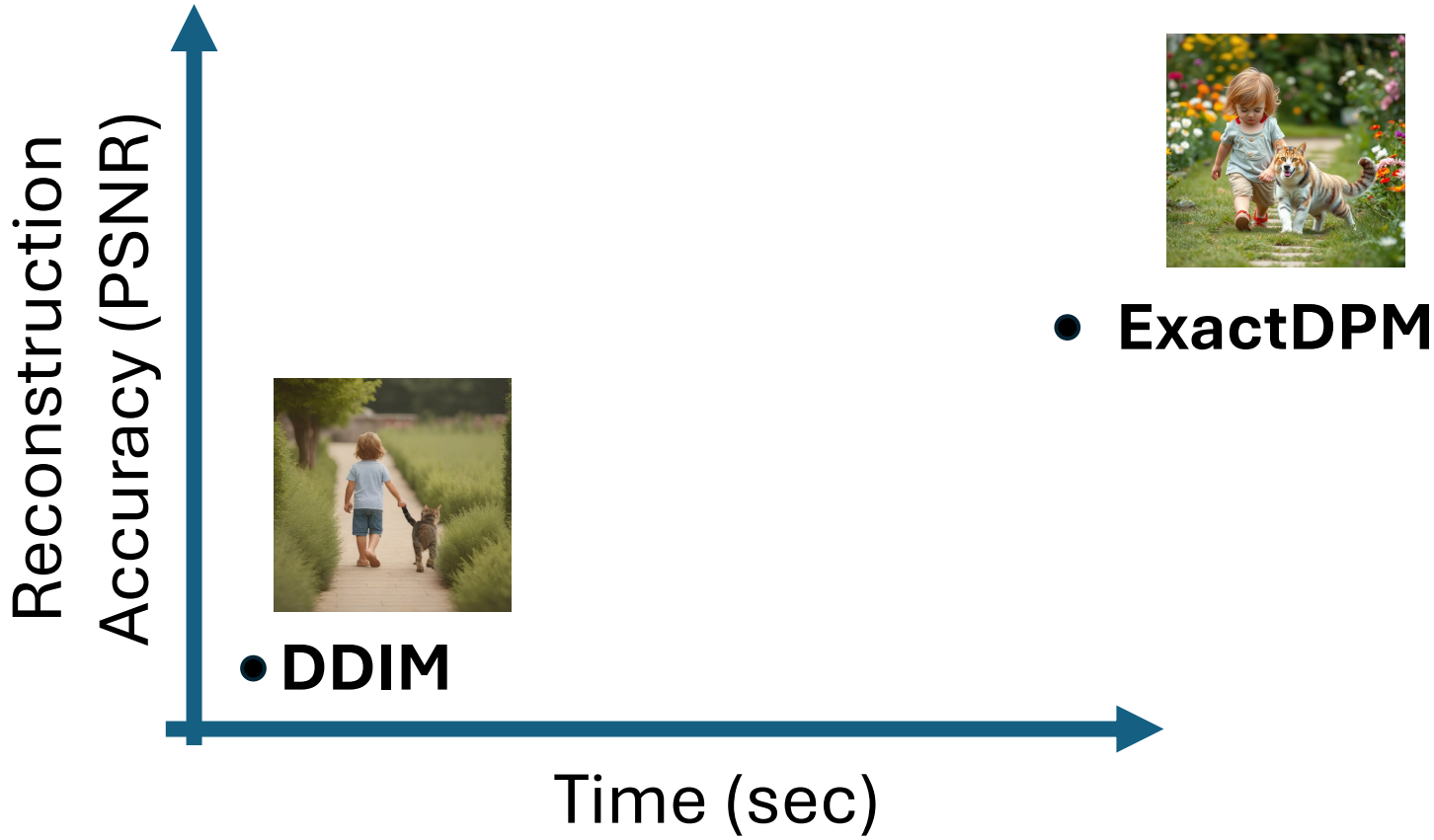


Image Inversion



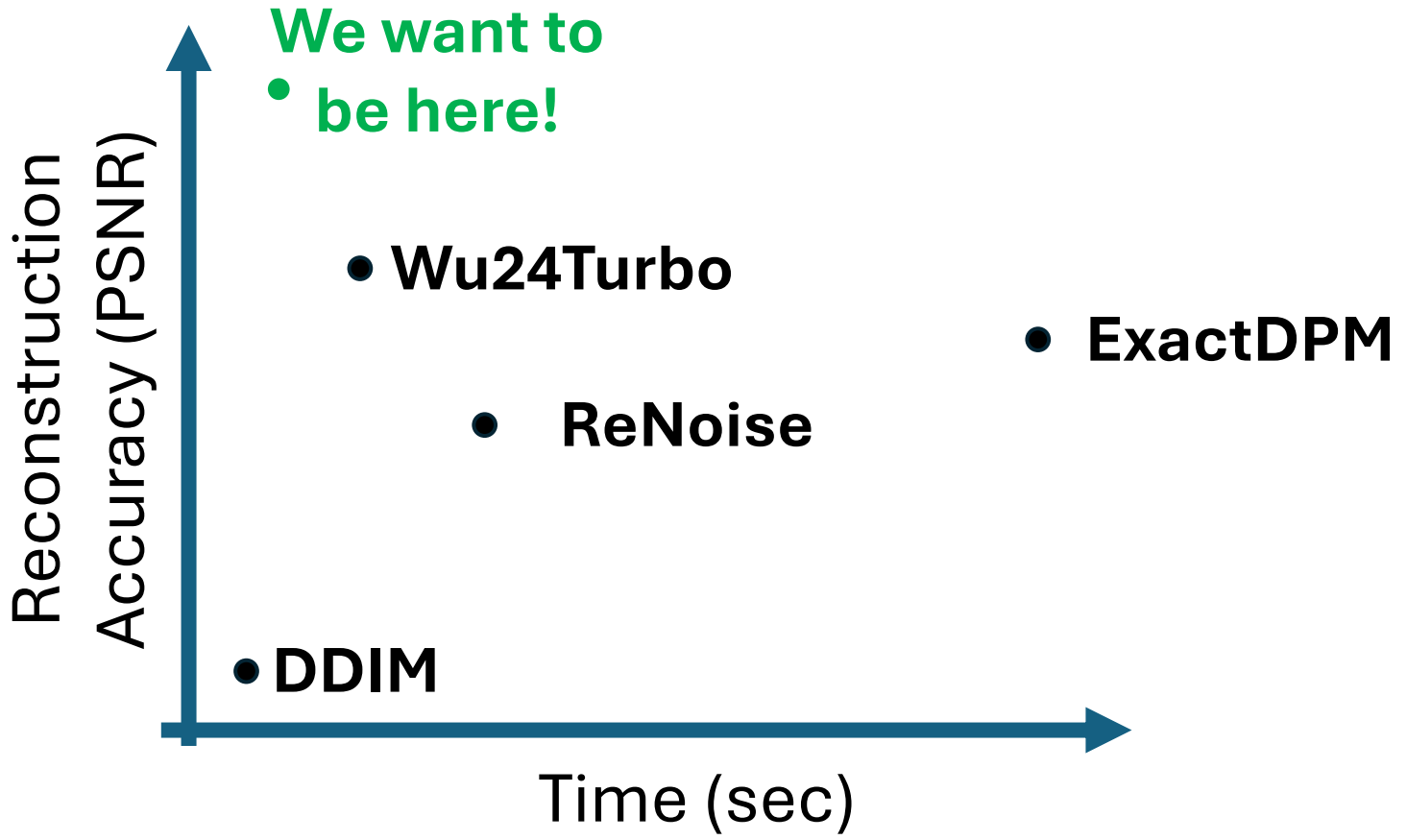
Current Inversion Methods

Input Image

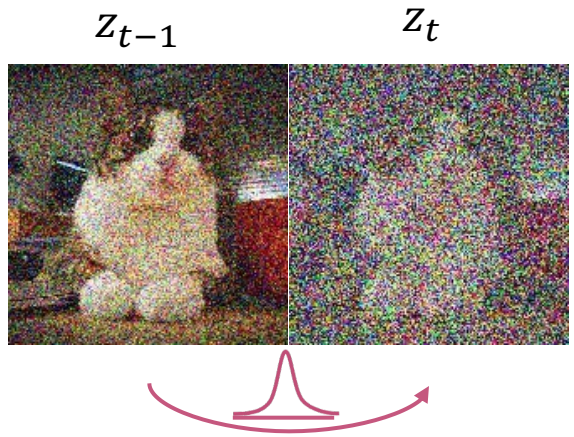


Current Inversion Methods

Input Image



Diffusion Models as Markov Chains



The noising process is a markov chain where each step is Gaussian transition:

$$q(z_t|z_{t-1}) = \mathcal{N}(z_t, \sqrt{\alpha_t} z_{t-1}, (1-\alpha_t)I)$$

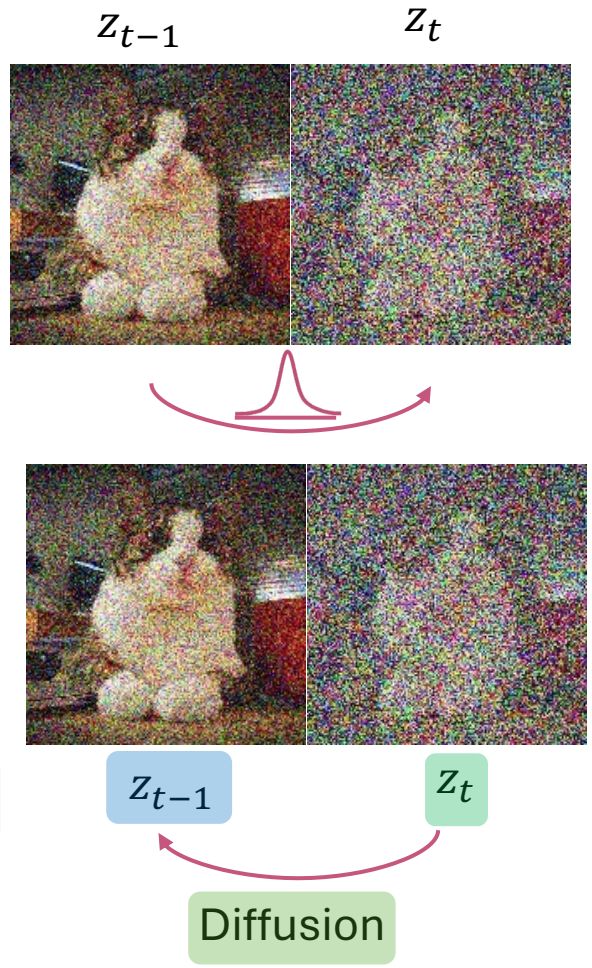
DDIM

The noising process is a markov chain where each step is Gaussian transition:

$$q(z_t|z_{t-1}) = \mathcal{N}(z_t, \sqrt{\alpha_t} z_{t-1}, (1-\alpha_t)I)$$

The denosing process (DDIM):

$$z_{t-1} = \sqrt{\frac{\alpha_{t-1}}{\alpha_t}} z_t - \sqrt{\alpha_{t-1}} \cdot \Delta\psi(\alpha_t) \cdot \epsilon_\theta(z_t, t, p)$$



Diffusion Denoising Implicit Models, Song et al, ICLR21

DDIM Inversion

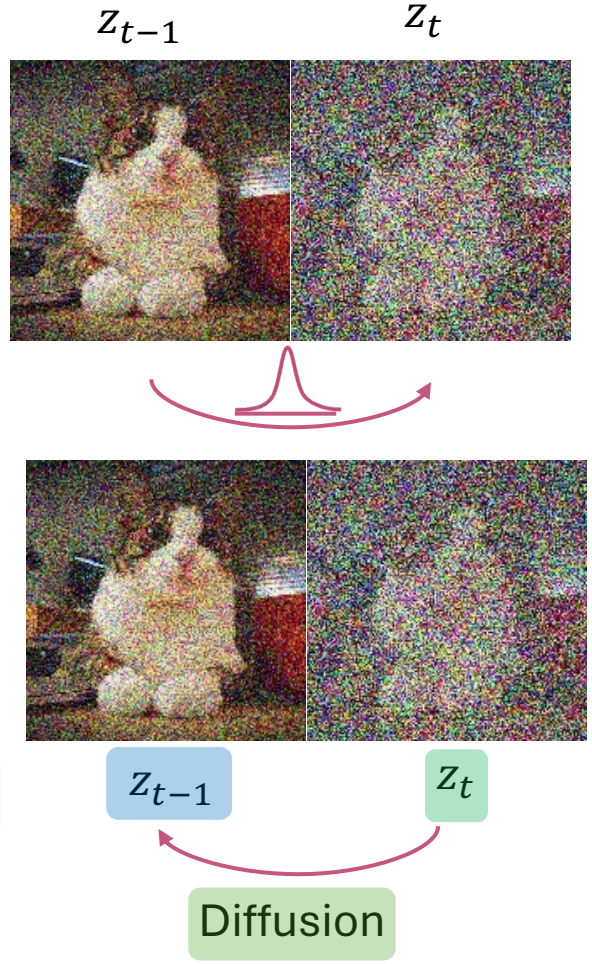
The noising process is a markov chain where each step is Gaussian transition:

$$q(z_t|z_{t-1}) = \mathcal{N}(z_t, \sqrt{\alpha_t} z_{t-1}, (1-\alpha_t)I)$$

The denosing process (DDIM):

$$z_{t-1} = \sqrt{\frac{\alpha_{t-1}}{\alpha_t}} z_t - \sqrt{\alpha_{t-1}} \cdot \Delta\psi(\alpha_t) \cdot \epsilon_\theta(z_t, t, p)$$

Rewrite it as: $z_t = \sqrt{\frac{\alpha_t}{\alpha_{t-1}}} z_{t-1} - \sqrt{\alpha_{t-1}} \cdot \Delta\psi(\alpha_t) \cdot \epsilon_\theta(z_t, t, p)$



Diffusion Denoising Implicit Models, Song et al, ICLR21

DDIM Inversion

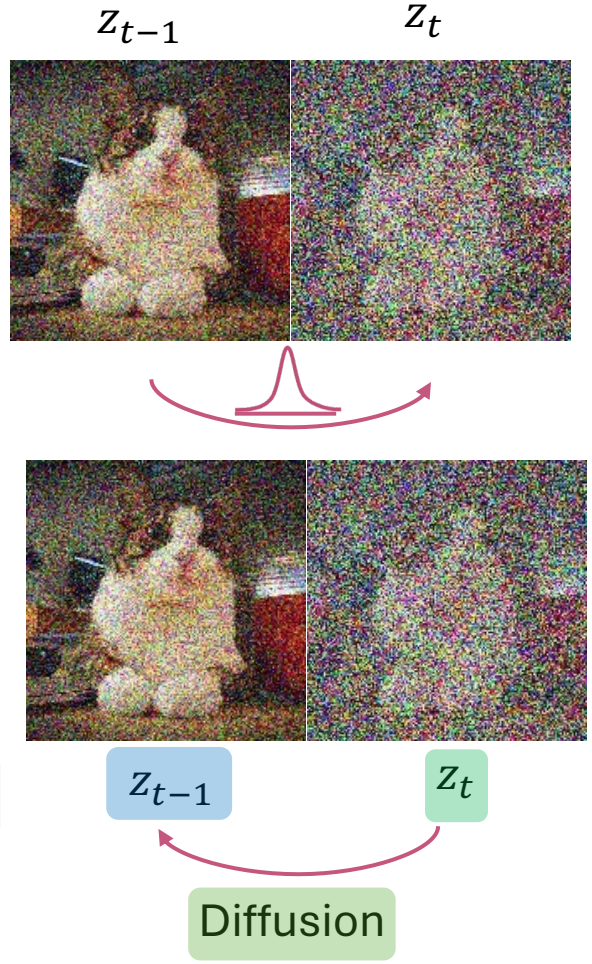
The noising process is a markov chain where each step is Gaussian transition:

$$q(z_t|z_{t-1}) = \mathcal{N}(z_t, \sqrt{\alpha_t} z_{t-1}, (1-\alpha_t)I)$$

The denosing process (DDIM):

$$z_{t-1} = \sqrt{\frac{\alpha_{t-1}}{\alpha_t}} z_t - \sqrt{\alpha_{t-1}} \cdot \Delta\psi(\alpha_t) \cdot \epsilon_\theta(z_t, t, p)$$

Rewrite it as: $z_t = \sqrt{\frac{\alpha_t}{\alpha_{t-1}}} z_{t-1} - \sqrt{\alpha_{t-1}} \cdot \Delta\psi(\alpha_t) \cdot \epsilon_\theta(z_t, t, p)$



Diffusion Denoising Implicit Models, Song et al, ICLR21

DDIM Inversion

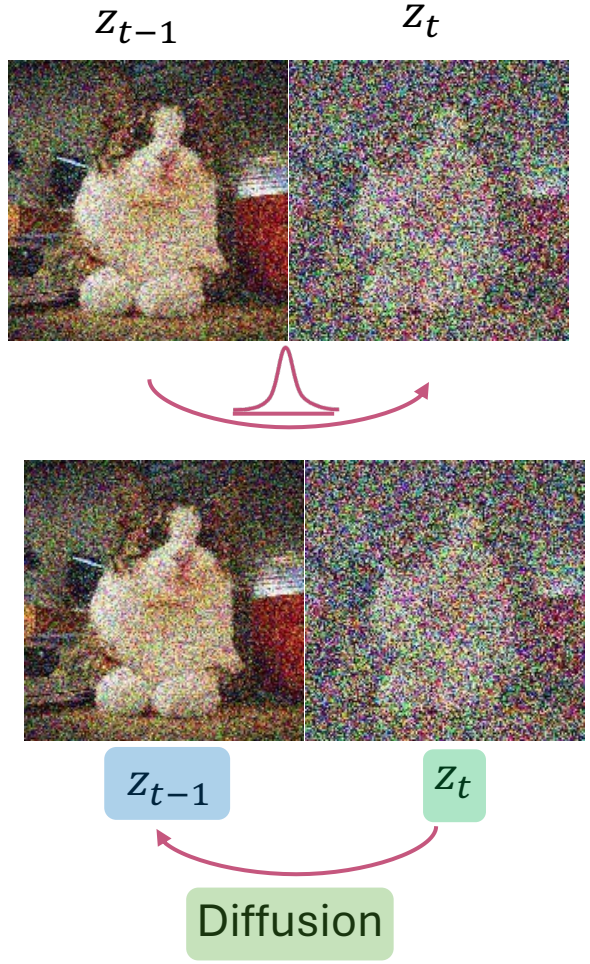
The noising process is a markov chain where each step is Gaussian transition:

$$q(z_t|z_{t-1}) = \mathcal{N}(z_t, \sqrt{\alpha_t} z_{t-1}, (1-\alpha_t)I)$$

The denosing process (DDIM):

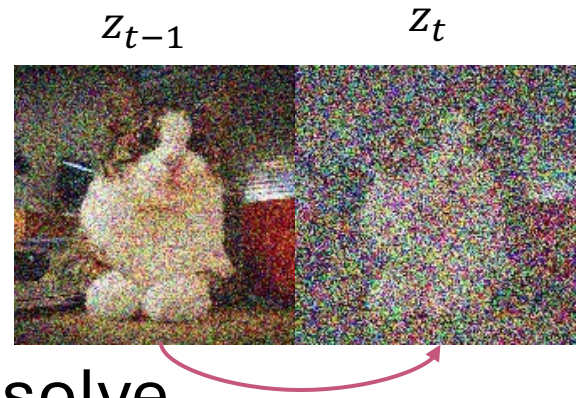
$$z_{t-1} = \sqrt{\frac{\alpha_{t-1}}{\alpha_t}} z_t - \sqrt{\alpha_{t-1}} \cdot \Delta\psi(\alpha_t) \cdot \epsilon_\theta(z_t, t, p)$$

Rewrite it as: $z_t = \sqrt{\frac{\alpha_t}{\alpha_{t-1}}} z_{t-1} - \sqrt{\alpha_{t-1}} \cdot \Delta\psi(\alpha_t) \cdot \epsilon_\theta(z_{t-1}, t, p)$



Diffusion Denoising Implicit Models, Song et al, ICLR21

Recent Inversion Methods



Frame inversion a root a root-finding problem and solve $r(z_t) = z_t - f(z_t)$ using:

$$z_t = \underbrace{\sqrt{\frac{\alpha_t}{\alpha_{t-1}}}\, z_{t-1} - \sqrt{\alpha_{t-1}} \cdot \Delta\psi(\alpha_t) \cdot \epsilon_\theta(z_t, t, p)}_{f(z_t)}$$

Solve it using:

- Fixed-point iteration Pan et al., ICCV23 and Garibi et al., ECCV24
- Gradient Descent Hong et al., CVPR24

Our key ideas

1. **Use an Efficient Newton-Raphson Iteration**

   Apply Newton Raphson on the norm of r: $\hat{r}(z_t) = ||z_t - f(z_t)||$. This reduces the Jacobian matrix into a vector, making it easy and fast for computation.

Our key ideas

1. **Use an Efficient Newton-Raphson Iteration**

   Apply Newton Raphson on the norm of r: $\hat{r}(z_t) = ||z_t - f(z_t)||$. This reduces the Jacobian matrix into a vector, making it easy and fast for computation.

2. **Guide Newton-Raphson** to in-distribution solutions

   We add a prior over the value of z_t:

$$\mathcal{F}(z_t) \coloneqq \|z_t - f(z_t)\| - \lambda \log(q(z_t \mid z_{t-1}))$$

Known by model definition

   λ is a weighting parameter for the guidance term.

Our key ideas

1. **Use an Efficient Newton-Raphson Iteration**

   Apply Newton Raphson on the norm of r: $\hat{r}(z_t) = ||z_t - f(z_t)||$. This reduces the Jacobian matrix into a vector, making it easy and fast for computation.

2. **Guide Newton-Raphson** to in-distribution solutions

   We add a prior over the value of z_t:

$$\mathcal{F}(z_t) \coloneqq \|z_t - f(z_t)\| - \lambda \log(q(z_t \mid z_{t-1}))$$

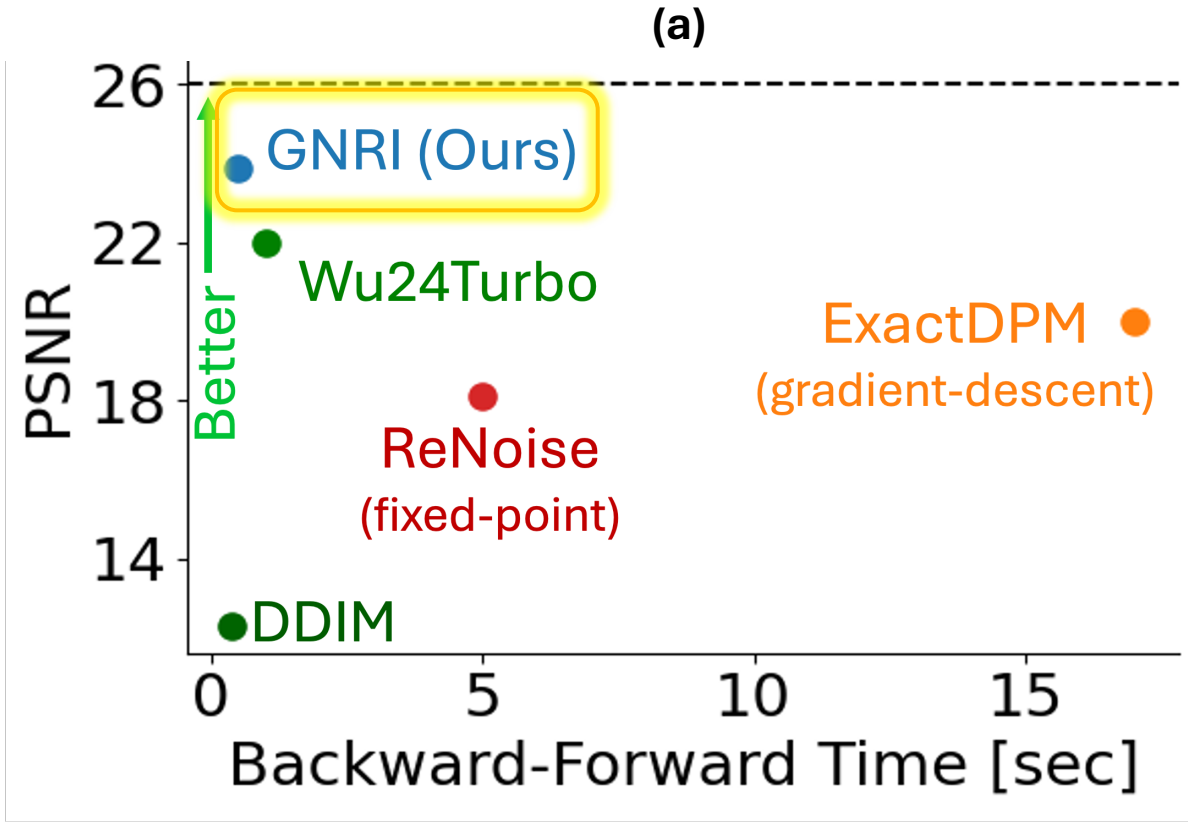
Known by model definition

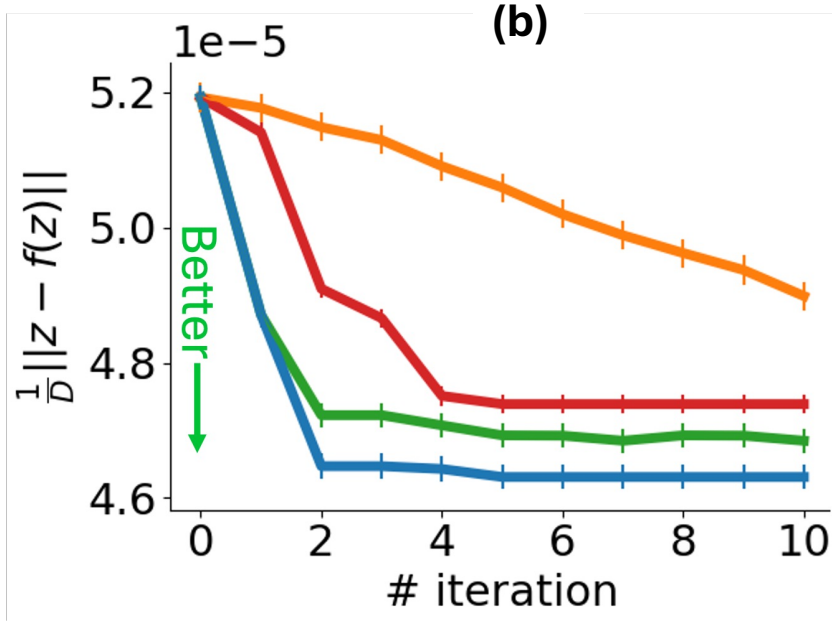
λ is a weighting parameter for the guidance term.

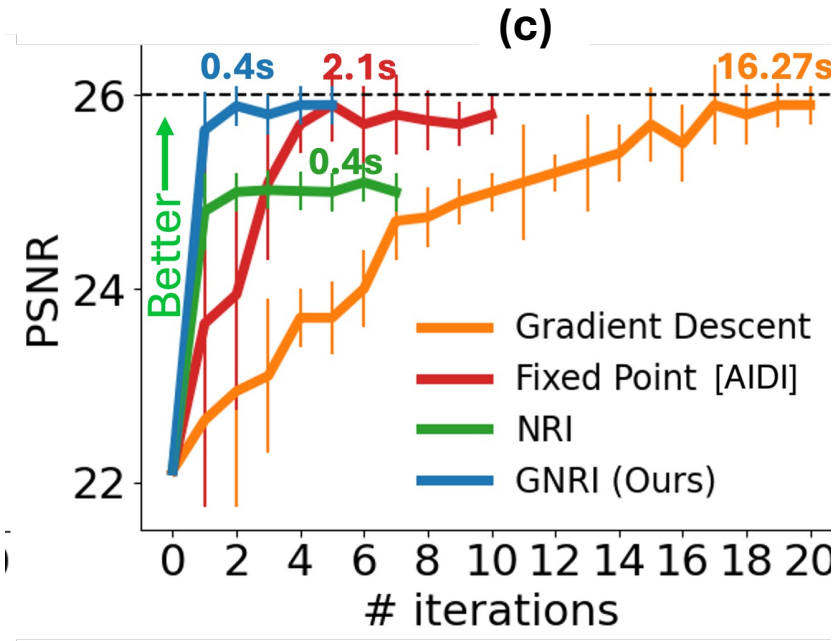
→ **G**uided **N**ewton **R**aphson **I**nversion

Reconstruction Quality

GNRI converges within 1-2 iterations!

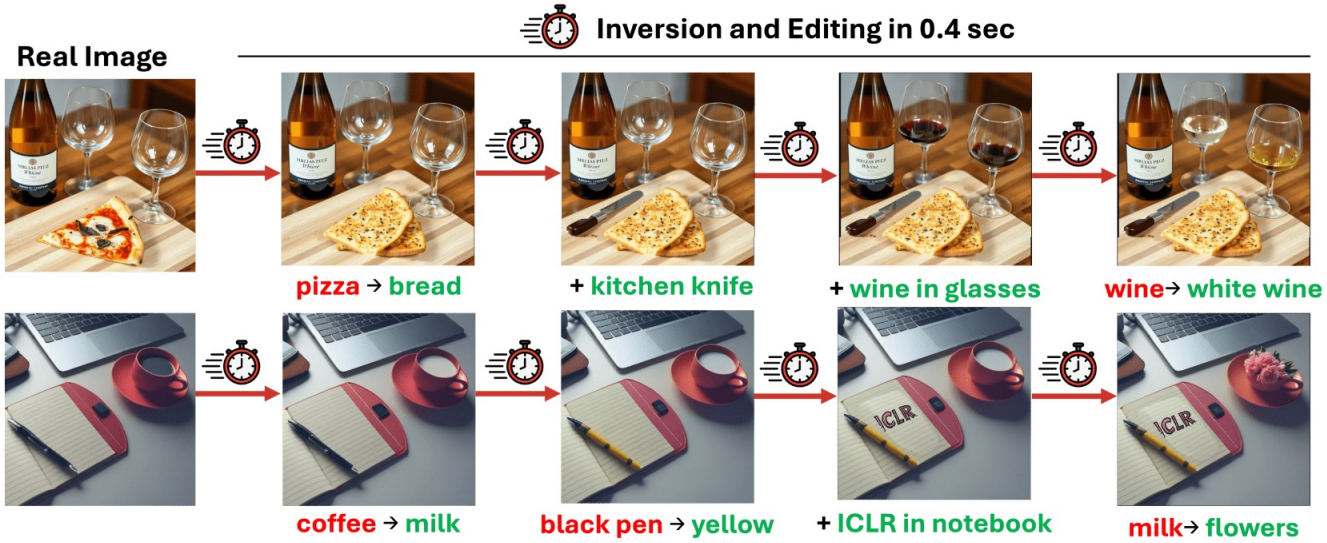


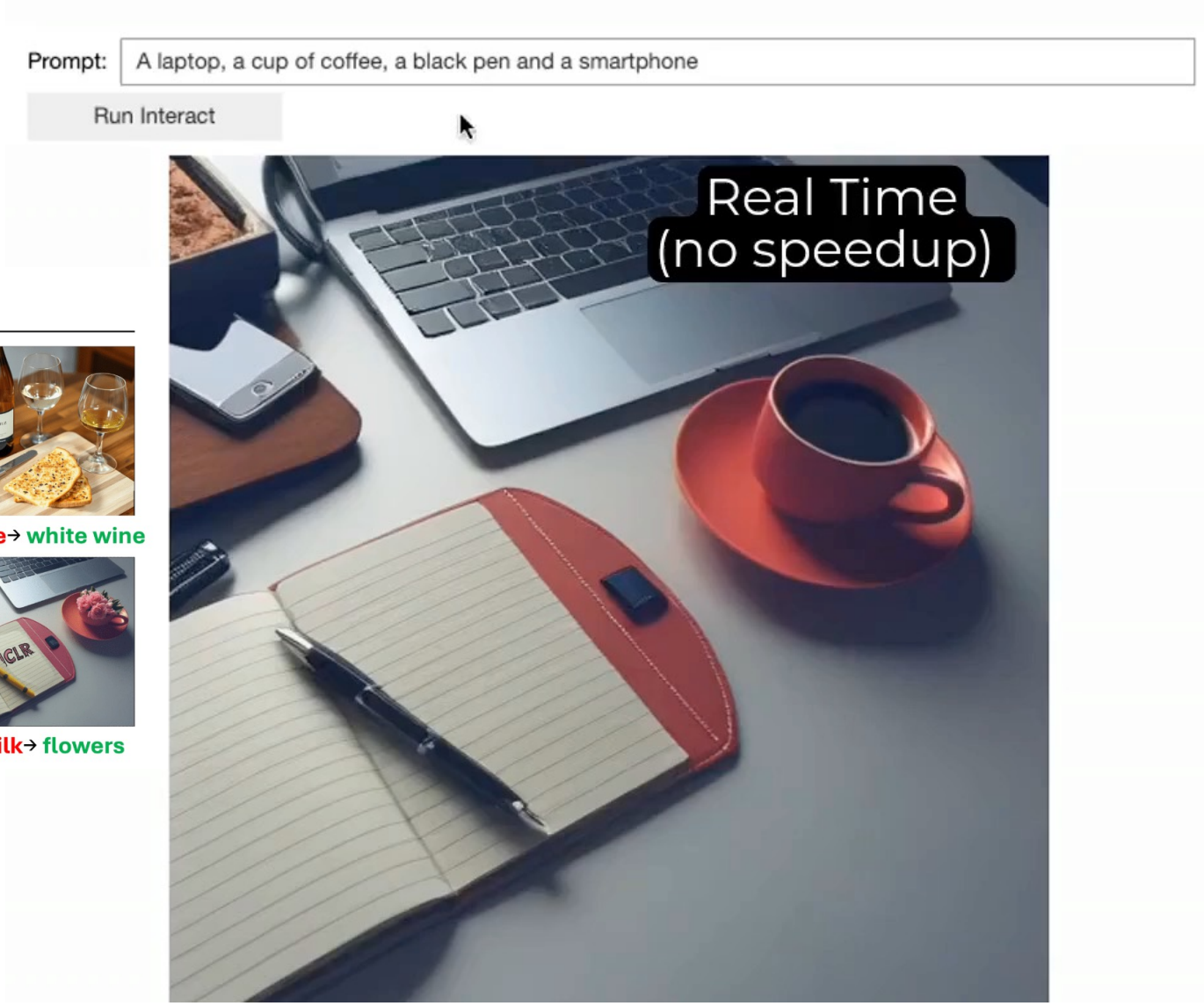




Real-time Image Editing

GNRI enables interactive editing!!





Project Page

Demo

Thank you!