Noise Map Guidance: Inversion with Spatial Context for Real Image Editing

Hansam Cho[1,2], Jonghyun Lee[1,2], Seoung Bum Kim[2], Tae-Hyun Oh[3,4], Yonghyun Jeong[2]

1 School of Industrial and Management Engineering, Korea University, 2 Naver Cloud,
3 Dept. of Electrical Engineering and Grad. School of Artificial Intelligence, POSTECH,
4 Institute for Convergence Research and Education in Advanced Technology, Yonsei University

Introduction

Real Image Editing with Diffusion Models

- Diffusion models emerge as a powerful generation tool and apply to real image editing

Stable Diffusion
(Rombach et al. CVPR 2022)

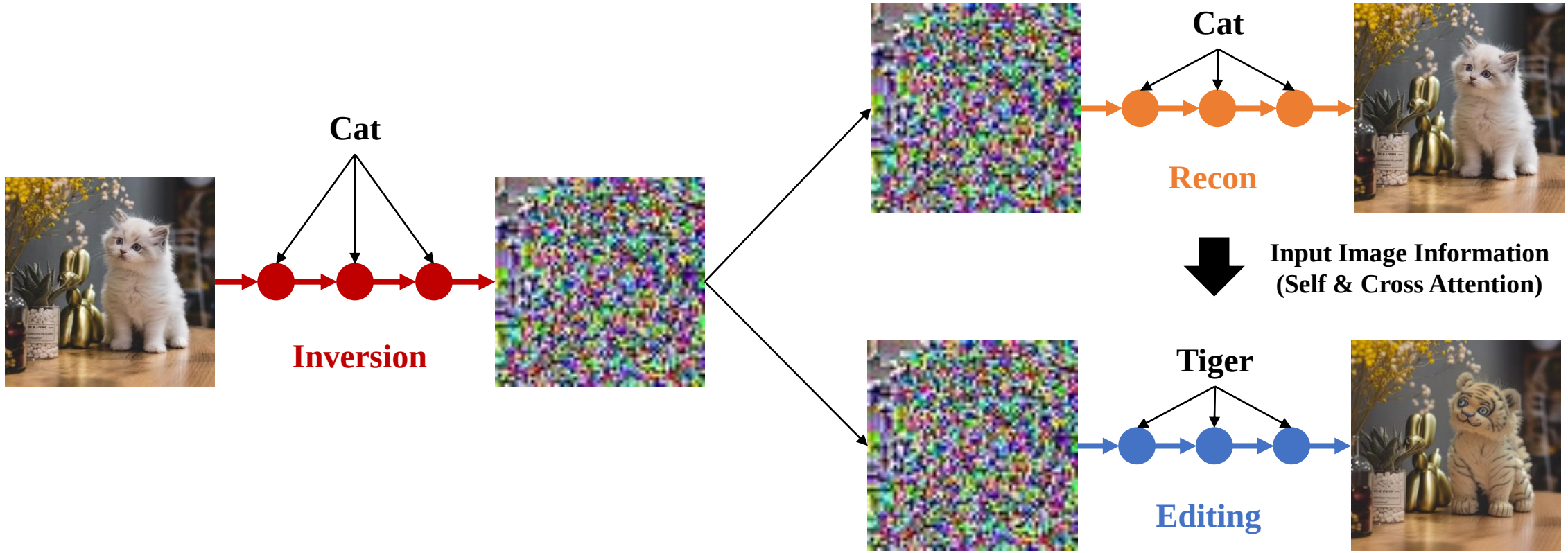
Flower → cupcake

Null-text Inversion
(Mokady et al. CVPR 2023)

Introduction

Real Image Editing with Diffusion Models

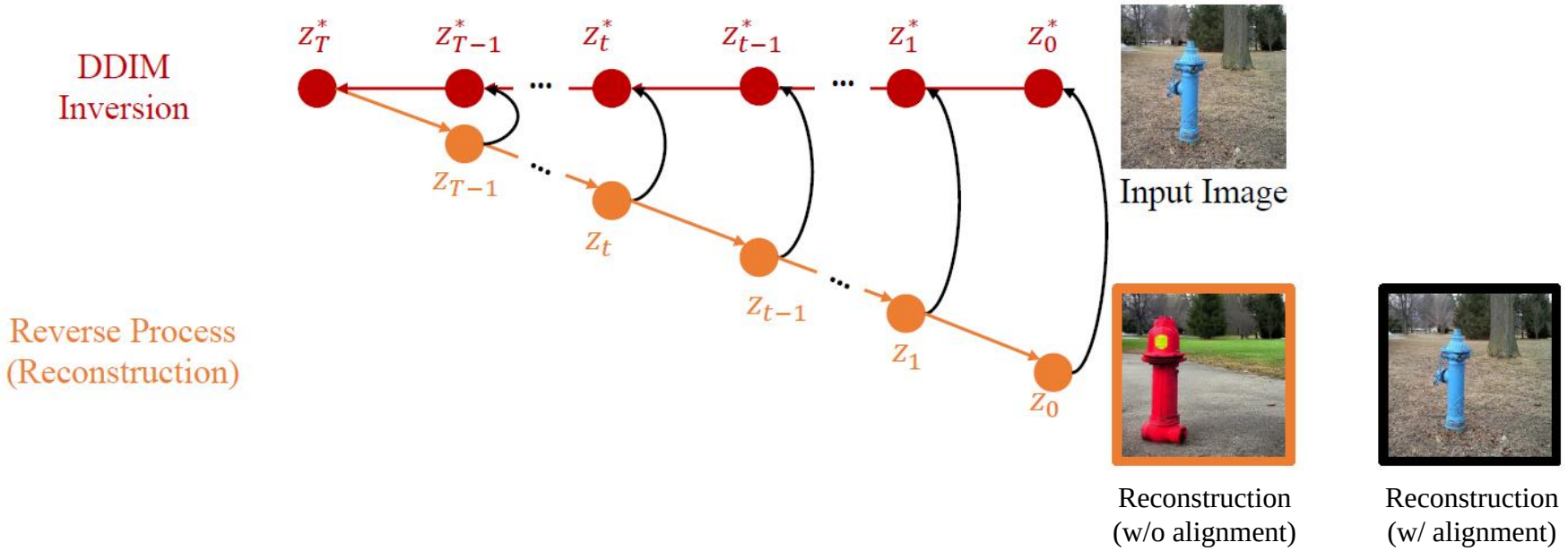
- 1) Inversion 2) Reconstruction and Editing
- Reconstruction is crucial for precise real image editing



Introduction

Real Image Editing with Diffusion Models

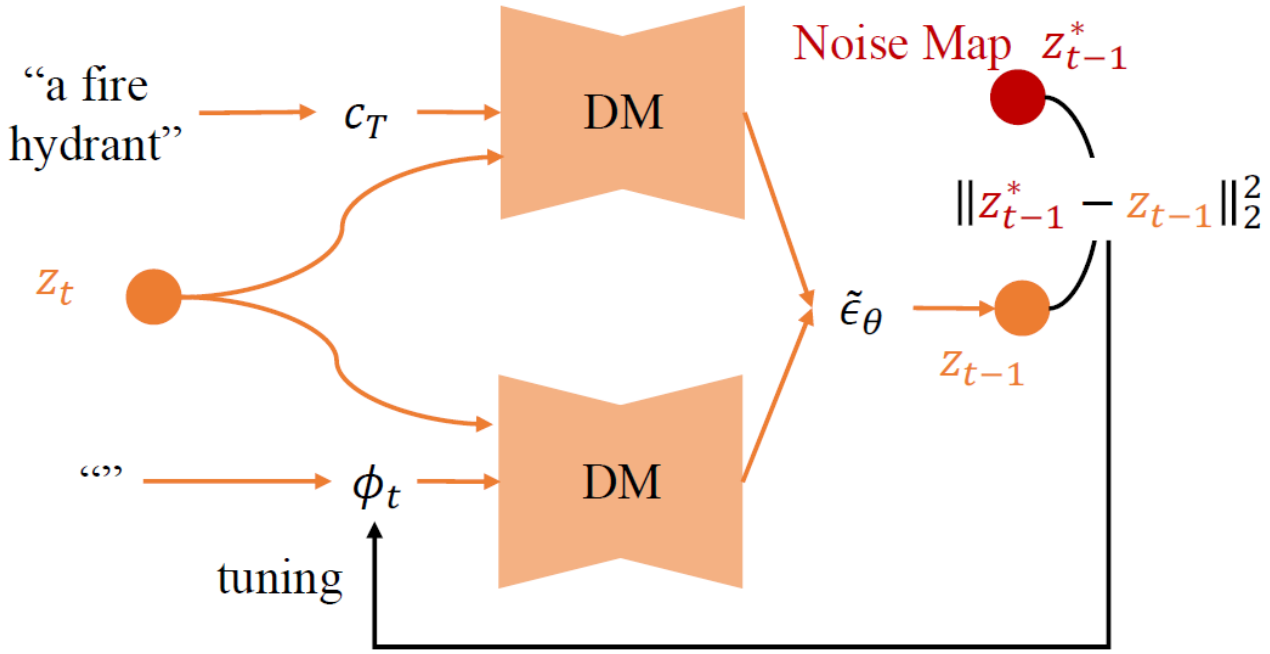
- Naive reconstruction is diverged from the inversion path because of Classifier-Free Guidance (CFG)



Related Work

Inversion with Diffusion Models

- The existing inversion method optimizes the null-text embedding to mitigate divergence
- However, text-embedding is one-dimensional vector → hard to preserve spatial context
- Optimization approach → time-consuming

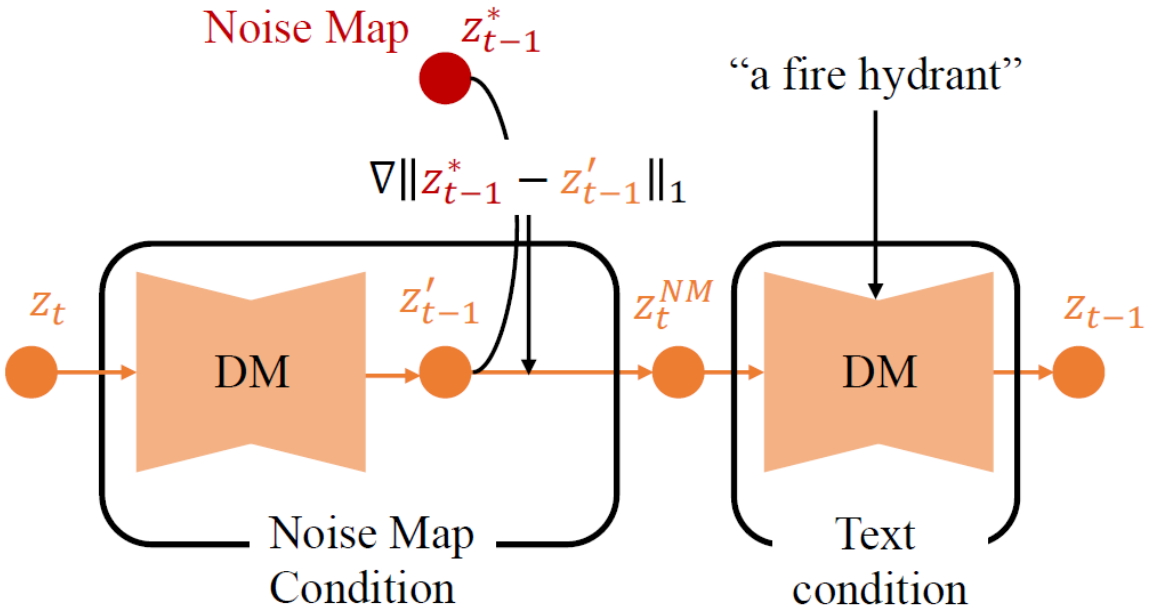


Null-text Inversion (Optimization)

Method

Noise Map Guidance (Ours)

- Preserve the spatial context → use noise map directly
- Fast editing → conditional approach (optimization-free)



Noise Map Guidance (Condition)

Method

Noise Map Guidance (Ours)

- Preserve the spatial context → use noise map directly
- Fast editing → conditional approach (optimization-free)

$$z_{t-1} = \sqrt{\frac{\alpha_{t-1}}{\alpha_t}} z_t + \sqrt{\alpha_{t-1}} \left(\sqrt{\frac{1}{\alpha_t - 1} - 1} - \sqrt{\frac{1}{\alpha_t} - 1} \right) \tilde{\epsilon}_\theta(z_t, c_T)$$

Reverse Process

$$z_{t-1} = \sqrt{\frac{\alpha_{t-1}}{\alpha_t}} z_t^{NM} + \sqrt{\alpha_{t-1}} \left(\sqrt{\frac{1}{\alpha_t - 1} - 1} - \sqrt{\frac{1}{\alpha_t} - 1} \right) \tilde{\epsilon}_\theta(z_t^{NM}, c_T)$$

Conditional Reverse Process

$$z_t^{NM} = \sqrt{\frac{\alpha_{t-1}}{\alpha_t}} z_t + \sqrt{\alpha_{t-1}} \left(\sqrt{\frac{1}{\alpha_t - 1} - 1} - \sqrt{\frac{1}{\alpha_t} - 1} \right) \tilde{\epsilon}_\theta(z_t, c_N)$$

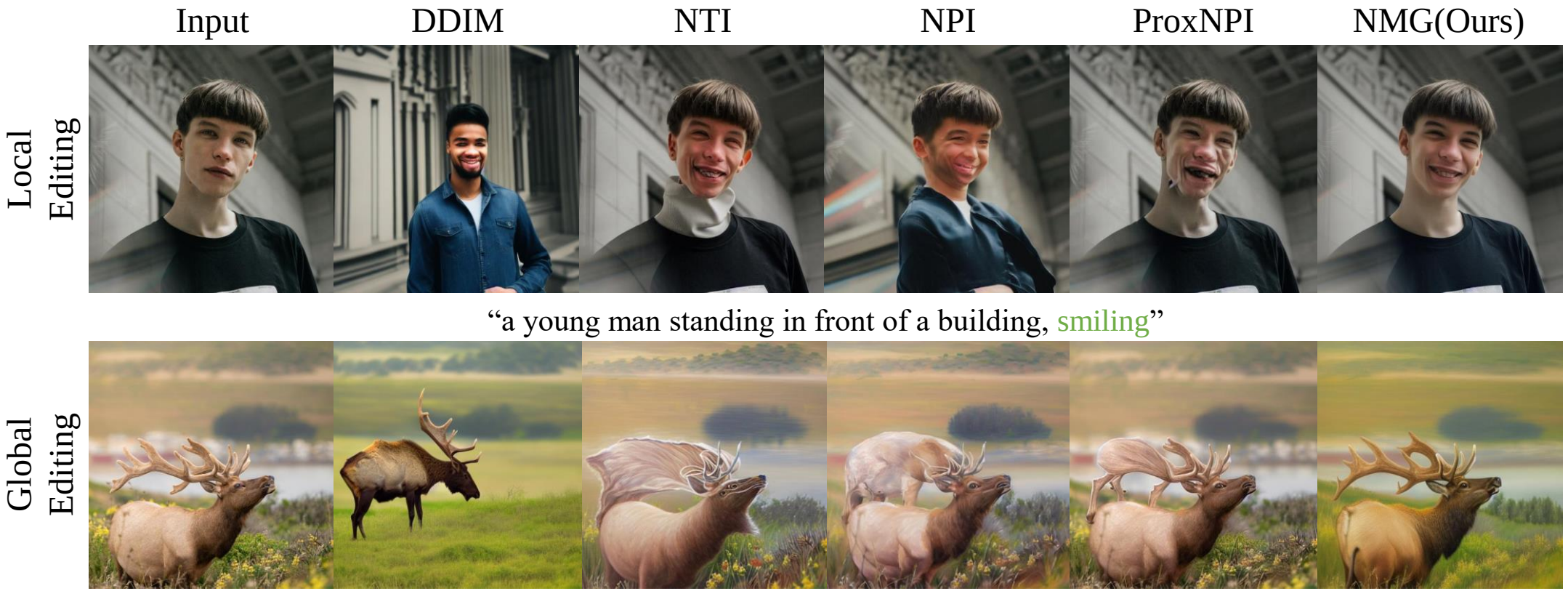
$$\tilde{\epsilon}_\theta(z_t, c_N) = -\sqrt{1 - \alpha_t} (\nabla_{z_t} \log p(z_t) + s_g \cdot \nabla_{z_t} \| z'_{t-1} - z^*_{t-1} \|_1)$$

Recon **Inversion**

Experiments

Noise Map Guidance + Prompt-to-Prompt

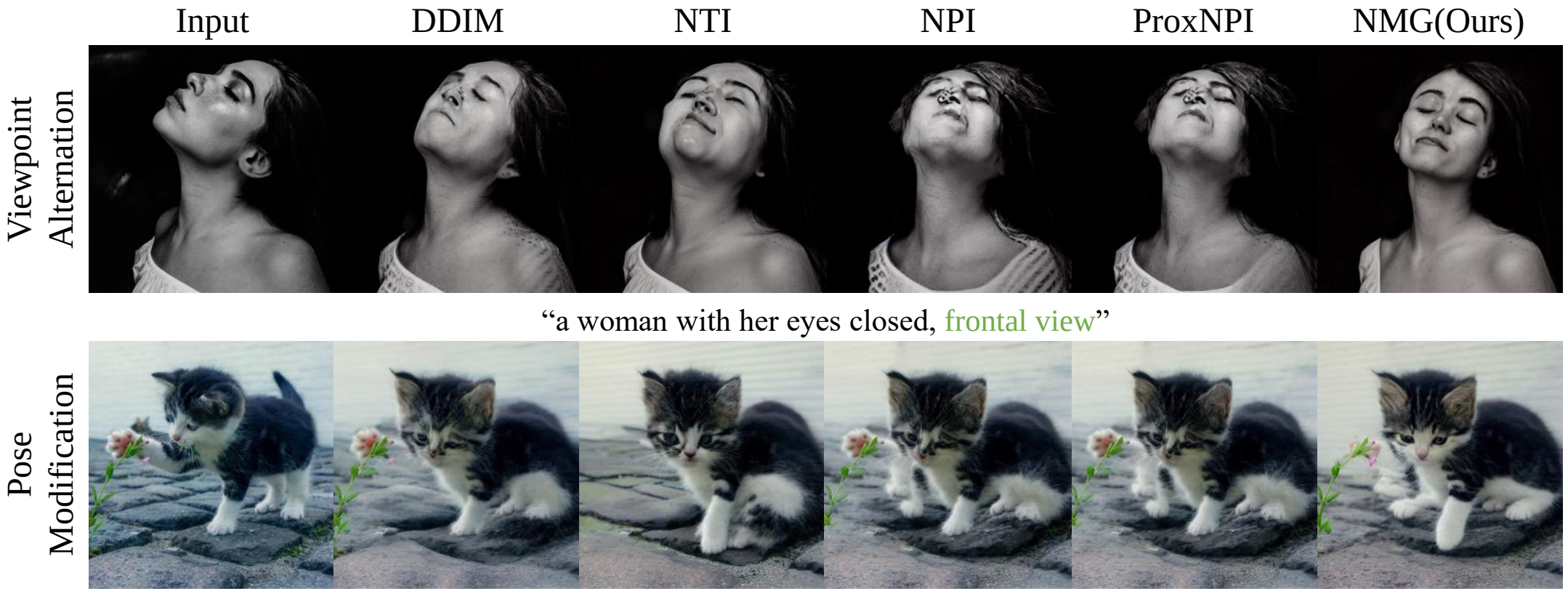
- Local and global editing (editing method: **Prompt-to-Prompt**)



Experiments

Noise Map Guidance + MasaCtrl

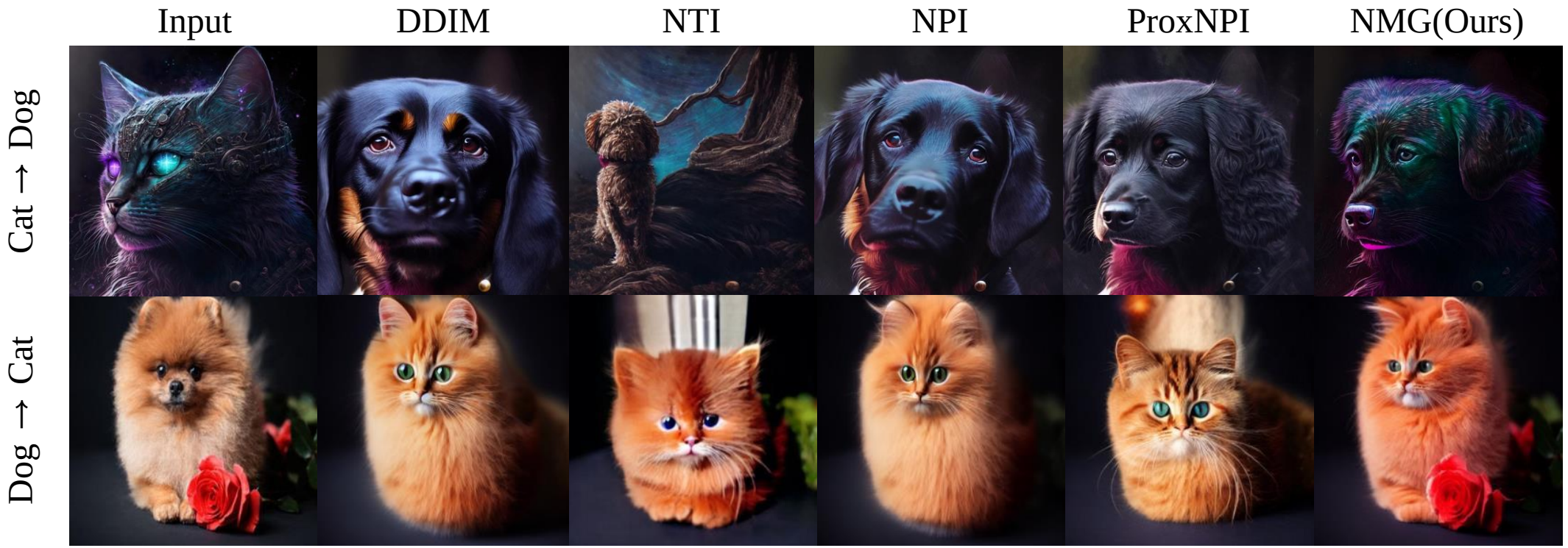
- Non-rigid editing (editing method: **MasaCtrl**)



Experiments

Noise Map Guidance + pix2pix-zero

- Variations of DDIM inversion (editing method: **pix2pix-zero**)



Thank You!!

Code: https://github.com/hansam95/NMG