Object-aware Inversion and Reassembly for Image Editing

Zhen Yang[1], Ganggui Ding[1], Wen Wang[1], Hao Chen[1], Bohan Zhuang[2], Chunhua Shen[1]

1. Zhejiang University, 2. Monash University

Project Website

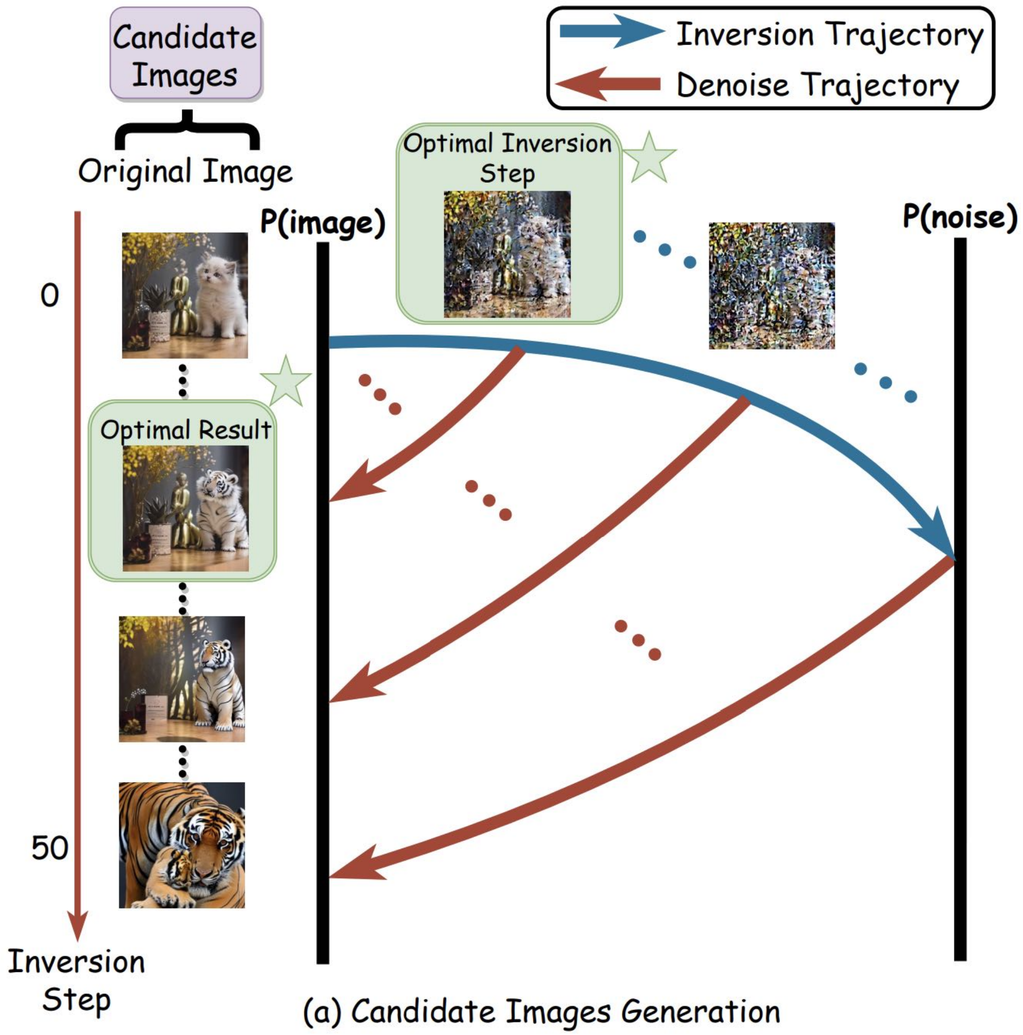
OpenReview

WeChat

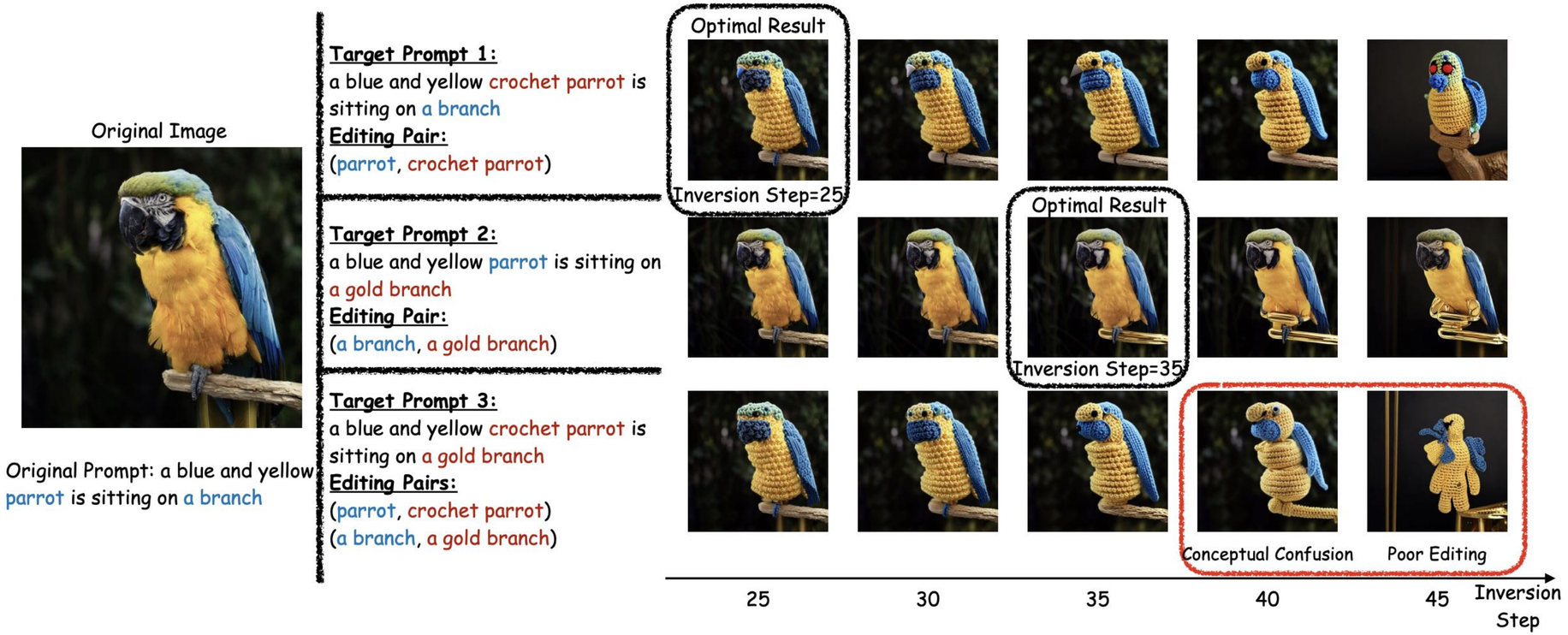
Twitter

Start Point

- Collect candidate images and manually select the optimal editing result.
- From t=0 to t=50, the editing result ranges from **"fidelity"** to **"editability"**.
- **People ignore "t".**
- Two technical problems: **speed** and **manually**. 🐵



Motivation



- 1️⃣ Different editing pairs require **different inversion steps** !!! → Search Metric
- 2️⃣ Editing multiple pairs simultaneously can result in **concept mismatch** and **poor editing** !!! → OIR 🐵

1 Different editing pairs require different inversion steps !!!

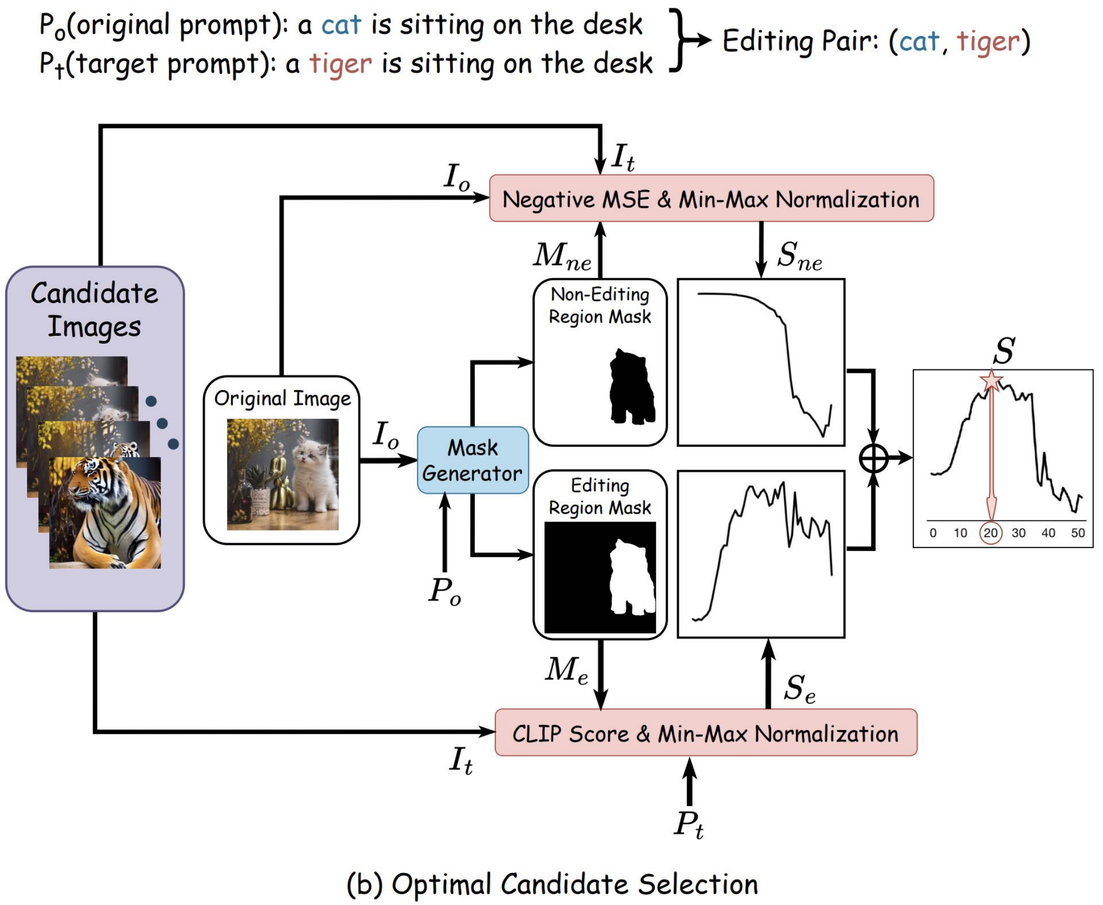
How to automatically select the optimal inversion step ? (Search metric)

- Inspired by human observation, we hope that the **editing region will be aligned with the target prompt**, and the **non-editing region will be consistent to the original image**.
- Editing region metric: CLIP Score (↑ ,0→50)
- Non-editing region metric: Negative MSE(↓ ,0→50)

$$S_e\left(I_t, P_t, M_e\right) = \text{normalize}\left(\frac{CLIP_{image}\left(I_t, M_e\right) \cdot CLIP_{text}\left(P_t\right)}{\left\| CLIP_{image}\left(I_t, M_e\right) \right\|_2 \cdot \left\| CLIP_{text}\left(P_t\right) \right\|_2}\right)$$

$$S_{ne}\left(I_t, I_o, M_{ne}\right) = \text{normalize}\left(-\left\| \left(I_t - I_o\right) \odot M_{ne} \right\|_2^2\right)$$

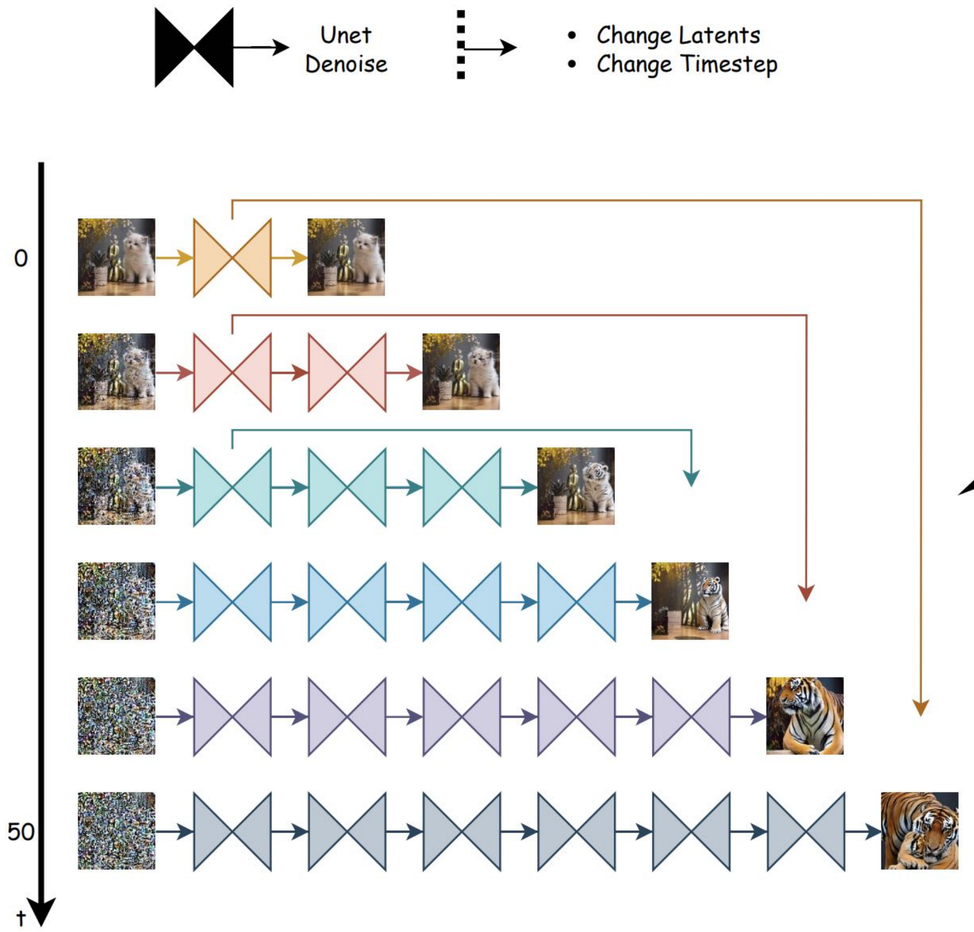
$$Search\ Metric = 0.5 \cdot \left(S_e + S_{ne}\right)$$



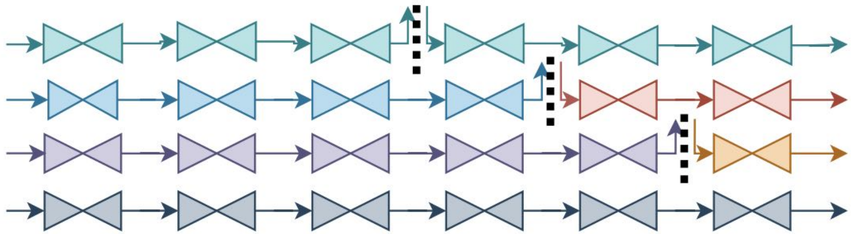
(b) Optimal Candidate Selection

1 Different editing pairs require different inversion steps !!!

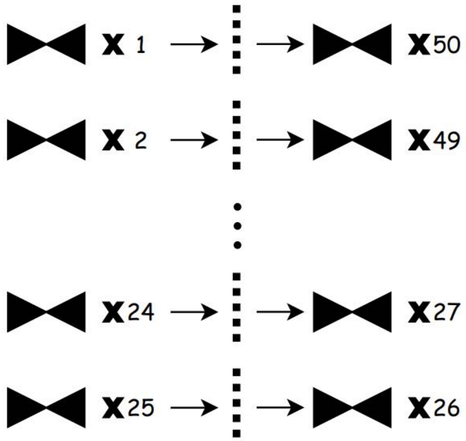
How to speed up ?



(a) Serial Denoise Methods



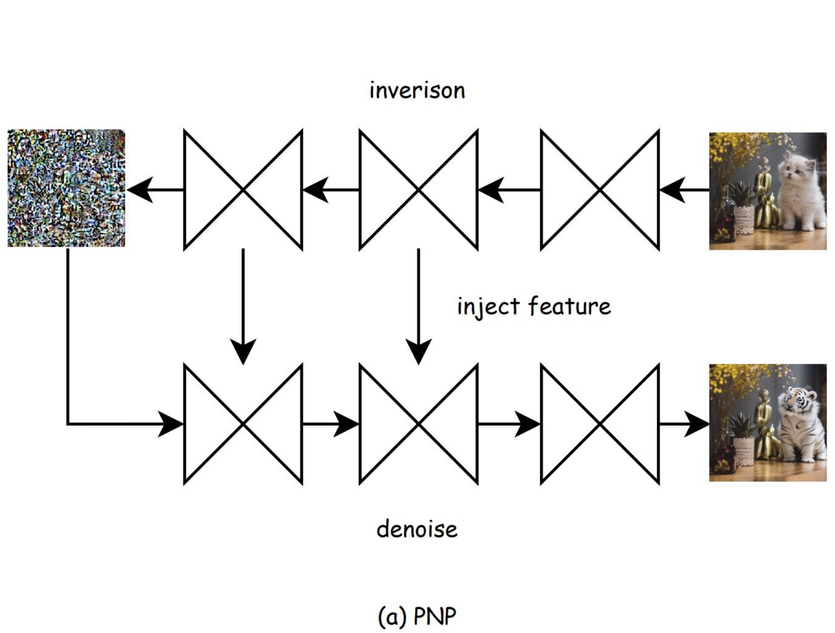
(b) Parallel Denoise Methods



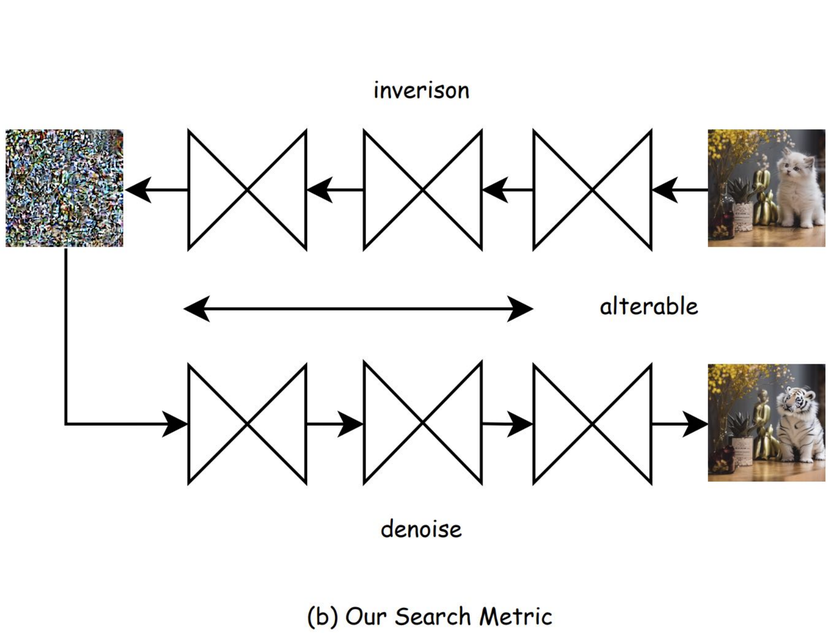
(c) Parallel Denoise method Extends To 51-Step Denoise

1 Different editing pairs require different inversion steps !!!

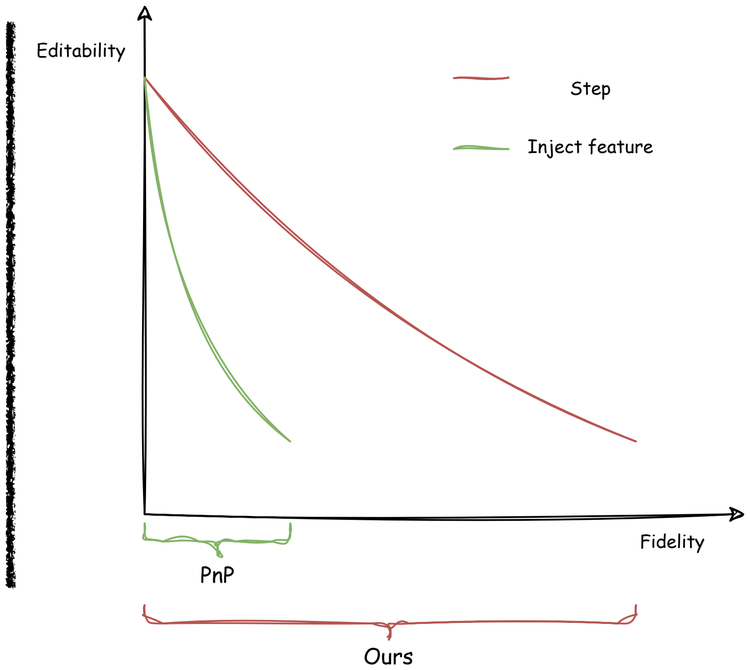
Why our novel method can work ?



(a) PNP



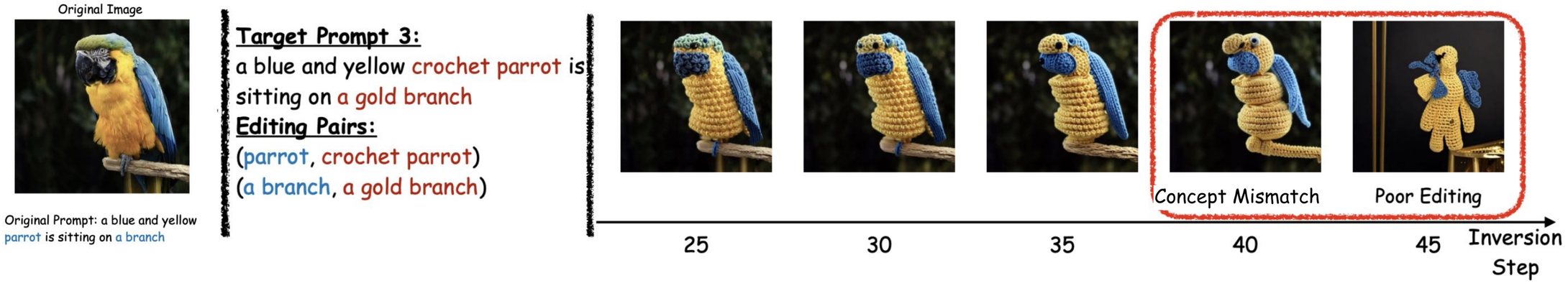
(b) Our Search Metric



- PNP **injects features** to control "editability" and "fidelity".
- Search metric **changes the timestep** to control "editability" and "fidelity".
- When **PNP injects nothing**, PNP is equivalent to the **search metric with timestep=0**. When **PNP injects all features**, PNP is equivalent to the **search metric with timestep=50**.

2 How to solve "concept mismatch" and "poor editing"!!!

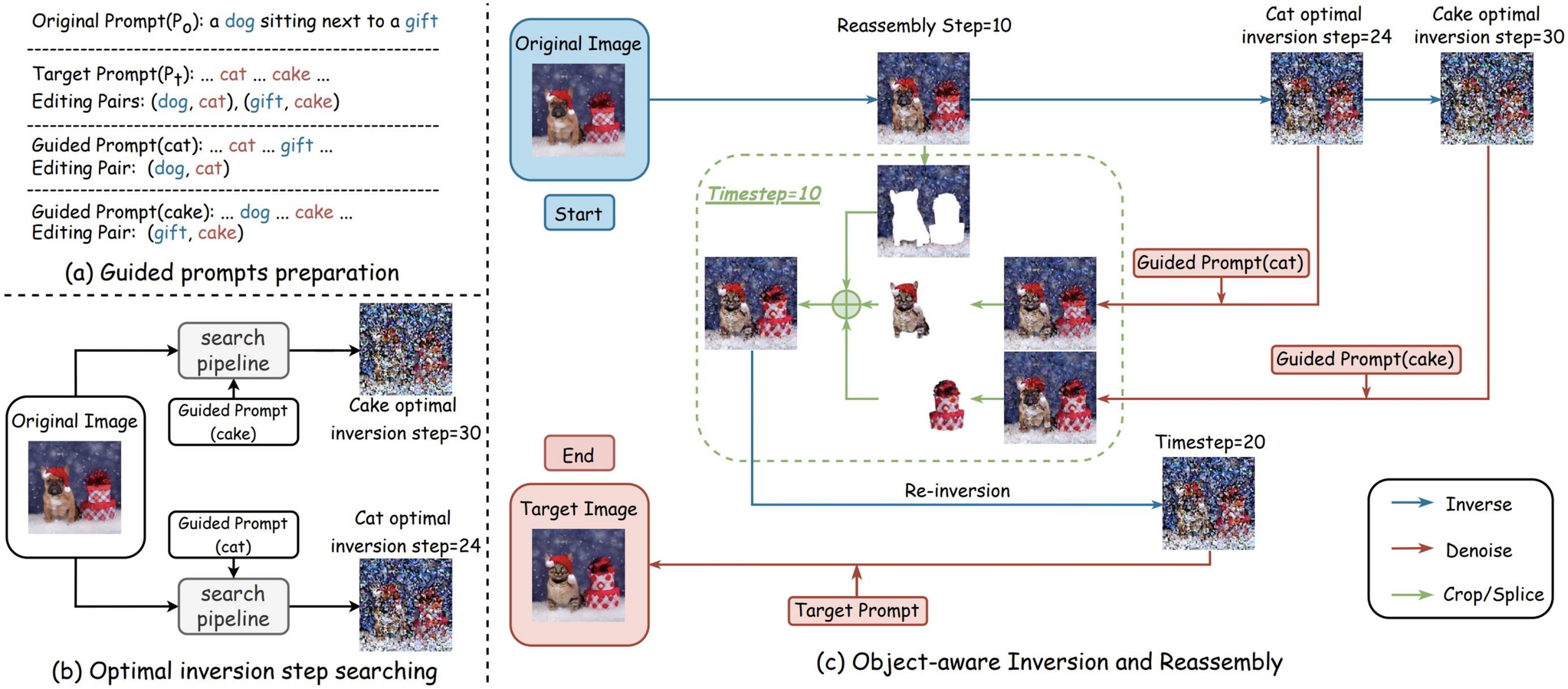
What is the "concept mismatch" and "poor editing" and where are they from?



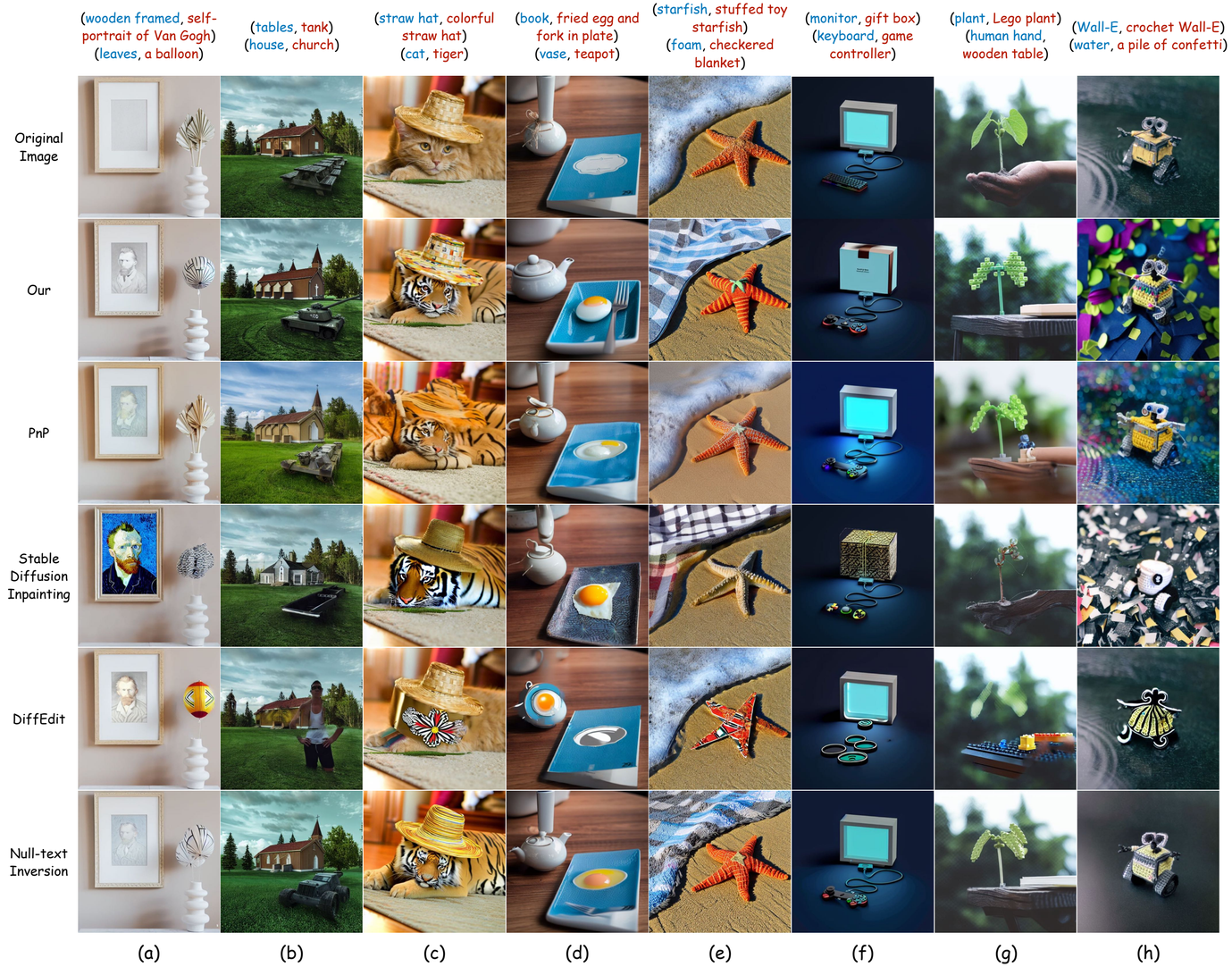
- **Concept mismatch:** mismatch the feature of crochet and gold.
  - Possible sources: There are overlapping features **in the early stage of** the denoising process.
- **Poor editing:** don't follow the objective of editing task (fidelity and editability).
  - Possible sources: Use the **same inversion step** for different editing pairs.

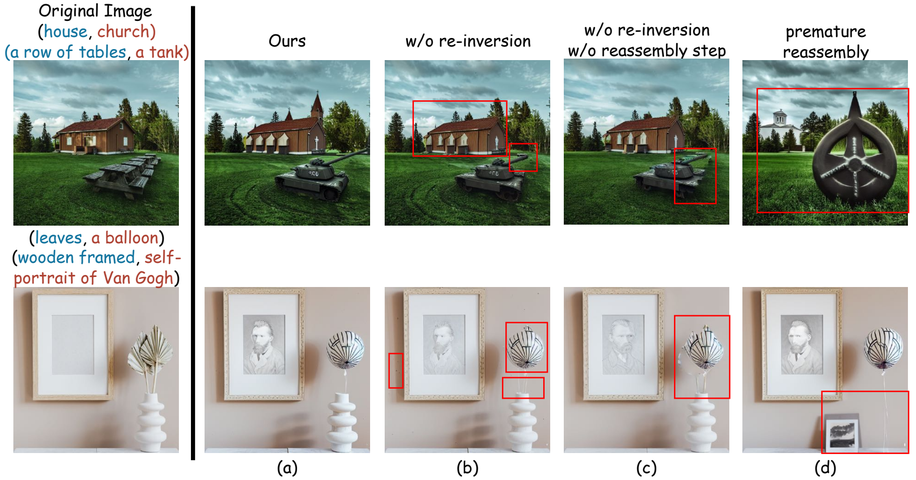
2 How to solve "concept mismatch" and "poor editing"!!!

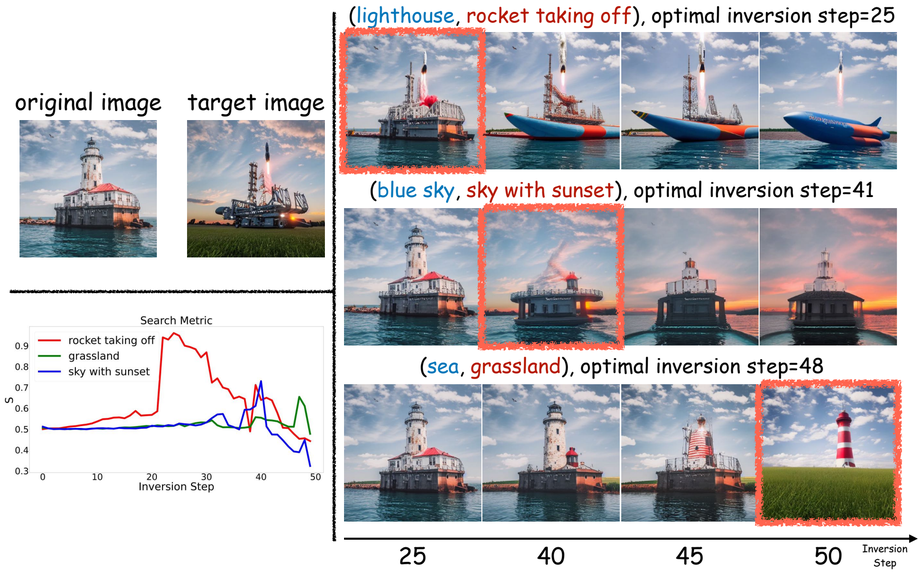
How to solve these two problems ? (OIR)



Results


(wooden framed, self-portrait of Van Gogh) (leaves, a balloon)
(tables, tank) (house, church)
(straw hat, colorful straw hat) (cat, tiger)
(book, fried egg and fork in plate) (vase, teapot)
(starfish, stuffed toy starfish) (foam, checkered blanket)
(monitor, gift box) (keyboard, game controller)
(plant, Lego plant) (human hand, wooden table)
(Wall-E, crochet Wall-E) (water, a pile of confetti)
Original Image
Our
PnP
Stable Diffusion Inpainting
DiffEdit
Null-text Inversion
(a)
(b)
(c)
(d)
(e)
(f)
(g)
(h)


Original Image
(house, church)
(a row of tables, a tank)
(leaves, a balloon)
(wooden framed, self-portrait of Van Gogh)
Ours
w/o re-inversion
w/o re-inversion
w/o reassembly step
premature reassembly
(a)
(b)
(c)
(d)


original image
target image
Search Metric
rocket taking off
grassland
sky with sunset
S
Inversion Step
(lighthouse, rocket taking off), optimal inversion step=25
(blue sky, sky with sunset), optimal inversion step=41
(sea, grassland), optimal inversion step=48
25
40
45
50
Inversion Step

I am currently seeking a Ph.D. supervisor.
If you are interested, please contact me. 🤗

WeChat

Twitter

zheny.cs@gmail.com