PixelMan@AAAI2025

liyaojiang1998.github.io

Keywords:
Generative AI, Diffusion Model, Image Editing, Non-Rigid Editing, Consistent Object Editing, High Consistency, High Efficiency, Inversion-Free, Training-Free

AAAI-25 / IAAI-25 / EAAI-25
FEBRUARY 25 – MARCH 4, 2025 | PHILADELPHIA, USA

HUAWEI

UNIVERSITY OF ALBERTA

PixelMan: Consistent Object Editing with Diffusion Models via Pixel Manipulation and Generation

Liyao Jiang[1,2], Negar Hassanpour[2], Mohammad Salameh[2], Mohammadreza Samadi[2], Jiao He[3], Fengyu Sun[3], Di Niu[1]

[1]*Dept. ECE, University of Alberta* [2]*Huawei Technologies Canada* [3]*Huawei Kirin Solution*

Background

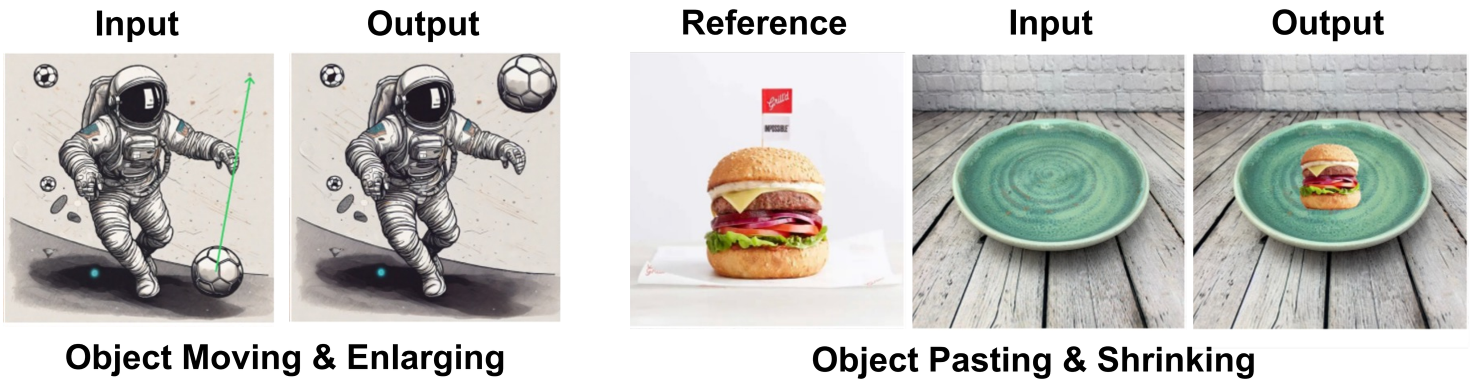
Promising results on using Diffusion Model for **text-guided rigid image editing** (i.e., editing color, texture, attributes, and style)

Our focus: Consistent object editing

- Before & after editing, preserve consistency for object and background
- Only edit non-rigid object attributes (e.g. position, size, composition)
- Typical tasks: object repositioning (moving), resizing, pasting

A challenging task involving multiple sub-tasks

1. Faithful reproduction of source object at the target location
2. Maintain background scene details
3. Harmonization of the new object into its surrounding context
4. Inpainting the vacated area with cohesive background



Challenges

1. **Low efficiency**
   - Rely on DDIM Inversion to reconstruct original image, which requires many (e.g., at least 50) steps, compromising quality when reducing # steps
2. **Low object and background consistency**
   - Altered object identity, inconsistent background
3. **Incomplete & incoherent inpainting**
   - Fail to inpaint vacated area with cohesive background

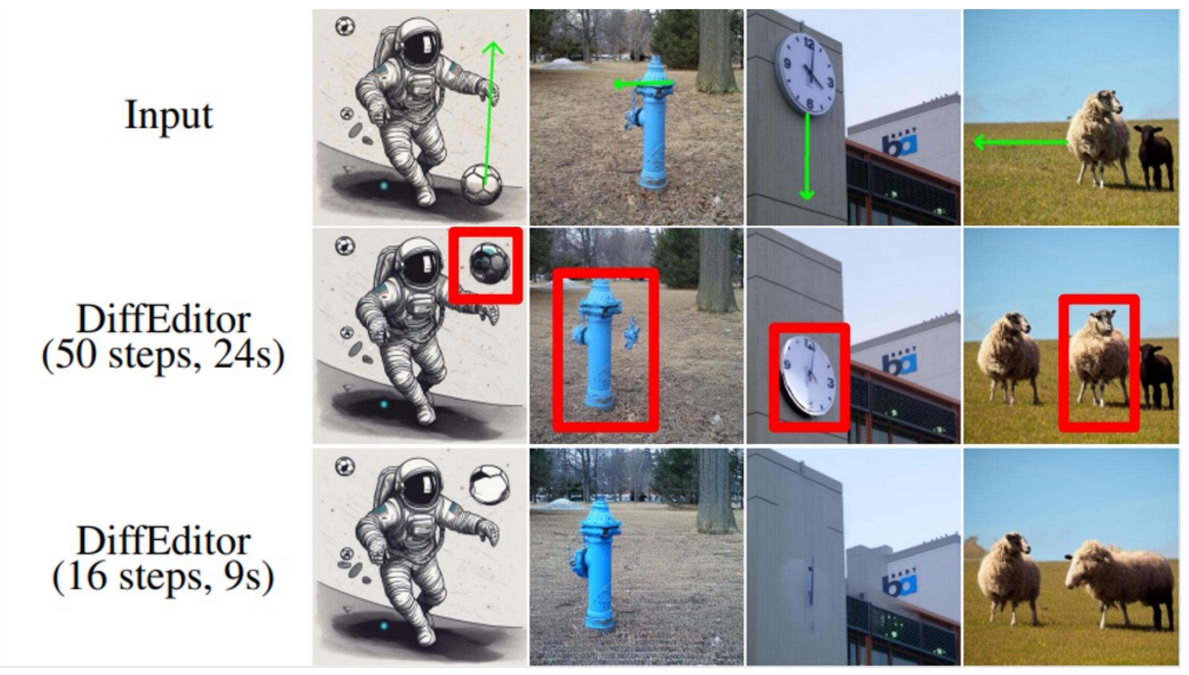


Figure: Issues faced by existing methods.

Results

PixelMan improves editing quality

- Object is consistent to the source (attributes and identity)
- Background is preserved after editing (texture and color)
- Original object is inpainted with cohesively background

While having better efficiency

- PixelMan@16 steps outperforms other methods@50steps
  - Reduce latency: 24s -> 9s; Reduce #NFEs: 176 -> 64
- Consistently outperform other methods at 8,16,50 steps (when using the same #Steps)

Aspects: IQA, Object Consistency, Background Consistency, Semantic Consistency

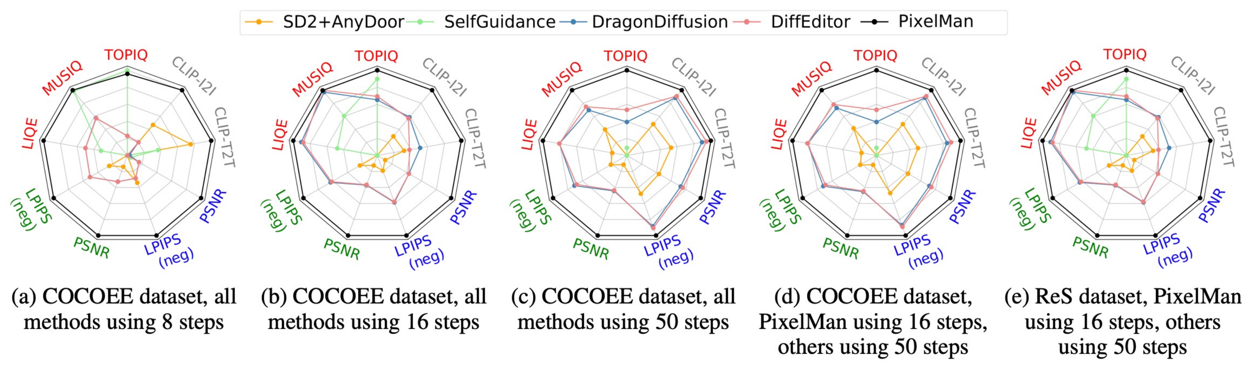


Figure: Visual comparison examples (on COCOEE dataset).

Methodology

1. Three-branched inversion-free sampling

Pixel manipulation helps to reproduce the object and background with high consistency, while being inversion-free which improves efficiency

a) Pixel-manipulated branch: copy the source object to target location in pixel space

b) Target branch: at each step, always anchor the target latents to the pixel-manipulated latents

c) Source branch: preserve uncontaminated K, V as context for generating harmonization effects (e.g., lighting, shadow, edge blending)

2. Editing guidance techniques

$$z_0^{out} = z_0^{man} + (\hat{z}_0^{tgt} - \hat{z}_0^{man}) \times (1 - m_{new})$$

Output = Anchor + Delta Edit Direction x Mask

Generation helps to find the delta editing direction to be added on top of the anchor (i.e., generate harmonization and inpainting edits)

a) Editing guidance based on energy functions with latents optimization (update z instead of ϵ, reduces #NFE)

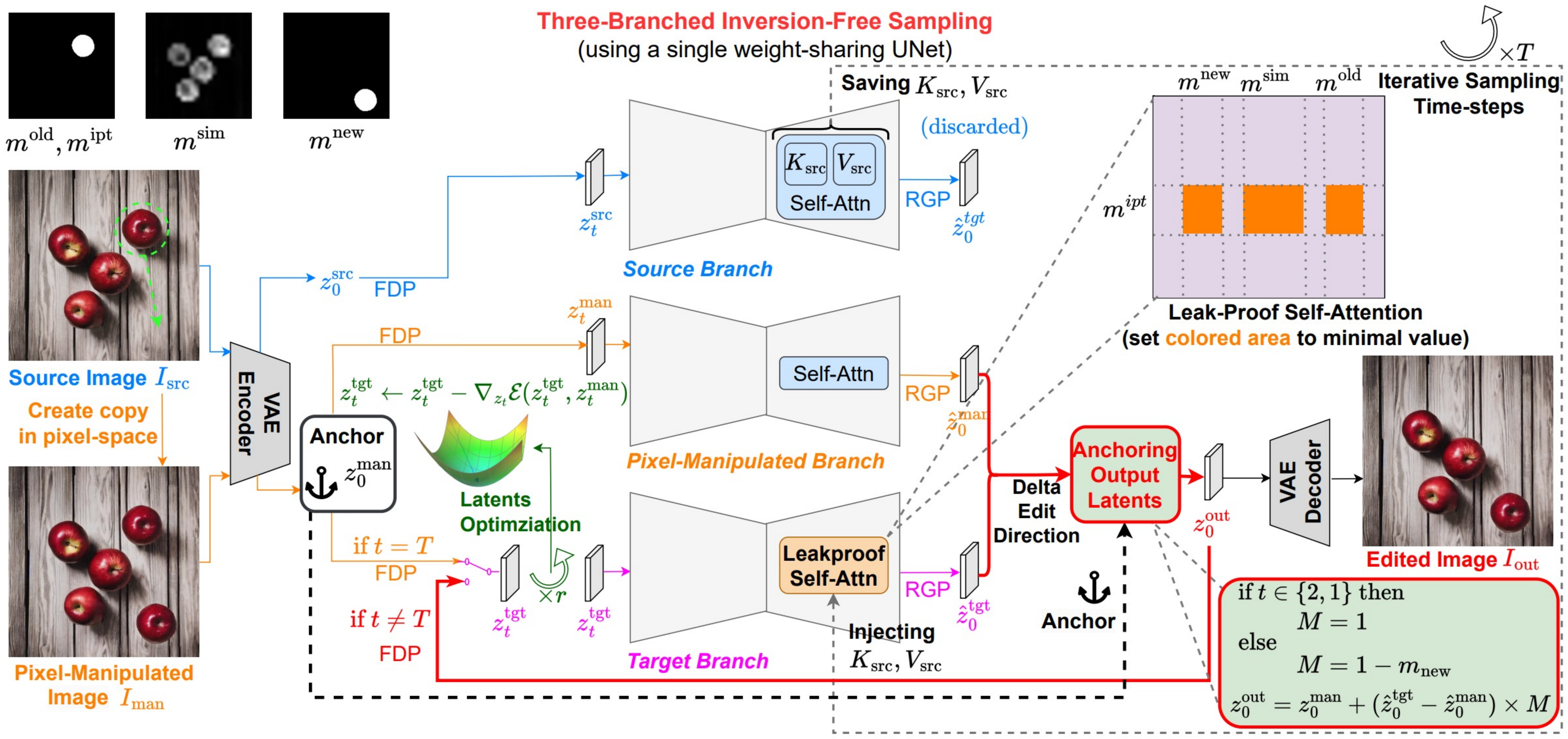
b) Injection of source K, V into the target branch

c) Apply leak-proof self-attention in target branch

3. Leak-proof self-attention

To achieve complete and cohesive inpainting

- Root cause of inpainting failure
  - Information leakage from similar objects through the self-attention
- Leak-proof self-attention: prevent attention to source, target, and similar objects
  - Set the corresponding QKˆT elements to minimal values



Conclusion

- PixelMan is an inversion-free and training-free method for high quality consistent object editing. It improves editing quality and enables faster editing, outperforming methods requiring 50 steps with only 16 steps
- It preserves consistency in the object and background
  - We utilize pixel manipulation, i.e., duplicate the source object to the target location in pixel space to serve as consistency anchor
  - We design a three-branched sampling approach to compute the delta edit direction, enabling seamless harmonization with lighting, shadows, and edges
- By introducing a leak-proof self-attention technique, our method prevents attention leakage, ensuring cohesive inpainting of the original object location
- Validated on COCOEE and ReS datasets with superior performance in object, background, and semantic consistency metrics. Achieves higher or comparable overall image quality while reducing latency