NEP: Autoregressive Image Editing via Next Editing Token Prediction

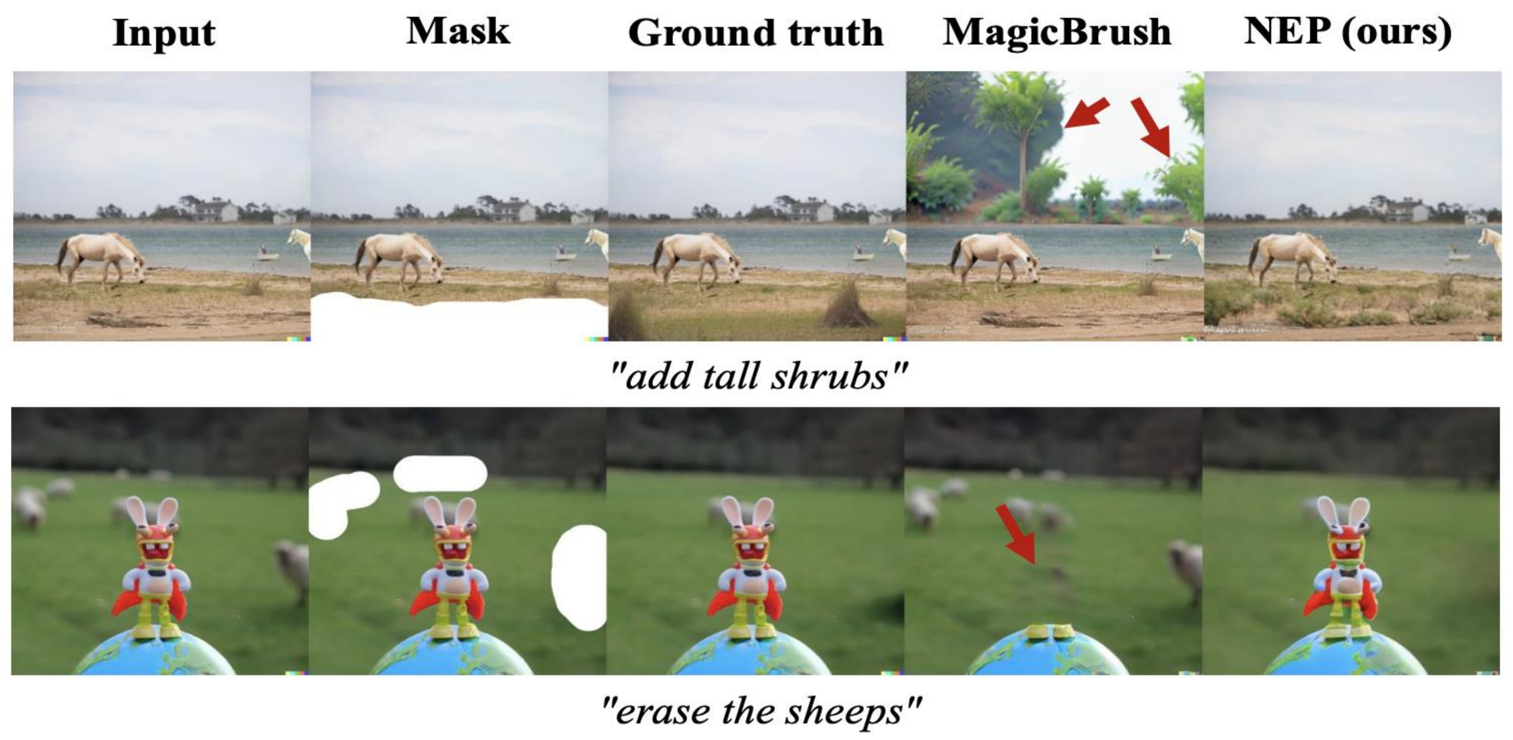
Huimin Wu, Xiaojian Ma, Haozhe Zhao, Yanpeng Zhao, Qing Li ✉

北京通用人工智能研究院
Beijing Institute for General Artificial Intelligence

NEURAL INFORMATION PROCESSING SYSTEMS

Introduction

TL;DR: We propose NEP, a targeted and efficient image editing method. It allows regeneration solely at edited regions, which avoids unintended modification and enhances efficiency. Besides, it makes self-improvement during the image generation process possible.

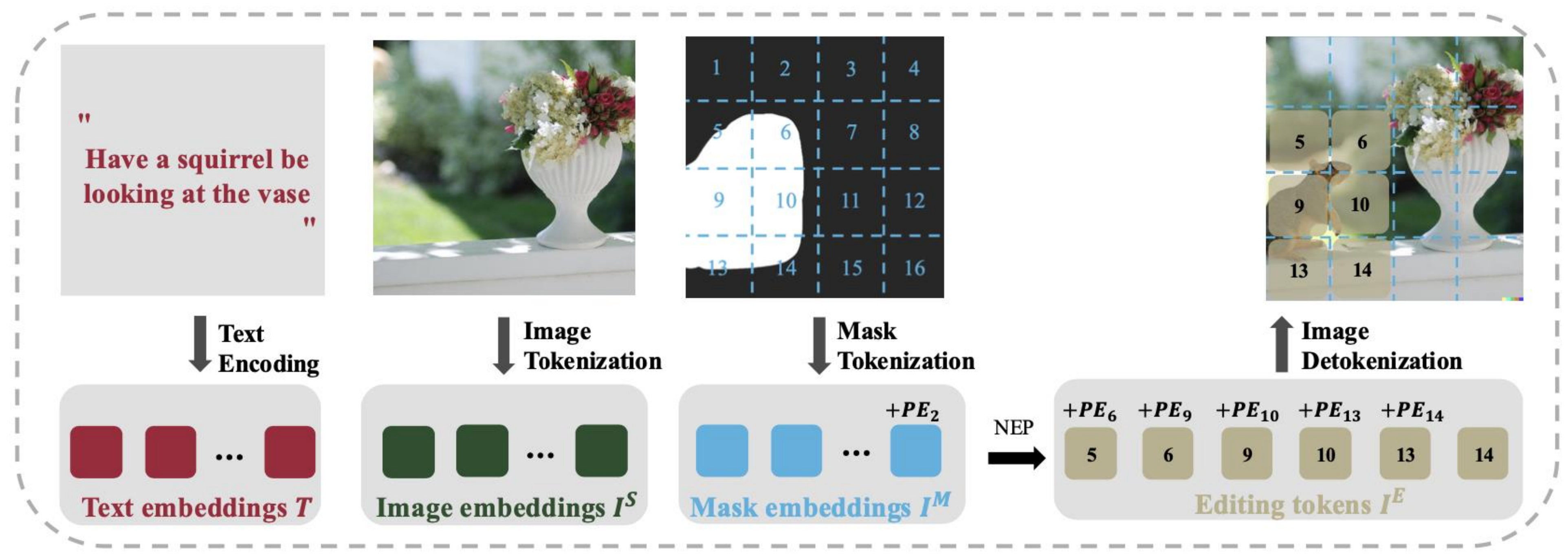


Visualized editing results

Next Editing-token Prediction

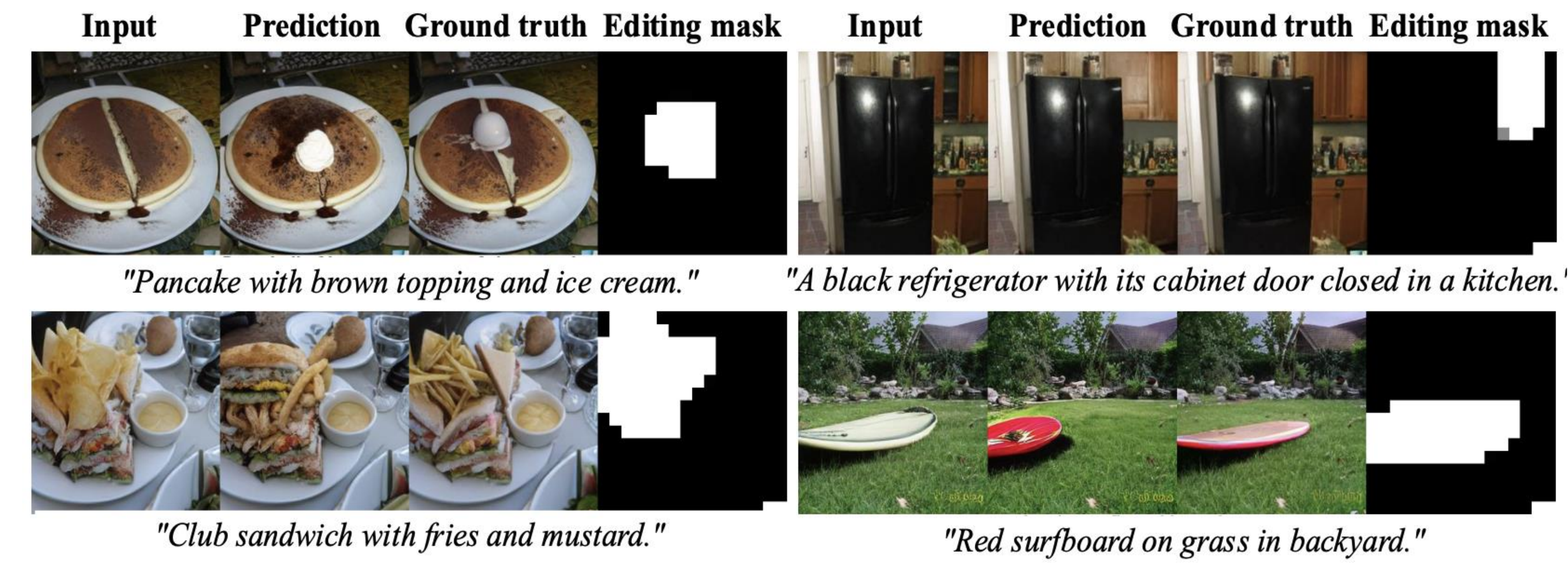
RLlamaGen: First stage

NEP: Second stage



Test-time Scaling with NEP

Zero-shot image editing



Text-to-image generation by integrating NEP into a self-improving loop

1. Revision region proposal
2. Image region revision
3. Reward model decision: to reject or accept

Text-to-image generation results

Methods	CLIP↑	FID↓
LlamaGen	0.320	15.07
LlamaGen ft.	0.326	12.00
RLlamaGen	0.325	11.49
TTS w/ NEP	**0.330**	**10.18**

# Revision rounds	CLIP↑	FID↓
0	0.325	11.49
1	**0.332**	9.94
2	**0.332**	9.93
3	**0.332**	9.85
4	**0.332**	**9.82**

Project page

Paper

Code

Quantitative editing results

Settings	Methods	L1↓	L2↓	CLIP-I↑	DINO↑
Single-turn	*Global Description-guided*				
	SD-SDEdit	0.1014	0.0278	0.8526	0.7726
	Null Text Inversion	0.0749	0.0197	0.8827	0.8206
	GLIDE	3.4973	115.8347	0.9487	0.9206
	Blended Diffusion	3.5631	119.2813	0.9291	0.8644
	Instruction-guided				
	HIVE	0.1092	0.0380	0.8519	0.7500
	InstructPix2Pix (IP2P)	0.1141	0.0371	0.8512	0.7437
	IP2P w/ MagicBrush	0.0625	0.0203	<u>0.9332</u>	<u>0.8987</u>
	UltraEdit	<u>0.0575</u>	<u>0.0172</u>	0.9307	0.8982
	FireEdit	0.0701	0.0238	0.9131	0.8619
	AnySD	0.1114	0.0439	0.8676	0.7680
	EditAR	0.1028	0.0285	0.8679	0.8042
	Ours	**0.0547**	**0.0163**	**0.9350**	**0.9044**

Method	CLIPdir↑	CLIPout↑	L1↓	CLIPimg↑	DINO↑
InstructPix2Pix	0.0784	0.2742	0.1213	0.8518	0.7656
MagicBrush	0.0658	0.2763	0.0652	0.9179	**0.8924**
Emu Edit	0.1066	0.2843	0.0895	0.8622	0.8358
UltraEdit	**0.1076**	0.2832	0.0713	0.8446	0.7937
MIGE	0.1070	0.3067	0.0865	0.8714	0.8432
AnyEdit	0.0626	0.2943	**0.0673**	**0.9202**	**0.8919**
Ours	0.1064	**0.3078**	0.0781	0.8710	0.8440

For the first time, autoregressive models can achieve top performance on well-recognized editing benchmarks.
without resorting to editing masks, our approach still achieves comparable or better editing performance.