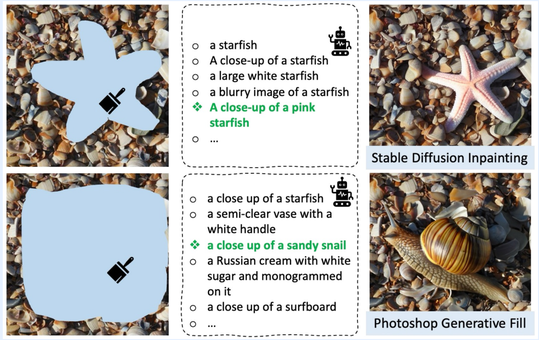
Brush2Prompt: Contextual Prompt Generator for Object Inpainting

Mang Tik Chiu[1,2], Yuqian Zhou[2], Lingzhi Zhang[2], Zhe Lin[2], Connelly Barnes[2], Sohrab Amirghodsi[2], Eli Shechtman[2], Humphrey Shi[1,3]

[1]UIUC, [2]Adobe, [3]Georgia Tech

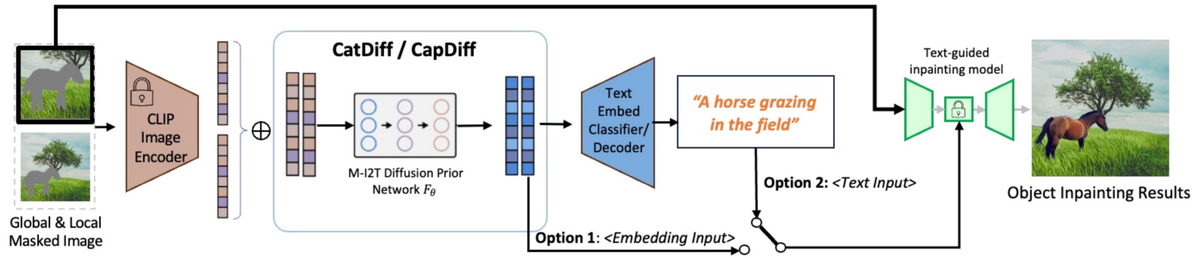
CVPR SEATTLE, WA JUNE 17-21, 2024

Introduction



- **Masked-Image-to-Text** generation
- **Prompt recommendation** for inpainting
- Leverage **context** and **mask shape** for diversity

Method

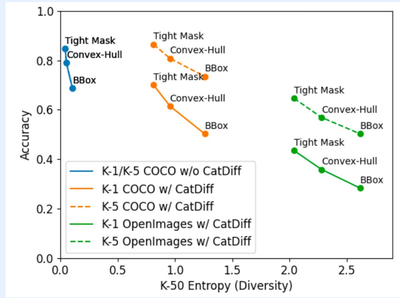


1. Encode global and local masked images using CLIP.
2. Use a diffusion transformer to regress text embeddings of object Classes/Captions.
3. Decode text embeddings into Class/Caption by separately finetuning a text decoder.
4. Perform text-to-image inference with any latent diffusion models using predicted text/embeddings.
5. (Optional) Perform prompt completion with masked image & prefix text from user

Results

CatDiff (Category diffusion):

- Our model achieves good tradeoff between classification accuracy and diversity. We evaluate our model with classification accuracy and entropy.



CapDiff (Caption diffusion):

- Our model generates context-aware prompts that are also diverse, when compared to other LLMs.

Mask shape	Method	BLEU ↑	BERTScore ↑	Self-BLEU ↓	Div-4 ↑
Tight mask	InstructBLIP [4]	0.071	0.704	0.466	0.570
	LLaVA-5-Prompt [18]	0.023	0.656	0.163	0.785
	CapDiff (Ours)	0.177	0.732	0.169	0.858
BBox	InstructBLIP [4]	0.060	0.691	0.438	0.597
	LLaVA-5-Prompt [18]	0.020	0.648	0.128	0.881
	CapDiff (Ours)	0.149	0.715	0.141	0.885

Prompting other LLMs is quite challenging:

InstructBlip
Prompt: Write a text prompt that describes a reasonable object to be inserted in the gray area.

The man is holding an accordion
A man is holding an accordion while smiling
A man is holding an accordion while smiling
The man is playing an accordion
The man is playing an accordion

LLAVA-5-Prompt
Prompt: Write five text prompts that describe reasonable objects to be inserted in the gray area

A blue and white umbrella
A pair of black sunglasses
A small, colorful kite
A red and white sports ball
A small, decorative wind chime

Ours
(no prompt needed)

A pair of jeans that are sitting on a table
A pair of black shorts with a white belt
A pair of blue jeans with a brown background
A pair of jeans with a belt
A pair of red jeans with holes in them

InstructBlip
Prompt: Write a text prompt that describes a reasonable object to be inserted in the gray area.

A yellow fire hydrant sits in front of a brick wall
A yellow fire hydrant sitting in front of a brick wall
A yellow fire hydrant next to a brick wall
A yellow fire hydrant sitting on the sidewalk in front of a brick building
There is a fire hydrant on the sidewalk in front of a

LLAVA-5-Prompt
Prompt: Write five text prompts that describe reasonable objects to be inserted in the gray area

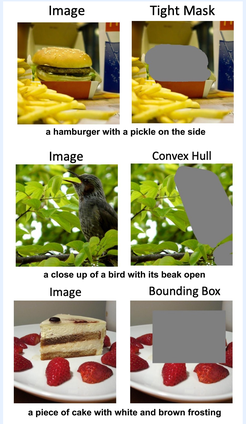
A metal bench
A large sculpture
A colorful mural
A set of stairs
A group of potted plants

Ours
(no prompt needed)

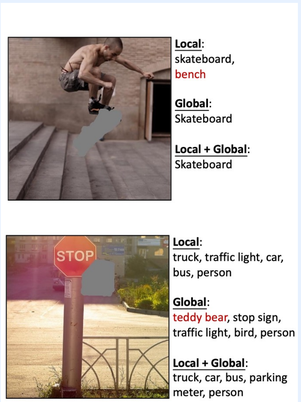
An orange dog with a tag on it
A bicycle wheel is wrapped in a cage
A man in a baseball uniform holding a bat
An orange car parked in a parking lot
A guitar with a wooden body

Visualizations

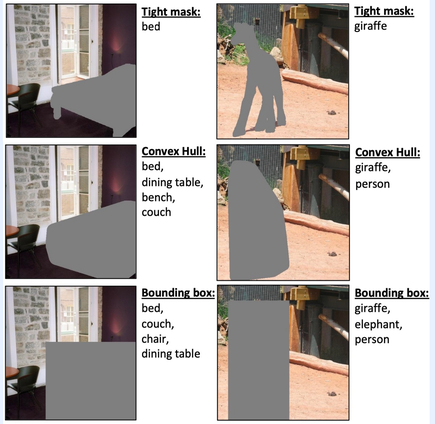
Masked-image-Caption dataset:



Global / Local context encoding:



Mask shape vs. Prompt diversity:



Text Prompt & Inpainting results using embedding/text: