



Input



"It is now midnight"

InstructPix2Pix: Learning to Follow Image Editing Instructions

Рецензия

Постановка/рекап

- Редактирование изображения на основе краткой инструкции (описания изменения) на естественном языке

До этого было:

- Style transfer, inpainting
- Prompt2Prompt
- SDEdit
- Text2LIVE

Всё без инструкций

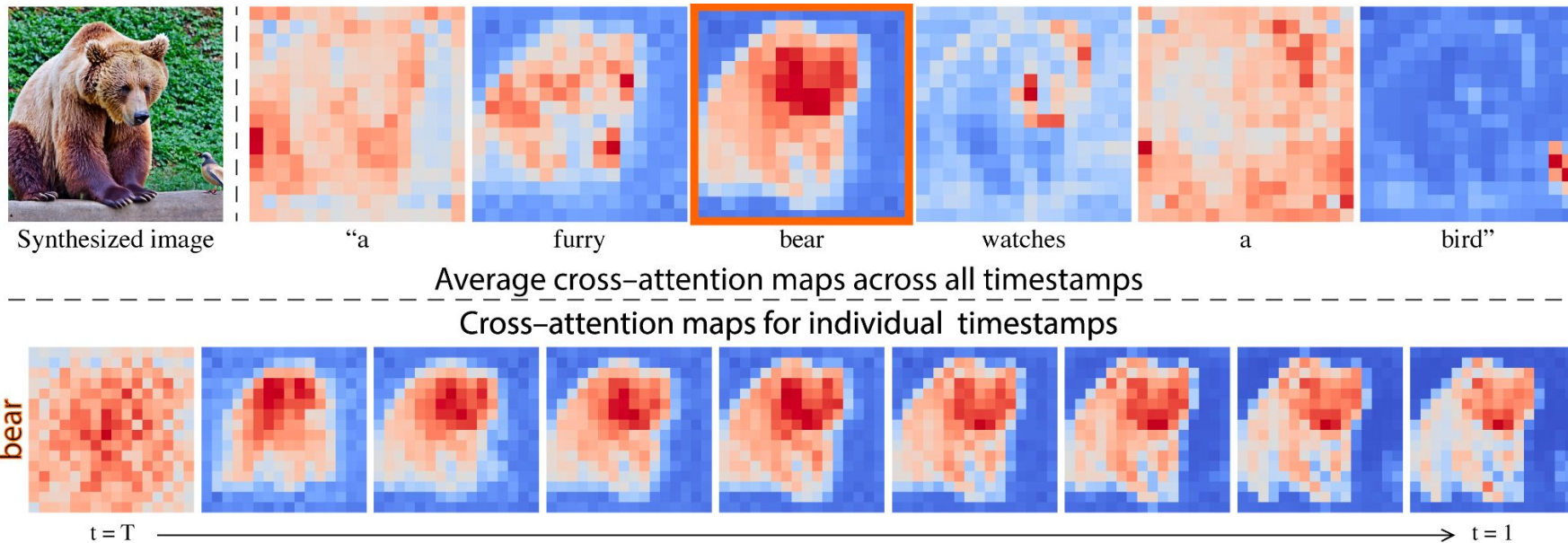
Публикация

- Статья не опубликована на конференциях или в журналах на данный момент
- Первый сабмит в ноябре 2022 года

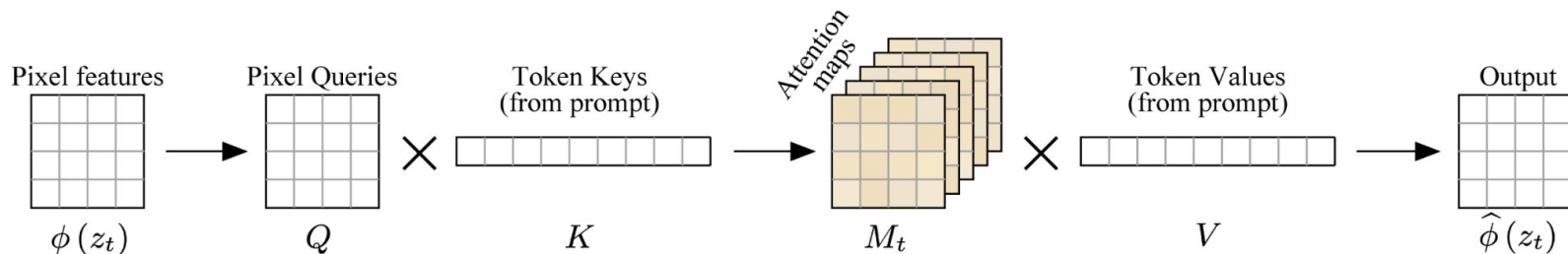
Авторы

- Tim Brooks* - PhD at Berkeley AI Research, now OpenAI
 - Обновление весов нейросети диффузией для обучения/оптимизации лосса
 - ГАНЫ для генерации длинных консистентных видео
 - Генерация подходящего окружения по позе человека
- Aleksander Holynski* - PhD, now postdoctoral scholar at Berkeley AI Research
 - По одному селфи-ролику с вращением и соответственно разным освещением (от солнца) получение геометрии и последующий рендер в других позах и с кастомным освещением
 - Анимирование картинок с помощью Eulerian Motion Fields
- Alexei A. Efros - professor at UC Berkeley
 - Image-to-Image translation: Pix2Pix, CycleGAN

Prompt-to-prompt

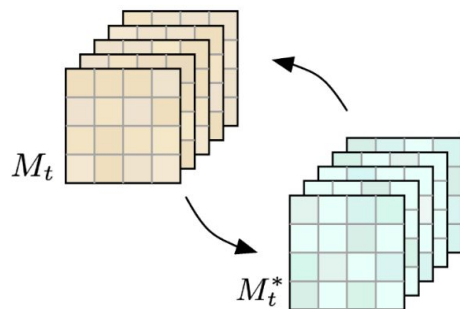


Prompt-to-prompt

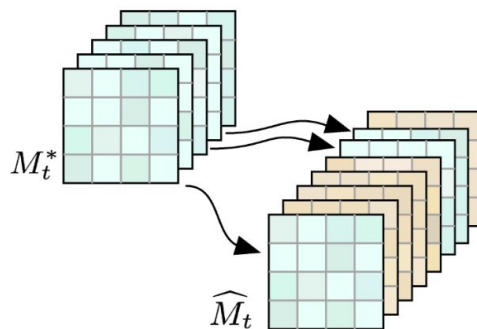


Text to Image Cross Attention

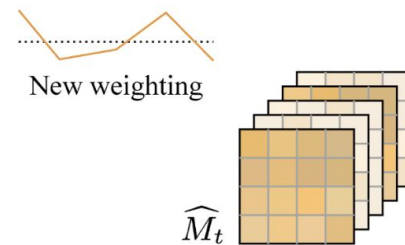
Cross Attention Control



Word Swap



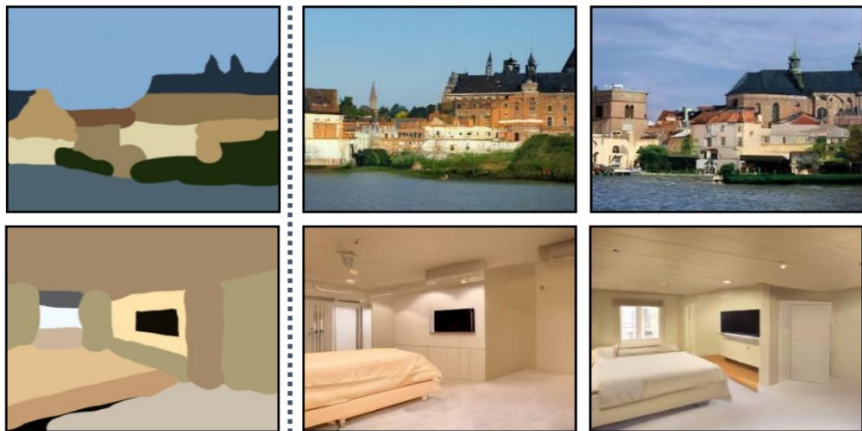
Prompt Refinement



Attention Re-weighting

Stochastic Differential Editing

Stroke Painting to Image



Input (guide)

Output

Image Compositing



Source

Input (guide)

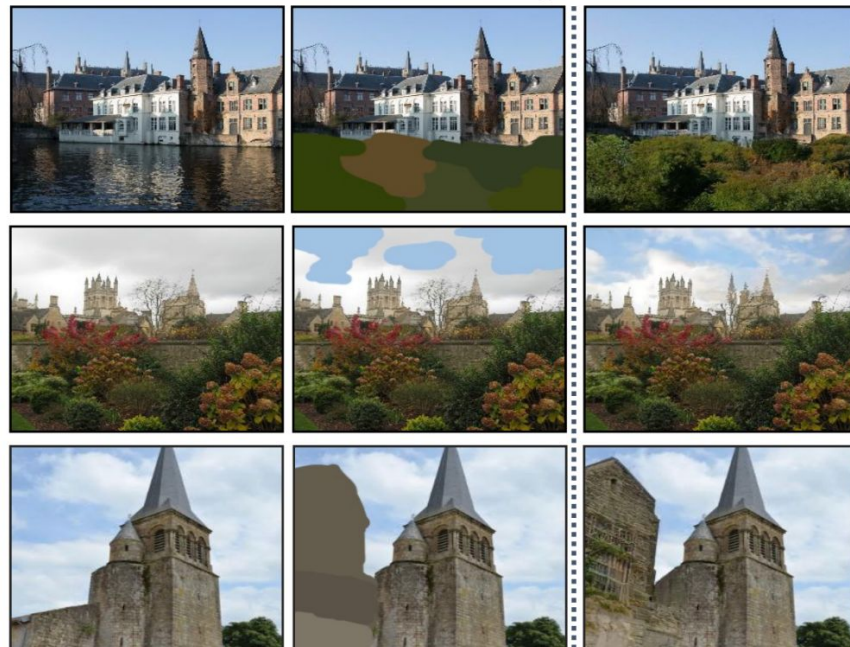
Output

Source

Input (guide)

Output

Stroke-based Editing

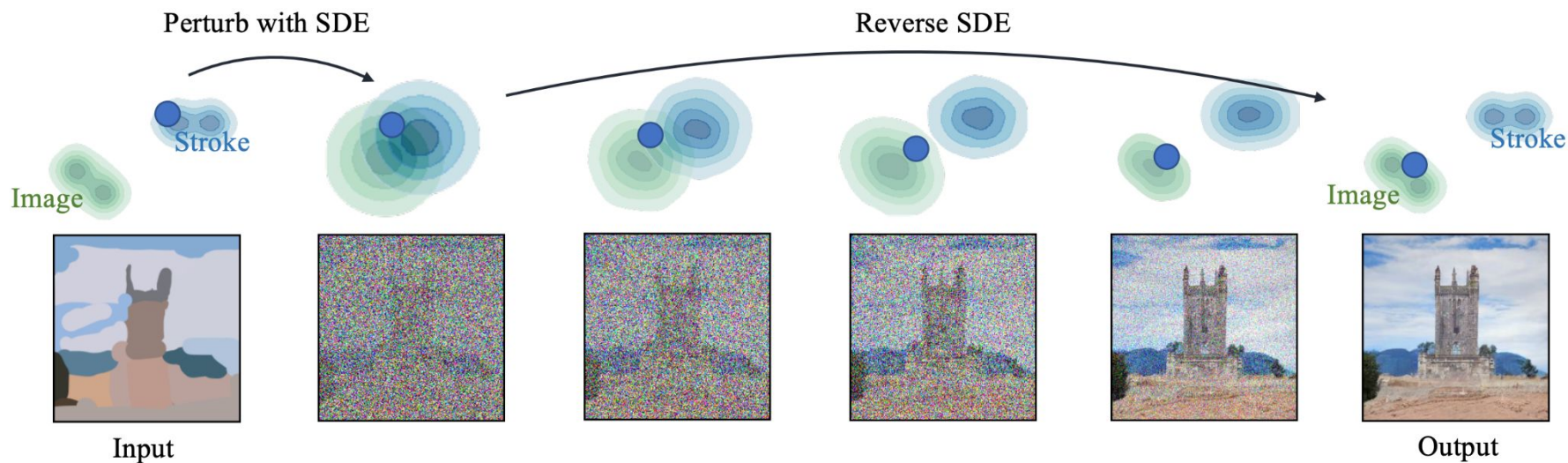


Source

Input (guide)

Output

Stochastic Differential Editing



Text2LIVE

Original Input



Edit Layer

"fire"



Composition

"fire out of bear's mouth"



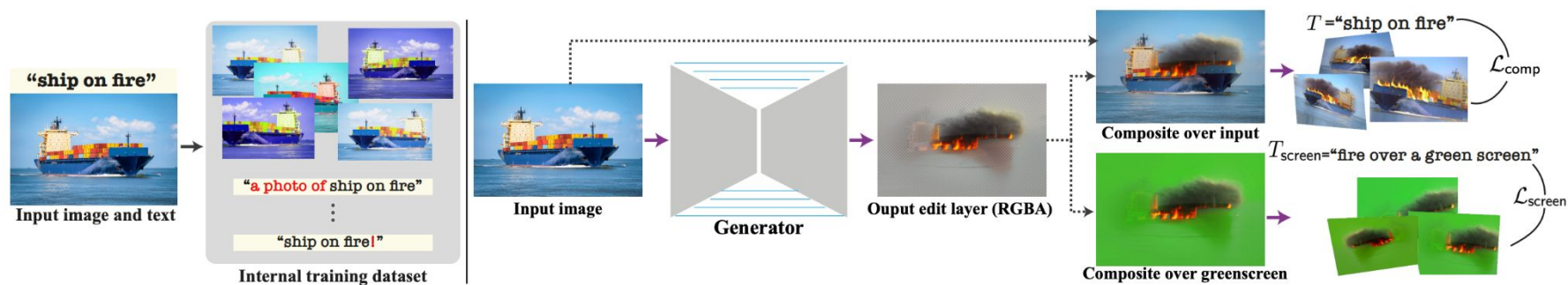
"latte art heart pattern"



"latte art heart pattern"



Text2LIVE



Плюсы и минусы

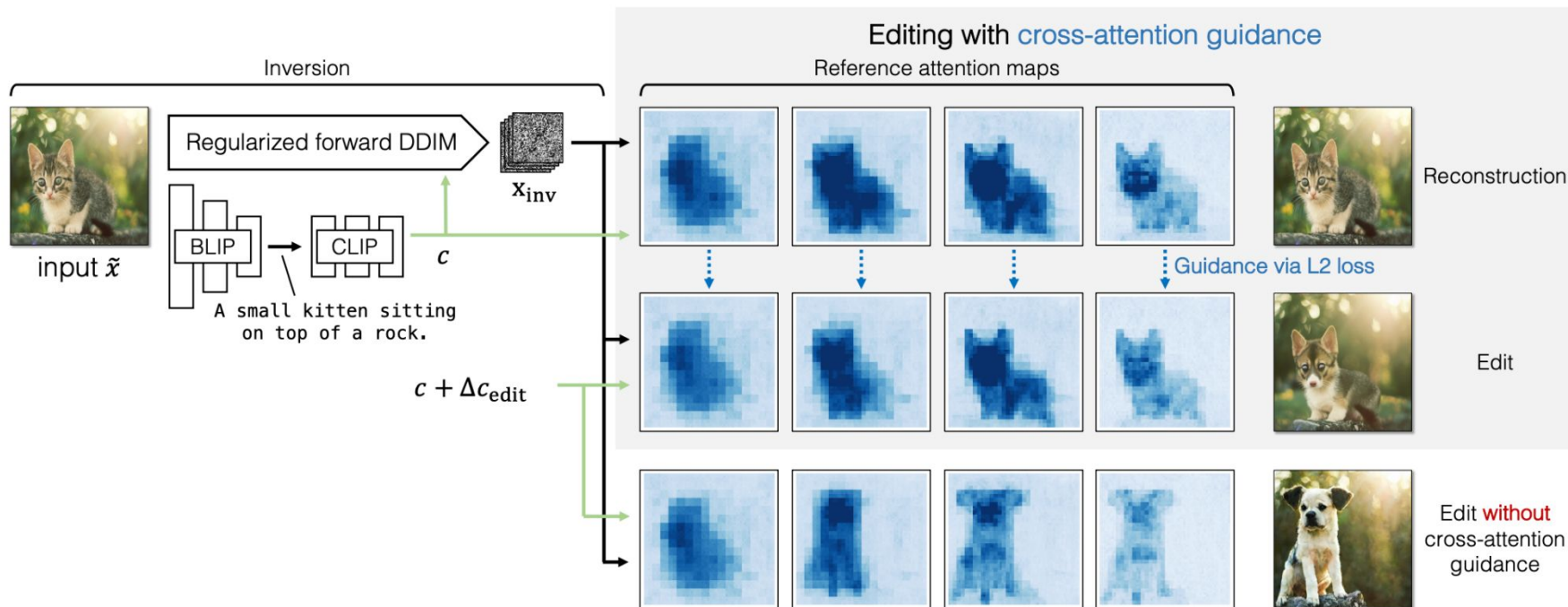
- + новизна - инструкции (как в RLHF, InstructGPT), только теперь для картинок
- + идея пайплайна построения datasets
- + open source
- + детали реализации в приложении, хорошо написанные подробные объяснения пайплайнов
- непонятно насколько diverse были промпты в трейне и тесте, инструкции очень простые и короткие - нет анализа работы на чём-то вне их ручного+gpt домена
- есть сравнение промтов Text2Live и SDEdit, но нет промтов Prompt-to-Prompt

Open source

- [Код](#)
- [Данные](#)
- [Веса модели](#)
- [HF Space / демо](#)



pix2pix-zero



Возможные продолжения

- Редактирование инструкциями видео и специфичных доменов (например, спектрограмм)
- Более сложные изменения геометрии и композиции изображений - в представленных работах всё статично, интересно было научиться качественно менять позы/эмоции/положение в пространстве