

Разработка обучаемых подходов автоматизированной разметки изображений для сегментации объектов

Измельцева У.А.
izmesteva.ua@phystech.edu

1 курс магистратуры «Методы и технологии искусственного интеллекта» Центр
когнитивного моделирования МФТИ

2024 – Научно-исследовательский семинар

Мотивация

Основания разработки:

- **Экономия времени.** Ручная разметка требует значительных временных затрат.
- **Снижение стоимости.** Ручная разметка дорогостоящая и подвержена человеческим ошибкам.
- **Расширение данных.** Автоматическая разметка позволяет увеличить объем данных для обучения.

Цель: разработать инструмент автоматизированной разметки на основе нейрорестевой модели, способной автоматически генерировать маски (возможны другие модальности), отражающие желаемые объекты на изображении, и синтетические данные в заданном домене.

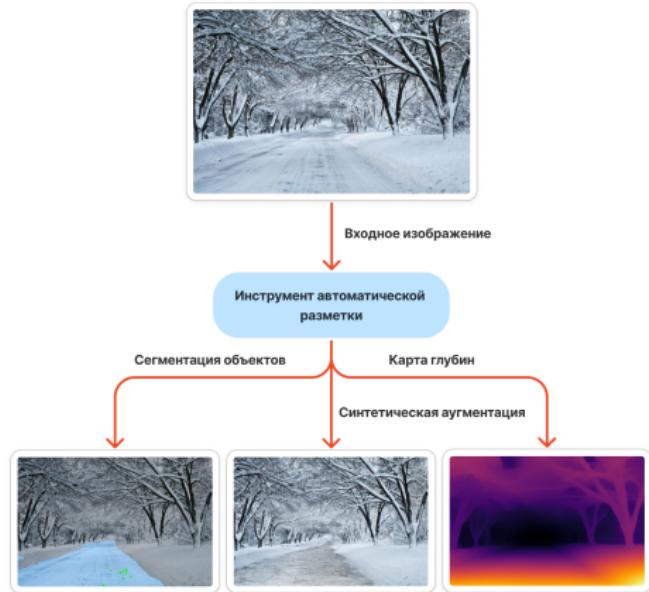
Постановка задачи

Входные данные:

- набор изображений;
- запрос в форме точки, области на изображении или перечня категорий для осуществления разметки;
- карты глубин (опционально);
- текстовые описания объектов (опционально).

Выходные данные:

- список масок, представляющих объекты на входных изображениях;
- карты глубин (опционально);
- новые сгенерированные размеченные данные (опционально).



Постановка задачи

Задача состоит в разработке модели F , которая по входным данным X генерирует выходные данные Y . $F: X \rightarrow Y$, где:

- X содержит изображения (I), запросы в виде точек (P), пространственных областей (A) или перечня категорий (C), карты глубины (D) и текстовые описания объектов (T);
- Y представляет список масок объектов на изображении (M) и карты глубины (D').

Математическая формулировка

Модель F может быть обучена на обучающем наборе данных $(X_1, Y_1), (X_2, Y_2), \dots, (X_n, Y_n)$, где n - количество доступных пар. Оптимальные параметры модели F могут быть получены:

$$\min_F \frac{1}{n} \sum_{i=1}^n L(F(X_i), Y_i) \quad (1)$$

где:

- $F(X_i)$ - предсказанная моделью маска для входных данных X_i ;
- Y_i - реальная маска, соответствующая входным данным X_i ;
- L - комбинированная функция потерь для генерации и сегментации.

Постановка экспериментов



Схема направлений экспериментов

В качестве первых экспериментов были выбраны направления работы с моделями сегментации, генеративными моделями и реализация простого интерфейса для их использования.

Используемые методы

Для сегментации объектов на изображении:

- Segment Anything¹

Для генерации изображений в режиме inpainting:

- Kandinsky 2.2²
- Stable Diffusion 2³

Для генерации карты глубин по изображению:

- Depth Anything⁴

¹Kirillov A. et al. Segment anything. – 2023.

²Razzhigaev A. et al. Kandinsky: an improved text-to-image synthesis with image prior and latent diffusion. – 2023.

³Rombach R. et al. High-resolution image synthesis with latent diffusion models. – 2022.

⁴Yang L. et al. Depth anything: Unleashing the power of large-scale unlabeled data. – 2024.

Используемые методы. Segment Anything

Сегментация объектов на изображении.

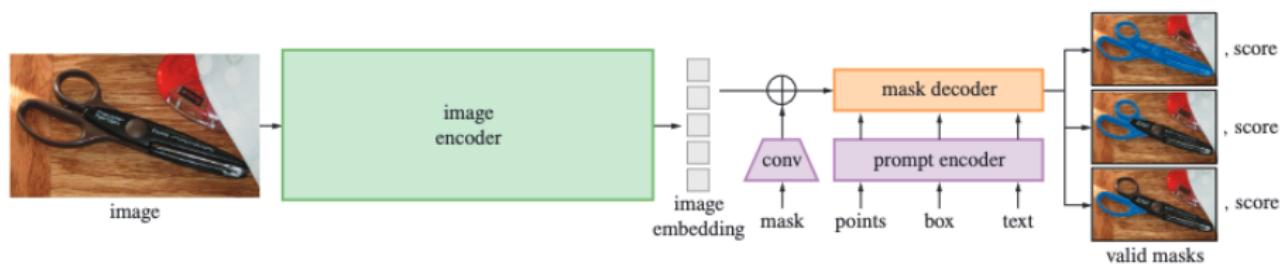


Схема модели SAM

Используемые методы. Kandinsky 2.2

Генерация изображений в режиме inpainting - процесс восстановления или восполнения участков изображения по выделенной маске.

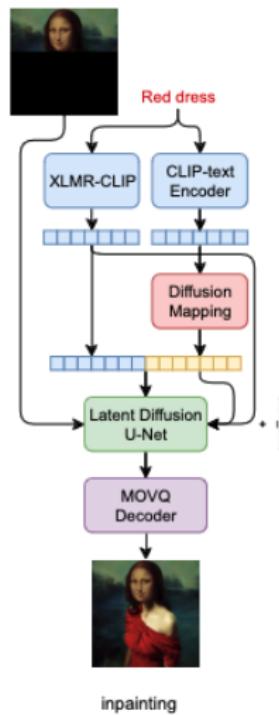


Схема модели Kandinsky 2.2

Используемые методы. Stable Diffusion 2

Генерация изображений в режиме inpainting - процесс восстановления или восполнения участков изображения по выделенной маске.

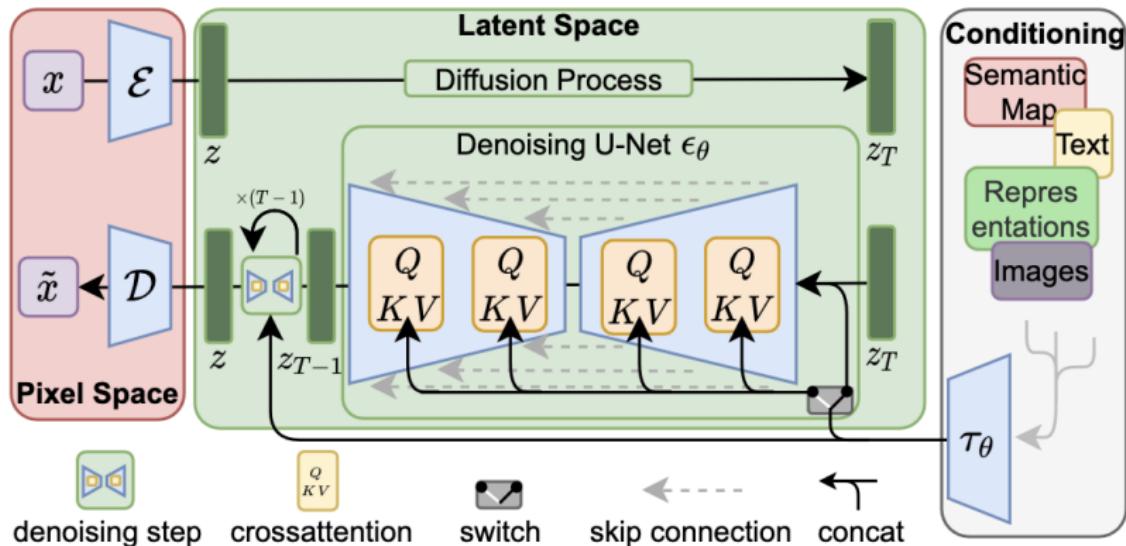


Схема модели Stable Diffusion 2

Используемые методы. Depth Anything

Генерация карты глубин по изображению.

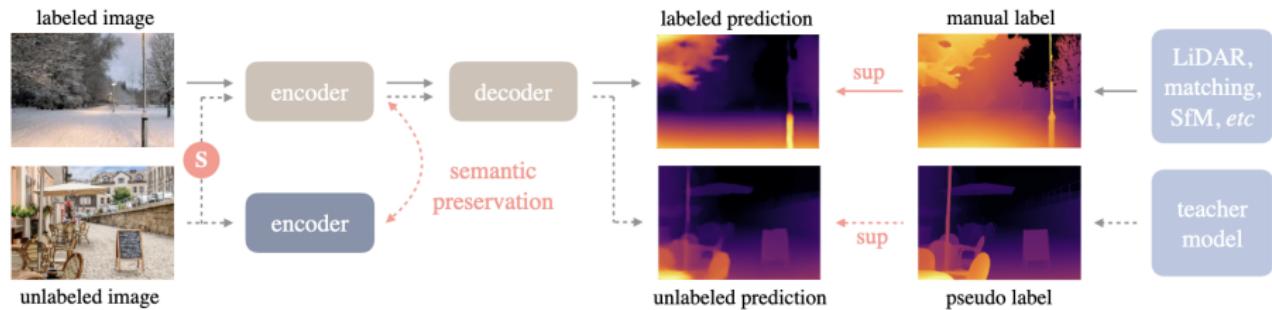


Схема модели Depth Anything

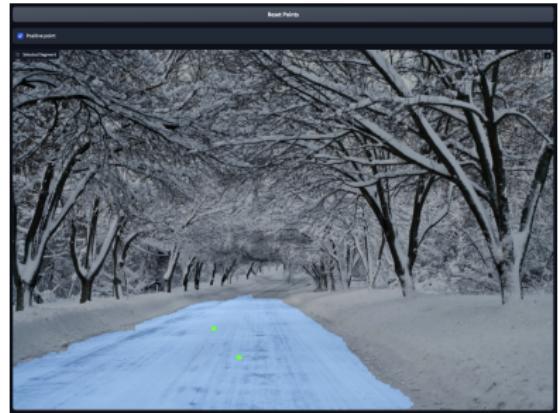
Стек технологий

- Gradio¹ - библиотека для создания веб-приложений
- Diffusers² - библиотека для работы с диффузионными моделями

¹<https://www.gradio.app/>

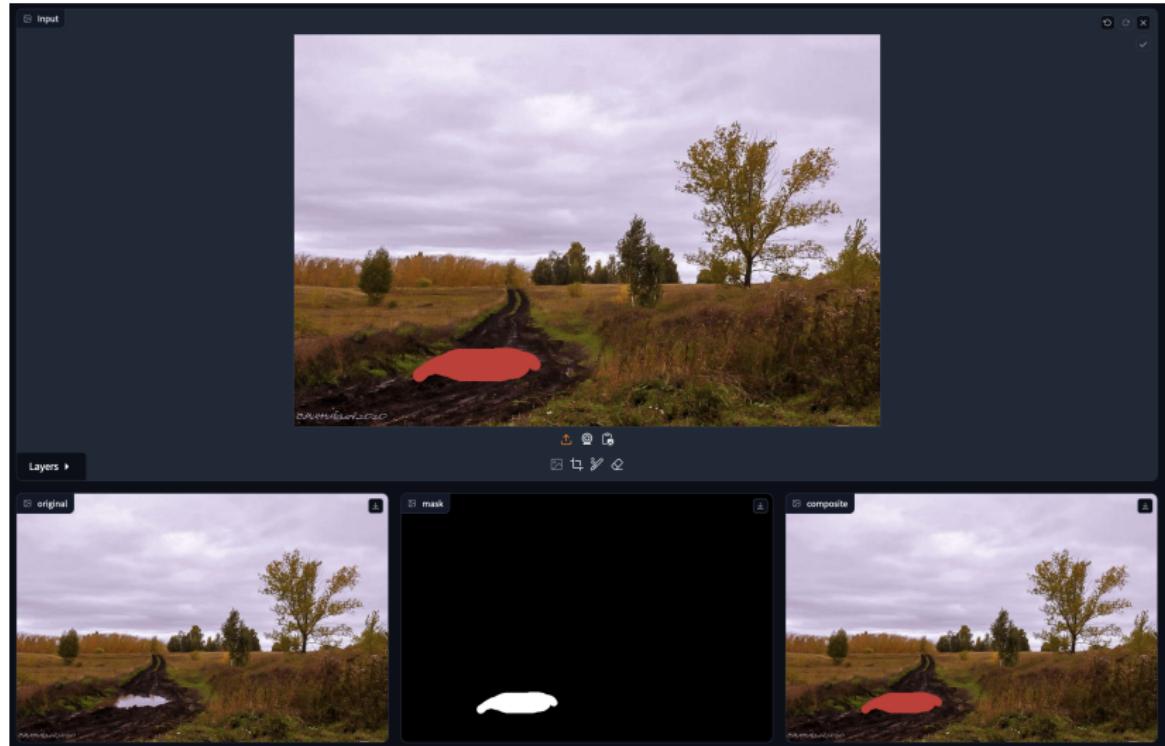
²<https://huggingface.co/docs/diffusers/>

Эксперименты



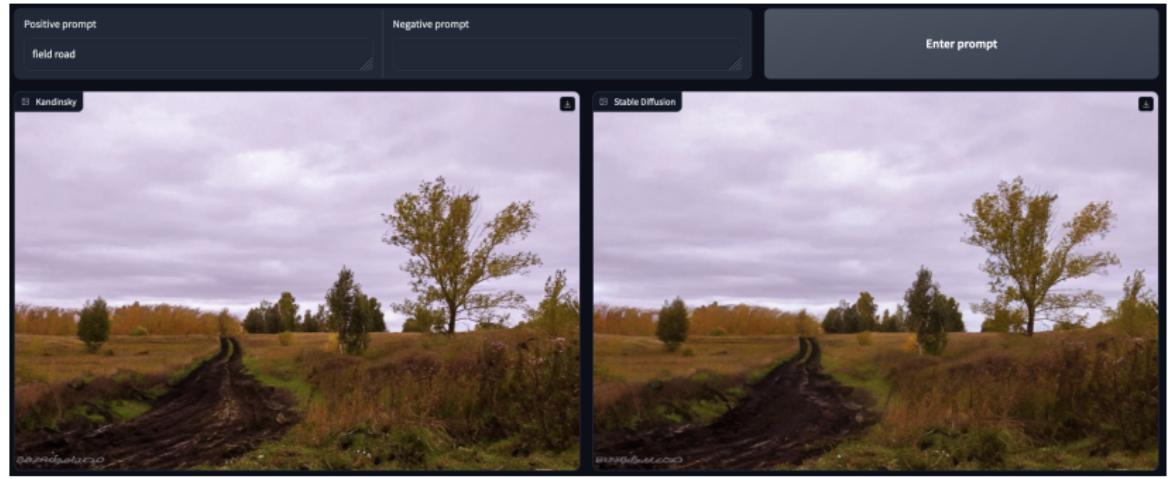
В интерфейсе на основе библиотеки gradio реализован инструмент выделения маски объекта средствами модели SAM

Эксперименты



Реализован инструмент мануального выделения маски

Эксперименты



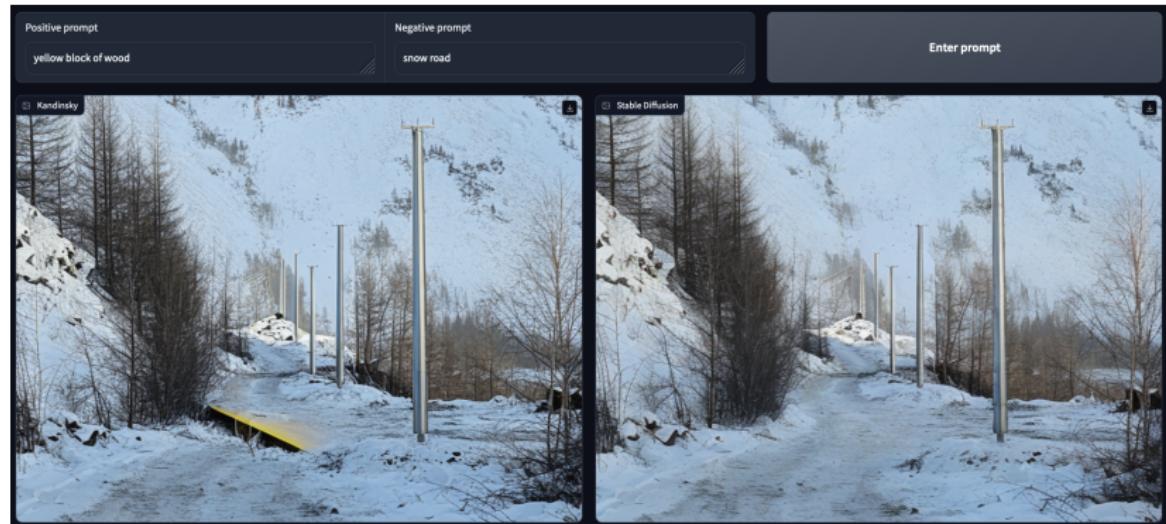
Реализован инструмент для генерации объектов на изображении с помощью моделей Kandinsky и Stable Diffusion. Запрос - "field road"

Эксперименты



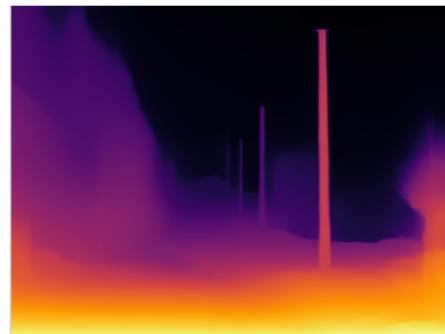
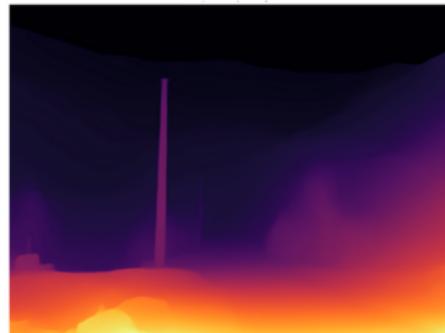
Интерфейс инструмента рисования маски

Эксперименты



Интерфейс генерации объектов на изображении. Запрос - "yellow block of wood". Негативный запрос - "snow road"

Эксперименты



Результаты работы Depth Anything

Выводы

- Предобученные веса SAM справляются с сегментированием объектов общего домена, возникают сложности с более сложными формами объектов
- Предобученные веса выбранных генеративных моделей хорошо работают на задаче удаления объекта с изображения
- Предобученные веса Stable Diffusion не восполняет отсутствующие объекты на изображении
- В контрасте с этим модель Kandinsky более успешно справляется с данной задачей
- Модель Depth Anything работает удовлетворительно, однако отсутствует информация об абсолютной величине глубины

Заключение

Были осуществлены следующие этапы:

- Реализация базового интерфейса для использования моделей
- Запуск выбранных генеративных моделей и модели сегментации в качестве baseline
- Тестирование модели преобразования изображения в карту глубин

Дальнейшие исследования:

- Подготовка датасета для обучения моделей в определенном домене
- Дообучение генеративных моделей на выбранном домене
- Тестирование более точных моделей сегментации
- Тестирование моделей преобразования изображения в карту глубин в абсолютных единицах

Спасибо за внимание!

izmesteva.ua@phystech.edu