Текстовое описание окружения автомобиля

Смоляр Родион

Бирюков Григорий

Ментор: Ковалева Маргарита

Попов Артемий

Зайцев Илья

Студкемп по компьютерному зрению и автономному транспорту

Яндекс Образование

01

Представление команды

Наша команда

Смоляр Родион

Ведущий разработчик

Попов Артемий

ML-инженер

Бирюков Григорий

ML-инженер

Зайцев Илья

ML-инженер

Ковалева Маргарита

Ментор

02

Цель и задачи проекта

Цель

Разработать модель, которая автоматически генерирует краткое текстовое описание окружающей среды автомобиля на фотографии, чтобы упростить поиск транспортных средств при сбоях GPS и повысить эффективность взаимодействия клиентов и сотрудников поддержки.

Задачи

01

Обучить несколько моделей, которые выделяют текстовые признаки из фотографии

02

Развернуть мультимодальную модель, принимающую на вход полученные признаки и фотографию, и генерирующую описание

Подзадачи

Первичная обработка изображения

1. Загрузить изображение
2. Правильно повернуть фото
3. Определить, есть ли автомобиль на фото
4. Удалить небо и автомобиль (необходимо для некоторых моделей)

Подзадачи

Выделение текстовых признаков

Обучить несколько моделей, которые:

1. Классифицируют объекты
2. Классифицируют тип окружения (парковка, заправка и т. д.)
3. Кратко (1-2 слова) описывают окружение

Подзадачи

Итоговая генерация описания

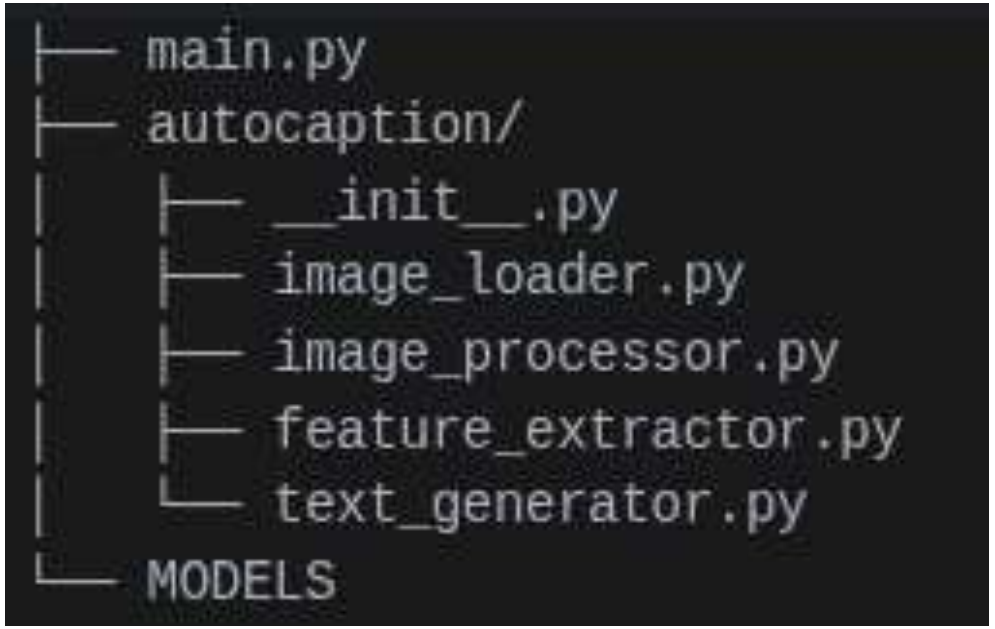
1. Развернуть мультимодальную модель
2. Оценить ее работу и исправить накопившиеся неточности

03

Ход работы

Планирование архитектуры проекта

- Принято решение сделать библиотеку autocaption, состоящую из четырех файлов: image_loader, image_processor, feature_extractor, text_generator
- В каждом классе содержатся методы работы с картинками
- Точка входа — файл main.py: изображение обрабатывается и проходит через модели, в конце получается текстовое описание



Файловая структура проекта

Модель, поворачивающая изображение

- Не все картинки в датасете были изначально повёрнуты правильно
- Дообучена модель ResNet50
- Из датасета отобраны картинки, имеющие изначально правильную ориентацию, в процессе обучения они поворачивались на определенный угол

Пример картинки с неправильной ориентацией

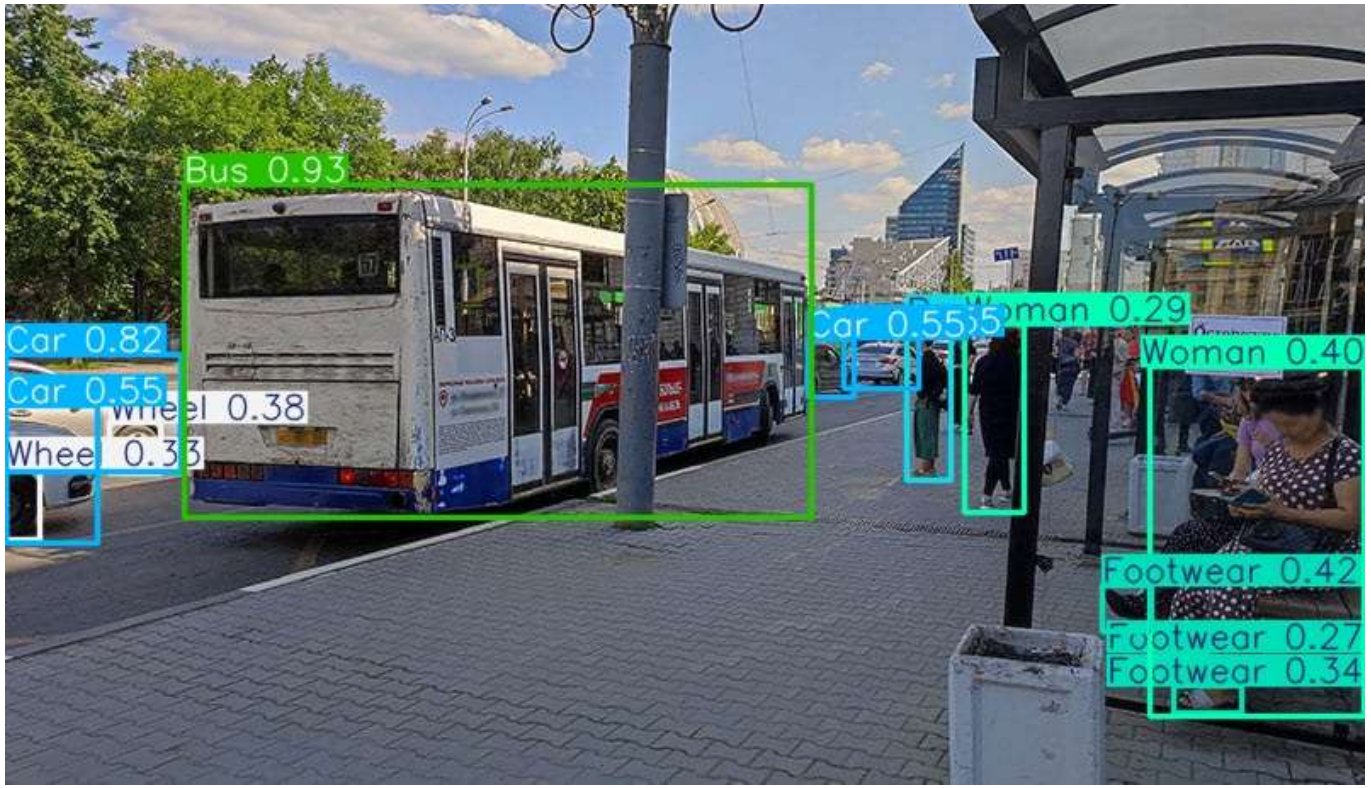
Модель, определяющая автомобиль на фото

- На некоторых фото нет автомобиля
- Обычно это некачественные фотографии
- В любом случае, решать задачу описания окружения автомобиля без автомобиля нецелесообразно

Пример фото, на котором нет автомобиля

Модель, которая находит предметы на фото

- Иногда по предметам можно понять окружение
- В обычных датасетах есть не все подходящие классы
- Была обучена модель YOLO на датасете Google Open Images, содержащим 60 классов



Найденные классы

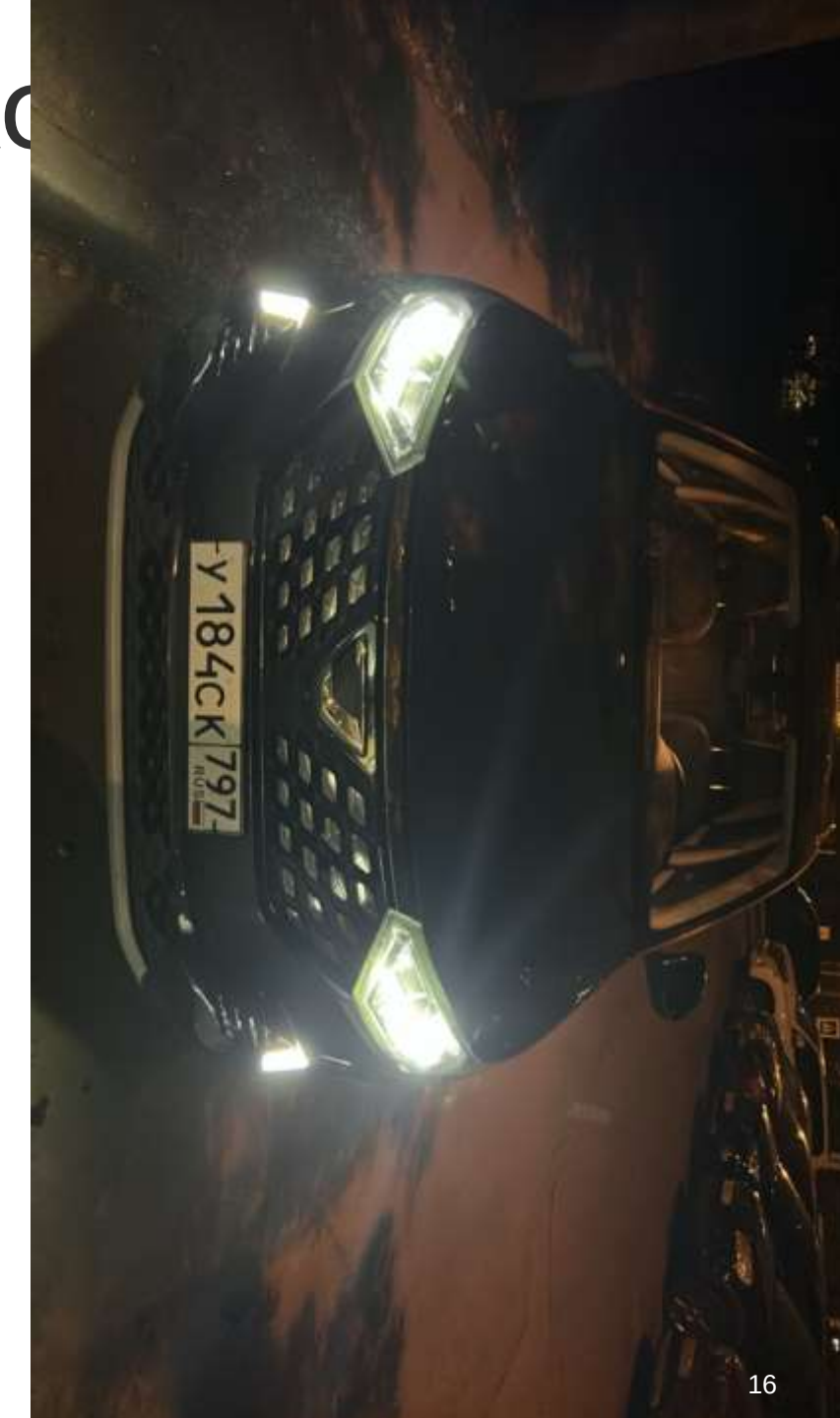
Модель, классифицирующая сцену

- Попытка получить информацию о сцене, решая задачу multi-label классификации
- Применена модель ResNet18, обученная датасете BDD100k, специально собранном для подобных задач

{'other': 0.08122928440570831, 'highway': 0.12126690149307251, 'residential': 0.8982969522476196, 'city street': 0.06356442719697952, 'parking lot': 0.7530761361122131, 'gas stations': 0.11552013456821442}

Модель, генерирующая кратко описание

- Попытка получить краткое описание картинки с помощью модели BLIP
- Использовано две модели: в одну подается вопрос, в другую - нет
- Получаем четыре разных описания: три от первой модели, и одно от второй




```
Базовое описание: the front view of a black suv parked on a street at night
Подробное описание: the front view of a black suv parked in a parking lot at
night
Альтернативное описание: the front end of a black car
Описание по вопросу (VQA): street
```

Мультимодальная модель

- Получает на вход все текстовые признаки, которые мы получили на прошлых этапах
- Генерирует описание картинки

«a car parked on a side of the road»

Эксперименты

То, что не вошло в итоговый проект, но заслуживает упоминания

04

Результаты

- У нас получилось выполнить поставленную задачу. Для каждой валидной картинки успешно генерируется описание
- В нашем датасете, к сожалению, даже человеческий глаз зачастую не может определить точное описание картинки: зачастую небо и автомобиль занимают большую часть фото

```
{'Car': 0.34714019298553467, 'Wheel': 0.3243175148963928}
{'other': 0.08122928440570831, 'highway': 0.12126690149307251, 'residential': 0.8982969522476196, 'city
street': 0.06356442719697952, 'parking lot': 0.7530761361122131, 'gas stations': 0.11552013456821442}
Базовое описание: a car parked in a parking lot
Подробное описание: a car parked in a parking lot with a parking meter on the side of the road and a
parking meter on the other side of the road
Альтернативное описание: a parked vehicle in a parking lot
Описание по вопросу (VQA): parking lot

Итоговое описание: The image features a silver car parked in a parking lot. The car is positioned in the
middle of the parking lot, and there are several other cars parked around it.
The location appears to be a city street, with buildings and infrastructure nearby.
```

05

Дальнейшие планы на проект

- Улучшить качество моделей
- Интегрировать в бизнес-процесс
- Разработать метрики для отслеживания полезности описаний в реальных сценариях
- Регулярно обновлять датасет новыми фотографиями и дообучать модель

Спасибо!

Смоляр Родион

Бирюков Григорий

Ментор: Ковалева Маргарита

Попов Артемий

Зайцев Илья

Студкемп по компьютерному зрению и автономному транспорту

Яндекс Образование