

Индексация сцен, снятых с беспилотного автомобиля, по текстовому запросу

Руководитель

Шахуро Владислав Игоревич

Получил степень специалиста на ВМК МГУ в 2015 году, в 2018 году закончил аспирантуру ВШЭ по компьютерным наукам. Защитил диссертацию в 2021 году, кандидат компьютерных наук. С 2015 по н.в. года работает в лаборатории компьютерной графики и мультимедиа ВМК МГУ. С 2021 по 2024 – ведущий инженер в Samsung AI Center Moscow. С весны 2024 является старшим научным сотрудником, руководителем группы “ИИ для роботов” лаборатории FusionBrain в AIRI. Автор 18 публикаций, включая 2 публикации на конференциях уровня A* AAAI и NeurIPS. С 2014 является семинаристом, а с 2021 лектором курса “Компьютерное зрение”, читаемом в Школе Анализа Данных Яндекса, в магистратуре ФПМИ МФТИ и на ВМК МГУ. Область научных интересов включает задачи компьютерного зрения и робототехники.

Зачем делаем?

Управление беспилотным автомобилем — хорошо поставленная задача, которая до сих пор не решена в силу “длинного хвоста” сложных случаев, которые необходимо корректно обрабатывать. Беспилотники собирают много данных, которые дорого размечать, поэтому развитие методов автоматической фильтрации собираемых данных выглядит хорошим способом дополнения обучающей выборки редкими сценами. Для этого нужны автоматические методы фильтрации данных для поиска сложных случаев в дорожных сценах

Что делаем?

Будем разрабатывать метод, который на вход принимает текстовый запрос (напр., “найти автомобиль, поворачивающий направо”) и проводит поиск в базе данных, состоящих из видео.

Как достигнем цель?

Нужно:

1. Провести обзор литературы, выбрать базовые методы поиска сцен в данных, снятых с беспилотных автомобилей
2. Выбрать бенчмарки, на которых будем проверять качество методов. Если такие бенчмарки отсутствуют или обладают существенными недостатками, нужно будет подготовить свой бенчмарк в партнерстве со NAVIO (бывш. Сбер.Автотех)
3. На основе базовых методов предложить улучшенный метод с использованием VLLM. Базовая гипотеза — VLLM за счет обучения на большом корпусе данных должны хорошо работать как для часто встречаемых сцен, так и редких сцен
4. Показать улучшение качества работы в сравнении с базовым, апробировать на данных NAVIO

Как измерить достижение результата?

Измерить точность поиска сцен по тестовой выборке бенчмарка