Contribution ID: **44** Type: **not specified**

Применение мультимодальных языковых моделей к задаче визуального вопросно-ответного анализа на видеоданных

Tuesday, 20 May 2025 13:28 (12 minutes)

В данной работе рассматривается применение мультимодальных языковых моделей (MLLM) к задаче визуального вопросно-ответного анализа (Video Question Answering, VideoQA) на основе видеоданных. Предложеный модульный подход включает отбор ключевых кадров с использованием CLIP, построение графа сцены по пространственно-семантическим отношениям между объектами с помощью MLLM и генерацию ответа на вопрос пользователя. Проведено экспериментальное сравнение различных MLLM и методов представления визуальных объектов.

Primary authors: Mr YUDIN, Dmitry (Заведующий лабораторией интеллектуального транспорта МФТИ - НКБ ВС); Mr LINOK, Sergey (Научный работник лаборатории интеллектуального транспорта МФТИ - НКБ ВС); SEMENOV, Vadim

Presenter: SEMENOV, Vadim

Session Classification: 20-Машинное обучение и нейросети

Track Classification: Машинное обучение и нейросети