



Исследование TRASSIR и AIRI: применение мультимодальных моделей в задачах видеоаналитики

TRASSIR и Институт AIRI завершили первый этап совместного исследовательского проекта, направленного на изучение потенциала мультимодальных моделей в анализе видеоданных. Проект сосредоточен на разработке и апробации эффективных пайплайнов видеоаналитики с использованием современных ИИ-архитектур.

Проект был направлен на создание эффективного пайплайна видеоаналитики с использованием мультимодальных моделей для расследования краж товара в магазинах. В процессе работы особое внимание уделялось снижению уровня ложных срабатываний и обеспечению устойчивости алгоритмов при отсутствии дополнительного обучения.

В исследовании использовались мультимодальные модели, работающие в zero-shot-режиме, то есть без дообучения на специализированных выборках. Такой подход позволил оценить универсальность моделей и определить границы применимости к задачам видеоаналитики.

Разработанный пайплайн учитывает и ограничения — объём видеопамати и скорость обработки, а также хорошо справляется с анализом сцен. Проблемной зоной являются короткие, быстрые эпизоды. В ответ на этот вызов предложены методы повторной обработки видео с фокусировкой на детализацию значимых фрагментов.

Были реализованы следующие методические улучшения:

- Механизмы рассуждения, позволяющие лучше распознавать сложные действия (например, "взятие+возврат" товара).
- Детализация эпизодов интереса для более точного анализа коротких событий.
- Поддержка потокового видео с учётом длинного временного контекста.

Так, в рамках одного из тестовых сценариев система изначально ошибочно определила резкий наклон покупателя как попытку кражи. После внедрения механизма рассуждения и анализа последовательности действий система корректно распознала, что покупатель просто наклонялся за упавшим товаром.

Анализ решений западных коллег, включая Amazon, показал, что основными ограничениями зарубежных систем являются отсутствие адаптации моделей к специфике целевого домена и слабая эффективность при анализе краткосрочных событий. Учёт этих факторов позволил создать более устойчивую архитектуру.

Результаты исследования задали направление для последующих работ в области прикладной видеоаналитики. Один из будущих проектов предусматривает использование мультимодальных моделей для оценки соблюдения стандартов обслуживания в предприятиях общественного питания, включая: анализ аудиокomпоненты на предмет соблюдения речевых скриптов, визуальный контроль за соблюдением санитарных норм и

стандартов внешнего вида персонала, а также детекцию нарушений, связанных с условиями допродаж и чистотой торгового пространства.

Особенность проекта заключается в интеграции речевых и визуальных данных в рамках единого аналитического пайплайна — одного из первых коммерчески ориентированных решений на основе мультимодальных ИИ-моделей.

TRASSIR и AIRI планируют продолжать исследовательскую деятельность в направлении разработки интеллектуальных и адаптивных систем видеоанализа.

Вопросы: pr@airi.net

Институт [AIRI](#) — автономная некоммерческая организация, занимающаяся фундаментальными и прикладными исследованиями в области искусственного интеллекта. На сегодняшний день более 200 научных сотрудников AIRI задействовано в исследовательских проектах Института для работы совместно с глобальным сообществом разработчиков, академическими и индустриальными партнерами.

[TRASSIR](#) – крупнейший российский разработчик открытой цифровой платформы и производитель оборудования для автоматизации безопасности и бизнес-процессов. Создает востребованные рынком продукты интеллектуальных систем видеонаблюдения. В портфеле компании: программное обеспечение, видеоаналитика, камеры, регистраторы и т.п. Продукция сертифицирована МинЦифры и МинПромТорга.