

Санкт-Петербургский политехнический университет Петра Великого
Институт компьютерных наук и технологий
Высшая школа программной инженерии

ОТЧЁТ
« ПРИМЕНЕНИЕ НЕЙРОННЫХ СЕТЕЙ ДЛЯ ОЦЕНКИ
ТРАФИКА МАГАЗИНА »

Ли Ицзя

3530904.90102

Санкт-Петербург
2023

Оглавление

| | |
|---|----|
| Реферат | 4 |
| Abstract | 5 |
| Введение..... | 5 |
| ГЛАВА 1 Обзор литературы..... | 7 |
| Распознавание пешеходов..... | 7 |
| Алгоритм распознавания объектов YOLO..... | 9 |
| Отслеживание (Трекинг) людей в виде последовательности | 10 |
| Алгоритм трекинга множества объектов DeepSORT..... | 11 |
| Повторная идентификация людей..... | 13 |
| Алгоритм Fast reid..... | 15 |
| Метод подсчета количества людей в видеопоследовательности..... | 18 |
| Основные исследования этой статьи | 18 |
| ГЛАВА 2 Обучение модели для отслеживания людей в нескольких видеопотоках | 19 |
| Обучение модели для повторной идентификации FastReID..... | 19 |
| Набора данных | 19 |
| Принцип обучения..... | 20 |
| Построить модель | 21 |
| Функции потери и алгоритмы оптимизации..... | 22 |
| Ходы обучения..... | 23 |
| Тренировочный процесс | 24 |
| Результаты обучения | 24 |
| Реализация алгоритма DeepSORT с помощью FastReID и yolov5..... | 28 |
| Устраивание модели | 28 |
| Набор тестовых данных MOT-16 | 28 |
| Эксперимент 1: Сравнительное тестирование улучшенного алгоритма DeepSORT и исходного алгоритма DeepSORT. | 30 |
| Эксперимент 2: сравнение улучшенного алгоритма DeepSORT с несколькими основными алгоритмами..... | 32 |
| ГЛАВА 3 Система оценки трафика магазина основа на нейронных сетях | 33 |
| Цели разработки системы | 33 |

| | |
|---|----|
| Описание системы | 33 |
| Общая архитектура системы | 33 |
| Модуль обнаружения пешеходов..... | 34 |
| Модуль извлечения признаков | 34 |
| Модуль отслеживания пешеходами..... | 35 |
| Модуль подсчета посетителей..... | 35 |
| Модуль хранения | 37 |
| Среда разработки | 37 |
| Демонстрация и тестирование системы | 38 |
| Алгоритм статистики трафика..... | 38 |
| Тестовая среда..... | 39 |
| Междукамерное отслеживание пешеходов..... | 39 |
| Метрики оценки системы..... | 45 |
| ГЛАВА 4 Анализ полученных результатов | 51 |
| Экспериментальная оценка модели FastReID на открытой базе данных | 51 |
| Экспериментальная оценка улучшенной модели DeepSORT на открытой базе данных | 52 |
| ЗАКЛЮЧЕНИЕ | 56 |
| Список Литературы..... | 57 |

Реферат

На - с., - рисунков, - таблицы, - приложение

KEY WORDS: Computer vision, Neural networks, Object detection, Multiple object tracking, Pedestrian re-identification

КЛЮЧЕВЫЕ СЛОВА: КОМПЬЮТЕРНОЕ ЗРЕНИЕ, НЕЙРОННЫЕ СЕТИ, ОБНАРУЖЕНИЕ ОБЪЕКТОВ, ОТСЛЕЖИВАНИЕ НЕСКОЛЬКИХ ОБЪЕКТОВ, ПЕРЕИДЕНТИФИКАЦИЯ ПЕШЕХОДОВ, YOLO, Улучшенный DeepSORT алгоритм

Выпускная квалификационная работа на тему: «Применение нейронных сетей для оценки трафика магазина».

Данная статья посвящена реализации алгоритмов нейронных сетей для оценки трафика в магазине. Для решения поставленных задач были использованы методы глубокого обучения и компьютерного зрения.

Задачи, решаемые в ВКР:

1. Для обнаружения пешеходов был использован алгоритм yolov5
2. Для установления связи между персонажами при работе с несколькими камерами был реализован алгоритм повторного распознавания (reid).
3. Для отслеживания локальных зон был реализован алгоритм DeepSORT он основе повторного распознавания.
4. На основе методики, описанных в данной статье, была разработана система подсчета посетителей для супермаркета.

В результате исследования была разработана система подсчета посетителей для супермаркетов, которая может идентифицировать входящих и выходящих людей с точностью до 95,5% и повторную идентификацию пешеходов с точностью до 85,0%. В конце, на других больших открытых наборах данных мы сравнили реализацию DeepSORT, FastReID с другими популярными алгоритмами и добились неплохих результатов.

Abstract

Введение

С развитием розничной торговли и усилением конкуренции многие отечественные торговые центры осознали важность данных о потоках посетителей для принятия бизнес-решений. Поток покупателей – это количество покупателей, входящих и выходящих из торгового центра в единицу времени. Благодаря углубленному анализу пассажиропотока он может обеспечить научную основу для управления всеми аспектами торговых центров. Анализ потока посетителей играет важную роль в повышении научности ежедневных бизнес-решений торговых центров, комфортности торговой среды и рациональности распределения человеческих ресурсов.

Самым ранним методом подсчета является метод искусственного подсчета. Торговый центр отправляет несколько сотрудников на вход и выход и непрерывно подсчитывает посетителей, входящих в торговый центр, визуально в течение определенного периода времени. Преимущество этого метода в том, что он не требует вложений в оборудование для сбора информации, а данные обследования получаются более полными и гибкими, а недостаток в том, что он требует трудовых и материальных ресурсов, подвержен пропускам.

С появлением механического сенсорного оборудования, инфракрасного сенсорного оборудования и оборудования для камер наблюдения метод подсчета людей, в свою очередь, прошел две стадии: статистику физического прикосновения и статистику инфракрасного излучения. Физическая сенсорная статистика в основном используется на входах и выходах, где пешеходы проходят по очереди, но есть недостатки, такие как высокие требования к установке оборудования, неудобное использование, низкая скорость движения и невозможность использования в широком диапазоне; инфракрасная статистика имеет низкие требования к установке и относительно удобна в использовании, но их статистика погрешности велика и не может быть использована в сценариях большой площади, кроме того, эти методы имеют весьма очевидные недостатки: невозможно узнать время входа и выхода каждого покупателя из магазина. Данные, полученные этим методом, могут быть использованы только для качественного понимания и не имеют реальной ценности для статистического анализа. В сегодняшней жесткой

экономической конкуренции эти статистические методы совершенно не отвечают потребностям рынка.

Традиционные ручные методы проверки неэффективны и дорогостоящи, поэтому для удовлетворения быстро меняющихся требований рынка необходим более эффективный, точный и автоматизированный метод. Исходя из этого, мы провели данное исследование, целью которого было реализовать алгоритм мониторинга и расчета пассажиропотока супермаркетов с использованием метода нейронных сетей.

Основная цель этой статьи — предложить метод, который сочетает в себе алгоритмы обнаружения пешеходов, отслеживания пешеходов и повторной идентификации пешеходов для получения статистики посещаемости магазина. Реализованы алгоритм yolov5, алгоритм отслеживания DeepSORT и алгоритм повторной идентификации пешеходов FastReID. Даже выполнили точное отслеживание и подсчет покупателей, входящих и выходящих из магазина под разными углами и в разных условиях освещения, а также разработали систему статистики пассажиропотока для супермаркетов.

С точки зрения анализа **осуществимости**, мы считаем, что метод, предложенный в этой статье, имеет высокую осуществимость. Во-первых, используемые алгоритмы являются относительно зрелыми и широко используемыми технологиями в современной области глубокого обучения и полностью проверены экспериментально. Во-вторых, с точки зрения аппаратной и программной среды конфигурация, требуемая в этой статье, относительно невелика и может работать на обычных ПК. В-третьих, мы также проводим всестороннюю экспериментальную оценку точности, эффективности и стабильности предложенного метода, демонстрируя его осуществимость.

Для достижения цели этой статьи нам необходимо выполнить следующие задачи:

- 1) Собрать видеонаблюдения супермаркета и выполнить предварительную обработку;
- 2) Разработать и реализовать модуль обнаружения пешеходов на основе алгоритма yolov5;
- 3) Разработать и реализовать алгоритм отслеживания DeepSORT на основе алгоритма повторной идентификации пешеходов FastReID, который применяется для отслеживания пешеходов в видеопотоке;

4) Разработайте и реализовать алгоритм повторной идентификации пешеходов FastReID для ассоциации персонажей с назначенными посетителями под несколькими камерами;

5) Хранить и записывать фотографии и времени входа и выхода посетителей, количество посетителей за определенный период времени.

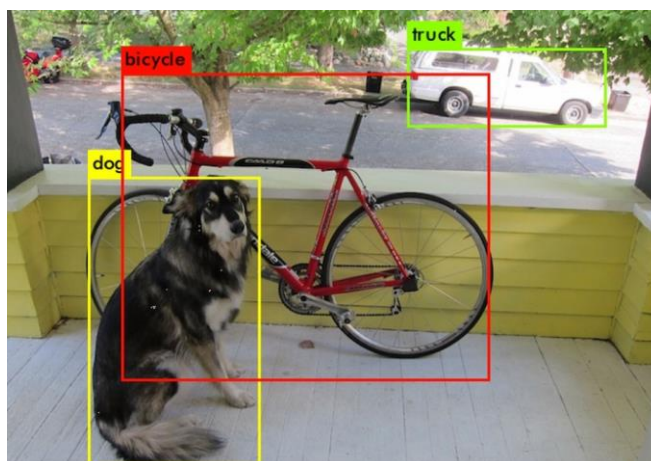
На этой основе мы разрабатываем практическую систему статистики потоков посетителей, ориентированную на супермаркеты, которая может обеспечить эффективное управление и поддержку принятия решений для супермаркетов.

ГЛАВА 1 Обзор литературы

Подсчет людей — это комплексная задача, объединяющая технологии обнаружения, отслеживания и повторной идентификации пешеходов. Ввиду этого в данной статье сначала описывается исследование обнаружения и отслеживания людей, затем вводится алгоритм повторной идентификации пешеходов и, наконец, вводится исследование подсчета людей на видеопоследовательности. Повторная идентификация пешеходов является самой важной технологией для решения проблемы ассоциации траекторий между камерами, а также основой данной исследовательской работы. Поэтому мы сосредоточимся и подробно представим реализованный нами алгоритм повторной идентификации пешеходов: FastReID.

Распознавание пешеходов

Суть обнаружения пешеходов заключается в обнаружении объектов. Обнаружение объектов лежит в основе задач многоцелевого сопровождения, основная цель которого — отметить объект, подлежащий обнаружению, прямоугольной рамкой (bounding box) из статической картинки каждого кадра и получить конкретное положение цели на картинке. Если при обнаружении объекта происходит пропущенное обнаружение, цель не будет отслеживаться позже, поэтому обнаружение объекта играет очень важную роль в задачах отслеживания нескольких целей.



Появление сверточных нейронных сетей значительно повысило эффективность методов обнаружения объектов. Такие алгоритмы можно условно разделить на следующие две категории. Одна из них представляет собой двухэтапный алгоритм обнаружения, который сначала выбирает области-кандидаты на изображении, а затем выполняет задачи классификации этих областей. Этот тип алгоритма характеризуется высокой точностью, но медленной скоростью. Например, R-CNN, предложенная Россом и др. [5], сначала предварительно определяет области, которые должны быть обнаружены на изображении, с помощью операции выборочного поиска, а затем извлекает признаки этих областей с помощью сверточной нейронной сети и, наконец, классифицирует их. Хотя алгоритм R-CNN значительно повышает точность обнаружения объектов, он имеет проблему вычислительной избыточности. Алгоритм sppnet, предложенный Хе и др. [6], устраняет влияние повторяющегося масштабирования изображений R-CNN путем добавления слоя объединения пространственных пирамид. Кроме того, Гиршик и др. [7] предложили алгоритм Fast R-CNN, который дополнительно оптимизировал алгоритмы R-CNN и sppnet, добавив слой объединения регионов. Рен и др. [8] предложили алгоритм Faster R-CNN, который заменил выборочный поиск сетью предложений регионов и обучил сеть от начала до конца, что значительно повысило скорость детектора. Другой представляет собой одноэтапный алгоритм обнаружения, который выполняет регрессию для вывода положения и категории объекта обнаружения при создании области-кандидата. Репрезентативным алгоритмом является алгоритм серии YOLO [9].

Серия алгоритмов YOLO была впервые предложена Редмоном и др. В 2016. Он отказался от двухэтапной парадигмы обнаружения «обнаружение предложения + проверка» и напрямую предсказал вероятность категории и положение каждой сетки, разделив изображение на несколько сеток. , информация.

Из приведенного выше введения мы делаем вывод, что традиционные подходы к обнаружению объектов обычно требуют ручного проектирования

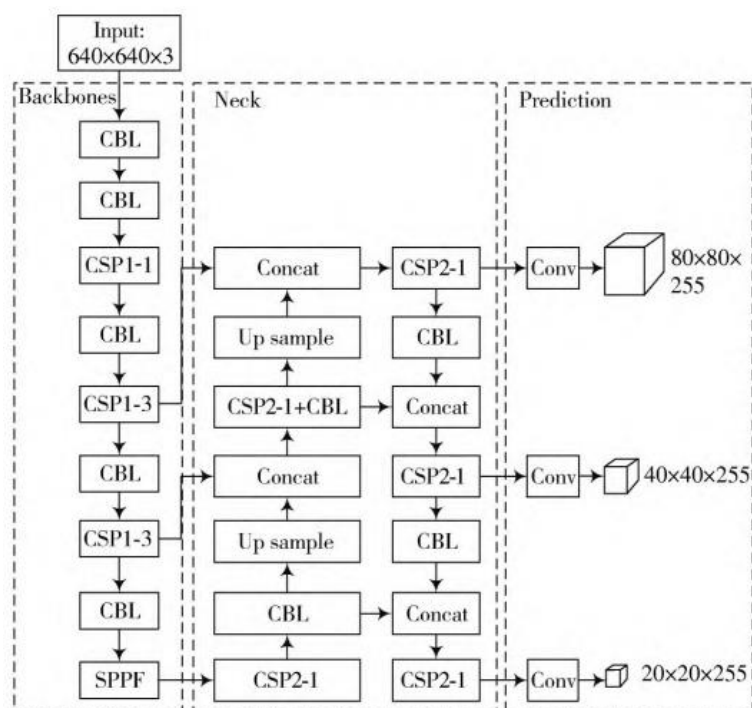
функций и классификаторов, которые плохо работают в сложных сценах и требуют большого опыта и времени для настройки параметров. Напротив, технология обнаружения объектов на основе глубокого обучения использует глубокие нейронные сети для автоматического изучения функций и классификаторов, что обеспечивает более высокую точность и надежность и подходит для различных сложных сцен. Обнаружение объектов является предпосылкой отслеживания объектов. Мы выбрали yolov5 в качестве детектора объектов для исследовательского проекта. Основная причина в том, что yolov5 отличается высокой точностью, высокой эффективностью, простотой в использовании и широкими возможностями настройки.

Исходя из вышеперечисленных преимуществ, yolov5 больше подходит для наших сценариев использования.

Алгоритм распознавания объектов YOLO

Yolov5 был предложен автором Glenn Jocher. Существует четыре размера моделей, а именно yolov5s, yolov5m, yolov5l и yolov5x. Среди них yolov5s имеет наименьшее количество сверточных слоев, самую быструю скорость обнаружения и самую низкую точность обнаружения, для остальных трех моделей количество сверточных слоев расположено в порядке возрастания. По мере увеличения сложности модели обнаружение. Скорость постепенно снижается, а точность обнаружения постепенно увеличивается.

Структура yolov5 состоит из четырех частей: Input, Backbones, Neck и Prediction, как показано на рисунке.



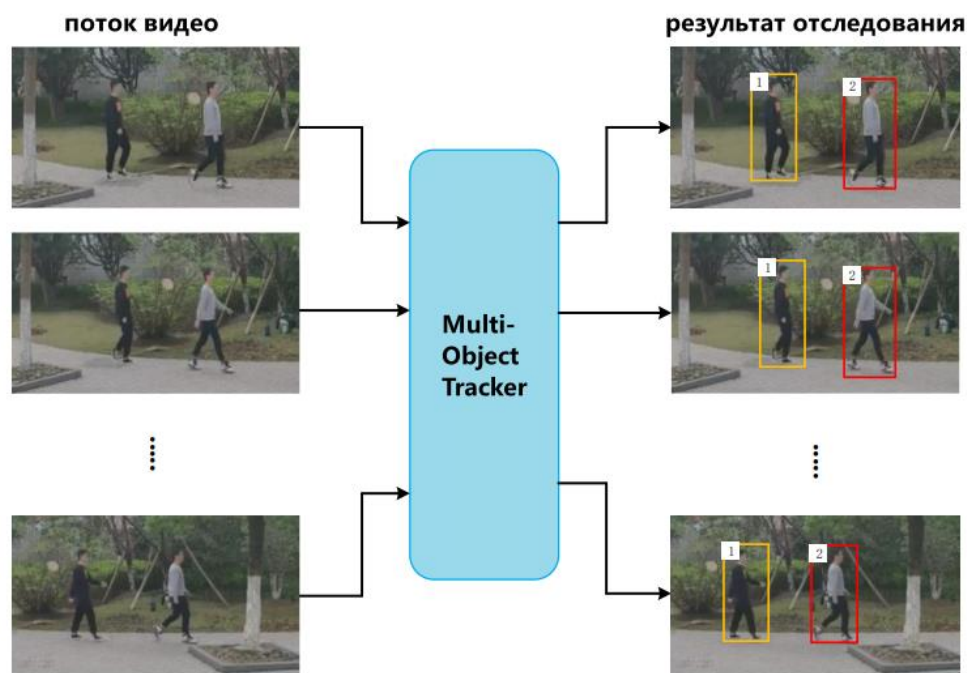
Сторона ввода включает в себя улучшение Mosaic, динамическую опорную рамку, адаптивную обработку изображений и т. Д.; Магистраль включает в себя кросс-этапную локальную сеть CSP, пространственный пирамидный пул SPPF, первый помогает сократить объем вычислений, а второй повышает точность обнаружения; Neck использует FPN+ Структура PAN, понижающая дискретизация улучшает семантическую информацию, повышающая дискретизация улучшает информацию о местоположении; Прогноз является конечным результатом обнаружения.

Чтобы добиться более высокой скорости, мы выбираем предварительно обученную модель yolov5s, выпущенную веб-сайтом с открытым исходным кодом yolov5 [Примечания].

Отслеживание (Трекинг) людей в виде последовательности

Задача трекинга объектов в видеопотоке заключается в сопоставлении детекция объекта на последовательности кадров видеопотока треку объекта. Задача трекинга множества объектов (Multi-Object Tracking, MOT) заключается в трекинге нескольких различных объектов. **【ОТСЛЕДОВАНИЕ ЛЮДЕЙ ПО ВИДЕО ПОСЛЕДОВАТЕЛЬНОСТИ Н.С. Захаров】** В частности, алгоритм трекинга множества объектов в основном решает проблему обнаружения объектов, которые мы хотим отслеживать, в каждом кадре видео, получения позиции на кадре и присвоения идентификатора каждому объекту. В процедуру трекинга идентификатор каждого объекта должен оставаться неизменным.





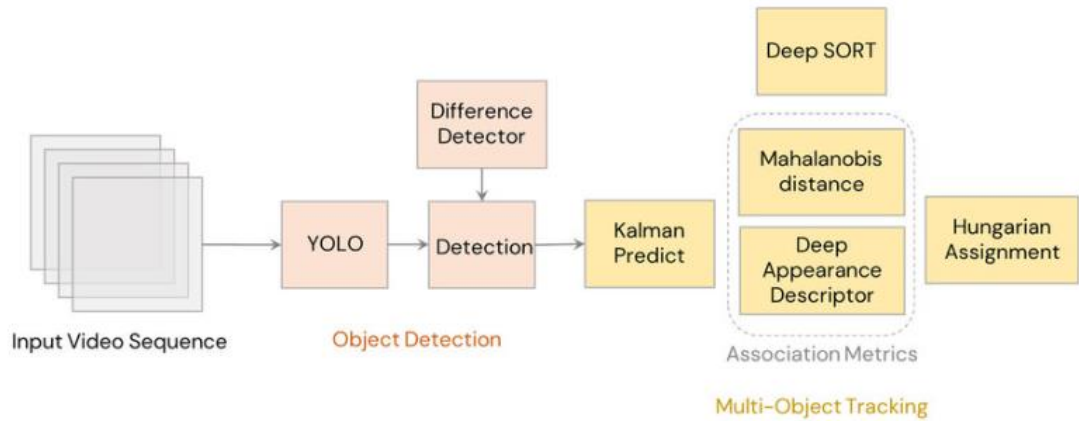
Как показано на рисунке, после ввода последовательности кадров, собранных камерой, в алгоритм MOT выводятся результаты отслеживания нескольких целей в кадре, включая положение цели на изображении и соответствующую идентификационную информацию.

По сравнению с отслеживанием одной цели задачи отслеживания нескольких целей должны не только сталкиваться с такими проблемами, как изменение освещения, окклюзия и размытие движения, но также решать такие проблемы, как неопределенное количество целей и прерывание траекторий целей во время процесса отслеживания. С непрерывным ростом исследований многоцелевого отслеживания в последние годы, связанные с ними методы появляются бесконечно. Далее рассматривается популярный алгоритм MOT – DeepSORT

Алгоритм трекинга множества объектов DeepSORT

DeepSORT является усовершенствованием алгоритма отслеживания SORT. Основная цель DeepSORT — генерировать непрерывные траектории движения множества персонажей в видеопотоке, а общий рабочий процесс показан на рисунке ниже. Чтобы уменьшить количество переключателей идентификатора цели отслеживания (переключение идентификатора, ID switch), DeepSORT вводит функции внешнего вида и каскадное сопоставление. Алгоритм DeepSORT принимает координаты кадра обнаружения и достоверность результатов обнаружения сети обнаружения объектов в качестве входных данных, предсказывает положение пешеходов в следующем кадре с помощью алгоритма фильтра Калмана, выполняет каскадное

сопоставление, а затем использует венгерский алгоритм для ассоциации данных. В конце, фильтр Калмана обновлен.



DeepSORT вычисляет сходство, используя информацию о движении и внешнем виде цели. Для информации о движении расстояние Махаланобиса используется для оценки корреляции между прогнозируемой целью и обнаруженной целью. Выражение расстояния Махаланобиса:

$$d^{(1)}(i, j) = (d_j - y_i)^T S_i^{-1} (d_j - y_i)$$

Где d_j – позиция bounding box j , y_i – позиция, предсказанное фильтром Калмана, S_i – ковариационная матрица обнаруженных и предсказанных позиций.

Когда цель закрыта в течение длительного времени или угол обзора колеблется, необходимо ввести информацию о внешнем виде и использовать косинусное расстояние для решения проблемы переключения идентичности, вызванного окклюзией.

Выражение косинусного расстояния:

$$d^{(2)}(i, j) = \text{Min} \left\{ 1 - r_j^T r_k^{(i)} \mid r_k^{(i)} \in R_i \right\}$$

Где r_j — собственный вектор bounding box d_j ; $r_k^{(i)}$ — набор собственных векторов, соответствующих последним 100 кадрам трека i ; R_i — набор векторов признаков внешнего вида.

Чтобы в полной мере использовать два вида информации, для суммирования используется линейно-взвешенный метод:

$$c_{i,j} = \lambda d^{(1)}(i, j) + (1 - \lambda) d^{(2)}(i, j)$$

В формуле: λ является гиперпараметром тогда и только тогда, когда значение измерения $c_{i,j}$ Существует между $d^{(1)}(i,j)$ И $d^{(2)}(i,j)$, только в этом случае он считается относящимся к цели.

Стоит заметить, что, поскольку алгоритм фильтра Калмана основан на линейном равномерном движении, алгоритм DeepSORT может предсказывать состояние цели только в линейной среде и может быть не в состоянии хорошо предсказать, когда траектория пешехода нелинейна.

Повторная идентификация людей

Повторная идентификация множества людей (Person Re-identification, также известная как Ре-идентификация человека или междуканальное сопровождение) — это новая технология, появившаяся в области интеллектуального анализа видео в последние годы. Ре-идентификация множества людей может быть определена как задача присвоения одного и того же имени или индекса всем образам одного и того же человека, получаемым с пространственно-разнесенных камер, области видимости которых не пересекаются друг с другом, на основе выделения и анализа признаков его изображений. **【 СОПРОВОЖДЕНИЕ И ПОВТОРНАЯ ИДЕНТИФИКАЦИЯ ЛЮДЕЙ Р. П. Богуш】** В видеонаблюдении из-за разрешения камеры и угла съемки обычно невозможно получить очень качественные изображения лиц. Когда распознавание лиц не работает, reid становится очень важной замещающей технологией. Очень важной особенностью reid является кроссканальность, поэтому при оценке производительности в научных работах необходимо получать одни и те же изображения пешеходов с разных камер.

Такие задачи характеризуется высокой сложностью реализации и требуют точной локализации людей в кадрах и правильной идентификации на текущем кадре или на кадре другой видеоканалы относительно предыдущих. **Одной из основных проблем является выбор дескриптора, описывающего человека.** **【Ye, S. Person Tracking and Re-Identification in Video for Indoor Multi-Camera Surveillance Systems / S. Ye, R. Bohush, C. Chen, I. Zakharava, S. Ablameyko // Pattern Recognition and Image Analysis, 2020. - Vol. 30, №4 – P. 827-837】** Для ее решения необходимо выявить отличительные признаки и путем сопоставления изображений людей из разных кадров или выполнения запроса сравнить их между собой или с признаками из имеющейся выборки изображений множества людей (галереи для ре-идентификации). Поиск и выделение набора наиболее отличительных особенностей объектов на изображениях, в том числе и людей, не формализован. Следовательно, требуется эмпирический поиск признаков, который в большинстве случаев является долгим и трудоемким процессом. Для сопровождения и повторной идентификации людей, из-за неоднозначности внешнего вида с разных ракурсов, вариаций освещения, различных разрешений камер, окклюзий такой

подход требует нерационально большое количество времени. Поэтому долгое время значимые результаты для указанных задач не достигались. Совершенствование средств вычислительной техники и открытия в области глубокого обучения, в частности, развитие сверточных нейронных сетей (СНС) позволили автоматизировать процесс извлечения признаков изображений людей и обеспечить значительное увеличение точности повторной идентификации, однако в полной мере решение не получено в настоящее время. **【 СОПРОВОЖДЕНИЕ И ПОВТОРНАЯ ИДЕНТИФИКАЦИЯ ЛЮДЕЙ Р. П. Богуш】**

Наиболее часто используемая концепция повторной идентификации пешеходов:

При выполнении задачи повторной идентификации пешеходов мы обычно разделяем все изображения пешеходов на две группы: набор запросов (Query) и набор галерей (Gallery).

Набор запросов (Query): коллекция изображений, которые мы хотим использовать для поиска подобных изображений в наборе галерей.

Набор галерей (Gallery): коллекция всех доступных изображений для поиска.

Probe (или query image): одно изображение или группа связанных изображений для поиска. В нашем случае это изображение человека.

Обычно мы используем "probe" для представления каждого изображения в наборе запросов, которое является целью нашего поиска, а "галерея" представляет собой набор изображений, в которых мы ищем соответствующие изображения. Цель повторной идентификации пешеходов заключается в том, чтобы найти все изображения пешеходов в наборе галереи, которые похожи или совпадают с изображениями в наборе запроса, и могут быть точно аутентифицированы или идентифицированы.



Задача повторной идентификации пешеходов в основном состоит из двух этапов: извлечение признаков и измерение сходства. Извлечение признаков - это обучение особенностей, способных справляться с изменениями пешеходов

под разными камерами. Обучение измерения - это отображение обученных признаков в пространство более высокой размерности, таким образом, чтобы одинаковые люди находились ближе друг к другу, а разные люди - дальше.

Традиционный метод заключается в ручном извлечении признаков изображения, например, цвета, HOG (гистограмма ориентированных градиентов)[4], SIFT (масштабно-инвариантное преобразование функций)[5], LOMO (локальный максимальный возникновение) и т. Д. Затем используется XQDA (кросс-видовой квадратичный дискриминантный анализ)[6] или KISSME (простое и прямолинейное изучение метрики)[7] для изучения лучшего измерения сходства. Однако традиционные методы ручного описания особенностей имеют ограниченные возможности и трудно применяются к задачам с большим объемом данных в сложных сценах. Кроме того, при обработке больших объемов данных традиционные методы обучения измерений также становятся очень сложными.

В последние годы глубокое обучение, представленное сверточными нейронными сетями, достигло больших успехов в области компьютерного зрения, победив традиционные методы в многих задачах и даже в какой-то мере превзойдя уровень человека [8-9]. В случае повторной идентификации пешеходов методы на основе глубокого обучения могут автоматически извлекать сложные описания признаков и использовать простые евклидовы расстояния для измерения сходства, что дает хорошие результаты. Другими словами, глубокое обучение может реализовать задачу повторной идентификации пешеходов полностью, что делает задачу более простой. В настоящее время методы повторной идентификации пешеходов на основе глубокого обучения значительно превосходят традиционные методы по производительности. Эти преимущества делают глубокое обучение популярным в области повторной идентификации пешеходов.

Мы выбрали алгоритм FastReID, основанный на метрическом обучении, в качестве метода повторной идентификации пешеходов исследовательского проекта, и эксперименты показали, что она сыграла очень хорошую роль в нашем проекте. Ниже мы представим его подробно.

Алгоритм Fast reid

Благодаря построению нейронной сети повторной идентификации WRN, DeepSORT способен эффективно решать проблему перекрытия. WRN состоит главным образом из остаточных блоков, имеет много параметров и требует больших вычислительных затрат. Чтобы улучшить скорость и способность идентификации повторной идентификации сети, мы вводим модель повторной идентификации человека FastReID взамен WRN. Это позволяет снизить вычислительную сложность, увеличить экспрессивность и достичь более быстрой скорости и лучшего эффекта идентификации. FastReID в целом

состоит из четырех частей: предварительная обработка изображения, основная сеть, интеграция функций и головные сети. Общая архитектура показана на следующей рисунке:

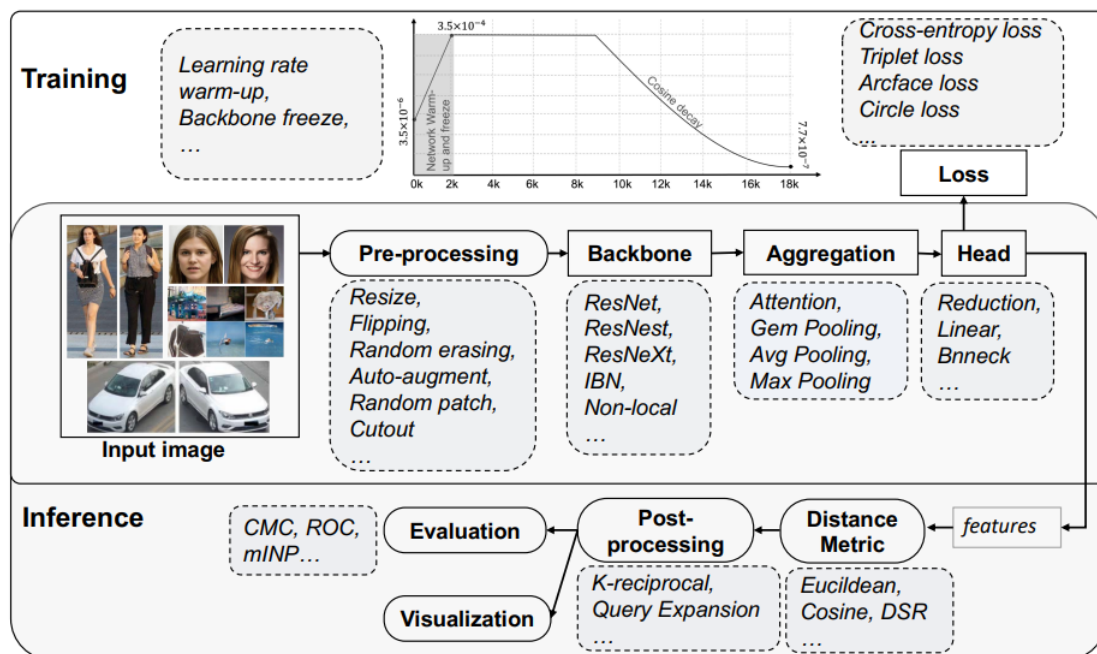


Figure 1. The Pipeline of FastReID library.

Training

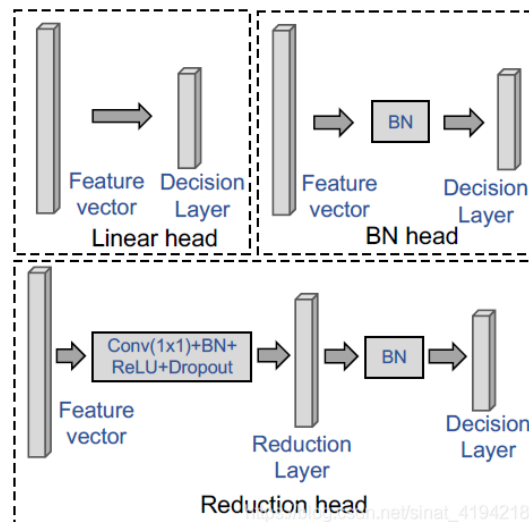
Модуль предварительной обработки данных (Pre-processing) : В модуле предварительной обработки данных FastReID используются различные методы увеличения данных, такие как Resize, Flipping, Random erasing, Auto-augment, Random patch, Cutout и т.д.; эти приемы повышают устойчивость функций.

Модуль основной сети (Backbone) Внутри модуля основной сети реализована модель извлечения черт resnet. Мы также добавили модуль instance batch normalization (IBN) в основную часть, чтобы изучить более устойчивые функции.

Модуль интеграции функций (Aggregation) Модуль интеграции характеристик используется для агрегации функций, сгенерированных основной сетью, в глобальную характеристику. Реализованы четыре операции пулинга: максимальное пулирование, среднее пулирование, Gem пулирование и Attention пулирование.

Модуль заголовков (Head) Модуль заголовка - это часть, которая обрабатывает глобальный вектор, созданный модулем агрегации, включая BN Head (batch normalization head), Linear head и Reduction head. Три типа заголовков показаны на рисунке 3: линейный заголовок содержит только один

слой decision, BN заголовок содержит слой BN и слой decision, а упрощенный заголовок содержит операции conv+BN+relu+dropout, слой reduction и слой decision.



Batch Normalization применяется для решения проблемы внутреннего сдвига ковариации, так как модели с насыщенной нелинейностью трудно обучать. Упрощенный слой (Reduction layer) направлен на уменьшение размерности высокомерных характеристик до 512 из 2048. Слой решений (Decision layer) выводит вероятности различных классов, отличает различные классы, используется для последующего обучения модели.

Функции потерь: Авторы реализовали четыре функции потерь, включая:

- Потерю кросс-энтропии (Cross-entropy loss)
- Тройную потерю (Triplet loss)
- Arcface loss
- Circle loss, предложенная Megvii в 2020 году, считается лучшей в настоящее время для всех видов задач машинного обучения.

Inference

Модуль расстояний (Distance Metric):

После обработки изображений модулями pre-processing, backbone, aggregation и head получается характеристический вектор размерности 512. Модуль расстояний выполняет сравнение вектора функций и галерейного векторного хранилища и выдает результат поиска: набор изображений, похожих на оригинальное изображение.

Модуль последующей обработки (Post-processing): Послеобработка относится к обработке результатов поиска, включая два способа повторной классификации (re-rank): k-reciprocal coding [11] и Query Expansion [12].

Модуль оценки производительности (Evaluation)

Для оценки эффективности мы использовали стандартные показатели, которые можно найти в большей части литературы по повторной идентификации личности, а именно кумулятивную кривую соответствия (СМС) и среднюю среднюю точность (mAP). Кроме того, мы добавляем две метрики: кривую рабочих характеристик приемника (ROC) и средний обратный отрицательный штраф (minp).

Модуль визуализации (Visualization)

Предоставляется инструмент рейтингового списка для поиска результатов, который полезен для отображения результатов повторной идентификации человека.

Метод подсчета количества людей в видеопоследовательности

Под визуальным подсчетом людей понимается возможность оценить количество людей в видео наблюдения с помощью определенной обработки. Визуальная статистика потоков людей широко используется во многих областях, в то же время широкое внедрение камер наблюдения и постоянное расширение сценариев применения статистики потоков людей привлекли большое внимание исследователей к проблеме статистики потоков людей. .

Многие компании в мире провели множество исследований и экспериментов с прикладными продуктами для подсчета людей, например, серия продуктов для подсчета людей Wise Count от компании Fei Ruisi имеет статистическую точность более 95% и может работать непрерывно в течение всего времени. День.В практических приложениях Он имеет высокую точность в реальном времени и статистическую точность.

Что касается статистики трафика покупателей, мы изучили алгоритм, основанный на сверточной нейронной сети, который может регистрировать конкретное время входа и выхода каждого покупателя из магазина для принятия бизнес-решений и анализа.

Основные исследования этой статьи

Основное содержание исследования этой статьи заключается в разработке метода, сочетающего алгоритм yolov5, алгоритм отслеживания

DeepSORT и алгоритм повторной идентификации пешеходов FastReID для реализации статистики пассажиропотока магазина. Путем оптимизации алгоритмов обнаружения, отслеживания и повторной идентификации пешеходов мы добились точного отслеживания и подсчета покупателей под разными углами и в различных условиях освещения, а также разработали систему статистики потока покупателей, ориентированную на магазины.

ГЛАВА 2 Обучение модели для отслеживания людей в нескольких видеопотоках

Обучение модели для повторной идентификации FastReID Набора данных

Набор данных MSMT17

На конференции CVPR2018 был предложен новый крупномасштабный набор данных MSMT17, который ближе к реальной сцене, а именно Multi-Scene Multi-Time, охватывающий несколько сцен и несколько периодов времени.

Набор данных MSMT17 использует сеть из 15 камер видеонаблюдения в кампусе, включая 12 наружных камер и 3 внутренние камеры. Для сбора необработанного видео наблюдения было выбрано 4 дня в месяц с разными погодными условиями. Каждый день собирается 3 часа видео, охватывающих три временных периода: утро, полдень и день. Таким образом, общая продолжительность необработанного видео составляет 180 часов.



Три комментатора-человека потратили два месяца на просмотр обнаруженных ограничивающих прямоугольников и пометку пешеходов. Наконец, получается 126 441 ограничительная рамка 4 101 пешехода.

Сравнение и статистическая информация с другими наборами данных показаны на рисунке ниже

| Dataset | <i>MSMT17</i> | <i>Duke</i> [41, 27] | <i>Market</i> [39] | <i>CUHK03</i> [20] | <i>CUHK01</i> [19] | <i>ViPeR</i> [8] | <i>PRID</i> [10] | <i>CAVIAR</i> [3] |
|------------|------------------------|----------------------|--------------------|--------------------|--------------------|------------------|------------------|-------------------|
| BBoxes | 126,441 | 36,411 | 32,668 | 28,192 | 3,884 | 1,264 | 1,134 | 610 |
| Identities | 4,101 | 1,812 | 1,501 | 1,467 | 971 | 632 | 934 | 72 |
| Cameras | 15 | 8 | 6 | 2 | 10 | 2 | 2 | 2 |
| Detector | Faster RCNN | hand | DPM | DPM, hand | hand | hand | hand | hand |
| Scene | outdoor, indoor | outdoor | outdoor | indoor | indoor | outdoor | outdoor | indoor |

Преимущества MSMT17 по сравнению с другими наборами данных заключаются в следующем:

- (1) Больше пешеходов, ограничивающих рамок и камер;
- (2) Сложные сцены и фоны;
- (3) охват нескольких периодов времени, поэтому есть сложные изменения освещения;
- (4) Лучший детектор пешеходов (Faster- RCNN), ограничительная рамка более точная.

Протокол оценки

Набор данных случайным образом делится в соотношении обучение-тестирование 1:3, а не поровну, как другие наборы данных. Целью этого является поощрение эффективных стратегий обучения, поскольку маркировка данных очень дорога в реальных приложениях.

Наконец, обучающий набор содержит 1041 пешехода с 32621 ограничивающей рамкой, а тестовый набор содержит 3060 пешеходов с 93820 ограничивающими рамками. Для тестового набора 11659 ограничивающих рамок выбираются случайным образом в качестве запроса, а остальные 82161 ограничивающие рамки используются в качестве галереи.

Показатели теста - кривая СМС и mAP. Для каждого запроса может быть несколько положительных совпадений.

Принцип обучения

Принцип обучения модели FastReID заключается в следующем:

1. Предварительная обработка данных: сначала изображения в наборе данных предварительно обрабатываются с вероятностью 0,5 обрезки, масштабирования, симметрии и улучшения данных, чтобы повысить устойчивость модели к изображениям людей в разных позах и условиях освещения.

2. Извлечение признаков: FastReID использует обученную сверточную нейронную сеть серии resnet в качестве экстрактора признаков для извлечения 512-мерного вектора признаков из входного изображения. В то же время FastReID также вводит некоторые новые методы извлечения признаков, такие как выравнивание с учетом масштаба (SAA) и случайное стирание, для дальнейшего улучшения способности выражения признаков.

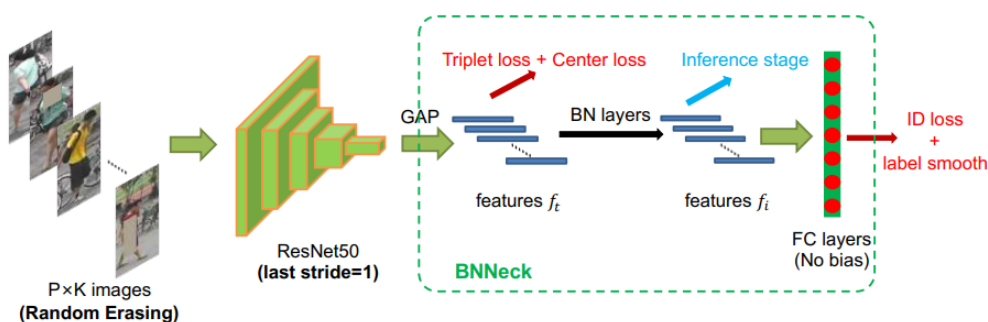
3. Метричное обучение: вектор признаков сопоставляется с низкоразмерным пространством посредством метрического обучения, а функции потери триплетов и перекрестной энтропийной потери используются для оптимизации модели, так что образцы с одной и той же идентичностью находятся ближе в пространстве признаков, а разные идентичности образцы находятся дальше друг от друга в пространстве признаков.

4. Оптимизация модели: используйте оптимизатор Adam для оптимизации параметров модели, чтобы повысить эффективность обучения и точность модели.

Построить модель

Чтобы ускорить сходимость моделей и повысить производительность модели, мы используем resnet50, модель сверточной нейронной сети, предварительно обученную на наборе данных imagenet. Imagenet — это широко используемый набор данных для классификации изображений, который содержит 1,4 миллиона изображений в более чем 1000 категориях. Модель resnet50 использует глубокую остаточную сетевую структуру, которая может эффективно справляться с крупномасштабными задачами классификации изображений. После завершения обучения веса модифицированной модели можно использовать в других задачах компьютерного зрения для повышения производительности и эффективности задачи, например, в нашей повторной идентификации пешеходов.

Размерность выходных признаков предварительно обученной модели resnet50 составляет 2048. Мы продолжаем вставлять средний слой пула и линейный классификатор после магистральной сети. Выходная размерность линейного классификатора равна N. N представляет собой идентификационный номер обучающих данных.



Функции потери и алгоритмы оптимизации

Функция потерь:

Наша модель Ре-идентификация дает два выхода: признак f и предсказанные логиты p . Признак f используются для расчета триплетных потерь, а логиты p используются для расчета энтропии перекрестных потерь.

Cross Entropy Loss

Crossentropyloss — это широко используемая функция потери классификации, которая обычно используется для обучения глубоких нейронных сетей задачам классификации изображений. В частности, для входной выборки x и ее истинной метки y формула расчета перекрестной энтропийной потери выглядит следующим образом:

$$\text{Cross Entropy Loss} = - \sum_{i=1}^C y_i \text{Log}(\hat{y}_i)$$

Где C — количество категорий, y_i Представляет вероятность i -й категории (то есть истинной метки), а \hat{y}_i Представляет вероятность того, что модель предсказывает, что x принадлежит i -й категории.

Цель crossentropyloss — минимизировать разрыв между предсказанием и реальной меткой, чтобы модель могла более точно предсказывать категорию, к которой относится каждая выборка (каждый человек в обучающей выборке рассматривается как отдельная категория).

Triplet Loss

Triplet Loss — одна из функций потерь, широко используемых в области распознавания лиц, назначение которой — приблизить векторы встраивания изображений (embedding) одного и того же человека и отдалить векторы встраивания изображений разных людей. В частности, Triplet Loss изучит вектор встраивания для каждого образца, что может сделать изображения одного и того же человека ближе, а изображения разных людей — дальше.

Для триплета, включающего три изображения привязки, положительное и отрицательное, формула расчета tripletloss выглядит следующим образом:

$$\text{Triplet Loss} = \text{Max}(0, d(a, p) - d(a, n) + m)$$

Где $d(a, p)$ представляет собой евклидово расстояние между привязкой и положительным значением, $d(a, n)$ представляет собой евклидово расстояние между привязкой и отрицательным значением, а m представляет собой отступ, который является предустановленным гиперпараметром A , обычно

положительным числом. Смысл этой формулы таков: если расстояние между текущим якорем и плюсом минус расстояние между якорем и минусом плюс маржа меньше или равно 0, это означает, что текущий вектор встраивания достаточно хорош и не нуждается в оптимизации. ; в противном случае необходимо обновить параметры модели для лучшего встраивания векторов.

Triplet Loss обычно используется в сочетании с Batch Hard Triplet Mining, то есть для обучения в каждой партии выбирается самый сложный триплет. Конкретный метод заключается в выборе изображений пешеходов, очень похожих на якорь, но разных категорий, в качестве отрицательных образцов, и выборе изображений пешеходов, которые сильно отличаются от якоря, но принадлежат к той же категории, что и положительные образцы, чтобы улучшить производительность модели.

Ходы обучения

1. Используем resnet50 (инициализированные веса взяты из предварительно обученной модели imagenet), а затем измените ее полносвязный слой на N. N представляет собой идентификационный номер обучающих данных.

2. Мы случайным образом отбираем P идентификаторов и собираем K изображений для каждого идентификатора, а размер последней партии $B = P * K$. В этой статье мы устанавливаем $P = 16$, $K = 4$.

3. Мы изменили размер каждого изображения на 256×128 и заполнили 10 пикселей значениями 0, а затем повторно обрезали изображение размером 256×128 , используя случайное кадрирование.

4. Каждое изображение случайным образом переворачивается по горизонтали с вероятностью 0,5.

5. Каждое изображение декодируется как 32-битное необработанное значение пикселя с плавающей запятой в $[0,1]$, а затем мы нормализуем каналы RGB, вычитая 0,485, 0,456, 0,406 и разделяя на 0,229, 0,224, 0,225.

6. Предсказать логиты p в соответствии с функциями reid f, выводимыми моделью.

7. Функции reid f используются для расчета триплетных потерь, логиты p используются для расчета энтропии перекрестных потерь, а запас m триплетных потерь устанавливается равным 0,3.

8. Для оптимизации модели используется метод Адама. Начальная скорость обучения установлена на 0,00035 и уменьшается на 0,1 в 40-ю и 70-ю эпохи соответственно. Всего 120 эпох.

[illegible]

```

[05/08 08:38:22 fastreid.data.dataset.bases] No Loaded Metadata in csv format:
| subset | # ids | # images | # cameras |
|-----|-----|-----|-----|
| query | 10000 | 11609 | 150 |
| gallery | 10000 | 10251 | 135 |
[05/08 08:38:22 fastreid.evaluation.evaluator] Start inference on 93620 images
[05/08 08:38:46 fastreid.evaluation.evaluator] Inference done 117/33. 0.6099 s / batch. ETA=0:08:04
[05/08 08:39:15 fastreid.evaluation.evaluator] Inference done 50/33. 0.6764 s / batch. ETA=0:07:19
[05/08 08:39:42 fastreid.evaluation.evaluator] Inference done 101/33. 0.6745 s / batch. ETA=0:07:07
[05/08 08:40:16 fastreid.evaluation.evaluator] Inference done 160/33. 0.6764 s / batch. ETA=0:06:37
[05/08 08:40:48 fastreid.evaluation.evaluator] Inference done 191/33. 0.6759 s / batch. ETA=0:06:06
[05/08 08:41:18 fastreid.evaluation.evaluator] Inference done 230/33. 0.6797 s / batch. ETA=0:05:36
[05/08 08:41:48 fastreid.evaluation.evaluator] Inference done 283/33. 0.6797 s / batch. ETA=0:05:06
[05/08 08:42:19 fastreid.evaluation.evaluator] Inference done 325/33. 0.6760 s / batch. ETA=0:04:36
[05/08 08:42:50 fastreid.evaluation.evaluator] Inference done 359/33. 0.6780 s / batch. ETA=0:04:07
[05/08 08:43:20 fastreid.evaluation.evaluator] Inference done 413/33. 0.6782 s / batch. ETA=0:03:37
[05/08 08:43:50 fastreid.evaluation.evaluator] Inference done 453/33. 0.6795 s / batch. ETA=0:03:08
[05/08 08:44:31 fastreid.evaluation.evaluator] Inference done 563/33. 0.6808 s / batch. ETA=0:02:18
[05/08 08:44:51 fastreid.evaluation.evaluator] Inference done 564/33. 0.6820 s / batch. ETA=0:02:09
[05/08 08:45:21 fastreid.evaluation.evaluator] Inference done 587/33. 0.6836 s / batch. ETA=0:01:40
[05/08 08:45:52 fastreid.evaluation.evaluator] Inference done 631/33. 0.6841 s / batch. ETA=0:01:10
[05/08 08:46:32 fastreid.evaluation.evaluator] Inference done 679/33. 0.6845 s / batch. ETA=0:00:39
[05/08 08:46:53 fastreid.evaluation.evaluator] Inference done 719/33. 0.6851 s / batch. ETA=0:00:09

```

```
[09/08 09:30:26] fastreid.evaluation.evaluator: Total inference time: 0:08:45.339901 (0.721615 s / batch per device, on 1 devices)
[09/08 09:30:26] fastreid.evaluation.evaluator: Total inference time: 0:08:45.339901 (0.721615 s / batch per device, on 1 devices)
[09/08 09:30:26] fastreid.evaluation.evaluator: Total inference time: 0:08:44 (0.720197 s / batch per device, on 1 devices)
WARNING:fastreid.fast_reid_fast_reid_fastreid.evaluation.rank.py:14: UserWarning: Cython rank evaluation (very fast so highly recommended) is unavailable, now use python evaluation.
warnings.warn()
[09/08 10:10:04] fastreid.engine.defaults: Evaluation results for MSMT17 in csv format:
[09/08 10:10:04] fastreid.evaluation.testing: Evaluation results in csv format:
| Dataset | Rank-1 | Rank-5 | Rank-10 | AP | mAP | mRec |
|-----|-----|-----|-----|-----|-----|-----|
| MSMT17 | 84.19 | 98.95 | 99.10 | 62.89 | 14.96 | 65.84 |
```

Метрики оценки эффективности модели

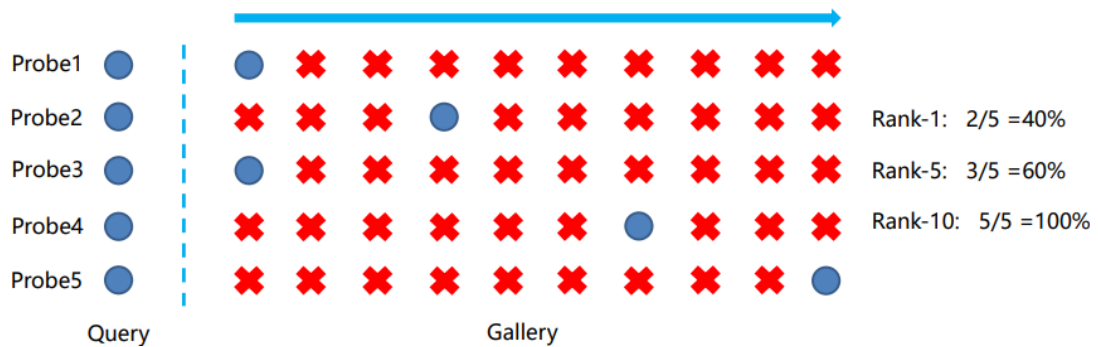
Кумулятивная кривая соответствия (СМС)

24 / 57

$$Rank - k = \frac{\sum_{probe \in Q} f(k, index_{probe})}{m}$$

$$f(k, index_{probe}) = \begin{cases} 0, & k \geq index_{probe} \\ 1, & k < index_{probe} \end{cases}$$

Где $index_{probe}$ Указывает местоположение первого изображения probe, принадлежащего тому же пешеходу, что и зонд изображения, который должен быть запрошен в результатах поиска, а m указывает количество изображений, которые должны быть извлечены в базе данных Q для извлечения. Вообще говоря, Ранг-1 является наиболее важным показателем для оценки эффективности модели повторной идентификации человека, который может быть выражен как вероятность того, что первое изображение в полученном отсортированном списке является тем же пешеходом, что и изображение, которое нужно получить. В реальных ситуациях, помимо использования ранга-1, другие часто используемые ранг-k включают ранг-5, ранг-10, ранг-20 и т. Д.



Средняя средняя точность (mAP)

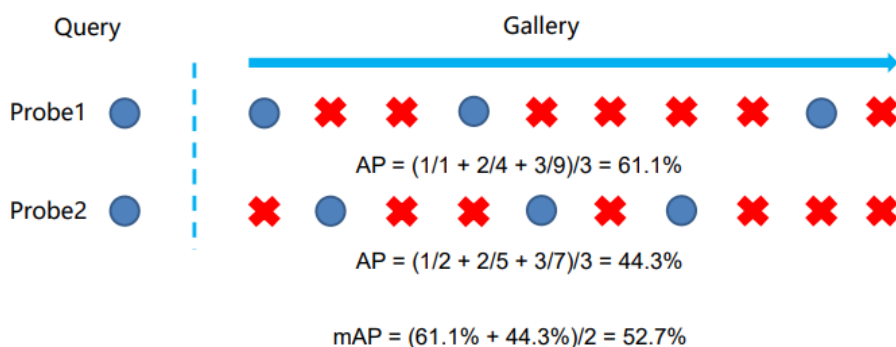
В ранних наборах данных повторной идентификации пешеходов в библиотеке изображений-кандидатов есть только одно изображение пешехода *probe*, которое совпадает с извлеченным изображением *probe*, но с введением больших наборов данных, таких как Market-1501 и DukeMTMC-ReID, изображение *probe* обычно можно найти несколько совпадающих изображений с одинаковым идентификатором в библиотеке изображений-кандидатов G , и трудно оценить влияние труднодоступных образцов на производительность модели, используя только индекс оценки СМС. Следовательно, для более полной оценки эффективности модели повторной идентификации человека к индексу оценки модели повторной идентификации человека добавляется средняя точность. Средняя точность — это индекс, который может оценить результаты ранжирования всех положительных образцов. Только когда все изображения извлеченного человека ранжируются вверху в библиотеке кандидатов, индекс mAP будет высоким, поэтому он может более полно отражать пешеходов. При расчете mAP сначала

вычисляется средняя точность (Average Precision, AP), соответствующая каждому извлекаемому изображению *probe*, которая используется для измерения точности распознавания модели на одном образце запроса. Процесс расчета показан в формуле:

$$AP(q) = \frac{\sum_{k \in k_1, k_2, \dots, k_S} \frac{k_r}{k}}{S}$$

Где S указывает количество положительных образцов, соответствующих полученному изображению *probe* в библиотеке кандидатов G , $\{k_1, k_2, \dots, k_S\}$ — позиция индекса S положительных образцов в результатах сортировки, а k_r указывает количество положительных образцов в первых k результатах. Наконец, после вычисления средней точности всех изображений пешеходов в библиотеке Q , содержащей m изображений, можно использовать среднее значение значений AP всех выборок для получения mAP. Процесс расчета показан в формуле:

$$mAP = \frac{\sum_{q_i \in Q} AP(q_i)}{m}$$

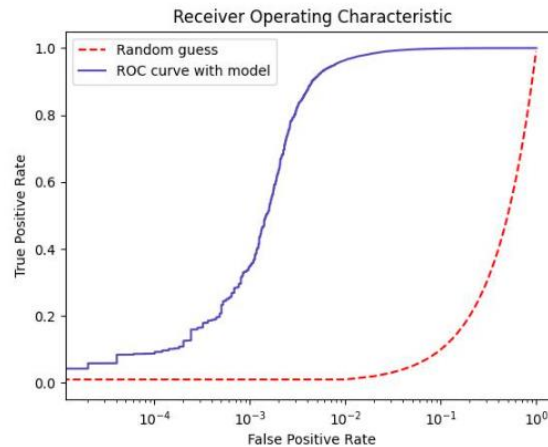


Среднее обратное отрицательное наказание (minp): указывает стоимость поиска всех правильных совпадений, чем ниже, тем лучше.

Результаты обучения на наборе данных MSMT17

Мы используем 1/3 данных в MSMT17 в качестве образцов для обучения модели FastReID, а затем используем оставшиеся 2/3 данных в качестве проверочного набора для проверки возможностей модели, и результаты следующие:

| Dataset | Rank-1↑ | Rank-5↑ | Rank-10↑ | MAP↑ | Minp↓ |
|---------|---------|---------|----------|-------|-------|
| MSMT17 | 84.19 | 90.95 | 95.10 | 62.89 | 14.94 |



По приведенным данным видно, что модель FastReID добилась хороших результатов на наборе данных MSMT17. В частности, она достиг 84,19% точности на 1-м ранге, 90,95% и 95,10% на 5-м и 10-м рангах соответственно. При этом средняя точность (mAP) также достигла 62,89%, что означает, что производительность модели на всем наборе данных относительно стабильна. Следует отметить, что модель составляет всего 14,94% с точки зрения обратного среднего коэффициента штрафа за отрицательную выборку (minp), Следует отметить, что с точки зрения обратного среднего отрицательного штрафа за выборку (minp) модель составляет 14,94%, что означает, что в некоторых трудно идентифицируемых случаях производительность модели может снижаться. В целом эти результаты показывают, что модель FastReID хорошо работает с набором данных MSMT17, но все еще есть возможности для улучшения при применении к особенно сложным сценариям распознавания.

Сравним производительность других моделей Reid в этом наборе данных:

| Methods | MSMT17 | |
|----------------------|--------|------|
| | Rank-1 | MAP |
| Ianet(IVPR'19) | 75.7 | 45.8 |
| Auto-reid(ICC'19) | 78.2 | 52.5 |
| Osnet(ICC'19) | 78.7 | 52.9 |
| Abdnet(ICC'19) | 82.3 | 60.8 |
| Circle Loss(CVPR'20) | 76.9 | 52.1 |

| | | |
|-------------|--------------|--------------|
| Ours | 84.19 | 62.89 |
|-------------|--------------|--------------|

Через сравнение мы знаем, что точность нашей модели намного выше, чем у других моделей, выпущенных в то же время, и показатели точности являются лучшими.

Реализация алгоритма DeepSORT с помощью FastReID и yolov5

Устраивание модели

1. Используем DeepSORT в качестве сети отслеживания, чтобы отслеживать каждого человека и сопоставлять его с предыдущими траекториями.
2. Используем yolov5s в качестве части обнаружения DeepSORT для обнаружения людей в каждом кадре изображения и вывода кадра обнаружения.
3. Используем FastReID в качестве экстрактора признаков для извлечения признаков внешнего вида пешеходов, чтобы DeepSORT мог различать разных людей.
4. Определим гиперпараметры модели. Параметры обучающей модели следующие:

```
DEEPSORT:
  REID_CKPT: "./fast-reid/checkpoint/model-final.pth"
  MAX_DIST: 0.2
  MIN_CONFIDENCE: 0.3
  NMS_MAX_OVERLAP: 0.5
  MAX_IOU_DISTANCE: 0.7
  MAX_AGE: 140
  N_INIT: 3
  NN_BUDGET: 100
```

Набор тестовых данных MOT-16

Мы используем общедоступный набор данных о пешеходах MOT-16 для тестирования нашей модели отслеживания нескольких объектов (DeepSORT после улучшения модуля повторной идентификации пешеходов). Этот набор данных является общим набором данных, используемым для оценки производительности алгоритмов отслеживания нескольких целей. Существуют различные пешеходные сцены, в том числе 14 видеопоследовательностей, а 7 видеопоследовательностей помечены подробно с идентификаторами пешеходов и позициями кадров, которые используются для обучать многоцелевых пешеходов Алгоритм слежения, еще 7 видеофрагментов служат тестовой выборкой.

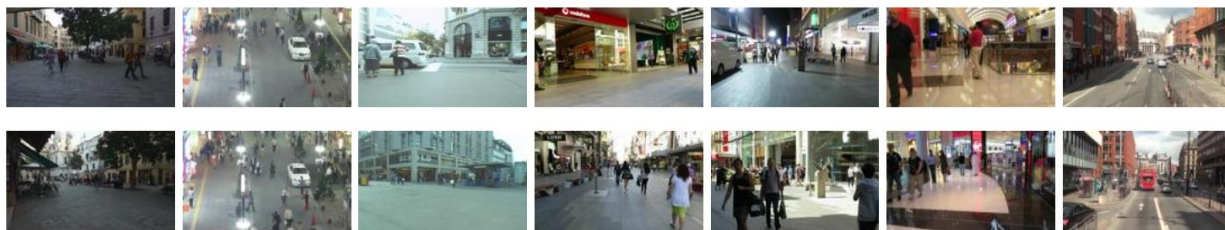


Figure 1 An overview of the MOT16 dataset. Top: Training sequences; bottom: test sequences.

Критерии оценки, использованные в эксперименте: точность сопровождения (MOTA), точность сопровождения (MOTP), количество правильно сопровождаемых траекторий цели выше 80% (MT), количество правильно сопровождаемых траекторий цели ниже 20% (ML), преобразование идентификатора цели Время (IDSW), точность отслеживания и точность отслеживания рассчитываются следующим образом:

$$MOTA = 1 - \frac{\sum_t (m_t + n_t + s_t)}{\sum_t g_t}$$

$$MOTP = \frac{\sum_{i,t} d_i^t}{\sum_t c_t}$$

Где t представляет t -й кадр, m_t представляет количество пропущенных целей обнаружения в t -кадре; n_t представляет количество ложных целей обнаружения в t -кадре; s_t представляет количество переключений идентичности в t -кадре; g_t представляет общее количество целей в кадре t ; расстояние между прогнозируемой позицией цели кадра i и реальной позицией; c_t представляет количество успешно сопоставленных целей в кадре t .

Все эксперименты проводились на видеопоследовательности тестового набора MOT- 16. Чтобы лучше подчеркнуть производительность алгоритма в этой статье, для сравнения были разработаны два разных эксперимента.

Эксперимент 1: Сравнительное тестирование нашего улучшенного алгоритма DeepSORT (yolov5+FastReID+DeepSORT) и оригинального алгоритма DeepSORT. Протестируйте оба варианта на тестовом видео, чтобы проанализировать плюсы и минусы улучшенного алгоритма DeepSORT.

Эксперимент 2. Анализ надежности улучшенного алгоритма в различных сценариях. Сравните улучшенный алгоритм с несколькими основными алгоритмами и проанализируйте преимущества и недостатки улучшенного алгоритма.

Эксперимент 1: Сравнительное тестирование улучшенного алгоритма DeepSORT и исходного алгоритма DeepSORT.

Процедура эксперимента

Мы протестируем все видеопоследовательности в наборе данных MOT16, в которых идентификаторы пешеходов и положения границ отмечены вручную (gt.txt), всего 7. Это: MOT16-02, MOT16-04, MOT16-05, MOT16-09, MOT16-10, MOT16-11, MOT16-13. Здесь в качестве примера взят MOT16-13, а операции для других видео аналогичны.

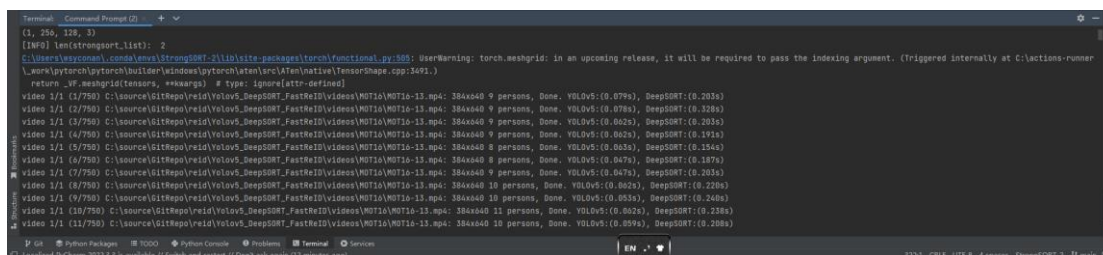
1. Загрузите набор данных MOT16, получите видеофрагменты и соответствующие им файлы gt.txt. Файл gt.txt в MOT16-13 выглядит следующим образом (частично), а параметры разделены запятыми:

```
1,1,1376,485,37,28,0,11,1
2,1,1379,486,37,28,0,11,1
3,1,1382,487,38,29,0,11,1
4,1,1386,488,38,29,0,11,1
5,1,1389,490,38,29,0,11,1
```

Файл gt.txt представляет собой текстовый файл CSV, каждая строка содержит объект, описывающий отслеживаемый объект в одном из фреймов, с 9 значениями, разделенными запятыми. Trackeval использует только первые 6, номер кадра, идентификатор цели и 4 координаты кадра отслеживания, а последние 3 (надежность цели, категория цели, видимость) не участвуют в расчетах и могут быть проигнорированы следующим образом:

<frame>, <id>, <bb_left>, <bb_top>, <bb_width>, <bb_height>, <conf>, <class>, <visibility>

2. Запустите программу track.py в созданной нами модели DeepSORT.



```
Terminal - Command Prompt (2)
(1, 256, 388, 3)
[INFO] len(strongsort_list): 2
C:\Users\yasyanov\conda\envs\strongsort_2\lib\site-packages\torch\functional.py:305: UserWarning: torch.meshgrid: in an upcoming release, it will be required to pass the indexing argument. (Triggered internally at C:\actions-runner
...\work\python\pytorch\builder\windows\pytorch\aten\src\aten\native\TensorShape.cpp:3493.)
return _VF.meshgrid(tensors, **kwargs) # type: ignore[attr-defined]
video 1/1 (1/750) C:\source\GitRepo\reid\VolovS_DeepSORT_FastReID\videos\MOT16\MOT16-13.mp4: 384x640 9 persons, Done. VLOvS:(0.679), DeepSORT:(0.2034)
video 1/1 (2/750) C:\source\GitRepo\reid\VolovS_DeepSORT_FastReID\videos\MOT16\MOT16-13.mp4: 384x640 9 persons, Done. VLOvS:(0.678), DeepSORT:(0.3284)
video 1/1 (3/750) C:\source\GitRepo\reid\VolovS_DeepSORT_FastReID\videos\MOT16\MOT16-13.mp4: 384x640 9 persons, Done. VLOvS:(0.662), DeepSORT:(0.2034)
video 1/1 (4/750) C:\source\GitRepo\reid\VolovS_DeepSORT_FastReID\videos\MOT16\MOT16-13.mp4: 384x640 9 persons, Done. VLOvS:(0.662), DeepSORT:(0.1914)
video 1/1 (5/750) C:\source\GitRepo\reid\VolovS_DeepSORT_FastReID\videos\MOT16\MOT16-13.mp4: 384x640 8 persons, Done. VLOvS:(0.6634), DeepSORT:(0.1942)
video 1/1 (6/750) C:\source\GitRepo\reid\VolovS_DeepSORT_FastReID\videos\MOT16\MOT16-13.mp4: 384x640 9 persons, Done. VLOvS:(0.647), DeepSORT:(0.1974)
video 1/1 (7/750) C:\source\GitRepo\reid\VolovS_DeepSORT_FastReID\videos\MOT16\MOT16-13.mp4: 384x640 9 persons, Done. VLOvS:(0.647), DeepSORT:(0.2034)
video 1/1 (8/750) C:\source\GitRepo\reid\VolovS_DeepSORT_FastReID\videos\MOT16\MOT16-13.mp4: 384x640 10 persons, Done. VLOvS:(0.662), DeepSORT:(0.2284)
video 1/1 (9/750) C:\source\GitRepo\reid\VolovS_DeepSORT_FastReID\videos\MOT16\MOT16-13.mp4: 384x640 10 persons, Done. VLOvS:(0.655), DeepSORT:(0.2484)
video 1/1 (10/750) C:\source\GitRepo\reid\VolovS_DeepSORT_FastReID\videos\MOT16\MOT16-13.mp4: 384x640 11 persons, Done. VLOvS:(0.662), DeepSORT:(0.2284)
video 1/1 (11/750) C:\source\GitRepo\reid\VolovS_DeepSORT_FastReID\videos\MOT16\MOT16-13.mp4: 384x640 11 persons, Done. VLOvS:(0.6094), DeepSORT:(0.2084)
```



```

Terminal - Command Prompt (ID)
video 1/1 (734/750) C:\source\GitRepo\reid\Volov5_DeepSORT_FastReID\videos\MOT16-13.mp4: 384x640 7 persons, Done. VLODs: (0.0626), DeepSORT: (0.1116)
video 1/1 (735/750) C:\source\GitRepo\reid\Volov5_DeepSORT_FastReID\videos\MOT16-13.mp4: 384x640 6 persons, Done. VLODs: (0.0478), DeepSORT: (0.1566)
video 1/1 (736/750) C:\source\GitRepo\reid\Volov5_DeepSORT_FastReID\videos\MOT16-13.mp4: 384x640 4 persons, Done. VLODs: (0.0668), DeepSORT: (0.0946)
video 1/1 (737/750) C:\source\GitRepo\reid\Volov5_DeepSORT_FastReID\videos\MOT16-13.mp4: 384x640 4 persons, Done. VLODs: (0.0478), DeepSORT: (0.0896)
video 1/1 (738/750) C:\source\GitRepo\reid\Volov5_DeepSORT_FastReID\videos\MOT16-13.mp4: 384x640 6 persons, Done. VLODs: (0.0626), DeepSORT: (0.1412)
video 1/1 (739/750) C:\source\GitRepo\reid\Volov5_DeepSORT_FastReID\videos\MOT16-13.mp4: 384x640 6 persons, Done. VLODs: (0.0626), DeepSORT: (0.1256)
video 1/1 (740/750) C:\source\GitRepo\reid\Volov5_DeepSORT_FastReID\videos\MOT16-13.mp4: 384x640 4 persons, Done. VLODs: (0.0488), DeepSORT: (0.0896)
video 1/1 (741/750) C:\source\GitRepo\reid\Volov5_DeepSORT_FastReID\videos\MOT16-13.mp4: 384x640 3 persons, Done. VLODs: (0.0578), DeepSORT: (0.0646)
video 1/1 (742/750) C:\source\GitRepo\reid\Volov5_DeepSORT_FastReID\videos\MOT16-13.mp4: 384x640 3 persons, Done. VLODs: (0.0518), DeepSORT: (0.0606)
video 1/1 (743/750) C:\source\GitRepo\reid\Volov5_DeepSORT_FastReID\videos\MOT16-13.mp4: 384x640 4 persons, Done. VLODs: (0.0508), DeepSORT: (0.0946)
video 1/1 (744/750) C:\source\GitRepo\reid\Volov5_DeepSORT_FastReID\videos\MOT16-13.mp4: 384x640 4 persons, Done. VLODs: (0.0618), DeepSORT: (0.1066)
video 1/1 (750/750) C:\source\GitRepo\reid\Volov5_DeepSORT_FastReID\videos\MOT16-13.mp4: 384x640 4 persons, Done. VLODs: (0.0578), DeepSORT: (0.1216)
Speed: 0.5ms pre-process, 56.0ms inference, 2.0ms NMS, 122.7ms strong sort update per image at shape (1, 3, 640, 640)
Results saved to run\track\exp15
1 tracks saved to run\track\exp15\tracks

```

Получите файл данных отслеживания MOT16-13.txt, совместимый с форматом MOT16. Файл MOT16-13.txt:

```

3 7 726 247 12 40 -1 -1 -1 0
3 8 269 276 16 45 -1 -1 -1 0
3 9 744 256 13 37 -1 -1 -1 0
4 1 1 341 13 93 -1 -1 -1 0
4 2 773 279 22 57 -1 -1 -1 0

```

Формат MOT16-13.txt немного отличается от gt.txt, заявляя, что он «совместим» с форматом MOT16. Формат «совместимость» имеет всего 10 значений, а его параметры разделены пробелами. Первые 6 параметров, участвующих в расчете MOT16, такие же, как и gt.txt, а последние 4 параметра в расчете индикатора не участвуют, все они равны -1. Формат следующий:

<frame>, <id>, <bb_left>, <bb_top>, <bb_width>, <bb_height>, <conf>, <x>, <y>, <z>

3. Запускаем программу оценки и получаем результат:

| | | | | | | | | | | | | | | | |
|---|--------|---------|--------|--------|--------|--------|--------|--------|--------|---------|---------|-------------|------|----|----|
| All sequences for deepsort finished in 0.91 seconds | | | | | | | | | | | | | | | |
| HOTA: deepsort-pedestrian | HOTA | DetA | AssA | DetRe | DetPr | AssRe | AssPr | LocA | OWTA | HOTA(0) | LocA(0) | HOTALocA(0) | | | |
| MOT16-13 | 57.313 | 59.474 | 56.029 | 66.611 | 73.942 | 63.281 | 75.428 | 81.584 | 66.957 | 74.316 | 76.452 | 56.816 | | | |
| COMBINED | 57.313 | 59.474 | 56.029 | 66.611 | 73.942 | 63.281 | 75.428 | 81.584 | 66.957 | 74.316 | 76.452 | 56.816 | | | |
| CLEAR: deepsort-pedestrian | HOTA | MOTP | MODA | CLR_Re | CLR_Pr | MTR | PTR | MLR | sMODA | CLR_TP | CLR_FN | CLR_FP | IDSW | MT | PT |
| MOT16-13 | 72.154 | 78.316 | 72.338 | 81.212 | 90.149 | 50 | 31.25 | 18.75 | 54.544 | 2654 | 614 | 290 | 6 | 8 | 5 |
| COMBINED | 72.154 | 78.316 | 72.338 | 81.212 | 90.149 | 50 | 31.25 | 18.75 | 54.544 | 2654 | 614 | 290 | 6 | 8 | 5 |
| Count: deepsort-pedestrian | Dets | GT Dets | Ids | GT Ids | | | | | | | | | | | |
| MOT16-13 | 2944 | 3268 | 18 | 16 | | | | | | | | | | | |
| COMBINED | 2944 | 3268 | 18 | 16 | | | | | | | | | | | |

4. Выполните ту же операцию для MOT16-02, MOT16-04, MOT16-05, MOT16-09, MOT16-10, MOT16-11 соответственно и просмотрите весь набор данных MOT16. Получите окончательный результат на этом наборе данных MOT16.

Результаты эксперимента

Сравните производительность исходного алгоритма DeepSORT, см. Таблицу ниже:

| Algorithm | MOTA ↑ | MOTP ↑ | MT / % ↑ | ML / % ↓ | Ids / % ↓ |
|-------------------|--------|--------|----------|----------|-----------|
| Original DeepSORT | 61.4 | 79.1 | 32.8 | 18.2 | 781 |

| | | | | | |
|-------------|-------------|-------------|-------------|-------------|------------|
| Ours | 66.2 | 80.8 | 35.3 | 17.6 | 760 |
|-------------|-------------|-------------|-------------|-------------|------------|

Видно, что после улучшения алгоритма DeepSORT все показатели немного улучшаются. Точность сопровождения (MOTA) увеличилась на 5,2%, точность сопровождения (MOTP) увеличилась на 1,7%, количество правильно сопровождаемых траекторий целей более 80% (MT) увеличилось на 3,5%, а количество правильно сопровождаемых траекторий целей ниже 20%. (ML) сократилось на 0,6%, а общее количество целевых переключателей идентификации (ID) сократилось в 21 раз. Видно, что улучшенный DeepSORT может уменьшить количество пешеходных переключений.

Эксперимент 2: сравнение улучшенного алгоритма DeepSORT с несколькими основными алгоритмами.

Результаты эксперимента 2 представлены в таблице.

| Algorithm | MOTA ↑ | MOTP ↑ | MT / % ↑ | ML / % ↓ | Ids / % ↓ | FPS / Hz ↑ |
|-------------------|-------------|-------------|-------------|-------------|------------|------------|
| SORT | 59.8 | 79.6 | 25.4 | 22.7 | 1423 | 8.6 |
| Original DeepSORT | 61.4 | 79.1 | 32.8 | 18.2 | 781 | 6.4 |
| JDE | 64.4 | - | 35.4 | 20 | 1544 | 18.5 |
| Ours | 66.2 | 80.8 | 35.3 | 17.6 | 760 | 5.8 |

Мы усовершенствовали алгоритм повторной идентификации пешеходов (FastReID) на основе DeepSORT, благодаря чему была повышена точность, точность и возможность сохранения идентификатора пешехода модели. По сравнению с другими моделями наша модель имеет комплексные преимущества: точность сопровождения (MOTA) и точность сопровождения (MOTP) самые высокие, а количество правильно сопровождаемых траекторий целей более чем на 80% (MT) практически равно JDE и делят первое место. Количество правильно отслеженных траекторий целей ниже 20% (ML) и общее количество переключений идентификации целей (ID) являются самыми низкими.

Кроме того, замечено, что производительность в реальном времени немного хуже, чем у алгоритма JDE, потому что JDE является одноэтапным алгоритмом отслеживания, а отслеживание в реальном времени относительно высоко, но переключение идентификаторов относительно частые, что обусловлено взаимным перекрытием мишеней. В розничной среде часто бывают сцены с большим количеством плотных пешеходов, поэтому наш алгоритм является подходящим выбором в это время.

ГЛАВА 3 Система оценки трафика магазина основа на нейтронных сетей

Цели разработки системы

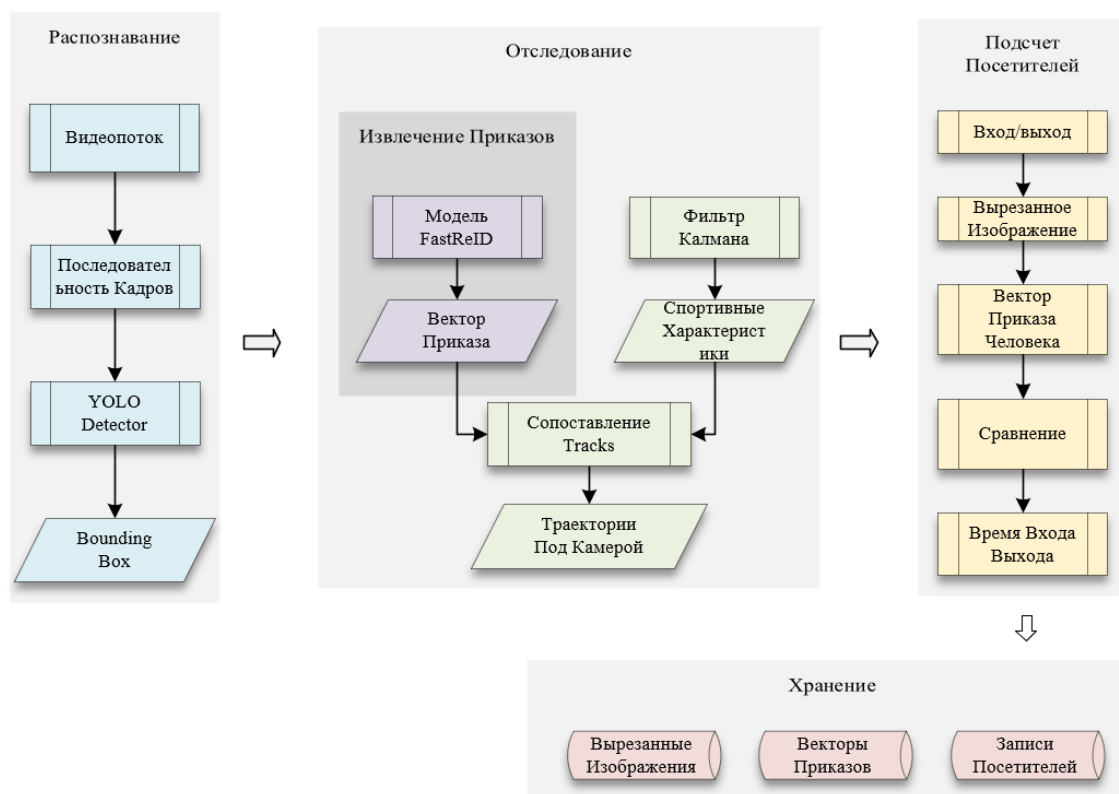
Подсчет людей на основе видеонаблюдения является основой для сбора бизнес-информации и выполнения задач интеллектуального управления, а также важным функциональным компонентом современного интеллектуального видеонаблюдения. В этом проекте будут использоваться передовые технологии компьютерного зрения и глубокого обучения, такие как yolov5, DeepSORT и FastReID, для реализации автоматической идентификации, отслеживания и анализа покупателей в магазине. В частности, система будет получать видеопоток в магазине через камеру, использовать yolov5 для автоматического обнаружения и идентификации посетителей, появляющихся на видео, затем использовать DeepSORT для отслеживания цели покупателя и, наконец, использовать FastReID для идентификации и анализа поведение покупателя, чтобы реализовать автоматическую статистику и анализ посещаемости магазина.

Благодаря исследованиям и практике этого проекта он может помочь продавцам более точно понять привычки и предпочтения поведения посетителей, сформулировать более научные и эффективные стратегии продаж, а также повысить прибыльность и рыночную конкурентоспособность магазинов. В то же время система также может предоставлять продавцам функции мониторинга и прогнозирования потоков посетителей в режиме реального времени, помогая продавцам еще больше оптимизировать управление операциями, повышать уровень обслуживания и удобство работы пользователей.

Описание системы

Общая архитектура системы

На основе FastReID и DeepSORT в этой статье разрабатывается система статистики потока покупателей, ориентированная на магазины. Система состоит из четырех модулей: модуль обнаружения пешеходов, модуль слежения за пешеходами, модуль статистики потоков людей и модуль хранения. Общая структура системы показана на рисунке.



Модуль обнаружения пешеходов

На вход модуля обнаружения пешеходов подается видеопоток с одной камеры, в процессе реализации в данной работе используется алгоритм обнаружения целей YOLO. В частности, в части обнаружения пешеходов алгоритм обнаружения целей YOLO используется для обнаружения пешеходов в каждом кадре видео, и пешеходы помечаются прямоугольным кадром из статического изображения каждого кадра, а конкретное положение пешехода в возвращается картинка, определение которой приведено в формуле:

$$b = [x, y, w, h]$$

Где (x, y) — пиксельные координаты верхнего левого угла кадра обнаружения, а (w, h) — ширина и высота кадра обнаружения.

Модуль извлечения признаков

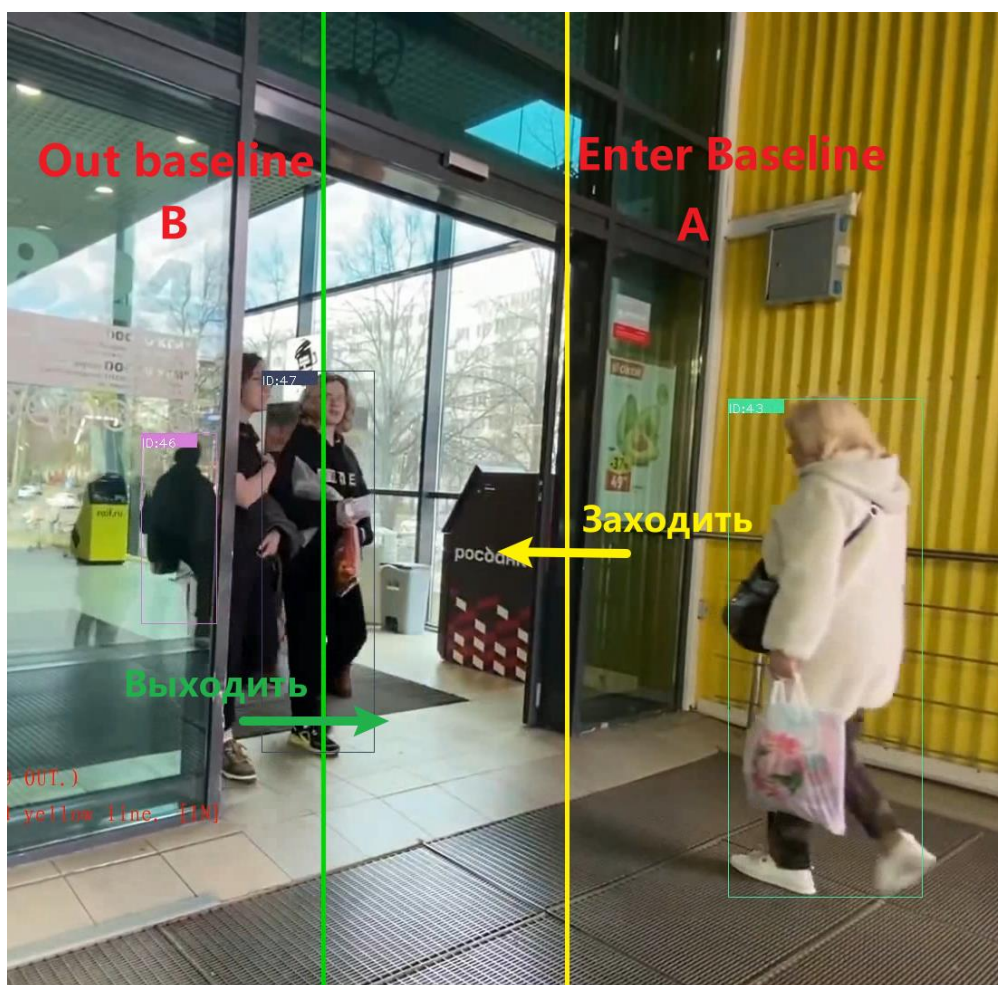
Модуль извлечения признаков предназначен для извлечения признаков внешнего вида пешеходов. Здесь мы будем использовать алгоритм FastReID и алгоритм фильтра Калмана. Входными данными модуля извлечения признаков является кадр обнаружения пешеходов, выдаваемый модулем обнаружения пешеходов. После модели FastReID и фильтра Калмана получаются признаки внешнего вида и признаки информации о движении пешеходов. Информация об этих двух функциях будет служить ориентиром

для последующего модуля отслеживания пешеходов, который будет соответствовать траектории и кадру обнаружения.

Модуль отслеживания пешеходами

Модуль отслеживания пешеходов использует модель DeepSORT, и его входными данными является кадр обнаружения, предоставленный модулем обнаружения пешеходов, и функции пешеходов, извлеченные модулем извлечения функций. Цель состоит в том, чтобы связать кадр обнаружения пешеходов в текущем кадре с существующим движением. Траектория в предыдущем кадре. После получения кадра обнаружения, предоставленного модулем обнаружения пешеходов, сначала используйте модуль извлечения признаков и модуль извлечения информации о движении, чтобы извлечь признаки внешнего вида и признаки движения текущего кадра обнаружения кадра и существующей дорожки движения, а затем вычислить матрицу расстояний между ними. , и, наконец, использовать венгерский алгоритм. Алгоритм связывает кадры-кандидаты обнаружения в текущем кадре с существующими траекториями движения в предыдущем кадре, чтобы построить траектории всех пешеходов в видео.

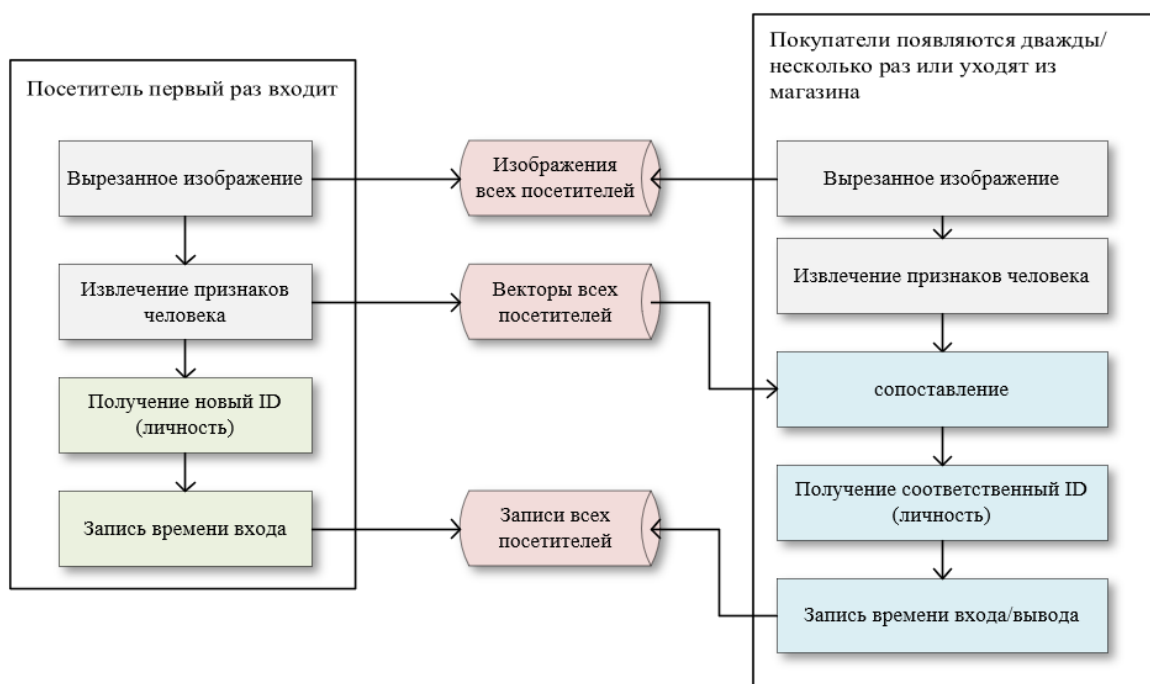
Модуль подсчета посетителей



Вход модуля подсчета посетителей — это траектория движения всех людей под одной камерой, и его цель — точно зафиксировать количество людей, входящих и выходящих из магазина. После рассмотрения реалистичных требований к задаче в этом исследовании используется «метод двух линий» для завершения подсчета людей. Конкретный принцип **двухлинейного метода** заключается в следующем: в магазине устанавливается стационарная камера и получается видеопоток наблюдения, причем в видеопотоке отмечаются две счетные линии, которые записываются как А и В соответственно. На основе траектории движения тела на пересечении траектории и линии указывается конкретное количество проходящих людей.

В качестве примера предположим, что магазин находится слева от линии В, как показано в рисунке. Видно, что траектория входа посетителя будет сначала соприкасаться с линией А, а затем с контактом с В, это называется **поведением входа**. И наоборот, траектория ухода пешехода должна сначала коснуться линии В, а затем линии А, это называется **поведением выхода**. Нам нужно только записать, касается ли траектория пешехода сначала линии А или линии В, а затем мы уже можем определить, входит ли пешеход в магазин или выходит из него.

Когда происходит вход или выход, модули обнаружения пешеходов, отслеживания пешеходов и извлечения функций будут работать последовательно, чтобы получить вырезанное изображение и вектор приказов посетителя, а также записать времени в журнал. А когда посетитель уходит под другой камерой, мы сравниваем его характеристики с набором характеристик в базе данных, чтобы получить информацию о личности посетителя, и на этой основе фиксируем время ухода. Таким образом, мы можем точно фиксировать время входа и выхода всех покупателей в супермаркете.



Модуль хранения

Модуль хранения отвечает за регистрацию времени, когда покупатели заходят в магазин, появляются под камерой и выходят из магазина.

Каждый раз, когда кто-то входит в зону подсчета, записывайте время появления человека и в то же время перехватывайте изображение соответствующего человека и извлекайте вектор признаков на основе изображения. Если камера является камерой в магазине или на выходе, личность пешехода также учитывается в соответствии с моделью повторной идентификации. Наконец, личность человека, в каком объективе он появляется, время появления и изображение человека сохраняются на диске для последующего статистического анализа.

Среда разработки

Аппаратные среды, использованные для разработки, показаны в Таблице 1:

Таблица 1

| Предмет | Технические характеристики | Комментарий |
|---------|---|-------------|
| Чип | AMD Ryzen 5 5600H with Radeon Graphics 3.30 ghz | - |

| | | |
|----------------------|-------------------------|-------------------------|
| Памяти | 16.0 GB | ПО использует до 2.1 GB |
| Камер | 1080P | 720P минимум |
| Операционная системы | Windows 10 Professional | - |

Программные среды, использованные в разработке, показаны в Таблице 2:

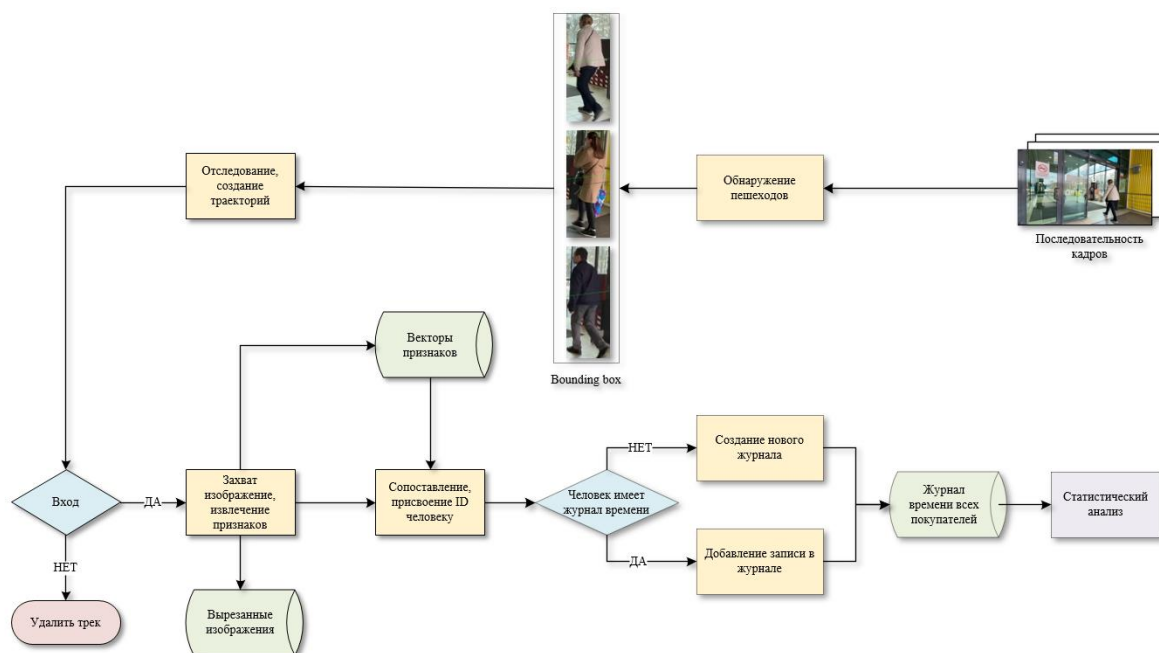
Таблица 2

| Предмет | Версия | Комментарий |
|---------|--------------------------------|-----------------|
| Python | 3.8.16 | Данная или выше |
| Pycharm | 2022.1.2 (Educational Edition) | Данная или выше |
| Opencv | 4.5.3 | Данная или выше |
| Pytorch | 1.13.0 | Данная или выше |

Демонстрация и тестирование системы

Алгоритм статистики трафика

Общий процесс статистики пассажиропотока показан на рисунке ниже:



Тестовая среда



Для проверки работоспособности различных функций в этой системе тест проведен в крупном супермаркете ОКЕЙ. Мы развернули 3 фиксированные камеры по траектории пешеходов, входящих в супермаркет, для получения видео. План этажа супермаркета и расположение 3-х камер показаны на рисунке выше. Продолжительность видеозаписи около 2 минут.

Междукамерное отслеживание пешеходов

Отслеживание людей под одной камерой

Следующие три изображения выбираются соответственно из изображений 300-го, 330-го, 360-го и 390-го кадра под камерой 3, а временной интервал выделения составляет около 1,5 секунды. На рисунке ниже в основном показано отслеживание человека 881.

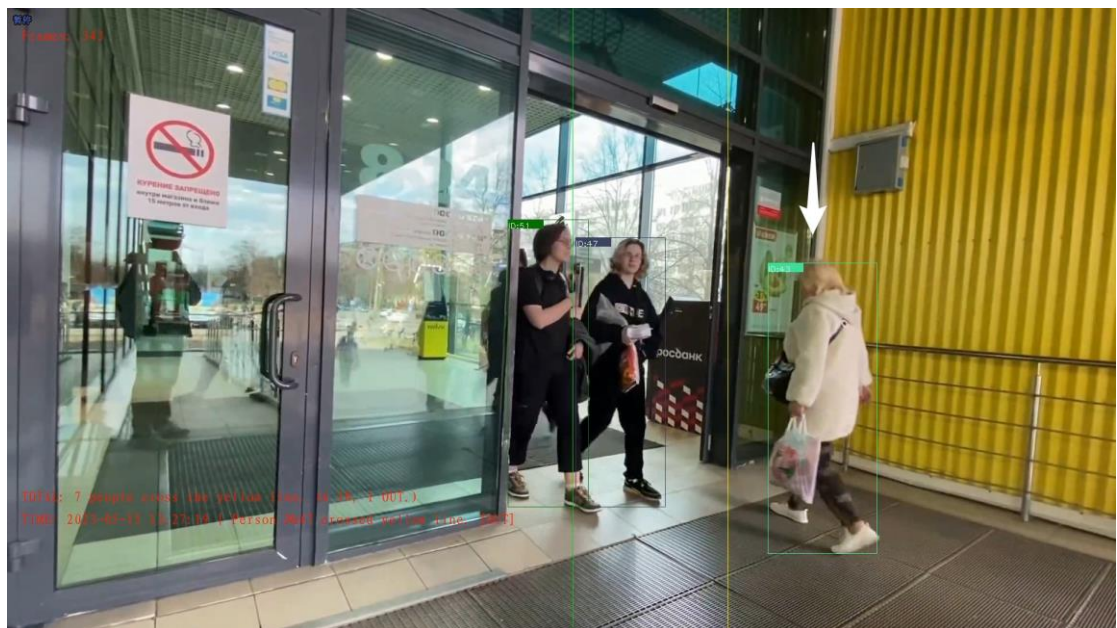


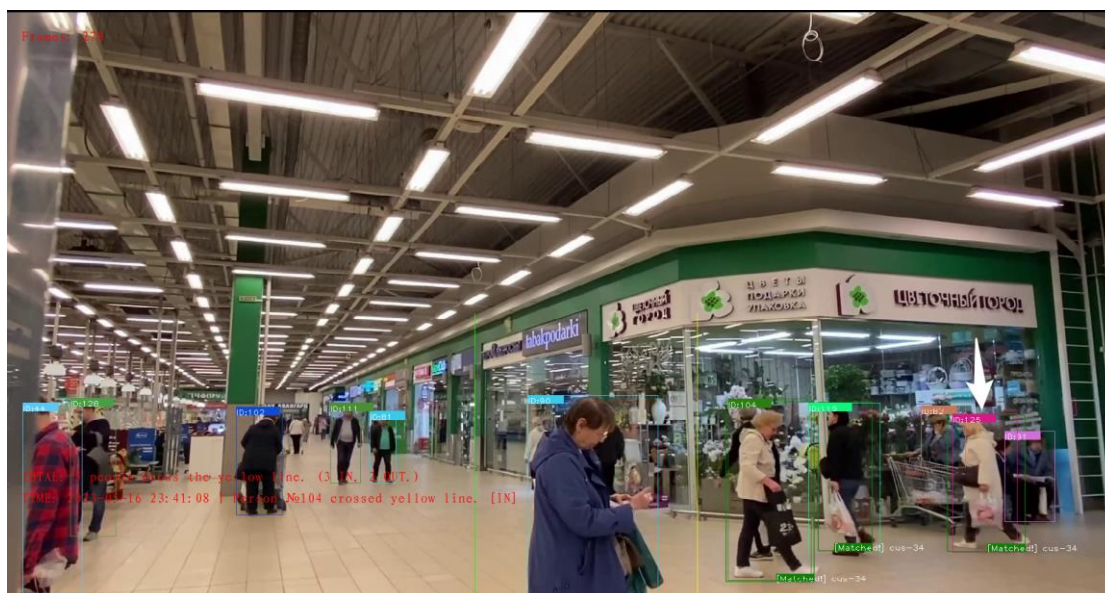


Видно, что детекция человека у № 881 хорошая в пределах 90 кадров, а идентификатор всегда сохраняется без переключения. Даже если окклюзия возникает в третьем кадре, идентификатор пешехода остается равным 881 в четвертом кадре, когда окклюзия исчезает в четвертом кадре. Это роль ассоциации признаков внешнего вида, предоставляемая FastReID. Кроме того, другие пешехода, отличные от 881, такие как 803, 819, 869, также хорошо отслеживаются.

Результаты испытаний доказывают, что система может непрерывно отслеживать нескольких пешеходов под одной камерой.

Отслеживание людей под несколькими камерами





На трех вышеприведенных снимках показаны результаты отслеживания пешехода *cus-43*, проходящего через три камеры в хронологическом порядке. Из результатов теста видно, что при первом появлении пешехода в зоне наблюдения камеры 1 система выдает ей идентификатор: *cus-43*, который является глобально уникальным. Когда пешеход *cus-43* исчез из поля зрения камеры 1 и вошел в камеру 2, система извлекла черты его внешности через FastReID для сравнения, узнала идентификационную информацию пешехода *cus-43* и восстановила отслеживание и повторно идентифицировала личность. Отображается информация в правом нижнем углу окна обнаружения. Точно так же, когда она вышла из камеры 2 и вошла в камеру 3, система также успешно идентифицировала ее в зоне наблюдения камеры 3.

На кадрах существуют желтая и зеленая виртуальные линии, они не реалины. Эти линии используются для подсчета качества людей.

Когда пешеходы проходят через линии, в левом нижнем углу экрана печатаются подсказки времени прохождения, личности входящего персонала и общего количества входящих и выходящих.



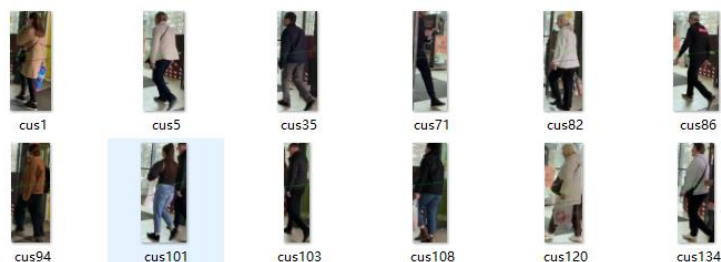
Запись времени входа и выхода покупателей

| PERSON-ID | CAMERA-ID | TIME | CAMERA-ID | TIME | CAMERA-ID | TIME |
|-----------|-----------|-----------------------|-----------|-----------------------|-----------|-----------------------|
| cus-1 | cam-1 | 2023-05-10 03: 31: 27 | cam-2 | 2023-05-10 03: 37: 59 | cam-3 | 2023-05-10 03: 46: 04 |
| cus-5 | cam-1 | 2023-05-10 03: 31: 37 | cam-2 | 2023-05-10 03: 50: 03 | | |
| cus-35 | cam-1 | 2023-05-10 03: 32: 03 | cam-2 | 2023-05-10 03: 38: 10 | cam-3 | 2023-05-10 03: 46: 19 |
| cus-71 | cam-1 | 2023-05-10 03: 32: 11 | cam-2 | 2023-05-10 03: 38: 15 | cam-3 | 2023-05-10 03: 46: 23 |
| cus-82 | cam-1 | 2023-05-10 03: 32: 37 | cam-2 | 2023-05-10 03: 38: 26 | cam-3 | 2023-05-10 03: 46: 35 |
| cus-94 | cam-1 | 2023-05-10 03: 32: 48 | cam-2 | 2023-05-10 03: 38: 32 | cam-3 | 2023-05-10 03: 46: 42 |
| cus-101 | cam-1 | 2023-05-10 03: 32: 53 | cam-2 | 2023-05-10 03: 41: 28 | | |
| cus-103 | cam-1 | 2023-05-10 03: 32: 53 | cam-2 | 2023-05-10 03: 41: 39 | cam-3 | 2023-05-10 03: 50: 03 |
| cus-108 | cam-1 | 2023-05-10 03: 33: 00 | cam-2 | 2023-05-10 03: 41: 02 | cam-3 | 2023-05-10 03: 49: 20 |
| cus-120 | cam-1 | 2023-05-10 03: 33: 10 | cam-2 | 2023-05-10 03: 42: 06 | cam-3 | 2023-05-10 03: 50: 23 |
| cus-134 | cam-1 | 2023-05-10 03: 33: 14 | cam-2 | 2023-05-10 03: 42: 30 | cam-3 | |
| cus-147 | cam-1 | 2023-05-10 03: 33: 21 | cam-2 | 2023-05-10 03: 42: 57 | cam-3 | 2023-05-10 03: 51: 18 |
| cus-203 | cam-1 | 2023-05-10 03: 34: 15 | cam-2 | 2023-05-10 03: 42: 59 | cam-3 | 2023-05-10 03: 51: 16 |
| cus-253 | cam-1 | 2023-05-10 03: 35: 21 | cam-2 | | cam-3 | |
| cus-248 | cam-1 | 2023-05-10 03: 35: 24 | cam-2 | | cam-3 | |
| cus-257 | cam-1 | 2023-05-10 03: 35: 28 | cam-2 | | cam-3 | |
| cus-306 | cam-1 | 2023-05-10 03: 36: 17 | cam-2 | | cam-3 | |
| cus-314 | cam-1 | 2023-05-10 03: 36: 30 | cam-2 | | cam-3 | |
| cus-331 | cam-1 | 2023-05-10 03: 36: 53 | cam-2 | | cam-3 | |
| cus-335 | cam-1 | 2023-05-10 03: 36: 56 | cam-2 | | cam-3 | |

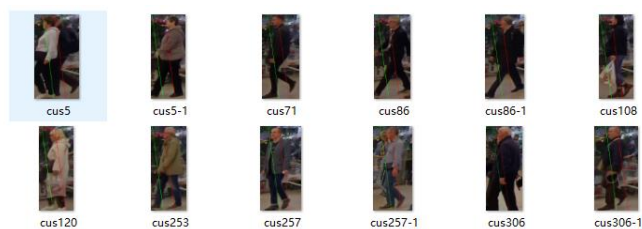
Мы используем структуру данных в Python для записи входных и выходных записей всех посетителей за определенный период времени. Распечатаем его, как показано выше.

Вырезанные изображения покупателей

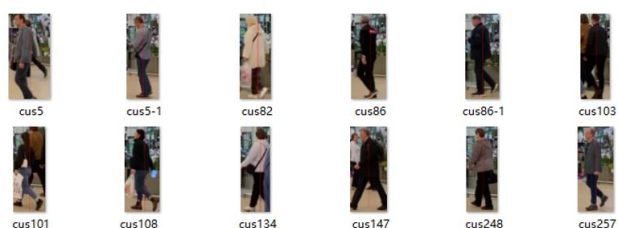
Вырезанные изображения покупателей из 1-ой камерой (частично):



Вырезанные изображения покупателей из 2-ой камерой (частично):



Вырезанные изображения покупателей из 3-ой камерой (частично):



Метрики оценки системы

Для оценки эффекта статистического метода в данной работе используются 3 камеры для выборки видео крупного супермаркета ОКЕЙ. Каждое из этих видео длится примерно 2 минуты. Результаты статистического теста представлены в таблице ниже:

Модуль подсчета входов и выходов пешеходов

| Камера | Ground Truth | | Count | | Точность |
|--------|--------------|------------------------|-------------|------------------------|---------------|
| | Направление | Количество посетителей | Направление | Количество посетителей | |
| Cam1 | Вход | 30 | Вход | 30 | 100.00% |
| | Выход | 16 | Выход | 15 | 93.75% |
| | Всего | 46 | Всего | 45 | 97.83% |
| Cam2 | Вход | 30 | Вход | 28 | 93.33% |
| | Выход | 23 | Выход | 20 | 86.96% |
| | Всего | 53 | Всего | 48 | 90.57% |
| Cam3 | Вход | 28 | Вход | 27 | 96.43% |
| | Выход | 25 | Выход | 25 | 100.00% |
| | Всего | 53 | Всего | 52 | 98.11% |
| Всего | | | | Вход | 96.59% |
| | | | | Выход | 93.57% |
| | | | | Всего | 95.50% |

Эти данные получают путем подсчета потока людей двухстрочным методом, включающим реальные и статистические значения трех камер. Видно, что каждая камера имеет разную точность в направлении входа и

выхода, но в целом средняя точность выше, 93,33% (вход), 93,57% (выход) и 93,59% (вход и выход всего).

При детальном анализе положения каждой камеры можно обнаружить, что точность обнаружения камеры 2 в направлении выхода низкая, всего 86,96%. Это связано с настройками камеры, окружающей средой и другими факторами. Потому что в in2 область для обнаружения уходящих людей меньше. Более того, в течение некоторого времени большое количество людей внезапно вошло и вышло из магазина одновременно (10 человек в течение 5 секунд), поэтому между персоналом произошло серьезное совпадение, и было 3 пропущенных инспекционных персонала, которые понизил среднюю точность.

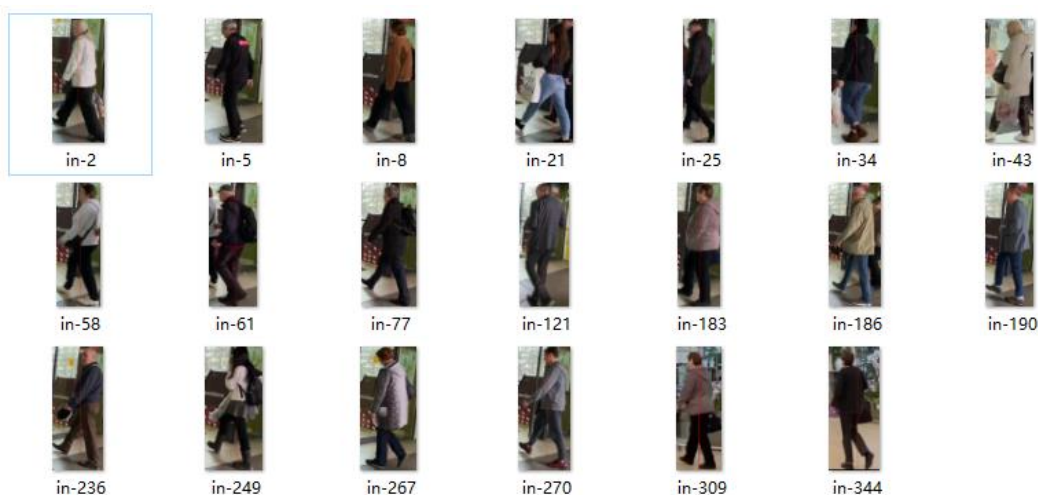
Точность кулачка 1 и кулачка 3 относительно высока, выше 97%. Это связано с тем, что в этих двух сценариях почти нет больших толп людей, входящих и выходящих из магазина.

В целом, несмотря на то, что еще есть возможности для повышения точности, общая производительность относительно стабильна и точна, что может помочь руководителям объектов лучше понимать поток людей.

Модуль межкамерной Ре-идентификации пешеходов

Всего в этом раунде мы оценили 20 человек. Эти 20 человек идут от камеры 1 к камере 2 через камеру 2, то есть имеют полную траекторию, пересекающую камеру.

В sam1 20 человек, формируют 20 скриншотов посетителей , на данный момент в пуле фич 20 векторов внешности.



В sam2 по очереди прошло 20 человек, а скриншотов сформировалось всего 19, потому что было два пешехода, идущих параллельно и один из них

покрыт, поэтому два кадра детектирования не могли быть правильно сформированы. Таким образом, происходит пропущенное обнаружение. Среди 19 изображений, подлежащих повторной идентификации, 18 человек получили правильную идентичность после повторной идентификации, а 1 человек не смог получить идентичность (внешнее сходство с исходной идентичностью было слишком низким, поэтому образец был отфильтрован как отрицательный). Следовательно, точность повторной идентификации по кулачку-2 составляет $18/20=90,0\%$.



Figure 2 Два пешехода, идущих параллельно и один из них не обнаружен



Figure 3 Человек не смог получить идентичность



Figure 4 Его правильная личность должна быть

В камере-3 последовательно прошли 20 человек, из которых сформировалось только 19 скриншотов, потому что рядом шли два пешехода, блокируя друг друга, и два кадра детектирования не могли быть корректно сформированы. Таким образом, происходит пропущенное обнаружение. Среди 19 изображений, подлежащих повторной идентификации, 19 получили правильную идентичность после повторной идентификации, а 0 получили неправильную идентичность. Следовательно, точность повторной идентификации по кулачку-3 составляет $19/20=95,0\%$.



Figure 5 Три человека, но один из них не обнаружен

Полный результат эксперимента показан следующей таблицей:

| ID чел | Cam-1 | Cam-2 | Cam-3 |
|--------|----------------------------|--|--|
| 2 | Успешно вошли в статистику | Успешно повторно идентифицирован и получил правильную личность | Успешно повторно идентифицирован и получил правильную личность |
| 5 | Успешно вошли в статистику | Успешно повторно идентифицирован и получил правильную личность | Успешно повторно идентифицирован и получил правильную личность |
| 8 | Успешно вошли в статистику | Успешно повторно идентифицирован и получил правильную личность | Скрыт и не может быть повторно идентифицирован |
| 21 | Успешно вошли в статистику | Успешно повторно идентифицирован и получил правильную личность | Успешно повторно идентифицирован и получил правильную личность |
| 25 | Успешно вошли в статистику | Успешно повторно идентифицирован и получил правильную личность | Успешно повторно идентифицирован и получил правильную личность |

| | | | |
|-----|----------------------------|--|--|
| 250 | Успешно вошли в статистику | Успешно повторно идентифицирован и получил правильную личность | Успешно повторно идентифицирован и получил правильную личность |
| 267 | Успешно вошли в статистику | Успешно повторно идентифицирован и получил правильную личность | Успешно повторно идентифицирован и получил правильную личность |
| 270 | Успешно вошли в статистику | Успешно повторно идентифицирован и получил правильную личность | Успешно повторно идентифицирован и получил правильную личность |





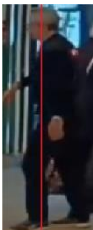

И записали времени, когда они входят в магазин (виртуальное время в экспериментальной среде):

```

1  {'customer-2': {'cam-1': '2023-05-15 13: 26: 36', 'cam-2': '2023-05-15 13: 33: 20', 'cam-3': '2023-05-15 13: 40: 24'
2  }, 'customer-5': {'cam-1': '2023-05-15 13: 26: 45', 'cam-2': '2023-05-15 13: 33: 43', 'cam-3': '2023-05-15 13: 42: 01'
3  }, 'customer-8': {'cam-1': '2023-05-15 13: 26: 55', 'cam-2': '2023-05-15 13: 33: 31', 'cam-3': '2023-05-15 13: 45: 21'
4  }, 'customer-21': {'cam-1': '2023-05-15 13: 26: 59', 'cam-2': '2023-05-15 13: 36: 31', 'cam-3': '2023-05-15 13: 45: 21'
5  }, 'customer-25': {'cam-1': '2023-05-15 13: 27: 00', 'cam-2': '2023-05-15 13: 33: 37', 'cam-3': '2023-05-15 13: 40: 38'
6  }, 'customer-34': {'cam-1': '2023-05-15 13: 27: 11', 'cam-2': '2023-05-15 13: 33: 47', 'cam-3': '2023-05-15 13: 40: 55'
7  }, 'customer-43': {'cam-1': '2023-05-15 13: 27: 25', 'cam-2': '2023-05-15 13: 33: 56', 'cam-3': '2023-05-15 13: 41: 01'
8  }, 'customer-58': {'cam-1': '2023-05-15 13: 27: 28', 'cam-2': '2023-05-15 13: 34: 10', 'cam-3': '2023-05-15 13: 44: 10'
9  }, 'customer-61': {'cam-1': '2023-05-15 13: 27: 30', 'cam-2': '2023-05-15 13: 34: 10', 'cam-3': '2023-05-15 13: 41: 49'
10 }, 'customer-77': {'cam-1': '2023-05-15 13: 27: 39', 'cam-2': '2023-05-15 13: 39: 09', 'cam-3': '2023-05-15 13: 41: 28'
11 }, 'customer-123': {'cam-1': '2023-05-15 13: 28: 39', 'cam-2': '2023-05-15 13: 39: 09', 'cam-3': '2023-05-15 13: 42: 44'
12 }, 'customer-183': {'cam-1': '2023-05-15 13: 30: 16', 'cam-2': '2023-05-15 13: 37: 35', 'cam-3': '2023-05-15 13: 45: 21'
13 }, 'customer-186': {'cam-1': '2023-05-15 13: 30: 19', 'cam-2': '2023-05-15 13: 37: 51', 'cam-3': '2023-05-15 13: 44: 04'
14 }, 'customer-188': {'cam-1': '2023-05-15 13: 30: 40', 'cam-2': '2023-05-15 13: 38: 04', 'cam-3': '2023-05-15 13: 43: 25'
15 }, 'customer-190': {'cam-1': '2023-05-15 13: 30: 26', 'cam-2': '2023-05-15 13: 37: 00', 'cam-3': '2023-05-15 13: 43: 09'
16 }, 'customer-236': {'cam-1': '2023-05-15 13: 31: 37', 'cam-2': '2023-05-15 13: 38: 30', 'cam-3': '2023-05-15 13: 44: 41'
17 }, 'customer-249': {'cam-1': '2023-05-15 13: 31: 56', 'cam-2': '2023-05-15 13: 39: 02', 'cam-3': '2023-05-15 13: 47: 47'
18 }, 'customer-267': {'cam-1': '2023-05-15 13: 32: 27', 'cam-2': '2023-05-15 13: 45: 59', 'cam-3': '2023-05-15 13: 48: 49'
19 }, 'customer-270': {'cam-1': '2023-05-15 13: 32: 32', 'cam-2': '2023-05-15 13: 45: 57', 'cam-3': '2023-05-15 13: 53: 03'
20 }
21

```

Всего было успешно повторно идентифицировано 17 человек из 20(частичная таблица):

| ID | Cam-1 | Cam-2 | Cam-3 |
|----|---|---|---|
| 2 |  |  |  |
| 5 |  |  |  |

| | | | |
|-----|---|---|---|
| 21 |  |  |  |
| 25 |  |  |  |
| 270 |  |  |  |

Экспериментальное явление: наблюдалось вышеописанное явление: 20 пешеходов прошли через камеру-1, камеру-2 и камеру-3 за определенный промежуток времени, 17 из них были полностью отслежены и имели записи с 3-х камер, и 2 пешехода прошли мимо. Другие Пешеходы закрыты и не могут быть обнаружены, а один человек не может получить первоначальную личность после повторной идентификации, поэтому у них есть только неполные записи времени под двумя камерами, которые не включены в полное отслеживание.

Экспериментальный вывод: точность отслеживания нашей системы статистики трафика достигает примерно $17/20=85,0\%$.

ГЛАВА 4 Анализ полученных результатов

Экспериментальная оценка модели FastReID на открытой базе данных

После завершения обучения модели мы решили протестировать производительность нашей модели на другом наборе данных Market1501 и DukeMTMC и сравнили ее с другими моделями ReID, появившимися в тот же период:

| Methods | Market1501 | | DukeMTMC | |
|----------------------|-------------|-------------|-------------|-------------|
| | Rank-1 | mAP | Rank-1 | mAP |
| IANet(IVPR'19) | 94.4 | 83.1 | 87.1 | 73.4 |
| Auto-ReID(ICCV'19) | 94.5 | 85.1 | - | - |
| OSNet(ICCV'19) | 94.8 | 84.9 | 88.6 | 73.5 |
| ABDNet(ICCV'19) | 95.6 | 88.3 | 89.0 | 78.6 |
| Circle Loss(CVPR'20) | 96.1 | 87.4 | 89.0 | 79.6 |
| ours | 95.7 | 88.4 | 90.1 | 81.3 |

Согласно приведенным выше данным, мы видим, что модель FastReID хорошо работает как с наборами данных Market1501, так и с наборами данных DukeMTMC. Его точность Rank-1 в наборе данных Market1501 составляет 95,7%, mAP — 88,4%, а его точность Rank-1 в наборе данных DukeMTMC — 90,1%, mAP — 81,3%.

По сравнению с другими алгоритмами FastReID показал очень хорошую производительность. Например, в наборе данных Market1501 точность Rank-1 FastReID и mAP превзошли такие алгоритмы, как IANet, Auto-ReID и OSNet соответственно; даже в наборе данных DukeMTMC его точность Rank-1 и mAP превзошли все другие алгоритмы. Это показывает, что FastReID обладает высокой надежностью и способностью к обобщению в задачах повторной идентификации личности.

В заключение, основываясь на этих оценочных метриках, мы можем считать FastReID эффективной моделью повторной идентификации личности, которую предполагается применять в практических сценариях.

Экспериментальная оценка улучшенной модели DeepSORT на открытой базе данных

Чтобы проверить эффективность улучшенного метода многоцелевого отслеживания DeepSORT на основе FastReID, мы провели сравнительный эксперимент на более сложном новом наборе данных MOT20.

Этот набор данных аннотирует 8 сложных видеопоследовательностей (4 для обучения и 4 для тестирования) в неограниченных условиях, снятых с помощью статических и движущихся камер. По сравнению с предыдущей серией наборов данных MOT Challenge, MOT20 фокусируется на сценах с плотным скоплением людей, а его видео могут охватывать до 246 человек в одном кадре. Таким образом, отслеживание становится более сложным. В нем сравнивается способность модели слежения за несколькими целями отслеживать небольшие цели.



Мы протестировали алгоритм многоцелевого отслеживания на четырех подмножествах набора данных MOT20 и сравнили результаты с результатами 10 лучших победителей всемирного конкурса MOT Challenge (в 2020 году). Результаты следующие:

| Traker | MOTA ↑ | IDF1↑ | MT/% ↑ | ML/% ↓ | IDs ↓ | FPS ↑ |
|-------------|--------|-------|--------|--------|-------|-------|
| ByteTrack | 67 | 70.2 | 47.7 | 21.2 | 680 | 17.7 |
| OUTrack | 65.4 | 65.1 | 49.5 | 13.3 | 2,885 | 5.1 |
| UTM | 64.4 | 65.9 | 65 | 8.7 | 2,592 | 22.4 |
| RFTTracker | 62.4 | 53 | 48.7 | 15.5 | 3,804 | 0 |
| SUSHI | 61.6 | 71.6 | 47.6 | 19.2 | 1,053 | 5.3 |
| TransCenter | 61 | 49.8 | 48.4 | 15.5 | 4,493 | 1 |
| MPTC | 60.6 | 59.7 | 51.1 | 16.7 | 4,533 | 0.7 |

| | | | | | | |
|-------------|-------------|-------------|-------------|-------------|-------------|------------|
| ТМОН | 60.1 | 61.2 | 46.7 | 17.8 | 2,342 | 0.6 |
| OCSORT | 59.9 | 67 | 38.5 | 26.6 | 554 | 27.6 |
| MFI | 59.3 | 59.1 | 41.1 | 17.3 | 191 | 0.5 |
| ours | 60.4 | 60.9 | 35.2 | 17.4 | 2021 | 8.6 |

Во-первых, следует отметить, что в наборе данных MOT20 производительность трекера измеряется такими показателями, как MOTA, IDF1, MT и ML. Среди них MOTA является основной оценочной метрикой для отслеживания нескольких объектов, которая сочетает в себе точность и полноту обнаружения, сопоставления и отслеживания. IDF1 - показатель, уделяющий больше внимания точности следящего устройства. MT и ML соответственно представляют собой долю количества правильно сопровождаемых целей и долю количества пропущенных сопровождаемых целей в общем количестве обработанных кадров.

Из данных видно, что лучше всего показал себя ByteTrack, за ним следуют OUTrack и UTM, а MOTA этих трех трекеров превысил 64. Другие трекеры работали относительно плохо, но были и трекеры, которые преуспели по определенным показателям. Например, SUSHI и OCSORT показывают более высокие результаты по IDF1 и MOTA, но хуже по ML, в то время как ТМОН лучше по IDF1.

Наш трекер имеет следующие сильные и слабые стороны по сравнению с другими 10 мировыми победителями:

Преимущества:

1. Трекер хорошо работает по индикатору IDF1 с оценкой 60,4, что указывает на то, что он в основном может выполнять многоцелевые задачи отслеживания.

2. Трекер также относительно хорошо показывает FPS, достигая 8,6 кадров/с, что означает, что он может обрабатывать высокоскоростные динамические сцены.

Недостатки:

1. По индикатору MOTA трекер набрал всего 60,4 балла, что указывает на то, что его производительность в наборе данных MOT20 относительно слабая.

2. Индекс ML/% относительно высок, 17,4, что показывает, что он не имеет хорошего контроля над пропорцией количества пропущенных целей к общему количеству.

3. Производительность этого трекера по двум индикаторам идентификаторов не так хороша, как у некоторых других трекеров. Идентификаторы 2021, а MT/% составляет 35,2, что показывает, что у него есть некоторые проблемы, такие как ложное обнаружение, пропущенные обнаружение и ложное отслеживание.

Хотя наш трекер хорошо работает по показателям FPS и IDF1, и имеет определенную производительность в реальном времени, он относительно плохо показывает соотношение количества правильно сопровождаемых целей к общему количеству и отношение количества пропущенных сопровождаемых целей к общему числу. Другими словами, DeepSORT может пропустить обнаружение (то есть не правильно идентифицировать и отследить цель) при работе с плотными скоплениями людей. После анализа я думаю, что это может быть вызвано следующими причинами:

1. Проблема извлечения признаков: DeepSORT использует сверточную нейронную сеть для извлечения признаков, но когда толпа плотная, изменения пикселей вокруг каждого пешехода малы, и трудно выделить уникальные признаки пешеходов. Это может привести к тому, что несколько пешеходов будут использовать одни и те же функции, из-за чего DeepSORT не сможет их различить.

2. Проблема перекрытия пешеходов: в плотных толпах разные пешеходы могут перекрывать, перекрывать или частично перекрывать друг друга, что также затрудняет для DeepSORT правильное отслеживание каждого пешехода.

3. Проблема ассоциации идентичности: в плотной толпе из-за большого количества пешеходов и их близости сходство между пешеходами может стать очень высоким. Это может привести к тому, что DeepSORT ошибочно сопоставит двух разных пешеходов с одним и тем же идентификатором или ошибочно разделит одного и того же пешехода на двух или более пешеходов.

4. Проблема визуального шума. В плотном скоплении людей может быть много факторов помех, таких как свет, фон, положение камеры и т. д.

Эти шумовые факторы могут повлиять на производительность DeepSORT, что приведет к пропущенным обнаружениям.

Для решения этих проблем в будущем могут быть предприняты следующие шаги:

1. Проблема выделения признаков: в случае перекрывающихся целей можно ввести больше представлений признаков, чтобы различать разные цели. Например, в качестве дополнительных признаков может использоваться такая информация, как контур, цвет, текстура, скорость и т. д. цели. Кроме того, для

извлечения более репрезентативных признаков можно использовать более мощные модели сверточных нейронных сетей.

2. Проблема перекрытия пешеходов: технология сегментации объектов может использоваться для точного определения границ каждого объекта, что позволяет избежать путаницы с идентификаторами или пропущенного обнаружения.

3. Проблема отслеживания объектов: при наличии нескольких перекрывающихся объектов следует использовать более точный алгоритм отслеживания объектов, такой как CenterTrack. CenterTrack — это алгоритм отслеживания нескольких целей, основанный на обнаружении центральной точки, который может выполнять точное отслеживание нескольких целей в сценах с плотным скоплением людей. Он использует обнаружение центральной точки вместо традиционного обнаружения ограничительной рамки и использует легкую структуру сверточной нейронной сети для достижения эффективного отслеживания объекта. Он оптимизирован для плотных целевых сцен и может адаптироваться к взаимному перекрытию или частичному перекрытию целей для лучшего отслеживания объектов.

4. Проблема визуального шума: При возникновении проблем с визуальным шумом вы можете уменьшить помехи, отрегулировав положение камеры, изменив окружающее освещение и используя разницу фона.

ЗАКЛЮЧЕНИЕ

Работа данной статьи является исследованием применения технологий нейронной сети для оценки трафика магазина. Реализован алгоритм YOLOv5 для обнаружения пешеходов. Реализован алгоритм отслеживания DeepSORT для отслеживания пешеходов. Реализован двухстрочный метод для подсчета количества посетителей магазина. Хранения и записи изображений посетителей, характеристика и времени входа и выхода. В конце, разработала систему оценки трафика магазина на сцене супермаркета.

Основная работа данной статьи включает в себя следующие аспекты:

1. Повторная идентификация пешехода

Реализован алгоритм повторной идентификации пешехода FastReID, который используется для установления связи между идентификаторами пешеходов под несколькими камерами. Модель хорошо зарекомендовала себя на наборах данных Market1501 и DukeMTMC со средней точностью (mAP) 88,4% и 81,3%.

2. Отслеживание пешеходов

На основе алгоритмов YOLOv5, FastReID и DeepSORT разработан и реализован трекер пешеходов, который может выполнять многоцелевое отслеживание под одной и той же камерой, а также может идентифицировать личности между камерами. Результаты тестирования на MOT16: точность отслеживания (MOTA) 66,2% и точность отслеживания (MOTP) 80,8%.

3. Применение системы подсчета посетителей в супермаркете

В данной статье разработана система подсчета потоков людей на основе видео наблюдения, которая фиксирует количество посетителей, изображения посетителей и время, когда посетители появляются в каждой зоне супермаркета в течение определенного периода времени. Точность идентификации входящих и выходящих людей достигла 95,%, а точность повторной идентификации пешеходов достигла 85%.

4. Сравните с другими продвинутыми алгоритмами на других наборах данных и выясните преимущества и недостатки модели FastReID и модели DeepSORT.

Хотя система в этой статье имеет хорошую производительность в реальном времени, когда более 50 человек входят и выходят одновременно, на нее легко влияет взаимный охват персонала, и точность отслеживания будет снижена. В такой ситуации могут быть внесены дополнительные улучшения, такие как улучшение выделения признаков пешеходов, внедрение более точного алгоритма отслеживания объектов и т. д., чтобы повысить точности.

Список Литературы