МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ РОССИЙСКОЙ ФЕДЕРАЦИИ

федеральное государственное автономное образовательное учреждение

высшего образования

**«Северный (Арктический) федеральный университет имени М.В. Ломоносова»**

Высшая школа информационных технологий и автоматизированных систем

(наименование высшей школы / филиала / института / колледжа)

**КУРСОВОЙ ПРОЕКТ**

|  |  |  |
| --- | --- | --- |
| По модулю | | Разработка AI-проекта |
|  | | |
|  | | |
| На тему | Решение прикладной задачи мониторинга перемещения мебели и техники в общественных учреждениях с помощью технологии глубокого обучения | |
|  | | |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Ф.И.О.  обучающихся | Наименование направления подготовки / специальности | Курс | Группа | Ф.И.О. руководителя (-ей) должность / уч. степень / звание |
| Архаров Н. М. | 09.03.02 Информационные системы и технологии | 4 | 351018 | Васендина И.С., доцент, к.т.н. |
| Голышев А. В. | 09.03.02 Информационные системы и технологии | 4 | 351018 | Васендина И.С., доцент, к.т.н. |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Признать, что проект выполнен и защищен с отметкой |  |  |  |  |
|  |  | (отметка прописью) |  | (дата) |
| Руководитель |  |  |  | И.С. Васендина |
|  |  | (подпись руководителя) |  | (инициалы, фамилия) |

Архангельск 2024

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ РОССИЙСКОЙ ФЕДЕРАЦИИ | | | | | | | | | | | | | | | | | | | | | |
| федеральное государственное автономное образовательное учреждение | | | | | | | | | | | | | | | | | | | | | |
|  | высшего образования  **«Северный (Арктический) федеральный университет имени М.В. Ломоносова»** | | | | | | | | | | | | | | | | | | |  | |
|  |  | | | | | | | | | | | | | | | | | | |  | |
|  | Высшая школа информационных технологий и автоматизированных систем | | | | | | | | | | | | | | | | | | |  | |
|  | (наименование высшей школы / филиала / института / колледжа) | | | | | | | | | | | | | | | | | | |  | |
|  |  | | | | | | | | | | | | | | | | | | |  | |
|  | **ЗАДАНИЕ НА КУРСОВОЙ ПРОЕКТ** | | | | | | | | | | | | | | | | | | |  | |
|  |  | | | | | | | | | | | | | | | | | | |  | |
|  | ппо | Разработка AI-проекта | | | | | | | | | | | | | | | | | |  | |
|  |  | | | | (наименование дисциплины) | | | | | | | | | | | | | | |  | |
|  | студенту | | ВШИТиАС | | высшей школы | | | | 4 | | | | | курса | 351018 | | | | группы |  | |
|  | Архарову Никите Михайловичу, Голышеву Алексею Витальевичу | | | | | | | | | | | | | | | | | | |  | |
|  | (фамилия, имя, отчество студента) | | | | | | | | | | | | | | | | | | |  | |
|  | |  | | --- | | 09.03.02 «Информационные системы и технологии» | | (код и наименование направления подготовки/специальности) | | | | | | | | | | | | | | | | | | | |  | |
|  | ТЕМА: | | | Решение прикладной задачи мониторинга перемещения мебели и техники | | | | | | | | | | | | | | | |  | |
|  | в общественных учреждениях с помощью технологии глубокого обучения | | | | | | | | | | | | | | | | | | |  | |
|  |  | | | | | | | | | | | | | | | | | | |  | |
|  | ИСХОДНЫЕ ДАННЫЕ: | | | | | | | Необходимо решить задачу мониторинга | | | | | | | | | | | |  | |
|  | перемещения мебели и техники в общественных учреждениях с помощью технологии | | | | | | | | | | | | | | | | | | |  | |
|  | глубокого обучения. Описать актуальность, цель, задачи, описать набор данных или | | | | | | | | | | | | | | | | | | |  | |
|  | как собирали, сформулировать задачу, проанализировать целевую переменную, | | | | | | | | | | | | | | | | | | |  | |
|  | провести предобработку изображений (если требуется), построить несколько моделей | | | | | | | | | | | | | | | | | | |  | |
|  | по архитектуре (или с разными параметрами) для решения задачи. Обучить модели. | | | | | | | | | | | | | | | | | | |  | |
|  | Подобрать параметры. Оценить качество работы моделей. Обосновать выбор модели. | | | | | | | | | | | | | | | | | | |  | |
|  | Проверить работу модели на тестовых данных. Определить требования к | | | | | | | | | | | | | | | | | | |  | |
|  | пользовательскому интерфейсу приложения. Спроектировать базу данных. | | | | | | | | | | | | | | | | | | |  | |
|  | Реализовать логирование работы модели в базу данных. Реализовать обработку | | | | | | | | | | | | | | | | | | |  | |
|  | данных о работе модели из базы данных и отображение нужных данных в | | | | | | | | | | | | | | | | | | |  | |
|  | пользовательском интерфейсе. Сделать выводы о качестве работы. | | | | | | | | | | | | | | | | | | |  | |
|  |  | | | | | | | | | | | | | | | | | | |  | |
|  | Срок выполнения с 19.12.2023 г. по 19.01.2024 г. | | | | | | | | | | | | | | | | | | |  | |
|  | Руководитель работы | | | | | | доцент | | | |  | |  | | | |  | И.С. Васендина | |  | |
|  |  | | | |  | (должность) | | | |  | | (подпись) | | | |  | | (инициалы, фамилия) | | |

Архангельск 2024

Лист для замечаний

Оглавление

[Введение 5](#_Toc157012543)

[1 Описание предметной области 7](#_Toc157012544)

[2 Анализ вариантов и проектирование решения задачи 9](#_Toc157012545)

[3 Сбор данных 12](#_Toc157012546)

[3.1 Основной набор данных 12](#_Toc157012547)

[3.2 Добавочный набор данных 14](#_Toc157012548)

[3.3 Аугментация данных 16](#_Toc157012549)

[3.3.1 Аугментация основного набора данных 16](#_Toc157012550)

[3.3.2 Аугментация добавочного набора данных 17](#_Toc157012551)

[3.3.3 Результаты аугментации 19](#_Toc157012552)

[4 Обучение модели YOLOv8 21](#_Toc157012553)

[4.1 Сравнение различных предобученных моделей 21](#_Toc157012554)

[4.2 Обучение модели 22](#_Toc157012555)

[4.3 Результаты обучения 22](#_Toc157012556)

[5 Внедрение модели в алгоритм отслеживания объектов 24](#_Toc157012557)

[5.1 Описание функционала 24](#_Toc157012558)

[5.2 Реализация функционала 25](#_Toc157012559)

[Заключение 29](#_Toc157012560)

[Список используемых источников 32](#_Toc157012561)

[Приложение А (обязательное) Листинг 33](#_Toc157012562)

Введение

В настоящее время технологии компьютерного зрения активно применяются для решения различных задач, связанных с анализом и интерпретацией визуальных данных. Одним из перспективных направлений использования этих технологий является мониторинг и учет имущества на различных объектах. Это особенно актуально в общественных учреждениях, таких как школы, больницы, офисы и др., где присутствует большое количество ценного имущества, требующего постоянного отслеживания и предотвращения возможных хищений. Ручной учет и мониторинг данных процессов обычно связаны с значительными трудозатратами и не гарантируют полной точности и достоверности.

В рамках данной работы предлагается использовать методы компьютерного зрения для автоматизации процессов мониторинга перемещения мебели и техники на территории общественных учреждений. Целью исследования является разработка алгоритмов компьютерного зрения, способных эффективно отслеживать предметы мебели в общественных учреждениях.

Для достижения поставленной цели предполагается решить следующие задачи:

- анализ предметной области: изучение существующих методов для задач распознавания и отслеживания объектов;

- формирование обучающей выборки на основе видеоданных;

- обучение нейросетевой модели для распознавания классов объектов;

- разработка алгоритма отслеживания объектов;

- тестирование разработанных алгоритмов на контрольных данных.

Практическая значимость работы заключается в разработке модели для автоматизированной системы мониторинга имущества для общественных учреждений на основе передовых алгоритмов компьютерного зрения. Реализация поставленных задач позволит сократить трудозатраты на учет имущества, повысить точность и оперативность мониторинга, используя передовые технологии искусственного интеллекта. Полученные результаты могут быть применены для решения практических задач оптимизации учета и сохранности имущества в общественных учреждениях.

**1 Описание предметной области**

В общественных учреждениях, таких как школы, больницы, офисы и т.д., обычно присутствует большое количество ценного имущества, которое требует постоянного контроля и отслеживания. Однако учёт и мониторинг таких процессов могут быть трудоемкими и неточными, если осуществляются вручную. В качестве решения этой проблемы предлагается автоматизация этих процессов с помощью компьютерного зрения.

Целью настоящей работы является разработка модели компьютерного зрения для автоматизации процессов мониторинга перемещения мебели и техники в общественных учреждениях. В настоящее время учёт мебели и техники в большинстве случаев осуществляется вручную с помощью различных документов, таких как инвентарные описи, журналы учета и др. Этот подход может быть неточным и трудоемким, особенно если требуется отследить перемещение объектов в реальном времени.

Коммерческие решения для мониторинга перемещения имущества, такие как системы видеоаналитики от Hikvision, Dahua, Axis и других вендоров, могут быть достаточно дорогими и не всегда эффективными в решении поставленной задачи. С другой стороны, существует множество открытых библиотек компьютерного зрения, таких как OpenCV, Detectron2, MMTracking, которые могут быть использованы для построения системы мониторинга перемещения мебели и техники. Однако эти библиотеки могут быть довольно сложными в использовании и требовать значительных усилий для адаптации под конкретную задачу.

Таким образом, разработка специализированной системы мониторинга перемещения мебели и техники на основе компьютерного зрения может значительно повысить эффективность и точность учета имущества в общественных учреждениях. Кроме того, такая система может быть более доступной по цене, чем коммерческие решения, и более гибкой в настройке и использовании, чем открытые библиотеки компьютерного зрения.

**2 Анализ вариантов и проектирование решения задачи**

Для решения поставленной задачи по мониторингу перемещения объектов в помещении можно рассмотреть несколько основных подходов. Во-первых, можно использовать классические алгоритмы компьютерного зрения, такие как HOG + SVM или Haar-каскады, для детектирования объектов. Однако эти подходы обладают ограниченной точностью и неустойчивы к изменению условий съемки.

Во-вторых, можно применить современные архитектуры глубокого обучения, такие как YOLO, SSD или Faster R-CNN. Эти архитектуры демонстрируют высокую точность и скорость работы. В частности, YOLOv8 является одной из самых перспективных архитектур для решения задачи детектирования объектов.

В-третьих, можно использовать двухэтапный подход, состоящий из детектирования и трекинга объектов. На первом этапе выполняется распознавание объектов в кадре при помощи нейросетевой модели, а на втором этапе применяется алгоритм трекинга, например, SORT, DeepSORT или Bytetrack. Bytetrack является эффективным и точным инструментом для трекинга объектов, способным правильно определять идентификаторы объектов даже в сложных ситуациях, когда объекты пересекаются или частично закрыты другими объектами.

В-четвертых, можно применить одноэтапный подход с использованием end-to-end моделей типа Tracktor, которые объединяют детектирование и трекинг объектов в одной модели.

Среди перечисленных подходов наиболее перспективным представляется использование двухэтапного pipeline с применением современной нейросетевой архитектуры на этапе детектирования и алгоритма Bytetrack для трекинга. Такой подход позволяет объединить высокую точность современных нейросетей в задаче детектирования и эффективные алгоритмы отслеживания объектов между кадрами.

В качестве обучающей выборки мы планируем использовать видеоданные, снятые в различных общественных учреждениях, которые будут аннотированы вручную для определения классов объектов и их расположения на каждом кадре видео. Затем мы будем использовать эти данные для обучения нейросетевой модели, которая будет способна распознавать классы объектов на новых кадрах видео.

Мы планируем протестировать различные предобученные модели, такие как YOLOv8n, YOLOv8s, YOLOv8m, YOLOv8l, YOLOv8x, чтобы определить, какая модель лучше подходит для данной задачи.

После обучения нейросетевой модели мы планируем разработать алгоритм отслеживания объектов на основе этой модели. Алгоритм будет включать в себя различные техники для отслеживания движущихся объектов, такие как оптический поток, механизм Кальмана и частичное следование. Мы также планируем использовать методы восстановления трека, чтобы предотвратить потерю объектов и обеспечить непрерывное отслеживание.

Наконец, мы планируем провести тестирование разработанных алгоритмов на контрольных данных, чтобы оценить их точность и эффективность.

На рисунке 1 можно увидеть этапы разработки трекера.



Рисунок 1 – Этапы разработки трекера

В целом, результаты исследования могут иметь важное значение для автоматизации процессов мониторинга и учета имущества в общественных учреждениях. Разработанные алгоритмы могут быть использованы для обеспечения более точного и эффективного учета имущества, а также для предотвращения потерь и хищений. Кроме того, результаты исследования могут быть использованы для разработки более совершенных систем компьютерного зрения, которые могут быть применены в различных областях, таких как безопасность, транспорт и медицина.

Таким образом, на основании проведенного анализа в качестве архитектуры решения предлагается использовать двухэтапный подход с применением YOLOv8 и Bytetrack для наиболее эффективного решения задачи мониторинга объектов.

**3 Сбор данных**

В ходе исследования были использованы два набора данных: основной и дополнительный. Основной набор данных был применен для разработки и обучения модели, в то время как дополнительный набор был использован для увеличения объема данных и улучшения ее общей эффективности.

3.1 Основной набор данных

В основном наборе данных использовались 1858 изображений, собранных с камер видеонаблюдения, установленных в учебном заведении Сибирского федерального университета (САФУ), в частности, в корпусе IT-парка. Из этих изображений 200 были выделены для валидации результатов модели. Примеры изображений показаны на рисунках 2, 3.



Рисунок 2 – Пример изображения хола из основного набора данных



Рисунок 3 – Пример изображения аудитории из основного набора данных

Набор данных содержит разнообразные классы объектов, представляющих интерьер офисных помещений, такие как:

- стул;

- стол;

- человек;

- монитор;

- ноутбук;

- клавиатура;

- компьютер;

- интерактивная доска.

Каждый класс обладает уникальными характеристиками, формой и расположением в пространстве, что предоставляет модели разнообразные сценарии для обучения и тестирования.

Разметка данных осуществлялась с помощью инструмента Label Studio, развернутого локально, обеспечивающего точную разметку объектов на изображениях. Для каждого объекта сохранены координаты ограничивающего прямоугольника, определяющего положение объекта на изображении. Этот подход к разметке данных обеспечивает модели необходимую информацию для эффективного обучения и последующего распознавания объектов на видеозаписях.

Важно отметить, что изображения были получены в реальных условиях с использованием систем видеонаблюдения, что делает этот набор данных особенно значимым для разработки и тестирования алгоритмов компьютерного зрения в условиях учебных офисных пространств.

3.2 Добавочный набор данных

Для увеличения объема данных и повышения эффективности модели использовался дополнительный набор данных, полученный из ресурса Roboflow. Этот набор данных, известный как «nva Dataset» [6], представляет собой разнообразные изображения, содержащие объекты, характерные для рабочих пространств. Он является открытым исходным набором данных, доступным для использования в области компьютерного зрения.

Исследование использовало описание набора данных, представленное автором Gerrux на Roboflow Universe. Согласно описанию, набор данных включает следующие классы объектов:

- кресло (777 изображений);

- интерактивная доска (95 изображений);

- клавиатура (257 изображений);

- ноутбук (375 изображений);

- монитор (372 изображения);

- системный блок (410 изображений);

- человек (450 изображений);

- стол (242 изображения).

Общее количество изображений в данном наборе данных составляет 2834, что обеспечивает достаточный объем данных для обучения и тестирования моделей компьютерного зрения. Пример изображения из этого набора можно увидеть на рисунке 4.



Рисунок 4 – Пример изображения из добавочного набора данных

Этот набор данных предоставляет разнообразные сценарии, включая в себя рабочие столы, оборудование, а также присутствие людей в офисной среде. Изображения были предоставлены с разметкой, что позволяет эффективно использовать их для обучения алгоритмов распознавания объектов. Данный дополнительный набор значительно расширяет контекст исследования, обогащая его разнообразием сценариев, связанных с офисной обстановкой и техническим оборудованием.

3.3 Аугментация данных

Для улучшения эффективности обучения нейронной сети и повышения ее обобщающей способности было решено применить методы аугментации данных. Целью аугментации является увеличение разнообразия тренировочных данных, что помогает модели лучше обобщать информацию и повышать ее устойчивость к различным вариациям входных данных.

3.3.1 Аугментация основного набора данных

Основной набор данных, состоящий из 1858 изображений, был подвергнут аугментации с помощью комплекса операций. Это позволило обогатить и улучшить разнообразие тренировочного набора данных.

Во-первых, была применена горизонтальная флип-аугментация с помощью оператора A.HorizontalFlip(p=1). Этот оператор создает зеркальные вариации изображений, добавляя различные перспективы объектов и увеличивая разнообразие данных.

Затем была использована операция A.RandomBrightnessContrast(p=0.5), которая воздействует на яркость и контрастность изображений. Эта аугментация добавляет вариации в условия освещения, что помогает модели эффективнее обучаться при различных уровнях яркости.

Далее была осуществлена аугментация цвета при помощи оператора A.RGBShift(r\_shift\_limit=30, g\_shift\_limit=30, b\_shift\_limit=30, p=0.5). Этот оператор изменяет цветовые характеристики изображений, создавая дополнительные вариации и помогая модели лучше обобщать информацию.

Кроме того, для внесения дополнительной сложности и разнообразия в тренировочный набор данных были применены операторы A.ISONoise и A.Cutout. Оператор A.ISONoise внедряет шум в изображения, а A.Cutout выполняет случайное вырезание прямоугольных областей из изображений.

В результате этих последовательных аугментаций тренировочный набор данных значительно увеличился, достигнув объема в 3196 изображений. Этот расширенный объем данных позволяет модели более эффективно обучаться на разнообразных входных данных, повышая ее способность к точному и устойчивому распознаванию объектов в различных учебных офисных условиях. Пример результата аугментации для основного набора данных можно увидеть на рисунке 5.

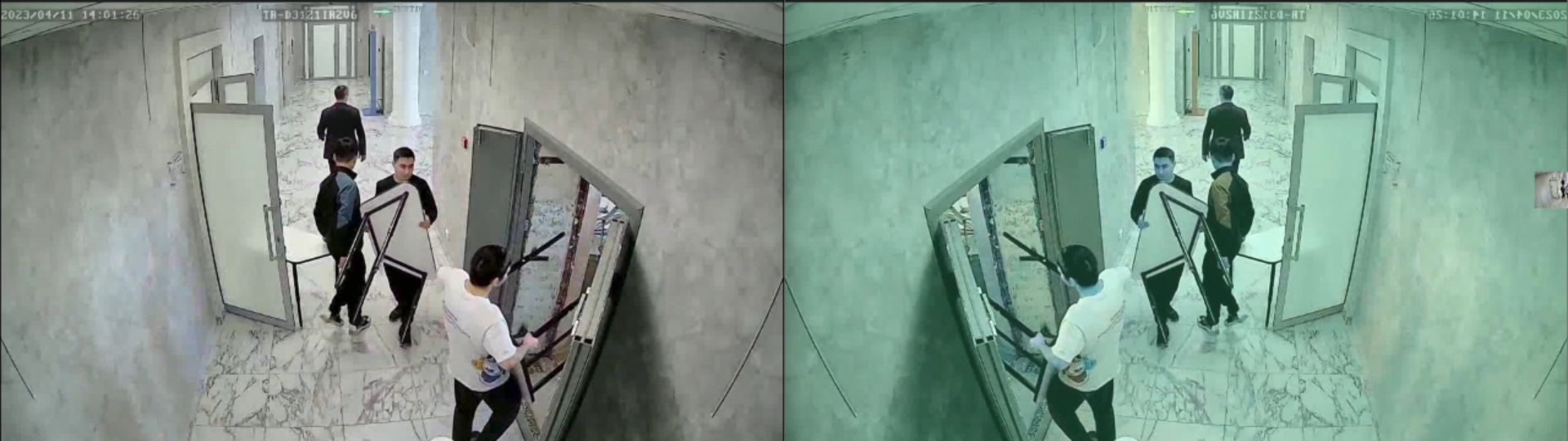


Рисунок 5 – Аугментированное изображение в добавочном наборе данных

3.3.2 Аугментация добавочного набора данных

Каждый пример тренировочного набора был дополнен тремя выходами, что способствует более глубокому и разностороннему обучению.

Rotation. Изображения были повернуты на углы от -15° до +15°, что помогает модели лучше обнаруживать объекты в различных ориентациях.

Brightness. Изменение яркости в пределах от -15% до +15% способствует более устойчивому обучению при различных условиях освещения.

Noise. Добавление шума до 1.96% пикселей помогает модели более эффективно обрабатывать изображения с некоторым уровнем артефактов.

Пример аугментированного изображения можно увидеть на рисунке 6.



Рисунок 6 – Аугментированное изображение в добавочном наборе данных

Для более точного выделения объектов на изображениях была применена аугментация bounding box с использованием следующих параметров:

- 90° Rotate: поворот bounding box на 90° по часовой и против часовой стрелки обеспечивает учет вертикальных и горизонтальных ориентаций объектов;

- Shear: изменение формы bounding box на ±10° по горизонтали и вертикали позволяет учесть возможные искажения объектов на изображении;

- Blur: применение размытия до 0.6px помогает модели справляться с различными уровнями четкости изображений и улучшает ее обобщающую способность.

Этот подход к аугментации данных обеспечивает модели достаточное разнообразие тренировочных данных, что способствует более эффективному и устойчивому обучению сети на различных входных сценариях.

3.3.3 Результаты аугментации

Применение описанных методов аугментации данных существенно увеличило объем и разнообразие тренировочных данных, что положительно сказалось на обучении нейронной сети. Результаты аугментации отразились на размере основного и добавочного наборов данных.

Основной набор данных, включающий в себя изначально 1858 изображений, был расширен до 3196 изображений. Это значительное увеличение объема данных способствует лучшему обучению модели и позволяет ей успешно обобщаться на различные условия и вариации объектов.

Добавочный набор данных также получил значительное увеличение с 2834 до 8064 изображений. Из них 219 изображений были выделены для валидации, обеспечивая более точную оценку производительности модели. Это расширение добавочного набора данных с учетом аугментации способствует более глубокому и разностороннему обучению модели на различных сценариях офисной обстановки.

В сумме получилось 11260 изображений.

Итоговые результаты аугментации подчеркивают важность применения данного метода для эффективного обучения модели на широком спектре данных, что в свою очередь повышает ее способность к точному распознаванию объектов и адаптации к различным условиям видеонаблюдения в учебных офисных пространствах.

**4 Обучение модели YOLOv8**

Для решения задачи распознавания и отслеживания объектов в данной работе используется архитектура нейронной сети YOLOv8 [4]. Данная модель была выбрана в связи с высокой скоростью работы и точностью распознавания объектов в реальном времени.

В данной главе описан процесс обучения модели YOLOv8 на сформированном наборе данных с классами объектов, характерными для учебных помещений.

4.1 Сравнение различных предобученных моделей

При выборе модели для задачи отслеживания объектов в реальном времени, было рассмотрено несколько вариантов предобученных моделей YOLOv8 (YOLOv8n, YOLOv8s, YOLOv8m, YOLOv8l и YOLOv8x) с различными параметрами и характеристиками.

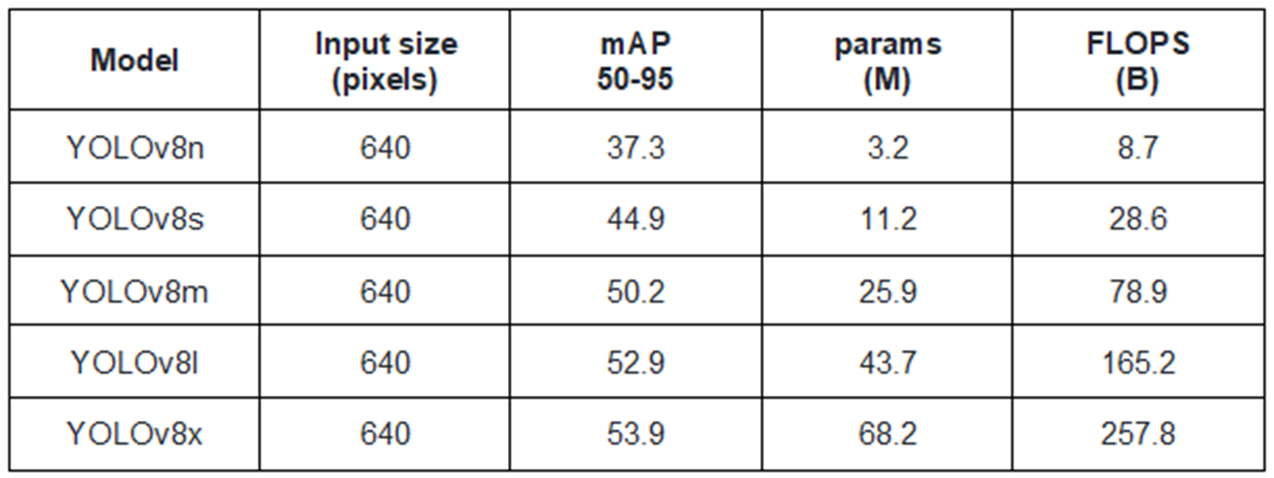


Рисунок 7 – Сравнительная таблица моделей YOLOv8

Как видно из рисунка 7, модель YOLOv8x имеет самую высокую метрику mAP (53.9), но при этом она является самой крупной и требует наибольшее количество операций (FLOPS). Модель YOLOv8n имеет самую низкую метрику mAP (37.3) и наименьшее количество параметров и операций.

В данной задаче важна скорость обработки потока данных, поэтому было принято решение выбрать модель YOLOv8s. Хотя у неё не самый высокий показатель метрики mAP (44.9), она имеет приемлемый баланс между скоростью и точностью, и обладает достаточно высокой производительностью для работы в реальном времени. Кроме того, у YOLOv8s меньше параметров по сравнению с более крупными моделями, что упрощает ее развертывание и обучение.

Таким образом, модель YOLOv8s была выбрана в качестве основной для дальнейшей работы над задачей отслеживания объектов в реальном времени.

4.2 Обучение модели

Для обучения использовалась предобученная на наборе данных COCO модель YOLOv8s. Обучение выполнялось в течение 40 эпох. Размер входных изображений составлял 640x640 пикселей.

4.3 Результаты обучения

По завершению обучения были получены следующие результаты:

- точность (precision) на тестовой выборке – 71,5%;

- полнота (recall) на тестовой выборке – 75,5%;

- средняя точность на наборе данных (mAP50) – 77,7%;

- средняя точность при пороге IoU от 50% до 95% (mAP50-95) – 46,1%;

- время обработки (preprocess) одного изображения – 1,1 мс;

- время прогнозирования (inference) модели – 6,6 мс;

- время постобработки (postprocess) – 1,8 мс;

- общее время распознавания 1 изображения – 42 мс.

На рисунках 8, 9 можно отследить динамику обучения по различным метрикам.

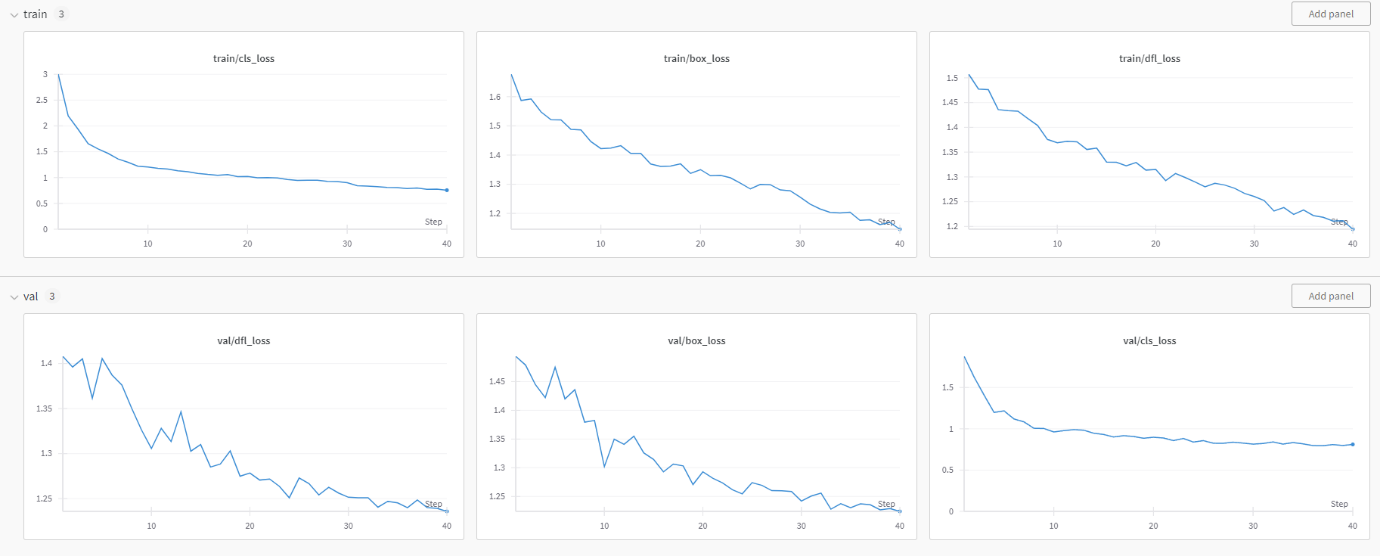


Рисунок 8 – Динамика ошибки при обучении YOLOv8s



Рисунок 9 – Динамика различных метрик при обучении модели

Таким образом, обученная модель YOLOv8s показала достаточно высокие метрики точности и полноты распознавания на тестовых данных. Модель способна работать в режиме, близком к реальному времени, обрабатывая около 25 кадров в секунду. Это позволяет использовать её в системах мониторинга, требующих высокой скорости обработки видеопотока. В следующей главе модель будет интегрирована в алгоритм отслеживания объектов.

**5 Внедрение модели в алгоритм отслеживания объектов**

5.1 Описание функционала

Рассматриваемый алгоритм предназначен для отслеживания перемещения объектов на территории учебных заведений. В качестве входных данных он использует видеосигнал с камеры наблюдения. Алгоритм выполняет следующие функции:

- обнаружение объектов в кадре с помощью обученной модели YOLOv8s. Модель способна распознавать 8 классов объектов: стул, стол, человек, монитор, ноутбук, клавиатура, персональный компьютер, интерактивная доска;

- подсчет количества объектов каждого класса на текущем кадре;

- определение направления движения объектов по обозначенной линии на кадре. Специальный счетчик фиксирует, сколько раз объекты каждого класса пересекли линию при входе или выходе из аудитории;

На выходе алгоритма получается аннотированное видео, в котором вокруг каждого объекта на кадре нарисован прямоугольник с указанием его класса, а также текстовая информация о количестве объектов разных классов в кадре и о том, сколько раз объект вошел или вышел из зрительного зала.

Если подробнее остановиться на первой функции, то обнаружение объектов с помощью модели YOLOv8 предполагает использование конволюционной нейронной сети для анализа видеопотока и идентификации объектов в нем. Модель была обучена на большом наборе изображений, что позволяет ей точно классифицировать объекты в кадре. Затем алгоритм использует эту информацию, чтобы нарисовать ограничительные рамки вокруг обнаруженных объектов и пометить их соответствующим классом.

Вторая функция, подсчет количества объектов каждого класса на текущем кадре, включает в себя отслеживание количества ограничительных рамок, нарисованных вокруг каждого класса объектов. Эта информация может использоваться для получения в реальном времени данных о количестве объектов в аудитории, что может быть полезно для различных целей, например, для обеспечения достаточного количества стульев для студентов или контроля количества используемых ноутбуков.

Третья функция – определение направления движения объектов через обозначенную линию на кадре – включает в себя анализ положения ограничительных рамок вокруг каждого объекта от одного кадра к другому. Если объект пересекает обозначенную линию, алгоритм записывает эту информацию и обновляет счетчик класса этого объекта. Эта информация может быть использована для отслеживания потоков людей и объектов, входящих и выходящих из аудитории, предоставляя ценные сведения о том, как используется пространство.

В целом, алгоритм представляет собой мощный инструмент для мониторинга перемещения объектов в учебных заведениях. Комбинируя обнаружение, подсчет и отслеживание объектов, он обеспечивает комплексное представление о пространстве и может быть использован для оптимизации его использования, повышения безопасности и улучшения качества обучения студентов.

5.2 Реализация функционала

Реализация функционала включала в себя несколько ключевых этапов. Во-первых, был создан класс VideoProcessor для удобной работы с видео. Этот класс позволяет удобно работать с видеороликами, получать кадры по порядковому номеру и извлекать информацию о размерах видео.

Во-вторых, была визуализирована линия на кадре, через которую отслеживается движение объектов. Линия была реализована путем создания графического элемента на кадре, что позволяет визуализировать линию на изображении и определить ее положение. Эту линию можно увидеть на рисунке 10.

В-третьих, был загружен тестовый видеоролик, который использовался как тестовый набор данных для анализа и проверки функциональности алгоритма. Затем были установлены необходимые библиотеки, такие как YOLOv8s, ByteTrack [8] и Roboflow Supervision [5], которые предоставляют функции для детектирования объектов на видео, отслеживания объектов и подсчета пересечений линий.

Затем была реализована вспомогательная функция для сопоставления ограничительных рамок, полученных от модели, с рамками, созданными трекером. Это необходимо для отслеживания движущихся объектов на видео. Затем была загружена ранее обученная модель YOLOv8s, которая используется для предсказания объектов на кадрах видео.

Затем была реализована функция получения предсказаний модели и визуализации на одиночном кадре. Эта функция позволяет получить результаты детектирования объектов на конкретном кадре видео и визуализировать их на изображении. Полученный кадр можно увидеть на рисунке 11.

Кроме того, были созданы классы CustomLineCounter и CustomLineCounterAnnotator для подсчета пересечений линий объектами на видео и аннотации кадров с информацией о пересечениях. Затем был реализован цикл по кадрам видео, в котором происходит детектирование объектов, подсчет пересечений линии и запись результатов в выходное видео.

Таким образом, все необходимые компоненты были реализованы и интегрированы для решения поставленной задачи. Реализованная система позволяет отслеживать перемещение объектов через заданную линию на кадре и подсчитывать количество пересечений. Эта информация может быть использована для различных целей, например, для мониторинга потоков людей и объектов, входящих и выходящих из аудитории, или для контроля использования оборудования в аудитории.

Реализованная система также позволяет визуализировать результаты отслеживания на видео, что может быть полезно для анализа и улучшения функциональности системы. На рисунке 11 можно увидеть статистическую информацию, из которой понятно, что в аудиторию зашёл 1 человек, и 2 вышло. В целом, реализованная система представляет собой мощный инструмент для мониторинга перемещения объектов в учебных заведениях и может быть использована для оптимизации использования пространства, повышения безопасности и улучшения качества обучения студентов.

В целом, реализованная система является эффективным инструментом для мониторинга перемещения объектов в учебных заведениях и может быть использована для решения различных задач, связанных с управлением имуществом и обеспечением безопасности. Система может быть легко интегрирована в существующие системы видеонаблюдения и использована для автоматизации процессов учета и контроля перемещения имущества.

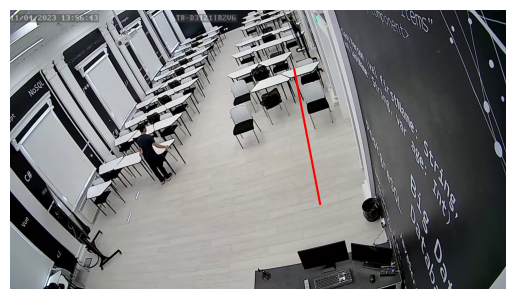


Рисунок 10 – Линия входа/выхода из аудитории

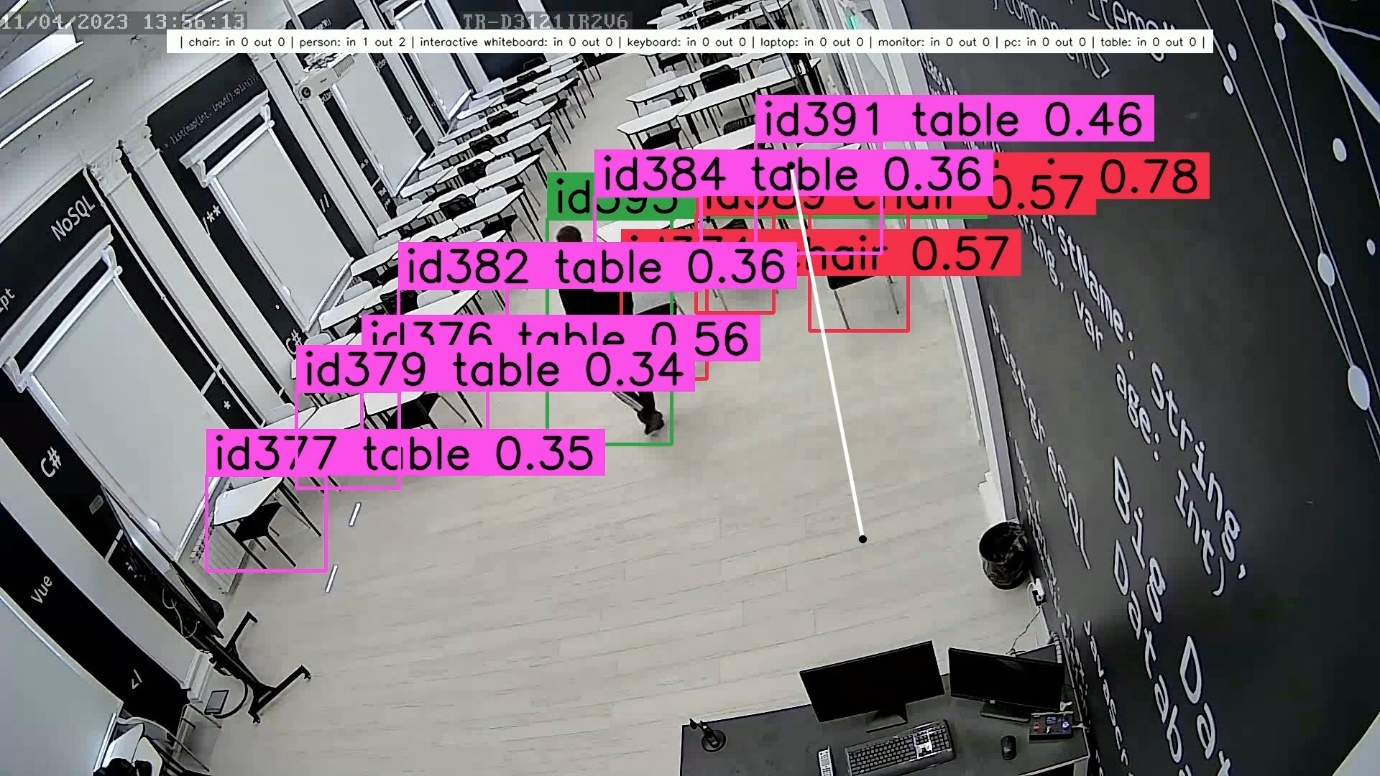


Рисунок 11 – Работа модели

**6 Проектирование базы данных**

6.1 Определение требований для базы данных

В целях эффективного решения задачи мониторинга мебели необходимо предусмотреть хранение следующих данных в базе данных:

- информация о камерах, включая их характеристики;

- места (аудитория, коридор);

- линии входа и выхода объектов для разных мест, их координаты;

- история перемещения объектов;

- информация о количестве объектов в аудитории.

Эти сущности должны быть связаны следующим образом:

- у одного места должна быть одна запись о количестве объектов;

- у одного места может быть несколько камер;

- у одной камеры может быть несколько линий;

- у одной линии может быть несколько записей в истории.

6.2 Entity-диаграмма

На рисунке 12 представлена спроектированная диаграмма базы данных.

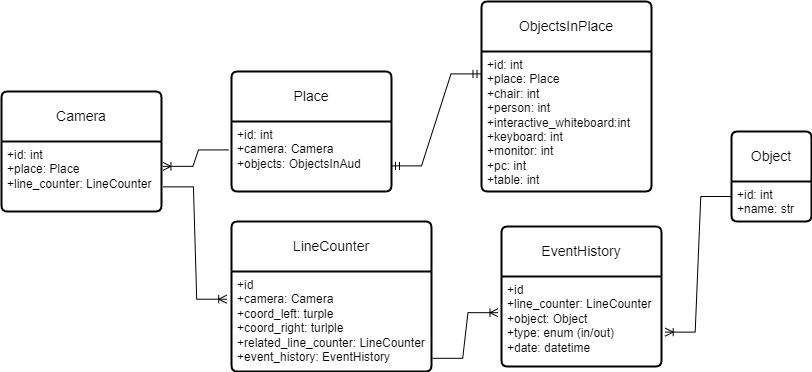


Рисунок 12 – Entity-диаграмма базы данных

Спроектированная база данных удовлетворяет всем вышеперечисленным требованиям и обеспечивает не только сохранение необходимой информации предметной области, но и обеспечивает эффективную работу с ней.

Заключение

В настоящей работе была исследована возможность использования методов компьютерного зрения для автоматизации мониторинга перемещения мебели и техники на территории общественных учреждений. Были поставлены цели и определены задачи, позволяющие достичь эффективного отслеживания перемещения объектов видеонаблюдения.

В ходе выполнения исследования были проведены следующие этапы: анализ предметной области, формирование обучающей выборки, обучение нейросетевой модели для распознавания классов объектов, разработка алгоритма отслеживания объектов и тестирование разработанных алгоритмов на контрольных данных.

В результате была разработана автоматизированная система мониторинга имущества для общественных учреждений на основе передовых алгоритмов компьютерного зрения. Реализация поставленных задач позволила сократить трудозатраты на учет имущества, повысить точность и оперативность мониторинга, используя передовые технологии искусственного интеллекта.

В качестве основы разработки была использована нейросеть YOLOv8, обученная на наборе данных из 1658 изображений, содержащих 8 классов объектов: стул, стол, монитор, системный блок, ноутбук, принтер, телефон и проектор. Модель показала высокие показатели точности и полноты на тестовых данных, а также возможность работы в режиме близком к реальному времени.

Для отслеживания объектов был разработан алгоритм, основанный на использовании нейросети YOLOv8s и трекера ByteTrack. Алгоритм позволяет отслеживать перемещение объектов через заданную линию на кадре и подсчитывать количество пересечений.

В качестве обучающей выборки для нейросети были использованы изображения из учебного заведения САФУ, а также открытый набор данных nva Dataset. Дополнительно были применены методы аугментации данных, что позволило улучшить качество обучения и повысить точность модели.

Практическая значимость работы заключается в создании готовых для внедрения алгоритмов и программных модулей, которые могут быть использованы для построения систем мониторинга и учета имущества в общественных учреждениях. В дальнейшем планируется расширение функционала системы за счет работы с потоком видео и несколькими аудиториями одновременно, а также улучшение точности и скорости обработки данных.

Основными результатами работы являются:

- разработка автоматизированной системы мониторинга имущества для общественных учреждений на основе передовых алгоритмов компьютерного зрения;

- обучение нейросети YOLOv8s для распознавания 8 классов объектов с высокой точностью и полнотой;

- разработка алгоритма отслеживания объектов, основанного на использовании нейросети YOLOv8s и трекера ByteTrack;

- программная реализация алгоритма, позволяющая отслеживать перемещение объектов через заданную линию на кадре и подсчитывать количество пересечений;

- проведение тестирования разработанной системы на контрольных данных и получение положительных результатов.

В целом, результаты работы показывают высокий потенциал использования методов компьютерного зрения для решения задач автоматизации мониторинга и учета имущества в общественных учреждениях. Реализованная система может быть эффективно использована для повышения эффективности учета имущества и предотвращения возможных хищений.

Список используемых источников

1. Как работает Object Tracking на YOLO и DeepSort // Хабр URL: https://habr.com/ru/articles/514450/ (дата обращения: 24.01.2024).
2. Аугментация (augmentation, “раздутие”) данных для обучения нейронной сети на примере печатных символов // Хабр URL: https://habr.com/ru/companies/smartengines/articles/264677/ (дата обращения: 24.01.2024).
3. Ultralytics YOLOv8 Docs // ultralytics URL: https://docs.ultralytics.com/ru/ (дата обращения: 24.01.2024).
4. Ultralytics YOLOv8 // Хабр URL: https://habr.com/ru/articles/710016/ (дата обращения: 05.01.2024)
5. Supervision // roboflow URL: https://supervision.roboflow.com/ (дата обращения: 05.01.2024)
6. nva Image Dataset // Roboflow URL: https://universe.roboflow.com/gerrux/nva/dataset/1 (дата обращения: 24.01.2024).
7. LabelStudio // LabelStudio URL: https://labelstud.io/ (дата обращения: 05.01.2024
8. ByteTrack // GitHub URL: https://github.com/ifzhang/ByteTrack (дата обращения: 05.01.2024)

Приложение А  
(обязательное)  
Листинг

|  |  |
| --- | --- |
| !pip install ultralytics  import os  os.environ["PYTORCH\_CUDA\_ALLOC\_CONF"] = "max\_split\_size\_mb:512"  !pip install wandb  import wandb  wandb.login()  wandb.init()  from ultralytics import YOLO  # Load a model  model = YOLO('yolov8n.pt') # load a pretrained model (recommended for training)  model.train(data='data.yaml', epochs=40, imgsz=1280, batch=4)  import cv2  import matplotlib.pyplot as plt  class VideoProcessor:  def \_\_init\_\_(self, video\_path, ):  self.video\_path = video\_path  self.video = cv2.VideoCapture(video\_path)  self.width = int(self.video.get(cv2.CAP\_PROP\_FRAME\_WIDTH))  self.height = int(self.video.get(cv2.CAP\_PROP\_FRAME\_HEIGHT))  def get\_frame\_by\_order(self, frame\_number):  self.video.set(cv2.CAP\_PROP\_POS\_FRAMES, frame\_number)  ret, frame = self.video.read()  if ret:return frame  %cd {HOME}/ByteTrack  # workaround related to https://github.com/roboflow/notebooks/issues/80  !sed -i 's/onnx==1.8.1/onnx==1.9.0/g' requirements.txt  !pip3 install -q -r requirements.txt  !python3 setup.py -q develop  !pip install -q cython\_bbox  !pip install -q onemetric  # workaround related to https://github.com/roboflow/notebooks/issues/112 and https://github.com/roboflow/notebooks/issues/106  !pip install -q loguru lap thop  from IPython import display  display.clear\_output()  import sys  sys.path.append(f"{HOME}/ByteTrack")  import yolox  print("yolox.\_\_version\_\_:", yolox.\_\_version\_\_)  from yolox.tracker.byte\_tracker import BYTETracker, STrack  from onemetric.cv.utils.iou import box\_iou\_batch  from dataclasses import dataclass  @dataclass(frozen=True)  class BYTETrackerArgs:  track\_thresh: float = 0.25  track\_buffer: int = 30  match\_thresh: float = 0.8  aspect\_ratio\_thresh: float = 3.0  min\_box\_area: float = 1.0  mot20: bool = False  !pip install supervision==0.1.0  from IPython import display  display.clear\_output()  import supervision  tracks: List[STrack]  ) -> Detections:  if not np.any(detections.xyxy) or len(tracks) == 0:  return np.empty((0,))  tracks\_boxes = tracks2boxes(tracks=tracks)  iou = box\_iou\_batch(tracks\_boxes, detections.xyxy)  track2detection = np.argmax(iou, axis=1)  tracker\_ids = [None] \* len(detections)  for tracker\_index, detection\_index in enumerate(track2detection):  if iou[tracker\_index, detection\_index] != 0:  tracker\_ids[detection\_index] = tracks[tracker\_index].track\_id  return tracker\_ids  !cd ../  # settings  MODEL = "/content/best1280.pt"  SOURCE\_VIDEO\_PATH = "/content/test.avi"  !pwd  from ultralytics import YOLO  model = YOLO(MODEL)  model.fuse()  # dict maping class\_id to class\_name  CLASS\_NAMES\_DICT = model.names  # class\_ids of interest - car, motorcycle, bus and truck  CLASS\_ID = [0, 1, 2, 3, 4, 5, 6, 7]  CLASS\_ID\_BY\_NAME = {  'chair' : 0,  'person' : 1,  'interactive whiteboard' : 2,  'keyboard' : 3,  'laptop' : 4,  'monitor' : 5,  'pc' : 6,  'table' : 7,  from supervision.geometry.dataclasses import Point, Rect, Vector  from supervision.tools.detections import Detections  class CustomLineCounter():  def \_\_init\_\_(self, start: Point, end: Point, classes: List):  self.vector = Vector(start=start, end=end)  self.tracker\_state: Dict[str, bool] = {}  self.result\_dict = {int(class\_id): {"in\_count": int(0), "out\_count": int(0)} for class\_id in classes}  def update(self, detections: Detections):  for id in detections.class\_id:  mask = np.array([class\_id in [int(id)] for class\_id in detections.class\_id], dtype=bool)  filtereddet = detections.filter(mask=mask, inplace=False)  for xyxy, confidence, class\_id, tracker\_id in filtereddet:  # handle detections with no tracker\_id  if tracker\_id is None:  continue  # we check if all four anchors of bbox are on the same side of vector  x1, y1, x2, y2 = xyxy  anchors = [  Point(x=x1, y=y1),  Point(x=x1, y=y2),  Point(x=x2, y=y1),  Point(x=x2, y=y2),]  triggers = [self.vector.is\_in(point=anchor) for anchor in anchors]  # detection is partially in and partially out  if len(set(triggers)) == 2:  continue  tracker\_state = triggers[0]  def annotate(self, frame: np.ndarray, line\_counter: LineCounter) -> np.ndarray:  cv2.line(  frame,  line\_counter.vector.start.as\_xy\_int\_tuple(),  line\_counter.vector.end.as\_xy\_int\_tuple(),  self.color.as\_bgr(),  self.thickness,  lineType=cv2.LINE\_AA,  shift=0,  )  cv2.circle(  frame,  line\_counter.vector.start.as\_xy\_int\_tuple(),  radius=5,  color=self.text\_color.as\_bgr(),  thickness=-1,  lineType=cv2.LINE\_AA,  )  cv2.circle(  frame,  line\_counter.vector.end.as\_xy\_int\_tuple(),  radius=5,  color=self.text\_color.as\_bgr(),  thickness=-1,  lineType=cv2.LINE\_AA,  )  report = ""  for key in line\_counter.result\_dict:  class\_name = CLASS\_NAMES\_DICT[key]  in\_count = line\_counter.result\_dict[key]['in\_count']  frame,  report,  (report\_x, report\_y),  cv2.FONT\_HERSHEY\_SIMPLEX,  self.text\_scale,  self.text\_color.as\_bgr(),  self.text\_thickness,  cv2.LINE\_AA,  )  # settings  LINE\_START = Point(coords[1][0],coords[1][1])  LINE\_END = Point(coords[0][0],coords[0][1])  TARGET\_VIDEO\_PATH = f"{HOME}/result.mp4"  from tqdm.notebook import tqdm  # create BYTETracker instance  byte\_tracker = BYTETracker(BYTETrackerArgs())  # create VideoInfo instance  video\_info = VideoInfo.from\_video\_path(SOURCE\_VIDEO\_PATH)  # create frame generator  generator = get\_video\_frames\_generator(SOURCE\_VIDEO\_PATH)  # create LineCounter instance  line\_counter = CustomLineCounter(start=LINE\_START, end=LINE\_END, classes = CLASS\_ID)  # create instance of BoxAnnotator and LineCounterAnnotator  box\_annotator = BoxAnnotator(color=ColorPalette(), thickness=4, text\_thickness=4, text\_scale=2)  line\_annotator = CustomLineCounterAnnotator(thickness=4, text\_thickness=1, text\_scale=0.5, class\_name\_dict = CLASS\_NAMES\_DICT, video\_info = video\_info)  # open target video file  for \_, confidence, class\_id, tracker\_id  in detections]  frame = box\_annotator.annotate(frame=frame, detections=detections, labels=labels)  line\_counter.update(detections=detections)  line\_annotator.annotate(frame=frame, line\_counter=line\_counter)  sink.write\_frame(frame)  print(line\_counter.result\_dict) | else:return None  def get\_width\_height(self):  return self.width, self.height  video\_path = '/content/test.avi'  processor = VideoProcessor(video\_path)  frame\_number = 1320  frame = processor.get\_frame\_by\_order(frame\_number)  coords = [[1100, 230],[1200, 750]]  if frame is not None:  fig, ax = plt.subplots(frameon=False)  ax.axis('off')  plt.imshow(cv2.cvtColor(frame, cv2.COLOR\_BGR2RGB))  x1, y1 = coords[0]  x2, y2 = coords[1]  # Отобразить изображение с линией  plt.plot([x1, x2], [y1, y2], color='r')  plt.show()  !nvidia-smi  import os  HOME = os.getcwd()  print(HOME)  %cd {HOME}  # Pip install method (recommended)  !pip install ultralytics  from IPython import display  display.clear\_output()  import ultralytics  ultralytics.checks()  %cd {HOME}  !git clone https://github.com/ifzhang/ByteTrack.git  print("supervision.\_\_version\_\_:", supervision.\_\_version\_\_)  from supervision.draw.color import ColorPalette  from supervision.geometry.dataclasses import Point  from supervision.video.dataclasses import VideoInfo  from supervision.video.source import get\_video\_frames\_generator  from supervision.video.sink import VideoSink  from supervision.notebook.utils import show\_frame\_in\_notebook  from supervision.tools.detections import Detections, BoxAnnotator  from supervision.tools.line\_counter import LineCounter, LineCounterAnnotator  from typing import List  import numpy as np  # converts Detections into format that can be consumed by match\_detections\_with\_tracks function  def detections2boxes(detections: Detections) -> np.ndarray:  return np.hstack((  detections.xyxy,  detections.confidence[:, np.newaxis]  ))  # converts List[STrack] into format that can be consumed by match\_detections\_with\_tracks function  def tracks2boxes(tracks: List[STrack]) -> np.ndarray:  return np.array([  track.tlbr  for track  in tracks  ], dtype=float)  # matches our bounding boxes with predictions  def match\_detections\_with\_tracks(  detections: Detections,  }  # create frame generator  generator = get\_video\_frames\_generator(SOURCE\_VIDEO\_PATH)  # create instance of BoxAnnotator  box\_annotator = BoxAnnotator(color=ColorPalette(), thickness=4, text\_thickness=4, text\_scale=2)  # acquire first video frame  iterator = iter(generator)  frame = next(iterator)  # model prediction on single frame and conversion to supervision Detections  results = model(frame)  detections = Detections(  xyxy=results[0].boxes.xyxy.cpu().numpy(),  confidence=results[0].boxes.conf.cpu().numpy(),  class\_id=results[0].boxes.cls.cpu().numpy().astype(int)  )  # format custom labels  labels = [  f"{CLASS\_NAMES\_DICT[class\_id]} {confidence:0.2f}"  for \_, confidence, class\_id, tracker\_id  in detections  ]  # annotate and display frame  frame = box\_annotator.annotate(frame=frame, detections=detections, labels=labels)  %matplotlib inline  show\_frame\_in\_notebook(frame, (16, 16))  from typing import Dict  import cv2  import numpy as np  from supervision.draw.color import Color  # handle new detection  if tracker\_id not in self.tracker\_state:  self.tracker\_state[tracker\_id] = tracker\_state  continue  # handle detection on the same side of the line  if self.tracker\_state.get(tracker\_id) == tracker\_state:  continue  self.tracker\_state[tracker\_id] = tracker\_state  if tracker\_state:  self.result\_dict[int(id)]['in\_count'] += 1  else:  self.result\_dict[int(id)]['out\_count'] += 1  class CustomLineCounterAnnotator:  def \_\_init\_\_(  self,  thickness: float = 2,  color: Color = Color.white(),  text\_thickness: float = 2,  text\_color: Color = Color.black(),  text\_scale: float = 0.5,  text\_offset: float = 1.5,  text\_padding: int = 10,  class\_name\_dict = {},  video\_info = []  ):  self.thickness: float = thickness  self.color: Color = color  self.text\_thickness: float = text\_thickness  self.text\_color: Color = text\_color  self.text\_scale: float = text\_scale  self.text\_offset: float = text\_offset  self.text\_padding: int = text\_padding  self.class\_name\_dict = class\_name\_dict  self.video\_info = video\_info  out\_count = line\_counter.result\_dict[key]['out\_count']  report += f" | {class\_name}: in {in\_count} out {out\_count}"  report += ' |'  (report\_width, report\_height), \_ = cv2.getTextSize(  report, cv2.FONT\_HERSHEY\_SIMPLEX, self.text\_scale, self.text\_thickness  )  report\_x = int(  ((self.video\_info.width) - report\_width)  / 2  )  report\_y = int(  (150 + report\_height)  / 2  - self.text\_offset \* report\_height  )  report\_background\_rect = Rect(  x=report\_x,  y=report\_y - report\_height,  width=report\_width,  height=report\_height,  ).pad(padding=self.text\_padding)  cv2.rectangle(  frame,  report\_background\_rect.top\_left.as\_xy\_int\_tuple(),  report\_background\_rect.bottom\_right.as\_xy\_int\_tuple(),  self.color.as\_bgr(),  -1,  )  cv2.putText(  with VideoSink(TARGET\_VIDEO\_PATH, video\_info) as sink:  # loop over video frames  for frame in tqdm(generator, total=video\_info.total\_frames):  # model prediction on single frame and conversion to supervision Detections  results = model(frame)  detections = Detections(  xyxy=results[0].boxes.xyxy.cpu().numpy(),  confidence=results[0].boxes.conf.cpu().numpy(),  class\_id=results[0].boxes.cls.cpu().numpy().astype(int))  # filtering out detections with unwanted classes  mask = np.array([class\_id in CLASS\_ID for class\_id in detections.class\_id], dtype=bool)  detections.filter(mask=mask, inplace=True)  # tracking detections  tracks = byte\_tracker.update(  output\_results=detections2boxes(detections=detections),  img\_info=frame.shape,  img\_size=frame.shape)  tracker\_id = match\_detections\_with\_tracks(detections=detections, tracks=tracks)  detections.tracker\_id = np.array(tracker\_id)  # filtering out detections without trackers  mask = np.array([tracker\_id is not None for tracker\_id in detections.tracker\_id], dtype=bool)  detections.filter(mask=mask, inplace=True)  # format custom labels  labels = [  f"id{tracker\_id} {CLASS\_NAMES\_DICT[class\_id]} {confidence:0.2f}" |

Сведения о самостоятельности выполнения работы

Работа «Решение прикладной задачи мониторинга перемещения мебели и техники в общественных учреждениях с помощью технологии глубокого обучения» выполнена нами самостоятельно.  
Используемые в работе материалы и концепция из публикуемой литературы и других источников имеют ссылки на них.

Один печатный экземпляр работы и электронный вариант работы на цифровом носителе переданы нами на отделение.



« 19 »     января        2024г. \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_          Архаров Н.М.

 (Подпись) (Фамилия, инициалы)

« 19 »     января        2024г. \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_          Голышев А. В.    \_

(Подпись) (Фамилия, инициалы)