

СОДЕРЖАНИЕ

ВВЕДЕНИЕ	2
1 МЕТОДЫ ПОВЫШЕНИЯ ТОЧНОСТИ ДЕТЕКТИРОВАНИЯ ОБЪЕКТОВ НА ИЗОБРАЖЕНИЯХ	3
1.1 УЛУЧШЕНИЕ ПРЕДВАРИТЕЛЬНОЙ ОБРАБОТКИ ИЗОБРАЖЕНИЙ	3
1.2 ОПТИМИЗАЦИЯ ГИПЕРПАРАМЕТРОВ И ФУНКЦИЙ ПОТЕРЬ	4
1.3 ПРИМЕНЕНИЕ СИНТЕТИЧЕСКИХ И ДОПОЛНЕННЫХ ДАННЫХ	5
2 МЕТОДОЛОГИЯ ИССЛЕДОВАНИЯ	7
2.1 РЕШЕНИЕ ЗАДАЧИ ДЕТЕКЦИИ	7
2.1.1 ПРИМЕНЕНИИ YOLOV8 ДЛЯ ЗАДАЧ ОБНАРУЖЕНИЯ	8
2.1.2 ДООБУЧЕНИЕ ПОЛЬЗОВАТЕЛЬСКОЙ МОДЕЛИ ДЕТЕКЦИИ	8
2.1.3 ПРИВЕДЕНИЕ ПОЛЬЗОВАТЕЛЬСКИХ ДАТАСЕТОВ К ОДНОЙ АННОТАЦИИ	11
2.2 РЕШЕНИЕ ЗАДАЧИ СЕГМЕНТАЦИИ	15
2.2.1 ПРИМЕНЕНИЕ МОДЕЛИ SAM ДЛЯ СЕГМЕНТАЦИИ ОБЪЕКТОВ	16
2.2.2 СОЗДАНИЕ НАБОРА ДАННЫХ СЕГМЕНТАЦИИ, ИСПОЛЬЗУЯ МОДЕЛЬ ОБНАРУЖЕНИЯ	19
ЗАКЛЮЧЕНИЕ	22
СПИСОК ИСПОЛЬЗОВАННЫХ ИСТОЧНИКОВ	ОШИБКА! ЗАКЛАДКА НЕ ОПРЕДЕЛЕНА.

ВВЕДЕНИЕ

Целью данной работы является исследование методов повышения точности детектирования объектов на изображениях и интеграция техник сегментации в контексте складской логистики. Компьютерное зрение играет ключевую роль в автоматизации и оптимизации процессов. Складская логистика сталкивается с многочисленными вызовами, включая необходимость быстрого и точного распознавания и сортировки товаров, требующих инновационных подходов в области компьютерного зрения. Данная работа ориентирована на адаптацию алгоритмов глубокого обучения для улучшения точности детекции и сегментации объектов в условиях, характерных для складских помещений, что включает в себя как улучшение предварительной обработки изображений, так и оптимизацию гиперпараметров и функций потерь. Особое внимание уделяется возможности применения синтетических и дополненных данных для увеличения эффективности обучения моделей.

В рамках методологического подхода рассматривается решение задачи детекции с использованием современной модели YOLOv8, которая демонстрирует впечатляющие результаты в обнаружении объектов в различных условиях. Помимо применения этой модели, в работе освещается процесс дообучения пользовательской модели детекции для специфических нужд складской логистики, а также методы приведения пользовательских датасетов к единой аннотации, что обеспечивает согласованность и повышает качество обучения.

Далее, исследование переходит к задаче сегментации. Здесь ключевым моментом является использование модели SAM для сегментации объектов. SAM, или Segment Anything Model, представляет собой передовой подход в выделении и сегментации интересующих объектов, что особенно важно для точного и эффективного определения габаритных характеристик. В рамках работы также рассматривается создание набора данных для сегментации на основе обнаруженных моделью YOLOv8 объектов.

1 Методы повышения точности детектирования объектов на изображениях

1.1 Улучшение предварительной обработки изображений

Нормализация и стандартизация данных. Этот шаг включает в себя масштабирование пиксельных значений изображений для улучшения стабильности и скорости обучения нейронных сетей. Нормализация часто включает преобразование диапазона пиксельных значений изображения к $[0, 1]$ или $[-1, 1]$ путем деления на 255 (максимальное значение пикселя). Стандартизация включает в себя центрирование распределения пикселей изображения вокруг нуля и нормализацию его дисперсии.

Аугментация данных используется для увеличения разнообразия обучающих данных и предотвращения переобучения, используются техники аугментации изображений. Это включает в себя:

- Геометрические преобразования, такие как вращение, масштабирование, и сдвиги.
- Изменение цветового пространства, включая коррекцию яркости, контрастности, насыщенности, и оттенка.
- Применение случайных обрезаек и перспективных трансформаций.
- Использование зеркального отображения (flip) и случайных искажений.
- Введение шума, такого как гауссовский или «соль-перец».

На рисунке 1 изображен пример преобразованного входного изображения

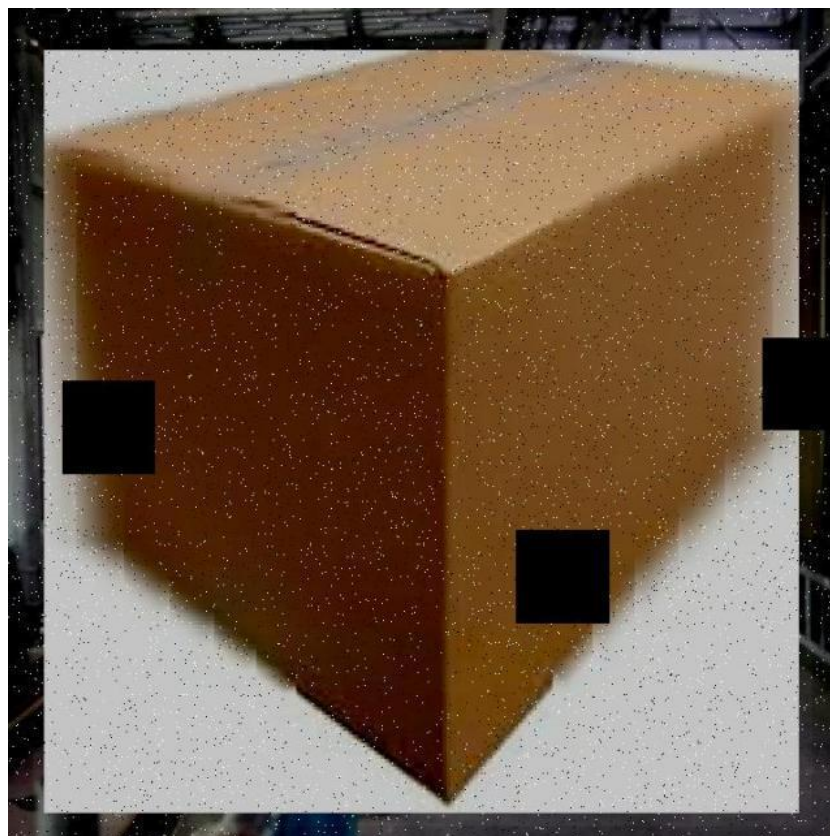


Рисунок 1 – Пример аугментации данных

Фильтрация и улучшение качества изображения: Применение различных фильтров для улучшения визуальных особенностей изображений и уменьшения шума:

- Фильтры размытия (например, Гауссово размытие) для сглаживания шумов.
- Фильтры улучшения резкости для выделения краев и деталей изображения.
- Морфологические операции (эрозия, дилатация) для улучшения структуры объектов.

1.2 Оптимизация гиперпараметров и функций потерь

Тюнинг гиперпараметров – это эффективный выбор и настройка гиперпараметров существенно влияют на производительность модели.

Важные гиперпараметры включают:

- **Скорость обучения.** Определение начальной скорости обучения и стратегии ее изменения в процессе обучения (например, с помощью планировщиков скорости обучения или адаптивных оптимизаторов).
- **Размер пакета.** Выбор подходящего размера пакета, который сбалансирован скорость обучения и стабильность.
- **Количество эпох.** Определение оптимального числа эпох для обучения модели без переобучения.

Кастомизация функций потерь – это адаптация функций потерь под специфику задачи детектирования:

- **Функции потерь для классификации.** Например, кросс-энтропия для определения вероятности принадлежности к классу.
- **Функции потерь для локализации объектов.** Использование потерь, таких как Mean Squared Error (MSE) или Intersection over Union (IoU), для оптимизации точности локализации объектов.
- **Комбинированные функции потерь.** Сочетание различных функций потерь для одновременной оптимизации классификации и локализации.

1.3 Применение синтетических и дополненных данных

Генерация синтетических данных используется для повышения обобщающей способности модели и уменьшения переобучения, генерация синтетических данных может быть использована для дополнения обучающего набора:

- **Использование GAN (Generative adversarial network,).** Генеративно-состязательные сети могут создавать реалистичные изображения для увеличения разнообразия данных.
- **Симуляции.** В некоторых случаях, особенно в робототехнике, симуляции могут быть использованы для генерации данных обучения.

Дополнение реальных данных синтетическими примерами и интеграция синтетических данных с реальными изображениями для создания более обширного и разнообразного набора данных:

- Разработка сценариев для синтетических данных. Создание сценариев, которые эффективно дополняют реальные данные и улучшают обучение в специфических случаях.
- Смешивание реальных и синтетических данных. Использование различных стратегий для смешивания синтетических и реальных данных, чтобы улучшить обучение и обеспечить реалистичность обучающего набора.

Симуляции и 3D моделирование цифровых двойников играют важную роль в обучении алгоритмов компьютерного зрения, особенно в условиях, где сбор реальных данных затруднителен, дорог или опасен. Симуляции позволяют создавать контролируемые условия, такие как различные погодные условия и освещение, что особенно полезно в робототехнике и промышленности. Они также автоматически предоставляют точные аннотации, например, ограничивающие рамки или пиксельные маски, что значительно экономит время и усилия по сравнению с ручной аннотацией. Кроме того, симуляции позволяют легко манипулировать объектами и их свойствами, создавая обширные и разнообразные наборы данных, которые масштабируемы и позволяют генерировать большие объемы данных без значительных дополнительных затрат.

С другой стороны, 3D моделирование цифровых двойников создает виртуальные копии реальных объектов или систем, которые могут быть интегрированы в симуляции для улучшения качества и реалистичности генерируемых данных. Это позволяет создавать детализированные и точные виртуальные копии реальных объектов, что особенно ценно для задач, требующих высокой степени точности, например, в медицинских приложениях или промышленном дизайне.

Цифровые двойники могут взаимодействовать с другими виртуальными объектами и средами в симуляциях, создавая динамические и реалистичные сценарии. Это помогает в обучении алгоритмов компьютерного зрения работать в условиях, близких к реальным. Кроме того, они используются для анализа и оптимизации производительности систем, позволяя проводить эксперименты и тестирования в безопасной и контролируемой среде, а также могут быть использованы для обучения и тестирования алгоритмов машинного обучения перед их применением в реальном мире.

Таким образом, сочетание симуляций и 3D моделирования цифровых двойников предоставляет мощные инструменты для разработки, тестирования и улучшения алгоритмов компьютерного зрения, обеспечивая более широкий спектр обучающих данных и возможности для тестирования в разнообразных и контролируемых средах.

2 Методология исследования

2.1 Решение задачи детекции

В складской логистики задача детекции играет критически важную роль, так как она лежит в основе автоматизации процессов учета и управления запасами. Автоматическое распознавание товаров и их характеристик через компьютерное зрение позволяет значительно ускорить и оптимизировать логистические процессы.

2.1.1 Применении YOLOv8 для задач обнаружения

YOLOv8, будучи одной из последних итераций в серии моделей YOLO (You Only Look Once), представляет собой мощный инструмент для детекции объектов в реальном времени. Благодаря своей способности быстро и точно локализовать объекты на изображениях, YOLOv8 идеально подходит для применения в складской логистике. Она может использоваться для идентификации товаров, их местоположения на полках, при движении по конвейерной ленте, а также для считывания штрих-кодов или других идентификаторов. Важно отметить, что YOLOv8 обладает высокой точностью и скоростью, что критически важно для временно чувствительных операций функционирования фулфилмента.

В таблице 1 представлены основные показатели доступных в YOLOv8 моделей

Таблица 1 – Показатели производительности YOLOv8

Модель	Размер	mAPval 50-95	Скорость CPU ONNX(мс)	СкоростьA100 TensorRT(мс)	params(M)	FLOPs(B)
YOLOv8n	640	37.3	80.4	0.99	3.2	8.7
YOLOv8s	640	44.9	128.4	1.20	11.2	28.6
YOLOv8m	640	50.2	234.7	1.83	25.9	78.9
YOLOv8l	640	52.9	375.2	2.39	43.7	165.2
YOLOv8x	640	53.9	479.1	3.53	68.2	257.8

2.1.2 Дообучение пользовательской модели детекции

Для достижения максимальной эффективности в специфических условиях, требуется дообучение модели детекции. Дообучение (fine-tuning) YOLOv8 на специализированных данных склада позволяет адаптировать модель к конкретным условиям и особенностям работы, таким как особенности освещения, рас-

положения товаров и их внешний вид. Это улучшает точность детекции и снижает количество ложных срабатываний, что особенно важно для автоматизированных систем управления складом.

В YOLOv8 существуют предварительно обученные модели Detect. Модели Detect были предварительно обучены на наборе данных COCO.

Набор данных COCO (Common Objects in Context) — это масштабный набор данных по обнаружению, сегментации объектов и созданию надписей. Он предназначен для стимулирования исследований самых разных категорий объектов и обычно используется для бенчмаркинга моделей компьютерного зрения.

Для детекции кастомных объектов требуется дообучить модель. Для дообучения будем использовать подходящий датасет из открытого источника Roboflow, для использования требуется экспортировать датасет в формате YOLOv8.

Моделью старта возьмем версию весов YOLOv8l.pt.

В результате обучения стартовой модели 10-ю эпохами получаем обновлённый файл весов best.pt с метриками изображенными на рисунке 2.

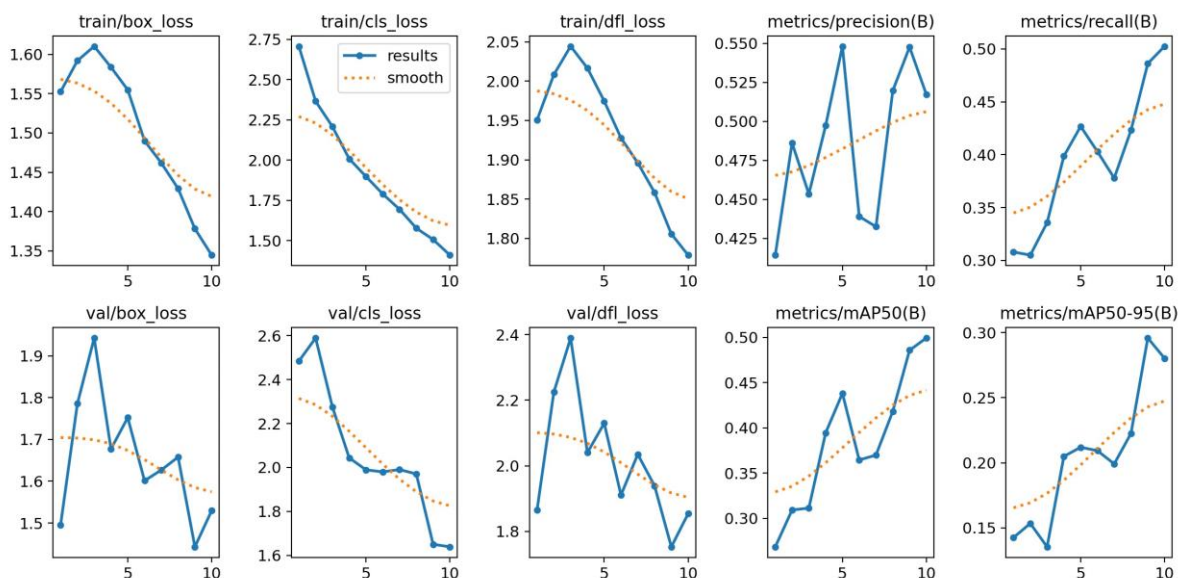


Рисунок 2 – Метрики обучения первых 10 эпох

После производим тестовое детектирования и проверяем на наличии в выходной видеозаписи артефактов и подозрительного поведения.

Далее приступаем к дообучению модели. В качестве датасета можно использовать как предыдущий датасет, так и другой подходящий с такой разметкой. Информация по сведению датасетов к одной аннотации рассказано в следующем пункте.

Для дообучения модели в функции обучения в качестве аргумента весов используем `best.pt` из предыдущего запуска.

В результате дообучения модели 15-ю эпохами получаем обновленный файл весов `best.pt` с метриками изображенными на рисунке 3.

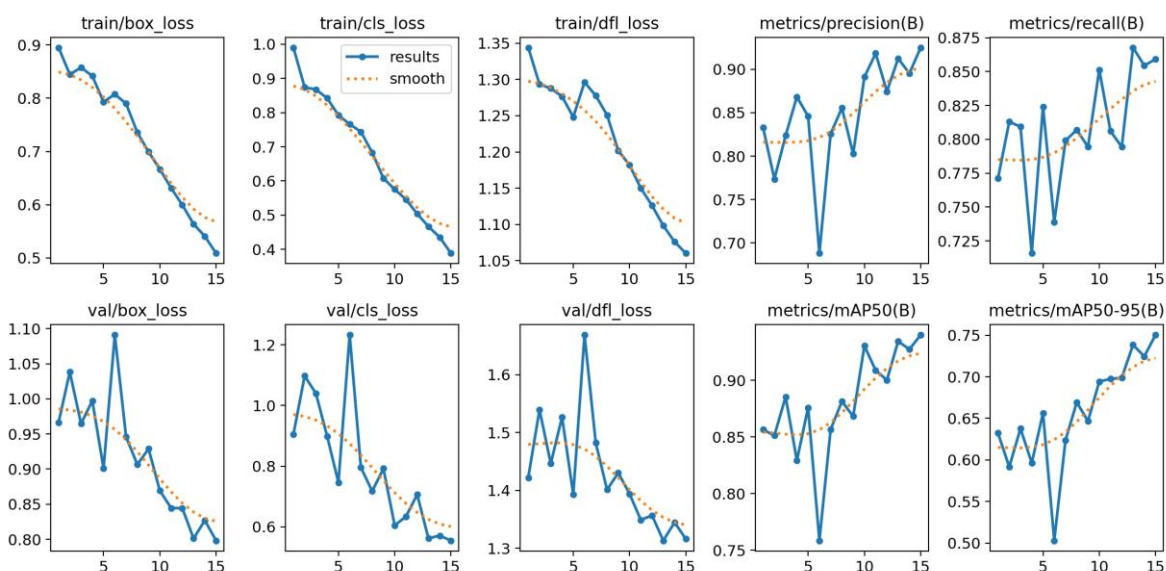


Рисунок 3 – Метрики дообучения 15 эпох

На рисунке 4 изображены детекции упаковок обновленной моделью.

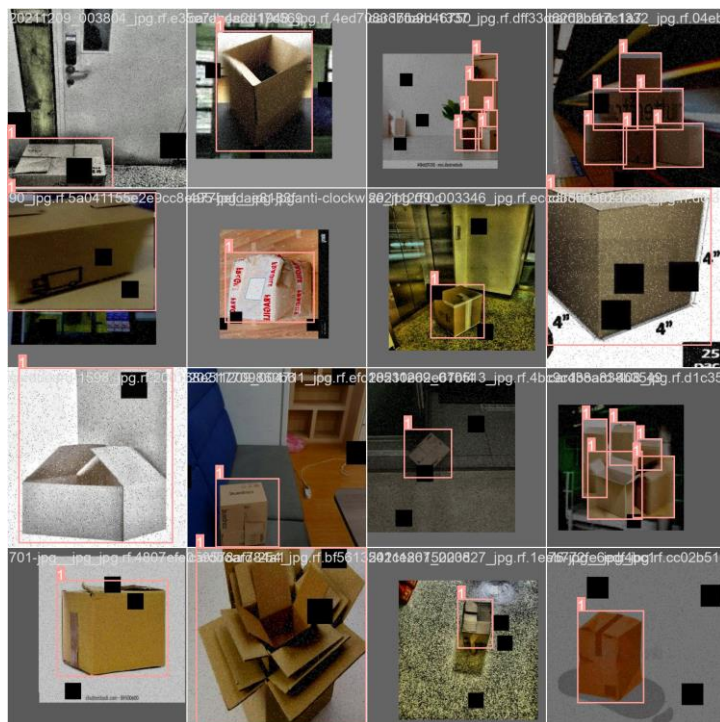


Рисунок 4 – Детекция дообученной модели

2.1.3 Приведение пользовательских датасетов к одной аннотации

Создание единого формата аннотаций для пользовательских датасетов является одним из ключевых шагов в процессе дообучения модели. Разнообразие форматов аннотаций в различных источниках данных может привести к несоответствиям и затруднениям в процессе обучения. Стандартизация аннотаций упрощает процесс интеграции данных из разных источников и повышает качество обучения модели. Это может включать унификацию форматов ограничивающих рамок, меток классов и других параметров, что позволяет создать согласованный и высокоэффективный датасет для обучения и дообучения моделей компьютерного зрения.

Таким образом, эффективное решение задачи детекции на складах с помощью YOLOv8, адаптация и стандартизация данных обеспечивают значительные

преимущества для автоматизации и оптимизации логистических процессов, повышая точность работы системы управления складом и снижая вероятность ошибок.

Для корректного последовательного обучения на нескольких датасетах требуется привести их к одной аннотации. Обратимся к `yaml` файлам в которых указаны пути к данным, количество и имена классов и прочая информация необходимая для начала обучения.

Наблюдаем что в файлах из листингов 1 и 2, разное количество классов и названия их различны. На это стоит обратить внимание так как это может привести к некорректному обучению что может пагубно повлиять на всю цепочку обучения.

Листинг 2 – Содержание файла dataset1.yaml

```
train: ../train/images
val: ../valid/images
test: ../test/images

nc: 2
names: ['label', 'package']

roboflow:
  workspace: larry-motalavigne
  project: package-and-label-detector
  version: 3
  license: CC BY 4.0
```

Листинг 2 – Содержание файла dataset2.yaml

```
train: ../train/images
val: ../valid/images
test: ../test/images

nc: 1
names: ['0']

roboflow:
  workspace: ieeeroboticswsu
  project: box-detector-weefm
  version: 1
  license: CC BY 4.0
```

Еще стоит обратить внимания что в первом датасете класс label стоит под номером 0, класс package под номером 1, а во втором класс label отсутствует, а класс package имеет название «0» и находится он под номером 0. Это значит что помимо аннотации требуется еще поменять набор лейблов изображения находящихся в папках имеющих «../train/labels», «../valid/labels», «../test/labels».

Для корректного изменения классов координат в файлах из папок labels используется скрипт на языке Python из листинга 3

Листинг 3 – Скрипт для изменения индексов координат в файлах images.txt

```
import os
# На вход подается путь к директории изменений
def modify_files_in_directory(directory):
    for filename in os.listdir(directory):
        file_path = os.path.join(directory, filename)

        # Проверяем, является ли файл обычным файлом
        if os.path.isfile(file_path):
            with open(file_path, 'r') as file:
                lines = file.readlines()

            with open(file_path, 'w') as file:
                for line in lines:
                    # Меняем первый символ с '0' на '1', если он существует
                    if line.startswith('0'):
                        line = '1' + line[1:]
                    file.write(line)
    print("Успешно изменено в " + directory)
```

После редактирования классов координат в файла labels, закончим изменения yaml файла второго датасета

Листинг 4 – Обновленное содержание файла dataset2.yaml

```
train: ../train/images
val: ../valid/images
test: ../test/images

nc: 2
names: ['label', 'package']

roboflow:
  workspace: ieeeroboticswsu
  project: box-detector-weefm
  version: 1
  license: CC BY 4.0
```

Теперь можно производить последовательное дообучение с произвольным чередованием датасетов. Результаты детекции обученной модели изображено на рисунке

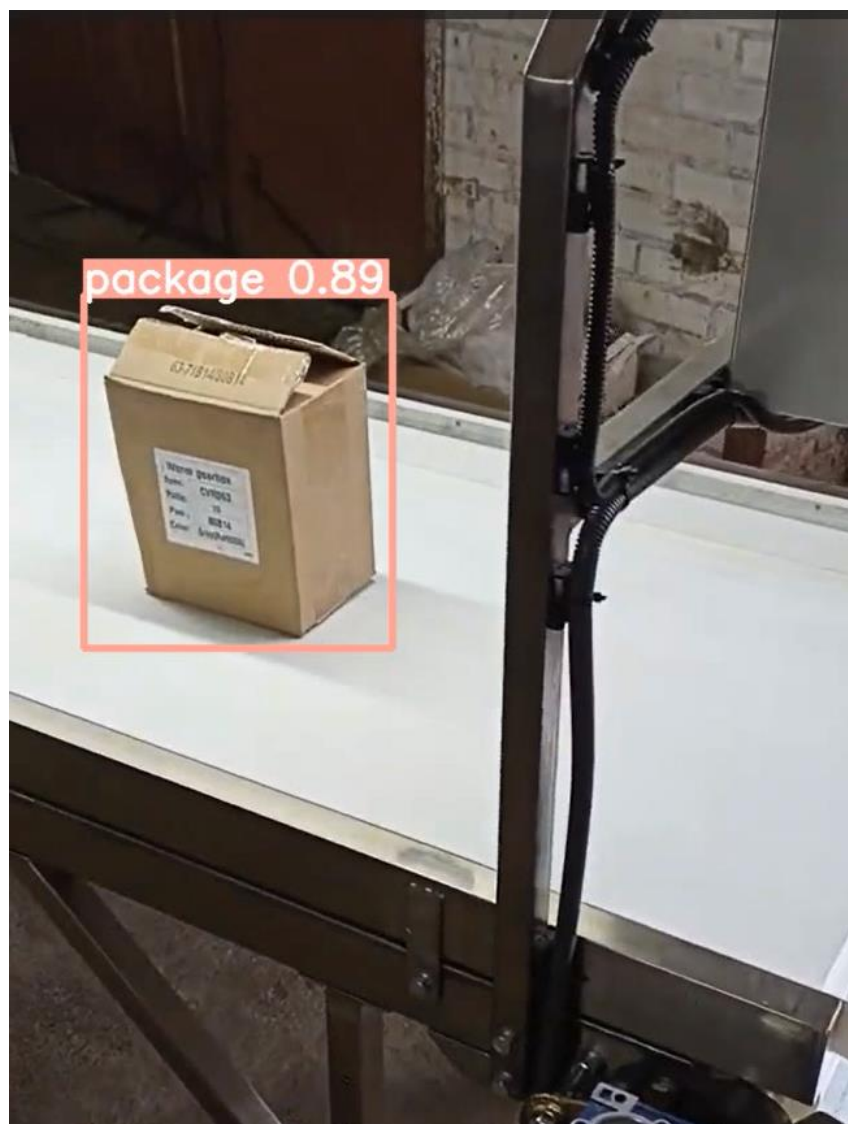


Рисунок 5 – Результаты детекции

2.2 Решение задачи сегментации

Сегментация изображений представляет собой процесс разделения цифрового изображения на несколько сегментов (набор пикселей, также называемых суперпикселями) для упрощения или изменения представления изображения для улучшения его анализа. Сегментация целевых объектов на изображении может существенно улучшить последующие задачи, такие как распознавание объектов,

трекинг и классификация, что, в свою очередь, способствует повышению эффективности автоматических систем управления на складах. Позволяет в дальнейшем определить габаритные характеристики груза.

В таблицах 2, 3 представлены метрики популярных моделей сегментации

Таблица 2 – Показатели производительности YOLOv8 для задач сегментации

Модель	Размер (пикселей)	mAPbox 50-95	mAPmask 50-95	Скорость CPU ONNX(мс)	Скорость A100 TensorRT(мс)	params(M)	FLOPs(B)
YOLOv8n-seg	640	36.7	30.5	96.1	1.21	3.4	12.6
YOLOv8s-seg	640	44.6	36.8	155.7	1.47	11.8	42.6
YOLOv8m-seg	640	49.9	40.8	317.0	2.18	27.3	110.2
YOLOv8l-seg	640	52.3	42.6	572.4	2.79	46.0	220.5
YOLOv8x-seg	640	53.4	43.4	712.1	4.02	71.8	344.1

Таблица 3 – Показатели производительности YOLOv8 и SAM

Модель	Размер	Параметры	Скорость (процессор)
Meta's SAM-b	358 МБ	94.7 М	51096 ms/im
MobileSAM	40,7 МБ	10.1 М	46122 ms/im
FastSAM-s с магистралью YOLOv8	23,7 МБ	11.8 М	115 ms/im
Ultralytics YOLOv8n-seg	6,7 МБ	3,4 М	59 ms/im

Стоит обратить внимание что при сравнении наиболее упрощенных и быстродействующих моделей, показатели YOLOv8n-seg многократно обгоняют своих оппонентов, например, размер в 53,4 раза меньше, параметры в 27,9 раз меньше, а время обработки одного образа в 866 раз быстрее

2.2.1 Применение модели SAM для сегментации объектов

Среди различных доступных подходов к сегментации объектов на изображениях одним из точных является применение модели Segment Anything Model (SAM), которая способна выделять почти любые объекты в различных условиях

освещения и фона. Эта гибкость особенно важна в складских условиях, где товары могут иметь разнообразные формы, размеры и располагаться в сложно структурированных средах. Модель SAM использует методы глубокого обучения, сверточные нейронные сети, что позволяет ей обучаться на основе аннотированных изображений и выделять семантические границы интересующих объектов. Данная модель может быть интегрирована с системами компьютерного зрения на складе для выполнения задач сортировки, инвентаризации и управления запасами.

Сегментация является одним из ключевых этапов в обработке изображений и компьютерном зрении и позволяет:

- Улучшить эффективность Обработки. Сегментация уменьшает объем обрабатываемых данных, выделяя важные объекты и их контуры, что делает последующий анализ более эффективным.
- Распознавание объектов. Она помогает в точном определении положения и формы различных объектов на изображении, что важно для их распознавания и классификации.
- Анализ изображений. Сегментация может использоваться для анализа медицинских изображений, аэрофотоснимков, исследований по биологии, метеорологии и многих других областей.
- В робототехнике сегментация используется для интерпретации сенсорных данных для облегчения процессов принятия решений.
- Изменение и улучшение изображений. Сегментация может использоваться для редактирования изображений, например, для изменения фона или для более детальной обработки определенных участков изображения.
- Трекинг объектов. В видеопотоках сегментация позволяет отслеживать движение объектов.
- В автономных системах, таких как беспилотные автомобили, сегментация помогает в интерпретации и понимании сцен в реальном времени для навигации и взаимодействия с окружающей средой.

Пример сегментированного изображения SAM представлен на рисунке 6

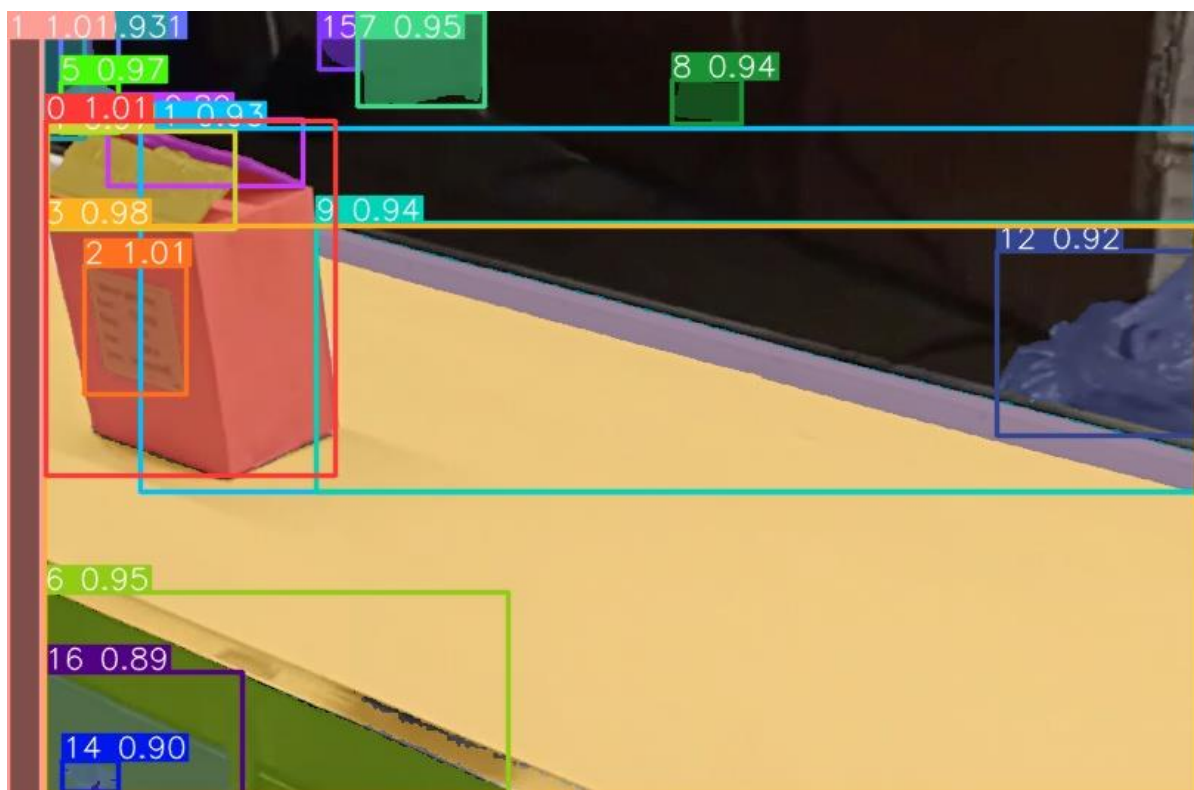


Рисунок 6 – Пример сегментации SAM

Также в процессе сегментации можно выделить и вырезать каждый слой, это позволяет в дальнейшем обрабатывать изображение по кускам, пример фрагмента изображен на рисунке 7, небольшой фрагмент более удобно обрабатывать, производить расчеты и разделять на более мелкие фрагменты, например для нахождения габаритных характеристик.



Рисунок 7 – Фрагмент сегментированного изображения

2.2.2 Создание набора данных сегментации, используя модель обнаружения

В процессе разработки систем компьютерного зрения часто возникает необходимость в создании наборов данных для специфических задач, таких как сегментация изображений. Однако, сбор и аннотация данных для сегментации может быть трудоемким процессом. Рассмотрим подход, который позволяет автоматизировать создание датасета для сегментации на основе существующего датасета, предназначенного для задачи детекции. Используя модель обнаружения объектов YOLOv8 и модель сегментации SAM (Segment Anything Model), предлагается метод автоматической аннотации, который значительно упрощает и ускоряет процесс подготовки данных.

Процесс начинается с использования модели YOLOv8, обученной на датасете для детекции. Эта модель способна эффективно идентифицировать и локализовать объекты на изображениях, предоставляя ограничивающие рамки вокруг каждого обнаруженного объекта. Однако, для задачи сегментации требуются более детальные аннотации, включающие пиксельные маски, которые точно определяют границы объектов. Здесь на помощь приходит модель SAM,

которая способна генерировать такие маски на основе областей, выделенных YOLOv8.

Используя функцию `auto_annotate` из библиотеки Ultralytics, процесс преобразования датасета для детекции в датасет для сегментации автоматизируется. Эта функция принимает путь к изображениям (`path/to/images`), модель YOLOv8 (`yolov8x.pt`), и модель SAM (`sam_b.pt`). Алгоритм проходит через каждое изображение в исходном датасете, использует YOLOv8 для обнаружения объектов, а затем применяет SAM для создания точных пиксельных масок для каждого обнаруженного объекта. Результатом является датасет, содержащий изображения с соответствующими масками сегментации, который может быть использован для обучения и оценки моделей сегментации.

На рисунке 8 изображен пример сегментации для модели YOLOv8l-seg

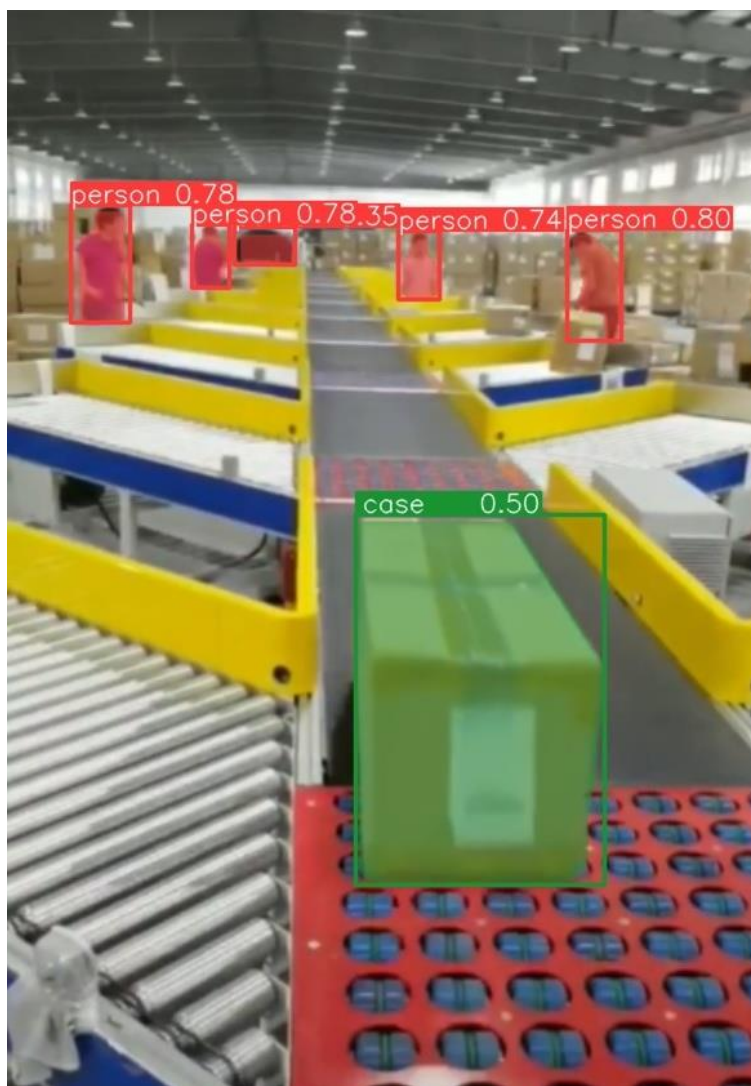


Рисунок 8 – Пример сегментации модели YOLOv8-seg

Такой подход значительно сокращает время и усилия, необходимые для подготовки датасетов сегментации, и предлагает эффективный способ повторного использования и расширения существующих датасетов для детекции. Кроме того, он позволяет создавать высококачественные наборы данных сегментации, которые могут быть адаптированы к различным условиям и требованиям конкретных приложений компьютерного зрения.

ЗАКЛЮЧЕНИЕ

В рамках научно-исследовательской работы были изучены и адаптированы методы повышения точности детектирования объектов на изображениях в сфере складской логистики. Рассмотрены улучшения предварительной обработки изображений, оптимизации гиперпараметров и функций потерь, а также применение синтетических и дополненных данных позволило углубить понимание влияния этих аспектов на эффективность моделей глубокого обучения.

Особое внимание было уделено применению модели YOLOv8 для задач обнаружения объектов, что демонстрирует значительные перспективы в повышении скорости и точности детекции в реальных условиях работы склада. Разработка методологии дообучения пользовательской модели и стандартизация пользовательских датасетов под одну аннотацию позволили не только адаптировать модели под конкретные требования, но и значительно повысить их обрабатывающую способность.

В области сегментации было установлено, что использование модели SAM в тандеме с YOLOv8 является эффективным решением для точного выделения объектов на изображениях, что открывает новые возможности для улучшения процессов инвентаризации и управления на складах. Также была рассмотрена возможность создания набора данных сегментации с использованием модели обнаружения, что обеспечивает более высокую точность и универсальность в разработке систем компьютерного зрения.

В заключение, достигнутые результаты подчеркивают значимость интеграции современных технологий компьютерного зрения в логистические процессы и открывают пути для дальнейших исследований в этом направлении. Предложенные подходы и разработанные методики способствуют повышению эффективности, надежности и точности автоматизированных систем складской логистики, что имеет важное значение для динамично развивающейся сферы современной индустрии.