

Дмитриев Андрей Глебович,
магистрант факультета информационных технологий, Московский
Политехнический университет, г. Москва

СЕГМЕНТАЦИЯ ЭКЗЕМПЛЯРОВ ТОВАРНЫХ ОБЪЕКТОВ НА ПОЛОЧНЫХ ИЗОБРАЖЕНИЯХ КАК ЭТАП ПОДГОТОВКИ К ИДЕНТИФИКАЦИИ ТОВАРНЫХ ПОЗИЦИЙ

Аннотация

В статье рассматривается сегментация экземпляров товарных объектов на полочных изображениях как подготовительный этап перед идентификацией товарных позиций. Основная проблема заключается в том, что фрагмент, сформированный по прямоугольной рамке, часто содержит фон, соседние товары, ценники и элементы полки. Эти области могут ухудшать последующее сопоставление с товарной галереей. Для проверки подхода используется YOLO-Seg и программный комплекс ShelfVision. Экспериментальная часть включает сегментационный набор из 300 изображений, 847 объектов и 60 категорий, оценку качества масок, сравнение фрагментов по рамке и по маске, а также дополнительную проверку влияния подготовленных фрагментов на демонстрационный контур SKU-сопоставления. На тестовой выборке получены $\text{mask precision} = 0,7000$, $\text{mask recall} = 0,9032$, $\text{mask F1} = 0,7887$, $\text{mean mask IoU} = 0,9336$ и $\text{APmask50-95} = 0,6481$. Сравнение 160 объектов показало, что прямоугольный фрагмент в среднем содержит 45,0035% области вне маски товара, тогда как масочное представление нейтрализует эту область. Дополнительная проверка полного контура на 300 полочных изображениях показала формирование 200 демонстрационных SKU, 1360 эталонов галереи и сопоставление 17101 проверяемого объекта со статусами *matched*, *matched_uncertain* и *unknown*. Полученные результаты подтверждают, что сегментация экземпляров объектов является обоснованным промежуточным этапом между локализацией товара и последующей идентификацией товарных позиций.

Ключевые слова: сегментация экземпляров объектов, товарная полка, YOLO-Seg, вырезанный фрагмент товара, маска объекта, SKU-галерея, идентификация товарных позиций, компьютерное зрение.

Annotation

The article considers the segmentation of instances of commodity objects on shelf images as a preparatory stage before the identification of commodity items. The main problem is that a fragment formed by a rectangular frame often contains a background, neighboring products, price tags, and shelf elements. These areas may worsen the subsequent comparison with the product gallery. YOLO-Seg and the ShelfVision software package are used to test the approach. The experimental part includes a segmentation set of 300 images, 847 objects and 60 categories, an assessment of the quality of masks, a comparison of fragments by frame and mask, as well as an additional check of the effect of prepared fragments on the demonstration contour of SKU matching. The test sample yielded mask precision = 0.7000, mask recall = 0.9032, mask F1 = 0.7887, mean mask IoU = 0.9336, and APmask50-95 = 0.6481. A comparison of 160 objects showed that the rectangular fragment contains, on average, 45.0035% of the area outside the product mask, while the masked representation neutralizes this area. An additional check of the full contour on 300 shelf images showed the formation of 200 demo SKUs, 1360 gallery standards and the comparison of 17101 checked objects with the statuses matched, matched_uncertain and unknown. The results obtained confirm that the segmentation of object instances is a reasonable intermediate step between the localization of goods and the subsequent identification of product items.

Keywords: segmentation of object instances, commodity shelf, YOLO-Seg, cut-out fragment of a product, object mask, SKU gallery, identification of product items, computer vision.

Введение

Автоматический анализ изображений товарных полок является актуальной задачей для розничной торговли. От качества выкладки зависят контроль наличия товаров, своевременное пополнение запасов, проверка видимости продукции и дальнейшая сверка с правилами размещения. Однако полочная

сцена заметно сложнее обычного изображения одного объекта: товары расположены близко друг к другу, частично перекрываются, имеют похожую форму и часто отличаются только мелкими деталями упаковки. Дополнительные сложности создают блики, тени, ценники, глянцевые поверхности и изменения ракурса съемки. В исследованиях, связанных с печатной информацией, покрытиями и маркировкой, отмечается, что внешний вид объекта может зависеть от материала, освещения и условий наблюдения [1-5]. Для товарной полки это означает, что даже один и тот же продукт может выглядеть по-разному на разных изображениях.

Большинство практических систем сначала решает задачу детекции: модель находит товар и строит вокруг него прямоугольную рамку. Такой подход удобен для подсчета объектов и первичной локализации, но для последующей идентификации товарных позиций его часто недостаточно. Если объект вырезается только по рамке, во фрагмент попадают не только пиксели товара, но и фон, соседние упаковки, ценники или элементы стеллажа. Для системы визуального сопоставления это становится источником шумовых признаков, особенно когда нужно различать похожие SKU.

Сегментация экземпляров объектов решает эту проблему точнее. Модель формирует не только рамку, но и пиксельную маску отдельного объекта. Благодаря этому можно сохранить область товара и нейтрализовать пространство вне маски. Такой фрагмент потенциально лучше подходит для дальнейшего сопоставления с товарной галереей. Классическим примером подхода к сегментации экземпляров объектов является Mask R-CNN [6], а современные модели, включая YOLO-Seg, развивают эту логику в сторону более удобного прикладного применения [12].

Цель статьи - оценить применимость сегментации экземпляров товарных объектов как подготовительного этапа перед идентификацией товарных позиций. В отличие от работ, где основное внимание уделяется только детекции или только классификации, в статье рассматривается промежуточный слой: качество подготовки вырезанного фрагмента товара. Для этого оцениваются метрики масок, сравниваются фрагменты по рамке и по маске, а также приводятся результаты интеграции подготовленных фрагментов в демонстрационный контур SKU-сопоставления ShelfVision.

1. Анализ существующих подходов

1.1. Детекция и сегментация экземпляров объектов

В задачах компьютерного зрения для анализа товарных полок используются два близких, но разных подхода: детекция объектов и сегментация экземпляров объектов. Детекция определяет расположение объекта через прямоугольную рамку. Она подходит для подсчета товаров и грубой локализации, однако прямоугольная рамка не учитывает реальную форму упаковки. В плотной сцене внутри рамки часто оказываются соседние объекты или фон. Для задач последующей идентификации это ограничение принципиально.

Сегментация экземпляров объектов формирует отдельную маску для каждого объекта. Это позволяет отделять товар от фона и соседних упаковок, а не только ограничивать его прямоугольником. Mask R-CNN показала эффективность такого подхода за счет добавления ветви предсказания маски к модели детекции [6]. Более поздние методы, такие как YOLACT, SOLO, Mask2Former и Segment Anything, развили эту идею в сторону ускорения, универсальности и работы с различными постановками сегментации [8-11].

Для практических систем анализа полок важен компромисс между качеством и скоростью. Полочная сцена может содержать десятки и сотни объектов, поэтому модель должна выделять товары достаточно точно и при этом быть пригодной для пакетной обработки изображений. YOLO-Seg интересна именно как прикладной вариант: она совмещает локализацию объектов и построение масок, а результаты можно использовать для дальнейшего извлечения вырезанных фрагментов товаров [12].

1.2. Полочные изображения и идентификация товарных позиций

Анализ товарных полок относится к отдельному классу прикладных задач компьютерного зрения. Известные наборы данных подтверждают специфику этой области: MVTeс D2S содержит плотную сегментационную разметку объектов супермаркета [13], а SKU-110K ориентирован на точную детекцию большого количества товаров в плотно заполненных сценах [14]. Эти наборы показывают, что полочная сцена не сводится к изображению одного объекта: важны плотность размещения, повторяемость упаковок, частичные перекрытия и визуальная похожесть товаров.

Идентификация товарной позиции является следующим уровнем после локализации и сегментации. Если детекция отвечает на вопрос, где расположен товар, то идентификация должна определить, какой именно товар изображен. Для этого используется классификационная модель, товарная галерея или система визуального поиска. В работах по распознаванию продукции на полках подчеркивается, что качество входного изображения товара существенно влияет на последующее распознавание [15-20]. Если фрагмент содержит соседние товары или ценник, система сопоставления анализирует не только товар, но и случайные элементы сцены.

Сегментация экземпляров объектов может рассматриваться как промежуточный этап между обнаружением и идентификацией. Она не решает задачу распознавания SKU полностью, но формирует более чистое входное представление объекта. Это особенно важно для товаров с похожим дизайном упаковки, где различия между SKU могут быть связаны с небольшим текстом, цветовой полосой, объемом или изображением вкуса.

2. Постановка задачи и методика эксперимента

В статье проверяется гипотеза о том, что фрагмент товара, сформированный по маске объекта, лучше подходит для подготовки к идентификации товарных позиций, чем фрагмент по прямоугольной рамке. Под улучшением понимается не рост точности распознавания SKU как таковой, а уменьшение доли визуальной области, не относящейся к товару. Полная оценка точности идентификации требует отдельной эталонной SKU-разметки, поэтому в данной статье результаты сопоставления с демонстрационной галереей трактуются как проверка работоспособности контура, а не как строгая ассигасу-оценка.

Практическая часть выполнена в программном комплексе ShelfVision. Комплекс поддерживает подготовку данных, применение модельных веток, сохранение предсказаний, извлечение вырезанных фрагментов товаров, формирование SKU-галереи, сопоставление объектов с эталонами и подготовку отчетных файлов. В статье используются результаты трех групп экспериментов: оценка YOLO-Seg на сегментационном наборе, сравнение фрагментов по рамке и по маске, а также дополнительная проверка полного контура SKU-сопоставления.

2.1. Экспериментальный набор данных

Экспериментальная часть по сегментации выполнена на наборе изображений товарных объектов с разметкой в формате COCO. Формат COCO удобен тем, что позволяет хранить не только прямоугольные рамки, но и сегментационную разметку объектов [7]. Наличие масок необходимо для оценки качества контурного выделения и сравнения фрагментов по рамке и по маске.

Набор включает 300 изображений, 847 объектов и 60 категорий. Он разделен на обучающую, валидационную и тестовую части. Характеристики набора представлены в таблице 1.

Таблица 1 - Характеристики сегментационного набора данных

Часть набора	Изображений	Объектов	Масок	Категорий	Среднее объектов
Обучающая	210	629	629	60	2,9952
Валидационная	45	94	94	60	2,0889
Тестовая	45	124	124	60	2,7556
Всего	300	847	847	60	2,8233

2.2. Модель и показатели качества

В качестве основной модели сегментации используется YOLOv8s-seg. Модель формирует прямоугольную рамку, оценку уверенности, класс объекта и маску. Это позволяет использовать одни и те же предсказания для двух задач: оценки качества сегментации и подготовки фрагментов товаров для дальнейшего сопоставления.

Для оценки качества сегментации используются mask precision, mask recall, mask F1, mean mask IoU, APmask50, APmask75 и APmask50-95. Эти показатели позволяют оценить как факт нахождения объекта, так и точность совпадения предсказанной маски с эталонной. Также фиксируются метрики рамок, поскольку BBox остается базовым способом локализации и используется для сравнения с масочным вариантом подготовки фрагмента.

2.3. Выбор параметров локализации

Для проверки влияния параметров первичной локализации была выполнена серия абляционных экспериментов YOLO. Она нужна для того, чтобы не выбирать модельную конфигурацию произвольно. В экспериментах изменялись размер модели, размер входного изображения, параметры обучения, оптимизатор и аугментации. Основным критерием сравнения выступало mAP50-95 при дополнительном учете precision, recall, mAP50 и времени обучения.

Таблица 2 - Результаты абляционного анализа YOLO

Эксперимент	Изменяемый параметр	P	R	mAP50	mAP50-95	Время, мин
BASE_yolov8s_img1024	Базовая конфигурация	0,8611	0,8174	0,8816	0,5561	62,12
E01_yolov8n	Более легкая модель	0,8671	0,7781	0,8663	0,5494	48,24
E02_yolov8m	Более крупная модель	0,8587	0,8112	0,8769	0,5542	119,61
E03_imgsz_640	Размер изображения 640	0,8694	0,8153	0,8878	0,5631	44,15
E04_imgsz_1280	Размер изображения 1280	0,8680	0,7927	0,8703	0,5484	86,14

E05_lr_0p005	Скорость обучения 0,005	0,8611	0,8174	0,8816	0,5561	63,96
E06_wd_0p001	Регуляризация 0,001	0,8629	0,8009	0,8720	0,5496	65,05
E07_AdamW	Оптимизатор AdamW	0,8513	0,7755	0,8605	0,5434	63,79
E08_mosaic_0	Отключение mosaic	0,8690	0,7984	0,8779	0,5538	61,16
E09_mixup_0p2	MixUp 0,2	0,8646	0,7834	0,8662	0,5441	64,55
E10_close_mosaic	Отключение mosaic в конце	0,8611	0,8174	0,8816	0,5561	63,38

Лучшей конфигурацией в этой серии стала E03_imgsz_640: она показала максимальное $mAP_{50-95} = 0,5631$ при $precision = 0,8694$, $recall = 0,8153$ и времени обучения 44,15 минуты. Этот результат не означает универсальное превосходство конфигурации для любых задач, но обосновывает ее использование в дальнейшем экспериментальном контуре ShelfVision.

2.4. Сравнение фрагментов по рамке и по маске

Ключевой эксперимент статьи связан со сравнением двух вариантов подготовки фрагмента товара. Первый вариант - фрагмент по прямоугольной рамке. В этом случае из изображения вырезается область BBox, которая

неизбежно содержит часть фона. Второй вариант - фрагмент по маске. Внутри ВВох сохраняются пиксели объекта, а область вне маски заменяется нейтральным белым фоном. Для количественного сравнения рассчитываются чистота объекта, доля фона и доля визуальной области, нейтрализуемой маской.

3. Результаты

3.1. Качество модели YOLO-Seg

Обучение YOLOv8s-seg выполнялось в течение 30 эпох. Итоговые показатели обучения и дополнительной оценки на тестовой выборке представлены в таблице 3. Значения показывают, что модель способна не только локализовать объекты, но и формировать маски, пригодные для дальнейшего анализа.

Таблица 3 - Метрики YOLO-Seg и дополнительной оценки масок

Группа оценки	Показатель	Значение
Обучение	Точность рамок	0,88011
Обучение	Полнота рамок	0,58831
Обучение	mAP50 рамок	0,75064
Обучение	mAP50-95 рамок	0,71896
Обучение	mAP50 масок	0,75064
Обучение	mAP50-95 масок	0,71763
Тестовая оценка	Истинно положительные маски	112
Тестовая оценка	Ложноположительные маски	48
Тестовая оценка	Пропущенные объекты	12

Тестовая оценка	mask precision	0,7000
Тестовая оценка	mask recall	0,9032
Тестовая оценка	mask F1	0,7887
Тестовая оценка	mean mask IoU	0,9336
Тестовая оценка	APmask50-95	0,6481

Особое значение имеет $\text{mean mask IoU} = 0,9336$. Этот показатель означает, что в случаях корректного сопоставления предсказанные маски хорошо совпадают с эталонной формой объекта. При этом mask precision ниже mask recall, что указывает на наличие ложноположительных предсказаний. Для практического контура это означает необходимость фильтрации по уверенности, размеру объекта или дополнительным правилам.

3.2. Чистота фрагмента по рамке и по маске

Сравнение фрагментов выполнено на 160 объектах с валидными масками. Для каждого объекта рассчитывались площадь прямоугольного фрагмента, площадь маски, доля объекта внутри рамки и доля визуального фона. Результаты представлены в таблице 4.

Таблица 4 - Сравнение фрагмента по рамке и по маске

Метод	Объектов	Средняя площадь фрагмент а	Средняя площадь объекта	Чистота объекта	Доля фона	Удалено пикселей
По рамке	160	288930,17 5	150251,70 62	0,549965	0,450035	0
По маске	160	288930,17 5	150251,70 62	1,000000	0,000000	22 188 555

Обычный прямоугольный фрагмент в среднем содержит только 54,9965% области самого товара. Остальные 45,0035% приходятся на фон, соседние объекты или элементы сцены. При использовании маски эта область нейтрализуется. Суммарно по 160 объектам было удалено или заменено нейтральным фоном 22 188 555 пикселей лишней визуальной области. Это количественно подтверждает, что масочное выделение улучшает качество входного изображения перед потенциальным SKU-сопоставлением.

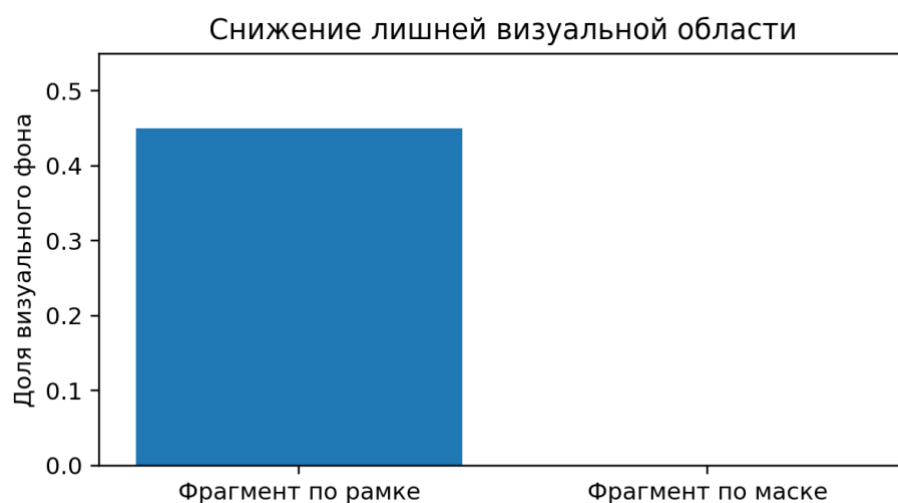


Рисунок 1 - Доля визуального фона при подготовке фрагмента товара

3.3. Интеграция с демонстрационным контуром SKU-сопоставления

Для проверки связи подготовленных фрагментов с дальнейшей идентификацией использовался демонстрационный контур ShelfVision. Он не является промышленной оценкой точности SKU-распознавания, но позволяет проверить, как система работает с галерейной и проверяемой частями данных. В этой части использовалось 300 изображений: 160 изображений для формирования галерейной части и 140 изображений для проверяемой части. На галерейной части было найдено 18855 объектов, на проверяемой - 17101 объект. Были сформированы 200 демонстрационных SKU и 1360 эталонов галереи.

Дополнительно была проведена ночная серия из 16 запусков полного контура. В ней изменялись параметры формирования галереи, дедупликации, числа эталонов и правил сопоставления. Лучшие варианты приведены в таблице 5. Показатель доли сопоставленных объектов не является точностью по реальным SKU: он показывает долю объектов, для которых в демонстрационной галерее найден кандидат.

Таблица 5 - Лучшие конфигурации демонстрационного SKU-сопоставления

Конфигурация	Доля сопоставленных	unknown	Средняя оценка сходства	Объекты	SKU	Эталоны
09_dedup082	0,9081	0,0919	0,7725	17101	150	1017
08_dedup086	0,9041	0,0959	0,7693	17101	200	1067
16_ref20	0,8790	0,1210	0,7621	17101	150	1163
07_gallery120	0,8698	0,1302	0,7552	17101	150	775
10_dedup090	0,8329	0,1671	0,7352	17101	150	497

Лучшим общим вариантом стала конфигурация 09 с порогом дедупликации 0,82. Она дала долю сопоставленных объектов 0,9081 и долю unknown 0,0919. Конфигурация 08 уступила незначительно, но является более консервативной: она использует более строгую дедупликацию 0,86 и большее число демонстрационных SKU. Эти результаты показывают, что качество галереи и правила отбора эталонов существенно влияют на дальнейшее сопоставление.

3.4. Итоговая проверка полного контура

Итоговая проверка полного контура ShelfVision была выполнена на 300 изображениях. После формирования галерейной и проверяемой частей 17101 проверяемый объект был сопоставлен с демонстрационной SKU-галереей. Распределение статусов и сводные показатели представлены в таблице 6.

Таблица 6 - Итоговые результаты полного контура ShelfVision

Показатель	Значение	Интерпретация
------------	----------	---------------

Изображений всего	300	Объем итоговой проверки
Галерейная часть	160 изображений, 18855 объектов	Основа для демонстрационной SKU-галереи
Проверяемая часть	140 изображений, 17101 объект	Объекты для SKU-сопоставления
Демонстрационные SKU	200	Количество SKU в галерее
Эталоны галереи	1360	Количество эталонных фрагментов
matched	8473	Уверенные сопоставления
matched_uncertain	7496	Спорные сопоставления
unknown	1132	Объекты без надежного совпадения
Доля объектов с назначенным кандидатом	0,9338	Не является точностью реального SKU-распознавания
Средняя оценка сходства	0,7848	Средняя визуальная близость к кандидатам
Средний margin	0,0438	Разница между ближайшими кандидатами

Полученные значения показывают, что система способна сопоставить большую часть проверяемых объектов с демонстрационной галереей. При этом высокая доля matched_uncertain указывает на значительное число визуально

близких кандидатов. Для полочной сцены это ожидаемо: похожие упаковки, блики, ценники и частичные перекрытия могут делать различие между SKU неустойчивым. Поэтому статус `matched_uncertain` является важной частью безопасной логики системы.

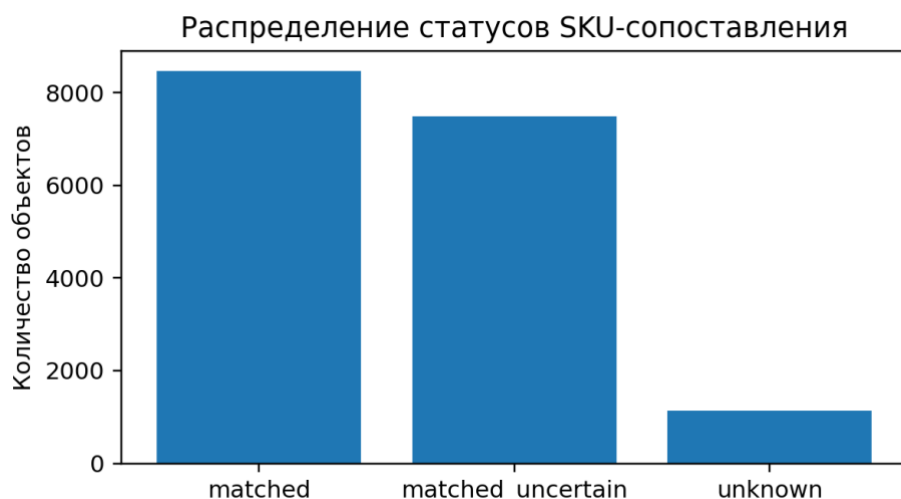


Рисунок 2 - Распределение статусов SKU-сопоставления

Отдельно был проанализирован порог сходства. При снижении порога система назначает кандидатов большему числу объектов, но становится менее осторожной. При повышении порога растет доля `unknown`. Результаты приведены в таблице 7 и на рисунке 3.

Таблица 7 - Влияние порога сходства на распределение результатов

Порог	С назначенным кандидатом	<code>unknown</code>	Доля кандидатом	Доля <code>unknown</code>
0,55	16810	291	0,9830	0,0170
0,60	16630	471	0,9725	0,0275
0,65	15969	1132	0,9338	0,0662
0,70	14412	2689	0,8428	0,1572
0,75	11605	5496	0,6786	0,3214

0,80	7913	9188	0,4627	0,5373
------	------	------	--------	--------

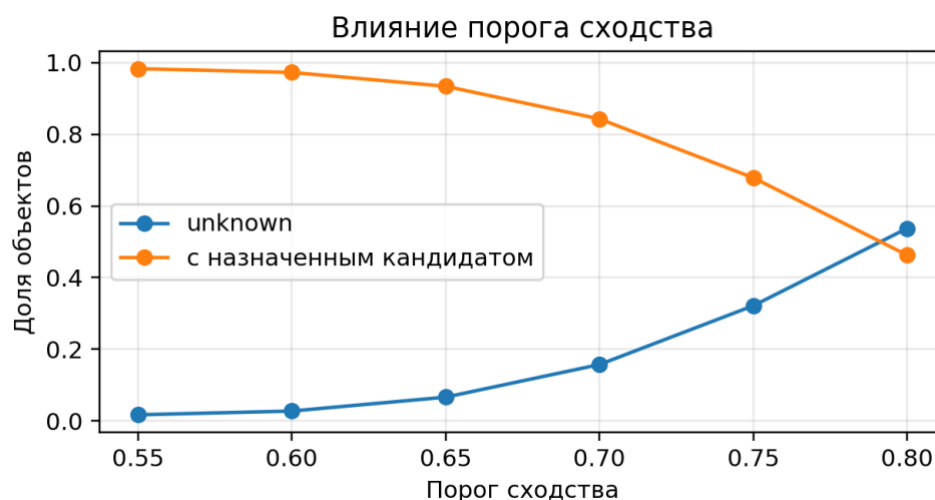


Рисунок 3 - Влияние порога сходства на долю объектов unknown

Рабочий порог 0,65 дает компромиссный режим: доля объектов с назначенным кандидатом составляет 0,9338, а доля unknown - 0,0662. Повышение порога до 0,80 увеличивает долю unknown до 0,5373. Это подтверждает, что порог сходства должен выбираться с учетом сценария: для полуавтоматической проверки можно допускать больше кандидатов, а для более строгой автоматической обработки следует повышать осторожность и выделять больше объектов для ручной проверки.

4. Обсуждение

Полученные результаты подтверждают, что сегментация экземпляров объектов полезна не только как способ более точной локализации, но и как этап подготовки изображения к идентификации товарных позиций. Главный численный результат состоит в том, что почти 45% площади прямоугольного фрагмента в среднем не относится к самому товару. Для визуального сопоставления это существенная величина: признаки фона, соседних упаковок и ценников могут смешиваться с признаками товара.

Масочное представление уменьшает этот риск. Оно не делает идентификацию автоматически точной, но повышает качество входного материала. Именно поэтому сегментацию следует рассматривать как

промежуточный слой между детекцией и SKU-сопоставлением. На уровне полного контура ShelfVision это проявляется в необходимости сохранять не только статус сопоставления, но и диагностические данные: вырезанный фрагмент, ближайших кандидатов, оценку сходства и `margin`.

Результаты демонстрационного контура показывают, что система способна обрабатывать большое количество объектов и сопоставлять их с галереей. Однако высокая доля `matched_uncertain` показывает, что задача остается сложной. Основные причины связаны с визуальной похожестью упаковок, качеством эталонов, неполнотой галереи и сложными условиями съемки. Поэтому корректная интерпретация результатов требует осторожности: доля объектов с назначенным кандидатом не является точностью реального SKU-распознавания без эталонной разметки.

Практический вывод состоит в том, что маски следует использовать не изолированно, а вместе с системой отчетности. Если объект получил спорный статус, необходимо иметь возможность открыть исходное изображение, фрагмент по рамке, фрагмент по маске, ближайших кандидатов и их оценки. Такая проверяемость делает систему пригодной для дальнейшего развития: можно улучшать галерею, фильтровать плохие фрагменты, добавлять OCR и собирать эталонную SKU-разметку.

5. Ограничения и дальнейшее развитие

Первое ограничение исследования связано с тем, что сегментационный эксперимент оценивает подготовительный этап, а не полную точность идентификации товарных позиций. Для расчета `top-1 accuracy` и `top-k accuracy` требуется эталонная SKU-разметка, где каждому проверяемому объекту заранее сопоставлен правильный SKU. В текущей работе результаты демонстрационного сопоставления показывают работоспособность контура, но не заменяют строгую оценку качества распознавания.

Второе ограничение связано с объемом сегментационного набора. 300 изображений и 847 объектов достаточны для проверки исследовательской

гипотезы, но для промышленного применения нужно больше данных: разные категории товаров, условия освещения, ракурсы, плотность выкладки и варианты перекрытий. Третье ограничение связано с использованием одной основной сегментационной архитектуры YOLOv8s-seg. В дальнейшем целесообразно сравнить ее с YOLOv8n-seg, YOLOv8m-seg, Mask R-CNN, YOLACT, Mask2Former и Segment Anything [6, 8-12].

Дальнейшее развитие должно включать сбор товарной галереи, прямое сравнение идентификации по фрагментам по рамке и по маске, анализ устойчивости к бликам и размытию, подключение OCR и ручную проверку спорных случаев. В полном прикладном контуре сегментация может также использоваться для анализа видимой площади товара, перекрытий и качества заполненности полки.

Заключение

В статье рассмотрена сегментация экземпляров товарных объектов на полочных изображениях как этап подготовки к идентификации товарных позиций. Показано, что прямоугольная рамка остается полезной для первичной локализации, но при формировании фрагмента товара часто включает значительную долю визуальной области, не относящейся к объекту. Для задач сопоставления с товарной галереей такая область может ухудшать устойчивость визуальных признаков.

Эксперимент с YOLOv8s-seg подтвердил применимость масочного выделения товарных объектов. На тестовой выборке получены $\text{mask recall} = 0,9032$, $\text{mask F1} = 0,7887$ и $\text{mean mask IoU} = 0,9336$. Сравнение 160 объектов показало, что прямоугольный фрагмент содержит в среднем 45,0035% области вне маски товара. При использовании фрагмента по маске эта область нейтрализуется; суммарно по всем объектам было удалено 22 188 555 пикселей лишней визуальной информации.

Дополнительные эксперименты полного контура ShelfVision показали, что подготовленные фрагменты могут использоваться в демонстрационном SKU-

сопоставлении. На 300 изображениях было сформировано 200 демонстрационных SKU и 1360 эталонов галереи; из 17101 проверяемого объекта 15969 получили назначенного кандидата. При этом результат корректно трактуется как проверка демонстрационного контура, а не как точность промышленного распознавания SKU. Полученные данные подтверждают, что сегментация экземпляров объектов является обоснованным промежуточным этапом между локализацией товара и последующей идентификацией товарных позиций.

Список источников

1. Верещагин В. Ю. Исследование параметров и разработка рекомендаций по воспроизведению печатной информации в устройствах с электронной бумагой : дис. ... канд. техн. наук. — Москва, 2016.
2. Верещагин В. Ю., Погиба А. Ю., Кондратов А. П. Плеохроизм бесцветных полимерных покрытий на светоотражающей поверхности // Ползуновский вестник. — 2025. — № 2. — С. 214-219. — DOI: 10.25712/ASTU.2072-8921.2025.02.033.
3. Погиба А. Ю., Верещагин В. Ю., Род В. А., Кондратов А. П. Многослойные полимерные оболочки для скрытой маркировки зеркальных изделий // Вестник Воронежского государственного университета инженерных технологий. — 2025. — Т. 87, № 2(104). — С. 167-176.
4. Kondratov A. P., Vereshchagin V. Y., Pogiba A. Y., Volinsky A. A. Transparent Multilayer Polymer Films for Hidden Marking of Reflective Containers // Optics Letters. — 2026. — Vol. 51, Issue 5. — P. 1215-1218.
5. Пухова Е. А., Верещагин В. Ю. Принтмедиа технологии : учебно-методическое пособие. — М. : Московский Политех, 2021.
6. He K., Gkioxari G., Dollar P., Girshick R. Mask R-CNN // Proceedings of the IEEE International Conference on Computer Vision. — 2017. — P. 2961-2969.
7. Lin T.-Y., Maire M., Belongie S. et al. Microsoft COCO: Common Objects in Context // European Conference on Computer Vision. — 2014. — P. 740-755.

8. Bolya D., Zhou C., Xiao F., Lee Y. J. YOLACT: Real-time Instance Segmentation // IEEE International Conference on Computer Vision. — 2019. — P. 9157-9166.
9. Wang X., Zhang R., Kong T., Li L., Shen C. SOLO: Segmenting Objects by Locations // European Conference on Computer Vision. — 2020.
10. Cheng B., Misra I., Schwing A. G., Kirillov A., Girdhar R. Masked-attention Mask Transformer for Universal Image Segmentation // CVPR. — 2022.
11. Kirillov A., Mintun E., Ravi N. et al. Segment Anything // IEEE/CVF International Conference on Computer Vision. — 2023.
12. Ultralytics. YOLO Documentation: Segmentation Tasks [Электронный ресурс]. — URL: <https://docs.ultralytics.com/tasks/segment/>
13. Follmann P., König R., Härtinger P., Klostermann M., Böttger T. MVTec D2S: Densely Segmented Supermarket Dataset // European Conference on Computer Vision Workshops. — 2018.
14. Goldman E., Herzig R., Eisenschtat A., Goldberger J., Hassner T. Precise Detection in Densely Packed Scenes // CVPR. — 2019.
15. Tonioni A., Serra E., Di Stefano L. A Deep Learning Pipeline for Product Recognition on Store Shelves // IEEE International Conference on Image Processing Applications and Systems. — 2018.
16. George M., Floerkemeier C. Recognizing Products: A Per-exemplar Multi-label Image Classification Approach // European Conference on Computer Vision Workshops. — 2014.
17. Wei X.-S., Cui Q., Yang L., Wang P., Liu L. RPC: A Large-Scale Retail Product Checkout Dataset // arXiv. — 2019.
18. Chen F., Zhang H., Li Z. et al. UniTail: Detecting, Reading, and Matching in Retail Scene // European Conference on Computer Vision. — 2022.
19. Melek C. G., Battini Sönmez E., Varli S. Datasets and Methods of Product Recognition on Grocery Shelf Images Using Computer Vision and Machine Learning Approaches: An Exhaustive Literature Review // Engineering Applications of Artificial Intelligence. — 2024. — Vol. 133.

20. Pietrini R., Paolanti M., Mancini A., Frontoni E., Zingaretti P. Shelf Management: A Deep Learning-Based System for Shelf Visual Monitoring // Expert Systems with Applications. — 2024. — Vol. 246.
21. Ren S., He K., Girshick R., Sun J. Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks // Advances in Neural Information Processing Systems. — 2015.
22. Rezatofighi H., Tsoi N., Gwak J., Sadeghian A., Reid I., Savarese S. Generalized Intersection over Union: A Metric and A Loss for Bounding Box Regression // CVPR. — 2019.
23. Everingham M., Van Gool L., Williams C. K. I., Winn J., Zisserman A. The Pascal Visual Object Classes Challenge: A Retrospective // International Journal of Computer Vision. — 2015. — Vol. 111. — P. 98-136.
24. Solovyev R., Wang W., Gabruseva T. Weighted Boxes Fusion: Ensembling Boxes from Different Object Detection Models // Image and Vision Computing. — 2021. — Vol. 107.
25. Liu J., Wang Z., Ma L., Fang C., Bai T., Zhang X., Liu J., Chen Z. Benchmarking Object Detection Robustness against Real-World Corruptions // International Journal of Computer Vision. — 2024.