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

Верещагин Владислав Юрьевич,
кандидат технических наук, доцент, Московский Политехнический университет, город Москва

МНОГОКРИТЕРИАЛЬНАЯ ОЦЕНКА МОДЕЛЕЙ ОБНАРУЖЕНИЯ ТОВАРОВ НА ПОЛКАХ РОЗНИЧНЫХ МАГАЗИНОВ С УЧЁТОМ НАРЕЗКИ ИЗОБРАЖЕНИЙ НА ФРАГМЕНТЫ И УХУДШЕНИЯ КАЧЕСТВА ИЗОБРАЖЕНИЙ

Аннотация. В статье рассматривается задача выбора модели обнаружения товаров для анализа полочных сцен розничных магазинов. Объединены сравнение моделей разных архитектурных парадигм, анализ нарезки изображений на фрагменты и проверка устойчивости к ухудшению качества входных изображений. Экспериментальная часть выполнена на подготовленном поднаборе SKU-110K: версия без нарезки содержит 300 изображений и 43 947 размеченных объектов, версия с нарезкой содержит 3 699 фрагментов и 87 715 объектов. Сравниваются YOLOv8s, RT-DETR-L, Faster R-CNN и WBF-ансамбль. Оценка строится по AP50-95, AP50, AP75, скорости обработки и падению качества при размытии, шуме, затемнении, JPEG-сжатии и снижении разрешения. По сравнению на изображениях без искажений RT-DETR-L немного превосходит YOLOv8s по AP50-95, однако YOLOv8s работает быстрее и показывает лучший AP75. Парная оценка YOLOv8s на тестовой части показала преимущество представления с нарезкой на фрагменты по абсолютным метрикам: AP50-95 составил 0,53813 против 0,31085 для варианта без нарезки. При ухудшении качества изображений наиболее заметное снижение качества в режиме с нарезкой вызвали размытие и шум. Результаты позволяют рассматривать YOLOv8s как рациональную базовую конфигурацию для быстрого контроля полок, а нарезку

изображений на фрагменты — как важный элемент подготовки данных для плотных сцен розничной торговли.

Ключевые слова: компьютерное зрение, обнаружение объектов, полочные сцены, розничная торговля, SKU-110K, нарезка изображений на фрагменты, YOLOv8, RT-DETR, Faster R-CNN, устойчивость моделей, ухудшение качества изображения.

Annotation. The article considers the problem of choosing a product detection model for analyzing the shelf scenes of retail stores. The comparison of models from different architectural paradigms, the analysis of image slicing into fragments, and the verification of resistance to deterioration in the quality of input images are combined. The experimental part was performed on a prepared subset of SKU-110K: the uncut version contains 300 images and 43,947 marked-up objects, the sliced version contains 3,699 fragments and 87,715 objects. YOLOv8s, RT-DETR-L, Faster R-CNN and the WBF ensemble are compared. The assessment is based on AP50-95, AP50, AP75, processing speed and quality loss due to blurring, noise, dimming, JPEG compression and resolution reduction. Compared to distortion-free images, the RT-DETR-L is slightly superior to the YOLOv8s in terms of AP50-95, but the YOLOv8s is faster and shows a better AP75. The paired YOLOv8s score on the test part showed the advantage of the representation with slicing into fragments according to absolute metrics: the AP50-95 was 0.53813 versus 0.31085 for the variant without slicing. As the image quality deteriorates, the most noticeable decrease in quality in slicing mode was caused by blurring and noise. The results allow us to consider YOLOv8s as a rational basic configuration for quick shelf control, and image slicing as an important element of data preparation for dense retail scenes.

Keywords: computer vision, object detection, shelf scenes, retail, SKU-110K, image slicing, YOLOv8, RT-DETR, Faster R-CNN, model stability, image quality degradation.

Введение

Автоматический анализ полочных изображений становится одной из практических задач компьютерного зрения в розничной торговле. По фотографии или видеокадру полки можно оценивать наличие товара, заполненность стеллажа, ошибки выкладки, пустые зоны и качество оформления и выкладки товаров. При ручной проверке такие операции требуют времени и зависят от внимательности сотрудника, поэтому торговые сети заинтересованы в алгоритмах, которые способны превращать изображение полки в структурированные данные.

Полочная сцена отличается от многих стандартных задач обнаружения объектов высокой плотностью объектов. В одном кадре может находиться несколько десятков или сотен товарных экземпляров. Упаковки расположены вплотную, частично перекрываются, имеют похожую форму и отличаются мелкими элементами дизайна. Даже небольшое смещение рамки может привести к захвату соседнего товара, а значит, к ошибке в подсчёте или подготовке объекта к дальнейшей идентификации SKU [3], [6].

Практическая система должна быть не только точной, но и устойчивой. Изображения из магазина редко бывают идеальными: встречаются блики, тени, шум камеры, смаз, JPEG-артефакты, снижение разрешения и различия в ракурсе съёмки. Модель, которая хорошо работает на изображениях без искажений, может заметно проседать при небольшом ухудшении изображения. Поэтому сравнение моделей обнаружения только по одной метрике AP50-95 не даёт полной картины. Требуется одновременно учитывать точность, скорость, полноту обнаружения и устойчивость к ухудшениям качества изображения [5], [7].

Дополнительную сложность создают свойства самой товарной упаковки. В работах В. Ю. Верещагина и соавторов рассматриваются оптические эффекты полимерных покрытий, светоотражающих поверхностей, стеклянной, пластиковой и металлической тары, а также скрытая маркировка изделий [4]. Для задач

компьютерного зрения это важно, потому что упаковка может менять цветовой контраст, создавать блики и нестабильные визуальные признаки при разных условиях освещения.

Цель статьи — выполнить многокритериальную оценку моделей обнаружения товаров на полочных изображениях с учётом архитектуры модели, скорости обработки изображения, подготовки данных через нарезку на фрагменты и устойчивости к типовым искажениям изображения. Отдельная задача состоит в проверке того, как один и тот же базовый алгоритм обнаружения ведёт себя на тестовой части данных без нарезки и с нарезкой на фрагменты. В статье эти элементы рассматриваются в единой экспериментальной постановке: оценивается не только максимальное значение одной метрики, но и практическая пригодность процесса обработки для плотных сцен розничной торговли.

1. Особенности полочных сцен и постановка задачи

Полочная сцена представляет собой плотную композицию товарных упаковок, ценников, полочного оборудования и рекламных элементов. В отличие от изображений общего назначения, где объект часто хорошо отделён от фона, здесь границы между соседними экземплярами могут быть почти незаметны. Это делает задачу особенно сложной для прямоугольная рамка-детекции: рамка должна охватить конкретный товар, не захватывая соседние упаковки и лишний фон.

Малый размер объектов является одной из ключевых причин использования нарезки изображений на фрагменты. Если весь снимок стеллажа приводится к фиксированному входу модели, мелкие товары теряют детали. При разбиении изображения на локальные фрагменты отдельная упаковка занимает большую долю входного изображения, а модель получает больше информации о границах, цветовых переходах и локальных признаках. Такой подход особенно важен для плотно заполненных сцен, к которым относится SKU-110K [3], [9].

Нарезка изображений на фрагменты не является автоматической гарантией роста качества. Она увеличивает число анализируемых фрагментов, требует пересчёта координат, объединения предсказаний и борьбы с дублями на границах фрагментов. Поэтому в эксперименте представление с нарезкой не только описывается как способ подготовки данных, но и отдельно проверяется на изображениях без искажений и при искусственном ухудшении качества. Такой подход позволяет отделить методическую роль нарезки от реальных числовых результатов.

Для выбора модели важно сравнивать разные архитектурные парадигмы. Одностадийные модели обнаружения, такие как YOLO, привлекательны скоростью и простотой внедрения. Двухстадийные подходы, например Faster R-CNN, важны как точный контрольный вариант и классическая схема с кандидатными областями [8]. Модели на основе трансформеров, такие как RT-DETR, интересны способностью учитывать более широкий контекст сцены и развивать идеи сквозного обнаружения объектов [12].

2. Материалы и методика эксперимента

Экспериментальная часть выполнена на подготовленном поднаборе SKU-110K. Этот набор данных выбран потому, что он ориентирован на плотные полочные сцены, где товары расположены близко друг к другу и требуют точной локализации. В рамках подготовки были сформированы две версии набора: исходная версия без нарезки и версия с нарезкой изображений на локальные фрагменты.

Версия без нарезки содержит 300 изображений и 43 947 размеченных объектов. Разбиение составляет 240 изображений для обучения, 30 для проверки и 30 для тестирования. Версия с нарезкой содержит 3 699 фрагментов и 87 715 объектов, а разбиение составляет 2 946 / 369 / 384 для обучения, проверки и тестирования соответственно. Связь фрагментов с исходными изображениями фиксируется в файле tile_map.json. Оба варианта имеют один класс объекта — товар.

Таблица 1 — Характеристики подготовленных данных

Параметр	Без нарезки	С нарезкой
Версия данных	sku110k_small/v1	sku110k_small/v1_tiled
Изображения / фрагменты	300 изображений	3 699 фрагментов
Количество объектов / рамок	43 947	87 715
Количество классов	1	1
Обучение / проверка / тест	240 / 30 / 30	2 946 / 369 / 384
Связь с исходным изображением	не требуется	tile_map.json, 3 699 фрагментов

В эксперимент включены YOLOv8s, RT-DETR-L, Faster R-CNN на Detectron2 и ансамбль WBF, объединяющий предсказания YOLO и RT-DETR. Такой набор позволяет сравнить быстрый одностадийный базовый вариант, модель на основе трансформера, классический двухстадийный подход и постобработку предсказаний через взвешенное объединение рамок.

Для оценки качества используются AP50-95, AP50 и AP75. AP50-95 даёт строгую усреднённую оценку по нескольким IoU-порогам, AP50 показывает способность модели находить объект в целом, AP75 сильнее штрафует неточную локализацию. Для практического применения дополнительно учитывается скорость обработки в миллисекундах на изображение.

Устойчивость проверялась на пяти сценариях искусственного ухудшения изображения. Размытие имитирует смаз и потерю фокусировки, шум — помехи камеры, затемнение — недостаточное освещение, JPEG-сжатие — артефакты сжатия, снижение разрешения — потерю детализации. Падение качества рассчитывается

относительно изображения без искажений отдельно для представлений тестовой части без нарезки и с нарезкой на фрагменты.

Таблица 2 — Сценарии ухудшения качества изображения

Сценарий	Что моделирует	Параметр
Без искажений	изображение без искусственного ухудшения	0.0
Размытие	смаз, расфокусировка	2.0
Шум	шум камеры, слабое освещение	10.0
Затемнение	затемнение сцены	0.6
JPEG-сжатие	потеря текстур при сжатии	50.0
Снижение разрешения	уменьшение разрешения	0.5

3. Результаты сравнения моделей

Первый блок результатов связан со сравнением моделей на изображениях без искажений. Он показывает базовый уровень качества и скорости при одинаковой постановке задачи. Для полочных сцен этот этап уже является сложным, поскольку даже без искусственных искажений модель работает с плотной выкладкой, большим числом мелких объектов и визуально похожими упаковками.

RT-DETR-L показал максимальное значение AP50-95 — 0,53879. YOLOv8s практически не уступил по AP50-95 — 0,53835, но оказался быстрее: 37,88 мс/изобр. против 52,28 мс/изобр. у RT-DETR-L. При этом YOLOv8s показал более высокий AP75, что важно для плотных сцен, где точность границ рамки влияет на разделение

соседних товаров. Faster R-CNN в этом сравнении уступил обеим моделям по качеству и скорости.

Таблица 3 — Сравнение моделей на изображениях без искажений

Модель	Парадигма	AP50-95	AP50	AP75	Скорость
YOLOv8s	одностадийная	0.53835	0.88806	0.59693	37.88 мс/изобр.
RT-DETR-L	на основе трансформера	0.53879	0.89088	0.58845	52.28 мс/изобр.
Faster R-CNN	двухстадийная	0.47307	0.85756	0.47615	88.44 мс/изобр.
WBF (YOLO + RT-DETR)	ансамбль	0.51891	0.87535	0.55422	—

Второй блок результатов связан с парной оценкой YOLOv8s на тестовых данных без нарезки и с нарезкой на фрагменты. Для тестовой части без нарезки использовалось 30 изображений и 4 526 размеченных объектов, для тестовой части с нарезкой — 384 фрагмента и 9 063 объекта. На сценарии без искажений представление с нарезкой показало более высокий AP50-95: 0,53813 против 0,31085 у варианта без нарезки. Такой результат показывает, что выбранная модель обнаружения и параметры обработки лучше согласованы с локальными фрагментами, где товар занимает большую часть входного изображения.

При искусственном ухудшении качества изображений режим с нарезкой сохранил более высокий абсолютный уровень метрик. Для представления с нарезкой наибольшее падение AP50-95 вызвало размытие: 0,01236 относительно уровня без

искажений. Шум дал второе по величине снижение — 0,00717. Для представления без нарезки наиболее заметным фактором оказался шум: падение AP50-95 составило 0,00789. Отрицательные значения ΔAP в отдельных строках снижения разрешения или JPEG-сжатия следует трактовать как небольшие колебания оценки на ограниченной тестовой части, а не как реальное улучшение качества от искажения.

Таблица 4 — Парное сравнение YOLOv8s на тестовых данных без нарезки и с нарезкой при ухудшении качества изображения

Режим данных	Сценарий	AP50-95	AP50	AP75	ΔAP50-95
Без нарезки	Без искажений	0.31085	0.54850	0.32629	0.00000
Без нарезки	Размытие	0.30787	0.54725	0.32425	0.00298
Без нарезки	Шум	0.30297	0.53868	0.31462	0.00789
Без нарезки	Затемнение	0.30761	0.53671	0.32241	0.00324
Без нарезки	JPEG-сжатие	0.30874	0.54007	0.32255	0.00211
Без нарезки	Снижение разрешения	0.31436	0.54994	0.33451	-0.00351
С нарезкой	Без искажений	0.53813	0.88830	0.59477	0.00000
С нарезкой	Размытие	0.52577	0.87269	0.57309	0.01236
С нарезкой	Шум	0.53096	0.87418	0.58572	0.00717
С нарезкой	Затемнение	0.53636	0.88764	0.58973	0.00177
С нарезкой	JPEG-сжатие	0.53837	0.88741	0.59491	-0.00024

С нарезкой	Снижение разрешения	0.53707	0.88784	0.59373	0.00106

4. Обсуждение результатов

Полученные результаты показывают, что выбор модели для анализа полок нельзя делать только по максимальному AP50-95. RT-DETR-L немного лидирует по общей метрике, но YOLOv8s работает быстрее и показывает лучший AP75. Для регулярного контроля полок это делает YOLOv8s более рациональной базовой конфигурацией: она сохраняет близкое качество при меньшей вычислительной стоимости.

Faster R-CNN важен как методически понятный двухстадийный контрольный вариант, но в выполненном эксперименте он проигрывает по скорости и строгим метрикам. Это не означает бесполезность двухстадийного подхода в целом, однако для выбранного программного контура и ограниченного поднабора SKU-110K его практическая роль скорее контрольная, чем основная.

WBF-ансамбль также не стал очевидным улучшением. Объединение предсказаний YOLO и RT-DETR может быть полезно в отдельных случаях, когда модели дополняют друг друга, но на плотных полках усреднение близких, но смещённых рамок способно ухудшать строгую локализацию. Поэтому WBF целесообразно рассматривать как дополнительный пакетный режим, а не как основной путь для быстрого анализа полки.

Нарезка изображений на фрагменты в этой постановке играет роль не только способа расширения рабочей выборки, но и практически значимого режима представления плотной сцены. Переход от 300 исходных изображений к 3 699 фрагментам увеличивает число локальных областей, в которых товары занимают большую площадь кадра. Парная оценка показала, что на тестовой части

представление с нарезкой обеспечивает более высокий абсолютный уровень AP50-95, AP50 и AP75, чем вариант без нарезки. Это согласуется с особенностями SKU-110K: мелкие и плотно расположенные товары лучше анализируются в локальном масштабе.

При этом результаты по нарезке изображений на фрагменты нужно интерпретировать аккуратно. Выигрыш представления с нарезкой получен в конкретной конфигурации YOLOv8s, выбранного размера входа и подготовленного поднабора данных. Чем больше фрагментов создаётся из одного изображения, тем выше вычислительная нагрузка. Поэтому для промышленного применения важно считать не только метрики на одном входе модели, но и полное время обработки исходного изображения с учётом нарезки, обработки фрагментов моделью и объединения результатов.

Работы В. Ю. Верещагина и соавторов важны для интерпретации результатов не как источники по YOLO или RT-DETR, а как предметное обоснование визуальной сложности упаковки. Отражающие поверхности, полимерные покрытия и скрытая маркировка могут менять видимость деталей при разном освещении [4]. В реальном магазине это проявляется в бликах, изменении контраста, потере читаемости мелких элементов и нестабильности признаков, поэтому проверка устойчивости является не дополнительной формальностью, а необходимым этапом выбора модели.

5. Практические рекомендации

Для быстрого базового контроля полок наиболее рационально использовать YOLOv8s. Эта модель почти не уступает RT-DETR-L по AP50-95, показывает лучший AP75 и заметно выигрывает по скорости обработки. Она подходит для сценариев, где требуется регулярно обрабатывать изображения полок и получать первичный отчёт о найденных товарных объектах.

RT-DETR-L можно рассматривать как альтернативу для случаев, где допустима более высокая вычислительная стоимость и требуется проверить потенциальный

выигрыш подхода на основе трансформера. На выбранных данных модель даёт максимальные AP50-95 и AP50, но уступает YOLOv8s по скорости и AP75, поэтому её преимущество не является безусловным.

Faster R-CNN следует оставить как контрольную модель для сравнения архитектурных парадигм. В рамках этой постановки он оказался тяжелее и слабее по основным метрикам на изображениях без искажений. Его использование оправдано при необходимости методического контрольного варианта или при дальнейших экспериментах с более точной настройкой двухстадийных моделей.

Для подготовки плотных данных рекомендуется сохранять версию набора с нарезкой изображений на фрагменты и файл соответствий tile_map.json. Парная оценка показала, что в выбранной конфигурации тестирование с нарезкой даёт более высокий абсолютный уровень AP50-95, чем тестирование без нарезки. Поэтому представление с нарезкой целесообразно рассматривать как основной режим подготовки плотных полочных сцен при условии контроля времени обработки и качества объединения результатов.

При эксплуатации системы особенно важно контролировать резкость изображения и шум. В режиме с нарезкой именно размытие и шум дали наиболее заметное падение AP50-95, а в режиме без нарезки наиболее чувствительным сценарием оказался шум. Поэтому качество камеры, стабилизация, фокусировка и условия освещения могут быть не менее важны, чем выбор конкретной архитектуры модели обнаружения.

6. Ограничения исследования и направления развития

Главное ограничение выполненной оценки связано с тем, что экспериментальная часть опирается на подготовленный поднабор SKU-110K, а не на полный набор изображений из реального магазина. Такой вариант достаточен для воспроизводимого сравнения моделей, однако он не закрывает все возможные

условия эксплуатации: разные камеры, высоту съёмки, освещение, категории товаров, стеклянные витрины и нестандартные промо-материалы.

Второе ограничение относится к интерпретации нарезки изображений на фрагменты. В работе выполнено парное сравнение YOLOv8s на тестовых данных без нарезки и с нарезкой, и оно показывает преимущество представления с нарезкой по абсолютным метрикам. Однако этот результат не означает, что любой вариант нарезки автоматически улучшит качество. Для более строгого вывода нужно дополнительно сравнить разные размеры фрагмента, перекрытие соседних фрагментов, правила объединения предсказаний и полное время обработки исходного изображения.

Третье ограничение связано с устойчивостью разных архитектур. В статье приведена подробная проверка устойчивости для базового YOLOv8s в двух режимах представления данных. Она показывает, какие искажения сильнее влияют на качество и как это связано с нарезкой изображений на фрагменты, но не заменяет полноценное сравнение устойчивости YOLOv8s, RT-DETR-L и Faster R-CNN по одинаковому протоколу ухудшения качества изображений.

Отдельного внимания требует скорость. В таблице сравнения на изображениях без искажений скорость указана в мс/изобр., что удобно для базовой интерпретации. Однако в режиме с нарезкой практическое время обработки одного исходного изображения складывается из времени обработки нескольких фрагментов, сборки координат и удаления дублей. Поэтому для промышленного применения нужно считать не только время на один вход модели, но и полное время обработки исходного изображения полки.

Дальнейшее развитие исследования можно связать с сегментационной веткой анализа. Обнаружение по прямоугольным рамкам отвечает на вопрос о расположении товара, но для последующей идентификации SKU более полезны маски объектов.

Маска позволяет отделить упаковку от фона, соседних товаров и ценников, а значит, получить более чистый вырезанный фрагмент для классификации, распознавания текста или поиска по векторным признакам.

Отдельным направлением развития является совершенствование программного контура ShelfVision: унификация формата предсказаний, автоматическая генерация отчётов, расширение оценки прямоугольных рамок и масок, анализ плотности полочных сцен и подготовка результатов к интеграции с системами розничной торговли.

Практически важным направлением остаётся идентификация SKU. Модель обнаружения находит товар как объект, но для магазина требуется понять, какой именно артикул находится на полке. Для этого можно использовать каскад: прямоугольная рамка или маска, вырезанный фрагмент, визуальные векторные признаки, распознавание текста по упаковке и поиск по товарному справочнику. Такая постановка не повторяет статью о моделях обнаружения, а переводит исследование на следующий уровень — от локализации к распознаванию конкретной продукции.

Ещё одно направление связано с анализом ошибок. Для полочных сцен полезно не только считать AP, но и классифицировать ошибки: пропуск мелких товаров, объединение соседних упаковок, ложные срабатывания на ценниках, смещение рамки, потеря качества при размытии или шуме. Такой анализ может быть оформлен как отдельный экспериментальный блок и использован для улучшения данных, вариантов расширения обучающей выборки и правил постобработки.

Таблица 5 — Возможные направления дальнейших экспериментов

Направление	Что измерять	Практический смысл

Устойчивость по всем моделям	ΔAP для YOLOv8s, RT-DETR-L, Faster R-CNN	Сравнить устойчивость архитектур при одинаковых сценариях ухудшения качества
Полное время нарезки и обработки	мс на исходное изображение, число фрагментов, время объединения	Оценить применимость нарезки изображений на фрагменты в реальном контроле
YOLO-Seg и оценка масок	mask IoU, APmask50, APmask75, APmask50-95	Оценить пригодность масок для подготовки объектов к SKU-идентификации
Сопоставление с товарным справочником	точность среди первых k вариантов, оценка распознавания текста и сходства по векторным признакам	Оценить качество сопоставления найденных объектов с товарным справочником

Заключение

В статье выполнена объединённая оценка моделей обнаружения товаров для полочных сцен розничных магазинов с учётом подготовки данных, архитектуры модели, скорости и устойчивости к ухудшению качества изображения. Такой формат объединяет две близкие постановки: сравнение одностадийных, двухстадийных и трансформерных моделей обнаружения, а также анализ роли нарезки изображений на фрагменты и сценариев ухудшения качества в плотных изображениях полок.

Подготовленный набор данных представлен в двух версиях. Версия без нарезки содержит 300 изображений и 43 947 размеченных объектов, версия с нарезкой содержит 3 699 фрагментов и 87 715 объектов. Парная проверка на тестовой части показала, что представление с нарезкой в выбранной конфигурации YOLOv8s обеспечивает более высокий уровень AP50-95: 0,53813 против 0,31085 для варианта без нарезки. Это подтверждает практическую значимость локального представления плотной полочной сцены.

Сравнение моделей показало близость YOLOv8s и RT-DETR-L по AP50-95: 0,53835 и 0,53879 соответственно. При этом YOLOv8s оказался быстрее и показал лучший AP75, что важно для точной локализации товаров в плотной выкладке. Faster R-CNN уступил по качеству и скорости, а WBF не обеспечил прироста строгой точности по сравнению с одиночными моделями.

Проверка устойчивости показала, что для представления с нарезкой наиболее чувствительным сценарием является размытие: падение AP50-95 составило 0,01236. Шум дал второе по величине снижение — 0,00717. Для представления без нарезки наиболее заметным сценарием стал шум с падением AP50-95 на 0,00789. Затемнение, JPEG-сжатие и снижение разрешения в выбранных параметрах оказали меньшее влияние или дали колебания в пределах ограниченной тестовой части.

С практической точки зрения YOLOv8s можно рассматривать как основной базовый вариант для быстрого анализа полок, RT-DETR-L — как сильную альтернативу при допустимом росте вычислительной стоимости, Faster R-CNN — как контрольный двухстадийный вариант, а WBF — как дополнительный пакетный режим. Представление с нарезкой в выполненной оценке показало заметное преимущество по абсолютным метрикам и может использоваться как основной режим подготовки плотных изображений полок. Дальнейшее развитие эксперимента целесообразно связать с оценкой устойчивости всех архитектур и с отдельной

сегментационной веткой, позволяющей готовить более чистые вырезанные фрагменты товаров для последующей идентификации SKU.

Список источников

1. Пухова Е. А., Верещагин В. Ю. Принтмедиатехнологии [Электронный ресурс] : учебно-методическое пособие для студентов магистратуры по направлению 09.04.01 «Информатика и вычислительная техника». — М. : Московский Политех, 2021. — № госрегистрации 0322200838. — URL: https://catalog.inforeg.ru/Inet/GetEzineByID/334980 (дата обращения: 02.05.2026).

2. Albumentations. Documentation [Электронный ресурс]. — URL: https://albumentations.ai/docs/ (дата обращения: 02.05.2026).

3. Goldman E., Herzig R., Eisenschtat A., Goldberger J., Hassner T. Precise Detection in Densely Packed Scenes [Электронный ресурс] // Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. — 2019. — URL: https://openaccess.thecvf.com/content_CVPR_2019/papers/Goldman_Precise_Detection_in_Densely_Packed_Scenes_CVPR_2019_paper.pdf (дата обращения: 02.05.2026).

4. Kondratov A. P., Vereshchagin V. Y., Pogiba A. Y., Volinsky A. A. Transparent Multilayer Polymer Films for Hidden Marking of Reflective Containers [Электронный ресурс] // Optics Letters. — 2026. — Vol. 51, № 5. — P. 1215–1218. — DOI: 10.1364/OL.585073. — URL: https://opg.optica.org/ol/abstract.cfm?uri=ol-51-5-1215 (дата обращения: 02.05.2026).

5. Liu J., Wang Z., Ma L., Fang C., Bai T., Zhang X., Liu J., Chen Z. Benchmarking Object Detection Robustness against Real-World Corruptions [Электронный ресурс] // International Journal of Computer Vision. — 2024. — URL: https://wangzhijie.me/assets/pubs/ijcv24-benchmark.pdf (дата обращения: 02.05.2026).

6. 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 [Электронный ресурс] // Engineering Applications of Artificial Intelligence. — 2024. — Vol. 133. — Art. 108452. — URL: https://www.sciencedirect.com/science/article/pii/S0952197624006109 (дата обращения: 02.05.2026).

7. Pietrini R., Paolanti M., Mancini A., Frontoni E., Zingaretti P. Shelf Management: A Deep Learning-Based System for Shelf Visual Monitoring [Электронный ресурс] // Expert Systems with Applications. — 2024. — Vol. 246. — Art. 124635. — URL: https://www.sciencedirect.com/science/article/pii/S0957417424015021 (дата обращения: 02.05.2026).

8. Ren S., He K., Girshick R., Sun J. Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks [Электронный ресурс] // Advances in Neural Information Processing Systems. — 2015. — URL: https://arxiv.org/abs/1506.01497 (дата обращения: 02.05.2026).

9. Rong T., Zhu Y., Xiong Y., Cai H. A Solution to Product Detection in Densely Packed Scenes [Электронный ресурс]. — 2020. — URL: https://arxiv.org/abs/2007.11946 (дата обращения: 02.05.2026).

10. Solovyev R., Wang W., Gabruseva T. Weighted Boxes Fusion: Ensembling Boxes from Different Object Detection Models [Электронный ресурс] // Image and Vision Computing. — 2021. — Vol. 107. — Art. 104117. — URL: https://arxiv.org/abs/1910.13302 (дата обращения: 02.05.2026).

11. Tonioni A., Serra E., Di Stefano L. A Deep Learning Pipeline for Product Recognition on Store Shelves [Электронный ресурс] // 2018 IEEE International Conference on Image Processing, Applications and Systems. — 2018. — URL: https://arxiv.org/abs/1810.01733 (дата обращения: 02.05.2026).

12. Ultralytics. Explore Ultralytics YOLOv8 [Электронный ресурс]. — URL: https://docs.ultralytics.com/models/yolov8/ (дата обращения: 02.05.2026).

13. Zhao Y., Lv W., Xu S. et al. DETRs Beat YOLOs on Real-Time Object Detection [Электронный ресурс] // Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. — 2024. — URL: https://openaccess.thecvf.com/content/CVPR2024/html/Zhao_DETRs_Beat_YOLOs_on_Real-time_Object_Detection_CVPR_2024_paper.html (дата обращения: 02.05.2026).