

ИНФОРМАЦИОННЫЕ ТЕХНОЛОГИИ И ТЕЛЕКОММУНИКАЦИИ

УДК 004.93

РАСПОЗНАВАНИЕ И ОЦЕНКА РАСПОЛОЖЕНИЯ ОБЪЕКТОВ ПРИ ПОМОЩИ НЕЙРОННЫХ СЕТЕЙ

С.А. ГУЛЯЕВ, асп., М.В. ТАРАЧКОВ, канд. техн. наук,
О.В. ТОЛСТЕЛЬ, канд. техн. наук, А.Е. ШИРКИН, ассистент

Балтийский федеральный университет имени Иммануила Канта,
236041, Калининград, ул. Александра Невского, 14, e-mail: shirkin39@mail.ru

© Гуляев С.А., Тарачков М.В., Толстель О.В., Ширкин А.Е., 2025

Разработана практическая база для реализации 3D-оценки позиции. В первой части статьи проанализирована система автоматизации минимаркетов, которая использует YOLOv8 для распознавания товаров. Во второй части исследованы методы распознавания и сегментации объектов: применение моделей YOLOv8 для распознавания объектов и U-Net для семантической сегментации 3D-моделей. Сделан вывод, что данные работы являются составляющими для решения задачи 3D-оценки позиции и представляют собой обзор решений нескольких небольших задач с использованием нейронных сетей.

Ключевые слова: нейронные сети, YOLOv8, U-Net, распознавание объектов, оценка расположения, сегментация изображений, автоматизация, минимаркет, детекция.

DOI: 10.46573/2658-5030-2025-2-66-77

ВВЕДЕНИЕ

Современные нейронные сети, задействованные в области компьютерного зрения, могут решать обширный диапазон задач в таких сферах, как торговля, робототехника и промышленная автоматизация. Для решения более сложных вопросов, (например, для оценки 3D-объекта по изображению) необходимо решить несколько подзадач (а именно детектирования, сегментации и непосредственной локализации объектов), что дает возможность получить требуемый результат. Для этого проанализируем в пределах статьи такие подходы, как использование YOLOv8 [1] для автоматизации минимаркетов, что обеспечит распознавание и классификацию товаров, и применение YOLOv8 для детекции и U-Net [2] с целью семантической сегментации механических деталей и последующего нахождения расстояния до объекта, расположенного на плоскости. Полученная база и проведенное исследование подходов позволяют перейти к решению задачи локализации, или оценки, положения объектов в 3D-пространстве. Таким образом, целью данной статьи является разработка практической базы для реализации 3D-оценки позиции. Для достижения цели нужно решить две практические задачи, применив при этом составляющие 3D-оценки позиции: 2D-детекцию и сегментацию объектов.

МАТЕРИАЛЫ И МЕТОДЫ

Решение поставленных задач состоит из трех этапов:

1. Автоматизации работы минимаркета и решения задачи детектирования.

2. Детектирования элементарных геометрических фигур на изображении при помощи нейронных сетей.

3. Детектирования и сегментации механических деталей.

В свою очередь, первый этап разбивается на два:

1. Автоматизации работы минимаркета.

2. Использования YOLOv8 для распознавания товаров.

В рамках второго подхода рассматривается решение задачи детектирования простых геометрических фигур при помощи YOLOv8. При третьем подходе решается задача семантической сегментации реальных промышленных деталей, смоделированных в 3D и проецированных на изображение. В нем используется нейронная сеть U-Net. Далее с учетом эвристического подхода и того, что объекты находятся на плоскости, производится геометрический расчет расстояния до объекта.

1. АВТОМАТИЗАЦИЯ РАБОТЫ МИНИМАРКЕТА

Проведем разработку информационной системы управления бесконтактным минимаркетом, осуществляющим распознавание товаров и автоматическое проведение платежей, а также сборку работоспособного стенда для демонстрации работы системы (рис. 1).

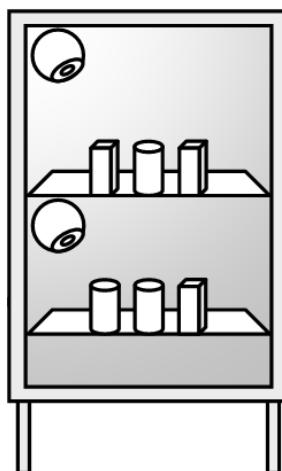


Рис. 1. Макет минимаркета

Микромаркет – это умный торговый холодильник с товарами, который работает по принципу самообслуживания, т. е. без участия продавца. Основная идея проекта заключалась в минимизации действий со стороны покупателя, а также времени, затраченного на покупку: покупатель открывает микромаркет с помощью веб-приложения, берет необходимые товары и закрывает его (в данном случае дополнительные действия не требуются) (рис. 2). Подсчет и определение взятых товаров происходят автоматически при помощи нейронной сети, затем покупателю выставляется счет на оплату в приложении.

Минимаркет – это магазин, который состоит из нескольких таких микромаркетов (умных холодильников), т. е. торговая точка с полным самообслуживанием.



Рис. 2. Схема работы микромаркета

Основными задачами здесь являлись:

1. Разработка информационной системы управления минимаркетом, а именно интерфейса для покупателей, серверной части, программного обеспечения для аппаратной части, а также обучение и настройка работы нейросети.
 2. Сборка тестового стенда для демонстрации работоспособности системы.
- Использование модели YOLOv8 для распознавания товаров было одной из составных частей нашей работы.

Использование YOLOv8 для распознавания товаров

Система управления минимаркетом использует YOLOv8 для распознавания предметов, за счет чего оптимизируется процесс покупки. Таким образом устраняется необходимость в ручном сканировании товаров (сейчас рынок представлен только микромаркетами с вышеназванным сканированием).

Особое внимание при реализации проекта уделялось тому, чтобы система могла распознавать товары на полках микромаркета, классифицировать их и передавать информацию о покупках в систему учета в режиме реального времени (рис. 3).

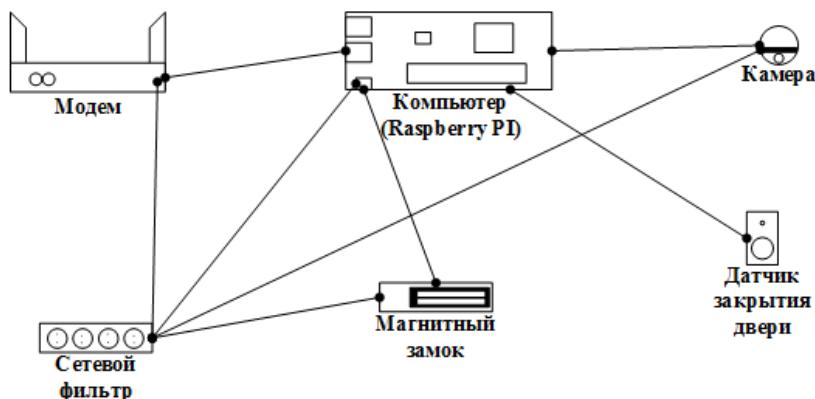


Рис. 3. Аппаратная часть микромаркета

Алгоритм YOLOv8 был выбран из-за его высокой скорости работы и точности детекции объектов.

После открытия дверцы микромаркета камера передает видеопоток на сервер, где YOLOv8 обрабатывает изображения, выделяя и классифицируя товары, взятые клиентом. Это позволяет системе минимаркета автоматически регистрировать, какие

товары были взяты покупателем, что значительно ускоряет процесс покупки и дает возможность клиентам брать нужные товары самостоятельно.

Для успешного применения YOLOv8 в системе минимаркета необходимо было подготовить качественный датасет с изображениями товаров. При разметке использовался инструмент с открытым исходным кодом CVAT [3]. В процессе работы был создан специализированный набор данных, включающий иллюстрации товаров в различных ракурсах и условиях освещения. Это позволило модели обучиться на множестве сценариев и повысить свою точность при детекции.

С целью оценки качества модели использовались метрики Precision, Recall и F1 Score [4]. Итоговые результаты составили 0,95, 0,93 и 0,94 соответственно. Высокие значения этих метрик гарантируют, что система будет распознавать товары с минимальной погрешностью.

2. РАСПОЗНАВАНИЕ И СЕГМЕНТАЦИЯ ОБЪЕКТОВ С ИСПОЛЬЗОВАНИЕМ НЕЙРОННЫХ СЕТЕЙ

Вторая часть статьи посвящена изучению методов нейросетевого распознавания и оценки положения объектов. Задачи охватывают применение нейросетевых архитектур YOLOv8 и U-Net для решения вопросов детекции и семантической сегментации, а также разработку подхода для оценки расстояния до реальных объектов на основе их 3D-моделей.

Целью данной работы являлась разработка и оценка эффективности методов машинного обучения для распознавания объектов и семантической сегментации.

2.1. Применение YOLOv8 для распознавания простейших фигур различных форм

YOLOv8 была выбрана для распознавания простейших фигур (детских игрушек) как одна из самых точных и быстрых нейросетевых моделей, с помощью которой осуществляют детекцию объектов. Основной задачей на данном этапе исследования была разработка системы, которая способна распознавать и классифицировать игрушки различных форм (круглые, квадратные и треугольные) по представленным изображениям.

Для обучения модели был подготовлен специальный набор данных, содержащий фотографии игрушек, снятых с разных углов и на различном фоне (рис. 4). Аннотации изображений производились с использованием инструмента Roboflow [5], что позволило четко обозначить контуры объектов и создать обучающий датасет высокого качества. Модель YOLOv8 прошла обучение на указанном наборе данных и показала высокие результаты во время тестирования (рис. 5).

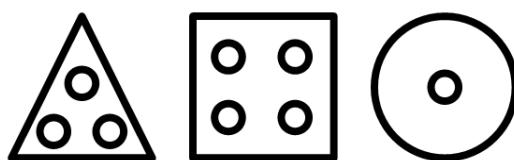


Рис. 4. Примеры фигур для распознавания

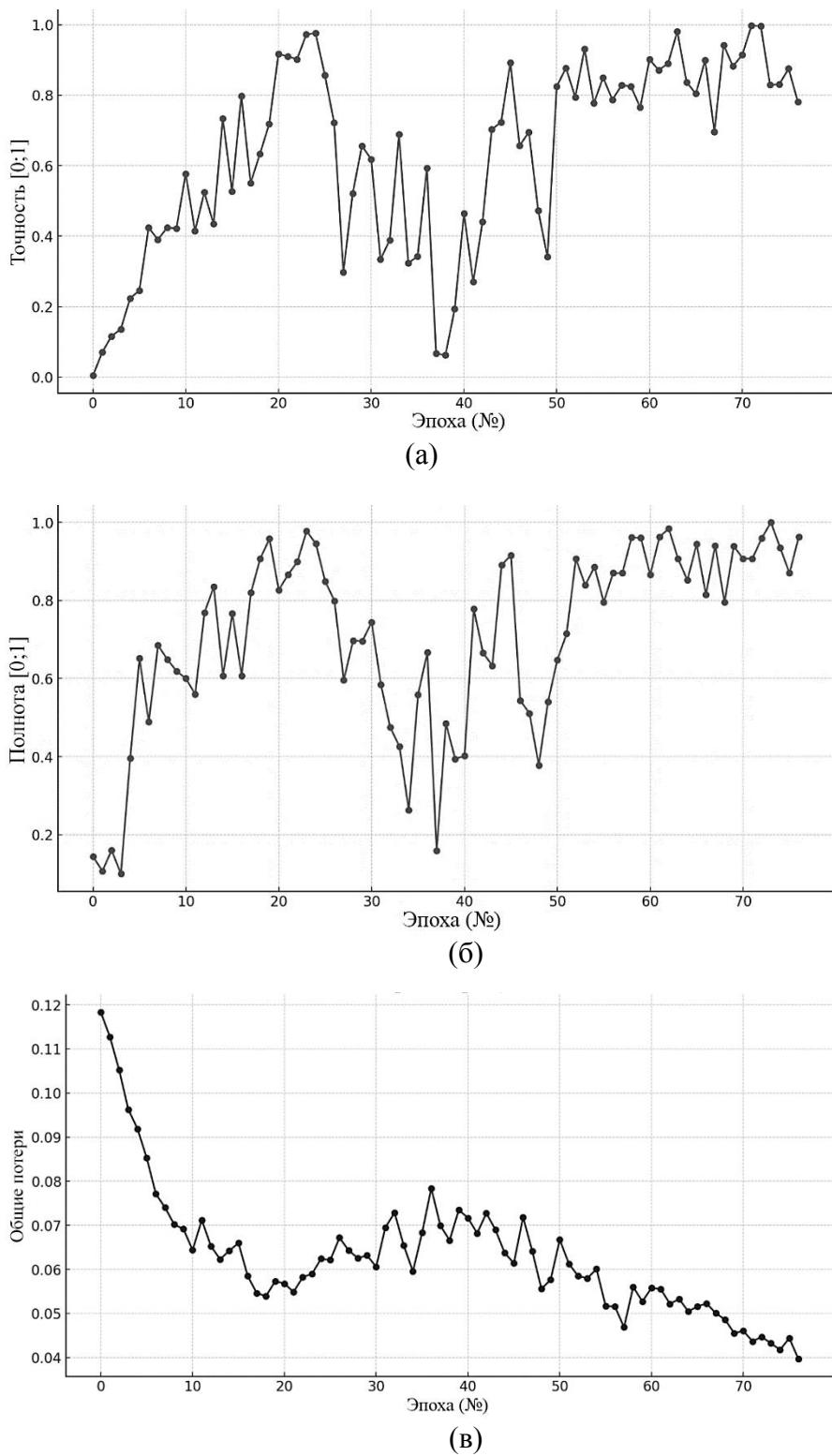


Рис. 5. Результаты обучения модели: динамика изменений точности (а) и показателя полноты (б); общие потери за каждую эпоху (в); динамика изменения средней точности предсказаний (г)

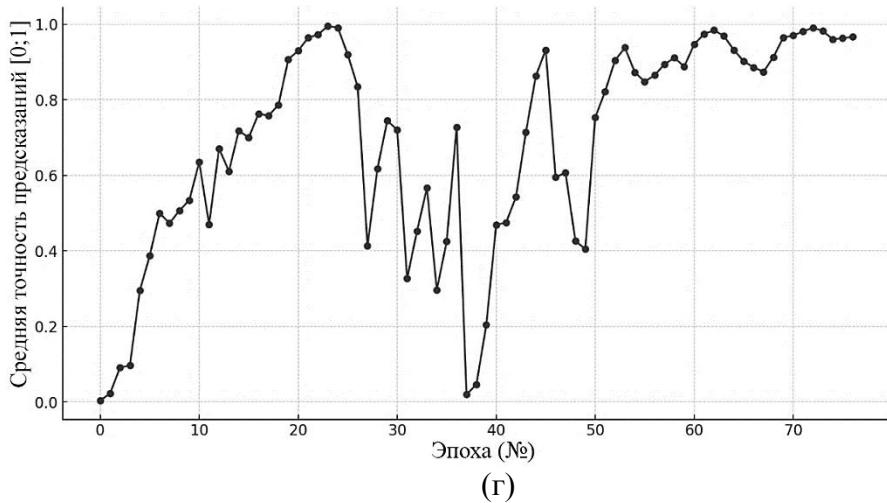


Рис. 5. Продолжение

Во время работы модель успешно распознавала объекты на изображениях, характеризующихся варьирующимся освещением и сложным фоном, что подтвердило ее надежность при функционировании в действительности. Благодаря своей архитектуре YOLOv8 обеспечивает детекцию объектов в режиме реального времени, что делает ее удобной для применения в системах мониторинга и инвентаризации.

2.2. Семантическая сегментация 3D-моделей с помощью U-Net

На следующем этапе исследования была применена модель U-Net, которая показала себя в качестве эффективного решения задач, требующих точного выделения контуров и сегментов объектов.

Основной целью на данном этапе было изучение потенциала использования алгоритмов семантической сегментации для анализа искусственно созданных объектов, что дает возможность оценить их результативность при работе с виртуальными моделями.

Для формирования 3D-моделей различных объектов (таких как гайка, стержень, опора) задействовали программное обеспечение «КОМПАС-3D» [6]. Модели были визуализированы в Blender [7] (рис. 6), что обеспечило получение изображений объектов с различных ракурсов и в разных условиях освещения.

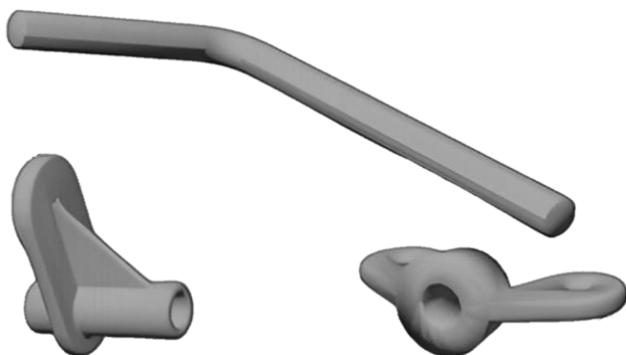


Рис. 6. Представление объектов в Blender

Для обучения модели U-Net использовался аннотированный датасет, в котором были обозначены все части и особенности формы каждого объекта. Благодаря уникальной архитектуре U-Net модель способна точно определять границы объектов и классифицировать их, что подтверждает ее эффективность при работе с виртуальными 3D-моделями (рис. 7).

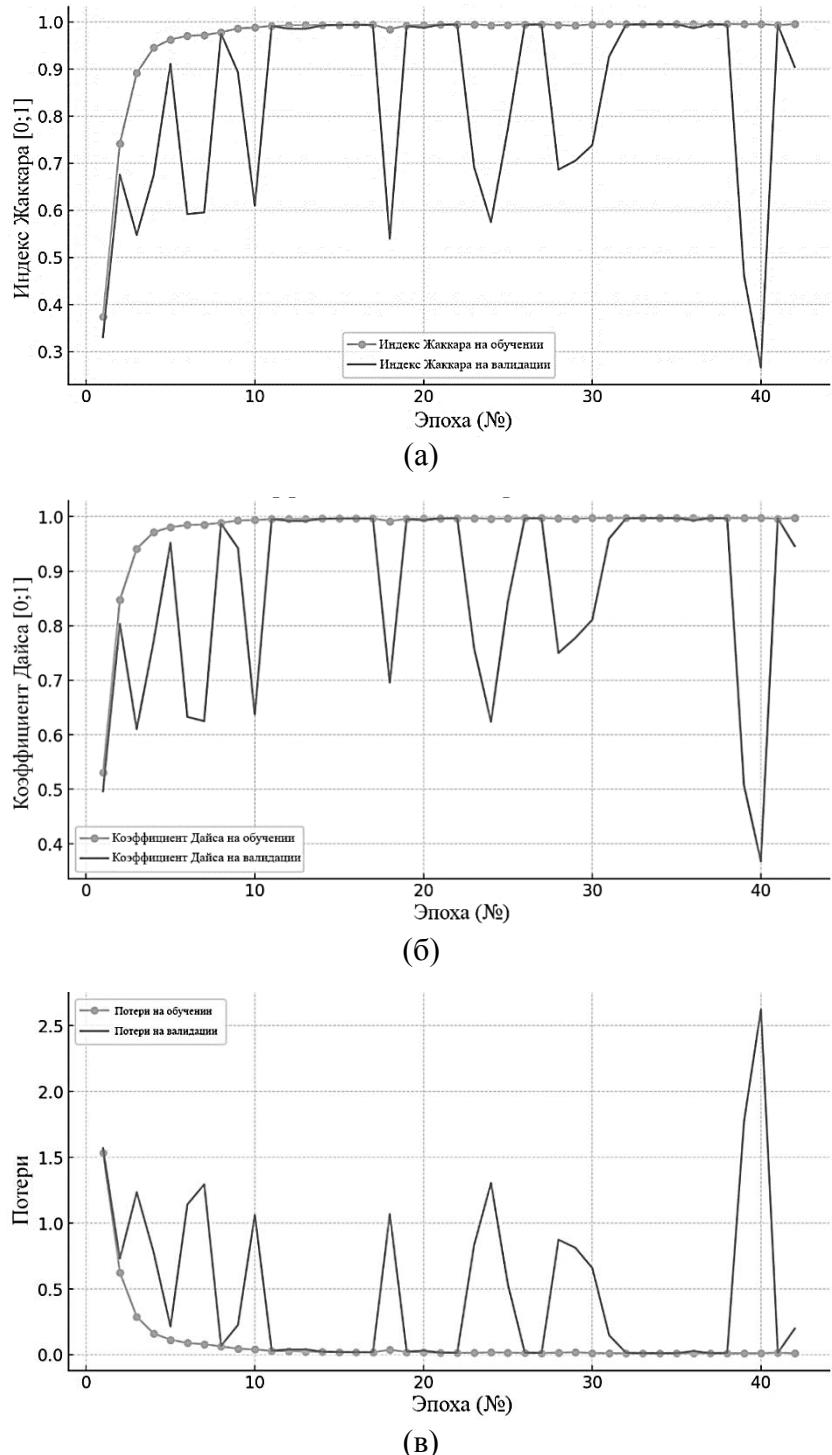


Рис. 7. Результаты обучения модели: график индекса Жаккара (а); коэффициента Дайса (б); потерь (в)

2.3. Семантическая сегментация и нахождение расстояния до реальных деталей на основе 3D-моделей

Далее были использованы результаты сегментации 3D-моделей при работе с реальными объектами. Этот процесс включал адаптацию моделей к условиям реальных съемок и определение расстояний до объектов на основе их визуального представления. В качестве объектов были выбраны реальные детали, такие как рым-гайки и мебельные уголки. Для создания цифровых копий указанных деталей были применены 3D-модели, полученные с помощью программного обеспечения «КОМПАС-3D» (рис. 8).



Рис. 8. Модель и реальная рым-гайка M8

Были исследованы также модели ResUNet [8] и VGG-UNet [9]. Первичное тестирование и анализ выявили необходимость доработки набора данных по причине неудовлетворительных итогов. После нее указанный набор стал давать хорошие результаты (рис. 9).

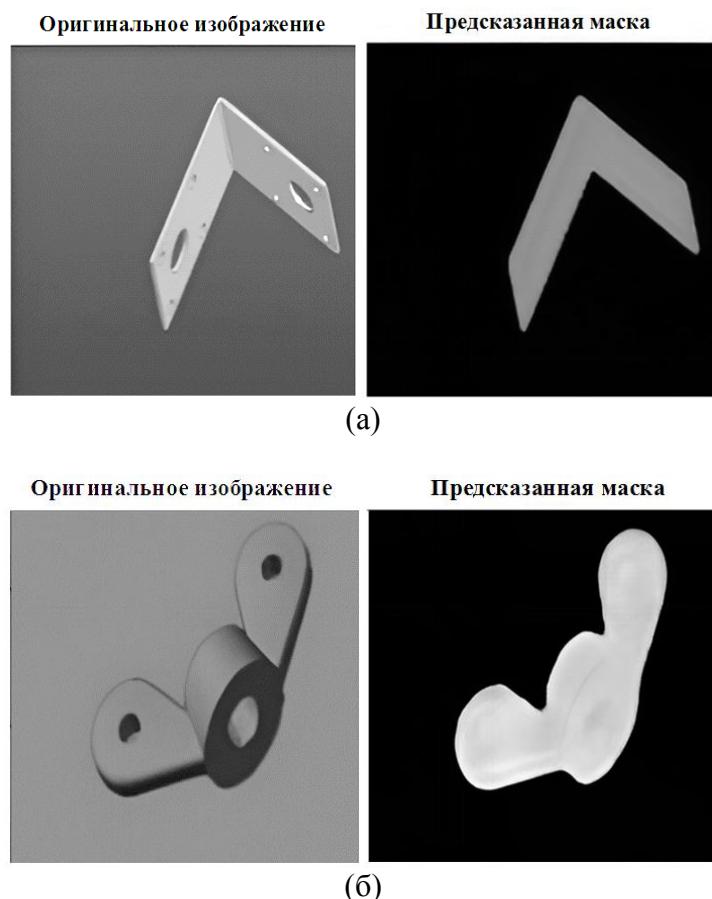


Рис. 9. Распознавание детали, отрисованной в программе «КОМПАС-3Д» (а); и детали, отрисованной в той же программе (дополнительный пример) (б)

Заключительным шагом на данном этапе стало определение расстояния до предмета, находящегося на плоскости. Для этого использовался метод триангуляции, основанный на знании высоты камеры относительно стола и угла ее наклона.

Расчет расстояния выполнялся с применением геометрических формул, учитывающих высоту камеры и угол наклона. Это позволило модели не только определить класс объекта, но и оценить его положение в пространстве, что важно для создания систем, способных функционировать в реальных условиях.

С целью нахождения центра массы выделенной маски объекта использовали моменты изображения. Это дало возможность определить координаты центра объекта на изображении. Вычисление моментов осуществлялось с помощью функции cv2.moments, которая предоставляет требуемые данные для нахождения координат центра массы (center_x, center_y).

Поскольку объект может находиться в любом месте изображения, необходимо рассчитать угловые смещения относительно центра изображения по горизонтальной и вертикальной осям (angle_offset_x, angle_offset_y). Эти смещения зависят от положения центра объекта и размеров иллюстрации (рис. 10).

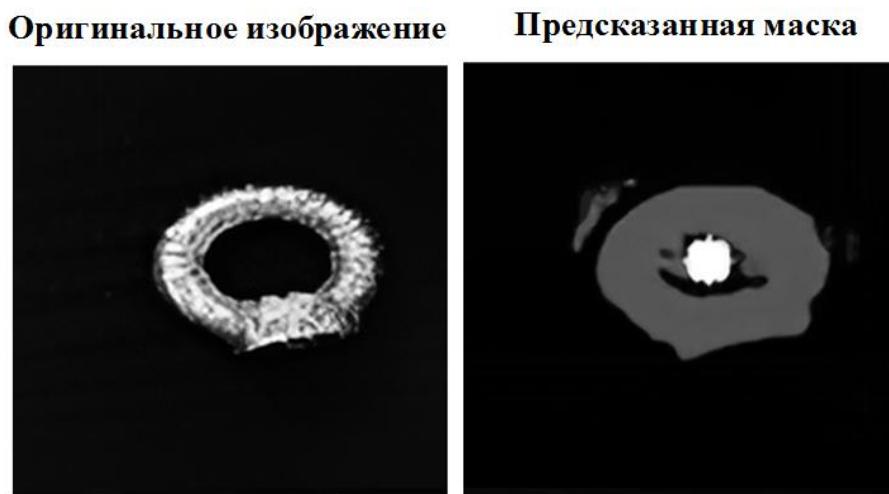


Рис. 10. Результаты распознавания расстояния до детали: горизонтального (28,26 см), прямого (39,98 см) (центр объекта находится в координатах (135, 120))

С учетом смещений производилась корректировка исходного угла наклона камеры (angle_degrees) для получения новых углов (angle_radians_x, angle_radians_y). Эти углы учитывали фактическое положение объекта на изображении.

Для каждого из скорректированных углов рассчитывались горизонтальные и прямые расстояния до объекта. Горизонтальное расстояние (d_x , d_y) определяли с использованием функции тангенса:

$$d = h \cdot \operatorname{tg} \theta,$$

где h – высота камеры; θ – угол наклона камеры.

Прямое расстояние (D_x , D_y) находили, применяя функцию косинуса:

$$D = \frac{h}{\cos \theta}.$$

Итоговые горизонтальное и прямое расстояния получали путем объединения рассчитанных компонент по осям x и y :

$$\text{horizontal_distance} = \sqrt{d_x^2 + d_y^2}, \quad \text{direct_distance} = \sqrt{D_x^2 + D_y^2}.$$

РЕЗУЛЬТАТЫ ИССЛЕДОВАНИЯ

Были получены результаты решения четырех задач:

1. Автоматизации работы минимаркета, где представлено интересное решение с использованием детектирования на изображении.
2. Детектирования простейших фигур на изображении.
3. Создания 3D-модели деталей, обучения и тестирования модели на фотографиях деталей.
4. Исследования возможностей анализа пространственных объектов.

Эти результаты подтвердили, что важно использовать высококачественные, разнородные данные для успешного обучения моделей машинного обучения и выявили проблемы, имеющиеся у существующих моделей семантической сегментации.

Использованный метод триангуляции позволил точно определить расстояние от камеры до объекта с учетом его положения на изображении и параметров съемки. В итоге была обеспечена высокая точность вычислений.

Примененный к генерации данных с 3D-деталями подход будет включаться в дальнейшие исследования, посвященные оценке 3D-позиции объекта по изображению [10–12].

ЗАКЛЮЧЕНИЕ

Анализ использования нейросетей для автоматизации торговли, распознавания и сегментации объектов выявил широкий спектр возможностей и эффективных решений задач в реальном времени.

Можно сделать вывод, что YOLOv8 оптимизирует процесс взаимодействия с микромаркетом за счет автоматического распознавания товаров, что ускоряет их приобретение и делает процесс интуитивно понятным для клиента.

YOLOv8 и U-Net позволяют распознавать и сегментировать объекты, обеспечивая детальную оценку их положения и формы. Это делает их полезными для снятия различных прикладных вопросов, когда нужно различать и локализовать множество мелких деталей.

Отличительная черта YOLOv8 – универсальность, поэтому эта нейросеть может быть востребована при решении задач детекции в реальном времени. Однако в зависимости от специфики последних и требований к точности могут использоваться различные модификации и версии данной модели. Все примененные подходы демонстрируют значение нейросетевых технологий для автоматизации в сфере торговли, промышленности и других сферах, где необходимо быстро и точно распознавать объекты.

ЛИТЕРАТУРА

1. Varghese R., Sambath M. YOLOv8: A Novel Object Detection Algorithm with Enhanced Performance and Robustness. URL: https://www.researchgate.net/publication/380840426_YOLOv8_A_Novel_Object_Detection_Algorithm_with_Enhanced_Performance_and_Robustness (дата обращения: 22.01.2025).

2. Ronneberger O., Fischer P., Brox T. U-Net: Convolutional Networks for Biomedical Image Segmentation. URL: <https://arxiv.org/pdf/1505.04597> (дата обращения: 22.01.2025).
3. CVAT.ai Corporation. (2023). Computer Vision Annotation Tool (CVAT) (Version 2.25.0) [Computer software]. URL: <https://github.com/cvat-ai/cvat> (дата обращения: 22.01.2025).
4. Powers D. Evaluation: From precision, recall and F-measure to ROC, informedness, markedness and correlation. URL: https://www.researchgate.net/publication/276412348_Evaluation_From_precision_recall_and_F-measure_to_ROC_informedness_markedness_correlation (дата обращения: 22.01.2025).
5. Roboflow. Roboflow Python Package [Computer software]. URL: <https://github.com/roboflow/roboflow-python> (дата обращения: 22.01.2025).
6. Слепова С.В., Шахина М.А. Система автоматизированного проектирования Компас-3D (Мультимедийный Курс Лекций) // Международный журнал прикладных и фундаментальных исследований. 2014. № 3 (часть 2). С. 207–208.
7. Коротцов И.Г., Казанкова А.А. Сравнительный анализ программного средства Blender для 3D-моделирования // Актуальные проблемы авиации и космонавтики. 2021. Т. 3. С. 1268–1270.
8. Jha D., Smedsrud P.H., Riegler M.A., Johansen D., Lange T., de , Halvorsen P., Johansen H.D. ResUNet++: An Advanced Architecture for Medical Image Segmentation. URL: https://www.researchgate.net/publication/337337131_ResUNet_An_Advanced_Architecture_for_Medical_Image_Segmentation (дата обращения: 22.01.2025).
9. Daniel J., Rose J.T., Vinnarasi F.S., Rajinikanth V. VGG-UNet/VGG-SegNet Supported Automatic Segmentation of Endoplasmic Reticulum Network in Fluorescence Microscopy Images. URL: https://www.researchgate.net/publication/361179414_VGG-UNet_VGG-SegNet_Supported_Automatic_Segmentation_of_Endoplasmic_Reticulum_Network_in_Fluorescence_Microscopy_Images (дата обращения: 22.01.2025).
10. Толстель О.В., Ширкин А.Е., Калабин А.Л. Проецирование технических объектов на изображении в метрическое пространство при помощи глубоких нейронных сетей для задачи детектирования // Программные продукты и системы. 2020. Т. 33. № 3. С. 464–470.
11. Толстель О.В., Ширкин А.Е., Калабин А.Л. Построение системы технического зрения для выравнивания содержимого упаковок дельта-манипулятором на пищевом производстве // Программные продукты и системы. 2023. Т. 36. № 2. С. 197–206.
12. Ширкин А.Е., Калабин А.Л., Толстель О.В. Оценка 6D-позиции объекта с выбором целевого объекта на изображении для робототехнических задач // Гибридные и синергетические интеллектуальные системы: Сборник статей по материалам VII Всероссийской Постполовской конференции / отв. ред. А.В. Колесников. СПб.: РХГА, 2024. С. 391–398.

Для цитирования: Гуляев С.А., Тарачков М.В., Толстель О.В., Ширкин А.Е. Распознавание и оценка расположения объектов при помощи нейронных сетей // Вестник Тверского государственного технического университета. Серия «Технические науки». 2025. № 2 (26). С. 66–77.