Научная статья

DOI: 10.15593/2224-9397/2024.1.09
УДК 004.89

А.Н. Кокоулин, А.А. Южаков

Пермский национальный исследовательский политехнический университет, Пермь, Российская Федерация

ДВУХСТУПЕННАЯ СХЕМА ОБНАРУЖЕНИЯ ОБЪЕКТОВ В ПОДСИСТЕМЕ МАШИННОГО ЗРЕНИЯ СЕРВИСНЫХ РОБОТОВ

Задача разработки и анализа эффективной подсистемы машинного зрения, способной обеспечить высокую точность классификации объектов, является одной из ключевых при создании сервисных роботов. Наиболее эффективным путем решения данной проблемы является повышение точности классификации и сегментации объектов. Недостатками традиционной одноступенной схемы классификации объектов на изображениях являются игнорирование контекста (структуры сцены) при поиске объектов и отсутствие жесткой привязки размеров объекта на изображении к параметрам перспективы сцены. В результате количество ложных обнаружений объектов в недопустимых позициях и ошибок сегментации с выходом за пределы объекта является неприемлемым. **Цели исследования:** разработка двухступенной схемы обработки изображений независимыми нейронными сетями с разделением системы классов и её практическая реализация в программно-аппаратном антропоморфном стоматологическом тренажере. **Методы:** основной принцип разработанной двухступенной схемы – разделение множества классов на «суперобъекты» и «вложенные объекты», при котором выполняется условие обязательного расположения вложенного объекта в границах суперобъекта. В рассматриваемой проблематике такими классами являются классы изображений зубов и результатов лечения – пломб. Нейронная сеть первой ступени обучена на множестве изображений зубов с признаками лечения и без них. Результатами ее работы являются поиск области интереса (ROI) и обрезка этой области. Нейронная сеть второго этапа обучена на качественных и некачественных изображениях пломб, и на переданном изображении ROI обнаруживает, классифицирует и сегментирует пломбу. **Результаты:** показана лучшая обнаруживающая способность в сравнении с традиционными схемами в реальных условиях (улучшение на 25 %). Подсистема машинного зрения внедрена и прошла тестовые испытания, подтвердившие теоретические результаты. **Практическая значимость:** предложенный подход может применяться для решения многих задач машинного зрения и интеллектуального видеонаблюдения, в которых можно выделить вложенные объекты, например, при анализе сложных составных объектов, при распознавании людей с оружием.

Ключевые слова: машинное зрение, обнаружение объектов, нейронные сети, YOLO, классификация, сегментация, двухступенная схема.

A.N. Kokoulin, A.A. Yuzhakov

Perm National Research Polytechnic University, Perm, Russian Federation

TWO-STAGE OBJECT DETECTION SCHEME IN COMPUTER VISION SYSTEMS OF SERVICE ROBOTS

The problem of developing and analyzing an effective machine vision subsystem, capable of providing high classification accuracy, is one of the most important for service robot design. The most efficient solution of this problem is the accuracy of classification and segmentation improvement. The disadvantages of the traditional one-stage scheme for classifying objects in images are ignoring the context (scene structure) when searching for objects and the lack of a strict link between the size of the object in the image and the perspective parameters of the scene. As a result, the number of false detections of objects in invalid positions and segmentation errors that go beyond the boundaries of the object is unacceptable. **Purpose:** the work proposes a two-stage scheme for processing photographs by independent neural networks with a division of the class system and practical implementation results in anthropomorphic dental simulator's soft- and hardware. **Methods:** the main principle of the developed two-stage scheme is the division of a set of classes into "superobjects" and "nested objects" sets, under which the condition of the mandatory location of the nested object within the boundaries of the superobject is met. In the dental simulator case these classes are classes of images of teeth and treatment results - fillings. The first stage neural network was trained on a variety of images of teeth with and without signs of treatment. The result of its work is to search for an area of interest (ROI), crop this area, and project the ROI onto the original image to obtain an image of the tooth in maximum quality. The second stage neural network is trained on high-quality and low-quality images of fillings, and on the transmitted image the ROI detects, classifies and segments the filling. **Results:** the best detecting ability is shown in comparison with traditional single-level schemes in real cases (up to 25 % improvement). The machine vision subsystem was implemented and passed tests showing the results that confirm the theory. **Practical relevance:** the proposed approach can be used to solve other problems related to computer vision and intelligent video surveillance systems, in which nested objects can be identified, for example, in complex assembled objects analysis, in weapon detection.

Keywords: computer vision, object detection, neural networks, YOLO, classification, segmentation, two-stage scheme.

Введение

Создание перспективных систем машинного зрения, обладающих высокими характеристиками, требует глубокого анализа предметной области и решения ряда научно-практических и технических проблем. Среди них ключевыми являются вопросы исследования и построения эффективной архитектуры, включающей программно-аппаратурные средства получения, предобработки и накопления исходных данных, и синтез математических методов обработки данных, оптимизированных для решения конкретных практических задач. Объектом исследования является подсистема машинного зрения (ПМЗ) для антропоморфного стоматологического тренажера [1], решающая задачу оценки качественных и количественных показателей, необходимых для приня-

тия решения о качестве проведенного лечения. Цель статьи – разработка и анализ методов, позволяющих повысить эффективность ПМЗ за счет обеспечения высокой точности классификации и сегментации результатов препарирования зубов в ротовой полости тренажера в условиях изменяющихся параметров освещения и ракурсов съемки.

На результатах, полученных от подсистемы машинного зрения, базируется выполнение алгоритма принятия решения о качестве проведенного лечения. В частности, оцениваются такие показатели качества лечения, как выбор правильного зуба и метода лечения, размеры, глубина и ориентация фрезеровки, качество выполнения работы. Наиболее перспективными методами анализа изображений являются методы искусственного интеллекта, в частности, нейронные сети (НС). Но первые попытки непосредственного использования нейронных сетей, обученных для обнаружения классов зубов и результатов лечения, не обеспечили приемлемой точности классификации и сегментации.

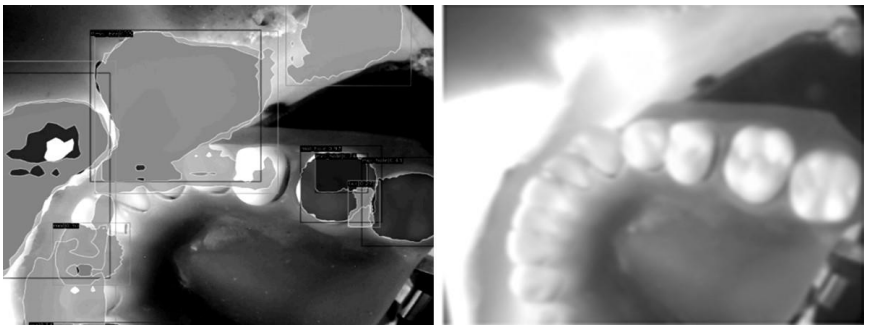


Рис. 1. Низкая точность обнаружения объектов одноступенной схемой

На рис. 1 приведены примеры ложного обнаружения и локализации объектов в ПМЗ (левое изображение), ошибочно размещенных нейронной сетью в произвольных местах полости рта. Поиск оптимальных гиперпараметров нейронных сетей [2, 3] и создание большой обучающей базы фотографий с применением техники аугментации [2] также не обеспечили устойчивости сети к ложным обнаружениям объектов. Одна из причин некачественной классификации и сегментации – сжатие исходного изображения до размеров, используемых при обучении модели, и соответствующая потеря качества исходных данных. Некоторое увеличение точности принесла технология обнаружения малых объектов SAHI [4, 5], но заметно снизилось быстродействие ПМЗ.

Анализ предметной области позволил сделать вывод о возможности представления фотографий зубов и результатов лечения как иерархии вложенных объектов. В работе предлагается двухступенная схема обработки фотографий независимыми нейронными сетями с разделением системы классов по принципу иерархии (или вложенности) объектов. Рассмотрена возможность использования двухступенной схемы для классификации, сегментации и качественной оценки результатов препарирования зубов тренажера студентами-стоматологами. Использование принципа иерархии позволило учесть контекст сцены при обнаружении объектов, а реализация этого принципа позволила повторно использовать уже созданные модели и накопленные результаты для создания качественной подсистемы машинного зрения.

Современные технологии компьютерного зрения

Развитие методов машинного обучения (AI/ML) позволяет решать широкий класс задач компьютерного зрения [6]. В частности, для решения задач обнаружения и распознавания объектов успешно используются две основные архитектуры нейронных сетей (НС):

– одноэтапные подходы, такие как SSD [7], YOLO [8], RFBNet [9];

– двухэтапные подходы, такие как Faster-RCNN [10], Mask-RCNN [11], Reasoning-RCNN [12].

При двухэтапном подходе модель определяет набор областей интереса (ROI). В следующий этап переходят только кандидаты из этого набора. Остальные области изображения считаются фоном. В одноэтапных подходах (например, архитектурах SSD и YOLO) этап выбора ROI отсутствует. Обнаружение объектов происходит непосредственно в возможных местах, которые определяются архитектурой нейронной сети. Одноэтапные подходы требуют меньше вычислительных ресурсов и позволяют обрабатывать изображения быстрее.

Анализ опубликованных работ позволяет сделать вывод, что нейросетевая модель, обученная на универсальном наборе классов, может неправильно распознавать объекты, если происходит изменение качества или ракурса изображения, или соотношение размеров объекта и самой фотографии сильно отличается от изображений в наборе данных [13]. Результаты классификации будут содержать множество ложноположительных (FP) и ложноотрицательных (FN) ошибок, что отрицательно отразится на точности всей системы.

Известны исследования по использованию пирамид характеристик изображений для сетей YOLO [8]. Feature Pyramid Network (FPN) дает значительные улучшения в качестве универсального средства извлечения функций. Нисходящая архитектура с латеральными связями разработана для построения высокоуровневых карт семантических признаков в некоторых масштабах объектов. Например, в статье [16] описывается архитектура сети, содержащей 3 головных узла (prediction heads) в YOLO v5, которые способны обнаруживать объекты, вписанные в квадраты (80×80, 40×40, 20×20) пикселей, на входном изображении размером 640×640 пикселей. Основным недостатком является то, что эти масштабы фиксированы – архитектура сети обычно состоит из 3–4 масштабов, и все очень маленькие или очень большие объекты пропускаются при обнаружении. Кроме того, не учитывается контекст сцены, и объекты могут быть ошибочно локализованы в недопустимых местах. Тем не менее имеет смысл использовать подход FPN в предлагаемой схеме на первой ступени поиска.

Некоторые исследования предлагают использовать ансамбль сетей, обученных обнаруживать объекты разных масштабов на одном и том же входном изображении [6]. Например, можно обучить три сети на трех наборах данных, первый из них содержит крупномасштабные объекты, второй – объекты среднего размера, третий – мелкие объекты. Тогда конечный результат обнаружения комбинируется из трех наборов выходных данных ансамбля НС. Этой схеме присущи те же недостатки, что и для FPN-подхода, и, кроме того, проведение трех независимых обработок троекратно увеличивает время обработки.

Кроме того, существует группа методов под названием SAHI (Slicing Aided Hyper Inference) и существует документация Ultralytics о том, как использовать YOLO с SAHI [4,17]. SAHI – это перспективная библиотека, предназначенная для оптимизации алгоритмов обнаружения объектов для крупномасштабных изображений с высоким разрешением. Основной принцип подхода заключается в разделении изображений в высоком качестве на окна-фрагменты, обнаружении объектов на каждом фрагменте и в последующем объединении результатов. Из-за проблем с производительностью и отсутствия анализа контекста этот подход из сравнения исключен.

Предлагаемое решение состоит в том, чтобы попытаться ограничить возможность присвоения объектам отдельных классов по пара-

метрам оптической перспективы вместо использования фиксированных наборов классов. Эту опцию можно реализовать, если анализируемое изображение содержит «эталонные» объекты, по которым можно оценить расстояние и размеры объектов в сцене. В задаче по обнаружению и сегментации пломб необходимо сравнивать размер пломбы с размером обработанного зуба.

Стоматологический антропоморфный тренажер

«Пропедевтика стоматологических заболеваний» – дисциплина, задачами которой является обучение студента-стоматолога широкому спектру мануальных навыков по основным разделам стоматологии, а также приобретение связанных с ними теоретических знаний. Следовательно, робот, преподающий данный курс, должен обучать теоретическим аспектам дисциплины, уметь проводить опросы по темам практических заданий и лекций, оценивать тестовые задания, а также объективно оценивать мануальные навыки студентов. В качестве базы для реализации тренажера выступает роботизированный комплекс производства «Промобот» [1].

Робот имеет свою управляющую систему на базе ROS (Robotic operational system), выполняемую в среде Ubuntu Linux на компактном компьютере внутри «тела» робота. Разрабатываемая ПМЗ и система принятия решений – это надстройка над роботом. Взаимодействие с компонентами робота осуществляется через программные и сетевые интерфейсы API с дополнительного компьютера, оснащенного графическим процессором GPU для ускорения работы нейронной сети ПМЗ. Разрабатываемый тренажер [18] предназначен для симуляции типового посещения пациентом стоматологического кабинета, поэтому, большое внимание уделяется диалоговой компоненте. Реализованы 4 сценария лечения (далее – «кейсы») – лечение кариеса, выполнение коронки, лечение канала зуба и удаление зуба. Каждый кейс имеет свой алгоритм проведения лечения, включающий, во-первых, сценарий диалога с пациентом, в результате которого будет выбран конкретный тип лечения, сторона челюсти и номер зуба, и, во-вторых, алгоритмы сбора и протоколирования видеоинформации о процессе лечения и методы оценки качества лечения.

Основа ПМЗ – программный сервис, включающий в себя HTTP-сервер для обработки запросов от подсистем робота и подсистем

«умной челюсти» с набором датчиков для управления программой опроса камер и для управления программой детекции объектов с помощью нейронной сети. За отображение результатов отвечает программа
с графическим интерфейсом, отображающая текущий вид полости рта с соответствующих ракурсов и фотографии препарируемого зуба с подсветкой результатов лечения. В полости рта установлены 5 компактных камер и система подсветки. Камеры и подсветка включаются и опрашиваются программой только в определенные моменты времени, поскольку до начала кейса у ПМЗ нет представления о том, какой сектор челюсти будет препарироваться и какие модели нейронных сетей необходимо использовать для анализа, а подсветка может мешать стоматологу проводить лечение.

Схема программно-аппаратного комплекса стоматологического тренажера представлена на рис. 2.

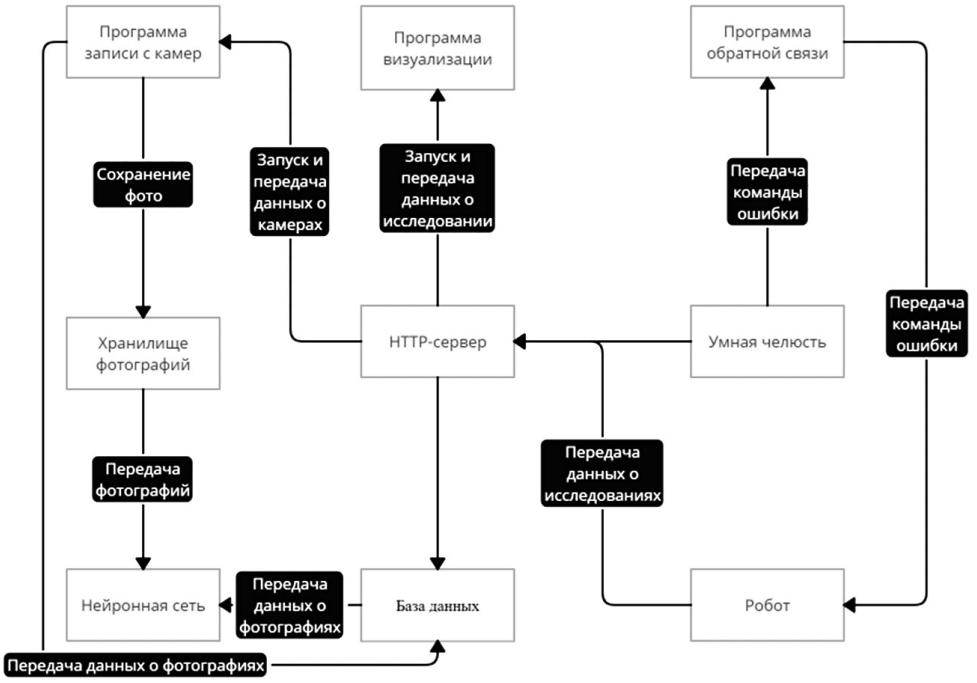


Рис. 2. Схема программно-аппаратного комплекса стоматологического тренажера

Для реализации диалога в конфигурации робота создается словарь (таблица), в котором каждой распознанной фразе в определенном контексте кейса лечения назначается свой обработчик. В обработчик

включены команды для «эмоционального» реагирования робота и команды для отправки в сервис управления распознаванием. Так, на произнесенную студентом фразу «Откиньтесь на спинку и откройте рот», происходит реакция робота в виде отработки сервоприводов «тела» и «челюсти», и, одновременно с этим, на сервис управления поступает команда начала нового кейса с указанием в параметрах вызова одного из четырех видов лечения (номера кейса).

Сервис создает новую запись в БД в таблице кейсов с заполнением номера (идентификатора) кейса, вида кейса и времени начала кейса. После того как студент определит расположение зуба для лечения, происходит реакция робота в виде фразы «я готов» и отправка команды на сервис управления с номером зуба. В зависимости от номера зуба и номера кейса сервис должен активировать блок из двух левых или двух правых камер для съемки выбранного зуба в вертикальной и горизонтальной проекции. При этом создаются первые снимки с исходными размерами и формой зуба до начала лечения и сохраняются в БД. За съемку отвечает программа, написанная на языке C++, с использованием библиотеки для получения полноразмерных фотографий с камер. Серии фотографий, иллюстрирующих разные стадии обработки зуба, сохраняются в БД. Выбор в пользу собственной библиотеки вместо популярной библиотеки OpenCV объясняется тем, что библиотека OpenCV не работает со специфическими размерами и форматами фотокамер, в частности, невозможно получить изображение в максимальном доступном разрешении 2592×1944 (5 мегапикселей) с используемых компактных камер, а изображение HD не устраивает с точки зрения качества распознавания. Кроме получения фотографий с камер программа отбраковывает фотографии, имеющие смазанные детали и проблемы с расфокусировкой камер; это достигается благодаря вычислению Лапласиана [5]. Помимо запуска программы получения фотографий с камер необходимо запустить и инициировать программу детекции, использующую свою нейронную сеть для каждого кейса.

Нейронная сеть (в первом прототипе – Mask R-CNN, в последующем – YOLO), являющаяся основой ПМЗ, была обучена на 4 больших наборах фотографий результатов правильного лечения и лечения, выполненного с ошибками, например, неровные края, непостоянная глубина, низкое качество выполнение углов и поверхностей. Специфика каждого кейса заключается в особенностях визуального анализа каче-

ственных и количественных показателей, получаемых при сравнении исходного изображения зуба и конечного результата лечения.

В частности, при лечении кариеса и лечении канала необходимо анализировать не только глубину и прочие линейные размеры препарирования, но и форму «пломбы», а при удалении зуба оценивать правильность наложения щипцов. В случае фрезерования зуба для установки коронки требуется анализировать и верхнюю, и боковую проекции зуба, чтобы оценить качество поверхности и объем снятого материала. При этом работают две нейронные сети, обученные на фотографиях результатов лечения в двух ракурсах. Удаление и лечение канала опираются не только на результаты распознавания соответствующих нейронных сетей, но и на показания разнообразных датчиков, встроенных в smart-зубы и челюсть.

Постановка задачи

Как было показано выше, на качество обработки входных изображений ПМЗ в первую очередь негативно влияет снижение разрешения изображения при его приведении к размерам модели. При невысоком соотношении размеров объектов к размеру изображения это приводит к значительному сокращению размерности множества характерных признаков объектов. Результаты классификации будут содержать большое число FP и FN ошибок. Для случая стоматологического тренажера, как, например, показано на рис.1, пломбы были ложно распознаны и локализованы вне зубов, на нёбе тренажера. Необходимо уточнить, что тривиальные решения, например, жесткое задание допустимых координат для расположения зубов и соответствующих пломб на фотографии, не работают, так как челюсть подвижна относительно системы камер, и в тренажере могут использоваться макеты челюсти как с 14, так и с 16 зубами.

Поэтому необходимы поиск и реализация методов, учитывающих специфику предметной области и базирующихся на современных технологиях компьютерного зрения.

Отсюда направлениями исследования для достижения поставленной цели являются:

– исследование возможности проведения классификации объектов с учетом контекста сцены. При этом должна быть ограничена возможность присвоения классов некоторым видам объектов путем ана-

лиза параметров оптической перспективы сцены или использован принцип иерархичности (вложенности) некоторых объектов и проведение поуровневой классификации объектов в соответствии с иерархией. На каждом уровне используется своя система классов;

– исследование возможности сохранения исходного качества изображения при передаче данных между ступенями детектора.

Прочие факторы, негативно влияющие на качество распознавания, такие как пылевое загрязнение камер, расфокусировка, изменение освещения и ракурса, могут быть устранены традиционными для нейронных сетей способами, и не рассматриваются в данной статье.

Предлагаемая новая схема двухступенной обработки изображений использует две независимые нейронные сети с разделением системы классов между ступенями распознавания, учитывает контекст сцены при классификации и сегментации объектов и адаптируется к степени удаленности объектов или изменению ракурса съемки.

Требования к ПМЗ стоматологического тренажера:

– обнаружение объектов искомых категорий вне зависимости от их размеров на изображении;

– устойчивость к изменению освещения и ракурсов съемки при открытии и закрытии челюсти тренажера;

– обработка исходных изображений в высоком разрешении с минимальной потерей качества между ступенями обработки;

– ограничение на размер модели и время обучения (модель можно обучить на GPU с объемом памяти 8–12 Гб);

– сохранение приемлемой скорости обнаружения объектов, сравнимую с скоростью исходной нейронной сети.

Область применения и перспективы использования подхода

В данной работе предлагается универсальное решение задачи классификации и сегментации объектов для таких видов предметной области, в которых к классам объектов можно применить принцип иерархии (или вложенности) с выделением классов «суперобъектов» и вложенных объектов, расположенных в границах соответствующих суперобъектов.

В случае рассматриваемого стоматологического тренажера суперклассами являются зубы разных типов, а искомыми вложенными классами – результаты лечения (пломбы, коронки, и т.д.), которые обя-

зательно располагаются в пределах зуба. Тогда нейронная сеть первой ступени распознавания будет обучена для обнаружения зубов, а нейронная сеть второй ступени будет обучена для распознавания результатов лечения только на искомом зубе, с обрезкой фона.

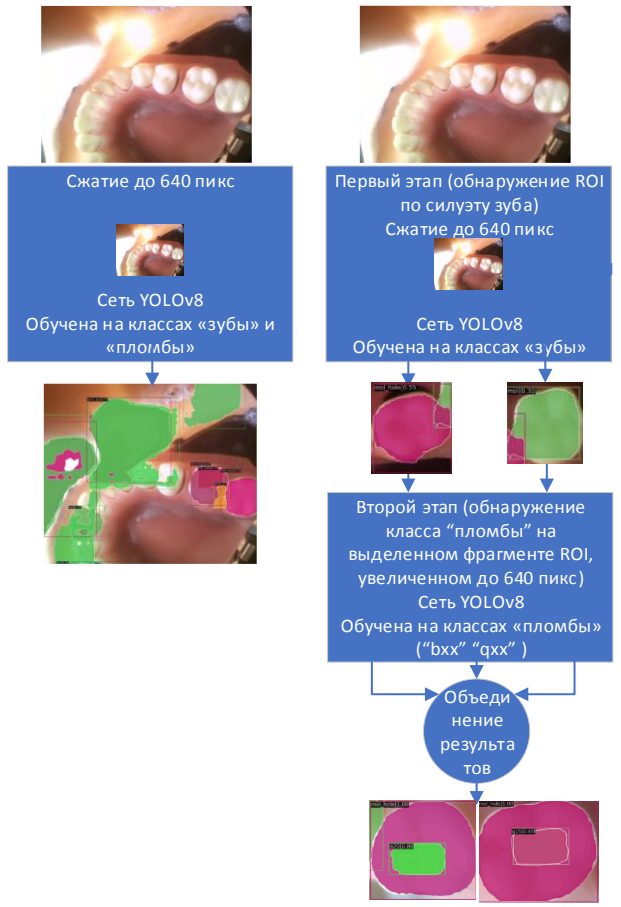
Разрабатываемый подход можно применить не только к сфере стоматологии. Существует множество производств и видов деятельности человека, где наблюдается иерархичность объектов, и к которым применим подход. Например, системы видеонаблюдения с возможностью обнаружения оружия и опасных предметов в толпе, когда люди находятся на разном удалении от камер, и, соответственно, носимое оружие может быть не различимо нейронной сетью на очень удаленных или очень приближенных позициях сцены. Но если предварительно обнаружить силуэты всех людей на снимке, нормализовать их по размерам (в пикселях), то детекция оружия будет более точной. Другие примеры – распознавание различных разноудаленных конструкций, механизмов со множеством деталей, аэрофотосъемка.

Двухступенная схема обнаружения объектов в ПМЗ

Разобьём задачу обнаружения и сегментации объектов на два этапа: на первом происходят поиск и выделение ROI «суперобъектов», как в двухэтапных нейронных сетях, наподобие Mask R-CNN, а на втором – обнаружение и сегментация искомых объектов. Основное отличие от сетей типа Mask R-CNN – это разделение системы классов и обработка данных в парадигме «coarse-fine» (от грубого анализа к точному). Mask R-CNN, взятая для примера, использует одну систему классов как для поиска ROI, так и для поиска объектов, поэтому, мелкие объекты могут быть ошибочно локализованы в произвольных местах изображения. Схема на рис. 3, *а* (далее – схема *а*) демонстрирует типовую одноуровневую архитектуру ПМЗ для обнаружения объектов.

В предлагаемой схеме (далее – схема *б*, рис. 3, *б*) отсекается возможность ложного обнаружения искомого класса объектов в произвольных местах, ограничивая работу второй нейронной сети только изображением ROI более крупного суперобъекта с отсечением фона. За счет своего более крупного размера суперобъект занимает значительную площадь на изображении и, соответственно, содержит большее множество характерных признаков, и вероятность его правильной локализации повышается. При этом суперобъекты могут не входить в множество иско-

мых классов, а относиться к множеству сопутствующих классов. Принимается, что искомые объекты могут быть только вложенными и не могут быть локализованы в пространстве вне суперобъектов, что позволяет полностью устранить ошибки, продемонстрированные на рис. 1. Кроме того, можно использовать некоторые «опорные» объекты сцены для оценки удаленности и габаритов искомых объектов.



а *б*

Рис. 3. Схемы одноступенной обработки изображений (*а*), двухступенной обработки (*б*)

Анализ эффективности двухступенной схемы

Для анализа эффективности предложенного подхода с использованием схем *а* и *б* были созданы 10 коллажей с вставкой изображения зуба с размерами в 100, 90 … 10 % от исходного. Была проведена экспери-

ментальная проверка гипотезы об эффективности двухступенной схемы *б* по сравнению с традиционными одноуровневыми подходами на базе сети YOLO. Результаты эксперимента сведены в табл. 1. Все необнаруженные пломбы относятся к сжатым до 30–10 % изображениям зубов; следует заметить, что на реальном тренажере изменение ракурса съемки до таких масштабов не может возникнуть. Нетрудно заметить, что качество обнаружения действительно увеличилось в сравнении с схемой *а*, что подтверждает гипотезу об эффективности подхода и о перспективности дальнейших исследований в этом направлении.

Таблица 1

Результаты сравнения схем *а* и *б*

№ коллажа	Двухступенная схема (схема *б*)		Схема *а*
	1. Обнаруженные ROI (зубы)	2. Обнаруженные пломбы	
1	9	**8**	6
2	10	**9**	7
3	9	**9**	8
4	10	**10**	8
5	10	**9**	7
6	10	**8**	6
7	10	**9**	7
8	10	**8**	**8**
9	9	**8**	7
10	10	**10**	9
Итог	97 из 100	**88 из 100**	73 из 100

Всего рассмотрено 10 синтетических изображений (коллажей), суммарно 100 объектов с пломбами и без них. В ячейках проставлено количество объектов, распознанных на соответствующем коллаже. В правый столбец помещены результаты распознавания результатов лечения зубов на коллажах по одноступенной схеме *а* с использованием нейронной сети YOLO, обученной на соответствующих классах. В столбец «1. Обнаруженные ROI (зубы)» помещены результаты работы первой ступени распознавания схемы *б* – поиск ROI «зубы», которые выступают в роли суперобъектов. Можно убедиться, что точность определения ROI 97 % гораздо выше, чем точность определения пломб одноступенной схемой *а* (73 %). В столбец «2. Обнаруженные пломбы» помещены результаты обнаружения результатов лечения – иско-

мых классов «пломба» внутри суперобъекта (ROI). Во всех случаях получен лучший результат по сравнению со схемой *а* – 88 %.

Таким образом, требование по обнаружению объектов искомых категорий вне зависимости от их размеров и требование к устойчивости схемы к изменению условий съемки при изменении положения челюсти выполняются, так как точность ПМЗ по обработке разномасштабных изображений выросла.

Главный недостаток схемы *в* – потенциальная потеря качества изображения в ROI, обусловленная изменением размеров исходного изображения до размеров, заданных при создании модели нейронной сети. Типовой размер для YOLO v8: 640×640, хотя можно использовать и собственный с условием, что размер в пикселях кратен 32 по высоте и ширине. Таким образом, исходное изображение высокого качества (HQ) сжимается до изображения в 640×640 пикселей (LQ). После обрезки фрагментов ROI из сжатого изображения, полученные с потерей качества изображения отправляются на вход второй ступени, что негативно отражается на точности распознавания.

В то же время, эти фрагменты ROI соответствуют областям на исходном изображении, которые имеют хороший уровень детализации. Этот факт можно использовать для повышения эффективности схемы.

Улучшенная двухступенная схема

После подтверждения экспериментальным путем перспективности подхода была разработана новая схема (рис. 4, далее – схема *в*), лишенная недостатков базовой схемы *в*.

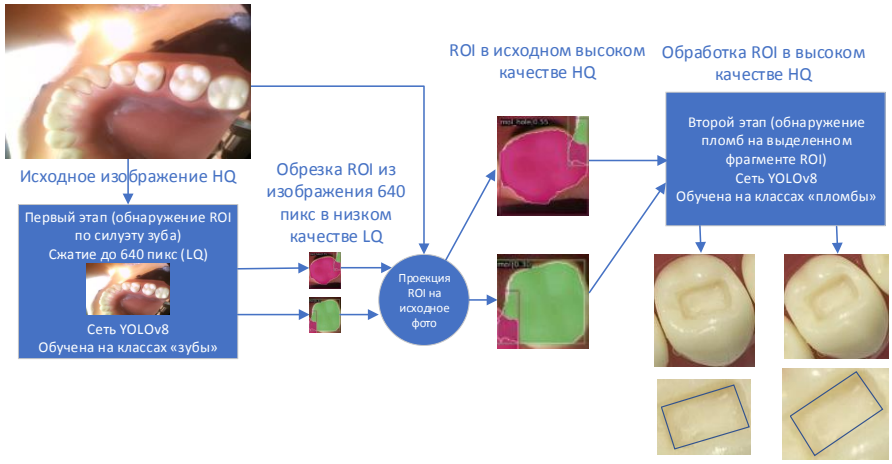


Рис. 4. Улучшенная схема двухступенной обработки изображений (схема *в*)

В схеме были добавлены промежуточные операции проецирования ROI на исходное изображение с целью получения высококачественного изображения ROI и повышения точности распознавания и сегментации на второй ступени распознавания и сегментации.

В результате это позволило выполнить требование минимальной потери качества между ступенями обработки выполняется. Оценка быстродействия схемы: по сравнению с одноступенной схемой требуется выполнить большее количество операций, но это не приводит к двукратному росту задержки.

В одноступенной схеме НС необходимо выполнить две операции на разных слоях сети: детекцию области интереса и ее сегментацию, а в двухступенной схеме можно реализовать для первой ступени только операцию детекции (обнаружения) объекта, а на второй ступени – операции детекции и сегментации. Операции детекции выполняются быстрее, чем операции сегментации, в которых необходимо каждому пикселю ROI присвоить класс объекта. Кроме того, для сегментации результата лечения можно использовать нейронную сеть с меньшим размером входного изображения, что пропорционально уменьшит задержку сегментации. Поэтому требование сохранения быстродействия ПМЗ на уровне одноступенной схемы выполняется.

Обучение нейронных сетей двухступенной схемы

Обучение нейронной сети YOLO v8 [19, 20] первой ступени проходило на основе базовой НС с применением переносимого обучения (transfer learning, TL). Такой подход позволяет получить более высокие показатели точности распознавания за меньшее количество итераций [6]. Кроме того, для сравнения была обучена сеть YOLO v8 только на классах зубов и зубов с пломбами. Оба процесса обучения выполнялись на компьютере с GPU nVidia 3080 в течение примерно одинакового срока (40 и 42 часа) на одинаковых аннотированных наборах обучающих и валидационных изображений. Для выбора наилучшей модели проведен эксперимент с использованием тестовой выборки в 500 размеченных фотографий.

Были произведёны подсчёт и сравнение результатов тестирования с доверительным порогом (confidence threshold) [21], равным 0,25. Результаты представлены в табл. 2. Для классов «зубы без пломб» и «зубы с пломбами» была оценена точность распознавания как процент

корректных распознаваий. Для анализа была также использована метрика mAP (mean average precision, средняя точность, усредненная по всем классам) [22, 23]. Кроме того, был оценен процент ложных обнаружений от общего количества кадров как один из показателей качества нейронной сети.

Таблица 2

Результаты тестирования НС первой ступени

Нейросеть	Точность, % (зубы без пломб)	Точность, % (зубы с пломбами)	mAP, %	Ошибки (ложные обнаружения), %
Сеть базовая с TL	88,14	91,42	91,26	5,58
Сеть класс «зубы»	96,50	96,25	96,38	1,55

Как видно из табл. 2, точность распознавания у нейросети, обученной только на классах зубов, выше практически на 5 %. Она лучше распознает ROI по сравнению с базовой НС.

С целью уменьшения количества ложных срабатываний было применено дополнение набора данных отрицательными примерами (negative samples), т.е. изображениями без искомых объектов – зубов, но с объектами, на которых произошло ложное срабатывание (блики, инструменты и т.д.).

Отрицательные примеры были добавлены в набор данных, который использовался для обучения новой нейронной сети [24, 25]. В результате тестирования нейронной сети процент ложных срабатываний снизился до 0,56 % .

По результатам тестирования было принято решение выбрать модель, обученную только на классах зубов, в качестве сети первой ступени. Размер входного изображения для сети был выбран стандартный 640×640 [19].

Для сети второй ступени была выбрана такая же сеть YOLO, но с размером входного изображения 416×416. Такой размер был выбран по двум причинам:

– размер зуба на исходной фотографии челюсти, как правило, имеет габариты менее 400×400 пикселей, и приведение его к стандартному размеру 640×640 не даст положительного результата, а может привести к возникновению артефактов зернистости, ступенчатости контуров;

– ускорение сегментации примерно в два раза по сравнению с входным изображением 640×640 пикселей.

Сеть второй ступени была обучена на фотографиях зубов с лечением и без него и имеет классы с названиями вида “qxx” и “bxx”, где xx – это глубина заглубления внутрь зуба. Признак хорошо выполненного лечения – буква “q” в начале названия класса, лечения с ошибками – буква ‘b” в начале имени класса. В обучающую и валидационную выборки были включены отрицательные примеры.

Для сравнения были обучены сети для схем *а*, *б* и *в* на одинаковых наборах фотографий (отличались только наборы классов). Обучающие выборки для сетей второй ступени (схем *б* и *в*) были синтезированы из фотографий для первой ступени путем обрезки ROI. Для оценки качества распознавания НС была использована та же система критериев, что и в табл. 2. Результаты тестирования представлены в табл. 3.

Таблица 3

Результаты тестирования двухступенной ПМЗ

Нейросеть	Точность, % (зубы без пломб)	Точность, % (зубы с пломбами)	mAP, %	Ложные срабатывания, %
Сеть схема A	76,44	64,19	71,31	7,58
Сеть схема B	92,93	92,02	92,87	5,57
Сеть схема C	97,60	97,12	97,20	0,56

Как видно из таблицы, последовательное улучшение структуры ПМЗ привело к увеличению точности распознавания объектов даже без тонкой настройки гиперпараметров нейронной сети, увеличения размерности модели или усложнения ее структуры. В итоговой двухступенной схеме используется исходная сеть YOLO в обеих ступенях, а увеличение точности в 25 % на реальных фотографиях происходит за счет искусственного ограничения назначения классов и локализации объектов в контексте сцены обрабатываемого изображения.

Заключение

В работе исследована двухступенная схема обработки фотографий независимыми нейронными сетями с разделением системы классов. Рассмотрена возможность использования двухступенной схемы для

классификации, сегментации и качественной оценки результатов препарирования зубов тренажера студентами-стоматологами. Недостатками традиционной одноступенной схемы классификации объектов на изображениях являются игнорирование контекста (структуры сцены) при поиске объектов и отсутствие жесткой привязки размеров объекта на изображении к параметрам перспективы сцены. В результате количество ложных обнаружений объектов в недопустимых позициях и ошибок сегментации с выходом за пределы объекта является неприемлемым. Основной принцип разработанной двухступенной схемы – разделение множества классов на «суперобъекты» и «вложенные объекты», при котором выполняется условие обязательного расположения вложенного объекта в границах суперобъекта. В рассматриваемой проблематике такими классами являются классы изображений зубов и результатов лечения – пломб. Результатом работы НС первой ступени является поиск области интереса ROI для получения изображения зуба в максимальном качестве. Нейронная сеть второго этапа на переданном изображении ROI обнаруживает и сегментирует пломбу. Показана лучшая обнаруживающая способность (улучшение до 25 %) в сравнении с традиционными схемами на тестовых примерах фотографий ротовой полости тренажера для разных ракурсов и параметров освещения. Предложенный подход может применяться также и для решения иных задач, связанных с системами машинного зрения и интеллектуального видеонаблюдения, в которых можно выделить вложенные объекты.

Библиографический список

1. Разработка комплекса «Антропоморфный стоматологический робот» с элементами искусственного интеллекта для имитации врачебных манипуляций и коммуникации «врач – пациент» / Н.Б. Асташина, А.А. Байдаров, С.Д. Арутюнов [и др.] // Пермский медицинский журнал. – 2022. – Т. 39, № 6. – С. 62–70. DOI: 10.17816/pmj39662-70

2. Akhmetzyanov, K.R. Neural network development based on knowledge about environmental influence / K.R. Akhmetzyanov, A.A. Yuzhakov, A.N. Kokoulin // Proceedings of the 2020 IEEE Conference of Russian Young Researchers in Electrical and Electronic Engineering, EIConRus 2020, St. Petersburg and Moscow, 27–30.01.2020. – St. Petersburg and Moscow: Institute of Electrical and Electronics Engineers Inc., 2020. – P. 218–221. DOI: 10.1109/EIConRus49466.2020.9039226

3. Оптимизация вычислений нейронной сети / К.Р. Ахметзянов, А.И. Тур, А.Н. Кокоулин, А.А. Южаков // Вестник Пермского национального исследовательского политехнического университета. Электротехника, информационные технологии, системы управления. – 2020. – № 36. – С. 117–130. DOI: 10.15593/2224-9397/2020.4.07

4. Akyon, F.C. Slicing aided hyper inference and fine-tuning for small object detection / F.C. Akyon, S.O. Altinuc, A. Temizel // 2022 IEEE International Conference on Image Processing (ICIP). – IEEE, 2022. – P. 966–970.

5. The optical method for the plastic waste recognition and sorting in a reverse vending machine / A.N. Kokoulin, A.A. Yuzhakov, A.I. Tur, A.I. Knyazev // International Multidisciplinary Scientific GeoConference Surveying Geology and Mining Ecology Management, SGEM. – 19 (4.1). – pp. 793–800.

6. Object detection with deep learning: A review / Z.Q. Zhao, P. Zheng, S.T. Xu, X. Wu // IEEE transactions on neural networks and learning systems. – 2019. – 30 (11). – P. 3212–3232.

7. Ssd: Single shot multibox detector / W. Liu, D. Anguelov, D. Erhan, C. Szegedy, S. Reed, C.Y. Fu, A.C. Berg // European conference on computer vision, Springer, Cham, 2016. – P. 21–37.

8. You only look once: Unified, realtime object detection / J. Redmon, S. Divvala, R. Girshick, A. Farhadi // Proceedings of the IEEE conference on computer vision and pattern recognition. – 2016. – P. 779–788.

9. RFBNet: deep multimodal networks with residual fusion blocks for RGB-D semantic segmentation / L. Deng, M. Yang, T. Li, Y. He, C. Wang. – 2019, arXiv: 1907.00135

10. Faster r-cnn: Towards real-time object detection with region proposal networks / S. Ren, K. He, R. Girshick, J. Sun // Advances in neural information processing systems. – 2015. – Vol. 28.

11. Mask r-cnn / K. He, G. Gkioxari, P. Dollár, R. Girshick // Proceedings of the IEEE international conference on computer vision. – 2017. – P. 2961–2969.

12. Reasoning-rcnn: Unifying adaptive global reasoning into largescale object detection / H. Xu, C. Jiang, X. Liang, L. Lin, Z. Li // Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. – 2019. – P. 6419–6428.

13. Real-time gun detection in cctv: An open problem / J. Gonzalez, C. Zaccaroa, J. Alvarez Garcia, L. Morilloa, F. Caparrinib // Neural Networks. – 2020. – Vol. 132. – P. 297–308.

14. A dataset and system for real-time gun detection in surveillance video using deep learning / Q. Delong, T. Weijun, L. Zhifu, Y. Qi, L. Jingfeng. – 2022. arXiv:2105.01058

15. Redmon, J. Yolov3: An incremental improvement / J. Redmon, A. Farhadi. – 2018. arXiv preprint. arXiv:1804.02767.

16. Tang, Sh. HIC-YOLOv5: Improved YOLOv5 for Small Object Detection / Sh. Tang, Y. Fang, S. Zhang. – 2023. arXiv:2309.16393v1 [cs.CV].

17. NICE: CVPR 2023 Challenge on Zero-shot Image Captioning / T. Kim, P. Ahn, S. Kim [et al.]. – 2023. arXiv:2309.01961v3 [cs.CV]. DOI: https://doi.org/10.48550/arXiv.2309.01961

18. Разработка антропоморфного стоматологического симулятора на базе Robo-C / А.А. Южаков, С.Д. Арутюнов, Н.Б. Асташина, А.А. Байдаров, И.И. Безукладников, С.А Сторожев // Вестник ИжГТУ им. М.Т. Калашникова. – 2023. – Т. 26, № 4. – С. 13–22. DOI: 10.22213/2413-1172-2023-4-13-22

19. Real-time object detection and tracking for mobile robot using YOLOV8 and STRONG SORT / Le Ba Chung, Nguyen Duc Duy // Universum: технические науки. – 2023. – № 11-6 (116). DOI: 10.32743/UniTech.2023.116.11.16223

20. What is YOLOv8? The Ultimate Guide [Электронный ресурс]. – URL: https://blog.roboflow.com/whats-new-in-yolov8/ (дата обращения: 13.06.2023).

21. ImageNet large scale visual recognition challenge / O. Russakovsky, J. Deng, H. Su, J. Krause, S. Satheesh, S. Ma, Z. Huang, A. Karpathy, A. Khosla, M. Bernstein, A.C. Berg, L. Fei-Fei // IJCV, 2015

22. mAP (mean Average Precision) в детекции объектов // Neti ML – Machine Learning Neti. Умная видеоаналитика [Электронный ресурс]. – URL: http://ml.i-neti.ru/map-mean-average-precision (дата обращения: 06.06.2022).

23. Shrivastava, A. Training regionbased object detectors with online hard example mining / A. Shrivastava, A. Gupta, R. Girshick // CVPR. – 2016.

24. Weapon detection in real-time CCTV videos using deep learning / M.T. Bhatti, M.G. Khan, M. Aslam, M.J. Fiaz // IEEE Access. – 2021. – Vol. 9. – P. 34366–4382. DOI: 10.1109/ACCESS.2021.3059170

25. Zahrawi, M. Improving video surveillance systems in banks using deep learning techniques / M. Zahrawi, K. Shaalan // Scientific Reports. – 2023. – Vol. 13. – P. 7911. DOI: 10.1038/s41598-023-35190-9

References

1. Astashina N.B., Baidarov A.A., Arutiunov S.D. et al. Razrabotka kompleksa “Antropomorfnyi stomatologicheskii robot” s elementami iskusstvennogo intellekta dlia imitatsii vrachebnykh manipuliatsii i kommunikatsii “vrach - patsient” [The development of the complex “Antropomorphic dental robot” with elements of the artificial intellect for clinical manipulations and “doctor-patient” communications]. *Permskii meditsinskii zhurnal*, 2022, vol. 39, no. 6, pp. 62-70. DOI: 10.17816/pmj39662-70

2. Akhmetzyanov K.R., Yuzhakov A.A., Kokoulin A.N. Neural network development based on knowledge about environmental influence. *Proceedings of the 2020 IEEE Conference of Russian Young Researchers in Electrical and Electronic Engineering, EIConRus 2020, St. Petersburg and Moscow, 27-30.01.2020*. St. Petersburg and Moscow: Institute of Electrical and Electronics Engineers Inc., 2020, pp. 218-221. DOI: 10.1109/EIConRus49466.2020.9039226

3. Akhmetzianov K.R., Tur A.I., Kokoulin A.N. Iuzhakov, A.A. Optimizatsiia vychislenii neironnoi seti [Optimization of neural networks]. *Vestnik Permskogo natsional'nogo issledovatel'skogo politekhnicheskogo universiteta. Elektrotekhnika, informatsionnye tekhnologii, sistemy upravleniia*, 2020, no. 36, pp. 117-130. DOI: 10.15593/2224-9397/2020.4.07

4. Akyon F.C., Altinuc S.O., Temizel A. Slicing aided hyper inference and fine-tuning for small object detection. *2022 IEEE International Conference on Image Processing (ICIP)*. IEEE, 2022, pp. 966-970.

5. Kokoulin A.N., Yuzhakov A.A., Tur A.I., Knyazev A.I. The optical method for the plastic waste recognition and sorting in a reverse vending machine. *International Multidisciplinary Scientific GeoConference Surveying Geology and Mining Ecology Management, SGEM*, 19 (4.1), pp. 793-800.

6. Zhao Z.Q., Zheng P., Xu S.T., Wu X. Object detection with deep learning: A review. *IEEE transactions on neural networks and learning systems*, 2019, 30 (11), pp. 3212-3232.

7. Liu W., Anguelov D., Erhan D., Szegedy C., Reed S., Fu C.Y., Berg A.C. Ssd: Single shot multibox detector. *European conference on computer vision, Springer, Cham*, 2016, pp. 21-37.

8. Redmon J., Divvala S., Girshick R., Farhadi A. You only look once: Unified, realtime object detection. *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 779-788.

9. Deng L., Yang M., Li T., He Y., Wang C. RFBNet: deep multimodal networks with residual fusion blocks for RGB-D semantic segmentation, 2019, arXiv: 1907.00135

10. 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, vol. 28.

11. He K., Gkioxari G., Dollár P., Girshick R. Mask r-cnn. *Proceedings of the IEEE international conference on computer vision*, 2017, pp. 2961-2969.

12. Xu H., Jiang C., Liang X., Lin L., Li Z. Reasoning-rcnn: Unifying adaptive global reasoning into largescale object detection. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2019, pp. 6419-6428.

13. Gonzalez J., Zaccaroa C., Alvarez Garcia J., Morilloa L., Caparrinib F. Real-time gun detection in cctv: An open problem. *Neural Networks*, 2020, vol. 132, pp. 297-308.

14. Delong Q., Weijun T., Zhifu L., Qi Y., Jingfeng L. A dataset and system for real-time gun detection in surveillance video using deep learning, 2022. arXiv:2105.01058

15. Redmon J., Farhadi A. Yolov3: An incremental improvement, 2018. arXiv preprint. arXiv:1804.02767.

16. Tang Sh., Fang Y., Zhang S. HIC-YOLOv5: Improved YOLOv5 for Small Object Detection, 2023. arXiv:2309.16393v1 [cs.CV].

17. Kim T., Ahn P., Kim S. et al. NICE: CVPR 2023 Challenge on Zero-shot Image Captioning, 2023. arXiv:2309.01961v3 [cs.CV]. DOI: https://doi.org/10.48550/arXiv.2309.01961

18. Iuzhakov A.A., Arutiunov S.D., Astashina N.B., Baidarov A.A., Bezukladnikov I.I., Storozhev S.A. Razrabotka antropomorfnogo stomatologicheskogo simuliatora na baze Robo-C [The development of anthropomorphic dental simulator based on Robo-C]. *Vestnik Izhevskogo gosudarstvennogo tekhnicheskogo universiteta imeni M.T. Kalashnikova*, 2023, vol. 26, no. 4, pp. 13-22. DOI: 10.22213/2413-1172-2023-4-13-22

19. Le Ba Chung, Nguyen Duc Duy. Real-time object detection and tracking for mobile robot using YOLOV8 and STRONG SORT. *Universum: tekhnicheskie nauki*, 2023, no. 11-6 (116). DOI: 10.32743/UniTech.2023.116.11.16223

20. What is YOLOv8? The Ultimate Guide, available at: https://blog.roboflow.com/whats-new-in-yolov8/ (accessed 13 June 2023).

21. Russakovsky O., Deng J., Su H., Krause J., Satheesh S., Ma S., Huang Z., Karpathy A., Khosla A., Bernstein M., Berg A.C., Fei-Fei L. ImageNet large scale visual recognition challenge. *IJCV*, 2015.

22. mAP (mean Average Precision) в детекции объектов [mAP (mean Average Precision) in Object Detection]. Neti ML - Machine Learning Neti. Smart video analytics, available at: http://ml.i-neti.ru/map-mean-average-precision (accessed 06 June 2022).

23. Shrivastava A., Gupta A., Girshick R. Training regionbased object detectors with online hard example mining. *CVPR*, 2016.

24. Bhatti M.T., Khan M.G., Aslam M., Fiaz M.J. Weapon detection in real-time CCTV videos using deep learning. *IEEE Access*, 2021, vol. 9, pp. 34366-4382. DOI: 10.1109/ACCESS.2021.3059170

25. Zahrawi M., Shaalan K. Improving video surveillance systems in banks using deep learning techniques. *Scientific Reports*, 2023, vol. 13, 7911 p. DOI: 10.1038/s41598-023-35190-9

Сведения об авторах

Южаков Александр Анатольевич (Пермь, Российская Федерация) – доктор технических наук, профессор, профессор кафедры «Автоматика и телемеханика» Пермского национального исследовательского политехнического университета (614990, Пермь, Комсомольский пр., 29, e-mail: uz@at.pstu.ru).

Кокоулин Андрей Николаеввич (Пермь, Российская Федерация) – кандидат технических наук, доцент кафедры «Автоматика и телемеханика» Пермского национального исследовательского политехнического университета (614990, Пермь, Комсомольский пр., 29, e-mail: a.n.kokoulin@at.pstu.ru).

About the authors

Aleksandr A. Uzhakov (Perm, Russian Federation) – Doctor of Technical Sciences, Professor, Department of Automation and Telemechanics Perm National Research Polytechnic University (614990, Perm, 29, Komsomolsky pr., e-mail: uz@at.pstu.ru).

Andrey N. Kokoulin (Perm, Russian Federation) – Ph. D. in Technical Sciences, Ass. Professor, Department of Automation and Telemechanics Perm National Research Polytechnic University (614990, Perm, 29, Komsomolsky pr., e-mail: a.n.kokoulin@at.pstu.ru).

Поступила: 06.03.2024. Одобрена: 18.03.2024. Принята к публикации: 20.04.2024.

Финансирование. Исследование не имело спонсорской поддержки.

Конфликт интересов. Авторы заявляют об отсутствии конфликта интересов по отношению к статье.

Вклад авторов. Авторы сделали равноценный вклад в подготовку статьи.

Просьба ссылаться на эту статью в русскоязычных источниках следующим образом:

Кокоулин, А.Н. Двухступенная схема обнаружения объектов в подсистеме машинного зрения сервисных роботов / А.Н. Кокоулин, А.А. Южаков // Вестник Пермского национального исследовательского политехнического университета. Электротехника, информационные технологии, системы управления. – 2024. – № 49. – С. 176–199. DOI: 10.15593/2224-9397/2024.1.09

Please cite this article in English as:

Kokoulin A.N., Yuzhakov A.A. Two-stage object detection scheme in computer vision systems of service robots. *Perm National Research Polytechnic University Bulletin. Electrotechnics, information technologies, control systems*, 2024, no. 49, pp. 176-199. DOI: 10.15593/2224-9397/2024.1.09