Инженерный вестник Дона, №9 (2025)
ivdon.ru/ru/magazine/archive/n9y2025/10396

Анализ глубоких нейронных сетей для обнаружения человека на земле с высоты полета квадрокоптера

Р.Р. Ахметзянова, Н.В. Андреянов

Казанский национальный исследовательский технический университет имени А.Н. Туполева, Казань

Аннотация: В современном мире, когда технологии развиваются с невероятной скоростью, компьютеры обрели способность «видеть» и воспринимать окружающий мир подобно человеку. Это привело к революции в анализе и обработке визуальных данных. Одним из ключевых достижений стало применение компьютерного зрения для поиска объектов на фотографиях и видео. Благодаря этим технологиям можно не только находить такие объекты как люди, автомобили или животные, но и точно указывать их положение с помощью ограничивающих рамок или масок для сегментации. В данной статье подробно рассматриваются современные модели глубоких нейронных сетей, применяемые для детекции человека на изображениях и видео, снятых с высоты и большого расстояния на сложном фоне. Анализируются архитектуры Faster Region-based Convolutional Neural Network (Faster R-CNN), Mask Region-based Convolutional Neural Network (Mask R-CNN), Single Shot Detector (SSD) и You Only Look Once (YOLO), сравниваются их точность, скорость и способность эффективно выявлять объекты в условиях неоднородного фона. Особое внимание уделено изучению особенностей каждой модели в конкретных практических ситуациях, где важны и высокое качество обнаружения целевых объектов, и скорость обработки изображений.
Ключевые слова: машинное обучение, искусственный интеллект, глубокое обучение, сверточные нейронные сети, детекция человека, компьютерное зрение, обнаружение объектов, обработка изображений.

В век стремительного развития технологий машины научились «видеть» и понимать окружающую среду так же, как это делает человек. И это абсолютно изменило процесс анализа и работы с визуальными данными. Стало важным направлением развития использование технологий компьютерного зрения для обнаружения объектов на изображениях и видео. Данная технология позволяет не только находить целевые объекты (например, людей, транспортные средства, животных), но и определять их точное местоположение с помощью ограничивающих рамок или масок сегментации.

Существует множество подзадач обнаружения объектов:

1. Классификация – присвоение единой метки или категории всему изображению на основе его визуального содержания.

2. Локализация – определение точного местоположения объекта на изображении или видеокадре. Локализация позволяет установить, где именно находится объект, выделяя его с помощью ограничивающей рамки.

3. Сегментация – метод разделения цифрового изображения на отдельные части, которые называются сегментами. Такой подход позволяет упростить изображение, что облегчает его последующую обработку и анализ [1].

В этой работе рассматривается задача обнаружения человека на изображениях, сделанных сверху и снятых с большого расстояния, с использованием глубоких нейронных сетей. Существует два основных подхода к обнаружению объектов: двухэтапные и одноэтапные методы [2].

Двухэтапные методы (их ещё называют методами на основе регионов) сначала выделяют на изображении области с объектами. Это делается с помощью селективного поиска или специальных слоев нейронной сети. Затем эти области анализируются: классификатор определяет класс объекта, а регрессор уточняет положение ограничивающих рамок.

В одноэтапных методах нет отдельного шага по поиску регионов. Вместо этого сеть обрабатывает все входное изображение один раз, предсказывает координаты ограничивающих рамок, а также определяет класс объекта и баллы доверия, после чего корректирует положение рамок.

К двухэтапным детекторам относятся следующие модели: Fast Region-based Convolutional Neural Network (Fast R-CNN), Faster Region-based Convolutional Neural Network (Faster R-CNN) и Mask Region-based Convolutional Neural Network (Mask R-CNN) и другие [2].

Fast Region-based Convolutional Neural Network – это усовершенствованный подход к обнаружению объектов, пришедший на

смену классическому Region-based Convolutional Neural Network (R-CNN). Его основное преимущество – извлечение признаков из всего изображения, а не из отдельных выделенных областей, как это делалось в Region-based Convolutional Neural Network. Это позволяет уменьшить количество вычислений. Также в Fast Region-based Convolutional Neural Network обучение нейронной сети, классификатора и регрессора ограничивающих рамок происходит одновременно с помощью единой функции потерь.

Faster R-CNN – усовершенствованная версия Fast Region-based Convolutional Neural Network, которая решает проблему медленной генерации региональных предложений [3,4]. В Faster R-CNN включен отдельный модуль, который отвечает за определение потенциальных областей на изображении, в которых с высокой вероятностью находятся объекты для последующего анализа [3,4].

Mask R-CNN – это улучшенная версия алгоритма Faster R-CNN. Данный метод был предложен в 2017 году и стал популярным благодаря возможности одновременно выполнять задачи обнаружения объектов и семантической сегментации [5]. В Mask R-CNN добавляется ветвь, которая предсказывает маски сегментации для каждого объекта, обнаруженного на изображении. Это позволяет выделять контуры на снимках, что полезно в задачах, требующих детальной информации о форме объектов [5].

К одноэтапным детекторам относятся модели You Only Look Once (YOLO), Single Shot Detector (SSD), и другие.

YOLO – одна из самых известных и простых моделей для решения задачи детекции объектов в режиме реального времени. Модель YOLO была разработана Джозефом Редмоном и Али Фархади в 2015 году. В последующие годы появлялись новые версии: YOLOv2, YOLOv3, YOLOv4, YOLOv5, YOLOv6 и др. [6,7]. YOLO обрабатывает изображение за один проход и это значительно ускоряет процесс обнаружения объектов [6,7].

Принцип работы модели заключается в том, что сначала меняется размер изображения, потом оно разбивается на сетку, происходит извлечение признаков, далее каждая ячейка предсказывает несколько ограничивающих рамок и вероятности принадлежности объектов к определенным классам. Далее модель объединяет все предсказания для вывода конечных результатов и также отсеиваются лишние предсказания. YOLO оптимизирован для работы в реальном времени, поэтому модель подходит для систем, требующих быстрой обработки, таких как видеонаблюдение или автономные системы и транспортные средства [6,8].

Single Shot Detector – это модель для обнаружения объектов на визуальных данных, выполняющая детекцию за один проход по входному изображению. SSD выделяется среди других моделей тем, что использует несколько уровней свёрточных слоёв с разным пространственным разрешением, что позволяет определять объекты разных размеров: более глубокие слои отвечают за крупные объекты, а более ранние – за мелкие [9,10]. Изображение проходит через базовую сеть и дополнительные сверточные слои, генерируя карты признаков разного масштаба. На каждой карте признаков применяются свертки для предсказания ограничивающих рамок и классов. После предсказания используется модуль немаксимального подавления для удаления дублирующихся ограничивающих рамок. В результате остаются только наиболее уверенные предсказания [9,10]. Эта модель широко применяется в системах видеонаблюдения, робототехнике и других областях, где важна быстрая и точная детекция объектов в реальном времени.

Эти модели хорошо справляются с задачей обнаружения человека на изображениях, благодаря современным архитектурам. Также эти модели обучены на больших наборах данных, где люди изображены в разных позах, с всевозможных ракурсов, на сложных фонах. Модель Faster R-CNN

обеспечивает высокую точность детекции объектов благодаря двухступенчатой архитектуре, а Mask R-CNN расширяет её возможности, добавляя сегментацию объектов. В SSD300 одноступенчатая архитектура, поэтому модель можно применять в режиме реального времени. YOLOv8 демонстрирует высокую скорость и точность, поэтому подходит для приложений, требующих мгновенного обнаружения объектов. Все модели адаптируемы и способны работать с объектами разных размеров, что делает их универсальными для различных задач в компьютерном зрении.

Для выбора оптимального алгоритма детекции людей на изображениях было принято решение протестировать предобученные модели на тестовых изображениях в количестве 500 штук. При проверке на тестовых изображениях сравнивались метрики: точность, полнота, F1-мера, средняя уверенность и среднее время обработки одного изображения (см. табл. 1).

Таблица № 1

Сравнительный анализ предобученных моделей на тестовой выборке 1

	Faster R-CNN	Mask R-CNN	SSD300	YOLOv8
Точность	0,57	0,59	0,01	0,62
Полнота	0,47	0,51	0,28	0,46
F1-мера	0,51	0,54	0,02	0,51
Средняя уверенность	0,37	0,46	0,12	0,45
Среднее время (в секундах)	3,4	3,6	0,9	0,2

Точность показывает долю правильно обнаруженных людей среди всех предсказанных боксов, полнота определяет долю найденных людей среди всех реальных людей на изображениях, F1-мера – это гармоническое среднее между точностью и полнотой [11]. Средняя уверенность – это среднее значение уверенности модели для всех предсказанных ограничивающих рамок, относящихся к классу "человек" [11].

Опираясь на таблицу 1, можно сделать вывод, что Faster R-CNN пропускает много людей, так как низкое значение полноты, у модели SSD300 наименьшая точность (1%) и уверенность (0.12), что означает слишком много ложных срабатываний. Mask R-CNN имеет лучшее значение F1-меры (0.54), но модель достаточно медленная (7.6 сек). YOLOv8 имеет хорошую точность (62%), но низкое значение полноты (43%), что означает пропуски людей на изображениях. При этом YOLOv8 продемонстрировала самое быстрое время обработки изображения.

При тестировании предобученных моделей на снимках, снятых камерой с близкого расстояния, модели показали хорошие результаты (см. табл. 2).

Таблица № 2

Сравнительный анализ предобученных моделей на тестовой выборке 2

	Faster R-CNN	Mask R-CNN	SSD300	YOLOv8
Точность	0.9017	0.9297	0.8889	0.7353
Полнота	0.9623	0.9811	0.9623	0.9615
F1-мера	0.9310	0.9547	0.9241	0.8333
Средняя уверенность	0.8945	0.9326	0.8977	0.7579
Среднее время (в секундах)	2.4222	2.6009	0.2746	0.2

В данном случае Mask R-CNN демонстрирует наилучшие результаты по точности, полноте и F1-мере, что делает её наиболее эффективной моделью для решения задачи. Но по времени обработки одного изображения уступает модели YOLOv8. Faster R-CNN и SSD300 показывают схожие результаты по точности и полноте, но SSD300 значительно быстрее (0.27 сек против 2.42 сек). YOLOv8 в сравнении с другими моделями уступает по точности и F1-мере, однако имеет высокую полноту и самое низкое время обработки (0.2 сек), поэтому YOLOv8 хорошо подходит для задач, требующих обработки визуальных данных в режиме реального времени.

Таким образом, для повышения точности детекции человека на фотографиях, сделанных с верхней точки обзора и с дальнего расстояния на сложном фоне, исходя из результатов, были выбраны модели Mask R-CNN, так как показывает более высокую точность, но в то же время тратит довольно много времени на обработку изображения, и YOLOv8 в том случае, если нужна высокая скорость обработки. Далее будут обучены и протестированы эти модели. Также для улучшения результатов обучения будет применена аугментация данных и наложение фильтров для улучшения видимости человека.

Литература

1. Лукашик, Д. В. Анализ современных методов сегментации изображений // Экономика и качество систем связи. – 2022. – № 2(24). – С. 57-65.

2. Дэвис Рой, Терк Мэтью Компьютерное зрение. Современные методы и перспективы развития / пер. с англ. В. С. Яценкова. – М.: ДМК Пресс, 2022. – 690 с.: ил.;

3. Shaoqing Ren, Kaiming He, Ross Girshick, and Jian Sun. Faster R-CNN: Towards RealTime Object Detection with Region Proposal Networks URL: arxiv.org/pdf/1506.01497.pdf (дата обращения: 25.06.2025).

4. Орлов Р. М., Симонов В. Л. Методы и алгоритмы обнаружения объектов в реальном времени для систем видеонаблюдения // Наука. Производство. Образование - 2024: Материалы IV Всероссийской научно-технической конференции, Москва, 19 ноября 2024 года. – Российский государственный социальный университет, 2025. – С. 264-266.

5. Сулицкий М. В., Зеленский И. С., Садовникова Н. П., Финогеев А.Г., Катеринина С.Ю. Разработка интеллектуальной системы распознавания объектов для решения задач ситуационного управления в городе //

Современные наукоемкие технологии. – 2023. – № 7. – С. 104-109. – DOI 10.17513/snt.39702.

6. Вильданов, А. Н. Генерация датасетов для учебных задач компьютерного зрения // Инженерный вестник Дона. – 2023. – № 4. URL: ivdon.ru/ru/magazine/archive/n4y2023/8320/.

7. Алеворян, А. И., Боровик И. Г., Муратова Д. Э. Современные алгоритмы обнаружения малых объектов с их частичным перекрытием // Наукосфера. – 2024. – № 5-2. – С. 229-235. – DOI 10.5281/zenodo.11490219

8. Ерохин, Д. Ю., Ершов М. Д. Современные сверточные нейронные сети для обнаружения и распознавания объектов // Цифровая обработка сигналов. – 2018. – № 3. – С. 64-69.

9. Wei Liu, Dragomir Anguelov, Dumitru Erhan, Christian Szegedy, Scott Reed, Cheng-Yang Fu, Alexander C. Berg SSD: Single Shot MultiBox Detector. In: Leibe, B., Matas, J., Sebe, N., Welling, M. (eds) Computer Vision – ECCV 2016. ECCV 2016. Lecture Notes in Computer Science(), vol 9905. Springer, Cham.

10. Визильтер, Ю. В., Горбацевич В. С., Моисеенко А. С. Однопроходный алгоритм обнаружения и распознавания лиц на основе сверточных нейронных сетей // Вестник компьютерных и информационных технологий. – 2021. – Т. 18, № 4(202). – С. 11-20. – DOI 10.14489/vkit.2021.04.pp.011-020.

11. Лебедев Б. К., Лебедев О. Б., Черкасов Р. И. Использование нейронных сетей для решения задач компьютерного зрения // Инженерный вестник Дона. – 2025. – № 2. – URL: ivdon.ru/ru/magazine/archive/n2y2025/9870/.

References

1. Lukashik, D. V. Ekonomika i kachestvo sistem svyazi. 2022. № 2(24). pp. 57-65.

2. Devis R., Terk M. Komp'yuternoe zrenie. Sovremennye metody i perspektivy razvitiya [Modern methods and development prospects]. Per. s angl. V. S. Yacenkova. M.: DMK Press, 2022. 690 p.: il.

3. Shaoqing Ren, Kaiming He, Ross Girshick, and Jian Sun. URL: arxiv.org/pdf/1506.01497.pdf (date assessed 25.06.2025).

4. Orlov R. M., Simonov V. L. Nauka. Proizvodstvo. Obrazovanie 2024: Materialy IV Vserossijskoj nauchno-tekhnicheskoj konferencii, Moskva, 19 noyabrya 2024 goda. Rossijskij gosudarstvennyj social'nyj universitet, 2025. pp. 264-266.

5. Sulickij M. V., Zelenskij I. S., Sadovnikova N. P., Finogeev A.G., Katerinina S.Yu. Sovremennye naukoemkie tekhnologii. 2023. № 7. pp. 104-109. DOI 10.17513/snt.39702.

6. Vil'danov A. N. Inzhenernyj vestnik Dona, 2023, № 4. URL: ivdon.ru/ru/magazine/archive/n4y2023/8320/.

7. Alevoryan, A. I., Borovik I. G., Muratova D. E. Naukosfera. 2024. № 5-2. pp. 229-235. DOI 10.5281/zenodo.11490219

8. Erohin, D. Yu., Ershov M. D. Cifrovaya obrabotka signalov. 2018. № 3. pp. 64-69.

9. Wei Liu, Dragomir Anguelov, Dumitru Erhan, Christian Szegedy, Scott Reed, Cheng-Yang Fu, Alexander C. Berg In: Leibe, B., Matas, J., Sebe, N., Welling, M. (eds) Computer Vision ECCV 2016. ECCV 2016. Lecture Notes in Computer Science (), vol 9905. Springer, Cham.

10. Vizil'ter, Yu. V., Gorbacevich V. S., Moiseenko A. S. Vestnik komp'yuternyh i informacionnyh tekhnologij. 2021. T. 18, № 4(202). pp. 11-20. DOI 10.14489/vkit.2021.04. pp.011-020.

11. Lebedev B. K., Lebedev O. B., Cherkasov R. I. Inzhenernyj vestnik Dona. 2025. № 2. URL: ivdon.ru/ru/magazine/archive/n2y2025/9870/.

Дата поступления: 13.07.2025

Дата публикации: 25.08.2025