

## МЕТОДЫ ИСКУССТВЕННОГО ИНТЕЛЛЕКТА

УДК 004.932.2+004.8

### ИСПОЛЬЗОВАНИЕ МЕТОДОВ КОМПЬЮТЕРНОГО ЗРЕНИЯ ДЛЯ ОБНАРУЖЕНИЯ ЛЮДЕЙ В ЛЕСОСТЕПНОЙ МЕСТНОСТИ

© А. Л. Осипов, С. Н. Терещенко

*Новосибирский государственный университет экономики и управления,  
630099, г. Новосибирск, ул. Каменская, 56  
E-mail: alosip@mail.ru*

Исследован подход применения методов обучения для настройки моделей нейронных сетей, которые используются для автоматического обнаружения людей в лесостепной местности на изображениях. Для обнаружения применялись модели нейронных сетей YOLOv51, DenseNet, CenterNet, InceptionV3, Xception, Faster R-CNN. Среди представленных моделей наилучшие результаты продемонстрировала дообученная нейронная сеть YOLOv51, которая позволила с точностью 0,8795 на тестовой выборке по метрике F1-score с пороговым значением 0,5 определять людей по графическим изображениям.

*Ключевые слова:* нейронные сети, искусственный интеллект, беспилотный летательный аппарат, лесной массив, поиск человека, глубокое обучение, аугментация, детекция объектов.

DOI: 10.15372/AUT20250213

EDN: XKOPNE

**Введение.** Спасение заблудившихся, больных или раненых людей часто включает в себя поиск по густо заросшей лесом местности, где солнечный свет в основном загораживается деревьями и другой растительностью, а лесная почва отражает мало света. Большинство поисково-спасательных работ в лесу выполняется с помощью вертолётов. Поиск осуществляется над районами, в которых с наибольшей вероятностью могут находиться пропавшие люди. Для обнаружения человека с вертолёта применяются тепловизионные камеры [1], предназначенные для выделения различий в температуре тела, что позволяет спасателям отличать людей от окружающей их среды. Однако иногда устройства выходят из строя в жаркую погоду, а также не всегда способны распознать человека по некоторым позам и в маскировочной униформе [2]. Всё это создаёт проблему точности обнаружения человека в лесу.

В последнее время всё более распространёнными становятся алгоритмы обнаружения лиц людей по изображениям с видеокамеры в реальном времени [3–5] с помощью беспилотных летательных аппаратов (БПЛА). Однако обычно используемые классификаторы реального времени (например, YOLO) [6–8] дают полезные выходные данные для целых изображений и в то же время плохие результаты для отдельных изображений, такие как невысокие средние значения точности и скорости обнаружения [4].

Существует потребность в надёжной и точной компьютерной модели, способной распознать очертание человека по изображениям. Эта модель позволит снизить количество ошибок, вызванных человеческим фактором, и быстрее находить пропавших в лесах людей.

Проведён анализ научных исследований в области обнаружения людей с помощью технологий компьютерного зрения.

В работе [9] рассматривается усовершенствованный алгоритм обнаружения людей YOLOv5 с использованием двух новых эталонных наборов данных: HERIDAL и SARD,

специально разработанных для БПЛА SAR при поисково-спасательных операциях в горах и лесистой местности Хорватии и Боснии и Герцеговины. Данный алгоритм позволяет обнаруживать людей за очень короткий промежуток времени, обеспечивая тем самым быструю организацию спасения.

В [3] исследуется комбинация расширенных свёрток с сетью агрегации путей (PAN) в качестве нового алгоритма обнаружения человека на основе глубокой нейронной сети в режиме реального времени. Подход обеспечивает как высокую точность (средняя точность (mAP) 88,0 %), так и работу в режиме реального времени (67,0 кадр/с (FPS)) на готовой коммерческой платформе ПК. Предлагаемый алгоритм является быстрым, эффективным с точки зрения памяти и менее затратным в вычислительном отношении для достижения высокой производительности обнаружения.

В работе [1] рассматривается применение интеграции изображений с помощью бортового оптического секционирования (AOS) — метода визуализации с синтетической апертурой, в котором используются беспилотные камеры для захвата неструктурированных полей теплового света аппаратов на территории Австрии. Метод подразумевает точность и полноту определения 96 и 93 % соответственно. Алгоритм может быть применён в сочетании с автономными или пилотируемыми летательными аппаратами.

В [10] предлагается использовать рой БПЛА, летающих под кронами деревьев. Рой представляет собой централизованную систему управления с лидером — основным беспилотником, передающим остальным информацию в режиме реального времени через порты MAVLink. Местоположение GPS основного дрона корректируется и отправляется другим беспилотным летательным аппаратам в виде серии динамических путевых точек. В качестве системы обнаружения данного подхода используется свёрточная нейронная сеть, основанная на одноступенчатых алгоритмах YOLO. Качественные характеристики такого подхода в данном исследовании не представлены.

В работе [2] выдвигается подход к обнаружению людей на аэрофотоснимках высокого разрешения HERIDAL. Глубокая нейронная сеть EfficientDET обучается с применением созданной базы данных для поисково-спасательной операции. Предложенный метод сравнивался с системой, используемой хорватскими поисковыми группами в горах, где результат достигал максимальной точности 93,29 % mAP.

Исследование [11] посвящено сравнительному анализу алгоритмов обнаружения человека на изображении с помощью нейронных сетей YOLOv3 и YOLOv5. Здесь проведены эксперименты на основе публичного набора данных Human Detection Dataset [12]. Средняя точность распознавания всех изображений из данного набора для YOLOv3 составила 70,2 %, а для YOLOv5 — 71,3 %.

В представленной работе для решения задачи детектирования человека в лесостепной местности применяется современный подход, основанный на использовании компьютерного зрения и глубокого обучения.

Цель работы состоит в применении моделей свёрточной нейронной сети для распознавания людей по их графическим изображениям в лесостепной местности.

**Материалы и методы исследования.** Основным исходным материалом для проведения научных исследований стали изображения людей, находящихся в лесостепной местности, которые были собраны из открытых источников на платформе «Цифровой прорыв» [13]. Эти фотографии имеют размеченные координаты, что позволяет точно определять местоположение объектов на изображениях.

Методы исследования включают в себя несколько ключевых аспектов. Во-первых, мы используем теорию проектирования и разработки систем искусственного интеллекта, что позволяет систематизировать процесс создания алгоритмов и моделей. Это включает в себя выбор архитектуры нейронной сети, оптимизацию её структуры и определение необходимых компонентов для достижения наилучших результатов.

Во-вторых, аугментация изображений играет важную роль в задачах компьютерного зрения. Данный процесс включает в себя модификацию исходных изображений для увеличения объёма обучающего набора данных.

Кроме того, применены алгоритмы настройки гиперпараметров для обучения моделей нейронной сети. Гиперпараметры, такие как скорость обучения, количество слоёв и размер пакета, существенно влияют на качество обучения. Используя методы (например, сеточный и случайный поиск), мы можем найти оптимальные значения гиперпараметров, что, в свою очередь, приведёт к улучшению производительности модели.

В результате сочетание тщательно подобранных методов и качественного исходного материала позволяет достигать высоких показателей в области распознавания объектов и анализа изображений.

**Результаты исследования.** В работе использовался набор данных в виде записей с видеокамеры с видом сверху на перекрёстках кампуса Стэнфордского университета с разметкой пешеходов [14]. Общий объём набора данных составляет 69 Гб. Данные содержат видеозаписи с аннотацией расположения пешеходов и иных объектов на каждом кадре. Выбрано 10 000 изображений с наличием пешеходов. Ракурс камеры схож с ракурсом БПЛА, поэтому такой набор данных подходит для целей исследования.

Также применялся открытый набор данных из размеченных изображений лесистой местности, снятых с БПЛА для участников хакатона «Цифровой прорыв» [15].

Изображения размечены на ограничивающие области (bounding boxes) для обнаружения человека, координаты и размеры которых задаёт прямоугольник, охватывающий объект (человека) на изображении. Такие изображения представлены в формате JPG с разрешением  $1024 \times 1024$  пикселей. К выборке данных прилагается файл в формате CSV с информацией о разметке каждого изображения, содержащей координаты ограничивающих областей. Для обучения использовалась библиотека YOLOv5. Пример изображений из набора данных представлен на рис. 1.

Как показано на рис. 1, площадь с изображением двух человек относительно всего изображения крайне мала. Остальная часть изображения содержит множество неоднородных признаков, что затрудняет обучение модели и подбор данных для обучения и тестирования.

Следующая особенность задачи — изображения людей сняты с верхнего и удалённого ракурса. Большинство известных предобученных моделей обучены на изображениях

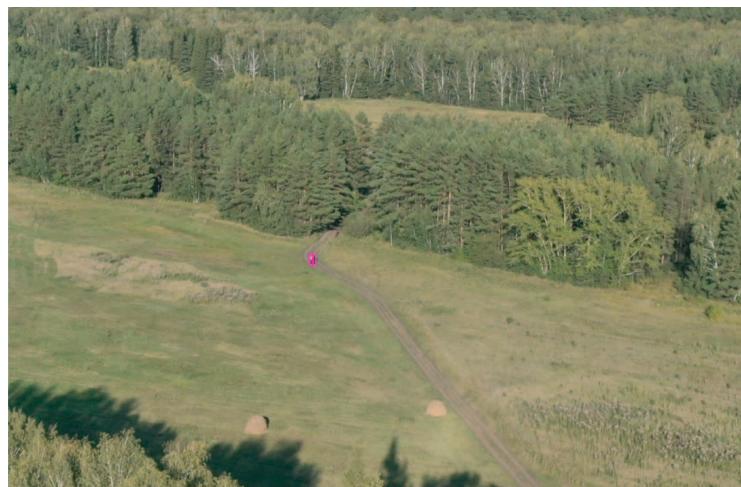


Рис. 1. Размеченные изображения людей, зафиксированные с БПЛА

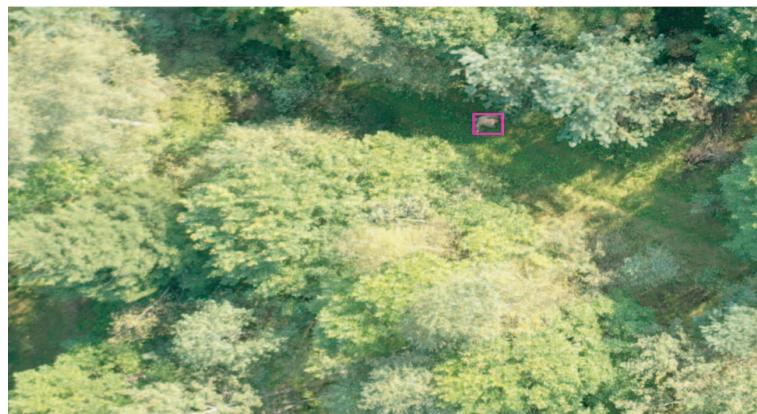


Рис. 2. Вид человека в маскировочной одежде, снятый с БПЛА



Рис. 3. Изображение с наличием человека и животных в степной зоне

людей более крупного плана и преимущественно с бокового ракурса. Также при виде человека сверху существенную роль играет его одежда и головное покрытие, которое может полностью скрыть признаки человека. В лесистой местности часто носят маскировочную одежду, зрительно схожую с окружающей средой. В разные времена года одежда сильно варьируется. Требуется обучение модели на дополнительном наборе данных. Пример изображения человека с верхнего ракурса показан на рис. 2.

Также необходимо увеличить обучающую выборку за счёт изображений в другие времена года.

Предварительное обучение выявило следующие проблемы. Поскольку в обучающей выборке мало изображений с наличием людей и животных, то модель в тестовой выборке классифицировала животных как человека, причём с высоким уровнем уверенности (выше, чем людей на тех же изображениях). Показатель уверенности достигает 0,98 при пороговом значении 0,5. Для уменьшения ошибочной классификации животных в обучающем наборе данных добавлено 1500 изображений с наличием животных в различных климатических условиях. Пример размеченного изображения человека с животными показан на рис. 3.

В общей сложности для улучшения обучающего набора добавлены 1500 изображений с наличием людей в разных местностях. Изображения подбирались из открытых источников. Критерием отбора изображений являлось наличие людей, преимущественно на фоне растительности. Также особое внимание уделялось формированию пул изображений с на-

личием животных (собак, коров, коз и др.). Выборка данных размечена в сервисе Roboflow. Весь набор данных в итоге составил 15 000 изображений из разных источников.

Современные нейронные сети для детекции объектов, как правило, строятся на основе анализа простейших графических примитивов на нижних слоях. Однако в контексте сложных задач, таких как компьютерное зрение, концепция transfer learning показала себя очень эффективной. Она предлагает использовать уже обученную нейронную сеть, тренированную на огромном массиве изображений (в несколько миллионов), подобных тем, которые мы хотим проанализировать. Эта предобученная модель предоставляет значительное преимущество в экономии ресурсов и ускорении.

Для обнаружения объектов применялись следующие модели нейронных сетей: YOLOv51, DenseNet, CenterNet, InceptionV3, Xception, Faster R-CNN.

Для увеличения общего количества данных использована технология аугментации изображений и применена технология зеркалирования по вертикальной центральной оси и поворот изображения на  $90^\circ$  в обе стороны. Также применялась технология поворота только ограничивающего объекта на  $90$  и  $180^\circ$  без поворота изображения. Аугментация данных проводилась на основе библиотеки Albumentations. Обучение составило 30 эпох, размер батча — три изображения. Размер изображения при обучении задан как  $1920 \times 1920$  пикселей. Использовался метод стохастического градиентного спуска (SGD). В качестве настройки гиперпараметра шага обучения подобрано значение  $lr = 0,01$ , вычисленное опытным путём. В качестве показателя функции потерь выбран показатель среднего отклонения (IoU) данных разметки человека в обучающей выборке. Обучение производилось на видеокарте Tesla T4.

В ходе эксперимента набор данных разделён на обучающую выборку, которая включала 80 % элементов, и тестовую, включающую 20 % элементов.

В таблице представлена точность обнаружения (детекции) людей, достигнутая при применении различных моделей с аугментацией и без неё на тестовой выборке (20 %) по метрике F1-score с пороговым значением 0,5.

Из шести представленных моделей наилучший результат показала модель YOLOv51 с точностью классификации 0,8795 с аугментацией на тестовой выборке. Близкий результат представлен нейронной сетью Xception с точностью классификации 0,869 с аугментацией на тестовой выборке. Остальные четыре модели продемонстрировали хорошие результаты, но несколько хуже, чем две предыдущие.

Следует отметить, что поиск людей в лесной местности весьма затруднён из-за плохой видимости сквозь густую крону деревьев. Потерявшийся человек обычно стремится выйти на открытое место, включая опушки и просеки, и поэтому предлагаемый способ на практике пригоден. Также данную технологию можно использовать с поисковыми дронами, которые умеют перемещаться под кронами деревьев. С помощью бортовой свёрточной нейронной сети YOLOv51 дрон для поиска объектов распознаёт человека и передаёт обработанное изображение и GPS-координаты ретранслятору.

Таблица  
Классификация по моделям

Модели	Точность классификации	
	Без аугментации	С аугментацией
DenseNet	0,732	0,858
CenterNet	0,702	0,832
InceptionV3	0,741	0,855
Xception	0,747	0,869
Faster R-CNN	0,728	0,821
YOLOv51	0,78	0,8795

**Заключение.** Выполнены исследования точности обнаружения (детекции) людей по изображениям с применением нескольких моделей нейронных сетей. Используя общедоступные наборы данных более чем из 15 000 изображений людей, глубокая свёрточная нейронная сеть обучается определять местоположение человека, достигая точности распознавания 0,8795. Показано, что наиболее эффективной оказалась нейронная сеть YOLOv51. Описана методика подготовки изображений для обучения и валидации, аугментации, процедуры обучения свёрточной нейронной сетью и настройкой гиперпараметров обучения.

## СПИСОК ЛИТЕРАТУРЫ

1. **Schedl D. C., Kurmi I., Bimber O.** Search and rescue with airborne optical sectioning // Nature Machine Intell. 2020. **2**. P. 783–790. DOI: 10.1038/s42256-020-00261-3.
2. **Dousai N. M. K., Loncaric S.** Detection of humans in drone images for search and rescue operations // Proc. of the 3rd Asia Pacific Information Technol. Conf. (APIT). Bangkok, Thailand, 15–17 Jan., 2021. P. 69–75. DOI: 10.1145/3449365.3449377.
3. **Golcarenarenji G., Martinez-Alpiste I., Wang Q., Alcaraz-Calero J. M.** Efficient Real-Time Human Detection Using Unmanned Aerial Vehicles Optical Imagery // Int. Journ. Remote Sens. 2021. **42**, Iss. 7. P. 2440–2462. DOI: 10.1080/01431161.2020.1862435.
4. **Telçeken M., Kutlu Y.** Detecting Tagged People in Camera Images // Journ. Intell. Syst. with Applications. 2022. **5**, Iss. 1. P. 27–32. DOI: 10.54856/jiswa.202205197.
5. **Yu X., Gong Y., Jiang N. et al.** Scale Match for Tiny Person Detection // ArXiv. 2019. DOI: 10.48550/arXiv.1912.10664.
6. **Bachir N., Memon Q.** Investigating YOLOv5 for Search and Rescue Operations Involving UAVs: Investigating YOLO5 // Proc. of the 5th Int. Conf. Control and Computer Vision (ICCCV). Xiamen, China, 19–21 Aug., 2022. P. 200–204. DOI: 10.1145/3561613.3561644.
7. **Небаба С. Г., Марков Н. Г.** Свёрточные нейронные сети семейства YOLO для мобильных систем компьютерного зрения // Компьютерные исследования и моделирование. 2024. **16**, вып. 3. С. 615–631. DOI: 10.20537/2076-7633-2024-16-3-615-631.
8. **Jung H.-K., Choi G.-S.** Improved YOLOv5: Efficient Object Detection Using Drone Images under Various Conditions // Appl. Sci. 2022. **12**, Iss. 14. 7255. DOI: 10.3390/app12147255.
9. **Yong S.-P., Yeong Y.-C.** Human Object Detection in Forest with Deep Learning based on Drone's Vision // Proc. of the 4th Int. Conf. Computer and Information Sciences (ICCOINS). Kuala Lumpur, Malaysia, 13–14 Aug., 2018. DOI: 10.1109/ICCOINS.2018.8510564.
10. **Meshcheryakov R. V., Trefilov P. M., Chekhov A. V. et al.** An application of swarm of quadcopters for searching operations // IFAC-PapersOnLine. 2019. **52**, Iss. 25. P. 14–18. DOI: 10.1016/j.ifacol.2019.12.438.
11. **Глебов В. В., Марьенков А. Н.** Сравнительный анализ алгоритмов обнаружения человека на изображении // Прикаспийский журнал: управление и высокие технологии. 2023. № 2, вып. 62. С. 97–106.
12. **Human** Detection Dataset: CCTV footage of humans. URL: <https://www.kaggle.com/datasets/constantinwerner/human-detection-dataset> (дата обращения: 23.09.2024).
13. **Цифровой** прорыв. URL: <https://www.hacks-ai.ru/> (дата обращения: 23.09.2024).
14. **Stanford** Aerial Pedestrian Dataset. URL: [https://cs.stanford.edu/~anenberg/uav\\_data/](https://cs.stanford.edu/~anenberg/uav_data/) (дата обращения: 23.09.2024).
15. **Поиск** пропавших людей на снимках лесного массива, полученных с помощью БПЛА или ещё один разбор задачи Цифрового Прорыва. URL: <https://habr.com/ru/articles/699076/> (дата обращения: 23.09.2024).

Поступила в редакцию 23.09.2024

После доработки 29.11.2024

Принята к публикации 17.03.2025