

К.В. БЕЛОКРЫЛОВ, младший научный сотрудник (КузГТУ)

Ю.М. ПОРОХИН, техник (КузГТУ)

И.С. СЫРКИН, к.т.н., доцент (КузГТУ)

В.Ю. САДОВЕЦ, к.т.н., доцент (КузГТУ)

г. Кемерово

ВЛИЯНИЕ ПОДГОТОВКИ ДАТАСЕТА НА ТОЧНОСТЬ МОДЕЛИ YOLO ПРИ ОБНАРУЖЕНИИ ЛЮДЕЙ НА АЭРОФОТОСНИМКАХ

В докладе приведены результаты серии экспериментов, посвящённых обнаружению людей на аэрофотоснимках с беспилотных воздушных судов. Эксперименты направлены на оценку влияния способа подготовки обучающего датасета на качество обнаружения мелких объектов. Актуальность исследования определяется практической значимостью задач поиска людей в лесной и пересечённой местности, мониторинга территорий и поддержки поисково-спасательных операций, где пропуск объекта может иметь критические последствия [1–3].

Целью работы является выявление конфигурации обучения, обеспечивающей наилучший баланс между полнотой обнаружения и точностью при детекции людей на снимках с беспилотного воздушного судна. Для достижения поставленной цели рассмотрены различные схемы подготовки исходных изображений, проанализировано влияние фоновых фрагментов на поведение модели.

1. Материалы и методы

Предметной областью исследования является задача поиска людей на аэрофотоснимках, полученных с беспилотного воздушного судна. Исходные данные представлены в формате YOLO и включают изображения формата JPG и файлы разметки TXT. Для обучения и оценки модели использовались пять вариантов датасета, различающихся способом формирования обучающих выборок.

Исходный датасет включал 32 992 полноразмерных изображения. Его недостатком является значительная вариативность размеров кадров от 6000×6000 до 8000×8000 . При подаче в модель изображения сжимаются до фиксированного входного разрешения 640×640 пикселей, из-за чего мелкие объекты теряют детализацию [2, 5]. Во втором варианте исходные изображения были разбиты на фрагменты размером 640×640 пикселей без перекрытия; итоговый набор составил 80 359 изображений, около 37% которых содержали только фон. Третий вариант был сформирован аналогичной простой нарезкой, но включал только фрагменты, на которых присутствовал хотя бы один человек; объём набора составил 50 378

изображений. В четвертом варианте использовался метод скользящего окна размером 640×640 пикселей с шагом 320 пикселей, что позволило формировать перекрывающиеся фрагменты; итоговый размер составил 169 118 изображений, из которых около 40% были фоновыми. Пятый вариант также строился методом скользящего окна с шагом 320 пикселей, но содержал только фрагменты с полностью видимыми людьми; итоговый размер составил 101 585 изображений.

Для детекции использовалась предобученная модель YOLOv11 типа *l*. Данный выбор обусловлен тем, что архитектура семейства YOLO обеспечивает одностадийную обработку изображения, а версия YOLOv11 демонстрирует высокую эффективность в задачах компьютерного зрения, в том числе при обнаружении мелких объектов [3–5]. Кроме того, использование предобученных весов и моделей семейства Ultralytics ранее показало свою практическую применимость для задачи обнаружения людей на аэрофотоснимках [3]. Обучение проводилось с использованием оптимизатора SGD при эффективном размере пакета 16. Для сравнительного анализа влияния структуры датасета использовался единый набор гиперпараметров и параметров аугментации, предварительно полученный в отдельной серии экспериментов по тюнингу [6, 7].

Поскольку задача предполагает обнаружение малых объектов и использование фрагментов, в которых человек виден полностью, функции *mosaic* и *erasing* для аугментации данных были отключены, так как они могут обрезать или существенно искажать целевые объекты [8, 9].

Итоговый набор параметров, использованный в сравнительных экспериментах, приведён в таблице 1.

Таблица 1

Использованные гиперпараметры обучения и аугментации

Параметр	Значение
lr0	0,00785
lrf	0,01
momentum	0,6
weight_decay	0,00063
warmup_epochs	3
warmup_momentum	0,49851
box	11,31475
cls	0,49909
dfl	2,03668
mosaic	0
erasing	0
hsv_h	0,01258
hsv_s	0,2
hsv_v	0,425
degrees	17,90097
translate	0,15
scale	0,2
shear	3,5
perspective	0,0001
flipud	0,5
fliplr	0,5

Качество моделей оценивалось по метрикам Precision, Recall и F1. Метрика Precision характеризует долю корректно обнаруженных объектов среди всех положительных срабатываний модели. Метрика Recall показывает долю найденных объектов среди всех реально присутствующих объектов. Метрика F1 используется как интегральный показатель, позволяющий сбалансированно оценить качество модели при наличии компромисса между точностью и полнотой.

2. Результаты и обсуждение

Сопоставление пяти вариантов подготовки датасета показало, что решающее влияние на итоговое качество оказывает не только объём обучающей выборки, но и способ представления объектов в изображении. Результаты сравнительного эксперимента приведены в таблице 2.

Таблица 2

Результаты сравнения вариантов подготовки датасета

Датасет	Эпоха лучшего результата	F1	Precision	Recall
1	50	0,4838	0,9604	0,3233
2	150	0,8072	1,0000	0,6767
3	200	0,8996	0,9729	0,8367
4	150	0,8008	0,9950	0,6700
5	150	0,9040	0,9487	0,8633

Наихудший результат продемонстрировал исходный датасет без предварительной нарезки. При обучении на полноразмерных изображениях модель показала $F1 = 0,4838$ и $Recall = 0,3233$. Основная причина состоит в том, что при приведении больших изображений к входному размеру модели мелкие объекты теряют значимую часть признаков, что ослабляет обучающий сигнал и ускоряет достижение плато в обучении [2, 5].

Все варианты, в которых изображения были предварительно приведены к разрешению 640×640 пикселей, обеспечили заметный прирост качества. Это подтверждает, что локальная чёткость объекта и отсутствие дополнительного масштабирования перед подачей в сеть имеют ключевое значение для детекции людей на аэрофотоснимках [2, 5].

Лучший результат по совокупности метрик был получен для пятого варианта датасета, сформированного методом скользящего окна с шагом 320 пикселей и содержащего только фрагменты с полностью видимыми людьми. Для этого варианта были получены значения $F1 = 0,9040$, $Precision = 0,9487$ и $Recall = 0,8633$. Данный результат указывает на то, что перекрытие фрагментов и исключение частично обрезанных объектов создают наиболее благоприятные условия для обучения.

Отдельно было установлено, что включение фоновых примеров влияет на характер ошибок модели. Варианты 2 и 4, содержавшие значительное количество пустых фрагментов, показали очень высокую точность ($Precision = 1,0$ и $0,995$ соответственно), однако уступили по полноте обнаружения. Напротив, варианты 3 и 5, исключавшие фоновую составляющую, обеспечили более высокие значения Recall. Для задач поисково-спасательного характера такой сдвиг в сторону полноты является оправданным, поскольку пропуск человека значительно критичнее, чем дополнительная проверка ложного срабатывания [1, 3].

Заключение

Обобщение результатов экспериментов показало, что качество обнаружения людей на аэрофотоснимках определяется влиянием структуры датасета. Наилучшей стратегией подготовки данных является нарезка изображений методом скользящего окна с шагом 320 пикселей и отбором только тех фрагментов, на которых человек представлен полностью. Такая схема обеспечивает наилучшие показатели полноты обнаружения и интегральной метрики F1.

Практический результат работы заключается в формировании рекомендованной конфигурации обучения модели YOLO для задач поиска людей с беспилотных воздушных судов: использование фрагментов 640×640 пикселей с перекрытием и исключение частично видимых объектов.

Список литературы

1. Tang G. и др. A Survey of Object Detection for UAVs Based on Deep Learning // Remote Sensing. 2024. Т. 16. № 1. С. 149.
2. Nikouei M. и др. Small object detection: A comprehensive survey on challenges, techniques and real-world applications // Intelligent Systems with Applications. 2025. № 27. С. 200561.
3. Белокрылов К.В., Порохин Ю.М., Сыркин И.С., Садовец В.Ю. Влияние типа модели Ultralytics и размера batch size на точность обнаружения людей на аэрофотоснимках // Инфокоммуникационные технологии. 2025. С. 116–122. [Электронный ресурс]. URL: <https://elibrary.ru/item.asp?id=87685166>.
4. Khanam R., Hussain M. YOLOv11: An Overview of the Key Architectural Enhancements // arXiv.org [Электронный ресурс]. URL: <https://arxiv.org/abs/2410.17725v1>.
5. Tariq M.F., Javed M.A. Small Object Detection with YOLO: A Performance Analysis Across Model Versions and Hardware // arXiv.org [Электронный ресурс]. URL: <https://arxiv.org/abs/2504.09900>.
6. Bischl B. и др. Hyperparameter Optimization: Foundations, Algorithms, Best Practices and Open Challenges // arXiv.org [Электронный ресурс]. URL: <https://arxiv.org/abs/2107.05847>.
7. Ultralytics. Hyperparameter Tuning // Ultralytics [Электронный ресурс]. URL: <https://docs.ultralytics.com/guides/hyperparameter-tuning/>.

8. Yang S. и др. Image Data Augmentation for Deep Learning: A Survey // arXiv.org [Электронный ресурс]. URL: <https://arxiv.org/abs/2204.08610>.
9. Cubuk E.D. и др. AutoAugment: Learning Augmentation Policies from Data // arXiv.org [Электронный ресурс]. URL: <https://arxiv.org/abs/1805.09501>.

Информация об авторах

Белокрылов Кирилл Викторович, младший научный сотрудник Научного центра «Цифровые технологии», студент группы ИИМ-241 кафедры ИиАПС ИИТМА КузГТУ, ФГБОУ ВО «Кузбасский государственный технический университет имени Т.Ф. Горбачёва», 650000, г. Кемерово, ул. Весенняя, д. 28, belokrylovkv1@kuzstu.ru.

Порохин Юрий Максимович, техник Научного центра «Цифровые технологии», студент группы ИТб-221 кафедры ИиАПС ИИТМА КузГТУ, ФГБОУ ВО «Кузбасский государственный технический университет имени Т.Ф. Горбачёва», 650000, г. Кемерово, ул. Весенняя, д. 28, porohin.jum@kuzstu.ru.

Сыркин Илья Сергеевич, кандидат технических наук, доцент кафедры ИиАПС ИИТМА КузГТУ, ФГБОУ ВО «Кузбасский государственный технический университет имени Т.Ф. Горбачёва», 650000, г. Кемерово, ул. Весенняя, д. 28, syrkin@kuzstu.ru.

Садовец Владимир Юрьевич, кандидат технических наук, доцент кафедры ИиАПС ИИТМА КузГТУ, ФГБОУ ВО «Кузбасский государственный технический университет имени Т.Ф. Горбачёва», 650000, г. Кемерово, ул. Весенняя, д. 28, svyu.pmh@kuzstu.ru.