

ДАТАСЕТ ДЛЯ РЕШЕНИЯ ЗАДАЧ ОБЪЕКТНОГО ДЕТЕКТИРОВАНИЯ ЛЕТАЮЩИХ ОБЪЕКТОВ НА ТЕПЛОВИЗИОННЫХ ИЗОБРАЖЕНИЯХ

Евсюткин И.В.¹, Небаба С.Г.², Курило Ю.Ю.³, Кушнарев А.А.⁴

¹НИ ТПУ, ИШИТР, к.т.н., доцент ОИТ, ive1@tpu.ru

²НИ ТПУ, ИШИТР, к.т.н., доцент ОИТ, stepanlfx@tpu.ru

³НИ ТПУ, ИШИТР, студент гр. 8ПМ4Л, aak404@tpu.ru

⁴НИ ТПУ, ИШИТР, студент гр. 8ПМ4Л, yyk10@tpu.ru

Аннотация

В работе рассмотрен процесс создания датасета для объектного детектирования летающих объектов на тепловизионных изображениях. Проведена разметка положения объектов на изображениях двух классов: беспилотные летательные аппараты (БПЛА) самолётно-го типа и БПЛА вертолётно-го типа. Приведены результаты сравнения эффективности разных подходов к формированию датасета.

Ключевые слова: объектное детектирование объектов, обработка изображений, сверточная нейронная сеть YOLO, датасет, тепловизионное изображение, беспилотный летательный аппарат.

Введение

Известно, что практически важными и актуальными являются задачи мониторинга воздушного пространства в режиме реального времени с целью объектного детектирования (обнаружения, позиционирования и классификации) летающих объектов (ЛО) различных классов [1]. Наибольший интерес сегодня среди таких объектов представляют беспилотные летательные аппараты (БПЛА). Традиционно с этой задачей должен справляться оператор видеокамеры в составе систем мониторинга. Однако, чтобы уменьшить число ошибок, связанных с человеческим фактором, и обеспечить масштабирование и поддержку принятия решения, применяются системы компьютерного зрения для частичной или полной автоматизации процесса мониторинга.

При разработке решений, автоматизирующих объектное детектирование в задачах мониторинга воздушного пространства, широко применяются телевизионные камеры, фиксирующие изображение в оптическом RGB-диапазоне видимого цвета, а также тепловизионные камеры, фиксирующие изображение в ближнем инфракрасном диапазоне. Телевизионные камеры для такой задачи подходят лишь частично, так как обеспечивают съёмку с требуемыми для оптического объектного детектирования характеристиками изображения только в дневное время. По этой причине важно рассмотрение именно тепловизионных камер, которые позволяют ориентироваться не на оптические характеристики объектов, а на их тепловые контуры, что потенциально способно обеспечить круглосуточный мониторинг воздушного пространства.

Наиболее активно развивающимся средством автоматизации систем мониторинга являются искусственные нейронные сети (ИНС) [2]. Свёрточные ИНС способны решить задачу обнаружения и классификации объектов в кадре, обозначив для оператора интересующие объекты прямоугольником и название класса рядом. Но чтобы такая ИНС обрела способность это делать, её нужно обучить, а для этого, в свою очередь, необходим датасет (набор размеченных изображений), который в открытом доступе для тепловизионных изображений ЛО, по-видимому, не существует. Поэтому актуальной становится задача создания датасета для решения задач объектного детектирования ЛО на тепловизионных изображениях.

В данной статье рассматриваются этапы разработки датасета на основе тепловизионных изображений и приводятся результаты сравнения эффективности подходов к формированию такого датасета. Показаны его основные характеристики.

Постановка задачи

Необходимо решить задачу создания датасета по изображениям, полученным с помощью приемников тепловизионных сигналов (тепловизионных камер). Этот датасет предназначен для обучения, валидации и тестирования свёрточных нейронных сетей (СНС), позволяющих решать задачи объектного детектирования ЛО в воздушном пространстве на тепловизионных изображениях. В первоначальной формулировке ставится задача детектировать на изображениях ЛО двух классов: БПЛА самолетного типа и БПЛА вертолетного типа. С точки зрения задачи детектирования объектов также важны относительные размеры объектов на изображении, так как объекты малых размеров чаще всего обладают меньшим числом деталей, которые могут служить их отличительными признаками при классификации. То, насколько полно представлены в датасете объекты различного размера, влияет и на конечную эффективность детектирования. Традиционно объекты на изображениях разделяют на объекты малых размеров (не более 32*32 пикселей по площади), средних размеров (от 33*33 до 96*96 пикселей) и крупные (97*97 пикселей и больше) [3]. Однако в поставленной задаче такого разделения ЛО по размерам не проводилось. На одном изображении может находиться один и более объектов разных категорий по размерам, а также принадлежащих разным классам. В датасете должны быть представлены изображения с несколькими объектами, в том числе разных классов, чтобы гарантировать эффективное детектирование нескольких ЛО различного класса, попадающих в кадр, обученной на этом датасете нейронной сетью.

Подбор изображений должен осуществляться с учетом следующих требований: исходные изображения должны максимально точно отражать реальные сценарии расположения объектов в воздушном пространстве (наличие одного или нескольких летающих объектов из указанных классов, фон в виде участков земной поверхности или облаков и т. д.).

Разметка изображений проводится с помощью системы Roboflow [4]. Эта система позволяет упростить процесс разметки. Большим её преимуществом является возможность удобной работы в команде, формирующей датасет.

Этапы создания датасета

Качество датасета оценивается по способности обученной на нём СНС решать задачи детектирования и классификации (метрики mAP, precision, accuracy и т. д.). Поэтому было решено создать несколько различных по способу получения и характеристикам датасетов, предназначенных для детектирования объектов на тепловизионных изображениях, и сравнить эффективность различных подходов к формированию датасета и его аугментации по метрике точности детектирования объектов mAP50 и mAP50-95 путем тестирования модели СНС YOLOv8s [5].

Первый рассмотренный подход к формированию датасета – использование имеющегося датасета из RGB-изображений, перевод изображений в градации серого цвета (grayscale) и инвертирование. Таким образом имитируется визуальное отображение тепловизионных снимков, хотя одна из ключевых характеристик объектов, фиксируемая при реальной тепловизионной съемке, температура объектов, при таком подходе на изображениях не представлена.

Для второго подхода к формированию датасета использовались тепловизионные видеозаписи, выполненные в реальных условиях по заранее спланированным сценариям. Эти сценарии содержали перечень условий, гарантирующих высокую вариативность изображений БПЛА: записи проводились с различных дистанций, на разном фоне (ясное или облачное небо, лес), в разное время суток (день, ночь), с различной траекторией полёта и т. п. Большая вариативность изображений в итоговом датасете гарантирует лучшие результаты обучения нейронных сетей, минимизируя риски переобучения [5].

Первичный анализ этих видео показал, что для части из них необходимо преобразование формата и нарезка на фрагменты, так как в системе Roboflow есть ограничения на объём загружаемого видео и его формат. Некоторые фрагменты видео были подготовительными или содержали окончание съёмки, где отсутствовали объекты интереса, и их рассмотрение не имело смысла. После учета этих особенностей видео были загружены в систему Roboflow для дальнейшей разметки.

Нарезка видео на отдельные кадры происходила на основе исходной частоты кадров видео и резкости маневрирования ЛО при полётах. Это связано с тем, что выбор слишком высокой частоты нарезки может привести к большому числу похожих изображений в выборке, что увеличит нагрузку на специалистов, которые будут формировать датасет. С другой стороны, выбор малой частоты разбиения может привести к потере ценных кадров, что уменьшит качество датасета.

Задача разметки состоит в том, чтобы вручную с поддержкой системы Roboflow производить анализ изображений в исходной выборке, оценивать необходимость включения изображения в датасет и в случае необходимости выделять прямоугольник, описывающий объект интереса на изображении. Данный процесс производился по ряду критериев: изображения должны содержать отличающиеся ЛО (по ракурсу, размерам и т. п.), кадры с высокой степенью схожести должны быть удалены. Также, по рекомендации к процессу обучения нейронных сетей, датасет должен включать в себя до 10 % изображений, на которых отсутствуют объекты [5]. Такие изображения помечаются как «null» (не содержат объектов ни одного из искомых классов). Всё это важно для последующего обучения СНС. Пример интерфейса разметки изображений при работе с системой Roboflow показан на рисунке 1, здесь показан кадр с размеченными двумя БПЛА вертолётного типа.

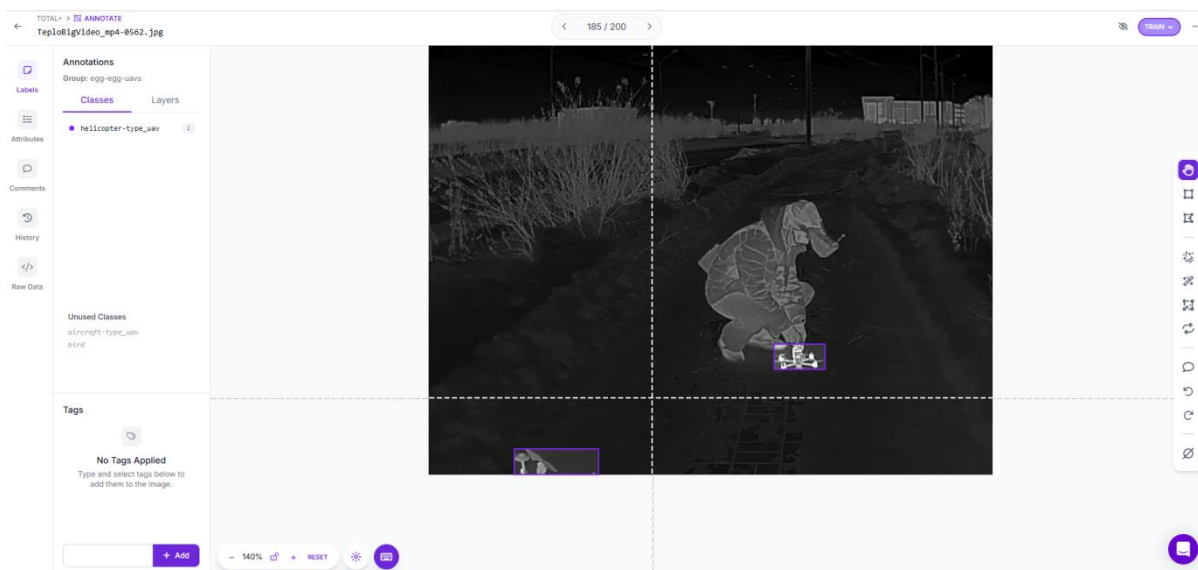
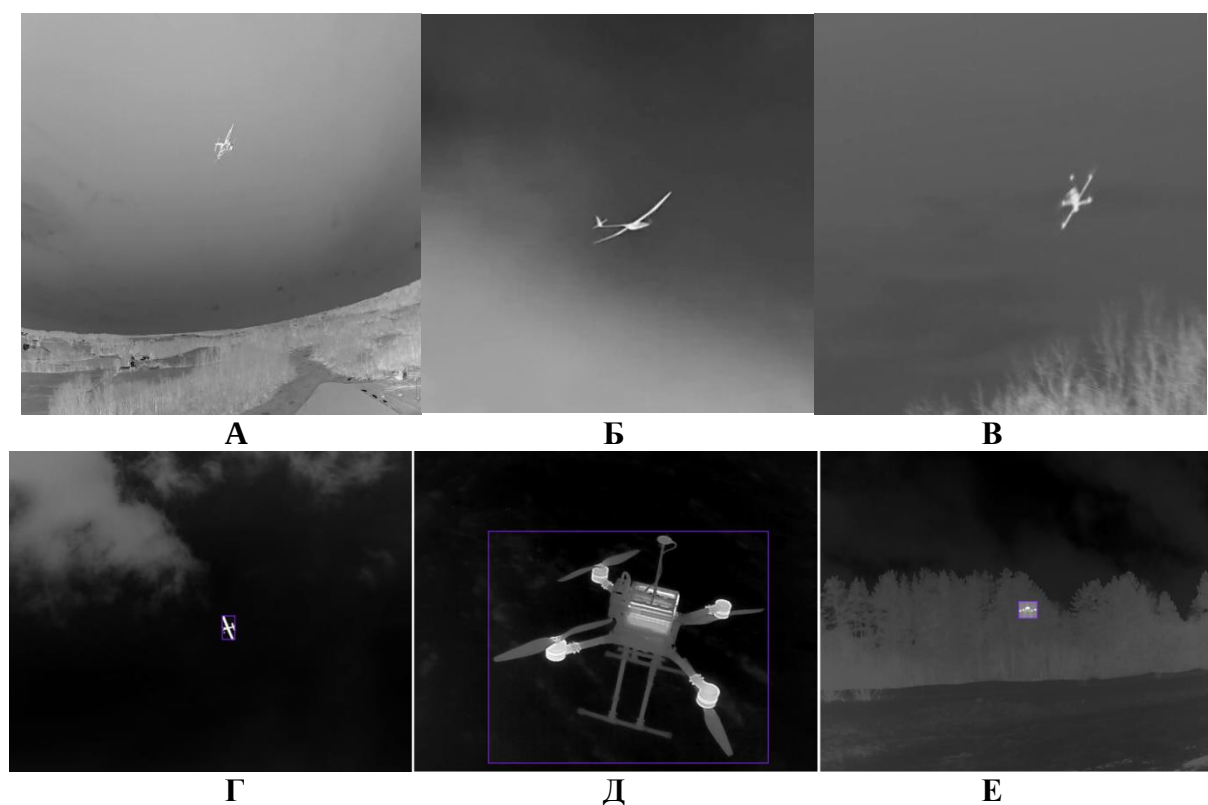


Рис. 1. Пример размеченного кадра в интерфейсе пользователя системы Roboflow

Характеристики сформированных датасетов

Примеры размеченных изображений в рамках первого и второго подходов создания датасетов показаны на рисунке 2.



*Рис. 2. Примеры размеченных изображений из датасета
А, Б, В – датасет из обесцвеченных инвертированных RGB-изображений ЛО,
Г, Д, Е – датасет из тепловизионных видеозаписей ЛО*

В датасете, составленном из обесцвеченных инвертированных RGB-изображений, количество размеченных объектов 27224 на 34335 изображениях. Во втором датасете, составленном на основе тепловизионных видеозаписей, количество размеченных объектов 4786 объектов на 6272 изображениях. По классам распределение ЛО в этом датасете следующее:

- «БПЛА самолетного типа» – 592 объекта.
- «БПЛА вертолетного типа» – 4191 объектов.

Можно заметить, что у второго полученного датасета выборка получилась несбалансированной, то есть количество объектов класса БПЛА самолётного типа по сравнению с классом БПЛА вертолётного типа значительно ниже. Этот недостаток исправляется путем применения методов аугментации и по мере поступления дополнительных видеоматериалов.

Все размеченные изображения датасетов были физически разделены на три выборки: обучающую (70 % от общего объема изображений в датасете), валидационную (20%) и тестовую (10 %).

Распределение числа изображений во втором датасете в зависимости от числа представленных на них объектов показано на рисунке 3. Первые три столбчатые диаграммы (отсчет слева направо) характеризуют этот датасет до аугментации, а четвертая диаграмма получена по результатам мозаичной аугментации датасета (на полученных дополнительных изображениях находятся по четыре ЛО).

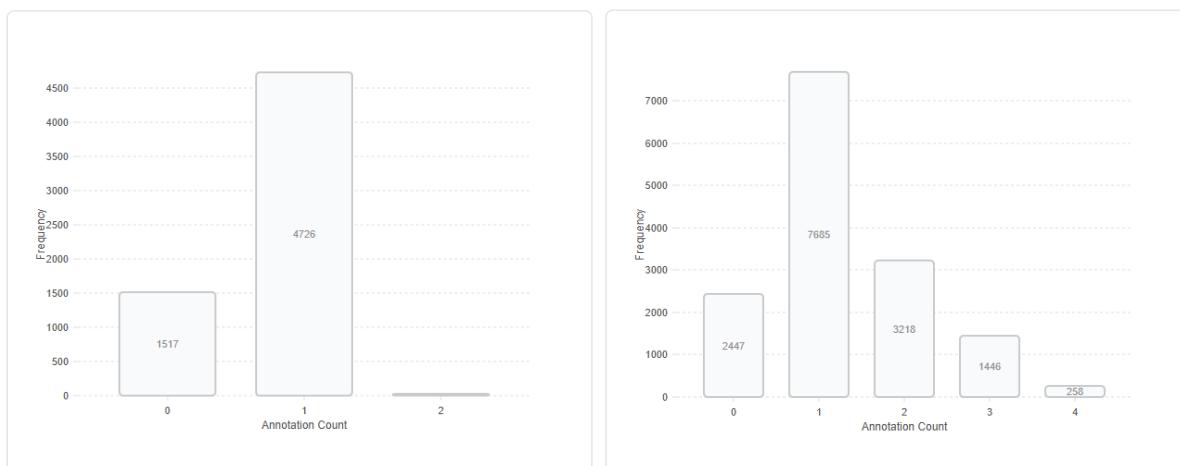


Рис. 3. Зависимость числа изображений от количества объектов на одном изображении исходного (слева) и аугментированного (справа) датасета с тепловизионными изображениями

Сравнение эффективности подходов к формированию датасетов

Итоговое тестирование и сравнение эффективности СНС YOLOv8s по точности детектирования, обученных на подготовленных датасетах, проводилось на отдельной тестовой выборке. Эта тестовая выборка составлена из 262 изображений, полученных с помощью тепловизионной камеры. Результаты тестирования датасетов по метрикам точности детектирования mAP50 и mAP50-95 на примере обученной СНС YOLOv8s приведены в таблице 1.

Таблица 1. Эффективность СНС YOLOv8s, обученной на подготовленных датасетах

Датасет	mAP50	mAP50-95
Инвертированные обесцвеченные RGB-изображения	0,29	0,109
Тепловизионные снимки	0,759	0,386

Сравнение результатов из таблицы 1 показывает, что использование имитации тепловизионных снимков для обучения нейронных сетей, даже снимков, которые показывают свою эффективность в обучении нейронных сетей на RGB-снимках, показывает существенно более низкую эффективность СНС YOLOv8 при тестировании на тепловизионных тестовых выборках по метрикам mAP50 и mAP50-95.

Исходя из этих результатов, в дальнейшей работе использовался датасет, сформированный на основе реальных тепловизионных снимков.

Аугментация датасета и сравнение результатов аугментации

Была проведена аугментация датасета (создание дополнительных изображений на основе имеющихся данных) для увеличения объема выборок и повышения эффективности детектирования, обученной на этом датасете нейронной сети. Использовались следующие приёмы и их сочетания с увеличением объема выборки в 2 или 3 раза:

- инвертирование изображений.
- поворот до 10°.
- зеркальное отражение по вертикали.
- мозаичная аугментация.
- увеличение и уменьшение яркости.
- увеличение насыщенности.
- копирование объектов по контуру и их дублирование на нескольких изображениях.

Это позволило иметь возможность исследовать полученные датасеты, чтобы получить максимальную точность детектирования ЛО на изображениях с помощью СНС YOLOv8s.

Количество аугментированных изображений в полученных таким образом датасетах соответственно кратно выше (в 2 и 3 раза). Важно отметить, что аугментации подлежат изображения только из тренировочной выборки, валидационная и тестовая выборки остаются неизменными, что позволяет сравнить обученные на их основе СНС в равных условиях.

Результаты тестирования датасетов, полученных с использованием различных методов аугментации, по метрикам точности детектирования mAP50 и mAP50-95 на примере обученной СНС YOLOv8s приведены в таблице 2.

Таблица 2. Эффективность СНС YOLOv8s, обученной на подготовленных датасетах

Методы аугментации	mAP50	mAP50-95
Без аугментации	0,759	0,386
Поворот до 10° (число изображений x2)	0,733	0,352
Поворот до 10° (число изображений x3)	0,726	0,352
Зеркальное отражение (число изображений x2)	0,676	0,247
Инвертирование (число изображений x2)	0,596	0,261
Мозаичная аугментация (число изображений x2)	0,957	0,599
Мозаичная аугментация (число изображений x3)	0,969	0,645
Уменьшение яркости (число изображений x2)	0,625	0,221
Увеличение яркости (число изображений x2)	0,593	0,241
Увеличение насыщенности (число изображений x2)	0,635	0,218

Сравнение результатов из таблицы 2 показывает, что мозаичная аугментация - единственный метод аугментации из нескольких рассмотренных, который показывает свою эффективность: использование этого метода аугментации датасета дает существенный прирост точности по метрике mAP50 (до 21 %) и по метрике mAP50-95 (до 26 %) при исследовании СНС YOLOv8s на тестовой выборке.

Заключение

Показаны два различных подхода к формированию датасета по тепловизионным изображениям и этапы его разработки: сценарии видеосъемки, предварительная обработка видео, разметка изображений, аугментация и получение выборок. Готовый датасет содержит 6272 размеченных тепловизионных изображений.

Получены результаты сравнения эффективности СНС YOLOv8s при использовании разных подходов к формированию датасета и его аугментации. Сделан вывод о том, что наиболее эффективным подходом является подготовка датасета из реальных тепловизионных снимков с последующей мозаичной аугментацией его изображений.

Список использованных источников

1. Nebaba S.G. Effectiveness of Moving Objects Detecting and Tracking in Airspace by Images in Nearinfrared / S.G. Nebaba, N.G. Markov // *Light & Engineering*. – 2022. – Vol.30, No. 2. – P. 62-69.
2. Гудфеллоу Я., Бенджио И., Курвилль А. Глубокое обучение. – М: ДМК-Пресс. – 2018. – 652 с.
3. Microsoft COCO: Common objects in context / T.Y. Lin, M. Maire, S. Belongie, J. Hays, P. Perona, D. Ramanan, C.L. Zitnick // *Computer Vision–ECCV 2014: 13th European Conference, Zurich, Switzerland, September 6-12, 2014, Proceedings, Part V 13*. – Springer International Publishing. – 2014. – P. 740–755. DOI: 10.48550/arXiv.1405.0312.
4. Roboflow [Электронный ресурс]. – URL: roboflow.com (дата обращения: 24.03.2025).
5. Reis D. Real-time flying object detection with YOLOv8 / D. Reis, J. Kupec, J. Hong, A. Daoudi // *arXiv preprint arXiv:2305.09972*. – 2023.