

ИССЛЕДОВАНИЕ МОДЕЛЕЙ СВЕРТОЧНЫХ НЕЙРОННЫХ СЕТЕЙ YOLOV5 И YOLOV8 ПРИ ДЕТЕКТИРОВАНИИ ЛЕТАЮЩИХ ОБЪЕКТОВ НА ИЗОБРАЖЕНИЯХ

Клековкин В. А.

Томский политехнический университет

ИШИТР, группа А2-39

E-mail: vak37@tpu.ru

Аннотация

Данная работа посвящена исследованию моделей сверточных нейронных сетей (СНС) YOLOv5s и YOLOv8s при объектном детектировании на изображениях летающих объектов (ЛО) в случае их малых размеров (площадь объекта не более 32x32 пикселя) и в случае ЛО на изображениях разных размеров (малые, средние и большие по площади). Созданы два датасета с изображениями ЛО трех классов: птицы, беспилотные летательные аппараты вертолетного и самолетного типов. Первый датасет сформирован по изображениям с ЛО только малых размеров, а второй – по изображениям объектов разных размеров. На каждом из датасетов проведено обучение моделей СНС YOLOv5s и YOLOv8s. Обе модели исследованы на тестовых выборках датасетов. Результаты исследований показали, что обе модели демонстрируют схожие показатели по точности детектирования по метрикам AP_{0,5} и mAP_{0,5} для всех классов объектов. Тем не менее, YOLOv8s демонстрирует небольшое преимущество в точности обнаружения объектов по сравнению с YOLOv5s, особенно в случае малых размеров объектов. Обе модели имеют высокие значения усредненной скорости вычислений, но модель YOLOv5s превосходит YOLOv8s по скорости вычислений как при детектировании ЛО малых размеров, так и объектов различных размеров.

Ключевые слова: сверточные нейронные сети YOLOv5 и YOLOv8, точность и скорость детектирования летающих объектов, датасет, птица, беспилотные летательные аппараты вертолетного и самолетного типа

Введение

На сегодняшний день задачи мониторинга воздушного пространства для объектного детектирования (обнаружения, позиционирования и классификации) летающих объектов (ЛО) приобретают особую важность. В данной области обычно выделяют три ключевых класса объектов: птицы, беспилотные летательные аппараты (БПЛА) самолётного типа и БПЛА вертолетного типа.

Несмотря на то, что особенности детектирования подвижных объектов, в том числе ЛО, на изображениях изучаются уже достаточно долго, исследования в этой сфере продолжают оставаться актуальными и востребованными, поскольку обеспечение детектирования с высокой точностью объектов малого размера (площадью не более 32x32 пикселя) является сложной задачей [1]. Известно, что для решения задачи объектного детектирования ЛО перспективными считаются модели СНС класса YOLO [2, 3].

В данной работе исследуются модели СНС YOLOv5s и YOLOv8s, входящие в класс YOLO. Разработаны два датасета с размеченными изображениями объектов малых размеров и разномасштабных объектов. Изображения содержат ЛО трех классов: птицы, беспилотные летательные аппараты самолетного и вертолетного типов. С использованием этих датасетов модели СНС YOLOv5s и YOLOv8s обучены и исследованы в части точности детектирования изображений с малыми и разномасштабными ЛО, а также по скорости вычисления этих моделей.

Постановка задачи

Будем рассматривать задачу объектного детектирования ЛО в воздушном пространстве на изображениях, учитывая, что все объекты подразделяются на три класса: БПЛА самолетного типа, БПЛА вертолетного типа и Птица. По размерам (по занимаемой ими площади на изображениях) объекты, не зависимо от их класса и опираясь на результаты работы [4], разделим на три категории:

- малые (площадь объекта на изображении меньше или равна 32×32 пикселей);
- средние (площадь от 33×33 до 96×96 пикселей включительно);
- большие (площадь 97×97 пикселей и более).

На одном изображении может находиться один и более объектов разных категорий по размерам, а также принадлежащих к разным классам.

Модели СНС YOLO, начиная с YOLOv4, зарекомендовали себя как одни из лучших решений для задач объектного детектирования объектов благодаря своей скорости и точности обнаружения и классификации объектов. С тех пор были выпущены современные версии YOLO, такие как YOLOv5, YOLOv7 и YOLOv8, которые продолжают улучшать производительность, добавляя новые функции и оптимизации. В результате проведенного авторами [3] анализа моделей класса YOLO при детектировании ЛО выявлено, что модели семейства YOLOv5 несколько уступают по точности детектирования ЛО нескольким классам на изображениях более новым моделям, таким как YOLOv7 и YOLOv8. Однако модели этого семейства значительно превосходят новые модели по производительности. Модель YOLOv5s – это одна из самых легких и быстрых моделей в семействе YOLOv5. Она оптимизирована для работы в реальном времени и может быть эффективно использована для решения задач, где важна скорость анализа изображений. Однако ее производительность может снижаться при детектировании объектов малых размеров, так как архитектура модели не всегда позволяет точно идентифицировать такие объекты. Модель YOLOv8s [3] – это хорошо себя зарекомендовавшая модель семейства YOLOv8, которая предлагает улучшенные алгоритмы анализа изображений и поэтому может обеспечивать более высокую точность детектирования объектов на изображении, и в тоже время она имеет высокую производительность. Эта модель использует более сложные механизмы, такие как улучшенные слои свертки, что может позволить ей лучше справляться с задачами детектирования ЛО. Эти аргументы способствовали выбору моделей YOLOv5s и YOLOv8s в качестве перспективных для их исследований при решении задачи детектирования малоразмерных и разномасштабных ЛО.

Целью данной работы является обучение и проведение сравнительных комплексных исследований моделей СНС YOLOv5s и YOLOv8s на сформированных датасетах по точности детектирования ЛО на изображениях и по скорости вычисления моделей.

Для оценки эффективности моделей СНС (точности классификации объектов) использовались такие известные метрики как Average Precision (AP) 0.5 и mean Average Precision (mAP) 0.5 [5]. Итоговую точность обнаружения ЛО на изображениях можно считать приемлемой, если значения метрики AP_{0,5} для каждого класса ЛО и метрики mAP_{0,5} для всех классов объектов превышают установленное пороговое значение. Согласно работе [6], это пороговое значение было установлено на уровне 0,9. В практических задачах, связанных с мониторингом воздушного пространства, способность одновременно обнаруживать ЛО различных масштабов критически важна для систем компьютерного зрения (СКЗ).

Для достижения поставленной цели следует создать два датасета, обучить и валидировать модели YOLOv5s и YOLOv8s на обучающих и валидационных выборках этих наборов данных, а также провести исследование моделей на тестовых наборах данных датасетов для оценки точности детектирования ЛО и скорости их вычислений.

Формирование датасетов

Для формирования датасетов для обучения, валидации и тестирования предложенных моделей СНС проводился поиск изображений с летающими объектами, относящимися к трем классам: БПЛА самолетного типа, включая «летающее крыло», БПЛА вертолетного типа и Птица. Учитывался ряд требований к подбору изображений: датасет должен включать только те изображения, которые максимально точно отражают реальные сценарии расположения объектов в воздушном пространстве (наличие одного или нескольких ЛО из указанных классов, разный фон, включающий облака, земную поверхность и т. д.). Для увеличения объема выборок датасетов часть изображений была подвергнута аугментации (созданы дополнительные изображения на основе имеющихся).

Пример исходных изображений для создаваемого датасета с объектами малых размеров демонстрируется на рисунке 1, а. Изображение слева содержит объект класса БПЛА вертолетного типа, среднее – БПЛА самолетного типа, на правом изображении можно увидеть четыре объекта класса Птица. Все изображения каждого из датасетов были масштабированы до единого размера 416x416 пикселей и размечены с использованием инструментов системы Roboflow [7]. На рисунке 1, б представлен пример размеченных изображений из созданного датасета с объектами малых размеров, которые соответствуют исходным изображениям на рисунке 1, а.

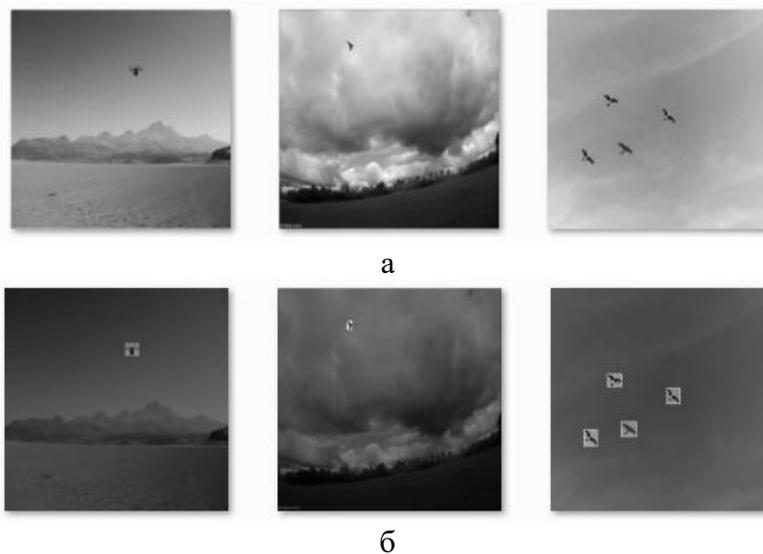


Рис. 1. Пример изображений из датасета: а – исходные, б – размеченные

В результате все размеченные изображения каждого из датасетов были физически разделены на три выборки: обучающую (70% от общего объема), валидационную (20%) и тестовую (10%). В датасете с изображениями ЛО малых размеров количество размеченных объектов составляет 4540. По классам распределение следующее:

- «БПЛА самолетного типа» – 1663 объекта;
- «Птица» – 1440 объектов;
- «БПЛА вертолетного типа» – 1437 объектов.

Датасет с разномасштабными ЛО содержит 7419 размеченных на изображениях объектов. Распределение по классам выглядит следующим образом:

- «БПЛА самолетного типа» – 2560 объектов;
- «Птица» – 2084 объекта;
- «БПЛА вертолетного типа» – 2775 объектов.

Количественные характеристики датасетов по классам представлены на рисунке 2.

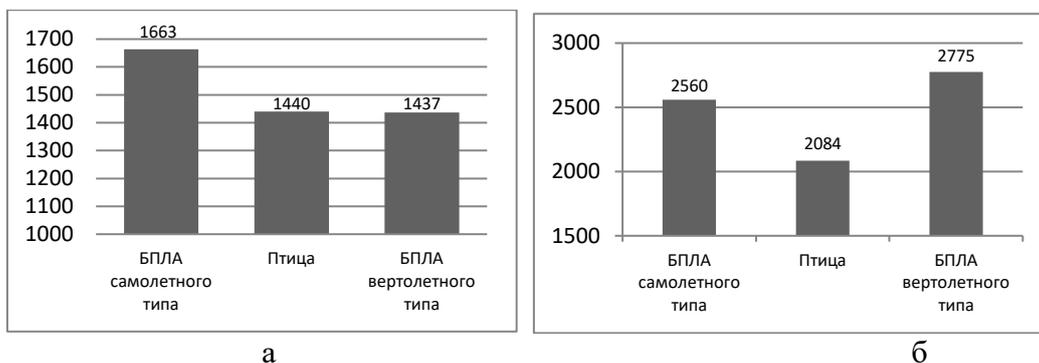


Рис. 2. Количество объектов по классам в датасетах: а – датасет с изображениями ЛО малых размеров; б – датасет с изображениями разномасштабных ЛО

Обучение и валидация моделей СНС

Программная реализация предложенных моделей СНС была выполнена на языке Python версии 3.6 с применением библиотеки PyTorch. Эксперименты проведены на компьютере с графическим процессором Tesla T4, 15102MiB.

В ходе обучения и валидации моделей YOLOv5s и YOLOv8s были проведены эксперименты с целью поиска оптимальных значений гиперпараметров `batch_size` и количества эпох обучения. Выявлено, что наилучшие результаты по точности детектирования ЛО по метрикам AP_{0,5} и mAP_{0,5} достигаются при значении `batch_size`, равном 32, и 100 эпохах обучения. Следует отметить, что при 100 эпохах обучения наблюдается эффект плато, что и стало основанием для выбора именно этого количества эпох.

Результаты исследования моделей СНС и их анализ

Обученные и прошедшие валидацию модели YOLOv5s и YOLOv8s сначала исследовались по точности детектирования малоразмерных объектов на изображениях, а затем в случае изображений с объектами разных размеров. Число изображений в тестовой выборке датасета с объектами малых размеров – 279, тестовая выборка датасета с разномасштабными объектами содержит 417 изображений.

Результаты исследования представлены в таблице 1.

Таблица 1. Результаты тестирования моделей YOLOv5s и YOLOv8s по метрикам AP и mAP на датасетах с ЛО малых и разномасштабных размеров

| Класс | Размеры объектов | | | |
|------------------------|------------------|---------|-----------------|---------|
| | малые | | разномасштабные | |
| | YOLOv5s | YOLOv8s | YOLOv5s | YOLOv8s |
| БПЛА самолетного типа | 0,935 | 0,937 | 0,951 | 0,950 |
| Птица | 0,930 | 0,935 | 0,919 | 0,927 |
| БПЛА вертолетного типа | 0,961 | 0,966 | 0,982 | 0,978 |
| Все классы | 0,942 | 0,946 | 0,951 | 0,952 |

На рисунке 3 приведен пример детектированных объектов на изображениях из тестового набора датасета с объектами малых размеров. Можем видеть, что на левом изображении детектирован объект класса Птица с точностью 0,94, на среднем – БПЛА вертолетного типа с точностью 0,92, а на правом изображении детектирован объект класса БПЛА самолетного типа с точностью 0,92. Важно отметить, что точность всех детектированных объектов превышает заданное пороговое значение 0,9.



Рис. 3. Пример детектированных объектов на изображениях из тестового набора датасета с объектами малых размеров

Комплексность исследований выбранных перспективных моделей СНС требует проведения второго эксперимента для оценки усредненной скорости вычислений каждой из моделей. Результаты этого эксперимента представлены в таблице 2. При вычислении усредненной скорости в виде показателя FPS учитывались такие значения как Inference – отражает среднее время, необходимое для обработки одного изображения моделью, и NMS (англ. Non-Maximum Suppression) – среднее время, затрачиваемое на обработку и устранение дублирующихся обнаружений [8].

Таблица 2. Результаты исследования моделей YOLOv5s и YOLOv8s по усредненной скорости вычисления их на датасетах с объектами малых и разномасштабных размеров

| Модель СНС | Размеры объектов | | | | | |
|------------|------------------|------------------|---------|------------------|-------|------------------|
| | малые | разных масштабов | малые | разных масштабов | малые | разных масштабов |
| | Inference, мс | | NMS, мс | | FPS | |
| YOLOv5s | 3,2 | 2,6 | 2,4 | 3,6 | 179 | 161 |
| YOLOv8s | 4,6 | 4,1 | 1,8 | 2,3 | 156 | 156 |

Исследования показали, что модель YOLOv8s демонстрирует небольшое преимущество по точности детектирования объектов по сравнению с моделью YOLOv5s по всем классам объектов в случае малых размеров этих объектов. Например, для класса БПЛА самолетного типа значение метрики $AP_{0,5}$ у модели YOLOv8s составляет 0,937, в то время как у YOLOv5s – 0,935.

Класс БПЛА вертолетного типа имеет наивысшие значения по точности детектирования среди объектов других классов и размеров. Это может свидетельствовать о том, что обе модели лучше справляются с распознаванием объектов этого класса. Точность детектирования объектов класса Птица ниже, чем у объектов других классов, что может указывать на сложность распознавания объектов данного класса.

Точность детектирования объектов всех классов по метрике $mAP_{0,5}$ с помощью модели YOLOv8s для малых объектов несколько выше, чем у модели YOLOv5s. Это может говорить о том, что модель YOLOv8s предпочтительней для детектирования объектов малых размеров.

Анализируя результаты исследования моделей YOLOv5s и YOLOv8s, представленные в таблице 1, можно сформулировать следующие выводы. Модель YOLOv8s демонстрирует более высокую точность детектирования по сравнению с YOLOv5s для класса Птица в случае и малых, и разномасштабных объектов. Однако модель YOLOv5s незначительно эффективнее модели YOLOv8s в случае разномасштабных объектов классов БПЛА самолетного и вертолетного типа. Важно учитывать при этом, что выбор модели для СКЗ реального времени может зависеть от конкретных классов объектов, которые необходимо распознавать с помощью СКЗ, а также от требований к скорости вычисления модели.

На основании результатов из таблицы 2 можно сделать следующий вывод. В целом, обе модели демонстрируют высокие значения усредненной скорости вычислений. Однако модель YOLOv5s превосходит по метрике FPS модель YOLOv8s как при детектировании малых объектов, так и при детектировании объектов разных размеров.

Результаты проведенных комплексных исследований моделей YOLOv5s и YOLOv8s, а также выводы, основанные на них, позволяют сформулировать практические рекомендации для применения данных моделей в СКЗ реального времени. Для детектирования объектов на изображениях, вне зависимости от категории их размеров, можно применять любую из рассмотренных моделей СНС. Однако если при проектировании СКЗ реального времени выставляется жесткое требование к скорости вычисления модели СНС, то предпочтение следует отдать модели YOLOv5s.

Заключение

Анализ ряда исследований, посвященных точности обнаружения ЛО на изображениях, подтвердил, что проблема низкой точности обнаружения ЛО на изображениях актуальна, особенно для объектов с малыми размерами. Проанализированы основные классы моделей СНС для детектирования объектов на изображениях, в результате были выбраны модели YOLOv5s и YOLOv8s как наиболее перспективные для реализации в СКЗ реального времени. Сформированы два датасета с размеченными изображениями малых ЛО и ЛО различных размеров.

Результаты исследований показали, что обе модели СНС могут быть эффективно применены в СКЗ реального времени для детектирования на изображениях ЛО малых и разных размеров. Однако, учитывая требования к производительности и точности детектирования объектов, выдвигаемые при проектировании конкретной СКЗ, может быть осуществлен выбор той или иной из исследованных моделей. Так, модель YOLOv8s, обладая более высокой точностью детектирования объектов, может быть предпочтительней в СКЗ, которые предназначены для решения задач, где критически важна точность обнаружения и распознавания объектов. Особенно это важно в случае объектов малого размера. В то же время, модель YOLOv5s, демонстрируя более высокую скорость вычисления, может быть более подходящей в условиях, когда для СКЗ реального времени скорость детектирования объектов является приоритетом.

Таким образом, результаты данного исследования подчеркивают важность выбора модели в зависимости от специфики решения задач мониторинга воздушного пространства. Для достижения наилучших результатов детектирования ЛО необходимо учитывать как размеры этих объектов, так и требования к вычислительным ресурсам проектируемой СКЗ реального времени.

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. Nelson J. Your Comprehensive Guide to the YOLO Family of Models [Электронный ресурс]. – URL: blog.roboflow.com/guide-to-yolo-models/ (дата обращения: 14.03.2025).
3. Небаба С.Г. Сверточные нейронные сети семейства YOLO для мобильных систем компьютерного зрения / С.Г. Небаба, Н.Г. Марков // *Компьютерные исследования и моделирование*. – 2024. – Т. 16, № 3. – С. 615–631. DOI: 10.20537/2076-7633-2024-16-3-615-631
4. Accurate UAV Small Object Detection Based on HRFPN and EfficientViT / S. Wu, X. Lu, C. Guo, H. Guo // *Sensors* 2024. – 2024. – Vol. 24, No. 5. – P. 4966. DOI: 10.3390/s24154966.
5. mAP (mean Average Precision) в детекции объектов [Электронный ресурс]. – URL: <https://ml.i-neti.ru/map-mean-average-precision/> (дата обращения: 21.03.2025).
6. Review of deep learning: Concepts, CNN architectures, challenges, applications, future directions / L. Alzubaidi, J. Zhang, A.J. Humaidi, A. Al-Dujaili, Y. Duan, O. Al-Shamma, J. Santamaría, M.A. Fadhel, M. Al-Amidie, L. Farhan // *Journal of Big Data*. – 2021. – Vol. 8, No. 53. – P. 1–77. DOI: 10.1186/s40537-021-00444-8.
7. Roboflow [Электронный ресурс]. – URL: roboflow.com/ (дата обращения: 21.03.2025).
8. Object Detection: Non-Maximum Suppression (NMS) [Электронный ресурс]. – URL: kikaben.com/object-detection-non-maximum-suppression/ (дата обращения: 22.03.2025).