

ОБУЧЕНИЕ МОДЕЛЕЙ СВЕРТОЧНЫХ НЕЙРОННЫХ СЕТЕЙ YOLOv5 и YOLOv7 ПРИ РЕШЕНИИ ЗАДАЧИ МУЛЬТИКЛАССИФИКАЦИИ ОБЪЕКТОВ НА ИЗОБРАЖЕНИЯХ

Тятюшкин И.М.¹, Кравченко А.А.²

¹Томский политехнический университет, ИШИТР, студент гр. 8ИИ1, imt12@tpu.ru,

²Томский политехнический университет, ИШИТР, студент гр. 8ИИ1, aak310@tpu.ru

Аннотация

В работе проводится сравнительный анализ результатов обучения и валидации моделей сверточных нейронных сетей YOLOv5 и YOLOv7 для решения задач мультиклассификации летающих объектов на изображениях.

Ключевые слова: сверточная нейронная сеть YOLO, обучение и валидация нейронных сетей, мультиклассификация летающих объектов на изображениях

Введение

Из-за быстрого роста объема визуальной информации, генерируемой ежедневно, автоматическое распознавание объектов на изображениях становится все более востребованным направлением исследований в области компьютерного зрения и машинного обучения. Одной из наиболее важных задач в этом направлении является классификация нескольких объектов на изображении, принадлежащих нескольким классам (задача мультиклассификации объектов). Модели свёрточных нейронных сетей (СНС) являются одними из наиболее эффективных моделей для решения задач обнаружения и классификации объектов на изображениях. Однако решение задачи мультиклассификации объектов с заданной точностью с помощью таких моделей является весьма нетривиальной задачей, особенно на этапах обучения и валидации используемых моделей СНС[1].

В последние годы были разработаны различные модели СНС. Для решения задач мультиклассификации подвижных объектов наиболее подходящим из современных моделей СНС является класс моделей YOLO (You Only Look Once), поскольку модели этого класса позволяют выполнять распознавание объектов в реальном времени [1, 1]. Для исследователей, решающих задачи мультиклассификации подвижных объектов, большой интерес представляют перспективные модели YOLOv5 и YOLOv7, входящие в класс YOLO.

Целью данной работы является определение по результатам обучения и валидации наиболее эффективной модели СНС из исследуемых моделей YOLOv5 и YOLOv7 при решении задачи мультиклассификации летающих объектов на изображениях. Полученные результаты могут быть использованы при разработке систем компьютерного зрения, способных решать сложные задачи мультиклассификации летающих объектов.

Модели СНС YOLOv5 и YOLOv7

Класс моделей YOLO является сегодня одним из наиболее популярных и современных классов СНС для решения задач обнаружения и классификации объектов на изображениях. Модели СНС этого класса, как и другие модели СНС с одноэтапной архитектурой, анализируют изображение за один проход (одномоментно выполняются этапы обнаружения, локализации и классификации объектов), что значительно повышает скорость анализа изображения. При этом такие модели СНС обладают весьма высокими показателями по точности детектирования объектов[1-3]. Кратко рассмотрим две наиболее перспективные модели из класса YOLO - YOLOv5 и YOLOv7.

YOLOv5 - это модель СНС, разработанная для решения задач распознавания объектов в реальном времени. Эта модель основана на архитектуре YOLOv4, но имеет ряд улучшений, которые позволяют увеличить точность и скорость распознавания объектов[4,5]. Архитектура YOLOv5 (Рис. 1) состоит из трех основных компонентов: backbone, neck и head. Backbone - это основная часть модели, которая отвечает за извлечение признаков объектов на изображении. В модели YOLOv5 в качестве backbone используется модель CSPDarknet53, которая является улучшенной версией модели Darknet53[4]. Neck - это часть модели, которая отвечает за объединение признаков объектов из различных слоев backbone. В YOLOv5 в качестве компоненты neck используется модуль FPN (Feature Pyramid Network), который позволяет объединять признаки из различных масштабов изображения. Head - это часть модели, которая отвечает за выдачу конечного результата – обнаружение

и классификацию объектов на изображении. В YOLOv5 в качестве head используется модуль модели YOLOv3, который позволяет определять координаты и классы объектов на изображении[5].

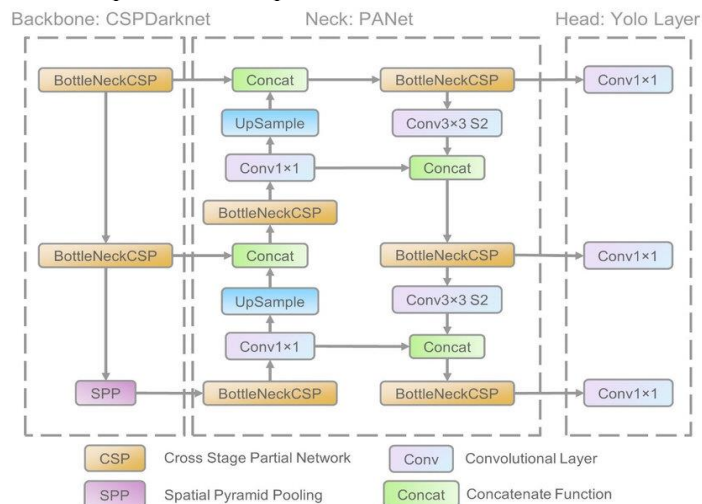


Рис. 1. Архитектура модели YOLOv5

Модель YOLOv7 - это модель СНС, разработанная для решения задач классификации объектов с высокой точностью[6]. Она также основана на архитектуре YOLOv4, но имеет ряд улучшений, которые позволяют увеличить точность классификации объектов на изображениях с высоким разрешением. Архитектура YOLOv7 состоит из трех основных компонентов: backbone, neck и head. Backbone - это основная часть модели, которая отвечает за извлечение признаков изображения. В YOLOv7 в качестве backbone используется модель EfficientNet, которая является высокоэффективной моделью для извлечения признаков изображения. Neck - это часть модели, которая отвечает за объединение признаков из различных слоев backbone. В модели YOLOv7 в качестве компоненты neck используется модуль BiFPN (англ. Bi-directional Feature Pyramid Network), который позволяет объединять признаки из различных масштабов изображения в двух направлениях. Head - это часть модели, которая отвечает за выдачу конечного результата - обнаружение и классификацию объектов на изображении. В YOLOv7 в качестве head используется модуль YOLOv4, который позволяет определять координаты и классы объектов на изображении.

Как следует из описания моделей, обе они имеют схожую архитектуру, но имеют особенности, которые позволяют им эффективно решать задачи обнаружения и классификации объектов в реальном времени и выполнять это с высокой точностью классификации.

Постановка задачи исследований

Необходимо решить задачу мультиклассификации летающих объектов четырех классов на изображениях. Для ее решения предлагается использовать две перспективные модели СНС класса YOLO: модели YOLOv5 и YOLOv7. Каждая из этих моделей, в свою очередь, является семейством моделей, отличающихся друг от друга, в первую очередь, числом сверточных слоев.

Целью исследования является проведение сравнительного анализа точности классификации объектов и скорости вычисления (обучения) двух моделей из каждого из этих семейств моделей YOLOv5 и YOLOv7 на этапах их обучения и валидации при решении задачи мультиклассификации летающих объектов на изображениях. Необходимо оценить точность классификации таких объектов по известным метрикам, а также время, необходимое для анализа одного изображения (время вычисления модели) для каждой из моделей.

При проведении исследования используются известные метрики оценки точности классификации объектов на изображениях, такие как Precision, Recall, AP0.5 для каждого класса (средняя по всем классам точность mAP0.5) и AP0.5:0.95 (средняя по всем классам mAP0.5:0.95) [1]. Также оценивается время, необходимое для анализа изображений в случае каждой модели. После этого необходимо сравнить полученные результаты и сделать выводы. В итоге определяется наиболее эффективная на этапах обучения и валидации модель СНС класса YOLO для решения задачи мультиклассификации летающих объектов.

Для проведения исследования был сформирован набор данных - датасет, содержащий размеченные и исходные изображения летающих объектов четырех классов: Птица (bird), Беспилотный летательный аппарат (БПЛА) вертолетного типа (helicopter-type_uav), БПЛА самолетного типа (aircraft-type_uav) и Неизвестный объект (unknown). В качестве неизвестных объектов на изображениях могут быть самолеты, вертолеты, воздушные шары и т.п. Датасет состоит из 15200 изображений размером 416x416 пикселей. В нем 70 % от общего объема составляют изображения обучающей выборки, 20 % - валидационной и 10 % - тестовой выборки. Изображения были выбраны таким образом, чтобы на них присутствовали как одиночные объекты, так и несколько объектов разных классов.

Обучение и валидация моделей YOLOv5 и YOLOv7

Обучение и валидация этих моделей СНС проводились с использованием фреймворка PyTorch [1, 2]. В качестве среды обучения использовался Kaggle Kernels, который предоставляет ограниченный, но бесплатный доступ к вычислительным ресурсам с возможностью обучать модели СНС на двух графических процессорах Tesla T4. Так как организация – владелец вычислительной среды накладывает определенные ограничения на время сеанса и используемые вычислительные единицы, обучение каждой модели проходило для 10 эпох.

Модели семейства YOLOv5

Было выбрано две модели из семейства YOLOv5 для обучения – YOLOv5s и YOLOv5x. Эксперименты показали, что обучение модели YOLOv5s в течение 10 эпох заняло 1.829 часа, что соответствует примерно 0.43 секунды на анализ одного изображения. Соответственно, обучение модели YOLOv5x в течение 10 эпох заняло 1.847 часа, что соответствует примерно 0.44 секунды на анализ одного изображения из обучающей выборки. Видим, что модель YOLOv5x вычисляется несколько дольше, чем модель YOLOv5s. На наш взгляд, это связано, в первую очередь, с тем, что обучающая выборка используемого размера с изображениями небольшого размера (416x416 пикселей) не может полностью загрузить графический ускоритель и тогда общая длительность процесса обучения определяется ресурсами процессора, который подготавливает данные для обучения и является в данном случае «узким горлышком».

Во время обучения получены оценки точности классификации летающих объектов на изображениях по четырем выбранным метрикам. Результаты представлены на рис. 2 в виде графиков зависимостей значений этих метрик от количества эпох обучения для каждой из моделей СНС. Следует отметить, что даже при небольшом числе эпох точность классификации объектов весьма высока.

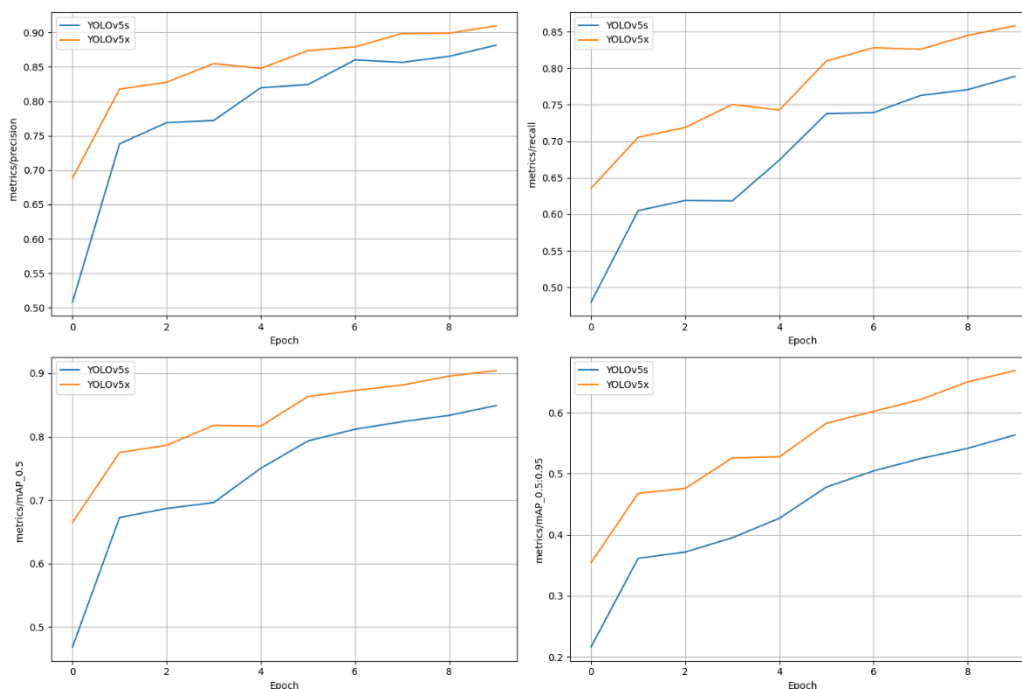


Рис. 2. Графики зависимости значений метрик от числа эпох обучения

Как можно видеть из графиков, модель YOLOv5x дает более высокую точность классификации по всем метрикам, чем модель YOLOv5s.

В таблице 1 приведены результаты валидации этих моделей. Голубым цветом показаны значения метрик для модели YOLOv5s, а оранжевым - полученные с помощью модели YOLOv5x. Из анализа этих результатов следует, что, как и на обучающей выборке, точность классификации летающих объектов по всем метрикам на валидационной выборке с помощью модели YOLOv5x выше, чем в случае использования модели YOLOv5s. Причем это имеет место для всех классов объектов.

Таблица 1

Точность классификации объектов по классам, полученная на валидационной выборке

	Precision	Recall	AP0.5, mAP0.5	AP0.5:0.9 mAP0.5:0.95
all	0.853 / 0.897	0.840 / 0.890	0.877 / 0.923	0.566 / 0.667
bird	0.841 / 0.890	0.771 / 0.830	0.817 / 0.887	0.548 / 0.675
helicopter-type_uav	0.826 / 0.860	0.833 / 0.905	0.877 / 0.914	0.515 / 0.598
aircraft-type_uav	0.942 / 0.966	0.927 / 0.949	0.960 / 0.971	0.697 / 0.795
unknown	0.802 / 0.872	0.830 / 0.876	0.855 / 0.919	0.505 / 0.602

Также по результатам валидации моделей можно сказать, что наиболее хорошо моделям из семейства YOLOv5 удастся обнаруживать и классифицировать БПЛА самолетного типа. Возможно, это связано с их хорошо отличимыми признаками, такими как геометрия фиксированного крыла и т.п.

Модели семейства YOLOv7

Из семейства YOLOv7 также было выбрано две модели для обучения – YOLOv7 и YOLOv7x. Экспериментальным путем были получены следующие результаты: обучение модели YOLOv7 заняло 4.8 часа на 10 эпох, что соответствует примерно 1.14 секунды на анализ одного изображения, с другой стороны, модель YOLOv7x обучалась 5 часов для 10 эпох, это соответствует примерно 1.19 секунды на анализ одного изображения. Разница во времени обучения этих моделей осталась, как и в случае моделей семейства YOLOv5, незначительной. Это обусловлено той же причиной, что и при обучении моделей семейства YOLOv5: процессор не успевает подготавливать достаточно данных для обучения модели на графических ускорителях Tesla T4.

При обучении этих моделей также были получены оценки точности классификации объектов на изображениях по четырем выбранным метрикам. Результаты представлены на рис. 3 в виде графиков зависимостей значений этих метрик от количества эпох обучения для каждой из моделей СНС. Следует отметить, что даже при небольшом числе эпох точность классификации объектов весьма высока.

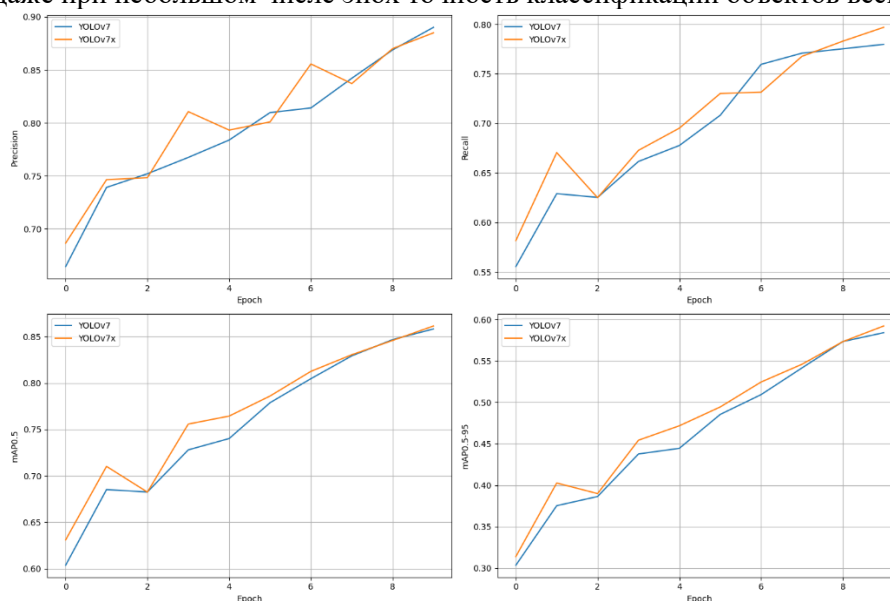


Рис. 3. Графики зависимости значений метрик от числа эпох обучения

Из графиков на рисунке 3 следует, что модель YOLOv7x показывает лучшие результаты, чем модель YOLOv7. Однако разница получилась небольшой, если сравнивать с разницей результатов по точности классификации объектов моделями семейства YOLOv5. Это объясняется выбранными для исследований из семейств моделями: отличие архитектуры моделей YOLOv5s и YOLOv5x по числу сверточных слоев значительно больше, чем это имеет место по числу таких слоев в архитектуре моделей YOLOv7 и YOLOv7x.

В таблице 2 приведены результаты валидации этих моделей. Голубым цветом показаны значения метрик для модели YOLOv7, а оранжевым - полученные с помощью модели YOLOv7x. Из анализа этих результатов следует, что, как и на обучающей выборке, точность классификации летающих объектов на валидационной выборке примерно одинаковая для этих моделей.

Таблица 2

Точность классификации объектов по классам, полученная на валидационной выборке

	Precision	Recall	AP0.5, mAP0.5	AP0.5:0.9 mAP0.5:0.95
all	0.890 / 0.885	0.780 / 0.797	0.858 / 0.861	0.584 / 0.592
bird	0.860 / 0.852	0.674 / 0.680	0.758 / 0.752	0.550 / 0.543
helicopter-type_uav	0.915 / 0.904	0.875 / 0.883	0.928 / 0.925	0.671 / 0.677
aircraft-type_uav	0.925 / 0.942	0.820 / 0.835	0.917 / 0.929	0.671 / 0.651
unknown	0.860 / 0.843	0.749 / 0.790	0.830 / 0.840	0.491 / 0.497

Также по результатам валидации этих моделей можно сказать, что наиболее хорошо им удастся обнаруживать и классифицировать БПЛА вертолетного типа. Этим они отличаются от моделей семейства YOLOv5, которые лучше распознают БПЛА самолетного типа.

Заключение

В данной работе было проведено исследование процессов обучения и валидации четырех моделей СНС из семейств YOLOv5 и YOLOv7 при решении задачи мультиклассификации летающих объектов на изображениях. Для обучения и валидации этих моделей использовался датасет, содержащий изображения летающих объектов четырех классов на 15200 изображений размером 416x416 пикселей.

По результатам обучения и валидации моделей семейства YOLOv5 можно сказать, что модель YOLOv5x значительно точнее решает задачи мультиклассификации летающих объектов, чем YOLOv5s при сравнительно близком времени на обучение. Таким образом, можно сделать вывод о том, что при обучении моделей семейства YOLOv5 на датасетах, которые не могут обеспечить полную загрузку графического ускорителя (используются малые по размерам изображения), предпочтительнее будет использовать модель YOLOv5x, так как сравнительно за такое же время можно получить значительно лучшие по точности классификации результаты обучения.

Результаты по времени обучения моделей семейства YOLOv7 получились похожими на результаты для семейства моделей YOLOv5, поскольку датасет для обучения и валидации использовался тот же – получили ту же проблему: графический ускоритель простаивает, так как не получает достаточного количества данных для полной загрузки. То есть, по времени обучения выбранные модели семейства YOLOv7 близки. Более того, отличия по точности классификации объектов на обучающей и валидационной выборках незначительны, но в пользу модели YOLOv7x. Таким образом, при работе с датасетами такого размера также будет предпочтительнее модель YOLOv7x.

При сравнении по времени обучения старших моделей YOLOv5x и YOLOv7x этих семейств лучшей является модель YOLOv5x. Однако о точности классификации летающих объектов с помощью этих моделей однозначный вывод сделать пока нельзя, поскольку требуются дополнительные результаты обучения моделей при значительно большем числе эпох. При сравнении младших моделей семейств по точности классификации получаем примерно одинаковые результаты при значительно более долгом обучении модели YOLOv7, что делает модель YOLOv5s предпочтительнее.

В итоге можно считать, что выбранные четыре модели из семейств YOLOv5 и YOLOv7 способны решать задачу мультиклассификации летающих объектов на изображениях. При этом модели из семейства YOLOv5 обучаются значительно быстрее, чем модели семейства YOLOv7. Однако

окончательный выбор модели зависит от конкретной решаемой задачи мультиклассификации и требований к точности классификации летающих объектов.

Результаты данного исследования могут быть использованы при разработке систем компьютерного зрения, предназначенных для решения задач мультиклассификации летающих объектов на изображениях.

Список использованных источников

1. Alzubaidi L., Zhang J., Humaidi A.J., Al-Dujaili A., Duan Y., Al-Shamma O., Santamaría J., Fadhel M.A., Al-Amidie M., Farhan L. Review of deep learning: Concepts, CNN architectures, challenges, applications, future directions // Journal of big Data. – 2021. – Vol. 8. – No. 1. – P. 53.
2. Свёрточная нейронная сеть // Хабрахабр. – URL: <https://habr.com/ru/articles/309508/> (дата обращения: 15.03.2024).
3. Как работает Object Tracking на YOLO и DeepSort // Хабрахабр. – URL: <https://habr.com/ru/articles/514450/> (дата обращения: 15.03.2024).
4. ULTRALYTICS // YOLOv5 // Документация. – URL: <https://docs.ultralytics.com/models/yolov5/>
5. GITHUB // YOLOv5 in PyTorch > ONNX > CoreML > TFLite // Репозиторий. – URL: <https://github.com/ultralytics/yolov5> (дата обращения: 15.03.2024).
6. ULTRALYTICS // YOLOv7 // Документация. – URL: <https://docs.ultralytics.com/models/yolov5/>
7. GITHUB // Implementation of paper - YOLOv7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors // Репозиторий. – URL: <https://github.com/WongKinYiu/yolov7> (дата обращения: 11.03 2024).
8. Знакомимся с методом обратного распространения ошибки // Хабрахабр. – URL: <https://habr.com/ru/companies/otus/articles/483466/> (дата обращения: 15.03.2024).
9. KAGGLE // Datasets // Сайт. – URL: <https://www.kaggle.com/datasets> (дата обращения: 20.03.2024).