

Детекция разметки взлётно-посадочной полосы с использованием нейронных сетей

А. Р. Музалевский¹, Е. В. Серых, М. М. Копычев, Е. В. Дряян
Санкт-Петербургский государственный электротехнический университет
«ЛЭТИ» им. В.И. Ульянова (Ленина)
¹muzal2010@yandex.ru

Аннотация. В статье рассматриваются вопросы использования свёрточной нейронной сети для решения задачи детекции горизонтальной разметки на взлётно-посадочной полосе на основе данных камеры. Для обучения сети был создан набор данных с использованием открытых источников. В качестве базовой архитектуры нейронной сети была выбрана YOLOv5.

Ключевые слова: детекция; нейронная сеть; разметка; взлётно-посадочная полоса

I. ВВЕДЕНИЕ

Задача детекции объектов на изображении – задача машинного обучения, в рамках которой на изображении выделяются один или несколько объектов целевых классов посредством нахождения координат прямоугольников, описывающих эти объекты [1].

В качестве метрик в задаче детекции чаще всего используются IoU (англ. Intersection over Union) – метрика степени пересечения между двумя ограничивающими рамками и mAP (англ. Mean Average Precision) – усреднённая по всем целевым классам величина средней точности.

Модели детекции применяются в различных областях:

- Беспилотные транспортные средства. Автомобилю необходимы различные модели для детекции дорожной разметки, дорожных знаков, автомобилей и прочих целевых классов, чтобы планировать в соответствии с этой информацией дальнейший безопасный маршрут.
- Медицина. Определение злокачественных клеток, поражений органов, аномалий.
- Промышленность. Контроль соблюдения зон и периметра; учёт и контроль товаров; контроль качества, обнаружение дефектов.
- Системы распознавания и видеонаблюдения. Поиск людей, системы обнаружения дорожных пробок и аварий.

Целью данной работы являются проверка осуществимости решения задачи детекции разметки на взлётно-посадочной полосе на небольшой выборке данных, созданной на базе открытых источников, и оценка влияния наличия в обучающей выборке фоновых кадров (не содержащих объекты целевых классов) на качество работы модели.

II. ОБЗОР РАБОТ СХОЖЕЙ ТЕМАТИКИ

Детекция дорожной разметки является известной задачей машинного обучения, для решения которой существуют различные методы и подходы.

Аналитический подход. В типичной задаче детекции разметки первым шагом является цветовая сегментация исходного изображения, поскольку во всем мире разметка полос движения преимущественно окрашена в желтый или белый цвета. Второй шаг – обнаружение границ. Затем используются различные техники уменьшения шума, после чего применяется преобразование Хафа [2], позволяющее выявить различные линии на изображении. В конце производится анализ выявленных линий в соответствии с задачей. Применимость такого подхода ограничена его небольшой устойчивостью к изменению изображения (например, изменение освещённости).

Методы глубокого обучения всё чаще используются для решения задачи детекции объектов, поскольку развитие существующих архитектур нейронных сетей направлено как на повышение точности, так и на снижение требований к вычислительным ресурсам, что позволяет получить компактную и точную систему детекции.

В работе [3] предложен собственный метод решения задачи детекции разметки LaneNet. В основе этого метода лежит последовательная работа двух нейронных сетей, первая сеть основана на архитектуре энкодер-декодер и предназначена для определения границ линий разметки, выход этой сети подаётся на вход второй, осуществляющей локализацию линий разметки.

В работе [4] предложена оригинальная архитектура нейронной сети VPGNet, предназначенной для детекции дорожной разметки. В ней для повышения точности в условиях слабой видимости используется положение точки схода линий разметки.

В [5] описана структура PointLaneNet, в которой свёрточная нейронная сеть извлекает признаки из исходного изображения с фронтальной камеры, которые затем поступают на вход следующей сети – ConvLaneNet, оценивающей вероятность существования линий разметки в каждом пикселе исходного кадра.

В [6] для детекции разметки используется комбинация признаков, обнаруженных на изображении с фронтальной камеры и на виде сверху. К виду сверху применяется фильтр, позволяющий выявить кандидатов в линии, которые поступают на вход свёрточной нейронной сети.

III. ОПИСАНИЕ АРХИТЕКТУРЫ НЕЙРОННОЙ СЕТИ

В основе данной работы лежит модель свёрточной нейронной сети YOLOv5 (You Only Look Once [7]). Основной задачей, для решения которой предназначена эта архитектура, является детекция объектов.

Данная модель относится к архитектуре One-Stage Detector. В рамках этого подхода осуществляется предсказание координат определённого количества описывающих прямоугольников с результатами классификации и вероятностью нахождения объекта. В дальнейшем модель корректирует положение описывающих прямоугольников. На рис. 1 представлена архитектура сети в обобщённом виде.

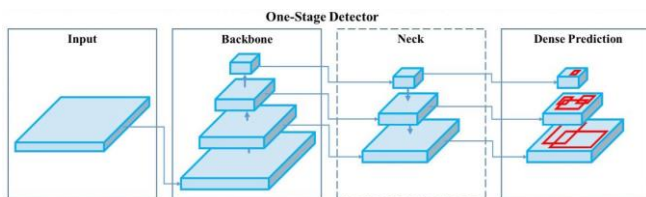


Рис. 1. Архитектура YOLOv5 в обобщённом виде

Данная архитектура отличается своей скоростью и точностью работы, а также простотой использования. Исходное изображение сжимается таким образом, чтобы получилась квадратная матрица, в каждой клетке которой записана информация о наличии объекта и классе этого объекта на соответствующей части кадра. Модель YOLO анализирует изображение только один раз, что существенно увеличивает скорость обработки в сравнении с другими архитектурами.

IV. ОПИСАНИЕ НАБОРА ДАННЫХ ДЛЯ ОБУЧЕНИЯ

Детекция линий разметки на взлётно-посадочной полосе является довольно узкоспециализированной задачей. Линии разметки по внешнему виду совпадают с разметкой на дорогах общего пользования, но стандарты её нанесения сильно отличаются. Что приводит к невозможности использования наборов данных, предназначенных для создания беспилотных автомобилей, так как пропорции и ракурсы разметки в значительной степени отличаются.

В результате анализа открытых источников данных с взлётно-посадочных полос, подходящей информации найдено не было, поэтому специально для данной задачи был создан собственный набор данных. Разметка для задачи детекции является менее трудоёмкой, чем для задачи сегментации, что позволяет в короткие сроки получить данные для проверки работоспособности выдвинутой гипотезы.

В качестве исходных данных были выбраны шесть любительских видео от пилотов самолётов (ракурс из кабины самолёта), которые были сделаны при посадке и выложены в открытый доступ. Все видео были разбиты на отдельные кадры. В программной среде CVAT [8] данные кадры были обогащены разметкой двух классов: «сплошная белая линия» и «прерывистая белая линия». Данные классы представляют собой две наиболее распространённые категории разметки на взлётно-посадочной полосе. В дальнейшем список классов может быть расширен и дополнен такими категориями как «жёлтая сплошная линия», «начало полосы», «конец полосы» и другие более специфические обозначения, характерные для аэродромного полотна.

Из всего набора данных была создана выборка, размер которой составил 14 650 кадров. 59 % кадров обладают размером (1080, 1920) пикселей, 41 % кадров – (1080, 1440). Поскольку исходные видео были взяты из открытых источников, они обладают рядом особенностей: наложение текста и таблиц в части

кадров, резкая смена ракурса или смазывание границ при тряске.

Данный объём данных был разделён на три выборки:

- обучающая выборка: 10562 кадра (~70 %), 7337 кадров содержат целевые классы;
- валидационная выборка: 2136 кадра (~15 %), 1980 кадров содержат целевые классы;
- тестовая выборка: 1952 кадра (~15 %), все кадры содержат целевые классы.

Кроме того, отдельно был выделен дополнительный тестовый набор (7036 кадров), содержащий только изображения без целевых классов разметки. Данный датасет был введён, чтобы оценить степень устойчивости модели, а именно количество ложных срабатываний.

При разделении выборок во избежание утечек информации, кадры из одного видео были изолированы и не могли попасть в разные выборки. Таким образом, в обучающей выборке находится 3 видео, в валидационной и в тестовой по одному, последнее видео использовалось для дополнительного тестового набора.

Пример подготовленного кадра с истинной разметкой представлен на рис. 2.

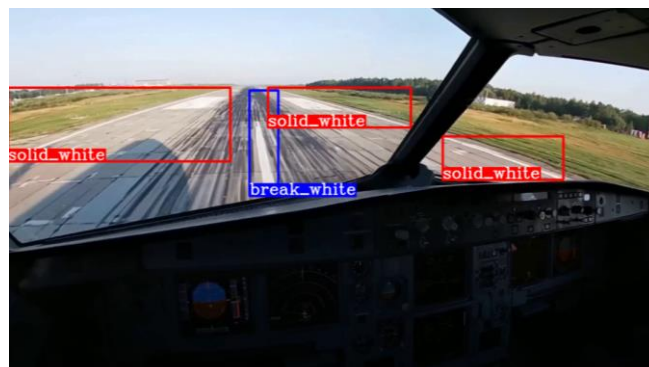


Рис. 2. Пример подготовленного кадра с истинной разметкой

На рис. 3 представлено распределение экземпляров разметки в каждой выборке. На одном кадре может быть сразу несколько линий разметки одного класса.

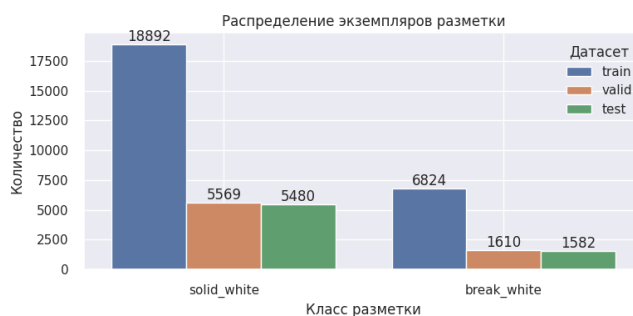


Рис. 3. Распределение экземпляров разметки по датасетам

Можно видеть, что в разметке присутствует дисбаланс, класс «сплошная линия» представлен мажорным числом экземпляров разметки, что объясняется тем, что зачастую сплошные линии разметки расположены по обеим сторонам взлётно-посадочной полосы.

Каждый кадр представлен RGB-изображением, на вход нейронной сети подавалось изображение,

отмасштабированное к размеру (640, 640). В качестве начальных весов модели использовались веса yolov5s, содержащие 7,2 миллиона параметров. Нейронная сеть обучалась в течение 100 эпох, с использованием GPU Nvidia 2080ti, для мониторинга процесса обучения использовался ClearML.

V. РЕЗУЛЬТАТЫ

В данной работе представлены результаты двух моделей. Первая была обучена без фоновых кадров (кадров без линий разметки), вторая была обучена с добавлением фоновых кадров.

На рис. 4–7 представлены графики обучения нейронной сети с фоновыми кадрами. По графикам видно, что модель не переобучилась и, начиная с 80 эпохи, вышла на плато.

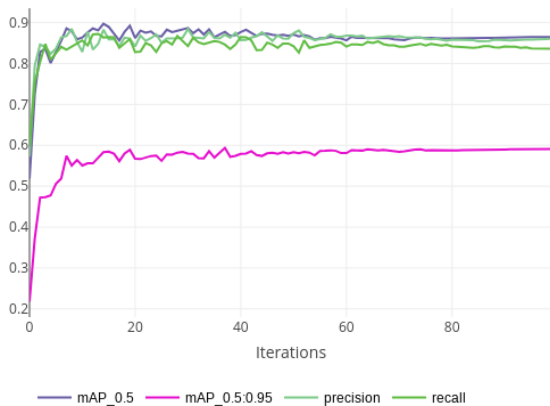


Рис. 4. График изменения метрик качества в процессе обучения

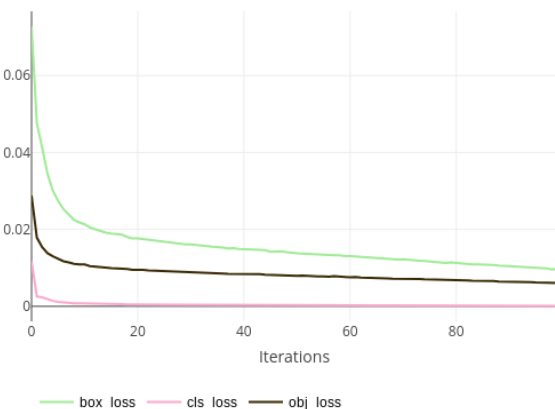


Рис. 5. График изменения функции потерь (на обучающей выборке)

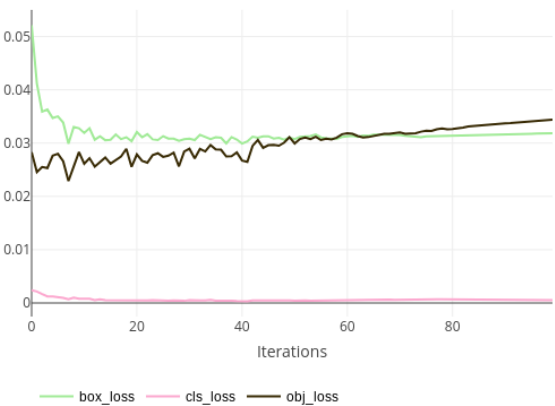


Рис. 6. График изменения функции потерь (на валидационной выборке)

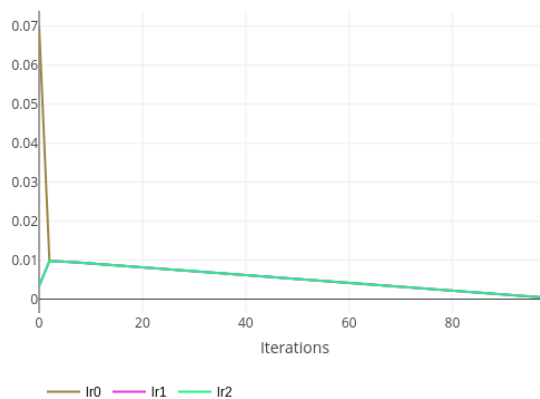


Рис. 7. График изменения коэффициента скорости обучения

В табл. 1 представлены метрики качества работы на тестовой выборке модели №1, обученной без использования фоновых кадров.

ТАБЛИЦА I. МЕТРИКИ МОДЕЛИ

Целевой класс	Показатели			
	Precision	Recall	mAP50	mAP50-95
all	0.862	0.861	0.884	0.602
solid white	0.952	0.869	0.93	0.698
break white	0.773	0.854	0.839	0.507

В табл. 2 представлены метрики качества работы на тестовой выборке модели №2, обученной с использованием фоновых кадров.

ТАБЛИЦА II. МЕТРИКИ МОДЕЛИ

Целевой класс	Показатели			
	Precision	Recall	mAP50	mAP50-95
all	0.855	0.855	0.879	0.606
solid white	0.959	0.843	0.93	0.7
break white	0.751	0.867	0.828	0.512

Можно заметить, что обе модели детектируют прерывистую линию хуже, чем сплошную. В этом случае дисбаланс разметки может негативно сказываться на качестве обучения моделей.

Кроме того, значения метрик mAP50-95 находятся на относительно низком уровне, но поскольку в данной задаче точность положения детектируемых областей не является приоритетом, в качестве основной метрики качества модели рассматривается mAP50.

В ходе визуального анализа результатов моделей №1 и №2 было замечено, что определяемые области находятся достаточно близко друг к другу, что также видно в представленных выше метриках. Поэтому далее будут приведены примеры работы только модели №2.

На рис. 8 представлен пример успешно произведённой детекции линий разметки с помощью модели №2 (с фоновыми кадрами).

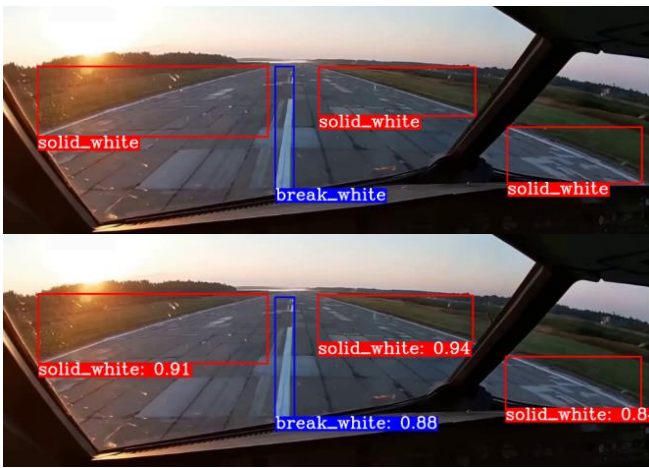


Рис. 8. Пример кадра, на котором модель №2 успешно определила все целевые линии разметки (верхний кадр – целевая разметка, нижний кадр – предсказания модели)

На рис. 9 представлен пример, когда модель №2 не смогла определить левую сплошную линию из-за наличия тумана в кадре.



Рис. 9. Пример кадра, на котором модель №2 не смогла определить левую сплошную линию разметки (верхний кадр – целевая разметка, нижний кадр – предсказания модели)

Можно видеть, что низкое качество взлётно-посадочной полосы не вводит модель в заблуждение, однако различные погодные условия значительно влияют на её работоспособность. Обогащение обучающей выборки кадрами с различным временем суток и погодными условиями может положительно сказаться на способности модели в дальнейшем обрабатывать более широкий диапазон входных данных.

В табл. 3 представлены результаты тестирования моделей на наличие ложных срабатываний на второй тестовой выборке.

ТАБЛИЦА III. МЕТРИКИ МОДЕЛИ

Имя модели	FP	FP (%)
Модель №1	759	11%
Модель №2	206	3%

На рис. 10 представлен пример кадра, где модель №1 совершила ложное срабатывание, в то время как модель №2 – нет.



Рис. 10. Пример ложного срабатывания модели №1

На рис. 11 представлен пример кадра, где модель №2 совершила ложное срабатывание, в то время как модель №1 – нет.



Рис. 11. Пример ложного срабатывания модели №2

На рис. 12 представлен пример кадра, где обе модели совершили ложные срабатывания.



Рис. 12. Пример ложных срабатываний обеих моделей

Обе модели успешно справились с детектированием линий разметки на взлётно-посадочной полосе, что подтверждает гипотезу о реализуемости данной задачи. Разница в метриках качества между двумя моделями незначительна, однако модель №1 совершила в 3,7 раз больше ложных срабатываний, чем модель №2 на второй тестовой выборке, что показало положительное влияние добавления фоновых кадров в обучающую выборку на качество работы модели.

VI. ЗАКЛЮЧЕНИЕ

В данной работе была осуществлена проверка осуществимости детекции линий разметки на взлётно-посадочной полосе на небольшом количестве данных. Результаты подтверждают осуществимость данной задачи. Обучение модели с использованием фоновых кадров (без целевых классов разметки) положительно сказывается на устойчивости модели в целом (количество ложных срабатываний уменьшается). Стоит отметить, что часть ложных срабатываний была

получена на кадрах с линиями рулёжных дорожек (тонкая сплошная белая линия), не представленными в обучающей выборке.

В качестве развития работы в дальнейших исследованиях планируется:

- увеличить объём данных для обучения, валидации и тестирования,
- увеличить разнообразие данных (различные времена года, погодные условия и время суток), чтобы повысить степень устойчивости модели,
- расширить набор классов разметки дополнительными категориями (при условии их достаточного количества в собранных данных),
- использовать в качестве начальных весов модели большего размера (medium, large, extra large),
- рассмотреть альтернативный подход к разметке данных, а именно размечать отрезками линии и использовать подходы на основе предсказания ключевых точек объектов.

СПИСОК ЛИТЕРАТУРЫ

- [1] Курс лекций по компьютерному зрению. Лекция 11. Детекция и сегментация. Стэнфорд, 2017. Онлайн-ресурс. URL: http://cs231n.stanford.edu/slides/2017/cs231n_2017_lecture11.pdf
- [2] Hough Transform. Онлайн-ресурс. URL: <https://homepages.inf.ed.ac.uk/rbf/HIPR2/hough.htm>
- [3] Wang, Ze, Weiqiang Ren and Qiang Qiu. "LaneNet: Real-Time Lane Detection Networks for Autonomous Driving." ArXiv abs/1807.01726 (2018): n. pag.
- [4] S. Lee, J. Kim, J. S. Yoon, S. Shin, O. Bailo, N. Kim, T.-H. Lee, H. S. Hong, S.-H. Han, I. S. Kweon, Vpynet: Vanishing point guided network for lane and road marking detection and recognition, in: Proceedings of the IEEE international conference on computer vision, 2017, pp. 1947–1955.
- [5] Z. Chen, Q. Liu and C. Lian, "PointLaneNet: Efficient end-to-end CNNs for Accurate Real-Time Lane Detection," 2019 IEEE Intelligent Vehicles Symposium (IV), Paris, France, 2019, pp. 2563–2568, doi: 10.1109/IVS.2019.8813778.
- [6] He, Bei & Ai, Rui & Yan, Yang & Lang, Xianpeng. (2016). Accurate and robust lane detection based on Dual-View Convolutional Neural Network. 1041-1046. 10.1109/IVS.2016.7535517.
- [7] Redmon, Joseph & Divvala, Santosh & Girshick, Ross & Farhadi, Ali. (2016). You Only Look Once: Unified, Real-Time Object Detection. 779-788. 10.1109/CVPR.2016.91.
- [8] Computer Vision Annotation Tool (CVAT). Онлайн-ресурс. URL: <https://www.cvat.ai/>