

Облегченная модифицированная сеть YOLO для детектирования объектов дорожной сцены

Бобков А.В., Ду Кэхао, Дай Ифань, Ван Чжун, Чэнь Хао

Московский государственный технический университет имени Н. Э. Баумана

Аннотация: В работе рассмотрена облегченный модифицированный вариант нейронной сети YOLO-v5, который используется для распознавания объектов дорожной сцены в задаче управления беспилотным автомобилем. В предложенной модели слой субдискретизации (пулинга) заменён на модуль ADown с целью снижения сложности модели. Добавлен модуль C2f в качестве модуля извлечения признаков, чтобы повысить точность за счет объединения признаков. Приведены эксперименты с использованием сцен заснеженных дорог и показана эффективность предложенной модели для распознавания объектов.

Ключевые слова: распознавание объектов дорожной сцены, YOLOv5, ADown, C2f, глубокое обучение, слой субдискретизации, нейронной сети, облегченная сети, набор данных

Введение

Использование частично или полностью автономных автомобилей ставит целью улучшение дорожно-транспортной ситуации и повышения безопасности дорожного движения путем повышения скорости реакции, снижения усталости и нагрузки на человека-оператора, повышение его ситуационной осведомлённости. Основой любой современной автоматической системы вождения является система распознавания дорожных объектов, определения их типа, положения, параметров движения и совместных маневров. Поэтому именно распознавание объектов дорожной сцены является приоритетным направлением современных исследований.

По способу поиска объекта методы машинного обучения можно разделить на однопроходные и двухпроходные. Двухпроходные алгоритмы сначала генерируют набор областей, в которых потенциально может находиться объект интереса, а затем модуль распознавания выполняет классификацию и идентификацию объектов в отобранных областях. Сюда можно отнести такие методы, как R-CNN (Region Convolution Neural Network,

«Нейронная сеть свертки регионов» [1]), Fast R-CNN [2], Faster R-CNN [3] и аналогичные. Двухпроходные алгоритмы позволяют добиться высокой точности обнаружения объекта, однако скорость их работы остается низкой и недостаточна для практических приложений. Однопроходные алгоритмы не требуют предварительного отбора областей интереса, операции выделения признаков, обнаружения объектов и их распознавания объединены и выполняются за один проход. Представителями данного направления являются алгоритмы SSD (Single Shot Detector, «Детектор с одним выстрелом») и YOLO (You Look Only Once, «Посмотри на изображение один раз»), к которым относится YOLOv1 [4], YOLO9000 [5], YOLOv3 [6], YOLOv4 [7]. Алгоритмы на основе сети YOLO работают гораздо быстрее, чем многопроходные алгоритмы, что делает их удобным инструментом для многих практических задач, в частности для беспилотных автомобилей.

Однако в реальных инженерных реализациях по-прежнему существует проблема, заключающаяся в том, что алгоритм распознавания и обнаружения цели отнимает много системных ресурсов, что влияет на дальнейшее развитие технологии автономных автомобилей. В этой статье рассмотрен улучшенный алгоритм YOLO, который позволяет значительно сократить объем вычислений нейросетевых моделей при сохранении точности распознавания.

1. Архитектура предлагаемого модуля

Семейство поисковых сетей YOLO представляет собой набор одноэтапных алгоритмов, которые можно использовать для распознавания объектов. Сеть YOLO позволяет определить класс объекта на изображении, его размер и положение в виде ограничивающей рамки, а также вероятность обнаружения объекта в данном положении.

YOLOv5s является наиболее часто используемой моделью в бортовых

приложениях реального времени, а объем её параметров составляет 7,25 МБ. Скорость её работы относительно высока, но в сценариях управления транспортными средствами с ограниченными ресурсами хранения, ограниченном вычислительном ресурсом и относительно высокими требованиями к работе в режиме реального времени, зависимость глубоких сверточных нейронных сетей от системных ресурсов сильно ограничивает её развертывание и применение на практике.

2. Модуль ADown

Модуль ADown (рис. 1) представляет собой блок понижающей субдискретизации (пулинга). Он снижает сложность модели за счет уменьшения количества параметров, что помогает повысить эффективность работы модели, особенно в средах с ограниченными ресурсами.

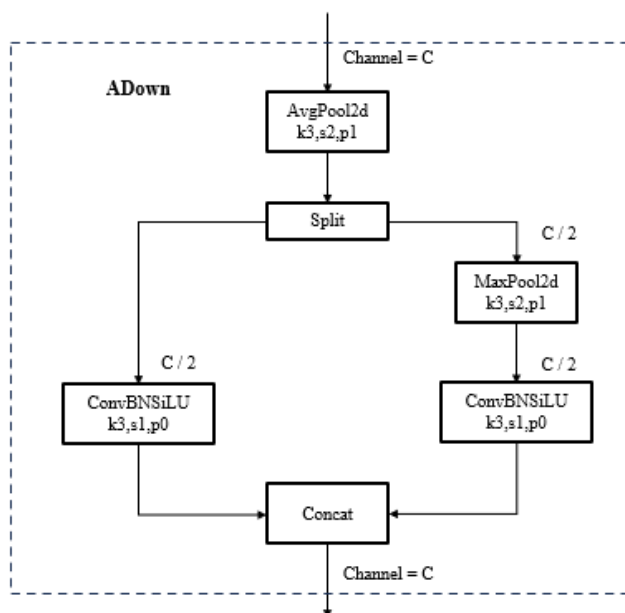


Рис. 1. – Структура модуля ADown

Конструкция модуля ADown направлена на сохранение как можно большего количества информации об изображении, чтобы модель могла более точно обнаруживать цели и в то же время обладала определенной обучаемостью, что означает, что ее можно настраивать в соответствии с

различными сценариями обработки данных для оптимизации ее производительности.

3. Модуль C2f

Модуль C2f (рис. 2а) — это модуль извлечения признаков объектов в YOLOv8, который в основном используется для более эффективного представления признаков объектов и объединения признаков объектов. Модуль C2f наследует идею структуры C3 (рис. 2б), и благодаря более эффективным стратегиям объединения признаков объектов улучшает способность сети к выражению объектов при сохранении вычислительной эффективности.

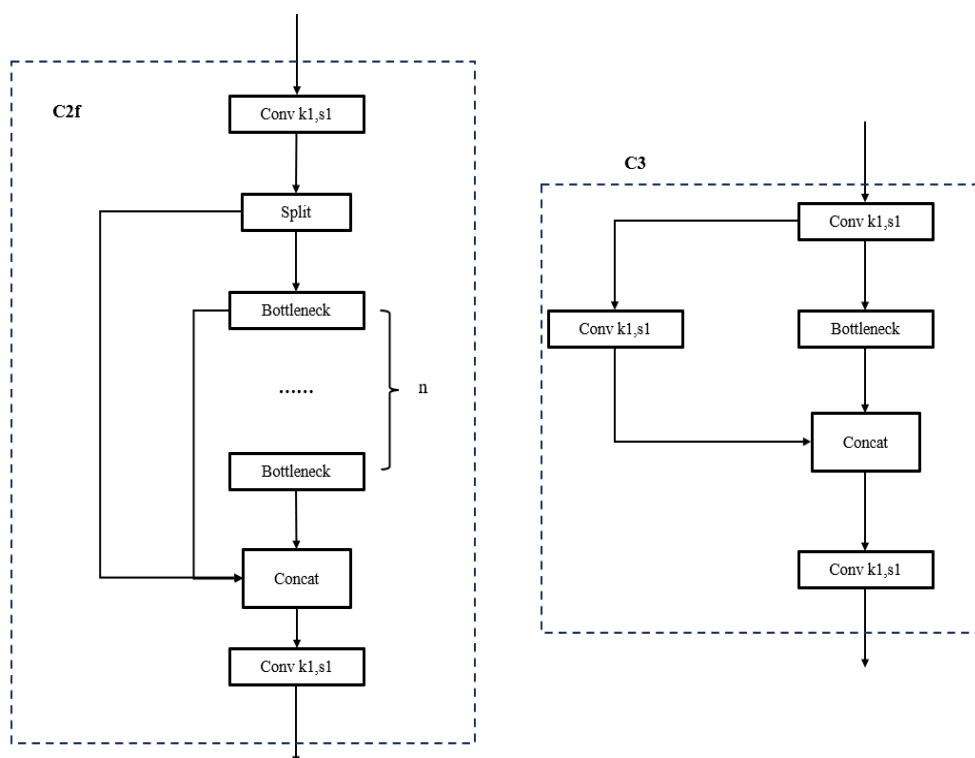


Рис. 2. – слева) Структура C2f, справа) Структура C3

Модуль C2f использует методы объединения признаков объектов, такие как объединение каналов (concat), в сочетании с методом облегченного внимания ELAN (Efficient Lightweight Attention Network, «Эффективная легкая сеть внимания» [9]) для улучшения передачи информации о

градиентах, которое улучшает скорость сходимости и эффект сходимости модели, а также повышает эффективность обучения.

По сравнению с модулем C3, модуль C2f имеет ряд существенных преимуществ в задачах обнаружения объектов. Облегченная конструкция модуля C2f особенно подходит для сценариев, где используются мобильные устройства или вычислительные ресурсы существенно ограничены.

4. Результаты экспериментов

Для проведения экспериментов использовался набор данных COCO-128[10] и предварительно обученная сеть YOLOv5s.pt для ускорения процесса обучения. При обучении использовались следующие гиперпараметры:

- размер изображения – 640x480;
- метод обучения – SGD
- размер пакета – 32;
- начальная скорость обучения – 0,01;
- затухание – 0,001;
- импульс 0,9;
- количество итераций – 400.

Остальные параметры были взяты по умолчанию (порог IoU – 0,45, порог доверия – 0,25).

Обучение сети выполнялось на графическом процессоре NVIDIA GeForce GTX 1050Ti с 4 ГБ памяти.

В таблице 1 приведены стандартные метрики: Сложность вычислений, точность (precision), полнота (recall), mAP@.5 (средняя точность AP при IOU не менее 0,5) и mAP@.5:.95 (порог IOU в диапазоне от 0,5 до 0,95 с шагом 0,5 среднего значения). «ADOWN backbone» значит модуль ADOWN только используется в основной сети.

Таблица № 1

Результаты экспериментов

Модель сети	Сложность вычислений	Точность	Полнота	map0.5	map0.5-0.95
C3	16.6 GFLOPs	0.794	0.724	0.81	0.5
C3+ADOWN backbone	13.9 GFLOPs	0.849	0.748	0.849	0.519
C3+ADOWN	13.2 GFLOPs	0.837	0.789	0.856	0.528
C2f	16.1 GFLOPs	0.851	0.808	0.88	0.558
C2f+ADOWN backbone	13.4 GFLOPs	0.827	0.853	0.89	0.58
C2f+ADOWN	12.7 GFLOPs	0.822	0.861	0.895	0.587

Проведем оценку эффективности предложенной модели на тестовом наборе данных (рис. 3).



Рис. 3. – Пример сравнения работы алгоритмов: слева) исходная сеть YOLO v5, справа) предложенная сеть с модулями C2f+ADOWN

Для качественного сравнения результатов был использован набор дорожных видео, снятых в горных районах Швейцарии. В экспериментах сравнивалась исходная сеть YOLO v5 и предложенная сеть с модулями C2f+ADOWN.

На паре кадров представлена дорожная сцена. Здесь предложенная сеть смогла обнаружить большинство объектов, которые исходная сеть не смогла

обнаружить. При этом, как видно из табл. 1 сложность вычислений предложенной модели меньше.

В целом, предложенная сеть практически везде превосходит оригинальную сеть YOLOv5 при детектировании объектов в дорожных сценах. Это говорит об эффективности предложенной модели, и возможности ее использования в практических задачах.

Выводы

В работе предложена усовершенствованная модель сети на основе YOLOv5, в которой добавлен модуль Adown в качестве модуля пулирования с целью снижения сложности модели. Добавлен модуль C2f в качестве модуля извлечения признаков, чтобы повысить точность за счет объединения признаков. Проведенные эксперименты показывают, что предложенная модель превосходит исходную YOLOv5 для рассматриваемого типа сцен, и может использоваться в задачах распознавания дорожных сцен.

Литература (References)

1. Girshick R, Donahue J, Darrell T, et al. Rich feature hierarchies for accurate object detection and semantic segmentation // Conference on Computer Vision and Pattern Recognition. IEEE, 2013, Pp. 580-587.
2. Girshick R. Fast R-CNN // International Conference on Computer Vision. IEEE, 2015, Pp. 1440-1448.
3. Ren S, He K, Girshick R, et al. Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks // Pattern Analysis and Machine Intelligence. IEEE Transactions. 2017. V. 39. № 6. Pp. 1137–1149.
4. Redmon J, Divvala S, Girshick R, et al. You Only Look Once: Unified, Real-Time Object Detection // Computer Vision and Pattern Recognition. IEEE, 2016. Pp. 779–788.

5. Redmon J, Farhadi A. YOLO9000: Better, Faster, Stronger // Computer Vision and Pattern Recognition. IEEE, 2017. Pp. 7263-7271.
6. Redmon J, Farhadi A. YOLOv3: an incremental improvement // arXiv, 2018, URL: arxiv.org/abs/1804.02767.
7. Bochkovskiy A, Wang C, Liao M. YOLOv4: Optimal Speed and Accuracy of Object Detection // arXiv, 2020, URL: arxiv.org/abs/2004.10934.
8. Wang C, Yeh I, Liao M. YOLOv9: Learning What You Want to Learn Using Programmable Gradient Information // arXiv, 2024, URL: arxiv.org/abs/2402.13616.
9. Zhang X. Efficient Long-Range Attention Network for Image Super-resolution // European Conference on Computer Vision. Springer Nature Switzerland, 2022. Pp. 649-667.
10. Lin T, Maire M, Belongie S. Microsoft COCO: common objects in context // European Conference on Computer Vision. Springer Nature Switzerland, 2014. Pp. 740–755.

Дата поступления: 5.06.2025

Дата публикации: 25.07.2025