УДК 543.257.2

УЛУЧШЕННАЯ МОДЕЛЬ YOLOV8 ДЛЯ ОБНАРУЖЕНИЯ МЕЛКИХ ОБЪЕКТОВ НА ИЗОБРАЖЕНИЯХ

Ли Чжиюань[1)], С. В. Абламейко[2)]

[1)]Белорусский государственный университет,
пр. Независимости, 4, 220030, г. Минск, Беларусь, lil877422480@gmail.com
[2)]Белорусский государственный университет,
пр. Независимости, 4, 220030, г. Минск, Беларусь, ablameyko@bsu.by

Обнаружение мелких объектов – ключевая проблема в области компьютерного зрения, имеющая важное применение в сферах обороны, мониторинга дорожного движения и промышленной автоматизации. В данной работе на основе модели YOLOv8n внедряется механизм согласованного внимания Coordinate Attention (CA) и заголовок SEResNeXtBottleneck. для улучшения способности модели обнаруживать малоразмерные объекты. Улучшенная модель была обучена и оценена на наборе данных DOTAv1,\ и результаты показали, что точность обнаружения мелких целей значительно повысилась, а показатели mAP50 и mAP50-95 значительно улучшились.

Ключевые слова: компьютерное зрение; обнаружение мелких объектов; YOLOv8; механизм канального внимания.

1. Введение

В современную эпоху быстрого технологического развития обнаружение мелких объектов играет все более важную роль в области компьютерного зрения. В области компьютерного зрения под мелкими объектами обычно понимаются объекты, занимающие небольшое количество пикселей на изображении. Согласно определению Международного общества оптической инженерии, мелкий объект – это объект с площадью изображения менее 80 пикселей на изображении размером 256 × 256 пикселей. То есть, если размер объекта составляет менее 0,12 % от исходного изображения, его можно считать мелким объектом. Сложность обнаружения мелких объектов заключается в малой площади охвата мелких объектов, меньшем охвате информации о признаках и слабой способности к выражению.

За основу был взят YOLOv8n [1], представляющий собой последнее достижение для обнаружения объектов. YOLOv8n внедряет новейшие архитектуры и методы оптимизации для повышения скорости и точности обнаружения, особенно для мелких объектов. В архитектуре YOLOv8n используется структура C2f для улучшения объединения признаков и модуль SPPF из YOLOv5 для расширения рецептивного поля. Головная часть разделяет головки классификации и обнаружения, а в качестве потерь классификации используется BCE Loss, применяя метод сопоставления по назначению для улучшения сопоставления образцов.

Однако, несмотря на значительный прогресс, достигнутый YOLOv8n в обнаружении объектов, модель все еще сталкивается с некоторыми проблемами при обнаружении мелких объектов. Эти проблемы в основном связаны с ограниченным охватом информации о признаках мелких объектов и небольшой зоной покрытия, что затрудняет представление и обнаружение таких объектов. Кроме того, использование фреймворка C2f и модуля SPPF, хотя и расширяет сенсорное поле, не всегда компенсирует эти недостатки, особенно в сложных сценах с большим количеством мелких объектов.

Мы предлагаем модель YOLOv8n-CA-OBB_SEResNeXtBottleneck, которая улучшает YOLOv8n и показывает, как эти улучшения могут быть использованы для значительного повышения эффективности обнаружения мелких объектов на изображениях.

2. Улучшение модели YOLOv8n

Для улучшения обнаружения и распознавания мелких объектов на спутниковых и аэрофотоснимках в данной работе предлагается усовершенствованная модель обнаружения YOLOv8n-CA-OBB_SEResNeXtBottleneck, основанная на модели YOLOv8n, которая включает механизм координатного внимания Attention mechanism Coordinate (CA) [2] в магистрали YOLOv8n и заменяет заголовок Detect в YOLO-v8n на заголовок обнаружения SEResNeXtBottleneck, чтобы повысить эффективность обнаружения и точность модели. Улучшенная модель показана на рис. 2. Модель нарисована с помощью инструментария визуализации TensorBoard.

Традиционные механизмы внимания, такие как сжимающее и побуждающее внимание SE (Squeeze-and-Excitation block) [3] и модуль конволюционного блочного внимания CBAM (Convolutional Block Attention Module) [4], имеют много недостатков. SE внимание фокусируется только на построении взаимозависимостей между каналами и игнорирует пространственные особенности. CBAM вводит крупномасштабное конво-люционное ядро. В CBAM для извлечения пространственных признаков используется крупномасштабное сверточное ядро, но проблема дальних зависимостей игнорируется. Хотя другие модули внимания без этой проблемы показывают хорошие результаты, количество параметров слишком велико для развертывания приложений. Модуль CA (Coordinate Attention) [2] – это новый модуль внимания, предложенный для канального внимания. Механизм координатного внимания не только захватывает кросс-канальную информацию, но и учитывает восприятие ориентации и местоположения, что помогает модели более точно находить и идентифицировать интересующую цель. Механизм внимания CA, названный C2f_CA в структуре нашей модели, используется для замены модуля C2f в оригинальной модели YOLOv8n. Как показано на рис. 1 C2f_CA[*].

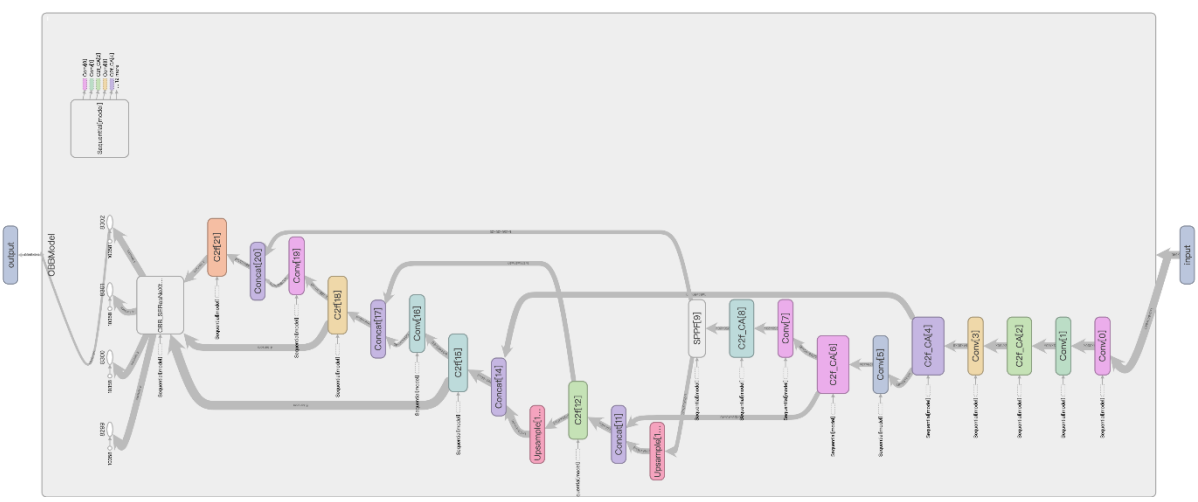


Рис. 1. Улучшенная структура сети YOLOv8n-CA-OBB_SEResNeXtBottleneck

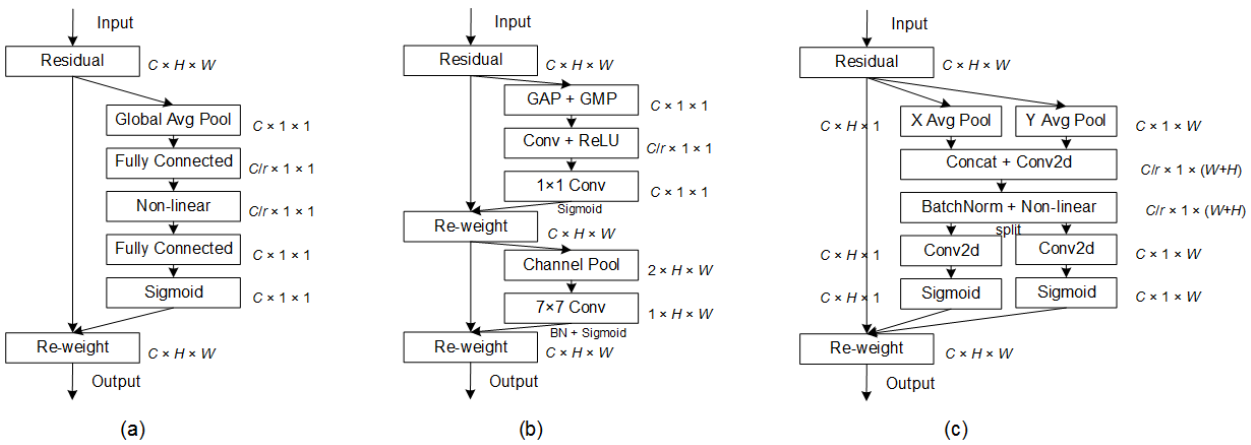
Сравним три модели внимания (рис.2):

(a) Squeeze-and-Excitation block (SE): Этот модуль сначала применяет глобальное усреднение по пространству (Global Avg Pool), чтобы сжать пространственные измерения и получить 1D вектор признаков на канал. Затем два полностью связанных слоя (Fully Connected) создают отношения между каналами, используя нелинейность для активации и сигмоидную функцию для создания весов, которые используются для повторного взвешивания исходных карт признаков. Этот подход улучшает представление канала, но, как вы упомянули, не учитывает пространственную информацию.

(b) Convolutional Block Attention Module (CBAM): Этот модуль развивает идею SE, вводя отдельные механизмы внимания для канального и пространственного измерений. Вначале применяется глобальное усреднение и максимальное пулинг, за которым следует свертка с

ReLU активацией для взвешивания по каналам. Для пространственного внимания используется пулинг по каналам и большое сверточное ядро (7x7), что позволяет модели захватывать пространственные зависимости, но не решает проблему дальних зависимостей.

(c) Coordinate Attention block (CA): Это новый подход, который предлагает решение упомянутых недостатков, сосредотачиваясь как на канальном, так и на пространственном внимании. Вместо одного глобального усреднения используются два отдельных усреднения по разным осям (X и Y), что помогает улавливать информацию о местоположении и ориентации. После конкатенации и свертки эти признаки объединяются, затем применяется BatchNorm и нелинейность. Пространственные веса получаются раздельно для каждой оси через Conv2d и сигмоидные функции, что обеспечивает более точное пространственное внимание с учетом глобального контекста и ориентации объектов.



(a) Squeeze-and-Excitation block (b) CBAM (c) Coordinate attention block

Рис. 2. Сравнение моделей внимания

Заголовок Detect в YOLOv8n играет роль обнаружения и распознавания изображений, но скорость распознавания мелких целей очень низкая в сложных условиях, поэтому мы предлагаем SEResNeXtBottleneck для YOLOv8n, который объединяет особенности SENet и ResNeXt, и интегрирует модуль SE в структуру узкого места ResNeXt. Этот подход использует особенности заголовка ResNeXt для увеличения широты и выразительности сети и улучшает понимание узкого места SEResNeXtBottleneck через канальный механизм модуля SE. Сеть демонстрирует отличную производительность, особенно при работе с признаками, обнаружении объектов и тонком распознавании, которые требуют умного разграничения с другими задачами.На рис. 3 представлена схема специфической структуры заголовка SEResNeXtBottleneck, которую мы построили с помощью инструментария визуализации TensorBoard.

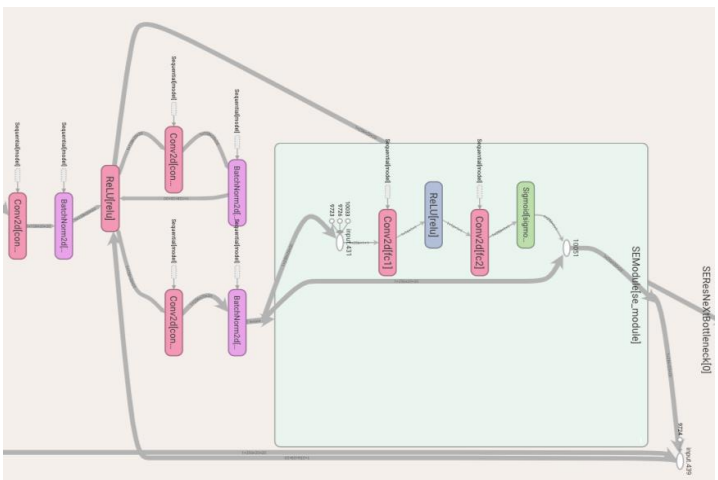


Рис. 3. Структура головки SEResNeXtBottleneck в модели

3. Эксперименты и результаты

3.1 Наборы данных

DOTA (Dataset for Object Detection in Aerial images) v1.[5]- это крупномасштабный набор данных для обнаружения объектов на аэроснимках. Набор данных DOTAv1 содержит более 2 800 изображений. Он содержит аннотационную информацию для более чем 188 000 экземпляров объектов. Существует 15 различных категорий объектов, включая самолеты, корабли, резервуары, бейсбольные поля, теннисные корты, баскетбольные площадки, спортивные площадки, гавани, мосты, крупные транспортные средства (например, грузовики и автобусы), вертолеты, плавательные бассейны, футбольные стадионы, круговые перекрестки и небольшие транспортные средства. На рис. 4 представлены категории тегов, включенных в набор данных. Набор данных предназначен для решения ряда задач по обнаружению объектов на аэрофотоснимках, включая, помимо прочего, обнаружение мелких объектов, обнаружение плотных объектов и крупномасштабных вариаций.

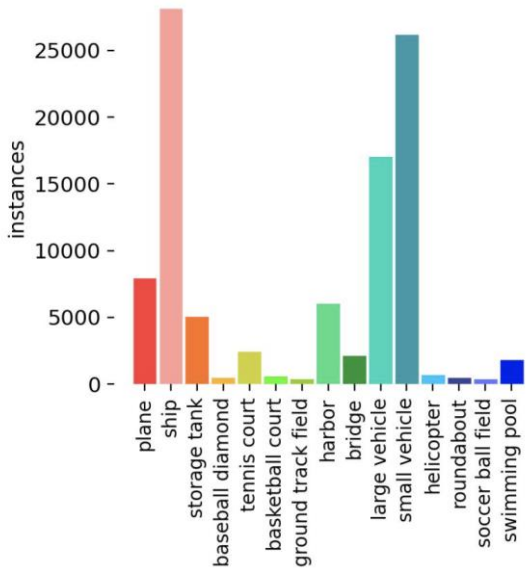


Рис. 4. Категории изображений для DOTAv1

3.2 Экспериментальное оборудование и показатели оценки

Эксперименты проводились на облачной вычислительной платформе AutoDL с конфигурацией среды: Python 3.10 (ubuntu22.04), PyTorch 2.1.0, Cuda 12.1, GPU RTX 4090(24GB) * 2, CPU 32 vCPU AMD EPYC 9654 96-Core Processor, RAM. Для обучения входное изображение имеет размер 640×640, а модель обучается с использованием SGD в качестве оптимизирующей функции. Период обучения модели (эпоха) составляет 300, размер партии – 18, начальная скорость обучения – 0,01. В этом эксперименте используется тот же алгоритм улучшения данных, что и в оригинальном алгоритме YOLOv8n.

В качестве оценочных показателей в данной работе используются F1 score, средняя точность (mAP), количество параметров (Params), гига операций с плавающей запятой в секунду (GFLOPs) и кадров в секунду (FPS). Корректность модели измеряется с помощью показателей корректности и полноты в качестве базовых показателей, а F1 score и mAP, вычисленные из показателей корректности и полноты – в качестве итоговых показателей оценки.

3.4 Оценка результатов

Результаты экспериментов показывают, что наша модель превосходит оригинальную модель YOLOv8n как на mAP50%, так и на mAP50-95%. В частности, модель улучшает свои показатели с 41,0 до 42,7 % на mAP50 % и с 24,6 до 25,5 % на mAP50-95 %, указывая на то, что добавленные механизм внимания CA и заголовок SEResNeXtBottleneck могут значительно улучшить производительность модели.

В следующей таблице представлены результаты прогнозирования на тестовом наборе DOTAv1-val после 300 эпох в каждом модельном фреймворке экспериментальной среды.

Сравнение результатов различных моделей

Модель	mAP50(%)	mAP50-95(%)
Yolov8n	41,0	24,6
Yolov8n-obb	42,3	24,9
YOLOv8n-SEResNeXtBottleneck head	42,2	24,6
YOLOv8n-CA-OBB	42,1	25,1
YOLOv8n-CA-OBB_SEResNeXt	42,7	25,5

На рис. 5 показаны результаты обнаружения, слева – оригинальное изображение из набора данных DOTAv1, среднее – результат обнаружения с помощью YOLOv8n, крайнее правое – результат обнаружения модели YOLOv8n-CA-OBB_SEResNeXtBottleneck. Согласно сравнению результатов, наша модель не только обнаружила больше мелких целей при обнаружении мелкой цели, но и точность обнаружения крупной цели была улучшена.

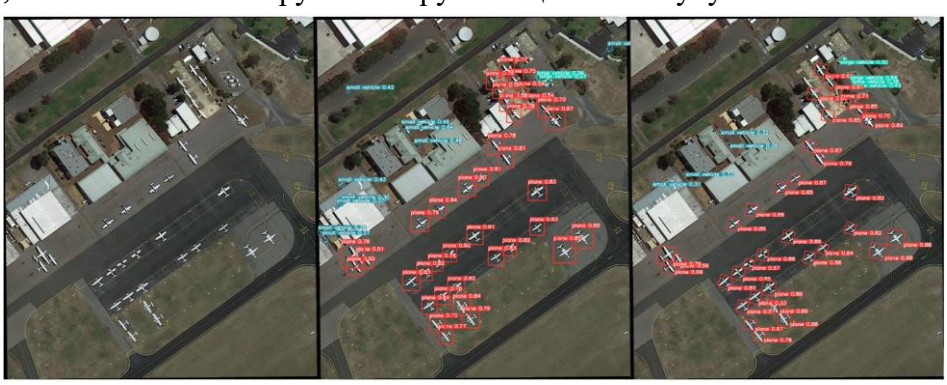


Рис. 5. Результаты исходного изображения, YOLOv8n и нашей модели.

4. Заключение

Предложенная модель позволила повысить точность и производительность обнаружения мелких объектов, включив механизм внимания CA в модель YOLOv8n и используя заголовок обнаружения SEResNeXtBottleneck вместо заголовка YOLOv8n. Благодаря экспериментальной проверке на наборе данных DOTAv1 наша улучшенная модель демонстрирует более точную способность к обнаружению мелких объектов, а результаты эксперимента показывают, что благодаря добавлению механизма внимания CA модель способна более эффективно фокусироваться на ключевых областях изображения, что повышает точность обнаружения.

Библиографические ссылки

1. Jocher G., Chaurasia A., Qiu J. “Ultralytics YOLOv8.” 2023. [Online]. Available: https://github.com/ultralytics/ultralytics.

2. Hou Q., Zhou D., Feng J. Coordinate attention for efficient mobile network design // Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2021: 13713-13722.

3. Hu J., Shen L., Sun G. Squeeze-and-excitation networks[C] // Proceedings of the IEEE conference on computer vision and pattern recognition. 2018: 7132-7141.

4. Aggregated residual transformations for deep neural networks[C] / S. Xie [et al.] // Proceedings of the IEEE conference on computer vision and pattern recognition. 2017: 1492-1500.

5. URL: https://captain-whu.github.io/DOTA/index.html.