Научная статья
Original article
УДК 004.89

ИССЛЕДОВАНИЕ МОДЕЛЕЙ СВЕРТОЧНЫХ НЕЙРОННЫХ СЕТЕЙ YOLOV3 И RETINANET ДЛЯ ЗАДАЧИ ДЕТЕКТИРОВАНИЯ ЛИЦА ЧЕЛОВЕКА НА ИЗОБРАЖЕНИИ

RESEARCH OF MODELS OF CONVOLUTIONAL NEURAL NETWORKS YOLOV3 AND RETINANET FOR THE TASK OF DETECTING A HUMAN FACE IN THE IMAGE

Болховитина Елена Игоревна, студентка 2 курса магистратуры по направлению подготовки «Информатика и вычислительная техника» (Искусственный интеллект), ФГБОУ ВО «Иркутский национальный исследовательский технический университет» (664074 Россия, г. Иркутск, ул. Лермонтова, д. 83)

Bolkhovitina Elena Igorevna, 2nd year Master's student in the field of study «Informatics and Computer Engineering» (Artificial Intelligence), Federal State Budget Educational Institution of Higher Education «Irkutsk National Research Technical University» (83 Lermontova st., Irkutsk, 664074 Russia)

Аннотация. В статье представлено исследование двух популярных моделей детектирования объектов на изображении на основе сверточных нейронных сетей для реализации задачи детектирования лиц на изображении – YOLOv3 и RetinaNet. Каждая модель была обучена на собственном наборе данных изображений лиц, состоящим из 200 различных изображений людей.

Осуществлено сравнение основных параметров производительности моделей, таких как потери при обучении, проверка точности распознавания (mAP) и среднее время вывода.

В ходе анализа двух моделей сверточных нейронных сетей архитектура RetinaNet показала конкурентоспособные результаты, превысив 90% точности. Архитектура YOLOv3 несмотря на то, что имеет наилучшие показатели в скорости обработки изображений и подходящим для решения задач в реальном времени, при обнаружении лица на изображении имеет худшие результаты по сравнению с RetinaNet.

Annotation. The article presents a study of two popular models for detecting objects in an image based on convolutional neural networks for implementing the task of detecting faces in an image - YOLOv3 and RetinaNet. Each model was trained on its own face image dataset of 200 different face images. The main performance parameters of the models are compared, such as training loss, recognition accuracy test (mAP), and average inference time.

In the course of analysis of two models of convolutional neural networks, the RetinaNet architecture showed competitive results, exceeding 90% accuracy. The YOLOv3 architecture, despite the fact that it has the best performance in image processing speed and is suitable for solving problems in real time, when detecting a face in an image, has worse results compared to RetinaNet.

Ключевые слова: сверточная нейронная сеть, one-stage детекторы, RetinaNet, YOLO, детектирование объектов, точность распознавания.

Keywords: convolutional neural network, one-stage detectors, RetinaNet, YOLO, object detection, recognition accuracy.

Введение

В последние годы машинное обучение (machine learning) получило широкое распространение. Алгоритмы машинного обучения способны решать задачи классификации, прогнозирования, кластеризации и регрессии в различных сферах жизни человека. Исследования, посвященные

идентификации и распознаванию лиц становятся сегодня особенно значимыми в связи с развитием нового уровня систем безопасности на основе распознавания лиц. Появление все более сложных алгоритмов позволяет надеяться на дальнейшее развитие и увеличение эффективности архитектур сверточных нейронных сетей для реализации задачи распознавания лиц.

Точность модели является основным показателем, определяющим может ли эта модель помочь для решения поставленной задачи. Кроме того, важна эффективность модели. Например, если работа модели занимает слишком много времени, она уже не будет пригодна для использования в бизнес-среде.

Современные подходы с использованием методов глубокого обучения для целевой классификации и регрессии можно разделить на две категории. Одна из них представляет собой двухэтапный алгоритм, представленный такими архитектурами, как R-CNN, Fast R-CNN и Faster R-CNN. Этот тип алгоритма обычно выполняется в два шага. Первый – использует выборочный поиск или «Сеть предложения области (RPN)» для создания возможной цели в области, а затем реализуется классификация и регрессия по предложенной области. Этот метод имеет высокую точность, но также ограничивает скорость обнаружения.

Другой алгоритм one-stage подход – одноэтапный алгоритм детектирования объектов на изображении. Представителями данного алгоритма являются архитектуры сверточных нейронных сетей RetinaNet и YOLO. Одноэтапные алгоритмы используют единую сеть для прямого прогнозирования ограничивающих рамок объекта и оценки вероятности класса из изображений. Скорость обнаружения повышается за счет отказа от использования RPN. Сверточная нейронная сеть применяется ко всему изображению сразу и прогнозирует вероятность существования объекта внутри привязки на основе пересечения по объединению (IoU) между привязкой и местоположением объекта.

One-stage детекторы

RetinaNet – один из типичных одноэтапных алгоритмов, который базируется на архитектурах сетей ResNet и Feature Pyramid Net (FPN). На основе структуры FPN добавляются связи «путь сверху вниз» и «горизонтальный». Каждый уровень FPN подключен к полностью сверточной сети, которая включают две независимые подсети, которые используются для классификации и регрессии. Главное нововведение в архитектуре RetinaNet является добавление функции кросс-энтропийных потерь (Focal Loss) к подсети классификации. Поскольку дисбаланс числа положительных и отрицательных образцов при целевом обнаружении одноэтапного алгоритма будет влиять на потери при обучении, Focal Loss присваивает разные веса объектам, что эффективно решает проблему дисбаланса классов в модели распознавания лиц. Исследование [3], проведенное в 2017 г., показало, что RetinaNet может достичь скорости обнаружения, аналогичной одноэтапным алгоритмам, а точность обнаружения намного превосходящую многие двухэтапные алгоритмы.

Архитектура сверточной нейронной сети YOLO (You Only Look Once) предлагает новую идею для распознавания объектов на изображении путем преобразования задачи в проблему регрессии [4]. Для непосредственного завершения регрессии обнаружения объекта с прогнозированием положения ограничивающей рамки и класса рамки объекта, весь фреймворк должен использовать только относительно простую структуру CNN. Структура магистральной сети YOLO v3 не имеет объединения и полносвязанных слоев, и сверточное преобразование изображения достигается за счет изменения размера шага сверточного ядра. YOLO v3 использует Darknet-53 в качестве архитектуры сети, что делает структуру сети более глубокой и эффективной для извлечения функций, что подтверждается улучшенной точностью по сравнению с YOLO v1 и YOLO v2. Даркнет-53 широко использует остаточную структуру ResNet, которая помогает избежать

проблем исчезновения градиента, даже когда структура сети достаточно глубокая.

RetinaNet использует магистраль ResNet [9] и чтобы проверить влияние магистрали на результаты обнаружения, обучим RetinaNet с различными вариантами ResNet такими как ResNet-34 и ResNeXt-101.

Для того чтобы произвести сравнительный анализ архитектур сверточных нейронных сетей RetinaNet и YOLO будем изменять магистрали или количество якорей, чтобы определить влияние этих изменений на полученные результаты. Имея такие данные, можно установить в будущем минимальную архитектуру необходимую для использования и обладающую приемлемой эффективностью.

Набор данных

Наш набор данных состоит из 200 фотографий с различными изображениями людей. Изображения случайным образом разделены на 140 изображений для обучения, 30 – для проверки (валидационная выборка) и 30 – для тестирования. Помимо слоев пакетной нормализации, к входным изображениям не применялась регуляризация или нормализация до нулевого среднего и единичной дисперсии. Помимо этого в наборе данных присутствуют файлы формата txt для каждого изображения, где изображены лица людей, которые содержат координаты ограничивающих рамок для каждого лица на изображении.

Валидация и метрики

Средняя точность (AP) – это оценка площади под кривой графика «Precision vs Recall». AP получается путем интерполяции точности на каждом уровне. Затем обученные модели оценим с помощью тестового набора с использованием средней точности (mAP). Кроме того, будем считать истинно положительными обнаружения, ограничивающие прямоугольники которых имеют IoU с истинностью основания больше 0,5.

Результаты тестирования моделей сверточных нейронных сетей

Тесты проводились на пользовательском компьютере с графическим процессором NVIDIA GeForce RTX 2060 SUPER. В таблице 1 представлены результаты обучения моделей сверточных нейронных сетей, их потери при обучении, проверка mAP и среднее время вывода. Так для архитектуры RetinaNet потери никогда не превышают 1, тогда как потери YOLO начинаются со значений выше 1. Лучшее сравнение архитектур можно провести, используя проверочный набор mAP.

Что касается времени обработки изображения, то увеличение глубины магистрали в RetinaNet влияет на увеличение времени обработки до 50 %.

Таблица 1 – Результаты обучения моделей CNN

Модель			Функция потерь Loss	Точность предсказания mAP (%)	Среднее время обработки изображения
Detector	Backbone	Anchors			
RetinaNet	ResNet-34	9	0.595	90.4%	89 ms
	ResNeXt-101	9	0.401	92.8%	121 ms
YOLOv3	Darknet-53	6	9.311	42.6%	76 ms
		9	7.360	59.8%	76 ms

Архитектура RetinaNet показала конкурентоспособные результаты по отношению к YOLOv3. Самая «легкая» модель RetinaNet превзошла результаты тестирования YOLOv3 с большим преимуществом. RetinaNet с архитектурой ResNet-34 всего на 7% медленнее, чем YOLOv3, но превосходит лучшее значение mAP сети YOLOv3 на 30,6 процентов.

Выводы

В работе было осуществлено обучение одноэтапных детекторов моделей сверточных нейронных сетей и анализ некоторых параметров работы таких моделей как RetinaNet и YOLOv3 для решения задачи обнаружения лиц на изображении. Результаты показали, что модели CNN для обнаружения объектов могут обнаруживать лица на изображении без переобучения.

Модель RetinaNet зарекомендовала себя как надежный одноэтапный детектор, способный улучшать mAP при увеличении глубины магистрали.

RetinaNet имеет высокий mAP (92.8%), но средняя скорость обработки изображения недостаточно высока для реального времени. Хотя YOLO v3 имеет меньший mAP (59.8%), система способна значительно улучшить скорость распознавания и достичь эффективности в режиме реального времени. В местах большого скопления объектов распознавание лиц требует не только достаточно высокого mAP, но и высокой скорости распознавания. Для задачи, где необходимо наиболее высокое значение точности распознавания, например, в случае задачи идентификации лица человека, RetinaNet может быть более эффективной.

Литература

1. Liu Jiaye, et al. Community transmission of severe acute respiratory syndrome coronavirus 2, shenzhen, China. Emerg Infect Dis 2020; 26(6):1320–3. https://doi.org/10.3201/eid2606.200239. 2020 – (дата обращения: 22.05.22).
2. Thaper R. Transmission of SARS-CoV-2 through the air. Curr Med Res Pract 2020; 10(4):196–7. https://doi.org/10.1016/j.cmrp.2020.07.005. – (дата обращения: 22.05.22).
3. Humphreys J. The importance of wearing masks in curtailing the COVID-19 pandemic. J Fam Med Prim Care Jun. 2020; 9(6):2606–7. https://doi.org/10.4103/ jfmpc.jfmpc_578_20. – (дата обращения: 22.05.22).
4. Esposito S, Principi N, Leung CC, Migliori GB. Universal use of face masks for success against COVID-19: evidence and implications for prevention policies. Eur Respir J Jun. 2020; 55(6). https://doi.org/10.1183/13993003.01260-2020. – (дата обращения: 22.05.22).
5. Zhao Z-Q, Zheng P, Xu S-T, Wu X. Object detection with deep learning: a review. IEEE Transactions on Neural Networks and Learning Systems Nov. 2019; 30(11): 3212–32. https://doi.org/10.1109/TNNLS.2018.2876865. – (дата обращения: 22.05.22).

6. Verma GK, Dhillon A. A handheld Gun detection using faster R-CNN deep learning. In: Proceedings of the 7th international conference on computer and communication technology, New York, NY, USA; Nov. 2017. p. 84–8. https://doi. org/10.1145/3154979.3154988. – (дата обращения: 22.05.22).
7. Fong S, Yang X-S. White learning: a white-box data fusion machine learning framework for extreme and Fast automated cancer diagnosis. IT Professional Sep. 2019; 21(5):71–7. https://doi.org/10.1109/MITP.2019.2931415. – (дата обращения: 22.05.22).
8. Wang Zhongyuan, et al. Masked face recognition dataset and application. arXiv: 2003.09093 [cs], Mar. 2020, [Online]. Available: http://arxiv.org/abs/2003.09093. – (дата обращения: 22.05.22).
9. K. He, X. Zhang, S. Ren, J. Sun. (2016) «Deep residual learning for image recognition.» 2016 Conference on Computer Vision and Pattern Recognition (CVPR): 770–778. – (дата обращения: 22.05.22).

References

1. Liu Jiaye, et al. Community transmission of severe acute respiratory syndrome coronavirus 2, shenzhen, China. Emerg Infect Dis 2020; 26(6):1320–3. https://doi.org/10.3201/eid2606.200239. 2020 – (date of the application: 22.05.22).
2. Thaper R. Transmission of SARS-CoV-2 through the air. Curr Med Res Pract 2020; 10(4):196–7. https://doi.org/10.1016/j.cmrp.2020.07.005. – (date of the application: 22.05.22).
3. Humphreys J. The importance of wearing masks in curtailing the COVID-19 pandemic. J Fam Med Prim Care Jun. 2020; 9(6):2606–7. https://doi.org/10.4103/ jfmpc.jfmpc_578_20. – (date of the application: 22.05.22).
4. Esposito S, Principi N, Leung CC, Migliori GB. Universal use of face masks for success against COVID-19: evidence and implications for prevention policies. Eur Respir J Jun. 2020; 55(6).

https://doi.org/10.1183/13993003.01260-2020. – (date of the application: 22.05.22).

5. Zhao Z-Q, Zheng P, Xu S-T, Wu X. Object detection with deep learning: a review. IEEE Transactions on Neural Networks and Learning Systems Nov. 2019; 30(11): 3212–32. https://doi.org/10.1109/TNNLS.2018.2876865. – (date of the application: 22.05.22).
6. Verma GK, Dhillon A. A handheld Gun detection using faster R-CNN deep learning. In: Proceedings of the 7th international conference on computer and communication technology, New York, NY, USA; Nov. 2017. p. 84–8. https://doi. org/10.1145/3154979.3154988. – (date of the application: 22.05.22).
7. Fong S, Yang X-S. White learning: a white-box data fusion machine learning framework for extreme and Fast automated cancer diagnosis. IT Professional Sep. 2019; 21(5):71–7. https://doi.org/10.1109/MITP.2019.2931415. – (date of the application: 22.05.22).
8. Wang Zhongyuan, et al. Masked face recognition dataset and application. arXiv: 2003.09093 [cs], Mar. 2020, [Online]. Available: http://arxiv.org/abs/2003.09093. – (date of the application: 22.05.22).
9. K. He, X. Zhang, S. Ren, J. Sun. (2016) «Deep residual learning for image recognition.» 2016 Conference on Computer Vision and Pattern Recognition (CVPR): 770–778. – (date of the application: 22.05.22).

©Болховитина Е.И., 2022 Научно-образовательный журнал для студентов и преподавателей «StudNet» №6/2022.

Для цитирования: Болховитина Е.И. ВЫБОР ОПТИМИЗАТОРА ДЛЯ ОБУЧЕНИЯ СВЕРТОЧНОЙ НЕЙРОННОЙ СЕТИ: ЗАДАЧА КЛАССИФИКАЦИИ ЛЕСНЫХ ПОЖАРОВ// Научно-образовательный журнал для студентов и преподавателей «StudNet» №6/2022.