

Распознавание цифр маркировки грузового контейнера с использованием алгоритма Faster-RCNN

Р.А. Макаров

*Казанский национальный исследовательский технический университет
им. А.Н. Туполева-КАИ, Казань*

Аннотация: В данной статье описывается обучение нейронной сети для распознавания цифр номера грузового контейнера. В связи с отсутствием набора данных, содержащих грузовой контейнер, обучение производилось на наборе данных Street View House Numbers. Перед обучением модели набор данных был проанализирован и построены гистограммы, отражающие общую информацию о наборе данных. Обучение нейронной сети производилось на языке программирования Python с использованием библиотеки Tensorflow. Полученные результаты оценки точности работы модели позволили сделать вывод о возможности использования набора данных и нейронной сети для решения задачи распознавания цифр грузового контейнера.

Ключевые слова: машинное обучение, распознавание цифр, грузовые контейнеры, нейронная сеть, классификация изображений.

Введение

В последнее время активно используются смешанные структуры нейронных сетей, которые помимо непосредственного распознавания объекта на изображении также способны найти и выделить область с распознаваемым объектом [1]. Данная архитектура нейронной сети, помимо сверточных слоев и классификатора, включает в себя сеть предложения регионов, которая формирует ограничивающие прямоугольники для каждого распознаваемого объекта на изображении.

В данной статье рассматривается процесс обучения нейронной сети Faster Rcnн Resnet 50 для распознавания цифр номера грузового контейнера. Основной целью работы являются проверка и оценка возможности обучения модели на готовом наборе данных для решения задачи распознавания цифр номера грузового контейнера.

Задача распознавания образов на естественных изображениях является актуальной по сей день. Основной сложностью при обработке подобных изображений является наличие шума, а также наличие независимых

искажений объектов на изображении. В то же время, существует множество готовых решений для распознавания текста на документах, что является частным случаем распознавания символов на изображении, т.к. символы имеют четкие границы, и чаще всего имеют черный цвет на белом фоне [2].

Анализ данных для обучения модели

На данный момент в открытом доступе отсутствует набор данных, содержащий грузовые контейнеры. В связи с этим, задачу распознавания цифр грузового контейнера можно решить следующими способами:

- 1) Создание синтетических данных;
- 2) Использование стороннего набора данных.

В первом случае необходимо реализовать генерацию искусственных данных для последующего обучения.

В данной работе используется второй метод решения задачи, для этого использовался набор данных Street View House Numbers. Он содержит 33 402 тренировочных изображений и 13 068 тестовых изображений.

Пример изображений для обучения модели представлен на рисунке 1.



Рис. 1. – Пример изображений из набора данных

Тренировочные изображения используются при непосредственном обучении нейронной сети и корректировке весов модели. Поэтому данные изображения не используются при оценке точности обученной модели.

Тестовые изображения используются для получения промежуточных значений точности модели при обучении, но они не используются для корректировки весов модели [3].

Для визуализации параметров набора данных использовались гистограммы. В данном случае можно получить следующую информацию:

- 1) Распределение классов набора данных;
- 2) Распределение ширины и высоты изображений;
- 3) Распределение ширины и высоты цифр на изображении.

Данная информация позволит сделать предварительные выводы о наборе данных, а также спрогнозировать дальнейший режим работы нейронной сети [4].

На рисунке 2 представлено распределение классов тренировочного набора данных. Исходя из него, видно, что наибольшее количество имеет класс с цифрой один.

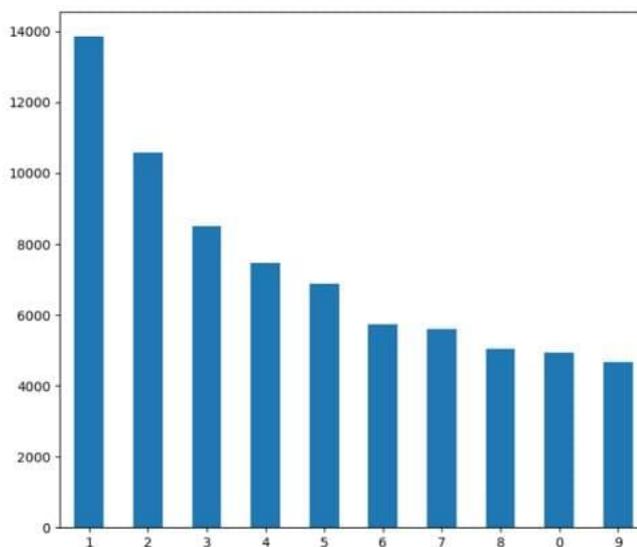


Рис. 2. – Гистограмма распределения классов набора данных

На рисунке 3 представлена гистограмма распределения ширины и высоты исходных изображений. Исходя из нее, видно, что наибольшее количество изображений имеет высоту от 15 до 60 пикселей, а ширину от 50 до 130 пикселей.

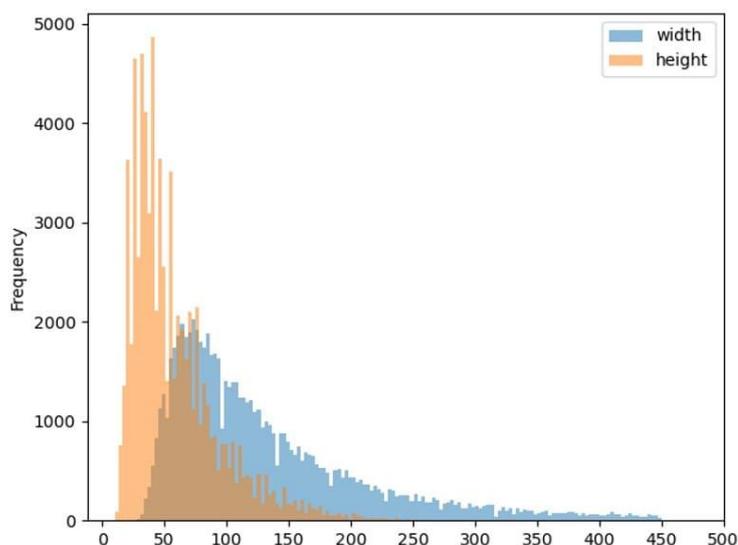


Рис. 3. – Гистограмма распределения ширины и высоты изображения

На рисунке 4 представлена гистограмма распределения размера цифр на изображении. Исходя из нее, видно, что наибольшее количество цифр имеют ширину от 7 до 20 пикселей, а высоту от 17 до 37 пикселей.

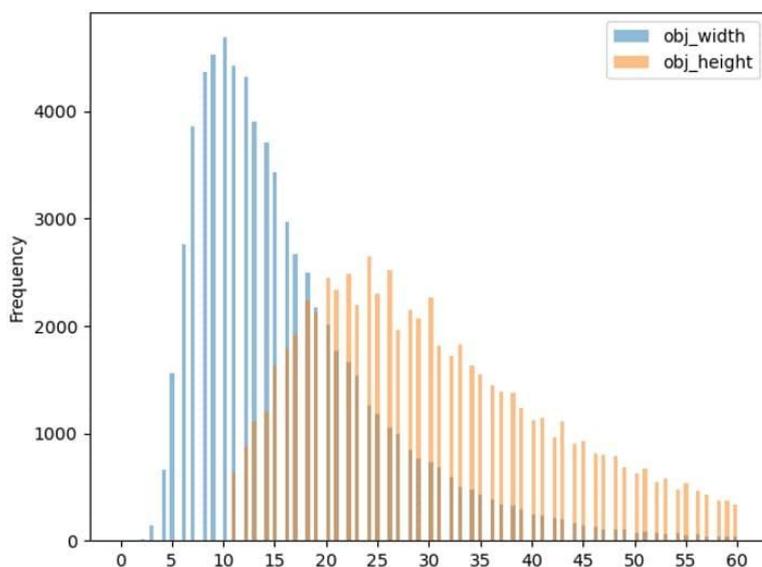


Рис. 4. – Гистограмма распределения ширины и высоты цифр

Тренировочный набор данных напрямую влияет на качество и точность работы нейронной сети, помимо непосредственно объектов распознавания, важно также соблюдать размерность входных данных [5]. В данном случае, при обучении модели на выбранном наборе данных, модель не сможет успешно обработать изображение большей или меньшей размерности.

работы сверточных слоев на выходе формируется карта признаков изображения. Базовой нейронной сетью для формирования признаков является Resnet 50. В данном случае, сеть разделена на две части: первая часть включает в себя слои от conv1 до conv4 и используется для формирования карты признаков изображения, вторая часть включает в себя слои от conv5 и используется для результирующей классификации объекта на изображении.

После слоев извлечения признаков, следует сеть предложения региона, имеющая следующий алгоритм работы: окно размером $n \times n$ скользит по карте признаков, при достижении центра объекта карты признаков, сеть формирует k прямоугольных областей разных форм и размеров [7]. В результате прохождения слоя регрессии, для каждого объекта карты признаков остается только 6 параметров: 4 координаты ограничивающего прямоугольника, вероятность нахождения объекта в ограничивающем прямоугольнике и вероятность классификации объекта.

Для обучения модели используется следующая функция потерь (1):

$$L(p, u, t^u, v) = L_{cls}(p, u) + \lambda[u \geq 1]L_{loc}(t^u, v), \quad (1)$$

где u – истинное значение распознаваемого объекта; v – регрессионное смещение координат ограничивающего прямоугольника; p – массив $(p_0, p_1 \dots p_k)$ вероятностей принадлежности объекта каждому из классов; t^u – предсказанное смещение координат ограничивающего прямоугольника.

Обучение модели

Обучение модели производилось с использованием библиотеки Tensorflow на языке программирования Python. Данная библиотека позволяет использовать графический процессор для произведения вычислений, что ускоряет процесс обучения и работы модели [8]. Для обучения использовался графический процессор Nvidia RTX 3060. Процесс обучения модели

ограничивается количеством итераций. В данном случае модель обучалась 1200 итераций, в результате чего промежуточная точность распознавания составила около 90%. График точности в процессе обучения представлен на рисунке 6.

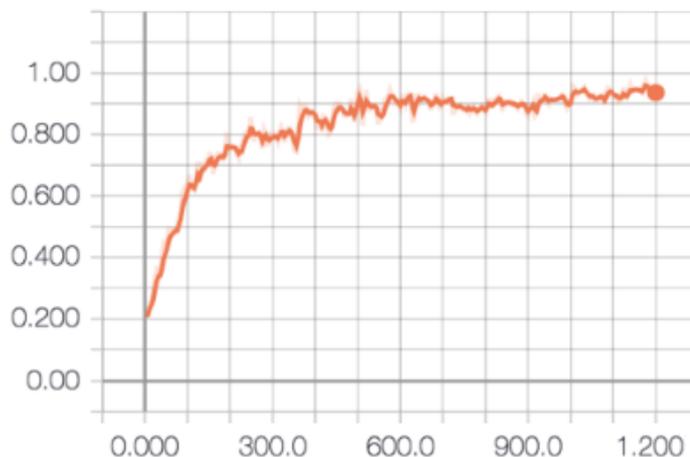


Рис. 6. – График изменения точности в процессе обучения.

Тестирование модели

Тестирование модели производилось на собранном вручную наборе данных, содержащих номера грузовых контейнеров. В данном случае удалось собрать около 150 изображений. Для оценки точности работы модели использовалось два параметра: точность локализации объекта, точность классификации объекта.

Точность локализации рассчитывалась, как отношение площади пересечения к объединенной площади предсказанной и размеченной области [9]. В данном случае, пороговым значением коэффициента пересечения было принято решение взять 0.8, если рассчитанное значение больше порогового, то считается, что объект локализован успешно.

В результате оценки работы модели, точность локализации на тестовом наборе данных составила 82%, точность распознавания на тестовых данных составила 86%.

Пример работы модели представлен на рисунке 7.



Рис. 7. – Пример работы модели по распознаванию цифр.

Исходя из графика на рисунке 6 видно, что модель достигла точность около 90% примерно на 950 итерации, в данном случае при дальнейшем обучении модели (более 1200 итераций) возможно возникновение эффекта переобучения модели. Данный эффект заключается в том, что модель продолжает корректировать вес согласно тренировочной выборке, и в это же время ухудшается качество распознавания изображений, не состоящих в тренировочной выборке [10].

Заключение

В результате данного исследования, была получена модель по распознаванию цифр номера грузового контейнера. В этом случае, вместо ручного сбора и разметки данных, использовался готовый набор данных Street View House Numbers.

Основной целью работы являлась проверка возможности использования готового набора данных для решения задачи по распознаванию цифр номера грузового контейнера. В результате оценки точности работы, можно сделать вывод, что модель является работоспособной.

В качестве возможных улучшений работы модели – можно рассмотреть аугментацию исходного набора данных [11] и применение весовых моделей изображения [12-14] для распознавания границ образов и сжатия данных.



Применение указанных технологий позволит повысить качество распознавания цифр на изображении.

Литература

1. R-CNN, Fast R-CNN, Faster R-CNN, YOLO – Object Detection Algorithms. URL: towardsdatascience.com/r-cnn-fast-r-cnn-faster-r-cnn-yolo-object-detection-algorithms-36d53571365e/ (дата обращения: 03.03.2023).

2. Золотарев О.В., Юрчак В.А. Инструменты решения проблем распознавания и кластеризации данных из документов методами машинного обучения // Инженерный вестник Дона, 2023, №2. URL: ivdon.ru/ru/magazine/archive/n2y2023/8215.

3. Nigar N., Faisal H.M., Shahzad M.K. An Offline Image Auditing System for Legacy Meter Reading Systems in Developing Countries: A Machine Learning Approach // Journal of Electrical and Computer Engineering. 2022. V. 2022. URL: doi.org/10.1155/2022/4543530.

4. Николенко С.И., Кадурын А.А., Архангельская Е.В. Глубокое обучение. Погружение в мир нейронных сетей. СПб.: Питер, 2018. 480 с.

5. Girshick R. Fast R-CNN // Computer Vision and Pattern Recognition, 2015. URL: arxiv.org/abs/1504.08083v2.

6. Евсина В.А., Широбокова С.Н., Жжонов В.А., Евсин В.А. Моделирование и реализация процесса распознавания знаков дорожного движения при определении ситуации на дороге с помощью искусственных нейронных сетей // Инженерный вестник Дона, 2022, №4. URL: ivdon.ru/ru/magazine/archive/n4y2022/7587.

7. Wang, Y., Wang, C., Zhang, H. Automatic Ship Detection Based on RetinaNet Using Multi-Resolution Gaofen-3 Imagery // Remote Sens, 2019, №11. DOI: doi.org/10.3390/rs11050531.

8. Ren S., He K., Girshick R. Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks // Computer Vision and Pattern Recognition, 2015. URL: arxiv.org/abs/1506.01497v3.

9. Girshick R., Donahue J., Darrell T. Rich feature hierarchies for accurate object detection and semantic segmentation // Computer Vision and Pattern Recognition, 2015. URL: arxiv.org/abs/1311.2524v5.

10. Ren, Y.; Zhu, C.; Xiao, S. Small Object Detection in Optical Remote Sensing Images via Modified Faster R-CNN // Appl. Sci., 2018, №8. DOI: doi.org/10.3390/app8050813.

11. Бережнов Н.И., Сирота А.А. Универсальный алгоритм улучшения изображений с использованием глубоких нейронных сетей. // Вестник Воронежского государственного университета. Серия: Системный анализ и информационные технологии. 2022. №2. С. 81-92.

12. Гизатуллин З.М., Ляшева С.А., Морозов О.Г., Шлеймович М.П. Метод обнаружения контуров на основе весовой модели изображения. // Компьютерная оптика. 2022. Т.44. № 3. С. 393-400.

13. Ляшева С.А., Морозов О.Г., Шлеймович М.П. Анализ весовой модели для сжатия изображений на основе вейвлет-преобразования. // Прикаспийский журнал: управление и высокие технологии. 2020. № 3 (51). С. 9-22.

14. Lyasheva S., Morozov O., Shleymovich M. Analysis of energy characteristics for issuing areas of significance when compressing images in cyber-physical systems. // Studies in Systems, Decision and Control. 2021. Т. 350. pp. 259-270.

References

1. R-CNN, Fast R-CNN, Faster R-CNN, YOLO – Object Detection Algorithms. URL: towardsdatascience.com/r-cnn-fast-r-cnn-faster-r-cnn-yolo-object-detection-algorithms-36d53571365e/ (accessed: 03.03.2023).

2. O.V. Zolotarev, V.A. Yurchak Inzhenernyj vestnik Dona, 2023, №2. URL: ivdon.ru/ru/magazine/archive/n2y2023/8215.

3. Nigar N., Faisal H.M., Shahzad M.K. Journal of Electrical and Computer Engineering. 2022. V. 2022. URL: doi.org/10.1155/2022/4543530 .

4. Nikolenko S.I., Kadurin A.A., Arhangel'skaya E.V. Glubokoe obuchenie. Pogruzhenie v mir nejronnyh setej [Deep learning. Loading into the world of neural networks]. Saint Petersburg: Piter, 2018. 480 p.

5. Girshick R. Computer Vision and Pattern Recognition, 2015. URL: arxiv.org/abs/1504.08083v2.

6. V.A. Evsina, S.N. Shirobokova, V.A. Zhzhonov, V.A. Evsin Inzhenernyj vestnik Dona, 2022, №4. URL: ivdon.ru/ru/magazine/archive/n4y2022/7587.

7. Wang, Y., Wang, C., Zhang, H. Remote Sens, 2019, №11. DOI: doi.org/10.3390/rs11050531.

8. Ren S., He K., Girshick R. Computer Vision and Pattern Recognition, 2015. URL: arxiv.org/abs/1506.01497v3.

9. Girshick R., Donahue J., Darrell T. Computer Vision and Pattern Recognition, 2015. URL: arxiv.org/abs/1311.2524v5.

10. Ren, Y.; Zhu, C.; Xiao, S. Appl. Sci., 2018, №8. DOI: doi.org/10.3390/app8050813.

11. Berezhnov N.I., Sirota A.A. Vestnik Voronezskogo gosudarstvennogo universiteta. Seria Sistemnyj analiz i informacionnye tehnologii. 2022. №2. pp. 81-92.

12. Gizatullin Z.M., Lyasheva S.A., Morozov O.G., Shleymovich M.P. Komp'uternaa optika. 2022. T.44. № 3. pp. 393-400.

13. Lyasheva S.A., Morozov O.G., Shleymovich M.P. Prikaspijskij zurnal: upravlenie i vysokie tehnologii. 2020. № 3 (51). pp. 9-22.

14. Lyasheva S., Morozov O., Shleymovich M. Studies in Systems, Decision and Control. 2021. T. 350. pp. 259-270.