Научная статья
Original article
УДК 519.673

DOI 10.55186/26584964_2022_5_6_2

ОБНАРУЖЕНИЕ ПРОСТРАНСТВЕННЫХ ОБЪЕКТОВ ДЛЯ АВТОНОМНЫХ ТРАНСПОРТНЫХ СРЕДСТВ

3D OBJECT DETECTION FOR AUTONOMOUS VEHICLES

Гришкин Валерий Михайлович, кандидат технических наук, доцент кафедры «компьютерного моделирования и многопроцессорных систем», Санкт-Петербургский государственный университет, Россия, г. Санкт-Петербург

Дягилев Михаил Сергеевич, студент 4 курс, факультет «Прикладная математика – процессов управления», Санкт-Петербургский государственный университет, Россия, г. Санкт-Петербург

Grishkin Valery Mikhailovich, candidate of technical sciences, associate professor Department of "computer modeling and multiprocessor systems", Saint Petersburg State University, Russia, St. Petersburg

Diaghilev Mikhail Sergeevich, 4th year student, Faculty of Applied Mathematics - Control Processes, St. Petersburg State University, Russia, St. Petersburg

Аннотация: Статья посвящена задаче детектирования объектов в различных дорожных ситуациях. В работе рассматривается практическое использование модели свёрточной нейронной сети класса «YOLO» различных

версий для поставленной задачи. Приминяется алгоритм тонкой настройки сети – используются готовые веса предобученной модели на стадии первоначального обучения сети. Работа алгоритма производится на собственноручной разметке данных. Более того, используются различные значения параметров модели в процессе обучения сети. Производится ряд опытов, способствующие улучшению показателя точности детектирования. В данной статье представлен сравнительный анализ экспериментов.

S u m m a r y: The article is devoted to the problem of detecting objects in various traffic situations. The paper considers the practical use of a convolutional neural network model of the YOLO class of various versions for the task. The network fine-tuning algorithm is used. Ready-made weights of the pre-trained model are implemented at the stage of initial network training. The algorithm works on handwritten data markup. Moreover, different values of model parameters are used in the process of network training. A number of experiments are being carried out to improve the detection accuracy. This article presents a comparative analysis of experiments.

Ключевые слова: Компьютерное зрение, свёрточные нейронные сети, искусственный интеллект, класс нейронных сетей YOLO, беспилотный автомобиль, детектирование объектов на изображениях.

Keywords: Computer vision, convolutional neural networks, artificial intelligence, neural networks of YOLO class, autonomous vehicles, detection system of 3D objects.

Технология беспилотных транспортных средств на данный момент не завоевала мирового признания. Несмотря на это, множество специалистов работают в этом направлении, так как результаты способны предоставить новые возможности и широкий функционал выполняемых задач.

Одной из современных целей XXI века является полномасштабный запуск беспилотных автомобилей. Крупные мировые компании, связанные с автомобильной сферой такие, как например: Yandex, Tesla Motors, Apple,

Honda, Volkswagen, Toyota, Lyft и многие другие последние года усиленно работают над оптимизацией программной части проектов. В частности, повышают качество определения объектов на дорогах. Успешные решения предприятий позволят значительно сократить финансовые расходы в автомобильной отрасли и претворить в жизнь абсолютно новую идею господства беспилотных автомобилей на дорогах всего мира.

В статье используются изображения, предоставленные Калифорнийской компанией Lyft [1]. Рассматриваемые данные содержат 192276 изображений различных дорожных сцен, полученных из снятого видеоряда автономного транспортного средства.

Для проведения экспериментов была выбрана модель свёрточной нейронной сети «YOLO» (You Only Look One). «YOLO» – новейшая система обнаружения объектов. Важнейшим преимуществом рассмотренной модели семейства нейронных сетей над другими является скорость. «YOLO» обладает рядом преимуществ: позволяет проводить классификацию, а также локализацию объектов всего лишь за один проход входного изображения через нейронную сеть. Именно поэтому она превосходит другие модели, в том числе Region-Based Convolutional Neural Network (R-CNN) [2]. Модель «YOLO» обрабатывает изображения в режиме реального времени со скоростью 45 кадров в секунду. Уменьшенная версия сети, Fast YOLO, обрабатывает 155 кадров в секунду [3].

На этапе обучения сверточная нейронная сеть «YOLO» предварительно на вход принимает подготовленные данные. В их состав входит информация о расположении объекта каждого класса на изображении. Во время обучения модель обрабатывает информацию, которая включает в себя само изображение и данные о нем – указание класса объекта и соответствующему ему границы обрамляющего прямоугольника. Он включает в себя значения: центра прямоугольника по оси абсцисс, ординате, а также ширины и высоты прямоугольника [4].

Обучающая и валидационная выборка из исходного датасета была размечена вручную. В общей сложности было размечено 1000 изображений, которые включают в себя 800 образов тренинговой выборки, и 200 проверочной выборки. В ходе работы создавались обрамляющие прямоугольники для каждого объекта сцены и указывалась вся вышеперечисленная необходимая информация. В следствии особенности исходного датасета для достижения сбалансированной выборки выбирались различные по сценам группы изображений. На основе статистики размера и положения центра объекта было принято решение увеличивать размер обрамляющего прямоугольника на 2 пикселя. Более того, для выполнения условия сбалансированности выборки данные содержат следующий ряд классов объектов на изображениях:

1. Легковой автомобиль;
2. Пешеход;
3. Велосипедист;
4. Мотоциклист;
5. Общественный транспорт;
6. Грузовой автомобиль.

Визуализация разметки изображений представлена на рисунке Рисунок *1*.



Рисунок 1 – Визуализация разметки изображений

После разметки данных происходил процесс обучения. Данный этап необходим для оптимизации параметров весов модели.

Обучение происходит за определённое количество эпох. Данный параметр является настраиваемым. Первоначальное обучение инициализирует веса по умолчанию, которые имеют значения, соответствующие нормальному распределению в некотором сегменте. Такой подход позволяет «усреднить выбросы» показателей ошибки.

Было обучено две модели: «YOLOv3», «YOLOv5». В процессе использовался подход «переноса обучения». Его смысл заключается в использовании «оптимизированных» весов готовой модели для первоначального этапа обучения рассматриваемой. В следствие того, что необходимые классы содержатся в исходной общедоступной выборки COCO [5], это позволяет применить ее готовый набор весов. В силу использования «тонкой настройки» минимальным значением для каждого класса объектов на этапе обучения является 200 изображений, в противном случае потребовалось 2000.

В ходе эксперимента изменялись такие параметры как: число тренировочных данных в одном батче, количество эпох обучения, скорость обучения и показатель нижней границы метрики IoU [6]. Последние два параметра были зафиксированы и соответствовали значениям 1e-5 и 50%. В свою очередь, сравнение моделей для одинаковых параметров приведены в следующем ряде таблиц.

В таблице 1 Сравнительный анализ результатов №1 приведены результаты точности детектирования для моделей с размером батча 16 и количеством эпох 60.

Таблица 1: Сравнительный анализ результатов №1

Модель	Класс объектов	Размер батча	Количество эпох	Время обучения сек	GFLOPs	Точность
«YOLOv3»	Л.а.	16	60	6:14:0	154.8	0.688
«YOLOv5»	Л.а.	16	60	1:02:0	15.8	0.627
«YOLOv3»	Пешеход	16	60	6:06:0	154.8	0.395

Модель	Класс объектов	Размер батча	Количество эпох	Время обучения сек	GFLOPs	Точность
«YOLOv5»	Пешеход	16	60	1:02:0	15.8	0.328
«YOLOv3»	Велосипедист	16	60	6:06:0	154.8	0.5577
«YOLOv5»	Велосипедист	16	60	1:02:0	15.8	0.576
«YOLOv3»	Мотоциклист	16	60	6:06:0	154.8	0.5
«YOLOv5»	Мотоциклист	16	60	1:02:0	15.8	0.5
«YOLOv3»	О.т.	16	60	6:06:0	154.8	0.178
«YOLOv5»	О.т.	16	60	1:02:0	15.8	0.153
«YOLOv3»	Г.а.	16	60	6:06:0	154.8	0.483
«YOLOv5»	Г.а.	16	60	1:02:0	15.8	0.461

В таблице 2 Сравнительный анализ результатов №2 приведены результаты точности детектирования для моделей с размером батча 2 и количеством эпох 30.

Таблица 2: Сравнительный анализ результатов №2

Модель	Класс объектов	Размер батча	Количество эпох	Время обучения	GFLOPs	Точность
«YOLOv3»	Л.а.	2	30	0:32:0	154.8	0.69
«YOLOv5»	Л.а.	2	30	0:22:0	15.8	0.623
«YOLOv3»	Пешеход	2	30	0:32:0	154.8	0.414
«YOLOv5»	Пешеход	2	30	0:22:0	15.8	0.318
«YOLOv3»	Велосипедист	2	30	0:32:0	154.8	0.577
«YOLOv5»	Велосипедист	2	30	0:22:0	15.8	0.48
«YOLOv3»	Мотоциклист	2	30	0:32:0	154.8	0.5
«YOLOv5»	Мотоциклист	2	30	0:22:0	15.8	0.5
«YOLOv3»	О.т.	2	30	0:32:0	154.8	0.209
«YOLOv5»	О.т.	2	30	0:22:0	15.8	0.141
«YOLOv3»	Г.а.	2	30	0:32:0	154.8	0.511
«YOLOv5»	Г.а.	2	30	0:22:0	15.8	0.454

В таблице 3 Сравнительный анализ результатов №3 приведены результаты точности детектирования для моделей с размером батча 2 и количеством эпох 100.

Таблица 3: Сравнительный анализ результатов №3

Модель	Класс объектов	Размер батча	Количество эпох	Время обучения	GFLOPs	Точность
«YOLOv3»	Л.а.	2	100	1:55:0	154.8	0.691
«YOLOv5»	Л.а.	2	100	1:12:0	15.8	0.63
«YOLOv3»	Пешеход	2	100	1:55:0	154.8	0.345
«YOLOv5»	Пешеход	2	100	1:12:0	15.8	0.278
«YOLOv3»	Велосипедист	2	100	1:55:0	154.8	0.576
«YOLOv5»	Велосипедист	2	100	1:12:0	15.8	0.577
«YOLOv3»	Мотоциклист	2	100	1:55:0	154.8	0.5
«YOLOv5»	Мотоциклист	2	100	1:12:0	15.8	0.5
«YOLOv3»	О.т.	2	100	1:55:0	154.8	0.207
«YOLOv5»	О.т.	2	100	1:12:0	15.8	0.106
«YOLOv3»	Г.а.	2	100	1:55:0	154.8	0.478
«YOLOv5»	Г.а.	2	100	1:12:0	15.8	0.424

Нетрудно заметить, что третья версия модели обучается в 1,5 раза, а в первом случае почти в 5 раз больше по времени, чем пятая. «YOLOv3» требует б**о**льших вычислительных мощностей по сравнению с «YOLOv5». Более того, низкая точность для всех классов, исключая легковые автомобили, может быть объяснена недостаточным количеством объектов в валидационной выборке (менее двухсот).

В тоже время результат 70% с небольшим разбросом детектирования легковых автомобилей для каждой модели можно считать достаточно хорошим, но требующим улучшения. Для этого класса объектов выполнено условие наличия более 200 экземпляров класса.

В качестве результата предсказания третьей версии модели «YOLO» на 30 эпохах с размером батча 2 можно рассмотреть на рисунке Рисунок *2*.



Рисунок 2 – Визуализация результатов детектирования объектов

Визуализация средней точности по всем классам для каждой апробации приведены на рисунке Рисунок *3*.

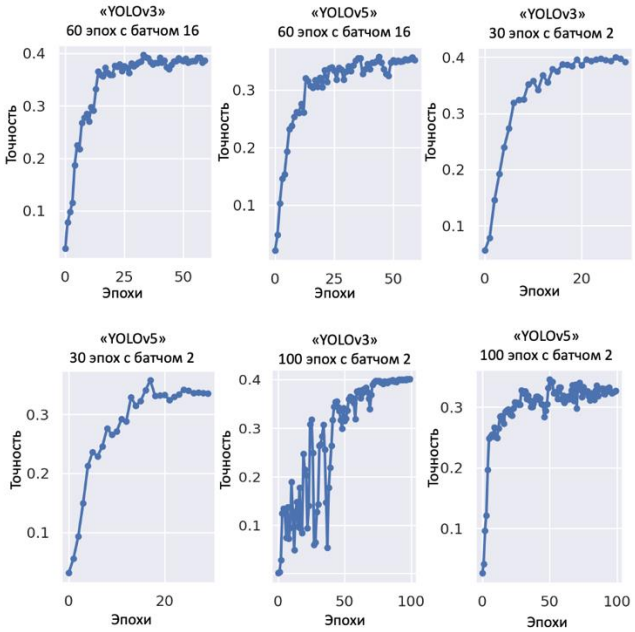


Рисунок 3 – Эмпирическая визуализация средней точности сетей

На основе экспериментов, можно сделать вывод о важности объёмного наличия предметов каждого класса объектов в выборке. В тоже время, эмпирические результаты свидетельствуют об оптимальной вариативности сцен изображений.

Литература

1. Сайт организации конкурсов по исследованию данных Kaggle. [Электронный ресурс]. URL: https://www.kaggle.com/competitions/3d-object-detection-for-autonomous-vehicles/data (дата обращения: 25.04.2022).
2. Rekha B. S, Dr. G N Srinivasan, Chandana Ep3, Achala N Gowda. Object Detection using Region based Convolutional Neural Network: A Survey // IJRASET. 2020. No 8. P. 1927 -1932.
3. Joseph Redmon, Santosh Divvala, Ross Girshick, Ali Farhadi. YOLO: Real-Time Object Detection. [Электронный ресурс]. URL: https://arxiv.org/pdf/1506.02640.pdf (дата обращения: 27.04.2022).
4. Joseph Redmon, Santosh Divvala, Ross Girshick, Ali Farhadi. You Only Look Once: Unified, Real-Time Object Detection. [Электронный ресурс]. URL: https://arxiv.org/pdf/1506.02640v1.pdf (дата обращения: 29.04.2022).
5. Сайт с датасетом COCO. [Электронный ресурс]. URL: https://cocodataset.org/#home (дата обращения: 01.05.2022).
6. Floris van Beers, Arvid Lindstrom, Emmanuel Okafor, Marco A. Wiering. Deep Neural Networks with Intersection over Union Loss for Binary Image Segmentation. [Электронный ресурс]. URL: https://www.researchgate.net/publication/331275482_Deep_Neural_Networks_with_Intersection_over_Union_Loss_for_Binary_Image_Segmentation (дата обращения: 02.05.2022).

Literature

1. Web-site of the organization of contests for data research Kaggle. [Electronic resource]. URL: https://www.kaggle.com/competitions/3d-object-detection-for-autonomous-vehicles/data (accessed: 25.04.2022).
2. Rekha B. S, Dr. G N Srinivasan, Chandana Ep3, Achala N Gowda. Object Detection using Region based Convolutional Neural Network: A Survey // IJRASET. 2020. No 8. P. 1927 -1932.

3. Joseph Redmon, Santosh Divvala, Ross Girshick, Ali Farhadi. YOLO: Real-Time Object Detection. [Electronic resource]. URL: https://arxiv.org/pdf/1506.02640.pdf (accessed: 27.04.2022).
4. Joseph Redmon, Santosh Divvala, Ross Girshick, Ali Farhadi. You Only Look Once: Unified, Real-Time Object Detection. [Electronic resource]. URL: https://arxiv.org/pdf/1506.02640v1.pdf (accessed: 29.04.2022).
5. Web-site for COCO dataset. [Electronic resource]. URL: https://cocodataset.org/#home (accessed: 01.05.2022).
6. Floris van Beers, Arvid Lindstrom, Emmanuel Okafor, Marco A. Wiering. Deep Neural Networks with Intersection over Union Loss for Binary Image Segmentation. [Electronic resource]. URL: https://www.researchgate.net/publication/331275482_Deep_Neural_Networks_with_Intersection_over_Union_Loss_for_Binary_Image_Segmentation (accessed: 02.05.2022).

© Гришкин В.М., Дягилев М.С. 2022 Научно-образовательный журнал для студентов и преподавателей «StudNet» №6/2022.

Для цитирования: Гришкин В.М., Дягилев М.С. ОБНАРУЖЕНИЕ ПРОСТРАНСТВЕННЫХ ОБЪЕКТОВ ДЛЯ АВТОНОМНЫХ ТРАНСПОРТНЫХ СРЕДСТВ// Научно-образовательный журнал для студентов и преподавателей «StudNet» №6/2022.