

УДК 004.932.2

С.В. Воронов, Р.Н. Мухометзянов, И.В. Воронов, В.А. Шрамов

ОБНАРУЖЕНИЕ И РАСПОЗНАВАНИЕ ДОРОЖНЫХ ЗНАКОВ В РЕАЛЬНОМ ВРЕМЕНИ НА МОБИЛЬНЫХ УСТРОЙСТВАХ¹

Воронов Сергей Васильевич, кандидат технических наук, окончил радиотехнический факультет Ульяновского государственного технического университета, доцент кафедры «Радиотехника» УлГТУ. Имеет статьи и монографии в областях обработки и анализа сигналов, изображений и их последовательностей, а также компьютерного зрения. [e-mail: valmedia@yandex.ru].

Мухометзянов Ринат Наилевич, аспирант университета Ватерлоо, Канада, окончил радиотехнический факультет УлГТУ. Имеет статьи в области компьютерного зрения. [e-mail: mukhometzyanov@mail.ru].

Воронов Илья Васильевич, окончил радиотехнический факультет УлГТУ, аспирант того же университета. Имеет статьи в области обработки и рекуррентного оценивания параметров цифровых изображений и их последовательностей. [e-mail: ilvo1987@gmail.com].

Шрамов Вадим Андреевич, окончил радиотехнический факультет УлГТУ, аспирант того же университета. Инженер-исследователь ФНПЦ АО «НПО «Марс». [e-mail: vadim_shramov@mail.ru].

Аннотация

Системы автоматического распознавания дорожных знаков выполняют локализацию дорожных знаков и их распознавание на изображениях, поступающих с бортовых камер автотранспортных средств. Подобные системы направлены на помощь водителям во время движения, а также являются неотъемлемой частью систем автоматического управления или продвинутой поддержки. Статья посвящена детектированию и распознаванию дорожных знаков на основе применения технологий глубокого обучения, адаптированных для мобильных систем с малым энергопотреблением. Предложен подход к распознаванию знаков, состоящий из двух последовательных этапов: детектирование дорожного знака и распознавание класса обнаруженного знака. Данные для анализа были взяты из трех открытых наборов изображений. С целью анализа эффективности полученного решения достигнутые результаты сравнивались с результатами известных подходов к решению задачи детектирования объектов, основанных на применении глубоких сверточных нейронных сетей. Проведенное исследование показало, что предложенный подход для всех использованных наборов данных дал как лучшее качество распознавания, так и максимальное быстроедействие.

Ключевые слова: глубокое обучение, алгоритм Виолы-Джонса, сверточные нейронные сети, распознавание дорожных знаков, обнаружение объектов.

TRAFFIC SIGN DETECTION AND RECOGNITION IN REAL TIME ON MOBILE DEVICES

Sergei Vasilevich Voronov, Candidate of Engineering; graduated from the Faculty of Radioengineering of Ulyanovsk State Technical University; Associate Professor of the Radioengineering Department of UISTU; an author of articles, monographs in the field of digital signal and image processing and computer vision. e-mail: valmedia@yandex.ru.

Rinat Nailevich Mukhometzianov, Postgraduate Student at the University of Waterloo, Canada; graduated from the Faculty of Radioengineering of Ulyanovsk State Technical University; an author of articles in the field of computer vision. e-mail: mukhometzyanov@mail.ru.

Ilya Vasilevich Voronov, Postgraduate Student at Ulyanovsk State Technical University; graduated from the Faculty of Radioengineering of Ulyanovsk State Technical University; an author of articles in the field of digital signal and image processing. e-mail: ilvo1987@gmail.com.

Vadim Andreevich Shramov, Postgraduate Student at Ulyanovsk State Technical University; graduated from the Faculty of Radioengineering of Ulyanovsk State Technical University; Research-Engineer at FRPC JSC 'RPA 'Mars'. e-mail: vadim_shramov@mail.ru.

¹ Исследование выполнено при финансовой поддержке РФФИ, проекты № 16-41-732001 и 18-41-730006.

Abstract

An automatic traffic sign recognition system localizes road signs from images captured by an onboard camera of a vehicle and determine what road signs depicted. Such systems may support the driver on the road, be part of the self-driving cars or advanced driver assistance systems. This paper proposes an approach for solving traffic sign recognition problem using deep learning methods adopted to mobile devices with low power consumption. The approach consists of two consecutive stages: detection of a traffic sign and recognition of a class of the detected sign. Data for analysis were taken from three open sets of images. In order to analyze the effectiveness of the solution obtained, the results achieved were compared with the results of well-known approaches to object detection based on the use of deep convolutional neural networks. The results showed that the proposed algorithm provides the best recognition quality for all used data sets, as well as the highest recognition rate.

Key words: deep learning, Viola-Jones detector, convolutional neural networks, traffic sign recognition, object detection.

ВВЕДЕНИЕ

Распознавание дорожных знаков является актуальной задачей в различных приложениях, таких как системы автоматизированной помощи водителю, системы автоматического управления, понимание сцен городской среды или даже мониторинг технического состояния знаков. Эту задачу можно считать относительно ограниченной, так как сами знаки являются твердыми телами, спроектированными таким образом, чтобы их было легко заметить водителям, а также они имеют лишь небольшие отклонения по внешнему виду.

Решению указанной задачи посвящено немало число исследований [1]. Практически во всех методах в данном направлении можно выделить 3 этапа, а именно: детектирование дорожного знака, распознавание формы знака и классификацию [2] – первые два при этом зачастую объединены. Существуют методы детектирования дорожных знаков, основанные на использовании информации об их цвете и форме [3, 4], а также основанные на применении методов машинного обучения, например AdaBoost [5], и глубокого обучения [6]. Не менее важной является и задача классификации обнаруженных знаков, качество которой во многом определяет надежность всей системы распознавания. Исследования, проведенные в работе [7], показали, что методы, основанные на применении глубоких сверточных нейронных сетей, обеспечивают высокую точность даже в ситуациях с недостаточной видимостью. Этот факт очень важен для систем распознавания дорожных знаков в реальном времени, так как возникают различные мешающие факторы, затрудняющие работу данных систем, например, искажение перспективы, вариации освещения, частичные блокировки, окклюзии, а также плохие погодные условия [2]. Также стоит отметить, что применение систем данного вида связано с необходимостью работы в режимах низкого энергопотребления и небольших вычислительных мощностей по причине требования к их мобильности.

Последние результаты в области разработки мобильных устройств показывают значительный прогресс. Такие устройства как iPhone, iPad, современные устройства на системе Android обладают достаточными

вычислительными ресурсами, чтобы производить распознавание образов в реальном времени. На сегодняшний день нормой являются смартфоны или планшеты с процессором, имеющие четыре ядра процессора. Мобильные устройства старого поколения недостаточно эффективны, так как требуют загрузки изображений на высокоскоростные облачные платформы, где и происходит распознавание сцен, что создает дополнительные проблемы, связанные с задержкой при передаче информации, а также требованием стабильной связи и затратами на обслуживание серверов. Поэтому обработка изображений непосредственно на мобильных платформах для ряда задач является более предпочтительной.

За последние годы, ведущее место в области машинного обучения заняли методы глубокого обучения благодаря их эффективности при решении крупномасштабных задач [8]. Методы данной группы достигли лидирующих показателей при решении большинства задач компьютерного зрения, таких как распознавание объектов [9], классификация видео [10] и понимание сцены [11]. Указанные методы хорошо себя зарекомендовали и при решении задачи распознавания дорожных знаков [12]. Тем не менее, несмотря на свои значительные преимущества, методы глубокого обучения остаются довольно затратными с точки зрения вычислительных и мощностных ресурсов. Поэтому многие из них реализуются в виде облачных сервисов [13], что, как отмечалось, вносит дополнительные сложности и делает их трудно применимыми в системах реального времени. Указанные факты в значительной мере затрудняют применение данных методов в системах распознавания дорожных знаков и требуют их адаптации. Настоящая работа посвящена детектированию и распознаванию дорожных знаков на основе применения технологий глубокого обучения, адаптированных для мобильных систем с малым энергопотреблением.

Для обучения и оценки эффективности исследуемых подходов использовались три публичных набора данных: набор немецких дорожных знаков GTSRB, коллекция дорожных знаков LISA и набор бельгийских дорожных знаков BelgiumTS [14].

ГЛУБОКИЕ НЕЙРОННЫЕ СЕТИ ДЕТЕКТИРОВАНИЯ И КЛАССИФИКАЦИИ ОБЪЕКТОВ

Signe-Shot Multibox Detector (SSD)

Модель глубокой нейронной сети SSD была предложена в работе [15]. На рисунке 1 приведена обобщенная архитектура сети SSD. Первой частью модели SSD является сверточная нейронная сеть классической архитектуры, которая применяется при решении задачи классификации изображений. Ее принято называть базовой сетью (base network). После базовой сети добавлены несколько сверточных слоев, которые и отвечают за формирование предсказаний ограничивающих прямоугольников, потенциально содержащих объекты, а также определяют вероятности принадлежности этих объектов к заданным классам. После чего, для уточнения местоположений объектов и удаления «сдвинутых» прямоугольников с одним и тем же объектом, применяется процедура подавления немаксимумов [15].

Для мобильных систем наиболее перспективной архитектурой базовой сети является MobileNet [16] с множителем глубины, равным 1. Нейронная сеть данной архитектуры была разработана специально для применения в мобильных системах, занимает относительно небольшое пространство в памяти, а также характеризуется высоким быстродействием при хорошей точности классификации [16]. Архитектура сети MobileNet основана на применении раздельных поканальных сверток (depthwise separable convolutions), которые состоят из двух слоев: сверточный слой с пространственным размером ядра 1×1 для разделения каждого канала карты признаков и отдельный для каждого канала сверточный слой с пространственным размером ядра 3×3, результаты которых объединяются в единую карту признаков [16].



Рис. 1. Обобщенная архитектура модели SSD

Функция потерь для модели нейронной сети SSD выглядит следующим образом [14]:

$$L(x, c, l, g) = 1/N (L_{yB}(c, x) + \alpha L_{лок}(x, l, g)),$$

где N – заданное число ограничивающих прямоугольников, x – индикатор совпадения предсказанного (l)

и реального (g) ограничивающих прямоугольников объектов, c – значение уверенности в предсказанном классе объекта, $L_{лок}$ – сглаженная на множитель α функция потерь на детектирование объекта; L_{yB} – кросс-энтропия для классификации объектов [15].

Модель SSD хорошо зарекомендовала себя при решении задачи детектирования объектов и превосходит большинство известных моделей, например, таких как R-CNN, Fast R-CNN [15, 16], по точности и быстродействию.

При обучении данной модели на трех наборах данных, указанных выше, были использованы следующие значения гиперпараметров: алгоритм оптимизации RMSProp [17] с начальным значением шага обучения 0,0001, затуханием 0,9 и значением момента 0,9. Шаг обучения выбран экспоненциально затухающим с показателем 0,95 при 30000 итераций. Все изображения были приведены к одному размеру 300×300 пикселей. Общее число итераций составляло 100000. Для тестирования был использован телефон марки Xiaomi Redmi Note 4X на базе системы Android 7.0. В качестве меры оценки качества была выбрана макроусредненная средняя точность (mean average precision) с порогом на пересечение над объединением 0,5 (maP@0.5IOU или просто maP) [17]. Полученные результаты исследуемой модели для трех наборов данных представлены в таблице 1.

Таблица 1
Результаты качества детектирования для модели SSD с базовой сетью MobileNet

Набор данных	maP	Число кадров в секунду	Размер модели, Мб
GTSRB	0,98	~3	25,9
LISA	0,56	~3	26,2
BelgiumTS	0,17	~3	27,2

Можно заметить, что лучшие результаты показаны на наборе данных GTSRB, худшие – на наборе данных BelgiumTS. Одна из главных причин этого – разное число классов и число экземпляров каждого класса. В наборе данных GTSRB имеется наименьшее число классов (43) и наибольшее среднее число изображений для каждого из классов. Плохие результаты на данных из коллекции BelgiumTS могут быть следствием наибольшего числа классов (61) при наименьшем среднем числе изображений для каждого класса. Более того, в наборе данных BelgiumTS имеется сильный дисбаланс классов, т. е. примеров одних классов гораздо больше, чем примеров других классов, что усложняет обучение. Для примера на рисунке 2 приведен график зависимости maP от номера итерации для одного из классов, имеющего всего 10 изображений. Из графика видно, что точность детектирования по началу немного возрастает, но затем падает.

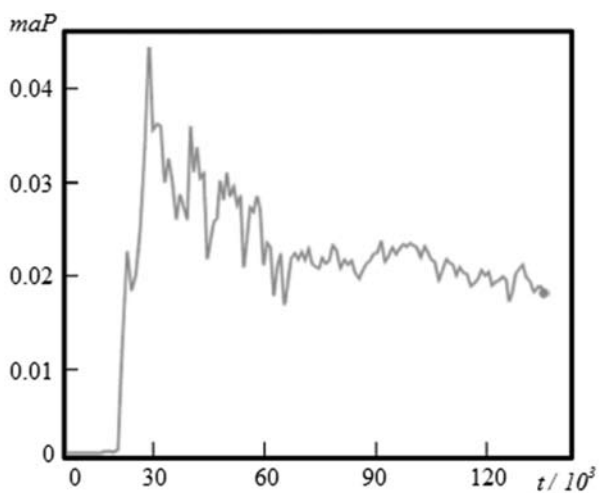


Рис. 2. Значения точности maP для одного класса с малым числом изображений из набора данных BelgiumTS

You Only Look Once (YOLO)

Модель YOLO [18] позволяет детектировать потенциальные области интереса, а также предсказывать к какому классу эти области относятся, используя информацию из всего изображения [19, 20]. Архитектура YOLO (рис. 3) представляет из себя композицию ячеек $S \times S$. Каждая такая ячейка состоит из регрессии для предсказания ограничивающих прямоугольников объектов и классификатора на заданное число классов C , который на основе выделенных ранее признаков предсказывает принадлежность объектов к заданным классам. Число ограничивающих прямоугольников объектов B , т. е. число регрессий в ячейках, может изменяться. В оригинальной работе [18] предлагается использовать значе-

ние $B = 2$. Регрессии ячеек предсказывают 5 значений для каждого ограничивающего прямоугольника x, y, w, h, p , где (x, y) – координаты центра прямоугольника относительно координат ячейки, w, h – ширина и высота прямоугольника соответственно, p – уверенность в нахождении объекта в данном прямоугольнике, а также в параметрах самого прямоугольника. Формально $p = P * IOU$, где P – вероятность отношения объекта к классу, а IOU – пересечение над объединением предсказанного и истинного ограничивающих прямоугольников. Таким образом, каждое выходное значение модели YOLO представляет из себя многомерный тензор размером $S \times S \times (5B + C)$.

При исследованиях была использована модель YOLO из библиотеки Darknet [21]. Число итераций обучения составило 240000, а изображения приведены к размеру 448×448 пикселей. Результаты исследования, представленные в таблице 2, показывают, что для качественного обучения модели YOLO необходим достаточно большой и полный набор данных. Используемые в данной работе наборы данных к таковым не относятся, что и повлияло на очень слабые результаты, так как модель явно получилась недообученной.

Таблица 2
Результаты качества детектирования для модели YOLO

Набор данных	maP	Число кадров в секунду	Размер модели, Мб
GTSRB	0,036	~2	203,9
LISA	0,043	~2	203,9
BelgiumTS	0,033	~1	203,9

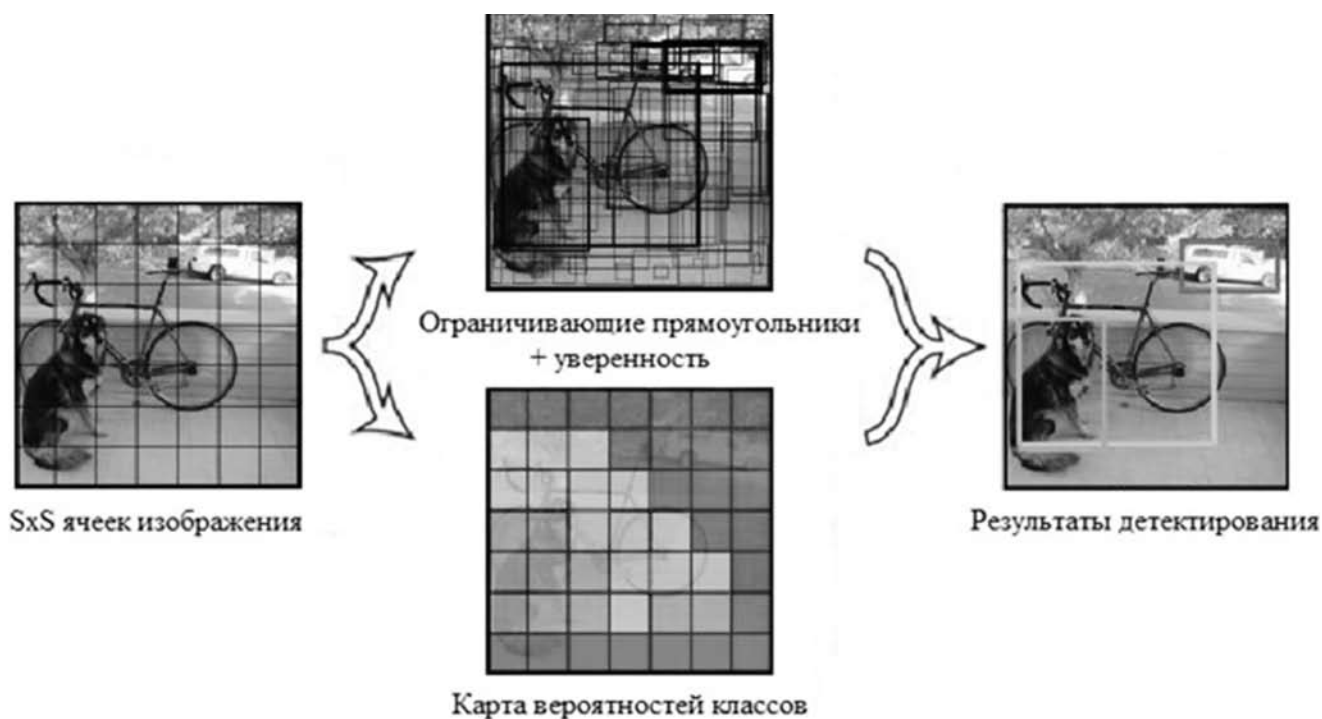


Рис. 3. Обобщенная модель YOLO

ДЕТЕКТИРОВАНИЕ НА ОСНОВЕ КОМБИНАЦИИ АЛГОРИТМА ВИОЛЫ-ДЖОНСА И СВЕРТОЧНОЙ НЕЙРОННОЙ СЕТИ

Решение задачи распознавания дорожных знаков представим в виде двух последовательных этапов: детектирование (локализация) дорожного знака и распознавание класса обнаруженного знака. На первом этапе на основе анализа всего изображения алгоритм детектирования генерирует области интереса, т. е. ограничивающие прямоугольники, которые потенциально содержат дорожный знак. Далее полученные области подаются на некоторый классификатор, который определяет класс дорожного знака или указывает на факт отсутствия в данной области знака.

При решении задачи детектирования дорожного знака перспективным является применение алгоритма Виолы-Джонса, который хорошо себя зарекомендовал при детектировании объектов благодаря сочетанию качества и высокого быстродействия, в том числе и для встраиваемых систем реального времени, так как позволяет обнаруживать объекты со скоростью до 15 кадров в секунду при наличии одноядерного процессора частой всего 700 МГц [22].

Детектор Виолы-Джонса представляет собой сильный бинарный классификатор, построенный на нескольких слабых детекторах (рис. 4). Каждый слабый детектор является очень простым бинарным классификатором. Обычно для таких классификаторов используются простые признаки, генерируемые из участков изображений, например, такие как признаки Хаара (рис. 5). Такие признаки легко и быстро вычисляются, особенно при использовании интегральных изображений [22]. Для обучения каскадного классификатора Виолы-Джонса было использовано 22000 изображений дорожного знака размером 80×80 пикселей и 11000 изображений фона размером 150×150 пикселей.

В качестве классификатора целесообразно использовать сверточные нейронные сети, так как они являются очень гибкими в проектировании, позволяя регулировать качество и требуемый объем вычислительных затрат [23]. Большинство современных архитектур сверточных нейронных сетей, таких как AlexNet, VGG Net, GoogLeNet, Microsoft ResNet [24], были сконструированы для решения задачи классификации объектов для 1000 разных классов, которые содержатся в известном

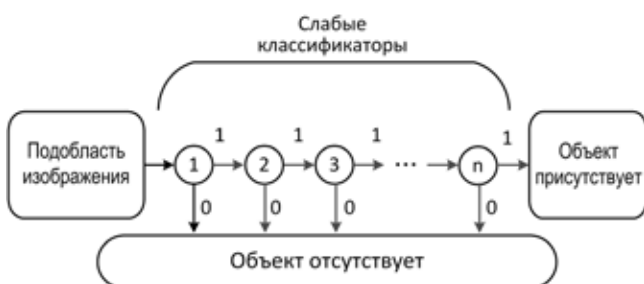


Рис. 4. Структура каскадного классификатора Виолы-Джонса

открытом наборе данных Imagenet [24]. Для решения поставленной в данной работе задачи необходимо построить классификатор на 40–60 классов, что значительно меньше. Кроме того, типичные знаки имеют всего лишь порядка 9 различных форм (круг, треугольник, прямоугольник и пр.), на которых используется порядка 7 основных цветов [35].

Поэтому для достижения наилучшей эффективности была разработана собственная неглубокая архитектура сверточной нейронной сети, имеющая для входных изображений размером 32×32 пикселя относительно небольшое число параметров. Ее структура для набора данных GTSRB приведена в таблице 3. Сеть состоит всего лишь из двух сверточных и пяти полносвязных слоев. Для обучения классификатора использовался алгоритм оптимизации Adam [26] с начальным шагом обучения 0,0001 и экспоненциальным затуханием 0,00000001 после каждой итерации, общее число эпох обучения составило 10000.

Таблица 3
Архитектура сверточной нейронной сети, используемая при распознавании знаков из набора GTSRB

№ слоя	Тип слоя	Глубина	Размерность выхода	Размер ядра
0	Входной	3	32x32x3	-
1	Сверточный	30	28x28x30	5x5
2	Пулинг с функцией максимума	30	14x14x30	2x2
3	Сверточный	200	10x10x200	5x5
4	Пулинг с функцией максимума	200	10x10x200	2x2
5	Полносвязный	-	2200	1x1
6	Полносвязный	-	1000	1x1
7	Полносвязный	-	500	1x1
8	Полносвязный	-	120	1x1
9	Полносвязный	-	43	1x1



Рис. 5. Пример каскадов Хаара.

Результаты качества детектирования для предложенного подхода представлены в таблице 4. При этом скорость детектирования составила порядка 25 кадров в секунду, а с распознаванием – 8 кадров в секунду. Также отметим, что данный подход обеспечил лучшее качество распознавания на всех наборах данных по сравнению с моделями SSD и YOLO.

Таблица 4

Результаты качества детектирования для алгоритма на основе детектора Виолы-Джонса и классификатора в виде сверточной нейронной сети

Набор данных	mAP	Число кадров в секунду	Размер модели, Мб
GTSRB	0,99	~8	54,9
LISA	0,86	~8	55,0
BelgiumTS	0,70	~8	54,9

ЗАКЛЮЧЕНИЕ

Исследованы подходы к распознаванию дорожных знаков на мобильной платформе в реальном времени с применением методов глубокого обучения. Данные для анализа были взяты из трех открытых наборов изображений: набор немецких дорожных знаков GTSRB, коллекция дорожных знаков LISA и набор бельгийских дорожных знаков BelgiumTS. В качестве платформы для проверки использовался смартфон Xiaomi Redmi Note 4X на базе системы Android

Предложен подход к распознаванию, состоящий из двух последовательных этапов: детектирование дорожного знака и распознавание класса обнаруженного знака. В качестве детектора предложено использовать алгоритм Виолы-Джонса на основе каскадов Хаара. Для решения задачи классификации была разработана архитектура сверточной нейронной сети с малым числом параметров. Для анализа эффективности данного решения полученные результаты сравнивались с известными подходами к решению данной задачи, основанными на применении моделей SSD и YOLO.

Проведенное исследование показало, что для поставленной задачи предложенный подход для всех использованных наборов данных дал как лучшее качество распознавания, так и максимальное быстродействие. Худшие результаты показала модель YOLO, что может быть связано с недостаточным объемом имеющихся данных для качественного обучения.

СПИСОК ЛИТЕРАТУРЫ

1. Gudigar A., Chokkadi S. and U. R. *Multimed Tools Appl* (2016) 75: 333. – URL: <https://doi.org/10.1007/s11042-014-2293-7>.
2. Escalera S., Baró X., Pujol O., Vitrià J., and P. Radeva *Background on Traffic Sign Detection and Recognition // Traffic-Sign Recognition Systems Springer Briefs in Computer Science*. 2011. – pp. 5–13.
3. Zumra Malik and Imran Siddiqi *Detection and Recognition of Traffic Signs from Road Scene Image // 12th*

IEEE International Conference on Frontiers of Information Technology. 2014. – pp. 330–335.

4. Paulo C.F. and Correia P.L. *Automatic detection and classification of traffic signs in Image Analysis for Multimedia Interactive Services (WIAMIS '07) // Eighth International Workshop on*. 6–8 June 2007. – pp. 1–11.

5. Timofte R., Zimmermann K. & Van Gool L. *Machine Vision and Applications* (2014) 25: 633.

6. Xiong C.Z., Wang, C., Ma W.X., Shan Y.M. *A traffic sign detection algorithm based on deep convolutional neural network // In Proceedings of the IEEE International Conference on Signal and Image Processing, Beijing, China, 13–15 August, 2016*. – pp. 676–679.

7. Ciresan D., Meier U., Masci J. and J. Schmidhuber *Multi-column deep neural network for traffic sign classification // Neural Networks*. 2012. – Vol. – 32. – pp. 333–338.

8. LeCun Y., Bengio Y., and G. Hinton *Deep learning // Nature*, 521(7553). 2015. pp. 436–444.

9. Krizhevsky A., Sutskever I., and G.E. Hinton *Imagenet classification with deep convolutional neural networks // In Advances in neural information processing systems*. – 2012. – pp. 1097–1105.

10. Karpathy A., Toderici G., Shetty S., Leung T., Sukthankar R. and L. Fei-Fei *Large-scale video classification with convolutional neural networks // In Proceedings of the IEEE conference on Computer Vision and Pattern Recognition*. – 2014. – pp. 1725–1732.

11. Zhou B., Lapedriza A., Xiao J., Torralba A. and A. Oliva *Learning deep features for scene recognition using places database // In Advances in neural information processing systems*. – 2014. – pp. 487–495.

12. Jin J., Fu K., Zhang C. *Traffic Sign Recognition with Hinge Loss Trained Convolution Neural Networks // IEEE Trans. Intell. Transp. Syst.* 2014; 15:1991–2000.

13. Chen T. Y.-H., Ravindranath L., Deng S., Bahl P. and H. Balakrishnan *Glimpse: Continuous, real-time object recognition on mobile devices // In Proceedings of the 13th ACM Conference on Embedded Networked Sensor Systems*. – 2015. – pp. 155–168.

14. Mukhometzianov R., Wang Y. *Review. Machine learning techniques for traffic sign detection arXiv preprint arXiv: 1712.04391*, 2017.

15. Liu W. et al. (2016) *SSD: Single Shot MultiBox Detector*. In: Leibe B., Matas J., Sebe N., Welling M. (eds) *Computer Vision – ECCV 2016. Lecture Notes in Computer Science*, Vol. 9905. Springer, Cham.

16. Howard A. G., Zhu M., Chen B., Kalenichenko D., Wang W., Weyand T., Andreetto M. and H. Adam *Mobilenets: Efficient convolutional neural networks for mobile vision applications // CoRR*. 2017. abs/1704.04861.

17. Tang S. and Y. Yuan *Object Detection based on Convolutional Neural Network*. Stanford University. – 2015.

18. Romero Aquino, Nelson & Gutoski, Matheus & Hattori, Leandro & Silvério Lopes, Heitor. (2017) *The Effect of Data Augmentation on the Performance of Convolutional Neural Networks // In Proceedings of the Brazilian Society of Computational Intelligence Conference*. At Niterói, Rio de Janeiro, October 2017.

19. Redmon J., Divvala S., Girshick R. and A. Farhadi. You Only Look Once: unified, real-time object detection arXiv preprint arXiv:1506.02640. – 2015.

20. Redmon J., Farhadi A. YOLO9000: better, faster, stronger. arXiv preprint arXiv:1612.08242, 2016.

21. Redmon J., Darknet: Open source neural networks in c.2013–2016. – URL: <http://pjreddie.com/darknet/>.

22. Viola P. and M. Jones. Robust Real-time Object Detection // *International Journal of Computer Vision*. – 2004. Vol. – 57 (2). – pp. 137–154.

23. Goodfellow I., Bengio Y. Deep Learning (Adaptive Computation and Machine Learning series) // *Adaptive Computation and Machine Learning series*. – 2016. – pp. 330–371.

24. Rawat, Waseem & Wang, Zenghui. (2017). Deep Convolutional Neural Networks for Image Classification: A Comprehensive Review. *Neural Computation*. 1-98. 10.1162/NECO_a_00990.

25. How to Read Traffic Signs: Recognizing Shapes and Colors | DMV.org, DMV.org Articles, 2017. – URL: <https://www.dmv.org/articles/how-to-read-traffic-signs-recognizing-shapes-and-colors/>.

26. Kingma D. and J. Ba. Adam: A method for stochastic optimization arXiv preprint arXiv:1412.6980, 2014.

REFERENCES

1. Gudigar A., Chokkadi S. and U. R. *Multimedia Tools and Applications*. (2016) 75: 333. Available at: <https://doi.org/10.1007/s11042-014-2293-7>.

2. Escalera S., Baró X., Pujol O., Vitrià J., and P. Radeva. Background on Traffic Sign Detection and Recognition. *Traffic-Sign Recognition Systems Springer Briefs in Computer Science*, 2011, pp. 5–13.

3. Zumra Malik and Imran Siddiqi. Detection and Recognition of Traffic Signs from Road Scene Image. *12th IEEE International Conference on Frontiers of Information Technology*. 2014, pp. 330-335.

4. Paulo C.F. and Correia P.L. Automatic Detection and Classification of Traffic Signs in Image Analysis for Multimedia Interactive Services (WIAMIS '07). *Eighth International Workshop on*. 6–8 June 2007, pp. 1–11.

5. Timofte R., Zimmermann K. and Van Gool L. *Machine Vision and Applications*, 2014, no. 25, pp. 633–647.

6. Xiong C.Z., Wang, C., Ma W.X., Shan Y.M. A Traffic Sign Detection Algorithm Based on Deep Convolutional Neural Network. *In Proceedings of the IEEE International Conference on Signal and Image Processing*, Beijing, China, 13–15 August, 2016, pp. 676–679.

7. Ciresan D., Meier U., Masci J. and J. Schmidhuber. Multi-Column Deep Neural Network for Traffic Sign Classification. *Neural Networks*, 2012, vol. 32, pp. 333–338.

8. LeCun Y., Bengio Y., and G. Hinton. Deep Learning. *Nature*, 2015, no. 521 (7553), pp. 436–444.

9. Krizhevsky A., Sutskever I., and G.E. Hinton. Imagenet Classification with Deep Convolutional Neural Networks. *In Advances in Neural Information Processing Systems*, 2012, pp. 1097–1105.

10. Karpathy A., Toderici G., Shetty S., Leung T., Sukthankar R. and L. Fei-Fei. Large-Scale Video Classification

with Convolutional Neural Networks. *In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2014, pp. 1725–1732.

11. Zhou B., Lapedriza A., Xiao J., Torralba A. and A. Oliva. Learning Deep Features for Scene Recognition Using Places Database. *In Advances in Neural Information Processing Systems*, 2014, pp. 487–495.

12. Jin J., Fu K., Zhang C. Traffic Sign Recognition with Hinge Loss Trained Convolution Neural Networks. *IEEE Trans. Intell. Transp. Syst.*, 2014, 15:1991–2000.

13. Chen T. Y.-H., Ravindranath L., Deng S., Bahl P. and H. Balakrishnan. Glimpse: Continuous, Real-Time Object Recognition on Mobile Devices. *In Proceedings of the 13th ACM Conference on Embedded Networked Sensor Systems*, 2015, pp. 155–168.

14. Mukhometzianov R., Wang Y. *Review. Machine Learning Techniques for Traffic Sign Detection arXiv preprint arXiv: 1712.04391*, 2017.

15. Liu W. et al. (2016) SSD: Single Shot MultiBox Detector. In: Leibe B., Matas J., Sebe N., Welling M. (eds) *Computer Vision – ECCV 2016. Lecture Notes in Computer Science*, vol. 9905. Springer, Cham.

16. Howard A. G., Zhu M., Chen B., Kalenichenko D., Wang W., Weyand T., Andreetto M. and H. Adam. Mobilenets: Efficient Convolutional Neural Networks for Mobile Vision Applications. *CoRR*, 2017, abs/1704.04861.

17. Tang S. and Y. Yuan. *Object Detection based on Convolutional Neural Network*. Stanford University. 2015.

18. Romero Aquino, Nelson & Gutoski, Matheus & Hattori, Leandro & Silvério Lopes, Heitor. The Effect of Data Augmentation on the Performance of Convolutional Neural Networks. *In Proceedings of the Brazilian Society of Computational Intelligence Conference*. At Niterói, Rio de Janeiro, October 2017.

19. Redmon J., Divvala S., Girshick R. and A. Farhadi. *You Only Look Once: Unified, Real-Time Object Detection arXiv preprint arXiv:1506.02640*, 2015.

20. Redmon J., Farhadi A. *YOLO9000: Better, Faster, Stronger. arXiv preprint arXiv:1612.08242*, 2016.

21. Redmon J. *Darknet: Open Source Neural Networks in c.2013–2016*. Available at: <http://pjreddie.com/darknet/>.

22. Viola P. and M. Jones. Robust Real-time Object Detection. *International Journal of Computer Vision*, 2004, vol. 57 (2), pp. 137–154.

23. Goodfellow I., Bengio Y. Deep Learning (Adaptive Computation and Machine Learning Series). *Adaptive Computation and Machine Learning Series*. 2016, pp. 330–371.

24. Rawat, Waseem & Wang, Zenghui. Deep Convolutional Neural Networks for Image Classification: A Comprehensive Review. *Neural Computation*, 2017. 1-98. 10.1162/NECO_a_00990.

25. *How to Read Traffic Signs: Recognizing Shapes and Colors*. DMV.org, DMV.org Articles, 2017. Available at: <https://www.dmv.org/articles/how-to-read-traffic-signs-recognizing-shapes-and-colors/>.

26. Kingma D. and J. Ba. Adam. *A Method for Stochastic Optimization arXiv preprint arXiv:1412.6980*, 2014.