

# ПРИМЕНЕНИЕ СВЕРТОЧНОЙ НЕЙРОННОЙ СЕТИ ДЛЯ ОБНАРУЖЕНИЯ АВТОМОБИЛЕЙ ЭКСТРЕННЫХ СЛУЖБ С ПОМОЩЬЮ ВИДЕОКАМЕР В УСЛОВИЯХ ИНТЕНСИВНОГО ДОРОЖНОГО ДВИЖЕНИЯ

## USING A CONVOLUTIONAL NEURAL NETWORK TO DETECT EMERGENCY VEHICLES USING VIDEO CAMERAS IN HEAVY TRAFFIC

**I. Zikratov  
I. Shakhsuvarova**

*Summary.* An automated system for detecting emergency vehicles from surveillance camera recordings using a deep convolutional neural network was proposed. The developed method has shown good results in detecting and classifying ambulances, repair and fire trucks. This technology can be used in an automated system that can detect a special vehicle on a high-traffic road and notify the dispatcher or automatically redirect other vehicles to make way.

*Keywords* emergency vehicles, convolutional neural network, automated system, high-traffic road, detecting.

**В** настоящее время с ростом количества автомобилей как в нашей стране, так и во всем мире, возникает множество острых проблем, которые безотлагательно необходимо решать. Возросшая интенсивность дорожного движения, сложность дорожных условий, многополосность, развязки и перекрестки — всё это приводит к росту дорожно-транспортных происшествий и возникновению автомобильных пробок. В этих сложных условиях машины специальных и экстренных служб, такие как скорая помощь, пожарные машины, застревают в пробках, что во многих случаях угрожает человеческой жизни. В таких ситуациях очень важно предоставить приоритет и помочь освободить путь проезду таких автомобилей. Но автоинспекции трудно, а иногда и невозможно справиться с этой задачей. По этой причине возникает острая необходимость в наличии автоматизированной системы, способной обнаружить автомобиль скорой помощи на дороге с интенсивным движением, сообщить об этом диспетчеру или автоматически перенаправить другие автомобили, чтобы освободить путь. Поэтому особую актуальность приобретает автоматизированная система обнаружения автомобилей скорой помощи по записям камер видеонаблюдения с использованием глубокой сверточной нейронной сети.

В свое время для обнаружения транспортных средств экстренных служб, таких как машины скорой помощи, полицейские машины, пожарные машины и т.д. было создано несколько систем. Неллоре и другие [1] рассчитали расстояние до автомобиля скорой помощи с использованием видеокamеры с последующей передачей данных в центр управления движением. Они использовали метод визуального зондирования. Была использована камера для записи видео 1920\*1080 пикселей с частотой 30 кадров в секунду. Сначала изображения преобразуются из графического формата RGB в градации серого, а затем производится их обработка. К каждому изображению были применены некоторые морфологические операции, после чего измерялось расстояние между камерой и автомобилем с помощью различных методов. Затем данные о расстоянии, скорости и количестве перемещенных передавались в центр управления дорожным

**Зикратов Игорь Алексеевич**

*Д.т.н., профессор, Санкт-Петербургский государственный университет телекоммуникаций им. проф. М.А. Бонч-Бруевича*  
zikratov.ia@spbgut.ru

**Шахсуварова Илона Витальевна**

*Аспирант, Санкт-Петербургский государственный университет телекоммуникаций им. проф. М.А. Бонч-Бруевича*  
shakhsuvarova.ilona@mail.ru

*Аннотация.* Предложена автоматизированная система для обнаружения автомобилей экстренных служб по записям камер видеонаблюдения с использованием глубокой сверточной нейронной сети. Разработанный метод показал хорошие результаты в обнаружении и классификации автомобилей скорой помощи, ремонтных и пожарных машин. Эта технология может быть использована в автоматизированной системе, которая сможет обнаружить специальный автомобиль на дороге с интенсивным движением, сообщить об этом диспетчеру или автоматически перенаправить другие автомобили, чтобы освободить путь.

*Ключевые слова:* автомобили экстренных служб, сверточная нейронная сеть, автоматизированная система, дорога с интенсивным движением, обнаружение.

движением, чтобы более эффективно управлять автомобилем скорой помощи. С другой стороны, в работе [2] разработали усовершенствованную систему управления дорожным движением, которая минимизирует уровень заторов для автомобилей экстренных служб. Архитектура контроллера управления трафиком была разработана с помощью контроллера нечеткой логики. Этот метод позволил управлять изменением сигнала светофора, изменением ограничения скорости, освобождением полосы движения и т.д. Нечеткая логика определяет наиболее точную оценку низкого, среднего и высокого уровня загруженности дорог. Другая идея была высказана в работе [3] о создании интеллектуальной системы управления дорожным движением, которая реализует некоторые встроенные системы для предоставления большего приоритета автомобилю экстренной службы в системе управления дорожным движением. Измерялась плотность потока транспортных средств с помощью инфракрасного детектора, но у авторов не получилось работать эффективно в сценариях реального времени. Для создания прототипа модели трафика был использован игрушечный автомобиль. Сначала изображения были обрезаны в интересующей области, на изображениях были выделены только красный и синий цвета и измерено расстояние между каждым красным и синим цветом. Если расстояние очень мало, то высока вероятность того, что это светодиодный фонарь на крыше автомобиля скорой помощи. Изображения были сделаны только в верхней части, при этом еще измерялась плотность потока транспортных средств. Изображения были преобразованы в полутоновые и к ним применены некоторые морфологические операции. Шум изображений был удален с помощью гауссовского фильтра и подсчитан набор связанных пикселей.

Автомобили скорой помощи играют важную роль в любой опасной для жизни ситуации. В соответствии со статистическими данными дорожные пробки уносят более 20% жизней пациентов в машине скорой помощи, причем, если состояние пациентов очень тяжелое, процент смертельного исхода увеличивается [4]. Это экстраординарные ситуации, когда экстренному пациенту необходимо немедленно отправиться в больницу, а машина скорой помощи застряла в пробке. Такой сценарий особо опасен в случае с кардиобольными, которых необходимо своевременно доставить в больницу. В пробках многие водители, и в особенности, автолюбители, не удосуживаются пропустить машину скорой помощи, а сотрудники автоинспекции не видят, какую полосу они должны освободить для проезда этого автомобиля. Поэтому многие пациенты погибают, не доехав до больницы [4].

Ситуация с пробками на дорогах также представляет собой серьезную проблему для пожарных команд.

По статистическим данным Международной ассоциации пожарно-спасательных служб в России в одной трети случаев вызова пожарная команда опаздывала из-за дорожных осложнений в виде пробок и отсутствия объездных путей, в частности, загроможденных дворовых проездов [5]. В США, по крайней мере, в 90% случаев, когда требуется пожарная команда, люди ожидают, что пожарные приедут в течение четырех минут [6]. В 2013 году в США пожарные команды получали вызовы более миллиона раз. Хотя пожарные реагировали на них очень быстро, за год погибло более трех тысяч гражданских лиц, свыше пятнадцати тысяч получили ранения, а материальный ущерб составил более одиннадцати миллиардов долларов [7]. Приведенные данные показывают, что для своевременного прибытия пожарной команды очень ценна каждая секунда. Таким образом, из-за задержки прибытия экстренных служб к месту чрезвычайной ситуации погибают люди и велики материальные потери.

Одним из возможных решений этих проблем может быть внедрение в практику интеллектуальной автоматизированной системы, интегрированной с системой управления дорожным движением, которая будет обнаруживать машины экстренных служб и предоставлять им приоритет. Необходимо создать систему для обнаружения автомобилей и классификации их как экстренных или обычных. Архитектуры глубокого обучения, такие как глубокие сверточные нейронные сети, применяются в компьютерном зрении и дают хорошие результаты, не только сравнимые с возможностями человека, но и превосходящие человеческий опыт. Человеческий мозг может легко обнаружить любое транспортное средство и визуально обработать его за малую долю времени. Но человеческий мозг имеет некоторые ограничения, он может отслеживать или удерживать внимание не более получаса. Для решения задачи возможно получение записей с камер видеонаблюдения на дороге и обнаружение специального автомобиля экстренной службы. С помощью камеры видеонаблюдения изображения делаются каждую секунду и на каждом изображении обнаруживается транспортное средство. После обнаружения каждого автомобиля, система классифицирует их на автомобиль экстренной службы и обычный автомобиль. Если обнаружен специальный автомобиль, компьютер может уведомить автоинспекцию или автоматизированную систему, чтобы освободить ему движение по трассе.

Для обнаружения объектов с помощью видеокамеры и классификации изображения как автомобиля скорой помощи можно использовать алгоритм обнаружения цели YOLO-V3 [8, 9]. You Only Look Once (YOLO) — это современная система обнаружения объектов в реальном времени. На компьютере с моделью графического ускорителя от NVIDIA Pascal Titan X с процессором ви-

деокарты на архитектуре Pascal YOLO-V3 обрабатывает изображения со скоростью 30 кадров в секунду и имеет mAP 57,9% на тест-разработчике COCO. YOLO-V3 очень быстрый и точный алгоритм, находится на одном уровне с Focal Loss, но примерно в 4 раза быстрее [9]. Кроме того, можно легко найти баланс между скоростью и точностью, просто изменив размер модели, не применяя переобучения. Архитектура YOLO-V3 очень быстрая и может обрабатывать 45 изображений в секунду на компьютере с хорошим процессором. YOLO-V3 преобразует изображение и делит его на фиксированные сетки. Затем он прогнозирует множество ограничивающих рамок и определяет вероятность появления объекта в этих рамках. Таким образом, он не требует сложной конвейеризации и напрямую оптимизирует производительность обнаружения.

Что касается сверточных нейронных сетей на основе регионов (region-based), то в последнее время для обнаружения транспортных средств часто используются такие методы, как Regions with CNN features (R-CNN), Spatial Pyramid Pooling Network (SSP-Net), Fast R-CNN и Faster R-CNN. Эти подходы, хотя и достигают самой современной точности, слишком требовательны к вычислительным ресурсам и слишком медленны в режиме реального или близкого к реальному времени.

Сверточная нейронная сеть R-CNN [10], ее конвейер работает медленнее, чем YOLO и YOLO допускает меньше фоновых ошибок. Метод YOLO обладает высокой обобщающей способностью для обнаружения объектов.

Первый шаг на пути к обнаружению автомобиля экстренной службы — это обнаружение объекта. Для нашей цели задача детектора объектов заключается в построении и обучении такого детектора автомобилей, который будет выдавать изображения только автомобилей. С помощью YOLO-V3 по этим изображениям мы будем классифицировать и производить отбор, является ли автомобиль экстренным, автомобилем скорой помощи или обычным автомобилем. Хотя YOLO-V3 это не самый совершенный алгоритм обнаружения объектов, но зато это самая быстрая модель. Поскольку мы должны прогнать много изображений для модели, которая будет создана на основе записей камер видеонаблюдения, алгоритм обнаружения должен быть очень быстрым.

ImageNet — это база данных изображений, организованная в соответствии с иерархией WordNet, в которой каждый узел иерархии представлен сотнями и тысячами изображений. Этот проект сыграл важную роль в продвижении исследований в области компьютерного зрения и глубокого обучения. Данные его доступны бесплатно исследователям для некоммерческого использования.

В YOLO-V3 использована глубокая сверточная нейронная сеть с архитектурой Darknet-53. Это 53-слойная нейронная сеть, обученная на ImageNet для классификации в сочетании со слоями обнаружения, что делает общую глубину сети 106 слоев. Она обеспечивает обнаружение в трех различных масштабах. Сначала изображение делится на клетки-боксы с помощью сетки выбранного шага. Каждый бокс представляет собой один объект, обнаруженный в заданном центре. Каждый бокс содержит координаты объекта, оценки объективности и оценки вероятности класса. Координаты бокса состоят из центра, высоты и ширины. Оценка объективности — это оценка вероятности того, что сетка обнаружит объект. Вероятность класса — это вероятность каждого класса, на котором обучена модель. В данном случае модель обучена на наборе данных Common Object in Context (COCO) [11], который содержит 80 классов объектов и, таким образом, оценка класса дает вероятность 80 объектов. COCO — это набор данных, который представляет собой крупномасштабное обнаружение объектов, создание подписей и сегментацию. Этот набор данных сегментирует и обнаруживает объекты, встречающиеся в повседневной жизни в нашей естественной среде. В наборе данных COCO имеются следующие классы: «грузовик», «автобус» и «автомобиль». Все изображения, отнесенные к этим классам, проходят через классификатор для отбора экстренных и обычных автомобилей.

Для классификации автомобилей экстренных служб используется два класса: автомобиль экстренной службы и обычный автомобиль. Для класса обычных автомобилей пригоден набор данных Stanford Universitys cars dataset [12]. В нем 8144 изображений обычных автомобилей, спортивных автомобилей, фургонов, внедорожников, грузовиков и т.д. Из этого набора данных можно выбрать 1500 изображений различных типов автомобилей для обучения и тестирования глубокой нейронной сети. Для класса автомобилей экстренных служб собирают изображения скорой помощи и пожарных или иных грузовиков из поиска Google. В Википедии есть много изображений машины скорой помощи или грузовика, которые используются для их описания. Все эти изображения представляют набор экспериментальных данных. Для машин скорой помощи отобрано около 800 изображений и 700 изображений для пожарных машин. Эти 1500 изображений автомобилей также разделены случайным образом на 90% для обучения и 10% для тестирования.

Затем для лучшего обучения модели проводится расширение объема данных. Это процесс создания дополнительных данных из имеющихся путем поворота, масштабирования, перевода, обрезки и других преобразований. Встроенный загрузчик данных выполняет эту работу во время обучения. В процессе загрузки данных

Таблица 1. Точность для различных размеров входных данных

Размер входного изображения	Точность определения
(64 x 64)	97,48%
(48 x 48)	96,93%
(224 x 224)	96,93%
(360 x 360)	96,93%
(128 x 128)	96,45%

Таблица 2. Сравнение точности обнаружения с различными предварительно обученными моделями

Модель с предварительным обучением	Точность
2-слойная сверточная нейронная сеть	96,93%
VGG-16 [17]	98,67%
Inception-v3 [18]	96,85%
Xception [19]	97,78%

он применяет трансформацию и каждый раз генерирует немного другое изображение. Поскольку изображения в нашем наборе данных собраны хорошо, достаточно применить лишь небольшое количество расширений. Выбраны следующие типы и количество расширяющих трансформаций: вращение от  $-20^\circ$  до  $20^\circ$  градусов; горизонтальный переворот; и масштабирующий множитель 1,1.

Для обучения модели вместо обычного стохастического градиентного спуска был использован алгоритм оптимизации Адама [13]. Этот метод использует не фиксированную, а адаптируемую скорость обучения и два параметра  $\beta_1$  и  $\beta_2$  для управления скоростью обучения, это, так называемые, импульсные параметры. Параметры оптимизации использованы следующие: скорость обучения 0,001,  $\beta_1=0,9$  и  $\beta_2=0,999$ .

Далее, прежде чем приступить к экспериментам с основной моделью, полезно построить базовую простую двухслойную сверточную нейронную сеть для получения базовой точности и провести несколько экспериментов с некоторыми гиперпараметрами и методами регуляризации. Точности различных экспериментов для различных размеров входных данных приведены в таблице 1.

Как видно из приведенной выше Таблицы 1, простая базовая модель дала наилучший результат при размерности 64 x 64. Хотя большие изображения содержат больше информации для обработки и обычно дают лучший результат, простая неглубокая сверточная нейронная сеть не может воспользоваться этим преимуществом, так как у нее недостаточно параметров и сложной структуры, чтобы уловить тонкие паттерны на большом изображении.

Глубокая сверточная нейронная сеть [14] является самой передовой технологией для классификации и обнаружения объектов на изображении. В настоящее время возможности обнаружения и классификации объектов значительно улучшились. Сверточная нейронная сеть использует концепцию нейронных ядер для обработки изображения. Параметры этих ядер изучаются в процессе обучения. Глубокая сверточная нейронная сеть строится из множества таких сверточных слоев. Но у нее есть два недостатка: такую сеть трудно обучить, и для ее обучения требуется большое количество данных. Для решения такого рода проблем можно использовать трансфертное обучение [15, 16]. Идея заключается в возможности использовать данные, полученные сетью в одном приложении, в другом приложении, в разных приложениях. Поэтому можно провести эксперименты с несколькими различными предварительно обученными моделями, включая VGG-16 [17], Inception-v3 [18] и Xception [19]. Для таких экспериментов можно выбрать не самый маленький размер изображения, например 224 x 224; использовался набор данных KDEF. После проверки работоспособности нейронных сетей и сборки всех шагов алгоритма, были получены окончательные программы для разных исследованных вариантов. Предварительные работы, исследование отдельных блоков программ производилось на обычном персональном компьютере со следующими характеристиками: ОС Windows 10, видеокарта Nvidia Geforce 840m 2020 Мбайт, процессор Intel core i3-4010U CPU1,7 ГГц, оперативная память 4 Гбайт. Для окончательных работ привлекались услуги аутсорсинга с профессиональным компьютером на базе платформы визуальных вычислений TITAN RTX: ОС Windows 10, видеокарта Nvidia Titan RTX 24 Гб, процессор Intel Core i9-10900X, два модуля памяти 16gb Kingston FURY Renegade 3000 МГц. В итоге проделанных

исследований можно получить и сравнить в Таблице 2 результаты для всех использованных моделей:

В проведенных исследованиях предложена модель, которая может обнаружить автомобили скорой помощи на дороге с интенсивным движением. В густонаселенных агломератах на дорогах с интенсивным движением слишком много транспорта и из-за этого машины экстренных служб, такие как скорая помощь, ремонтная служба и пожарные машины испытывают проблемы

в пути посреди дороги. Предложенная модель достаточно эффективно решает эту проблему, проявляет хорошие результаты в обнаружении и идентификации автомобилей экстренных служб всех видов. Она может быть встроена в систему видеонаблюдения для отслеживания специальных автомобилей и предоставления приоритета на дороге для проезда машин экстренных служб. Благодаря этому автоматизированному процессу не потребуется усилий со стороны человека для ручного управления в сложных дорожных условиях.

#### ЛИТЕРАТУРА

1. K. Nellore and G.P. Hancke, Traffic management for emergency vehicle priority based on visual sensing, *Sensors*, vol. 16, no. 11, p.1892, 2016.
2. S. Djahel, N. Smith, S. Wang, and J. Murphy, "Reducing emergency services response time in smart cities: An advanced adaptive and fuzzy approach," in *Smart Cities Conference (ISC2)*, 2015 IEEE First International. IEEE, 2015, pp. 1–8.
3. V. Parthasarathi, M. Surya, B. Akshay, K.M. Siva, and S.K. Vasudevan, Smart control of traffic signal system using image processing, *Indian Journal of Science and Technology*, vol. 8, no. 16, 2015.
4. Исследование показало рост числа смертей в машинах скорой помощи в России [Электронный ресурс] URL: <https://russian.rt.com/russia/news/646919-issledovanie-smert-skoraya-pomosch> (дата обращения 27.02.2022).
5. CTIF — Международная ассоциация пожарно-спасательных служб [Электронный ресурс] URL: <https://www.ctif.org/ctif-international-association-fire-rescue-services> (дата обращения 27.02.2022).
6. L.C. Smeby Jr. et al., *Fire and emergency services administration: Management and leadership practices*. Jones & Bartlett Publishers, 2013.
7. M.J. Karter, *Fire loss in the United States during 2010*. National Fire Protection Association Quincy, MA, 2011.
8. J. Redmon and A. Farhadi, Yolov3: An incremental improvement, arXiv preprint arXiv:1804.02767, 2018. [Электронный ресурс] URL: <https://pjreddie.com/media/files/papers/YOLOv3.pdf> (дата обращения 27.02.2022).
9. YOLO: Real-Time Object Detection [Электронный ресурс] URL: <https://pjreddie.com/darknet/yolo/> (дата обращения 27.02.2022).
10. R. Girshick, Fast R-CNN, in *Proceedings of the IEEE international conference on computer vision*, 2015, pp. 1440–1448. [Электронный ресурс] URL: [https://www.cv-foundation.org/openaccess/content\\_iccv\\_2015/papers/Girshick\\_Fast\\_R-CNN\\_ICCV\\_2015\\_paper.pdf](https://www.cv-foundation.org/openaccess/content_iccv_2015/papers/Girshick_Fast_R-CNN_ICCV_2015_paper.pdf) (дата обращения 27.02.2022).
11. T.-Y. Lin, M. Maire, S. Belongie, J. Hays, P. Perona, D. Ramanan, P. Dollár, and C.L. Zitnick, Microsoft coco: Common objects in context, in *European conference on computer vision*. Springer, 2014, pp. 740–755.
12. C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich, Going deeper with convolutions, in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2015, pp. 1–9.
13. D.P. Kingma and J. Ba, Adam: A method for stochastic optimization, arXiv preprint arXiv:1412.6980, 2014.
14. Y. LeCun, Y. Bengio et al., Convolutional networks for images, speech, and time series, *The handbook of brain theory and neural networks*, vol. 3361, no. 10, p. 1995, 1995.
15. G. Bradski, "The OpenCV Library," *Dr. Dobb's Journal of Software Tools*, 2000.
16. J. Yosinski, J. Clune, Y. Bengio, and H. Lipson, How transferable are features in deep neural networks? in *Advances in neural information processing systems*, 2014, pp. 3320–3328.
17. K. Simonyan and A. Zisserman, Very deep convolutional networks for large-scale image recognition, arXiv preprint arXiv:1409.1556, 2014.
18. C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich, Going deeper with convolutions, in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2015, pp. 1–9.
19. F. Chollet, "Xception: Deep learning with depthwise separable convolutions," arXiv preprint, pp. 1610–02 357, 2017.