

СПЛОШНОЕ АВТОНОМНОЕ ВОЖДЕНИЕ НА ОСНОВЕ КОМПЬЮТЕРНОГО ЗРЕНИЯ И КОНВОЛЮЦИОННЫХ НЕЙРОННЫХ СЕТЕЙ

CONTINUOUS AUTONOMOUS DRIVING BASED ON COMPUTER VISION AND CONVOLUTIONAL NEURAL NETWORKS

*Wei Xiaoyu
Zheng Jingyi*

Summary. The article provides an overview of the main approaches to the organization of continuous autonomous driving. Particular attention is paid to computer vision and deep learning technologies. The logic of organizing the infrastructure for driving cars of high autonomy is considered. The paper proposes a new approach to the construction of neural networks based on structural and functional analysis. A software solution to the problem of continuous steering angle prediction for autonomous driving in Python based on convolutional neural networks is described.

Keywords: unmanned vehicle, autonomous driving, telematics, artificial intelligence, artificial neural networks, segmented images, hypergraphs, multi-agent system, rational agent, computer vision, pattern recognition, smart city, road infrastructure architecture, V2V-, V2I-, V2X-interaction, odometry, telematics, HAD systems, CAN bus.

Вэй Сяоюй

*МГТУ им. Н.Э. Баумана
569006420@mail.ru*

Чжэн Цзини

*МГТУ им. Н.Э. Баумана
sofazjy@gmail.com*

Аннотация. В статье дается обзор основных подходов к организации сплошного автономного вождения. Особое внимание уделяется технологиям компьютерного зрения и глубокого обучения. Рассмотрена логика организации инфраструктуры вождения автомобилей высокой автономности. В работе предложен новый подход к построению нейронных сетей, на базе структурно-функционального анализа. Описывается программное решение задачи непрерывного прогнозирования угла поворота руля для автономного вождения на языке Python на основе конволюционных нейронных сетей.

Ключевые слова: беспилотное транспортное средство, автономное вождение, телематические средства, искусственный интеллект, искусственные нейронные сети, сегментированные изображения, гиперграфы, многоагентная система, рациональный агент, компьютерное зрение, распознавание образов, «умный город», архитектура дорожной инфраструктуры, V2V-, V2I-, V2X-взаимодействие, одометрия, телематика, HAD-системы, CAN-шина.

Автопилоты (в авиационном смысле) происходят не от слова «автомобиль», а от слова «автоматический» — они существуют уже несколько десятилетий и прекрасно себя зарекомендовали на практике. Прежде всего они избавляют пилотов от лишней нагрузки и рутинных операций, предоставляя им дополнительное время на восстановление сил и внимания и снижая тем самым потенциальную аварийность. Дроны и квадрокоптеры как материальное воплощение автопилотов возникли немного позже, но их эффективность (прежде всего в военном деле) поражает воображение. Беспилотные автомобили пока не получили такого распространения, но этот час не за горами и его с нетерпением ждут сотни миллионов автолюбителей и профессионалов по всему свету.

Теоретически беспилотным можно сделать любой автомобиль, оснащенный CAN-шиной [1,2]. Для этого к CAN-шине подключается система беспилотного

управления рулем, тормозами и акселератором, а также сенсоры, в результате чего машина оказывается способна ездить самостоятельно. В простейшем случае потребуется добавить только бортовой компьютер, видеокамеру и дисплей в кабину водителя (на всякий случай). По этой схеме российская фирма Cognitive Pilot модернизирует комбайны для работы в беспилотном режиме [2, 3]. Беспилотный автомобиль такого класса вполне может эксплуатироваться в закрытых зонах, в сельском хозяйстве, на железных дорогах.

При переходе к интенсивному беспилотному и безостановочному дорожному движению требования к аппаратуре возрастают многократно. Поскольку для этого нужна автоматическая отказоустойчивая система, способная работать в любых условиях вождения (то есть в режиме реального времени) с экстремальными требованиями к быстродействию при крайне низких показателях допустимых ошибок. Автономные системы такого класса являются омниканальными и образуют

Таблица 1. Основные сенсоры беспилотных автомобилей

Тип устройства	Назначение	Количество	Примечание
Видеокамера	Распознавание объектов по видео-изображениям		
Радар	Угол обзора 90–120 градусов Замер расстояний до объектов	2–4	Кроме беспилотников Tesla [6]
Лидар	Угол обзора 360 градусов Замер расстояний до объектов	2–5	Кроме беспилотников Tesla [6]
Микрофон	Звуки: сирены, скорая помощь	1	Только в решениях фирмы Waymo [1]

Таблица 2. Основные датчики беспилотных автомобилей

Обозначения	Назначение
IMU	Блок инерциальных датчиков
GPS	Одометрия
RTK	Телематика

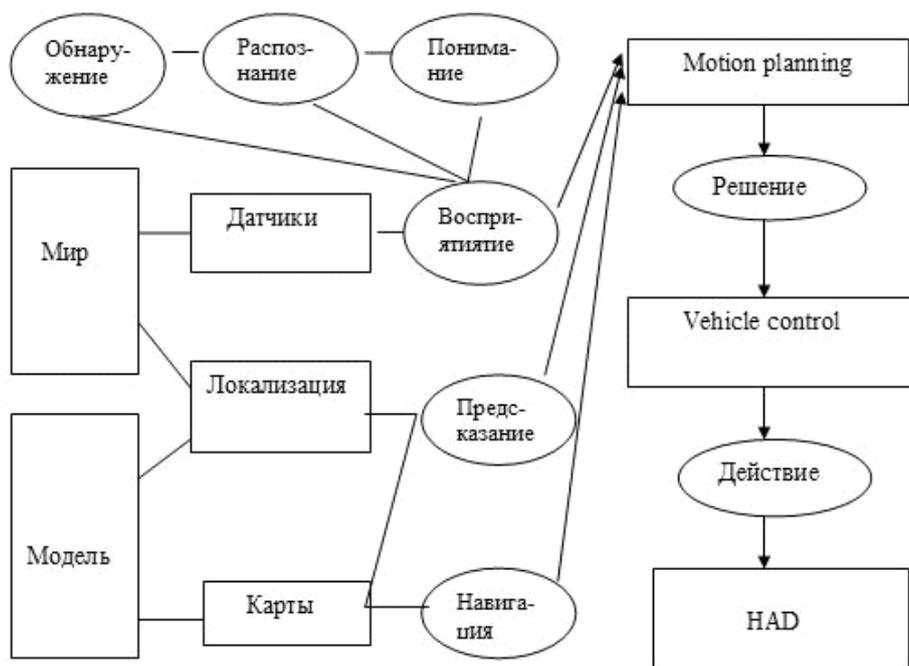


Рис. 1. Логика организации инфраструктуры вождения автомобилей высокой автономности

и потребляют огромные объемы данных [4]. Для их обработки требуются колоссальные вычислительные мощности [5].

Принцип действия наземных беспилотников представлен на рисунке 1.

Беспилотный автомобиль взаимодействует с окружающей действительностью двумя способами — непосредственно — через сенсоры (perceptions) и опосредованно — через сопоставление данных своих

датчиков с параметрами моделей мира (прежде всего картографических и математических).

Характеристики основных сенсоров беспилотных автомобилей представлены в таблице 1.

Характеристики основных датчиков, применяемых в беспилотных автомобилях, представлены в таблице 2.

Сенсоры беспилотного автомобиля собирают информацию об окружающем мире и передают ее в ком-

Таблица 3. Основные интерфейсы «Умного города»

Обозначение	Класс	Расшифровка
V2V	vehicle-to-vehicle	автомобили обмениваются информацией напрямую друг с другом
V2I	vehicle-to-infrastructure	автомобили обмениваются информацией с дорожной инфраструктурой
V2P	vehicle-to-pedestrian	автомобили обмениваются информацией с пешеходами (например, автомобиль видит смартфон пешехода и понимает, что тут находится человек)
V2H	vehicle-to-home	автомобили обмениваются информацией с квартирами
V2C	vehicle-to-cloud	автомобили обмениваются информацией с облаками

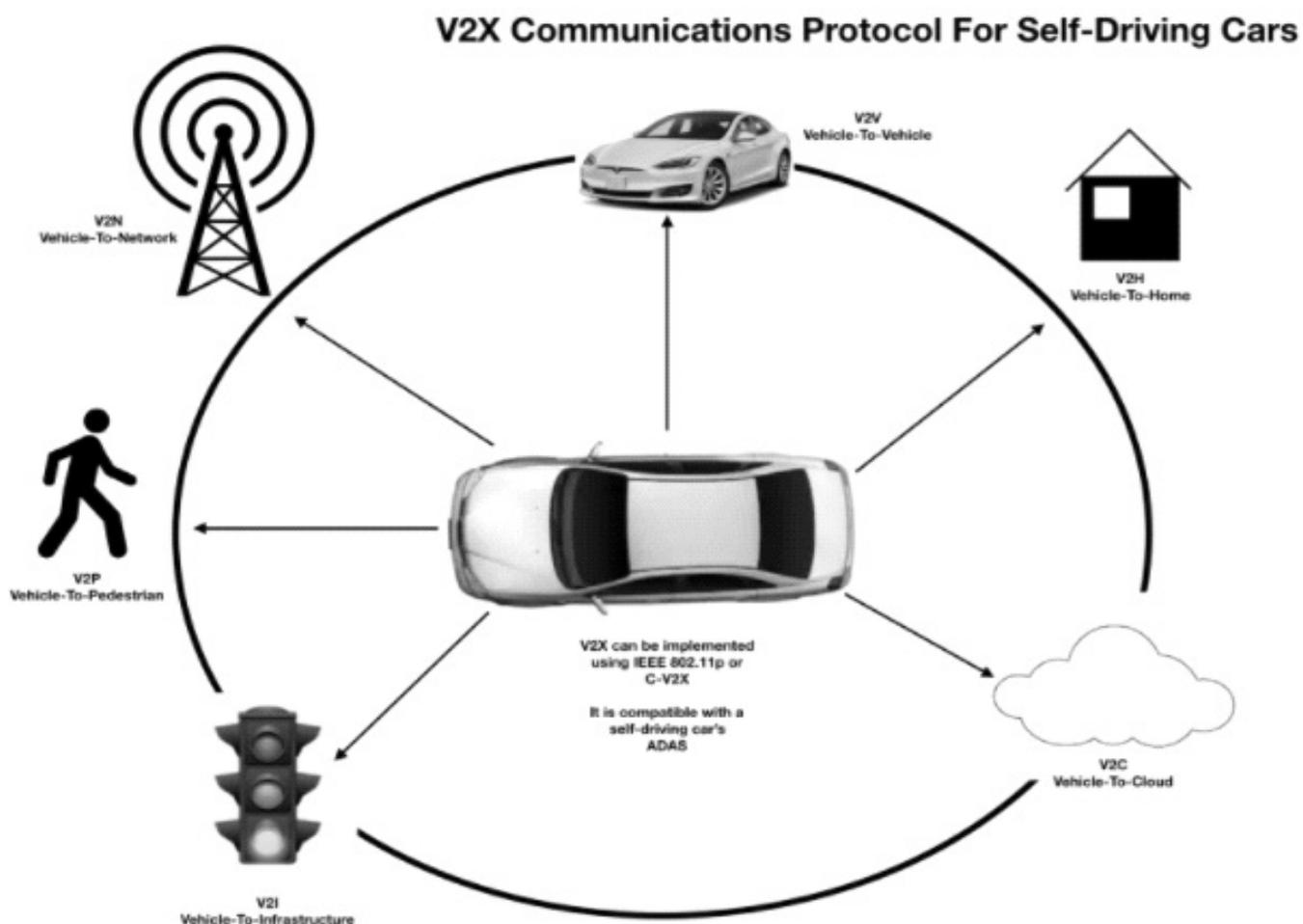


Рис. 2. Архитектура дорожной инфраструктуры «умного города» [6]

понент системы управления (motion planning), который планирует действия на основе этой информации. Туда же поступают данные карт и локализация по данным датчиков. При этом достигается сантиметровая точность навигации (в отличие от метровой, доступной в GPS). А затем motion planning передаёт принятые решения компоненту «управление автомобилем» (vehicle

control), который отдаёт команды исполнительным механизмам автомобиля и направляет его по заданной траектории.

HAD-система представляет собой многоагентную транспортную систему, в которую могут входить различные виды агентов. Простейшими рефлексными

агентами могут быть, например, активные технические средства организации дорожного движения — «умное» периферийное оборудование автоматизированной системы управления дорожным движением (знаки, светофоры и пр.).

Беспилотные автомобили в такой системе реализуют взаимодействие с любым другим объектом (автомобилями, дорожной инфраструктурой, пешеходами, пассажирами и т.д.) посредством технологий беспроводной связи. Такое взаимодействие называется V2X-взаимодействием [7,8].

Основные интерфейсы, задействованные в HAD-и ADAS-технологиях, представлены в таблице 3 [3,9].

Архитектура дорожной инфраструктуры «умного города» представлена на рисунке 2.

Реализация подсистемы vehicle control базируется на механизмах Computer Vision, технологической основой которого является CNN (Convolutional Neural Networks) и глубокое обучение (Deep Learning) для анализа и распознавания нерегулярных, неоднородных или сложно структурированных данных (нерегулярные массивы; данные различного размера и различной природы; деревья; скелеты; графовые структуры; 2D-, 3D- и ND-облака точек; триангулированные поверхности; сегментированные изображения; гиперграфы, описывающие марковские модели т.п.).

Научное сообщество сейчас находится на этапе порождения эвристик, практика намного опережает теорию. Научные группы и прикладные разработчики используют глубокое обучение и глубокие конволюционные нейронные сети, получают при этом state-of-the-art результаты, но никакой теории построения и обучения CNN не существует. Даже там, где попытки построить структурные CNN предпринимаются, получающиеся архитектуры оказываются весьма экзотичными, поскольку на сегодняшний день не существует единых принципов, позволяющих применять CNN к нерегулярным данным напрямую. В научной литературе справедливо отмечается, что никаких предпосылок к созданию такой теории пока нет и даже неясен математический аппарат, который мог бы стать базой такой теории [10, с. 887].

Одна из первых попыток использовать структурно сложные CNN была предпринята в работе [13]. В ней предлагалось использовать в конволюционных слоях вместо конволюционных фильтров простые классические нейронные сети, состоящие из нескольких слоёв. Дальнейшим развитием такого подхода стало появление «модульных» сетей (GoogleNet [11], SqueezeNet [14],

Inception v4 [15]). Например, в работе [11] был предложен т.н. Inception module, представляющий из себя набор параллельно выполняемых конволюционных слоёв, слоёв пулинга с различными параметрами и слоя конкатенации, объединяющего их результаты, а CNN состоит из последовательности таких «модулей».

Другой подход к формированию сложной архитектуры сети был развит в работе [16], где CNN предлагается формировать как последовательность блоков, каждый из которых построен по принципу фрактальной структуры из классических конволюционных слоёв.

Во всех этих случаях сложная архитектура CNN была связана со сложной логикой обработки данных, а не с их собственной пространственной структурой. Между тем графовое представление является естественным представлением первичных (анализируемых) данных для огромного количества задач, начиная от задач анализа химических элементов ([17,18,19]), заканчивая задачами понимания текста [20] и описания сцены для решения задач обучения с подкреплением [21].

В работе [22] предлагается представить облака точек в воксельном виде (Voxel networks) с последующей их обработкой средствами CNN с трёхмерными конволюционными фильтрами.

Аналогичный подход используется и при обработке видеопоследовательностей в работе [23].

Основной проблемой такого подхода является очень большая размерность входных данных, что ведёт к неприемлемой скорости реакции таких CNN и большим потребностям в памяти. Дальнейшим развитием такого подхода явилось использование иерархических решёток, например, в [24,25,26]. В [27] была предложена модификация данного подхода, при которой отклики фильтров считались только по активным (непустым) вокселям, что в теории позволяет повысить скорость.

Несмотря на существенные успехи этого направления, проблема большого объёма входных данных и низкой скорости трёхмерных конволюций является главной проблемой воксельного подхода.

Нами предлагается иной подход на базе структурно-функционального анализа и синтеза CNN, лишенный перечисленных недостатков.

Для его реализации необходимо системно рассмотреть типы и структуры данных с точки зрения того, каковы их структурные (геометрические и топологические) особенности (внутренние метрики, симметрии,

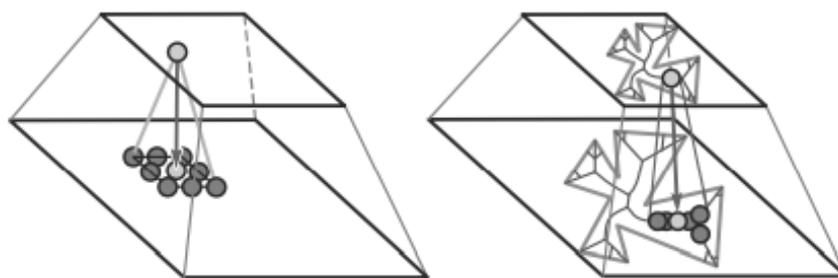


Рис. 3. Реализация функции выборки локального многообразия (окрестности) для регулярных массивов (слева) и нерегулярных структур (справа)

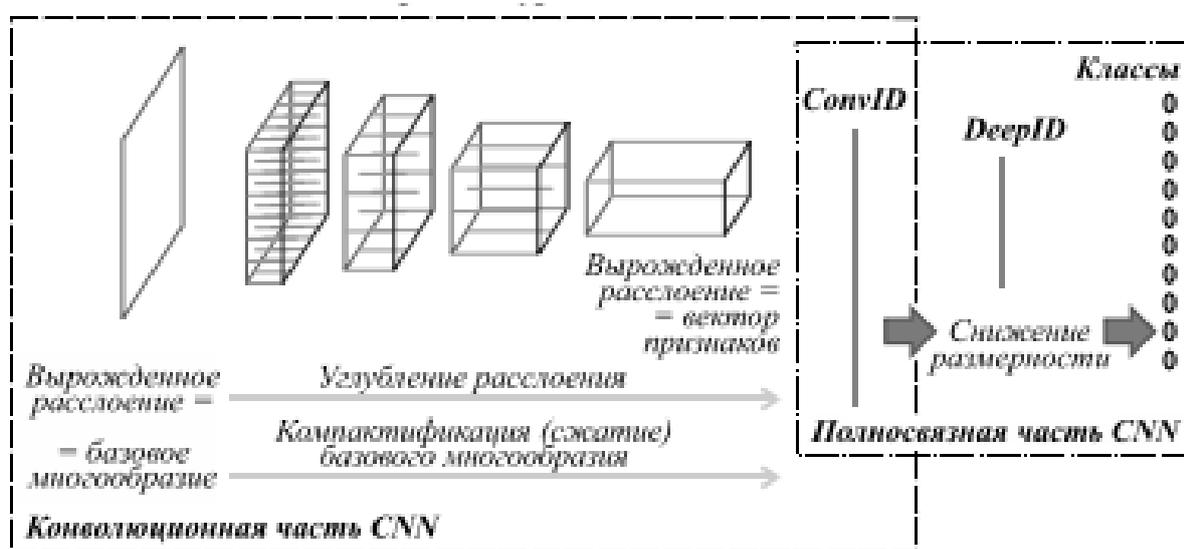


Рис. 4. Расслоение базового многообразия

связность, фрактальность/ масштабируемость и т.п.). После этого можно предложить различные способы заполнения соответствующих ячеек структурной таблицы.

Рассмотрим различия в реализации функции выборки локального многообразия (окрестности) для регулярных и нерегулярных структур. На рисунке 3 эти различия проиллюстрированы для случая скелетного описания формы фигуры.

Предложенная методика позволяет создавать, конфигурировать и обучать CNN, в том числе обладающие подобной нетривиальной и переменной (зависящей от самого анализируемого изображения или сигнала) структурой (рисунок 4).

Для практической реализации такой методики синтеза CNN предлагается построить и использовать алгоритм, представленный на рисунке 5.

Предлагается программное решение задачи непрерывного прогнозирования угла поворота руля для автономного вождения на языке Python.

Задача состоит в том, чтобы, используя CNN, сделать так, чтобы транспортное средство управляло собой само, избегая препятствий, аварийных ситуаций и нарушения правил дорожного движения. Выходными параметрами модели является ускорение (торможение) и угол поворота руля транспортного средства.

Каждое мгновение модельного времени нейронная сеть решает, как изменить свою скорость управляемого транспортного средства и направление его движения в зависимости от состояния окружающей среды. Нейронная сеть видит только то, что видел бы человек, если бы находился за рулем, включая показания приборов и датчиков на панели управления, только реагирует несколько быстрее.

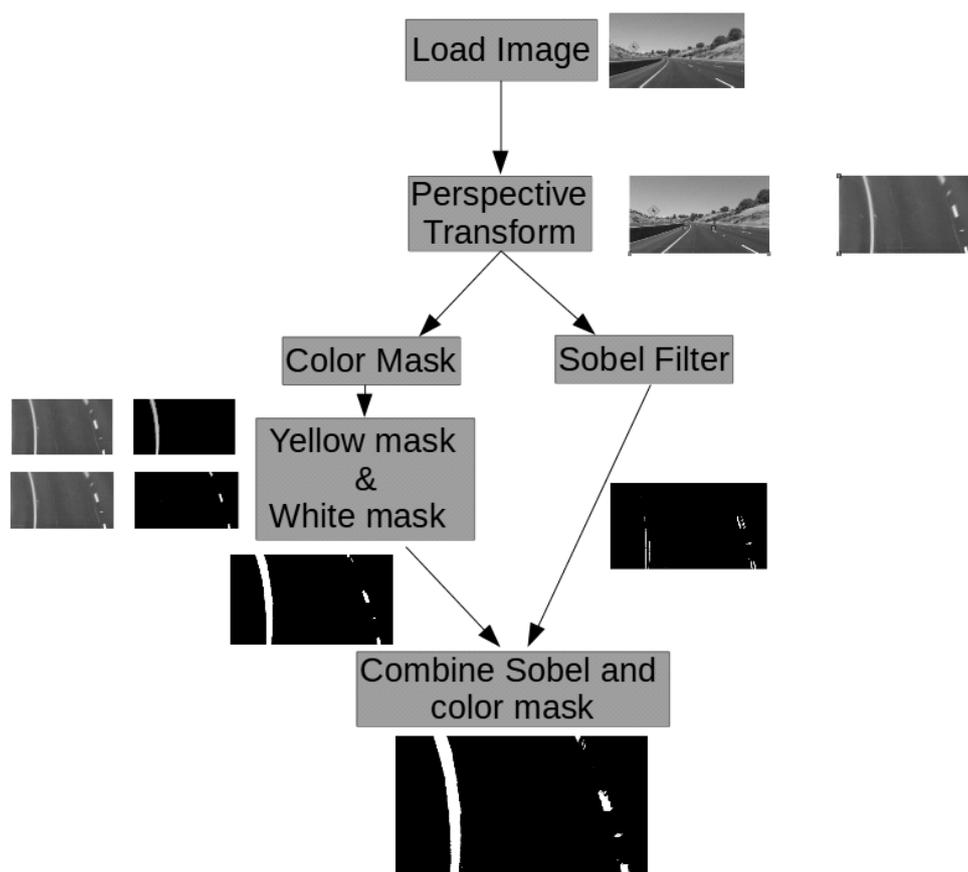


Рис. 5. Алгоритм синтеза CNN

Ценность предлагаемого в данной статье подхода, как нам представляется, заключается в его системности и последовательности. Поскольку научное сообщество сейчас находится на этапе порождения эвристик, мы

предлагаем искать и конструировать эти эвристики на регулярной основе и в рамках некоторой систематической терминологии, пригодной для описания широкого класса архитектур CNN.

ЛИТЕРАТУРА

1. Каланов Глеб. Как устроены беспилотные автомобили и кто их делает // 15.02.2021 [Электронный ресурс] — Режим доступа: <https://mag.auto.ru/article/selfdrivingsetup/>. (дата обращения: 02.02.2022).
2. Черногоров Андрей. Как мы сделали автопилот для сельхозкомбайна на видеоаналитике первыми в мире // 14 апреля 2020 [Электронный ресурс] — Режим доступа: <https://habr.com/ru/company/cognitiverobot/blog/497098/>. (дата обращения: 02.02.2022).
3. Савицкий Геннадий. Взгляд на ADAS изнутри: когда поедет робот? // 21.10.2020 [Электронный ресурс] — Режим доступа: <https://habr.com/ru/company/cognitiverobot/blog/524236/>. (дата обращения: 02.02.2022).
4. Блог компании Hewlett Packard Enterprise. Разработка инфраструктуры вождения автомобилей высокой автономности (HAD) // 12.03.2021 [Электронный ресурс] — Режим доступа: <https://habr.com/ru/company/hpe/blog/546616/>. (дата обращения: 02.02.2022).
5. Автопилот. Беспилотный автомобиль. // 07.07.2021 [Электронный ресурс] — Режим доступа: <https://www.tadviser.ru/index.php/%D0%A1%D1%82%D0%B0%D1%82%D1%8C%D1%8F...> (дата обращения: 02.02.2022).
6. Нигматулин Руслан. Беспилотные автомобили для начинающих. // 01.12.2018 [Электронный ресурс] — Режим доступа: <https://habr.com/ru/post/431758/>. (дата обращения: 02.02.2022).
7. Душкин Р.В. I2V-взаимодействие — основа беспилотного движения. // Дорожная держава. — 2018. — № 85.
8. Душкин Р.В., Андронов М.Г. Использование методов искусственного интеллекта для организации беспилотного движения // Тренды и управление. — 2020. — № 1 [Электронный ресурс] — Режим доступа: <https://cyberleninka.ru/article/n/ispolzovanie-metodov-iskusstvennogo-intellekta-dlya-organizatsii-bespilotnogo-dvizheniya>. (дата обращения: 02.02.2022).

9. Пичугин Артем. Распознавание дорожных знаков с помощью CNN: Spatial Transformer Networks [Электронный ресурс] Режим доступа: <https://habr.com/ru/company/newprolab/blog/339484/>. (дата обращения: 02.02.2022).
10. Визильтер Ю.В. и др. Структурно-функциональный анализ и синтез глубоких конволюционных нейронных сетей. // Компьютерная оптика. 2019. № 5.
11. Krizhevsky, A. ImageNet classification with deep convolutional neural networks / A. Krizhevsky, I. Sutskever, G. Hinton // Proceedings of the 25th International Conference on Neural Information Processing Systems. — 2012. — Vol. 1. — P. 1106–1114.
12. Szegedy, C. Going deeper with convolutions / C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, A. Rabinovich // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. — 2015. — P. 1–9.
13. Wu, X. Learning robust deep face representation / X. Wu. — 2015. [Электронный ресурс] — Режим доступа: <https://arxiv.org/pdf/1507.04844.pdf> (дата обращения: 02.02.2022).
14. Lin, M. Network in network / M. Lin, Q. Chen, S. Yan. — 2013. [Электронный ресурс] — Режим доступа: <https://arxiv.org/pdf/1312.4400.pdf> (дата обращения: 02.02.2022).
15. Forrest, N. SqueezeNet: AlexNet-level accuracy with 50x fewer parameters and <0.5MB model size / N. Forrest, H. Song, W.M. Matthew, A. Khalid, J.D. William, K. Kurt. — 2016. [Электронный ресурс] — Режим доступа: <https://arxiv.org/abs/1602.07360> (дата обращения: 02.02.2022).
16. Szegedy, C. Inception-v4, Inception-ResNet and the Impact of Residual Connections on Learning / C. Szegedy, S. Ioffe, V. Vanhoucke. — 2016. [Электронный ресурс] — Режим доступа: <https://arxiv.org/abs/1602.07261> (дата обращения: 02.02.2022).
17. Larsson, G. Fractalnet: ultra-deep neural networks without residuals / G. Larsson, M. Maire, G. Shakhnarovich. — 2017. [Электронный ресурс] — Режим доступа: <https://arxiv.org/abs/1605.07648> (дата обращения: 02.02.2022).
18. Duvenaud, D. Convolutional networks on graphs for learning molecular fingerprints / D. Duvenaud, D. Maclaurin, J. Aguilera-Iparraguirre, R. Gomez-Bombarelli, T. Hirzel, A. Aspuru-Guzik, R.P. Adams. — 2015. [Электронный ресурс] — Режим доступа: <https://arxiv.org/pdf/1509.09292.pdf> (дата обращения: 02.02.2022).
19. De Cao, N. MolGAN: An implicit generative model for small molecular graphs / N. De Cao, T. Kipf. — 2018. [Электронный ресурс] — Режим доступа: <https://arxiv.org/pdf/1805.11973.pdf> (дата обращения: 02.02.2022).
20. Gomes, J. Atomic convolutional networks for predicting protein-ligand binding affinity / J. Gomes, B. Ramsundar, E.N. Feinberg, V.S. Pande. — 2017. [Электронный ресурс] — Режим доступа: <https://arxiv.org/pdf/1703.10603.pdf> (request date 14.09.2018).
21. Yao, L. Graph convolutional networks for text classification / L. Yao, C. Mao, Y. Luo. — 2018. [Электронный ресурс] — Режим доступа: <https://arxiv.org/pdf/1809.05679.pdf> (дата обращения: 02.02.2022).
22. Xiong, W. DeepPath: A reinforcement learning method for knowledge graph reasoning / W. Xiong, T. Hoang, W.Y. Wang — 2018. [Электронный ресурс] — Режим доступа: <https://arxiv.org/pdf/1707.06690.pdf> (дата обращения: 02.02.2022).
23. Maturana, D. VoxNet: A 3D convolutional neural network for real-time object recognition / D. Maturana, S. Scherer // IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS). — 2015. — P. 922–928. — DOI:10.1109/IROS.2015.7353481.
24. Tran, D. Learning spatiotemporal features with 3D convolutional networks / D. Tran, L. Bourdev, R. Fergus // IEEE International Conference on Computer Vision (ICCV). — 2015. — P. 4489–4497. — DOI: 10.1109/ICCV.2015.510.
25. Riegler, G. OctNet: Learning deep 3D representations at high resolutions / G. Riegler, A.O. Ulusoy, A. Geiger // CVPR. — 2017. — P. 6620–6629.
26. Riegler, G. OctNetFusion: Learning depth fusion from data. / G. Riegler, A.O. Ulusoy, H. Bischof, A. Geiger // 2017 International Conference on 3D Vision (3DV). — 2017. — P. 57–66.
27. Klovov, R. Escape from cells: Deep KdNetworks for the recognition of 3D point cloud models / R. Klovov, V. Lempitsky // ICCV. — 2017 — P. 863–872.