

Разработка интеллектуального программного модуля распознавания изображений для звуковых очков

*А.Р. Вильданов, А.Р. Тухфатуллин, М. Л. Шустрова,
Н. А. Староверова*

Казанский национальный исследовательский технологический университет, г. Казань

Аннотация: В статье проведен обзор носимых устройств, повышающих качество распознавания пользователем объектов окружающего мира, раскрыт процесс разработки интеллектуального модуля распознавания объектов городской инфраструктуры для «звуковых очков» - устройства, предназначенного для незрячих пользователей. Охарактеризован функционал разрабатываемого модуля, его архитектура, описан процесс и результаты обучения нейросети для распознавания объектов, а также приведен алгоритм идентификации цвета светофоров. Приведены результаты тестирования разработанного программного модуля. Программный модуль, созданный в настоящей работе, представляют собой неотъемлемую составляющую полноценного функционирующего устройства «звуковые очки», применение которого позволит людям с ограниченными зрительными возможностями получить информацию о происходящих событиях и окружающей обстановке и ориентироваться даже в незнакомой среде.

Ключевые слова: компьютерное зрение, нейронная сеть, умные очки, звуковые очки, инвалиды по зрению, YOLOv5, распознавание объектов.

По данным [1], на сегодняшний день в России проживает более 2,5 миллиона человек с ослабленным зрением, 300 000 из которых полностью слепы. По крайней мере четверть из них являются активными пользователями смартфонов [2-4], и во многих регионах России мобильные приложения официально включены в перечень реабилитационных средств для слабовидящих. Программное обеспечение для чтения с экрана и специальные разговорные устройства позволяют людям с нарушением зрения использовать компьютеры, сотовые телефоны и прочее оборудование [5, 6].

В настоящее время в России и за рубежом существует достаточно много разработок «умных очков» с расширенным функционалом, однако подобные разработки зачастую только расширяют возможности использования устройства пользователями с нормальным зрением.

Количество разработок для пользователей с ограничением по зрению довольно ограничено. В работе [5] представлена система, преобразующая изображения, полученные с помощью камеры, в звуки. Аппаратная реализация предполагает использование возможности портативного устройства, такого, как КПК или мобильный телефон, или специальной встроенной системы с микроконтроллером, а изображения принимаются через небольшую веб-камеру. Портативное устройство с искусственным интеллектом Элада Моисеева [3], в свою очередь, крепится к очкам человека и может распознавать текст, денежные номиналы и лица людей. Очки E-sight [7] снабжены специальной камерой, способной увеличивать изображение в 14 раз и записывать все происходящее. Полученная информация преобразовывается в сигналы, понятные слабовидящему человеку. Разработка ученого Оксфордского университета Стивена Хикса – очки Smart [8] с инфракрасным проектором и миникамерами. ИК - проектор способен измерять расстояние до объектов, информация подается человеку в доступной для него форме. Разработка иерусалимского ученого Амира Амеди [9], в свою очередь, работает по звуковому принципу и применима в том числе пользователями с полностью утраченным зрением. Камера очков преобразует визуальный сигнал в звуковой.

Как можно видеть из проведенного анализа, принципы функционирования устройств для людей со сниженным зрением несколько варьируются, при этом перечень технологий, применимых для людей с полным отсутствием зрения, фактически ограничен в настоящее время «звуковыми очками», которые можно считать наиболее применимыми для данной категории населения устройствами.

Актуальность настоящей работы обусловлена тем, что разработки функциональных звуковых очков ведутся преимущественно за пределами Российской Федерации, подобные устройства нерусифицированы и

практически недоступны жителям нашей страны. В то же время, устройство, для которого разрабатывается данное программное обеспечение, способно обеспечить нуждающихся в нем людей возможностью беспрепятственного и безопасного перемещения в городской среде.

Одним из ключевых компонентов умных звуковых очков, преобразующих визуальную информацию в звуковой сигнал, является их программное обеспечение, реализующее функции захвата изображения и его интерпретации. Целью настоящей работы является разработка интеллектуального программного модуля для «умных» очков для распознавания ключевых объектов городской инфраструктуры.

В рамках концепции разрабатываемого программного модуля, основной его функцией является непрерывный анализ поступающего видеопотока и определение объектов, потенциально являющихся препятствием для пользователя. Алгоритм работы программного обеспечения предполагает, что после распознавания большинства объектов нейронной сетью информация о названии объекта, его направление и расположение относительно пользователя будут переводиться на звуковую дорожку с помощью специализированного модуля программного обеспечения. В случае распознавания светофоров необходимо предусмотреть не только определение вида объекта, но и идентификацию его цвета.

Реализация программного модуля включает формирование датасетов и обучение нейронной сети для распознавания пяти объектов: люди, машины, столбы, деревья и светофоры, а также разработку алгоритма определения цвета светофора [10]. Нейронная сеть должна осуществлять детектирование поступающих изображений в режиме реального времени. Полученные результаты оцениваются по уровню приоритетности [11, 12]. Все объекты, за исключением светофоров, сразу направляются на выход модуля. Если в кадре

находится светофор, данный фрагмент изображения направляется в программный модуль распознавания цвета светофора.

Программное обеспечение для распознавания объектов реализовано на языке python с использованием библиотек PyTorch, NumPy, OpenCV, Time, Pyttsx3. Поскольку анализируемый видеопоток содержит большое количество движущихся объектов, для снижения нагрузки на микропроцессор при его обработке, применен метод вычитания фона, как наиболее эффективный для подобных задач [13]. Для обучения нейронной сети, распознающей типы объектов, использованы обучающие выборки изображений людей, столбов, машин и деревьев, созданных авторами работы и аккумулярованными из Интернет – источников [14]. Общий объем обучающих выборок составляет 6813 изображений, где на каждый объект приходится около 1300 изображений. Разметка изображения реализована с помощью ресурса roboflow.com (рис.1).

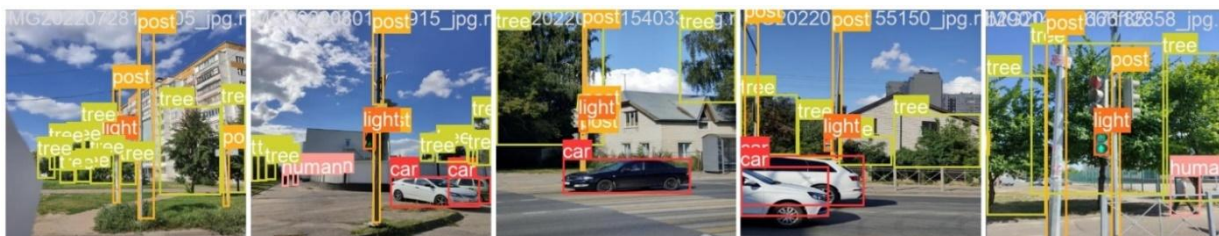


Рис. 1. – Пример реализованной разметки изображений

Для реализации процесса обучения, использована нейронная сеть YOLOv5s, преимуществами которой являются ее свободное распространение и доступность в формате модуля для Python3, ее принцип детектирования изображений, существенно увеличивающий скорость обработки визуального сигнала, а также возможность применения на небольших по мощности компьютерах, что актуально в рамках решаемой задачи [15].

Решаемая задача относится к мультиклассовой классификации. Поэтому для оценки качества работы нейронной сети целесообразно использовать показатели точности (precision) и полноты распознавания

нейросети (recall). Результат распознавания тестовых выборок в виде соотношения «точность – полнота» для каждого класса объектов приведен на рис. 2а, матрица ошибок (confusion matrix) – на рис. 2б, где по оси ординат представлены вероятности, с которыми нейросеть распознала объекты пяти рассматриваемых классов.

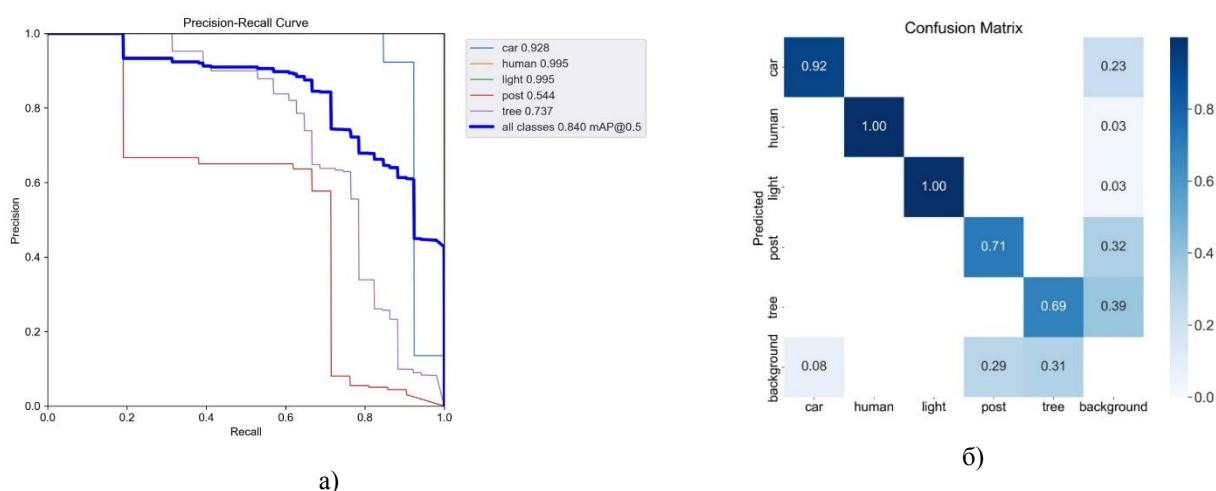


Рис. 2. – Качество распознавания: а - precision-recall, б - матрица ошибок

Наилучшие результаты распознавания обученной нейронной сети получены у групп объектов «светофоры» и «люди», со средней точностью распознавания 99,5%. Для категории «машины» 92% объектов на исходном изображении было распознано верно и 8% автомобилей было принято нейросетью за фон. Основной причиной подобных расхождений является удаленность нераспознанных автомобилей, и по мере приближения объекта он будет адекватно распознан. Согласно результатам тестирования обученной в рамках данной работы нейросети, возникли некоторые сложности в распознавании столбов и деревьев (71% и 69% корректно распознанных объектов соответственно). Результаты по данным категориям объектов указывают на целесообразность дополнительных исследований методов распознавания подобных объектов и способах доработки продукта.

Захват видео осуществляется с помощью библиотеки OpenCV и команды `cv2.VideoCapture()`, после чего изображение поступает в нейронную сеть, где определяются координаты пешеходного светофора, совершается его обрезка, определяется и выводится доминирующий на изображении цвет (красный или зелёный). В случае, если количество цветowych пикселей для обеих масок не превышает 1000 пикселей в течение 3 секунд, происходит генерация сообщения "Светофор сломан".

Основным результатом настоящей работы является функционирующий программный модуль, производящий захват изображения с видеопотока, его разметку и идентификацию объектов, находящихся в кадре (рис.3).



Рис. 3. – Процесс функционирования нейронной сети

Тестирование функционирования нейросети показало вероятности корректного определения объектов на уровнях 100% для людей и светофоров, 99% - для автомобилей, 71% для деревьев и 69% для столбов.

Заключение

Настоящая работа раскрывает вопрос разработки интеллектуального модуля распознавания объектов городской инфраструктуры, являющегося частью программного обеспечения звуковых очков, предназначенных для незрячих пользователей. Полученные в работе результаты можно в целом назвать положительными, однако относительно невысокое качество распознавания обученной в рамках настоящей работы нейросетью деревьев и фонарных столбов указывает на необходимость дополнительных исследований методов распознавания указанных объектов. Программный модуль, созданный в настоящей работе, представляет собой неотъемлемую

составляющую полноценного функционирующего устройства «звуковые очки», применение которого позволит людям с ограниченными зрительными возможностями получить информацию о происходящих событиях и окружающей обстановке и ориентироваться даже в незнакомой среде. В целом, исследование доказывает перспективность дальнейших разработок в данной области.

Литература

1. Умные цифровые устройства для незрячих // URL: evercare.ru/news/umnye-cifrovye-ustroystva-dlya-nezryachikh (дата обращения: 13.10.2022).
 2. Зиятдинова, А. Староверова Н. А. Аналитический обзор и сравнение возможностей операционных систем для мобильных устройств // Фундаментальные исследования, 2015. №9-2. С. 227-231
 3. Moisseiev E, Mannis MJ. Evaluation of a Portable Artificial Vision Device Among Patients With Low Vision. JAMA Ophthalmol. 2016 Jul 1;134(7):748-52. doi: 10.1001/jamaophthalmol.2016.1000. PMID: 27148909.
 4. Корнатко Р. Ч. Новые средства спутниковой навигации для людей с проблемами зрения // Материалы XIX Международной научно-практической конференции по вопросам социальной политики. Выпуск 2. – С.-Пб.: СПбГИПСР, 2021. - С. 40-50.
 5. Кесада Э. Портативная система, обеспечивающая восприятие слепыми или слабовидящими людьми окружающего пространства посредством звука или касания. Патент на изобретение № 2719025 С2 // Бюллетень, 2020. №11. Стр. 1-19.
 6. Вилаират В., Эден И., Тукрал В., Нистер Д., Прадип В Невизуальная обратная связь визуального изменения в способе и устройстве отслеживания взгляда. Патент на изобретение №2691589 С2 //Бюллетень, 2019. №17. Стр 1-26.
 7. eSight 3: AR-очки для людей с плохим зрением // planetvrar.com URL:planetvrar.com/esight-3-ar-ochki/ (дата обращения: 23.11.2022).
-

8. Николайкова А.Н., Ковалёва Д.О. Особенности работы «умных очков» – Google Glass // Актуальные проблемы энергетики. Материалы 74-й научно-технической конференции студентов и аспирантов. М. Минск: Белорусский национальный технический университет, 2018. - С. 566-567.

9. Дмитренко А. В., Губницкая Ю. С. Анализ и исследование мультимедийных устройств для людей с ограниченными возможностями // Полиграфические, мультимедийные и web-технологии: материалы Молодежной школы-семинара 2-й Международной научно-технической конференции. Харьков: ООО «Типография Мадрид», 2017. - С. 41-44.

10. Артемьев, И.С., Лебедев, А.И., Долгий, А.И., Хатламаджиян, А.Е., Меерович, В.Д. Метод блочного оптического распознавания инвентарных номеров железнодорожных подвижных единиц на основе комитетной нейроиммунной модели классификации // Инженерный вестник Дона, 2014, №1. URL: ivdon.ru/ru/magazine/archive/n1y2014/2259

11. Staroverova N. A, Shustrova M.L., Staroverov S.A., Dykman L.A. Development of a Neurocomputer Modular Information System for Cancerous Diseases Diagnostics in Animals // Herald of the Bauman Moscow State Technical University. Series Instrument Engineering. 2020. №2(131). pp. 75-84.

12. Файзрахманов А.Ф., Тузанкин Д.С., Шустрова М.Л., Староверова Н.А. Машинное обучение в медицине: эволюция и перспективы // Южно-Сибирский научный вестник. 2021. № 4 (38). С. 43-49.

13. Баймухаметова А.И. Методы вычитания фона для обнаружения объекта в видеопотоке // Международный журнал гуманитарных и естественных наук. - 2022. - №7-1. - С. 74-78.

14. Галушка В.В., Фатхи В.А. Формирование обучающей выборки при использовании искусственных нейронных сетей в задачах поиска ошибок баз данных // Инженерный вестник, 2013, №12. URL: ivdon.ru/magazine/archive/n2y2013/1597

15. Hao, L., Jinguo, Zh., Xiaohu, J., Xizhou Zh., Hongsheng, L., Chun Y., Xiaohua, W., Yu Q., Xiaogang W., Wenhai W., Jifeng D. Uni-Perceiver v2: A Generalist Model for Large-Scale Vision and Vision-Language Tasks//arXiv preprint arXiv:2211.09808. – 2022.

References

1. Umnye cifrovye ustrojstva dlja nezrjachih [Smart digital devices for the blind]. URL:evercare.ru/news/umnye-cifrovye-ustrojstva-dlya-nezryachikh.
2. Zijatdinova, A., Staroverova, N.A., Fundamental'nye issledovanija. 2015. № 9-2. pp. 227-231.
3. Moisseiev E, Mannis MJ. Evaluation of a Portable Artificial Vision Device Among Patients With Low Vision. JAMA Ophthalmol. 2016 Jul 1;134(7):748-52. doi: 10.1001/jamaophthalmol.2016.1000. PMID: 27148909.
4. Kornatko R. Ch. Materialy XIX Mezhdunarodnoj nauchno-prakticheskoy konferencii po voprosam social'noj politiki. Vypusk 2. Sankt-Peterburgskij gosudarstvennyj institut psihologii i social'noj raboty [Materials of the XIX International Scientific and Practical Conference on Social Policy. Issue 2], 2021. pp. 40-50.
5. Kesada E. Portativnaya sistema, obespechivayushchaya vospriyatie slepymi ili slabovidyashchimi lyud'mi okruzhayushchego prostranstva posredstvom zvuka ili kasaniya. Patent na izobrenie № 2719025 C2 [A portable system that allows blind or visually impaired people to perceive the surrounding space through sound or touch. Patent for an invention №2719025 C2]. Byulleten', 2020. №11. pp. 1-19.
6. Vilairat V., Eden I., Tukral V., Nister D., Pradip V Nevizual'naya obratnaya svyaz' vizual'nogo izmeneniya v sposobe i ustrojstve otslezhivaniya vzglyada. Patent na izobrenie №2691589 C2 [Non-visual feedback of a visual change in the way and device of gaze tracking. Patent for an invention №2691589 C2]. Byulleten', 2019. №17. pp 1-26.



7. eSight 3: AR-ochki dlja ljudej s plohim zreniem. planetvrrar.com URL: planetvrrar.com/esight-3-ar-ochki.

8. Nikolajkova A.N., Kovaljova D.O. Aktual'nye problemy jenergetiki. Materialy 74-j nauchno-tehnicheskoy konferencii studentov i aspirantov (Actual problems of energy. Materials of the 74th Scientific and Technical Conference of Students and Postgraduates). Minsk: Belorusskij nacional'nyj tehničeskij universitet, 2018. pp. 566-567.

9. Dmitrenko A. V., Gubnickaja Ju. S. Poligraficheskie, mul'timedijnye i web-tehnologii: materialy Molodezhnoj shkoly-seminara 2-j Mezhdunarodnoj nauchno-tehnicheskoy konferencii [Printing, multimedia and web technologies: materials of the Youth School-seminar of the 2nd International Scientific and Technical Conference.]. Har'kov: OOO «Tipografija Madrid», 2017. pp. 41-44.

10. Artem'ev, I.S., Lebedev, A.I., Dolgij, A.I., Hatlamadzhijan A.E., Meerovich, V.D. Inzhenernyj vestnik Dona, 2014, №1 URL ivdon.ru/ru/magazine/archive/n1y2014/2259

11. Staroverova, N.A., Shustrova, M.L., Staroverov, S.A., Dykman, L.A. Herald of the Bauman Moscow State Technical University. Series Instrument Engineering. 2020. No 2(131). pp. 75-84.

12. Fajzrahmanov A.F., Tuzankin D.S., Shustrova M.L., Staroverova N.A. Juzhno-Sibirskij nauchnyj vestnik. 2021. № 4 (38). pp. 43-49.

13. Bajmuhametova A.I. Mezhdunarodnyj zhurnal gumanitarnyh i estestvennyh nauk, 2022. №7-1. pp. 74-78.

14. Galushka, V. V., Fathi, V.A., Inzhenernyj vestnik Dona, 2013, №12. URL: ivdon.ru/magazine/archive/n2y2013/1597



15. Hao, L., Jinguo, Zh., Xiaohu, J., Xizhou Zh., Hongsheng, L., Chun Y., Xiaohua, W., Yu Q., Xiaogang W., Wenhai W., Jifeng D. Uni-Perceiver v2: A Generalist Model for Large-Scale Vision and Vision-Language Tasks. arXiv preprint arXiv:2211.09808. 2022.