

ИСПОЛЬЗОВАНИЕ НЕЙРОННОЙ СЕТИ ДЛЯ СОЗДАНИЯ ЦИФРОВОГО ПОМОЩНИКА СЛАБОВИДЯЩИМ ЛЮДЯМ

Получено 14.07.2022 Доработано после рецензирования 18.08.2022 Принято 26.08.2022

УДК 004 JEL I19 DOI <https://doi.org/10.26425/2658-3445-2022-5-3-73-82>

Плотников Сергей Олегович

Студент, Государственный университет управления, г. Москва, Российская Федерация

ORCID: 0000-0001-9067-0363

E-mail: s112968@guu.ru

Сметанин Дмитрий Юрьевич

Студент, Государственный университет управления, г. Москва, Российская Федерация

ORCID: 0000-0002-4642-2044

E-mail: s112464@guu.ru

Басова Анжелика Валерьевна

Студент, Государственный университет управления, г. Москва, Российская Федерация

ORCID: 0000-0001-5590-3467

E-mail: s112966@guu.ru

Львутин Илья Александрович

Студент, Государственный университет управления, г. Москва, Российская Федерация

ORCID: 0000-0002-9011-1294

E-mail: s114237@guu.ru

Белоусова Мария Николаевна

Канд. экон. наук, доц. каф. информационных систем, Государственный университет управления, г. Москва, Российская Федерация

ORCID: 0000-0002-0072-5656

E-mail: maryzver@gmail.com

АННОТАЦИЯ

Опыт проводимых исследований обработки изображений наглядно демонстрирует огромную сферу для разработки новых нейронных сетей, способных помогать людям в большом спектре задач. В работе было выбрано направление, связанное с помощью людям, которые имеют проблемы со зрением. В статье рассматривается сверточная нейронная сеть модели Mask R-CNN для сегментации объектов на изображении. В процессе исследования было изучено большое количество алгоритмов способных быстро и точно обрабатывать изображения, например Faster R-CNN, который являлся наиболее результативным в 2020 г. В ходе анализа было выявлено, что использование технологии Mask R-CNN позволяет существенно увеличить эффективность выполнения поставленных задач, так как данный алгоритм является новейшей версией модели машинного обучения. В результате исследования была разработана нейронная сеть, способная определять и различать большое количество объектов на изображении. Следующим этапом планируется доработать алгоритм и использовать дополнительные средства взаимодействия с аппаратным обеспечением систем для увеличения скорости работы нейронной сети. В дальнейшем будет осуществлена интеграция полученной нейронной сети в приложение «Цифровой помощник для слепых и слабовидящих». Данное приложение гарантированно улучшит повседневную жизнь людей-инвалидов, которые испытывают определенные неудобства из-за их особенностей, и может стать основой других, более крупных, проектов связанных, например, с беспилотными устройствами, а также сервисами, работа которых напрямую строится на обработке изображений.

КЛЮЧЕВЫЕ СЛОВА

Цифровой помощник, машинное обучение, нейронные сети, архитектура Mask R-CNN, сверточная сеть, алгоритм поиска, сегментация, классификация, селективный поиск, обработка изображений, проблемы со зрением, нарушение зрения, помощь инвалидам, электронный помощник

ДЛЯ ЦИТИРОВАНИЯ

Плотников С.О., Сметанин Д.Ю., Басова А.В., Львутин И.А., Белоусова М.Н. Использование нейронной сети для создания цифрового помощника слабовидящим людям//E-Management. 2022. Т. 5, № 3. С. 73–82.

© Плотников С.О., Сметанин Д.Ю., Басова А.В., Львутин И.А., Белоусова М.Н., 2022.

Статья доступна по лицензии Creative Commons «Attribution» («Атрибуция») 4.0. всемирная (<http://creativecommons.org/licenses/by/4.0/>).



USE OF A NEURAL NETWORK IN CREATING A DIGITAL ASSISTANT FOR BLIND AND VISUALLY IMPAIRED PEOPLE

Received 14.07.2022 Revised 18.08.2022 Accepted 26.08.2022

Sergey O. Plotnikov

Student, State University of Management, Moscow, Russia

ORCID: 0000-0001-9067-0363

E-mail: s112968@guu.ru

Dmitry Yu. Smetanin

Student, State University of Management, Moscow, Russia

ORCID: 0000-0002-4642-2044

E-mail: s112464@guu.ru

Angelika V. Basova

Student, State University of Management, Moscow, Russia

ORCID: 0000-0001-5590-3467

E-mail: s112966@guu.ru

Ilya A. Lvutin

Student, State University of Management, Moscow, Russia

ORCID: 0000-0002-9011-1294

E-mail: s114237@guu.ru

Maria N. Belousova

Cand. Sci. (Econ.), Assoc. Prof. at the Information Systems Department, State University of Management, Moscow, Russia

ORCID: 0000-0002-0072-5656

E-mail: maryzver@gmail.com

ABSTRACT

The experience of ongoing research in image processing clearly demonstrates the huge scope for the development of new neural networks that can help people in a wide range of tasks. The authors chose the direction of work related to helping people who have vision problems. The article considers a convolutional neural network of the Mask R-CNN model for segmenting objects in an image. During the research the authors study a large number of algorithms that can quickly and accurately process images, such as Faster R-CNN, which was the most efficient in 2020. During the analysis, it was revealed that the use of Mask R-N technology can significantly increase the efficiency of performing tasks, since this algorithm is the latest version of the machine learning model. As a result of the study, a neural network was developed that is capable of identifying and distinguishing a large number of objects in an image. The next step is to refine the algorithm and use additional means of interaction with the hardware of the systems to increase the speed of the neural network. In the future, the resulting neural network will be integrated into the Digital Assistant for the Blind and Visually Impaired Persons application. This application is guaranteed to improve the daily life of people with disabilities who experience certain inconveniences due to their features, and can become the basis for other, larger projects related, for example, to unmanned devices, as well as services whose work is directly based on image processing.

KEYWORDS

Digital assistant, machine learning, neural networks, Mask R-CNN architecture, convolutional network, search algorithm, segmentation, classification, selective search, image processing, vision problems, visual impairment, assistance to the disabled, electronic assistant

FOR CITATION

Plotnikov S.O., Smetanin D.Yu., Basova A.V., Lvutin I.A., Belousova M.N. (2022) Use of a neural network in creating a digital assistant for blind and visually impaired people. *E-Management*, vol. 5, no. 3, pp. 73–82. DOI: 10.26425/2658-3445-2022-5-3-73-82

© Plotnikov S.O., Smetanin D.Yu., Basova A.V., Lvutin I.A., Belousova M.N., 2022.

This is an open access article under the CC BY 4.0 license (<http://creativecommons.org/licenses/by/4.0/>).



ВВЕДЕНИЕ / INTRODUCTION

В современных реалиях сфера компьютерного зрения стала неотъемлемой частью повседневной жизни. Она используется в огромном количестве процессов, от работы в обычном браузере и использования дома робота-пылесоса до управления беспилотными автомобилями.

Компьютерное зрение – область искусственного интеллекта, выполняющая анализ изображений. С помощью машинного обучения накапливается огромное количество информации, при помощи которой алгоритм по обработке изображений совершенствуется.

В рамках исследования было выбрано направление, связанное с использованием компьютерного зрения для помощи людям, имеющим проблемы со зрением, с помощью технологии нейронной сети Mask R-CNN.

Актуальность темы подтверждается тем, что в России, по данным независимых источников, каждый второй житель имеет какое-либо нарушение зрения. Главный офтальмолог Российской Федерации, директор МНИИ глазных болезней имени Гельмгольца Владимир Нероев на пресс-конференции в 2009 г. сообщил, что уровень слепоты и слабовидения в России достигает 19 человек на 10 тысяч населения. Всего же количество учтенных слепых и слабовидящих в России составляет 218 тысяч человек, из них абсолютно слепых – 103 тысячи. По словам Нероева, в контингенте инвалидов по зрению 22 % составляет молодежь¹.

В процессе исследования были изучены особенности повседневной жизни людей с нарушениями зрения [Белокурова, 2015; Болдинова, 2015].

Существует множество подходов и решений для задач распознавания образов объектов, однако все они уступают нейронным сетям, применение которых позволяет получить более точный результат за малый промежуток времени. Это подтверждается практическими работами авторов [Маркова, 2017; Сикорский, 2017; Сирота, 2019; Ковалевский, 2018].

В настоящее время предложено большое количество архитектур нейросетей для распознавания объектов. Анализ предлагаемых решений показывает, что до сих пор не существует такой модели, которая была бы лучшей среди всех результирующих показателей работы.

С алгоритмами нейросети MASK R-CNN (англ. Region Based Convolutional Neural Networks – региональные сверточные нейронные сети) функционал приложения для слабовидящих людей значительно расширяется. Например, вместо простого озвучивания названий предметов, окружающих пользователя, можно внедрить цифрового помощника, который будет активно взаимодействовать с человеком, детально описывать место, в котором он находится, предлагать ему наиболее комфортные маршруты, а также своевременно реагировать на изменения обстановки вокруг него, что повышает не только уровень комфорта, но и безопасность.

Разные возрастные группы инвалидов объединяют проблемы равного доступа к образованию, услугам, рынку труда и информации. Независимо от возраста, незрячие и слабовидящие нуждаются в помощи, так как ее отсутствие серьезно сужает возможности развития и означает ограничение мобильности, падение доходов, а также понижение социального статуса. Таким образом, приложение будет являться неотъемлемой частью жизни этих людей.

МЕТОДЫ ИССЛЕДОВАНИЯ / RESEARCH METHODS

Проанализировав существующие варианты модели машинного обучения для компьютерного зрения, была выбрана сеть с архитектурой Mask R-CNN, которая позволяет выделять на фотографиях маски экземпляров различных объектов, даже в случае частичного перекрытия одного объекта другим или различного размера объектов. Для обучения нейронной сети был выбран набор данных MS COCO (англ. Common Objects in Context) от Microsoft, содержащий больше 200 000 изображений и предоставляющий возможность определять 81 категорию объектов. Также данный набор данных был расширен нашими собственными проклассифицированными изображениями.

Mask R-CNN позволяет осуществлять следующие задачи компьютерного зрения:

- классификация;
- семантическая сегментация;
- обнаружение объекта;
- сегментация экземпляров.

¹Демоскоп Weekly (Среда 7 окт. 2009). Около 45 тысяч россиян ежегодно становятся инвалидами по зрению. Режим доступа: <http://www.demoscope.ru/weekly/2009/0393/rossia01.php#17> (дата обращения: 10.05.2022).

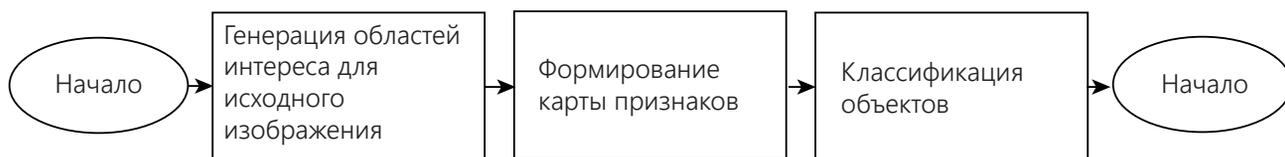
Для приложения, распознающего объекты на изображении и оповещающего о них пользователя, достаточно классификации изображения по типу содержащегося в нем объекта и определения объектов, но для проверки результатов обучения нейронной сети и визуализации результатов необходимо было изучить и освоить оставшиеся 2 задачи компьютерного зрения.

Архитектура нейронной сети

Mask R-CNN – это региональная сверточная нейронная сеть, что отличает ее от любой другой сети наличием методов обработки данных, имеющих сетчатую топологию.

В упрощенном виде работу региональной сверточной сети можно описать следующим образом.

В сеть поступает изображение, оно преобразуется в читаемые для алгоритма данные, затем на основе этих данных формируются гипотезы с помощью алгоритма поиска (в нашем случае это выборочный поиск – selective search), затем данные об этих теориях переходят на сверточный слой, в котором происходит определение признаков, преобразуемых в векторы, которые затем сопоставляются с данными, полученными из обучения с помощью методов опорных векторов, и утверждаются для выделения с регионами расположения с помощью регрессии (рис. 1).

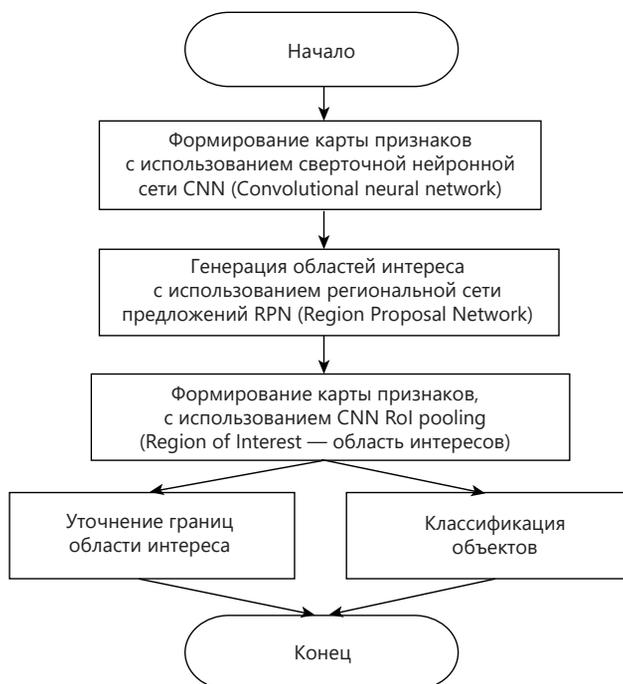


Составлено авторами по материалам исследования / *Compiled by the authors on the materials of the study*

Рис. 1. Архитектура нейронной сети упрощенно
Figure 1. Simplified neural network architecture

Далее будут освещены те алгоритмы, которые отличают Mask R-CNN от других нейронных сетей.

Для сегментации объектов на изображении используется схема поиска из Faster R-CNN. Архитектура работы алгоритма R-CNN представлена на рисунке 2. По сравнению с предыдущей моделью сегментации объектов R-CNN, в Faster R-CNN вместо использования внешнего алгоритма для генерации областей интереса, предположительно содержащих объект, используется специальная сеть Region Proposal Network (RPN).



Составлено авторами по материалам исследования / *Compiled by the authors on the materials of the study*

Рис. 2. Алгоритм работы R-CNN
Figure 2. R-CNN operation algorithm

Выделим этапы, которые выполняет сверточная нейронная сеть [Сирота, 2019, Rafael, 2018].

1. Формируется карта признаков на основе исходного изображения.
2. Осуществляется генерация областей интереса, возможно содержащих объект. С использованием региональной сети предложений Region Proposal Network (далее – RPN) сформированная карта признаков обрабатывается передвигающимся ядром (англ. kernel).
3. Выполняется преобразование вектора признаков области интереса из исходного изображения в вектор признаков фиксированной размерности, с помощью слоя RoI pooling (англ. Region of Interest — область интересов).
4. Осуществляется уточнение границ области интереса при помощи регрессионной модели (англ. Bounding Box Regression). Полученные области интереса и полученные вектора признаков подаются на вход двум новым полностью связанным слоям.
5. Выполняется классификация объектов, содержащихся в предполагаемых областях интереса. Последние два этапа выполняются параллельно.

Алгоритм поиска объектов

Нейронная сеть Mask R-CNN подразумевает использование алгоритма селективного поиска (англ. Selective search), который использует 4 меры сходства, основанные на совместимости цвета, текстуры, размера и формы. Рассмотрим эти сходства.

1. Сходство по цвету.

Цветовое сходство двух областей основано на пересечении гистограмм и может быть рассчитано как:

$$s_{color}(r_i, r_j) = \sum_{k=1}^n \min(c_i^k, c_j^k), \quad (1)$$

где r_i и r_j – две области или сегмента на изображении, c_i^k – значение гистограммы для ячейки k^{th} в дескрипторе цвета

Сходство текстур двух областей также вычисляется с использованием пересечений гистограмм:

$$s_{texture}(r_i, r_j) = \sum_{k=1}^n \min(t_i^k, t_j^k), \quad (2)$$

где t_i^k – значение гистограммы для ячейки k^{th} в дескрипторе текстуры.

2. Сходство размеров стимулирует более мелкие регионы (англ. bounding boxes) к раннему объединению. Это гарантирует, что предложения регионов во всех масштабах будут сформированы во всех частях изображения. Если эта мера сходства не будет принята во внимание, один регион будет продолжать поглощать все 34 более мелких соседних региона один за другим, и, следовательно, предложения по регионам в нескольких масштабах будут создаваться только в этом месте. Сходство размеров определяется как:

$$s_{size}(r_i, r_j) = 1 - \frac{size(r_i) + size(r_j)}{size(im)}, \quad (3)$$

где $size(im)$ – размер изображения в пикселях.

Совместимость форм определяется как:

$$s_{fill}(r_i, r_j) = 1 - \frac{size(BB_{ii}) - size(r_i) - size(r_j)}{size(im)}, \quad (4)$$

где $size(BB_{ii})$ является ограничивающей рамкой вокруг t_i и r_j .

3. Итоговое сходство.

Окончательное сходство между двумя регионами определяется как линейная комбинация вышеупомянутых 4 сходств:

$$s(r_i, r_j) = a1s_{color}(r_i, r_j) + a2s_{texture}(r_i, r_j) + a3s_{size}(r_i, r_j) + a4s_{fill}(r_i, r_j), \quad (5)$$

где $a_i \in \{0, 1\}$ обозначает, используется ли мера подобия или нет.

Сверточный слой

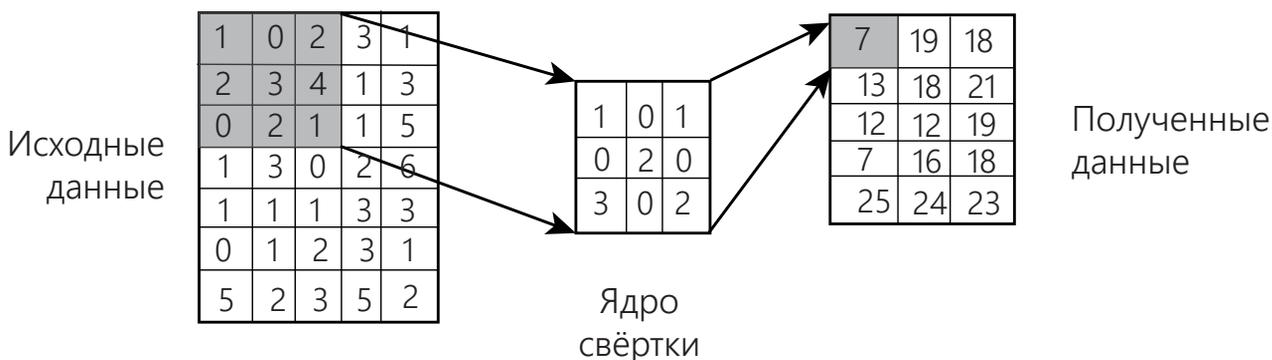
Следующая особенность работы нейронной сети R-CNN это сверточный слой, который выделяет признаки расположенных рядом пикселей из найденных алгоритмом селективного поиска областей в вектор. Работа алгоритма происходит следующим образом: по изображению последовательно перемещается ядро (kernel) размером $n \times n$ пикселей (чем меньше n , тем точнее карта слоя). У каждого элемента ядра есть свой весовой коэффициент, который умножается на значение того пикселя, на который в данный момент наложен.

Последовательные перемещения по горизонтали и вертикали с помощью разных фильтров создают каналы, по которым далее появляются новые карты для обработки. А далее характеристики, полученные при возникновении карт, записываются в вектор и отправляются на слой классификации, что можно видеть на рисунке 2.

Для оценки качества классификации используется показатель IoU (англ. Intersection over Union – пересечение, деленное на объединение). Ниже приведена формула расчета показателя IoU:

$$IoU = \frac{AoO}{AoU}, \tag{6}$$

где AoO (англ. Area of Overlap) – площадь пересечения истинного ограничивающего прямоугольника и предсказанного; AoU (Area of Union) – площадь объединения истинного ограничивающего прямоугольника и предсказанного.



Составлено авторами по материалам исследования / *Compiled by the authors on the materials of the study*

Рис. 3. Сверточный слой нейронной сети
 Figure 3. Convolutional layer of a neural network

Для алгоритма сверточной нейронной сети CNN существует несколько моделей, наиболее распространенные из них это ResNet50 и VGG16. С данными анализа моделей алгоритмов можно ознакомиться в таблице 1.

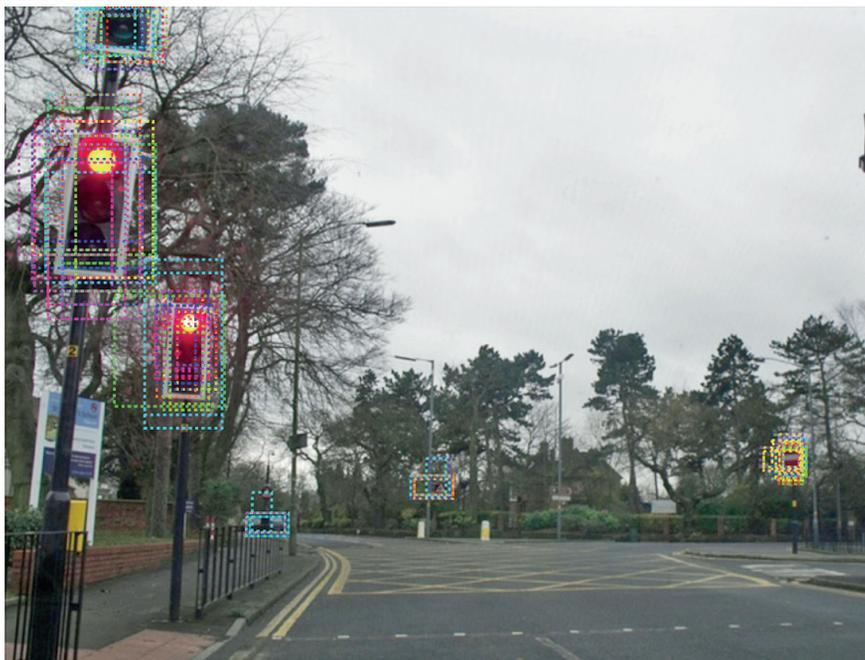
Таблица 1. Сравнение моделей алгоритма CNN

Table 1. Comparison of CNN algorithm models

Модель CNN	Количество эпох обучения	Точность распознавания на тестовой выборке	Время обнаружения объекта, сек
VGG	50	0,8867	1,363
	100	0,9054	1,317
	150	0,9647	1,407
ResNet50	50	0,9234	1,248
	100	0,9648	1,243
	150	0,9832	1,313

Составлено авторами по материалам исследования / *Compiled by the authors on the materials of the study*

Перечисленные выше алгоритмы качественно отличают региональные сверточные нейронные сети на фоне других. По результатам аналитики существующих сетей, алгоритмов и моделей машинного обучения была написана нейронная сеть для сегментации объектов на изображении. Реализованные алгоритмы селективного поиска сверточной нейронной сети CNN представлены на рисунках 4 и 5. Визуализированный результат работы нейронной сети можно увидеть на рисунке 6.



Составлено авторами по материалам исследования / *Compiled by the authors on the materials of the study*

Рис. 4. Результат работы алгоритма селективного поиска
Figure 4. The result of the selective search algorithm



Составлено авторами по материалам исследования / *Compiled by the authors on the materials of the study*

Рис. 5. Результат работы алгоритма CNN
Figure 5. The result of the CNN algorithm



Составлено авторами по материалам исследования / *Compiled by the authors on the materials of the study*

Рис. 6. Результат работы нейронной сети
Figure 6. The result of the neural network

Оценка скорости работы нейронной сети

Количество объектов на изображении не влияет на скорость работы нейронной сети. На мобильной версии графической карты Nvidia GTX 1050 скорость обработки изображения составляет 1,3 секунды (табл. 1). На более новых видеокартах Nvidia RTX 30 серии скорость работы алгоритма выше в 3 раза, и будет составлять 0,4 секунды. Также на скорость обработки изображений влияет и объем набора используемых данных. Файл, в котором хранятся веса полносвязного слоя нейронной сети, равен 250 мегабайтам, что позволяет использовать его на мобильных устройствах, он также не замедляет обработку изображения, при достаточной точности.

ЗАКЛЮЧЕНИЕ / CONCLUSION

В статье рассмотрены вопросы, связанные с внедрением искусственного интеллекта для помощи людям, имеющим проблемы со зрением, а именно использование сверточной нейронной сети модели Mask R-CNN для сегментации объектов на изображении. Выявлено, что использование такой технологии существенно увеличит уровень комфорта, безопасности и мобильности в повседневной жизни.

По результатам аналитики существующих сетей, алгоритмов и моделей машинного обучения была написана нейронная сеть для сегментации объектов на изображении. Реализованы алгоритмы селективного поиска и сверточной нейронной сети. В результате анализа было выявлено, что выбранная нейронная сеть Mask R-CNN позволит существенно увеличить быстродействие приложения и количество выполняемых им задач. Сеть с архитектурой Mask R-CNN позволяет выделять на фотографиях маски экземпляров различных объектов, даже в случае частичного перекрытия одного объекта другим или различного размера объектов. Для обучения нейронной сети используется dataset MS COCO (Common Objects in Context) от компании Microsoft, содержащий больше 200 000 изображений и предоставляющий возможность определять 91 категорию объектов. Таким образом, достигается максимальная эффективность создаваемого приложения.

Данная работа позволяет понять принцип технологии обработки изображения с помощью машинного обучения и нейронных сетей и выявить целесообразность ее использования, как помощника для слепых и слабовидящих людей.

СПИСОК ЛИТЕРАТУРЫ

- Белоглазова А.А. (2015). Образование и социализация детей с нарушениями зрения // *Коррекционная педагогика: теория и практика*. № 1. С. 83–86.
- Болдинова О.Г. (2015). Социализации дошкольников с нарушениями зрения в инклюзивном образовании // *Вестник Череповецкого государственного университета*. Т. 5, № 66. С. 87–91.
- Ковалевский А.М. (2018). Алгоритмы профилирования пользователя посредством нейронных сетей: Автореф. дисс. ... магистра техн. наук: 1–40 80 02. Минск: БГУИР. 14 с.
- Маркова С.В., Жигалов К.Ю. (2017). Применение нейронной сети для создания системы распознавания изображений // *Фундаментальные исследования*. № 8. С. 60–64.
- Сикорский О.С. (2017). Обзор сверточных нейронных сетей для задачи классификации изображений // *Новые информационные технологии в автоматизированных системах*. № 20. С. 37–42.
- Сирота А.А., Митрофанова Е.Ю., Милованова А.И. (2019). Анализ алгоритмов поиска объектов на изображениях с использованием различных модификаций сверточных нейронных сетей // *Вестник ВГУ*. №3. С. 123–137.
- Forsyth D.A., Ponce J. (2004). *Computer vision. Modern approach*. Upper Saddle River: Prentice Hall. 935 p.
- Girshick R., Donahue J., Darrell T., Malik J. (2013). Rich Feature Hierarchies for Accurate Object Detection and Semantic Segmentation // *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, Columbus, USA, 23–28 June 2014*. New York: IEEE. P. 580–587. <http://dx.doi.org/10.1109/CVPR.2014.81>
- Gonzalez R., Faisal Z. (2019). *Digital Image Processing*. 2nd ed. London: Pearson plc. 1192 p.
- Gu J., Wang Zh., Kuen J., Ma L., Shahroudy A., Shuai B., Liu T., Wang X., Wang G. (2018). Recent Advances in Convolutional Neural Networks // *Pattern Recognition*. V. 77. P. 354–377. <https://doi.org/10.1016/j.patcog.2017.10.013>
- He K., Gkioxari G., Dollár P. and Girshick R. (2017). Mask R-CNN // *Proceedings of IEEE International Conference on Computer Vision, ICCV 2017, Venice, 22–29 October 2017*. New York: IEEE. P. 2980–2988. <https://doi.org/10.1109/iccv.2017.322>
- Li Z., Liu F., Yang W., Peng S., Zhou J. (2021). A Survey of Convolutional Neural Networks: Analysis, Applications, and Prospects // *IEEE Transactions on Neural Networks and Learning Systems*. P. 1–21. <https://doi.org/10.1109/TNNLS.2021.3084827>
- Lin T.Y., Maire M., Belongie S., Bourdev L., Girshick R., Hays J., Perona P., Ramanan D., Zitnick L.C., Dollár P. (2014). Microsoft COCO: Common Objects in Context // *Proceedings of 13th European Conference on Computer Vision ECCV, Zurich, Switzerland, 6–12 September 2014*. Springer. 16 p.
- Uijlings J., Sande K., Gevers T., Smeulders A.W.M. (2013). Selective Search for Object Recognition // *International Journal of Computer Vision*. V. 104. P. 154–171. <https://doi.org/10.1007/s11263-013-0620-5>
- Vuletić G., Šarlija T., Benjak T. (2016). Quality of life in blind and partially sighted people // *Journal of Applied Health Sciences*. V. 2. P. 101–112. <http://dx.doi.org/10.24141/1/2/2/3>
- Welp A., Woodbury R.B., McCoy M.A. (2016). *Making Eye Health a Population Health Imperative: Vision for Tomorrow*. Washington (DC): National Academies Press. 586 p.
- Wu J. (2017). *Introduction to Convolutional Neural Networks*. Nanjing: Nanjing University. 31 p.

REFERENCES

- Beloglazova A.A. (2015), “Education and socialization of children with visual impairment”, *Correctional Pedagogy: Theory and Practice*, no 1, pp. 83–86.
- Boldinova O.G. (2015), “Socialization of preschool children with visual impairments in inclusive education”, *Bulletin of Cherepovets State University*, vol. 5, no. 66, pp. 87–91.
- Forsyth D.A., Ponce J. (2004), *Computer vision. Modern approach*, Prentice Hall, Upper Saddle River, US.
- Girshick R., Donahue J., Darrell T., Malik J. (2013), Rich Feature Hierarchies for Accurate Object Detection and Semantic Segmentation, In: *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, Columbus, USA, 23–28 June 2014*. IEEE, New York, US, pp. 580–587, <http://dx.doi.org/10.1109/CVPR.2014.81>
- Gonzalez R., Faisal Z. (2019), *Digital Image Processing*, 2nd ed., Pearson plc, London, UK.
- Gu J., Wang Zh., Kuen J., Ma L., Shahroudy A., Shuai B., Liu T., Wang X., Wang G. (2018), “Recent Advances in Convolutional Neural Networks”, *Pattern Recognition*, vol. 77, pp. 354–377, <https://doi.org/10.1016/j.patcog.2017.10.013>

- He K., Gkioxari G., Dollár P. and Girshick R. (2017), Mask R-CNN, In: *Proceedings of IEEE International Conference on Computer Vision, ICCV 2017, Venice, 22–29 October 2017*. IEEE, New York, US, pp. 2980–2988, <https://doi.org/10.1109/iccv.2017.322>
- Kovalevskij A.M. (2018), “User profiling algorithms using neural networks”: Abstr. diss. ... Mgr. Sci. (Tech.): 1–40 80 02, BSUIR, Minsk, Belarus.
- Li Z., Liu F., Yang W., Peng S., Zhou J. (2021), “A Survey of Convolutional Neural Networks: Analysis, Applications, and Prospects”, *IEEE Transactions on Neural Networks and Learning Systems*, pp. 1–21, <https://doi.org/10.1109/TNNLS.2021.3084827>
- Lin T.Y., Maire M., Belongie S., Bourdev L., Girshick R., Hays J., Perona P., Ramanan D., Zitnick L.C., Dollár P. (2014), Microsoft COCO: Common Objects in Context, In: *Proceedings of 13th European Conference on Computer Vision ECCV, Zurich, Switzerland, 6–12 September 2014*. Springer. 16 p.
- Markova S.V., Zhigalov K.Yu. (2017), “Application of a neural network to create an image recognition system”, *Basic Research*, no. 8, pp. 60–64.
- Sikorskij O.S. (2017), “Overview of convolutional neural networks for the problem of image classification”, *New information technologies in automated systems*, no. 20, pp. 37–42.
- Sirota A.A., Mitrofanova E.YU., Milovanova A.I. (2019), “Analysis of algorithms for searching objects in images using various modifications of convolutional neural networks”, *Proceedings of Voronezh State University. Series Systems analysis and information technologies*, no. 3, pp. 123–137.
- Uijlings J., Sande K., Gevers T., Smeulders A.W.M. (2013), “Selective Search for Object Recognition”, *International Journal of Computer Vision*, vol. 104, pp. 154–171, <https://doi.org/10.1007/s11263-013-0620-5>
- Vuletić G., Šarlija T., Benjak T. (2016), “Quality of life in blind and partially sighted people”, *Journal of Applied Health Sciences*, vol. 2, pp. 101–112, <http://dx.doi.org/10.24141/1/2/2/3>
- Welp A., Woodbury R.B., McCoy M.A. (2016), *Making Eye Health a Population Health Imperative: Vision for Tomorrow*, National Academies Press, Washington (DC), US.
- Wu J. (2017), *Introduction to Convolutional Neural Networks*, Nanjing University, Nanjing, China.