
Разработка моделей и методов распознавания русской дактильной азбуки

И.П. Печенов, И.В. Гуляев, И.Н. Шабанова, А.А. Астрелин, Д.М. Курляк

Волгоградский государственный технический университет

Аннотация: Данная статья посвящена разработке моделей и методов для распознавания русской дактильной азбуки, используемой глухими и слабослышащими людьми для коммуникации. В статье проводится исследование в области распознавания жестов и символов русской дактильной азбуки с применением современных методов машинного обучения и компьютерного зрения. Авторы представляют разработанные модели и алгоритмы, основанные на глубоком обучении и нейронных сетях для повышения точности распознавания жестов. Особое внимание уделено проблемам обучения на ограниченном объеме данных.

Ключевые слова: машинное зрение, распознавание жестов, классификация дактильных жестов, перенос обучения, русский дактильный алфавит, глубокое обучение, компьютеризация, программное обеспечение, ассистивные технологии, сверточные нейронные сети.

Введение

Количество людей, имеющих нарушения слуха, весьма значительно и продолжает расти как в мире, так и в России.

В мире, по оценкам Всемирной организации здравоохранения (ВОЗ) на 2022 год, свыше 5% населения мира (или около 466 миллионов человек) имеют проблемы со слухом. Это включает в себя людей с различными степенями нарушения слуха, начиная от легкой потери слуха до глухоты [1].

В России, по данным Национального общества глухих и слабослышащих РФ, около 10% населения России имеют какие-либо нарушения слуха. При этом, около 5% из них (то есть 7,25 миллионов) являются глухими, то есть имеют слуховую потерю более 90 децибел в наилучшем ухе.

Важно отметить, что данные могут различаться в зависимости от источников и методологии исследования. Тем не менее, достоверно можно сказать, что количество людей, имеющих нарушения слуха, является значительным в России, и требует принятия соответствующих мер для предоставления им необходимой поддержки.

Развитие эффективных систем распознавания дактильной азбуки является крайне важным, учитывая огромное количество людей, страдающих от проблем со слухом. Безусловно, в мире уже не первый год существуют системы распознавания жестов, однако подавляющее большинство из них предполагают использование специального оборудования для обнаружения основных точек руки. Поэтому разработка методов распознавания дактильной азбуки без посреднических устройств может значительно улучшить коммуникацию для этих людей, обеспечивая им равные возможности и повышая качество их жизни.

Анализ предметной области

Русский жестовый дактильный алфавит - это система жестов, которая используется для передачи алфавитных символов на русском языке с помощью рук и пальцев. Он позволяет людям с нарушением слуха и зрения общаться и передавать информацию.

Жестовый алфавит состоит из двух частей: знаков для букв и знаков для некоторых других символов. Знаки букв обычно создаются с помощью одной руки, в то время как знаки для других символов создаются с помощью двух рук.

Некоторые буквы имеют уникальные жесты, в то время как другие кабельные обозначения имеют сходство с буквенными сочетаниями или их звуками. Например, жест для буквы "А" создается путем разведения пальцев, как бы образующих "шапку", которая напоминает форму этой буквы, пример жеста на рисунке 1 [2].

При дактилировании важно следовать определенным правилам, чтобы обеспечить удобство для глухого собеседника. Рука должна быть расположена так, чтобы глухой мог легко видеть жесты и движения переводчика. Выбор правой или левой руки зависит от вашей доминирующей



Рис. 1. – Пример жеста буквы «А»

руки и положения говорящего. Важно контролировать положение ладони, чтобы избежать недопониманий или ненужных жестов. Во время паузы рекомендуется сохранять последнюю дактилему в стабильном положении для четкости передаваемых мыслей [3].

Описание метода

Для решения задачи распознавания жеста был выбран комбинированный метод на основе определения руки в кадре с помощью библиотеки `mediapipe` и распознавания жеста на основе нейронной сети, потому что это позволяет объединить преимущества обеих технологий для более эффективного и точного распознавания жестов.

Основное преимущество такого подхода заключается в том, что использование `mediapipe` для определения положения руки в кадре обеспечивает быструю и точную локализацию жестов, что повышает точность распознавания и устойчивость к шумам и искажениям изображения даже в условиях изменяющегося освещения или заднего фона.

Далее, применение нейронной сети для распознавания жестов позволяет адаптировать систему к различным жестам и улучшить общую точность распознавания. Нейронная сеть позволяет извлечь ключевые признаки из изображения и классифицировать его в соответствии с обученными данными.

Также, к общим преимуществам можно отнести обработку данных в реальном времени, система способна обеспечить отзывчивость и мгновенное распознавание жестов [4].

Набор данных фото русского жестового алфавита

Для сбора данных было написано простое десктопное приложение, которое было призвано оптимизировать процесс набора датасета - набор данных, используемый для обучения, валидации и тестирования нейронных сетей. Суть приложения в захвате видеопотока с камеры и определения на нем руки человека, что призвано уменьшить количество некорректно собранных данных. В процессе исполнения отдельных жестов пользователю достаточно в специальном окне указать букву русского алфавита, которая дактилируется в данный момент и кликнуть по значку на графическом интерфейсе приложения или же нажать горячую клавишу на клавиатуре для сохранения фото.

Готовое ПО было предоставлено четырем педагогам специальных образовательных учреждений для людей с нарушением слуха и десяти респондентам, имеющим навыки общения посредством жестовых языков.

По результатам исследования был собран датасет из более 12500 оригинальных изображений, который впоследствии увеличили до 25000 путем аугментации данных - расширения разнообразия обучающего набора данных. Могут применяться такие техники аугментации, как повороты, отражения, изменения масштаба и т.д. Панорама датасета показана на рисунке 2.

Выбор архитектуры ИНС

При выборе архитектуры нейронной сети для задачи классификации жестов, в первую очередь, необходимо учитывать сложность задачи, количество классов, размер датасета и доступные вычислительные ресурсы.



Рис. 2. – Панорама части датасета

В дальнейшем собранные данные были очищены от различных артефактов и шума, что могло негативно сказаться на точности модели [5].

Для начала, стоит определить, что задача классификации жестов относится к области компьютерного зрения и требует использования глубоких нейронных сетей. Учитывая, что в данном случае имеется большое количество классов и средний размер датасета, также задачу можно отнести к сложной.

Исходя из малого количества данных на класс, целесообразно использовать для решения данной задачи предобученную модель. Данный вид моделей уже имеют широкий набор знаний, которые они приобрели на больших объемах данных.

Учитывая доступность вычислительных ресурсов, было решено использовать глубокие нейронные сети, такие, как ResNet, DenseNet и Vgg которые известны своей высокой точностью и способностью обучаться на сложных наборах данных. Эти архитектуры позволяют обучить модель достаточно глубоко, чтобы она могла извлечь сложные признаки из изображений жестов и выполнить классификацию с высокой точностью.

Кроме того, для улучшения производительности и ускорения процесса обучения были использованы техники передачи обучения (transfer learning) и настройки гиперпараметров модели. Это позволит использовать уже

предобученную модель как базовую и дообучить ее на нашем датасете, что сократит время обучения и улучшит результаты.

Обучение производилось на PyTorch. Данный фреймворк обладает отличной поддержкой для обучения нейронных сетей на графических процессорах, что позволяет ускорить процесс обучения. Также он имеет архив (модельный зоопарк) необходимых для реализации метода моделей нейронных сетей [6].

Ранее был упомянут механизм передачи модели обучения (transfer learning), далее рассмотрим этот механизм более детально. Центральная концепция техники передачи знаний в обучении модели основана на принципе, что после обучения нейронной сети на обширном наборе данных, мы можем успешно применить эту модель к другим данным, которые не использовались при первоначальном обучении. Именно этим принципом и обусловлено наименование данной методики - передача обучения (transfer learning), где знания передаются с одного набора данных на другой [7].

Для успешной реализации методики передачи знаний в обучении необходимо адаптировать последний слой нашей сверточной нейронной сети. Эта процедура необходима в силу различного количества классов в различных наборах данных. Например, в базе данных ImageNet содержится 1000 уникальных классов, в то время, как наш набор данных для классификации включает в себя всего лишь 25 классов. Именно поэтому важно модифицировать выходной слой нашей сверточной нейронной сети, чтобы количество его выходов соответствовало количеству классов в наших данных [8]. Схематический пример замены классификатора представлен на рисунке 3.

Для того, чтобы не нарушить заранее обученную модель во время тренировки, важно правильно управлять процессом обновления параметров.

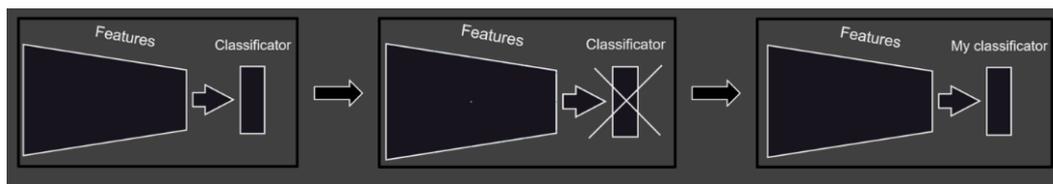


Рис. 3. – Визуализация процесса замены классификатора

Мы можем достигнуть этого, запретив алгоритму изменять значения параметров в процессе как прямого, так и обратного распространения [9]. Путем "замораживания" параметров предварительно обученной модели мы позволяем обучаться только последнему слою сети классификации, сохраняя при этом значения параметров предварительно обученной модели постоянными. Это позволяет значительно сократить время обучения за счет обучения только последнего слоя с намного меньшим количеством изменяемых переменных, а не всей модели. Такой подход является еще одним преимуществом использования предварительно обученных моделей [10].

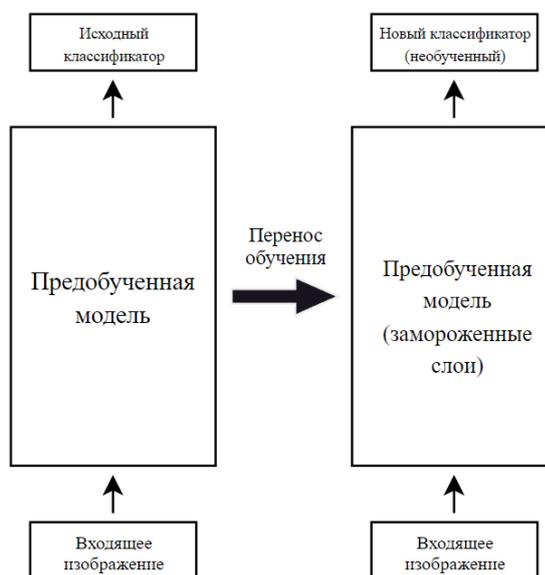


Рис. 4. – Визуализация процесса переноса обучения

Для оценки эффективности моделей была изучена способность нейросети выразить уверенность в своих прогнозах. С целью дальнейшего анализа результатов были выбраны следующие метрики оценки нейросети:

1. Accuracy (Acc) показывает общую точность модели, то есть долю правильно классифицированных примеров от общего числа примеров.

2. Specificity (Sp) показывает специфичность модели, то есть способность правильно определять негативные случаи (истинно отрицательные).

3. Sensitivity (Sn) показывает чувствительность модели, то есть способность правильно определять положительные случаи (истинно положительные).

4. Precision (Pr) показывает точность модели, то есть долю верно предсказанных положительных примеров среди всех прогнозов, отмеченных как положительные.

5. Matthews Correlation Coefficient (MCC) показывает коэффициент корреляции Мэтьюса, который учитывает истинно положительные и истинно отрицательные прогнозы, оценивая эффективность бинарной классификации.

6. F1-Score (F1-S) показывает средневзвешенное гармоническое между точностью и полнотой модели, дает баланс между ними и хорошо подходит для неравномерных классов [11].

Таблица № 1

Оценка метрик

| Модель | Датасет | Время обучения | Метрики(%) | | | | | |
|--------------|---------|----------------|------------|-------|-------|-------|-------|-------|
| | | | Acc | Sp | Sn | Pr | MCC | F1-S |
| ResNet 152 | Оригин. | 47 мин. | 96.37 | 96.78 | 97.99 | 96.14 | 90.78 | 99.83 |
| | Доп. | | 99.50 | 97.97 | 99.07 | 96.77 | 91.29 | 97.35 |
| DenseNet 161 | Оригин. | 38 мин. | 97.11 | 97.31 | 96.65 | 96.46 | 91.89 | 98.15 |
| | Доп. | | 98.27 | 99.64 | 96.81 | 98.72 | 93.91 | 99.58 |
| Vgg16 | Оригин. | 25 мин. | 93.77 | 95.38 | 94.68 | 93.43 | 91.90 | 97.60 |
| | Доп. | | 94.97 | 95.91 | 98.40 | 96.88 | 92.99 | 97.83 |

Полученные результаты также подтверждают, что обученные нейросети успешно генерализирует изученные паттерны на новых данных. Они проявляют высокую устойчивость к различным входным образцам, что

делает их эффективным инструментом для распознавания разнообразных жестов. В таблице 1 приведены краткие данные тестирования моделей.

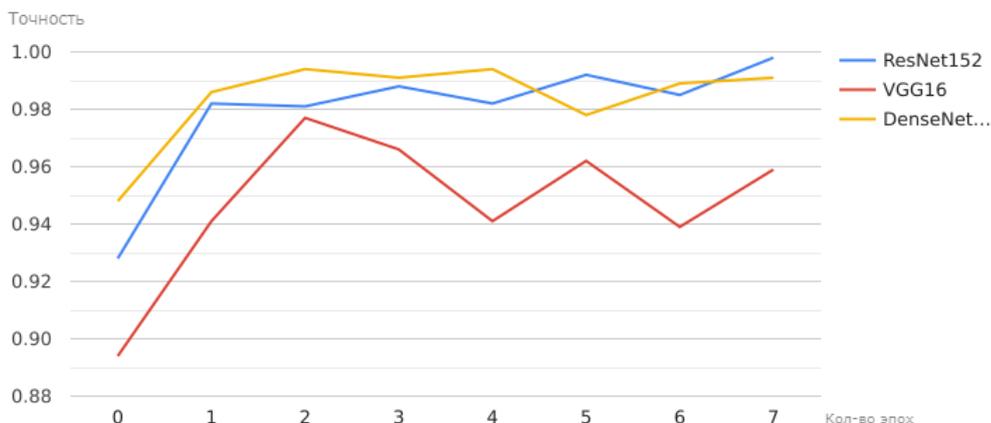


Рис. 5. – Точность обучения моделей на валидационной выборке

Также, благодаря аугментации данных, была проанализирована способность нейронных сетей преодолевать различные помехи, такие, как изменения освещения, размытие/изменение фона и другие артефакты на изображениях. Полученные результаты подтвердили её способность демонстрировать постоянную производительность даже в условиях переменных внешних факторов.



Рис. 6. – Аугментации данных на примере замены фона

Реализация метода

Далее необходимо разработать инференс (inference) - систему. Такая система включает в себя обработку входных данных с камеры, их предварительную обработку, передачу через нейронную сеть, и анализ результата для вывода ответа на основе прогноза модели.

При разработке приложения для распознавания жестов первым важным шагом стало определение руки в видеопотоке с использованием библиотеки Mediapipe. Данная библиотека предоставила необходимые инструменты для точного определения ключевых точек руки, таких, как кончики пальцев и ладонь. После выделения руки на кадре, мы осуществили её кадрирование, оставив только область, содержащую руку, с использованием точек, расположенных на ладони. Такой подход значительно упростил обработку изображения, позволив сосредоточиться исключительно на руке, и исключить ненужные фоны и элементы, что в итоге сделало изображение более пригодным для последующего анализа. По завершении изоляции изображения руки, необходимо приступить к его дальнейшей обработке, включая изменение размера изображения до требуемого разрешения, а также нормализацию значений пикселей для соответствия требованиям модели.

После успешной обработки изображения, оно поступает на вход нейронной сети для анализа и распознавания жестов. Нейронная сеть использует извлеченные ранее характеристики для детального анализа изображения и выдает результат распознавания в виде определенного жеста. Общая схема реализации метода представлена на рисунке 7.

И, в заключение, результаты распознавания жестов отображаются пользователю на экране устройства. Этот шаг позволит пользователям увидеть и взаимодействовать с результатами распознавания жестов, что составляет главную цель создания приложения.

Результаты и тестирование

Для проверки эффективности и точности программы по распознаванию русской тактильной азбуки в реальных условиях использования было проведено ручное тестирование. В тестировании участвовали как педагоги из специальных учреждений для глухих и слабослышащих, так и респонденты с опытом жестового языка.

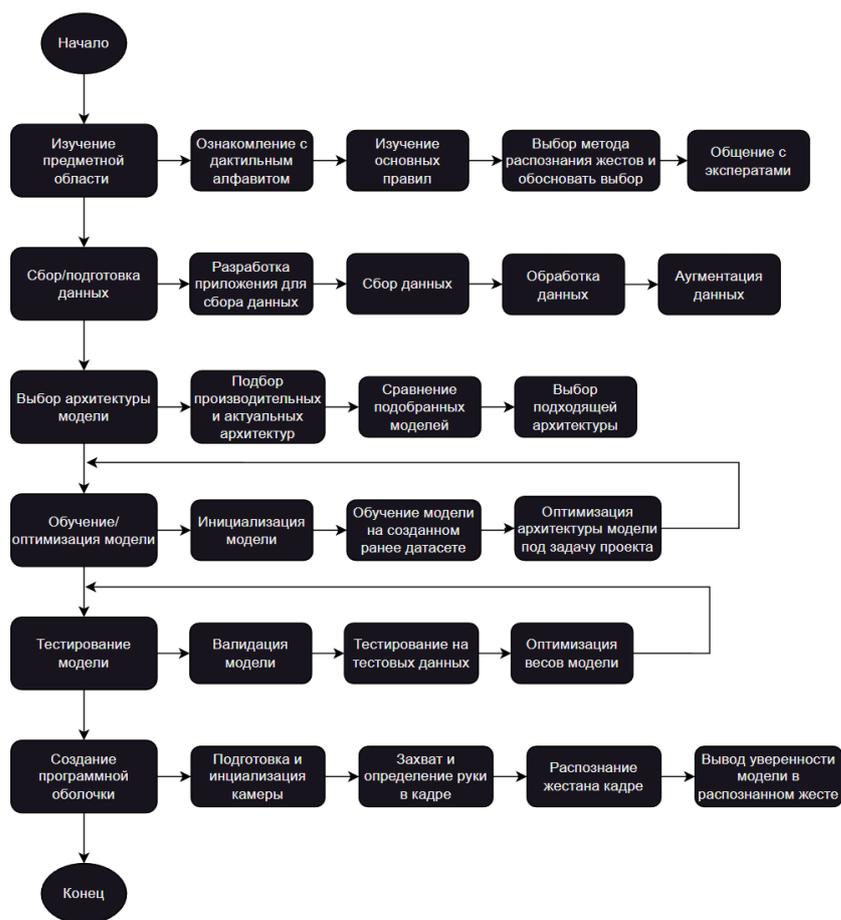


Рис. 7. – Схема реализации метода

Выборка включала в себя 4 педагогов из специальных учреждений и 10 респондентов с опытом жестового общения. Участники были разнообразны по возрасту, полу, цвету кожи и опыту общения на жестовом языке. Перед тестированием участникам была предоставлена информация о целях и процессе испытаний, а также объяснено, какие действия им необходимо выполнять для проведения жестов для распознавания.

По результатам тестирования было правильно распознано 96% общего числа тестов. Оценена была точность программы в распознавании различных жестов, включая степень совпадения с фактическими жестами. Были выявлены факторы, влияющие на качество распознавания жестов, такие, как фон, освещение и четкость изображения с вебкамеры [12].

После завершения тестирования участники дали обратную связь о процессе и качестве распознавания. Были выявлены возможные улучшения программы. Пример работы программы представлен на рисунке 8.

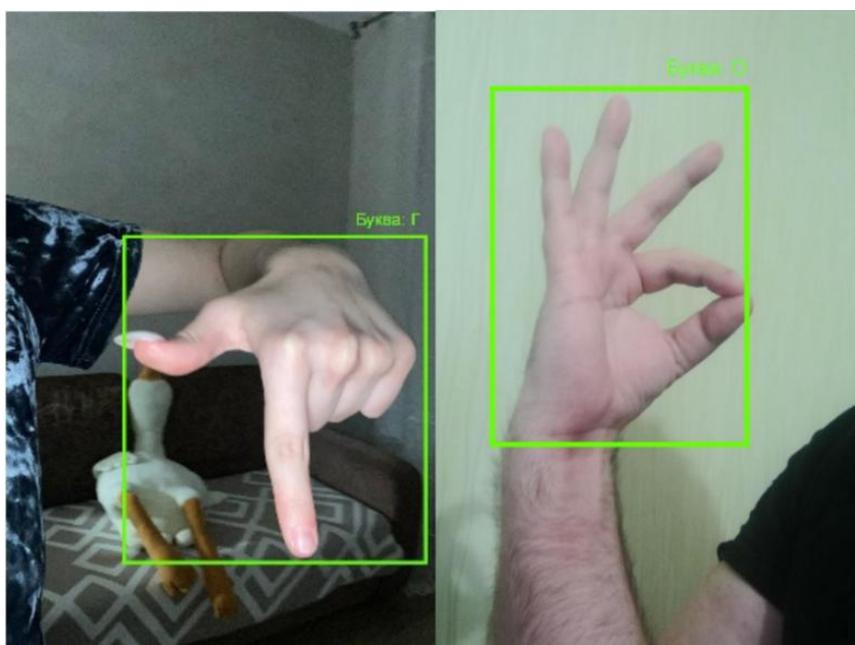


Рис. 8. – Пример распознавания буквы жеста

Заключение

В данной статье были представлены результаты исследований, посвященных разработке моделей и методов для распознавания русской дактильной азбуки, используемой глухими и слабослышащими для коммуникации. Работа сосредоточилась на исследовании распознавания жестов и символов дактильной азбуки, применения современных методов машинного обучения и компьютерного зрения.

Использование современных методов машинного обучения и компьютерного зрения в сочетании с алгоритмами глубокого обучения открывает новые возможности для развития систем распознавания тактильной азбуки без специализированного оборудования. Это важно для создания равных возможностей коммуникации и повышения качества жизни людей с нарушениями слуха.

Результаты исследования подтверждают эффективность обученных нейросетей в распознавании разнообразных жестов русской тактильной азбуки. Тестирование программы на реальных пользователях показало высокую точность распознавания (96%) и выявило факторы, влияющие на качество работы системы, такие, как освещение, фон и четкость изображения.

Полученные данные и результаты позволяют сделать вывод о перспективности разработанных моделей и методов в области распознавания тактильной азбуки. Обратная связь от участников тестирования и обозначенные возможности улучшений программы являются ценными рекомендациями для дальнейшей корректировки и совершенствования разработанной системы.

Таким образом, разработка моделей и методов распознавания русской тактильной азбуки представляет важный шаг в развитии ассистивных технологий и обеспечении качественной коммуникации для людей с нарушениями слуха, открывая новые горизонты для создания инновационных решений в данной области.

Литература

1. Глухота и потеря слуха / ВОЗ, 02 марта 2021 г. URL: who.int/ru/news-room/fact-sheets/detail/deafness-and-hearing-loss (дата обращения: 19.02.2024).

2. Харламенков, А.Е. Теория русского жестового языка: учебник. Москва: Издательство Юрайт, 2022. – 171 с.
 3. Гриф М.Г., Королькова О.О., Птушкин Г.С., Траулько Е.В. Особенности использования систем компьютерного сурдоперевода в инклюзивном образовании лиц с нарушением слуха: учебное пособие. Новосибирск: НГТУ, 2014. — 71 с.
 4. Liu Yu Han. Feature Extraction and Image Recognition with Convolutional Neural Networks // Journal of Physics Conference Series. 2018. P. 1–7. URL: iopscience.iop.org/article/10.1088/17426596/1087/6/062032/pdf (дата обращения: 12.11.2023).
 5. Галушка В.В., Фатхи В.А. Формирование обучающей выборки при использовании искусственных нейронных сетей в задачах поиска ошибок баз данных // Инженерный вестник Дона. 2013. № 2. URL: ivdon.ru/magazine/archive/n2y2013/1597.
 6. Романов Д.Е. Нейронные сети обратного распространения ошибки // Инженерный вестник Дона. 2009. №3. URL: ivdon.ru/ru/magazine/archive/n3y2009/143.
 7. Петров А.И., Белов Ю.С. Трансферное обучение модели VGG-16 // Региональная научно-техническая конференция. 2023. Т.2. С.57-61.
 8. Демин И.С., Белов Ю.С., Чухраев И.В. Обучение сверточной нейронной сети на базе архитектуры U-NET с использованием минимальных ресурсов // Электромагнитные волны и электронные системы. 2019. Т. 24. № 7. С. 24–29.
 9. Kensert A., Harrison P.J., Spjuth O. Transfer learning with deep convolutional neural network for classifying cellular morphological changes. PubMed.gov. 2018. pp. 1–11. URL: [biorxiv.org/content/10.1101/345728v1.full.pdf](https://www.biorxiv.org/content/10.1101/345728v1.full.pdf) (дата обращения: 12.11.2023).
-

10. Weiss K., Khoshgoftaar T.M., Wang D.D. A survey of transfer learning. *Journal of Big Data*. 2016. pp. 1–40. URL: journalofbigdata.springeropen.com/articles/10.1186/s40537-016-0043-6 (дата обращения: 19.02.2024).
11. Accuracy, Precision, Recall & F1 Score: Interpretation of Performance Measures. URL: blog.exsilio.com/all/accuracy-precision-recall-f1-score-interpretation-of-performance-measures. (дата обращения: 22.04.2024 г.).
12. Kaoutar B.A., Jelodar A. Fine-Tuning VGG Neural Network For Fine-grained State Recognition of Food Images. 2018. URL: arxiv.org/ftp/arxiv/papers/1809/1809.09529.pdf (дата обращения: 19.02.2024).

References

1. Gluxota i poterya sluxa [Deafness and hearing loss], VOZ, 02 marta 2021 g. URL: who.int/ru/news-room/fact-sheets/detail/deafness-and-hearing-loss (date accessed: 19.02.2024).
 2. Xarlamenkov, A.E. *Teoriya russkogo zhestovogo yazy`ka: uchebnik*. Moskva: Izdatel`stvo Yurajt, 2022. 171 p.
 3. Grif M.G., Korol`kova O.O., Ptushkin G.S., Traul`ko E.V. *Osobnosti ispol`zovaniya sistem komp`yuternogo surdoperevoda v inklyuzivnom obrazovanii licz s narusheniem sluxa: uchebnoe posobie*. Novosibirsk: NGTU, 2014. 71 p.
 4. Liu Yu Han. Feature Extraction and Image Recognition with Convolutional Neural Networks: *Journal of Physics Conference Series*. 2018. pp. 1–7. URL: iopscience.iop.org/article/10.1088/17426596/1087/6/062032/pdf (date accessed: 12.11.2023).
 5. Galushka V.V., Fatxi V.A. *Inzhenernyj vestnik Dona*, 2013, № 2. URL: ivdon.ru/magazine/archive/n2y2013/1597.
-

6. Romanov D.E. Inzhenernyj vestnik Dona, 2009, №3. URL: ivdon.ru/ru/magazine/archive/n3y2009/143.
7. Petrov A.I., Belov Yu.S. Transferное обучение модели VGG-16 Regional'naya nauchno-texnicheskaya konferenciya. 2023. T.2. pp. 57-61.
8. Demin I.S., Belov Yu.S., Chuxraev I.V. Obuchenie svertochnoj nejronnoj seti na baze arxitektury` U-NET s ispol`zovaniem minimal`ny`x resursov: E`lektromagnitny`e volny` i e`lektronny`e sistemy`. 2019. T. 24. № 7. pp. 24–29.
9. Kensert A., Harrison P.J., Spjuth O. Transfer learning with deep convolutional neural network for classifying cellular morphological changes. PubMed.gov. 2018. pp. 1–11. URL: biorxiv.org/content/10.1101/345728v1.full.pdf (date accessed: 12.11.2023).
10. Weiss K., Khoshgoftaar T.M., Wang D.D. A survey of transfer learning. Journal of Big Data. 2016. pp. 1–40. URL: journalofbigdata.springeropen.com/articles/10.1186/s40537-016-0043-6 (date accessed: 19.02.2024).
11. Accuracy, Precision, Recall & F1 Score: Interpretation of Performance Measures. URL: blog.exsilio.com/all/accuracy-precision-recall-f1-score-interpretation-of-performance-measures. (date accessed: 22.04.2024 г.).
12. Kaoutar B.A., Jelodar A. Fine-Tuning VGG Neural Network For Fine-grained State Recognition of Food Images. 2018. URL: arxiv.org/ftp/arxiv/papers/1809/1809.09529.pdf (date accessed: 19.02.2024).

Дата поступления: 19.03.2024

Дата публикации: 6.05.2024