

Научная статья
УДК 004.934
doi: 10.34822/1999-7604-2022-1-46-54

РАЗРАБОТКА МОДУЛЯ ПРОГРАММНОГО ОБЕСПЕЧЕНИЯ ПЕРЕВОДЧИКА СЛОВ ЖЕСТОВОГО ЯЗЫКА В ЗВУКОВОЙ ФОРМАТ

Валерия Олеговна Семенова^{1✉}, Лариса Леонидовна Семенова²

^{1,2}Сургутский государственный университет, Сургут, Россия

¹semenova_vo@bk.ru ✉, <http://orcid.org/0000-0003-3639-9036>

²semenova_ll@surgu.ru, <http://orcid.org/0000-0001-6384-0996>

Аннотация. В статье определены и сформулированы основные требования к нейронной сети с функциями распознавания жестов кистей рук, проведено исследование существующих подходов на соответствие этим требованиям. Составлен перечень для имитации реальных условий использования модуля распознавания жестов «Диалогового сурдопереводчика», дополнительно проведены измерения необходимых характеристик камеры для корректной работы нейронной сети с целью получения качества распознавания кисти руки в режиме реального времени не ниже 75 %. Разработан прототип нейронной сети, представлен успешный результат различения левой и правой руки в имитационных условиях. В качестве перспективы заявлена необходимость обеспечения работы программного обеспечения не только с жестами, но и с речью для дальнейшего развития технологий инклюзии людей с коммуникативными трудностями.

Ключевые слова: сурдоперевод, нейронные сети, жестовый язык, машинное зрение, машинное обучение, MediaPipe

Для цитирования: Семенова В. О., Семенова Л. Л. Разработка модуля программного обеспечения переводчика слов жестового языка в звуковой формат // Вестник кибернетики. 2022. № 1 (45). С. 46–54. DOI 10.34822/1999-7604-2022-1-46-54.

Original article

DEVELOPING A SOFTWARE MODULE FOR TRANSLATING WORDS FROM SIGN LANGUAGE INTO AN AUDIO FORMAT

Valeriya O. Semenova^{1✉}, Larisa L. Semenova²

^{1,2}Surgut State University, Surgut, Russia

¹semenova_vo@bk.ru ✉, <http://orcid.org/0000-0003-3639-9036>

²semenova_ll@surgu.ru, <http://orcid.org/0000-0001-6384-0996>

Abstract. The article determines and formulates the main requirements to a neural network that detects hand gestures. A study of existing approaches is conducted on the matter of their compliance with the requirements. A list of real conditions is compiled to imitate an application of a sign recognition module “Dialogue Sign Language Translator”. In addition to that, the required camera characteristics were measured for proper operation of the neural network in order to obtain a real-time palm recognition quality of at least 75 %. A prototype of a neural network was developed. The study shows the efficiency at differentiating left and right hands during imitation. From the prospect point of view, it is necessary to build the software operating both with signs and with speech in order to continue developing the technologies for inclusion of people with communication difficulties.

Keywords: sign language translation, neural networks, sign language, machine vision, machine learning, MediaPipe

For citation: Semenova V. O., Semenova L. L. *Developing a Software Module for Translating Words from Sign Language into an Audio Format // Proceedings in Cybernetics. 2022. No. 1 (45). P. 46–54. DOI 10.34822/1999-7604-2022-1-46-54.*

ВВЕДЕНИЕ

Одной из актуальных проблем инклюзии в современном обществе является наличие языкового барьера, возникающего у людей, вынуждено использующих для коммуникации жестовый язык. Поскольку людей, не имеющих коммуникативных трудностей и владеющих жестовым языком на необходимом для общения уровне, крайне мало, имеется только два варианта преодоления данного барьера: помощь близкого окружения либо услуги профессиональных сурдопереводчиков. Таким образом, создание переносного устройства, способного стать при необходимости посредником в диалоге, – важное направление в разработке технологий снижения потерь от социально значимых заболеваний и повышения скорости инклюзии людей с коммуникативными трудностями.

МАТЕРИАЛЫ И МЕТОДЫ

В процессе решения поставленных задач использованы принципы системного подхода, теории графов, методы машинного обучения (в частности, машинное зрение и распознавание визуальных образов в реальном времени), методы линейного программирования, а также теоретические и натурные экспериментальные методы исследования.

В результате исследования существующих подходов к распознаванию кисти руки установлено, что алгоритм, реализующий возможность представления объекта на изображении с описанием распространения интенсивности градиентов или направлением контуров, способствует уменьшению количества ошибок во время распознавания жестов [1]. При распознавании модулем жестов руки с помощью камеры в режиме реального времени возможно применение алгоритма, последовательно реализующего два взаимодополняющих метода: обнаружение жеста в видеопотоке осу-

ществляется на основе анализа движений с последующим распознаванием обнаруженного жеста с помощью вычитания фона [2].

По итогам теоретического анализа рассмотренных методов решения, установлено, что наиболее эффективной как в упрощении и удешевлении разработки, так и в целях повышения качества функционала будет интеграция фреймворка MediaPipe Hands, в основе которого лежит конвейер машинного обучения, состоящий из двух одновременно работающих моделей: модели обнаружения полного изображения ладони и возвращения ограничивающей рамки руки и модели ориентира детектором ладони, работающей с обрезанной областью изображения, и возвращения ее высокоточных трехмерных ключевых точек. Основным преимуществом использования машинного обучения является резкое снижение потребности в увеличении данных (например, вращении, перемещении и масштабировании), вместо этого сеть направляет большую часть своих возможностей на точность прогнозирования координат, что существенно снижает вероятность понижения эффективности распознавания в случае частичной видимости руки [3–5].

РЕЗУЛЬТАТЫ И ИХ ОБСУЖДЕНИЕ

Диалоговый цикл. В результате исследования установлено, что существующие устройства перевода жестового языка не способны обеспечить поддержку диалога и ориентированы на монолог говорящего на жестовом языке, т. е. компенсацию только половины языкового барьера. Данная проблема может быть решена устройством с функционалом поддержки диалогового цикла не только для перевода жестовых слов, но и для работы с речевым ответом говорящего собеседника (рис. 1).



Рис. 1. Общий вид диалогового цикла при использовании устройства сурдоперевода с поддержкой диалогового цикла
 Примечание: составлено авторами.

В случае, если каждая из данных частей будет работать по своему алгоритму, полностью реализующему соответствующую ветвь цикла, последовательное их использование позволит вести диалог в простом и понятном для обеих сторон виде.

Подсистема «речь – текст». Возможность использования устройства для сурдоперевода в случае отсутствия функционала помощи в понимании речевого ответа оправдана лишь в случае пользователя с коммуни-

кативными трудностями, но без проблем со слухом, мешающих восприятию устной речи в достаточной для понимания мере, что оставляет проблему коммуникации открытой для пользователей с нарушением не только речи, но и слуха. В таком случае необходимо использование подсистемы «речь – текст», задачей которой является получение устного сообщения через систему улавливания речи и последующее его преобразование в текст с выводом на экран устройства (рис. 2).

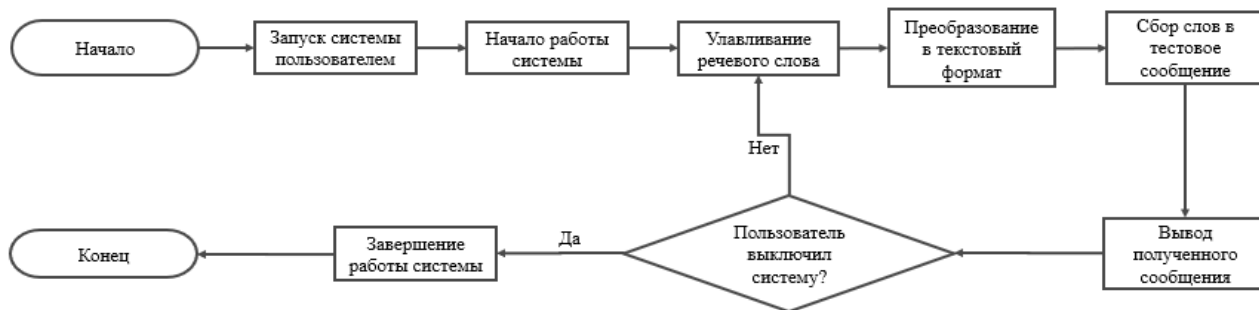


Рис. 2. Общий вид алгоритма работы подсистемы «речь – текст»
 Примечание: составлено авторами.

Данная подсистема реализует первую половину предполагаемого диалогового цикла, ответственную за получение и понимание устного сообщения человеком с коммуникативными трудностями.

Подсистема «жест – звук». Эффективность работы системы перевода жестовых слов, состоящей из программного комплекса и устройства считывания и вывода данных, определяется точностью и естественностью

работы итогового программного продукта. При этом большая роль в качественной оценке работоспособности отводится подсистеме «жест – звук», состоящей из двух основных частей: нейронной сети, распознающей жестовое слово и осуществляющей его перевод в текстовую форму, и программы синтеза голоса для чтения выводимого нейронной сетью текста, имитирующего естественную речь человека (рис. 3).

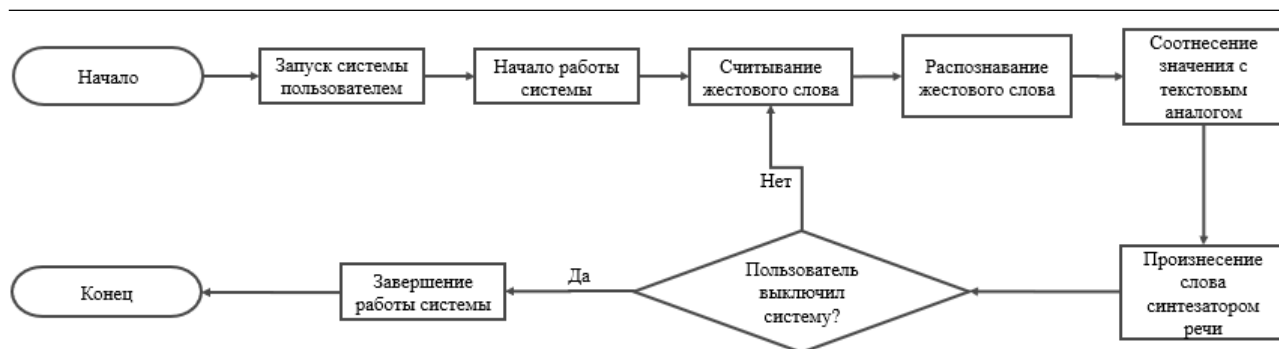


Рис. 3. Общий вид алгоритма работы подсистемы «жест – звук»

Примечание: составлено авторами.

Данная подсистема реализует вторую половину предполагаемого диалогового цикла, ответственную за передачу сообщения от человека с коммуникативными трудностями.

Исследования для определения оптимального типа нейронной сети. В ходе исследовательской работы определены следующие ключевые требования к распознаванию жестов кистей нейронной сетью:

1. Работа с безмаркерной системой, т. е. считывание жеста с помощью камеры устройства или оптического датчика.

2. Показ достаточного уровня распознавания жеста кисти руки как при усредненных, так и при отличных от усредненных условиях съемки, включая близкие к граничным. Под достаточным уровнем распознавания понимается процент считанных жестов, необходимый для сохранения смысла изначального сообщения.

3. Работа с камерами или оптическими датчиками с низкими техническими показателями для исключения критического влияния на себестоимость устройства.

4. Корреляция значения жеста в соответствии с показаниями гироскопа, т. е. не только жеста кисти руки, но и ее положения относительно камеры или оптического датчика.

В ходе экспериментальных исследований существующих нейронных сетей было решено интегрировать фреймворк с открытым кодом MediaPipe Hands распознавания жестов кистей рук в итоговый программный продукт. Данный фреймворк соответствует заявленным в проекте требованиям и имеет возможность изменения исходного кода при необходимости. Распознавание жестов при исполь-

зовании фреймворка происходит с помощью построения трехмерного скелета ладони, содержащего 21 точку наблюдения. Через положение этих точек в пространстве определяется жест.

Ключевые особенности отобранного для нейронной сети фреймворка:

1. Необходимость использования камеры для считывания жеста кисти руки.

2. Низкие требования к техническим характеристикам камеры при условиях сохранения считывания на достаточном уровне для использования потребителем.

3. Распознавание не только жеста кисти руки, но и положения кисти относительно камеры без использования гироскопа.

4. Распознавание жестов как одной, так и одновременно двух рук без изменения механизма распознавания и требований к камере.

Экспериментальные исследования для определения минимально необходимых характеристик камеры. В ходе экспериментальных исследований заранее отобранные камеры с различным фокусным расстоянием и качеством съемки были поочередно (в порядке уменьшения показателей) протестированы с использованием фреймворка MediaPipe Hands с целью определения минимальных характеристик камеры для комфортного использования проектируемого устройства перевода.

По результатам исследования установлено, что для работы достаточно использования камеры с разрешением не менее 1 МП и углом обзора не менее 100° (фокусное расстояние не менее 20 мм). Данные технические характеристики являются минимальными.

ми и обеспечивают достаточный уровень распознавания жестов кистей рук.

Необходимые условия и их комбинации для симуляции использования модуля распознавания жестов «Диалогового сурдопереводчика» потребителем в повседневной жизни:

1. Имитация низкой освещенности (0,1 лк) в темное время суток.

2. Имитация повышенной освещенности (10 000 лк) при солнечной погоде.

3. Имитация размытости границ в ситуации загрязнения объектива камеры.

4. Имитация оптического дефекта попадания на объектив камеры в процессе работы нейронной сети луча солнца или частицы грязи.

Разработка прототипа нейронной сети.

В качестве экспериментального образца был использован фреймворк MediaPipe Hands, интегрированный в приложение на смартфон с операционной системой Android. Данное решение было принято в связи с легкостью интеграции данного фреймворка в приложение, а также возможностью впоследствии использовать полученный экспериментальный образец как основу или составную часть итогового программного продукта.

Минимальные требования к смартфону для возможности запуска и корректной работы экспериментального образца нейронной сети:

1. Android 4.0.

2. Поддержка Bluetooth.

3. Процессор с поддержкой ARM64.

4. Минимум 26 МБ свободного места во внутренней памяти.

Итоговое приложение имеет следующие атрибуты:

- наименование исполняемого файла – handddetection.apk;

- размер исполняемого файла – 26 368 524 байт (25,1 Мб);

- версия файла – 1.0;

- версия продукта – 1.0;

- внутреннее имя – handddetection;

- исходное имя файла – handddetection.apk;

- название продукта – Hand Reader;

- описание версии файла – 1.0;

- язык – English (United States).

Экспериментальный образец модуля может работать как с фронтальной камерой смартфона, так и с камерой, подключенной по сети Bluetooth.

Минимальные требования к камере, предъявляемые приложением, соответствуют минимальным требованиям, заявленным в результате исследования.

Для проверки качества распознавания жестов прототипом и его взаимодействия с фреймворком было принято решение обучить нейронную сеть идентифицировать руку в кадре как правую или левую [6].

Поскольку экспериментальный образец нейронной сети в целях удобства считывает жесты кисти руки через фронтальную камеру смартфона, необходимо учесть зеркальность изображения. Для этого нейронная сеть, распознавая кисть руки, намерено дать ей противоположное название.

Сама модель нейронной сети представляет собой многослойный персептрон (MLP), являющийся многослойной структурой нейронов с полносвязным соединением двух соседних слоев. Структура такой нейронной сети характеризуется наличием нескольких типов нейронных слоев: входного, скрытого и выходного – для получения, обработки и вывода сигнала соответственно [7] (рис. 5).

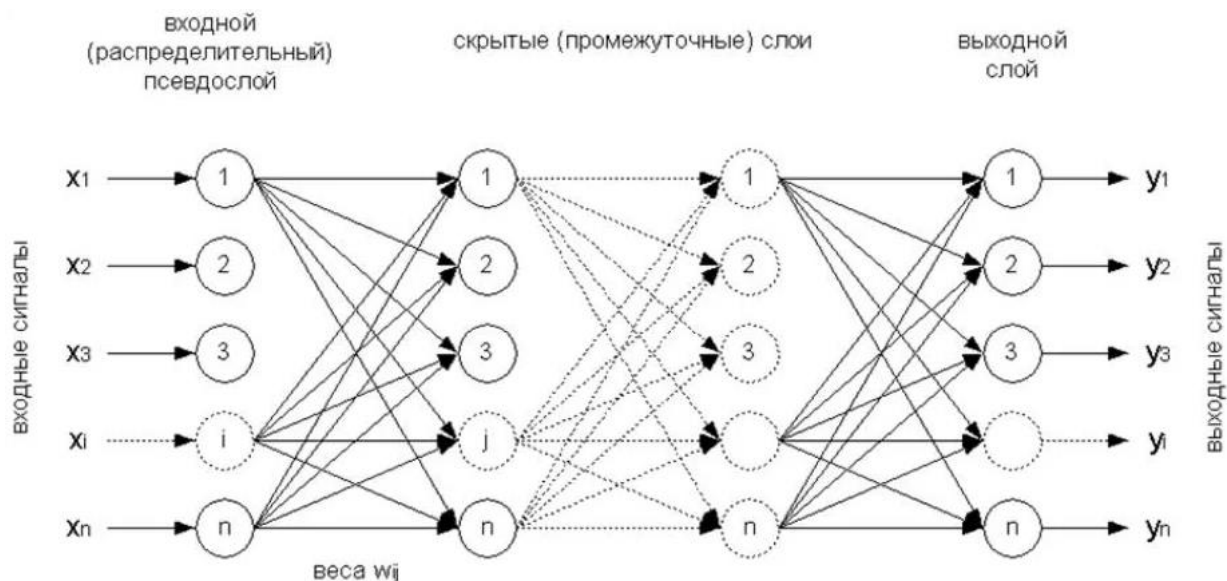


Рис. 5. Структура многослойного персептрона по Румельхарту

Примечание: составлено авторами.

Для рассматриваемого прототипа разработана модель многослойного персептрона, имеющая входной слой, получающий координаты суставов от фреймворка (рис. 6), два

скрытых слоя, в которых происходит соотношение координат и принятие решения, и выходной слой, где в качестве вывода используются утверждения left и right.

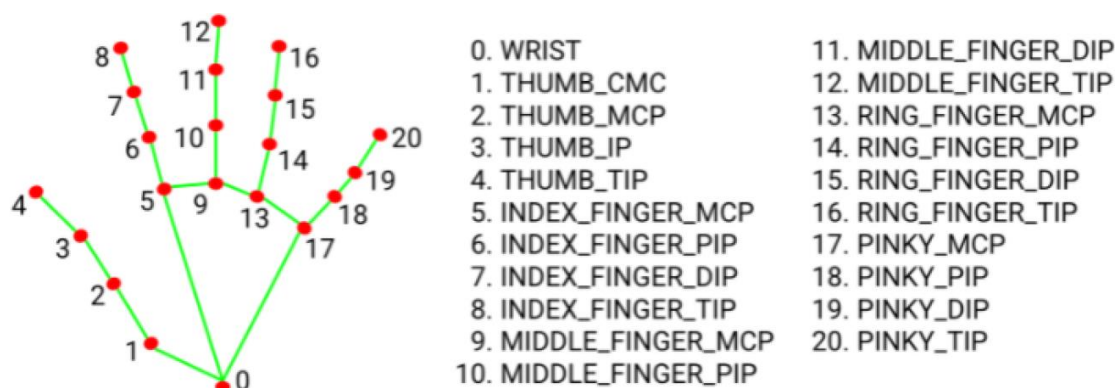


Рис. 6. Координаты ладони, получаемые из фреймворка MediaPipe Hands

Примечание: составлено разработчиками MediaPipe Hands.

Алгоритм, по которому нейронная сеть обрабатывает координаты, полученные от фреймворка MediaPipe Hands, для дальнейшего определения жеста, можно свести к четырем последовательным шагам (рис. 7):

1. Получение координат от фреймворка MediaPipe Hands, помещение их в массив.

2. Конвертация координат через приведение последнего (Т-15) значения к координатам [0, 0].

3. Нормализация координат, их перерасчет под заданные параметры экрана устройства, на который выводится изображение.

4. Разбиение каждой координаты [x, y] на пару [x] и [y].

Данный способ позволяет привести координаты к виду, при котором возможно распознавание их как жестов кисти руки.

① Координаты, полученные от фреймворка

T-15	T-14	T-13		T-2	T-1	T
[550, 165]	[526, 176]	[509, 188]		[644, 219]	[644, 196]	[642, 178]

② Конвертация в относительные координаты (T-15 приводится к [0, 0])

T-15	T-14	T-13		T-2	T-1	T
[0, 0]	[-24, 11]	[-17, 12]		[5, -16]	[0, -23]	[-2, -18]

③ Нормализация по высоте и ширине экрана

Ширина X и высота Y

T-15	T-14	T-13		T-2	T-1	T
[0.0, 0.0]	[-0.025, 0.0204]	[-0.0427, 0.0426]		[0.0979, 0.1]	[0.0979, 0.0574]	[0.0958, 0.024]

④ Сведение к одномерному массиву

T-15		T-14		T-13			T-2		T-1		T	
0.0000	0.0000	-0.0250	0.0204	-0.0427	0.0426		0.0979	0.1000	0.0979	0.0574	0.0958	0.0241

Рис. 7. Модель алгоритма обработки координат нейронной сети с примером

Примечание: составлено авторами.

Поскольку модель многомерного персептрона, лежащего в основе разработанной нейронной сети, представляет собой крайне сложное сплетение нейронов и связей между

ними, в качестве иллюстрации используется упрощенная схема, показывающая основные слои, а также входные и выходные данные (рис. 8).

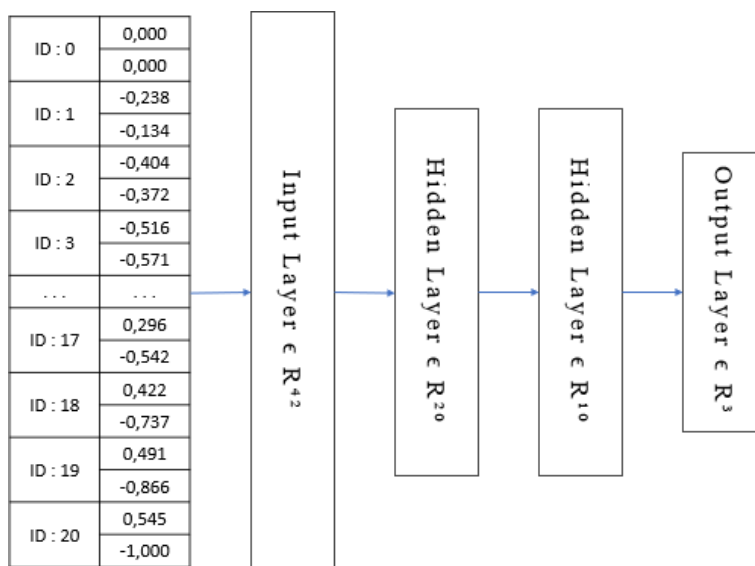


Рис. 8. Упрощенная модель многомерного персептрона, лежащего в основе прототипа

Примечание: составлено авторами.

Использование прототипа нейронной сети в задаче идентификации правой и левой руки. Для получения представления о работе экспериментального образца нейронной сети было решено использовать распознавание правой и левой руки.

Определение происходит через построение трехмерного скелета ладони [8], при этом бо-

лее яркими белыми кругами обозначены суставы кисти, находящиеся ближе остальных к камере (рис. 9).

Данный метод позволяет компенсировать зеркальность изображения и вывести истинное наименование руки в кадре. Успешность с учетом положений, в которых идентификация ладони невозможна, составляет 93 %.

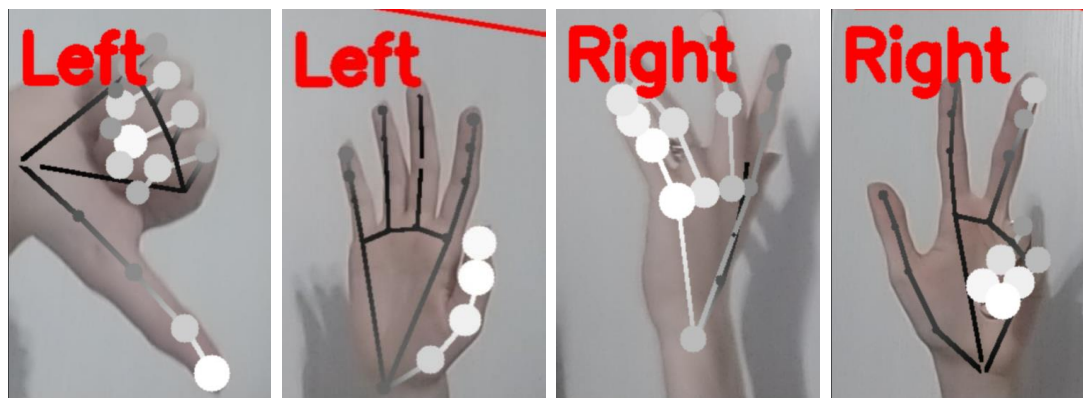


Рис. 9. Результаты использования прототипа нейронной сети

Примечание: фото авторов.

Результаты экспериментального исследования с использованием предельных имитационных условий могут быть обобщены в следующих утверждениях, а также проиллюстрированы примерами работы нейронной сети при комбинировании экспериментальных условий:

1. Нейронная сеть считывает жесты с помощью камеры в реальном времени, при этом отсутствует необходимость хранения в памяти изображений, выполненных в процессе считывания.

2. Нейронная сеть считывает жесты в условиях низкой (0,1 лк) и повышенной освещенности (10 000 лк), которая приводит к снижению уровня контрастности между фоном и распознаваемой ладонью.

3. Нейронная сеть считывает жесты в условиях нечеткости границ между фоном и ладонью.

Данные выводы свидетельствуют о соответствии выбранной нейронной сети заявленному требованию о достаточном уровне распознавания жестов кисти рук (рис. 10).

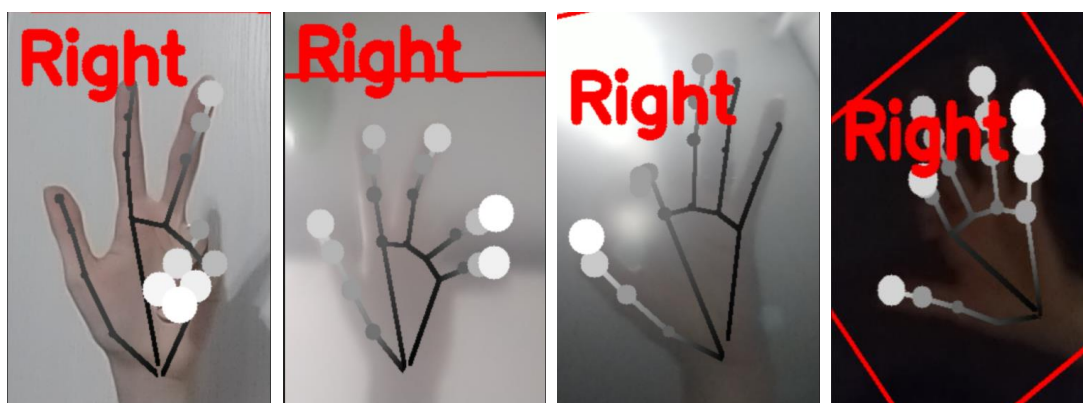


Рис. 10. Результаты использования прототипа нейронной сети в условиях, имитирующих реальные: 1 – нормальные условия; 2 – условия повышенной освещенности; 3 – условия размытости границ и оптического дефекта; 4 – условия низкой освещенности

Примечание: фотографии авторов.

ЗАКЛЮЧЕНИЕ

На данном этапе разработки программного обеспечения для «Диалогового сурдопереводчика» очевидно, что наибольшую сложность представляет создание подсистемы «жест – звук», особенно в части распознавания и перевода жеста. Разработанный

экспериментальный прототип показал верность выбранных методов и может быть использован при дальнейшем усложнении нейронной сети для работы как со статическими, так и с динамическими жестами.

Учет необходимости работы не только с жестами, но и с речью собеседника в данной

работе может считаться перспективным нововведением, способным дать толчок дальнейшему развитию технологий инклюзии людей с коммуникативными трудностями.

По результатам проектирования и разработки получено свидетельство о регистрации программы для ЭВМ.

Список источников

1. Бизюкин Г. А., Майков К. А. Адаптивный метод распознавания динамических жестов // Новые информ. технологии в автоматизир. системах. 2017. № 20. URL: <https://cyberleninka.ru/article/n/adaptivnyy-metod-raspoznavaniya-dinamicheskikh-zhestov> (дата обращения: 15.12.2021).
2. Тухбатуллин М. С., Кирпичников А. П., Ляшева С. А., Шлеймович М. П. Распознавание динамических жестов на основе вычитания фона // Вестник Казан. технолог. ун-та. 2016. № 18. URL: <https://cyberleninka.ru/article/n/raspoznavanie-dinamicheskikh-zhestov-na-osnove-vychitaniya-fona> (дата обращения: 15.11.2021).
3. MediaPipe Hands. URL: <https://google.github.io/mediapipe/solutions/hands.html> (дата обращения: 17.12.2021).
4. Zhang F., Bazarevsky V., Vakunov A. et al. MediaPipe Hands: On-device Real-time Hand Tracking // arXiv. 2020. URL: <https://arxiv.org/pdf/2006.10214.pdf> (дата обращения: 17.12.2021).
5. Lugaresi C., Tang J., Nash H. et al. MediaPipe: A Framework for Building Perception Pipelines // Third Workshop on Computer Vision for AR/VR at IEEE Computer Vision and Pattern Recognition (CVPR). 2019. URL: <https://arxiv.org/pdf/1906.08172.pdf> (дата обращения: 17.12.2021).
6. Маликов А. В. Адаптация структуры диагностической искусственной нейронной сети при появлении новых обучающих примеров // Тр. учеб. заведений связи. 2020. № 4. URL: <https://cyberleninka.ru/article/n/adaptatsiya-struktury-diagnosticheskoy-iskusstvennoy-neyronnoy-setipri-poyavlenii-novykh-obuchayushchih-primerov> (дата обращения: 20.02.2022).
7. Тормозов В. С., Василенко К. А., Золкин А. Л. Настройка и обучение многослойного перцептрона для задачи выделения дорожного покрытия на космических снимках города // Программные продукты и системы. 2020. № 2. URL: <https://cyberleninka.ru/article/n/nastroyka-i-obucheniye-mnogosloynogo-perseptrona-dlya-zadachi-vydeleniya-dorozhnogo-pokrytiya-na-kosmicheskikh-snimkah-goroda> (дата обращения: 20.02.2022).
8. Graph API. URL: <https://docs.opencv.org/4.x/d0/d1e/gapi.html> (дата обращения: 20.02.2022).

Информация об авторах

В. О. Семенова – бакалавр.
Л. Л. Семенова – старший преподаватель.

References

1. Bizyukin G. A., Maikov K. A. Adaptivnyi metod raspoznavaniia dinamicheskikh zhestov // Novye inform. tekhnologii v avtomatizir. sistemakh. 2017. No. 20. URL: <https://cyberleninka.ru/article/n/adaptivnyy-metod-raspoznavaniya-dinamicheskikh-zhestov> (accessed: 15.12.2021). (In Russian).
2. Tukhbatullin M. S., Kirpichnikov A. P., Lyasheva S. A., Shleymovich M. P. Raspoznavanie dinamicheskikh zhestov na osnove vychitaniia fona // Bulletin of the Technological University. 2016. No. 18. URL: <https://cyberleninka.ru/article/n/raspoznavanie-dinamicheskikh-zhestov-na-osnove-vychitaniya-fona> (accessed: 15.11.2021). (In Russian).
3. MediaPipe Hands. URL: <https://google.github.io/mediapipe/solutions/hands.html> (accessed: 17.12.2021).
4. Zhang F., Bazarevsky V., Vakunov A. et al. MediaPipe Hands: On-device Real-time Hand Tracking // arXiv. 2020. URL: <https://arxiv.org/pdf/2006.10214.pdf> (accessed: 17.12.2021).
5. Lugaresi C., Tang J., Nash H. et al. MediaPipe: A Framework for Building Perception Pipelines // Third Workshop on Computer Vision for AR/VR at IEEE Computer Vision and Pattern Recognition (CVPR). 2019. URL: <https://arxiv.org/pdf/1906.08172.pdf> (accessed: 17.12.2021).
6. Malikov A. V. Adaptation of the Diagnostic Artificial Neural Network Structure When New Training Examples Appear // Proceedings of Telecommunication Universities. 2020. No. 4. URL: <https://cyberleninka.ru/article/n/adaptatsiya-struktury-diagnosticheskoy-iskusstvennoy-neyronnoy-setipri-poyavlenii-novykh-obuchayushchih-primerov> (accessed: 20.02.2022). (In Russian).
7. Tormozov V. S., Vasilenko K. A., Zolkin A. L. Setting up and Training a Multilayer Perceptron for the Problem of Highlighting the Road Surface in the City Space Images // Software & Systems. 2020. No. 2. URL: <https://cyberleninka.ru/article/n/nastroyka-i-obucheniye-mnogosloynogo-perseptrona-dlya-zadachi-vydeleniya-dorozhnogo-pokrytiya-na-kosmicheskikh-snimkah-goroda> (accessed: 20.02.2022). (In Russian).
8. Graph API. URL: <https://docs.opencv.org/4.x/d0/d1e/gapi.html> (accessed: 20.02.2022).

Information about the authors

V. O. Semenova – Bachelor's Degree Student.
L. L. Semenova – Senior Lecturer.