Разработка алгоритма распознавания эмоций человека с использованием сверточной нейронной сети средствами Python

В.В. Семенюк[1], М.В. Складчиков[2]

[1]Южно-Российский государственный университет (НПИ) имени М.И. Платова, г. Новочеркасск
[2]Донецкий национальный технический университет

Аннотация: В данной статье рассматривается проблема анализа и распознавания эмоций человека с помощью обработки звуковых данных. Ввиду увеличения сфер применения, что в большей степени вызвано сложной эпидемиологической ситуацией в мире, решение описанной задачи является актуальным вопросом. Описаны основные этапы: аудиопоток данных записывается в аудиофайл и в соответствии с подходом «дактилоскопии звука» преобразуется в изображение, являющееся спектрограммой звукового набора данных. Описаны этапы обучения сверточной нейронной сети на заранее заготовленном наборе звуковых данных, а также описана структура алгоритма. Для валидации нейронной сети был отобран иной, не участвующий в тренировке, набор аудиоданных. В результате проведения исследования, были построены графики, демонстрирующие точность работы предлагаемого метода.
Ключевые слова: нейронная сеть, распознавание эмоций человека, сверточная нейронная сеть, дактилоскопия звука, Tenserflow, Keras, Matlab, Deep Network Toolbox.

Введение

Каждый человек выражает эмоции при возникновении определённых внешних или внутренних возбудителей. Существуют сложности в классификации эмоций из-за персонифицированности их выражения [1-3]. Известны различные подходы к их идентификации и классификации [4-6].

Сверточные нейронные сети (СНС) применяются для анализа изображений [7, 8]. Поэтому в данной работе оцифрованный звук преобразовывался в изображение с использованием методики, известной как «дактилоскопия звука» [9]. Такое преобразование исходного звукового сигнала позволило снизить временные затраты на обработку входных данных.

Исследования в области классификации эмоций проводятся уже длительное время, мотивированные низкой точностью текущих алгоритмов, что подчеркивает необходимость дальнейших исследований в данной сфере.

Большинство предыдущих работ, рассмотренных в контексте стратегии исследования, ориентировались на классификацию эмоций на основе видеопотока данных, используя опорные точки лица в качестве аттракторов. Основной задачей нейронной сети является построение карты точек лица, предоставляя необходимые данные для обучения. Из-за требования к сложным вычислениям для данной задачи широко применяются сверточные нейронные сети. В качестве исходных данных для СНС выступает база изображений. Сеть выделяет ключевые признаки, которые служат основой для классификации эмоций по характерным лицевым сегментам. Путем использования специализированных наборов данных, созданных в контролируемых условиях и направленных на использование в обучении нейронных сетей, достигается создание оптимального алгоритма для распознавания [10, 11].

Целью данной работы является разработка алгоритма идентификации эмоций человека на основании набора звуковых данных. Новизна предлагаемого метода заключается в том, что в качестве инструмента для идентификации выбрана сверточная нейронная сеть, являющаяся нестандартным вариантом для её применения.

Методики оценки эмоционального состояния

Если проанализировать текущие модели, направленные на распознавание эмоционального состояния, то большая их часть базируется на видеоданных. Несмотря на распространённость, такие модели имеют громоздкую архитектуру, что усложняет процесс обучения и делает систему менее быстродействующей. Альтернативным вариантом является использование голосового набора данных.

В качестве объекта исследования использовались аудиозаписи с характерными признаками различных эмоций. Сам процесс распознавания выполнялся с использованием сверточной нейронной сети,

оптимизированной для работы с аудиозаписями и текстом в рамках эксперимента с целью увеличения быстродействия. Была осуществлена трансформация входного звукового потока в изображение по методике "audio fingerprint". Таким образом, объектом для последующего анализа с помощью СНС служила спектрограмма.

Построение модели нейронной сети

На рис. 1 представлена блок схема классов эмоций человека.

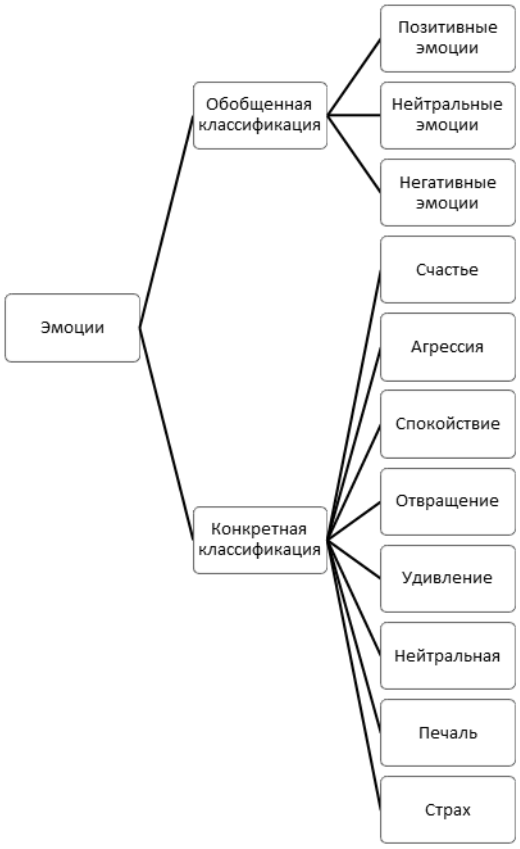


Рис. 1. – Иерархия классов эмоций

Для проведения эксперимента было решено разделить весь обучающий набор данных на две модели: малый набор данных (позитивные, негативные и нейтральные) и расширенный (агрессия, спокойствие, отвращение, страх, счастье, нейтральное состояние, печаль, удивление). Данное разделение обусловлено в первую очередь необходимостью понятия архитектурной

сборки модели и «механизма» обучения нейронной сети для более точной классификации. В качестве набора данных, необходимого для обучения, была выбрана библиотека голосовых команд, которая идеально подходит под поставленные задачи.

Для обучения нейронной сети, в соответствии с рис.1, был отобран набор аудиофайлов (табл. 1, 2).

Таблица №1

Количество аудио файлов для каждого класса обучающей и тестовой выборок (8 классов)

Класс		Количество изображений обучающей выборки	Количество изображений тестовой выборки
1.	Агрессия	665	167
2.	Спокойствие	299	75
3.	Отвращение	519	130
4.	Страх	665	167
5.	Счастье	665	167
6.	Нейтральное	244	113
7.	Печаль	346	87
8.	Удивление	200	50
Всего		3603	956

Таблица №2

Количество аудио файлов для обобщенных классов обучающей и тестовой выборок (3 класса)

Класс	Количество изображений обучающей выборки	Количество изображений тестовой выборки
Позитивные	865	217
Нейтральные	543	188
Негативные	2195	551
Всего	3603	956

Затем, с помощью специального алгоритма, аудиофайлы были преобразованы в набор спектрограмм (рис.2).

а) б)

Рис. 2. – Временное представление (а) и спектрограмма (б)

Структура нейронной сети, реализующая алгоритм, состоит из следующих слоев:

- сверточный слой (3);
- слой выделения признаков (1);
- слой коррекции (2);
- конвертер из сверточной структуры в многослойную (1);
- слой многоосной нейронной сети (2).

Были разработаны 2 модели СНС для 8 (подробных) и 3 (общих) классов. Структура моделей идентична за исключением выхода, количество нейронов на котором должно быть равно количеству классов.

Результаты исследования

Модели имеют полностью одинаковую структуру и состоят из практически одинакового набора файлов:

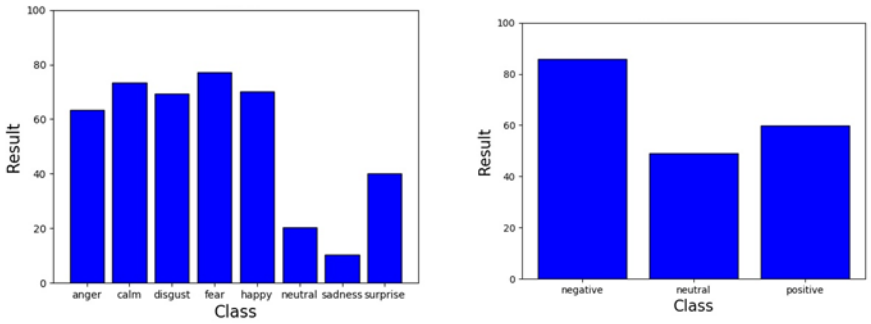
- train – необходим для обучения нейронной сети;
- test – используется для тестирования работоспособности модели на конкретной эпохе. Данный файл необходим для корректирования дальнейшего обучения модели;
- single_test – данный script необходим для проверки работоспособности готовой модели нейронной сети. Принципиальное отличие от предыдущего

файла состоит в том, что в данном случае для проверки используются изображения, которые не участвовали в обучении.

Описанные выше scrip-файлы идентичны для каждого случая распознавания. Единственное отличие – количество выходных слоёв (оно соответствует количеству распознаваемых эмоций).

Рассмотрим процесс оценки эмоционального состояния более подробно.

После обучения нейронной сети на вход поступали информационные данные, разделенные по группам. Каждая группа характеризовала конкретную эмоцию и состояла из 100 файлов, соответственно. На рис.3 показан результат тестирования.



а) б)

Рис. 3. – Результаты тестирования для 8 классов (а) и для 3 классов (б)

Тестирование для распознавания 3 эмоций является наиболее простым вариантом, которые показывает отличную точность. Это обусловлено тем, что каждая эмоция, с точки зрения физиологии, имеет общие параметры, которые присущи разным эмоциональным состояниям.

В тестировании участвовали эмоции, приведенные в таблице 1.

Результаты исследований показали, что корректно распознаются не все эмоции. Нейронная сеть показала хорошие результаты при распознавании эмоций № 1, 3, 4, 5 и 7 (см. табл. 1). Эмоции № 6 и 8 (см. табл. 1)

распознаются нейронной сетью некорректно, относя их к негативным эмоциям.

На основании результатов, представленных на рис.3, имеем следующие показатели точности:

1. Распознавание 3 эмоций:

• Negative – 85%;

• Neutral – 49, 3%;

• Positive – 57, 2%.

2. Распознавание 8 эмоций:

• Anger – 63,2%;

• Calm – 74,1%;

• Disgust – 68,9%;

• Fear – 75, 7%;

• Happy – 72, 1%;

• Neutral – 22, 06%;

• Sadness – 8, 9%;

• Surprise – 37, 9%.

Как видно из полученных результатов, точность была недостаточной, что требовало усовершенствования разработанного алгоритма. Для этого было решено изменить подход к реализации самого алгоритма. Первый вариант подразумевал изменение архитектуры нейронной сети, однако, вид набора данных оставался неизменным. На основании второго варианта требовалось изменить вид набора данных, а архитектуру оставить неизменной.

Для осуществления данного исследования было решено перейти в другой пакет для разработки нейронной сети. На основании проведенного литературного обзора, было выявлено несколько наиболее подходящих программ для разработки. Однако наиболее презентабельным и удобным

оказался язык программирования Matlab. В нём имеется специализированное дополнение для разработки нейронных сетей, которые подразумевают использование уже готовых архитектур или создание собственной модели искусственного интеллекта.

Данное расширение предназначено для изучения и создания моделей глубокого машинного обучения. Deep Learning Toolbox предоставляет мощный инструментарий, который позволяет создавать как новые, так и использовать предварительно обученные модели машинного обучения. Это дополнение обеспечивает возможность работы с сверточными нейронными сетями и сетями долгой кратковременной памяти (LSTM).

Преимуществом использования данного расширения является возможность создания нейронных сетей без глубокого погружения в математический аппарат конкретной архитектуры нейронной сети. По информационному сопровождению пользователя есть возможность узнать основные концепции построения и обучения нейронных сетей.

Для закрытия пробелов в области глубокого обучения корпорация MathWorks разработала специализированное образовательное дополнение, известное как Matlab Onramp. Это дополнение поставляется вместе с приобретенной библиотекой и позволяет пользователям ознакомиться с интерфейсом и базовыми операциями с нейронными сетями.

Корпорация MathWorks осознавала, что она не единственная компания, занимающаяся разработкой программных решений для обучения нейронных сетей. Поэтому основной упор был сделан на визуальную составляющую, обеспечивающую простую и интерактивную работу с алгоритмами построения и обучения моделей машинного обучения. Кроме того, внимание было уделено сокращению времени обучения.

В Matlab существует определенный метод, который условно позволяет изменять структуру нейронной сети. Этот этап позволяет создавать условные

модели машинного обучения и проводить с ними манипуляции. Путем создания временных слоев копируется архитектура нейронной сети и помещается в оперативную память. Важно отметить, что копируется не только сам слой, но и вся информация о модели.

Благодаря возможности клонирования отдельных участков или всей модели нейронной сети, можно изменять ее структуру и подстраивать под конкретные условия. Изначально было решено переобучить нейронную сеть так, чтобы размер входного изображения полностью соответствовал указанному во входном слое. Такой метод исследования был выбран с целью полного сохранения целостности структуры нейронной сети. Для сравнения и для получения большего количества практических исследований было проведено обучение с изменением размера входного слоя. На рис.4 показан результат обучения нейронной сети GoogleNet.

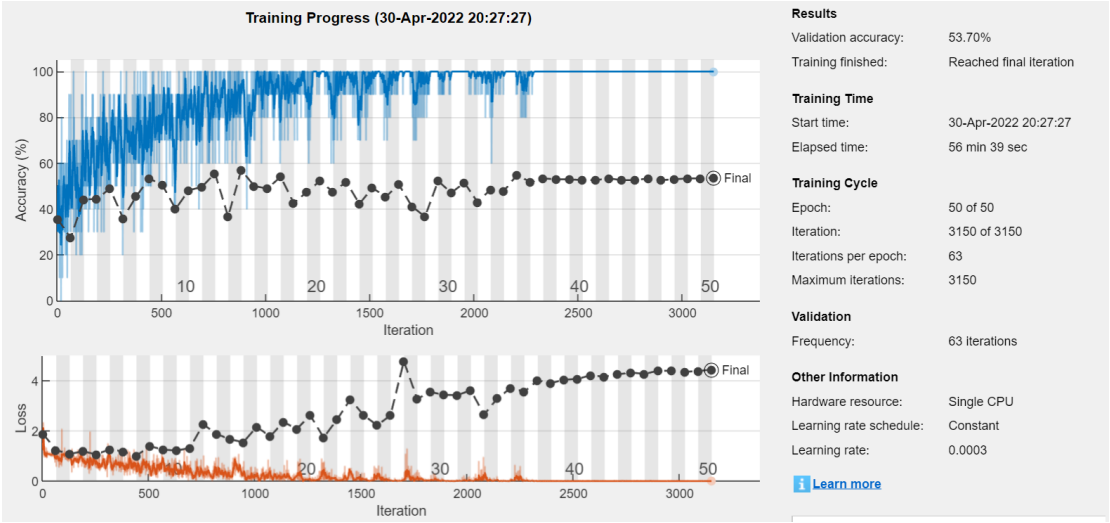


Рис. 4. – Результат обучения нейронной сети

Из рис.4 видно, что нейронная сеть не смогла достаточно обучиться для распознавания трех эмоций. Низкая точность и ухудшающийся график ошибок указывают на то, что данный метод обучения не пригоден. Та же самая ситуация наблюдается с графиком обучения ResNet-50. Из проведенного эксперимента можно сделать вывод: изменение структуры

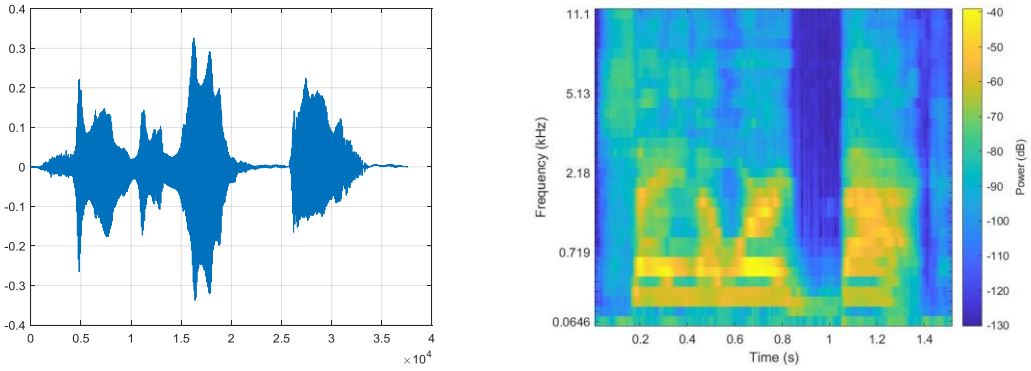
нейронной сети при использовании неизменного входного набора данных не привело к увеличению точности; в некоторых случаях она даже уменьшилась. Это свидетельствует, прежде всего, о том, что замена структуры нейронной сети не способствует достижению более высоких результатов.

Следовательно, было решено провести исследование, позволяющее оценить степень изменение входного набора данных при неизменной архитектуре нейронной сети. В качестве иного подхода к образованию пакета входного набора данных, было решено выбрать MFCC.

MFCC обладает рядом преимуществ перед своими аналогами, основным из которых является способность учитывать слуховые характеристики человека. Значительным достоинством также является возможность преобразования неоднородных частот в однородные благодаря проведенному анализу. Это преобразование позволяет перейти от нелинейного алгоритма описания слуховых характеристик к более простой и понятной линейной системе.

Основные вычисления осуществлялись с использованием встроенной команды mfcc. В роли операторов команда принимает массив данных из звукового файла, частоту дискретизации и параметры, устанавливающие ширину окна, ограничивающего обрабатываемый диапазон звукового файла. На рис.5 представлен периодический сигнал, описывающий отдельный звуковой файл, и его спектрограмма Мела.

После получения полного набора спектрограмм Мела, они были поданы на входной слой исследуемых нейронных сетей. Важно отметить, что время обучения значительно сократилось по сравнению с предыдущим вариантом, и точность работы алгоритма заметно увеличилась. График результатов обучения нейронной сети ResNet-50 представлен на рис.6, а GoogleNet на рис.7.



а) б)

Рис. 5. – Аудио сигнал (а), спектрограмма MFCC (б)

Как можно заметить из представленных графиков, точность алгоритма была улучшена. Результаты этого эксперимента подчеркивают, что повышение точности работы алгоритма в данном контексте напрямую зависит от характеристик входных данных. Важно отметить, что использование MFCC обеспечивает более точное и полное представление, что является значимым фактором для нейронной сети. Уточнение голосовых характеристик позволяет более точно выделять признаки на каждом изображении, что, в свою очередь, положительно сказывается на точности распознавания.

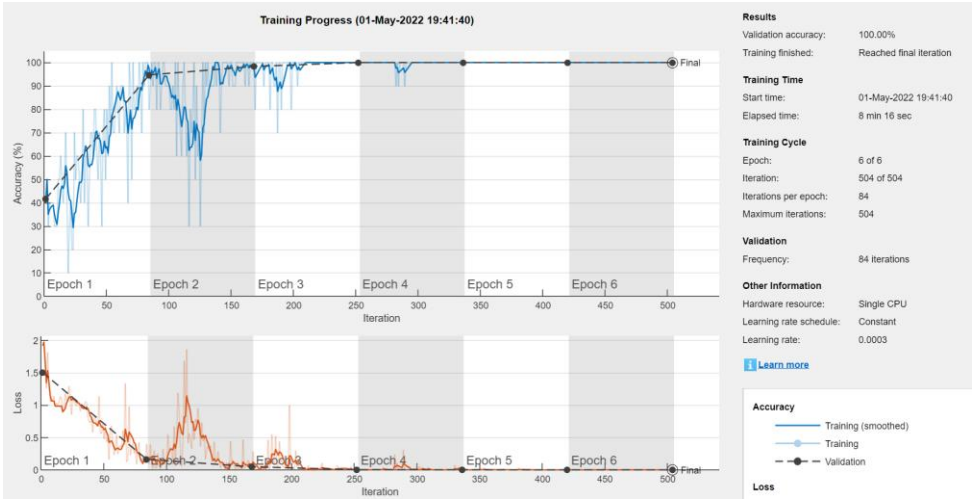


Рис. 6. – Результат обучения сети ResNet-50 с использованием MFCC

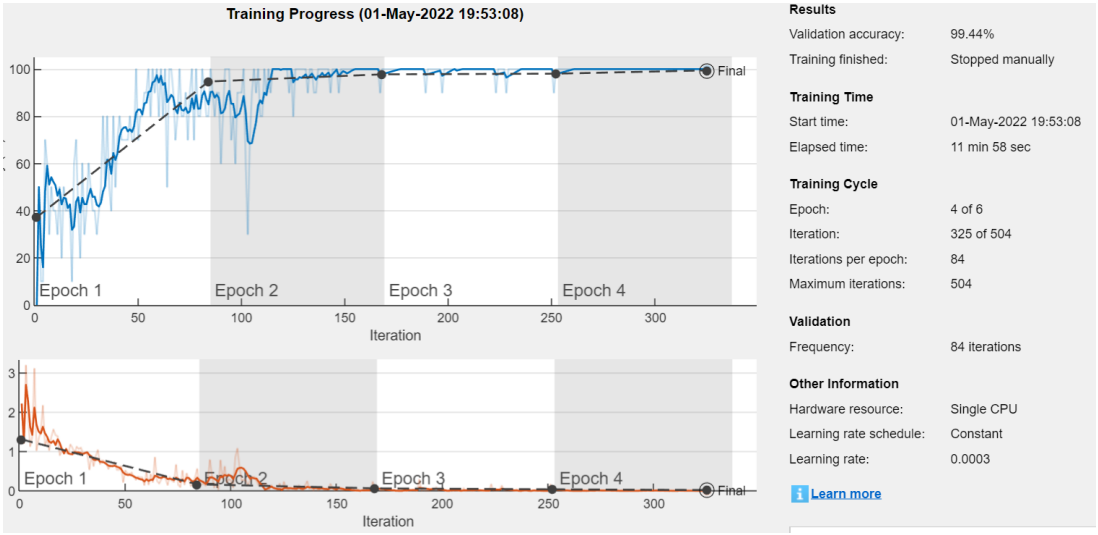


Рис. 7. – Результат обучения сети GoogleNet с использованием MFCC

Выводы

Как показано на рис.3, проведение тестирования представляет собой задачу неоднозначную, из-за трудности определения человеческих эмоций. Набор тестов не обеспечивает определенного результата, поэтому был проведен индивидуальный анализ для каждого образца. Применение обобщенных классов привело к более точным результатам, в отличие от использования конкретных классов. Следует отметить, что из-за меньшего числа классов скорость работы программы существенно выше при классификации 3 классов, чем при 8 классах. Таким образом, программа требует меньше процессорного времени для выполнения.

В дальнейшем планируется провести исследование различных математических пакетов для построения более сложной концептуальной модели. Это позволит улучшить качество распознавания и позволит обрабатывать более сложные наборы данных.

Литература

1. Марьев А.А. Метод интерпретации результатов измерений параметров речевого сигнала в задачах диагностики психоэмоционального состояния

человека по его речи // Инженерный вестник Дона, 2011, №4. URL: ivdon.ru/ru/magazine/archive/n4y2011/538

2. Изард К.Э. Психология эмоций. СПб.: Питер, 2012. 464 с.

3. Сидоров К.В., Ребрун И.А., Кожевников Д.Д., Соботницкий И.С. Диагностика психофизиологического и эмоционального состояния человека-оператора // Инженерный вестник Дона, 2012, №4(2). URL: ivdon.ru/ru/magazine/archive/n4p2y2012/1480

4. Перервенко Ю.С., Старченко И.Б. Эмоциональная речь: детерминированный хаос или нелинейный случайный процесс // Известия ЮФУ. Технические науки. 2008. №1(78). С. 100-101.

5. Галичий Д.А., Афанасьев Г.И., Нестеров Ю.Г. Распознавание эмоций человека при помощи современных методов глубокого обучения // E-SCIO. 2021. Т.5. №56. С. 316-329.

6. Zhang C., Xue L. Autoencoder with emotion embedding for speech emotion recognitio // IEEE access. 2021. V.9. pp. 51231-51241.

7. Li Z. Liu F., Yang W., Peng Sh., Zhou J. A survey of convolutional neural networks: analysis, applications, and prospects // IEEE Trans Neural Network Learn Systems. 2022. V.33(12). pp. 6999-7019.

8. Baxtiyarovich K. D. Convolutional neural networks for image recognition // International journal of advanced research in education, technology and management. 2023. V.2. №3. URL: ijaretm.com/index.php/ij/article/view/224

9. Cano P. A review of audio fingerprinting // Journal of VLSI signal processing systems for signal, image and video technology. 2005. V.41. pp. 271-284.

10. Hassani B., Mahoor M. Facial Expression Recognition Using Enhanced Deep 3D Convolutional Neural Networks. 2017 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW). 2017. URL: doi.org/10.1109/CVPRW.2017.282

11. Byoung C.K. A Brief Review of Facial Emotion Recognition Based on Visual Information // Sensors. 2018. V.18 (2). URL: doi.org/10.3390/s18020401

References

1. Mar'ev A.A. Inzhenernyj vestnik Dona, 2011, №4. URL: ivdon.ru/ru/magazine/archive/n4y2011/538

2. Izard K.E. Psikhologiya emotsiy [Psychology of emotions]. SPb.: Piter, 2012. 464 p.

3. Sidorov K.V., Rebrun I.A., Kozhevnikov D.D., Sobotnitskiy I.S. Inzhenernyj vestnik Dona, 2012, №4(2). URL: ivdon.ru/ru/magazine/archive/n4p2y2012/1480

4. Perervenko Yu.S., Starchenko I.B. Izvestiya YuFU. Tekhnicheskie nauki. 2008. №1(78). pp. 100-101.

5. Galichiy D.A., Afanas'ev G.I., Nesterov Yu.G. E-SCIO. 2021. V.5. №56. pp. 316-329.

6. Zhang C., Xue L. IEEE access. 2021. V.9. pp. 51231-51241.

7. Li Z. Liu F., Yang W., Peng Sh., Zhou J. IEEE Trans Neural Network Learn Systems. 2022. V.33(12). pp. 6999-7019.

8. Baxtiyarovich K. D. International journal of advanced research in education, technology and management. 2023. V.2. №3. URL: ijaretm.com/index.php/ij/article/view/224

9. Cano P. Journal of VLSI signal processing systems for signal, image and video technology. 2005. V.41. pp. 271-284.

10. Hassani B., Mahoor M. Facial Expression Recognition Using Enhanced Deep 3D Convolutional Neural Networks. 2017 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW). 2017. URL: doi.org/10.1109/CVPRW.2017.282

11. Byoung C.K. Sensors. 2018. V.18 (2). URL: doi.org/10.3390/s18020401

Дата поступления: 18.11.2023

Дата публикации: 28.12.2023