УДК 621.391

**АЛГОРИТМ РАСПОЗНАВАНИЯ ЭМОЦИЙ ПО ИЗОБРАЖЕНИЮ ЛИЦА ЧЕЛОВЕКА НА ОСНОВЕ СВЕРТОЧНОЙ НЕЙРОННОЙ СЕТИ**

**Л.И. Ивановский –** аспирант, Ярославский государственный университет им. П.Г. Демидова

E-mail: leonel-unknown@yandex.ru

**В.В. Хрящев** – к.т.н., доцент, Ярославский государственный университет им. П.Г. Демидова

E-mail: vhr@yandex.ru

**А.А.Лебедев** – аспирант, Ярославский государственный университет им. П.Г. Демидова

E-mail: lebedeves@gmail.com

**О.?.Степанова** – аспирантка, Ярославский государственный университет им. П.Г. Демидова

E-mail: ?

**А.Л.Приоров** – д.т.н., доцент, Ярославский государственный университет им. П.Г. Демидова

E-mail: andcat@yandex.ru

**АННОТАЦИЯ**

В данной работе описан алгоритм распознавания эмоций по изображению лица человека. В основе разработанного алгоритма лежит реализация сверточной нейронной сети. Данная нейронная сеть способна классифицировать изображения лиц людей на один из шести классов: спокойствие, улыбка, удивление, заинтересованность, отвращение, крик. Стадии обучения и тестирования осуществлялись на графическом процессоре суперкомпьютера, с использованием изображений из базы Multi-PIE. Для разработанной сверточной нейронной сети была получена матрица неточностей, были построены графики изменения доли правильных ответов и функции потерь в зависимости от количества проделанных обучающих итераций, а также были получены значения метрик оценки качества алгоритма.

***Ключевые слова****: глубокое машинное обучение, сверточная нейронная сеть, распознавание эмоций.*

**ANNOTATION**

This paper presents an algorithm for expression recognition using facial images. The developed algorithm is based on the implementation of convolutional neural network. The aim of this network is to classify facial images into one of the six types of emotions: neutral, smile, surprise, squint, disgust and scream. Learning and testing processes of algorithm were performed using face images from the CMU MultiPie database. To accelerate the neural network operations, the training and testing processes were performed parallel, on a large number of independent streams on GPU. For developed models there were given metrics of quality.

***Keywords:*** *deep machine learning, convolutional neural network, facial expression recognition.*

**ВВЕДЕНИЕ**

Известно, что человек способен определять эмоции, вне зависимости от возраста людей, их пола или расовой принадлежности. Однако для компьютера поставленная задача распознавания эмоций является трудно выполнимой. Сложности в данной задаче заключаются в разнообразии лиц, а также в степени и частоте выражения эмоций. На практике, в задачах компьютерного зрения дополнительные трудности могут также возникнуть в связи с наличием шума и оптических препятствий на изображениях, недостаточной освещенностью сцены или малым разрешением картинок.

Целью данной работы является разработка и исследование алгоритма классификации эмоций по изображению лица человека на основе сверточных нейронных сетей.

На сегодняшний день использование подобного рода алгоритмов популярно в области видеоаналитики, для оценки действия рекламы, для сбора статистики в масштабах торгового центра и для анализа работы персонала при общении с клиентом. Эмоции человека могут дать информацию о его внутреннем состоянии и переживаниях. Благодаря распознанным эмоциям возможно предсказание дальнейшего поведения человека. Такие результаты позволяют применить поставленную задачу в сфере безопасности и противодействия терроризму, для поиска злоумышленников и преступников по изображениям, поступающим с камер видеонаблюдения. Также совместное использование алгоритмов распознавания речи и эмоций может быть применено в области психологии, психиатрии и криминалистики [1].

Основная сложность поставленной задачи заключается в том, что некоторые типы эмоций мало отличаются друг от друга, порой всего несколькими, незначительными чертами, зависящими, в частности, от индивидуальных особенностей каждого человека. Кроме того, выражение лица человека может быть спонтанным. Идеальный алгоритм распознавания эмоций должен учитывать все эти факторы для получения точных результатов [2].

В большинстве работ, посвященных тематике поведения человека, представлено от 6 до 8 различных типов эмоций: спокойствие, улыбка, удивление, заинтересованность, отвращение, крик, грусть и злость. Примеры изображений с такими типами эмоций приведены на рис.1.

а) Спокойствие б) Улыбка в) Удивление

г) Заинтересованность д) Отвращение е) Крик

Рис. 1. 6 типов эмоций, представленных в базе изображений Multi-Pie

Задача алгоритма машинного обучения для решения поставленной задачи заключается в детектировании лица в видеопотоке и в выделении признаков для последующей классификации изображения в соответствии с обнаруженным типом эмоций.

Для детектирования лица человека могут быть использованы разнообразные подходы, от классического алгоритма Виолы-Джонса до вычислительно эффективного алгоритма PICO или алгоритмов, основанных на применении сверточных нейронных сетей. В большинстве ситуаций подобного рода алгоритмы способны детектировать более чем 90% лиц. На практике, для данной задачи существует два принципиально отличающихся друг от друга подхода по обнаружению лиц в видеопотоке: детектирование в каждый момент времени, на каждом отдельном кадре или детектирование на первом кадре и последующее отслеживание. Далее, распознавание эмоций может происходить как для каждого кадра по отдельности, так и для видеофрагмента в целом [3].

Для решения задачи классификации в данном случае используется сверточная нейронная сеть, которая позволяет полностью отойти от этапа формирования признаков, поскольку они формируются алгоритмом самостоятельно, в процессе обучения модели [4].

Использование такого подхода позволяет получить наилучшие результаты классификации во многих задачах. В статье [5] приведены различные типы глубокой сверточной архитектуры нейронной сети. Эти типы архитектуры позволяют решать проблемы распознавания пола и прогнозирования возраста человека по изображению его лица. В работе [6] описано использование сверточной нейронной сети в задаче классификации объектов разных типов в изображении. В статье [7] описывается двухфазная гибридная система, в которой объединены воедино сверточная нейронная сеть и логистический классификатор. На первом этапе сеть распознает изображения лиц. На втором этапе логистический классификатор используется для классификации картинок, полученных на предыдущем этапе. Доля правильных ответов такой модели составила 86,06%. В статье [8] также представлен двухфазный классификатор, чтобы классифицировать изображение лица человека в соответствии с одним из семи типов эмоций. Сверточная нейронная сеть используется для классификации нейтральных типов эмоций, а гауссова модель применяется для классификации оставшихся типов эмоций. Доля правильных ответов такой модели 71%. В статье [9] описывается алгоритм для задачи распознавания лиц. Сначала из входного изображения извлекаются значения признаков, а затем с помощью гауссовой модели и сверточной нейронной сети извлекается необходимая информация. Доля правильных ответов такой модели составила около 85%

Данная статья состоит из пяти частей. В первом части осуществляется постановка задачи. Также в ней рассматриваются основные трудности, возникающие в процессе распознавания компьютером эмоций человека. Во втором разделе обозреваются некоторые известные базы изображений с различными типами эмоций. Третья часть посвящена сверточным нейронным сетям. В ней рассказывается об особенностях использования данного подхода из области глубокого машинного обучения, а также об инструментах, с помощью которых можно реализовать и обучить классификатор такого типа. Кроме того, в данном разделе описывается архитектура сети, которая применялась для задачи распознавания эмоций по изображению лица человека, а также особенности обучения классификатора. В четвертой части приводятся результаты тестирования разработанного алгоритма. Для поставленной задачи была получена матрица неточностей, был построен график функции потерь, а также были посчитаны значения метрик оценки качества алгоритма. Затем, в заключении подводятся итоги и указываются дальнейшие планы по применению разработанной модели. В конце указывается список литературы, использованной в данной статье.

1. **ОБЗОР БАЗ ДАННЫХ С ИЗОБРАЖЕНИЯМИ**

Базы данных с изображениями являются важнейшей и неотъемлемой частью процессов обучения и тестирования, оценки эффективности и сравнительного анализа различных алгоритмов машинного обучения. На сегодняшний день, существует несколько доступных баз данных с изображениями и видеопоследовательностями.

База данных Cohn-Kanade состоит из черно-белых изображений 210 различных взрослых людей в возрасте от 18 до 50 лет, 69% которых составляют женщины, 31% - мужчины, 81% - европейцы, 13% - афроамериканцы и 6% - представители других этнических групп. Снимки были сделаны под разными углами обзора: 0° и 30° [10]. Примеры изображений из базы данных Cohn-Kanade приведены на рис.2.

База изображений JAFFE содержит 213 черно-белых изображений 10 разных японских девушек. В этой базе представлены 7 типов эмоций: спокойствие, грусть, удивление, счастье, страх, злость и отвращение [11]. Примеры снимков из базы данных JAFFE приведены на рис.3.

База данных BU-3DFE содержит 2500 цветных трехмерных моделей 100 людей, в возрасте от 18 до 70 лет, с 7 типами различных эмоций: спокойствие, гнев, отвращение, страх, счастье, грусть и удивление. Изображения сделаны с 3 разных ракурсов: +45°, 0°,−45°. Среди всех моделей 56% составляют мужские лица, а 44% женские [12]. Примеры изображений из базы данных BU-3DFE приведены на рис.4.



Рис. 2. Примеры изображений из базы данных Cohn–Kanade



Рис. 3. Примеры изображений из базы данных JAFFE



Рис. 4. Примеры изображений из базы данных BU-3DFE

База данных MMI содержит более 1500 примеров статических изображений и фотопоследовательностей 19 разных людей, выражающих различные типы эмоций. Снимки были сделаны в профиль и анфас [13]. Примеры изображений из базы данных MMI приведены на рис.5.

База данных MUG состоит из 86 последовательности фронтальных снимков людей, 35 из которых женщины и 51 – мужчины. В этой базе присутствуют картинки с 7 различными типами эмоций: злость, отвращение, страх, счастье, печаль, удивление и спокойствие. Каждый тип эмоции представлен в виде нескольких размеченных последовательностей изображений. Каждая такая последовательность содержит от 50 до 160 снимков, для каждого субъекта по отдельности [14]. Примеры изображений из базы данных MMI приведены на рис.6.

База данных Radbound содержит более 8000 цветных изображений 67 разных людей. В этой базе представлены 8 типов эмоций: спокойствие, грусть, презрение, удивление, счастье, страх, злость и отвращение. Снимки были сделаны под разными углами обзора: 0°, 45°, 90°, 135°, 180° [15]. Примеры изображений из базы данных Radbound приведены на рис.7.

База данных Multi-Pie содержит более чем 750000 цветных изображений 337 разных людей с 6 различными типами эмоций: спокойствие, улыбка, удивление, заинтересованность, отвращение и крик. Снимки были сделаны под разными углами обзора, не превосходящими 90°, и с разной освещенностью сцены [16]. Примеры изображений из базы данных Multi-Pie приведены на рис.8.

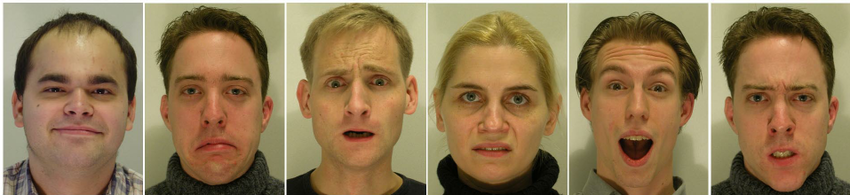


Рис. 5. Примеры изображений из базы данных MMI



Рис. 6. Примеры изображений из базы данных MUG



Рис. 7. Примеры изображений из базы данных Radboud



Рис. 8. Примеры изображений из базы данных Multi-Pie

1. **МОДЕЛИРОВАНИЕ СВЕРТОЧНОЙ НЕЙРОННОЙ СЕТИ**

В данной работе представлен алгоритм, основанный на применении сверточной нейронной сети – специальной архитектуры сети, способной достаточно быстро и качественно распознавать объекты на изображениях, а впоследствии и классифицировать их. Сверточные нейронные сети относятся к алгоритмам глубокого машинного обучения. Особенность таких сетей заключается в том, что дескрипторы изображений формируются не самим исследователем, а за счет операции двумерной свертки, при этом сами сверточные фильтры формируются в процессе обучения. Сгенерированные таким образом дескрипторы дают, как правило, наилучшие результаты [17].

Реализация архитектуры сети осуществлялась с помощью фреймворка Caffe [18] на основе модели, предложенной в [19]. Эта библиотека позволяет достаточно просто, описать сверточную нейронную сеть и параметры ее запуска в 2 файлах формата prototxt. Фреймворк Caffe может быть интегрирован в проекты, написанные на языках программирования С++ и Python. Эта библиотека позволяет также воспользоваться уже реализованными, готовыми алгоритмами машинного обучения. Фреймворк Caffe, на сегодняшний день, активно применяется для решения практических задач по предсказанию пола или возраста человека по изображению лица, а также по детектированию различных объектов на снимках [19].

Как показано на рис.9, разработанная нейронная сеть состоит из 6 сверточных слоев (3 с фильтрами и 3 с фильтрами ), 6 слоев с функцией активации , 6 слоев, реализующих процесс локальной нормализации и 3 слоев, описывающих процесс дискретизации с помощью операции max pooling. После всех проделанных операций получается одномерный вектор из 96 значений признаков. Полученный вектор передается на вход полносвязнному слою, который имеет выход из 6 значений, в соответствии с числом классов – типов различных эмоций. По этому набору значений, с помощью обобщенной логистической функции Softmax нетрудно вычислить апостериорные вероятности для каждого класса. Благодаря такому вектору значений вероятностей исходное изображение сопоставляется с соответствующим выражением, обнаруженным на лице человека. Другими словами, нейронная сеть классифицирует входное изображение в соответствии с наиболее вероятным типом эмоции.

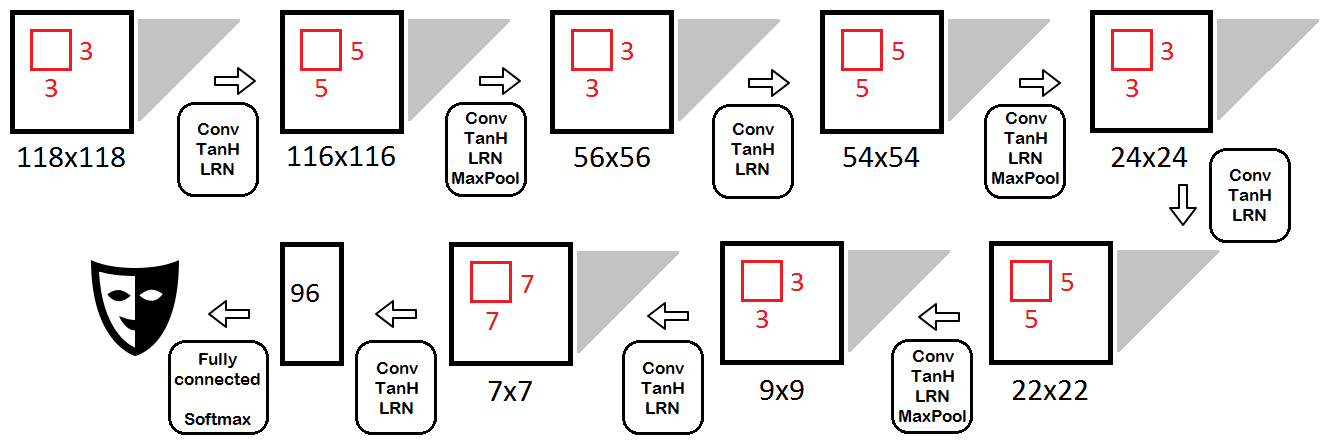


Рис. 9. Архитектура сверточной нейронной сети

Подход с использованием сверточных нейронных сетей имеет большую вычислительную сложность. В связи с этим, процессы обучения и тестирования разработанной модели осуществлялись на большом числе независимых потоков графического процессора видеокарты. Для этого была использована технология параллельных вычислений NVIDIA CUDA. CUDA позволяет решить проблему высокой ресурсоемкости алгоритмов с помощью распараллеливания вычислений на графическом процессоре видеокарты [20]. Эта технология является кроссплатформенной и поддерживается всеми современными видеокартами фирмы NVIDIA.

Помимо файла с описанием структуры сверточной нейронной сети, для обучения разработанной на Caffe модели необходим также файл с описанием значений некоторых параметров. Базовая (начальная) скорость обучения сети была равной 0,01. В качестве алгоритма численной оптимизации использовался стохастический градиентный спуск (SGD) с использованием добавочного моментного коэффициента равного 0,9. Для регуляризации модели в процессе обучения применялось правило обновления весов, в результате чего добавлялся коэффициент распада веса равный 0,0005. Классификатор заканчивал свое обучение после совершения 70000 итераций.

1. **РЕЗУЛЬТАТЫ ТЕСТИРОВАНИЯ**

Для выполнения численного эксперимента была подготовлена выборка из 210000 случайно отобранных снимков (35000 изображений для каждого из 6 типов эмоций, представленных в базе данных Multi-Pie), сделанных в различных условиях освещенности сцены, с углом обзора камеры не превышающим 45°. Эти изображения были помечены в соответствии с одним из 6 типов эмоций. Разметка выборки хранилась в текстовом файле. Над каждым элементом выборки осуществлялись следующие преобразования: из картинки вырезался участок, размером с изображением лица, и переводился в черно-белый формат. Такое преобразование осуществлялось с помощью вычислительно эффективного алгоритма PICO для детектирования лиц.

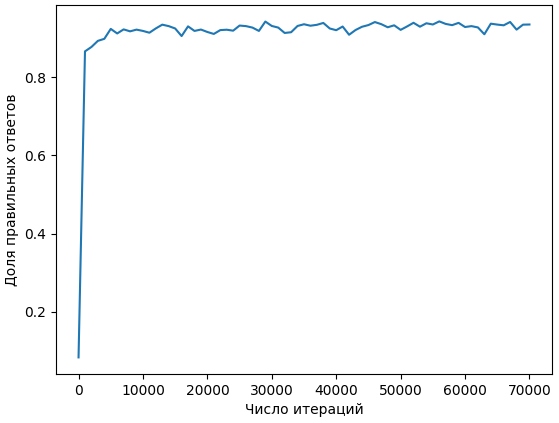
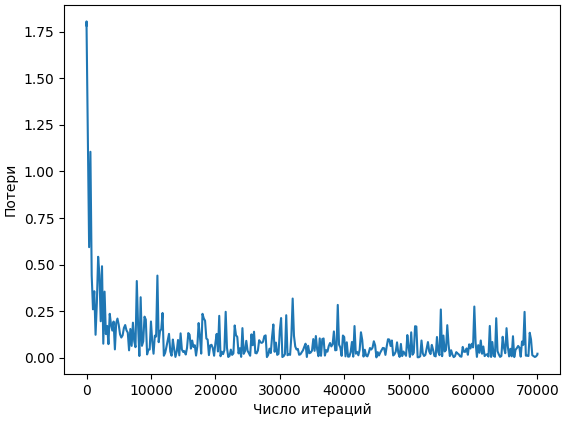


Рис. 10. Преобразование изображения из выборки перед фазой обучения и тестирования модели

Такое множество картинок в свою очередь разбивалось тренировочный и тестовый наборы данных в соотношении 80/20. Они не содержали одинаковых изображений. Более того, снимки с одним и тем же человеком не находились одновременно в тренировочном и тестовом наборе данных. Все данные были размечены в соответствии с классом, которому они принадлежали. Разметка изображений из тренировочной и тестовой выборки хранились в текстовых файлах.

Тестирование сверточной нейронной сети осуществлялось на суперкомпьютере NVIDIA DGX-1 и занимало в общей сложности около 30 минут. В результате численного эксперимента, доля правильных ответов (A) классификатора составила 94,41%. Данная величина рассчитывалась по формуле:

где - количество изображений, по которым классификатор принял верное решение, а - размер тестовой выборки [21]. Согласно рис. 11а, с ростом числа проделанных итераций обучения, значение доли правильных ответов существенно не менялось. Также, как видно из графика функции потерь на рис. 11б, алгоритм сходится, а величина ошибки незначительная.

а) график изменения значения доли правильных ответов б) график функции потерь

Рис. 11. Графики изменения величин в зависимости от числа проделанных обучающих итераций

Помимо этого, была построена также матрица неточностей, которая позволяет оценить качество работы алгоритма на основе сверточной нейронной сети.

Таблица 1

Матрица неточностей

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Категории | | Фактический класс | | | | | |
| **Улыбка** | **Удивление** | **Отвращение** | **Заинтересованность** | **Крик** | **Спокойствие** |
| Классы | **Улыбка** | 6514 | 37 | 53 | 6 | 0 | 145 |
| **Удивление** | 140 | 6767 | 2 | 0 | 16 | 8 |
| **Отвращение** | 72 | 2 | 6321 | 337 | 33 | 43 |
| **Заинтересованность** | 160 | 2 | 545 | 6476 | 9 | 172 |
| **Крик** | 21 | 156 | 18 | 7 | 6940 | 1 |
| **Спокойствие** | 93 | 36 | 64 | 174 | 1 | 6631 |

Как видно из табл. 1, на тестовой выборке обученная сверточная нейронная сеть достаточно хорошо справилась с задачей распознавания эмоций по изображению лица человека. Это подтверждают и посчитанные для каждого класса значения точности (P), полноты (R) и F-меры (F). Значения метрик качества классификатора рассчитывались по формулам

где - истинно-положительные решения, – ложно-положительгные ответы, – ложно-отрицательные решения [22]. Значения точности (P), полноты (R) и F-меры (F) для каждого класса отражены в табл.2.

Таблица 2

Анализ ошибок классификатора

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Категории | | Метрики | | |
| **Точность (P)** | **Полнота (R)** | **F-мера (F)** |
| Классы | **Улыбка** | 0,96 | 0,93 | 0,95 |
| **Удивление** | 0,98 | 0,97 | 0,97 |
| **Отвращение** | 0,93 | 0,9 | 0,92 |
| **Заинтересованность** | 0,88 | 0,93 | 0,9 |
| **Крик** | 0,97 | 0,99 | 0,98 |
| **Спокойствие** | 0,95 | 0,95 | 0,95 |

Согласно табл. 1,2 хуже всего определялись классы «отвращение» и «заинтересованность». Это объясняется тем, что и без помощи компьютерного зрения, на картинках из базы данных Multi-Pie, эти эмоции оказывались трудно различимыми [24].

а) Отвращение б) Заинтересованность

Рис.12. Пример наиболее трудно различимых картинок с эмоциями

Однако обученная сверточная нейронная сеть достаточно хорошо справляется с задачей распознавания эмоций по изображению лица человека: доля правильных ответов составляет чуть меньше 94,5%, а значение F-меры для каждого из классов превысило 0,9.

**ЗАКЛЮЧЕНИЕ**

В заключение необходимо отметить, что предложенный алгоритм является достаточно простым в реализации. С помощью разработанной модели можно распознавать эмоции по изображениям лиц людей. Тестирование показало достаточно высокий результат: доля правильных ответов классификатора составляет 94,01%, а значение F-меры для каждого из классов не меньше 0,89.

Для более качественной проверки робастности сверточной нейронной сети можно проверить качество ее работы на другой базе изображений. В ближайшее время планируется, что разработанный алгоритм будет внедрен и проверен на реальных снимках с камер видеонаблюдения.

Данная работа была выполнена при поддержке гранта УМНИК «Разработка алгоритмов прогнозирования индивидуального поведения на основе визуального распознавания эмоций», договор № 0033562, соглашение № 11320ГУ/2017.

**ЛИТЕРАТУРА**

1. Гудфеллоу Я., Бенджио И., Курвилль А. Глубокое обучение // ДМК Пресс, 2017, 652 с.
2. Y. Lewenberg, Y. Bachrach, S. Shankar, A. Criminisi, Predicting Personal Traits from Facial Images using Convolutional Neural Networks Augmented with Facial Landmark Information // Proceedings of the 13th AAAI Conference on Artificial Intelligence, 2016, pp. 4365 – 4366.
3. Плас Дж. Вандер Python для сложных задач. Наука о данных и машинное обучение // Питер, 2017, 576 с.
4. Y. Tian, T. Kanade, J.F. Cohn, Facial Expression Recognition // In Li S.Z., Jain A.K. Handbook of Face Recognition. Second Edition, Springer-Verlag London Limited, 2011, pp. 487 – 519.
5. Z. Niu, M. Zhou, L. Wang, X. Gao, G. Hua, Ordinal Regression with Multiple Output CNN for Age Estimation // 2016 IEEE Conference on Computer Vision and Pattern Recognition, 2016, pp. 4920 – 4928.
6. S. Paisitkriangkrai, J. Sherrah, P. Janney, A. Van-Den Hengel, Effective Semantic Pixel labeling with Convolutional Networks and Conditional Random Fields // 2015 IEEE Conference on Computer Vision and Pattern Recognition Workshops, 2015, pp. 36 – 43.
7. H. Khalajzadeh, M. Mansouri, M. Teshnehlab, Face Recognition using Convolutional Neural Network and Simple Logistic Classifier // In: Snášel V., Krömer P., Köppen M., Schaefer G. (eds) Soft Computing in Industrial Applications. Advances in Intelligent Systems and Computing, vol. 223. Springer, Cham, 2013, pp. 197 – 207.
8. Z. Wen, T. Huang, Capturing subtle facial motions in 3d face tracking // Proceedings of the 9th IEEE International Conference on Computer Vision, 2003, pp. 1340 – 1350.
9. H. Singh, R. Patel, Facial Expression Analysis using Deep Learning // International Research Journal of Engineering and Technology, vol. 4, issue 10, 2017, pp. 66 – 69.
10. The Cohn–Kanade AU-Coded Face Expression Database. Web: http://www.pitt.edu/~emotion/ck-spread.htm.
11. The Japanese Female Facial Expression (JAFFE) Database. Web: http://www.kasrl.org/jaffe.html.
12. The Binghamton University 3D Facial Expression (BU-3DFE) Database. Web: http://www.cs.binghamton.edu/~lijun/Research/3DFE/3DFE\_Analysis.html
13. The MMI Facial Expression Database. Web: https://mmifacedb.eu.
14. The Multimedia Understanding Group (MUG) Facial Expression Database. https://mug.ee.auth.gr/fed.
15. The Radboud Faces Database. Web: http://www.socsci.ru.nl:8180/RaFD2/RaFD?p=main.
16. The CMU Multi-PIE Face Database. Web: http://www.cs.cmu.edu/afs/cs/project/PIE/MultiPie/Multi-Pie/Home.html.
17. Бринк Х., Ричардс Д., Февероволоф М. Машинное обучение // Питер, 2017
18. Caffe Framework. Web: http://caffe.berkleyvision.org.
19. Ивановский Л.И., Хрящев В.В., Храбров Д.Е. Распознавание эмоций по изображению лица человека на основе сверточных нейронных сетей // Материалы 14-ой международной конференции «Телевидение: передача и обработка изображений», 2017, с. 130 – 132.
20. Флах П. Машинное обучение. Наука и искусство построения алгоритмов, которые извлекают знания из данных // М.: ДМК Пресс, 2015.
21. Bajpai S. Implementing Convolutional Neural Network with Parrallel Computing Using CUDA // International Journal of Innovative Science, Engineering & Technology, Vol. 2, Issue 11, 2015, pp. 517 – 520.
22. Scikit-learn. Web: http://scikit-learn.org/stable.
23. Рашка С. Python и машинное обучение // ДМК Пресс, 2017, 418 с.
24. L. Ivanovsky, V. Khryashchev, A. Lebedev, I. Kosterin, Facial Expression Recognition Algorithm Based on Deep Convolution Neural Network // Proceedings of the 21th Conference of Open Innovations Association FRUCT'21. Helsinki, Finland, 2017.