**ИСПОЛЬЗОВАНИЕ ГЛУБОКОГО ОБУЧЕНИЯ И СВЕРТОЧНЫХ НЕЙРОННЫХ СЕТЕЙ ДЛЯ АНАЛИЗА ВЫРАЖЕНИЯ ЛИЦА**

асп. Степанова О.А., асп. Ивановский Л.И., доцент, к.т.н. Хрящев В.В.

Ярославский государственный университет им. П.Г. Демидова

Автоматическое распознавание выражения человеческого лица является важным компонентом для создания систем взаимодействия человека и компьютера. На сегодняшний день одним из перспективных подходов к решению данной задачи является использование сверточных нейронных сетей – инструмента из области глубокого машинного обучения. В работе представлен алгоритм на основе сверточной нейронной сети, решающий задачу многоклассовой классификации (определение типа эмоции по изображению лица человека) и бинарной классификации (детектирование улыбки). Обучение и тестирование алгоритма осуществлялось на графическом процессоре суперкомпьютера с использованием изображений из базы Multi-PIE. Результаты тестирования показали высокий уровень точности модели как при решении задачи распознавания эмоций, так и при решении задачи детектирования улыбки. Использование разработанной модели возможно в приложениях реального времени или в специальных решениях для встроенных систем, таких как NVIDIA Jetson.

Автоматическое распознавание выражений лица человека – перспективная область современных алгоритмов компьютерного зрения и машинного обучения. На сегодняшний день системы, анализирующие эмоции, применяются в широком спектре коммерческих решений, таких как мониторинг эффективности маркетинговых и рекламных компаний, оценка работы персонала при общении с клиентом, повышение интеллектуализации взаимодействия компьютерных систем и человека [1]. Отдельным направлением в области анализа выражения лица человека является детектирование улыбки, находящее свое применение в системе Smile To Pay, в программном обеспечении фотокамер, а также в приложениях для выбора лучших фотографий [2]. В системах автоматического распознавания выражений человеческого лица, в общем случае реализуются следующие этапы [3]:

1. Захват изображений лица из базы данных или потокового видео.
2. Предварительная обработка изображений: снижение помех, фильтрация, повышение четкости.
3. Извлечение оптимального набора признаков. Данная стадия является одним из ключевых этапов, поскольку отобранные признаки должны минимизировать внутриклассовые отличия и максимизировать дисперсию между классами.
4. Классификация на заданное число классов.

Одним из перспективных подходов к классификации выражения лица человека является использование алгоритмов глубокого машинного обучения, таких как сверточные нейронные сети (convolutional neural networks, CNN). Особенность таких сетей заключается в том, что дескрипторы изображений вычисляются в результате операции двумерной свёртки, при этом сами свёрточные фильтры формируются в процессе обучения. Сгенерированные таким образом дескрипторы дают, как правило, наилучшие результаты во многих задачах классификации [4].

В данной работе представлена реализация сверточной нейронной сети для решения задачи бинарной классификации (детектирование улыбки) и многоклассовой классификации (распознавание различных типов эмоций).

Реализация архитектуры свёрточной нейронной сети осуществлялась с помощью фреймворка Caffe [5] на основе сети, предложенной в [6]. Как показано на Рисунке 1, разработанная нейронная сеть состоит из восьми сверточных слоев (четыре слоя с фильтрами , три – с фильтрами ), восьми слоев с функцией активации , восьми слоев, реализующих процесс локальной нормализации, и трех слоев, описывающих процесс дискретизации с помощью операции max pooling. После всех проделанных операций формируется одномерный вектор из 96 значений признаков, которые подаются на вход полносвязнному слою. Далее, на выходе с полносвязного слоя формируется вектор, с количеством элементов равным числу классов: 6 классов для задачи распознавания эмоций и 2 класса для задачи детектирования улыбки. По данному вектору, с помощью обобщенной логистической функции Softmax нетрудно вычислить апостериорные вероятности для каждого класса. Таким образом, сверточная нейронная сеть классифицирует входное изображение лица человека в соответствии с наиболее вероятным типом эмоций.

Базовая (начальная) скорость обучения сети была выбрана равной 0,01. В качестве алгоритма численной оптимизации использовался стохастический градиентный спуск 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.

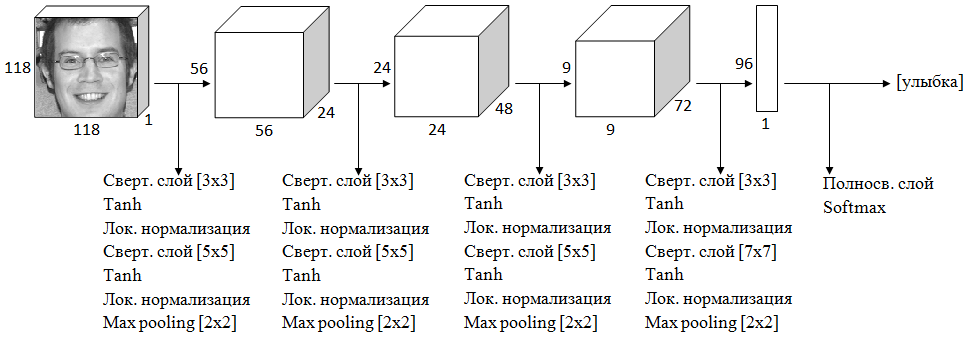


Рисунок 1 – Архитектура разработанной свёрточной нейронной сети.

Эксперимент по оценке точности алгоритма проводился на базе изображений MultiPIE [7]. Эта база содержит более 750000 изображений 337 различных людей. Изображения представляют собой снимки, сделанные на фотокамеру, под углом обзора, не превышающим 90 градусов, и с разной степенью освещенности. В базе данных MultiPIE содержатся изображения людей с 6 различными типами эмоций: спокойствие, улыбка, удивление, заинтересованность, отвращение и крик. Примеры изображений из базы данных Multi-PIE приведены на Рисунке 2.

|  |  |  |
| --- | --- | --- |
| a.png | b.png | c.png |
| а) Спокойствие | б) Улыбка | в) Удивление |
| d.png | e.png | f.png |
| г) Заинтересованность | д) Отвращение | е) Крик |

Рисунок 2 – Типы эмоций, представленные в базе изображений MultiPIE.

Для тестирования разработанного алгоритма из базы MultiPIE бралась выборка из случайно отобранных снимков (70000 картинок для задачи детектирования улыбки и 210000 картинок для задачи распознавания эмоций). Все отобранные изображения были сделаны под углом обзора камеры, не превышающим 45°. Над каждым элементом выборки осуществлялись следующие преобразования: из картинки вырезался участок размером с изображением лица и переводился в черно-белый формат. Выборка разбивалась тренировочный и тестовый наборы данных в соотношении 80/20. Тренировочный и тестовый набор 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 хранилась в текстовых файлах.

Тестирование осуществлялось на суперкомпьютере NVIDIA DGX-1 в Центре искусственного интеллекта Ярославского университета им. П.Г. Демидова. Для ускорения работы сети процессы обучения и тестирования свёрточной нейронной сети осуществлялись параллельно, на большом числе независимых потоков графического процессора видеокарты.

Результаты численного эксперимента представлены в Таблице 1.

Таблица 1

Результаты тестирования сверточной нейронной сети

|  |  |
| --- | --- |
| **Задача** | **Доля правильных ответов** |
| Распознавание эмоций | 94.48% |
| Детектирование улыбок | 96.32% |

Доля правильных ответов (A) классификатора рассчитывалась по формуле:

где - количество изображений, по которым классификатор принял верное решение, а - размер тестовой выборки [8].

На рисунке 3 представлена графическая интерпретация результатов тестирования модели сверточной нейронной сети для решения задачи распознавания эмоций. Из графиков функций потерь видно, что алгоритм классификации сходится, а величина ошибки незначительная [8].

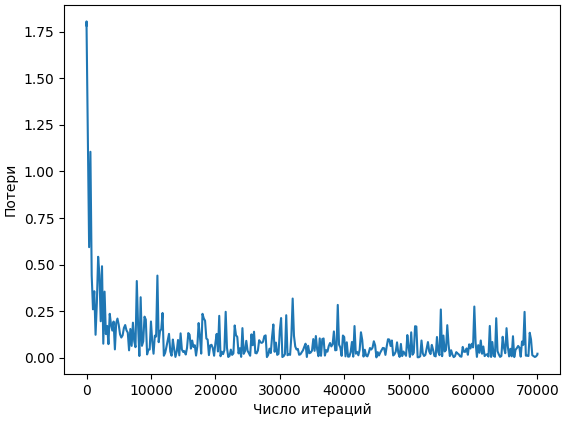


Рисунок 3 – Зависимость величины ошибки от числа проделанных обучающих итераций для решения задачи распознавания эмоций.

В заключение необходимо отметить, что предложенный алгоритм классификации является простым в реализации, а также позволяет эффективно детектировать улыбки и распознавать эмоции на изображениях лиц людей. Для увеличения качества классификации в ближайшем будущем планируется попробовать применить подход с использованием ансамбля сверточных нейронных сетей. Для многих задач классификации, такой подход позволяет снизить величину ошибки, а также уменьшить время, затраченное на обучение модели [9].

Использование разработанной модели возможно в приложениях реального времени или в специальных решениях для встроенных систем, таких как NVIDIA Jetson.

Исследование выполнено при поддержке гранта УМНИК-НТИ № 0033562 «Разработка алгоритмов прогнозирования индивидуального поведения на основе визуального распознавания эмоций».

**Литература**

1. Ахметшин Р.И., Кирпичников А.П., Шлеймович М.П. (2015) Распознавание эмоций человека на изображениях // Вестник Казанского технологического университета. 2015. №11.

2. Kaihao Zhang, Yongzhen Huang, Hong Wu, Liang Wang (2016) Facial smile detection based on deep learning features // Pattern Recognition (ACPR), 3rd IAPR Asian Conference, Kuala Lumpur, 2016.

3. Mehdi Ghayoumi (2017) A Quick Review of Deep Learning in Facial Expression // Journal of Communication and Computer 14, 2017. pp. 34-38.

4. Николенко С., Кадурин А., Архангельская А. (2017) Глубокое обучение. Погружение в мир нейронных сетей // Питер, 2018. 480 с.

5. Caffe Framework [Электронный ресурс]. URL: http://caffe.berkleyvision.org.

6. Niu Z., Zhou M., Wang L., Gao X., Hua G. Ordinal Regression with Multiple Output CNN for Age Estimation // IEEE CVPR, 2016.

7. The CMU Multi-PIE Face Database [Электронный ресурс]. URL: http://www.cs.cmu.edu/afs/cs/project/PIE/MultiPie/Multi-Pie/Home.html.

8. Вандер Плас Дж. Python для сложных задач: наука о данных и машинное обучение. СПб.: Питер, 2018, 576 с.

9. Sert E., Ertekin S., Halici U. Ensemble of convolutional neural networks for classification of breast microcalcification from mammograms // Conference proceedings: Annual International Conference of the IEEE Engineering in Medicine and Biology Society, 2017.

**DEEP LEARNING AND CONVOLUTIONAL NEURAL NETWORKS   
FOR FACIAL EXPRESSION ANALYSIS**

O. Stepanova, L. Ivanovsky, V. Khryashchev

P.G. Demidov Yaroslavl State University

Automatic facial expression recognition is an important component for the creation of human-computer interaction systems. To date, one of the most promising approaches to solving this task is the use of convolutional neural networks - a tool from the field of deep machine learning. The paper presents an algorithm based on the use of convolutional neural network, which solves the problems of multiclass classification (emotion recognition) and binary classification (smile detection). Training and testing of the algorithm was carried out on the supercomputer NVIDIA DGX-1 using images from the Multi-PIE database. The testing results showed a high level of accuracy of the proposed model both in solving the task of emotion recognition, and in solving the task of smile detection. It is planned that the developed algorithm will be verified on real images from CCTV cameras.