

УДК 159.9.072

ПРИМЕНЕНИЕ НЕЙРОННЫХ СЕТЕЙ ДЛЯ РАСПОЗНАВАНИЯ ЭМОЦИЙ ПО ИЗОБРАЖЕНИЯМ ЛИЦ



И.Г. Шупейко
доцент кафедры инженерной психологии и эргономики Белорусского государственного университета информатики и радиоэлектроники



Л. А. Хаткевич
Старший преподаватель кафедры теории вероятностей и математической статистики Белорусского государственного университета



Е.Е. Семёнов
Математик-программист

Белорусский государственный университет информатики и радиоэлектроники, Республика Беларусь
E-mail: shupeyko@bsuir.by, lyudmila.khatkevich@yandex.by

И.Г. Шупейко

Доцент кафедры инженерной психологии и эргономики Белорусского государственного университета информатики, и радиоэлектроники, кандидат психологических наук, доцент. Научные интересы: методы инженерно-психологического проектирования систем «человек-машина-среда», использование компьютерных технологий в инженерно-психологических исследованиях.

Л.А. Хаткевич

Старший преподаватель кафедры теории вероятностей и математической статистики Белорусского государственного университета. Окончила Белорусский государственный университет. Научные интересы: методы обработки и распознавания цифровых изображений.

Е.Е. Семёнов

Математик-программист, закончил факультет прикладной математики Белорусского государственного университета в 2021 г.

Аннотация. В статье показана актуальность задачи распознавания эмоций по изображениям лиц. Приведено описание архитектуры нейронной сети, примененной для решения задачи. Описаны результаты экспериментального исследования точности классификации на выборке изображений из интернет-источника.

Ключевые слова: задачи распознавания, обработка цифровых изображений, нейронные сети, точность классификации.

Введение

Технологии распознавания эмоций активно применяются во многих областях жизни. Распознавание эмоций используют для предупреждения насилия как в общественных местах, так и дома. Система распознавания эмоций может быть использована для оценки степени удовлетворенности клиентов фирмы (например, кафе, отделения банка) по данным

видеонаблюдения. Еще одна сфера приложения - в роли помощника в работе отдела кадров на этапе первичного отбора персонала.

Большинство существующих систем распознавания эмоций основано на анализе выражения лица, голоса человека, выделении ключевых слов. Так, при визуальном анализе приподнятые уголки рта свидетельствуют о том, что человек в хорошем настроении, а наморщенный нос — что он злится или испытывает отвращение.

В данной статье распознавание эмоций осуществляется по изображениям в цифровом формате. Для распознавания применены методы глубокого обучения [1, 2].

Архитектура сверточной нейронной сети

Сверточные сети – это нейронные сети, в которых вместо операции умножения входов на матрицу весов, по крайней мере в одном слое, используется свертка [3, 4].

Прототип модели нейронной сети, используемой для распознавания эмоций, является: CРCРССCРССCРFFF:

- C - сверточный слой(convolution layer),
- P - слой пулинга(pooling layer),
- B - слой групповой нормализации(batch normalization layer),
- F - соединенный слой(fully connected layer),
- D - слой отбрасывания(drop out layer).

Модель является представителем семейства VGG моделей и характеризуется большим количеством сверточных слоев с малым размером ядра 3x3.

Главной проблемой в работе определено переобучение, связанное с качеством набора данных [4].

Для устранения переобучения нейронной сети уменьшено количество нейронов на выходе первого соединенного слоя до 1000. Для устранения переобучения также был применен метод ранней остановки, удалены указанные слои. Для увеличения скорости вычислений и устранения внутреннего ковариационного сдвига[5] были добавлены блоки групповой нормализации. В целях улучшения временных показателей обучения модели, была осуществлена перемена мест слоев пулинга и ReLU. В последнем слое пулинга размер ядра был сменен на 3. Первый и последний пункты связаны с меньшим размером изображений в сравнении с теми, которые использовались вместе с прототипом модели.

После фиксирования архитектуры были проведены автоматизированные исследования с целью найти оптимальные гиперпараметры и получить лучшую модель. Использовался поиск по сетке для определения оптимального количества эпох и размера групп, а также случайный поиск для определения наилучших весов [6]. Были установлены следующие оптимальные параметры: 11 больших эпох, 1,2 малые эпохи и 128 изображений в группе. Большая эпоха - итерация с шагом оптимизатора 0.001. Малая эпоха - итерация с шагом оптимизатора 0.0001.

Во всех исследованиях использовался только графический процессор. Графические процессоры обладают большим числом ядер, что позволяет превосходить центральные процессоры в объеме производимых однотипных вычислений за единицу времени с максимальным использованием распараллеливания. В таблице1 представлено время обучения в минутах для CPU и GPU.

Таблица 1. Сравнение временных показателей

тип процессора	1 эпоха	5 эпох	10 эпох
CPU	5.52	27.25	54.67
GPU	0.25	1.25	2.5

Результаты практического исследования

Исходный набор данных FER-2013 (распознавание эмоций лица) [2] - состоит из 35,887 размеченных черно-белых изображений с разрешением 48×48. Набор данных является сложным, поскольку в нем присутствуют изображения разных полов и возрастов людей, находящихся в разнообразных позах. Такие особенности набора данных отражают реальные условия задачи распознавания. Набор изначально разделен на обучающую, проверочную и тестовую выборки, по 28,709, 3,589, и 3,589 изображений соответственно. Каждому изображению ставится в соответствие одна из семи эмоций: злость, отвращение, страх, счастье, грусть, удивление, нейтральность. Из набора данных удалена эмоция отвращения, поскольку она представлена небольшим количеством изображений.

На рисунке 1 представлена матрица ошибок распознавания эмоций для первой модели. На пересечении *i*-ой строки и *j*-ого столбца матрицы расположена плитка, цвет которой зависит от количества изображений, ответ модели для которых равен значению *i*, а метка из набора данных соответствует значению столбца *j*. Для каждой плитки указано точное количество таких изображений. Диагональные плитки соответствующие правильной классификации, являются самыми темными, что говорит о корректной работе модели и высоких показателях точности. Также можно заметить сложность дифференцирования между собой некоторых эмоций, например, нейтральности и грусти.

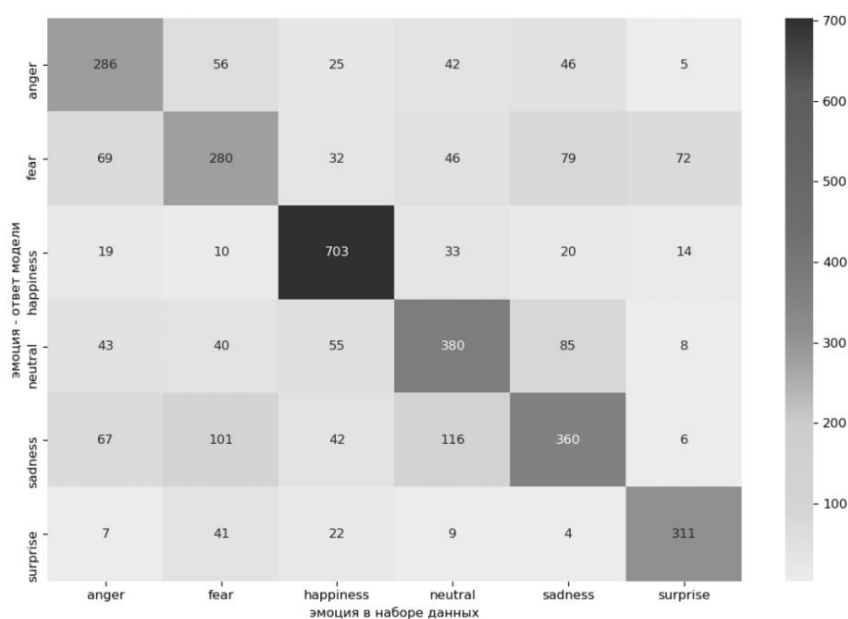


Рисунок 1. Матрица ошибок.

Главной проблемой набора данных было определено его качество, на рисунке 2 и в таблице 2 приведены несколько примеров.

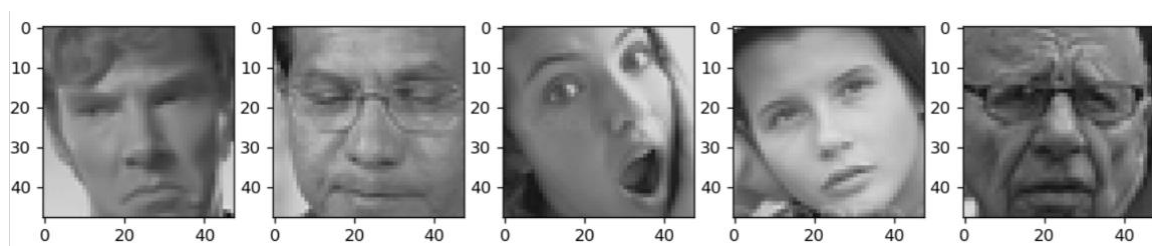


Рисунок 2. Примеры изображений

Таблица 2. Результаты распознавания эмоций

Модель1	грусть	грусть	удивление	нейтральность	злость
Данные	злость	страх	страх	счастье	злость

На первом и третьем изображении представлен типичный пример ошибки нейронной сети. На изображении присутствуют явные признаки другого класса (опущенные уголки рта, широко открытый рот), и сеть закономерно дает неверный ответ. Четвертое изображение демонстрирует низкое качество набора данных.

В таблице 3 представлены вероятности правильной классификации отдельно для каждой эмоции для двух лучших моделей. Отметим, что вероятности распознавания эмоций «счастье» и «удивление» достигают 80 %.

Таблица 3. Вероятности правильной классификации

эмоция	модель 1(%)	модель 2(%)
злость	58.2	59.9
страх	53.0	49.1
счастье	80.0	79.5
грусть	60.6	59.4
удивление	74.8	79.1
нейтральность	60.7	61.2

Точность распознавания 6 эмоций, полученная в данной работе, составляет 65.56% . Данная точность была зафиксирована для двух полученных моделей. Точность распознавания человеком составляет $65 \pm 5\%$, что практически совпадает с результатами, полученными на моделях.

Список использованных источников

- [1] Гудфеллоу Я., Бенджио И., Курвилль А. Г93 Глубокое обучение / пер. с англ. А. А. Слинкина. – 2-е изд., испр. – М.: ДМК Пресс, 2018. – 652 с.: цв.ил.
- [2] Khairuddin Y., Chen Z. Facial Emotion Recognition: State of the Art Performance on FER2013 / Yousif Khairuddin, Zhuofa Chen // Dept. of Electrical and Computer Engineering, Boston University, Boston, MA, USA, <https://arxiv.org/ftp/arxiv/papers/2105/2105.03588.pdf>
- [3] Simonyan K., Zisserman A. VERY DEEP CONVOLUTIONAL NETWORKS FOR LARGE-SCALE IMAGE RECOGNITION / Karen Simonyan, Andrew Zisserman // Visual Geometry Group, Department of Engineering Science, University of Oxford, <http://arxiv.org/abs/1409.1556v6>
- [2] . Pramerdorfer C., Kampel M. Facial Expression Recognition using Convolutional Neural Networks: State of the Art / Christopher Pramerdorfer, Martin Kampel // Computer Vision Lab, TU Wien Vienna, Austria, arXiv:1612.02903v1 [cs.CV] 9 Dec 2016
- [3] Ioffe S., Szegedy C. Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift / Sergey Ioffe, Christian Szegedy // <http://arxiv.org/abs/1502.03167v3> - 2 Mar 2015.
- [4] Kingma D., Ba J. ADAM: A METHOD FOR STOCHASTIC OPTIMIZATION / Diederik P. Kingma, Jimmy Lei Ba // arXiv:1412.6980v9 30 Jan 2017

THE EMOTIONS RECOGNITION FROM FACIAL IMAGES USING NEURAL NETWORKS

I.G. SHUPEYKO, PhD

*Associate professor of
the Department
of En-gineering
Psychology and Ergonomics
of the Belarusian State
University of Informatics and
Radioelectronics*

L.A. Khatkevich

*Senior Lecturer at the
Department of Probability Theory
and Mathematical Statistics of the
Belarusian State University.*

E.E. Semenov

Mathematician-programmer

Abstract.. The article shows the relevance of the problem of recognizing emotions from facial images. The neural network architecture description used to solve the problem is given. The results of the classification accuracy on the sample of Internet source images are described.

Keywords: recognition tasks, digital image processing, neural networks, classification accuracy