

Леонова Амелия Александровна, магистрант,
Санкт-Петербургский государственный электротехнический
университет «ЛЭТИ» им. В. И. Ульянова (Ленина),
г. Санкт-Петербург

ВЫБОР И ПОСТРОЕНИЕ АРХИТЕКТУРЫ НЕЙРОННОЙ СЕТИ ДЛЯ АНАЛИЗА ЭМОЦИЙ ЧЕЛОВЕКА

Аннотация: Изучение эмоционального состояния человека посредством нейросетевого моделирования проводится учеными разных стран мира, однако универсального подхода пока не существует. Для решения задачи анализа эмоций с оптимальным уровнем точности важно определиться с архитектурой нейронной сети. Наиболее подходящим вариантом является архитектура сверточной сети, так как она устойчива к сдвигу изображения, позволяет сократить количество параметров для обработки, устойчива к мелкому шуму, искажению изображения.

Abstract: The study of human emotional state by means of neural network modeling is carried out by scientists from different countries of the world, but there is no universal approach yet. To solve the problem of emotion analysis with an optimal level of accuracy, it is important to determine the architecture of the neural network. The most suitable option is the convolutional network architecture, as it is resistant to image shift, allows to reduce the number of parameters for processing, is resistant to small noise and image distortion.

Ключевые слова: нейросеть, архитектура, анализ, базовые эмоции, фотография лица
Keywords: neural network, architecture, analysis, basic emotion, facial photography

В современных условиях изучение эмоций людей составляет важную составляющую для всех отраслей, где в основе результативности работы субъектов хозяйствования лежит человеческий фактор. Настроение работников влияет на их работоспособность, соответственно и на итоговые результаты их труда. Определяющую роль в общении с клиентами также играет эмоциональная составляющая. Изучение эмоционального состояния проводится сегодня специалистами и учеными в разрезе не только индивидуальных исследований, но и в контексте международных конкурсов. Однако универсального подхода пока не существует. Ученые пытаются решить задачу определения эмоций с оптимальным уровнем точности, доступности и объективности исходных данных. Рассмотрим возможность определения эмоций по лицу человека посредством считывания информации как с цветных, так и черно-белых фотографий лиц людей разных возрастов (детей и взрослых), с возможным небольшим наклоном головы. Возьмем по каждой базовой эмоции порядка 70–100 изображений.

Алгоритм проведения анализа эмоций посредством нейросетей представим следующим образом:

1. Обоснование использования фотографий для анализа эмоций: уход от использования ключевых точек лица при обработке изображений.
2. Этап сбора базовых эмоций.
3. Этап предварительной обработки изображений.
4. Этап выбора и построения архитектуры нейронной сети.



1 этап. Обоснование использования фотографий без обозначения ключевых точек лица.

Использование ключевых опорных точек лица, по которым можно идентифицировать лицо в пространстве, и которые сегодня используются в системах распознавания лиц, может достигать 68. Но эти точки используют только для распознавания лица. Если же говорить об эмоции, то известно, что в мимических движениях лица в той или иной степени участвует около 100 мышц, каждая из которой влияет на расположение ключевых точек эмоции. Таким образом для того, чтобы идентифицировать правильно эмоцию, необходимо выбрать и обозначить на изображении лица все ключевые точки, характеризующую данную эмоцию. Однако, не всегда можно правильно обозначить ключевые точки лица, например, при размере изображения 48 на 48 точек для разных лиц в ряде случаев невозможно будет идентифицировать различия между некоторыми ключевыми точками лица. Нахождение точек мышц лица связано с погрешностью наложения сети ключевых точек на конкретное изображение лица (взрослый или ребенок), а также достаточно затратно. Причем, важно не количество используемых изображений лиц, (в конкурсах, указанных в первой главе работы – их было десятки тысяч с размером фотографий 48 на 48 точек), а качество, которое зависит от размера изображения, которое увеличим до 100 точек по вертикали и горизонтали. При этом количество изображений лиц сократим от 70 до 100 по каждой базовой эмоции. Также не будем отказываться от цветных изображений.

2 этап. Сбор базовых эмоций: радость, грусть, злость, удивление, отвращение, страх.

На данном этапе посредством данных, находящихся в открытом доступе в сети Интернет были собраны фотографии лиц по всем базовым эмоциям. По эмоции «радость» были найдены фотографии, доступные в сети Интернет посредством сервисов «Яндекс Браузер» и веб-браузера «Mozilla Firefox». По всем базовым эмоциям были собраны фотографии в основном цветные, иногда черно-белые. Фотографии лиц людей любого пола и возраста – детей и взрослых, мужчин и женщин, где ясно выражена данная эмоция. Пример исходного материала по базовой эмоции радость представлен на рис. 1.



Рисунок 1 – Исходный материал по эмоции радость

По аналогии был собран исходный материал по всем базовым эмоциям.

3 этап. Предварительная обработка изображений: выбор границ лица на изображении

Проведем поиск способов обнаружения границ лица на фотографии. При выборе способа идентификации лица основными критериями будем считать простоту и свободный доступ для использования. К такой модели можно отнести – модель, разработанную в Китае «Каскад с воронкообразной структурой для обнаружения лиц с нескольких точек зрения с учетом выравнивания», предложенную в 2016г. Шучже У, Мейной Кан, Чжэньлиан Хэ, Шигуан Шаном и Силян Чен [1]. Исходный код реализации данной модели опубликован на Github [2]. С использованием данной модели все собранные изображения были обработаны и получены наборы с изображением лиц по всем базовым эмоциям, размером 100 на 100 точек.



Пример фотографии лица до обработки приведен на рис. 2.



Рисунок 2 – Размер изображения 208 на 315 точек

Пример фотографии лица после обработки приведен на рис. 3.



Рисунок 3 – Размер изображения 100 на 100 точек

Аналогично был обработан весь исходный материал по всем базовым шести эмоциям: радость, грусть, злость, удивление, отвращение, страх. На следующем этапе эксперимента проведем выбор и построение архитектуры нейронной сети.

4 этап. Выбор и построение архитектуры нейронной сети

На основе известных архитектур нейронных сетей, применяемых для классификации изображений была выбрана облегченная модель сверточной нейронной сети. Наиболее подходящей моделью является модель сверточной сети, так как она устойчива к сдвигу изображения, позволяет сократить количество параметров для обработки, устойчива к мелкому шуму (например, родинка или веснушка на изображении), искажению изображения, устойчива к повороту (например, наклон головы).

Сегодня есть известные архитектуры сверточных сетей. Например:

1. Alexnet (2012г., применена для распознавания изображений с использованием активационной функции ReLU [7].
2. Vgg (2014г., применена для распознавания изображений, 19 слоев, свертки 3x3 – стали стандартом в индустрии с данного момента времени [9].
3. GoogleLeNet aka Inception (2014г., введен inception слой, который в данный момент уже не используется. Нейронная сеть состояла из 22 слоев [5].
4. Batch normalization (2015г., основная идея в стабилизации и увеличения скорости [4].
5. Kaiming weight initialization (2015г., предложено использовать особенность функции активации ReLU при активации весов [8].
6. ResNet (2015г. Нейронная сеть состояла из 152 слоев [6].
7. MobileNet (2017г. Модифицирована для обработки изображений на мобильных устройствах. Функции активации ReLU [3].
8. EfficientNet (2019г. Применено масштабирование (глубже, шире) слоев сверточной сети [10].

На основе анализа данных моделей была выбрана архитектура нейронной сети похожая на модель Vgg (2014г.), но в данной модели использовано только 6 сверточных слоев (это позволяет сэкономить ресурсы при обучении и работе сети – память, объем работы ядер



графических процессоров), применен max pooling с ядром 3x3, и strides 3x3 (снижает сложность изображения, что делает обучение более эффективным и ускоряет процесс обработки информации), применено 2 выходных линейных слоя. Pooling и dropout применены после каждой пары сверточных слоев. В дальнейшей работе, в процессе экспериментирования при последовательном изображении эмоций обнаружилась плохая сходимость процесса обучения для общепринятой функции ReLU для сверточных сетей. Поэтому в исследовании была заменена данная функция на гиперболический тангенс, что позволило получить в дальнейшем, при проведении эксперимента, положительный результат распознавания всех базовых эмоций радость, злость, грусть, удивление, отвращение, страх на тестовом наборе (точность составила 67 % правильных распознаваний).

Список литературы:

1. Шучже У, Мейной Кан, Ч. Хэ, Ш. Шаном, С. Чен Каскад с воронкообразной структурой для обнаружения лиц с нескольких точек зрения с учетом выравнивания: [Электронный ресурс], URL: <https://arxiv.org/pdf/1609.07304>. (Дата обращения 29.04.25).
2. Шучже У, Мейной Кан, Ч. Хэ, Ш. Шаном, С. Чен Исходный код реализации модели: [Электронный ресурс], URL: <https://github.com/atomashpolkiy/rustface>. (Дата обращения 19.04.25).
3. Andrew G. Howard, Menglong Zhu, Bo Chen, Dmitry Kalenichenko, Weijun Wang, Tobias Weyand, Marco Andreetto, Hartwig Adam, MobileNets: Efficient Convolutional Neural Networks for Mobile Vision Applications // MobileNets: эффективные сверточные нейронные сети для приложений мобильного зрения [Электронный ресурс], URL: <https://arxiv.org/pdf/1704.04861.pdf>. (Дата обращения 17.04.25).
4. Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift // Пакетная нормализация: ускорение обучения глубокой сети за счет уменьшения внутреннего ковариационного сдвига: [Электронный ресурс], URL: <https://arxiv.org/pdf/1502.03167.pdf>. (Дата обращения 13.05.25).
5. Christian Szegedy, Wei Liu, Yangqing Jia, Pierre Sermanet, Scott Reed, Dragomir Anguelov, Dumitru Erhan, Vincent Vanhoucke, Andrew Rabinovich, Going deeper with: [Электронный ресурс], URL: <https://arxiv.org/pdf/1409.4842.pdf>. (Дата обращения 13.03.25).
6. Delving Deep into Rectifiers: Surpassing Human-Level Performance on ImageNet Classification // Углубляемся в выпрямители: превосходим человеческий уровень производительности по классификации ImageNet: [Электронный ресурс], URL: <https://arxiv.org/pdf/1502.01852.pdf>. (Дата обращения 14.03.25).
7. ImageNet Classification with Deep Convolutional Neural Networks // Архитектура сети Alexnet: [Электронный ресурс], URL: <https://proceedings.neurips.cc/paper/2012/file/c399862d3b9d6b76c8436e924a68c45b-Paper.pdf>. (Дата обращения 13.05.25).
8. Kaiming He, Xiangyu Zhang, Shaoqing Ren, Jian Sun, Deep Residual Learning for Image Recognition // Глубокое остаточное обучение для распознавания изображений: [Электронный ресурс], URL: <https://arxiv.org/pdf/1512.03385.pdf>. (Дата обращения 14.03.25).
9. Karen Simonyan, Andrew Zisserman, Very deep convolutional networks For large-scale image recognition // Сеть Vgg: [Электронный ресурс], URL: <https://www.robots.ox.ac.uk/~vgg/publications/2015/Simonyan15/simonyan15.pdf>. (Дата обращения 13.05.25).
10. Mingxing Tan, Quoc V. Le EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks // EfficientNet: переосмысление масштабирования модели для сверточных нейронных сетей: [Электронный ресурс] URL: <https://arxiv.org/pdf/1905.11946.pdf>. (Дата обращения 18.05.25).

