

СТРУКТУРА НЕЙРОННОЙ СЕТИ ДЛЯ ОЦЕНКИ ЭМОЦИОНАЛЬНОГО СОСТОЯНИЯ ЧЕЛОВЕКА

студент 4 курса, 5КБ группы Федосеев В. С.

Научный руководитель - канд. техн. наук Садов В. С.

Белорусский государственный университет

Минск, Беларусь

Эмоции окрашивают все психические процессы человека, при этом являясь универсальным регулятором его внутренней и внешней жизни. Такая универсальная важность эмоций является залогом повышенного интереса к ним и высокой степенью их изученности.

Знание того, какие эмоции испытывает человек, может помочь вовремя среагировать на фактор, вызвавший соответствующую эмоцию. Например, если человек чувствует страх, очевидно, что он чего-то опасается, и, возможно, ему нужна помощь, чтобы устранить фактор, или, например, предостеречь от последствий тяжёлых психических болезней.

Задача распознавания эмоций требует в первую очередь выбор модели представления эмоций. В настоящее время существует огромное количество различных таких моделей, например, трёхмерное пространство эмоциональных состояний по В. Вундту, шкала Г. Шлоссберга, и др. [1], [2]. Однако наибольшую популярность завоевала модель Пола Экмана. Он предложил модель так называемых базовых (первичных) эмоций, которые присутствуют в репертуаре всех без исключения человеческих культур и которые в свою очередь относят к разряду человеческих универсалий. Они бывают положительными (радость и удивление) и отрицательными (гнев, грусть, страх, отвращение и, иногда, некоторые исследователи включают в этот список презрение) [3].

Базовые эмоции характеризуются универсальными мимическими выражениями, причём эти выражения стабильны. Их можно различить даже у представителей далёких друг от друга культур. Имеется прочная связь между структурами мозга, ответственными за данные эмоции, и мышечной системой [4].

По этой причине модель Пола Экмана может успешно применяться для оценки эмоционального состояния человека по его изображению лица (задача

классификации). Для этой задачи широко используются методы глубокого обучения, в частности, различные нейросетевые подходы.

Существует множество архитектур нейронных сетей. Их можно разделить на 4 основные группы: предсказывающие, генеративные, с памятью и операционные.

Для решения исходной задачи была выбрана операционная свёрточная нейронная сеть, так как она наилучшим образом подходит для работы с данными, имеющими чётко выраженную сеточную топологию, и способна хорошо масштабироваться к задачам очень большого размера. Данный подход особенно успешным оказался в применении к двумерным изображениям [5].

Также важной задачей стояло исследование характеристик разработанной структуры нейронной сети, так как правильный подбор параметров, оптимизируемых во время обучения, и настраиваемых гиперпараметров, позволяющих управлять процессом обучения нейронной сети, напрямую влияет на её качество работы.

Главной особенностью свёрточных нейронных сетей является то, что в её промежуточных слоях (нейронные сети обычно состоят из нескольких слоёв) используется такая математическая операция как кросс-корреляция (в контексте машинного обучения под свёрткой обычно понимают именно её). Для двухмерного случая:

$$S(i, j) = (K * I)(i, j) = \sum_m \sum_n I(i + m, j + n) K(m, n), \quad (1)$$

где K – ядро свёртки (фильтр), I – изображение, i и j – номера соответственно строки и столбца изображения, m и n – такие же номера, но для ядра.

Также не менее важным является использование следующих элементов в свёрточной нейронной сети:

субдискретизация (пулинг): помогает обучиться нейронной сети инвариантности, то есть вне зависимости от того, какие использованы аффинные преобразования к исходному изображению, нейронная сеть его правильно классифицирует.

паддинг: решает проблему сокращения размерностей при прохождении входного изображения через множество слоёв.

пакетная нормализация: противодействует внутреннему ковариационному сдвигу, который может негативно влиять на работу нейронной сети.

дропаут: в первую очередь предназначен для уменьшения переобучения сети.

В совокупности всего выше сказанного и принимая особенности свёрточных нейронных сетей, была разработана следующая структура сети, представленная на рисунке.

На рисунке 1 слева расположены два блока – блок из двух свёрточных слоёв с пакетной нормализацией и пулингом по максимальному значению и блок из полносвязного слоя с пакетной нормализацией, справа – структура самой нейронной сети.

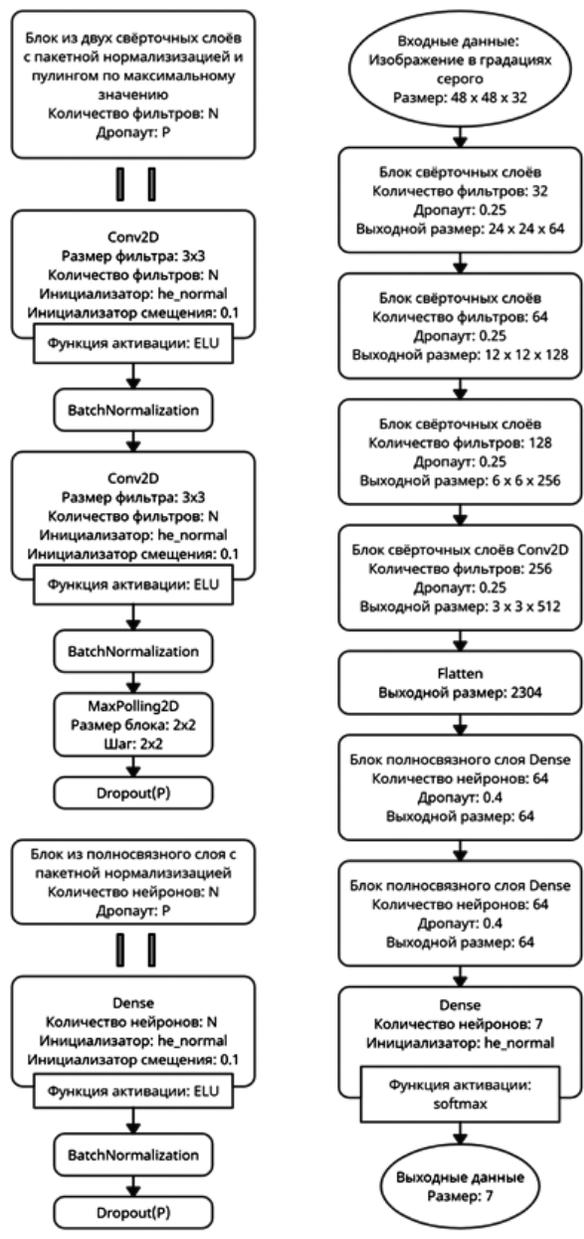


Рисунок 1. Структура нейронной сети

Первый блок ответственен за выделение различных признаков, причём как видно из структуры, таких блоков всего 4, и они идут последовательно, с увеличением количества фильтров, что позволяет выстраивать более сложные признаки из простых, выделенных на предыдущих блоках.

Второй блок необходим для моделирования сложной нелинейной функции, используемой для итоговой классификации.

Однако сама по себе структура бесполезна. Она не сможет выполнять поставленную задачу (распознавать эмоции). Данную структуру необходимо смоделировать. Для этой цели широко используется библиотека Keras – мощная и популярная библиотека для языка программирования Python, предназначенная для глубокого обучения.

Для обучения модели необходим датасет. В качестве его был выбран FER2013+ [6]. Данный датасет (рисунок 2) содержит 32298 изображений размером 48 на 48 в градациях серого, имеющий 7 категорий эмоций: злость, отвращение, страх, радость, нейтральное, грусть, удивление (что совпадает с моделью Пола Экмана за исключением нейтральной эмоции). Также была проведена аугментация данных (методы искусственного увеличения набора данных путём создания новых образцов из существующих данных, используя различные аффинные преобразования и др.) с помощью метода ImageDataGenerator в Keras.

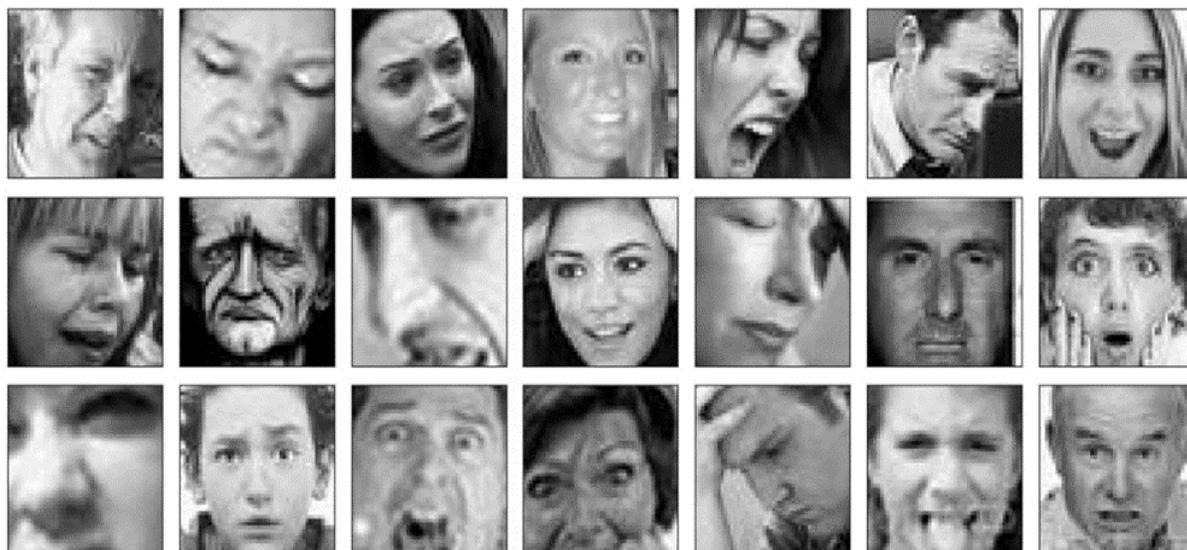


Рисунок 2. Пример изображений из FER2013+

После создания модели нейронной сети, перед её обучением, она требует компиляции, определяющей функцию потерь, с помощью которой минимизируются ошибки классификации; оптимизатор, он же – алгоритм минимизации функции потерь; метрики, оценивающей то, насколько хорошо обученная модель предсказывает класс для данных, которые она не видела. Для данной модели в качестве функции потерь служит кросс-энтропия, а в качестве оптимизатора – Адам с шагом обучения 0,001.

После компиляции следующим шагом является обучение. Модель обучалась с использованием CPU Intel I3 6006-U и GPU Nvidia MX-940. Среднее время обучения для одной эпохи составило 65 с. Конечные результаты следующие (рисунок 3):

- значение функции потерь на обучающем наборе данных: 0,414;
- ассигасу модели на обучающем наборе данных: 85,25%;
- значение функции потерь на валидационном наборе данных: 0,443;
- ассигасу модели на валидационном наборе данных: 85,23%.

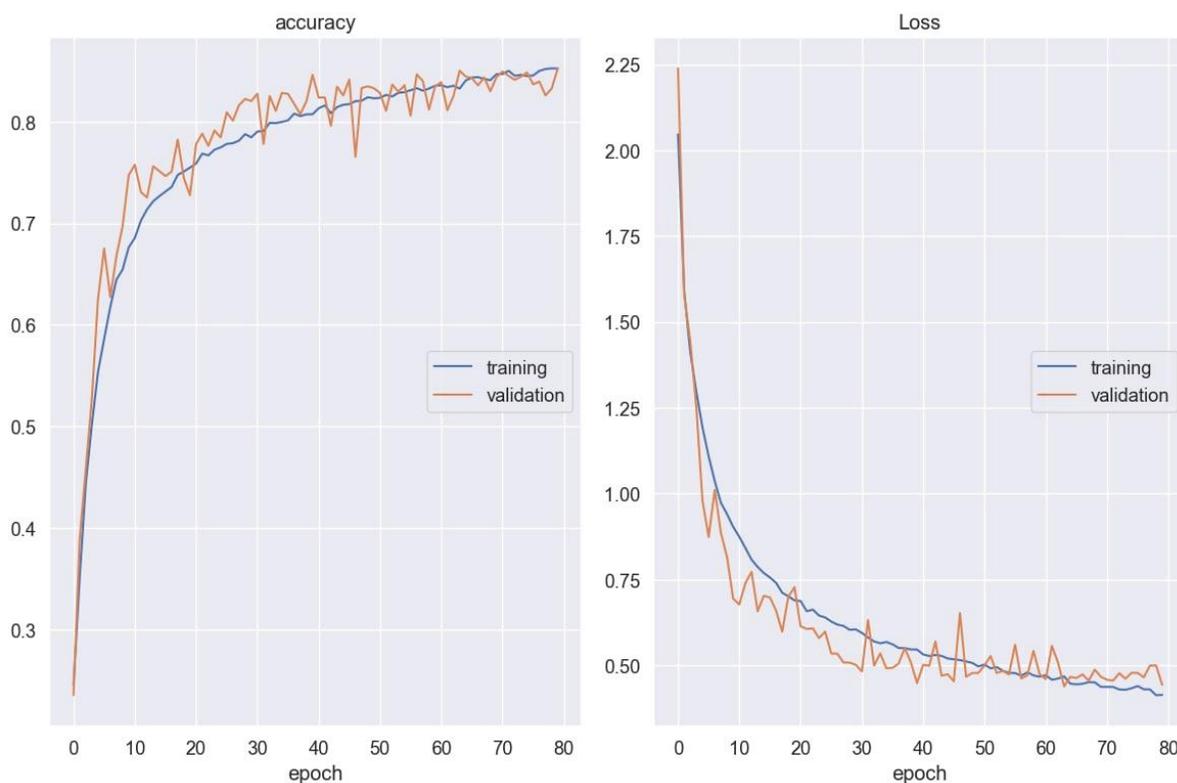


Рисунок 3. График изменения точности/ассигасу (отношение количества правильно классифицируемых изображений ко всем) и функции потерь/loss для обучающих и валидационных данных в зависимости от эпохи (1 эпоха – проход всех данных)

Далее полученный классификатор был испытан на тестовых данных, на которых были получены множество оценок обученной модели.

Из рисунка 4 видно, что точность определения нейтральной эмоции и грусти заметно просаживается относительно других классов. Это связано с тем, что данные эмоции достаточно похожи друг на друга, поэтому классификатор и ошибается заметно чаще.

В таблице 1 приведены метрики классификатора эмоций.

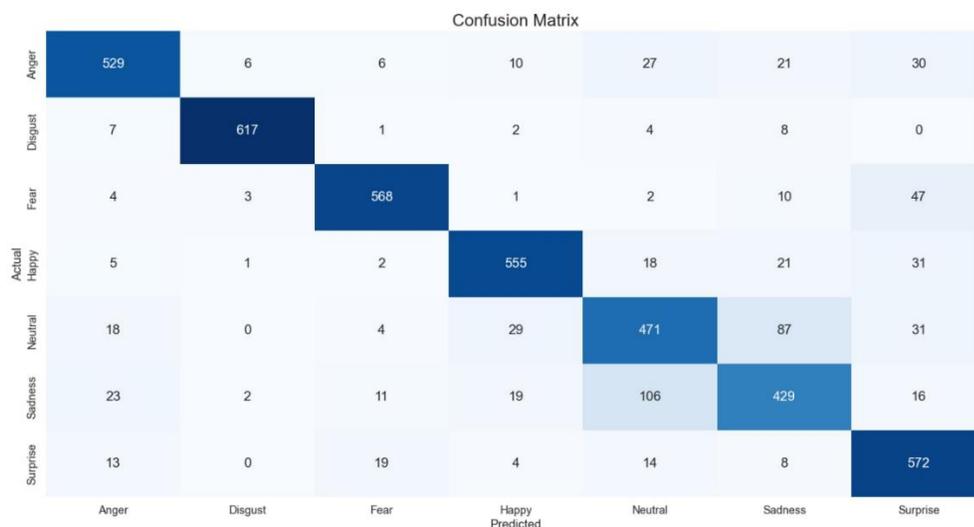


Рисунок 4. Матрица ошибок. Показывает точность прогнозирования модели (классификатора) в отношении 7 классов. Правильные прогнозы идут по диагонали от верхнего левого угла к нижнему правому

В таблице 1 приведены метрики классификатора эмоций.

Таблица 1. Метрики разработанного классификатора

Класс/Метрики	Precision	Recall	F1-score
Злость	0,88	0,84	0,86
Отвращение	0,98	0,97	0,97
Страх	0,93	0,89	0,91
Радость	0,90	0,88	0,89
Нейтральный	0,73	0,74	0,73
Грусть	0,73	0,71	0,72
Удивление	0,79	0,91	0,84

Исходя из рисунка 4 и таблицы 1 можно сделать вывод, что эмоция отвращения определяется классификатором наилучшим образом, в то же время как нейтральная эмоция и грусть – наихудшим.

Сравнение результатов классификации эмоций, полученные с помощью других наиболее известных структур нейронной сети представлены в таблице 2 [7].

Таблица 2. Сравнительные результаты характеристик классификаторов

Название модели	Точность	Функция потерь
Исходная	85 %	0,44
VGG16	60 %	1,10
VGG19	53 %	1,20
FerConvNet	65 %	1,07
EfficientNetB7	60 %	1,09
FerConvNet_Bilateral	65 %	1,09
FerConvNet_NonlocalMeans	65 %	1,09
FerConvNet_HDM	85 %	0,56

В ходе проделанной работы были сделаны следующие выводы: высокая точность модели обусловлена множеством факторов:

- использование сбалансированного, «чистого» датасета;
- предобработка изображений с помощью ImageDataGenerator;
- построение структуры свёрточной нейронной сети с использованием субдискретизации, паддинга, пакетной нормализации, дропаута и L-регуляризации.

- избежание эффекта переобучения;

- применение метрик для отслеживания качества обучения модели.

Литература

1. Вундт, В. – Очерк психологии / В. Вундт. – СПб.: Издание Ф. Павленкова, 1896. – 220 с.
2. Schlosberg H. Three dimensions of emotion. – Psychological Review. 61(2), p. 81–88.
3. Экман, П. Психология эмоций. Я знаю, что ты чувствуешь / Пол Экман. – СПб.: Питер, 2014. – 336 с.

4. Бутовская, М.Л. Язык тела: природа и культура / М.Л. Бутовская. – М.: Научный мир, 2004. – 440 с.

5. Гудфеллоу, Я. И. Глубокое обучение / Я. Гудфеллоу, И. Бенджио, А. Курвилль / пер. с англ. А. А. Слинкина. – 2-е изд., испр. – М.: ДМК Пресс, 2018. – 652 с.

6. Fer2013Plus [Электронный ресурс]. – Режим доступа: <https://www.kaggle.com/datasets/subhaditya/fer2013plus>. – Дата доступа: 15.11.2022.

7. Srinivas. Facial expression recognition for low resolution images using convolutional neural networks and denoising techniques / Pavan Nageswar Reddy Bodavarapu, Srinivas // Indian journal of science and technology. – Apr, 2021. – Vol. 14, № 12. – P. 971–983.