

СВЕРТОЧНАЯ НЕЙРОННАЯ СЕТЬ ДЛЯ ОПРЕДЕЛЕНИЯ ЭМОЦИОНАЛЬНОГО СОСТОЯНИЯ ЧЕЛОВЕКА

Шаймарданов Али Нурлыбекович

студент, Евразийский национальный университет им Л.Н. Гумилева, РК, г. Астана

Ламашева Жанар Бейбутовна

научный руководитель, PhD, старший преподаватель, Евразийский национальный университет им Л.Н. Гумилева, РК, г. Астана

Распознавание выражений лица является важной темой исследований в самых разных областях - от искусственного интеллекта и игр до взаимодействия человека и компьютера и психологии. Способность определить и понять эмоциональное состояние человека может иметь значительные последствия в различных областях, таких как психология, здравоохранение, обслуживание клиентов и даже взаимодействие человека и компьютера. В последние годы развитие нейронных сетей произвело революцию в области искусственного интеллекта, предоставив мощные инструменты для анализа сложных закономерностей и составления прогнозов. В этой статье мы рассмотрим процесс разработки нейронной сети для определения эмоционального состояния человека и ее потенциальное применение.

Первым шагом в разработке нейронной сети для определения эмоциональных состояний является сбор разнообразных и репрезентативных данных. Этот набор данных должен содержать примеры различных эмоций, выраженных с помощью различных модальностей, таких как изображения лица, аудиозаписи и текстовые данные. В качестве набора данных для исследования был выбран набор данных Facial Expression Recognition 2013 (рис. 1). Набор данных fer2013 представляет собой коллекцию изображений выражений лица, которая содержит 35 887 полутоновых изображений размером 48x48 пикселей. Каждое изображение представляет собой выражение лица человека и обозначено одной из семи категорий эмоций: гнев, отвращение, страх, счастье, грусть, удивление и нейтральность [1].



Рисунок 1. Набор данных FER 2013

Архитектура нейронной сети играет решающую роль в ее способности обучаться и обобщать данные. В случае обнаружения эмоций могут быть использованы различные типы нейронных сетей, такие как сверточные нейронные сети (CNN), рекуррентные нейронные сети (RNN) или их комбинация. В данной работе была выбрана архитектура MobileNetV2. Архитектура MobileNetV2 состоит из серии инвертированных остаточных блоков, где вход и выход блока соединены через слой узкого места (рис. 2). Такая структура позволяет значительно сократить количество необходимых параметров, сохраняя при этом высокую точность при решении задач классификации изображений. MobileNetV2 также включает функцию "линейные узкие места", которая позволяет улучшить поток информации через сеть и повышает ее способность к обобщению на новые данные [2].

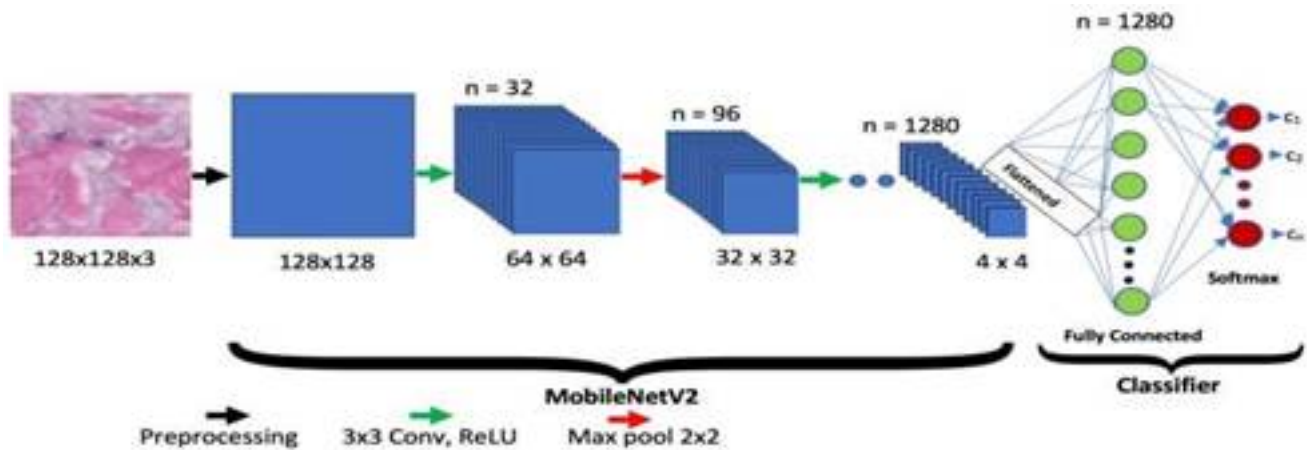


Рисунок 2. Архитектура сети MobileNetV2

После определения архитектуры нейронной сети набор данных делится на обучающий, проверочный и тестовый наборы. В данном исследовании использовалась библиотека Keras в python для приема изображений для обучения. Чтобы минимизировать потери нейронной сети во время обучения, в данном исследовании использовался алгоритм градиентного спуска Mini-Batch. Этот тип алгоритма градиентного спуска подходит из-за его возможностей в нахождении коэффициентов или весов нейронных сетей путем разделения набора данных для обучения на небольшие партии, т.е. обучающую и проверочную партии, используя технику увеличения данных. С увеличением количества сверточных слоев было также принято одновременное увеличение размера ядра, которое использовалось параллельно с функцией активации ReLU для обучения модели. Функция активации SoftMax и оптимизатор Адама были использованы для повышения эффективности классификации модели [3].

```
final_output = layers.Dense(128)(base_output) ## добавление нового слоя, после выхода слоя глобального объединения
final_output = layers.Activation('relu')(final_output) ## функция активации
final_output = layers.Dense(64)(final_output)
final_output = layers.Activation('relu')(final_output)
final_output = layers.Dense(7, activation='softmax')(final_output) ## my classes are 7, classification layer
```

Рисунок 3. Использование ReLu в качестве активационной функции

Алгоритм и модель распознавания эмоций по лицу, предложенные в данном исследовании, основаны на двух основных идеях: во-первых, использование высокопроизводительных глубоких конволюционных нейронных сетей для извлечения признаков и классификации эмоций, в которых используется один классификатор для обнаружения лиц с нескольких

точек зрения как в сценариях реального времени, так и на цифровых изображениях или видеокадрах. В данном исследовании также была предпринята попытка оптимизировать вычислительную сложность глубокой конволюционной нейронной сети (DCNN) путем изменения архитектуры за счет добавления слоев для улучшения идентификации образов в режиме реального времени или цифровых изображений. Дополнительные слои применяют больше фильтров свертки к изображению для обнаружения особенностей изображения. Для дальнейшего повышения эффективности и точности прогнозирования модели количество эпох обучения было увеличено до 25.

Алгоритм и модель распознавания эмоций по лицу, предложенные в данном исследовании, основаны на двух основных идеях: во-первых, использование высокопроизводительных глубоких конволюционных нейронных сетей для извлечения признаков и классификации эмоций, в которых используется один классификатор для обнаружения лиц с нескольких точек зрения как в сценариях реального времени, так и на цифровых изображениях или видеокадрах. В данном исследовании также была предпринята попытка оптимизировать вычислительную сложность глубокой конволюционной нейронной сети (DCNN) путем изменения архитектуры за счет добавления слоев для улучшения идентификации образов в режиме реального времени или цифровых изображений [4]. Дополнительные слои применяют больше фильтров свертки к изображению для обнаружения особенностей изображения. Для дальнейшего повышения эффективности и точности прогнозирования модели количество эпох обучения было увеличено до 25.

В распознавании эмоций в предложенной модели было использовано 3 этапа, т.е. обнаружение лица (рис. 4), извлечение признаков и классификация эмоций с помощью глубокого обучения, что обеспечивает более хорошие результаты по сравнению с предыдущей моделью. В предложенном методе, по мере уменьшения времени вычислений, точность валидации увеличивается, при этом значительно уменьшаются потери при валидации. Предложенная модель DCNN была протестирована на наборе данных FER-2013.



Рисунок 4. Результат алгоритма, предсказание эмоции счастье

Давайте теперь проанализируем нашу модель, определим точность распознавания и сделаем оценку. Сначала покажем кривую обучения модели, которая отображает эффективность обучения модели с течением времени или опыта (рис. 6). После каждого обновления модель оценивалась на тренировочном наборе данных и промежуточном проверочном наборе данных. Кривые обучения были выбраны в качестве графической метрики для данного исследования, так как они широко применяются в машинном обучении для моделей, которые оптимизируют

свои внутренние параметры постепенно. Из второго графика кривого обучения видно, что график потерь при обучении и потерь при проверке уменьшается до точки стабильности с минимальной разницей в обобщении. Кроме того, из первого графика можно сделать вывод, что график точности обучения и точности проверки резко возрастает с увеличением последовательности обучения и размера партии с минимальным разрывом обобщения. С тематической точки зрения, модель хорошо подходит и эффективно обобщается [5].

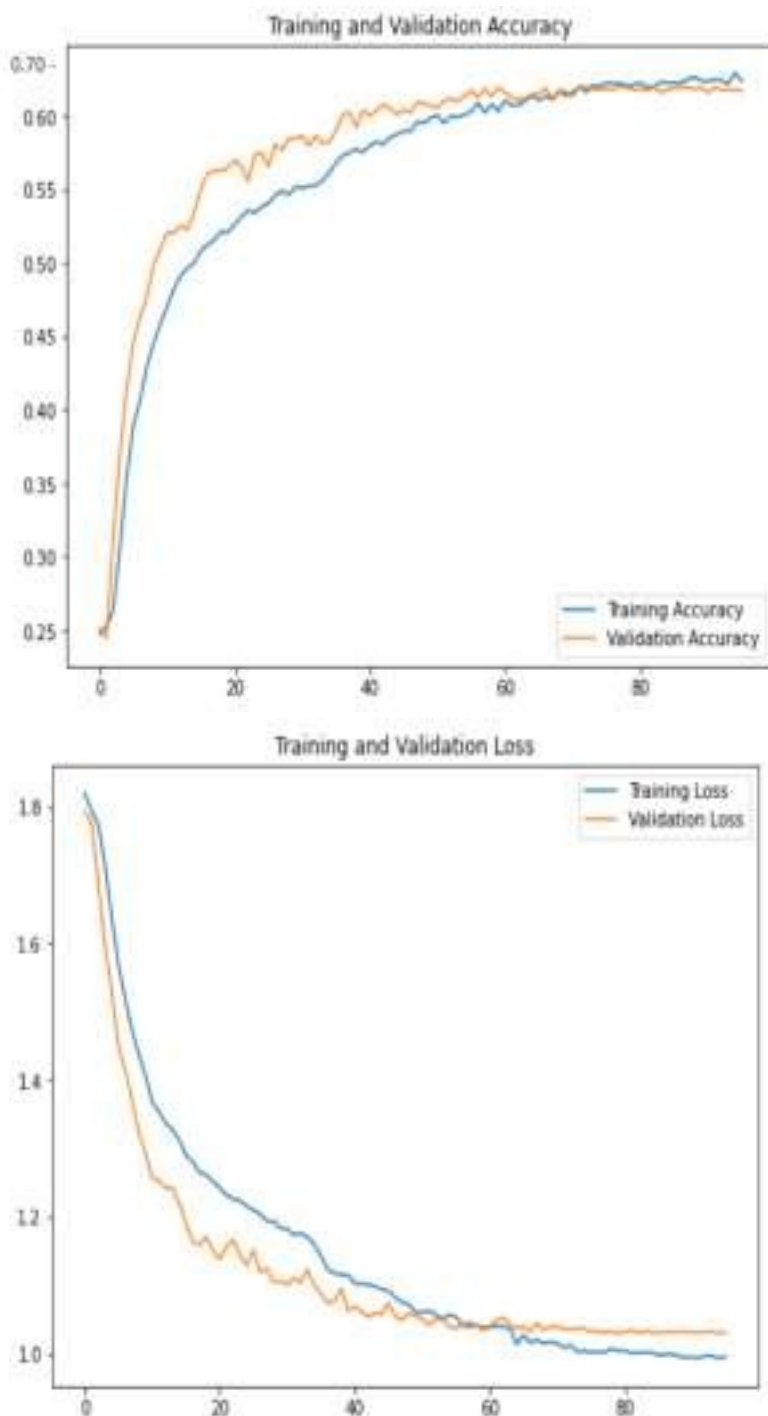


Рисунок 6. Графики точности и потери при обучении и валидации

В данной работе предложена глубокая сверточная-модель для распознавания эмоций на основе семи эмоциональных классов ("нейтральный", "счастливый", "грустный", "сердитый", "удивленный", "страх" и "отвращение"). Структура предложенной модели имеет хорошую

обобщенность и эффективность классификации. Во-первых, были получены различные высококачественные базы данных с хорошей классификацией. Затем область лица обнаруживается, вырезается и преобразуется в полутоновое изображение одного канала для удаления ненужной информации [6]. Для решения проблемы избыточной подгонки применяется расширение данных изображения, которое увеличивает количество и вариации обучающих изображений. Настройка гиперпараметров была использована для достижения современной точности классификации в 70%. В предложенной гибридной архитектуре была разработана оптимальная структура для сокращения времени выполнения и повышения эффективности классификации изображений в реальном времени и цифровых изображений. Это было сделано путем регулировки количества карт признаков в конволюционном слое, слоев в нейро-сетевой модели и множества эпох обучения. Эксперименты с перекрестной валидацией показали, что предложенная архитектура конволюционной нейронной сети имеет лучшую производительность классификации и универсальность, чем некоторые современные технологии. Каскадная модель Хаара, используемая для обнаружения лиц в видеопотоке или в изображении, показала лучшую производительность классификации по сравнению с другими реализациями. Экспериментальные результаты подтвердили эффективность методов предварительной обработки и дополнения данных [7].

Список литературы:

1. Srivastava N. Dropout: A Simple Way to Prevent Neural Networks from Overfitting / N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, R. Salakhutdinov // Journal of Machine Learning Research, 2014. – P. 1929-1958.
2. Sandler M., Howard A., Zhu M., Zhmoginov A., Chen L.C. Mobilenetv2: Inverted residuals and linear bottlenecks // IEEE conference on computer vision and pattern recognition: proc. (Salt Lake City, USA, 18–23 June, 2018). – USA, 2018. – P. 4520.
3. Флах П. Машинное обучение. Наука и искусство построения алгоритмов, которые извлекают знания из данных. – М.: ДМК Пресс, 2015. – 400 с.
4. Kar A. High Performance Human Face Recognition using Gabor Based Pseudo Hidden Markov Model // International Journal of Applied Evolutionary Computation (IJAEC), 2013. – P. 11-22.
5. Lopes AT de Aguiar E, Souza AFD, Oliveira-Santos T. Facial expression recognition with Convolutional Neural Networks: Coping with few data and the training sample order. Pattern Recognition. 2017. – P. 610-628.
6. Mayya V., Pai R.M., Pai M.M. Automatic Facial Expression Recognition Using DCNN. Procedia Computer Science. 2016. – P. 453-461.
7. Beale M.H., Hagan M. T., Demuth H.B. Neural Network Toolbox. User's Guide. — Natick: Math Works, Inc., 2015. — P. 406.