

УДК 004.85

## РЕШЕНИЕ ЗАДАЧИ РАСПОЗНАВАНИЯ ВЫРАЖЕНИЯ ЛИЦА С ПОМОЩЬЮ СВЕРТОЧНЫХ НЕЙРОННЫХ СЕТЕЙ

© Емельянов А.В., Солдатова О.П.

*Самарский национальный исследовательский университет  
имени академика С.П. Королева, г. Самара, Российская Федерация*

e-mail: sorryfor\_this@mail.ru

Распознавание выражений лица – сложная и многоаспектная задача, которая охватывает многие области, включая компьютерное зрение, машинное обучение и психологию. Это весьма сложное направление, поскольку выражения лица могут быть очень разнообразными и субъективными, а также они могут изменяться в зависимости от культурного и социального контекста. Кроме того, присутствуют технические сложности, такие как вариации в освещении, углах съемки и качестве изображений.

В данной работе проводится детальный анализ современных моделей нейронных сетей для распознавания выражений лица. Особое внимание уделяется сверточным нейронным сетям, которые в настоящее время являются одними из наиболее эффективных моделей в этой области.

Для улучшения точности распознавания и нормализации лица применяется метод каскадной регрессии. Этот метод позволяет точно определять особенности лица и может использоваться в режиме реального времени, что делает его полезным для многих приложений [1].

Архитектура используемой в работе сети основана на архитектуре VGG. Сеть содержит 29,9 миллиона параметров. Она состоит из 4 блоков и полносвязного классификатора.

Первый блок в этой сети отвечает за начальное извлечение признаков из входных данных. Он состоит из слоев свертки, за каждым из которых следует слой нормализации и слой макс-пулинга. К выходу свертки 1-го слоя применяется функция активации  $\text{relu}$ .

После свертки слой нормализации применяет преобразование, которое поддерживает среднюю активацию близкой к 0, а стандартное отклонение активации близкое к 1. Это приводит к стабилизации процесса обучения и резкому сокращению количества эпох обучения, необходимых для обучения глубоких сетей.

После другого набора слоев свертки и нормализации добавляется слой макс-пулинга для уменьшения пространственных размеров выходного объема. Он выполняет операцию максимального объединения, используя размер пула  $2 \times 2$ , эффективно уменьшая выборку карт объектов на половину их размера, что снижает стоимость вычислений для последующих слоев.

Далее в сети идут блоки 2, 3 и 4, каждый из которых следует той же схеме, что и блок 1, но с ключевым отличием: количеством фильтров в слоях свертки. Количество фильтров удваивается с каждым последующим блоком, увеличиваясь со 128 в блоке 2 до 256 в блоке 3 и до 512 в блоке 4. По мере того как сеть становится глубже, она изучает более сложные высокоуровневые функции.

Последней частью сети является классификатор, состоящий из двух полносвязных слоев со слоями дропаута между ними и выходного слоя. Перед этими слоями применяется слой преобразования 3D-вывода предыдущих слоев в 1D-вектор.

Первый полносвязный слой имеет 4096 единиц с функцией активации  $\text{relu}$ . Затем следует слой дропаута, который случайным образом с установленной вероятностью исключает нейроны из обучения, помогая предотвратить переобучение. Затем эти слои повторяются.

Наконец, сеть завершается выходным слоем с использованием функции активации «softmax», которая выводит распределение вероятностей по целевым классам. Функция «softmax» гарантирует, что сумма выходных вероятностей для всех классов равна 1, поэтому каждый выход можно интерпретировать как вероятность модели того, что входные данные принадлежат определенному классу.

При обучении модели используется стохастический градиентный спуск (SGD) с импульсом Нестерова. Этот подход направлен на решение двух проблем: плохой обусловленности матрицы Гессе и дисперсии стохастического градиента [2]. Матрица Гессе – это квадратная матрица вторых производных функции, играющая ключевую роль в оптимизации. При недостаточно хорошей обусловленности матрицы Гессе сходимость к минимуму функции ошибки может стать очень медленной, что неэффективно с точки зрения вычислительных ресурсов и времени. В SGD каждый шаг обновления весов происходит по одному, случайно выбранному примеру из набора данных, в отличие от обычного градиентного спуска, где шаги обновления происходят по всему набору данных. Это приводит к большой дисперсии в шагах обновления весов и может привести к менее стабильному обучению. Использование импульса Нестерова помогает уменьшить эту дисперсию и делает процесс обучения более сглаженным и стабильным.

Для обучения модели нейронной сети используется датасет FER2013, который содержит 28 709 тренировочных изображений, 3589 валидационных изображений и 3589 тестовых изображений с семью видами эмоций (гнев, отвращение, страх, радость, печаль, удивление и отсутствие эмоции). Обучение и тестирование на этом датасете позволили получить модель, способную эффективно распознавать выражения лица в реальном мире.

В рамках работы были построены графики обучения, которые позволяют визуализировать процесс обучения и проследить за его динамикой. Кроме того, были составлены матрицы ошибок, включающие показатели точности, полноты и оценку F1. Каждый из них равен 0,714. Эти показатели помогают оценить качество работы модели и выявить возможные направления для ее улучшения.

Также были созданы карты значимости, которые демонстрируют, на какие области изображения модель обращает особое внимание при принятии решений. Это помогает понять, какие особенности лица модель считает наиболее важными для определения эмоций.

Показатель аккуратности (accuracy) для разработанной модели составил 71,38 процента, что демонстрирует достаточно высокую эффективность в задаче распознавания выражений лица на основе датасета FER2013.

### Библиографический список

1. Xiong X., De la Torre, F. Supervised Descent Method and Its Applications to Face Alignment / X. Xiong, F. De la Torre et al. // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR). IEEE. 2013. P. 532–539.
2. Goodfellow I., Bengio Y., Courville A. Deep Learning. MIT Press, 2016. 781 p.