УДК 004.8

**© Кристина Сергеевна Светова**

студентка 4 курса, Институт математики и информационных технологий,

Петрозаводский государственный университет (Петрозаводск, Россия)

*Научный руководитель:* канд. физ.-мат. наук, доцент Дмитрий Жоржевич Корзун

**Задача распознавания эмоций человека по видеоданным**

**Аннотация.** В данной работе разрабатывается эффективное решение задачи распознавания эмоциональной оценки человека на основе методов искусственного интеллекта. Анализируется разработанное решение с существующими методами. В результате, на основе нейронных сетей и методов трансферного обучения, была построена модель, позволяющая распознать эмоциональную оценку с точностью = 97%.

**Ключевые слова:** распознавание эмоций, нейронные сети, глубокое обучение, методы дообучения модей

В настоящее время задача распознавания эмоций по видеоданным остаётся актуальной и важной, особенно в контексте разработки цифровых ассистентов [3]. Существующие методы искусственного интеллекта пока не достигли достаточной эффективности в этой области. Для более точного распознавания эмоций необходимо исследовать альтернативные подходы, которые учитывают множество признаков, способных влиять на принятие решений. Одним из таких подходов является применение глубокого обучения и нейронных сетей для улучшения результатов распознавания. В настоящее время считается, что эффективными методами являются использование глубокого обучения, таких как сверточные или рекуррентные нейронные сети, ансамблей моделей, комбинирующих несколько моделей на различных архитектурах, а также адаптивных признаков, которые учитывают контекстуальные особенности изображения. Дополнительно, аугментация данных, включающая генерацию новых примеров с использованием различных преобразований, таких как изменение яркости, контраста, добавление шума и других, также может существенно улучшить определение эмоциональной оценки в видеопотоке. Исследования в области распознавания эмоций приобретают все большую значимость в контексте разработки цифровых ассистентов, которые помогают мониторировать и поддерживать здоровье, а также автономные, когнитивные и двигательные функции человека [1 ; с. 86—87]. Постоянное совершенствование методов и алгоритмов для распознавания эмоций позволит создать более точные и эффективные системы, способные адекватно взаимодействовать с людьми и помогать им в повседневной жизни.

В качестве набора данных для проведения экспериментов используется набор FER2013, состоящий из 30 000 изображений лиц RGB формата, размером 48х48 пикселей, размеченных по 7 классам (гнев, отвращение, страх, счастье, грусть, удивление, нейтральное состояние) с содержанием по 5 000 образцов каждого класса за исключением класса, “отвращение”, содержащим 600 объектов.

Ранее [2 ; с. 258—259] строится сверточная модель нейронной сети с подбором параметров GridSearch позволяющим получить оценку равную 86% точности (accuracy), что будет считаться начальной оценкой для улучшения в текущем работе.

Для построения модели с более высокими метриками точности можно рассмотреть возможность использования предобученных моделей, имеющих лучшую обобщающую способность и эффективного использования ресурсов и времени, затрачиваемых на обучение модели. В качестве такой модели в данной задачи предлагается использовать вариант модели VGG — VGG19, использованной для решения задачи классификации изображений, построенной на основе сверточной нейронной сети и состоящей из 19 слоёв: 16 сверточных, 3 полносвязных, 5 слоев MaxPool для уменьшения размерности пространственного представления признаков и 1 слоя функции активации SoftMax, изначально обученной на наборе данных ImageNet задачи детектирования лиц на 14 млн изображений [4 ; с. 835—837].

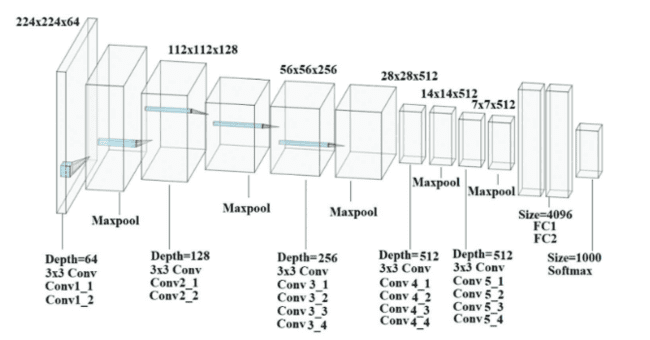


Рис. 1. Архитектура модели VGG-19

Представленную модель VGG-19 можно применить для прикладной задачи распознавания эмоций при условии выполнения трансферного обучения, при котором предварительно сохраняются параметры слоев и лишь последние n-слоёв проходят процесс обучения на основе новых данных, характерных для решаемой задачи. Данный метод обучения имеет право на существование только при схожести начальной задачи и решаемой (в текущем случае - задачи детекции и распознавания эмоции), в отличной ситуации может потребоваться подбор параметров на всех слоях из-за различия набора данных, вследствии чего подобранная архитекутура теряет свою производительность. Операция трансферного обучения будет заключаться в изменении предпоследних слоев, определяющих высокоуровневые признаки. Данные будут проходить через слой MaxPool для усреднения значения экземпляра объекта, а затем через полносвязный слой с функцией активации для вычисления вероятности принадлежности к уникальному классу, соответствующему одной из 7 эмоций. Процессу обучения подвергаются только последние измененные слои, остальные параметры слоев остаются фиксированными.

Для преодоления проблемы вариативности данных, которая может оказать влияние на недостаточную обобщающую способность модели, предлагается применить метод аугментации данных (генерирования на основе текущего набора новых изменённых фильтрами объектов), что в свою очередь может повлиять на увеличение эффективности распознавания. Так, для каждого из 30 000 образцов набора данных были использованы следующие фильтры:

* зеркальное отображение
* поворот на случайный угол
* масштабирование изображения
* увеличение шума
* изменение насыщенности цвета

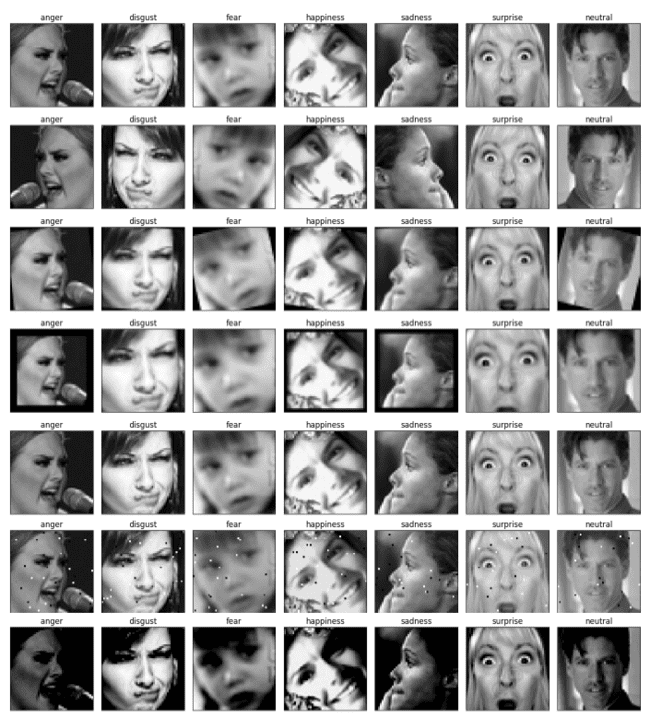


Рис. 2. Применение фильтров к входному набору данных

Модель проходила обучение со следующими параметрами:

* + использование выборки из 215 322 образцов
  + обучение на 32 образцах за одну итерацию обучения
  + прохождение 25 проходов по всему набору данных
  + использование Adam оптимизатора
  + использование функции потерь кросс-энтропии
  + прохождение тестирования на кросс-валидации

Используя данную модель, были получены результаты на тестовом наборе данных, превосходящие полученные ранее, тем самым определить алгоритм, способный совершать эффективные предсказания.

*Таблица*

**Результаты экспериментов**

|  |  |  |
| --- | --- | --- |
| Метод | Описание | Точность |
| CNN | Использование сверточной модели, подбор параметров GridSearch | 86% |
| CNN + VGG-19 | Использование предобученной сверточной модели с трансферным обучением | 92% |
| CNN + VGG-19 + аугментация | Использование предобученной сверточной модели с трансферным обучением и обучением с помощью расширенного набора данных | 97% |

В результате удалось разработать прототип модуля распознавания эмоций для прикладных задачей с целью получения результатов высокой точности - 97% с использованием предобученной модели, методов трансферного обучения и аугментации данных.

**Список литературы**

1. *Корзун Д.Ж.* О цифровом ассистенте для мониторинга жизнестойкости человека в условиях повседневной жизни // Большие данные и проблемы общества. 2022. С. 86—89.

2.*Светова К. С.* Возможности технологий искусственного интеллекта по распознаванию эмоционального состояния детей в умном доме // Научно-исследовательская работа обучающихся и молодых учёных : материалы 73-й Всероссийской (с международным участием) научной конференции обучающихся и молодых учёных. Петрозаводск, 2021. С. 257—260.

3. *Korzun D., Balandina E., Kashevnik A., Balandin S., Viola F.* Ambient Intelligence Services in IoT Environments: Emerging Research and Opportunities. IGI Global, 2019. 199 p.

4. *Sudha V., Ganeshbabu T. R.* A Convolutional Neural Network Classifier VGG-19 Architecture for Lesion Detection and Grading in Diabetic Retinopathy Based on Deep Learning // Computers, Materials and Continua (Tech Science Press). 2020. С. 827—842.