Министерство науки и высшего образования Российской Федерации

ФГАОУ ВО «Северо-Восточный федеральный университет

имени М.К Аммосова»

Институт математики и информатики

Кафедра «Информационные технологии»

Архитектура

на тему: Распознавание эмоций по фотографии

Выполнили: студенты 1 курса СВФУ ИМИ

гр. М-ИВТ-21 Лотов А.Р, Заморщикова Д.A. , Дегтярев А.В

Проверил: доцент научно-исследовательской кафедры ВТ

Григорьев Александр Виссарионович

Якутск, 2022

В проекте используется архитектура свёрточной нейронной сети. Данная архитектура часто используется в работах с изображениями.

Модель строится из 3 основных слоёв:

* Свертка (Convolution);
* Субдискретизация (pooling);
* Полносвязная нейронная сеть.

Свертка — это суммирование результата поэлементного произведения каждого фрагмента на ядро. Весовые коэффициенты ядра свёртки неизвестны и находятся по мере обучения модели (рис 1)[1].

Изображение выглядит как текст, электроника, мобильный телефон

Автоматически созданное описание

Рисунок 1. Свёртка

Субдискретизация – это уменьшение размерности сформулированных карт признаков. В данном проекте использовали MaxPool (рис 2)[].

Изображение выглядит как стол

Автоматически созданное описание

Рисунок 2. MaxPool

Нейронная сеть обучалась на датасете из Kaggle [2]. Датасет состоит из фотографий с размерностью 48x48 пикселей, и рассортированным по папкам (рис 3).

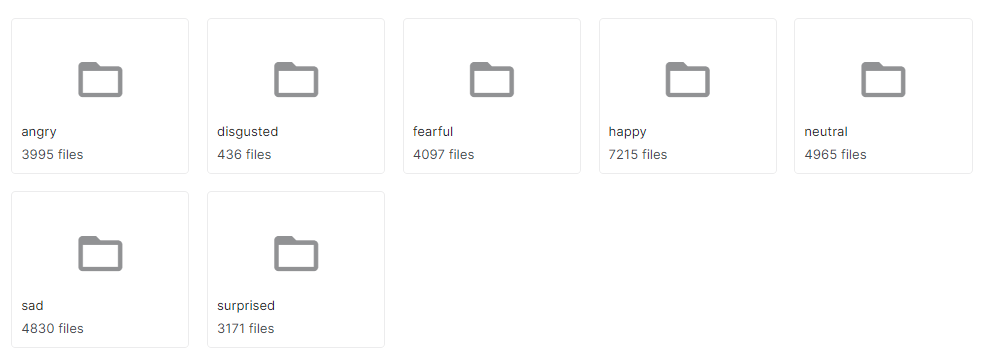


Рисунок 3. Датасет

Для тренировки нейронной сети использовали ImageDataGenerator где рандомно изменяет вытянутость или поворот по горизонтали, для уменьшения вероятности переобучения.

flow\_from\_directory для создания выборки (рис 4)

Изображение выглядит как текст, внутренний, снимок экрана, черный

Автоматически созданное описание

Рисунок 4. flow\_from\_directory

На вход подаётся изображение 48x48 пикселей и класс. А на выходе выходит массив из 7 элементов, где 1 элемент злость, 2 отвращение, и т. д.

В итоге модель выглядит так (рис 5).

Изображение выглядит как текст

Автоматически созданное описание

Рисунок 5. Модель

Dropout исключает некоторые случайные нейроны, для уменьшение переобучения сети.

Batchnormalization нормализует данные, нужен для ускорения обучения.

Flatten используется для конвертации входящих данных в меньшую размерность.

Для определения лица использовали библиотеку open-cv, модель ssd. Обрезаем фото по лицу и уменьшаем размер изображения до 48x48 пикселей, и прогоняем по нейронной сети. В ответе выходит массив из 7 элементов, оттуда находим самый вероятный элемент.

# Список литературы

1. <https://towardsdatascience.com/intuitively-understanding-convolutions-for-deep-learning-1f6f42faee1>
2. https://www.kaggle.com/ananthu017/emotion-detection-fer