

ОБУЧЕНИЕ И АНАЛИЗ РАБОТЫ СвёрТОЧНОЙ НЕЙРОННОЙ СЕТИ ДЛЯ РАСПОЗНАВАНИЯ ВЫРАЖЕНИЙ ЛИЦА

Белорусский государственный университет информатики и радиоэлектроники
г. Минск, Республика Беларусь

Вашкевич Г.С.

Азаров И.С. – д.т.н., доцент

Свёрточные нейронные сети широко применяются в задачах распознавания образов. В данной работе рассматривается их применение к задаче распознавания выражений человеческого лица. При этом современные методы визуализации работы свёрточных нейронных сетей позволяют оценить качество обучения сети, а также определить, какие регионы изображения и в какой степени влияют на результат работы сети.

Для обучения была выбрана свёрточная нейронная сеть, состоящая из трёх свёрточных слоёв и двух полносвязных слоёв. Архитектура сети представлена на рисунке 1. Фильтры всех свёрточных слоёв имеют размеры 5x5. В первом свёрточном слое 64 фильтра, во втором – 128, а в третьем – 256 фильтров. Первый 2 слоя max-объединения используют окно, размером 2x2 с шагом 2. Третий слой max-объединения использует окно размером 4x4 с шагом 4. Первый полносвязный слой имеет 300 нейронов. Во втором полносвязном слое 8 нейронов – по числу классов, данный слой использует SoftMax в качестве функции активации [1].

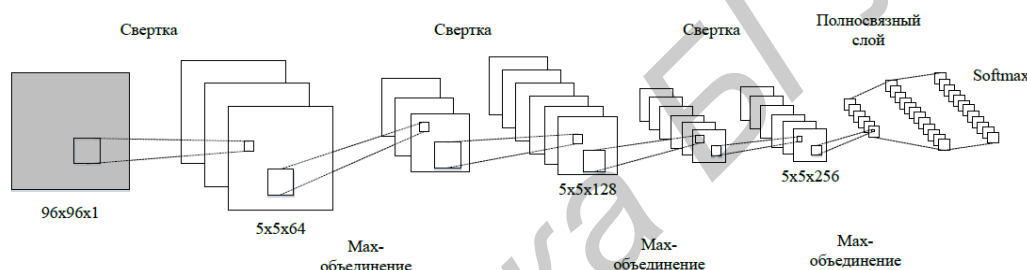


Рис. 1 – Архитектура свёрточной нейронной сети

В качестве набора данных для обучения использовалась база Extended Cohn-Kanade [2], которая состоит из 486 серий фотографий 97 человек. Каждая серия содержит изображения с нарастающим проявлением одной из доступных эмоций, начиная от нейтрального выражения лица и заканчивая максимальным выражением определенной эмоции. В данном наборе представлены 7 базовых эмоций: гнев, отвращение, презрение, страх, радость, грусть и удивление. Нами был введен дополнительный класс, отражающий нейтральное выражение лица. Элементами данного класса считались по два первых изображения из каждой последовательности фотографий.

Обучение сети проводилось при помощи метода Adagrad [3] со значением learning rate равным 0,001. Точность распознавания выражений лица на тестовом наборе данных составляет 99,6%.

Для визуализации сети использовался метод, вычисляющий карты заметности, суть которого заключается в вычислении градиента от выхода сети с максимальной активацией по отношению к входам сети [4]. Пример визуализации работы свёрточной сети показан на рисунке 2.

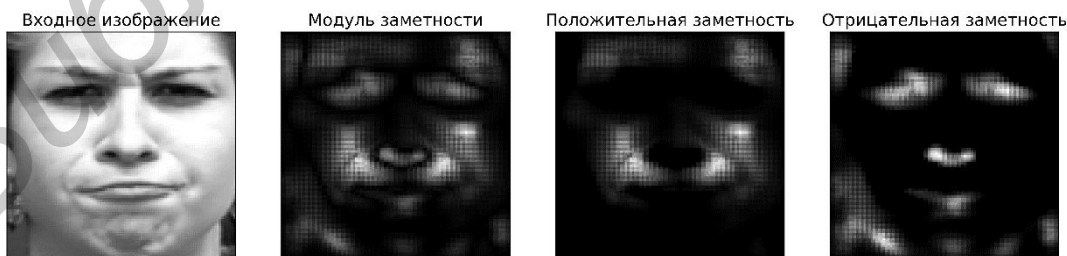


Рис. 2 – Визуализация работы свёрточной сети методом вычисления карт заметности

Список использованных источников:

1. P. Khorrami, T. Paine, T. Huang, Do Deep Neural Networks learn facial action units when doing expression recognition?
2. P. Lucey, J. Cohn, T Kanade, The Extended Cohn-Kanade Dataset (CK+): A complete dataset for action unit and emotion-specified expression, Computer Vision and Pattern Recognition, 2010.
3. J. Duchi, E. Hazan, Y. Singer, Adaptive subgradient methods for online learning and stochastic optimization.
4. K. Simonyan, A. Vedaldi, A. Zisserman, Deep Inside Convolutional Networks: Visualising Image Classification Models and Saliency Maps.
5. J. Springenberg, A. Dosovitskiy, T. Brox, Striving for Simplicity: The All Convolutional Net.