

# Методы для распознавания эмоций, настроения, пола и возраста человека

Д.Д. Прибавкин<sup>1</sup>, П.Ю. Якимов<sup>1,2</sup>

<sup>1</sup>Самарский национальный исследовательский университет им. академика С.П. Королева, Московское шоссе 34А, Самара, Россия, 443086

<sup>2</sup>Институт систем обработки изображений РАН - филиал ФНИЦ «Кристаллография и фотоника» РАН, Молодогвардейская 151, Самара, Россия, 443001

**Аннотация.** Распознавание на изображениях не только образов, но и метаданных становится все популярнее среди исследователей в области сверточных нейронных сетей и глубокого обучения. В данной статье приводится аналитический обзор современных программных решений, распознающих на изображениях эмоции, настроение, пол и возраст человека. Энтузиасты изобретают все новые и новые архитектуры сверточных нейронных сетей, позволяющие решать поставленные задачи с немалой точностью распознавания.

## 1. Введение

За последние годы наблюдается стремительное развитие технологий параллельной обработки данных, в особенности благодаря развитию графических процессоров, предназначенных теперь уже не только для компьютерной графики.

Это позволило обучать даже самые сложные по своим архитектурам нейронные сети и открыло целый горизонт нерешаемых прежде задач. Современные интеллектуальные системы сосредотачиваются не только на распознавании образов из входного изображения, но и учатся вычленять метаданные из распознанных объектов, такие как эмоции, настроение, пол или возраст человека.

Множество исследователей и энтузиастов в области машинного обучения и сверточных нейронных сетей разрабатывают и предлагают на обозрение общественности собственные, уникальные решения, отличающиеся как идейно, так и технически.

Данная статья предлагает аналитический обзор следующих программных решений в области распознавания эмоций, настроения, пола и возраста человека:

- Emotion recognition using Deep Convolutional Neural Network [1].
- A Compact Soft Stagewise Regression Network [2].
- Real-time Convolutional Neural Networks for Emotion and Gender Classification [3].
- Deep Expectation of Real and Apparent Age from a Single Image Without Facial Landmarks [4].

## 2. Обзор существующих решений

### 2.1. Emotion recognition using Deep Convolutional Neural Networks

Решение, представляющее собой обученную нейронную сеть, распознающую в реальном времени эмоции на человеческом лице, распознанном из входного видеопотока. Ее построение производилось с помощью программной библиотеки TFLearn для языка программирования python, основанной на известном фреймворке для машинного обучения TensorFlow, разработанным компанией Google в 2015 году [7]. Данный фреймворк упрощает разработку сети, так как требует описания только лишь самих слоев, вместо того, чтобы описывать каждый нейрон по отдельности, а также упрощает обучение сети, предоставляя обратную связь процесса в реальном времени и точность обучения. Более того, библиотека позволяет сохранять результат натренированной модели и переиспользовать его в дальнейшем.

Модель итоговой нейронной сети представлена на рисунке 1.

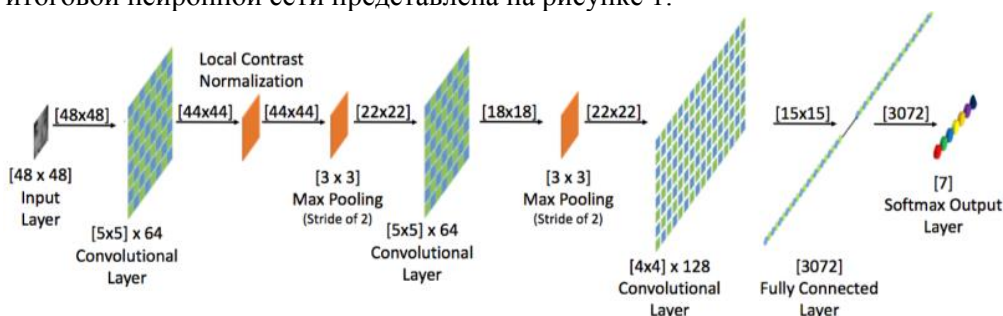


Рисунок 1. Модель нейронной сети.

В каждом кадре видеопотока производится попытка распознавания человеческого лица (лиц). Достигается это с помощью методов распознавания открытой библиотеки OpenCV [5]. Затем, если на изображении было распознано лицо, это лицо вырезается и масштабируется до размера 48x48 пикселей. Только после этого оно подается на вход нейронной сети. Таким образом, мы получаем оптимизированное программное обеспечение, затрагивающее ресурсы нейронной сети только, если в кадре присутствует хотя бы одно человеческое лицо.

Модель обучалась с помощью датасета FER-2013, насчитывающего порядка 20000 изображений, содержащих примеры проявления следующих эмоций: гнев, страх, счастье, грусть, удивление, безразличие и отвращение. Плотность распределения эмоций в данном датасете отражена на рисунке 2.

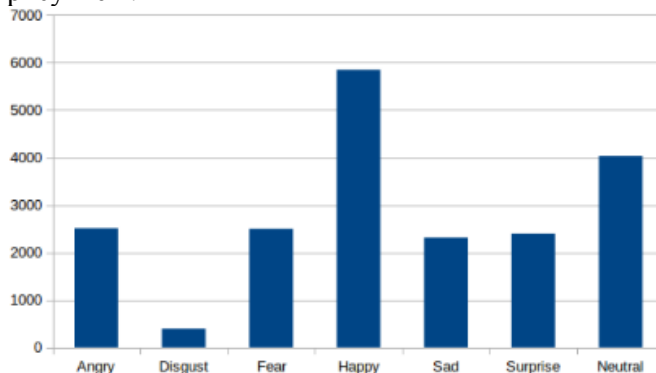


Рисунок 2. Плотность распределения эмоций в датасете FER-2013

По результатам обучения была достигнута точность распознавания эмоций в 67%, что довольно неплохо для ограниченной в ресурсах группы исследователей из MIT.

### 2.2. SSR-Net

Данное решение представляет из себя оригинальную со слов автора нейронную сеть с мягкой ступенчатой регрессией (soft stagewise regression network) для распознавания возраста и пола.

Сеть распознает возраст и пол по следующему принципу: на вход данной сети подаются изображения формата 64x64 пикселей, после производится многоуровневая классификация из нескольких классов, где каждый уровень служит для уточнения предыдущего результата, а затем с помощью регрессии обрабатывается результат классификации. Модель сама по себе очень компактна и занимает всего лишь 0,32 МБ. Но не смотря на свои компактные размеры, производительность SSR-Net приближается к характеристикам самых современных методов, размеры моделей которых больше в 1500 раз. Модель данной нейронной сети с тремя уровнями и размером пула 2 представлена на рисунке 3.

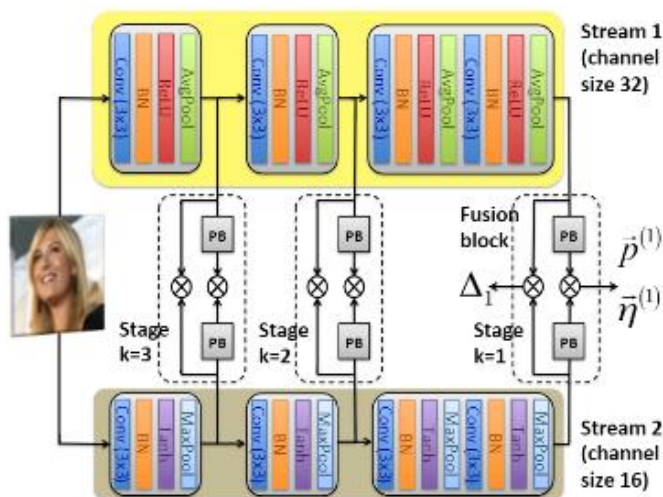


Рисунок 3. Модель нейронной сети SSR-Net.

Для обучения данной модели использовались такие датасеты как IMDB, WIKI и Morph2 [6]. Около 80% случайно выбранных изображений из датасетов были использованы для обучения сети, а остальные 20% - для тестирования.

Пример зависимости количества ошибок распознавания сетей SSR-Net, MobileNet и DenseNet, обученных на датасете Morph2, от количества эпох представлен на рисунке 4.

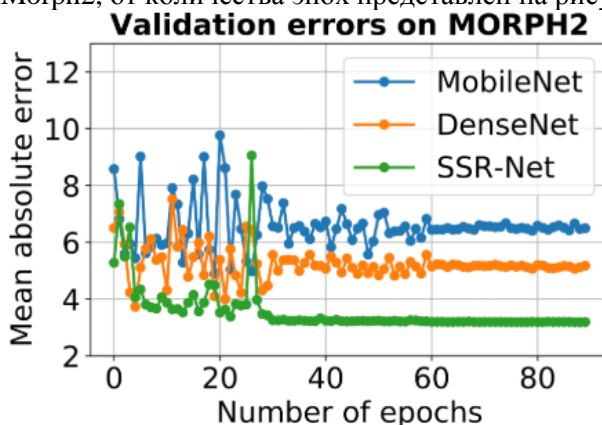


Рисунок 4. График зависимости ошибок распознавания от количества эпох.

### 2.3. Face classification and detection from the B-IT-BOTS robotics team

Представляет из себя классификатор эмоций и пола человека в реальном времени, основывающийся на сверточной нейронной сети и открытой библиотеки обработки изображений openCV [5]. Модель данной нейронной сети представлена на рисунке 5 и содержит 600,000 параметров.

Данная модель обучалась на датасете IMDB, насчитывающем порядка 460,723 RGB изображений, каждое из которых принадлежит одному из классов: мужчина или женщина [6].

На данном датасете была достигнута точность распознавания в 96%. Также данная модель валидировалась на датасете FER-2013, который включает в себя 35,887 изображений в серых тонах, каждое из которых принадлежит одному из классов эмоций: злость, отвращение, страх, радость, грусть, удивление и безразличие. На данном датасете была достигнута точность в 66%.

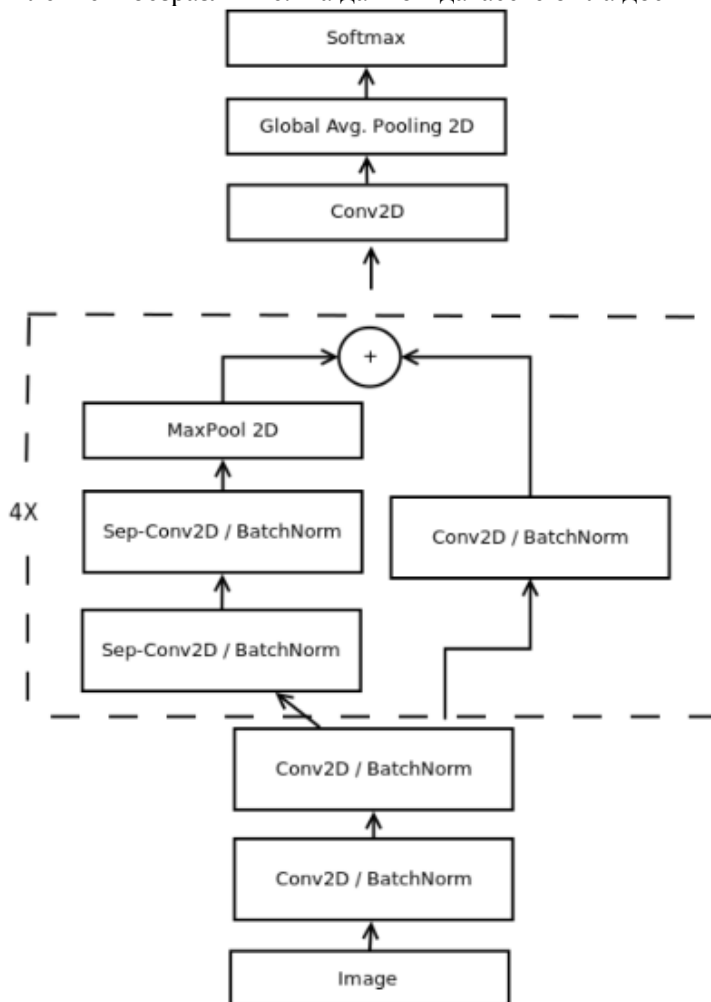


Рисунок 5. Модель нейронной сети.

#### 2.4. Age and gender estimation

Данное решение представляет из себя реализацию сверточной нейронной сети для распознавания пола и возраста человека по входному изображению. За основу была взята архитектура сети VGG-16 за счет ее глубины и управляемости. Данная сеть принимает на вход изображения размером 256x256 пикселей.

Обучение производилось на датасетах IMDB-WIKI, и была достигнута точность распознавания в 64% [6].

### 3. Заключение

В итоге, мы можем заключить, что такие задачи, как распознавание эмоций, настроения, пола и возраста пользуются большой популярностью у исследователей всего мира. Энтузиасты используют различные подходы к реализации интеллектуальных систем, способных решать такие задачи и достигают неплохих результатов в точности распознавания даже при ограниченных ресурсах. Основным же средством реализации поставленных задач являются сверточные нейронные сети различных архитектур, обученные на известных в сети датасетах изображений.

#### 4. Литература

- [1] Correa, E. Emotion Recognition using Deep Convolutional Neural Network / E. Correa, A. Jonker, M. Ozo, R. Stolk [Electronic resource]. – Access mode: [https://github.com/isseu/emotion-recognition-neural-networks/blob/master/paper/Report\\_NN.pdf](https://github.com/isseu/emotion-recognition-neural-networks/blob/master/paper/Report_NN.pdf) (1.11.2018).
- [2] Tsun-Yi, Y. SSR-Net: A Compact Soft Stagewise Regression Network for Age Estimation / Y. Tsun-Yi, H. Yi-Hsuan, Y.-Y. Lin, H. Pi-Cheng, Y.-Y. Chuang [Electronic resource]. – Access mode: [https://github.com/shamangary/SSR-Net/blob/master/ijcai18\\_ssnet\\_pdfa\\_2b.pdf](https://github.com/shamangary/SSR-Net/blob/master/ijcai18_ssnet_pdfa_2b.pdf) (14.11.2018).
- [3] Arriaga, O. Real-time Convolutional Neural Networks for Emotion and Gender Classification / O. Arriaga, P.G. Plöger, M. Valdenegro [Electronic resource]. – Access mode: [https://github.com/oarriaga/face\\_classification/blob/master/report.pdf](https://github.com/oarriaga/face_classification/blob/master/report.pdf) (8.10.2018).
- [4] Rothe, R. Deep Expectation of Real and Apparent Age from a Single Image Without Facial Landmarks / R. Rothe, R. Timofte, L. Van Gool // International Journal of Computer Cision. – 2018. – Vol. 126(2-4). – P. 144-157.
- [5] Библиотека обработки изображений OpenCV [Electronic resource]. – Access mode: <http://opencv.org> (04.10.2018).
- [6] IMDB-wiki dataset [Electronic resource]. – Access mode: <https://data.vision.ee.ethz.ch/cvl/rrothe/imdb-wiki/> (04.10.2018).
- [7] Библиотека для глубокого обучения TFLearn [Electronic resource]. – Access mode: <http://tflearn.org/> (04.10.2018).

## Methods for emotions, mood, gender and age recognition

D.D. Pribavkin<sup>1</sup>, P.Y. Yakimov<sup>1,2</sup>

<sup>1</sup>Samara National Research University, Moskovskoe Shosse 34A, Samara, Russia, 443086

<sup>2</sup>Image Processing Systems Institute of RAS - Branch of the FSRC "Crystallography and Photonics" RAS, Molodogvardejskaya street 151, Samara, Russia, 443001

**Abstract.** Recognition on images not only of images, but also of metadata is becoming increasingly popular among researchers in the field of convolutional neural networks and deep learning. This article provides an analytical overview of modern software solutions that recognize the images of emotions, mood, gender and age of a person. Enthusiasts invent all new and new architectures of convolutional neural networks, allowing to solve the tasks with considerable recognition accuracy.