

МИНОБРНАУКИ РОССИИ

Федеральное государственное бюджетное образовательное учреждение
высшего образования

**«САРАТОВСКИЙ НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ
ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ ИМЕНИ Н.Г. ЧЕРНЫШЕВСКОГО»**

Кафедра теоретических основ
компьютерной безопасности и
криптографии

Обнаружение лиц на изображениях с помощью нейросетевой модели

АВТОРЕФЕРАТ

дипломной работы

студента 6 курса 631 группы

специальности 10.05.01 Компьютерная безопасность

факультета компьютерных наук и информационных технологий

Кузьмина Владислава Олеговича

Научный руководитель

доцент

И. И. Слеповичев

22.01.2022 г.

Заведующий кафедрой

д. ф.-м. н., доцент

М. Б. Абросимов

22.01.2022 г.

Саратов 2022

ВВЕДЕНИЕ

В 1943 году У. Маккалок и У. Питтс формализуют понятие нейронной сети в фундаментальной статье о логическом исчислении идей и нервной активности. С тех пор эта идея получила огромное развитие – появлялись новые типы, улучшались старые; определялись классы задач, решаемые с помощью нейронных сетей. Множество математиков и по сей день продолжают работу в данной области.

Одним из направлений, в котором активно используются нейронные сети, является компьютерное зрение – теория и технология создания искусственных систем, получающих информацию из изображений и выполняющих определённые задачи: распознавание, оценка движения, восстановление сцены и изображений.

В данной работе продемонстрировано применение нейронных сетей в теории компьютерного зрения для решения такой задачи, как детектирование лиц на фотографии, которая является частным случаем задачи распознавания объектов. Актуальность работы заключается в повсеместном развитии и применении технологий детектирования и распознавания лиц, в том числе и в целях безопасности, например, распознавание лиц в метро и на предприятиях, работающих с конфиденциальной информацией.

Целью работы является создание нейросетевой модели для решения задачи обнаружения лиц на фотографии, и для её достижения требуется решить следующие подзадачи:

- 1) изучить существующие подходы к детектированию лиц;
- 2) выбрать оптимальный подход, реализовать его программно и объединить с методом распознавания лиц для создания практического приложения.

Дипломная работа состоит из введения, 3 разделов, заключения, списка использованных источников и 1 приложения. Общий объем работы – 47

страниц, из них 34 страницы – основное содержание, включая 26 рисунков, список использованных источников из 19 наименований.

КРАТКОЕ СОДЕРЖАНИЕ

Первый раздел содержит необходимые теоретические сведения, касающиеся нейронных сетей. Здесь приводятся основные определения и обозначения, используемые далее в дипломной работе, а также применения и этапы решения задач с помощью нейронных сетей.

Второй раздел «Детектирование лиц с помощью ИНС» начинается с описания истории решения задачи детектирования лиц с помощью таких методов, как алгоритм Виолы-Джонса, DPM, многослойных и радиально-базисных нейронных сетей, а также представляет структуру, позднее использованную в приложении – MTCNN. Она состоит из трех этапов: на первом она быстро определяет предполагаемые области с лицами с помощью свёрточной сети с небольшим количеством слоев. Затем она фильтрует полученные результаты, отбрасывая большое количество неправильно выбранных областей с помощью более сложной сети. Наконец, она использует ещё более мощную сеть для итогового уточнения результата и вывода позиций лицевых ориентиров.

В следующих подразделах представлены определения, касающиеся свёрточных нейронных сетей.

Свёртка (convolution) – операция над парой матриц A (размера $n_x \times n_y$) и B (размера $m_x \times m_y$), результатом которой является матрица

$$C = A * B$$

размера $(n_x - m_x + 1) \times (n_y - m_y + 1)$. Каждый элемент результата вычисляется как скалярное произведение матрицы B и некоторой подматрицы A такого же размера (подматрица определяется положением элемента в результате). То есть,

$$C_{ij} = \sum_{u=0}^{m_x-1} \sum_{v=0}^{m_y-1} A_{i+u, j+v} B_{u,v}$$

В свёрточной нейронной сети выходы промежуточных слоев образуют *матрицу* (изображение) или *набор матриц* (несколько слоёв изображения). Так,

например, на вход свёрточной нейронной сети можно подавать три слоя изображения (R-, G-, B-каналы изображения). Основными видами слоев в свёрточной нейронной сети являются *свёрточные слои (convolutional layer)*, *пулингговые слои (pooling layer)* и *полносвязные слои (fully-connected, слои из перцептронов)*.

Свёрточный слой нейронной сети представляет собой применение операции свёртки к выходам с предыдущего слоя, где веса ядра свёртки являются обучаемыми параметрами. Еще один обучаемый вес используется в качестве *константного сдвига*. При этом есть несколько важных деталей:

1) В одном свёрточном слое может быть несколько свёрток. В этом случае для каждой свёртки на выходе получится своё изображение. Например, если вход имел размерность $w \times h$, а в слое было n свёрток с ядром размерности $k_x \times k_y$, то выход будет иметь размерность $n \times (w - k_x + 1) \times (h - k_y + 1)$;

2) Ядра свёртки могут быть трёхмерными. Свёртка трехмерного входа с трехмерным ядром происходит аналогично, просто скалярное произведение считается еще и по всем слоям изображения. Например, для усреднения информации о цветах исходного изображения, на первом слое можно использовать свёртку размерности $3 \times w \times h$. На выходе такого слоя будет уже одно изображение (вместо трёх);

3) Можно заметить, что применение операции свёртки уменьшает изображение. Также пиксели, которые находятся на границе изображения, участвуют в меньшем количестве свёрток, чем внутренние. В связи с этим в свёрточных слоях используется *дополнение изображения (padding)*. Выходы с предыдущего слоя дополняются пикселями так, чтобы после свёртки сохранился размер изображения. Такие свёртки называют *одинаковыми (same convolution)*, а свёртки без дополнения изображения называются *правильными (valid convolution)*;

4) Еще одним параметром свёрточного слоя является *сдвиг (stride)*. Хотя обычно свёртка применяется подряд для каждого пикселя, иногда используется

сдвиг, отличный от единицы – скалярное произведение считается не со всеми возможными положениями ядра, а только с положениями, кратными некоторому сдвигу s . Тогда, если вход имел размерность $w \times h$, а ядро свёртки имело размерность $k_x \times k_y$ и использовался сдвиг s , то выход будет иметь размерность $\left\lfloor \frac{w-k_x}{s} + 1 \right\rfloor \times \left\lfloor \frac{h-k_y}{s} + 1 \right\rfloor$.

Пулинговый слой призван снижать размерность изображения. Исходное изображение делится на блоки размером $w \times h$ и для каждого блока вычисляется некоторая функция. Чаще всего используется функция *максимума* (*max pooling*) или *(взвешенного) среднего* (*(weighted) average pooling*). Обучаемых параметров у этого слоя нет.

В разделе 2.4 приведено теоретическое описание структуры МТСNN, используемой далее в практической части, рисунок 1.

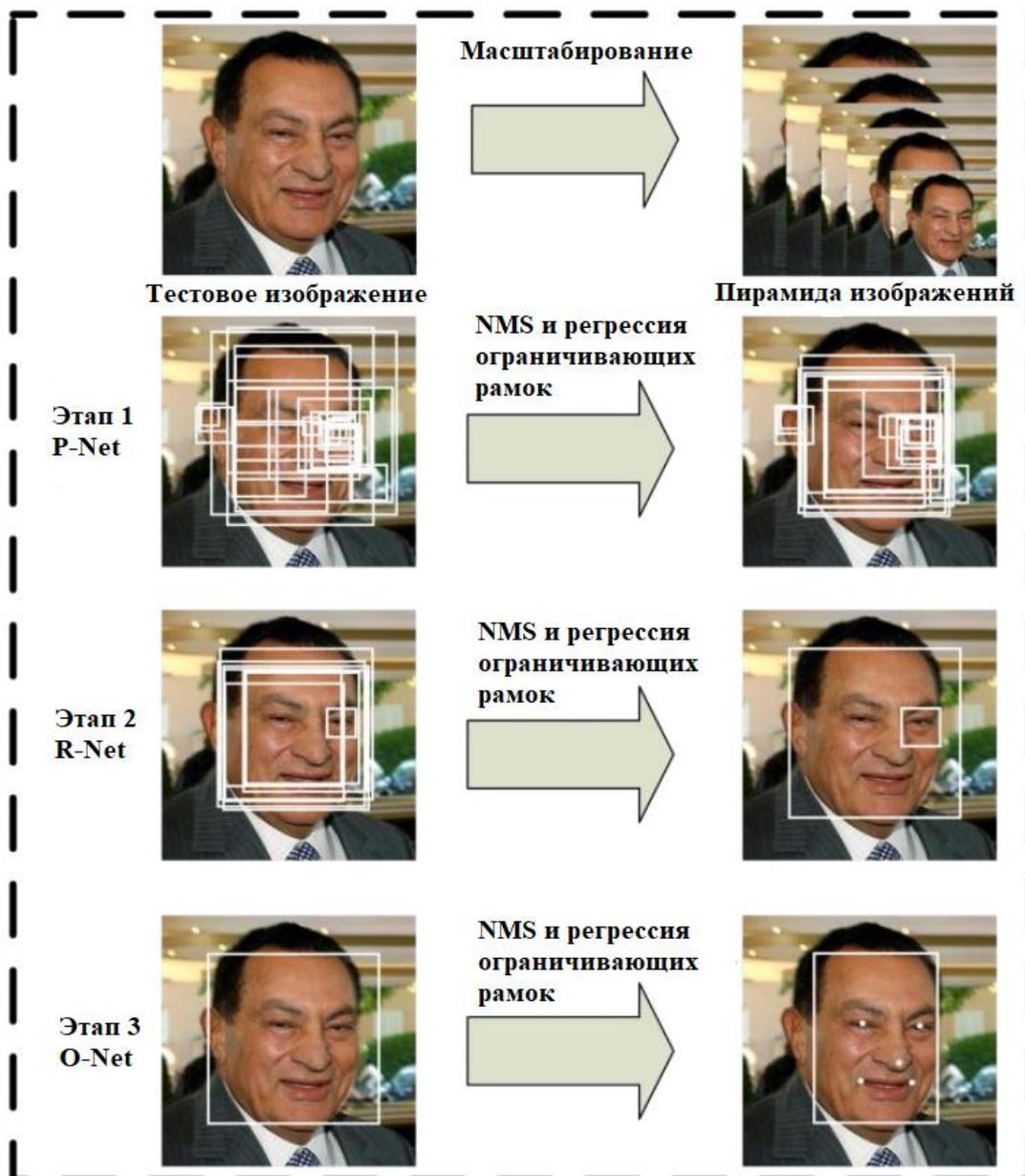


Рисунок 1 – Схема MTCNN, содержащая три этапа обработки многозадачными свёрточными сетями

Этап 1: Первая сеть, состоящая только из свёрточных/пулинговых слоёв, называемая *сетью гипотез (Proposal Net, P-Net)*, используется, чтобы получить области-кандидаты и их вектора регрессии ограничивающих прямоугольников (рамок). Затем для калибровки кандидатов используются оценочные векторы регрессии ограничивающих прямоугольников. После этого применяется «не максимальное подавление» (*non-maximum suppression, NMS*) для объединения сильно перекрывающихся кандидатов.

Этап 2: Выход первой сети подается в другую свёрточную сеть, называемую *уточняющей (Refine Network, R-Net)*, которая отклоняет большое

количество ложных кандидатов, выполняет калибровку с регрессией ограничивающих прямоугольников и слияние кандидатов с помощью NMS.

Этап 3: Этот этап аналогичен второму, но здесь *выходная сеть (Output Net, O-Net)* стремится более подробно описать лицо. В частности, она выводит позиции пяти ориентиров на лицах – левый и правый глаза, нос, левый и правый углы рта.

В следующем подразделе приведено математическое описание процесса обучения MTCNN.

Для обучения свёрточных сетей в MTCNN используются три задачи: задача классификации (есть на изображении лицо или нет), регрессия ограничивающих прямоугольников и локализация лицевых ориентиров.

1. Классификация лиц: Задача обучения формулируется как задача двухклассной классификации. Для каждого образца x_i в качестве функции потерь используется кросс-энтропия:

$$L_i^{det} = -(y_i^{det} \log(p_i) + (1 - y_i^{det})(1 - \log(p_i))),$$

где p_i – создаваемая сетью вероятность того, что образец является лицом. Обозначение $y_i^{det} \in \{0,1\}$ означает истинный ответ.

2. Регрессия ограничивающих прямоугольников: Для каждого окна-кандидата прогнозируется разница между ним и ближайшим истинным ответом (т. е. левый верхний угол, высоту и ширину ограничивающих прямоугольников). Цель обучения формулируется как задача регрессии, и в качестве функции потерь для каждого экземпляра x_i используется евклидова потеря:

$$L_i^{box} = \|\hat{y}_i^{det} - y_i^{box}\|_2^2,$$

где \hat{y}_i^{det} – цель регрессии, полученная из сети, и y_i^{box} – истинная координата. Так как координат – четыре, то $y_i^{box} \in \mathbb{R}^4$.

3. Локализация лицевых ориентиров: аналогично задаче регрессии ограничивающих прямоугольников, определение лицевых ориентиров

формулируется как задача регрессии, и для этого минимизируется евклидова потеря:

$$L_i^{landmark} = \|\hat{y}_i^{landmark} - y_i^{landmark}\|_2^2,$$

где $\hat{y}_i^{landmark}$ – координата лицевого ориентира, полученная из сети, и $y_i^{landmark}$ – истинная координата. Так как лицевых ориентиров – пять (левый и правый глаза, нос, левый и правый углы рта), то $y_i^{landmark} \in \mathbb{R}^{10}$.

В третьем разделе описана программная реализация полученной в ходе работы структуры, состоящей из двух частей:

1. Часть детектирования: написанный скрипт применяет ко всем файлам из указанной папки детектор MTCNN и сохраняет результаты его работы – изображения с вырезанными лицами – в результирующую папку;

2. Часть распознавания пола: подобранная модель нейронной сети обучается на результатах первой части структуры, либо выдает вероятности принадлежности человека на фотографии тому или другому полу, если сеть уже обучена.

Также в этом разделе приведена структура нейросети, использованной для определения пола человека по фотографии, а также скриншоты её работы на данных, полученных детектором.

ЗАКЛЮЧЕНИЕ

В данной работе были рассмотрены и изучены основы нейронных сетей, в том числе свёрточных. Улучшению практических навыков программирования поспособствовало изучение реализации детектора MTCNN и его комбинирование с моделью свёрточной нейронной сети для получения системы, научившейся распознавать пол человека на фотографии, с использованием API для глубокого обучения Keras на языке Python.

Таким образом, все поставленные задачи были полностью выполнены, цель работы можно считать достигнутой.