

Федеральное государственное автономное образовательное учреждение  
высшего образования «Национальный исследовательский университет  
«Высшая школа экономики»

Факультет компьютерных наук  
Основная образовательная программа  
Прикладная математика и информатика

## ГРУППОВАЯ КУРСОВАЯ РАБОТА

**ПРОГРАММНЫЙ ПРОЕКТ НА ТЕМУ**

**"КЛАССИФИКАЦИЯ ПОЛА И ВОЗРАСТА ЛЮДЕЙ НА ФОТОГРАФИИ"**

Выполнили студенты группы 171, 3 курса,  
Биршерт Алексей Дмитриевич,  
Шабалин Александр Михайлович

Руководитель КР: старший преподаватель  
Соколов Евгений Андреевич

Куратор: Магистр, руководитель группы нейросетевых технологий  
компьютерного зрения  
Овчаренко Сергей Александрович

Москва 2020

# Содержание

<b>1</b>	<b>Введение</b>	<b>4</b>
<b>2</b>	<b>Обзор литературы</b>	<b>6</b>
2.1	Детектирование лиц . . . . .	6
2.2	Классификация возраста и пола . . . . .	7
<b>3</b>	<b>Детектирование лиц</b>	<b>9</b>
3.1	Описание выбранного метода . . . . .	9
3.2	Подготовка данных и обучение . . . . .	10
3.3	Выравнивание . . . . .	11
3.4	Эксперименты . . . . .	11
3.5	Результаты . . . . .	12
3.6	Примеры работы модели . . . . .	13
<b>4</b>	<b>Классификация гендерных и возрастных групп</b>	<b>14</b>
4.1	Описание метода . . . . .	14
4.2	Описание данных . . . . .	15
4.3	Постановка задач обучения . . . . .	16
4.4	Эксперименты . . . . .	16
4.5	Результаты . . . . .	18
<b>5</b>	<b>Выводы</b>	<b>19</b>
<b>6</b>	<b>Планы дальнейшей работы</b>	<b>20</b>
<b>7</b>	<b>Список литературы</b>	<b>21</b>

## Аннотация

Автоматическое предсказание пола и возраста человека по фотографиям, полученным в самых разных условиях - это важная и сложная задача, находящая применение во многих областях жизнедеятельности людей. В своем проекте мы стремились воплотить подход к решению этой задачи, основанный на сверточных нейронных сетях. Свое решение мы разбили на две составных части - детектирование лица человека и ключевых точек на его лице и дальнейшее предсказание пола и возраста по признакам лица.

Для детектирования мы используем архитектуру RetinaFace с предобученным на ImageNet-1000 ResNet-18 в качестве основной модели. В качестве обучающих данных мы используем датасет WIDER FACE с добавленными к нему пятью ключевыми точками для каждого лица. На тестовой выборке наш детектор получает значение метрики Precision равное 83%.

Для классификации найденных лиц мы используем модель на основе двух предобученных на ImageNet-1000 ResNet-18. В качестве обучающих данных мы используем датасеты IMDB-WIKI-101 и FGNET. В качестве тестовых данных мы используем датасет Adience, наша модель получает для возраста значение метрики Exact-accuracy 45% и значение метрики One-off-accuracy 80%, для пола значение метрики Accuracy 85%.

В итоге мы получили рабочую модель, способную обрабатывать большие массивы фотографий, детектировать на них людей и предсказывать для их пол и возраст в автоматическом режиме. Ссылка на гитхаб с проектом - [https://github.com/birshert/age\\_gender\\_classification](https://github.com/birshert/age_gender_classification).

**Ключевые слова**—Определение возраста и пола, Детектирование лиц, Компьютерное зрение, Глубокое обучение

Automatically predicting real age and gender from face images acquired in unconstrained conditions is an important and challenging task in many real-world applications. In our project we intended to construct an approach to predicting a person's real age and gender from photograph based on convolutional neural networks. Our solution is divided into solving two separate subtasks - detecting one's face and it's landmarks and further age and gender estimation based on facial features.

For face detection we use RetinaFace architecture with a pretrained on ImageNet-1000 ResNet-18 as a backbone. For training we use WIDER FACE dataset with added five facial landmarks for every face. On testing subset we achieve 83% Precision result.

For real age and gender estimation we use a model based on two pretrained on ImageNet-1000 ResNet-18. For training we use IMDB-WIKI-101 dataset and FGNET dataset. For testing we use Adience dataset, our model achieves Exact-accuracy 45%, One-off-accuracy 80% for age estimation and 85% Accuracy for gender prediction.

As a result, we got a working model capable of processing large volumes of photos, detecting people faces on them and predicting their real age and gender in automatic mode. Github project link - [https://github.com/birshert/age\\_gender\\_classification](https://github.com/birshert/age_gender_classification).

**Keywords**—Age and gender estimation, Facial detection, Computer vision, Deep learning

# 1 Введение

Во многих сферах деятельности требуется получать информацию о статистике про людей, проживающих на некоторой территории или пользующихся каким-либо продуктом. Хорошим источником такой информации являются фотографии людей, запечатлённых в определенных местах. Однако ручная обработка десятков тысяч фотографий в попытке получить нужные данные крайне сложна и время- и трудозатратна, поэтому большинство современных решений являют собой автоматическую обработку массивов фотографий, классифицируя людей на фотографиях по полу и возрасту с помощью методов компьютерного зрения.

За время исследования проблемы классификации людей по фотографиям стало ясно, что использование алгоритмов, опирающихся на фиксированные выделенные людьми признаки, не приносит удовлетворительных результатов. Так происходит потому, что люди на фотографиях, запечатленные в неформальных условиях, имеют множество разных поз, форм и вариаций поворотов относительно камеры, которые не поддаются ручному анализу.

Все известные подходы к классификации возрастных групп людей по фотографии заключаются в анализе изображения лица. Самые ранние [1] основывались на различиях в пропорциях и размерах черт лица в зависимости от возраста - так называемые антропометрические модели [2]. Все они вычисляли координаты точек на лице и в дальнейшем их анализировали. Более поздние и более современные методы [3, 4] опирались на использование сверточных нейронных сетей небольшой глубины. Самые современные подходы используют глубокие сверточные нейронные сети [24].

Первые подходы к классификации пола использовали фотографии лица низкого разрешения - до 20 на 20 пикселей, и обучали на них различные типы классификаторов [5]. Позднее стали использовать LBP для выявления новых признаков [6], использовать сверточные нейронные сети [3, 4].

В ходе выполнения работы мы не будем предлагать каких-либо новых ме-

тодов, однако мы воспользуемся рядом различных уже существующих технологий машинного обучения, проверив их качество при применении к нашей задаче. В итоге мы получим алгоритм, способный перебирать большие объемы фотографий, находить на них лица людей, а затем классифицировать их по полу и возрасту.

Дальнейшая работа описана в следующих главах - Обзор литературы, Детектирование лиц, Классификация гендерных и возрастных групп, Выводы и Планы дальнейшей работы. "Детектирование лиц" выполнено Александром Шабалиным, "Классификация гендерных и возрастных групп" Алексеем Биршертом.

## 2 Обзор литературы

### 2.1 Детектирование лиц

Детектирование лиц на фотографии - одна из древнейших задач компьютерного зрения, возникшая в 1990-х годах. Первые хорошие результаты появились в 2004 году. Описанный в статье [7] метод находил лица с помощью признаков Хаара, используя каскад детекторов, обученных алгоритмом AdaBoost. В 2014 году в статье [8] был предложен метод, использующий Deformable Parts AgeGender (DPM). Его идея заключается в нахождении зависимостей между подвижными частями. Например, лицо представляется, как нечто, состоящее из глаз, носа, рта, расположенных в некотором антропоморфическом виде. Однако все описанные методы показывали довольно плохие результаты в сложных случаях, так как опирались на ограниченный, придуманный людьми набор признаков. Поэтому методы, основанные на сверточных нейронных сетях быстро вытеснили остальные. Лучшие известные на данный момент подходы описаны в статье [9]. Все из них используют нейронные сети (обычно ResNet [10]) для получения признаков. Так как модель сама находит признаки и зависимости, результат получается лучше.

Немаловажной задачей является и выравнивание лиц. Самый быстрый метод получить выровненное лицо - определить ключевые точки на лице и преобразовать изображение так, чтобы эти точки были на заранее определенных местах. В статье [12] описан метод вычисления координат основных точек на лице человека - окаймляющих лицо, глаза, нос, рот и брови. Вычисление точек происходит с помощью каскада регрессоров, обучаемых с помощью градиентного бустинга.

В нашей работе мы пользуемся архитектурой RetinaFace [11]. На данный момент она является state-of-the-art в задаче детектирования лиц и показывает впечатляющие результаты. Одна из ее отличительных черт - умение предсказывать пять ключевых точек, необходимых для выравнивания изображения.

## 2.2 Классификация возраста и пола

Задачи определения пола и возраста человека находят применение в разных сферах жизни человека, в наружном наблюдении, в антропологии, в биометрической идентификации. Современные подходы к этой задаче опираются на сверточные нейронные сети. Так, например, в статье [3] описано решение с помощью сверточной сети небольшой глубины. Для классификации возраста и пола используется одна и та же архитектура. Нейронная сеть состоит из трёх свёрточных слоёв и двух полносвязных, небольшой размер сети объясняется желанием быть физичным в распознавании лиц и нежеланием переобучиться. Точность по классификации пола была  $86.8 \pm 1.4\%$ , возрастных групп -  $50.7 \pm 5.1\%$  для точного попадания в группу и  $84.7 \pm 2.2\%$  для попадания в правильную или соседнюю для датасета Adience.

В статье [4] описан алгоритм анализа лица с помощью пяти сверточных нейросетей, получающих изображения лица целиком, левого и правого глаза, носа и рта соответственно. Итоговое решение принимается на основе выходов всех пяти нейросетей. Нейросеть, получающая на вход всё изображение лица, имеет три сверточных слоя, прочие по два. Точность по классификации пола достигла  $89.6 \pm 1.3\%$ , возрастных групп -  $54.3 \pm 3.5\%$  для точного попадания в группу и  $87.6 \pm 1.9\%$  для попадания в правильную или соседнюю на датасете Adience, что является значительным улучшением результата предыдущей статьи.

В статье ?? описан метод использования ансамбля глубоких сверточных нейронных сетей для предсказания реального возраста. Для этого авторы статьи собрали свой собственный датасет IMDB-WIKI-101, дообучали на нём модели с архитектурой VGG-16 предобученные на ImageNet-1000, потом классифицировали признаки, полученные сверточной нейронной сетью с помощью двух полносвязных слоев и использовали выход второго слоя для предсказания возраста от 0 до 100 включительно. В процессе они использовали различные улучшения в выравнивании лиц (вращали изображение вдоль центра и



выбирали детектированное лицо с максимальным показателем уверенности модели детектирования лиц), затем тренировали ансамбль из 20 моделей на датасете LAP ?? для получения оптимального результата. В итоге им удалось получить результат MAE по возрасту 3.2 на датасете IMDB-WIKI-101.

## 3 Детектирование лиц

Эта часть сделана Александром Шабалиным

### 3.1 Описание выбранного метода

Для детектирования лиц была использована архитектура нейронной сети RetinaFace. Она состоит из двух частей: основной модели и пирамиды признаков. Основная модель - это сверточная нейронная сеть, предназначенная для выявления признаков из изображения. В качестве основной модели может выступать MobileNet, VGG, ResNet и другие.

Идея пирамиды признаков довольно проста. При детектировании лиц мы хотим находить лица разных размеров. Для этого строятся две пирамиды из пяти карт признаков в каждой. Первая получается проходом снизу в верх, а вторая - сверху вниз. Слоями первой пирамиды являются выходы четырех слоев основной модели такие, что размер каждого следующего выхода в два раза меньше размера предыдущего. Так мы получаем карты разного размера, что позволяет детектировать как маленькие, так и большие лица. Вторая пирамида необходима из-за того, что на ранних слоях сверточной сети содержится значительно меньше семантической информации, а значит, предсказания лиц на ранних слоях менее точны и мы должны как-то компенсировать это. Слои второй пирамиды мы получаем, проходя сверху вниз следующим образом. Первый (верхний) слой получается с помощью наложения свертки с ядром размера  $1 \times 1$  на верхний слой первой пирамиды. Каждый следующий из пяти слоев получается путем поэлементного суммирования предыдущего слоя, увеличенного в два раза, со сверткой размера  $1 \times 1$  соответствующего слоя первой пирамиды. Таким образом нам удастся передать большим по размеру картам признаков семантическую информацию меньших карт, компенсировав ее нехватку. Предсказания получаются с помощью применения нескольких сверточных слоев к картам признаков второй пирамиды и конкатенированием результатов этих сверток.

В своей реализации в качестве основной модели используется ResNet18 предобученный на ImageNet. Выходы четырех блоков ResNet являются четырьмя картами признаков первой пирамиды, пятая карта признаков получается наложением свертки с ядром размера  $3 \times 3$  и шагом 2 на последнюю карту признаков. Так как датасет ImageNet содержит в себе картинки с различными изображениями, а в нашей задаче необходимо находить лица людей, мы замораживаем только первый слой ResNet, а остальные дообучаем. Выбор ResNet с 18 слоями обосновывается тем, что в нашей задаче очень важна скорость работы. С увеличением размера сверточной сети скорость ее работы заметно падает, а качество увеличивается не так сильно.

## 3.2 Подготовка данных и обучение

Для обучения модели был использован датасет WIDER FACE. Он содержит 32,203 фотографии с 393,703 лицами на них. Для каждого лица хранятся координаты пяти ключевых точек: две для глаз, одна для носа и две для рта, а также координаты прямоугольной рамки, ограничивающей лицо. В качестве аугментации мы пробовали делать случайный переворот изображения, однако он не улучшил результаты, поэтому мы решили от него отказаться. Входные изображения масштабируются до квадратных следующим образом: большая сторона становится равной 256, а к меньшей добавляется нулевой отступ с двух сторон так, чтобы размер итогового изображения составлял 256 на 256. При обучении был использован Adam оптимизатор со скоростью обучения  $10^{-3}$ . Модель обучается в течение 20-ти эпох с размером батча 32.

Ошибка модели считается по формуле:

$$L = L_{cls}(p, p^*) + 0.25L_{box}(t, t^*) + 0.1L_{pts}(l, l^*),$$

где  $p$  - вероятность лица,  $p^*$  - истинный ответ (0 или 1), для подсчета ошибки классификатора  $L_{cls}(p, p^*)$  используется софтмакс ошибка для бинарных

классов классов (лицо/не лицо).

$t = \{t_1, t_2, t_3, t_4\}$  и  $t^* = \{t_1^*, t_2^*, t_3^*, t_4^*\}$  предсказанные и истинные координаты ограничивающей рамки соответственно. Для подсчета ошибки  $L_{box}(t, t^*)$  используется функция Smooth-L<sub>1</sub> loss.

$l = \{l_{x1}, l_{y1}, \dots, l_{x5}, l_{y5}\}$  и  $l^* = \{l_{x1}^*, l_{y1}^*, \dots, l_{x5}^*, l_{y5}^*\}$  предсказанные и истинные координаты ключевых точек лица. Для подсчета ошибки  $L_{pts}(l, l^*)$  так же используется Smooth-L<sub>1</sub> loss.

### 3.3 Выравнивание

Перед передачей найденных лиц классификатору возраста и пола необходимо их выровнить. Выравнивание производится на основе двух ключевых точек для глаз, найденных вместе с лицами. Для этого находится угол отклонения прямой, проведенной с помощью точек глаз, от горизонтали. После этого мы домножаем матрицу изображения на матрицу поворота таким образом, что линия глаз становится горизонтальной. Мы не используем остальные точки, так как они избыточны, а в некоторых случаях даже мешают. Например, расположение точек рта у улыбающегося человека и у серьезного отличаются, но это отличие не должно никак влиять на выравнивание. Глаза же всегда располагаются на одном месте относительно лица, что позволяет их считать хорошей опорой для выравнивания.

### 3.4 Эксперименты

Для определения с выбором основной модели были проведены сравнения между ResNet18, ResNet50, VGG и MobileNet. По результатам этого эксперимента выяснилось, что ResNet50 получает лучший результат, но для предсказания ответа ему требуется гораздо больше времени, чем ResNet18. MobileNet работает быстрее остальных, но ошибка такой модели оказывается больше остальных. VGG показывает хороший результат, но работает медленнее всех. В результате выбор остановился на ResNet18, потому что баланс скорости и

качества детектирования лиц этой модели мы считаем оптимальным. Графики для этого эксперимента появятся чуть позже, потому что они не успели посчитаться :(

Для ускорения времени работы алгоритма мы пробовали уменьшать число карт признаков в пирамиде признаков, однако в таком случае некоторые слишком большие или слишком маленькие лица переставали обнаруживаться, что крайне негативно отражалось на результатах модели.

### 3.5 Результаты

Полученная модель хорошо справляется с поставленной задачей, получая значение метрики Precision равное 83%. Для нашей задачи эта метрика имеет крайне важное значение, так как ты не должны классифицировать несуществующих людей. Несмотря на то, что реализация авторов RetinaFace получала 91%, мы считаем это хорошим результатом, так как из-за нехватки вычислительных мощностей пришлось почти в 3 раза снизить размер стороны входного изображения, использовать ResNet18 вместо ResNet152, а также уменьшить количество эпох обучения, что заметно сказывается на результатах работы.

### 3.6 Примеры работы модели



(a) Модель смогла найти 50 лиц из 61 указанного в ответе.



(b) Модель хорошо справилась со своей задачей, найдя все повернутые к камере лица, однако она также выделила пустую часть стола.

## 4 Классификация гендерных и возрастных групп

Эта часть выполнена Алексеем Биршертом.

### 4.1 Описание метода

Для предсказания пола и возраста используются две модели, состоящих из основной и выходной моделей каждая. В качестве основной модели используется глубокая нейронная сеть ResNet-18, без последнего полносвязного слоя. В качестве выходной модели используется персептрон из двух полносвязных слоёв с нелинейностью ReLU и дропаутом между ними. Первая модель предназначена для классификации пола и имеет в выходной модели 512 и 256 входных и выходных нейронов в первом слое, 256 и 2 во втором соответственно. Вторая модель предназначена для классификации возраста и имеет в выходной модели 512 и 512 входных и выходных нейронов в первом слое, 512 и 101 во втором соответственно. На вход подаются фотографии лиц людей, выделенные и выровненные с помощью модели детектирования лиц, размером 227 на 227 пикселей, 3 канала цвета - R, G, B. Предсказанный пол определяется с помощью определения выходного нейрона соответствующей нейронной сети с максимальным значением - первый это "женский" второй "мужской". Предсказанный возраст определяется следующим образом: сначала для вектора значений выходных нейронов соответствующей нейронной сети применяется преобразование софтмакс, затем значения умножаются на соответствующий им возраст. После вектор суммируется - получаем математическое ожидание возраста при вероятностном распределении, выданном моделью.

$$AGE = \sum_{i=0}^{100} i \cdot softmax(x)_i, \quad softmax(x)_i = \frac{\exp(x_i)}{\sum_{j=0}^{100} \exp(x_j)}.$$

## 4.2 Описание данных

В качестве датасета для обучения двух вышеописанных моделей были избраны датасеты IMDB-WIKI-101 и FGNET. Распределение реального возраста в датасете IMDB-WIKI-101 имеет вид нормальной кривой со средним около 35 лет, имея малое количество объектов с возрастом меньше 10 лет или больше 90. Для восполнения данных по возрасту до 10 лет был избран датасет FGNET, в котором большая часть объектов это дети до 15 лет. Для улучшения сходимости нейронных сетей была произведена предобработка всех объектов - в итоговую выборку не были включены следующие объекты: объекты с плохо различимыми лицами (показатель уверенности модели распознавания лиц в том, что это лицо, ниже фиксированного значения), объекты с некорректно заполненными данными по полу/возрасту, объекты со слишком маленькими фотографиями. Для каждого объекта были выделены мета-данные про пол и возраст, а так же детектировано и выделено лицо. Отступ от границы лица был поставлен 40%, чтобы можно было полностью получить лицо внутри квадрата. Итого было получено около 200 тысяч объектов, которые были в дальнейшем поделены с сохранением баланса классов 1 к 19 на валидационную и обучающую выборки соответственно. В качестве датасета для тестирования был избран датасет Adience, по которому известно большое количество результатов различных моделей. Из него были исключены объекты с некорректным описанием пола или возраста. Итого было получено почти 11 тысяч объектов для тестовой выборки. В Adience метки возраста в формате 8 групп - 0: [0, 2], 1: [4, 6], 2: [8, 12], 3: [15, 20], 4: [25, 32], 5: [38, 43], 6: [48, 53], 7: [60, 100]. В связи с этим, необходимо было решить как относить к этим группам метки реального возраста от 0 до 100. Было принято решение относить к ближайшей группе - например, 22 года ближе к 20, чем к 25, следовательно относится к группе 3: [15, 20]. В случае одинакового расстояния выбиралась первая по порядку группа.



### 4.3 Постановка задач обучения

Задача классификации пола является задачей бинарной классификации, целевая переменная для одного объекта это число 0 или 1. Для обучения была выбрана перекрестная энтропия - функция ошибки со следующей формулой:  $loss(x, y) = -x_y + \log \left( \sum_{j=1}^K \exp(x_j) \right)$ , где  $x$  - вектор значений выходных нейронов нейронной сети,  $y$  - целевая переменная. Для оценки качества классификации использовались метрика доля правильных ответов (далее ассигасу), так как выборки сбалансированны по полу.

Задача классификации возраста является задачей многоклассовой классификации. Если смотреть на задачу предсказания возраста по фотографии с точки зрения человека, человек гораздо точнее способен угадать диапазон возраста, нежели точный возраст. Поэтому задача классификации возраста была интерпретирована как задача с множественными правильными ответами - каждому объекту может соответствовать набор правильных классов. Целевой переменной для одного объекта служил вектор из нулей и единиц, единицы на позициях правильных классов. Правильные классы определялись как значения возраста, которые отличаются по модулю от правильного не больше, чем на фиксированное число, которое было гиперпараметром (далее об этом в разделе эксперименты (4.4)). Для обучения была выбрана бинарная перекрестная энтропия -  $loss(x, y) = \text{sum}(L)$ ,  $L = \{l_1, \dots, l_N\}$ ,  $l_i = (y_i \log(x_i) + (1 - y_i) \log(1 - x_i))$ , где  $x$  это вектор значений выходных нейронов нейронной сети после применения сигмоидного преобразования,  $y$  - целевая переменная. Для оценки качества классификации использовались средний модуль отклонения (далее MAE) и доля объектов, у которых MAE не превышает фиксированного числа (далее CS-5).

### 4.4 Эксперименты

Первым необходимо было решить вопрос архитектуры базовой модели, было принято решение остановиться на ResNet-18. Всего было опробовано

четыре различных архитектуры - MobileNet, ShuffleNet, ResNet и нейронная сеть с тремя сверточными слоями с нормализацией [25] между слоями и активацией ReLU и последующими двумя полносвязными слоями. Лучшее всего себя проявила архитектура ResNet-18. MobileNet и ShuffleNet оказались не сильно быстрее, не смотря на реализацию, однако достаточно сильно проигрывали в качестве. Простая нейронная сеть из трёх слоёв оказалась медленнее ResNet в силу больших размеров изображений и огромного размера полносвязного слоя. Из-за этого она очень плохо и медленно сходилась и показывала самые плохие результаты. Графики сравнения появятся чуть позже.

Так как ResNet-18 предобучен на датасете ImageNet-1000, он имеет очень хорошую способность выделять признаки из изображения. Известно, что первые (входные) сверточные слои очень похожи между собой даже для различных задач. Таким образом, было решено первый сверточный слой ResNet-18 заморозить в процессе обучения, дообучать все остальные и полносвязные в конце.

Вторым необходимо было решить вопрос количества моделей - одна модель с общей базовой моделью и двумя параллельными выходными моделями или две отдельные модели из основной и выходной модели. В итоге лучше себя показала конструкция из двух отдельных моделей, в силу использования функций ошибки, которые имеют разный масштаб и решают частично разные задачи на моменте основной модели, и сложности в подборе гиперпараметров модель из одной общей основной модели очень плохо сходилась к локальным минимумам. Графики сравнения появятся чуть позже.

В процессе подбора гиперпараметров были выбраны следующие значения: темп обучения был выставлен на  $1e - 3$  для первых 20 эпох, далее  $1e - 4$  для модели предсказания возраста,  $1e - 4$  для первых 30 эпох и  $2e - 5$  далее для модели предсказания пола; коэффициент регуляризации был установлен на  $1e - 3$  для обеих моделей; размер батча изображений 128; размер входных картинок 227 на 227 пикселей; дропаут 0.1 перед 4 слоём ResNet, 0.2 перед

первым полносвязным слоём и 0.4 между слоями. В качестве размера окна для правильного возраста после сравнений было выбрано число в 5 лет.

В процессе выбора аугментации были выбраны следующие трансформации - для обучения каждое изображение преобразовывалось к квадрату 256 на 256 пикселей, затем из него выбирался случайный квадрат со стороной 227 пикселей, который с вероятностью 0.5 мог быть отзеркален вдоль вертикальной оси, проходящей через его центр. Для тестирования каждое изображение преобразовывалось к квадрату 256 на 256 пикселей, затем из него вырезался квадрат со стороной 227 пикселей из центра.

## 4.5 Результаты

В итоге получилось достичь следующих показателей метрик на обучающей и валидационной выборках: для возраста MAE 4 и 4.3 соответственно, CS-5 0.8 и 0.76 соответственно, для пола Accuracy 0.95 и 0.94 соответственно.

Модель показала следующие результаты на датасете Adience: для возраста Exact-accuracy 45% и One-off-accuracy 80%, для пола значение метрики Accuracy 85%.

В этой подглаве планируется показать примеры удачных и неудачных распознаваний и сравнить результаты с тремя статьями. Уже можно сказать, что одну результат одной статьи удалось превзойти, к результатам двух других приблизиться снизу.

Всего было написано более 2300 строк кода на языке программирования Python3, все модели были написаны и обучены с использованием фреймворка машинного обучения PyTorch.

## 5 Выводы

Решая поставленную задачу о классификации людей на фотографиях по полу и возрасту мы получили алгоритм, позволяющий быстро - у нас была цель суметь обрабатывать миллион фотографий за день на одной видеокарте NVIDIA RTX 2080 Ti, и мы ее выполнили - и качественно находить лица людей на фотографиях, а затем классифицировать их пол и возраст. При решении мы опробовали множество различных подходов и остановились на архитектуре RetinaFace для детектирования лиц и двух ResNet для классификации пола и возраста соответственно. К сожалению, мы были ограничены вычислительными и временными ресурсами, поэтому не сумели добиться действительно высокого качества.

## 6 Планы дальнейшей работы

Для улучшения качества классификации в дальнейшем у нас есть несколько идей, которые мы не успели реализовать. В своем решении для классификации пола и возраста мы используем только лицо. Безусловно, лицо содержит большую часть необходимых для классификации признаков, однако используя для классификации остальные части тела, можно добиться лучших результатов. Также мы не умеем детектировать и классифицировать людей со спины. Это довольно сложная задача, но ее решение внесло бы ощутимый вклад в работу нашего алгоритма. Кроме этого, в конечном варианте решения мы не отсекаем лица неживых людей, найденные детектирующей моделью. Мы пробовали отличать живые лица от фотографий и лиц на рекламных щитах с помощью LBP [13], но такой подход не дал ожидаемых результатов и мы отложили эту задачу из-за нехватки ресурсов на ее выполнение. Для классификации найденных лиц на памятники и живых людей мы пробовали два подхода: в первом мы пытались искать отличия на основе цветовой палитры, а во втором использовали сверточные нейронные сети. Первый метод не показал себя достаточно хорошо, второй не удалось довести до конца из-за нехватки данных для обучения и временных ресурсов.

## 7 Список литературы

### Список литературы

1. Age Classification from Facial Images. Young H. Kwon, Niels da Vitoria Lobo. In IEEE 1994
2. Age and Gender Estimation of Unfiltered Faces. Eran Eidinger, Roeen Enbar, Tal Hassner. In IEEE 2014
3. Age and Gender Classification using Convolutional Neural Networks. Gil Levi, Tal Hassner. In IEEE 2015
4. Age/gender classification with whole-component convolutional neural networks (WC-CNN). Chun-Ting Huang, Yueru Chen, Ruiyuan Lin, C.-C. Jay Kuo. In IEEE 2018
5. Learning Gender with Support Faces. Baback Moghaddam, Ming-Hsuan Yang. In IEEE 2002
6. Demographic Classification with Local Binary Patterns. Zhiguang Yang, Haizhou Ai. In ICB 2007
7. Robust Real-Time Face Detection. Paul Viola, Michael J. Jones. In IEEE 2003
8. Face detection without bells and whistles. Makrus Mathias, Rodrigo Beneson, Marco Pedersoli, Lus Van Gool. In ECCV 2014
9. Accurate Face Detection for High Performance. Faen Zhang, Xinyu Fan, Guo Ai, Jianfei Song, Yongqiang Qin, Jiahong Wu. In ArXiv 2019
10. Deep Residual Learning for Image Recognition. Kaiming He, Xiangyu Zhang, Shaoqing Ren, Jian Sun. In IEEE 2015

11. RetinaFace: Single-stage Dense Face Localisation in the Wild. Jiankang Deng, Jia Guo, Yuxiang Zhou, Jinke Yu, Irene Kotsia, Stefanos Zafeiriou. In ArXiv 2019
12. One Millisecond Face Alignment with an Ensemble of Regression Trees. Vahid Kazemi and Josephine Sullivan. In IEEE 2014
13. Face Anti-Spoofing Based on Color Texture Analysis. Zinelabidine Boulkenafet, Jukka Komulainen, Abdenour Hadid. In IEEE 2015
14. Learning Deep Models for Face Anti-Spoofing: Binary or Auxiliary Supervision. Yaojie Liu, Amin Jourabloo, Xiaoming Liu. In IEEE/CVF 2018
15. Deep Tree Learning for Zero-shot Face Anti-Spoofing. Yaojie Liu, Joel Stehouwer, Amin Jourabloo, Xiaoming Liu. In IEEE/CVF 2019
16. Primate Face Identification in the Wild. Ankita Shukla, Gullal Singh Cheema, Saket Anand, Qamar Qureshi, Yadvendradev Jhala. In PRICAI 2019
17. Deep Learning Face Representation by Joint Identification-Verification. Yi Sun, Xiaogang Wang, Xiaoou Tang. In NIPS 2014
18. A Discriminative Feature Learning Approach for Deep Face Recognition. Yandong Wen, Kaipeng Zhang, Zhifeng Li, and Yu Qiao. In ECCV 2016
19. Densely Connected Convolutional Networks. Gao Huang, Zhuang Liu, Laurens van der Maaten, Kilian Q. Weinberger. In IEEE 2016
20. NUAA Photograph Imposter Database
21. Labeled Faces in the Wild Database
22. IMDB-WIKI Database
23. 300 Faces In-the-Wild Database

24. Age Group and Gender Estimation in the Wild with Deep RoR Architecture. Ke Zhang, Ce Gao, Liru Guo, Miao Sun, Xingfang Yuan, Tony X. Han, Zhenbing Zhao and Baogang Li. In IEEE 2017
25. ImageNet Classification with Deep Convolutional Neural Networks. Alex Krizhevsky, Ilya Sutskever, Geoffrey E. Hinton. In Advances in neural information processing systems 2012
26. DEX: Deep EXpectation of apparent age from a single image. Rasmus Rothe, Radu Timofte, Luc Van Gool. In IJCV 2016
27. Sergio Escalera, Junior Fabian, Pablo Pardo, Xavier Baro, Jordi Gonzalez, Hugo J. Escalante, Marc Oliu, Dusan Misevic, Ulrich Steiner, Isabelle Guyon. In EFP 2015.