

Санкт-Петербургский государственный университет

КНЯЗЕВ Никита Андреевич

Выпускная квалификационная работа

*Применение нейронных сетей для распознавания пола и
возраста человека на основе фотографий*

уровень образования: бакалавриат

Направление: 02.03.02

«Фундаментальная информатика и информационные технологии»

Основная образовательная программа СВ.5003.2017

«Программирование и информационные технологии»

Профиль *«Автоматизация научных исследований»*

Научный руководитель:

доцент кафедры теории систем

управления электрофизической аппаратурой,

к. ф.-м. н.

Козынченко Владимир Александрович

Рецензент: профессор

кафедры теории управления,

д. ф.-м. н.

Котина Елена Дмитриевна

Санкт-Петербург

2021

Содержание

Введение	4
Обзор литературы	7
Постановка задачи	9
Глава 1. Нейронные сети	10
1.1. Некоторые теоретические выкладки о нейронных сетях .	10
1.2. Обучение нейронных сетей	12
1.3. VGG-Face	15
Глава 2. Построение и обучение нейронных сетей	16
2.1. Математическая постановка задачи	16
2.2. Выбор базы фотографий	17
2.3. Предобработка данных	18
2.4. Подготовка рабочей среды	19
2.5. Определение пола	19
2.6. Предсказание пола и возраста одним классификатором . .	20
2.7. Определение возраста регрессией без разделения по гендерному признаку	22
2.8. Определение возраста классификацией без разделения по гендерному признаку	23
2.9. Определение возраста классификацией с разделением по гендерному признаку	24
2.10. Определение возраста регрессией с разделением по гендерному признаку	27
Глава 3. Анализ полученных результатов	28
3.1. Результаты вычислительных экспериментов	28
3.2. Сравнение моделей	29
Выводы	30
Заключение	31
Список использованных источников	32
Приложение А. Существующие решения задачи	35

Приложение Б. Архитектура сети VGG-Face	38
Приложение С. Базы фотографий лиц людей	40

Введение

Возраст и пол человека играют важнейшую роль в социальной жизни индивида и в его взаимодействии с другими людьми. Например, в зависимости от возрастной группы человека во многих языках меняется форма обращения к нему.

Определение пола и возраста является популярной задачей машинного обучения, и в настоящий момент существуют различные подходы к ее решению [7, 12]. Решение данной задачи имеет множество практических приложений. К примеру, появляется возможность быстрого принятия решения о допуске лиц на мероприятие, имеющее возрастной ценз, или на продажу им некоторых товаров. В области маркетинга и анализа статистики сегментирование посетителей по социально-демографическому признаку позволяет лучше понять состояние рынка и предпочтения пользователей.

Рассмотрим подробнее некоторые возможные методы решения данной задачи:

1. Методы, основанные на активной модели внешнего вида (ААА) – группа статистических методов, оценивающая визуальные возрастные особенности. Учитываются как главные особенности, такие как глаза, рот, нос и подбородок, так и вторичные, например, морщины. На основе различий этих особенностей у лиц разных возрастных групп строится модель внешнего вида.
2. Mean-Variance Loss [17] – метод, который рассматривает целевую переменную как распределение вероятностей по возрастным и гендерным классам. Функция потерь состоит из комбинации перекрестной энтропии, квадрата отклонения среднего значения распределения от целевой переменной и стандартного отклонения полученного распределения вероятности. В результате подход позволяет уменьшить разброс предсказания и повысить среднюю точность.
3. SSR-Net [21] – компактная сеть с мягкой поэтапной регрессией для оценки возраста. Иерархическая структура позволяет использовать простые и быстрые нейронные сети.

4. Кроме нейронных сетей, для определения пола и возраста применяются и другие методы машинного обучения, например Random Forest [19]. В задаче классификации метод определяет целевой класс как наиболее распространенный предсказанный класс среди всех деревьев. Что касается регрессии, то целевая переменная определяется как среднее значение среди всех деревьев.
5. Модель типа CaffeModel – данное решение представлено в статье «Понимание и сравнение глубоких нейронных сетей для классификации по возрасту и полу». Модель решает задачу классификации. Возрастной диапазон поделен на определенные промежутки, каждый из которых является отдельным классом.

Перечисление работ, содержащих описание решения данной задачи приведено в Приложении А. На данный момент нет ответа на вопрос, какая методология прогнозирования пола и возраста является лучшей. В этом исследовании мы будем работать именно с нейросетевыми моделями.

Распознавание пола человека по фотографии относится к задаче бинарной классификации. Предсказание возраста – это регрессионная задача, но мы решим ее и как задачу классификации, путем группировки по различным возрастным диапазонам. В качестве основы для решения поставленных задач используем предобученную сверточную нейронную сеть.

Областью исследования являются нейронные сети. Предмет исследования – изображения лиц людей.

В данной работе предстоит решить такие проблемы как: выбор наилучшей предобученной сети, подходящей для распознавания пола и возраста; поиск базы фотографий; разработка и обучение нейронных сетей; тестирование и сравнение полученных моделей; анализ результатов.

В работе приведен обзор литературы по изучаемой теме и сформулированы цель и задачи исследования. В первой главе представлены некоторые теоретические сведения по сверточным нейронным сетям, сделан выбор архитектуры сети, которая станет основой последующих моделей. Вторая глава посвящена подготовке данных и обучению нейронных сетей. В третьей главе представлены полученные результаты проведенных вычислительных экспериментов, выполнено сравнение моделей. В разделе «Выводы» проводится анализ полученных результатов. В заключении подведены итоги проведенной работы.

Обзор литературы

В рамках работы был изучен материал по созданию, обучению нейронных сетей. Задачам определения пола и возраста по фотографии посвящено большое количество работ.

Возможные методы решения поставленной задачи:

В работе H. Pan, H. Han, S. Shan, X. Chen [17] описывается теория по применению метода Mean-variance loss, в том числе для задачи определения возраста.

Работа W. Shen, Y. Guo, Y. Wang [19] посвящена теории случайных лесов и применению ансамблей решающих деревьев в регрессионной задаче оценки возраста.

В работе T. Y. Yang, Y. H. Huang, Y. Y. Lin [21] описываются нейронные сети типа SSR-Net. Показано применение их компактной архитектуры на примере распознавания возраста.

Теория и практика по нейронным сетям:

В работе О. Жерона систематически изложена теория машинного обучения с использованием Scikit-Learn и TensorFlow [3]. Вопросу о применении нейронных сетей на основе метода кейсов посвящена работа У. Микелуччи [5]. Различные аспекты глубокого обучения рассматриваются и в работах других авторов, например С. Паттанаяка и Э. Траска [8, 11].

Некоторые существующие решения задачи определения пола и возраста человека по фотографии:

Среди работ, посвященных данной теме, следует отметить работы В. Кустиковой [4], Д. В. Пакулича, С. А. Якимова, С. А. Алямкина [7], В. Булдумака [1], А. Суилина [10].

Материалы для обучения сетей и проведения вычислительных экспериментов:

Наборы фотографий – IMDb & Wiki [16]; Adience [14]; FG-Net [15]; Mega Age и Mega Age Asian [18]; CARC [13].

VGG-Face [20] – сверточная нейронная сеть, используемая для задач распознавания лица.

Постановка задачи

Целью работы является разработка системы распознавания возраста и пола человека по фотографии лица. Основываясь на данной цели, были сформулированы основные задачи:

1. провести анализ существующих решений данной задачи;
2. найти подходящую базу фотографий лиц людей с метками для пола и возраста;
3. выполнить предобработку изображений;
4. подобрать архитектуры нейронных сетей для решения задачи;
5. задать параметры обучения;
6. обучить сети, проверить работоспособность на тестовом множестве и сравнить модели;
7. провести анализ полученных результатов.

Глава 1. Нейронные сети

1.1 Некоторые теоретические выкладки о нейронных сетях

Нейросеть представляет собой систему из множества нейронов (процессоров). Нейрон – вычислительная единица, получающая информацию и производящая над ней какие-либо вычисления. По отдельности эти процессоры достаточно просты, но, будучи соединенными, они способны выполнять очень сложные задачи. Будем использовать модель искусственного нейрона Мак-Каллока-Питтса (Питца) [6]. Искусственный нейрон содержит веса (параметр, благодаря которому входная информация изменяется), сумматор (элемент, осуществляющий суммирование сигналов, поступающих на его вход) и функцию активации, которая определяет выходное значение нейрона в зависимости от результата взвешенной суммы входов и порогового значения. Схема искусственного нейрона представлена на Рисунке 1.

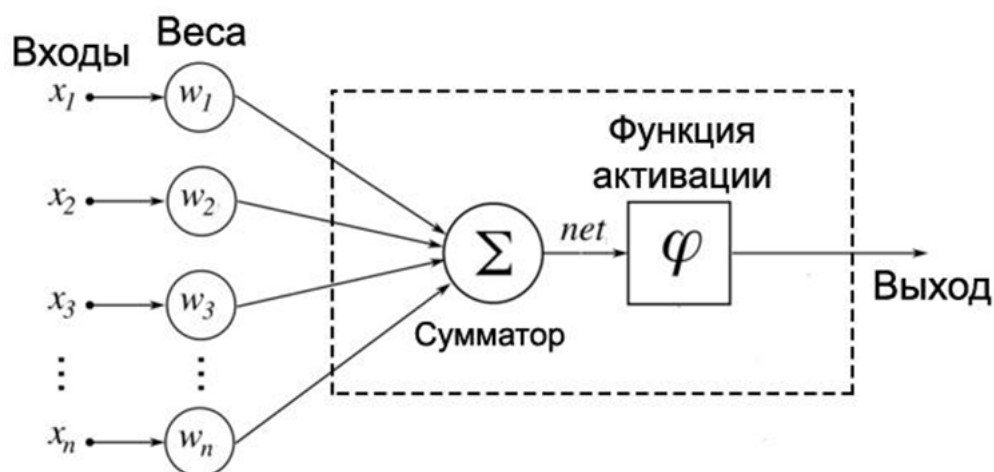


Рис. 1 – Схема искусственного нейрона

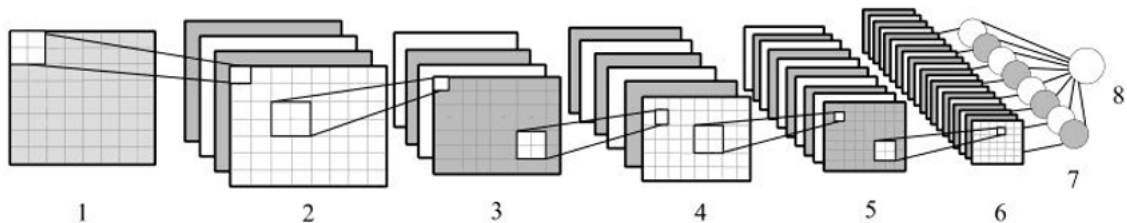
Принцип работы нейрона следующий: на каждый вход подаются значения. Далее взвешенные сигналы поступают на вход сумматора. Функция активации рассчитывает выходной сигнал.

Многослойными персептронами называют полносвязные нейронные сети прямого распространения. Входной сигнал в таких сетях распространяется в прямом направлении, от слоя к слою. Многослойный персептрон в общем представлении состоит из следующих элементов:

- множества входных узлов, которые образуют входной слой;
- одного или нескольких скрытых слоев нейронов;
- выходного слоя нейронов.

Результатом работы подобной сети будет вектор сигналов выходного слоя. К этому вектору применяется функция ошибки, которая будет показывать, насколько результат сети далек от эталонного.

В работе для распознавания пола и возраста будут использоваться нейросети сверточного типа. Как правило, архитектура данной сети состоит из чередующихся сверточных и подвыборочных слоев. Схема сверточной нейронной сети изображена на Рисунке 2.



1 – вход; 2, 4, 6 – сверточные слои; 3, 5 – подвыборочные слои;
7 – слой из обычных нейронов; 8 - выход

Рис. 2 – Схема сверточной нейронной сети [9]

Работа сверточной нейронной сети обычно интерпретируется как переход от конкретных особенностей изображения к более абстрактным деталям, и далее к еще более абстрактным.

Название архитектура получила из-за наличия операции свертки (1).

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

где A – матрица размера $(n_x \times n_y)$,

B – матрица (ядро) размера $(m_x \times m_y)$,

C – результирующая матрица размера $(n_x - m_x + 1 \times n_y - m_y + 1)$,

$i = 0, 1, \dots, n_x - m_x + 1$; $j = 0, 1, \dots, n_y - m_y + 1$.

Другой тип слоев данной сети – подвыборочный (pooling). Он выполняет уменьшение размерности сформированных карт признаков. В данной архитектуре считается, что информация о факте наличия искомого признака важнее точного знания его координат, поэтому из нескольких соседних нейронов карты признаков выбирается максимальный и принимается за один нейрон уплотненной карты признаков меньшей размерности. За счет данной операции, помимо ускорения дальнейших вычислений, сеть становится более инвариантной к небольшим изменениям входного изображения, таким как небольшие повороты, небольшое масштабирование. После сверточных и подвыборочных слоев следует классификатор в виде многослойного персептрона. Метод обратного распространения ошибки для сверточных и предвыборочных слоев имеет свои особенности [8].

Структура сверточной сети – однонаправленная (без обратных связей), принципиально многослойная.

1.2 Обучение нейронных сетей

Задача обучения нейронной сети состоит в настройке весов, таким образом, чтобы выходная ошибка была минимальна.

Существует много разных видов машинного обучения. Для удобства сгруппируем их в обширные категории на основе следующих признаков:

1. обучаются ли они с использованием меток (обучение с учителем, обучение без учителя, частичное обучение и обучение с подкреплением);

2. могут ли они обучаться «онлайн» («оффлайн» обучение, пакетное обучение или обучение на мини пакетах).

Перечисленные критерии не являются взаимоисключающими, их можно комбинировать. Более подробно рассмотрим обучение с учителем и пакетное обучение.

Обучение с учителем предполагает, что кроме обучающего множества входных данных у нас есть такое множество ответов (меток), что каждому примеру соответствует определенная метка. При обработке примера сетью, выход сравнивается с эталонным результатом и свободные параметры изменяются в зависимости от ошибки (насколько сильно вектор, выданный сетью, отличается от указанной метки).

Эпоха – набор итераций (шагов) обучения, в процессе которых предъявлены все примеры обучающей выборки. Шаг обучения – совокупность вычислений над одним примером или пакетом примеров из обучающего набора, приводящих к изменению весов. Для преодоления проблемы эффективной работы с большими объемами данных и избежания локальных минимумов используют пакетное обучение (обучение на «мини батчах») – выборку делят на части меньшего размера (batch), загружают их по очереди и обновляют веса нейросети в конце каждого шага.

Для обучения используют различные методы оптимизации первого порядка, для расчета градиента применяют метод обратного распространения ошибки. В основе идеи алгоритма лежит использование выходной ошибки сети для вычисления величин коррекции весов нейронов в скрытых слоях.

Алгоритм является итеративным и использует принцип обучения «по шагам», когда веса нейронов сети корректируются после подачи на ее вход одного обучающего примера или пакета примеров. На каждой итерации происходит два прохода сети – прямой и обратный. На прямом входной вектор распространяется от входов сети к ее выходам и формирует некоторый выходной вектор, соответствующий текущему состоянию весов. Затем вычисляется ошибка нейронной сети в соответствии с используемой функцией ошибки. На обратном проходе эта ошибка распространяется от выхода

сети к ее входам, и производится коррекция весов нейронов.

Алгоритм использует так называемый стохастический градиентный спуск, «продвигаясь» в многомерном пространстве весов в направлении антиградиента с целью достичь минимума функции ошибки. В данной работе используются исключительно алгоритмы мини-пакетного обучения с учителем на основе метода обратного распространения ошибки.

Для обучения моделей будем применять Adam – один из самых эффективных алгоритмов оптимизации в обучении нейронных сетей. Он сочетает в себе идеи RMSProp и оптимизатора импульса. Отличается он от них двумя идеями: во-первых, оценка первого момента вычисляется как скользящее среднее; во-вторых, из-за того, что оценки первого и второго моментов инициализируются нулями, используется небольшая коррекция, чтобы результирующие оценки не были смещены к нулю. Метод также инвариантен к масштабированию градиентов. При заданных гиперпараметрах $\beta_1 = 0,9$, $\beta_2 = 0,99$, η и $m_0 = 0$, $v_0 = 0$ правило пересчета следующее:

$$m_p = \beta_1 m_{p-1} + (1 - \beta_1) * \nabla L(w_{p-1}),$$

$$v_p = \beta_2 v_{p-1} + (1 - \beta_2) * (\nabla L(w_{p-1}))^2,$$

$$\hat{m}_p = \frac{m_p}{1 - \beta_1^p}, \hat{v}_p = \frac{v_p}{1 - \beta_2^p},$$

$$w_p = w_{p-1} - \frac{\eta * \hat{m}_p}{\sqrt{\hat{v}_p + \epsilon}},$$

где m_p – первый момент, v_p – второй момент, p – номер шага,

w_p – вектор весов на p -ом шаге,

ϵ – малое число (для избежания деления на нуль),

L – функция ошибки (бинарная или категориальная перекрестная энтропия для классификации и среднеквадратичная ошибка для регрессии),

$\nabla L(w_{p-1})$ – вектор градиента (частных производных $\frac{\partial L}{\partial w_{p-1,i}}$).

1.3 VGG-Face

Для решения задачи определения пола и возраста человека по фотографии воспользуемся предобученной сетью VGG-Face. Архитектура сети представлена на Рисунке 3 и в Таблице 9 [Приложение Б. Архитектура сети VGG-Face].

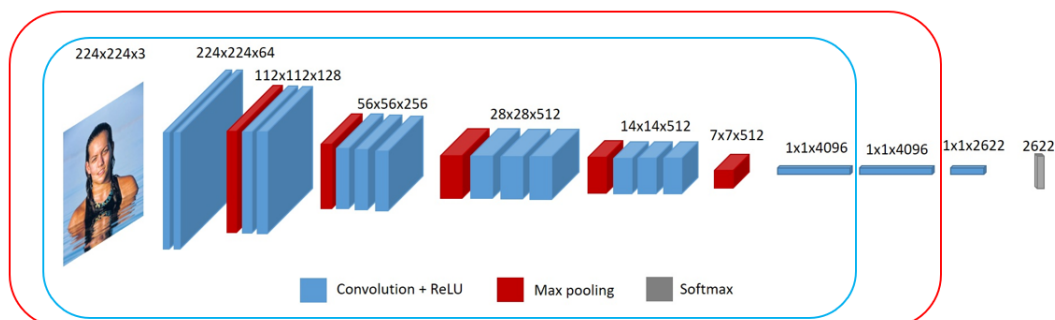


Рис. 3 – Схема архитектуры сети VGG-Face

VGG16 – модель сверточной нейронной сети, предложенная К. Simonyan и А. Zisserman из Оксфордского университета в статье «Very Deep Convolutional Networks for Large-Scale Image Recognition». Модель достигает точности 92.7% – при тестировании на ImageNet в задаче распознавания объектов на изображении. Этот датасет состоит из более чем 14 миллионов изображений, принадлежащих к 1000 классам. Сеть VGG16 обучалась на протяжении нескольких недель при использовании видеокарт NVIDIA TITAN BLACK.

Модель VGG-Face – это модификация VGG16, предназначенная для распознавания лиц.

Синим цветом на Рисунке 3 выделены слои, веса которых были заморожены, так как они уже могут находить какие-либо признаки. Красным выделена часть архитектуры, которая будет использоваться в последующих моделях без изменений.

Глава 2. Построение и обучение нейронных сетей

2.1 Математическая постановка задачи

Математически задачу можно поставить следующим образом:

- **Дано:** база фотографий людей различного пола и возраста, произвольного размера.
- **Требуется получить:** отображение множества фотографий в множество N классов, которое будет определено экспериментально на основе метрик и матрицы неточностей ($4 \leq N \leq 101$).
- Для оценки моделей будем использовать стандартные метрики качества классификации и регрессии:

$$accuracy = \frac{TP + TN}{TP + TN + FP + FN}; \quad (2)$$

$$recall = \frac{TP}{TP + FN}; \quad (3)$$

$$precision = \frac{TP}{TP + FP}; \quad (4)$$

$$F\text{-мера} = 2 \frac{precision * recall}{precision + recall};$$

$$mae = \frac{\sum_{i=1}^n |y_i - x_i|}{n},$$

где TP – количество предсказаний, когда классификатор верно отнес объект к рассматриваемому классу;

TN – количество предсказаний, когда классификатор верно утверждает, что объект не принадлежит к рассматриваемому классу;

FP – количество предсказаний, когда классификатор неверно отнес объект к рассматриваемому классу;

FN – количество предсказаний, когда классификатор неверно утверждает, что объект не принадлежит к рассматриваемому классу;

x_i – истинное значение возраста для i -го изображения; y_i – предсказание классификатора для i -го изображения.

Указанные формулы (2), (3), (4) и соответствующие обозначения для них взяты из стандартной задачи бинарной классификации, в случае большего количества классов будем вычислять среднее значение соответствующих показателей, где конкретный класс будет считаться положительным классом, а все другие – отрицательным.

2.2 Выбор базы фотографий

Для эффективного обучения и правильной работы сети необходимы размеченные данные, содержащие большое количество примеров. В свободном доступе существует некоторое множество таких наборов, но их разнообразность и наличие ошибочных меток могут сильно повлиять на результаты работы модели.

В связи с этим были проанализированы следующие наборы данных:

- IMDB + Wiki;
- LAP;
- Adience;
- Morph Album II;
- FG-NET;
- MegaAge и MegaAgeAsian;
- CARC.

На основе проведенного сравнения [Приложение С. Базы фотографий лиц людей] для исследования был выбран набор Wikipedia (Wiki), так как он находится в открытом доступе и содержит достаточное количество фотографий для проведения вычислительных экспериментов.

Примеры изображений из датасета Wiki и его распределение данных представлены на Рисунках 4 и 5.

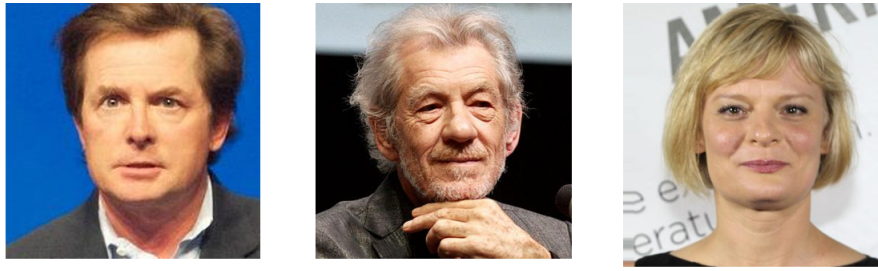


Рис. 4 – Примеры фотографий набора Wiki

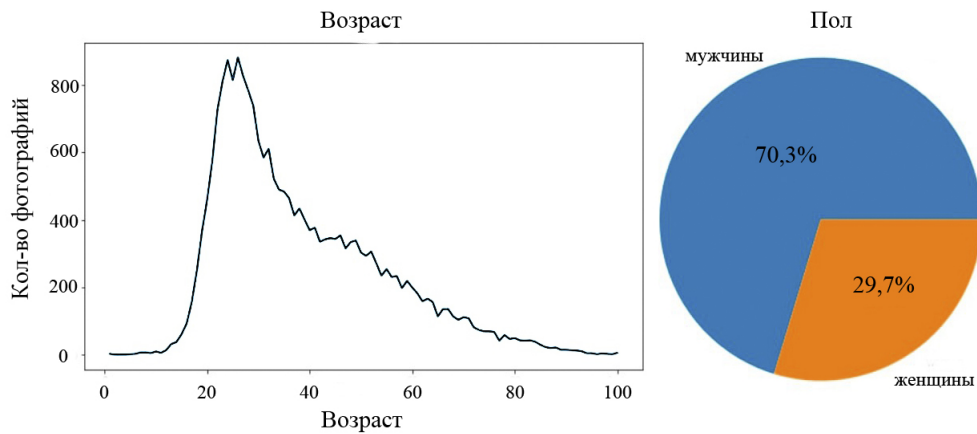


Рис. 5 – Распределение данных набора Wiki

2.3 Предобработка данных

При распаковке архива с базой фотографий создается 100 папок изображений и индексный файл (.mat). Преобразование в фрейм данных pandas упростит предобработку. Набор данных содержит дату рождения (dob) в формате datetime. Нам нужно преобразовать это в формат времени данных Python. Теперь у нас есть дата рождения и время съемки. Вычитание этих значений даст нам возраст.

На некоторых фотографиях Wiki нет людей, или на них изображены несколько человек. Для части людей получился возраст более 100 лет или меньше 0, либо информация о возрасте и вовсе отсутствует. Очистим датасет от подобных фотографий и ограничим диапазон возраста от 0 до 100 лет. Преобразуем фотографии под размер 224 × 224 пикселей. Также

разделим выборку на обучающую (70%), тестовую (20%) и валидационную (10%).

2.4 Подготовка рабочей среды

Архитектуры нейросетей и алгоритмы для работы с базой данных фотографий написаны на языке программирования Python. Нейронные сети реализованы в среде TensorFlow. Для обучения моделей и проведения вычислительных экспериментов использовался сервис SageMaker от Amazon, инстанс ml.p2.xlarge.

2.5 Определение пола

Распознавание пола довольно простая задача бинарной классификации. Добавим к выделенной части архитектуры VGG-Face слой свертки с двумя классами для гендерного признака, flatten для преобразования 2d-вектора признаков в одномерный и функцию активации softmax. Модифицированная архитектура указана в Таблице 1.

Таблица 1 – Архитектура сети для распознавания пола

№	Слой	Output Shape	Кол-во параметров
36	VGG-Face (1-36 слои)	(None, 1, 1, 4096)	0
37	predictions (Conv2D)	(None, 1, 1, 2)	8 194
38	Flatten	(None, 2)	0
39	Activation	(None, 2)	0

Зададим следующие параметры обучения:

- функция потерь – бинарная кросс-энтропия;
- оптимизатор – Adam;
- количество эпох обучения – 20;
- batch size – 256;
- шаг обучения – 10^{-3} .

Обучим модель. Графики обучения представлены на Рисунке 6.

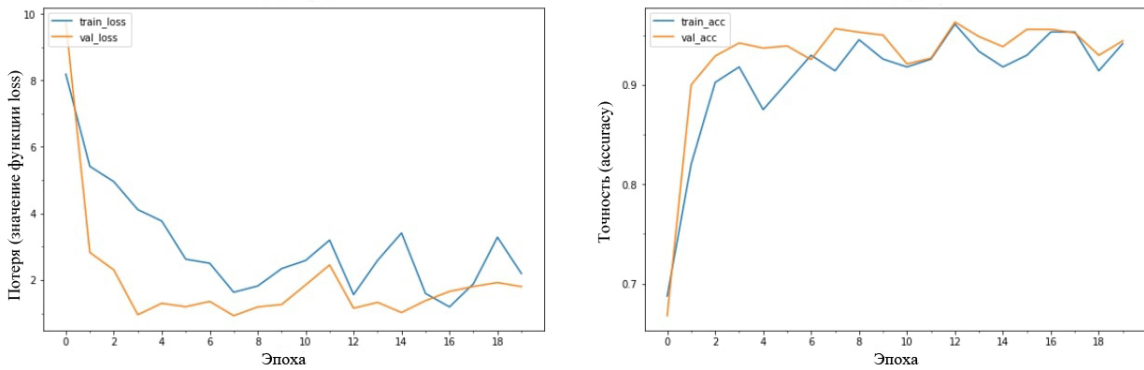


Рис. 6 – Графики обучения модели для определения пола

2.6 Предсказание пола и возраста одним классификатором

Определение возраста – это регрессионная задача, но мы решим ее и как задачу классификации, путем группировки по различным возрастным диапазонам, то есть каждый возрастной промежуток будет отдельным классом. Возрастное разбиение сделаем следующим: 15–30, 31–40, 41–60, 61–80 (основываясь на периодах Эриксона [2]). Фотографий с людьми моложе 15 и старше 80 мало, поэтому отбросим их. Точное количество фотографий в каждом возрастном диапазоне указано в Таблице 2. Сбалансируем выборки так, чтобы число изображений с мужчинами было таким же как с женщинами.

Таблица 2 – Распределение фотографий по возрастным промежуткам

	15–30	31–40	41–60	61–80
Кол-во фотографий	3 867	2 038	2 299	796

Модифицированная архитектура представлена в Таблице 3. В данном случае число классов сделаем равным шести (два для пола и четыре для возраста), функция активации – sigmoid.

Таблица 3 – Архитектура сети для распознавания пола и возраста

№	Слой	Output Shape	Кол-во параметров
36	VGG-Face (1-36 слои)	(None, 1, 1, 4096)	0
37	predictions (Conv2D)	(None, 1, 1, 6)	24 582
38	Flatten	(None, 6)	0
39	Activation	(None, 6)	0

Зададим следующие параметры обучения:

- функция потерь – бинарная кросс-энтропия;
- оптимизатор – Adam;
- количество эпох обучения – 20;
- batch size – 256;
- шаг обучения – 10^{-3} .

Обучим модель. Графики обучения представлены на Рисунке 7.

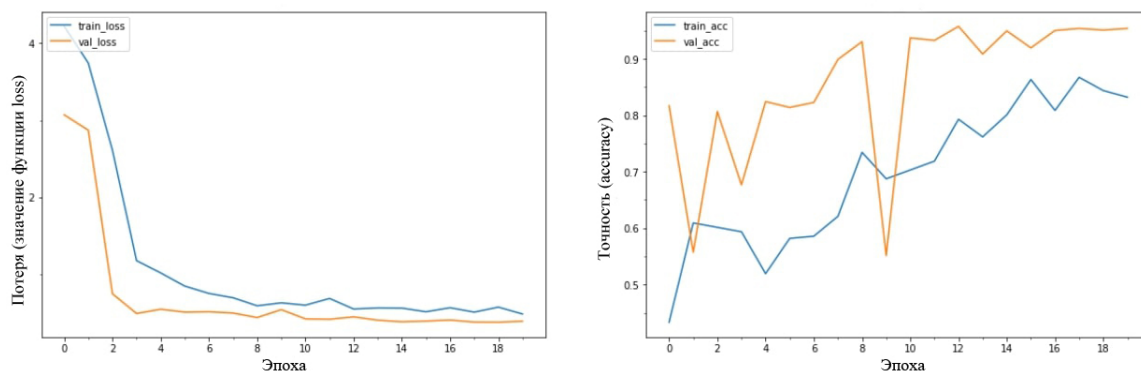


Рис. 7 – Графики обучения модели для определения пола и возраста

2.7 Определение возраста регрессией без разделения по гендерному признаку

Теперь решим задачу определения возраста как регрессионную. Сбалансируем выборки так, чтобы число изображений с мужчинами было таким же как с женщинами.

К выделенной части VGG-Face добавим два слоя свертки и полносвязный слой. Модифицированная архитектура сети представлена в Таблице 4.

Таблица 4 – Архитектура сети для распознавания возраста регрессией

№	Слой	Output Shape	Кол-во параметров
36	VGG-Face (1-36 слои)	(None, 1, 1, 4096)	0
37	Conv2D	(None, 1, 1, 128)	524 416
38	Conv2D	(None, 1, 1, 32)	4 128
39	Dense	(None, 1, 1, 1)	33

Зададим следующие параметры обучения:

- функция потерь – среднеквадратичная ошибка;
- оптимизатор – Adam;
- количество эпох обучения – 10;
- batch size – 256;
- шаг обучения – 10^{-5} .

Обучим модель.

2.8 Определение возраста классификацией без разделения по гендерному признаку

Поставим задачу определения возраста как классификационную. Возьмем такие же возрастные периоды, как и ранее. Архитектура сети будет аналогична, как у модели определения пола и возраста, за исключением числа классов и функции активации. Сбалансируем выборки так, чтобы число изображений с мужчинами было таким же как с женщинами.

Модифицированная архитектура сети представлена в Таблице 5. Число классов равно четырем, функция активации – softmax.

Таблица 5 – Архитектура сети для распознавания возраста классификацией

№	Слой	Output Shape	Кол-во параметров
36	VGG-Face (1-36 слои)	(None, 1, 1, 4096)	0
37	predictions (Conv2D)	(None, 1, 1, 4)	16 388
38	Flatten	(None, 1, 4)	0
39	Activation	(None, 1, 4)	0

Зададим следующие параметры обучения:

- функция потерь – категориальная кросс-энтропия;
- оптимизатор – Adam;
- количество эпох обучения – 20;
- batch size – 256;
- шаг обучения – 10^{-5} .

Обучим модель. Графики обучения представлены на Рисунке 8.

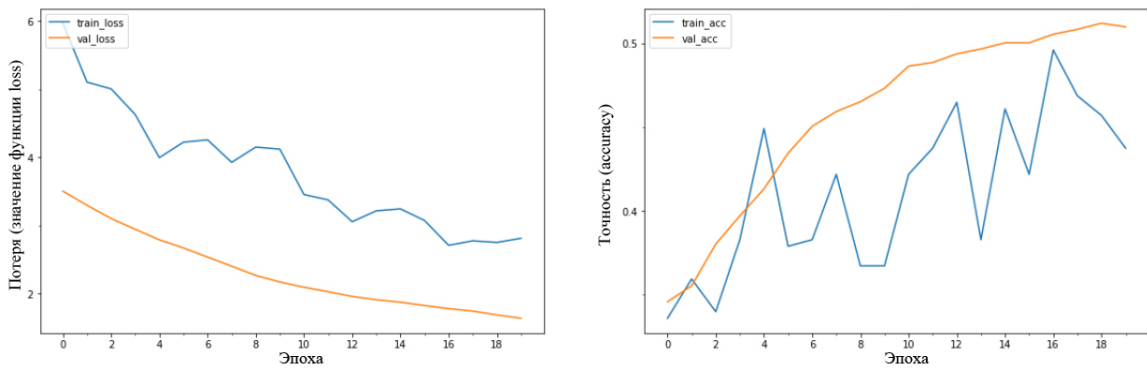


Рис. 8 – Графики обучения модели для определения возраста классификацией

2.9 Определение возраста классификацией с разделением по гендерному признаку

Предполагается, что сначала будет определяться пол, а затем – возраст.

Разделим набор Wiki на две части: фотографии, содержащие только мужчин, и фотографии, содержащие только женщин. Наборы train, test, validation сбалансируем по количеству изображений для каждого возрастного класса и сделаем распределение по возрастам близкое к гауссову. Пример такого распределения представлен на Рисунке 9.

К сожалению, набор Wiki содержит мало фотографий с женщинами. Увеличим их количество путем поворота изображений влево и вправо на небольшой угол и отражения по горизонтали. Пример преобразований представлен на Рисунке 10. Точное количество фотографий в каждом возрастном диапазоне указано в Таблице 6.

Таблица 6 – Распределение фотографий по возрастным промежуткам

	15–30	31–40	41–60	61–80
Кол-во фотографий муж.	319	370	412	368
Кол-во фотографий жен.	503	650	736	667

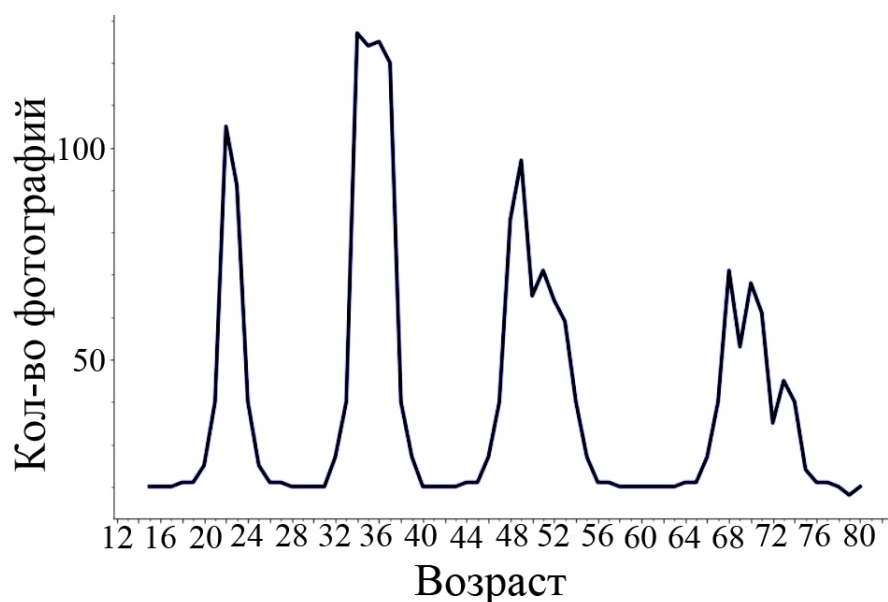


Рис. 9 – Модифицированное распределение данных для мужчин



Рис. 10 – Пример преобразований фотографий

Архитектура сети остается такой же, как и в случае классификации без разделения датасета по гендерному признаку (Таблица 5).

Зададим следующие параметры обучения:

- функция потерь – категориальная кросс-энтропия;
- оптимизатор – Adam;
- количество эпох обучения – 160 для мужчин и 240 для женщин;
- batch size – 256;
- шаг обучения – 10^{-5} .

Обучим модели. Графики обучения представлены на Рисунках 11 и 12.

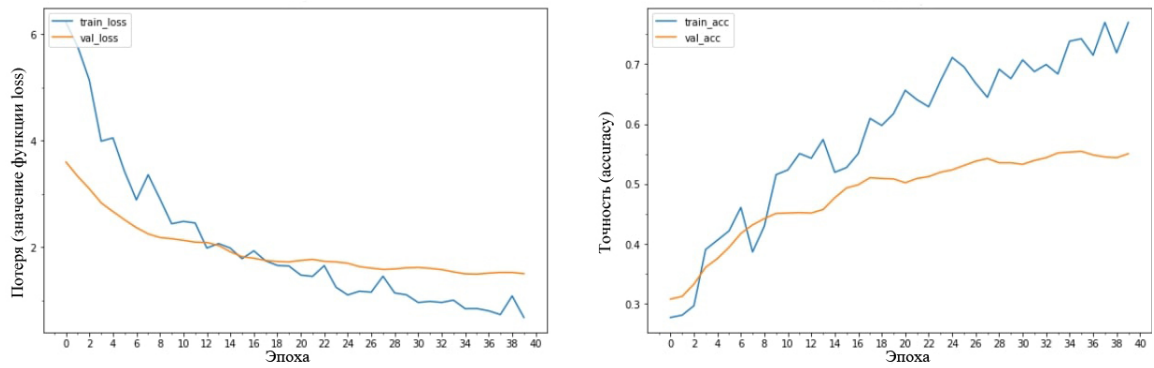


Рис. 11 – Графики обучения модели определения возраста классификацией для мужчин

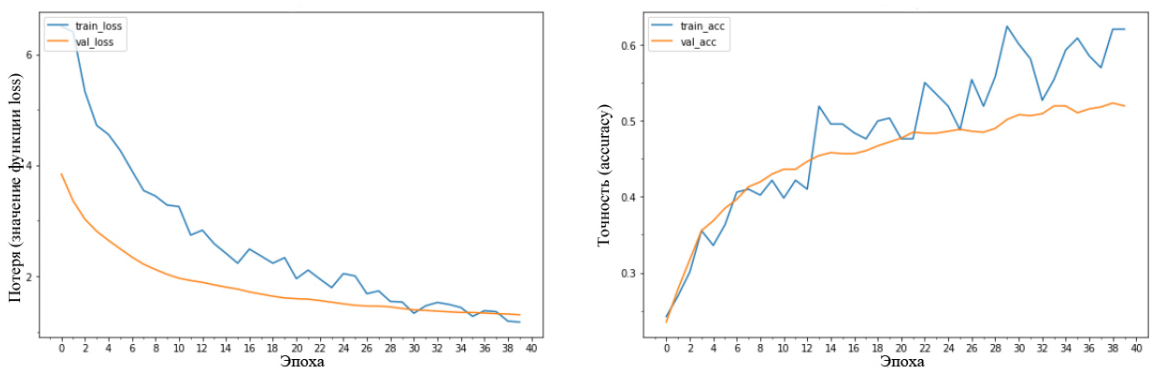


Рис. 12 – Графики обучения модели определения возраста классификацией для женщин (первые 40 эпох)

2.10 Определение возраста регрессией с разделением по гендерному признаку

Проделаем аналогичные действия с разделением и балансировкой датасета, как у прошлой модели. Архитектура сети будет такой же, как в случае регрессии без разделения по гендерному признаку (Таблица 4).

Зададим следующие параметры обучения:

- функция потерь – среднеквадратичная ошибка;
- оптимизатор – Adam;
- количество эпох обучения – 20 для мужчин и женщин;
- batch size – 256;
- шаг обучения – 10^{-5} .

Обучим модель.

Глава 3. Анализ полученных результатов

3.1 Результаты вычислительных экспериментов

Проведенные вычислительные эксперименты и полученные результаты приведены в Таблице 7.

Таблица 7 – Полученные результаты

Модель	Кол-во эпох	Train/Test	Accuracy	Recall	F-мера	Мае
Пол	20	9000/2600	0,95	0,96	0,96	-
Пол и возраст	20	9000/2600	Пол: 0,95 Возраст: 0,41	Пол: 0,96 Возраст: 0,42	Пол: 0,96 Возраст: 0,41	-
Возраст, регрессия без разделения по гендеру	10	15000/7000	-	-	-	22,26
Возраст, классификация без разделения по гендеру	20	9000/2600	0,52	0,45	0,45	-
Возраст, классификация для мужчин	160	1469/800	0,59	0,60	0,60	-
Возраст, классификация для женщин	240	2556/800	0,63	0,64	0,63	-
Возраст, регрессия для мужчин	20	1469/1000	-	-	-	10,76
Возраст, регрессия для женщин	20	2556/800	-	-	-	11,73

3.2 Сравнение моделей

Как можно увидеть из Таблицы 7, модель для определения пола и возраста имеет довольно низкую точность предсказания возрастного класса на тестовом множестве, точность для пола такая же, как у модели определения только пола. Сеть для распознавания возраста с помощью классификации, имея похожую архитектуру, показала более высокую точность, но все еще довольно низкую. После балансировки и разделения датасета по гендерному признаку точность классификационной модели для определения возраста заметно возросла (на 10–15%). Аналогичная ситуация возникает и с регрессией: при идентичной архитектуре точность увеличилась примерно вдвое.

Можно сделать вывод, что качество распределения данных в датасете имеет ключевое значение для обучения и правильной работы сети. Модель, обученная на сбалансированном и разделенном по гендерному признаку наборе фотографий имеет лучший результат, по сравнению с другими моделями.

Выводы

Целью данного исследования было построить нейросетевую модель, прогнозирующую пол и возраст человека по фотографии лица. Разработанная система определяет пол с точностью 95% и возрастной класс с точностью около 60% на тестовой выборке. Качество работы моделей росло с балансировкой и разделением обучающей выборки. Значит можно сделать вывод, что дальнейшее увеличение обучающей выборки данных, количества эпох обучения и более точная настройка возрастных классов позволит улучшить результат.

На основе полученных результатов можно отметить, что задачи распознавания пола и возраста эффективней решать по отдельности, предварительно разбив датасет по гендерному признаку и сбалансировав его по количеству фотографий для каждого возрастного класса.

Заключение

В выпускной квалификационной работе бакалавра были решены следующие задачи:

1. изучены возможные решения данной проблемы;
2. найдены необходимые для обучения и тестирования сети компоненты, а именно – набор фотографий Wikipedia и предобученная сеть VGG-Face;
3. выполнена предобработка собранных данных, также выборка поделена на 3 части – обучающую, тестовую и валидационную;
4. разработаны и обучены нейронные сети;
5. проанализированы полученные результаты, выполнено сравнение реализованных моделей и сделаны соответствующие выводы.

СПИСОК ИСПОЛЬЗОВАННЫХ ИСТОЧНИКОВ

- [1] Булдумак, В. Определяем возраст и пол человека, используя нейронную сеть [Электронный ресурс]: URL:<https://python-scripts.com/predict-age-and-gender> (дата обращения: 12.01.2021).
- [2] Возрастная периодизация Эриксона [Электронный ресурс]: URL:<https://mamsila.ru/post/vozrastnaya-periodizaciya-eriksona> (дата обращения: 10.02.2021).
- [3] Жерон, О. Прикладное машинное обучение с помощью Scikit-Learn и TensorFlow / О. Жерон; пер. с англ. – М. : СПб. ООО Альфа-книга, 2018. – 688 с.
- [4] Кустикова, В. Сверточные нейронные сети. Глубокие остаточные сети [Электронный ресурс]: URL:http://hpc-education.unn.ru/files/courses/intel-neon-course/Rus/Lectures/Presentations/4_CNN.pdf (дата обращения: 12.01.2021).
- [5] Микелуччи, У. Прикладное глубокое обучение. Подход к пониманию глубоких нейронных сетей на основе метода кейсов. / У. Микелуччи; пер. с англ. – М. : СПб. БХВ-Петербург, 2020. – 368 с.
- [6] Модель искусственного нейрона Мак-Каллока-Питсса [Электронный ресурс]: URL:<http://www.mathnet.ru/links/c380a2a18f452384395ac32be4f81eee/mais177.pdf> (дата обращения: 20.12.2020).
- [7] Пакулич, Д. В. Распознавание возраста по изображению лица с использованием сверточных нейронных сетей / Д. В. Пакулич, С. А. Якимов, С. А. Алямкин. – Т. 55 Вып. 3 изд. – М : Новосибирск. Автометрия, 2019. – с. 52–61.
- [8] Паттанаяк, С. Глубокое обучение и TensorFlow для профессионалов. Математический подход к построению систем искусственного интеллекта на Python / С. Паттанаяк; пер. с англ. – М. : СПб. ООО "Диалектика 2019. – 480 с.

- [9] Сверточная нейронная сеть, часть 1: структура, топология, функции активации и обучающее множество [Электронный ресурс]: URL:<https://habr.com/ru/post/348000/> (дата обращения: 10.02.2021).
- [10] Суилин, А. Определение пола и возраста по фото [Электронный ресурс]: URL:https://suilin.ru/project/gender_age/ (дата обращения: 15.02.2021).
- [11] Траск, Э. Грокаем глубокое обучение / Э. Траск; пер. с англ. – М. : СПб. Питер, 2019. – 352 с.
- [12] Apparent Age and Gender Prediction in Keras [Электронный ресурс]: URL:<https://sefiks.com/2019/02/13/apparent-age-and-gender-prediction-in-keras/> (дата обращения: 12.01.2021).
- [13] Cross-Age Reference Coding for Age-Invariant Face Recognition and Retrieval [Электронный ресурс]: URL:<https://bcsiriuschen.github.io/CARC/> (дата обращения: 10.02.2021).
- [14] Dataset Adience [Электронный ресурс]: URL:<https://talhassner.github.io/home/projects/Adience/Adience-data.html> (дата обращения: 10.02.2021).
- [15] Dataset FG-Net [Электронный ресурс]: URL:http://yanweifu.github.io/FG_NET_data/FGNET.zip (дата обращения: 10.02.2021).
- [16] IMDB-WIKI – 500k+ face images with age and gender labels [Электронный ресурс]: URL:<https://data.vision.ee.ethz.ch/cvl/rrothe/imdb-wiki/> (дата обращения: 10.02.2021).
- [17] Pan, H. Mean-variance loss for deep age estimation from a face / H. Pan, H. Han, S. Shan, X. Chen; Proc. of the Conf. on Computer Vision and Pattern Recognition. – М. : Salt Lake City, USA, 18–22 June, 2018. – p. 5285–5294.

- [18] Quantifying Facial Age by Posterior of Age Comparisons [Электронный ресурс]: URL:<http://mmlab.ie.cuhk.edu.hk/projects/MegaAge/> (дата обращения: 10.02.2021).
- [19] Shen, W. Deep regression forests for age estimation / W. Shen, Y. Guo, Y. Wang; Proc. Of the Conf. on Computer Vision and Pattern Recognition. – М. : Salt Lake City, USA, 18–22 June, 2018. – p. 2304–2313.
- [20] VGG-Face model for Keras // GitHub [Электронный ресурс]: URL:<https://gist.github.com/EncodeTS/6bbe8cb8bebad7a672f0d872561782d9> (дата обращения: 15.02.2021).
- [21] Yang, T. Y. SSR-Net: A compact soft stagewise regression network for age estimation / T. Y. Yang, Y. H. Huang, Y. Y. Lin; Proc. of the IEEE Intern. Joint Conf. on Artificial Intelligence. – М. : Stockholm, Sweden, 13–19 July, 2018. – p. 1078–1084.

Приложение А. Существующие решения задачи

Рассмотренные статьи представлены в Таблице 8.

Таблица 8 – Существующие решения задачи

Год	Автор	Статья	Dataset	Архитектура	Тип решаемой задачи
2017	В. Булдумак	Определяем возраст и пол человека, используя нейронную сеть	200 000 изображений	ResNet-10	Определение пола и возраста
2018	Артур Суилин	Определение пола и возраста по фото	IMDB + WIKI, Adience, UTKFace	ResNet-50	Определение пола и возраста
2019	Д. В. Пакулич, С. А. Якимов, С. А. Алямкин	Распознавание возраста по изображению лица с использованием сверточных нейронных сетей	IMDB + Wiki, LAP	MobileNetV2, ResNet50, SSR-Net	Определение возраста

Год	Автор	Статья	Dataset	Архитектура	Тип решаемой задачи
2018	Кустикова Валентина	Сверточные нейронные сети. Глубокие остаточные сети.	Не указан	модель из статьи	Определение пола
-	Gil Levi, Tal Hassner	Age and Gender Classification using Convolutional Neural Networks	Adience	Модель Caffee	Определение пола и возраста
-	Ari Ekmekji	Convolutional Neural Networks for Age and Gender Classification	Adience	модель из статьи	Определение пола и возраста

Год	Автор	Статья	Dataset	Архитектура	Тип решаемой задачи
2016	S. S. Liew, M. Khalil-Hani, S. A. Radzi, R. Bakhteri	Gender classification: a convolutional neural network approach	SUMS и AT&T	модель из статьи	Определение пола
2015	G. Antipova, S. A. Berrania, J. L. Dugelay	Minimalistic CNN-based ensemble model for gender prediction from face images	CASIA Web-Face и Labeled Faces in the Wild	модель из статьи	Определение пола
2018	Hongyu Pan, Hu Han, Shiguang Shan, Xilin Chen	Mean-Variance Loss for Deep Age Estimation from a Face	CLAP2016, MORPH Album II, FG-NET, AADB	Mean-Variance Loss, Alex-Net, VGG-16	Определение возраста

Приложение Б. Архитектура сети VGG-Face

Полная архитектура сети VGG-Face представлена в Таблице 9.

Таблица 9 – Архитектура сети VGG-Face

№	Слой	Output Shape	Кол-во параметров
1	Zero_Padding_2D (Input Layer)	(None, 224, 224, 3)	0
2	Zero_Padding_2D	(None, 226, 226, 3)	0
3	Conv_2D	(None, 224, 224, 64)	1 792
4	Zero_Padding_2D	(None, 226, 226, 64)	0
5	Conv_2D	(None, 224, 224, 64)	36 928
6	Max_Pooling_2D	(None, 112, 112, 64)	0
7	Zero_Padding_2D	(None, 114, 114, 64)	0
8	Conv_2D	(None, 112, 112, 128)	73 856
9	Zero_Padding_2D	(None, 114, 114, 128)	0
10	Conv_2D	(None, 112, 112, 128)	147 584
11	Max_Pooling_2D	(None, 56, 56, 128)	0
12	Zero_Padding_2D	(None, 58, 58, 128)	0
13	Conv_2D	(None, 56, 56, 256)	295 168
14	Zero_Padding_2D	(None, 58, 58, 256)	0
15	Conv_2D	(None, 56, 56, 256)	590 080
16	Zero_Padding_2D	(None, 58, 58, 256)	0
17	Conv_2D	(None, 56, 56, 256)	590 080
18	Max_Pooling_2D	(None, 28, 28, 256)	0
19	Zero_Padding_2D	(None, 30, 30, 256)	0
20	Conv_2D	(None, 28, 28, 512)	1 180 160
21	Zero_Padding_2D	(None, 30, 30, 512)	0

№	Слой	Output Shape	Кол-во параметров
22	Conv_2D	(None, 28, 28, 512)	2 359 808
23	Zero_Padding_2D	(None, 30, 30, 512)	0
24	Conv_2D	(None, 28, 28, 512)	2 359 808
25	Max_Pooling_2D	(None, 14, 14, 512)	0
26	Zero_Padding_2D	(None, 16, 16, 512)	0
27	Conv_2D	(None, 14, 14, 512)	2 359 808
28	Zero_Padding_2D	(None, 16, 16, 512)	0
29	Conv_2D	(None, 14, 14, 512)	2 359 808
30	Zero_Padding_2D	(None, 16, 16, 512)	0
31	Conv_2D	(None, 14, 14, 512)	2 359 808
32	Max_Pooling_2D	(None, 7, 7, 512)	0
33	Conv_2D	(None, 1, 1, 4096)	102 764 544
34	Dropout	(None, 1, 1, 4096)	0
35	Conv_2D	(None, 1, 1, 4096)	16 781 312
36	Dropout	(None, 1, 1, 4096)	0
37	predictions (Conv_2D)	(None, 1, 1, 2622)	10 742 334
38	Flatten	(None, 2622)	0
39	Activation	(None, 2622)	0

Приложение С. Базы фотографий лиц людей

Подробные характеристики рассмотренных наборов фотографий представлены в Таблице 10.

Таблица 10 – Сравнительная таблица датасетов

База данных	Доступность	Общее число фотографий	Разметка	Размер изображений	Метаданные	Фон
IMDB и Wiki	открытая	460 000, 62 000	есть	от 32×32 до 512×512	Формат .mat: dob – дата рождения, photo_taken – год съемки, full_path – путь, gender – пол, name – имя и другие	есть
LAP	закрытая	7 519	есть	неизвестно	неизвестно	неизвестно

База данных	Доступность	Общее число фотографий	Разметка	Размер изображений	Метаданные	Фон
Adience	открытая	26 580	есть	неизвестно	Формат .txt: user_id – папка с изображением, original_image – имя, face_id – Face ID, age – возраст, gender – пол, x, y, dx, dy – ограничивающая рамка лица	есть
Morph Album II	закрытая	55 134	есть	200 × 240 или 400 × 480	Формат .csv: id_num – идентификатор, dob – дата рождения, race – раса, gender – пол, age – возраст photo – имя файла	есть

База данных	Доступность	Общее число фотографий	Разметка	Размер изображений	Метаданные	Фон
FG-Net	открытая	1 002	есть	неизвестно	Метки в названии изображений	есть
Mega Age и Mega Age Asian	открытая	41 941 и 40 000	есть	неизвестно	Формат .txt: возраст, название изображения	есть
CARC	открытая	160 000	есть	неизвестно	Формат .mat: age – возраст, identity – идентификатор, year – год съемки, name – название фотографии	есть