

МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ РОССИЙСКОЙ ФЕДЕРАЦИИ
федеральное государственное бюджетное образовательное учреждение высшего образования
«Тольяттинский государственный университет»

Институт математики, физики и информационных технологий

(наименование института полностью)

Кафедра Прикладная математика и информатика

(наименование)

09.04.03 Прикладная информатика

(код и наименование направления подготовки)

Управление корпоративными информационными процессами

(направленность (профиль))

ВЫПУСКНАЯ КВАЛИФИКАЦИОННАЯ РАБОТА (МАГИСТЕРСКАЯ ДИССЕРТАЦИЯ)

на тему «Исследование и разработка методов компьютерного зрения для задач определения атрибутов личности по изображению лица»

Обучающийся

Г.И. Захаренко

(Инициалы Фамилия)

(личная подпись)

Научный
руководитель

д.т.н., доцент, С.В. Мкртычев

(ученая степень (при наличии), ученое звание (при наличии), Инициалы Фамилия)

Тольятти 2022

Оглавление

Глава 1 Анализ современного состояния исследований в области компьютерного зрения.....	9
1.1 Понятие и проблемы компьютерного зрения	9
1.2 Задачи компьютерного зрения.....	14
1.3 Актуальность задачи определения атрибутов личности по изображению лица	16
Глава 2 Анализ методов компьютерного зрения для решения задач определения атрибутов личности по изображению лица	22
2.1 Основные этапы определения атрибутов личности по изображению лица.....	22
2.2 Роль машинного обучения в определении атрибутов личности по изображению лица	26
2.3 Метод опорных векторов	29
2.4 Кластеризация изображения	31
2.5 Метод главных компонент (РСА)	34
2.6 Метод Виолы-Джонса.....	37
2.7 Нейросетевые методы.....	41
Глава 3 Разработка алгоритма определения атрибутов личности с помощью сверточных нейронных сетей.....	48
Глава 4 Экспериментальное исследование и апробация метода распознавания личности на основе нейросетевых алгоритмов.....	56
4.1 Описание программного комплекса и оборудования	56
4.2 Общая характеристика языка программирования Python	56
4.3 Формирование набора данных изображений лиц.....	59
4.4 Модели глубокого обучения.....	60
4.5 Методы оценки качества работы алгоритмов.....	61
4.6 Сравнительное исследование моделей Keras.....	63
4.7 Тестирование метода определения возраста и пола человека по изображению лица	66

Заключение	71
Список используемой литературы и используемых источников.....	73
Приложение А Фрагмент программного кода на языке программирования Python глубокого обучения модели определения атрибутов личности по изображению лица	77

Введение

Человечество с древних времен в процессе развития использует иллюстрации. Изображения возникли раньше письменности и несли в себе функцию сохранения информации. До нас дошли древние пещерные изображения, которые иллюстрируют события, происходящие в жизни обычного человека того времени (охота, стая животных, корабли). Кроме рисунков для фиксирования объектов на историческом пути становления компьютерного зрения применялись различные приборы и устройства.

Еще в 1970-е годы XX века ученые предлагали применять к изображению математические операции. Так, А. Розенфельд доказал, что любой матрице можно сопоставить дискретное изображение [18]. Вместе с тем изображение интересно не своими функциями и матрицами, а той конкретной информацией, которая в нем изначально заложена.

Прошло 50 лет, но несмотря на значительные успехи, достигнутые в области изучения искусственного интеллекта (ИИ) и машинного зрения, его роль в автоматизации является главной только на развитом производстве с высоким технологическим уровнем. В остальных случаях компьютерное зрение выступает как дополнительная вспомогательная опция к человеческому анализу.

Аналитики ведущего поставщика информации и консультационных услуг IDC в своем докладе «Эра данных 2025» (Data Age 2025) указывают, что уже к 2025 году объем всей имеющейся информации в мире составит 163 зеттабайт. Если сравнить с 10-летним предшествующим периодом, показатель превысит данные 2016 года в 10 раз. Другими словами, каждый год человечество создает почти столько же информации, сколько уже имеет.

Для эпохи продолжительного информационного взрыва характерна неразбериха и обесценивание информации. Из информационного водопада все сложнее выудить полезную релевантную часть. Человеку справиться с этой задачей не под силу.

Вследствие этого возникают сложности при решении таких задач, как выбор эффективных методов и средств сбора, обработки и анализа Big Data (большие объемы, высокоскоростные и/или вариативные информационные активы), а также препятствия для своевременного и обоснованного принятия управленческих решений [12].

Человек – самый распространенный объект на фотографиях, поэтому анализ изображений с людьми является первостепенной задачей компьютерного зрения. Лицо человека несет достаточно информации, которую можно обозначить как атрибуты. К ним относятся пол, возраст, эмоции, раса и другое, которые необходимо выделить для получения информации.

Чтобы решить эту задачу автоматически, следует привлечь к определению атрибутов личности по изображению лица интеллектуальные компьютерные системы. Информационный хаос смогут победить только самые инновационные разработки в такой области ИИ, как компьютерное зрение.

Об актуальности темы свидетельствует не только практическое внедрение системы технического зрения (СТЗ) в нетрадиционные для автоматизации сферы человеческой жизнедеятельности (мультимедиа, искусство и дизайн, геоинформационные системы и базы данных изображений) [24], но и проявление все большего интереса к ней со стороны научного сообщества.

Компьютерное зрение является довольно молодой наукой [16]. Примечательно, что свое основное развитие она получает в основном в последние годы в сторону практического использования. Например, изучение задач определения атрибутов личности по изображению лица с помощью компьютерного зрения связано с коммерческими интересами в большей степени, чем с научными. В связи с этим данная область прикладной информатики не может похвастаться большим количеством изданных научных произведений.

В ходе написания диссертации исследованы зарубежные и отечественные научные труды в области компьютерного зрения. По вышеуказанным причинам упор сделан на современные издания, однако рассмотрены также и классические монографии.

Объектом настоящего исследования является информационный процесс обработки графических изображений.

Предметом исследования являются методы компьютерного зрения для задач определения атрибутов личности по изображению лица.

Цель исследования: исследовать и разработать точные автоматические методы и алгоритмы распознавания таких атрибутов человека как пол и возраст, а также провести анализ их эффективности.

Гипотеза диссертации: следует ожидать, что разработанный в рамках диссертационного исследования алгоритм достаточно точно определит атрибуты личности по изображению лица.

В соответствии с поставленной целью выделены основные задачи диссертации:

- провести обзор современного состояния исследований в области методов компьютерного зрения;
- провести анализ наиболее эффективных методов определения атрибутов личности по изображению лица;
- выбрать методы, на основании которых разработать алгоритм определения возраста и пола лица по изображению;
- провести апробацию моделей, полученных в ходе обучения нейросетей с различными характеристиками, с целью определения более эффективных;
- провести практические эксперименты и оценить качество работы разработанного алгоритма.

В ходе работы применялись следующие методы научного исследования: визуализация данных, математическая статистика, ранжирование, шкалирование и нахождение математических величин.

При выполнении диссертации последовательно пройдены следующие этапы:

- постановочный (постановка проблемы, определение предмета и объекта исследования, изучение литературы в предметной области, формулировка цели, гипотезы и задач исследования);
- исследовательский (выбор методологии исследования, непосредственный процесс исследования, формулирование предварительных выводов и их уточнение, обоснование заключительных выводов);
- заключительный (практическое обоснование полученных результатов и их апробация, подтверждение или опровержение гипотезы).

Научная новизна диссертационной работы. В процессе исследования на основании метода машинного обучения был предложен алгоритм решения задачи автоматического определения по изображению таких атрибутов человека, как возраст и пол.

Теоретическая значимость. Теоретические положения и выводы, полученные в ходе диссертации, приведут к приращению научных знаний в области исследования компьютерного зрения, а также могут способствовать развитию научного интереса в изучении данного направления.

Практическая значимость диссертационной работы состоит в реализации теоретических аспектов в ходе практических экспериментов и в возможности применять получившиеся результаты в будущих достижениях области компьютерного зрения.

При проведении научного исследования принято участие в VII Международной научно-практической конференции (школы-семинара) молодых ученых «Прикладная математика и информатика: современные исследования в области естественных и технических наук» и Всероссийской студенческой научно-практической междисциплинарной конференции «Молодежь. Наука. Общество» с докладами, затрагивающими тему

диссертации: «Формирование изображений в системе технического зрения с помощью датчиков», «Цветовая фильтрация OpenCV для решения задач компьютерного зрения».

На защиту выносятся:

- алгоритм автоматического определения атрибутов личности (пол и возраст) по изображению лица на основе методов компьютерного зрения;
- результаты апробации и оценки эффективности разработанных методов определения атрибутов личности (пол и возраст) по изображению лица.

Диссертация состоит из введения, 4 глав, заключения, списка используемой литературы и используемых источников и приложения.

Работа изложена на 82 страницах, содержит 26 рисунков, 3 таблицы и 41 источник.

Использованные в диссертации изображения находятся в свободном доступе, являются общественным достоянием или собственными изображениями.

Глава 1 Анализ современного состояния исследований в области компьютерного зрения

1.1 Понятие и проблемы компьютерного зрения

Компьютерное зрение находится на одной из первых ступенек своего развития. Наука уже достигла ощутимых успехов в этой сфере, но окружающая среда изменчива, и объекты не всегда будут позировать для распознавания. Эффективное решение вопроса в этой области еще предстоит найти в ближайшем будущем [16].

Простыми словами, компьютерное зрение – автоматическое извлечение из изображения релевантной информации.

СТЗ внедряется в нашу повседневную жизнь: распознавание сетчатки глаз в целях безопасности, распознавание денежных купюр в банкомате, идентификация собственника, нарушившего Правила дорожного движения транспортного средства, вождение автомобиля автопилотом и многое другое.

Восприятие цвета – психофизиологическое свойство человеческого зрения, возникающее при наблюдении за объектом. Отраженный свет возникает за счет отражения света от поверхности объекта.

Вообще свет – электромагнитное излучение. Глаз воспринимает лишь небольшой фрагмент всего спектра излучения, но не весь.

Любой цвет возможно описать композицией трех цветов (красный, зеленый и синий). Цветность отличается размерами волн.

Как правило, в графических системах компьютера в основном используется трехцветная модель RGB (красный-зеленый-синий). Современные системы имеют основной трехбайтовый формат кодирования цвета, при котором на каждую компоненту цветности выделяется один байт памяти. Это дает возможность получать более 16 миллионов разных цветов [21].

Восприятие глаза построено следующим образом. Колбочки условно делятся на три группы по чувствительности к принимаемым световым волнам разной длины: длинных (красный цвет), средних (зеленый цвет) и коротких (синий цвет). Соответственно сетчатка принимает спектр в трех числах [14].

В итоге мозг получает усредненное значение в видимом диапазоне, и человек видит цвет с оттенком.

Глазное зрение человека отличается от фотоаппарата фокусом внимания. Конечно, зрение человека нельзя назвать совершенным.

При ухудшении видимости, например, в вечернее и ночное время, глаз не способен явственно различать цвета и очертания предметов. Зрение ухудшается и в связи с самочувствием человека и даже зависит от эмоционального состояния.

Наконец, даже если человек видит все четко, он способен ошибаться в даче оценки увиденному.

К сожалению, свои проблемы имеет и СТЗ.

Компьютерное зрение является подобластью ИИ. СТЗ должна не только «увидеть» изображение, но и предоставить нужные пользователю выводы о полученной информации.

Согласно «тесту Тьюринга» СТЗ должна ответить на все вопросы, касаемые полученного изображения, также качественно, как это сделал бы человек [23].

По какой причине несмотря на важность компьютерного зрения многие идеи до сих пор не могут быть технически воплощены?

Сложность обработки и снятия информации СТЗ с изображений объясняется следующими факторами:

- ракурс. Изображение одного и того же объекта с каждой стороны будет получено разное;
- освещение объекта влияет на полноту полученной информации;

- относительность объекта. В зависимости от расстояния до датчика объект может представляться разным;
- деформация объектов. Постоянство ракурса не является гарантией получения всей информации об объекте, который находится в движении;
- перекрытие. Объекты часто перекрывают друг друга, тем самым скрывая нужную информацию;
- слияние. Бывают ситуации, когда один объект растворяется в аналогичной цветовой гамме другого (например, черная кошка ночью);
- внутриклассовая изменчивость. Объекты одного класса и одинакового предназначения могут иметь совершенно разный внешний вид.

Как же распутать и сопоставить информацию, полученную по изображениям?

СТЗ может сделать умозаключения по изображению, опираясь на следующие свойства полученного изображения:

- цвет. Разные объекты могут иметь разную цветовую гамму;
- освещение. Границы объектов возможно определить по неоднородности освещения;
- текстура. Позволяет распознать объект по его визуальным свойствам.

На рисунке 1 отдаленные объекты вполне различимы, так как имеют отличную текстуру от текстуры поверхности воды и неба, хотя кажутся слитыми воедино;



Рисунок 1 – Картина «Впечатление. Восходящее солнце» (К. Моне, 1872 г.)

- контекст. СТЗ способна получить информацию об объекте из окружающей его обстановки. Сложно понять, что за объект расположен на рисунке 2.



Рисунок 2 – Неопознанный объект

Но если посмотреть на данный объект в контексте на рисунке 3, то можно с высокой уверенностью предположить, что этот объект – пистолет в руке дуэлянта;



Рисунок 3 – Картина «Дуэль Онегина и Ленского» (И.Е. Репин, 1899 г.)

- интерпретация объекта. Необходимо учитывать тематику предназначения объекта.

Двумерное изображение не позволяет получить полную информацию об объекте. Реальность искажается, так как линии объекта передаются в СТЗ с погрешностью.

Выше в работе рассмотрены основные факторы, способствующие сложности снятия информации с объектов. В свою очередь, данные факторы делятся на внешние и внутренние. Существуют довольно простые методы выделения интересующих объектов из сцены независимо от внешних факторов:

- положение камеры,
- внутренние свойства камеры,
- освещение.

Внешние факторы не зависят от свойств объекта, а включают характеристики устройства и навыки оператора.

Из внутренних факторов объекта, влияющих на качество получения информации, следует выделить:

- деформация объектов,
- внутриклассовая изменчивость.

Одним из основных методов снижения степени влияния факторов является введение ограничений, упрощающих задачи СТЗ:

- удобный ракурс съемки;
- освещение (в случае проведения съемки в помещении);
- типы объектов.

Существуют сцены, объекты с которых СТЗ распознает с довольно высоким качеством, независимо от внешних и внутренних факторов. Примеры простых задач для СТЗ:

- фиксация клеток крови;
- классификация монет и купюр;
- распознавание государственных номерных знаков на транспортных средствах;
- подсчет простых предметов.

Перед решением задачи определения атрибутов личности вводное изображение должно быть улучшено путем предобработки, в целях создания максимально идеальных условий для распознавания.

1.2 Задачи компьютерного зрения

Диапазон задач компьютерного зрения охватывает практически все сферы жизнедеятельности человека. В научной литературе нет однозначного понятия задачи компьютерного зрения. Одни авторы сужают задачу компьютерного зрения до одной цели, а другие, наоборот, включают в задачу смежные способы получения информации об изображении.

Так, Розенфельд А. считает, что работа с числовым массивом (матрицей), полученным из изображения, не является его обработкой [18], а значит не может входить в задачу компьютерного зрения.

Рейнхард Клетте дает общее понятие, что «задача компьютерного зрения — это применение фото- и видеокамер для анализа или понимания реальных сцен» [8].

Ряд авторов расширяют применение компьютерного зрения, включая в его задачи калибровку и самоориентацию приемников информации, обнаружение объектов, слежение за ними, работа с трехмерным пространством, высокоточные измерения объектов сцены, описание сцены и идентификация объектов, организация зрительной обратной связи [2].

Шапиро Л. под целью компьютерного зрения понимает формирование выводов относительно элементов на изображении посредством их анализа [24].

Селянкин В.В. классифицирует задачи компьютерного зрения в зависимости от их практического результата: обработка видеоизображения, сортировка деталей, обработка текста и другие [20]. При этом Селянкин В.В. не делает различий между специальными и общеприменимыми задачами компьютерного зрения.

Журавлев Ю. и Гуревич И. считают, что при получении вводного изображения перед СТЗ ставятся следующие задачи [6]:

- сопоставление. Полученное изображение сравнивается с имеющейся базой объектов и делается вывод о принадлежности изображения к определенному классу;
- поиск. На полученном изображении проводится поиск определенного фрагмента;
- восстановление. Собирается изображение из фрагментов;
- классификация. Из изображения извлекаются классы информации.

С утверждением А. Розенфельда в части того, что работа с числовым массивом не является обработкой изображений, нельзя согласиться. Именно с помощью вычислительных операций над матрицей цифрового изображения возможно его улучшить и отфильтровать. Тем самым, изображение предварительно обрабатывается с целью дальнейшей работы с ним на более сложных алгоритмах.

Понимание задачи компьютерного зрения только как необходимости достижения определенного результата (сопоставление, поиск, восстановление, классификация) также не может являться полным и многогранным.

Цель (задача) компьютерного зрения в общем смысле – получение определенных (полезно значимых) выводов из внешнего визуального источника информации с помощью математических алгоритмов, реализуемых программными средствами. Уже в зависимости от конкретной ситуации задача будет конкретизирована: задача приема изображения, задача поиска объекта на изображении, задача классификации и так далее.

Таким образом, под решением любой задачи компьютерного зрения следует понимать процесс, состоящий из комбинации программных методов и алгоритмов (подпроцессов), подбираемых в зависимости от требуемого результата и приводящих к достижению конкретной практической цели (задачи).

1.3 Актуальность задачи определения атрибутов личности по изображению лица

Автоматическое определение атрибутов личности с помощью компьютерного зрения является одной из наиболее актуальных задач, решаемой для множества сфер жизнедеятельности. Безопасность, розничная торговля, маркетинг, производство одежды, психология... Этот список будет

только расширяться, так как человек является постоянным объектом в фото- и видеофайлах.

Самые важные атрибуты человека, которые на практике необходимо распознать с помощью методов компьютерного зрения [10]:

- бинарная категория (мужской и женский пол);
- множественная нечисловая категория (раса: европеоидная, монголоидная, негроидная);
- числовая категория (возраст).

Конечно, существуют и другие атрибуты: эмоции, степень усталости, болевые ощущения [29]. Однако нынешние научные разработки находятся только в начале пути к такой тонкой идентификации.

Распространенные идеи практического применения автоматического определения атрибутов личности по изображению лица:

- оборот товаров, имеющих возрастные ограничения;
- реклама индивиду релевантных товаров и услуг;
- соблюдение правил доступа несовершеннолетних лиц к информации;
- поиск лиц, находящихся в розыске;
- адаптивность программного обеспечения («умное управление»).

Данные задачи целесообразно рассмотреть в соответствии с направлениями жизнедеятельности человека:

- розничная торговля. Современное общество в своем развитии поднимается на ступень цифровизации. Это проявляется во всех отраслях. Согласно прогнозу исследовательского агентства Data Insight, средний рост Интернет-торговли в нашей стране с 2019 по 2024 год составит 33,2%. Тем самым, оборот денежных средств в электронной коммерции, как показано на рисунке 4, вырастет с 1,7 до 7,2 триллиона рублей (более, чем в 4 раза) [5].

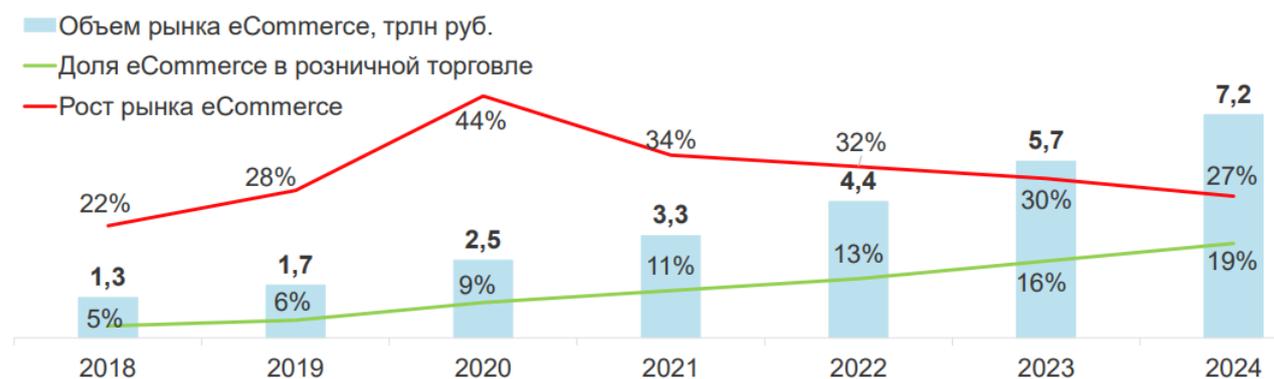


Рисунок 4 – Прогнозируемый рост интернет-торговли

Конечно, в основном продажи касаются непродовольственных товаров, которые реализуются сетями крупнейших игроков рынка (Wb, Ozon, Aliexpress, Beru, Goods, Аптека.ру). Вместе с тем, согласно сведениям Агентства инноваций города Москвы, оборот фудтеха (производства, приготовления и доставки еды с использованием IT-решений) в России в 2020 году достиг 120 млрд рублей (данные за 2021 год еще обрабатываются). Для сравнения, в 2019 году этот показатель был на уровне 44 млрд рублей. Такой высокий взлет связывается, прежде всего с пандемией.

Совсем скоро продукты питания встанут в один разряд с товарами общего пользования и на таких же условиях будут продаваться на электронных торговых площадках.

Законодательством установлены возрастные ограничения на некоторые группы изделий. С помощью компьютерного зрения возможно ограничить реализацию таких товаров несовершеннолетнему. Перед продажей товара можно проверять, кто именно осуществляет заказ, и при необходимости заблокировать сделку;

- маркетинг. На рекламирование продукции законодательством также наложены ограничения. Это касается вредных для здоровья продуктов. Наружная динамическая реклама, а также индивидуальная реклама должна быть избирательной. Методы,

встроенные в рекламные щиты, могут выявлять возрастные ограничения среди потока людей и передавать соответствующую правовым нормам информацию.

Эффективным маркетинговым ходом будет подача информации в зависимости от эмоционального состояния людей, их пола, расы и возраста;

- киноиндустрия. В Российской Федерации принята маркировка видеоматериалов по возрасту: 0+, 6+, 12+, 16+, 18+. Ограничить доступ к просмотру видеоконтента лиц, не достигших необходимого возраста, можно не только на этапе прохода в кинозал, но и в процессе приобретения билетов через онлайн-приложения.

Кроме того, автоматическое распознавание атрибутов личности применяется при подборе актеров в театре и кино для выполнения ролей, специфичных для определенного возраста;

- безопасность. Лиц, совершивших правонарушения и находящихся в розыске, условно можно разделить на 3 категории:
- преступники, личность которых установлена,
- преступники, личность которых не установлена, но имеется изображение их лица,
- правонарушители, личность которых и ее идентифицирующие признаки не установлены.

Было бы эффективным при обращении человека за покупками, снятии им наличных денежных средств, получении любых услуг проверять его по базе МВД России «Розыск». Принцип действия такой идеи представлен на рисунке 5;

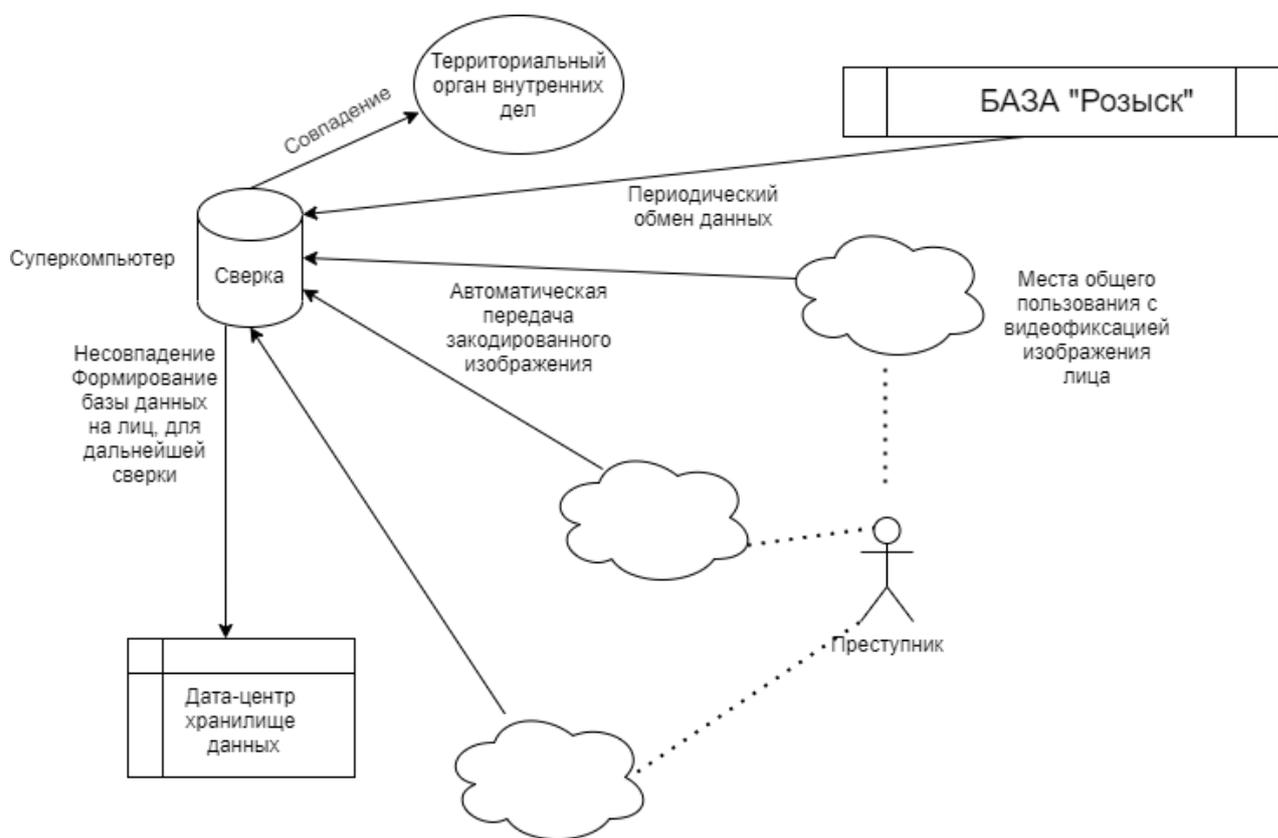


Рисунок 5 – Передача информации об атрибутах личности по распознанному изображению лица

- адаптивность программного приложения. Поведение системы с помощью обратной связи может подстраиваться под интересы конкретного человека. Здесь уже выявляются нетрадиционные атрибуты – эмоциональное состояние и реакции пользователя, которые нельзя однозначно выразить числовым или нечисловым выражением.

Таким образом, увеличивается спрос на интеллектуальные прикладные приложения, использование которых связано с извлечением биометрической информации из изображений или видео лиц. Использование таких приложений неразрывно связано с контролем доступа, взаимодействием человека и компьютера, наблюдением за участками местности.

Выводы по главе 1

В главе 1 магистерской диссертации представлен анализ современного состояния исследований задач компьютерного зрения, рассмотрена проблематика при реализации методов компьютерного зрения на практике, обусловлена актуальность решения задачи определения атрибутов личности по изображению лица.

Результаты проделанной работы позволили сделать следующие выводы:

- несмотря на цифровую трансформацию всех сфер жизнедеятельности человека компьютерное зрение является молодой наукой и требует дальнейшего углубленного изучения;
- эффективность методики компьютерного зрения в области распознавания объектов на изображении зависит от множества объективных и субъективных факторов;
- определение атрибутов личности по изображению лица – информационный процесс получения информации из изображения лица с помощью комплекса методов и алгоритмов;
- определение атрибутов личности по изображению лица – одна из важнейших задач компьютерного зрения, которая широко применяется и постоянно совершенствуется.

Таким образом, получено подтверждение теоретической и практической значимости проведенного исследования.

Глава 2 Анализ методов компьютерного зрения для решения задач определения атрибутов личности по изображению лица

2.1 Основные этапы определения атрибутов личности по изображению лица

С помощью методов компьютерного зрения по изображению лица возможно определить следующие атрибуты личности: пол, эмоции, возраст, расу и другие.

Работу СТЗ при решении такой задачи условно можно разделить на следующие этапы работы с изображением:

- формирование и представление;
- обработка;
- анализ;
- классификация полученных признаков.

Для каждого этапа характерны свои особенности и проблемы. Чтобы выполнить цель диссертации необходимо рассмотреть элементарные принципы действия СТЗ, систематизированные в современных научных трудах, сопоставив их с задачами определения атрибутов личности по изображению лица:

- формирование и представление. Для успешного проведения обработки и анализа изображений лица СТЗ необходимо как можно качественнее выполнить первый этап – формирование и представление. Это достижимо путем применения высокотехнологичного оборудования и создания условий, максимально приближенных к идеальным, в процессе захвата изображения. От этих факторов зависит точность и эффективность распознавания изображения программными алгоритмами.

Для примера взята аббревиатура «ТГУ» и с помощью программной утилиты преобразования изображений «можно посмотреть» на нее датчиком

с различным количеством пикселей. На рисунке 6 можно увидеть, что при увеличении размера пикселей и уменьшения их числа в матрице датчика точность фиксации изображения уменьшается вплоть до полной неопределенности.



Рисунок 6 – Фиксация изображения датчиком с разными размерами пиксела

В связи с погрешностью СТЗ и по объективным причинам входное изображение может быть получено темным, слабоконтрастным, шумным, нерезким, неравномерно освещенным или иметь неправильные цвета.

Основные проблемы возникают так как модуль воспринимающего устройства в зависимости от энергии освещения выдает тот или иной цифровой сигнал. Датчик имеет ограниченную техническими характеристиками чувствительность, что создает порог для восприятия яркости. Проблемы решаются на цикле предварительной обработки изображений.

Технология производства датчиков получения информации для СТЗ переживает бурный рост, поэтому не исключено появление их новых типов.

Большой перспективой обладает разработка технологий в области распознавания некачественных изображений. Вполне возможно, что в скором будущем компьютерный «глаз» сможет видеть в кромешной темноте в дождливую погоду и на большой скорости [7];

- обработка. Способов обработки изображений в современной прикладной информатике много. Клетте Р. предлагает

классифицировать операции по работе с изображениями в зависимости от выполненных преобразований на точечные, локальные и глобальные [8].

Улучшение изображения – процесс, при котором с входными данными проводятся операции с целью увеличения точности полученной цифровой информации. В результате для дальнейшей обработки используется измененное изображение. Улучшение изображения увеличивает вероятность обнаружения на нем важных деталей и объектов.

На вооружении СТЗ имеются ряд методов, которые могут существенно улучшить изображение: понижение шума, фильтрация изображения и выделение контраста между картиной и сценой. Следует отметить, что в ходе работы с изображением возможно скорректировать каждый его пиксель.

Этап обработки изображений важен для решения задачи компьютерного зрения. Как следует из примеров, информацию возможно извлечь даже из некачественного изображения. Только после детальной обработки изображения возможен его анализ СТЗ;

- анализ. После качественной обработки изображения СТЗ проводит анализ его участков. С помощью методов находят прямые и окружности, которые служат для распознавания конкретных образов в изображении [8]. Также выделяются грани (края) объектов изображения и угловые (особые) точки этих линий. Кроме особых точек имеются иные характерные черты изображения, учитываемые в ходе его анализа СТЗ.

На рисунке 7 обозначены характерные черты: точки, линии, области и структуры картины Рене Магритт. Эффективное распознавание таких частей изображения дает возможность провести анализ и точную классификацию объекта на сцене.



Рисунок 7 – Работа алгоритма на примере картины «Каникулы Гегеля» (Рене Магритт, 1958 г).

Исходя из задач СТЗ, анализируются следующие атрибуты входного изображения [2]:

- положение (вершины, концы и центры отрезков);
- геометрические (длина, площадь, расположение в пространстве, периметр, ширина линии, количество особых точек, диаметр);
- радиометрические (автокорреляция, яркость, контраст);
- текстурные (показатель однородности, моменты, результаты применения текстурных фильтров);
- топологические (соседство, связность, включение, параллельность);
- цветовые;
- динамические (статические и динамические свойства объекта);
- временные.

Методы компьютерного зрения позволяют решать задачи определения атрибутов личности по изображению лица с единых теоретических позиций. При правильно подобранной методике изменения параметров положения, поворота и масштаба изображения мало влияют на объем вычислений.

2.2 Роль машинного обучения в определении атрибутов личности по изображению лица

Классификация полученных признаков изображения – завершающий этап работы с информацией, который позволяет определить, то есть соотнести, атрибуты конкретной личности по атрибутам лица. Наиболее эффективно данная задача решается с помощью машинного обучения.

Машинное обучение и компьютерное зрение — это две смежные научные области, имеющие прикладное значение. Машинное обучение появилось практически с момента создания компьютера. Так, в 1958 году американский ученый Фрэнк Розенблатт создал вычислительный алгоритм «Марк-1». Компьютер мог обучаться в простейших задачах и был построен на перцептроне - нейронной сети, которую Розенблатт разработал тремя годами ранее [17]. Перцептрон – компьютерная модель восприятия информации мозгом. Фактически «Марк-1» имитировал взаимодействие человеческого глаза и мозга.

Машинное обучение – метод ИИ, целью которого является не прямое решение задачи, а обучение за счет применения ранее полученных решений подобных задач. Подход, при котором прошлые данные или примеры используются для первоначального формирования и совершенствования схемы предсказания, называется методом машинного обучения (Machine Learning).

В зависимости от методики программных алгоритмов машинное обучение может распознавать объекты или только классифицировать их.

В первом случае обучение компьютера проводится «без учителя». Входными данными является группа различных объектов, никак не разделенных между собой. Система самостоятельно определяет классы, основываясь на признаках объектов. Чтобы выполнить эту задачу компьютеру необходимо сформулировать вывод-правило, согласно которому объекты одного класса отличимые от остальных.

Во втором случае человек взаимодействует с нейросетью как учитель с учеником. Учитель определяет классы, а интеллектуальная система ищет их общие свойства. Получившийся вывод компьютер тестирует на новых примерах и таким образом обучается классифицировать элементы.

В таблице 1 указаны отличия данных методов.

Таблица 1 – Отличие самостоятельного машинного обучение от обучения «с учителем»

Машинное обучение				
с учителем		самостоятельное		
предсказать категорию	предсказать значение	разделить по схожести	найти зависимости	выявить последовательности
Классификация (разделить объекты по цвету)	Регрессия (разделить объекты по длине)	Кластеризация (разделить объекты по признакам)	Уменьшение размерности (взять из нескольких объектов общие признаки)	Ассоциация (принадлежность объекта другому)

В научной статье Горячкин Б.С. и Китов М.А. делят алгоритмы и методы компьютерного зрения на две большие группы [4]:

- классические. Алгоритм, «не думая», извлекает необходимую информацию из изображения, а на дополнительных этапах анализирует ее;

- системы машинного обучения. Такие методы имитируют человеческое мышление и, за счет сложности, требуют значительных вычислительных ресурсов.

Кроме запоминания информации СТЗ должна уметь делать выводы на новых данных по обучающим примерам. Этого можно добиться, используя машинное обучение.

Как бы не был идеально обучен классификатор, он не будет работать безупречно. В ходе машинного обучения учитываются допущенные ошибки первого и второго рода. Ошибка первого рода – когда искомый объект будет пропущен. Ошибка второго рода – вторичный класс (фон) будет принят за искомый объект.

Для оценки классификаторов количественный показатель ошибок как правило не учитываются, а важны точность и полнота. Точность – доля искомым объектов X среди всех объектов, классифицированных СТЗ как объекты X. Полнота – доля правильно распознанных объектов X среди всех искомым объектов X.

На представленном рисунке 8 точность распознавание данных методом составила 8 из 12 (66%), а полнота – 8 из 10 (80%).

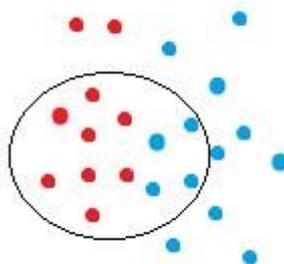


Рисунок 8 – Точность и полнота распознавания

Интегральный показатель классификатора высчитывается по следующей формуле: $F = \text{произведение точности и полноты}$ к сумме

точности и полноты. Чем больше интегральный показатель, тем классификатор лучше распознает.

Также при помощи экспериментальных показателей точности и полноты строятся ROC-кривые, которые позволяют наглядно выбрать параметр с оптимальным соотношением ошибок.

Большинство методов допускают возможность регулировки параметра (например, регулировка баланса), из-за которого соотношение ошибок может меняться.

2.3 Метод опорных векторов

Метод является прикладным алгоритмом и эффективно решает такую задачу компьютерного зрения, как классификация изображений.

Решением задач классификации образов определяется, к какому классу принадлежит элемент из множества. Каждый объект j имеет набор признаков x_j , каждому объекту можно приписать метку класса u_j . Тогда всю известную информацию можно представить в виде матрицы: $X = \{x_{ij}\}$ и $u = \{u_j\}$. Тем самым каждому вектор-признаку x задается метка u (класс объекта). Каждое сочетание (x, u) будет являться прецедентом.

Данный метод называется обучением по прецедентам. Для выполнения задачи определения класса метод использует решающее правило (классификатор). Классификатор делит общее множество определенных признаков на сектора в зависимости от принадлежности к классу.

Для классификации используется функция потерь, которая выдает значение «0», если объект имеет признаки, позволяющие отнести его к определенному классу, или «1», когда таких признаков не установлено. Наиболее хороший результат метод опорных векторов дает, когда потери для поступающих новых данных минимальны. Чтобы добиться таких результатов необходимо подобрать правильные значения классификаторов, которые бывают двух видов:

- линейный классификатор применяется для линейно делимых элементов и классифицирует признаки классов на гиперплоскости. Возможно бесконечное число вариантов такого деления.

Для выбора линейного классификатора учитывается, что новые данные могут намного отличаться от имеющихся. Чтобы минимизировать потери необходимо поставить одну линию (рисунок 9) так, чтобы расстояние (отступ, маркер) от ближайших к ней известных было максимальным.

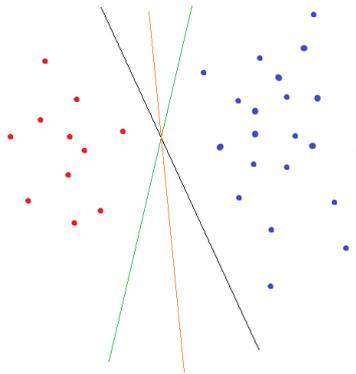


Рисунок 9 – Работа линейного классификатора

Эта задача решается методом множителей Лагранжа. На рисунке 2.4 оптимальному решению разделения признаков соответствует оранжевая линия. Опорные вектора – наиболее близкие к классификатору точки (известные признаки);

- когда элементы невозможно поделить линейно, применяются нелинейный классификатор. На рисунке 10 слева показан набор признаков изображения, который неравномерен, поделить его отрезком невозможно. Как вариант возможно возвести в квадрат имеющиеся значения, чтобы получить нелинейный двумерный график и классифицировать элементы розовой линией.

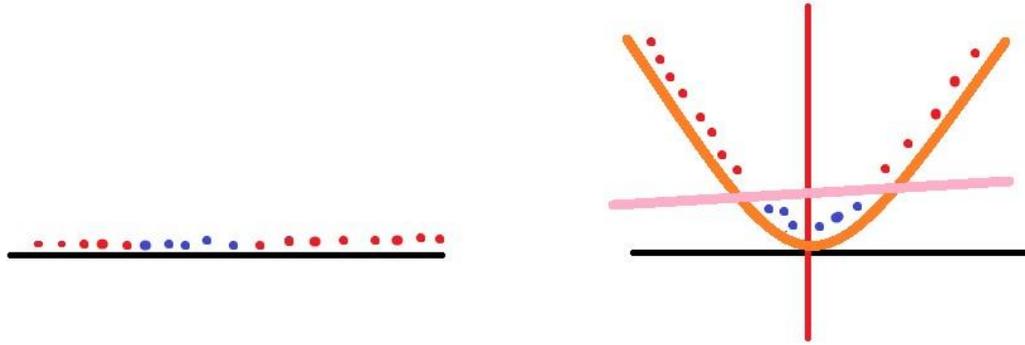


Рисунок 10 – Разделение сложных данных с помощью нелинейного классификатора

В ходе решения задачи определения атрибутов личности по изображению лица метод опорных векторов определяет гиперплоскость признаков, с помощью которой изображение лица отделяется от остальной области изображения. Как правило с этой целью применяется нелинейный классификатор.

Преимущество использования метода опорных векторов в том, что он эффективен при обработке Big Data, так как использует подмножество опорных векторов в функции принятия решений, являясь универсальным методом.

Вместе с тем такой сложный метод забирает много вычислительной мощности. Увеличение длины вектора многократно удлиняет время работы алгоритма. Эта проблема частично решается применением ядерной функции, которая отлично справляется с задачей классификации в одномерном пространстве и не увеличивает время исполнения алгоритма.

2.4 Кластеризация изображения

Кластеризация – процесс разбиения множества векторов-признаков, полученных из изображения на подмножества – кластера [24]. Такие элементы

относятся к разным классам и отличаются по своим признакам. Задача кластеризации заключается в разбиении этих объектов на группы на основе их сходства (общего признака).

Так, на рисунке 11 пиксели возможно условно сгруппировать на три кластера. Каждая группа сформирована на основе сходства определенного признака

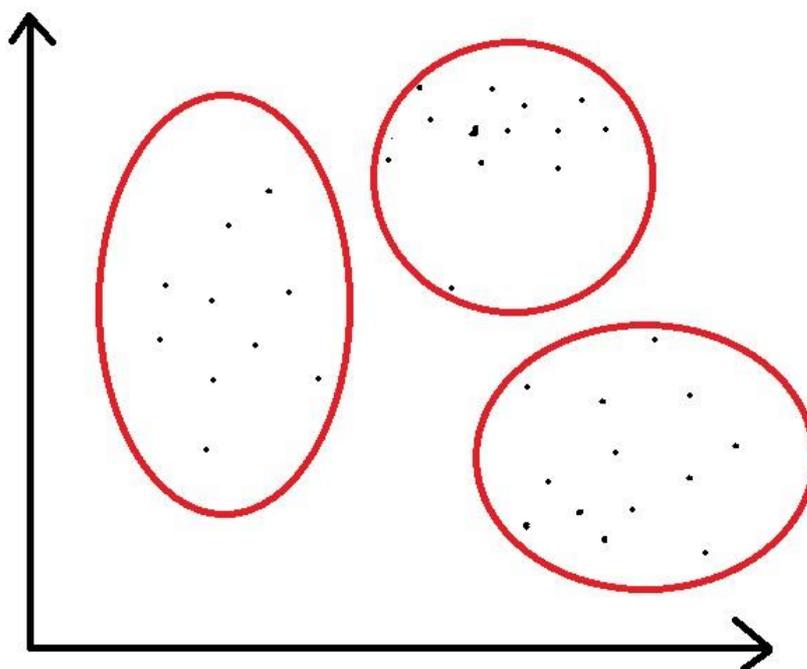


Рисунок 11 – Метод кластеризации

Кластеризация применяется для распознавания, для разделения наборов изображений, а также для организации и навигации [22]. Далее рассмотрены основные методы кластеризации.

2.4.1 Метод k-средних

Алгоритм работает следующим образом. Дано число n пикселей, которые необходимо разбить на k кластера. Определяются центры $k - g$ и к каждому k относят наиболее подходящие пиксели n . В качестве меры близости используется Евклидовое расстояние между пикселем n и g .

Определяется минимум суммы квадратов по формуле $D(k, g) = \sum_{k \text{ (кластеров)}} \sum (n - g)^2$.

Чтобы решить данную задачу, сначала произвольным образом выбираются центры g . Затем считаются расстояния между каждой n до ранее выбранных g . Каждый n соотносится с ближайшим к нему g . Затем g меняется и пересчитываются все n и так до тех пор, пока значения n не перестанут меняться.

На практике в качестве критерия для кластеризации применяется разбиение изображения по яркости или цвету. Из недостатков метода k -средних выделяют большую требуемую вычислительную мощность [28] и временные затраты на алгоритм. Кроме того, результат алгоритма зависит от начально выбранных g и не учитывает строение кластеров.

При квантовании k -средними также возникают трудности. Чем больше создается кластеров, тем меньше теряется информации. Тем самым, увеличивая число кластеров, возможно добиться большей точности. Однако увеличить количество кластеров по объективным причинам бывает невозможно, так как увеличится сложность алгоритма и время на его исполнение.

2.4.2 LSH (локально-чувствительное хеширование)

В процессе применения данного метода множество векторов случайно разбивается линией на две примерно равные плоскости. Векторам, находящимся в первом пространстве, присваивается код 0, а во втором – 1. Таких линий проводится несколько (рисунок 12). Соответственно каждой области пространства соответствует битовый код. Вектора с одинаковым битовым кодом находятся в одном пространстве [26]. Так выводится функция хеш. При увеличении областей хеш удлиняется, точность повышается, но скорость вычислений замедляется.

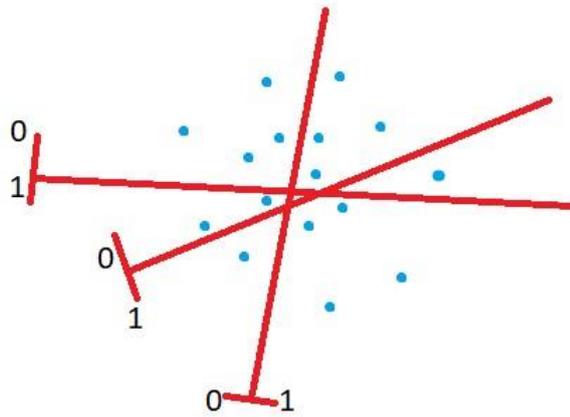


Рисунок 12 – Метод LSH в действии

Чтобы повысить точность без ущерба для скорости методы k-средних и LSH применяются в комбинации. Так, ячейки после кластеризации имеют большие размеры и элементы внутри них отличаются друг от друга. Сами ячейки разделяются с помощью метода LSH. Для каждого кластера вычисляется свой бинарный код, а каждый вектор помимо номера ячейки определяется своей бинарной подписью, тем самым кодируется расстояние между центром ячейки и местоположением вектора. В результате положение вектора внутри ячейки кодируется в хеш. Идентификатор изображения и его бинарная подпись хранятся в памяти RAM и не занимают много места, поэтому алгоритм работает достаточно быстро.

2.5 Метод главных компонент (PCA)

Данный метод позволяет уменьшить размерность данных, не допуская при этом потерю информации об изображении. Даже маленькое по размеру изображение имеет большое число измерений. При увеличении разрешения геометрически увеличивается количество измерений и усложняется алгоритм. Метод делает акцент на доминирующие направления.

Изображения лиц представляются в виде набора главных компонент – собственных лиц. Такие лица соответствуют определенным векторам.

Главные компоненты вычисляются путем нахождения собственных векторов и собственных значений ковариационной матрицы, полученной из изображений. Теперь можно получить реконструкцию изображения, которая является произведением суммы главных компонент и на соответствующие собственные вектора. Для каждого изображения вычисляются до 200 главных компонент.

При распознавании СТЗ сравнивает главные компоненты неизвестного изображения с компонентами всех известных изображений. В ходе решения задачи определения атрибутов личности по изображению лица изображения, соответствующие одному человеку, группируются в кластеры и из базы данных подбираются наиболее близкие к ним образцы [11].

Тем самым для решения задачи распознавания лиц вычисляется расстояние от проекции тестового вектора до среднего вектора тренировочного набора и расстояние от тестового вектора до его проекции в подпространство главных компонент. Согласно полученным данным, делается вывод о принадлежности объекта на изображении к конкретному атрибуту.

В машинном обучении PCA также используется для снижения размерности вектора-признака при условии сохранения важной информации. Данный метод целесообразно использовать для исключения переобучения модели.

Например, чтобы выразить на языке Python метод главных компонент необходимо прописать следующий алгоритм:

- открыть окно изображения;
- измерить изображение;
- создать матрицу со строчками изображений;
- применить метод главных компонент и после обработки матрицы из п. 3 получить на выходе матрицу проекции. Сначала данные центрируются, по каждому направлению вычисляется среднее. Находятся собственные вектора, соответствующие получившимся

значениям ковариационной матрицы. Функция `range()` проводит замену значений на целые числа;

- вывести на экран среднее из всех изображений и первые 5 мод, имеющие наибольшую дисперсию.

Для наглядного тестирования алгоритма применялись 462 изображения формата «.png» буквы русского алфавита «Ж», выполненные в распространенном графическом редакторе (рисунок 13).

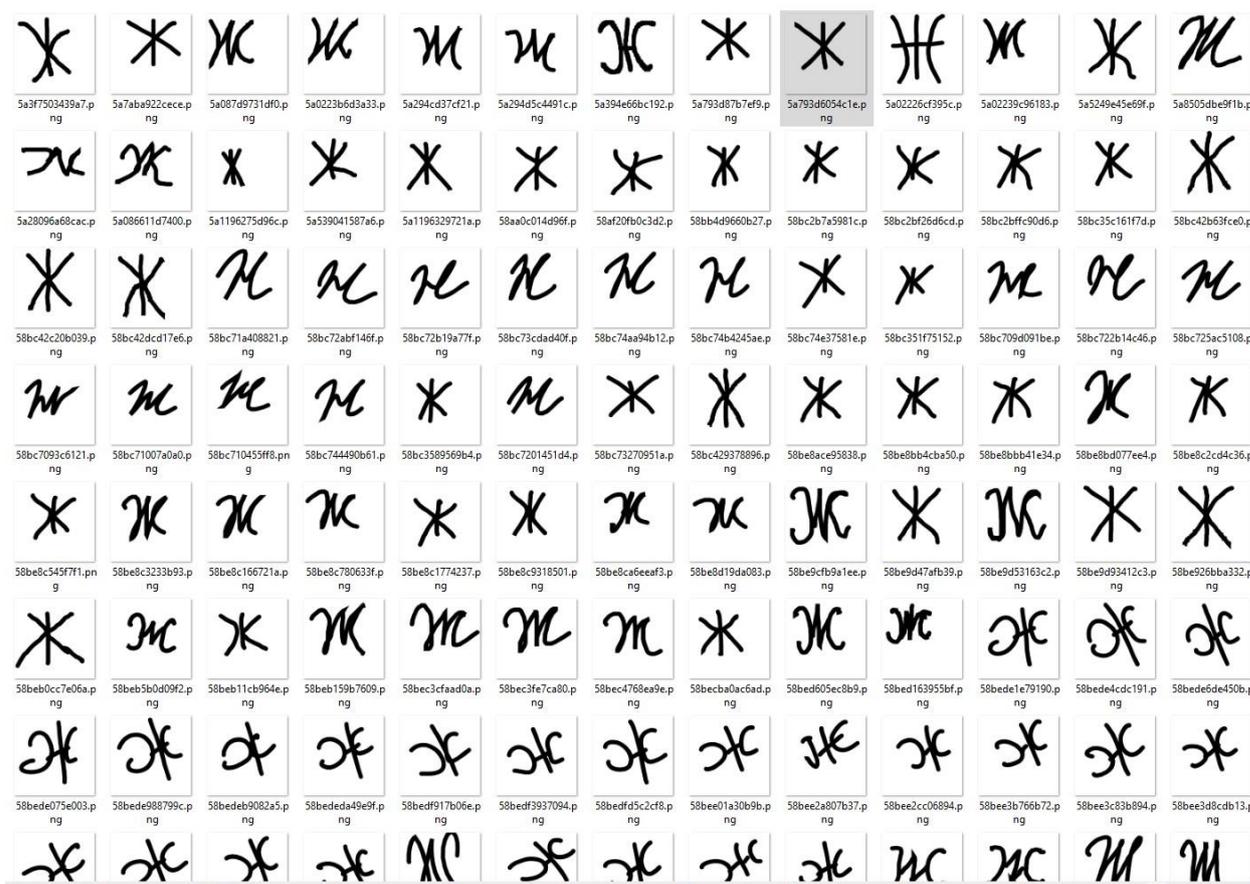


Рисунок 13 – Изображения буквы «Ж» для тестирования алгоритма PCA

В результате работы алгоритма (рисунок 14) на выходе с помощью функции `subplot()` получается усредненное изображение буквы Ж (выделено красной рамкой), а также 5 мод распределения изображений.

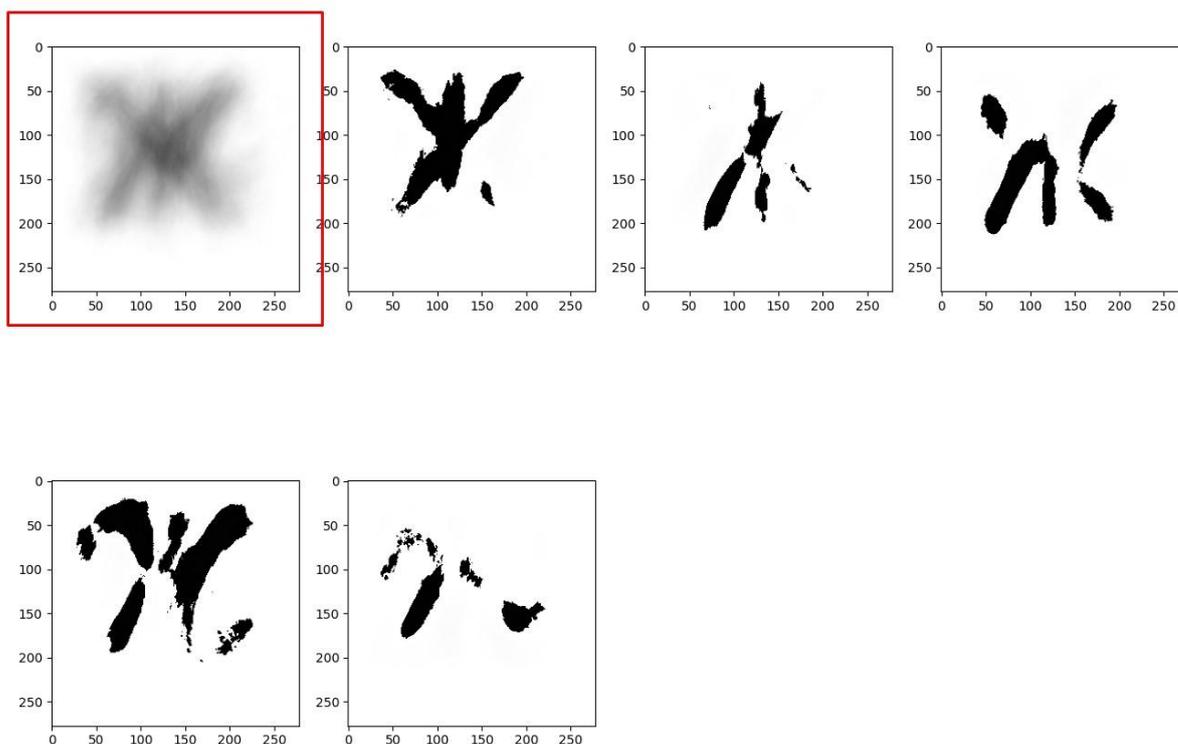


Рисунок 14 – Результаты работы PCA

Однако метод главных компонент не учитывает латентную структуру переменных, возможна потеря уникальных, не вписывающихся в среднее значение данных. Кроме того, при изменении масштаба изображения, освещенности метод осуществляет ложные срабатывания и делает неправильные выводы.

Таким образом, метод главных компонент является одним из мощных и универсальных средств анализа, который, отбрасывая некоторые признаки, позволяет учитывать лишь наиболее значимые комбинации их значений [13].

2.6 Метод Виолы-Джонса

На сцене можно выделить две категории элементов: объекты, имеющие определенные размер, и формы, материалы, имеющие однородный и повторяющийся шаблон мелких деталей без определенного размера.

Скользящее окно – подход, при котором изображение разбивается на части. Рамка сканирует изображение и анализирует его частями. Метод аналогичен разбивке изображения на множество прямоугольников.

Метод сопоставления шаблонов является простейшим классификатором. Чтобы применить его берется шаблон с искомым элементом. Скользящее окно «проходит» по изображению и каждый фрагмент сравнивается с шаблоном. Результатом работы метода будет карта отклика всего изображения, по которой можно сделать вывод, имеется ли на изображении объект (рисунок 15).



Рисунок 15 – Натюрморт с тарелкой лимонов (Гарцони Джованна, 1645-1650). Скользящее окно по шаблону ищет объект «осу» на изображении

Однако такой классификатор не сможет учитывать объекты, если они находятся под другим ракурсом или имеют другую расцветку.

Кроме того, следует выделить проблемы метода скользящего окна:

- размер окна – однотипные элементы на изображении не всегда одинаковы по размеру и могут не уместиться в окне. Проблема решается использованием окон разного размера или сканированием пирамиды изображения (окно одного размера проходит по изображениям разных размеров);

- пропорции объектов разные и по этой причине могут ошибочно определяться классификатором. Чтобы снизить ошибку применяются окна разных пропорций;
- множество откликов. Верстка определяет несколько откликов на сцене, где может быть изображение. Для преодоления этого ложные сигналы не учитываются путем подавления значений ниже экстремумов;
- объекты накладываются друг на друга. Является самой сложной проблемой, решение которой в разработке.

В человеческом мозге имеются группы клеток, каждая из которых занимается анализом только определенно расположенных краев и градиентов изображения. Так, после получения информации об изображении в мозге активируются те или иные участки в зависимости от ориентации линий изображения. В результате человек делает вывод об изображении.

СТЗ имеет похожую технику распознавания. Так, сначала оценивается градиент каждого пикселя. Градиент изображения – направление максимального изменения яркости и показатель изменения яркости (сила градиента). Полученные данные используются для создания гистограммы ориентаций градиентов.

Тем самым, статистика и выводы делаются по окрестности пикселей изображения. Если изображение будет немного сдвинуто или повернуто в отличие от шаблонного, СТЗ все равно сможет решить задачу распознавания, используя линейную комбинацию сложных шаблонов из тестовых изображений.

Для классификации сложных изображений в реальном мире разрабатываются высокоуровневые методы и алгоритмы:

Детектор Виолы-Джонса, предложенный в 2001 году, способен найти изображение в реальном времени. Обучение метода проводится долго, но поиск ведется достаточно быстро. Классификатор быстро вычисляет простые

признаки объектов. Так, для распознавания лица учитываются два смежных прямоугольника, лежащих на глазах и щеках (признаки Хаара) [41].

Для выделения объекта впервые был применен бустинг для выбора признаков. Бустинг позволяет выстроить множество простых классификаторов, собрав из них сильный («комитетный») классификатор.

Кроме того, каскадным методом фильтруются ложные срабатывания. Так, использованием простых классификаторов отбрасываются отрицательные окна, в которых не обнаружен элемент. На отфильтрованных таким образом положительных примерах применяется более сложный классификатор. Затем запускается третий еще более сложный классификатор и так далее, каскадом. Если элемент не был отброшен каскадными классификаторами, то делается вывод, что он искомым.

В результате каскадный фильтр позволяет сочетать в детекторе точность и скорость распознавания. Как правило, в результате работы алгоритм выделяет на сцене человеческие лица (рисунок 16).



Рисунок 16 – Сестры Шарлотта и Зинаида Бонапарт (Жак-Луи Давид, 1825).
Пример работы детектора Виолы-Джонса

Машинное обучение в методе Виолы-Джонса решает такую задачу как классификация.

Недостатки метода Виолы-Джонса:

- фильтры Хаара не являются самыми важными признаками изображения. Детектор не учитывает градиенты;
- точность распознавания детектора оказывается недостаточной на практике;
- метод дает множество откликов у правильных обнаружений, чем замедляется скорость работы;
- дискретность операции предполагает ее завершение в конце алгоритма;
- требуется долгое время на обучение.

Чтобы улучшить качество метода дополнительно учитываются цвета и градиенты изображения. По мере развития детектора предлагались следующие методы улучшения его эффективности: учет масштабов (использование изображения разного размера для подсчета признаков), мягкий каскад (возможность досрочно завершить каскадную фильтрацию путем более раннего выявления отрицательного примера), «шепчущие каскады» (функционирование приоритетного каскада за счет остановки других), работа с ракурсами (строятся обучающие модели с разными ракурсами искомого элемента), уточнение разметки (ложные срабатывания исключаются по ходу работы метода).

2.7 Нейросетевые методы

Модели сверточных нейронных сетей являются первым алгоритмом машинного обучения, который стали исследовать сразу же после создания компьютеров. Для понимания построения искусственной нейронной сети необходимо разобрать, как работает человеческий мозг.

Человеческий мозг состоит из отдельных клеток сложной структуры, которые связываются между собой в сеть. Каждый отдельный нейрон выглядит как клетка с отростками-дендритами, среди которых выделяется один длинный – аксон [15].

Время от времени нейрон вырабатывает электрический сигнал, который передает по аксону. При этом аксон является изолированным от внешнего пространства и может достигать значительной длины. Скорость сигнала может достигать до 100 метров в секунду [1]. После того, как электрический сигнал прошел, аксон вырабатывает вещество – нейромедиатор. Данное вещество воспринимается отростками других нейронов, отчего происходит процесс, в ходе которого импульсы начинают генерировать другие нейроны (рисунок 17). Так происходит человеческое мышление и именно так человек распознает объекты окружающего мира.

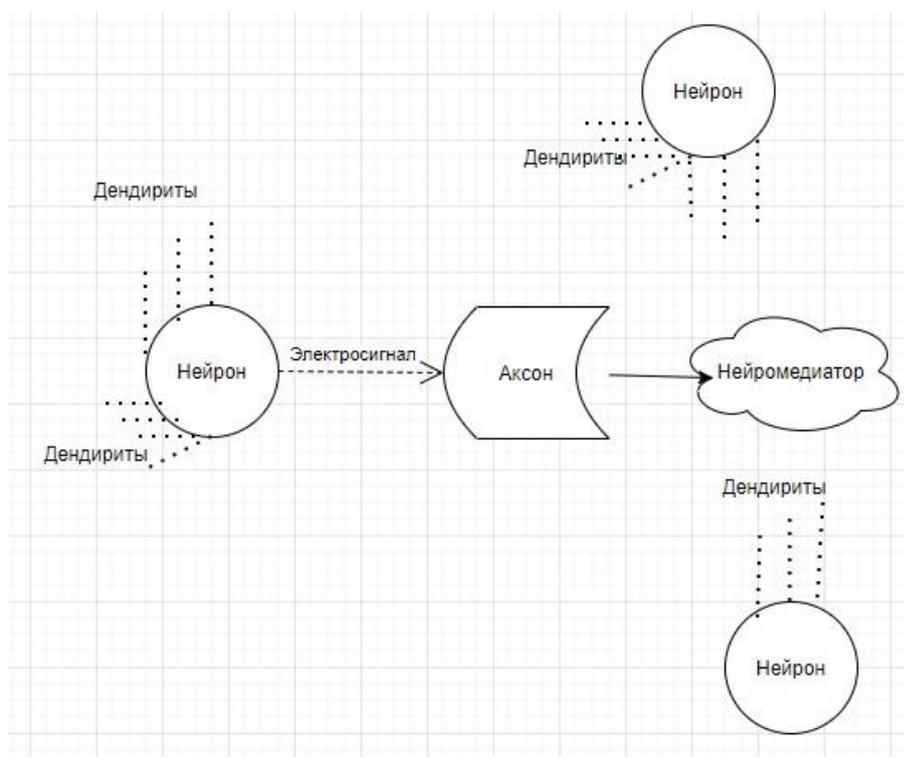


Рисунок 17 – Схема работы человеческого мозга

До сих пор учеными достоверно не установлено, как именно функционирует головной мозг человека и такое описание можно назвать лишь гипотезой. Вместе с тем, человек способен распознать изображение за один взгляд (100 миллисекунд).

Простая математическая модель нейрона была предложена учеными Мак-Каллоком и Питтсом еще в 1943 году [9]. Нейрон здесь представлен как линейная функция классификатора. На нейрон подаются базовые сигналы. Если сумма значений последних больше определенного порога, то нейрон возбуждается и подает ответ, если меньше, то нейрон бездействует (рисунок 18).

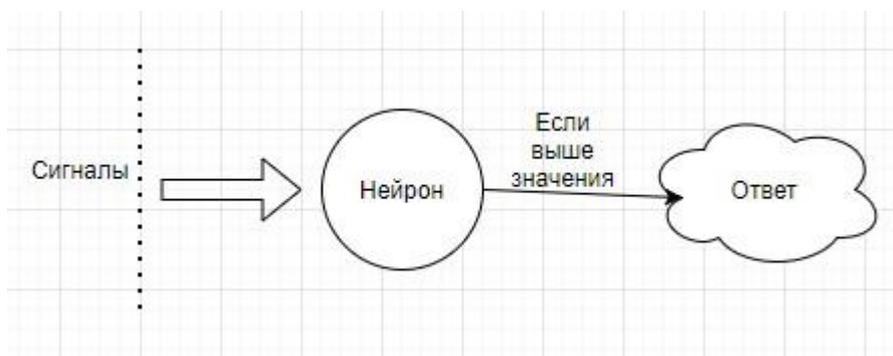


Рисунок 18 – Математическая модель нейрона Мак-Каллока-Питтса

В нейроне происходят определенные операции, то есть выполняется линейная классификация с целью принятия решения о подаче сигнала. Этот классификатор возможно настроить.

Поскольку нейрон – линейный классификатор, возможно подобрать такие его параметры, чтобы ошибка вывода была минимальна. Это возможно сделать с помощью градиентного спуска. Берется начальное понижение и меняются его параметры, чтобы суммарная ошибка стала минимальной. Каждый момент времени ошибка оценивается и определяется градиент ее изменения, то есть устанавливается в каком направлении менять параметры (делать градиентный шаг) в целях понижения ошибки [3].

К недостаткам метода относится то, что выборка не всегда должна быть линейно разделимой и неизвестен конец операций по понижению ошибки.

С помощью нейросети (комбинации нейронов) возможно представить любую непрерывную функцию с необходимой точностью. Однако конфигурация нейросети подбирается экспериментально. Каждая связь между нейронами имеет свой вес – параметр. Число нейронов и их связей определяет большое количество параметров.

Комбинация сверток и максимизации с нормализацией позволит реализовать все эвристические признаки с помощью нейросети. Сверточный слой сравнивает фрагмент изображением с шаблонами и элементом из словаря. В конце проводится максимизация данной области (рисунок 19). Слой из специальных нейронов будет нормализовать отклики в конкретной области изображения.



Рисунок 19 – Общая схема вычисления признаков изображения с помощью нейросетей

С такими задачами наиболее точно справляются специально обученные нейронные сети. Сверточные нейронные сети обеспечивают частичную устойчивость к изменениям масштаба, смещениям, поворотам, смене ракурса и прочим искажениям. Стоит отметить, что принципом работы нейросетей служит все та же свертка. Чем же отличается нейронная свертка от обычной?

В ходе действия нейронной сети ядра свертки определяются автоматически в процессе обучения. Сверточные нейронные сети состоят из чередующихся слоев свертки и подвыборки (усреднения изображения),

соединенных друг с другом. Тем самым, хорошо учитываются локальные особенности изображения.

При создании нейросети «с нуля» в нее включаются несколько слоев, подобранных в зависимости от конкретной задачи. После отработки алгоритм выдает обученную модель и показывает вероятность правильности классификации объектов по ее категориям.

Глубокое обучение (deep learning) — это подмножество методов машинного обучения, области изучения и создания машин, которые могут обучаться (иногда с целью достичь уровня искусственного интеллекта) [25]. Обучение модели нейросети занимает большое количество времени (часы и даже дни).

Недостатком применения нейросетей является необходимость ее индивидуального подбора и обучения для определенной задачи. Нельзя создать нейросеть на все случаи жизни.

Преимуществом нейронных сетей перед ранее рассмотренными методами заключается в способности захватывать более глубокие признаки в Big Data. В процессе обучения искусственные нейроны будут обобщать информацию, двигаясь от частного к общему. Тем самым, правильное обучение и последующее применение нейронных сетей позволяет эффективно решать задачи определения атрибутов личности по изображению лица.

Нейросетевые разработки в настоящее время все еще находятся на начальной стадии своего развития. Даже маленькие шаги в их освоении способны вызвать кардинальные мировые перемены.

В сети Интернет множество информационных ресурсов, на которых имеются миллионы изображений, распределенных по классам, которые используются для обучения нейросетей. Тенденция развития нейросетей в мире такова, что они становятся сложнее, в классификатор добавляются слои, нейронов становится больше, обучение нейросети постоянно усложняется.

В результате анализа по каждому методу компьютерного зрения в контексте решения задач определения атрибутов личности по изображению лица определены достоинства и недостатки (таблица 2).

Таблица 2 – Сравнение методов компьютерного зрения для задач определения атрибутов личности по изображению лица

Метод	Достоинства	Недостатки
Метод опорных векторов	Универсальный метод, позволяющий работать с Big Data	Длительное время срабатывания
Кластеризация	Алгоритм обучается очень быстро и при работе требует минимум вычислений	Результативность напрямую зависит от выбора значения k, непоследовательность работы метода
Метод главных компонент	Уменьшает размерность данных	Не всегда точный и может не учитывать детали
Детектор Виолы-Джонса	Быстрый поиск объектов в реальном времени	Долгое время на обучение, недостаточная точность
Нейросетевые методы	Захват глубоких признаков, обобщение результатов	Требуется индивидуальное обучение для конкретной задачи, длительность обучения

Выводы по главе 2

В главе 2 магистерской диссертации представлен анализ наиболее популярных методов компьютерного зрения в контексте решения задачи определения атрибутов личности по изображению лица: метод опорных векторов, метод главных компонент, детектор Виолы-Джонса, сверточная нейронная сеть. В таблице 2 выделены достоинства и недостатки каждого метода.

Результаты проделанной работы позволили сделать следующие выводы:

- единого подхода к решению задачи определения атрибутов личности по изображению лица в научной литературе не имеется. Каждый метод имеет как достоинства, так и недостатки;
- среди рассмотренных методов наиболее преимущественными являются нейросетевые методы, которые могут гарантировать достаточную точность для определения атрибутов личности даже при изменении объектом контроля своего местоположения. Длительная скорость обучения алгоритма в контексте решения текущей задачи исследования существенным недостатком не является. Качество работы указанных алгоритмов возможно регулировать с помощью тонкой настройки и подбора нейронных слоев. Кроме того, возможность захвата глубоких признаков изображения позволит получить приемлемую точность при решении задачи определения атрибутов личности по изображению;
- подходы к решению задачи определения атрибутов личности по изображению лица включают комбинации программных методов и алгоритмов (подпроцессов), подбираемых в зависимости от требуемого результата и приводящих к достижению конкретной практической цели (задачи). Поэтому для решения задачи определения атрибутов личности по изображению необходимо разработать алгоритм, включающий несколько методов в совокупности.

Таким образом, некоторые из рассмотренных методов послужили основой при разработке предлагаемых подходов, излагаемых в последующих главах.

Глава 3 Разработка алгоритма определения атрибутов личности с помощью сверточных нейронных сетей

При разработке алгоритма определения атрибутов личности следует учитывать, что методы должны решать задачу оценки пола и возраста в реальных условиях нахождения объекта, когда лицо может быть не выровнено, на сцене имеются проблемы с освещением и фоном.

Разработка методов компьютерного зрения для решения задач определения атрибута личности «пол» по изображению лица с помощью машинного обучения проходит следующие фазы (рисунок 20):

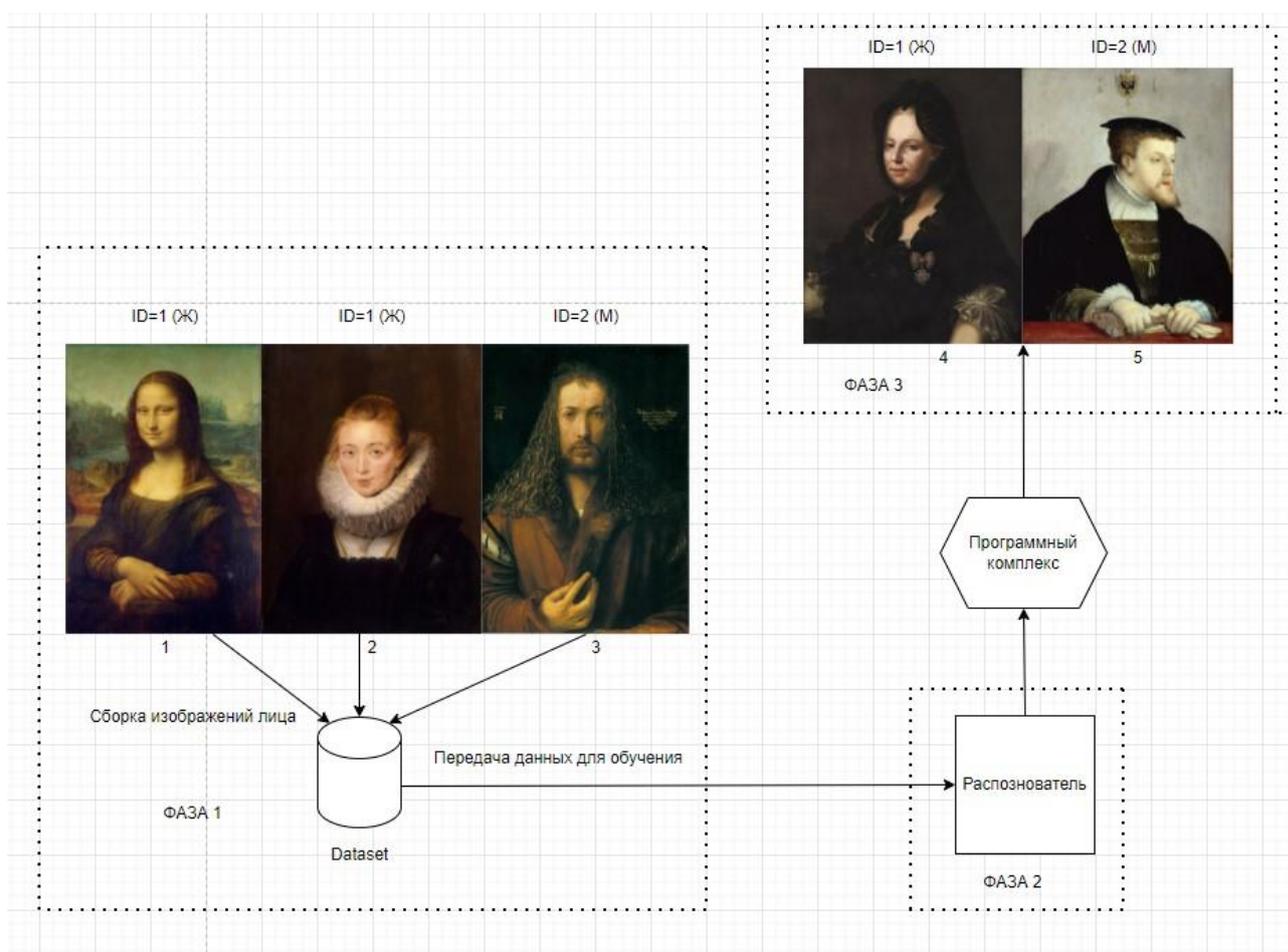


Рисунок 20 – Фазы разработки методов компьютерного зрения

- создание базы путем обработки имеющейся информации – тренировочных изображений,
- тренировка распознавания набором характеристик (признаков), полученных из входных изображений,
- распознавание путем точного определения и измерения черт лица по входному изображению с использованием модели, полученной после тренировочного процесса.

Для определения атрибута личности «возраст» по изображению лица предлагается алгоритм, построенный на базе метода глубокого ожидания (DEX) [39]. Суть методики заключается в обнаружении и распознавании лиц на изображениях при помощи глубокого обучения (рисунок 21).

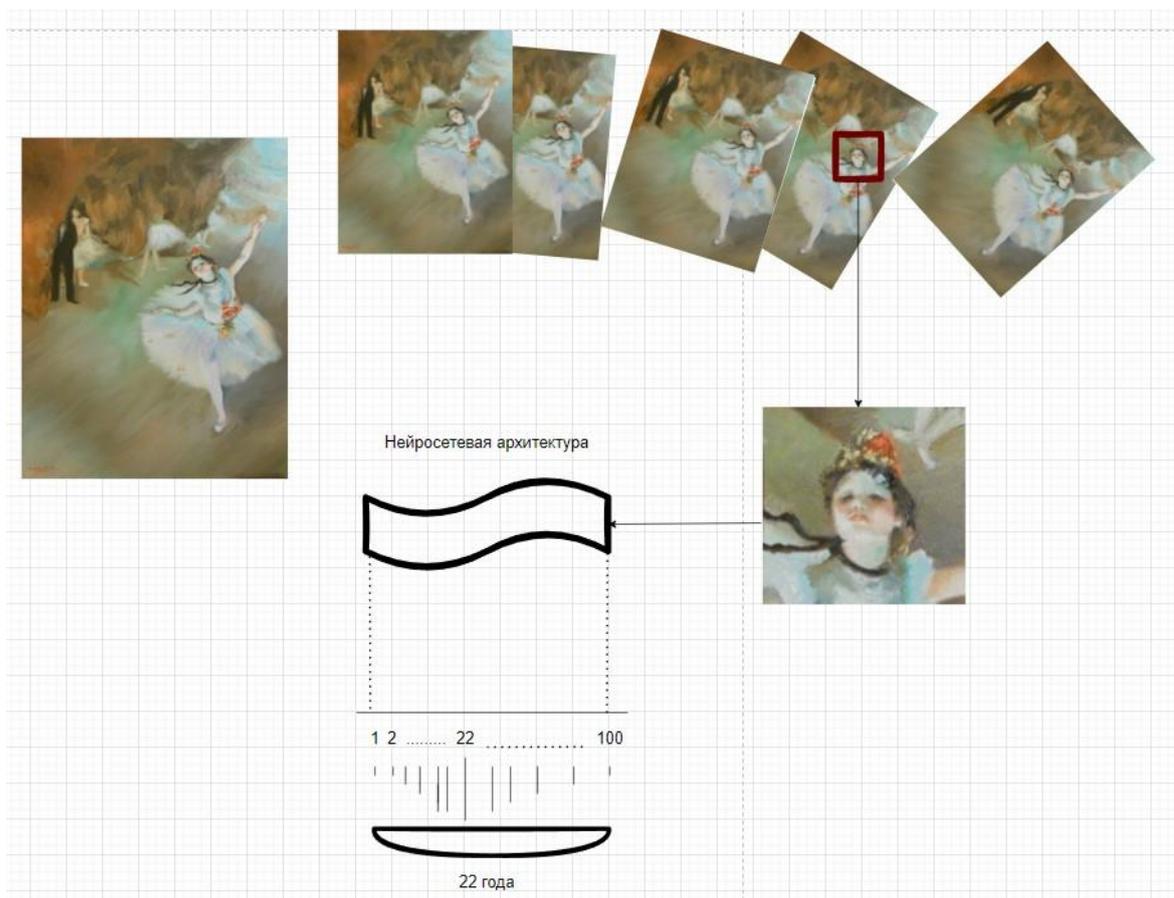


Рисунок 21 – Звезда балета (Э. Дега, 1876-1878 гг). Метод DEX

Чтобы качественно решить поставленную в исследовании задачу необходимо руководствоваться следующими принципами глубокого обучения:

- более глубокие нейронные сети имеют сложные параметры и способны значительно лучше моделировать нелинейные образования с требуемой глубиной [32];
- для качественного обучения требуется огромный и разнообразный набор тренировочных данных. Обученная модель будет более устойчивой и сможет точнее распознавать даже лица, имеющие редкие индивидуальные признаки.
- выравнивание лица на изображениях уменьшает вычислительную сложность алгоритма;
- процесс работы сверточной нейронной сети изучен не полностью и кроме теоретических предпосылок для ввода ее тонких настроек следует опираться на экспериментальные исследования ее поведения.

Для точного прогнозирования возраста и пола последовательно осуществляются этапы:

- создание обучающих данных. На данном этапе отфильтровываются шумовые данные, сериализуются изображения и их метки, создается база для последующего обучения. Таким образом, для улучшения качества обучения требуется отсеять изображения, не содержащие лиц или содержащие более 1 лица. Кроме того, отфильтровываются изображения лиц, пол/возраст которых не указан либо указан возраст вне установленного для исследования диапазона.

На выходе парсер получает из каждого изображения важную для дальнейшей реализации информацию: год рождения человека и дату фотоснимка, пол лица, размер изображения, а также иные данные;

- обнаружение лица на изображении. Для обнаружения местоположения лиц на обучающих изображениях запускается каскадный детектор лиц Dlib [34].

Данный метод использует алгоритм HOG, который обнаруживает особые точки, которые используются в компьютерном зрении и обработке изображений с целью распознавания лиц. HOG включает в себя следующие шаги:

- вычисление градиента,
- группировка направлений,
- блоки дескрипторов,
- нормализация блоков,
- SVM-классификатор (метод опорных векторов).

Кроме того, детектор дополнительно запускается на повернутом под разными углами тренировочном изображении и обнаруживает лицо в разных ракурсах. Это позволяет применить полученный метод к реальным жизненным условиям, в том числе в видеопоследовательности.

После обнаружения лицо берется в рамку, которая раздвигается во все 4 стороны на 40% шире от первоначальной величины. Часть изображения с лицом внутри рамки и будет являться входными данными для глубокой сверточной сети. Тем самым для обучения используются объекты лица, имеющие единый сопоставимый размер, выровненные и центрированные, с минимальным фоном;

- обучение нейронной сети.

Оценку возраста в методе DEX можно рассматривать как поэтапную регрессию путем глубокой классификации с выборкой значения в непрерывном диапазоне от 0 до 100 лет с последующим уточнением ожидаемого значения. Цель регрессии – выбор такой функции (модели) из множества всех моделей, использование которой будет минимизировать функцию потерь. Метод DEX позволяет свести к минимуму ошибку оценки возраста в данных проверки.

На блок-схеме представлен упрощенный алгоритм разработанного метода (рисунок 22).

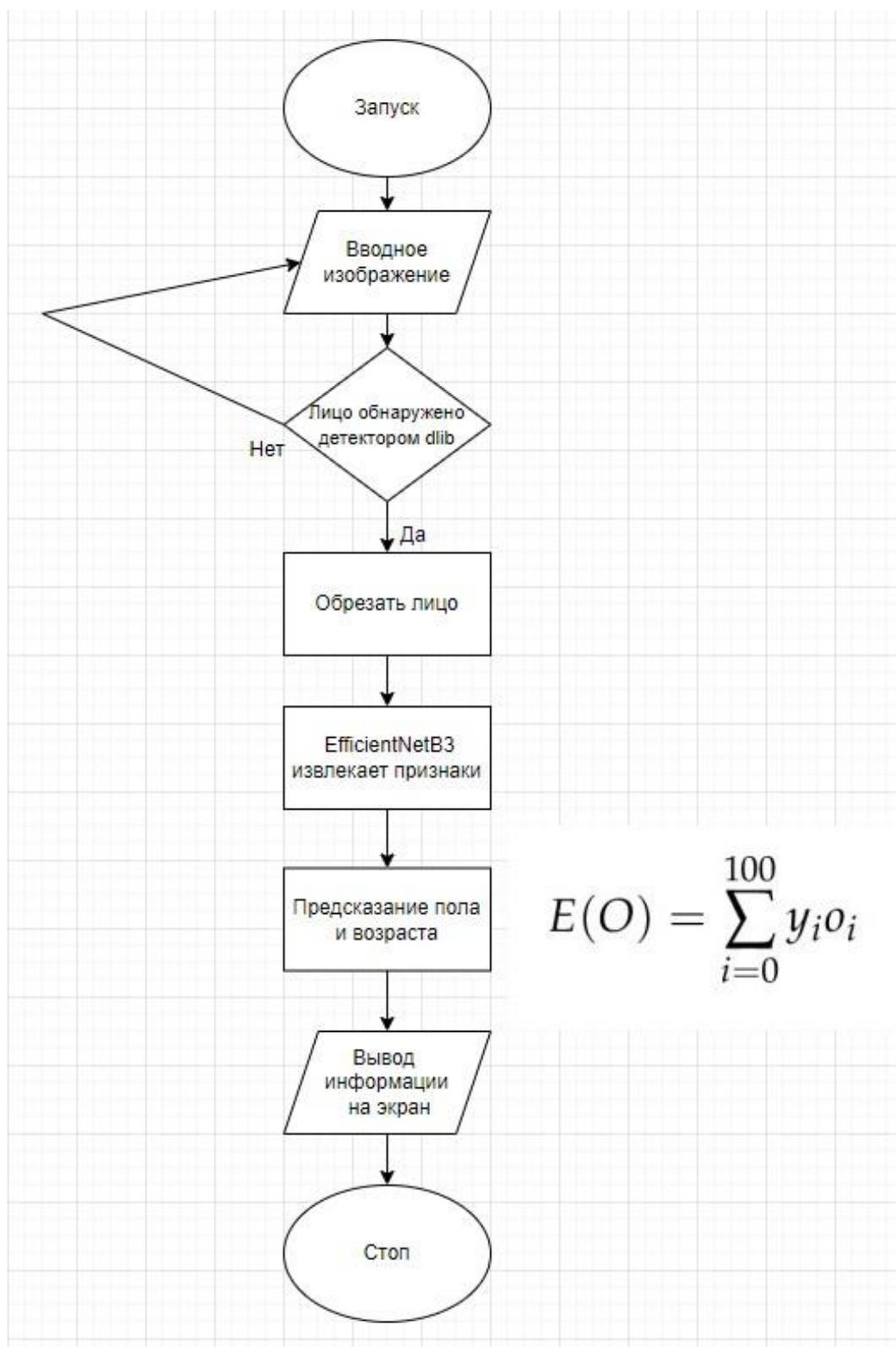


Рисунок 22 – Блок-схема разработанного метода на основе DEX

При этом u_i является вероятностью наличия i -возраста личности на изображении, а o_i отражает сам i -возраст. Сумма произведений каждого значения предсказания и соответствующего ему возраста будет определять итоговый прогнозируемый возраст.

Средняя абсолютная ошибка возраста вычисляется путем деления модуля разницы между прогнозом и реальным значением для всех объектов на количество данных объектов. При правильно подобранных параметрах нейросети обучение модели идет успешно и с каждой эпохой значение средней абсолютной ошибки снижается.

Оценка пола также проводится путем обучения модели по тренировочной выборке, имеющей n изображений лиц. Пусть U – одно изображение из n , y - пол человека (мужской или женский), а $F(U)$ будет являться функцией, вычисляющей признаки лица по одному изображению. Тогда, модель, полученная в процессе глубокого обучения, будет иметь вид: $y(F(U))$, то есть функцию и выявленный метод принятия решения для отнесения изображения к конкретному полу.

Сверточные нейронные сети обеспечивают частичную устойчивость к изменениям масштаба, смещениям, поворотам, смене ракурса и прочим искажениям. Стоит отметить, что принципом работы нейросетей служит все та же свертка.

Сверточные нейронные сети состоят из чередующихся слоев свертки и подвыборки (усреднения изображения), соединенных друг с другом. Тем самым, хорошо учитываются локальные особенности изображения.

При создании нейросети «с нуля» в нее включаются несколько слоев, подобранных в зависимости от конкретной задачи. На вход подается изображение в цветовой модели RGB определенного размера в пикселах.

Softmax является выходным слоем архитектуры нейросети и решает задачу предсказания возраста лица. Эта функция, превращает получившиеся наборы чисел в вероятности (принадлежности лица каждому возрасту в диапазоне $1 \dots 100$). Сумма всех вероятностей равна единице.

Нейронная сеть будет иметь один выходной узел для каждого непрерывного значения, которое пытается предсказать. Выводом алгоритма будет наиболее вероятный возраст лица.

После отработки алгоритм выдает обученную модель, которую возможно использовать для автоматического определения атрибутов личности по изображению лица на фотографиях или в видеопоследовательности.

Выводы по главе 3

В главе 3 предложен и изложен теоретический подход к решению задачи автоматического определения атрибутов личности по изображению лица, основанный на глубоком обучении с помощью сверточной нейронной сети. Распознавание пола человека по фотографии предлагается решать с помощью бинарной классификации, а предсказание возраста – путем регрессионного анализа.

Перед тренировочным процессом метаданные проверяются на допустимость и полноту, каждый объект контроля (лицо человека) обнаруживается с помощью детектора и центрируется в отдельном изображении. За счет этого алгоритм не будет иметь большую вычислительную сложность и позволит провести эффективное качественное обучение.

Технически обучение заключается в нахождении весов — коэффициентов связей между нейронами. В процессе обучения нейронная сеть выявляет сложные зависимости между входными параметрами и выходными, а также использует метод обобщения. Это дает возможность получить такую функцию, которая будет наиболее приближена к функции человеческого мышления по определению возраста лица. Конечно, как и человек, даже самая точная модель будет иметь погрешности определения возрастам и пола, которые могут уменьшиться в ходе процесса обучения.

Если обучение проведено успешно, результатом глубокого обучения будет модель, с помощью которой алгоритм сможет давать верный результат. В этом случае нейросеть правильно определит пол и возраст человека, которого не было в обучающей выборке, а также сможет работать с зашумленными, искаженными изображениями.

Когда имеет место переобучение модели, тренировочный процесс следует прекратить, так как его дальнейшее проведение не улучшит точность модели предсказывать.

В ходе распознавания возраста алгоритм решает задачу регрессии. Алгоритм предсказывает не один конкретный возраст, а рассчитывает вероятность каждого возраста в диапазоне от 1 до 100 у лица. Тем самым, итоговый возраст рассчитывается по формуле с учетом ста предсказаний, зависящих от множества входных параметров. В отличие от задачи классификации в задаче регрессии выходными могут быть только поля с непрерывным видом данных (1...100).

В следующей главе рассмотренные теоретические основы реализованы в программном комплексе.

Глава 4 Экспериментальное исследование и апробация метода распознавания личности на основе нейросетевых алгоритмов

4.1 Описание программного комплекса и оборудования

Операционная система: Windows-10-10.0.19041-SP0.

Среда разработки: PyCharm 2021.3.3.

Интерпретатор: Python version 3.8.12.

Сервис для ML-разработки: Yandex DataSphere с видеокартой Tesla V100-SXM2-32GB.

Визуализация обучения моделей проведена на платформе машинного обучения Weights & Biases [35].

4.2 Общая характеристика языка программирования Python

Программная реализация разработанного метода определения атрибутов личности по изображению лица выполнена на языке программирования Python.

Выбор этого языка обусловлен тем, что Python – объектно-ориентированный интерпретируемый язык программирования. После написания программистом кода интерпретатор выполняет его частями и тем самым преобразует в машинный код.

Принципами программирования на Python является его простота и обязательность предоставления всех необходимых отступов при написании кода. Последнее требование приучает программистов создавать структурированный, хорошо читаемый, понятный другим программный код. Python имеет способность работать практически с любыми аппаратными платформами и операционными системами.

В ходе любого объектно-ориентированного программирования объекты взаимодействуют между собой. Объект при этом принадлежит определенному классу, имеет свойства и методы.

Язык Python включает такие части, как процедуры и функции, которые отвечают за выполнение логически завершенных задач. Данные части состоят из заголовка с наименованием процедуры или функции и тела с программным кодом. Python – процедурный язык программирования. За счет введения процедур и функций сложная задача разбивается на логически связанные подзадачи. Код становится компактным и понятным.

Для эффективного решения задач компьютерного зрения язык программирования Python подходит больше всего. Этот вывод сделан на основании следующих свойств:

- язык на Python значительно короче других высокоуровневых языков. Тем самым, уменьшается время разработки и увеличивается скорость восприятия алгоритма третьими лицами;
- легкость чтения. Даже junior-программист сможет прочитать чужой код и сразу начать модифицировать его;
- расширяемость. Python имеет большое число библиотек, упрощающих работу с компьютерным зрением. Библиотеки находятся в свободном доступе и легко устанавливаются;
- взаимодействие. Python обладает широкими возможностями для промежуточного тестирования алгоритма;
- объектно-ориентированный, императивный и функциональный подходы. Другими словами, Python – мультипарадигменный язык программирования, позволяющий решать многогранные задачи из широкого диапазона;
- многоплатформенность. Программировать на Python возможно во всех основных операционных системах [19].

Изначально Python создан в 1991 году голландцем Гвидо ван Россумом и выложен им во всеобщий доступ [38]. Уже 30 лет язык Python непрерывно совершенствуется и дорабатывается профессиональными программистами и любителями-исследователями со всего мира.

4.2.1 Библиотека Keras

Библиотека Keras предназначена для проведения глубокого обучения и построения моделей для решения задач компьютерного зрения. Наиболее важным достоинством Keras является ее быстродействие.

Библиотека Keras позволяет применить сверточные нейронные сети для решения задачи распознавания объекта на изображении. Свертки как правило применяются с размером 3 на три пиксела. Значение интенсивности пиксела изображения умножается на соответствующий элемент ядра свертки. Затем все полученные значения складываются.

Следует отметить следующие возможности данной библиотеки:

- легко и быстро создает классы;
- поддерживает сверточные и рекуррентные сети;
- может функционировать на различном оборудовании ЭВМ.

При создании разработчики модуля руководствовались следующими идеями:

- юзабилити. Программный интерфейс библиотеки удобен для пользователей. Для решения задач компьютерного зрения не требуется писать длинные алгоритмы, а следует использовать специально предназначенные команды. Тем самым Keras минимизирует количество действий, требуемых от пользователя.
- модульность. Библиотека обладает обширным списком дополнительных функций. Так, можно подключить нейронные слои, различные оптимизаторы и регуляторы.

4.2.2 Библиотека OpenCV (Open Source Computer Vision Library)

OpenCV с открытым исходным кодом является бесплатной и распространяется под лицензией BSD [37]. Библиотека предназначена для

разработки алгоритмов компьютерного зрения на различных языках программирования, в том числе на Python. С помощью данной библиотеки возможно создавать сложнейшие и быстродействующие приложения.

OpenCV позволяет открывать и сохранять файлы с изображениями и работать с ними путем масштабирования, поворотов, кадрирования, отражения по осям.

Python имеет все возможности для качественного решения задачи компьютерного зрения. В зависимости от конкретной ситуации на практике используется как одна-две библиотеки, так и комбинация нескольких библиотек.

4.3 Формирование набора данных изображений лиц

На сегодняшний день самым большим общедоступным набором данных для прогнозирования возраста является датасет IMDB-WIKI [31]. Предлагаемый датасет имеет 523 051 изображений лиц (рисунок 23). Изображения взяты с информационного ресурса IMDb и из Википедии, а также являются кадрами из фильмов.



Рисунок 23 – Фрагмент базы IMDB-WIKI

К изображениям предоставляется отдельный файл .mat, содержащий всю метаинформацию:

- dob: дата рождения;
- photo_taken: год, когда была сделана фотография;
- full_path: путь к файлу;
- пол: 0 для женщин и 1 для мужчин, NaN, если неизвестно;
- имя: имя знаменитости;
- face_location: расположение лица;
- face_score: оценка детектора;
- second_face_score: оценка детектора второго лица;
- celeb_names: список всех имен знаменитостей;
- celeb_id: индекс имени знаменитости.

4.4 Модели глубокого обучения

Обучение нейросети «с нуля» занимает большое количество времени (часы и даже дни). В связи с этим разработчики Keras предлагают использовать на старте уже предварительно обученные нейронные сети для проведения трансферного обучения [33].

На момент проведения исследования на официальном сайте Keras представлено 40 моделей глубокого обучения. Каждая нейросеть имеет свои характеристики: размер, точность, параметры, глубина (слои активации, слой пакетной нормализации и другие слои) и время обработки. При этом точность модели указана на основании полученных результатов классификации изображений при обучении на наборе данных ImageNet [40] (включает десять миллионов URL с изображениями).

Модели можно использовать для прогнозирования, извлечения признаков и тонкой настройки. Для выполнения целей исследования предлагается взять три предварительно обученных модели:

- классическая модель VGG-16, которая использовалась при описании метода DEX разработчиками [39]. Размер ввода по умолчанию для этой модели — 224x224. Модель имеет 16 нейронных слоев и заявленную точность в размере 71,3 %. Это модель, имеющая наименьшее количество слоев в Keras;
- одна из самых точных моделей Keras - EfficientNetB3 имеет 210 нейронных слоев и заявленную точность в 81.6%. Модели присуща тщательная балансировка глубины, ширины и разрешения нейросети. Разработчиками модели заявлено, что EfficientNets хорошо достигают современной точности на многих популярных моделях данных [36]. Входное изображение может быть размером до 255x255;
- модель DenseNet201 с 402 слоями является одной из самых многослойных моделей Keras и может быть эффективна для глубокого обучения по распознаванию атрибутов личности по изображению лица. Модель имеет заявленную точность в размере 77.3%. Нейросеть содержит более короткие связи между слоями, близкими ко входу, и слоями, близкими к выходу. По информации, представленной разработчиками, у сетей DenseNet есть несколько неоспоримых преимуществ: они устраняют проблему исчезающего градиента, усиливают распространение признаков, применяют повторное использование функций и существенно сокращают количество используемых параметров [30]. На вход направляются изображения формата 224x224.

4.5 Методы оценки качества работы алгоритмов

Для измерения качества работы алгоритмов определения возраста личности по изображению лица предлагается использовать показатель

средней абсолютной ошибки (MAE). Абсолютная ошибка определяется с помощью формулы 1.

$$e = |y - \tilde{y}|, \quad (1)$$

где y отражает фактический возраст человека, а \tilde{y} – возраст, вычисленный алгоритмом.

Соответственно, средняя абсолютная ошибка вычисляется путем сложения показателей всех ошибок и деления полученной суммы на количество вариантов (в случае определения возраста – 100).

Для определения MAE предлагается использовать базу данных APPA-REAL [27], отличную от первоначальной базы обучения IMDB-WIKI, которая включает 7 591 изображение с соответствующими метками реального (биологического) и кажущегося (воспринимаемого внешними наблюдателями) возраста.

Кроме того, при оценке моделей необходимо учитывать значение рассчитанной функции потерь (val-loss), которая отражает как именно выходные данные связаны с исходными. В ходе обучения функция принимает два параметра: прогнозируемые и истинные выходные данные. Данные параметры сравниваются, и их разница будет являться потерей. Тем самым алгоритм проверяет насколько хорошо работает модель. Чем ниже значение потери, тем более вероятность истинности выражения 2.

$$\text{прогнозируемые} = \text{истинные выходные данные} \quad (2)$$

Для оценки модели по определению пола предлагается учитывать показатель точности, выраженный в процентах, который определяется при обучении модели.

4.6 Сравнительное исследование моделей Keras

Для алгоритмов Keras (VGG, EfficientNetB3, DenseNet201) выбрана низкая скорость обучения 0,001 и входной размер фотографий 224x224. При обработке Big Data использовался оптимизатор Adam, который сочетает в себе идею накопления движения и идею более слабого обновления весов для типичных признаков.

Количество обучающих примеров за один проход по нейросети вперед/назад (`batch_size`) составляло 32.

Для обучения каждой модели использовалось различное количество эпох (один проход вперед и один проход назад всех обучающих примеров), в зависимости от начала стагнации обучения.

Фрагмент программного кода на языке программирования Python глубокого обучения модели определения атрибутов личности по изображению лица приведен в Приложении А.

По каждой модели алгоритм отработал следующим образом:

- из обучающей базы после отфильтровки сформирован мета-файл формата `.csv`, включающий сведения о каждом изображении: `genders` (пол), `ages` (возраст), путь к изображению и его название;
- с помощью `OpenCV` каждое изображение прочитано (`cv2.imread`), детектором найдено лицо и командой `cv2.resize` изображение обрезано до плоскости, лежащей по периметру лица;
- в ходе тренировочного процесса выделенным векторам-признаком каждого изображения сопоставлен соответствующий пол и возраст. Чтобы не запутаться, алгоритм применяет мета-файл формата `.csv` для навигации. Сделаны соответствующие выводы по закономерностям определения пола и возраста. После каждой эпохи веса модели в случае ее улучшения сохранялись.

Параметр `val_loss` визуализирован при обучении (рисунок 24). Если `val_loss` не уменьшался в ходе нескольких эпох, глубокое обучение прерывалось, так как дальнейший процесс обучения не имел смысла.

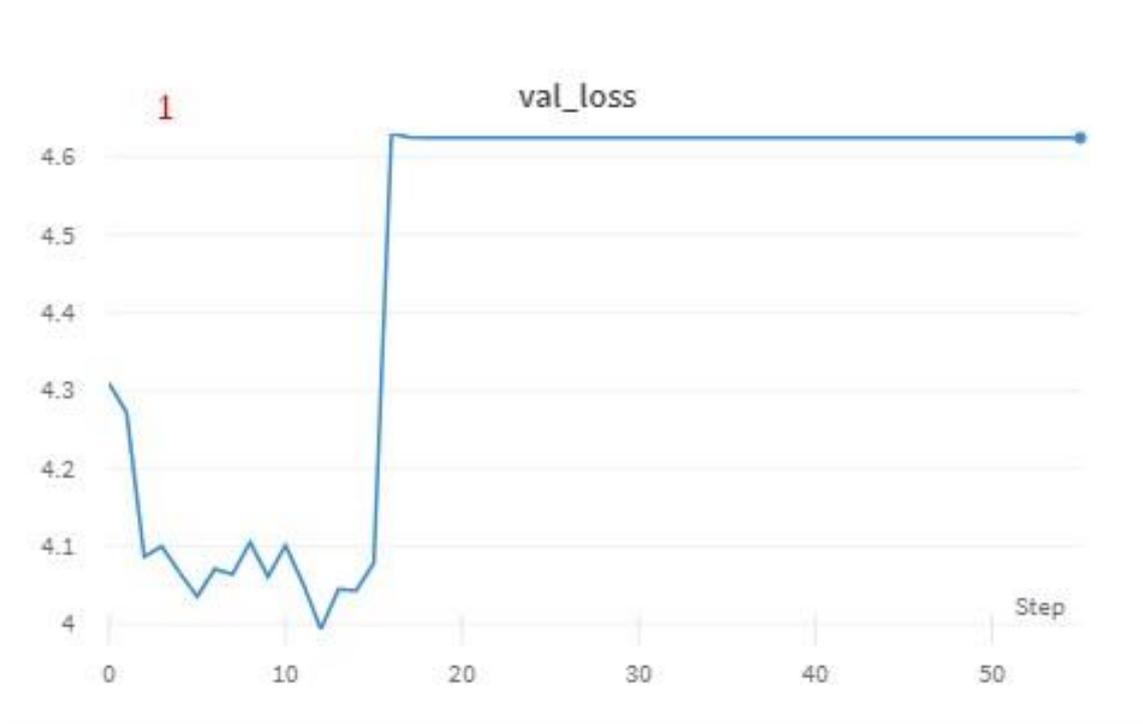
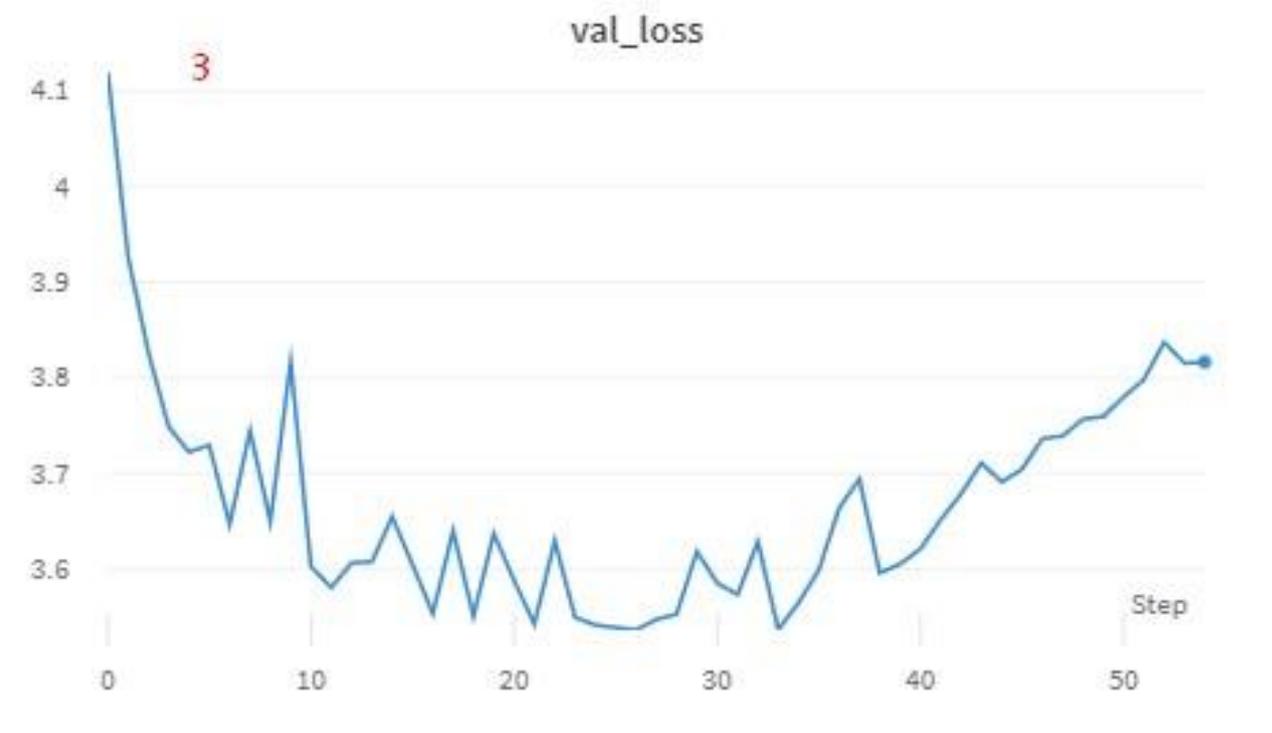


Рисунок 24 – Обучение нейросети по готовым моделям (1 - VGG, 2 – EfficientNetB3, 3 - DenseNet201)



Продолжение рисунка 24



Продолжение рисунка 24

Глубокое обучение классической модели VGG-16 не потребовало большие вычислительные мощности. Продолжительность обучения одной эпохи составила около 10 минут. После 13 эпохи эффективность обучения стала снижаться, а с 15 эпохи график резко ушел вверх и оставался на одном уровне, что говорит о переоснащении модели. То есть обученная нейросеть способна хорошо объяснить данные обучающей, но не может эффективно работать с информацией, полученной из других изображений.

При обучении модели EfficientNetB3 за основу взяты веса, полученные ранее на 11 эпохе обучения. Модель хорошо улучшалась до 6 эпохи, однако впоследствии график стал иметь восходящий тренд, что говорит об использовании всех возможностей для обучения модели.

В процессе обучения нейросети на DenseNet201 после 27 эпохи модель не улучшалась на протяжении последующих эпох. Напротив, линия графика

шла то вверх, то вниз, что говорит о переобучении модели. Продолжительность обучения каждой эпохи составляла 25 минут.

После получения обученных моделей-весов, лучшие из них с наименьшим val-loss протестированы на обучающем датасете APPA-REAL. Также в процессе обучения получены значения точности алгоритма при определении пола человека. Данные значения занесены в таблицу 3.

Таблица 3 – Показатели глубокого обучения

№	Название модели	Val-loss	MAE Apparent, возраст	MAE Real, возраст	Точность распознавания пола, %
1	VGG	3,99	10,06	11,15	86,28
2	EfficientNetB3	3,5	5,58	6,54	91,76
3	DenseNet201	3,54	5,62	6,56	91,86

Таким образом в ходе исследования экспериментальным путем установлено, что нейросеть, обученная на основе модели Keras – EfficientNetB3 дает наименьшие показатели val-loss, MAE Real, MAE Apparent возраста, а также высокую точность определения пола лица на изображении и может применяться в реальных условиях.

4.7 Тестирование метода определения возраста и пола человека по изображению лица

Для тестирования эффективности выбранной модели использована видеопоследовательность с веб-камеры и свое изображение лица (рисунок 25).

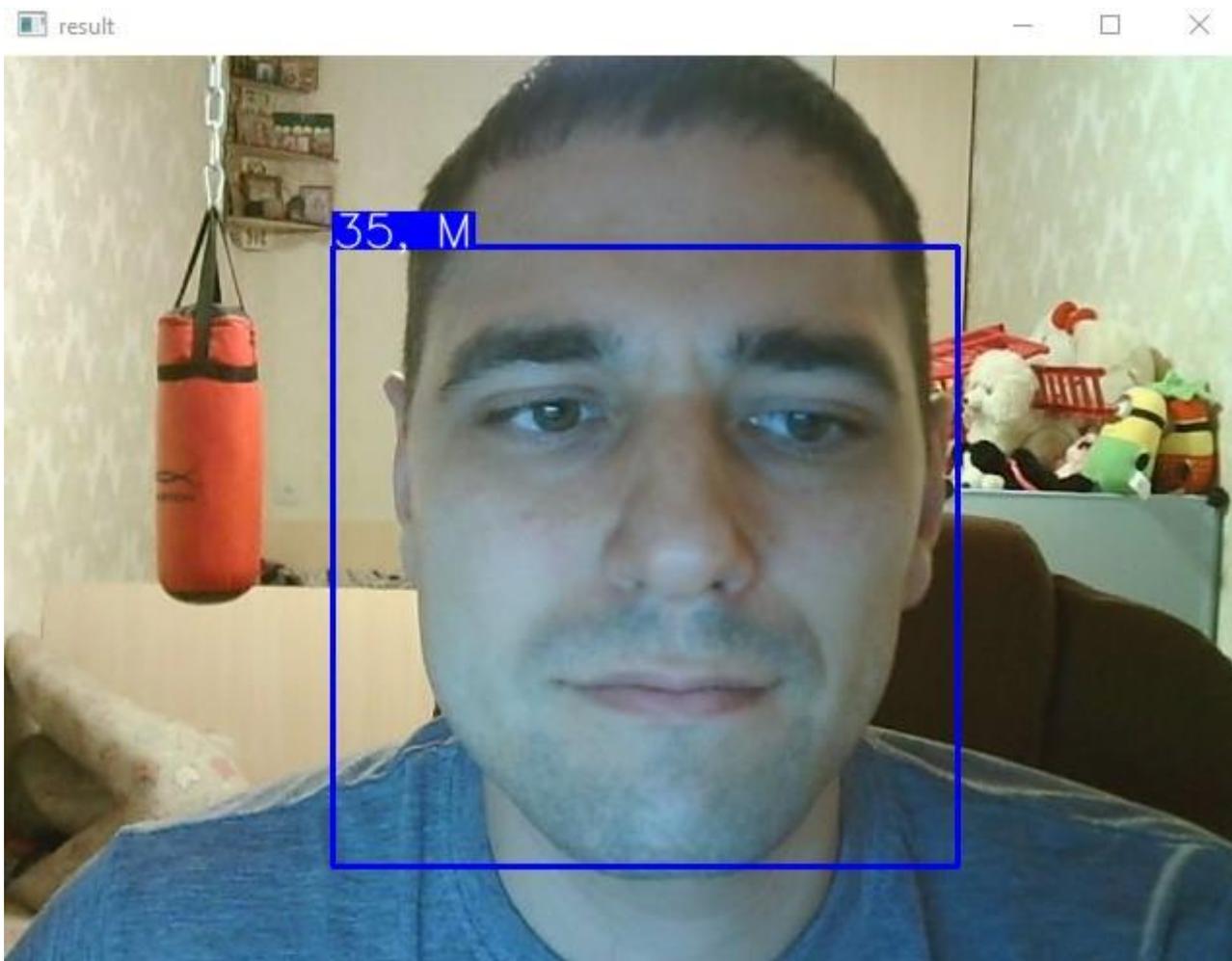


Рисунок 25 – Определение атрибутов личности по изображению лица

После фокусировки видеокамеры, в идеальных условиях (освещение, ракурс) алгоритм совершенно точно определил возраст и пол.

Кроме того, для дальнейшего тестирования алгоритма в реальных условиях произвольно выбраны 100 изображений спортсменов различных сборных России мужского и женского пола. Некоторые фотографии выполнялись в ходе спортивных игр, то есть в реальных, а не в специально созданных условиях, что значительно повысило сложность задачи.

За основу выбран показатель биологического возраста. При определении точности пола посчитан процент верных срабатываний, а для вычисления точности определения возраста вычислен показатель MAE.

Как следует из полученных результатов, алгоритм наиболее точно решает вопрос определения возраста, когда лицо на изображении зафиксировано в анфас. Точность снижается при поворотах головы спортсмена и изменении выражения лица. Кроме того, результативность алгоритма очень зависит от качества фотографии (рисунок 26).

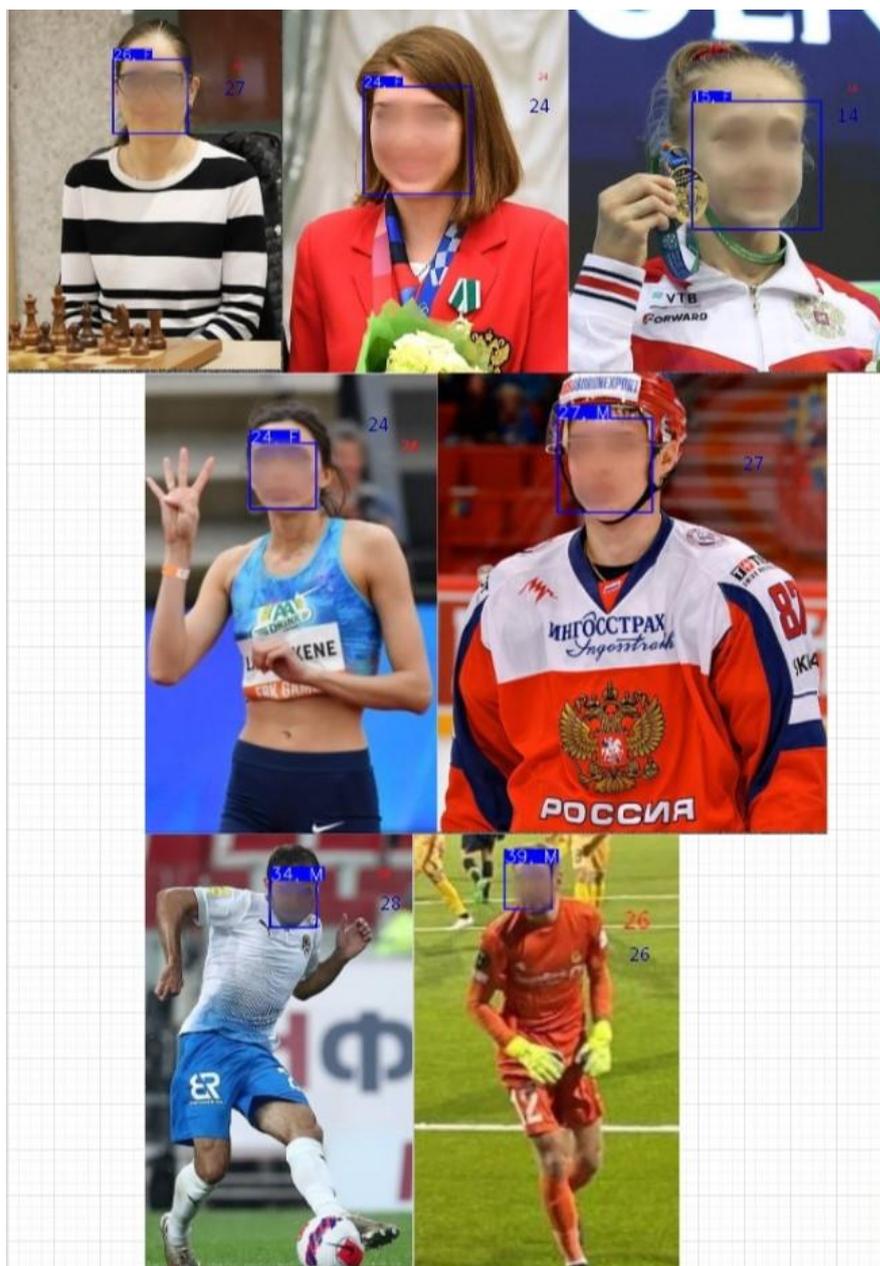


Рисунок 26 – Примеры работы алгоритма по определению изображений (с целью соблюдения ст. 152.1 Гражданского кодекса РФ лица размыты и не идентифицируются)

Таким образом, в ходе проведенного опыта два лица на изображении не были распознаны, точность определения пола составила 97 % (неправильно определен пол у трех лиц). Общая сумма отклонения предсказанного алгоритмом возраста от биологического составила 502 года на 98 изображений. В итоге средняя арифметическая ошибка (MAE) определения возраста определена как 5,12.

Полученные результаты превышают показатели точности модели EfficientNetB3, полученные ранее при тестировании на обучающем датасете APPA-REAL.

Тем самым, по результатам тестирования обученной модели в реальных условиях получен приемлемый результат с высокой точностью.

Выводы по главе 4

В ходе выполнения практической части ВКР был реализован разработанный в главе 3 метод глубокого ожидания (DEX). Программный код полностью написан на языке Python. Глубокое обучение проводилось на нескольких моделях разработчика Keras, имеющих различные характеристики.

По результатам тренировок самыми эффективными оказались модели, обученные на базе многослойных нейросетей, которые дают минимальное количество ошибок определения возраста и высокий показатель точности определения пола лица по изображению.

Нейросеть VGG-16 с минимальным количеством слоев оказалась слишком простой для решения поставленной задачи.

Нейросеть DenseNet201, имеющая более 400 слоев показала также приемлемые результаты при обучении. Однако при затрате больших вычислительных мощностей полученная модель допускает больше ошибок при определении возраста по сравнению с моделью, обученной на нейросети EfficientNetB3, имеющей 210 слоев.

Таким образом, заявленная разработчиками Keras точность прямо влияет на результаты глубокого обучения независимо от количества слоев моделей.

Проведенные тесты алгоритма в реальном времени и пространстве показали, что задача определения атрибутов личности по изображению лица решена. На эффективность алгоритма прямо влияет качество входящего изображения, а также ракурс съемки.

Алгоритм возможно применять в реальных неподготовленных условиях, с учетом возможной погрешности MAE при определении возраста лица. Если создать благоприятные условия съемки (освещенность, качество, правильный ракурс), возможно добиться достаточно высокой точности при определении возраста лица. Кроме того, даже в неподготовленных условиях алгоритм практически безошибочно классифицирует пол человека

Улучшение точности алгоритма возможно при наличии большего количества обучающих изображений, в том числе с добавлением лиц в профиль. Кроме того, для получения лучшего результата необходимо использовать высококачественные датчики получения информации.

Заключение

Компьютерное зрение – перспективное научное направление в области ИИ, имеющее прямое прикладное применение. Правильное построение программного кода зависит от полноты понимания теоретических основ и элементарных принципов действия СТЗ.

Магистерская диссертация посвящена актуальной проблеме автоматического определения атрибутов личности по изображению лица, решаемой при помощи методов компьютерного зрения.

Выполненные в работе научные исследования представлены следующими основными результатами.

В процессе проведенного исследования в главе 1 диссертации исследована широкая теоретическая база в области понимания задач компьютерного зрения и их решения, рассмотрены проблемы, возникающие при решении задач компьютерного зрения, доказана актуальность решения такой задачи, как определение атрибутов личности по изображению лица.

В главе 2 диссертации теоретически определены общие положения функционирования СТЗ по формированию, представлению, обработке и анализу изображений, а также рассмотрены связанные с данными операциями методы машинного обучения в контексте решения задач определения атрибутов личности по изображению лица. По итогам исследования определены достоинства и недостатки каждого метода, а для решения поставленной задачи предложено использовать нейросетевые алгоритмы.

В главе 3 разработан алгоритм автоматического определения возраста и пола лица на основе метода глубокого ожидания (DEX), определены способы предобработки изображений и метод обнаружения лица на изображении.

В главе 4 определен программный комплекс для решения задачи определения атрибутов личности по изображению лица, найдена подходящая

база фотографий лиц людей с метками для пола и возраста, подобраны архитектуры нейронных сетей для глубокого обучения и проведено их теоретическое и практическое сравнение. В ходе исследования обучены модели по заданным параметрам, проверена их работоспособность на тестовом множестве. Из полученных весов выбрана модель с наименьшими ошибками определения возраста, которая успешно применена в реальной ситуации. Так, был правильно определен пол и возраст личности по изображению лица на видеопоследовательности.

Тем самым, проведенным исследованием получены полезные теоретические и практические методики решения такой задачи компьютерного зрения, как определение атрибутов личности по изображению лица.

Для доказывания главной гипотезы работы исходные данные исследования получены на основании системного подхода к изучению научной отечественной и зарубежной литературы в области компьютерного зрения, а также моделирования работы СТЗ для решения конкретных задач.

Таким образом, в ходе работы решена актуальная научно-практическая проблема исследования и разработан алгоритм, который достаточно точно определил атрибуты личности по изображению лица.

Список используемой литературы и используемых источников

1. Блум Ф., Лейзерсон А., Хофстедтер Л. Мозг, разум и поведение. М. : Мир, 1988. 246 с.
2. Визильтер Ю.В. [и др.]. Обработка и анализ изображений в задачах машинного зрения: Курс лекций и практических занятий. М.: Физматкнига, 2010. 672 с.
3. Глебов Н.И., Кочетов Ю.А., Плясунов А.В. Методы оптимизации. Учеб. пособие. Новосибирск : Новосиб. ун-т, 2000. 105 с.
4. Горячкин Б.С., Китов М.А. Компьютерное зрение // E-Scio 2020. № 9. С. 317-345.
5. Динамика интернет-торговли, прогноз. Data Insight [Электронный ресурс]. URL: https://datainsight.ru/sites/default/files/DI_eCommerce2020_2024.pdf. Л. 3 (дата обращения: 28.10.2020).
6. Журавлев Ю.И., Гуревич И.В. Распознавание образов и анализ изображений / Искусственный интеллект. - В 3-х кн. - М. : Радио и связь, Кн. 2. Модели и методы: Справочник, 1990. 304 с.
7. Захаренко Г.И. Формирование изображений в системе технического зрения с помощью датчиков // VII Международная научно-практическая конференция (школа-семинар) молодых ученых: сборник материалов. 2021. С. 256-261.
8. Клетте Рейнхард. Компьютерное зрение. Теория и алгоритмы. М. : ДМК Пресс, 2019. 506 с.
9. Комарцова Л.Г., Максимов А.В. Нейрокомпьютеры: учеб. пособие для студентов вузов, обучающихся по специальности «Вычисл. машины, комплексы, системы и сети» направления подгот. дипломиров. специалистов «Информатика и вычисл. техника». М. : Изд-во МГТУ им. Н.Э. Баумана, 2004. 399 с.

10. Коршунов А. [и др.]. Определение демографических атрибутов пользователей микроблогов. Труды Института системного программирования РАН, 2013, том 25. С. 180.
11. Кузнецов Д.А. [и др.]. Сравнительный анализ популярных методов распознавания лиц на изображениях // Научный результат. Информационные технологии. Т. 4. № 4. 2019.
12. Мкртычев С.В. Большие данные: подходы к определению и классификации // Сборник докладов III Всероссийской научной конференции с международным участием (ТГУ). 2021. С. 253-258.
13. Мокеев А.В. О точности и быстродействии метода синтеза главных компонент. Бизнес-информатика № 3 (13). 2010.
14. Новаковский С.В. Цвет в цветном телевидении. М. : Радио и связь, 1988. 288 с.
15. Поляков Г.И. О принципах нейронной организации мозга. М. : Изд-во Моск. ун-та, 1965. 166 с.
16. Потапов А.С. Системы компьютерного зрения. Учебное пособие. СПб : Университет ИТМО, 2016. 161 с.
17. Розенблатт Ф. Принципы нейродинамики : Перцептроны и теория механизмов мозга. М. : Мир, 1965. 480 с.
18. Розенфельд А. Распознавание и обработка изображений с помощью вычислительных машин. М. : Мир, 1972. 230 с.
19. Сегаран Т. Программируем коллективный разум. СПб. : Символ-Плюс, 2008. 368 с.
20. Селянкин В.В. Решение задач компьютерного зрения: Учебное пособие. Таганрог : Изд-во ЮФУ, 2016. 92 с.
21. Селянкин В.В., Скороход С.В. Анализ и обработка изображений в задачах компьютерного зрения: учебное пособие. Таганрог : Изд-во ЮФУ, 2015. 82 с.
22. Содем Я. Э. Программирование компьютерного зрения на языке Python. М. : ДМК Пресс, 2016. 311 с.

23. Тьюринг А. М. Вычислительные машины и разум. М. : АСТ, 2018. 125 с.
24. Шапиро Л., Стокман Дж. Компьютерное зрение. М. : Лаборатория знаний, 2020. 763 с.
25. Шолле Ф. Глубокое обучение на Python. СПб. : Питер, 2019. 397 с.
26. Alexandr Andoni and Piotr Indyk. Near-Optimal Hashing Algorithms for Approximate Nearest Neighbor in High Dimensions. Proceedings of the 47th Annual IEEE Symposium on Foundations of Computer Science, October 2006. Pages 459-468.
27. APPA-REAL (real and apparent age) [Электронный ресурс]. URL: <https://chalearnlap.cvc.uab.cat/dataset/26/description/> (дата обращения: 12.05.2021).
28. David Arthur, Sergei Vassilvitskii. «How Slow is the k-means Method?». - Published in SCG, 2006.
29. Facial Features Underlying the Decoding of Pain Expressions. PLoS Genetich, 2019.
30. Gao Huang et al. Densely Connected Convolutional Networks. Weinberger, 2016.
31. IMDB-WIKI – 500k+ face images with age and gender labels. Rasmus Rothe, Radu Timofte, Luc Van Gool.
32. Kaiming He et al. Deep Residual Learning for Image Recognition, 2015.
33. Keras: the Python deep learning API [Официальный сайт разработчика Keras] // URL: <https://keras.io/api/applications/> (дата обращения: 23.09.2021).
34. Library dlib [Официальный сайт разработчика dlib]. URL: <http://dlib.net/> (дата обращения: 18.03.2021).
35. Machine learning platform Weights & Biases [Сервис wandb]. URL: <https://docs.wandb.ai/> (дата обращения: 04.12.2021).
36. Mingxing Tan, Quoc V. Le. EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks, 2019.

37. Open Source Computer Vision Library [Официальный сайт разработчика OpenCV]. URL: <https://opencv.org/> (дата обращения: 23.09.2021).
38. Python Software Foundation [Официальный сайт разработчика Python.]. URL: <http://python.org/> (дата обращения: 20.06.2021).
39. Rothe R., Timofte R. and Gool L.V., "DEX: Deep EXpectation of apparent age from a single image," in Proc. of ICCV, 2015.
40. Simonyan K. and Zisserman A. Very deep convolutional networks for large-scale image recognition. CoRR, abs/1409.1556, 2014.
41. Viola, Jones. Robust Real-time Object Detection. - International Journal of Computer Vision 57(2), 137–154, 2004.

Приложение А

Фрагмент программного кода на языке программирования Python глубокого обучения модели определения атрибутов личности по изображению лица

```
#импортирование нужных библиотек
import ...

#использование платформы Hydra для обработки файла конфигурации,
#включающего параметры обучения, модели, названия БД, проекта wandb
@hydra.main(config_path="table/configuration.yaml")
def main(configuration):
    if configuration.wandb.project:
        wandb.init(project=configuration.wandb.project)
        callbacks = [WandbCallback()]
    else:
        callbacks = []

#задается путь к мета-файлу с обучающими данными
meta_file = Path(r'C:\Users\frela\PycharmProjects\Copy\table\imdb.csv')

#чтение мета-данных
line = pd.read_csv(str(meta_file))
deep, val = train_test_split(line, random_state=42, test_size=0.1)

#задается класс для перебора последовательности изображений согласно
#файлу конфигурации
class ImageSequence(Sequence):
    def __init__(self, configuration, line, mode):
```

Продолжение Приложения А

```
self.line = line
self.index = np.arange(len(line))
self.example = configuration.train.example
self.db_file =
Path(r'C:\Users\freia\PycharmProjects\pythonProject6
\data\imdb_crop')
self.img_pixel = configuration.model.img_pixel
self.mode = mode

#функция чтения и обрезки тестовых изображений
def __getitem__(self, idx):
    sample_index = self.index[idx * self.example:(idx + 1) *
self.example]
    images = []
    genders = []
    ages = []

    for _, row in self.line.iloc[sample_index].iterrows():
        image = cv2.imread(str(self.db_file.joinpath
(row["img_paths"])))
        image = cv2.resize(image, (self.img_pixel, self.img_pixel))

    #аугментация обучающих изображений (их поворот под углом, по
#оси, масштабирование, изменение гаммы)
    if self.mode == "train":
        convert = A.Compose([
            A.ShiftScaleRotate(shift_limit=0.03125,
```

Продолжение Приложения А

```
scale_limit=0.20, rotate_limit=20,  
border_mode=cv2.BORDER_CONSTANT,  
value=0, p=1.0),  
A.RandomBrightnessContrast(brightness_limit=0.2,  
contrast_limit=0.2, p=0.5),  
A.HorizontalFlip(p=0.5)  
])  
  
image = convert(image=image)["image"]  
  
images.append(image)  
genders.append(row["genders"])  
ages.append(row["ages"])  
  
images = np.asarray(images)  
genders = np.asarray(genders)  
ages = np.asarray(ages)  
  
return images, (genders, ages)  
  
def __len__(self):  
    return len(self.line) // self.example  
  
def on_epoch_end(self):  
    np.random.shuffle(self.index)  
  
train_generation = ImageSequence(configuration, deep, "train")  
val_generation = ImageSequence(configuration, val, "val")
```

Продолжение Приложения А

```
#задается синхронное обучение на нескольких процессорах на одном  
#компьютере
```

```
mirrored = tf.distribute.MirroredStrategy()
```

```
with mirrored.scope():
```

```
def get_model(configuration):
```

```
    initial_model = getattr(applications,
```

```
    configuration.model.model_keras)(
```

```
    include_top=False,
```

```
    input_shape=(configuration.model.img_pixel,
```

```
    configuration.model.img_pixel, 3),
```

```
    pooling="avg"
```

```
)
```

```
    sign = initial_model.output
```

```
    pred_gender = Dense(units=2, activation="softmax",
```

```
    name="pred_gender")(sign)
```

```
    pred_age = Dense(units=101, activation="softmax",
```

```
    name="pred_age")(sign)
```

```
    model = Model(inputs=initial_model.input, outputs=[pred_gender,
```

```
    pred_age])
```

```
    return model
```

```
model = get_model(configuration)
```

```
def get_optimization(configuration):
```

```
    if configuration.train.optimization == "adam":
```

```
        return Adam(lr=configuration.train.speed)
```

Продолжение Приложения А

```
opt = get_optimization(configuration)

def get_planning(configuration):
    class Schedule:
        def __init__(self, nb_epochs, initial_speed):
            self.epochs = nb_epochs
            self.initial_speed = initial_speed

        def __call__(self, epoch_idx):
            if epoch_idx < self.epochs * 0.25:
                return self.initial_speed
            elif epoch_idx < self.epochs * 0.50:
                return self.initial_speed * 0.2
            elif epoch_idx < self.epochs * 0.75:
                return self.initial_speed * 0.04
            return self.initial_speed * 0.008

    return Schedule(configuration.train.epochs,
                    configuration.train.speed)

planning = get_planning(configuration)
model.compile(optimizer=opt,
              loss=["sparse_categorical_crossentropy",
                  "sparse_categorical_crossentropy"],
              metrics=['accuracy'])

#задается формат сохранения улучшенных весов-моделей
save = Path(r"C:\Users\fre1a\PycharmProjects\Copy\save")
```

Продолжение Приложения А

```
save.mkdir(exist_ok=True)
filename = "_".join([configuration.model.model_keras,
str(configuration.model.img_pixel),
"weights.{epoch:02d}-{val_loss:.2f}.hdf5"])
callbacks.extend([
LearningRateScheduler(schedule=planning),
ModelCheckpoint(str(save) + "/" + filename,
monitor="val_loss",
verbose=1,
save_best_only=True,
mode="auto")
])

#метод Keras обучает модель в течение фиксированного количества эпох
#(итераций в наборе данных)
model.fit(train_generation, epochs=configuration.train.epochs,
callbacks=callbacks, validation_data=val_generation,
workers=multiprocessing.cpu_count())

if __name__ == '__main__':
    main()
```