

Министерство науки и высшего образования Российской Федерации
Федеральное государственное автономное образовательное учреждение
высшего образования

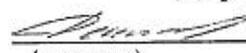


«Уральский федеральный университет
имени первого Президента России Б.Н. Ельцина»

Институт радиоэлектроники и информационных технологий – РТФ
Школа профессионального и академического образования

ДОПУСТИТЬ К ЗАЩИТЕ ПЕРЕД ГЭК

Директор ШПиАО

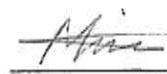

(подпись) Д.В. Денисов
(Ф.И.О.)

« 03 » июня 2024 г.

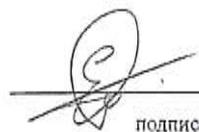
ВЫПУСКНАЯ КВАЛИФИКАЦИОННАЯ РАБОТА

ДЕТЕКЦИЯ ПОВТОРНОГО ПОСЕЩЕНИЯ КЛИЕНТАМИ МАГАЗИНА ПО
ПРОДАЖЕ АВТОМОБИЛЕЙ

Научный руководитель: Ронкин Михаил Владимирович
к. т. н., доцент


подпись

Нормоконтролер: Огуренко Егор Владимирович


подпись

Студент группы: РИМ-220962 Ротерман Виктор Андреевич


подпись

Екатеринбург
2024

Министерство науки и высшего образования Российской Федерации
Федеральное государственное автономное образовательное учреждение высшего образования
«Уральский федеральный университет
имени первого Президента России Б.Н. Ельцина»

Институт радиоэлектроники и информационных технологий – РТФ
Школа профессионального и академического образования
Направление подготовки 09.04.01 Информатика и вычислительная техника
Образовательная программа 09.04.01/33.03 Инженерия машинного обучения

ЗАДАНИЕ

на выполнение выпускной квалификационной работы

студента Ротерман Виктора Андреевича группы РИМ-220962
(фамилия, имя, отчество)

1. Тема выпускной квалификационной работы Детекция повторного посещения клиентами магазина по продаже автомобилей

Утверждена распоряжением по институту от «4» декабря 2023 г. № 33.02-05/298

2. Научный руководитель Ронкин Михаил Владимирович, доцент, к. т. н.

(Ф.И.О., должность, ученая степень, ученое звание)

3. Исходные данные к работе

Описание базового алгоритма работы модуля, 4 видеофрагмента

4. Перечень демонстрационных материалов _____

5. Календарный план

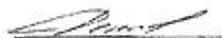
№ п/п	Наименование этапов выполнения работы	Срок выполнения этапов работы	Отметка о выполнении
1.	<i>1 раздел (глава)</i>	до 23.03.2024 г.	✓
2.	<i>2 раздел (глава)</i>	до 29.04.2024 г.	✓
3.	<i>3–4 раздел (глава)</i>	до 20.05.2024 г.	✓
4.	<i>ВКР в целом</i>	до 24.05.2024 г.	✓

Научный руководитель Ронкин Михаил Владимирович
Ф.И.О.

Студент задание принял к исполнению 12.02.24
дата

6. Допустить Ротерман Виктора Андреевича к защите выпускной квалификационной работы в экзаменационной комиссии

Директор ШПиАО



(подпись)

Д.В. Денисов

Ф.И.О.

РЕФЕРАТ

Выпускная квалификационная работа магистра 32 с., 12 рис., 6 табл., 14 источников.

ДЕТЕКЦИЯ ОБЪЕКТОВ НА ВИДЕО, ДЕТЕКЦИЯ ЛИЦ НА ВИДЕО, РАСПОЗНАВАНИЕ ЛИЦ НА ВИДЕО, АЛГОРИТМ РАСПОЗНАВАНИЯ ЛИЦ НА ВИДЕО.

Цель работы – разработать модуль программного обеспечения позволяющие определять, уникальных клиентов по видеофрагменту посещения клиентами, магазина по продаже автомобилей.

Объект исследования – системы распознавания лиц на фото и видео.

Методы исследования: анализ и синтез, эксперимент, измерение, сравнение.

Результаты работы: архитектура модуля «Детекция повторного посещения клиентами магазина по продаже автомобилей», алгоритм распознавания лиц и проверки их на уникальность, оптимальные параметры алгоритма распознавания лиц и проверки их на уникальность.

Выпускная квалификационная работа выполнена в текстовом редакторе Microsoft Word и представлена в твердой копии.

СОДЕРЖАНИЕ

ВВЕДЕНИЕ	5
1. Анализ систем распознавания лиц на фотографии и видео	7
1.1 Анализ опубликованных исследований в рамках задачи.....	7
1.1.1 Системы использующие классические методы машинного обучения	7
1.1.2 Системы с применением глубокого обучения нейронных сетей	9
1.2 Бейзлайн конвейера системы распознавания лиц	13
1.3 Перечень датасетов для обучения и тестирования системы.....	13
2. Разработка модуля «Детекции повторного посещения клиентами магазина по продаже автомобилей».....	14
2.1 Входные данные.....	14
2.2 Текущий уровень решения проблемы	15
2.3 Общая архитектура системы	16
2.4 Развертывание базы данных	16
2.5 Алгоритм работы распознавания лиц и проверки их на уникальность ..	17
2.5.1 Разложение видео на кадры.....	19
2.5.2 Распознавание и трекинг людей на кадре	19
2.5.3 Обрезка изображения	19
2.5.4 Распознавание лица на фрагменте изображения.....	20
2.5.5 Преобразование изображения в вектор	21
2.5.6 Проверка на уникальность.....	21
2.5.7 Запись в БД.....	22
2.6 Тестирование и отладка	22
3. Анализ результатов работы разработанного модуля.....	26
ЗАКЛЮЧЕНИЕ	28
БИБЛЕОГРАФИЧЕСКИЙ СПИСОК.....	30

ВВЕДЕНИЕ

Технологии распознавания лица в последние годы все больше и больше внедряются в нашу жизнь. Сегодня эта функция используется на каждом шагу: в телефонах, ноутбуках, паспортах и платежных приложениях. На цифровом рынке смартфоны, поддерживающие функцию распознавания лица, все больше вытесняют своих более устарелых конкурентов. В то же время технология идентификации лиц превращается в инструмент, который можно использовать с целью повышение эффективности принятия управленческих решений в бизнес-задачах.

В наши дни технология распознавания лиц применяется в следующих сферах: государственная безопасность (выявление нарушителей), маркетинг (выявление у людей заинтересованности в рекламе), мобильные приложения (использование лица для создания персонажей) и т.д.

Цель данной работы является разработанный модуль программного обеспечения, позволяющие определять уникальных клиентов по видеофрагменту посещения клиентами, магазина по продаже автомобилей.

Для достижения цели поставлены следующие задачи:

- Проанализировать текущие исследования в рамках задач распознавания лиц на фото и видео.
- Сформировать архитектуру, позволяющую определять уникальность клиента на видео по его лицу.
- Определить параметры и показатели, тестирования и отладки.
- Проанализировать полученные результаты и сформировать предложения по развитию модуля.

Актуальность работы обусловлена потребностью бизнеса автоматически получать информацию о количестве клиентов, повторно посетивших их магазин, для проведения последующего анализа с целью повышения объемов продаж.

Объектом исследования являются системы распознавания лиц на фото и видео.

Предметом являются нейросетевые архитектуры, применяемые в компьютерном зрении.

При написании данной работы использовались следующие методы: анализ и синтез, эксперимент, измерение и сравнение.

В качестве теоретической и информационно-аналитической базы были использованы труды российских и иностранных авторов по заявленной тематике, информационные материалы, размещенные на сайтах сети Интернет.

1. АНАЛИЗ СИСТЕМ РАСПОЗНАВАНИЯ ЛИЦ НА ФОТОГРАФИИ И ВИДЕО

1.1 Анализ опубликованных исследований в рамках задачи

Подходы к построению системы распознавания лиц (РЛ) можно разделить минимум на две группы – с использованием классических методов машинного обучения и с применением глубокого обучения нейронных сетей.

1.1.1 Системы использующие классические методы машинного обучения

Авторы статьи [1] рассматривают два варианта реализации системы РЛ на фото с использованием классического машинного обучения.

Первая система использует метод сравнения гибких графов (Flexible Graph Comparison).

Основная суть его работы – сравнение двух графов. Каждый из этих графов представляет собой описание лица, то есть хранит информацию в своих ребрах и вершинах. Первый из этих графов является примером, так называемым «эталонным» графом. Он не подлежит изменению. Второй, наоборот, изменчив и способен подстраиваться под шаблон. В этих вариациях графы могут представлять собой как обычные прямоугольные сетки (Рисунок 1), так и более сложные, построенные по форме лица (Рисунок 2).

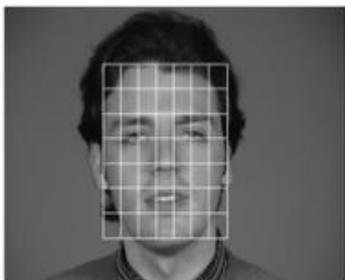


Рисунок 1 - Прямоугольная сетка

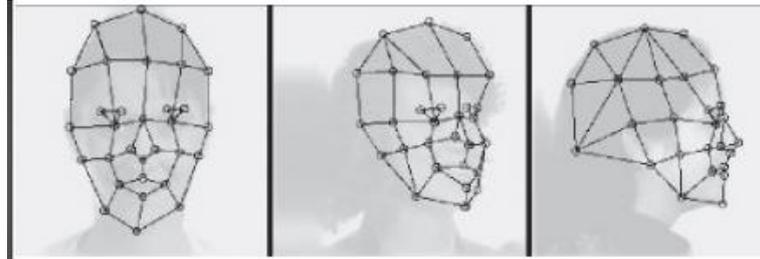


Рисунок 2 - Сетка по форме лица

Граф деформируется по следующему алгоритму.

Каждая из вершин графа смещается со своего исходного местоположения. При этом он перемещается в такое положение, чтобы в двух соответствующих вершинах каждого из графов различие признаков в вершинах стремилось оставаться минимальным. Как только все вершины сместятся на максимально возможное расстояние, при котором достигается минимальная разница между признаками графа, деформация считается завершенной.



Рисунок 3 – Преобразование сетки графа

Хотя этот метод чрезвычайно точен (точность распознавания 95–97%), он используется редко из-за высокой вычислительной сложности и длительного времени работы при большом количестве лиц.

Второй вид систем используют метод главных компонент (Principal Component Analysis).

Для корректной работы с этим методом создается матрица, в которую складываются все лица базы данных, по одному лицу на каждую строку. После этого все данные в матрице нормализуются и по строкам, приведенным к нулю и

первой дисперсии, рассчитывается ковариантная матрица. В полученной матрице решается задача о собственных значениях, после чего формируются группы векторов с присвоенными им собственными значениями.

Этот метод широко используется на практике в различных приложениях. Однако при плохом освещении или недостаточной точности изображения эффективность метода существенно падает.

1.1.2 Системы с применением глубокого обучения нейронных сетей

По мере увеличения требований к вычислительным ресурсам в классических методах распознавания лиц стали применять глубокое обучение, которое заменило вышеупомянутые методы извлечения признаков, разрабатываемые вручную.

В последние годы алгоритмы, основанные на глубоком обучении, стали основным методом исследования обнаружения объектов из-за их превосходной производительности. История их развития представлена на рисунке 4. [2]

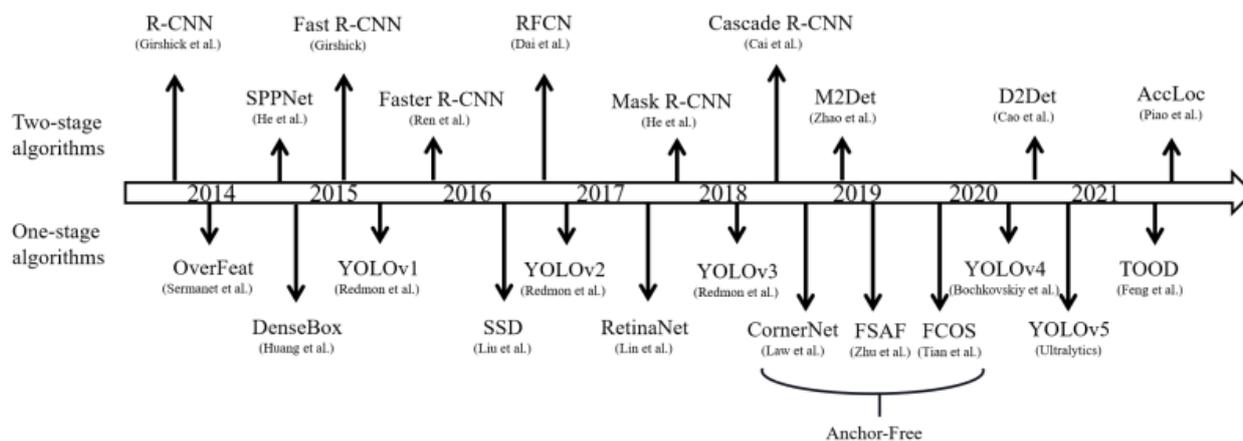


Рисунок 4 – История развития моделей обнаружения объектов на основе глубокого обучения

Далее рассмотрим конвейер системы, использующей нейронные сети для «обучения» и «вывода» идентификатора изображения лица.

В целом конвейер обучения модели содержит два этапа: предобработку изображения лица (предварительная обработка) и обучение модели (обучение). Получение результата РЛ с помощью обученной модели состоит из трех частей: предварительная обработка изображений лица (предварительная обработка), преобразование изображения в вектор (выделение признаков) и распознавание путем сопоставления признаков между тестовым изображением и изображениями с известными метками (распознавание). [3]

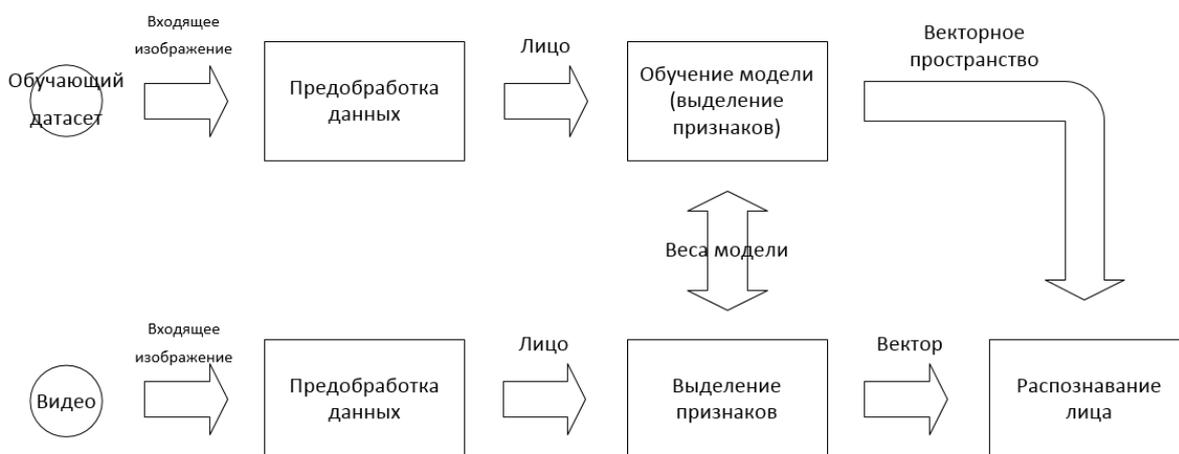


Рисунок 5 – Простая схема конвейера обучения модели и распознавания лица моделью

Предварительная обработка минимально подразделяется на обнаружение лица (Face detection) и выравнивание лица (Face alignment), также в зависимости от требований к системе может включать защиту от подделки лица (Anti-spoofing). Соответственно при наличие защите от подделки лица, блок предварительной обработки на этапе обучения и при РЛ будет отличаться.

При обнаружении лица рассматривая человеческое лицо как объект, можно использовать множество алгоритмов обнаружения объектов для получения отличных результатов обнаружения лиц, например, Faster-RCNN [4], SSD [5], YOLO (с различными версиями [6, 7, 8]).

Однако некоторые исследователи рассматривали человеческое лицо как особый объект и разработали архитектуру переменной глубины для поиска лиц в

реальной жизни с более высокой точностью — Mtcnn [9]. Mtcnn принял каскадную структуру с тремя этапами глубоких сверточных сетей (Р-сеть, R-сеть, O-net), которые прогнозируют местоположение как лица, так и ориентира от грубого до точного. Немаксимальное подавление (Non-maximum suppression (NMS)) используется для объединения сильно перекрывающихся кандидатов, которые создаются Р-сетью, R-сетью и O-сетью.

RetinaFace [10] — еще один детектор лиц, использующий FPN для извлечения многоуровневых признаков изображения. Однако в отличие от двухэтапных методов обнаружения, таких как SRN, RetinaFace представляет собой одноэтапный детектор лиц. Более того, RetinaFace выполняла многозадачное распознавание лиц. В частности, RetinaFace предсказала одновременно следующие 4 аспекта: оценку лица, рамку лица, пять ориентиров лица и плотные трехмерные вершины лица, проецируемые на плоскость изображения. Конвейер RetinaFace можно увидеть на рис. 3.

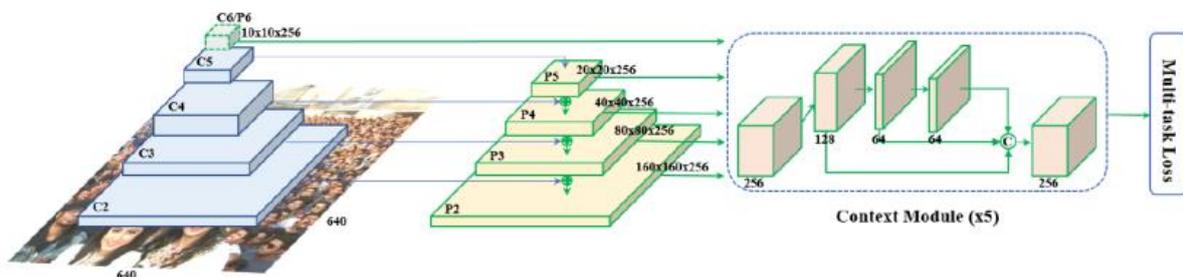


Рисунок 3 – Конвейер RetinaFace

Распространенный способ выравнивания изображений лиц — использование 2D-преобразования для калибровки ориентиров лица по заранее заданным фронтальным шаблонам или среднему режиму лица.

STN (Spatial Transform Networks) [11] представил обучаемый модуль, который позволяет манипулировать изображением внутри сети, а также может использоваться для выравнивания лиц. Объединённое выравнивание лица на основе STN и экстрактор функций лица содержит рекурсивный пространственный преобразователь (Recursive Spatial Transformer (ReST)) для

сквозного обучения при РЛ с выравниванием. 2D-выравнивание лица выполняется быстрее, чем 3D-выравнивание в STN и поэтому оно широко используется на этапе предварительной обработки РЛ.

АРА [12] предложил более общий метод 2D-выравнивания лица. Вместо того, чтобы выравнивать все лица по почти фронтальной форме, АРА адаптивно изучила несколько шаблонов, специфичных для поз, которые сохраняли внешний вид лица с меньшими артефактами и потерей информации.

При достаточном наличии вычислительных мощностей оборудования, можно использовать любую современную модель для построения системы РЛ. Различные модели с одинаковым количеством параметров имеют одинаковую точность по РЛ. В результате мы можете использовать Resnet, ResNext, SEResnet, Inception net, Densenet и т. д. для формирования нашей системы РЛ. Кроме того, NAS можно использовать для поиска лучших гиперпараметров сети.

Последняя часть конвейера РЛ — это распознавание лиц путем сравнения векторов лиц. Чтобы применить РЛ, необходимо построить галерею лиц. Сначала у нас есть набор идентификаторов лиц S , где каждый идентификатор содержит одно (или несколько) изображений лица. Все встраивания лиц из изображений галереи будут извлечены обученной моделью и сохранены в базе данных.

За исключением классификации, основанной на softmax, РЛ можно также рассматривать как извлечение особенностей лица и выполнение сопоставления признаков. Таким образом, обучение модели РЛ может стать процессом изучения компактного евклидова пространства признаков, где расстояние напрямую соответствует мере сходства лиц. Это базовое моделирование метрического обучения.

В работе [13] использовали косинусное сходство, чтобы получить часть наиболее похожих лиц изображения запроса. Затем для повторного ранжирования этих сходств и получения окончательных результатов идентификации было использовано поэтапное расстояние Earth Mover's Distance.

1.2 Бейзлайн конвейера системы распознавания лиц

По итогам анализа бейзлайн конвейера обучения будет выглядеть как представлено на рисунке 4.

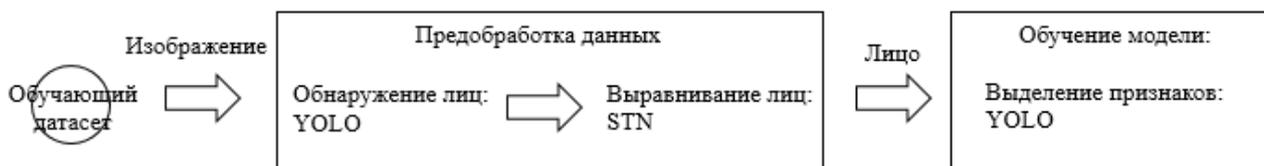


Рисунок 4 - Бейзлайн конвейера обучения

На рисунке 5 представлен бейзлайн конвейера РЛ.

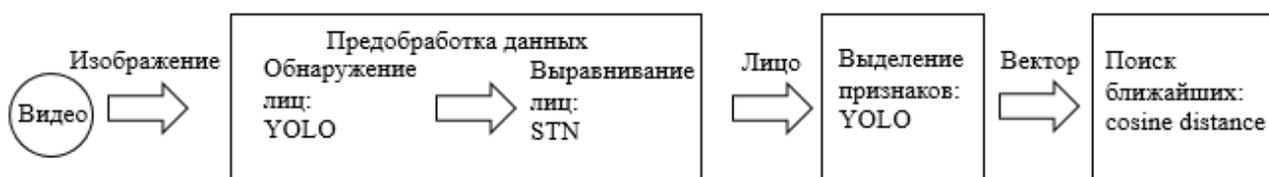


Рисунок 5 - Бейзлайн конвейера РЛ

1.3 Перечень датасетов для обучения и тестирования системы

Таблица 1 – Информация о тренировочных датасетах

Наименование	Кол-во уникальных лиц, шт	Изображений, шт
CASIA-Webface	10575	494414
UMDFaces	8277	367888
vggface2(VGG2)	9131	3,33 млн

Таблица 2 – Информация о тестовых датасетах

Наименование	Кол-во уникальных лиц, шт.	Изображений, шт.
LFW	5749	13233
MORPH Album 2	20000	78000

2. РАЗРАБОТКА МОДУЛЯ «ДЕТЕКЦИИ ПОВТОРНОГО ПОСЕЩЕНИЯ КЛИЕНТАМИ МАГАЗИНА ПО ПРОДАЖЕ АВТОМОБИЛЕЙ»

2.1 Входные данные

В ходе исследования исходных данных к решаемой задаче, был установлен ряд ограничений. Камера, с которой берется видеофрагмент, установлена достаточно далеко от входа и смотрит на весь шоурум в целом, пример кадра представлен на рисунке 6.



Рисунок 6 – Пример кадра с камеры в шоуруме автосалона

Из первого ограничения следует, что лица людей, находящихся в верхней половине кадра, получаются достаточно плохого качества. В кадре постоянно находятся автомобили, которые, загораживают обзор.

В ходе исследования требуется найти такое решение, которое, позволяет на основе алгоритма или применяемых фреймворков обойти эти ограничения.

2.2 Текущий уровень решения проблемы

Задача предоставлена партнером ООО «АйПиЭсТи», с описанием базового решения в формате – этот модуль распознает лица с камер и сопоставляет его с лицами в базе данных (БД), куда занесены менеджеры и хостес, тем самым рассчитывая количество клиентов.

Блок схема работы базового алгоритма представлена на рисунке 7.



Рисунок 7 – Блок схема работы базового алгоритма

Данные хранятся в БД в таблице, параметры которой указаны в таблице 3.

Таблица 3 – Информация о таблице, используемой в работе базового алгоритма

№	Наименование	Тип	Функция
1	id	uuid NOT NULL DEFAULT	gen_random_uuid()
2	URL	VARCHAR	
3	time_created	TIMESTAMP NOT NULL DEFAULT	current_timestamp
4	time_updated	TIMESTAMP NOT NULL DEFAULT	current_timestamp

Технологический стек компании партнера, в рамках данной задачи:

- YOLO;
- OpenCV;
- Kornia.

2.3 Общая архитектура системы

Обобщенное описание модуля сформулировано как – необходимо реализовать модуль проверки лиц на уникальность. В базе данных хранится список уникальных клиентов: если найденное в кадре лицо не совпадает ни с одним из лиц базы — то это уникальный клиент, предполагается в результате получить список новых клиентов и добавить их в базу.

На основании этого описания и теории изученной в 1ой главе, была сформирована архитектура модуля, представленная на рисунке 8.

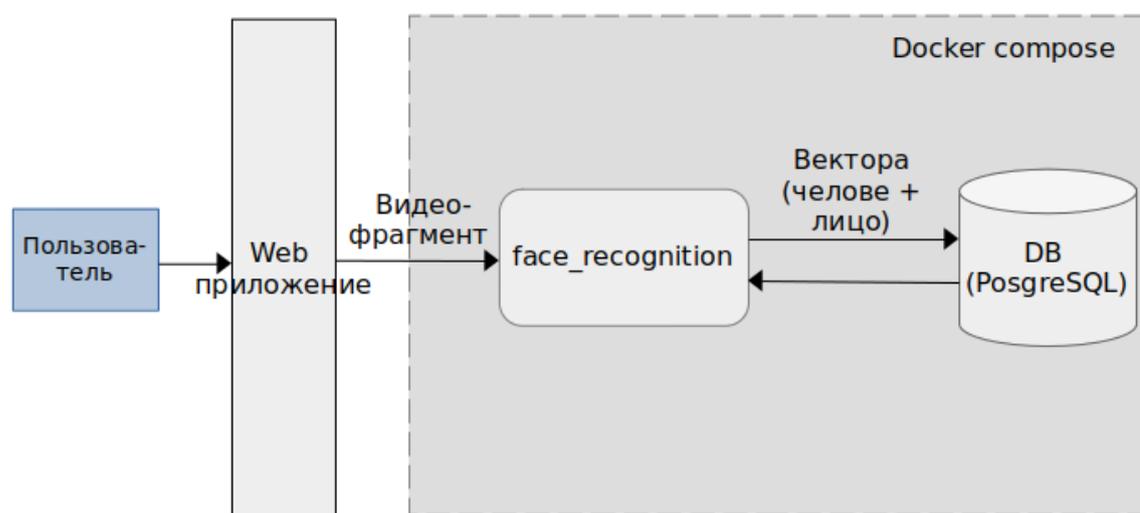


Рисунок 8 – Архитектура модуля детекции повторного посещения клиентами магазина по продаже автомобилей

Исходя из того, что на вход приходи уже готовы видеофрагмент, достаточно реализовать алгоритм РЛ, который напрямую общается с базой данных (БД), записывает туда уникальных клиентов и берет оттуда информацию для проверки клиентов на уникальность. Работа модуля разворачивается в docker compose [15].

2.4 Развертывание базы данных

Никаких технических требований к БД не предъявлялось, PostgreSQL [16] был выбран исходя из того, что сейчас становится все более распространенным инструментом и разработчик Российская компания.

В рамках установленного функционала, для реализации достаточно одной таблицы, её параметры указаны в Таблице 4.

Таблица 4 – Информация о таблице, используемой для хранения уникальных клиентов

№	Наименование	Тип	Функция
1	time_created	TIMESTAMP NOT NULL DEFAULT	current_timestamp
2	time_updated	TIMESTAMP NOT NULL DEFAULT	current_timestamp
3	id	uuid NOT NULL DEFAULT	gen_random_uuid()
4	person_id	uuid NOT NULL	
5	person_vector	public.vector	
6	face_vector	public.vector	
7	face_detection_conf	FLOAT	

Следует обратить внимание на столбцы person_vector и face_vector. Необходимо, чтобы им был присвоен тип public.vector, с целью дальнейшего хранения изображений сразу в формате вектора. Для использование типа public.vector, следует установить расширение к PostgreSQL - pgvector [17].

2.5 Алгоритм работы распознавания лиц и проверки их на уникальность

В ходе разработки алгоритма для модуля «Детекции повторного посещения клиентами магазина по продаже автомобилей», было принято решение отказаться от блока выравнивания лица (face alignment) из-за качества входных данных, выравнивание лица скорее вредило, чем улучшало результаты распознавания

Основной идеей всего алгоритма является — распознавание лица с привязкой к человеку в кадре и последующее сравнение его пар векторов, то есть мы в конечном итоге проверяем близость векторов не только лица человека на кадре, но и в целом всего человека.

Основные блоки алгоритма представлены на рисунке 9.

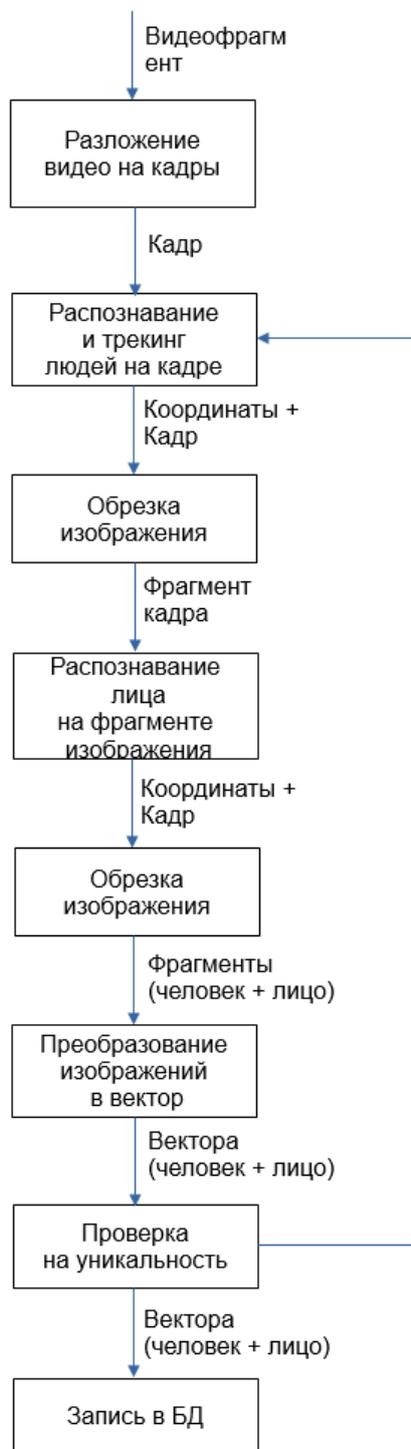


Рисунок 9 – Блок схема алгоритма распознавания лиц и проверки их на уникальность

2.5.1 Разложение видео на кадры

Для получения кадров использую библиотеку OpenCV [18], также тут запускаю цикл, в котором считываю видео покадрово и далее буду осуществлять все действия применимо к конкретному кадру.

2.5.2 Распознавание и трекинг людей на кадре

Для распознавания и трекинга использую предобученные модели YOLOv8 [19], с использованием ограничения по классам, устанавливаю только класс — person. На выходе получаю координаты боксов всех людей в кадре и их трек id, по которому отслеживаю нахождения человека в кадре. Пример результат показан на рисунке 10.



Рисунок 10 – Результат работы блока «Распознавание и трекинг людей на кадре», алгоритма распознавания лиц и проверки их на уникальность

2.5.3 Обрезка изображения

Для реализации данного блока алгоритма, снова использую OpenCV. Тут появляется еще один цикл, в котором я вырезаю фрагмент изображения распознанного человека. Результат работы представлен на рисунке 11.



Рисунок 11 – Результат работы блока «Обрезка изображения», алгоритма распознавания лиц и проверки их на уникальность

2.5.4 Распознавание лица на фрагменте изображения

Для данного распознавания также использую модель YOLOv8, только эта предварительно обучена на датасетах CASIA-Webface, UMDFaces, vggface2(VGG2). На выходе получаю только координаты боксов лица. Пример результата представлен на рисунке 12.



Рисунок 12 – Результат работы блока «Распознавание лица на фрагменте изображения», алгоритма распознавания лиц и проверки их на уникальность

Далее следует снова обрезка изображения, процесс и результат аналогичны п. 2.3.3

2.5.5 Преобразование изображения в вектор

Для преобразования изображения в вектор я использую стандартный метод YOLOv8. Единственно, что метод необходимо вызвать применимо к той модели с помощью, которой, был найден этот объект, то есть если мне надо преобразовать в вектор фрагмента с лицом, то и метод я вызываю от модели, которая распознает именно лицо. Результатом работы данного блока является вектор размерностью 1 x 658.

2.5.6 Проверка на уникальность

В этом блоке мной реализовано несколько условий проверки. В первую очередь проверяю на уникальность внутри самой программы, сверяя записывали ли я такой трэк id, это позволяет значительно уменьшить количество проверяемых изображений.

Вторым происходит проверка наличия схожих изображений в БД, вычисляя косинусное расстояние (1) между векторами изображений всего человека и лица в отдельности.

Косинусное расстояние между двух векторов считается как разница единицы и скалярного произведения этих векторов, деленное на произведение длин этих векторов.

$$\text{cosine dist} = 1 - \frac{(\vec{a}, \vec{b})}{|\vec{a}| * |\vec{b}|}, \quad (1)$$

где α – угол между двумя векторами;

\vec{a} – один вектор, в нашем случае это вектор фрагмента изображения, полученного в ходе распознавания;

\vec{b} – второй вектор, в нашем случае это вектор подставляемый из БД, либо алгоритма.

Третье условие проверки, проверяется насколько модель уверена, что этот именно лицо. Если похожий человек уже есть БД и модель нашла его лицо, которое, она с большей вероятностью характеризует именно как лицо, лучше поместить в БД новое лицо.

Исходя из этих условий и записываются результаты в БД.

2.5.7 Запись в БД

В это пункте стоит отметить, что запись в БД осуществляет в двух местах это в момент, когда мы нашли абсолютно нового человека и в конце выполнения алгоритма. Такая реализация сделана с целью уменьшить время выполнения и избежать постоянных запросов к БД.

2.6 Тестирование и отладка

Для тестирования и отладки работы алгоритма я выбрал 5 переменных максимально допустимы значения косинусных расстояний для векторов изображения человека в целом и изображения лица, уверенность модели в том, что это именно детектируемый объект, аналогично, отдельно для модели распознавания человека отдельно для модели распознавания лица человека и веса модели, которая, распознает людей.

Оптимальные значения косинусных расстояний должны обеспечить точность определения уникальности найденного моделью человека и его лица. Диапазон проверки значений от 0.05 до 0.25 с шагом 0.05.

Оптимальные значения уверенности модели и их комбинация с различными весами модели, должны обеспечить стабильность отслеживания отдельного человека в кадре и получение изображения человека и его лица более высокого качества. Диапазон проверки значений от 0.25 до 0.8 с шагом 0.05.

Для проверки результативности алгоритма использовались 4 видео, сводная информация по которым указана в таблице 5.

Таблица 5 – Сводная информация о видео, применяемых для проверки результативности алгоритма

Наименование	Длительность, сек	Кол-во уникальных людей, шт	Количество повторяющихся людей, шт
Видео 1	220	6	0
Видео 2	60	4	0
Видео 3	60	4	1
Видео 4	150	3	3

Оценивать будем по простым метрикам, нам важно, чтобы были обнаружены все уникальные лица на фрагменте, пусть эта метрика называется полнота (2) и является отношением фактического количества уникальных лиц, к количеству обнаруженных лиц алгоритмом.

$$p = \frac{x}{y}, \quad (2)$$

где p – полнота;

x – количество уникальных лиц, обнаруженных алгоритмом на видеофрагменте, шт;

y – фактическое количество уникальных лиц, обнаруженных алгоритмом на видеофрагменте, шт.

Значение меньше единицы свидетельствует о наличие лишних обнаруженных, больше единицы – значит алгоритм, кого-то не обнаружил.

Также введем показатель, который поможет нам оценить качество работы алгоритма в рамках определения уникальности человека, попавшего в кадр, назовем его точность (3) и будем рассчитывать его как отношение верно повторно записанных значений ко всему количество повторно записанных значений.

$$l = \frac{n}{m}, \quad (3)$$

где l – точность;

n – количество верно повторно записанных строк в БД, шт;

m – количество все повторно записанных строк в БД, шт.

Косвенно будем оценивать по времени обработки всех видео алгоритмом.

Введем коэффициент времени выполнения (4) равный отношению времени обработки алгоритмом видео в секундах к реальной длительности видео в секундах.

$$k = \frac{t_a}{t_r}, \quad (4)$$

где k – коэффициент времени выполнения;

t_a – времена обработки видео алгоритмом, сек;

t_r – реальная длительности времени, сек.

Для реализации в общей системе, чем меньше коэффициент, тем лучше, т.к. предполагается, что анализу будут подвергаться видео длительность несколько недель. В дальнейшем при демонстрации работы заказчику, может негативно повлиять то факт, что алгоритм обрабатывает видео столько же времени сколько длится само видео, или что хуже того - дольше.

По результатам работы алгоритма, данные сводились в таблицу, по которой уже визуальнo искал ошибки. Проверял на наличие повторных записей под одним и тем же id, также ошибкой признавалось, когда на видео присутствует человек, ранее попадавший в БД, а алгоритм при этом не перезаписал его.

Пример выполнения такой оценки рассмотрим на основе выдержки из результатов работы, представлен в таблице 5.

Таблица 5 – Выдержка из сводной таблицы оценки результатов работы алгоритма

Видео	Уникальный идентификатор человека	Фрагмент
Видео 1	f704f49d-d1b3-45f1-a1ca-34c5a6a3df3e	
	acb9333d-2843-419f-a4f1-7fdaa662a3d0	
	bd979b16-dd5f-4ba5-ad53-02d819638a51	
	<u>c1d670e0-13a8-11ef-a31a-b94d60cec96e</u>	

Видео 4	<u>c1d670e0-13a8-11ef-a31a-b94d60cec96e</u>	

В таблице 5 видно, уникальный идентификатор человека (выделен подчеркиванием) двух записей одинаковый, а на фрагменте изображены различные люди, что свидетельствует о ошибке возникшей в ходе сравнения косинусного расстояния.

3. АНАЛИЗ РЕЗУЛЬТАТОВ РАБОТЫ РАЗРАБОТАННОГО МОДУЛЯ

По итогам отладки и тестирования различных наборов параметров и весов, я свел всю полученную информацию и ранжировал в порядке убывания по метрике точность. Результат представлен в таблице 6.

Таблица 6 – Ранжированные результаты тестирования работы модуля «Детекция повторного посещения клиентами магазина по продаже автомобилей»

№	Модель yolo	Косинусное расстояние		Полнота	Точность	Коэффициент времени выполнения
		Для человека	Для лица			
1	yolo8n	0.1	0.1	1.133	0.833	1.002
2	yolo8n	0.2	0.05	0.895	0.692	0.963
3	yolo8n	0.15	0.05	0.85	0.637	0.967
4	yolo8n	0.1	0.05	0.708	0.556	0.984
5	yolo8n	0.05	0.2	0.654	0.429	1.049
6	yolo8n	0.05	0.15	0.654	0.429	1
7	yolo8n	0.05	0.1	0.667	0.333	0.984
8	yolo8n	0.05	0.05	0.362	0	0.967
9	yolo8s	0.1	0.1	0.515	0	1.12
10	yolo8s	0.15	0.05	0.378	-	1.143
11	yolo8s	0.2	0.05	0.395	-	1.131
12	yolo8m	0.1	0.1	0.378	-	1.214
13	yolo8m	0.2	0.05	0.236	-	1.356
14	yolo8m	0.15	0.05	0.287	-	1.296
- указан там где, не было возможности рассчитать данный показатель, из-за деления на 0.						

В качестве отображений результатов для моделей с большими весами, чем уolo8n я ограничился вывод трех результатов, для наглядности отображения, что увеличение размерности модели приводит к моментальному падению точности алгоритма, что сопровождается падением полноты и увеличением коэффициента времени исполнения.

На данном этапе реализации модуля «Детекции повторного посещения клиентами магазина по продаже автомобилей» достаточно полученной точности и полноты полученных с использованием модели uolo8n и косинусными расстояниями для вектора человека и вектора лица 0.1 и 0.1 соответственно.

При условии, что на текущем этапе развития системы, данные показатели модуля «Детекции повторного посещения клиентами магазина по продаже автомобилей» удовлетворяют потребностям, остаются и четко видны возможности улучшения и развития, такие как:

- Разметка кадров, получаемых с этой камеры и дообучение моделей на реальных данных (допустим с этой же камеры);
- Для детекции лица на фрагменте изображения использовать архитектуру mtcnn [9];
- Помимо векторов, хранить координаты 5 основных лицевых точек – центры глаз, кончик носа, уголки рта и использовать их для определения уникальности клиента.

ЗАКЛЮЧЕНИЕ

Результаты полученные при выполнении исследования в рамках данной ВКР, являются основой для построения архитектуры и разработки модуля «Детекция повторного посещения клиентами магазина по продаже автомобилей». Сформулированные и проверенные гипотезы позволят в дальнейшем грамотно развивать архитектуру и алгоритм заложенные в основе модуля.

По результатам выполнения поставленных задач были достигнуты следующие результаты:

- Разработана архитектура модуля детекции повторного посещения клиентами магазина по продаже автомобилей;
- Разработан алгоритм распознавания лиц и проверки их на уникальность;
- Определены 5 переменных, для проведения тестирования и отладки алгоритма;
- Определены 3 метрики, для проверки качества работы алгоритма;
- По результатам отладки и тестирования выделена оптимальная модель с параметрами:
 - а) Модели $u_0 \otimes n$;
 - б) Косинусное расстояний для векторов изображения человека в целом – 0.1;
 - б) Косинусное расстояний для векторов изображения лица человека – 0.1;
 - в) $konf$ модели распознавания человека – 0.25;
 - г) $konf$ модели распознавания лица – 0.5.

Результативность оптимальной модели:

- Полнота – 1.133;
- Точность – 0.733;

- Коэффициент времени выполнения - 1.002.

Код и рекомендации по сборке модуля опубликованы в открытом доступе по ссылке – https://github.com/ViktorRtm/diploma_face_recognition.

В итоге представлена техническое решение, готовое к интеграции в имеющуюся структуру проекта, а также определены направления развития этого решения, такие как:

- Разметка кадров, получаемых с этой камеры и дообучение моделей на реальных данных, с добавлением аугментации [20, 21];
- Для детекции лица на фрагменте изображения использовать архитектуру mtcnn [9];
- Помимо векторов хранить, маски людей/лиц полученные в результате сегментации [22], координаты 5 основных лицевых точек – центры глаз, кончик носа, уголки рта и использовать их для определения уникальности клиента [23].

БИБЛЕОГРАФИЧЕСКИЙ СПИСОК

1. I.L. Grinin. Analysis of Methods and Principles of Face Recognition Systems Based on Artificial Intelligence // Перспективы науки. – 2022. – №5(152). – С. 25 – 30.
2. Sheng Xu, Zhanyu Guo, Yuchi Liu, Jingwei Fan, Xuxu Liu. An Improved Lightweight YOLOv5 Model Based on Attention Mechanism for Face Mask Detection // arXiv 2022: <http://arxiv.org/abs/2203.16506>. – Дата обращения 26.04.2024.
3. Xinyi Wang, Jianteng Peng, Sufang Zhang, Bihui Chen, Yi Wang, Yandong Guo. A Survey of Face Recognition // arXiv 2022: <http://arxiv.org/abs/2212.13038>. (дата обращения: 26.04.2024).
4. S. Ren, K. He, R. Girshick, J. Sun. Faster r-cnn: Towards real-time object detection with region proposal networks // Advances in neural information processing systems. – 2015. – № 28. – С. 91–99.
5. W. Liu, D. Anguelov, D. Erhan, C. Szegedy, S. Reed, C.-Y. Fu, A. C. Berg. Ssd: Single shot multibox detector // European conference on computer vision. – 2016. – № 9905. – С. 21–37.
6. J. Redmon, S. Divvala, R. Girshick, A. Farhadi. You only look once: Unified, real-time object detection // Proceedings of the IEEE conference on computer vision and pattern recognition. – 2016. – С. 779–788.
7. J. Redmon, A. Farhadi. Yolo9000: better, faster, stronger // Proceedings of the IEEE conference on computer vision and pattern recognition. – 2017. – С. 6517–6525.
8. J. Redmon, A. Farhadi. Yolov3: An incremental improvement // arXiv 2018: <http://arxiv.org/abs/1804.02767> (дата обращения: 03.05.2024).
9. K. Zhang, Z. Zhang, Z. Li, Y. Qiao. Joint face detection and alignment using multitask cascaded convolutional networks // IEEE Signal Processing Letters. – 2016. – № 23. – С. 1499–1503.

10. J. Deng, J. Guo, Y. Zhou, J. Yu, I. Kotsia, S. Zafeiriou. Retinaface: Single-stage dense face localisation in the wild // arXiv 2019: <http://arxiv.org/abs/1905.00641> (дата обращения: 03.05.2024).
11. M. Jaderberg, K. Simonyan, A. Zisserman. Spatial transformer networks // Advances in neural information processing systems. – 2015. – № 28. – С. 2017–2025.
12. Z. An, W. Deng, Y. Zhong, Y. Huang, X. Tao. Apa: Adaptive pose alignment for robust face recognition // Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops 2019: <https://doi.org/10.1109/ACCESS.2019.2894162> (дата обращения: 03.05.2024).
13. H. Phan, A. Nguyen. Deepface-emd: Re-ranking using patch-wise earth mover's distance improves out-of-distribution face identification // Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. – 2022. – С. 20259–20269.
14. Xin Wang, Hui Guo, Shu Hu, Ming-Ching Chang, Siwei Lyu. GAN-generated Faces Detection: A Survey and New Perspectives // arXiv 2023: <http://arxiv.org/abs/2202.07145> (дата обращения: 03.05.2024).
15. Docker Compose overview: <https://docs.docker.com/compose/> (дата обращения: 19.04.2024).
16. PostgreSQL 15.7 Documentation: <https://www.postgresql.org/docs/15/index.html> (дата обращения: 03.05.2024).
17. Pgvector Documentation, Installation guide: <https://github.com/pgvector/pgvector> (дата обращения: 03.05.2024).
18. OpenCV 4.10.0 Documentation: <https://docs.opencv.org/4.x/> (дата обращения: 12.04.2024)
19. Ultralytics YOLOv8 Docs: <https://docs.ultralytics.com/> (дата обращения: 12.04.2024).

20. Boutros F, Klemm M, Fang M, Kuijper A, Damer N. Unsupervised Face Recognition using Unlabeled Synthetic Data // arXiv 2022: <http://arxiv.org/abs/2211.07371> (дата обращения: 17.05.2024).
21. Osama R. Shahin, Rami Ayedi, Alanazi Rayan, Rasha M. Abd El-Aziz, Ahmed I. Taloba. Human Face Recognition from Part of a Facial Image based on Image Stitching // International Journal of Advanced Computer Science and Applications. – 2021. – №12. – С. 456–463.
22. Hassani A, Shair ZE, Ud Duala Refat R, Malik H. Distilling Facial Knowledge with Teacher-Tasks: Semantic-Segmentation-Features for Pose-Invariant Face-Recognition // 2022 IEEE International Conference on Image Processing (ICIP). – 2022. – С. 741–745.
23. Mojtaba Kollahdouzi, Alireza Sepas-Moghaddam, and Ali Etemad. Face Trees for Expression Recognition // arXiv 2021: <https://arxiv.org/pdf/2112.02487> (дата обращения: 17.05.2024).