

Министерство науки и высшего образования Российской Федерации
федеральное государственное автономное
образовательное учреждение высшего образования
«Национальный исследовательский Томский политехнический университет» (ТПУ)

Школа Инженерная школа информационных технологий и робототехники
Направление подготовки 09.04.04 Программная инженерия
ООП/ОПОП Технологии больших данных
Отделение школы (НОЦ) Информационных технологий

ВЫПУСКНАЯ КВАЛИФИКАЦИОННАЯ РАБОТА МАГИСТРАНТА

Тема работы
Реализация методов классификации людей по полу и возрасту и их повторной идентификации в видеопотоке с помощью технологий глубокого обучения

УДК 004.421:004.415.2:004.62:61

Обучающийся

Группа	ФИО	Подпись	Дата
8ПМ11	Семенюта Антон Вадимович		10.06.2023 г.

Руководитель ВКР

Должность	ФИО	Ученая степень, звание	Подпись	Дата
доцент ОИТ ИШИТР	Аксёнов С. В.	К.Т.Н.		10.06.2023 г.

КОНСУЛЬТАНТЫ ПО РАЗДЕЛАМ:

По разделу «Финансовый менеджмент, ресурсоэффективность и ресурсосбережение»

Должность	ФИО	Ученая степень, звание	Подпись	Дата
профессор ОСГН ШБИП	Жиронкин С.А.	Д.Э.Н.		

По разделу «Социальная ответственность»

Должность	ФИО	Ученая степень, звание	Подпись	Дата
профессор ООД ШБИП	Федорчук Ю.М.	Д.Т.Н.		

ДОПУСТИТЬ К ЗАЩИТЕ:

Руководитель ООП, должность	ФИО	Ученая степень, звание	Подпись	Дата
доцент ОИТ ИШИТР	Губин Е. И.	к.ф.-м.н.		

Томск – 2023 г.

ПЛАНИРУЕМЫЕ РЕЗУЛЬТАТЫ ОСВОЕНИЯ ООП
по направлению 09.04.04 «Программная инженерия»

Код компетенции	Наименование компетенции
Универсальные компетенции	
УК(У)-1	Способен осуществлять критический анализ проблемных ситуаций на основе системного подхода, вырабатывать стратегию действий
УК(У)-2	Способен управлять проектом на всех этапах его жизненного цикла
УК(У)-3	Способен организовывать и руководить работой команды, вырабатывая командную стратегию для достижения поставленной цели
УК(У)-4	Способен применять современные коммуникативные технологии, в том числе на иностранном (-ых) языке (-ах), для академического и профессионального взаимодействия
УК(У)-5	Способен анализировать и учитывать разнообразие культур в процессе межкультурного взаимодействия
УК(У)-6	Способен определять и реализовывать приоритеты собственной деятельности и способы ее совершенствования на основе самооценки
Общепрофессиональные компетенции	
ОПК(У)-1	Способен самостоятельно приобретать, развивать и применять математические, естественно-научные, социально-экономические и профессиональные знания для решения нестандартных задач, в том числе в новой или незнакомой среде и в междисциплинарном контексте
ОПК(У)-2	Способен разрабатывать оригинальные алгоритмы и программные средства, в том числе с использованием современных интеллектуальных технологий, для решения профессиональных задач
ОПК(У)-3	Способен анализировать профессиональную информацию, выделять в ней главное, структурировать, оформлять и представлять в виде аналитических обзоров с обоснованными выводами и рекомендациями
ОПК(У)-4	Способен применять на практике новые научные принципы и методы исследований
ОПК(У)-5	Способен разрабатывать и модернизировать программное и аппаратное обеспечение информационных и автоматизированных систем
ОПК(У)-6	Способен самостоятельно приобретать с помощью информационных технологий и использовать в практической деятельности новые знания

Код компетенции	Наименование компетенции
	и умения, в том числе в новых областях знаний, непосредственно не связанных со сферой деятельности
ОПК(У)-7	Способен применять при решении профессиональных задач методы и средства получения, хранения, переработки и трансляции информации посредством современных компьютерных технологий, в том числе, в глобальных компьютерных сетях
ОПК(У)-8	Способен осуществлять эффективное управление разработкой программных средств и проектов
Профессиональные компетенции	
ПК(У)-1	Способен к созданию вариантов архитектуры программного средства
ПК(У)-2	Способен разрабатывать и администрировать системы управления базами данных
ПК(У)-3	Способен управлять процессами и проектами по созданию (модификации) информационных ресурсов
ПК(У)-4	Способен проектировать и организовывать учебный процесс по образовательным программам с использованием современных образовательных технологий
ПК(У)-5	Способен осуществлять руководство разработкой комплексных проектов на всех стадиях и этапах выполнения работ

Министерство науки и высшего образования Российской Федерации
федеральное государственное автономное
образовательное учреждение высшего образования
«Национальный исследовательский Томский политехнический университет» (ТПУ)

Школа Инженерная школа информационных технологий и робототехники
Направление подготовки 09.04.04 Программная инженерия
ООП/ОПОП Технологии больших данных
Отделение школы (НОЦ) Информационных технологий

УТВЕРЖДАЮ:

Руководитель ООП

_____ Губин Е. И.
(подпись) (дата) (Ф.И.О.)

ЗАДАНИЕ на выполнение выпускной квалификационной работы

Обучающийся:

Группа	ФИО
8ПМ11	Семенюта Антон Вадимович

Тема работы:

Реализация методов классификации людей по полу и возрасту и их повторной идентификации в видеопотоке с помощью технологий глубокого обучения	
Утверждена приказом директора (дата, номер)	№ 146-39/с от 26.05.2023 г.

Срок сдачи обучающимся выполненной работы:	10.06.2023 г.
--	---------------

ТЕХНИЧЕСКОЕ ЗАДАНИЕ:

Исходные данные к работе <i>(наименование объекта исследования или проектирования; производительность или нагрузка; режим работы (непрерывный, периодический, циклический и т. д.); вид сырья или материал изделия; требования к продукту, изделию или процессу; особые требования к особенностям функционирования (эксплуатации) объекта или изделия в плане безопасности эксплуатации, влияния на окружающую среду, энергозатратам; экономический анализ и т. д.)</i>	Объектом исследования является алгоритмы классификации людей по полу и возрасту и их повторной идентификации.
Перечень разделов пояснительной записки, подлежащих исследованию, проектированию и разработке <i>(аналитический обзор по литературным источникам с целью выяснения достижений мировой науки техники в рассматриваемой области; постановка задачи исследования, проектирования, конструирования; содержание процедуры исследования, проектирования, конструирования; обсуждение результатов выполненной</i>	1. Анализ существующих аналогов решения. 2. Разработка алгоритмов классификации пола и возраста и повторной идентификации людей. 3. Проектирование и разработка программных средств. 4. Разработка и разметка обучающих наборов данных. 5. Проведение испытаний алгоритмов.

работы; наименование дополнительных разделов, подлежащих разработке; заключение по работе)	6. Работа над разделом по финансовому менеджменту, ресурсоэффективности и ресурсосбережения. 7. Работа над разделом по социальной ответственности.
Перечень графического материала (с точным указанием обязательных чертежей)	1. Схемы пайплайнов обработки данных. 2. Примеры изображений-данных пайплайнов. 3. Блок-схемы алгоритмов. 4. Диаграммы классов. 5. Матрица SWOT-анализа. 6. Диаграмма Ганта.
Консультанты по разделам выпускной квалификационной работы (с указанием разделов)	
Раздел	Консультант
Основная часть	доцент ОИТ ИШИТР, к.т.н., доцент Аксёнов С. В.
Финансовый менеджмент, ресурсоэффективность и ресурсосбережение	профессор ОСГН ШБИП, д.э.н., профессор Жиронкин С.А.
Социальная ответственность	профессор ООД ШБИП, д.т.н., профессор Федорчук Ю.М.
Раздел на английском языке	старший преподаватель ОИЯ ШБИП, к.ф.н., старший преподаватель Куркан Н.В.
Названия разделов, которые должны быть написаны на иностранном языке:	
Literature Review	

Дата выдачи задания на выполнение выпускной квалификационной работы по линейному графику	1.03.2023 г.
---	--------------

Задание выдал руководитель ВКР:

Должность	ФИО	Ученая степень, звание	Подпись	Дата
доцент ОИТ ИШИТР	Аксёнов С. В.	к.т.н., доцент		1.03.2023 г.

Задание принял к исполнению обучающийся:

Группа	ФИО	Подпись	Дата
8ПМ11	Семенюта Антон Вадимович		1.03.2023 г.

Министерство науки и высшего образования Российской Федерации
 федеральное государственное автономное
 образовательное учреждение высшего образования
 «Национальный исследовательский Томский политехнический университет» (ТПУ)

Инженерная школа Информационных технологий и робототехники
 Направление подготовки (ООП / ОПОП) 09.04.04 Программная инженерия
 Уровень образования магистратура
 Отделение школы (НОЦ) Информационных технологий
 Период выполнения весенний семестр 2022 /2023 учебного года

КАЛЕНДАРНЫЙ РЕЙТИНГ-ПЛАН выполнения выпускной квалификационной работы

Обучающийся:

Группа	ФИО
8ПМ11	Семенюта Антон Вадимович

Тема работы:

Реализация методов классификации людей по полу и возрасту и их повторной идентификации в видеопотоке с помощью технологий глубокого обучения
--

Срок сдачи обучающимся выполненной работы:	10.06.2023 г.
--	---------------

Дата контроля	Название раздела (модуля) / вид работы (исследования)	Максимальный балл раздела (модуля)
10.06.2023	Основная часть	70
10.06.2023	Финансовый менеджмент, ресурсоэффективность и ресурсосбережение	10
10.06.2023	Социальная ответственность	10
10.06.2023	Раздел на английском языке	10

СОСТАВИЛ:

руководитель ВКР

Должность	ФИО	Ученая степень, звание	Подпись	Дата
доцент ОИТ ИШИТР	Аксёнов С. В.	к.т.н.		

СОГЛАСОВАНО:

руководитель ООП

Должность	ФИО	Ученая степень, звание	Подпись	Дата
доцент ОИТ ИШИТР	Губин Е. И.	к.ф.-м.н.		

Задание принял к исполнению обучающийся:

Группа	ФИО	Подпись	Дата
8ПМ11	Семенюта Антон Вадимович		1.03.2023 г.

**ЗАДАНИЕ К РАЗДЕЛУ
«ФИНАНСОВЫЙ МЕНЕДЖМЕНТ, РЕСУРСОЭФФЕКТИВНОСТЬ
И РЕСУРСОСБЕРЕЖЕНИЕ»**

Обучающемуся:

Группа	ФИО
8ПМ11	Семенюте Антону Вадимовичу

Школа	Инженерная школа информационных технологий и робототехники	Отделение школы (НОЦ)	Отделение информационных технологий
Уровень образования	Магистратура	Направление/ООП/ОПОП	09.04.04 «Программная инженерия»

Исходные данные к разделу «Финансовый менеджмент, ресурсоэффективность и ресурсосбережение»:

1. Стоимость ресурсов научного исследования (НИ): материально-технических, энергетических, финансовых, информационных и человеческих	Бюджет проекта – не более 672734,00 руб., в т.ч. затраты по оплате труда – не более 288397,38 руб.
2. Нормы и нормативы расходования ресурсов	Значение показателя интегральной ресурсоэффективности – не менее 4,46 баллов из 5,00
3. Используемая система налогообложения, ставки налогов, отчислений, дисконтирования и кредитования	Коэффициент отчислений на уплату во внебюджетные фонды - 30%

Перечень вопросов, подлежащих исследованию, проектированию и разработке:

1. Оценка коммерческого и инновационного потенциала НТИ	Проведение предпроектного анализа. Определение целевого рынка и проведение его сегментирования. Выполнение SWOT-анализа проекта
2. Разработка устава научно-технического проекта	Определение целей и ожиданий, требований проекта. Определение заинтересованных сторон и их ожиданий
3. Планирование процесса управления НТИ: структура и график проведения, бюджет, риски и организация закупок	Составление календарного плана проекта. Определение бюджета НТИ
4. Определение ресурсной, финансовой, экономической эффективности	Проведение оценки экономической эффективности разработки

Перечень графического материала:

1. Сегментирование рынка
2. Оценка конкурентоспособности технических решений
3. Матрица SWOT
4. Цели и результаты проекта
5. Перечень работ и исполнителей
6. График проведения НТИ
7. Бюджет НТИ
8. Оценка ресурсной, финансовой и экономической эффективности НТИ

Дата выдачи задания к разделу в соответствии с календарным учебным графиком	01.03.2023 г.
--	---------------

Задание выдал консультант по разделу «Финансовый менеджмент, ресурсоэффективность и ресурсосбережение»:

Должность	ФИО	Ученая степень, звание	Подпись	Дата
Профессор ОСГН ШБИП ТПУ	Жиронкин Сергей Александрович	д-р. экон. наук		

Задание принял к исполнению обучающийся:

Группа	ФИО	Подпись	Дата
8ПМ11	Семенюта Антон Вадимович		

ЗАДАНИЕ К РАЗДЕЛУ «СОЦИАЛЬНАЯ ОТВЕТСТВЕННОСТЬ»

Обучающемуся:

Группа	ФИО
8ПМ11	Семенюта Антон Вадимович

Школа	Инженерная школа информационных технологий и робототехники	Отделение школы (НОЦ)	Отделение информационных технологий
Уровень образования	магистратура	Направление/ООП/ОПОП	09.04.04 «Программная инженерия»

Исходные данные к разделу «Социальная ответственность»:

1. Характеристика объекта исследования (вещество, материал, прибор, алгоритм, методика, рабочая зона) и области его применения	Объект исследования: алгоритм и программные средства классификации людей по полу и возрасту и повторной идентификации в видеопотоке. Область применения: интеллектуальная видеоаналитика в городских условиях. Рабочее место: рабочее место оператора ПЭВМ в помещении офисного типа площадью 12 м², оснащённым системой отопления и кондиционирования воздуха, искусственными источниками света; в помещении присутствует окно.
--	--

Перечень вопросов, подлежащих исследованию, проектированию и разработке:

1. Производственная безопасность 1.1. Анализ выявленных вредных факторов: • Природа воздействия • Действие на организм человека • Нормы воздействия и нормативные документы (для вредных факторов) • СЗ коллективные и индивидуальные 1.2. Анализ выявленных опасных факторов: • Термические источники опасности • Электробезопасность • Пожаробезопасность	1. Вредные факторы: 1.1. Недостаточная освещенность; 1.2. Нарушения микроклимата, оптимальные и допустимые параметры; 1.3. Шум, ПДУ, СКЗ, СИЗ; 1.4. Повышенный уровень электромагнитного излучения, ПДУ, СКЗ, СИЗ; 2. Опасные факторы: 2.1. Электроопасность; класс электроопасности помещения, безопасные номиналы I, U, R_{заземления}, СКЗ, СИЗ; Приведен расчет освещения рабочего места; 2.2. Пожароопасность, категория пожароопасности помещения, марки огнетушителей, их назначение и ограничение применения; Приведена схема эвакуации.
2. Экологическая безопасность: • Выбросы в окружающую среду	Наличие промышленных отходов (бумага-черновики, вторцвет- и чермет, пластмасса, перегоревшие

Решения по обеспечению экологической безопасности	люминесцентные лампы, оргтехника) и способы их утилизации;
3. Безопасность в чрезвычайных ситуациях: 1. перечень возможных ЧС при разработке и эксплуатации проектируемого решения; 2. разработка превентивных мер по предупреждению ЧС; 3. разработка действий в результате возникшей ЧС и мер по ликвидации её последствий.	Рассмотрены 2 ситуации ЧС: 1) природная – сильные морозы зимой, (аварии на электро-, тепло-коммуникациях, водоканале, транспорте); 2) техногенная – несанкционированное проникновение посторонних на рабочее место (возможны проявления вандализма, диверсии, промышленного шпионажа), представлены мероприятия по обеспечению устойчивой работы производства в том и другом случае.
4. Перечень нормативно-технической документации.	– ГОСТы, СанПиНы, СНиПы

Дата выдачи задания к разделу в соответствии с календарным учебным графиком	01.03.2023 г.
--	---------------

Задание выдал консультант по разделу «Социальная ответственность»:

Должность	ФИО	Ученая степень, звание	Подпись	Дата
Профессор ТПУ	Федорчук Ю.М.	д.т.н.		

Задание принял к исполнению обучающийся:

Группа	ФИО	Подпись	Дата
8ПМ11	Семенюта Антон Вадимович		

Реферат

Выпускная квалификационная работа содержит 160 страниц, 44 рисунков, 38 таблиц, 27 формул, 5 приложений и 72 литературных источника.

Ключевые слова: видеоаналитика, глубокое обучение, предсказательная модель, классификация, повторное распознавание.

Цель работы: реализация методов классификации людей по полу и возрасту и их повторной идентификации в видеопотоке с помощью технологий глубокого обучения. Разработанные алгоритмы должны работать на данных видеозаписей в общих городских условиях, а также быть внедрены в ядро платформы, обеспечивающую детектирование и трекинг объектов на видеозаписях, снятых с помощью камер статического видеонаблюдения.

В первой главе представлены аналитический обзор предметной области и теоретические исследования методов распознавания пола и возраста и повторной идентификации.

Во второй главе описаны выбор методов для реализации алгоритмов и концепции решений.

В третьей главе описана разработка программных средств, реализующих алгоритмы.

В четвертой главе описаны разработка и разметка обучающих выборок, а также проведение испытаний алгоритмов.

Пятая глава представляет собой выполненное задание по разделу «Финансовый менеджмент, ресурсоэффективность и ресурсосбережение», где был проведён анализ в области проектного и финансового менеджмента.

Шестая глава представляет собой выполненное задание по разделу «Социальная ответственность», где были рассмотрены аспекты производственной и экологической безопасности, безопасности в чрезвычайных ситуациях, а также правовые вопросы организации труда.

Оглавление

Реферат	11
Список терминов, условных обозначений и сокращений	15
Введение	18
Глава 1. Аналитический обзор предметной области	20
1.1 Теоретические исследования методов распознавания пола и возраста	20
1.1.1 Методы распознавания пола и возраста по изображению лица	20
1.1.2 Методы распознавания пола и возраста по признакам силуэта и движения	23
1.2 Теоретические исследования методов повторной идентификации	31
1.2.1 Постановка задачи	31
1.2.2 Ограничения решения задачи	32
1.2.3 Этапы работы алгоритмов повторного распознавания	33
1.2.4 Этап извлечения графических признаков	33
1.2.5 Этап расчёта дескриптора схожести	36
1.2.6 Этап ранжирования	38
Глава 2. Разработка алгоритмов распознавания пола и возраста и повторной идентификации людей	40
2.1 Разработка алгоритма распознавания пола и возраста	40
2.1.1 Выбор методов для реализации алгоритма	40
2.1.2 Концепция алгоритма	41
2.2 Разработка алгоритма повторной идентификации	45
2.2.1 Выбор методов для реализации алгоритма	45
2.2.2 Концепция алгоритма	49
Глава 3. Разработка программных средств, реализующих распознавание пола и возраста и повторную идентификацию людей	54
3.1 Инструменты разработки	54
3.2 Реализация программного модуля распознавания пола и возраста	56
3.3 Реализация программного модуля распознавания пола и возраста	60
Глава 4. Проведение испытаний алгоритмов распознавания пола и возраста и повторной идентификации людей	67

4.1	Разработка и разметка обучающих выборок.....	67
4.1.1	Данные для задачи сегментации объектов	67
4.1.2	Данные для задачи определения пола и возраста	69
4.1.3	Данные для задачи создания индивидуальных отпечатков объектов по графическим признакам	72
4.2	Условия проведения испытаний.....	74
4.3	Испытания моделей сегментации для распознавания пола и возраста	74
4.4	Испытания алгоритмов распознавания пола и возраста	77
4.5	Испытания алгоритмов повторной идентификации.....	81
4.6.	Испытания трекера с использованием алгоритма повторного распознавания.....	94
Глава 5	Финансовый менеджмент, ресурсоэффективность и ресурсосбережение.....	97
5.1	Предпроектный анализ.....	97
5.1.1	Потенциальные потребители результатов исследования.....	97
5.1.2	Анализ конкурентных решений.....	98
5.1.3	SWOT-анализ.....	100
5.1.4	Цели и результаты проекта	106
5.2	Планирование научно-исследовательских работ	107
5.2.1	Структура работ в рамках научного исследования	107
5.2.2	План проекта.....	108
5.2.3	Бюджет научно-технического исследования	110
5.3	Оценка сравнительной эффективности исследования	113
5.3.1	Интегральный показатель финансовой эффективности.....	113
5.3.2	Интегральный показатель ресурсоэффективности.....	114
5.3.3	Интегральный показатель эффективности	114
5.4	Вывод	115
Глава 6.	Социальная ответственность	117
6.1	Введение	117
6.2	Производственная безопасность	117
6.2.1	Вредные факторы.....	117
6.2.2	Опасные факторы.....	123
6.3	Экологическая безопасность	126

6.4 Безопасность в чрезвычайных ситуациях	127
6.5 Вывод	128
6.6 Перечень НТД	128
Заключение.....	130
Список источников.....	131
ПРИЛОЖЕНИЕ А	137
ПРИЛОЖЕНИЕ Б	157
ПРИЛОЖЕНИЕ В.....	158
ПРИЛОЖЕНИЕ Г	159
ПРИЛОЖЕНИЕ Д.....	160

Список терминов, условных обозначений и сокращений

В данной работе применены следующие термины с соответствующими определениями.

Ограничивающее окно (bounding box) – простая фигура (прямоугольник), ограничивающая форму более сложной геометрической фигуры.

Траектория (track) – набор положений объекта на каждом кадре видеопоследовательности во времени.

Треклет (tracklet) – некоторая ограниченная часть траектории.

Трекинг (tracking) – процесс отслеживания объекта на видеопоследовательности, определение траектории объекта.

Трекер (tracker) – алгоритм трекинга.

Отслеживаемый объект (tracking object) – объект, траектория движения которого на видеопотоке отслеживается трекером.

Детектирование (detecting) – процесс локализации объектов определённых классов на изображении.

Детектор (detector) – алгоритм детектирования.

Детекция (detection) – результат работы детектора, представляющий собой совокупность ограничивающего окна и значений вероятностей принадлежности классам для объекта на изображении.

Аугментация (augmentation) – создание искусственного примера данных путём внесения изменений в реальный пример (например, искусственное изменение изображения для получения нового).

Сегментация (segmentation) – процесс разделения изображения на несколько семантических сегментов (наборов пикселей), например, сегмент «Человек» и сегмент «Фон».

Графические характеристики или графические признаки – структура данных, содержащая числа, компактно описывающие некоторое изображение.

Отпечаток объекта – структура уникальных данных, описывающих некоторый объект на изображении.

Дескриптор – в общем смысле это некоторый набор признаков, описывающих определённый объект или величину для её последующего распознавания и/или идентификации (примеры: дескриптор лица – набор признаков, описывающих изображение лица; дескриптор схожести – величина, определяющая схожесть объектов на основе поднаборов графических признаков).

Повторное распознавание объекта, ре-идентификация объекта (object re-identification, ReID) – процесс присвоения набору признаков, описывающих объект, идентификатора из набора ранее наблюдаемых объектов или объектов, наблюдаемых на видеопотоках других камер.

Поднабор признаков для ре-идентификации (query of images, query of features) – совокупность признаков, описывающих объект для задачи ре-идентификации.

Галерея (gallery) – совокупность признаков ранее наблюдавшихся объектов или признаков объектов, наблюдаемых на видеопотоках других камер, предназначенная для ранжирования относительно схожести с входным поднабором признаков объекта в задаче ре-идентификации.

Ре-идентификация объектов на одном примере (single-shot re-identification) – вид задачи ре-идентификации, в которой входной поднабор признаков для ре-идентификации состоит из одного примера (например, один кадр с изображением объекта).

Ре-идентификация объектов на множестве примеров (multi-shot re-identification) – вид задачи ре-идентификации, в которой входной поднабор признаков для ре-идентификации состоит из множества примеров (например, набор кадров с изображением объекта, соответствующих треклету).

Парсер (parser) – программа для чтения, разбора и структурирования изначально неструктурированной информации.

CNN (Convolutional Neural Networks) – свёрточная нейронная сеть – особый вид архитектур нейронных сетей, в которых используются свёрточные слои – слои, частично связанные с предыдущим слоем.

LSTM (Long Short-Term Memory) – долгая краткосрочная память – особый вид архитектур рекуррентных нейронных сетей, а также вид слоёв рекуррентных нейронных сетей, предназначенных для работы с временными данными.

GEI (Gait Energy Image) – энергетическое изображение походки – монохромное изображение, агрегирующее в себе множество представлений объекта, полученных во время его движения (походки) – дескриптор походки объекта.

BGEI (Back-filled Gait Energy Image) – энергетическое изображение походки с обратным закрашиванием – вид GEI, при котором часть, изображения, находящаяся на изображении за передним краем движущегося объекта, закрашена.

GEV (Gait Energy Volume) – энергетическое объёмное изображение походки – вид GEI, полученного с использованием информации о расстоянии от камеры до каждой точки объекта.

SOTA (State of the art) – соответствующий современному уровню развития – условное название моделей и/или алгоритмов, решающих определённую задачу с наилучшими на текущий момент характеристиками (например, с наилучшим качеством и/или производительностью).

IOU (Intersection Over Union) – метрика оценки степени перекрытия ограничивающих окон друг друга на изображении, учитывает площадь пересечения и площадь объединения ограничивающих окон.

DIOU (Distance Intersection Over Union) – метрика оценки степени перекрытия ограничивающих окон друг друга на изображении, учитывает IOU и расстояние между центрами ограничивающих окон.

CIOU (Complete Intersection Over Union) – метрика оценки степени перекрытия ограничивающих окон друг друга на изображении, учитывает DIOU и соотношение размеров ограничивающих окон.

ARC (Aspect Ratio Consistency) – метрика оценки степени различия в соотношении сторон ограничивающих окон на изображении.

Введение

В настоящее время всё большей популярностью пользуются интеллектуальные системы видеонаблюдения, способные автоматизировать выполнение задач, решаемых обычно человеком. Такие системы основаны на использовании алгоритмов компьютерного зрения и методах глубокого обучения и используются во множестве прикладных областей: робототехника, подсчёт пешеходного и автомобильного трафика, умные дома, охрана безопасности жизнедеятельности и других.

Зачастую интеллектуальные системы решают задачи детектирования и трекинга объектов на кадрах видеопотока. На данных, полученных в результате решения этих задач, проводится интеллектуальное описание объектов и событий на изображениях видеопоследовательности.

Наиболее полезными задачами интеллектуальной видеоаналитики являются определение пола и возраста людей, а также создание анонимных индивидуальных графических отпечатков объектов для обеспечения возможности повторной идентификации. Однако большинство решений, демонстрирующих хорошее качество видеоаналитики, используют в качестве признаков изображения лиц человека. Такие решения не способны работать с данными видеонаблюдения с произвольными внешними условиями в силу часто возникающих трудностей выделения лица на кадрах.

Целью данной работы является реализация методов классификации людей по полу и возрасту и их повторной идентификации в видеопотоке с помощью технологий глубокого обучения. Разработанные алгоритмы должны работать на данных видеозаписей в общих городских условиях, а также быть внедрены в ядро платформы, обеспечивающую детектирование и трекинг объектов на видеозаписях и видеопотоках, снятых с помощью камер статического видеонаблюдения.

Для достижения поставленной цели необходимо выполнить следующие задачи:

- Провести анализ существующих алгоритмов интеллектуального описания объектов на кадрах видеопотока;
- Разработать собственные алгоритмы интеллектуального описания объектов, работающих на данных, типичных для городской среды;
- Провести исследования и настройку алгоритмов с помощью разработанных программных средств;
- Выполнить внедрение и тестирование разработанных алгоритмов в состав ядра платформы.

Практическая значимость данной работы состоит в расширении функциональности решения для автоматизации подсчёта пешеходного и автомобильного трафика по части

интеллектуального описания объектов. Подсчитанный системой трафик используется для задачи оценки пассажиропотока и выбора локаций для открытия новых торговых точек на основе сравнительного анализа

При этом, научную новизну составляет собственная архитектура алгоритмов на базе графических признаков на выходе нейросетевого детектора.

Глава 1. Аналитический обзор предметной области

1.1 Теоретические исследования методов распознавания пола и возраста

1.1.1 Методы распознавания пола и возраста по изображению лица

Распознавание по лицу является популярным и точным подходом к задаче по определению пола и возраста по фото либо видео. Причиной этому служит наличие на лице всех визуальных особенностей, которые позволяют определять пол, возраст, расу, настроение и многие другие особенности человека, визуально видимые в основном (или исключительно) по внешнему виду и представлению лица [1]. На видео распознавание по лицу предпочтительно для определения пола и возраста в случаях, когда на кадрах лицо человека видно отчетливо.

В последовательности действий для решения описанной задачи можно выделить три этапа (Рисунок 1):

1. Определение людей на кадрах видеопотока и построение детекций.
2. Детектирование лиц.
3. Распознавание пола и возраста по лицу.

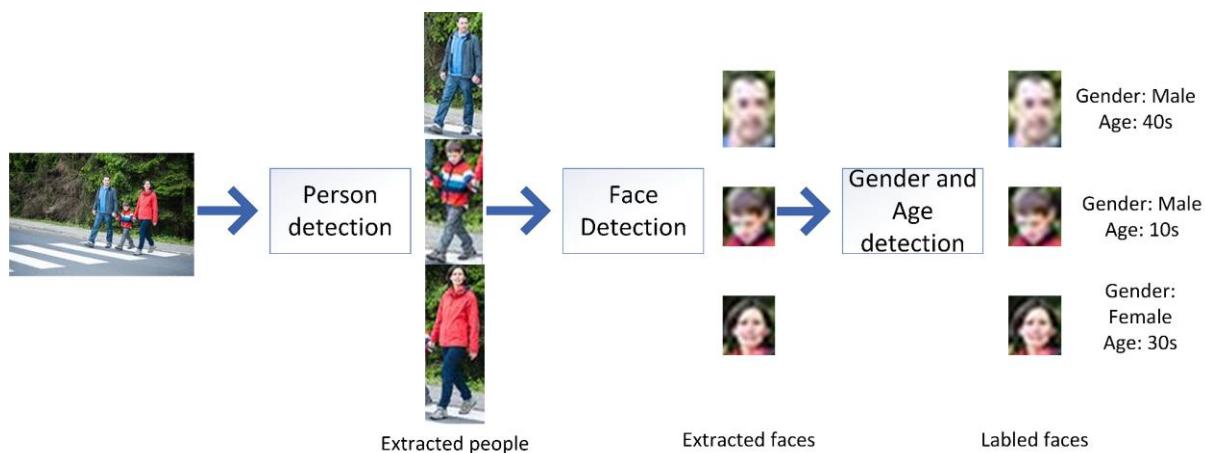


Рисунок 1. Последовательность этапов для определения пола и возраста

1.1.1.1 Обнаружение лица

Обнаружение лица (face detection) является достаточно известной задачей, у которой есть множество различных подходов [2]. Для работы системе необходимо ограничивать область поиска лица. Это необходимо для возможности однозначного сопоставления обнаруженного лица с человеком. Так же данное решение позволит ускорить разработку и общую скорость работы системы, так как отсутствует необходимость в написании и тестировании дополнительного модуля по сопоставлению ограничивающих окон детектированных лиц и ограничивающих окон пешеходов. Ниже расписаны подходы к задаче детектирования лица на кадре.

Классические методы по определению лиц

Подходы, широко использовавшиеся в начале 2000х годов. Использовались в том числе классические статистические методы, анализ главных компонент (РСА), линейные и нелинейные классификаторы [3]. Так же использовались методы, отображающие изображение на плоскость и при помощи нахождения расстояния Махаланобиса [4], находили степень близости текущего обнаруженного объекта с референсным вектором лица. Несмотря на хорошую теоретическую базу для соответствующих методов, они почти не применялись в реальных промышленных приложениях. Проблемой являлось крайне низкая способность подобных подходов адаптироваться к инвариантности среды.

Методы, основанные на применении нейронных сетей

Глубокие сверточные нейронные сети (deep CNN) хорошо себя зарекомендовали в задачах детектирования объектов на изображениях. Они способны определять характерные черты объекта напрямую из пикселей изображения [5]. К наиболее известным архитектурам сверточных нейронных сетей можно отнести AlexNet, GoogLeNet, VGGNet, ResNet. Для обучения CNN требуется огромный объем данных, поэтому для применения их к узконаправленным задачам обычно используют предобученные на известных датасетах модели, а затем, применяя метод переноса обучения (transfer learning), рассчитывают веса на последних слоях сети. Подходы на основе CNN являются наиболее точными и в то же время наиболее требовательными к вычислительным мощностям.

Методы, основанные на вейвлетах Хаара и Габора

При помощи вейвлетов из изображения получали наиболее сильные сигналы, фактически выделяя важнейшие признаки на изображении, отбрасывая остальное. Вейвлеты Габора были широко представлены в исследованиях алгоритмов распознавания лиц [6; 7]. Различные вейвлеты применялись с идеей выделения визуальных признаков вне зависимости от задачи. В дальнейшем данные визуальные признаки использовались в авторских алгоритмах, в том числе и для нахождения лица, как, например, в алгоритме Виолы-Джонса [8]. Использование признаков, полученных вейвлетом Габора является сложным и времязатратным [9]. Существующие упрощенные вейвлеты и алгоритмы на их основе работают быстрее, однако сильно зависят от освещения и других внешних факторов [10].

Методы, основанные на нахождении дескрипторов лица

Вместо использования всего изображения, алгоритмы разбивают его на подобласти в каждой из которой рассчитывают гистограммы и находят наиболее яркие и важные признаки для решаемой в данный момент задачи [11]. В дальнейшем полученные признаки

могут быть использованы для задачи определения лица, так и в задаче идентификации, если есть пример, для которого также можно получить признаки подобным способом. Например, авторы применяли подход на основе локальных дескрипторов, включая различные локальные бинарные шаблоны (LBP), для задачи идентификации по лицу [12]. При этом находились такие дескрипторы, которые были устойчивы к возрастным изменениям черт лица. Данные методы компактны и быстро работают, хорошо справляются с изменениями в освещении и различными эмоциями на лице. Однако эти подходы очень чувствительны к шуму на изображении.

Методы, основанные на трёхмерных изображениях и картах глубины

Многие современные камеры, и даже смартфоны, оснащены лидарами и обладают возможностью быстро и точно получать карту глубины изображения [13]. Данный метод съёмки позволяет значительно снизить чувствительность алгоритмов к освещению и внешним факторам съёмки. Модели, работающие с трёхмерными изображениями, намного более точны и инвариантны к изменениям чем работающие с двумерными [14]. Методы используют текстуры изображения [15], компенсируют вариативное освещение [16], а также компенсируют различные позы человека при помощи стремления к симметричному лицу при анализе [17]. Такие системы, требуют особого оборудования, тщательной калибровки и синхронизации данных, что делает их сложно применимыми в системах общего назначения.

1.1.1.2 Определение пола и возраста

Если определение пола является задачей бинарной классификации, то задачу определение возраста можно решать, как мультиклассовую классификацию по возрастным кластерам, так и как задачу регрессионного предсказания количества лет. Например, для определения возраста авторы [18] дообучили модель VGG-face методом переноса обучения на наборе данных с изображениями людей, для каждого из которых имелась метка о возрасте. На выходном слое нейронной сети было 8 выходов, соответствующих возрастным группам. В [19] для задачи определения возраста также использовались CNN (Ranking-CNN), но с бинарным выходом, а проставление метки о возрасте человека происходило агрегацией выходов “базовых” нейросетей. Каждая из базовых сверточных сетей была обучена на метках принадлежности человека к группе “старше” или “младше” определенного возраста.

Если для задачи определения лица на кадре методы не на основе нейросетей имеют достаточно хорошую точность и могут составить конкуренцию CNN, то для задачи определения возраста по лицу сверточные сети являются непревзойденными лидерами. Так, в 2015 году была представлена сеть AgeNet [20], демонстрирующая высокую точность

в задачах определения возраста по лицу и занявшая призовое место в соревновании «ChaLearn 2015 apparent age competition» [21]. Подход представляет собой глубокую сверточную нейронную сеть, в которой в качестве функции потерь используются как Евклидова норма, так и кросс-энтропия. Большинство современных методов для определения пола и возраста используют известные архитектуры глубоких свёрточных нейросетей, таких как VGG-16 [1; 22] или ImageNet [19].

1.1.2 Методы распознавания пола и возраста по признакам силуэта и движения

Потребность в определении пола и возраста человека на основе признаков движения или силуэта возникает в ситуациях, когда лицо человека на кадрах плохо различимо в связи с удаленностью камеры, либо не видно вовсе. Такие ситуации характерны для систем мониторинга, устанавливаемых в местах большого скопления людей (например, стадионы, аэропорты, вокзалы), чтобы охватить на видео как можно больше пространства.

Все алгоритмы распознавания пола и возраста используют результаты этапов детекции и трекинга объектов. Задача распознавания пола обычно представляется как задача бинарной классификации, а задача распознавания возраста – либо как задача регрессии [23], либо как задача многоклассовой классификации [24], когда выделяются возрастные группы (например, 0-10, 11-20, 21-30, и т.д.).

Ключевым признаком, который используется в подходах к определению пола и возраста на основе признаков формы тела или походки является GEI (Gait Energy Image). GEI – изображение, предоставляющее динамическую и статическую информацию о походке человека. Это один из способов получения биометрических данных на расстоянии. GEI обычно отражает пространственно-временные характеристики походки, такие как длина шага, скорость шага, ширина шага, время шага, время маха, время опоры (средние характеристики походки) и их соответствующую изменчивость и асимметрию (динамические характеристики походки). Пример GEI-изображения представлен на Рисунке 2.

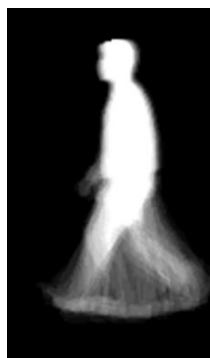


Рисунок 2. Пример GEI-изображения

Подходы на основе GEI показывают высокие метрики качества как в задачах определения пола [25; 26], так и в задачах определения возраста [27].

Распишем основные этапы подходов на основе GEI. Входными данными являются изображения ограничивающих окон объектов на протяжении наблюдения за ними, полученные на этапах детектирования и трекинга. Ниже перечислены основные этапы подходов.

- Обрезка и масштабирование ограничивающих окон;
- Сегментация объектов и бинаризация изображений;
- Классификация по углам наблюдения (опциональный);
- Конструирование GEI;
- Определение пола и возраста.

1.1.2.1 Обрезка и масштабирование ограничивающих окон

На данном этапе происходит обрезка объектов по координатам соответствующих ограничивающих окон. На Рисунке 3 представлен пример изображений, демонстрирующих результаты работы данного этапа.



Рисунок 3. Изображение с объектом, обрезанное по ограничивающему окну

После получения обрезанного изображения объекта происходит его масштабирование. В [26] все изображения приводятся к формату 224x224 пикселей. В [25] – 240x240 пикселей. Чтобы не исказить пропорции картинки, принято добавлять чёрные отступы (padding) по бокам.

В случае с низким качеством изображения дополнительно применяются подходы для увеличения разрешения. В [28] был применен двухступенчатый метод восстановления изображения на основе CNN. В ходе первого шага из изображения удаляются шум и размытие с помощью модели Image Restoration CNN (IRCNN). На втором шаге происходит улучшение разрешения изображения (Super-Resolution Reconstruction, SR) с помощью очень глубоких сверточных сетей SR (Very deep convolution networks SR, VDSR).

1.1.2.2 Сегментация объектов и бинаризация изображений

На этом этапе происходит вычитание фона, и таким образом изображение сегментируется. Методы для вычитания фона можно разделить на две группы: (1) методы основанные на нейронных сетях и (2) методы, не использующие нейросети. Последние широко применялись до 2010-х годов, до появления современных архитектур сверточных

нейронных сетей. Например, в [29] предложен эффективный подход для определения фона и удаления теней на черно-белом видео. Для определения движущихся объектов используется локальная пространственная когерентность. Подход основывается на предположении о том, что пиксели заднего фона на черно-белом видео, снятом статической камерой, имеют нормальное распределение. Тогда пиксели движущегося объекта будут создавать выбросы в распределении. Поэтому для определения параметров распределения используется метрически усредненное среднее. После первичного определения фона применяются морфологические операторы для удаления теней и изолированных ошибок подхода.

Наиболее эффективные подходы к вычитанию фона – подходы на основе сверточных нейронных сетей. Первыми исследователями, применившими такой подход к задаче выделения фона были Brahm и Van Droogenbroeck. Построенная ими нейронная сеть описана в [30] и носит название ConvNet. Для обучения модели использовался набор данных с видеозаписями CDnet 2014, используемый в качестве бенчмарка для соревнований [31; 32].

Авторы статьи [33] использовали разницу между эталонным фоновым изображением (до появления объекта) и текущее изображение и создали свою реализацию алгоритма вычитания фона. Для адаптации к динамическим изменениям сцены использовали изображения с инфракрасной камеры.

После отделения объекта от фона изображение бинаризуется: пиксели изображения, принадлежащие объекту, окрашиваются в белый цвет, а остальные в черный. Пример бинаризованных изображений приведен на Рисунке 4.



Рисунок 4. Пример бинаризованных изображений объекта

Результатом этапа является последовательность обрезанных и отмасштабированных бинарных изображений силуэта (как правило, белый силуэт объекта на чёрном фоне).

1.1.2.3 Классификация по углам наблюдения

Потребность в данном этапе зависит от специфики места, в котором ведется наблюдение. В таких местах как торговые центры люди склонны менять направление

движения в течение небольшого промежутка времени, поворачиваясь к камере разными сторонами. Для описания такого типа передвижения существует термин “вольная походка” (freestyle walk) – передвижение, при котором наблюдаемый объект меняет направление движения и, соответственно, угол движения относительно наблюдателя в течение всего временного промежутка наблюдения. Противоположность походке в одном направлении (fixed-direction walk) [26].

Данный этап призван увеличить точность распознавания пола и возраста при вольной походке. Вместо того, чтобы объединять все изображения силуэтов, извлеченные из всей прогулки в характеристику GEI, в ходе этого этапа происходит классификация силуэтных изображений на группы в соответствие с направлением движения объекта относительно камеры (угол наблюдения). Построение нескольких GEI (один GEI для одной группы силуэтных изображений) эффективно показывает уникальные характеристики человека. Это связано с тем, что это уменьшает помехи от изображений силуэтов, снятых под разными углами наблюдения. Кроме того, некоторые изображения силуэтов могут не отражать все тело субъекта из-за ограниченного поля зрения камеры. Эти частичные силуэтные изображения тела считаются шумом. Исключение таких силуэтных изображений из построения GEI приводит к более высокому качеству GEI.

В статье [26] используется классификационная модель для категоризации силуэта изображения на 10 групп углов наблюдения. Каждая группа представляет диапазон в 36 градусов ($360^\circ/10$) угла наблюдения. Также используется специальная группа для частичных изображений силуэта тела, которые не должны быть задействованы в получении GEI. В качестве классификационной модели используется VGG16. Функция активации выходного слоя – softmax на 11 классов. В статье также используется определение угла перед построением GEI. Для этой цели авторы используют CNN, обученную на датасетах CASIA-(B) [35], OU-ISIR [36], and OU-MVLP [37].

1.1.2.4 Конструирование GEI

После классификации силуэтных изображений прогулки на группы, соответствующие углам наблюдения, силуэтные изображения в каждой группе объединяются в усреднённые изображения (GEI). Для каждого GEI его значение в позиции (x, y) определяется Формулой 1.

$$GEI(x, y) = \frac{1}{N} \sum_{t=1}^N I_t(x, y), \quad (1)$$

где N – число изображений силуэтов в группе, $I_i(x, y)$ – бинарное значение, записанное в пикселе изображения на позиции (x, y) i-го силуэта, где 1 – пиксель принадлежит силуэту, 0 – пиксель не принадлежит силуэту.

Та же формула используется и в случае, когда шаг классификации по углам наблюдения был пропущен. В этом случае итоговое изображение будет иметь один канал.

В [25] после составления GEI улучшить качество модели распознавания пола помогло применение локальных бинарных шаблонов с нечеткой логикой (Fuzzification of Local Binary Patterns). Метод помог уменьшить влияние шумов на бинарные паттерны и создать более информативные признаки за счёт включения нечеткой логики в процедуру расчёта гистограммы переменных кодов.

Существуют подходы, использующие модификации GEI. К примеру, в [38] описывается подход, в котором рассчитывается характеристика, представляющая походку в трехмерном пространстве (Gait Energy Volume или GEV) с использованием изображений глубины (depth image). Это модификация GEI в трехмерном пространстве, где аналогом бинарных пикселей силуэта служат бинарные воксели. Получения таких изображений требует наличия камер глубины. На Рисунке 5 ниже представлены построенные 3D-изображения с вокселями и рассчитанная на их основе GEV.

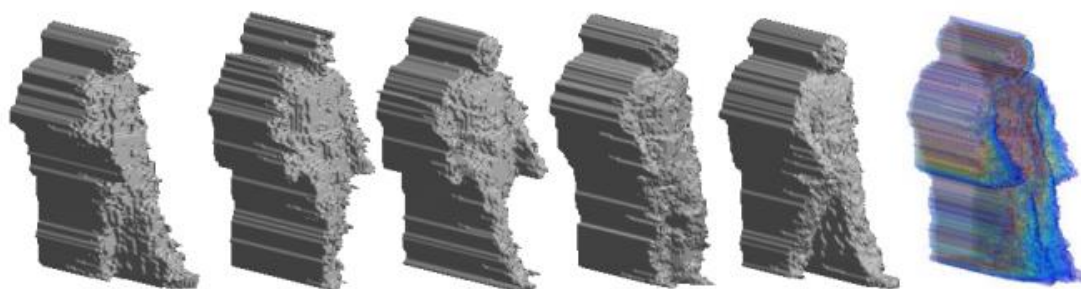


Рисунок 5. Примеры GEV

В [38] описан подход, где рассчитывается BGEI (Back-filled Gait Energy Image) – представление походки, использующее допущение, что фронтальная поверхность модели содержит всю информацию о походке. При использовании BGEI есть вероятность всё-таки потерять некоторую информацию, так как задняя нога больше не представлена в расчете характеристики, однако такая информация в некоторых случаях может быть ненужной или вносить минимальный вклад в точность распознавания пола и возраста. BGEI можно строить как на основе GEI, так и на основе GEV. Для силуэтов вида сбоку BGEI строится по методу обратной заливки бинарных силуэтов. Для этого находится пиксель, характеризующий фронтальную поверхность объекта на профильной проекции в каждой строке, и от него строка заполняется до конца изображения. Заполненные бинарные силуэты выравниваются по центроиду фронтальной поверхности. На Рисунках 6-7 продемонстрированы силуэты, построенные по приведённому методу на основе GEI и GEV соответственно.

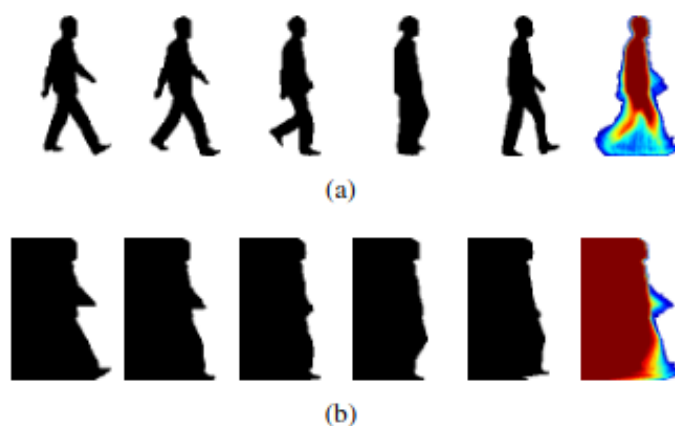


Рисунок 6. Примеры силуэтов с рассчитанными GEI – (a), и BGEI – (b)

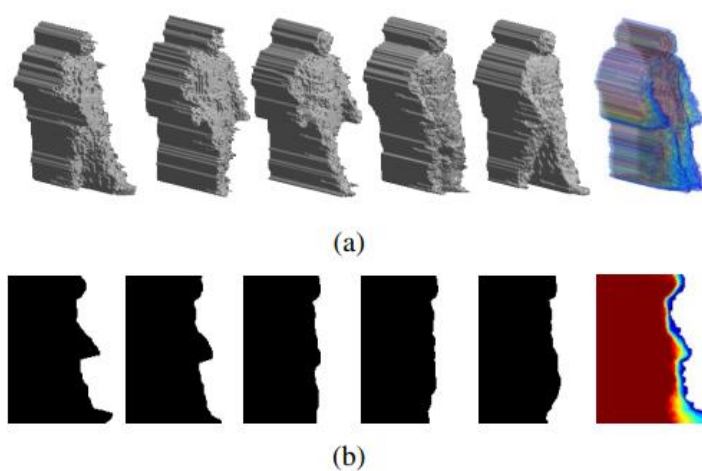


Рисунок 7. Примеры объёмных силуэтов с рассчитанными GEV – (a) и BGEI – (b)

1.1.2.5 Определение пола и возраста

Перед применением непосредственно алгоритмов классификации или регрессии для определения пола либо возраста построенные на предыдущих этапах N GEI изображений объединяются в одно изображение с N количеством каналов. Если классификация по углам наблюдений не производилась, то количество каналов будет равным одному. На Рисунке 8 приведен пример результирующего изображения размерностью $224 \times 224 \times 10$, где строилось 10 GEI [26].

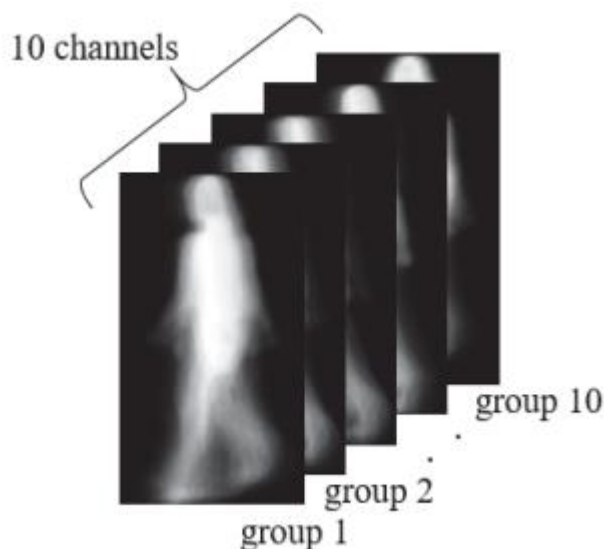


Рисунок 8. Входное изображение для алгоритма классификации

После получения результирующего изображения применяются методы классификации или регрессии. Это могут быть как алгоритмы машинного обучения, так и нейронные сети. Для задачи определения пола в [39] авторы использовали сверточную нейронную сеть VGG16. В [26] для классификации пола авторы разработали свою архитектуру CNN, имеющую 7 слоев. В [25] был применен метод опорных векторов с линейным ядром.

Для определения возраста авторы статьи [24] тестировали несколько подходов: как методы классификации, так и регрессии. Для классификации были выделены возрастные группы (0–5, 6–10, 11–15, 16–20, 21–30, 31–40, 41–50, 51–60, и более 60 лет). Затем рассчитав L2-расстояние на основе GEI между соседними парами групп, и экспериментально установив пороговое значение для такого расстояния, авторы объединили некоторые возрастные группы в одну. После объединения получилось пять групп. Для классификации изображений на группы применялась модификация метода опорных векторов под названием DAGSVM, позволяющий расширить применения классического SVM для задачи многоклассовой классификации.

Подход на основе продольных и поперечных проекций силуэта

GEI является хорошо зарекомендовавшей себя характеристикой применительно к задачам определения пола и возраста, отражающей динамические и образные особенности походки. Помимо подходов, базирующихся на GEI, есть также другие подходы, демонстрирующие высокие метрики качества для задачи определения возраста. Так, например, рассматривалась задача бинарной классификации людей на «молодых» и «престарелых» [27]. В качестве характеристик, позволяющих определить, к какой группе

относится человек, использовались продольные и поперечные проекции силуэта (Silhouette longitudinal projection и Silhouette traverse projection) SLP и STP соответственно. Для их построения сначала применялся подход к локализации вертикального положения точек тела без использования специальных датчиков. Далее анализировалась длина шага человека в течение отрывка видео его походки – Рисунок 9.

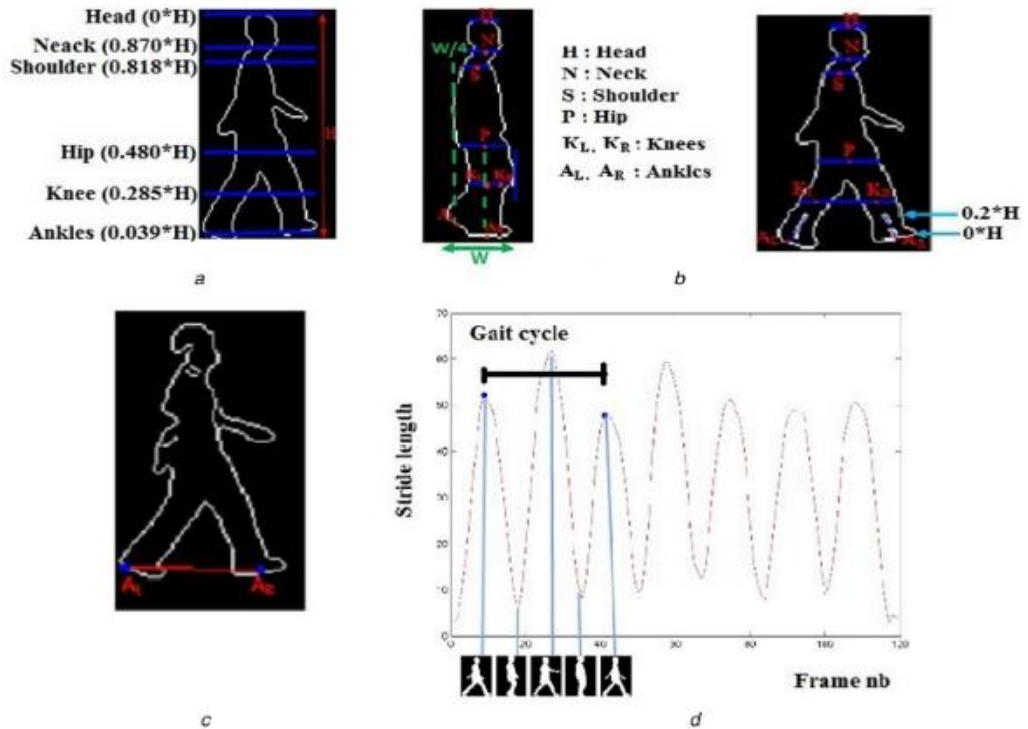


Рисунок 9. Определение длины шага

Затем на бинарных изображениях силуэтов строились SLP и STP с последующей агрегацией и получения результирующего дескриптора – Рисунок 10.

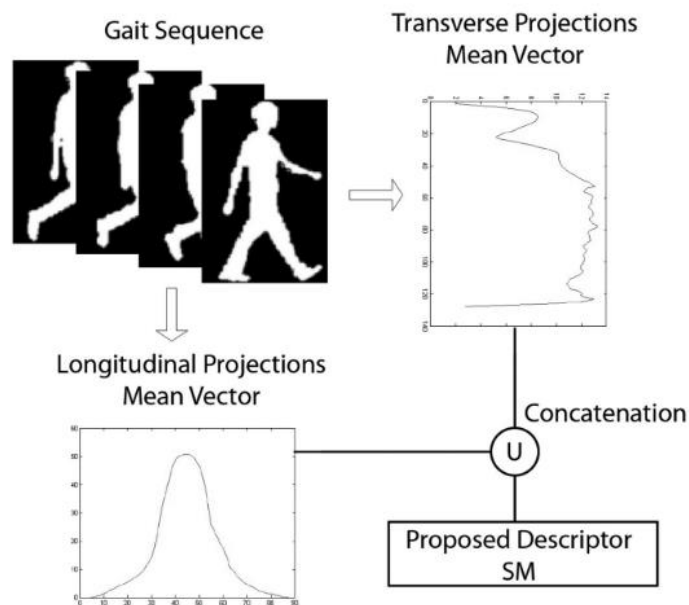


Рисунок 10. Схема результирующего дескриптора на основе SLP и STP

Авторами подтверждено, что данный дескриптор отражает особенности походки молодых и престарелых людей: широкий шаг, характерный для молодых людей, проявляется на продольной проекции, а более широкий силуэт, как правило характерный для людей старшего возраста, находит отражение на поперечной проекции.

Подход с использованием камер разного типа

Типичные изображения объектов ограничены в возможностях распознавания пола и возраста из-за таких факторов, как освещение, фон, тени, различные типы одежды и аксессуаров. Поэтому в статье [28] для определения пола использовали комбинацию изображений с камер видимого света и инфракрасных камер, основанных на температурной разнице между человеческим телом и фоном. В этой статье для изображений с каждой камеры авторы обучили две свёрточные нейронные сети ResNet-101 с входным разрешением изображений 197x447 пикселей, учитывающим пропорции человека. Для принятия решения о поле человека веса моделей на последних слоях сначала нормализовывались, затем агрегировались для получения результирующего показателя. Если значение показателя превышало экспериментальным образом подобранное пороговое значение, то человек на изображении классифицировался как мужчина, иначе – как женщина.

1.2 Теоретические исследования методов повторной идентификации

1.2.1 Постановка задачи

Создание индивидуальных отпечатков объектов по графическим признакам применяется для обеспечения возможности запомнить объект и распознать его повторно как на последующих кадрах видеопотока (для улучшения качества работы алгоритма трекинга), так и при появлении этого объекта в видеопотоке спустя длительный промежуток времени. Таким образом, с помощью создания таких отпечатков реализуется функциональность повторного анонимного распознавания объектов без необходимости иметь предварительно заполненную базу данных некоторых признаков объектов. В большинстве источников такая задача имеет название задачи ре-идентификации объекта в видеопотоке [40; 41]. Далее для обозначения данной задачи будут использоваться оба термина.

Задача ре-идентификации объекта (Person / Target / Object re-identification, ReID) определяется как задача поиска визуальных (графических) описаний объекта в галерее, представляющей собой набор таких же описаний множества объектов, наблюдавшихся в видеопотоке, по запрашиваемому поднабору визуальных (графических) описаний (query of images / query of features) того же объекта, взятому с других сцен и / или в другие моменты времени на той же сцене видеопотока. Также задача подразумевает ранжирование описаний

объектов галереи по убыванию степени схожести с входным / запрашиваемым объектом (object of interest) [42; 43]. На Рисунке 11 представлен пример наборов изображений (query of images) для людей на сцене видеопотока в различные моменты времени. Результатом работы алгоритма, решающего задачу ре-идентификации, является сопоставление запрашиваемых наборов изображений (справа) с набором изображений из галереи, полученным в другой момент времени (слева).



Рисунок 11. Пример результатов работы алгоритма ре-идентификации объектов

1.2.2 Ограничения решения задачи

Наиболее точные методы повторного распознавания людей основываются на биометрических признаках, которые можно извлечь из графической информации отображения объектов в видеопотоке (изображение лица, особенности походки и силуэта) [44]. Однако эти же методы сильно страдают от условий внешнего окружения: взаимные перекрытия людей, различные погодные условия и условия освещения, различное расстояние от камеры до объектов, различное качество видеоматериала. Такие условия могут создать значительные сложности для извлечения графических отпечатков объектов, основанных на биометрии. Учитывая то, что для качественной работы методов ре-идентификации графические признаки должны достаточно различаться для того, чтобы успешно различать одинаково выглядящие объекты, и одновременно нести общую информацию, чтобы быть тиражируемыми на различные условия, то графические признаки, определяемые внешним видом объектов, являются более предпочтительными для использования в работе алгоритмов повторного распознавания [45; 46]. Графические признаки, основанные на биометрических показателях человека, могут быть применены

для уточнения результатов ре-идентификации только в тех случаях, когда на камерах видеонаблюдения удаётся выделить соответствующие данные. Решение о рассмотрении данных методов для формирования индивидуальных отпечатков было перенесено на следующий этап работ.

Далее будут рассмотрены методы повторного распознавания (ре-идентификации) объектов, использующие в качестве графического отпечатка внешние визуальные признаки объектов без биометрической информации.

1.2.3 Этапы работы алгоритмов повторного распознавания

Все алгоритмы повторного распознавания по отпечатку графических признаков работают с результатами методов детектирования, т.е. с ограничивающими окнами объектов. Исходная платформа, включающая в себя детектор множества объектов YOLOv4, полностью удовлетворяет начальным требованиям для внедрения подобных алгоритмов. В общем случае конвейер работы алгоритмов ре-идентификации по графическим признакам объекта делится на 3 основных этапа [45]:

1. Извлечение графических признаков объекта из ограничивающего окна на кадре видеопотока – получение графического отпечатка;
2. Расчёт дескриптора схожести графических отпечатков – получение метрики для поиска в галерее;
3. Ранжирование галереи в порядке убывания схожести графических отпечатков объектов в ней и поиск наиболее близкого сочетания для получения результата – извлечение из галереи наиболее подходящего под входное графическое описание объекта.

Далее будут подробно рассмотрены методы, применяемые на различных этапах работы алгоритмов повторного распознавания.

1.2.4 Этап извлечения графических признаков

На данном этапе осуществляется расчёт графических отпечатков объекта, для которого необходимо осуществить поиск в галерее. Обычно такой отпечаток представляет собой некоторый вектор признаков, полученный из внешнего графического представления объекта на кадре или кадрах видеопотока.

Одними из первых для извлечения графических отпечатков объектов стали применяться признаки цвета в виде цветовых гистограмм [47; 48; 49], а также признаки текстуры (GLCM – Gray Level Co-occurrence Matrices, LBP – Local Binary Patterns, Haralik, Laws и другие) [47; 49; 50]. Данные методы позволяли достаточно успешно определять схожесть объектов при ре-идентификации на множестве камер (по готовым размеченным наборам данных), а также различать объекты в толпе. Однако, данные методы сильно

зависели от условий внешней среды, освещения и качества видеоматериала – они не могли быть легко тиражируемыми на общие условия, присущие камерам видеонаблюдения.

Одним из развитий упомянутых выше методов является борьба с ограничениями внешней среды. Например, в это происходит путём цветовой нормализации, которая позволяет сгладить влияние непостоянного освещения и затенения объектов на изображениях с различных камер [47]. Помимо этого, для задач мультикамерной ре-идентификации объектов применяются методы переноса яркости между несколькими камерами, что позволяет усреднить условия внешнего освещения между сравниваемыми отпечатками графических признаков [48; 50], улучшить общее качество алгоритма ре-идентификации и расширить применимость предложенных методов извлечения графических признаков для задачи повторного распознавания.

В качестве дополнительной информации для отпечатков графических признаков, призванной улучшить качество ре-идентификации объектов определённых классов, в [51] используется подход выделения областей важности на объектах класса человек. Для выделения таких областей используется алгоритм на графах, позволяющий выделить карту групп пикселей, значительно отличающихся от окружающих их других пикселей. Выделенные области используются как веса графических признаков, описывающих ограничивающее окно объекта. Другим подходом, широко используемым для графического описания людей в алгоритмах повторного распознавания, является использование признаков формы: в качестве дополнительной информации, описывающей графический отпечаток, используется информация о размерах объектов или данные о позе [47; 49; 52], принимаемой человеком. Такое обогащение признаков позволяет легче сопоставлять между собой объекты на близко расположенных кадрах одной камеры или на кадрах множества камер. Ещё более продвинутым методом описания графических признаков объекта класса человек на наборе кадров видеопотока является использование признаков GEI (Gait Energy Image), объединяющих динамическую и статистическую информацию о передвижении объекта в пределах треклета [53].

Другим направлением привнесения вспомогательной информации в отпечатки объектов на основе графических признаков является учёт среды, в которой находится объект. В [45; 52] упоминаются методы, которые помимо признаков непосредственно объекта используют графические признаки контекста сцены и опираются на них для получения более точной информации о местоположении объекта в различные временные периоды или на множестве камер. Более продвинутый подход применяется в работе [54], где вместо признаков только одного объекта класса человек, рассматриваются признаки группы людей, рядом с которыми находится этот человек. Такой подход исходит из

предположения, что движущийся в видеопотоке человек на соседних кадрах или различных камерах будет находиться в составе группы людей на протяжении относительно длительного промежутка времени. При этом, на последующих этапах работы алгоритма, сопоставление объектов производится многоступенчато – сначала ре-идентифицируется группа людей, а затем каждый человек в этой группе.

Наиболее популярными и одновременно наиболее точными методами извлечения графических отпечатков для повторного распознавания объектов на сегодняшний день являются нейросетевые алгоритмы, использующие глубокое обучение для расчёта наиболее информативных признаков из ограничивающих окон. Наиболее успешные архитектуры нейронных сетей для таких задач базируются на свёрточных слоях популярных архитектур, таких как VGG16 или ResNet [55; 56; 42], которые, имея предварительно обученные на больших наборах данных веса, после дополнительной тонкой настройки (Fine Tuning) на размеченных данных задачи ре-идентификации, показывают одни из наиболее лучших результатов при любых внешних условиях. Таким образом, основные исследования на текущий момент направлены на улучшение качества базовых архитектур нейронных сетей путём их более тонкой настройки [56; 57], генерации реалистичных изображений для расширения обучающей выборки [57], аугментации обучающих данных [58], разработки специальных слоёв для задачи ре-идентификации [59; 60] и устойчивых функций ошибки, например, таких как triplet-loss [61; 62], позволяющих одновременно обучать модели на схожих и значительно различающихся данных графических отпечатков объектов – Рисунок 12.

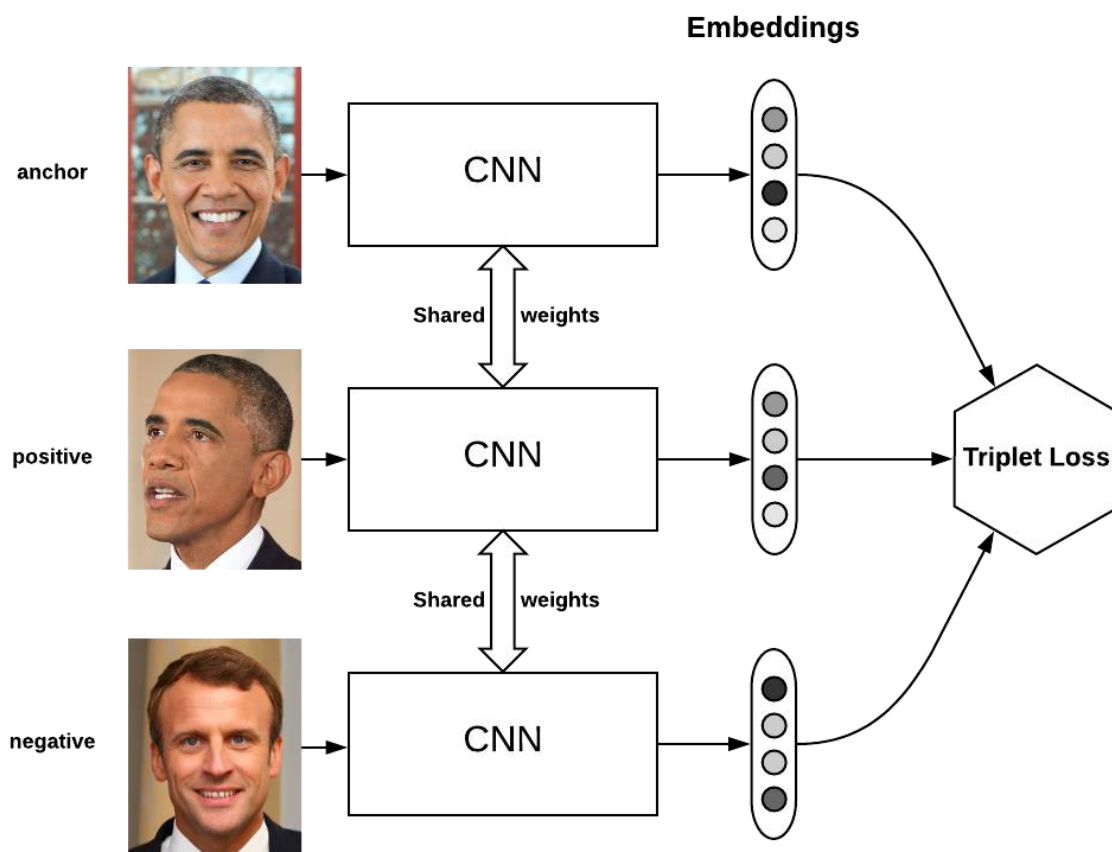


Рисунок 12 – Демонстрация работы функции ошибки triplet-loss

В современных моделях, показывающих наилучшее качество решения задачи ре-идентификации – SOTA-решениях (SOTA – State Of The Art) – применяются подходы отдельного извлечения графических признаков из различных частей ограничивающего окна, что позволяет успешно бороться с проблемами различного масштаба объектов (удаляющиеся и приближающиеся в видеопотоке объекты), различными условиями съёмки и различным качеством видеоматериалов [63; 64; 55]. Помимо этого, применяется обобщение методов извлечения графических признаков для видеоданных: графические отпечатки объектов извлекаются не с одного кадра, а агрегируются исходя из информации о наборе кадров перемещения объекта [64]; набор кадров перемещения объекта из треклета используется в качестве входных данных для обучения нейронных сетей с рекуррентными слоями [65].

1.2.5 Этап расчёта дескриптора схожести

На данном этапе конвейера алгоритмов ре-идентификации извлечённые графические признаки объекта используются для расчёта некоторой метрики схожести с аналогичными графическими признаками галереи. Рассчитанная метрика является основным инструментом для дальнейшего поиска схожих объектов в галерее.

Наиболее популярными подходами для создания дескриптора схожести являются метрики оценки расстояния между двумя векторами: L2-норма (Евклидово расстояние) или косинусное расстояние [45]. Данные метрики являются одними из базовых математических норм расчёта сходства векторов, которые настолько зарекомендовали себя в решении задач компьютерного зрения и, в частности, ре-идентификации, что применяются до сих пор в том числе и в SOTA-решениях [55].

Другим также достаточно распространённым на текущий момент подходом является применение моделей глубокого машинного обучения. Так в для расчёта метрик схожести применяются сиамские нейронные сети [67], позволяющие на основе двух предъявленных на вход векторов графических признаков объектов рассчитать для них дескриптор схожести. Развитием сиамских нейронных сетей для задачи ре-идентификации объектов в видеопотоке является применение архитектур рекуррентных нейронных сетей типа Long Short-Term Memory (LSTM) для извлечения графических признаков во временной зависимости с одновременным обучением на основе этих признаков нейронной сети для расчёта дескриптора схожести [66; 68]. Пример архитектуры таких сетей представлен на Рисунке 13.

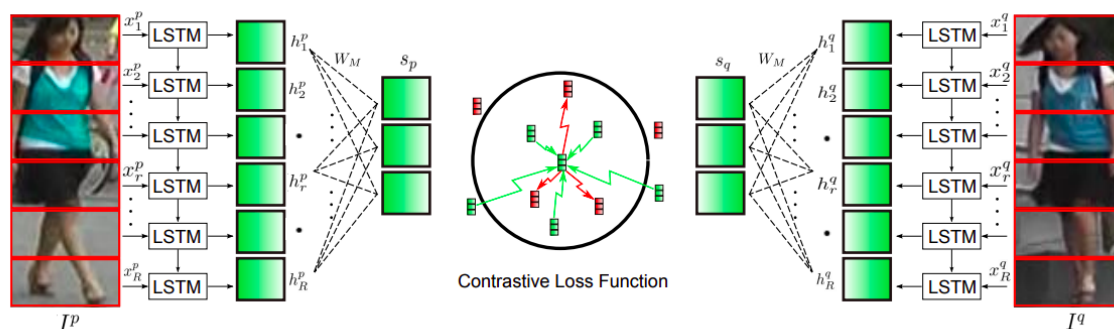


Рисунок 13. Пример архитектуры сиамской нейронной сети с применением слоёв LSTM

Одним из наиболее продвинутых подходов к расчёту метрики схожести имеет следующая архитектура нейронной сети [64]. В ней для расчёта метрики схожести графических признаков объекта используется механизм внимания (Attention Mechanism), позволяющий рассчитать набор пространственно-временных метрик между набором входных признаков объекта и набором признаков из галереи при обучении сети – Рисунок 14. Это позволяет значительно повысить качество повторного распознавания объектов в различных временных отрезках или на видеопотоках с множества камер [64].

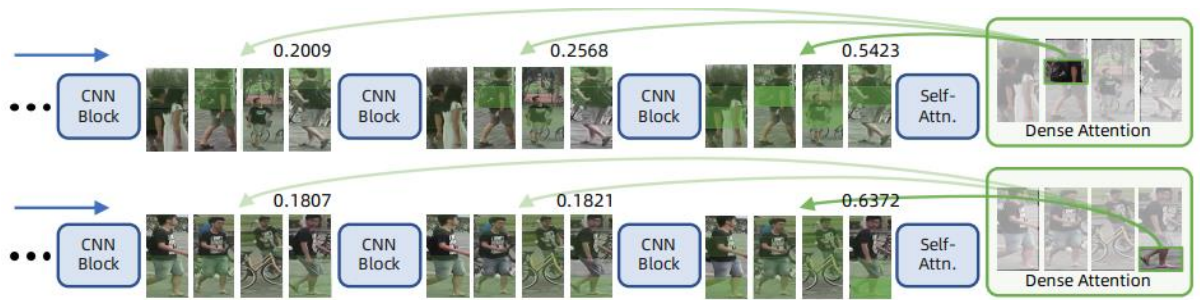


Рисунок 14 – Визуализация работы Механизма Внимания для обучения сопоставления отдельных признаков изображения признакам из галереи

1.2.6 Этап ранжирования

На данном заключительном этапе работы алгоритмов повторного распознавания объектов по графическим признакам на основе рассчитанной метрики схожести осуществляется ранжирование представленных в галерее векторов признаков и определение наиболее схожего объекта (если такой имеется) на основе первых K векторов галереи, где K – является гиперпараметром алгоритма.

Методы ранжирования делятся на несколько типов по способу сравнения входного набора признаков (query of images) с галереей и по способу обучения модели сравнивать эти наборы с галереей. Алгоритмы, которые для сравнения и ранжирования используют лишь один вектор признаков и осуществляют ранжирование также по одному вектору признаков из галереи образуют подмножество one-shot подходов. Алгоритмы, которые сравнивают и ранжируют наборы векторов признаков, представляют собой multi-shot подходы. По характеру обучения алгоритмы делятся на те, которые во время обучения сравнивают и ранжируют размеченные примеры обучающей выборки (обучение с учителем), а также те, которые во время обучения учатся лишь выделять сильные признаки хорошо отделяющие разные объекты и позволяющие сопоставлять одни и те же объекты (обучение без учителя), а само сопоставление и ранжирование осуществляется с помощью не нейросетевых подходов [45].

Для этапа ранжирования алгоритмы ре-идентификации предлагают как использование более оптимальных подходов к упорядочиванию галереи на основе метрик схожести, так и упрощению самого алгоритма упорядочивания путём группировки векторов признаков в галерее. Типичными методами, с помощью которых осуществляется ранжирование и сопоставление входного вектора признаков с признаками из галереи являются алгоритмы перебора, поиска ближайших соседей (например, k -nearest neighbours) и непосредственно ранжирования (например, Ranking SVM [45], где классический алгоритм SVM используется для попарного сравнения входного вектора с набором векторов из галереи). К более продвинутым подходам можно отнести алгоритм [69], где применяется метод инвариантного ранжирования опорных векторов (Invariant Ranking SVM),

предлагающий обучение алгоритма SVM на объединённых графических признаках объектов для обеспечения инвариантности ранжирования относительно различных поз, форм, видов представления объектов и ракурсов камеры. Помимо модификаций непосредственно ранжирования, в [70] применяется нейронная сеть для кластеризации объектов в семантические группы на основе признаков объектов с отдельных кадров и изменения этих признаков со временем в видеопотоке. Разделение объектов на группы позволяет улучшить качество и оптимизировать процесс ранжирования – сначала отбираются группы, затем объекты в этих группах.

Современные методы ранжирования, реализующие подходы обучения с учителем, чаще всего представляют собой нейросетевые методы, упоминавшиеся ранее [64; 65], либо подходы, реализующие все этапы работы алгоритмов ре-идентификации в одной модели [71; 72] – end-to-end подходы. В последнем методе реализуется автоматическое получение треклетов объектов, извлечение графических признаков и сравнение их с признаками треклетов других объектов в данных (multi-shot подход). Метод обобщает подходы использования размеченных данных для задачи ре-идентификации как в одном видеопотоке, так и на множестве камер: осуществляется ранжирование и поиск наиболее похожих треклетов объектов одновременно в нескольких видеопотоках.

Глава 2. Разработка алгоритмов распознавания пола и возраста и повторной идентификации людей

2.1 Разработка алгоритма распознавания пола и возраста

2.1.1 Выбор методов для реализации алгоритма

На основе проведённого литературного обзора на первом этапе НИОКР, а также учитывая дополнительный анализ существующих на сегодняшний день методов распознавания пола и возраста человека, не использующих изображение лица, для извлечения признаков из трека объекта был выбран алгоритм построения GEI-изображения, который требует проведения этапа сегментации и непосредственно создания самого изображения (карты признаков). Выбор в пользу именно этого подхода был сделан благодаря эффективности, показанной в нескольких успешных работах и соответствие подхода особенностям данных, загружаемых на платформу – видеоданные перемещений людей в определённом направлении (человек на протяжении своего трека двигается в одну сторону). Дополнительным преимуществом использования GEI-признаков является возможность построения изображения даже на основе кадров, где виден только силуэт человека (нет информации о цветах, затруднительно выделить ключевые точки или текстурные признаки) – такие данные в том числе относятся к общим условиям съёмки.

Для осуществления этапа сегментации ограничивающего окна человека, необходимого для построения GEI-изображения, было выбрано решение PaddleSeg, имеющее в своём составе данные и инструменты сегментации, включающие изображения людей в полный рост, в отличие от большинства других, где съёмки осуществлялись под менее подходящим под данные платформы острым углом к горизонту. Благодаря этому, модели сегментации не нужно обучать с нуля, для построения модели может использоваться подход переноса обучения.

Задачу распознавания возраста было принято решать не как задачу регрессии и расчёта точного числового значения, а как задачу классификации и определения одной из трёх возрастных категорий. Выбор в пользу такого решения был сделан на основании практически полного отсутствия открытых данных, условия съёмки которых приближены к данным, свойственным платформе детектирования и трекинга объектов, и большой трудоёмкости создания собственного набора, удовлетворяющего всем требованиям разработки алгоритма распознавания пола. Решение подзадачи классификации пола и возраста было принято осуществлять с помощью свёрточной нейронной сети, которая использует для извлечения признаков предварительно обученные веса нейронной сети

ResNet, показывающей хорошие результаты в различных задачах классификации на изображениях реального мира.

2.1.2 Концепция алгоритма

Предложенная версия разработанного алгоритма распознавания пола и возраста основывается на выбранных на текущем этапе работ методах, призванных решить проблемы, выявленные на этапе разработки первой экспериментальной версии алгоритма, и сделать подход более универсальным, работающим в наиболее общих внешних условиях, собственных данным, загружаемым на платформу детектирования и трекинга объектов.

Разработанная версия алгоритма распознавания пола и возраста учитывает особенности его внедрения в работу ядра платформы детектирования и трекинга объектов (учитывает доступность данных, их структуру и формат, а также интерфейсы имеющихся программных компонентов), а также возможности по дальнейшей оптимизации производительности для достижения требуемых показателей времени обработки видеопотоков.

Логика работы алгоритма, показанная для одной итерации обработки одного кадра видеопотока, с учётом данных и их доступности на различных этапах работы уже имеющихся в составе ядра компонентов для детектирования и трекинга представлена на Рисунке 15.

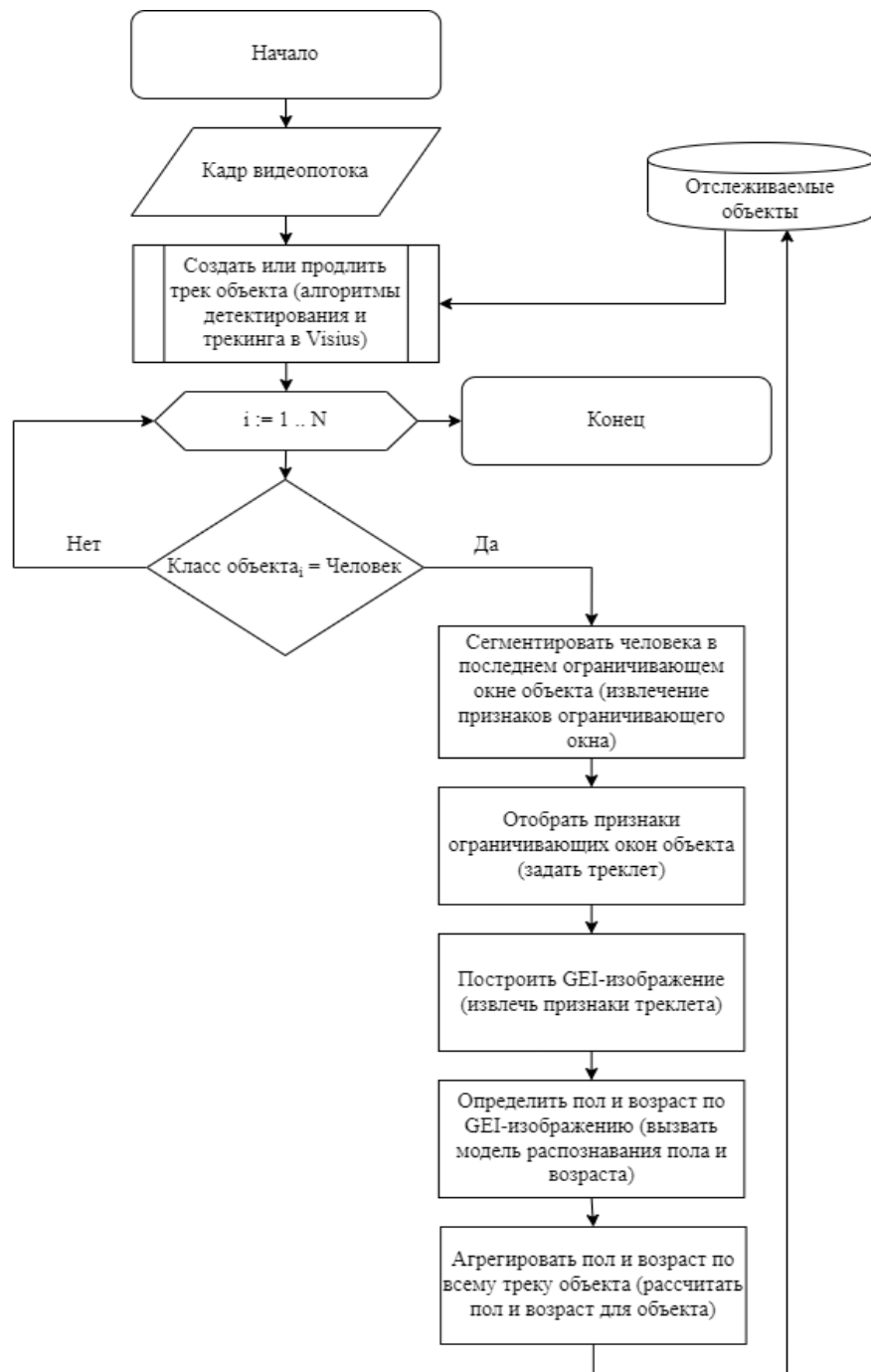


Рисунок 15. Блок-схема работы разработанных алгоритмов распознавания пола и возраста

Основная часть алгоритмов работает с результатом трекинга – базой отслеживаемых на видеопотоке объектов, для которых уже было произведено обновление состояния, соответствующее текущему обрабатываемому кадру. Таким образом, входными данными является набор обновлённых в соответствии с текущим обрабатываемым кадром объектов (треков), хранящих информацию о последнем ограничивающем окне. Разработанный алгоритм распознавания пола и возраста работает с каждым отслеживаемым объектом независимо и для каждого i -го объекта из N отслеживаемых на текущий момент выполняет последовательность действий:

Определяет принадлежность объекта к классу “Человек”. Только для объектов этого класса возможно и требуется определение пола и возраста. Фильтрация классов выступает этапом валидации входных данных для алгоритма.

Для каждого объекта класса “Человек” выбирается последнее ограничивающее окно, соответствующее текущему обрабатываемому кадру и, в терминах ядра платформы, осуществляется извлечение графических признаков из ограничивающего окна объекта, которое в данном случае имеет конкретную реализацию – сегментация человека на изображении ограничивающего окна – выделение пикселей, относящихся к человеку. Данная операция осуществляется с помощью модели сегментации MobileSeg из библиотеки PaddleSeg. Результатом работы модели сегментации является бинарное изображение, где значение интенсивности цвета пикселя, равное 1, говорит о том, что пиксель принадлежит сегменту силуэта, а значение 0 – не принадлежит. Извлечённые признаки привязываются к объекту, который хранит историю таких признаков, соответствующую предыдущим обработанным кадрам видеопотока.

Среди набора извлечённых признаков ограничивающих окон объектов выбирается поднабор, описывающий ограниченный по длительности треклет объекта (весь объект описывается полным треком). На текущем этапе работ используется правило выбора определённого количества последних ограничивающих окон объектов.

На основе признаков ограничивающих окон строятся признаки треклета, в качестве которых в данном случае выступает GEI-изображение, которое строится согласно Формуле 1, определяющей каждое значение изображения в позиции (x, y) .

Осуществляется вызов модели свёрточной нейронной сети, предварительно обученной методом переноса обучения на базе архитектуры ResNet50, для определения пола и возраста человека по признакам его треклета – по GEI-изображению. Предложенная архитектура сети, объединяющая решение задач как определения пола, так и определения возраста, представлена на Рисунке 16.

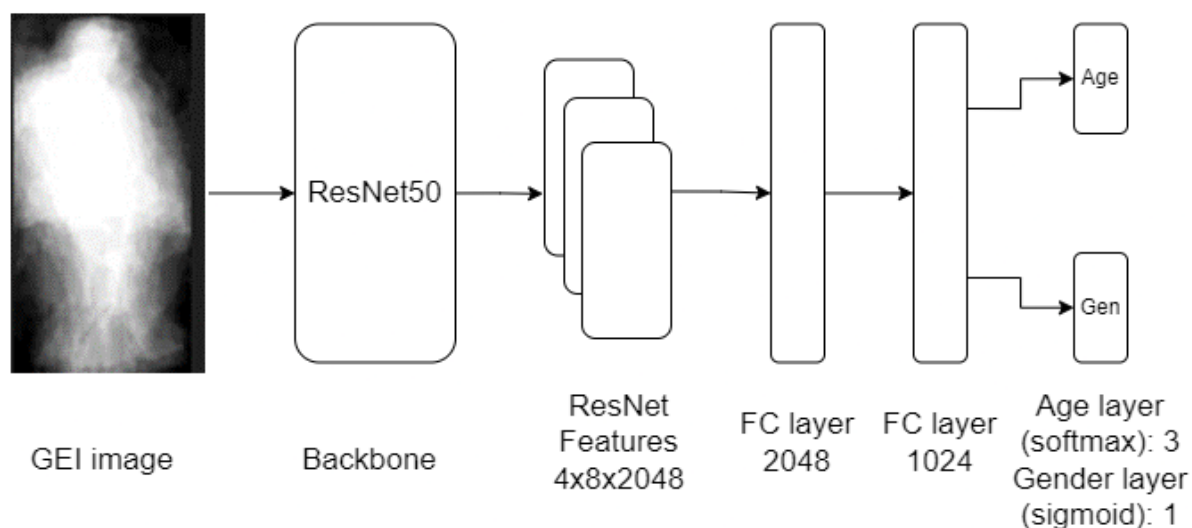


Рисунок 16. Архитектура свёрточной нейронной сети, использующейся в качестве классификатора для распознавания пола и возраста

Результатом работы модели классификации пола и возраста является два набора чисел: код класса пола человека с соответствующим значением вероятности отнесения примера к этому классу, рассчитанное моделью, а также код класса возраста человека с аналогичным значением вероятности. В общем случае, полученный в результате вызова модели результат является одним из многих для текущего рассматриваемого объекта – такие результаты могут рассчитываться для любого количества треклетов одного объекта и храниться в его истории.

6. На основе всей истории результатов распознавания пола и возраста текущего объекта осуществляется агрегация как выбор чаще всего встречающегося класса и принимается окончательное решение о значении пола и возраста для объекта. Агрегация необходима для того, чтобы всё время пытаться уточнить пол и возраст объекта на видеопотоке, т.к. в различные моменты времени это можно сделать с большей или меньшей уверенностью. Все промежуточные и финальные результаты работы алгоритма сохраняются в отслеживаемых объектах (треках) для использования в следующих итерациях обработки кадров видеопотока.

В общем виде, конвейер преобразований данных для одного поднабора ограничивающих окон объекта выглядит так, как это представлено на Рисунке 17.

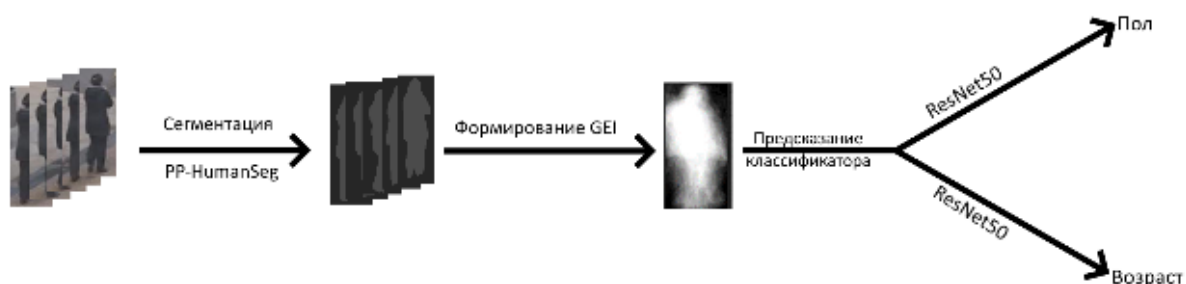


Рисунок 17 – Общий вид конвейера преобразований данных треклета объекта в алгоритме распознавания пола и возраста

Предложенная версия разработанного алгоритма распознавания пола и возраста адаптирована под архитектуру работы ядра платформы, использует наиболее характерную для данных платформы информацию о движении и силуэте человека, слабо зависит от внешних условий съёмки и может применяться в случаях, когда невозможно распознать лицо человека, или даже в случаях отсутствия информации о цветах, описывающих человека. Помимо этого, алгоритм открыт для оптимизаций с точки зрения производительности всей системы в целом: отбор ограничивающих окон треклета может осуществляться на основе их качества (выбирать только те из них, где человек отчётливо различим и не перекрывается другими объектами), обработка данных каждого объекта может выполняться параллельно, благодаря работе с треклетом объекта вызов алгоритма может осуществляться не на каждой итерации обработки очередного кадра видеопотока, а намного реже.

2.2 Разработка алгоритма повторной идентификации

2.2.1 Выбор методов для реализации алгоритма

Теоретический обзор возможных методов выполнения задачи создания индивидуальных отпечатков объектов по графическим признакам для последующего повторного распознавания, который был приведён в первом отчёте о НИОКР, наиболее полно рассматривает применяемые на сегодняшний день алгоритмы на различных этапах работы конвейера для повторного распознавания объектов. Выбор подходов, сделанный на первом этапе работ, является полностью обоснованным, однако разработка алгоритмов, необходимая для устранения недостатков экспериментальной версии, требует расширения исследований в выбранной области. Так, например, новые исследования должны коснуться выбора наиболее подходящей модели для извлечения графических признаков из ограничивающих окон объектов, способов представления признаков всего треклета, дескрипторов схожести графических признаков, а также возможных способов агрегации кандидатов, выбранных из галереи.

Выбор нейросетевой модели должен учитывать кандидатов из числа общеизвестных архитектур нейронных сетей, показывающих хорошее качество в задачах классификации на данных реального мира, например, распространённая модель ResNet50, нейронные сети типа MobileNet или ShuffleNet, последняя из которых применяет случайное перемешивание цветовых каналов, что позволяет значительно снизить вычислительную сложность с сохранением качества на уровне ближайших конкурентов. Помимо глубоких нейронных сетей общего назначения также следует рассмотреть специализированные архитектуры,

предназначенные для решения задачи повторного распознавания: Omni-Scale Net (OSNet) – архитектура нейронных сетей, нацеленная на извлечение разномасштабных признаков с их последующей агрегацией – Рисунок 18; Multi-Scale Attention Deep Net (MuDeep) – специальная архитектура, продолжающая идеи OSNet, но использующая механизм внимания для подбора весов карт признаков, описывающих разные масштабы объекта; или Multi-Level Factorisation Net (MLFN) – глубокая нейронная сеть, использующая свёрточные слои для извлечения характерных признаков объекта на низком уровне, а также их смешивание с высокоуровневыми признаками, выделяемыми на каждом уровне глубины сети, что позволяет описывать отличительные черты объекта независимо от условий съёмки – Рисунок 19.

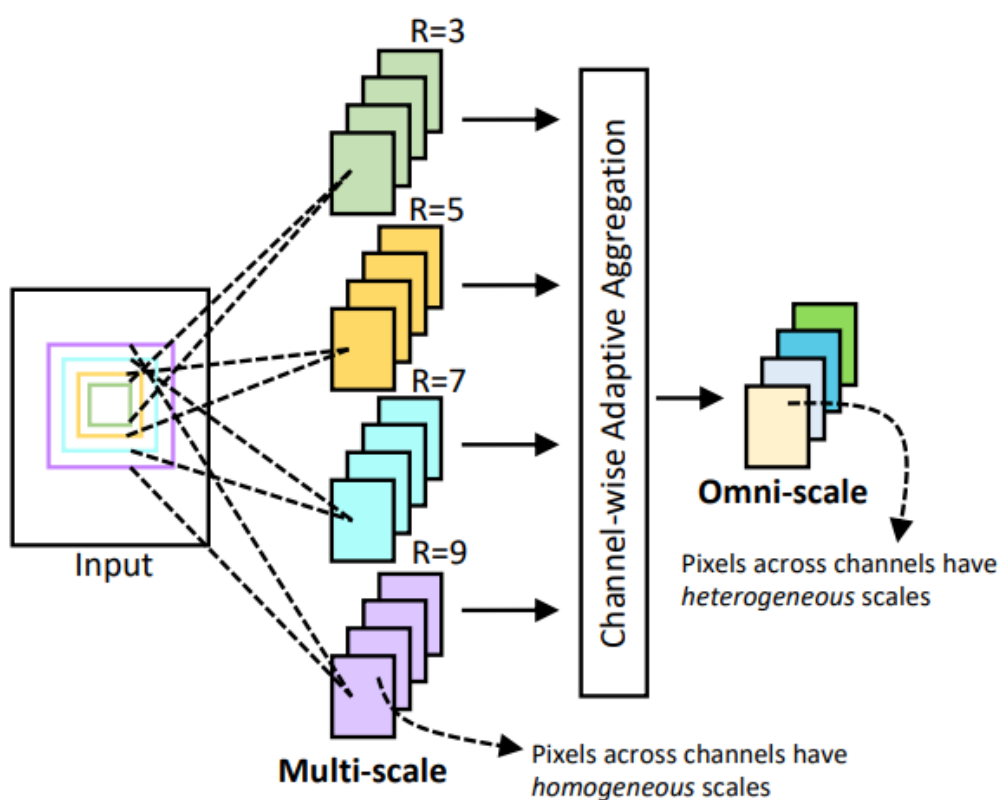


Рисунок 18. Схема особенностей архитектуры OSNet

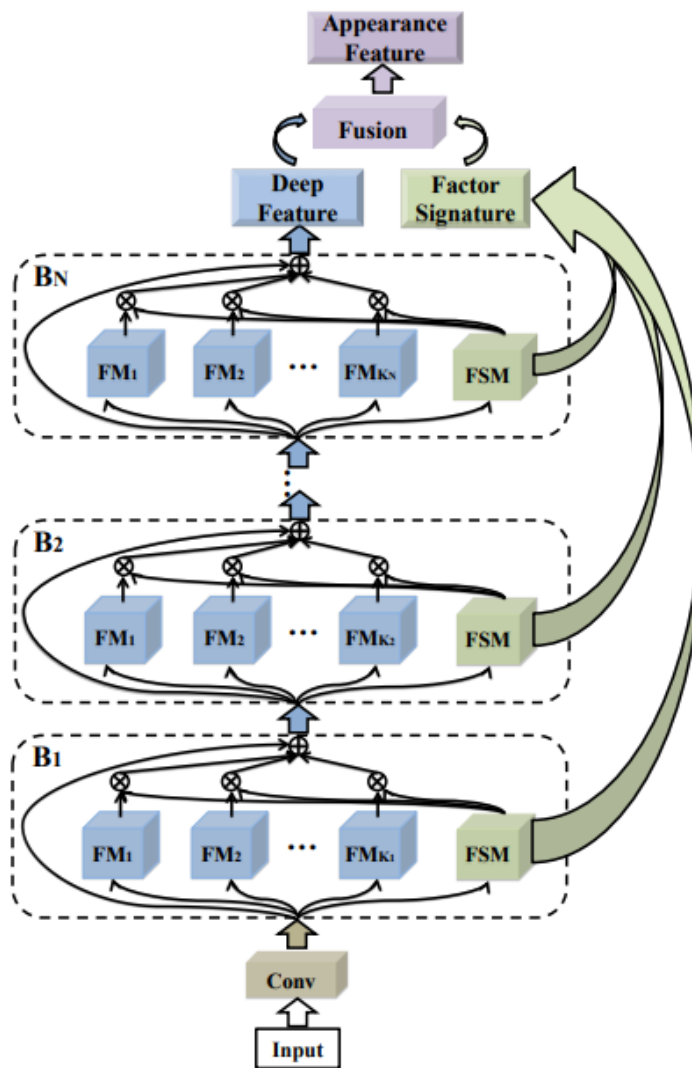


Рисунок 19. Общая схема части архитектуры MLFN, предназначенной для извлечения графических признаков

Способы извлечения графических признаков из треклета могут варьироваться от простых методов агрегации (расчёт среднего или максимального значения среди соответствующих элементов карт признаков, полученных из каждого ограничивающего окна) до использования рекуррентных архитектур нейронных сетей, сразу обрабатывающих все ограничивающие окна треклета.

На первом этапе работ в качестве дескриптора схожести применялось исключительно косинусное расстояние, наиболее часто встречающееся в существующих современных подходах. Однако, это не единственная возможная метрика. В качестве альтернатив на текущем этапе можно также рассмотреть такие меры расстояния как L1– и L2-нормы, расстояние Чебышева и расстояние Махаланобиса. Помимо математического расчёта, для определения расстояний между векторами графических признаков также применяются специальные архитектуры нейронных сетей, называемые сиамскими – модель учится различать векторы признаков объектов – Рисунок 20.



Рисунок 20. Принцип работы модели сиамской нейронной сети

Методы агрегации кандидатов из галереи, полученных на основе расчёта и сравнения дескрипторов схожести, не являются обязательным этапом алгоритма повторного распознавания. Здесь могут применяться подходы на основе различных правил голосования, как это было сделано на первом этапе работ, а также различные правила агрегации: среднее, максимальное, мода и т.д. Могут также применяться правила взвешивания кандидатов для расчёта агрегированного значения.

Учитывая выбранные направления по разработке алгоритмов создания индивидуальных отпечатков объектов по графическим признакам для повторного распознавания и возможности в выборе и реализации конкретных методов, для задачи извлечения наиболее информативных графических признаков из ограничивающих окон объекта был выбран ряд моделей (ResNet50, ShuffleNet, OSNet, MuDeep, MLFN) для дальнейшего проведения экспериментов; в качестве метода извлечения графических признаков треклета был выбран поэлементный расчёт среднего значения – от использования для этого архитектур рекуррентных нейронных сетей было решено отказаться в пользу скорости работы первого метода; для экспериментов с подбором наиболее подходящего алгоритма расчёта дескриптора схожести между векторами признаков были отобраны следующие меры (как часто применяемые в задачах повторного

распознавания): косинусное расстояние, L1- и L2-норма, расстояние Чебышева, расстояние Махаланобиса – от использования сиамских нейронных сетей было решено отказаться в силу наличия неудачного опыта их применения (при разработке первой версии платформы детектирования и трекинга объектов проводились эксперименты для улучшения алгоритма трекинга с их использованием) и экономии ресурсов для сохранения быстродействия; среднее значение, мода и взвешенное на количество одноимённых кандидатов среднее значение были выбраны для экспериментов с методами агрегации относительного всего трека объекта.

2.2.2 Концепция алгоритма

Предложенная версия разработанных алгоритмов создания индивидуальных отпечатков объектов по графическим признакам для повторного распознавания использует конвейер, включающий отобранные в результате экспериментов методы, показывающие наилучшую эффективность с точки зрения качества и производительности, избавляя решение от недостатков экспериментальной версии.

Разработанные алгоритмы учитывают как наиболее общие внешние условия съёмки, свойственные данным, загружаемым на платформу детектирования и трекинга, так и особенности их внедрения в модуль ядра платформы: учитывают доступность данных, их структуру и формат, а также интерфейсы имеющихся программных компонентов. Помимо этого, алгоритмы спроектированы таким образом, чтобы имелись широкие возможности по дальнейшей оптимизации производительности для достижения заявленных показателей скорости работы платформы.

Логика работы конвейера разработанных алгоритмов создания индивидуальных графических отпечатков для повторного распознавания объектов, показанная для одной итерации обработки одного кадра видеопотока, с учётом данных и их доступности на различных этапах работы уже имеющихся в составе ядра компонентов для детектирования и трекинга представлена на Рисунке 21.

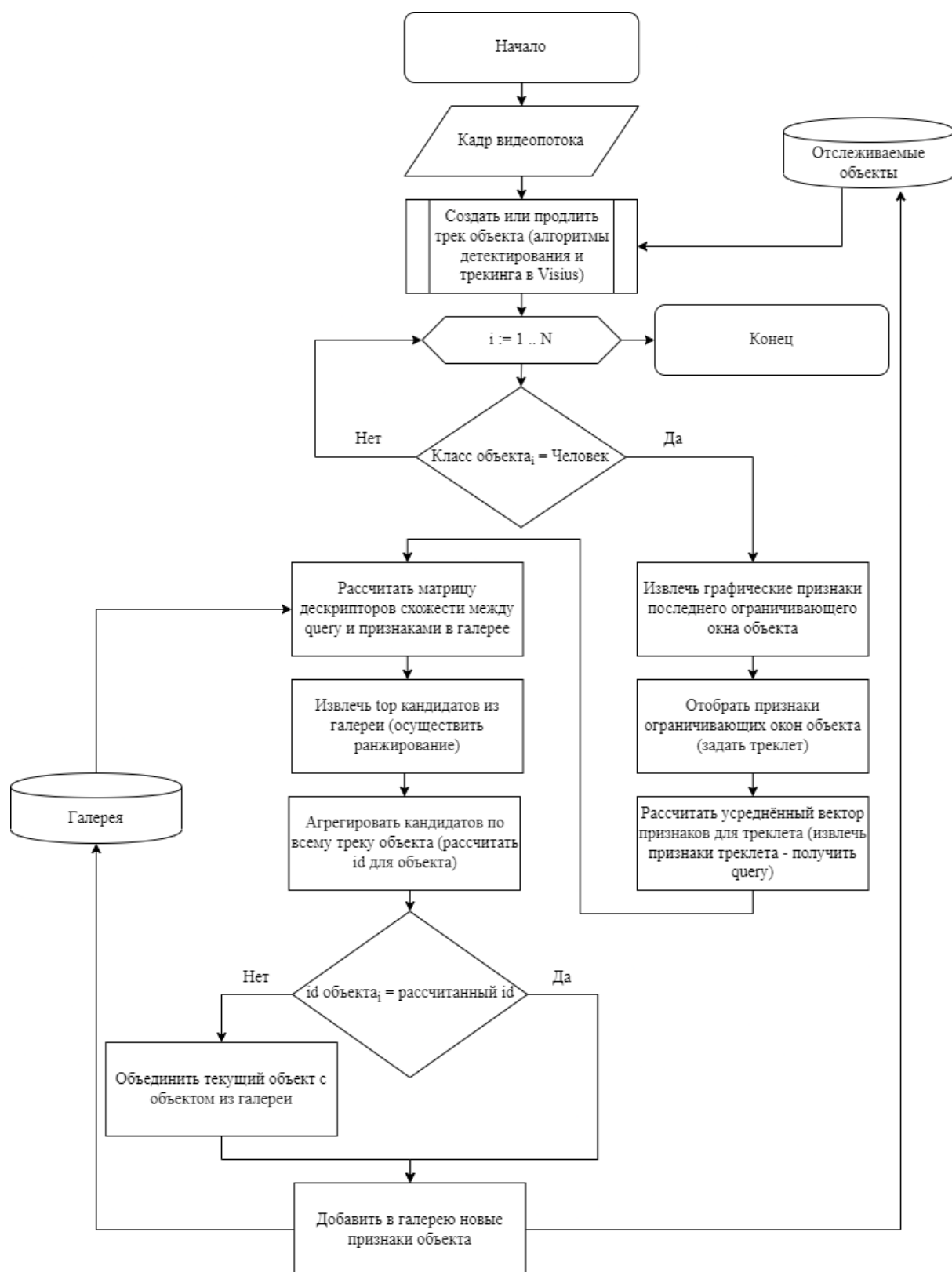


Рисунок 21. Блок-схема работы разработанных алгоритмов создания индивидуальных отпечатков по графическим признакам для повторного распознавания

Принцип работы конвейера разработанных алгоритмов повторного распознавания похож на принцип работы алгоритма распознавания пола и возраста. Он также работает с результатами обработки текущего кадра алгоритмами детектирования и трекинга – обновлёнными данными отслеживаемых объектов (треков), хранящих информацию о

последнем ограничивающем окне. Разработанные алгоритмы работают с каждым отслеживаемым объектом независимо и для каждого i -го объекта из N отслеживаемых на текущий момент выполняет последовательность действий:

1. Определяет принадлежность объекта к классу “Человек”. Только для таких объектов на текущий момент требуется реализация повторного распознавания, а также только на объектах этого класса были обучены модели для извлечения графических признаков. Фильтрация классов выступает этапом валидации входных данных;

2. Для каждого объекта отобранного класса осуществляется извлечение графических признаков из последнего ограничивающего окна. Эту операцию выполняет отобранная по результатам экспериментов и предварительно обученная с помощью переноса обучения на наборе специально подготовленных данных, модель MLFN. Результатом работы данной модели является вектор графических признаков, соответствующий последнему ограничивающему окну объекта. Извлечённые признаки добавляются к объекту для хранения вместе с историей таких же признаков, соответствующих предыдущим обработанным кадрам видеопотока;

3. Для отбора треклета осуществляется выбор графических признаков ограниченного количества ограничивающих окон объекта. На текущем этапе работ используется правило выбора последних ограничивающих окон объектов;

4. На основе отобранных признаков ограничивающих окон треклетов осуществляется расчёт признаков всего треклета путём усреднения соответствующих значений векторов признаков ограничивающих окон согласно Формуле 2 (формула показывает расчёт значения вектора усреднённых значений признаков TFV, находящегося в позиции x);

$$TFV(x) = \frac{1}{N} \sum_{i=1}^N I_i(x), \quad (2)$$

где N – количество векторов признаков ограничивающих окон треклета, $I_i(x)$ – значение признака в позиции x i -го вектора. Полученный усреднённый вектор признаков представляет собой запрос (query), необходимый для осуществления расчёта дескрипторов схожести с признаками треклетов из галереи, их последующего ранжирования и отбора, необходимого для повторного распознавания;

5. Имея усреднённый вектор графических признаков, описывающий треклет объекта, который представляет собой запрос (query) в задаче повторного распознавания, осуществляется расчёт матрицы дескрипторов схожести между query и признаками в галерее. В качестве дескриптора схожести используется косинусное расстояние между векторами признаков (Формула 3), отобранное в результате экспериментов как метод, показавший наилучшее качество;

$$D(u, v) = 1 - \frac{(u) \cdot (v)}{|u|_2 \cdot |v|_2}, \quad (3)$$

где u и v векторы. Матрица дескрипторов схожести и является результатом выполнения данного этапа алгоритма.

6. На основе полученной матрицы дескрипторов схожести осуществляется ранжирование галереи и извлечение ограниченного набора (top) наиболее подходящих кандидатов. Ранжирование осуществляется простым перебором и сравнением значений дескрипторов схожести. Результатом этапа ранжирования является top наиболее подходящих кандидатов графических признаков треклетов объектов из галереи, каждый из которых привязан к определённому идентификатору (id) объекта. Извлечённые кандидаты привязываются к объектам;

7. На основе всей истории извлечённых кандидатов для одного объекта с помощью метода агрегации, описанного Формулой 4, отобранной по результатам экспериментальных исследований, принимается решение об идентификаторе текущего объекта;

$$S^{id} = \frac{\frac{1}{N} \sum_{i=1}^N d_i^{id}}{\left(\frac{N^{id}}{N}\right)^c}, \quad (4)$$

где S^{id} – значение метрики для агрегации (чем она меньше для кандидата с определённым id , тем более подходящим является этот кандидат), d_i – значение дескриптора схожести i -го кандидата с $query$, имеющим идентификатор id , N – количество кандидатов, N^{id} – количество кандидатов с идентификатором, совпадающим с идентификатором id , принадлежащим $query$, c – регулировочный гиперпараметр. Рассчитанный с помощью агрегации идентификатор может совпадать с текущим id объекта, а может и не совпадать. Если id не совпадает, это означает, что в галерее уже хранится информация о текущем отслеживаемом объекте, что, в свою очередь, означает, что текущий объект появился на видеопотоке повторно и ему нужно присвоить идентификатор, соответствующий его первому появлению на видеопотоке – объединить текущий объект с объектом из галереи. Результатом данного этапа работы конвейера алгоритма является определение корректного текущего id объекта – ре-идентификация объекта.

8. После завершения этапа ре-идентификации, независимо от того, произошло ли объединение текущего объекта с объектом из галереи, признаки треклета текущего объекта добавляются в галерею для обеспечения возможности повторного распознавания этого же объекта на кадрах видеопотока.

В общем виде, процесс получения и преобразования данных, происходящий в разработанном алгоритме создания индивидуальных отпечатков объектов по графическим признакам для повторного распознавания, показан на Рисунке 22.



Рисунок 22 – Схема преобразования данных в конвейере разработанных алгоритмов создания индивидуальных графических отпечатков объектов для повторного распознавания

На Рисунке 22 в компоненте Predictor заключён процесс получения признаков треклета с последующим расчётом дескриптора схожести и ранжированием; компонент Aggregator определяет способ агрегации результатов ранжирования и выдает окончательное решение о значении идентификатора объекта.

Предложенная концепция работы разработанных алгоритмов создания индивидуальных отпечатков объектов по графическим признакам для их последующего повторного распознавания адаптирована под общие условия съёмки, свойственные данным, загружаемым на платформу детектирования и трекинга, и позволяет добиться лучших показателей качества (результаты экспериментов с оценкой качества работы алгоритмов приведены в разделе 5), устраняя недостатки экспериментальной версии. Дополнительным плюсом являются широкие возможности по оптимизации производительности алгоритмов: отбор ограничивающих окон треклета может осуществляться на основе их качества (выбирать только те из них, где человек отчётливо различим и не перекрывается другими объектами); возможна дальнейшая оптимизация алгоритмов ранжирования (например, с помощью алгоритмов k приближённых соседей или деревьев поиска); добавление признаков в галерею также может учитывать качество ограничивающих окон треклета, на основе которых он был выбран; обработка данных каждого объекта может выполняться параллельно; благодаря работе с треклетом объекта вызов всего конвейера повторного распознавания может осуществляться не на каждой итерации обработки очередного кадра видеопотока, а намного реже.

Глава 3. Разработка программных средств, реализующих распознавание пола и возраста и повторную идентификацию людей

3.1 Инструменты разработки

Для программной реализации алгоритмов распознавания пола и возраста, а также алгоритмов создания индивидуальных графических отпечатков объектов для повторного распознавания, которые используют современные нейросетевые подходы для решения задач, в качестве основного языка программирования был выбран Python. Выбор в пользу данного языка обоснован следующими причинами:

1. Python является основным языком программирования, на котором разработана исходная платформа детектирования и трекинга. Использование данного языка позволяет значительно упростить и повысить эффективность внедрения новых разработанных алгоритмов в состав модуля ядра платформы.

2. Язык Python является одним из наиболее используемых для проведения научных исследований и разработок, связанных с алгоритмами машинного и глубокого обучения, а также компьютерного зрения. Это объясняется простотой его синтаксиса и сравнительно высокой скоростью исполнения.

3. Для данного языка программирования существует множество фреймворков и библиотек, позволяющих как осуществлять научные исследования в сферах искусственного интеллекта, так и выполнять быстрое прототипирование разработанных алгоритмов для их тестирования и оценки.

4. Python позволяет быстро и легко внедрить результаты научных исследований (алгоритмы и модели) в работу сервисов и приложений без внесения изменений в исходный код, создавая тем самым единообразную среду для выполнения программ.

5. Большинство открытых программных инструментов, наиболее часто применяемых в задачах машинного обучения и компьютерного зрения, имеют свои реализации на языке Python.

6. Большинство современных научных разработок в сфере искусственного интеллекта имеют открытый исходный код на языке Python, что позволяет осуществлять их быстрое внедрение в проект.

7. Язык программирования Python находится на первом месте в мировом рейтинге наиболее популярных языков программирования ТЮВЕ по состоянию на декабрь 2022 года. Это позволяет во время разработки не только опираться на множество готовых

инструментов, созданных мировым сообществом учёных и разработчиков, но и рассчитывать на длительную поддержку и актуальность выбранных алгоритмов и моделей.

Для реализации алгоритмов распознавания пола и возраста был выбран подход, использующий модель глубокой нейронной сети для сегментации изображений ограничивающих окон объектов, создающий GEI-изображение по сегментированным изображениям треклета и использующий свёрточную нейронную сеть на базе ResNet50 для одновременной классификации пола и возраста.

В качестве инструментов для разработки модели сегментации была выбрана открытая библиотека PaddleSeg, предоставляющая большой выбор предварительно обученных моделей сегментации для различных задач. Выбор в пользу PaddleSeg был сделан благодаря тому, что среди найденных решений, данное позволяло решить одну из наиболее критичных проблем экспериментальной версии алгоритмов – недостаток данных, релевантных задачам, свойственным исходной платформе. Для решения данной проблемы PaddleSeg, в отличие от других решений, предлагает предварительно обученные модели на данных, близких к наиболее общим условиям съёмки, где люди попадают в кадр в полный рост. Использование моделей из PaddleSeg позволяет осуществить подход переноса обучения и улучшить показатели качества решения задачи сегментации человека на изображении. В качестве конкретной реализации была отобрана модель MobileSeg, использующая в качестве backbone-сети MobileNetV3 – это модель, которая обеспечивает наилучший компромисс между качеством и производительностью и может быть легко внедрена в модуль ядра платформы.

Для разработки модели классификации пола и возраста на основе предварительно обученных весов ResNet50 был выбран фреймворк TensorFlow с надстройкой Keras, т.к. он уже используется в работе ядра платформы, у разработчиков есть широкий опыт работы с ним, а Keras сразу предоставляет все необходимые инструменты для подгрузки весов предварительно обученной модели и реализации подхода переноса обучения.

Для разработки алгоритмов создания индивидуальных отпечатков объектов по графическим признакам для повторного распознавания используются модели и методы из открытой библиотеки Torchreid, написанной на популярном фреймворке для реализации моделей глубокого обучения PyTorch и включающей в себя множество методов для извлечения и обработки графических признаков для задачи повторного распознавания. Библиотека Torchreid была выбрана в силу того, что она поддерживает множество современных SOTA-моделей и наборов данных для задачи ре-идентификации; имеет открытую реализацию, активно поддерживаемую сообществом; поддерживает методы работы как с изображениями, так и с видеопотоками; имеет модульный дизайн с

возможностью его простого расширения собственной функциональностью. Все модели и методы, использующиеся для разработки алгоритмов повторного распознавания по графическим признакам, были взяты из библиотеки Torchreid.

3.2 Реализация программного модуля распознавания пола и возраста

Для реализации разработанных алгоритмов распознавания пола и возраста, а также их внедрения в работу модуля ядра платформы детектирования и трекинга использовался язык программирования Python, открытая библиотека PaddleSeg для модели сегментации человека на изображениях ограничивающих окон и фреймворк TensorFlow с надстройкой Keras для работы с нейросетевой моделью классификации пола и возраста по GEI-изображениям, описывающим треклет отслеживаемого объекта.

Программные компоненты алгоритмов распознавания пола и возраста реализовывались таким образом, чтобы быть затем легко встраиваемыми в конвейер алгоритмов обработки видеопотоков модуля ядра: компоненты реализовывались согласно модульной парадигме проектирования программных систем, используемые методы могут быть легко конфигурируемыми, вся архитектура подсистемы является поддерживаемой и открытой для дальнейших оптимизаций производительности и улучшения качества.

Ниже представлены диаграммы классов основных функциональных блоков алгоритма распознавания пола и возраста. Для данных диаграмм были сделаны следующие допущения:

- диаграммы описывают отдельные логические блоки, которые были получены путём условного разделения для удобства представления общей структуры программной системы;
- для классов, уже присутствовавших на приведённых ранее диаграммах или подробно описываемых в других диаграммах, принято упрощённое отображение, включающее в себя только название класса;
- на диаграммах показаны только наиболее важные с точки зрения логики работы алгоритма распознавания пола и возраста на основе GEI-признаков классы, их поля и методы; поля и методы классов, не относящиеся к прямой функциональности алгоритма опущены;
- в целях упрощения на диаграммах опущены некоторые библиотеки функций;
- опущены некоторые аргументы методов классов, не влияющие на основную логику работы;
- значения, определённые в классах-перечислениях, приведены в виде строковых переменных;

– на диаграммах принято следующее упрощение отображения конструкторов классов: опущены аргументы конструкторов, которые полностью совпадают по названию с полями класса.

На Рисунке 23 представлена диаграмма классов функционального блока извлечения признаков для распознавания пола и возраста, к которым относятся сегментация изображений ограничивающих окон объектов, получение треклета, создание GEI-изображения как операция извлечения признаков из треклета.

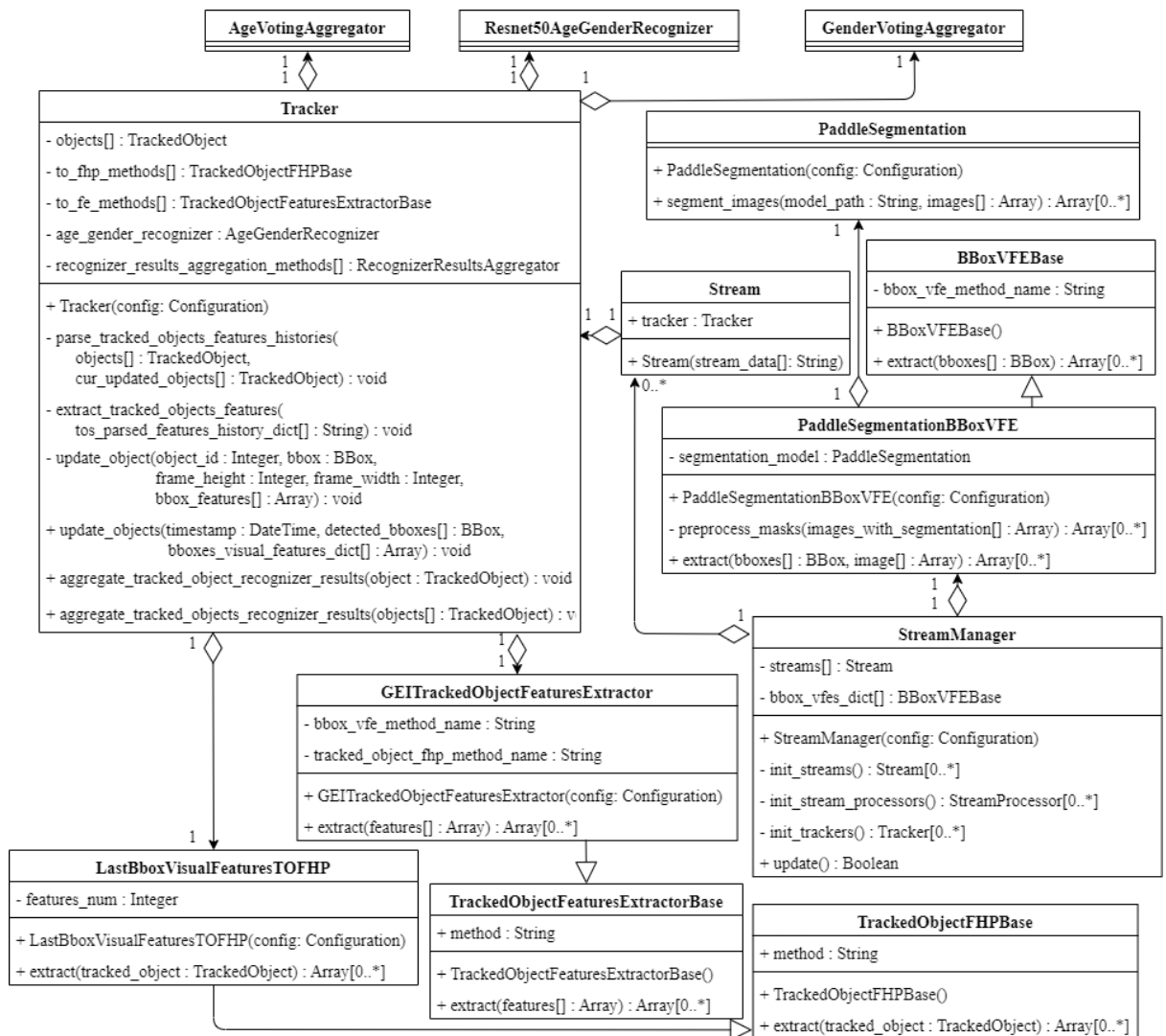


Рисунок 23. Диаграмма классов функционального блока извлечения признаков для распознавания пола и возраста

На диаграмме классов Рисунка 23 все сущности можно разделить на блоки извлечения графических признаков из изображения ограничивающего окна (классы PaddleSegmentationBBoxVFE и BBoxVFEBase, а также PaddleSegmentation), выбора треклета, описываемого признаками ограничивающих окон объектов (классы LastBboxVisualFeaturesTOFHP и TrackedObjectFeaturesExtractorBase), извлечения признаков из треклетов в виде GEI-изображения (GEITrackedObjectFeaturesExtractor и

TrackedObjectFeaturesExtractorBase), а также блок классов, связывающих работу алгоритмов распознавания пола и возраста с конвейером видеоаналитики модуля ядра платформы (классы StreamManager, Stream и Tracker). Отдельно показаны классы Resnet50AgeGenderRecognizer, AgeVotingAggregator и GenderVotingAggregator, которые более подробно описываются на следующей диаграмме.

С точки зрения конвейера обработки видеопотоков модуля ядра, функциональность извлечения признаков из ограничивающих окон представляет собой отдельный компонент, работа которого вызывается с помощью класса StreamManager в методе update, где сосредоточена вся логика обработки очередного кадра видеопотока. В методе update происходит вызов метода extract класса PaddleSegmentationBBoxVFE, наследника класса BBoxVFEBase, где производится обращение к методу segment_images класса PaddleSegmentation, который, в свою очередь, инкапсулирует работу с обученной моделью MobileSeg библиотеки PaddleSeg. Сегментированные изображения после обработки в методе preprocess_masks возвращаются в StreamManager, где привязываются к текущим отслеживаемым объектам. Так осуществляется извлечение графических признаков из ограничивающих окон объектов, необходимое для дальнейшего построения GEI-изображения по треклету. Вся последующая логика работы алгоритма распознавания пола и возраста осуществляется путём вызова методов соответствующих классов из класса Tracker, т.к. необходимым условием работы алгоритма является завершение обновления положения объектов на текущем обрабатываемом кадре – после получения результатов от алгоритма трекинга.

Нейросетевая модель классификации пола и возраста работает с GEI-изображениями, для их расчёта необходимо выделить признаки треклета отслеживаемого объекта. Данный механизм является двухэтапным. Для задания треклета используется метод extract класса LastBboxVisualFeaturesTOFHP (TOFHP – аббревиатура Tracked Object Features History Parser – парсер истории признаков отслеживаемого объекта), наследника класса TrackedObjectFHPBase, где берутся бинарные маски сегментации фиксированного количества последних ограничивающих окон объектов. Для непосредственно создания GEI-изображения предназначен класс GEITrackedObjectFeaturesExtractor, наследник класса TrackedObjectFeaturesExtractorBase, который реализует построение GEI в методе extract, используя полученные ранее признаки треклета.

Имея GEI-изображения, описывающие треклеты отслеживаемых объектов, могут быть запущены процессы распознавания пола и возраста, инкапсулированные в классах на диаграмме Рисунка 24.

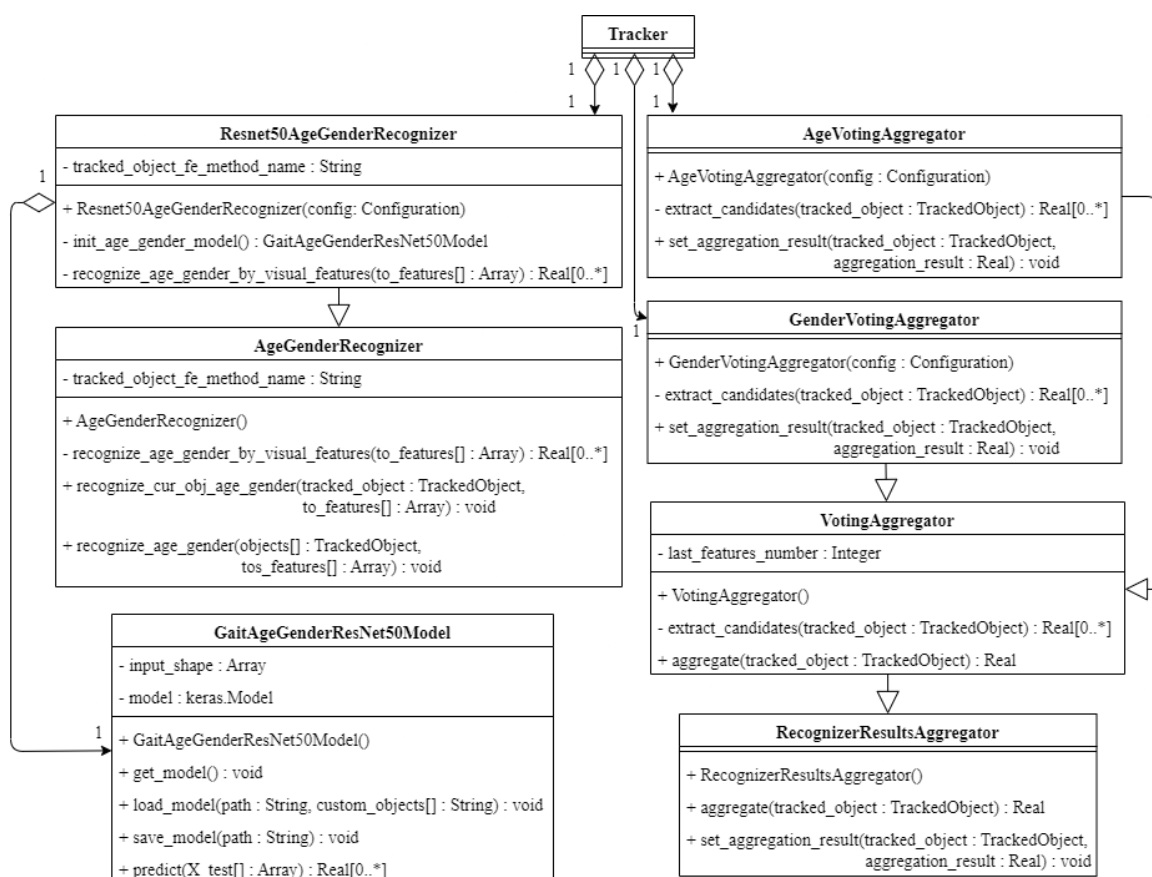


Рисунок 24. Диаграмма классов функционального блока непосредственно алгоритмов распознавания пола и возраста

Как уже упоминалось при описании концепции работы разработанного алгоритма распознавания пола и возраста, сам процесс распознавания является двухэтапным: классификация пола и возраста нейросетевой моделью на базе ResNet50 (за это отвечают классы `Resnet50AgeGenderRecognizer` и `AgeGenderRecognizer`, а также `GaitAgeGenderResNet50Model`), агрегация множества результатов классификации для всего трека объекта (инкапсулировано в классах `AgeVotingAggregator` и `GenderVotingAggregator`, наследниках `VotingAggregator` и `RecognizerResultsAggregator`).

Распознавание пола и возраста осуществляется в методе `recognize_age_gender_by_visual_features` класса `Resnet50AgeGenderRecognizer`, который во время своего создания загружает обученную нейросетевую модель с помощью метода `init_age_gender_model`, работа с которой инкапсулирована в классе `GaitAgeGenderResNet50Model` (методы `get_model`, `load_model`, `save_model`, `predict`). Распознавание осуществляется для всех текущих отслеживаемых объектов, что вызывается методами `recognize_age_gender` и `recognize_cur_obj_age_gender` базового класса `AgeGenderRecognizer`. Результатом распознавания является привязывание ко всем объектам информации о поле и возрасте для текущего треклета с соответствующими вероятностями, предоставленными нейросетевой моделью.

Агрегация результатов распознавания для всего трека отслеживаемых объектов осуществляется в классах `AgeVotingAggregator` и `GenderVotingAggregator`. Метод `aggregate` базового класса `VotingAggregator` вызывается для каждого отслеживаемого объекта, который хранит историю результатов распознавания пола и возраста. В методах `extract_candidates` классов-наследников реализуется корректное извлечение из объектов информации о поле и возрасте, а в методах `set_aggregation_result` для объекта устанавливается окончательное решение о поле и возрасте, актуальное на момент обработки текущего кадра видеопотока.

3.3 Реализация программного модуля распознавания пола и возраста

Для реализации разработанных алгоритмов создания индивидуальных отпечатков объектов по графическим признакам для повторного распознавания, а также внедрения этих алгоритмов в работу модуля ядра исходной платформы в качестве основного языка программирования использовался Python. Реализация алгоритмов осуществлялась с использованием библиотеки `Torchreid` с открытым программным кодом.

Программные компоненты разработанной версии алгоритмов создания графических отпечатков объектов для повторного распознавания проектировались и реализовывались таким образом, чтобы их можно было легко встроить в конвейер обработки видеопотоков модуля ядра исходной платформы, легко настраивать параметры их работы, а также легко модифицировать в том числе с целью оптимизации производительности и улучшения качества.

Ниже представлены диаграммы классов основных функциональных блоков алгоритма создания индивидуальных отпечатков объектов по графическим признакам для повторного распознавания. Для данных диаграмм были сделаны следующие допущения:

- диаграммы описывают отдельные логические блоки, которые были получены путём условного разделения для удобства представления общей структуры программной системы;
- для классов, уже присутствовавших на приведённых ранее диаграммах или подробно описываемых в других диаграммах, принято упрощённое отображение, включающее в себя только название класса;
- на диаграммах показаны только наиболее важные с точки зрения логики работы алгоритма создания графических отпечатков объектов для повторного распознавания классы, их поля и методы; поля и методы классов, не относящиеся к функциональности создания отпечатков или повторного распознавания опущены;
- в целях упрощения на диаграммах опущены некоторые библиотеки функций;

- опущены некоторые аргументы методов классов, не влияющие на основную логику работы;
- значения, определённые в классах-перечислениях, приведены в виде строковых переменных;
- на диаграммах принято следующее упрощение отображения конструкторов классов: опущены аргументы конструкторов, которые полностью совпадают по названию с полями класса.

На Рисунке 25 представлена диаграмма классов функционального блока извлечения графических признаков разработанного алгоритма.

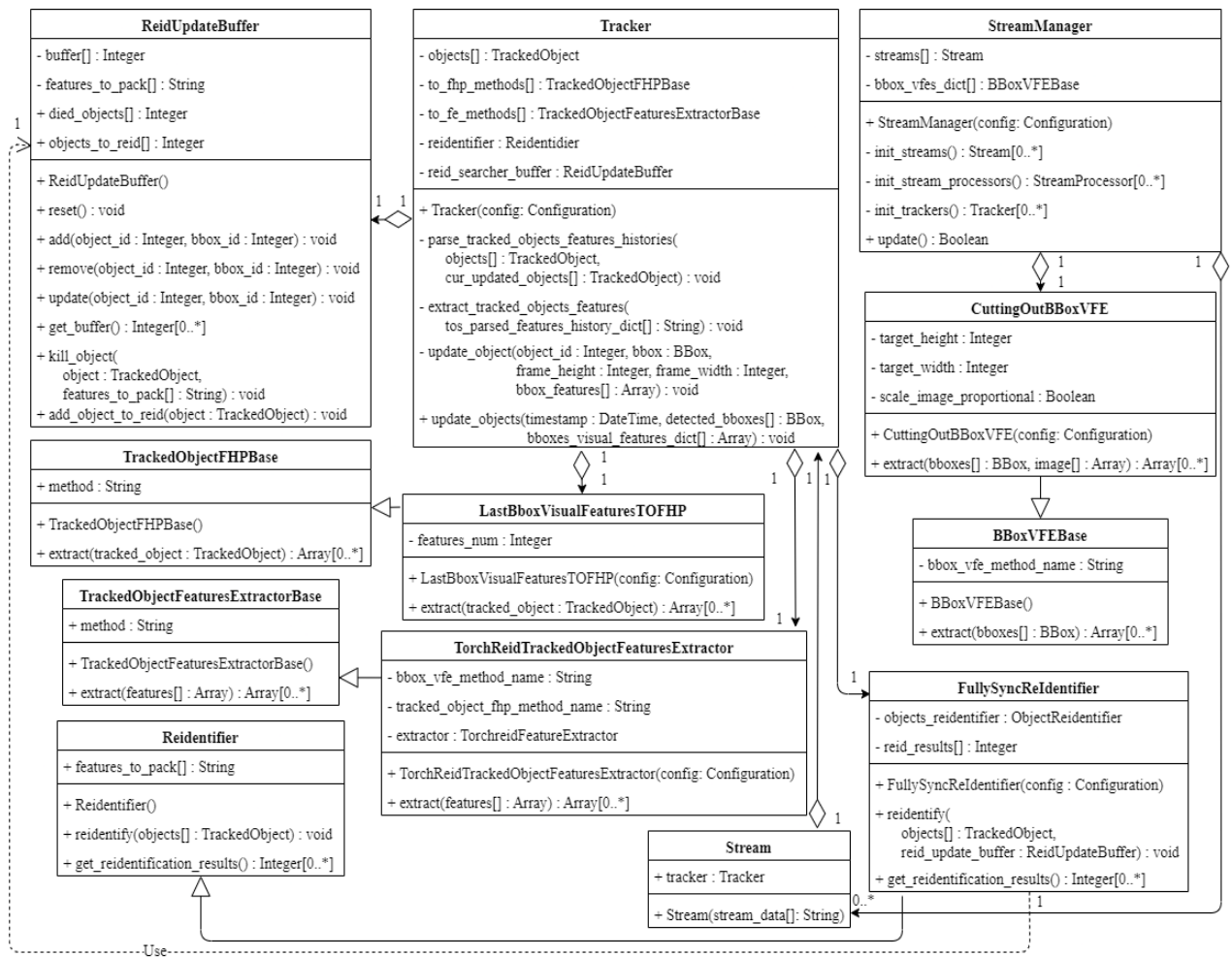


Рисунок 25. Диаграмма классов функционального блока извлечения графических признаков для повторного распознавания объектов

Классы, представленные на Рисунке 25, описывают функциональность извлечения графических признаков из ограничивающих окон (классы CuttingOutBBoxVFE и BBoxVFEBASE) и из треклетов объектов (классы LastBboxVisualFeaturesTOFHP и TrackedObjectFHPBase для выбора треклета и классы TorchReidTrackedObjectFeaturesExtractor и TrackedObjectFeaturesExtractorBase для извлечения признаков), а также связывают обработку видеопотока с функциональным

блоком повторного распознавания (ре-идентификации) объектов (классы `ReidUpdateBuffer`, `FullySyncReIdentifier` и `Reidentifier`, связанные с классом `StreamManager` через классы `Stream` и `Tracker`).

В текущей реализации извлечение графических признаков из ограничивающих окон реализуется с помощью метода `extract` класса `CuttingOutBBoxVFE`. При этом, функционально – это кадрирование всего изображения и получение его части, соответствующей ограничивающему окну объекта на текущем кадре видеопотока.

Извлечение графических признаков из трекета объекта функционально разделено на выбор трекета определённой длины (метод `extract` класса `LastBboxVisualFeaturesTOFHP`; `TOFHP` – аббревиатура `Tracked Object Features History Parser` – парсер истории признаков отслеживаемого объекта), представляющий собой набор признаков ограничивающих окон, и непосредственно извлечение графических признаков с помощью метода `extract` класса `TorchReidTrackedObjectFeaturesExtractor`, использующего обученную модель `MLFN` из библиотеки `Torchreid`. Инициация извлечения признаков из трекета осуществляется в классе `Tracker` с помощью последовательного вызова методов `parse_tracked_objects_features_histories` и `extract_tracked_objects_features`.

Повторное распознавание с использованием извлечённых графических признаков трекета на верхнем уровне осуществляется с помощью класса `FullySyncReIdentifier` – наследника `Reidentifier`. В данном случае, приставка `FullySync` в названии класса определяет специфику его работы – он работает синхронно с потоком обработки объектов на кадре видеопотока в трекере. В будущем, для оптимизации работы алгоритмов, возможен асинхронный запуск алгоритма повторного распознавания с накопленной в данных класса `ReidUpdateBuffer` историей отслеживаемых объектов (класс `ReidUpdateBuffer` хранит историю объектов и их признаков от предыдущего до текущего момента вызова алгоритма повторного распознавания, а также предоставляет программный интерфейс для управления этими данными). Алгоритм повторного распознавания реализуется в методе `reidentify`, который обновляет внутренние состояния компонентов ре-идентификации, после чего их можно получить с помощью метода `get_reidentification_results`.

Функциональный блок непосредственно алгоритмов повторного распознавания на основе отпечатков графических признаков объектов в виде диаграммы классов представлен на Рисунке 26.

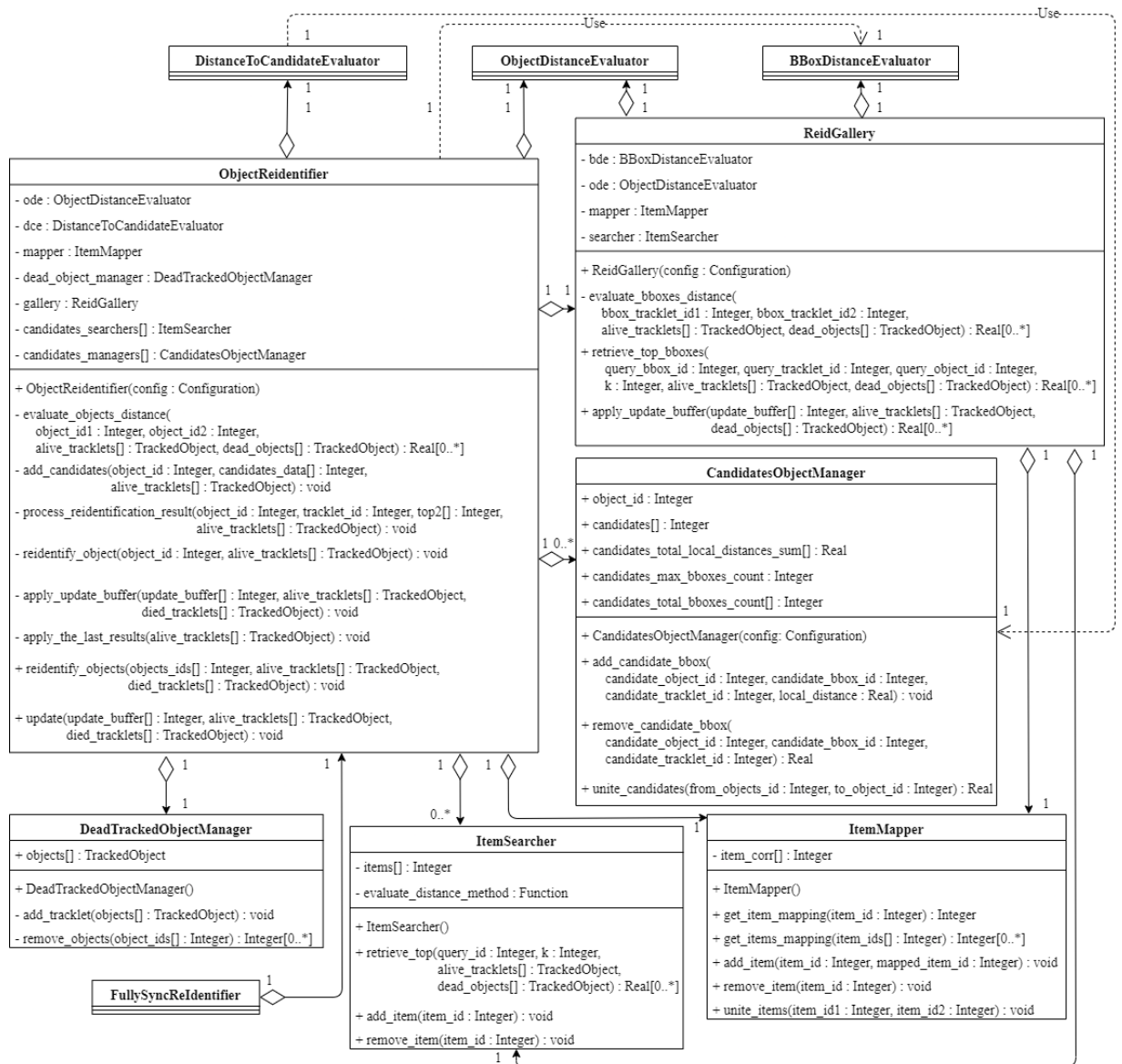


Рисунок 26. Диаграмма классов функционального блока непосредственных алгоритмов ре-идентификации объектов

Непосредственно функциональность повторного распознавания на уровне отслеживаемых объектов реализуется в классе `ObjectReidentifier`. Повторное распознавания или ре-идентификация осуществляется среди объектов, ссылки на которые хранятся в объектах классов `ReidGallery` и `DeadTrackedObjectManager` – в них содержится информация о текущих и уже покинувших сцену объектах. Для поиска и сопоставления треклетов объектов по значениям дескрипторов схожести между их графическими признаками используются классы `ItemSearcher` и `ItemMapper` соответственно (при этом, эти же классы используются на верхнем уровне в классе `ObjectReidentifier` для поиска и сопоставления треков объектов, но уже с другим дескриптором схожести). Класс `CandidatesObjectManager` инкапсулирует методы для управления данными отобранных в процессе ранжирования алгоритма повторного распознавания кандидатов – среди них осуществляется агрегация и

принятие окончательного решения об идентификаторе текущего объекта. Классы `DistanceToCandidateEvaluator`, `ObjectDistanceEvaluator` и `BBoxDistanceEvaluator` реализуют функции расчёта дескрипторов схожести для сравнения признаков текущего объекта и всех отобранных кандидатов, треклетов одного объекта и ограничивающих окон объекта соответственно. Их поля и методы на Рисунке 26 опущены для экономии места, более глубокое рассмотрение их реализации описывается далее.

Класс `ReidGallery` является центральным в реализации алгоритмов повторного распознавания и ответственен за извлечение `top`'а отсортированных признаков и соответствующих им треклетов (метод `retrieve_top_bboxes`) среди всех когда либо появлявшихся на видеопотоке объектов (текущих и покинувших сцену) с помощью расчёта дескрипторов схожести (метод `evaluate_bboxes_distance`) между признаками ограничивающих окон (используется объект `bde` класса `BBoxDistanceEvaluator`) или треклетов (используется объект `ode` класса `ObjectDistanceEvaluator`). Для поиска используется объект `searcher` класса `ItemSearcher`, в котором реализуется расчёт дескрипторов схожести между графическими признаками `query` и признаками, хранящимися в галерее, и их сортировка с целью формирования `top`'а наиболее близких к `query` (метод `retrieve_top` класса `ItemSearcher`). Для сопоставления идентификаторов треклетов и идентификаторов объектов используется объект `mapper` класса `ItemMapper` (методы `get_item_mapping` и `get_items_mapping` класса `ItemMapper`). Для обновления информации о существующих объектах, которое может выполняться с заданной периодичностью, используется метод `apply_update_buffer`, который, в свою очередь, инициирует обновление данных, хранящихся в объекте `searcher`.

Отобранные на уровне `ReidGallery` `top` идентификаторов объектов и признаков их треклетов, наиболее релевантных `query`, называемые кандидатами, передаются на верхний уровень – в класс `ObjectReidentifier`, где в первую очередь добавляются в списки `candidates_managers` (коллекция объектов класса `ItemSearcher`) и `candidates_searchers` (коллекция объектов класса `CandidatesObjectManager`) для определённого объекта, которому соответствовал `query` (метод `add_candidates`). Затем осуществляется поиск и сортировка признаков уже среди отобранных кандидатов на уровне всего объекта – производится этап агрегации нескольких решений алгоритма реидентификации с помощью функции взвешенного среднего (`weighted mean`). Окончательное решение принимается по двум наиболее релевантным решениям (метод `process_reidentification_result`): если идентификатор, не соответствующий идентификатору текущего объекта, оказывается среди этих двух решений, то окончательным решением является смена идентификатора (это позволяет исправлять ошибки трекера типа смена идентификаторов (`id switch`)), в

противном случае, если среди этих решений нет идентификаторов, отличных от текущего, данные объекта не меняются. Весь процесс ре-идентификации инициируется в методе `reidentify_objects` класса `ObjectReidentifier`, опускается до уровня расчётов в `ReidGallery`, а окончательный результат возвращается в самый верхнеуровневый класс `FullySyncReIdentifier` (всё происходит в методе `reidentify`). После завершения работы алгоритма повторного распознавания, в методе `reidentify` класса `FullySyncReIdentifier` вызывается метод `update` класса `ObjectReidentifier` для обновления данных в галерее и всех коллекций объектов классов `ItemSearcher` и `CandidatesObjectManager`, помимо этого, метод `apply_the_last_results` применяет результаты ре-идентификации к хранимым в коллекциях `candidates_managers` и `candidates_searchers` данным.

Функциональный блок расчёта дескрипторов схожести алгоритмов повторного распознавания представлен на диаграмме классов Рисунка 27.

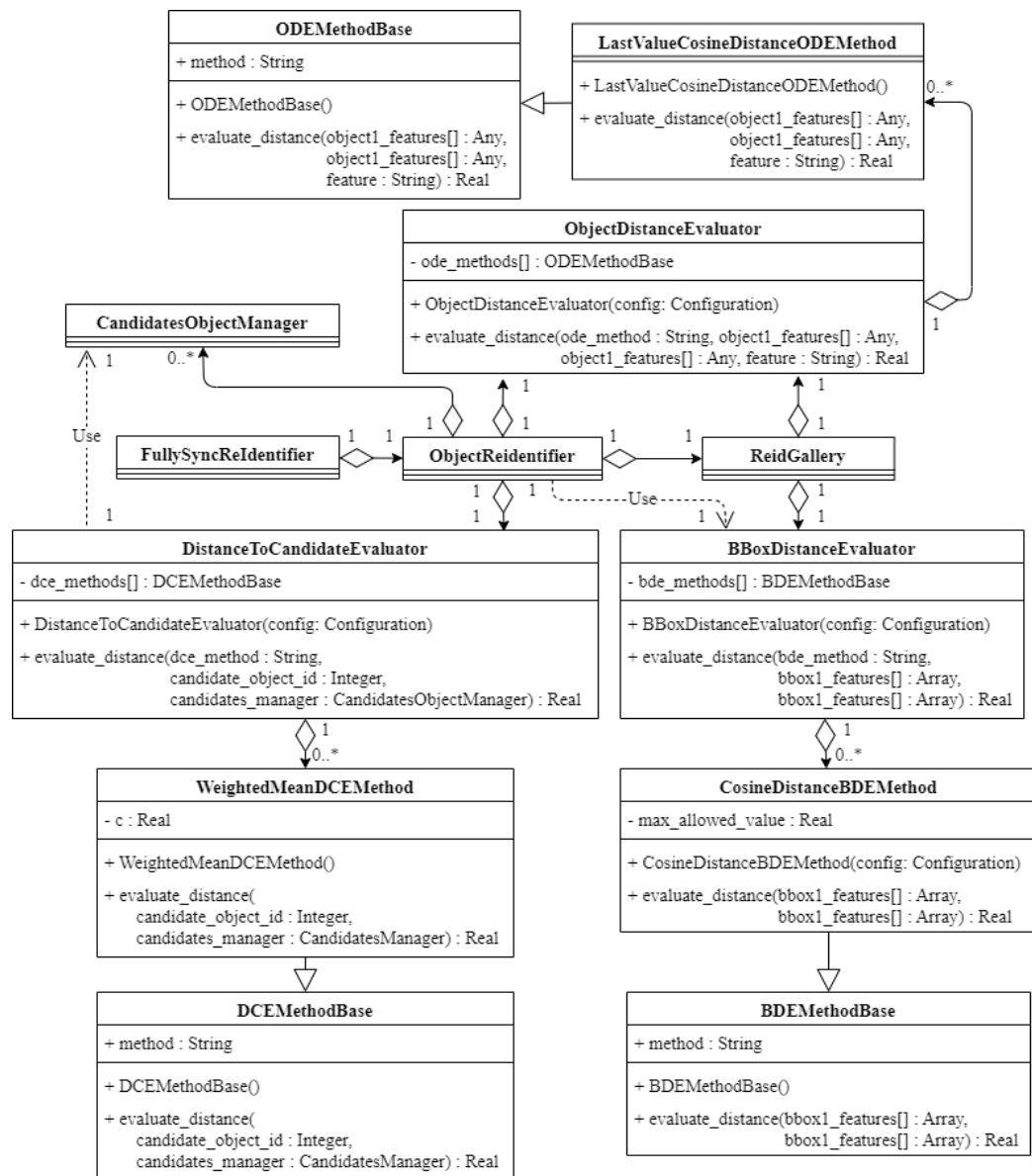


Рисунок 27. Диаграмма классов функционального блока расчёта дескрипторов схожести, используемого алгоритмами ре-идентификации объектов

На Рисунке 27 детально рассмотрены поля и методы классов для расчёта дескрипторов схожести ограничивающих окон (класс `BBoxDistanceEvaluator` с методом `CosineDistanceBDEMethod`, реализующим расчёт косинусного расстояния), треклетов и объектов (класс `ObjectDistanceEvaluator` с методом `LastValueCosineDistanceODEMethod`, реализующим расчёт косинусного расстояния между графическими признаками треклетов), извлечённых из галереи кандидатов для всего объекта (класс `DistanceToCandidateEvaluator` с методом `WeightedMeanDCEMethod`, реализующим взвешенное среднее). Все классы реализуют расчёт дескриптора схожести как расчёт расстояния между векторами графических признаков с помощью методов `evaluate_distance`.

Глава 4. Проведение испытаний алгоритмов распознавания пола и возраста и повторной идентификации людей

4.1 Разработка и разметка обучающих выборок

4.1.1 Данные для задачи сегментации объектов

Одним из важных выводов, сделанных по результатам разработки экспериментальной версии алгоритмов определения пола и возраста людей на первом этапе НИОКР, является необходимость расширения применяемых методов в пользу подходов, использующих информацию о силуэте или походке. Такие методы представляют собой один из способов решения задачи определения пола и возраста в случаях с плохим освещением и большой удаленностью объектов от камеры, когда распознать лицо не представляется возможным. Применение одного из таких методов, основанного на GEI-признаках, выбранного в качестве одного из перспективных для разработки алгоритмов определения пола и возраста и описанного в разделе 1, невозможно без модели сегментации объектов, которая должна обучаться на данных, имеющих специальную разметку для таких задач. Такие данные представляют собой исходные изображения, а также изображения или специальные файлы разметки, содержащие данные о полигонах на исходном изображении, которые описывают объект как сегмент изображения – ограниченный набор пикселей изображения, относящихся к данному объекту. Помимо описания полигонов, данные могут содержать дополнительную информацию о классе объектов, идентификаторе объекта и т.д. На Рисунке 28 приведён пример исходного изображения, сегмента объекта класса человек для этого изображения, а также полупрозрачное наложение сегмента на исходное изображение.

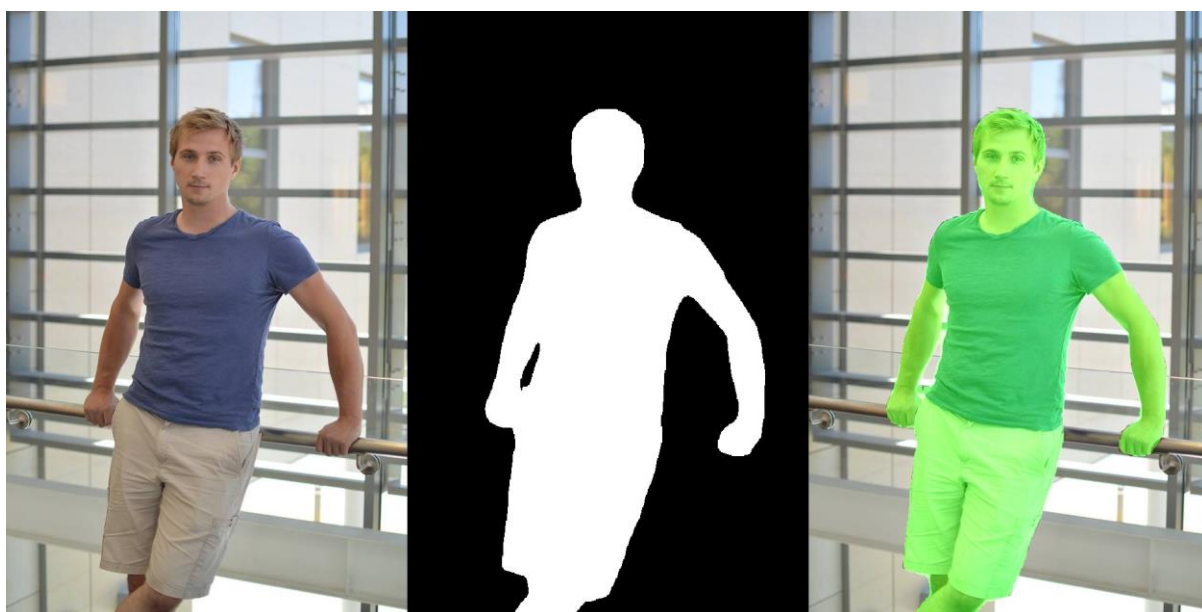


Рисунок 28. Пример изображения и разметки для задачи сегментации объектов

В настоящее время в открытом доступе находится огромное количество размеченных наборов данных, предназначенных для обучения и тестирования моделей, решающих различные подзадачи сегментации объектов на изображениях. Для подготовки модели сегментации, на базе результатов которой осуществлялось построение GEI-изображения объекта класса человек, необходимое для разработки алгоритмов определения пола и возраста людей, использовался набор данных UP-S31. Это специфический набор данных, предназначенный непосредственно для решения задачи сегментации тела человека на изображении и отделении его от остального фона. Особенность данного набора заключается в разметке сегментов не просто тела человека, но отдельных его частей – пример показан на Рисунке 29.

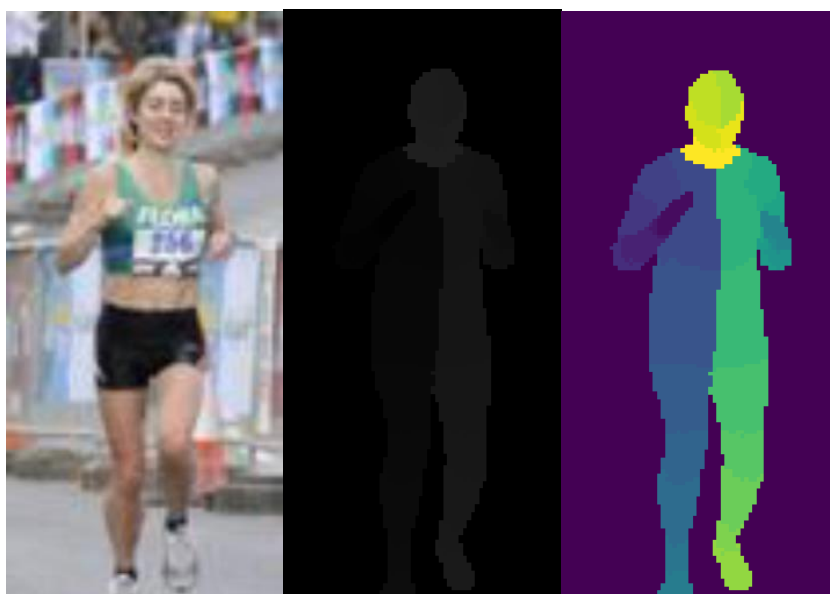


Рисунок 29. Пример исходного изображения и разметки сегментов тела человека из набора данных UP-S31

Исходный набор данных содержит в себе 8515 размеченных изображений. Для работы с набором не использовалась информация об отдельных частях тела – все части брались в совокупности для определения сегмента всего тела человека. В силу небольшого количество исходных данных, они подвергались аугментации с помощью следующих преобразований: изменение пропорций изображения, кадрирование изображения, эффект блюра, цветовая нормализация, цветокоррекция и аффинные преобразования. Аугментации применялись случайным образом на этапе обучения модели сегментации. Само обучение проводилось на 80 % данных от их изначального количества (6812 примеров в обучающей выборке), а качество валидировалось на оставшихся 20 % (1703 примеров в валидационной выборке).

4.1.2 Данные для задачи определения пола и возраста

На первом этапе НИОКР для оценки экспериментальных версий алгоритмов определения пола и возраста в качестве данных для тестирования использовался собственный набор, специально созданный для этой цели. Набор создавался в силу отсутствия в открытом доступе данных для распознавания лиц, условия съёмки которых были бы близкими к произвольным, которые являются типичными данными, загружаемыми пользователями на платформу детектирования и трекинга. Однако результаты экспериментов показали, что подготовленный набор имеет ряд недостатков, не позволяющих корректно оценить качество работы экспериментальных версий алгоритмов: отсутствие в выборке возрастов до 15 лет и более 45 лет; однородность с точки зрения внешнего вида людей и внешних условий съёмки; небольшое количество примеров в наборе, не позволяющее использовать его для целей обучения моделей по распознаванию пола и возраста. Одной из задач разработки алгоритмов определения пола и возраста являлась задача расширения обучающей и валидационной выборок данных примерами, близкими к требуемым произвольным внешним условиям съёмки с учётом проблем, присущих предыдущей выборке.

В силу особенностей методов определения пола и возраста, применяемых в разработанных версиях алгоритмов, которые используют GEI-признаки для решения задачи, каждый пример набора данных должен состоять из набора изображений движения одного и того же человека (треклет) в одном направлении, а разметка должна содержать информацию о поле и возрасте этого человека. Примеры таких наборов изображений представлены на Рисунке 30. Несмотря на то, что в открытом доступе существует множество наборов данных, содержащих такую информацию, все они требуют предварительной обработки и, возможно, дополнительной разметки для того, чтобы быть готовыми к использованию в рамках разработки алгоритмов распознавания пола и возраста.

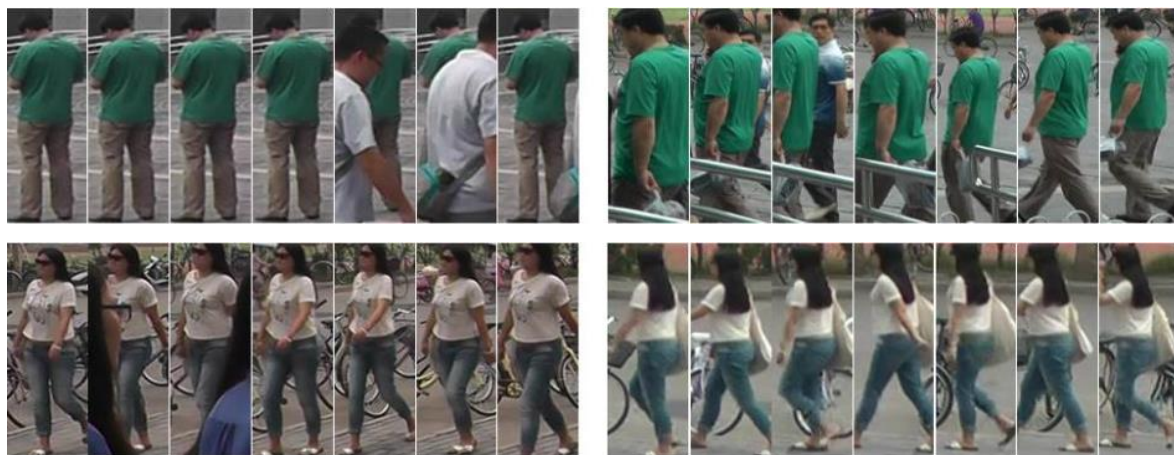


Рисунок 30. Примеры треклетов изображений одного человека

Для разметки и организации наборов данных из открытых наборов была предложена файловая структура организации треклетов изображений, включающая в себя разметку и представленная в следующем виде:

```
[dataset-1]
  [subset-1]
    [person-id_gender-id_age-id]
      img1.jpg
      img2.jpg
      ...
    ...
  [subset-2]
  ...
[dataset-2]
...
```

Структура размеченных данных представляет собой набор папок с наборами данных (dataset-1, dataset-2 и т.д.), внутри расположены подпапки поднаборов данных (subset-1, subset-2 – для отделения друг от друга, например, обучающей и валидационной выборок), подпапками поднаборов данных являются папки с треклетами, в названии которых присутствуют код-идентификатор человека (person_id), код-идентификатор пола человека (gender_id) и код-идентификатор возраста (age_id).

В качестве возможных значений идентификаторов пола использовались значения 0 – женский пол и 1 – мужской пол. В качестве возможных значений идентификаторов возраста использовались значения 1 – возраст в диапазоне до 15 лет, 2 – возраст в диапазоне от 16 до 45 лет, 3 – возраст более 45 лет.

В силу ошибок в разметке ограничивающих окон объектов многих открытых наборов данных, текущие наборы данных формировались с помощью дополнительных преобразований, которым предшествовала фильтрация для исключения некорректно размеченных ограничивающих окон людей (окно включает в себя несколько людей, окно некорректно размечает человека). Каждый треклет формировался путём кадрирования изображения и выделения на нём человека по разметке определённого ограничивающего окна. Эта разметка уточнялась с помощью предварительно обученной модели сегментации: выделялся силуэт человека, по крайним точкам силуэта определялись координаты итогового ограничивающего окна, а в треклет попадало изображение-результат кадрирования по заново размеченным ограничивающим окнам объектов. Пример одного из кадров треклета и результата предварительного преобразования представлен на Рисунке 31.



Рисунок 31. Пример исходного изображения человека из треклета набора данных для определения пола и возраста (слева) и результата автоматической предобработки (справа)

В Таблице 1 приведены использованные наборы данных, подготовленные для обучения и валидации разработанных алгоритмов распознавания пола и возраста, а также их основные характеристики.

Таблица 1 – Открытые наборы данных, подготовленные для построения моделей определения пола и возраста

Название набора данных	Количество изображений	Количество уникальных объектов	Соотношение классов пола: 0 / 1	Соотношение классов возраста: 1 / 2 / 3
PETA – Town Centre	6968	159	62 / 97	0 / 106 / 53
MARS	1191003	1028	605 / 423	56 / 900 / 72

С помощью указанных в таблице наборов были решены как проблемы недостатка данных для обучения, так и в большей степени проблемы однородности, однообразия внешних условий съёмки, соотношения классов пола и возраста; данные приближены к типичным для платформы детектирования и трекинга условиям – видео с камер уличного наблюдения или съёмок внутри здания. Помимо оригинальных изображений треклетов во время обучения моделей определения пола и возраста также использовались и аугментированные изображения, где в качестве методов аугментации применялись: зеркалирование, аффинные преобразования. Для обучения и тестирования моделей наборы данных были разделены на обучающую и валидационную подвыборки в соотношении 75 к 25 % соответственно.

4.1.3 Данные для задачи создания индивидуальных отпечатков объектов по графическим признакам

Данные, подготовленные для построения моделей, позволяющих решить задачу создания индивидуальных отпечатков объектов по графическим признакам для их повторного распознавания, были представлены открытыми наборами. Типичные данные для решения подобных задач представляют собой треклеты (последовательность изображений одного и того же человека), с которыми соотносится информация об идентификаторе объекта; при этом, в наборе содержится набор различных треклетов одного и того же человека как частей одного трека или как результатов записи с использованием нескольких камер. Примеры данных для задачи повторного распознавания представлены на Рисунке 32.

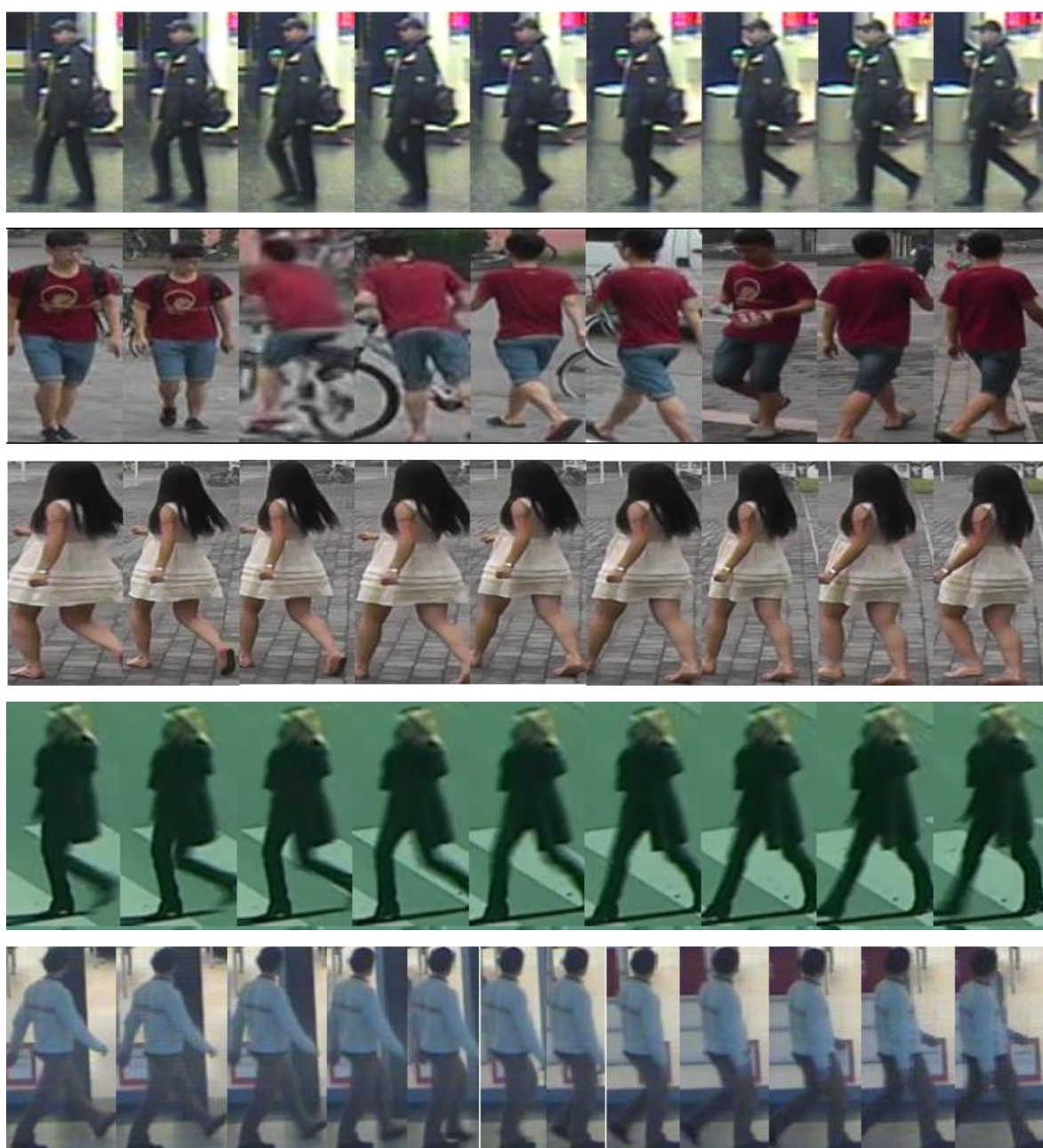




Рисунок 32. Примеры треклетов в данных для задачи повторного распознавания

Разметка данных для повторного распознавания обеспечивается файловой структурой их расположения, которая выглядит следующим образом:

```
[dataset-1]
  [subset-1]
    [person-id]
      [tracklet-id]
        img1.jpg
        img2.jpg
        ...
      ...
    [subset-2]
    ...
  [dataset-2]
  ...
```

Файловая структура наборов данных для повторного распознавания похожа на структуру данных для задачи определения пола и возраста, но на уровне поддиректории, относящейся к определённому человеку, дополнительно присутствует уровень, описывающий каждый треклет этого человека (tracklet-id).

К данным для повторного распознавания также как и к данным для определения пола и возраста применялся алгоритм предварительной обработки, устраняющий проблемы неточной разметки границ объектов.

В Таблице 2 представлены данные, используемые для разработки алгоритмов создания индивидуальных графических отпечатков объектов для обеспечения возможности повторного распознавания, а также их основные характеристики.

Таблица 2 – Открытые наборы данных, подготовленные для построения моделей создания индивидуальных отпечатков объектов по графическим признакам

Название набора данных	Количество изображений	Количество уникальных объектов
CAVIAR	177759	121
I-LIDS-VID	43059	300
Market-1501	32668	1259
MARS	1191003	1261
PRID2011	96121	934
DukeMTMC-VideoReID	811776	1812

Представленные в таблице открытые наборы значительно расширяют количество данных для разработки алгоритмов повторного распознавания людей на основе индивидуальных отпечатков по графическим признакам, диверсифицируют условия съёмки и приближают данные к тем общим условиям, которые свойственны загружаемым пользователями данным на платформу детектирования и трекинга. Для обучения представленные данные в равном соотношении разбивались на поднаборы для обучения, для описания галереи и для описания входного набора признаков (query). При этом, набор данных CAVIAR использовался для тестирования и проведения экспериментов, т.к. данные в этом наборе максимально приближены к внешним условиям съёмки, свойственным для входных данных исходной платформы.

4.2 Условия проведения испытаний

Все испытания и исследования осуществлялись на физических вычислительных машинах со следующими характеристиками:

- Процессор (CPU) Intel Core i5-7600, 4 x 3.50 ГГц;
- Графический ускоритель (GPU) Nvidia GeForce GTX 1080 Ti;
- Оперативная память (RAM) 16 (DDR4) Гб;
- Операционная система Ubuntu 18.04 (64 bit).

4.3 Испытания моделей сегментации для распознавания пола и возраста

Сегментация пикселей изображения ограничивающего окна объекта является обязательной стадией функционирования разработанного алгоритма распознавания пола и возраста людей, требующей отдельной специальной модели, базирующейся на глубоких

архитектурах нейронных сетей. В качестве наиболее подходящей нейронной сети для решения поставленной задачи была выбрана модель MobileSeg из открытой библиотеки PaddleSeg, использующей в качестве backbone-сети легковесную MobileNetV3, которая позволяет уменьшить требования к производительности модуля ядра при её внедрении в работу платформы. Несмотря на то, что модели из данной библиотеки были предварительно обучены на данных изображений людей в полный рост, для применения в составе разработанного алгоритма распознавания пола и возраста её необходимо дообучить на данных, приближенных к общим условиям съёмки уличных камер видеонаблюдения, свойственных платформе детектирования и трекинга.

Для донастройки весов модели сегментации использовались собранные данные. Оценка качества работы модели на валидационной подвыборке осуществлялась с помощью нескольких метрик: Mean Intersection Over Union (MIOU) – усреднённое значение метрики IOU для всех примеров валидационной выборки, Accuracy – доля верно выделенных пикселей изображения ограничивающего окна человека (Формула 5), Precision – величина, обратная доле выделенных пикселей, не соответствующих сегменту человека (доля отсутствия ложных срабатываний для пикселей, принадлежащих сегменту человека – Формула 6), Карра – коэффициент согласованности, учитывающий случайные совпадения (Формула 7).

$$Accuracy = \frac{TP + TN}{TP + FP + FN + TN'} \quad (5)$$

где TP – True Positives – количество пикселей изображения, принадлежащих сегменту человека и выделенных моделью, TN – True Negatives – количество пикселей изображения, не принадлежащих сегменту человека и не выделенных моделью, FP – False Positives – количество пикселей изображения, не принадлежащих человеку, но выделенных моделью, FN – False Negatives – количество пикселей изображения, принадлежащих человеку, но не выделенных моделью.

$$Precision = \frac{TP}{TP + FP} \quad (6)$$

где TP – True Positives – количество пикселей изображения, принадлежащих сегменту человека и выделенных моделью, FP – False Positives – количество пикселей изображения, не принадлежащих человеку, но выделенных моделью.

$$kappa = \frac{2(TP \cdot TN - FN \cdot FP)}{(TP + FP)(FP + TN) + (TP + FN)(FN + TN)} \quad (7)$$

где TP – True Positives – количество пикселей изображения, принадлежащих сегменту человека и выделенных моделью, TN – True Negatives – количество пикселей

изображения, не принадлежащих сегменту человека и не выделенных моделью, FP – False Positives – количество пикселей изображения, не принадлежащих человеку, но выделенных моделью, FN – False Negatives – количество пикселей изображения, принадлежащих человеку, но не выделенных моделью.

Все вышеприведённые метрики оценки качества имеют область значений от 0 до 1. Чем выше значение метрики, тем выше итоговое качество.

Цель эксперимента заключалась в оценке количества эпох дообучения нейронной сети для сегментации пикселей изображения ограничивающего окна, принадлежащих человеку, с помощью приведённых выше метрик. Результаты экспериментов представлены в Таблице 3.

Таблица 3 – Результаты экспериментов по оценке количества эпох дообучения нейронной сети для сегментации человека на изображениях ограничивающих окон

Кол-во эпох	МЮ	Accuracy	Precision	Kappa
500	0,804	0,936	0,794	0,773
1000	0,814	0,940	0,797	0,786
1250	0,825	0,944	0,822	0,800
1500	0,823	0,943	0,815	0,797
2000	0,827	0,946	0,832	0,803
3000	0,831	0,947	0,831	0,808
5000	0,834	0,948	0,830	0,812
10000	0,838	0,949	0,846	0,816
20000	0,839	0,949	0,844	0,813

По результатам, приведённым в Таблице 3, видно, что количество эпох для дообучения модели сегментации должно быть не менее 10000. При этом значении измеряемые метрики качества имеют наилучшие показания. При значении 20000 метрики становятся немного хуже – сеть переобучается. Визуализация результатов работы модели также показывает заметные улучшения после её дообучения – Рисунок 33.

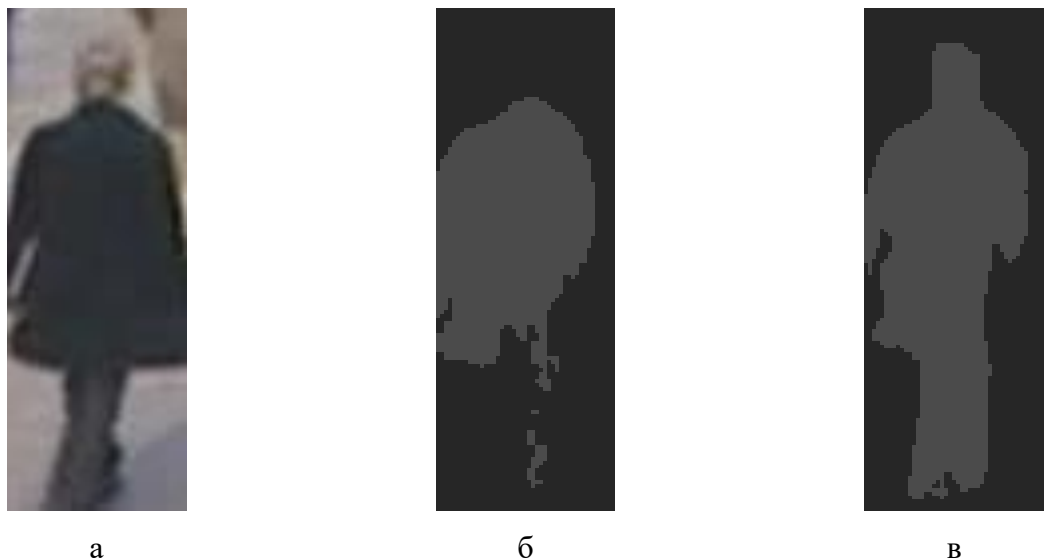


Рисунок 33. Пример сегментации с помощью выбранной модели: а – исходное изображение из набора данных, не участвовавшего в обучении и валидации, б – результат сегментации модели без дообучения, в – результата сегментации модели с дообучением на 10000 эпохах

Таким образом, результатом проведённых экспериментов является определение наилучшего количества эпох (10000) для дообучения модели сегментации человека на изображениях ограничивающих окон, приближенных к наиболее общим условиям съёмки, свойственным платформе детектирования и трекинга. От качества работы данной модели зависит качество работы всего разработанного алгоритма распознавания пола и возраста.

4.4 Испытания алгоритмов распознавания пола и возраста

Выбранные направления разработки алгоритмов распознавания пола и возраста, а также конкретный набор методов, определённый по результатам дополнительного теоретического исследования, включающий в себя работу с признаками треклета объекта в виде GEI-изображения, использование для задачи классификации свёрточных нейронных сетей на базе архитектуры ResNet, расширение данных в пользу разнообразия относительно размеченных классов, требуют проведения экспериментов для выбора наиболее подходящих и эффективных подходов для решения поставленных задач. Задачей проведения экспериментов является оценка качества работы разработанного алгоритма при использовании различных методов, входящих в его состав, и выбор наилучших из них для использования в составе модуля ядра исходной платформы.

Оценка качества работы алгоритмов распознавания пола и возраста осуществляется по трём основным метрикам, широко используемым в задачах классификации: точность (precision) – Формула 8, полнота (recall) – Формула 9 и мера F1 – Формула 10. Данные метрики рассчитываются для каждого класса в отдельности, который можно назвать целевым.

$$Precision = \frac{TP}{TP + FP}, \quad (8)$$

$$Recall = \frac{TP}{TP + FN}, \quad (9)$$

$$F1 = \frac{2 \cdot Precision \cdot Recall}{Precision + Recall}, \quad (10)$$

где TP – True Positives – количество примеров изображений людей целевого класса, определённых как принадлежащие целевому классу, TN – True Negatives – количество примеров изображений людей классов, отличных от целевого, которые не были определены как принадлежащие целевому классу, FP – False Positives – количество примеров изображений людей классов, отличных от целевого, но которые были определены как принадлежащие целевому классу, FN – False Negatives – количество примеров изображений людей целевого класса, которые были определены как принадлежащие к другому классу, отличному от целевого. Таким образом, метрика Precision показывает насколько мало ошибок типа ложное срабатывание допускает модель при распознавании определённого класса возраста, либо класса женского пола; метрика Recall показывает насколько много изображений людей целевого класса смогла распознать модель; мера F1 является средневзвешенной метрикой, объединяющей показатели Precision и Recall и показывающей общее качество работа алгоритма определения пола или возраста.

Для проведения испытаний разработанных алгоритмов распознавания пола и возраста использовались следующие методы с соответствующими гиперпараметрами:

- Архитектура модели свёрточной нейронной сети на базе ResNet50:
 1. ResNet50 с добавлением полносвязных слоёв и одного выхода для классификации пола или возраста;
 2. ResNet50 с добавлением полносвязных слоёв и двух выходов для одновременной классификации пола и возраста.
- Размер треклета отслеживаемого объекта для расчёта GEI-изображения, используемого в качестве признаков для моделей классификации пола и возраста (максимальный размер треклета определялся объёмом доступной памяти GPU для обучения и разумным временем обучения модели):
 1. Размер треклета не ограничен – для формирования GEI используется произвольное количество ограничивающих окон объекта;
 2. Размер треклета ограничен и равен числу в диапазоне от 1 до 6.
- Метод расчёта метрик для агрегации нескольких результатов работы алгоритма распознавания пола и возраста по треклету и принятия окончательного решения для всего отслеживаемого объекта:

1. Среднее значение вероятностей принадлежности к классу для каждого класса. Выбирается класс с наибольшей средней вероятностью (mean) – Формула 11;

2. Мода среди классов. В качестве окончательного решения выбирается наиболее часто встречающийся класс (mode) – Формула 12.

3. Среднее взвешенное значение вероятностей принадлежности к классу для каждого класса. Взвешивание среднего значения осуществляется на величину, равную частоте встречаемости самого популярного среди агрегируемых результатов класса (weighted mean) – Формула 13.

Формулы описанных метрик представлены в Таблице 4, где S^{cl} – значение метрики для агрегации (окончательное решение принимается по максимальному значению), p_i – значение вероятности принадлежности i -го агрегируемого результата к классу cl , N – общее количество агрегируемых результатов, N^{cl} – количество агрегируемых результатов, принадлежащих к классу cl , c – регуляровочный гиперпараметр.

Таблица 4 – Формулы расчёта метрик для агрегации решений моделей классификации пола и возраста, выбранных для проведения эксперимента

Метрика для агрегации	Формула	Номер
mean	$S^{cl} = \frac{1}{N} \sum_{i=1}^N p_i^{cl}$	(11)
mode	$S^{cl} = Mo(N)$	(12)
weighted mean	$S^{cl} = \frac{\frac{1}{N} \sum_{i=1}^N p_i^{cl}}{\left(\frac{N^{cl}}{N}\right)^c}$	(13)

Эксперименты условно разделены на 2 части: отбор моделей для классификации пола и возраста, отбор используемой с выбранной моделью функции агрегации.

Результаты экспериментов по отбору модели классификации пола и возраста с учётом различных подходов к формированию треклета представлены в Таблице 5.

Таблица 6 – Результаты экспериментов по отбору модели классификации пола и возраста

№	Кол-во вых-ов	Размер треклета	Precision				Recall				F1			
			пол	<15	15-45	>45	пол	<15	15-45	>45	пол	<15	15-45	>45
1	1	не огр.	0.75	0.40	0.50	0.63	0.65	0.48	0.22	0.35	0.69	0.43	0.40	0.50
2	1	6	0.76	0.46	0.50	0.65	0.70	0.53	0.25	0.35	0.72	0.51	0.46	0.51

3	2	не огр.	0.81	0.35	0.51	0.44	0.76	0.27	0.22	0.54	0.78	0.36	0.37	0.51
4	2	6	0.83	0.62	0.55	0.48	0.85	0.38	0.28	0.44	0.84	0.52	0.46	0.50

По результатам, приведённым в Таблице 6, видно, что более высокое качество как в задаче классификации пола, так и в задаче классификации возраста показывает модель с двумя выходами. Это может объясняться тем, что при обучении сразу на двух выходах, свёрточные слои архитектуры ResNet извлекают такие признаки, которые одновременно информативны как для задачи распознавания пола, так и для задачи распознавания возраста. Также это может говорить о том, что определить, например, возраст человека проще по признакам силуэта и движения, зная его пол. С точки зрения размера треклета, лучшее качество показывают модели, для обучения которых использовались треклеты фиксированного размера. Без ограничения размера треклета GEI-изображения, на которых производится классификация, могут значительно отличаться даже для примеров одинаковых классов пола и возраста, что не даёт нейронной сети выделить характерные признаки для того или иного класса. В результате, наилучшей моделью, показывающей метрики качества классификации пола 0.84 по мере F1 и возраста в среднем 0.49 по мере F1, является нейронная сеть с двумя выходами для одновременной классификации пола и возраста, использующая GEI-изображения, построенные по треклетам объектов фиксированного размера 6.

Результаты экспериментов по отбору методов агрегации для множества решений модели классификации пола и возраста в рамках истории одного отслеживаемого объекта в видеопотоке представлены в таблице 6. Для проведения данных экспериментов использовались значения параметров, отобранные по результатам экспериментов с моделями. Значение для показателя “ср. возр” из таблицы рассчитывалось как среднее среди трёх возрастных категорий.

Таблица 6 – Результаты экспериментов по отбору методов агрегации множества решений модели классификации пола и возраста

№	Ф-ция	Precision		Recall		F1	
		пол	ср. возр.	пол	ср. возр	пол	ср. возр
1	mean	0.79	0,77	0,89	0,63	0,84	0,69
2	mode	0,85	0,79	0,94	0,71	0,89	0,74
3	w. mean c=0.1	0,82	0,76	0,94	0,66	0,87	0,71

4	w. mean c=0.3	0,83	0,76	0,94	0,65	0,88	0,70
5	w. mean c=0.5	0,82	0,77	0,94	0,63	0,87	0,70
6	w. mean c=0.7	0,82	0,78	0,94	0,68	0,87	0,72
7	w. mean c=0.9	0,82	0,78	0,94	0,69	0,87	0,73
8	w. mean c=1.1	0,82	0,78	0,94	0,69	0,87	0,73

По результатам Таблицы 6 видно, что наилучшим методом агрегации множества решений модели классификации для одного объекта является выбор наиболее часто встречающегося класса. При этом, позволяет улучшить итоговые значения метрик качества: для распознавания пола – до 0,89 по мере F1, для распознавания возраста – до 0,74 по мере F1, значительно улучшая показатели метрик как Recall, так и Precision. Ещё большее улучшение качества работы модели возможно, например, с помощью дальнейшего расширения данных и их аугментации, но затруднительно из-за наличия пограничных случаев: например, для класса возраста – существуют люди, по признакам силуэта и движения выглядящие как более молодые или более пожилые, чем они есть на самом деле.

Таким образом, по результатам всех экспериментов, проведённых с моделями распознавания пола и возраста, а также с методами агрегации множества результатов для одного объекта, была отобрана нейросетевая модель на базе ResNet50 с двумя выходами для одновременной классификации пола и возраста человека по GEI-изображению, построенному на основе данных трека размера 6; был выбран метод агрегации по моде. Итоговый конвейер алгоритмов распознавания пола и возраста достигает метрик качества, равных 0,89 по мере F1 для пола и 0,74 по мере F1 для возраста.

4.5 Испытания алгоритмов повторной идентификации

Выбранные направления разработки алгоритмов создания индивидуальных отпечатков объектов по графическим признакам для повторного распознавания, включающие в себя выбор модели для извлечения графических признаков из ограничивающих окон трека отслеживаемого объекта, размера трека, методов расширения количества данных, дескриптора схожести, метода агрегации выбранных из галереи кандидатов, требуют проведения ряда экспериментов для принятия обоснованных решений при разработке. Задачей проведения экспериментов является оценка качества работы разработанного алгоритма при выборе различных значений параметров входящих в

него методов и выбор наилучших из них для использования в составе модуля ядра исходной платформы.

Оценка качества в экспериментах осуществлялась по двум основным метрикам, широко используемым в задаче повторного распознавания или ре-идентификации: Cumulative Match Characteristic (CMC) и Mean Average Precision (mAP).

Метрика CMC показывает среднюю долю корректных объектов, попавших в выборку из K примеров галереи алгоритма ре-идентификации. Для одного сопоставления query и галереи рассчитывается показатель точности Acc(k) – Формула 14.

$$Acc(k) = \begin{cases} 1, & \text{если } k \text{ векторов галереи содержат } GT ID \\ 0, & \text{в других случаях} \end{cases} \quad (14)$$

где k – ранг – количество первых отобранных примеров из ранжированной галереи. Тогда метрика CMC ранга k для всей выборки является средней величиной результатов всех сопоставлений – Формула 15.

$$CMC(k) = \sum_{i=1}^P Acc(k)_i, \quad (15)$$

где P – количество примеров query в выборке. Чем больше значение параметра k, тем больше метрика CMC и выше точность алгоритма. Поэтому при решении задач ре-идентификации для оценки качества используется либо CMC со значением k, равным 1 (ранг 1 – наиболее строгий вариант метрики), либо mAP, рассчитывающийся для заданного значения k.

Метрика mAP представляет собой усреднённое для всех примеров значение метрики Average Precision (AP), рассчитанное для выборки из k примеров ранжированной галереи. AP представляет собой интегральный показатель вероятности нахождения примера с идентификатором, соответствующим идентификатору query в выборке из k примеров галереи – Формула 16.

$$AP = \int_0^1 p(r)dr, \quad (16)$$

где r – размер выборки примеров из галереи, p(r) – вероятность нахождения в выборке r примера с идентификатором, соответствующим идентификатору query. Тогда mAP рассчитывается по Формуле 17.

$$mAP = \frac{1}{n} \sum_{i=1}^n AP_i, \quad (17)$$

где AP – значение метрики Average Precision для i-го примера query из n. Визуализация расчёта метрик AP и mAP представлена на Рисунке 34.

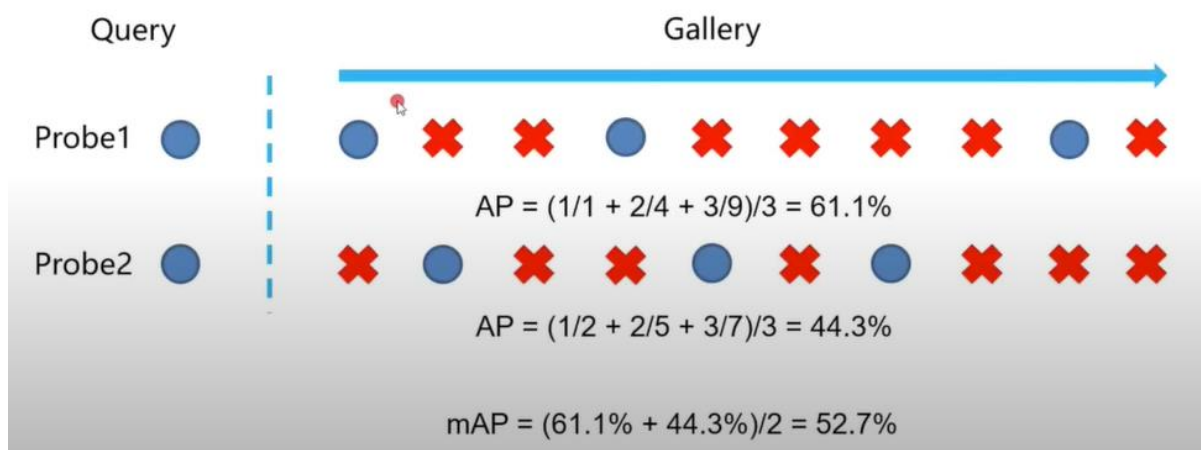


Рисунок 34 – Демонстрация расчёта метрик AP и mAP для задачи повторного распознавания объектов

Все эксперименты, проведённые с разработанным алгоритмом создания индивидуальных графических отпечатков объектов для их повторного распознавания разделены на 2 части: эксперименты с методами создания отпечатков (извлечения графических признаков) и расчёта дескрипторов схожести – эксперименты над логическим компонентом Predictor и эксперименты с методами агрегации кандидатов, отобранных из галереи по запросу (query) для принятия окончательного решения об идентификаторе объекта во время его повторного распознавания – эксперименты над логическим компонентом Aggregator.

Для проведения испытаний логического компонента Predictor использовались следующие гиперпараметры с возможными значениями:

- Модель извлечения графических признаков из ограничивающего окна объекта из библиотеки Torchreid: ResNet50, ShuffleNet, MuDeep, OSNet, MLFN;
- Размер трекета отслеживаемого объекта для расчёта усреднённого вектора признаков (максимальный размер трекета определялся объёмом доступной памяти GPU для обучения и разумным временем обучения модели): 1-6;
- Тип применяемых случайных аугментаций к изображениям ограничивающих окон трекетов во время обучения модели (аугментация применяется для расширения выборки и повышения обобщающей способности модели за счёт искусственно созданных примеров):
 1. Случайное отражение (RF);
 2. Случайное отражение, случайное кадрирование (RF + RC);
 3. Случайное отражение, случайное кадрирование, случайное изменение яркости контрастности и насыщенности (RF + RC + CJ);

4. Случайное отражение, случайное кадрирование, случайное затирание области изображения случайными цветовыми значениями (RF + RC + RE);
5. Случайное отражение, случайное кадрирование, случайное наложение друг на друга нескольких изображения (RF + RC + RP).
 - Количество эпох обучения модели извлечения графических признаков: от 30 до 80 эпох;
 - Deskриптор схожести признаков треклетов, а точнее метрика расстояния между векторами признаков:
 1. L2-норма (L2 norm) – Формула 18;
 2. Косинусное расстояние (cos) – Формула 19;
 3. Корреляция между двумя векторами (corr) – Формула 20;
 4. L1-норма (L1 norm) – Формула 21;
 5. Расстояние Чебышева (Chebyshev) – Формула 22.

Формулы описанных дескрипторов схожести представлены в таблице 7, где u и v – векторы.

Таблица 7 – Формулы дескрипторов схожести, выбранных для проведения экспериментов

Дескриптор схожести	Формула	Номер
L2 norm	$d(u,v) = \sqrt{\sum_i^k (u_i - v_i)^2}$	(18)
cos	$1 - \frac{(u) \cdot (v)}{\ u\ _2 \ v\ _2}$	(19)
corr	$1 - \frac{(u - \bar{u}) \cdot (v - \bar{v})}{\ (u - \bar{u})\ _2 \ (v - \bar{v})\ _2}$	(20)
L1 norm	$d(u,v) = \sum_i^k u_i - v_i $	(21)
Chebyshev	$d(u,v) = \max(i) u_i - v_i $	(22)

Для проведения испытаний компонента Aggregator использовались следующие гиперпараметры с возможными значениями:

- Метод расчёта метрик для агрегации кандидатов, отобранных как k первых примеров из ранжированной относительно query галереи:
 1. Среднее значение дескрипторов схожести для кандидатов, имеющих определённый идентификатор. Чем меньше рассчитанное значение для

определённого идентификатора, тем более привлекательным для принятия окончательного решения является рассматриваемый идентификатор (mean) – Формула 23.

2. Мода идентификаторов среди отобранных кандидатов. В качестве окончательного решения выбирается наиболее часто встречающийся идентификатор (mode) – Формула 24.

3. Среднее взвешенное значение дескрипторов схожести. Взвешивание среднего значения осуществляется на величину, равную отношению кандидатов с искомым идентификатором, к общему количеству кандидатов (weighted mean) – Формула 25.

Формулы описанных метрик представлены в Таблице 8, где S^{id} – значение метрики для агрегации (чем она меньше для кандидата с определённым id, тем более подходящим является этот кандидат), d_i – значение дескриптора схожести i -го кандидата с query, имеющим идентификатор id , N – количество кандидатов, N^{id} – количество кандидатов с идентификатором, совпадающим с идентификатором id , принадлежащим query, c – регуляризовочный гиперпараметр.

Таблица 8 – Формулы расчёта метрик для агрегации кандидатов, выбранные для проведения эксперимента

Метрика для агрегации	Формула	Номер
mean	$S^{id} = \frac{1}{N} \sum_{i=1}^N d_i^{id}$	(23)
mode	$S^{id} = Mo(N)$	(24)
weighted mean	$S^{id} = \frac{\frac{1}{N} \sum_{i=1}^N d_i^{id}}{\left(\frac{N^{id}}{N}\right)^c}$	(25)

- Размер треклета отслеживаемого объекта: 1-9;
- Количество кандидатов для метода агрегации: 1-4.

Эксперименты с компонентом Predictor в силу длительности их выполнения (в рамках данных экспериментов обучаются модели глубоких нейронных сетей) осуществлялись независимо – каждый гиперпараметр настраивался отдельно от других – значения остальных гиперпараметров оставались неизменными. Эксперименты с компонентом Aggregator проводились с использованием отобранных значений параметров компонента Predictor путём полного перебора возможных значений гиперпараметров.

Далее описываются результаты экспериментов над логическим компонентом Predictor. Для каждого эксперимента использовался набор значений по умолчанию, среди которых менялись только значения оцениваемых параметров:

- Модель извлечения графических признаков из ограничивающих окон объектов: MLFN;
- Модель использует предварительно обученные на наборе данных ImageNet веса;
- Размер треклета отслеживаемого объекта: 6;
- Типы применяемых случайных аугментаций данных во время обучения: RF + RC;
- Количество эпох обучения модели: 30;
- Дескриптор схожести признаков треклетов: косинусное расстояние;
- Агрегация кандидатов не осуществлялась (компонент Aggregator исключён из экспериментов).

Результаты экспериментов по отбору модели извлечения графических признаков представлены в Таблицах 9-10. Для получения результатов в Таблице 10 были взяты 2 лучшие относительно метрик качества модели из Таблицы 11.

Таблица 9 – Результаты экспериментов по выбору модели извлечения графических признаков (треклет размера 1)

Модель	CMC Rank-(1-10)										AP
	1	2	3	4	5	6	7	8	9	10	
ResNet50	0.77	0.83	0.86	0.89	0.90	0.90	0.91	0.92	0.93	0.94	0.56
ShuffleNet	0.54	0.64	0.71	0.76	0.78	0.80	0.82	0.83	0.85	0.86	0.34
MuDeep	0.73	0.80	0.84	0.86	0.88	0.89	0.90	0.90	0.91	0.92	0.46
OSNet	0.65	0.74	0.79	0.82	0.84	0.85	0.87	0.88	0.89	0.90	0.40
MLFN	0.82	0.88	0.90	0.92	0.93	0.94	0.95	0.95	0.95	0.96	0.63

Таблица 10 – Результаты экспериментов по выбору модели извлечения графических признаков (треклет размера 6)

Модель	CMC Rank-(1-10)										AP
	1	2	3	4	5	6	7	8	9	10	
ResNet50	0.44	0.53	0.60	0.70	0.78	0.78	0.83	0.83	0.86	0.86	0.32

MLFN	0.54	0.71	0.82	0.86	0.86	0.89	0.89	0.89	0.93	0.93	0.50
------	------	------	------	------	------	------	------	------	------	------	------

Результаты экспериментов по выбору модели показывают, что наилучшей глубокой нейронной сетью, извлекающей графические признаки из ограничивающих окон объектов для повторного распознавания, является MLFN. Причём, эта модель является лучшей как по метрике СМС любого ранга, так и по метрике mAP. Таким образом, результатом отбора модели является MLFN.

Результаты экспериментов по отбору размера треклета, из ограничивающих окон которого производится извлечение графических признаков, представлены в Таблице 12. Результаты в Таблице 11 приведены только для минимального и максимального размера треклетов, где максимальный размер выбирался исходя из объёма доступной памяти GPU и времени обучения моделей.

Таблица 11 – Результаты экспериментов по отбору размера треклета

Размер треклета	CMC Rank-(1-10)										AP
	1	2	3	4	5	6	7	8	9	10	
1	0.44	0.53	0.59	0.64	0.66	0.68	0.70	0.72	0.75	0.76	0.32
6	0.68	0.68	0.82	0.84	0.86	0.87	0.89	0.89	0.90	0.90	0.58

По результатам эксперимента, представленным в Таблице 11, видно, что чем больше размер треклета, тем выше итоговое качество решение задачи повторного распознавания как по метрике СМС, так и по метрике mAP. Дальнейшее увеличение размера треклета значительно увеличивает потребление памяти GPU и время обучения модели, поэтому в качестве значения размера треклета для текущего этапа работ было выбрано значение 6.

Результаты экспериментов по методам аугментации данных для улучшения обобщающих свойств модели представлены в Таблице 12.

Таблица 12 – Результаты экспериментов по методам аугментации данных, применяемых во время обучения модели

Тип аугментации	CMC Rank-(1-10)										AP
	1	2	3	4	5	6	7	8	9	10	
RF+RC	0.53	0.71	0.82	0.85	0.86	0.86	0.89	0.89	0.89	0.93	0.50
RF+RC+CJ	0.50	0.68	0.76	0.82	0.89	0.89	0.89	0.89	0.89	0.89	0.49
RF+RC+RE	0.50	0.68	0.71	0.75	0.89	0.93	0.93	0.96	0.96	0.96	0.47

RF+RC+RP	0.39	0.64	0.71	0.82	0.86	0.86	0.86	0.89	0.89	0.92	0.46
RF	0.42	0.43	0.61	0.71	0.79	0.82	0.86	0.86	0.86	0.90	0.46

Как видно из полученных результатов, методы аугментации, заключающиеся в случайном отражении и случайном кадрировании, позволяют улучшить метрики качества решения задачи повторного распознавания – это видно как по метрике СМС, так и по метрике mAP. Слишком сильные изменения данных, как видно из экспериментов для методов RF+RC+CJ, RF+RC+RE, RF+RC+RP могут вредить качеству. Таким образом, в качестве методов аугментации, применяющихся во время обучения модели, выбраны методы RF и RC.

Для выбора количества эпох для обучения модели было отобрано несколько диапазонов, результаты расчёта метрик для которых приведены в Таблице 13.

Таблица 13 – Результаты экспериментов по обучению модели для извлечения графических признаков на различном количестве эпох

Кол-во эпох	СМС Rank-(1-10)										AP
	1	2	3	4	5	6	7	8	9	10	
30	0.54	0.70	0.74	0.76	0.83	0.77	0.83	0.88	0.88	0.91	0.35
60	0.67	0.79	0.81	0.86	0.86	0.88	0.90	0.93	0.95	0.98	0.41
80	0.67	0.83	0.86	0.88	0.91	0.91	0.91	0.93	0.93	0.95	0.40

Из результатов, отображённых в Таблице 14, видно, что более подходящим является количество эпох, равное 60. При его увеличении до 80, по метрике mAP видно небольшое падение значения метрики mAP, что говорит о переобучении. Таким образом, для обучения модели извлечения графических признаков из ограничивающих окон объектов выбрано количество эпох, равное 60.

Результаты экспериментов, связанных с выбором дескриптора схожести, на основании значения которого осуществляется ранжирование примеров в галерее при решении задачи повторного распознавания объектов, представлены в Таблице 14.

Таблица 14 – Результаты экспериментов по выбору дескриптора схожести

Дескриптор схожести	СМС Rank-1										AP
	1	2	3	4	5	6	7	8	9	10	
L2 norm	0.65	0.79	0.86	0.86	0.86	0.88	0.88	0.93	0.95	0.95	0.40

cos	0.67	0.79	0.81	0.86	0.86	0.88	0.91	0.93	0.95	0.98	0.41
corr	0.56	0.79	0.85	0.85	0.96	0.88	0.88	0.91	0.95	0.95	0.40
L1 norm	0.65	0.79	0.86	0.86	0.88	0.88	0.91	0.95	0.95	0.95	0.40
Chebyshev	0.35	0.51	0.59	0.65	0.67	0.70	0.70	0.74	0.74	0.77	0.22

По результатам сравнения дескрипторов схожести видно, что наилучшими показателями качества обладает косинусное расстояние, которое также наиболее часто используют исследователи в работах, посвящённых решению задачи ре-идентификации. Таким образом, косинусное расстояние остаётся основным используемым дескриптором схожести в разработанном алгоритме создания индивидуальных отпечатков объектов по графическим признакам.

По результатам всех экспериментов, проведённых с компонентом Predictor разработанного алгоритма создания графических отпечатков объектов для их повторного распознавания, были отобраны наилучшие относительно метрик качества СМС-1 и mAP значения гиперпараметров: модель MLFN из библиотеки Torchreid для извлечения графических признаков из ограничивающих окон объектов; размер треклета, равный 6; использование аугментаций вида случайное отражение и случайное кадрирование изображений ограничивающих окон во время обучения модели; количество эпох обучения, равное 60; косинусное расстояние в качестве дескриптора схожести графических признаков треклетов. Конвейер разработанных алгоритмов повторного распознавания с отобранными значениями гиперпараметров без использования агрегации кандидатов, извлечённых из галереи (при использовании только компонента Predictor – без компонента Aggregator) достигает значений метрик качества СМС-1 и mAP, равных 0,8 и 0,57 соответственно.

Эксперименты с методами и гиперпараметрами логического модуля Aggregator разработанного алгоритма создания индивидуальных отпечатков объектов по графическим признакам для повторного распознавания разделены на два отдельных блока: блок выбора значения регуляровочного параметра с метрики агрегации weighted mean и блок одновременного выбора метрики агрегации, длины треклета и количества кандидатов.

Результаты экспериментов с выбором значения регуляровочного параметра с метрики weighted mean представлены на Рисунках 35-38.

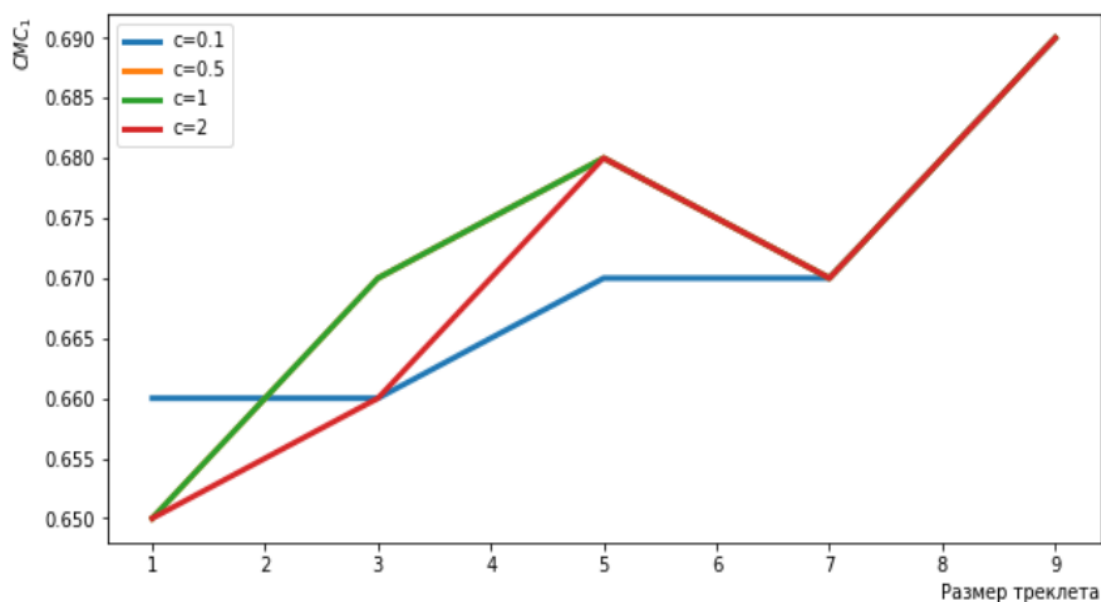


Рисунок 35. Зависимость значения метрики СМС-1 от размера треклета при различных значениях регуляровочного параметра с метрики агрегации weighted mean и количестве кандидатов, равном 1

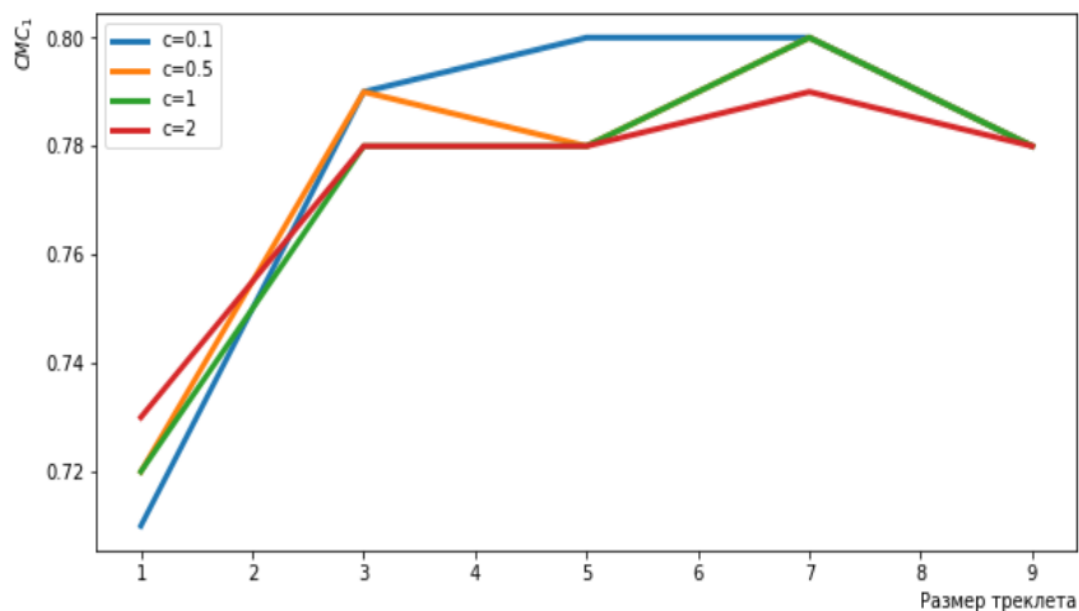


Рисунок 36. Зависимость значения метрики СМС-1 от размера треклета при различных значениях регуляровочного параметра с метрики агрегации weighted mean и количестве кандидатов, равном 2

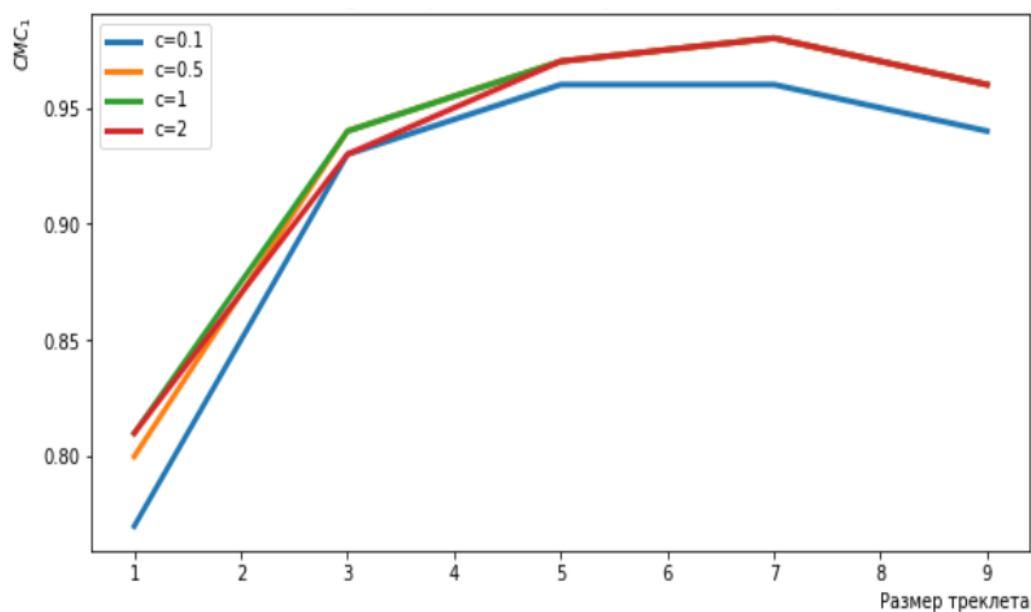


Рисунок 37. Зависимость значения метрики СМС-1 от размера треклета при различных значениях регуляровочного параметра с метрики агрегации weighted mean и количестве кандидатов, равном 3

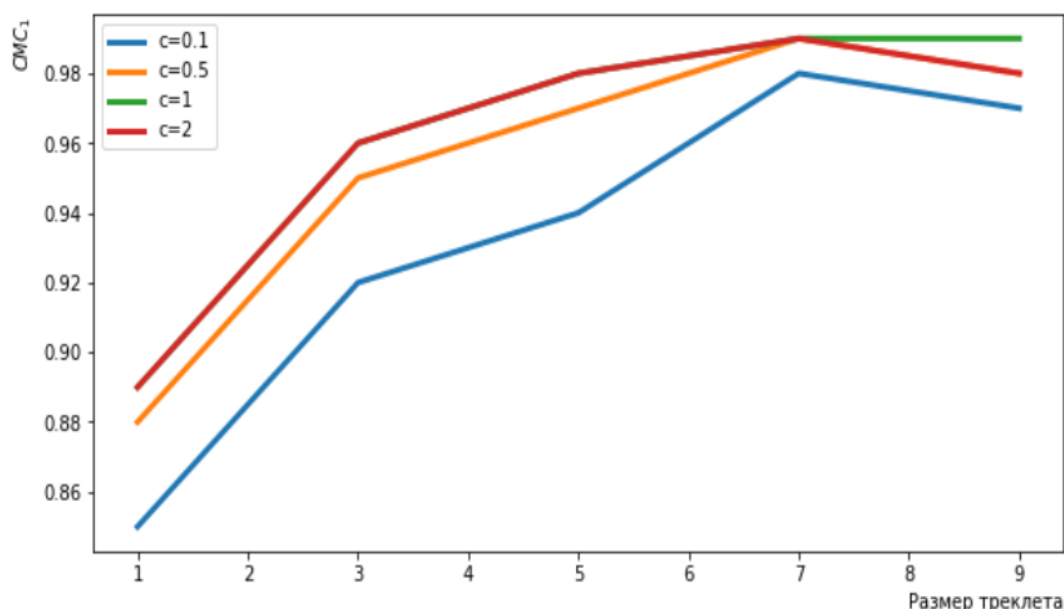


Рисунок 38 – Зависимость значения метрики СМС-1 от размера треклета при различных значениях регуляровочного параметра с метрики агрегации weighted mean и количестве кандидатов, равном 4

На Рисунках 35-38 видно, что при увеличении размера треклета и количества кандидатов, по которым производится агрегация, качество по метрике СМС-1 растёт, при этом, наилучшего качества во всех случаях позволяет достичь метрика агрегации weighted mean при значении регуляровочного параметра c , равном 1. Таким образом, по результатам экспериментов в качестве значения параметра c метрики агрегации weighted mean выбрано значение 1.

Эксперименты с отбором наилучших значений параметров размера треклета, количества кандидатов для агрегации и методов расчёта метрик для агрегации были

проведены с учётом настройки параметра с метода weighted sum, и их результаты приведены на Рисунках 39-42.

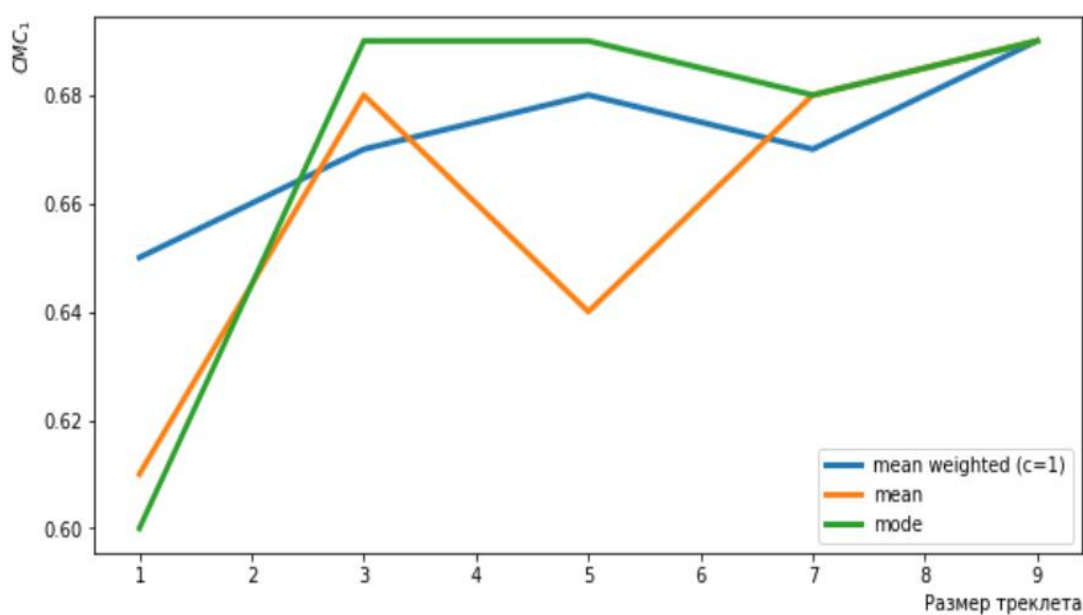


Рисунок 39. Зависимость значения метрики CMC-1 от размера треклета при использовании различных метрик для агрегации и количестве кандидатов, равном 1

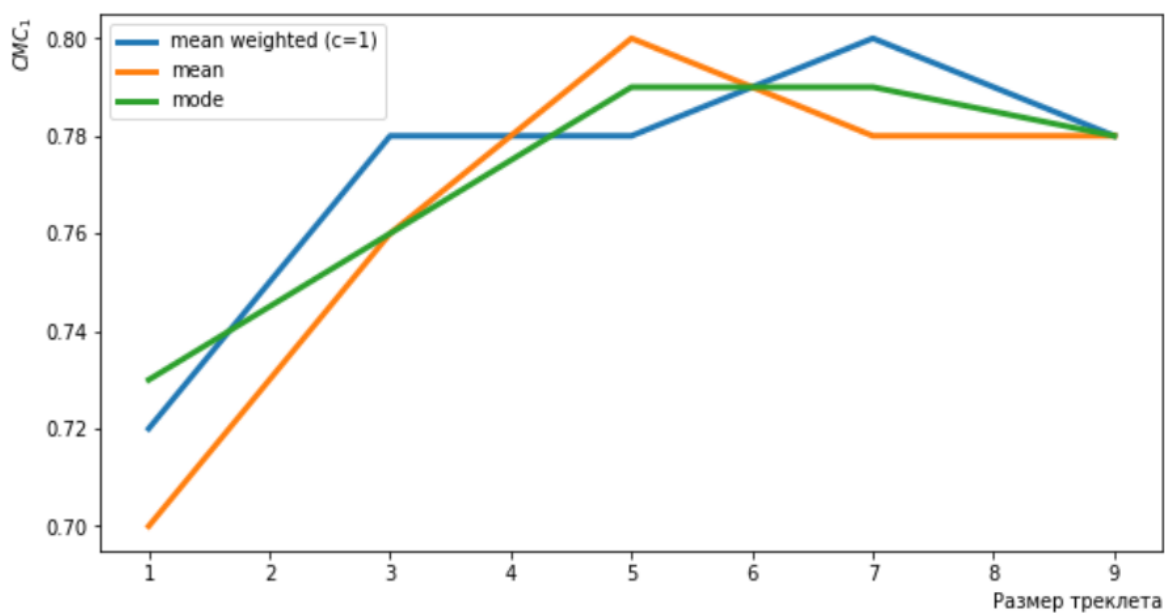


Рисунок 40. Зависимость значения метрики CMC-1 от размера треклета при использовании различных метрик для агрегации и количестве кандидатов, равном 2

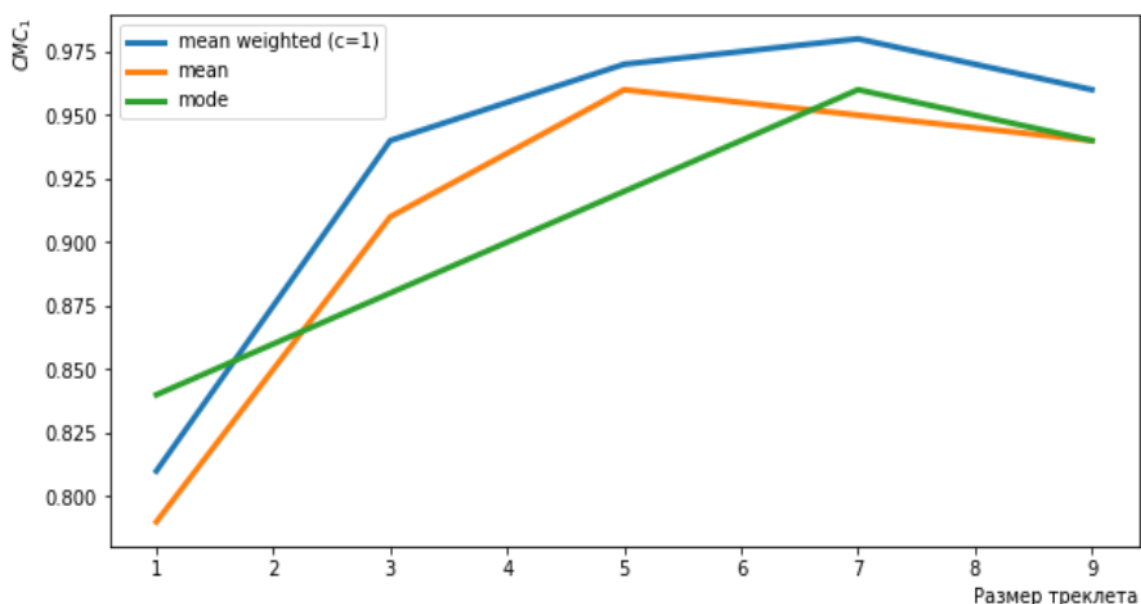


Рисунок 41. Зависимость значения метрики СМС-1 от размера треклета при использовании различных метрик для агрегации и количестве кандидатов, равном 3

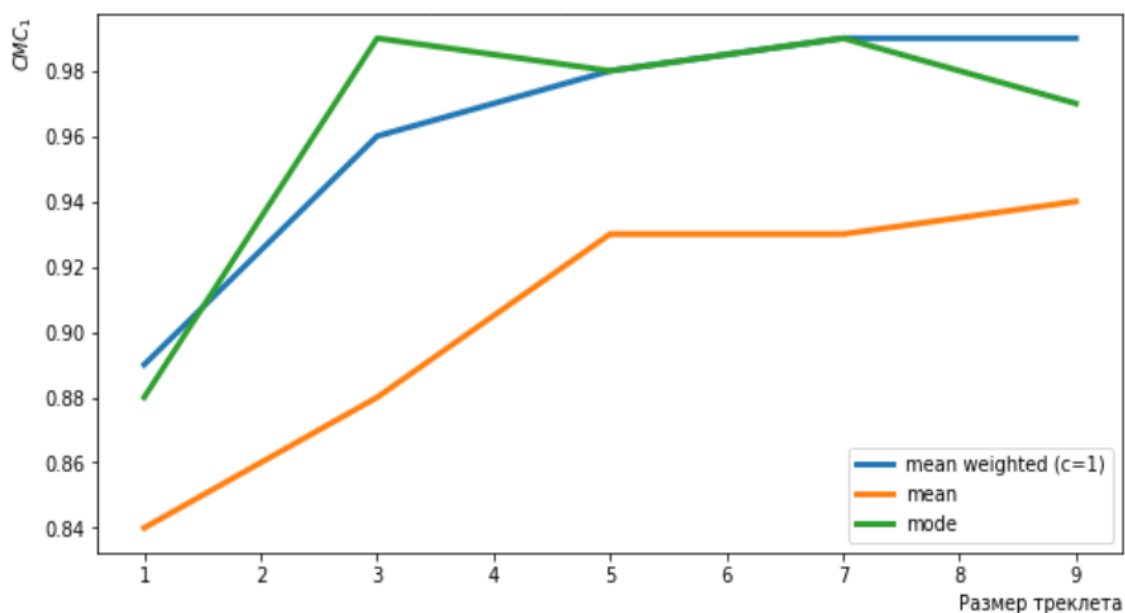


Рисунок 42. Зависимость значения метрики СМС-1 от размера треклета при использовании различных метрик для агрегации и количестве кандидатов, равном 4

Исходя из данных, представленных на Рисунках 39-42, можно заключить, что метод weighted mean показывает наилучшие значения метрики качества СМС-1, что особенно заметно при увеличении как размера треклета, так и количества кандидатов для агрегации. Оптимальный размер треклета равен 7, однако, уменьшение значений метрики качества на больших размерах может объясняться размером треклета, по которому строятся графические признаки, отобранному при проведении экспериментов над компонентом Predictor. Также видно, что качество тем выше, чем выше количество кандидатов, однако, при работе алгоритма в составе модуля ядра исходной платформы, количество кандидатов не может быть слишком большим из-за ограничений по памяти и скорости работы –

отобранное по результатам текущих экспериментов значение 4 может быть использовано и для работы в составе ядра платформы.

По результатам всех экспериментов, проведённых с компонентом Aggregator разработанного алгоритма создания графических отпечатков объектов для их повторного распознавания, были отобраны наилучшие относительно метрики качества СМС-1 значения гиперпараметров: метрика метода агрегации weighted mean со значением регулировочного параметра c , равным 1; максимальная длина треклета, равная 7; минимальное количество кандидатов для агрегации – 4. Конвейер разработанных алгоритмов повторного распознавания с отобранными значениями гиперпараметров достигает значения метрики качества СМС-1, равного 0,98.

4.6. Испытания трекера с использованием алгоритма повторного распознавания

Разработанный алгоритм создания индивидуальных отпечатков объектов по графическим признакам для повторного распознавания, внедрённый на текущем этапе в работу модуля ядра платформы детектирования и трекинга, позволяет не только решить задачу распознавания объектов, которые посещают сцену видеопотока повторно, но и, благодаря особенностям своей работы, улучшить работу алгоритма трекинга множества объектов за счёт возможности исправления ошибок трекера. Данный эксперимент предполагает оценку качества работы алгоритма трекинга с учётом работы внедрённого в модуль ядра разработанного алгоритма повторного распознавания объектов по их индивидуальным графическим отпечаткам.

Внедрение алгоритмов распознавания пола и возраста людей, а также создания графических отпечатков для повторного распознавания объектов позволило расширить алгоритм трекинга, функционирующий на уровне модуля ядра платформы, увеличив количество этапов его работы и добавив новые методы. Сам по себе алгоритм трекинга является гибко конфигурируемым и включает в себя несколько этапов: расчёт графических признаков для ограничивающих окон объектов, предсказание движения объекта, каскад ассоциаций для сопоставления объектов на кадре видеопотока, а также новые методы, добавленные с внедрением новых алгоритмов: извлечение признаков из треклета объекта, определение пола и возраста людей по треклету и всей истории существования объекта, повторное определение идентификатора объекта по треклету и всей истории существования объекта.

Для проведения экспериментов по оценке качества алгоритмов трекинга с учётом внедрённых в работу ядра платформы разработанных алгоритмов создания

индивидуальных графических отпечатков объектов для повторного распознавания, были созданы специальные конфигурации, представленные в Таблице 15.

Таблица 15 – Конфигурации экспериментов для оценки качества алгоритма трекинга

Настройка		Эксперимент 0	Эксперимент 1	Эксперимент 2	Эксперимент 3
метод расчёта графических признаков ограничивающих окон		Извлечение векторов графических признаков из изображений ограничивающих окон с помощью MLFN	Извлечение векторов графических признаков из изображений ограничивающих окон с помощью MLFN	Извлечение векторов графических признаков из изображений ограничивающих окон с помощью MLFN	Извлечение векторов графических признаков из изображений ограничивающих окон с помощью MLFN
ассоциация	фильтрация объектов	для всех объектов	для всех объектов	для всех объектов	для всех объектов
	приоритет методов продолжения треков	расчёт по вектору скорости с задержкой в 4 кадра, последнее ограничивающее окно объекта	расчёт по вектору скорости с задержкой в 4 кадра, последнее ограничивающее окно объекта	расчёт по вектору скорости с задержкой в 4 кадра, последнее ограничивающее окно объекта	расчёт по вектору скорости с задержкой в 4 кадра, последнее ограничивающее окно объекта
	взвешенная сумма метрик схожести	IOU с коэффициентом 0.5 ARC с коэффициентом 0.5	IOU с коэффициентом 0.5 AR С с коэффициентом 0.5	IOU с коэффициентом 0.5 AR С с коэффициентом 0.5	IOU с коэффициентом 0.5 AR С с коэффициентом 0.5
	ограничения	Сумма метрик ≥ 0.1	Сумма метрик ≥ 0.1	Сумма метрик ≥ 0.1	Сумма метрик ≥ 0.1
приоритет методов продолжения треков неассоциированных объектов		расчёт по вектору скорости с задержкой в 4 кадра, последнее ограничивающее окно объекта	расчёт по вектору скорости с задержкой в 4 кадра, последнее ограничивающее окно объекта	расчёт по вектору скорости с задержкой в 4 кадра, последнее ограничивающее окно объекта	расчёт по вектору скорости с задержкой в 4 кадра, последнее ограничивающее окно объекта
извлечение признаков из треклета объекта		усреднение графических признаков ограничивающих окон	усреднение графических признаков ограничивающих окон	усреднение графических признаков ограничивающих окон	усреднение графических признаков ограничивающих окон

повторное распознавание объектов	расчёт дескрипторов схожести между векторами графических признаков	нет	расчёт косинусного расстояния, ограничение ≤ 0.8	расчёт косинусного расстояния, ограничение ≤ 0.8	расчёт косинусного расстояния, ограничение ≤ 0.8
	агрегация отобранных кандидатов		взвешенное голосование	взвешенное голосование, ограничение ≤ 0.5	взвешенное среднее, ограничение ≤ 0.25

Для оценки качества работы алгоритма трекинга используется метрика MOTA, рассчитываемая с помощью специальной утилиты, работающей на подготовленных для оценки качества работы алгоритма трекинга данных. Утилита была создана при первичной разработке платформы детектирования и трекинга. Результаты оценки метрик качества работы алгоритма трекинга представлены в Таблице 16.

Таблица 16 – Результаты расчёта метрики MOTA на данных для оценки качества работы алгоритма трекинга

Набор данных	Детекции	Эксперимент т 0	Эксперимент т 1	Эксперимент т 2	Эксперимент т 3
MOT17	DPM	0,31495	0,30707	0,30427	0,30427
PETS2009	Детекции	0,13518	0,13693	0,13720	0,13609
Среднее		0,11253	0,11100	0,11037	0,11009

По представленным результатам можно заключить, что добавление методов повторного распознавания объектов оказывает как положительное, так и отрицательное влияние на метрики качества алгоритма трекинга. Положительное влияние отмечается при использовании набора данных PETS2009 – на всех экспериментах качество работы алгоритма трекинга возросло, на эксперименте 2 качество улучшилось на 0.2%. Отрицательное влияние отмечается при использовании данных из набора MOT17 – на всех экспериментах качество упало, на экспериментах 2 и 3 качество ухудшилось на 1%. Полученные результаты сильно зависят от особенностей данных. Как видно на рисунке 35 набор данных PETS2009 содержит видеопоследовательности движения людей в условиях хорошей видимости и редких случаях перекрытия одних объектов другими, набор данных MOT17 включает частое перекрытие объектов, что в случае работы методов повторного распознавания приводит к добавлению в галерею признаков некачественных изображений (небольших изображений объекта издалека, изображений с перекрытым объектом и пр.) ограничивающих окон и ошибок в распознавании.

Глава 5 Финансовый менеджмент, ресурсоэффективность и ресурсосбережение

5.1 Предпроектный анализ

5.1.1 Потенциальные потребители результатов исследования

Разработанные алгоритмы классификации пола и возраста людей и их повторного распознавания по признакам силуэта и движения планируется использовать для подсчёта пешеходного трафика в общих городских условиях и интеллектуальной аналитики полученных данных. Также такая аналитика может быть использована для изучения данных, собранных в ходе охранного видеонаблюдения, мониторинга дорожного трафика и работы систем домашней автоматизации.

Заинтересованными в продукте могут быть как крупные компании, имеющие большое количество локаций, требующих мониторинга и детального анализа трафика людей, так и мелкие компании, выбирающие новое место размещения точки розничной торговли.

Сегментировать рынок услуг можно по размеру компании-заказчика и сфере применения разработки. Результат сегментирования представлен в Таблице 17.

Таблица 17 – Карта сегментирования рынка

		Сфера применения разработки			
		Ритейл	Системы охранного видеонаблюдения	Системы мониторинга дорожного трафика	Системы домашней автоматизации
Размер компании	Крупные				
	Средние				
	Мелкие				
	Фирма А		Фирма Б		Фирма В

Таким образом, основными заказчиками, на которые ориентирован программный продукт, являются средние и мелкие компании, имеющие потребность в подсчёте переходного трафика для задач ритейла и в установке охранного видеонаблюдения. Также в будущем можно ориентировать продукт на решение задач мониторинга дорожного трафика для средних компаний.

5.1.2 Анализ конкурентных решений

В настоящий момент на рынке существуют несколько компаний-конкурентов, занимающиеся созданием решений использования технологий машинного обучения и компьютерного зрения для подсчёта трафика и получения детальной аналитики. Среди них рассмотрены следующие решения:

1. Orwell 2k – система видеонаблюдения с на основе компьютерного зрения, предназначенная для автоматического обнаружения и классификации целей и ситуаций и передачи в реальном времени видеоинформации оператору;
2. «Интеллект» – программная платформа, предназначенная для создания комплексных систем безопасности;
3. Macroscop Basic – интеллектуальный модуль распознавания лиц в видеопотоке;

Также были выделены следующие критерии оценки продуктов:

1. Технические критерии оценки ресурсоэффективности
 - a. облачный сервис, доступный для быстрого подключения;
 - b. возможно индивидуализации задач;
 - c. автоматизация детектирования объектов;
 - d. автоматизация трекинга объектов;
 - e. распознавание лиц;
 - f. определение пола;
 - g. определение возраста;
 - h. повторное распознавание объектов.
2. Экономические критерии оценки эффективности:
 - a. обслуживание;
 - b. цена.

Экспертная оценка основных технических характеристик данных продуктов представлена в Таблице 18.

Таблица 18 – Оценочная карта сравнения технических решений

№	Критерии оценки	Вес	Баллы				Конкурентоспособность			
			Бф	Бк1	Бк2	Бк3	Кф	К1	К2	К3
	1	2	3	4	5	6	7	8	9	10
Технические критерии оценки ресурсоэффективности										
1	облачный сервис, доступный для быстрого подключения	0,14	5	1	1	5	0,7	0,14	0,14	0,7
2	возможность индивидуализации решаемых задач	0,14	3	5	2	2	0,42	0,7	0,28	0,28
3	автоматизация детектирования объектов	0,15	5	5	4	5	0,7	0,7	0,56	0,7
4	автоматизация трекинга объектов	0,14	5	2	2	1	0,7	0,28	0,28	0,14
5	распознавание лиц	0,04	1	3	2	5	0,04	0,12	0,08	0,2
6	определение пола	0,06	5	3	1	1	0,3	0,18	0,06	0,06
7	определение возраста	0,06	5	3	1	1	0,3	0,18	0,06	0,06
8	повторное распознавание объектов	0,06	5	3	1	5	0,3	0,18	0,06	0,3
Экономические критерии оценки эффективности										
9	обслуживание	0,06	4	2	3	4	0,24	0,12	0,18	0,24
10	цена	0,15	4	3	4	3	0,6	0,45	0,6	0,45
	Итого	1	42	30	21	32	4,3	3,05	2,3	3,13

Исходя из данных, приведённых в Таблице 18, продукты компаний-конкурентов в основной массе не предоставляют такую детальную видеоаналитику, как определение пола, возраста, а также повторное распознавание объектов.

Кроме того, особенностью работы большинства компаний-конкурентов является обязательное условие по размещению собственных систем видеонаблюдения с модулями видеоаналитики на пользовательской стороне. Закрытое решение поставляется по составленному индивидуальному проекту, разработка и поставка которого занимает более нескольких месяцев.

Таким образом, преимуществами разрабатываемого продукта будут возможность использования клиентами широкого спектра областей (ритейл, безопасность, “умный город”) без обязательной интеграции с инфраструктурой заказчика, а также более детальная видеоаналитика.

5.1.3 SWOT-анализ

SWOT-анализ – метод планирования, который заключается в выявлении внутренних и внешних факторов среды объекта планирования, которые разделяются на следующие категории:

- Strength (сильные стороны);
- Weaknesses (слабые стороны);
- Opportunities (возможности);
- Threats (угрозы).

Для исследуемой разработки было проведено три этапа SWOT-анализа. Результаты первого этапа приведены в Таблице 19.

Таблица 19 – Результаты первого этапа SWOT-анализа

	Сильные стороны научно-исследовательского проекта: Си1.Хорошее качество видеоаналитики (0,74-0,89 F1 для классификации пола и возраста; 0,98 СМС-1 для повторного распознавания) Си2.Развёртывание системы в облаке - необходимость модификаций под отдельного пользователя Си3.Гибкая архитектура разработанных программных средств Си4.Поддержка государством разработки внутренних ИТ-продуктов Си5.Квалифицированная поддержка продукта	Слабые стороны научно-исследовательского проекта: Сл1.Сильная зависимости качества видеоаналитики от качества данных пользователя Сл2.Нестабильность работы на отдельных облачных решениях Сл3.Отсутствие большого количества собственных размеченных данных для моделей видеоаналитики Сл4.Сильная зависимость работоспособности продукта от внешних облачных решений Сл5.Необходимость закупки хороших видеокамер
Возможности: В1.Доработка функционала (детектирование лиц) В2.Улучшение качества классификации пола и возраста и повторного распознавания людей В3.Улучшение производительности для работы в реальном времени В4.Появление дополнительного спроса на программный продукт В5.Повышение стоимости подобных продуктов		
Угрозы: У1.Подорожание использования облачных решений У2.Появление более качественных аналогов У3.Появление более дешёвых аналогов У4.Падение спроса на интеллектуальную видеоаналитику У5.Введение дополнительных государственных требований к сертификации продукции		

В рамках второго этапа на основе предыдущих результатов были выявлены соответствия сильных и слабых сторон проекта внешним условиям окружающей среды. Интерактивная матрица проекта представлена в Таблице 20.

Таблица 20 – Интерактивная матрица проекта

Сильные стороны проекта						
Возможности		Си1	Си2	Си3	Си4	Си5
	B1	-	-	+	-	-
	B2	+	-	+	-	-
	B3	-	+	-	-	-
	B4	-	+	-	+	-
	B5	-	-	-	-	+
Слабые стороны проекта						
Возможности		Сл1	Сл2	Сл3	Сл4	Сл5
	B1	+	-	+	-	-
	B2	+	-	+	-	-
	B3	-	+	-	-	-
	B4	-	-	-	+	-
	B5	-	-	-	+	-
Сильные стороны проекта						
Угрозы		Си1	Си2	Си3	Си4	Си5
	У1	-	-	+	-	-
	У2	+	+	+	-	-
	У3	+	+	-	-	-
	У4	-	-	+	-	+
	У5	-	-	+	-	-
Слабые стороны проекта						
Угрозы		Сл1	Сл2	Сл3	Сл4	Сл5
	У1	-	+	-	-	-
	У2	-	-	+	-	-
	У3	+	-	-	-	+
	У4	-	-	-	-	-
	У5	-	-	-	+	-

Составленная в ходе выполнения третьего этапа SWOT-анализа итоговая матрица представлена в Таблице 21.

Таблица 21 – Итоговая матрица SWOT-анализа

	Сильные стороны научно-исследовательского проекта: С1.Хорошее качество видеоаналитики (0,74-0,89 F1 для классификации пола и возраста; 0,98 СМС-1 для повторного распознавания) С2.Развёртывание системы в облаке - необходимость модификаций под отдельного пользователя С3.Гибкая архитектура разработанных программных средств С4.Поддержка государством разработки внутренних IT-продуктов С5.Квалифицированная поддержка продукта	Слабые стороны научно-исследовательского проекта: Сл1.Сильная зависимости качества видеоаналитики от качества данных пользователя Сл2.Нестабильность работы на отдельных облачных решениях Сл3.Отсутствие большого количества собственных размеченных данных для моделей видеоаналитики Сл4.Сильная зависимость работоспособности продукта от внешних облачных решений Сл5.Необходимость закупки хороших видеокамер
Возможности: В1.Доработка функционала (детектирование лиц) В2.Улучшение качества классификации пола и возраста и повторного распознавания людей В3.Улучшение производительности для работы в реальном времени В4.Появление дополнительного спроса на программный продукт В5.Повышение стоимости подобных продуктов	В1Си3.Гибкая архитектура позволит сократить расходы на внедрение дополнительных функций интеллектуальной видеоаналитики В2Си1.Исследование зависимости качества работы алгоритмов от условий поможет оценить направление дальнейшей разработки В3Си2. Использование облачного подхода оставляет много возможностей для изменения конфигураций используемых машин В3Си3.Гибкая архитектура позволит сократить расходы на внедрение модификаций В4Си4.Поддержка государством внутренних продуктов может привести к повышению доверия к инновациям в обществе В5Си5.Повышение стоимости продукта позволит нанять более квалифицированный персонал	В1Сл3.Потребуется большие траты на добычу и разметку данных для обучения моделей В2Сл1.Появится возможность улучшить качество работы в произвольных условиях В2Сл3.Потребуется большие траты на добычу и разметку данных для дополнительного обучения моделей В3Сл2.Сложность в рассмотрении всех проблем, возникающих при работе платформы в облаках, из-за их частой недетерминированности В4Сл4.Появление дополнительного спроса может способствовать развёртыванию системы на альтернативных облачных решениях В5Сл4.Повышение стоимости продукта позволит уменьшить зависимость от конкретных облачных решений

Угрозы: У1.Подорожание использования облачных решений У2.Появление более качественных аналогов У3.Появление более дешёвых аналогов У4.Отсутствие спроса на интеллектуальную видеоаналитику У5.Введение дополнительных государственных требований к сертификации продукции	У1Си3.Гибкая архитектура позволит сократить расходы на внедрение модификаций, направленных на устранение влияния проблем облачных решений У2Си1.Высокая скорость работы алгоритма, напрямую влияющая на ценообразование, может помочь в конкурентной борьбе У2Си2.Универсальный облачный подход может помочь в конкурентной борьбе У2Си3.Гибкая архитектура позволит сократить расходы на внедрение дополнительных функций и модификаций У3Си1.Высокое качество работы алгоритмов может помочь в конкурентной борьбе У3Си2.Универсальный подход, удобный для всех пользователей, может помочь в конкурентной борьбе У4Си3.Гибкая архитектура позволит модифицировать продукт под потребности заказчиков У4Си5.Поддержка продукта поможет оперативно выявить новые потребности У5Си3.Гибкая архитектура позволит модифицировать продукт под новые стандарты	У1Сл2.Сложность в рассмотрении всех проблем, возникающих при работе платформы в облаках, из-за их частой недетерминированности У2Сл3.Потребуются большие траты на добычу и разметку данных для обучения моделей при необходимости повышения качества работы разработки У3Сл1.Недостатки алгоритма, проявляющиеся при работе в определённых условиях, могут быть нивелированы в конкурентных решениях У3Сл5.Необходимость закупки видеокамер препятствует снижению цены на продукт вслед за аналогами У5Сл4.Конкретные облачные сервисы могут не подходить под новые стандарты
---	--	---

Самым большим преимуществом данной разработки является её гибкая архитектура, а недостатком – сильная зависимость от предоставляемых пользователем данных для корректной работы и работы облачных сервисов.

5.1.4 Цели и результаты проекта

Информация о заинтересованных сторонах проекта представлена в Таблице 22.

Таблица 22 – Заинтересованные стороны проекта

Заинтересованные стороны	Ожидания
Компания-пользователь	Облегчение процесса анализа трафика и/или обеспечения безопасности
Разработчик	Получение заработной платы
Компания разработчика	Получение прибыли с полноценного комплекса продукта
Научный руководитель, студент	Выполненная выпускная квалификационная работа

Цели и результат проекта представлены в Таблице 23.

Таблица 23 – Цели и результат проекта

Цели проекта	– Изучить предметную область. – Разработать алгоритмы классификации пола и возраста и повторного распознавания людей. – Разработать программные средства. – Провести тестирование и настройку разработанных программных средств. – Внедрить разработанные программные средства в ядро платформы для подсчёта трафика.
Ожидаемые результаты	– Успешное внедрение разработки в ядро платформы. – Сдана выпускная квалификационная работа.
Критерии приёмки	– Успешное тестирование функционала в соответствии с функциональным требованием.
Требования к результату проекта	– Выполнены все пункты функционального требования

5.2 Планирование научно-исследовательских работ

5.2.1 Структура работ в рамках научного исследования

В Таблице 24 приведен порядок работ, выполняемых в ходе разработки, и исполнитель каждой работы.

Таблица 24 – Перечень работ и исполнителей при разработке модуля

№	Наименование работы	Исполнители работы
1	Выбор научного руководителя работы	Семенюта Антон Вадимович
2	Составление и утверждение темы работы	Аксёнов Сергей Владимирович, Семенюта Антон Вадимович
3	Составление календарного плана-графика выполнения работы	Аксёнов Сергей Владимирович, Семенюта Антон Вадимович
4	Подбор и изучение литературы по теме работы	Семенюта Антон Вадимович
5	Анализ предметной области	Семенюта Антон Вадимович
6	Исследование алгоритмов классификации пола и возраста и повторного распознавания	Семенюта Антон Вадимович
7	Разработка программных средств	Семенюта Антон Вадимович
8	Внедрение программных средств в ядро платформы интеллектуальной видеоаналитики	Семенюта Антон Вадимович
9	Согласование выполненной работы с научным руководителем	Аксёнов Сергей Владимирович
10	Выполнение других частей работы (финансовый менеджмент, социальная ответственность)	Семенюта Антон Вадимович
11	Подведение итогов, оформление работы	Аксёнов Сергей Владимирович, Семенюта Антон Вадимович

5.2.2 План проекта

Таблица 25 – Календарный план-график проведения НИОКР по теме

№	Вид	Исп	Т раб, дн.	Продолжительность														
				Янв.			Фев.			Март			Апр.			Май		
				1	2	3	1	2	3	1	2	3	1	2	3	1	2	3
1	Выбор научного руководителя работы	С	8															
2	Составление и утверждение темы работы	С	4															
		Р	2															
3	Составление календарного плана-графика выполнения работы	С	5															
		Р	1															
4	Подбор и изучение литературы по теме работы	С	18															
5	Анализ предметной области	С	14															
6	Исследование алгоритмов классификации пола и возраста и повторного распознавания	С	18															
7	Разработка программных средств	С	16															
8	Внедрение программных средств в ядро платформы интеллектуальной видеоаналитики	С	15															
9	Согласование выполненной работы с научным руководителем	Р	1															
10	Выполнение других частей работы (финансовый менеджмент, социальная ответственность)	С	7															
11	Подведение итогов, оформление работы	С	5															
		Р	1															

 – Руководитель(Р)

 – Студент (С)

Итого руководитель потратит на проект около 5 дней, а студент – 112 дней.

5.2.3 Бюджет научно-технического исследования

В процессе формирования бюджета научно-технического исследования используется следующая группировка затрат по статьям:

- материальные затраты НТИ;
- затраты на основное оборудование;
- основная заработная плата исполнителей темы;
- дополнительная заработная плата исполнителей темы;
- отчисления во внебюджетные фонды (страховые отчисления);
- накладные расходы.

5.2.3.1 Расчет материальных затрат на научное исследование

В материальные затраты научного исследования вошли затраты на канцелярские принадлежности в сумме 2000 рублей (Таблица 26).

Таблица 26 – Расчет затрат на по статье “Материальные затраты на научное исследование”

Наименование	Марка, размер	Кол-во	Цена за единицу, руб	Сумма, руб
Канцелярский набор	ErichKrause	2	1000	2000
Всего за материалы				2000
Транспортно-заготовительные расходы (3-5%)				100
Итого по статье C_m				2100

Таким образом, общая сумма материальных затрат составляет 2100 рублей.

5.2.3.2 Специальное оборудование для научных работ

Поскольку необходимые лицензии на программное обеспечение, которое было использовано при реализации разработки, уже были приобретены компанией, затраты на оборудование включают в себя только затраты на оборудование студента.

Во время проведения научного исследования использовался ПК стоимостью 65000 рублей. Расчёт затрат на амортизацию представлен в Таблице 27.

Таблица 27 – Расчет затрат на по статье “Спецоборудование для научных работ”

№ п/п	Наименование оборудования	Кол-во единиц оборудования	Цена единицы оборудования, тыс. руб	Общая стоимость оборудования, тыс. руб
1	Персональный компьютер	1	65000	65000
Итого				65000

Таким образом, общая сумма затрат на специальное оборудование для научных работ составляет 65000 рублей.

5.2.3.3 Основная заработная плата

Статья включает основную заработную плату научных и инженерно-технических работников, участвующих в выполнении работ по данной теме.

В Таблице 28 приведён баланс рабочего времени для участников разработки.

Таблица 28. Баланс рабочего времени участников разработки

Показатели рабочего времени	Руководитель	Инженер
Календарное число дней	365	365
Количество нерабочих дней - выходные дни - праздничные дни	104 14	104 14
Потери рабочего времени - отпуск - невыходы по болезни	24 —	24 —
Действительный годовой фонд рабочего времени	219	219

Расчёт основной заработной платы исполнителей представлен в Таблице 29.

Таблица 29 – Расчет основной заработной платы

Исполнители	З _{ок} , руб.	k _{пр}	k _д	k _р	З _м , руб	З _{дн} , руб.	T _р , раб. дн.	З _{осн} , руб.
Руководитель	30000	0,3	0,35	1,3	64350	3290,96	5	16454,80
Инженер	20000	0,3	0,35	1,3	42900	2193,97	112	245724,64
Итого по статье								262179,44

5.2.3.4 Дополнительная заработная плата научно-производственного персонала

В данную статью включается сумма выплат, предусмотренных законодательством о труде. Дополнительная заработная плата рассчитывается исходя из 10-15% от основной заработной платы. Заработная плата исполнителей НТИ представлена в Таблице 30.

Таблица 30 – Заработная плата исполнителей НТИ, руб

Заработная плата	Руководитель	Инженер
Основная зарплата	16454,80	245724,64
Дополнительная зарплата	1645,48	24572,46
Зарплата исполнителя	18100,28	270297,10
Итого по статье С _{зп}	288397,38	

5.2.3.5 Отчисления на социальные нужды

Статья включает в себя отчисления во внебюджетные фонды. Коэффициент отчислений на уплату во внебюджетные фонды составил 0,3.

Размер отчислений на социальные нужды представлены в Таблице 31.

Таблица 31 – Отчисления на социальные нужды, руб

Заработная плата	Руководитель	Инженер
Зарплата исполнителя	18100,28	270297,10
Отчисления на социальные нужды	5430,08	81089,13
Итого	86519,21	

5.2.3.6 Накладные расходы

В эту статью включаются затраты на управление и хозяйственное обслуживание.

Накладные расходы составляют 80-100% от суммы основной и дополнительной заработной платы.

Расчёт накладных расходов представлен в Таблице 32.

Таблица 32 – Накладные расходы, руб

Заработная плата	Руководитель	Инженер
Зарплата исполнителя	18100,28	270297,10
Накладные расходы	14480,22	216237,68
Итого	230717,90	

5.2.3.7 Формирование бюджета затрат научно-исследовательского проекта

Рассчитанная величина затрат научно-исследовательской работы является основой для формирования бюджета затрат проекта, который при формировании договора с заказчиком защищается научной организацией в качестве нижнего предела затрат на разработку научно-технической продукции. Все категории трат сведены в Таблице 33.

Таблица 33 – Бюджет затрат НТИ

Наименование статьи	Сумма, руб.
1. Материальные затраты	2100,00
2. Специальное оборудование	65000,00
3. Основная заработная плата	262179,44
4. Дополнительная заработная плата	26217,94
5. Отчисления на социальные нужды	86519,21
6. Накладные расходы	230717,90
7. Общий бюджет	672734,49

В результате было вычислено, что бюджет на разработку НТИ составит 672734,49 рублей.

5.3 Оценка сравнительной эффективности исследования

Определение эффективности происходит на основе расчета интегрального показателя эффективности научного исследования. Его нахождение связано с определением двух средневзвешенных величин: финансовой эффективности и ресурсоэффективности.

5.3.1 Интегральный показатель финансовой эффективности

Интегральный показатель финансовой эффективности научного исследования получают в ходе оценки бюджета затрат трёх (или более) вариантов исполнения исследования. Для этого наибольший интегральный показатель принимается за базу расчёта, с которым соотносятся финансовые значения по всем вариантам исполнения.

Расчёт интегрального показателя показателя финансовой эффективности приведен ниже.

$$I_{\Phi}^p = \frac{672734}{672734} = 1,00$$

$$I_{\Phi}^a = \frac{659279}{672734} = 0,98$$

$$I_{\Phi}^b = \frac{639097}{672734} = 0,95$$

$$I_{\Phi}^c = \frac{605460}{672734} = 0,90$$

Полученные величины интегрального финансового показателя эффективности разработки говорят о том, что альтернативные варианты исполнения исследования позволят сократить бюджет.

5.3.2 Интегральный показатель ресурсоэффективности

Расчёт интегрального показателя ресурсоэффективности приведен в Таблице 34.

Таблица 34. Сравнение характеристик вариантов исполнения проекта

Критерии \ ПО	Вес	Текущий проект	Аналог 1	Аналог 2	Аналог 3
Удобство использования	0,25	5	4	3	2
Масштабируемость	0,08	4	4	3	2
Требуемые ресурсы	0,08	4	5	4	4
Функциональность	0,14	4	4	4	4
Удобство обслуживания	0,08	5	4	3	3
Срок разработки	0,12	3	5	5	5
Надёжность	0,25	5	4	3	3
ИТОГО	1	4,46	4,22	3,46	3,13

5.3.3 Интегральный показатель эффективности

Сравнение интегрального показателя эффективности текущего проекта и аналогов позволит определить сравнительную эффективность проекта. Расчёты представлены в Таблице 35.

Таблица 35 – Сравнительная эффективность разработки

№	Показатели	Разработка	Аналог 1	Аналог 2	Аналог 3
1	Интегральный финансовый показатель разработки	1,00	0,98	0,95	0,90
2	Интегральный показатель ресурсоэффективности разработки	4,46	4,22	3,46	3,13
3	Интегральный показатель эффективности	4,46	4,31	3,64	3,48
4	Сравнительная эффективность аналогов и разработки	1	0,97	0,82	0,78

Таким образом, исходя из результатов сравнения значений интегральных показателей эффективности, можно сказать, что выбранный вариант исполнения технической задачи является наиболее приемлемым с позиции финансовой и ресурсной эффективности.

5.4 Вывод

В ходе данной работе были рассмотрены потенциальные потребители результатов исследования, также для анализа конкурентных технических решений. С позиции ресурсоэффективности и ресурсосбережения была составлена оценочная карта сравнения конкурентных технических решений, по результату которой разрабатываемая система имеет лучшие качества.

Далее был сформирован SWOT-анализ, в котором балы описаны сильные и слабые стороны проекта, в выявлении возможностей и угроз для реализации проекта, для выявления соответствия и несоответствия была составлена интерактивная матрица проекта.

В рамках процессов инициации определены внутренние и внешние заинтересованные стороны проекта с их ожиданиями от проекта, цели и результат проекта.

План проекта представлен на диаграмме Ганта, из которого видно какой исполнитель (студент или руководитель) какой вид работ осуществлял и в течении какого количества дней.

В бюджет инженерно-технического проекта занесены материальные затраты в размере 2100 рублей. Также добавлена стоимость оборудования, которая составила 65000 рублей. Была рассчитана основная и дополнительная заработная плата исполнителей проекта, сумма которых составила 288397,38 рублей. Вместе с отчислениями на социальные нужды и накладными расходами бюджет проекта получился равным 672734,49 рублям.

Был рассчитан интегральный финансовый показатель в ходе оценки бюджета затрат четырёх вариантов исполнения. Было выяснено, что альтернативные варианты исполнения могут удешевить разработку. Рассчитан интегральный показатель ресурсоэффективности для четырёх вариантов исполнения. Расчет интегрального показателя эффективности для разработки и аналогов позволил рассчитать сравнительную эффективность разработки, которая говорит о приемлемости существующего варианта решения поставленной в магистерской диссертации технической задачи с позиции финансовой и ресурсной эффективности.

Глава 6. Социальная ответственность

6.1 Введение

Социальная ответственность - ответственность отдельного ученого и научного сообщества перед обществом. Первостепенное значение при этом имеет безопасность применения технологий, которые создаются на основе достижений науки, предотвращение или минимизация возможных негативных последствий их применения, обеспечение безопасного как для испытуемых, как и для окружающей среды проведения исследований.

Разработка алгоритмов классификации пола и возраста людей и их повторного распознавания в видеопотоке, а также соответствующих программных средств осуществлялась на ПЭВМ.

В данном разделе рассматриваются опасные и вредные факторы, оказывающие влияние на производственную деятельность разработчика, воздействие объекта исследования на окружающую среду, правовые и организационные вопросы и мероприятия в чрезвычайных ситуациях.

6.2 Производственная безопасность

6.2.1 Вредные факторы

6.2.1.1 Отклонение показателей микроклимата в помещении

Проанализируем микроклимат в помещении, где находится рабочее место. Микроклимат производственных помещений определяют следующие параметры: температура, относительная влажность, скорость движения воздуха. Эти факторы влияют на организм человека, определяя его самочувствие.

Оптимальные и допустимые значения параметров микроклимата приведены в Таблице 36 и 37.

Таблица 37 – Оптимальные нормы микроклимата

Период года	Температура воздуха, С°	Относительная влажность воздуха, %	Скорость движения воздуха, м/с
Холодный	19-23	40-60	0.1
Теплый	23-25		0.2

Таблица 37 – Допустимые нормы микроклимата

Период года	Температура воздуха, С°		Относительная влажность воздуха, %	Скорость движения воздуха, м/с
	Нижняя допустимая граница	Верхняя допустимая граница		
Холодный	15	24	20-80	<0.5

Теплый	22	28	20-80	<0.5
--------	----	----	-------	------

Общая площадь рабочего помещения составляет 42м², объем составляет 147м³. По СанПиН 2.2.2/2.4.1340-03 санитарные нормы составляют 6,5 м² и 20 м³ объема на одного человека. Исходя из приведенных выше данных, можно сказать, что количество рабочих мест соответствует размерам помещения по санитарным нормам.

После анализа габаритных размеров рассмотрим микроклимат в этой комнате. В качестве параметров микроклимата рассмотрим температуру, влажность воздуха, скорость ветра.

В помещении осуществляется естественная вентиляция посредством наличия легко открываемого оконного проема (форточки), а также дверного проема. По зоне действия такая вентиляция является общеобменной. Основным недостаток - приточный воздух поступает в помещение без предварительной очистки и нагревания. Согласно нормам СанПиН 2.2.2/2.4.1340-03 объем воздуха необходимый на одного человека в помещении без дополнительной вентиляции должен быть более 40м³[1]. В нашем случае объем воздуха на одного человека составляет 42 м³, из этого следует, что дополнительная вентиляция не требуется. Параметры микроклимата поддерживаются в холодное время года за счет систем водяного отопления с нагревом воды до 100°С, а в теплое время года – за счет кондиционирования, с параметрами согласно [2]. Нормируемые параметры микроклимата, ионного состава воздуха, содержания вредных веществ должны соответствовать требованиям [3].

6.2.1.2 Превышение уровней шума

Одним из наиболее распространенных в производстве вредных факторов является шум. Он создается вентиляционным и рабочим оборудованием, преобразователями напряжения, рабочими лампами дневного света, а также проникает снаружи. Шум вызывает головную боль, усталость, бессонницу или сонливость, ослабляет внимание, память ухудшается, реакция уменьшается.

Основным источником шума в комнате являются компьютерные охлаждающие вентиляторы и. Уровень шума варьируется от 35 до 42 дБА. Согласно СанПиН 2.2.2 / 2.4.1340-03, при выполнении основных работ на ПЭВМ уровень шума на рабочем месте не должен превышать 80 дБА [4].

При значениях выше допустимого уровня необходимо предусмотреть средства индивидуальной защиты(СИЗ) и средства коллективной защиты (СКЗ) от шума.

Средства коллективной защиты:

1. Устранение причин шума или существенное его ослабление в источнике образования;

2. Изоляция источников шума от окружающей среды (применение глушителей, экранов, звукопоглощающих строительных материалов, например, любой пористый материал – шамотный кирпич, микропористая резина, поролон и др.);

3. Применение средств, снижающих шум и вибрацию на пути их распространения;
Средства индивидуальной защиты;

1. Применение спецодежды и защитных средств органов слуха: наушники, беруши, антифоны.

6.2.1.3 Повышенный уровень электромагнитных излучений

Источником электромагнитных излучений в нашем случае являются дисплей ПЭВМ. Монитор компьютера включает в себя излучения рентгеновской, ультрафиолетовой и инфракрасной области, а также широкий диапазон электромагнитных волн других частот. Согласно СанПиН 2.2.2/2.4.1340-03 напряженность электромагнитного поля по электрической составляющей на расстоянии 50 см вокруг ВДТ не должна превышать 25В/м в диапазоне от 5Гц до 2кГц, 2,5В/м в диапазоне от 2 до 400кГц [1]. Плотность магнитного потока не должна превышать в диапазоне от 5 Гц до 2 кГц 250нТл, и 25нТл в диапазоне от 2 до 400кГц. Поверхностный электростатический потенциал не должен превышать 500В [1]. В ходе работы использовалась ПЭВМ типа Acer VN7-791 со следующими характеристиками: напряженность электромагнитного поля 2,5В/м; поверхностный потенциал составляет 450 В (основы противопожарной защиты предприятий ГОСТ 12.1.004 и ГОСТ 12.1.010 – 76.) [5].

При длительном постоянном воздействии электромагнитного поля (ЭМП) радиочастотного диапазона при работе на ПЭВМ у человеческого организма возникают сердечно-сосудистые, респираторные и нервные расстройства, головные боли, усталость, ухудшение состояния здоровья, гипотония, изменения сердечной мышцы проводимости. Тепловой эффект ЭМП характеризуется увеличением температуры тела, локальным селективным нагревом тканей, органов, клеток за счет перехода ЭМП на теплую энергию.

Предельно допустимые уровни (ПДУ) облучения (по *ОСТ 54 30013-83*):

- a. до 10 мкВт/см² , время работы (8 часов);
- b. от 10 до 100 мкВт/см² , время работы не более 2 часов;
- c. от 100 до 1000 мкВт/см² , время работы не более 20 мин. при условии пользования защитными очками;
- d. для населения в целом ППМ не должен превышать 1 мкВт/см².

Защита человека от опасного воздействия электромагнитного излучения осуществляется следующими способами:

СКЗ:

1. защита временем;
2. защита расстоянием;
3. снижение интенсивности излучения непосредственно в самом источнике излучения;
4. заземление экрана вокруг источника;
5. защита рабочего места от излучения.

СИЗ:

1. Очки и специальная одежда, выполненная из металлизированной ткани (кольчуга). При этом следует отметить, что использование СИЗ возможно при кратковременных работах и является мерой аварийного характера. Ежедневная защита обслуживающего персонала должна обеспечиваться другими средствами;
2. Вместо обычных стекол используют стекла, покрытые тонким слоем золота или диоксида олова (SnO_2).

6.2.1.4 Недостаточная освещенность

Для обеспечения требуемой освещенности необходимо использовать совмещенное освещение, создаваемое сочетанием естественного и искусственного освещения. При данном этапе развития осветительной техники целесообразно использовать люминесцентные лампы, которые по сравнению с лампами накаливания имеют большую светотдачу на ватт потребляемой мощности и более естественный спектр.

Минимальный уровень средней освещенности на рабочих местах с постоянным пребыванием людей должен быть не менее 200 лк.

В расчётном задании должны быть решены следующие вопросы:

- выбор системы освещения;
- выбор источников света;
- выбор светильников и их размещение;
- выбор нормируемой освещённости;
- расчёт освещения методом светового потока.

В данном расчётном задании для всех помещений рассчитывается общее равномерное освещение (Таблица 38).

Таблица 38 – Габариты помещения.

Параметр	Обозначение	Значение, м
Длина	A	12
Ширина	B	10
Высота помещения	H	3,5

Расчёт общего равномерного искусственного освещения горизонтальной рабочей поверхности выполняется методом коэффициента светового потока, учитывающим световой поток, отражённый от потолка и стен.

Световой поток лампы определяется по Формуле 26:

$$\Phi_{\text{рас}} = E_{\text{н}} \cdot S \cdot K_3 \cdot Z / (N \cdot \eta) \quad (26)$$

Где $E_{\text{н}}$ – нормируемая минимальная освещённость по СНиП 23-05-95, лк; S – площадь освещаемого помещения, м^2 ; K_3 – коэффициент запаса, учитывающий загрязнение светильника (источника света, светотехнической арматуры, стен и пр., т. е. отражающих поверхностей), наличие в атмосфере цеха дыма, пыли (табл. 4.9); Z – коэффициент неравномерности освещения, отношение $E_{\text{ср}} / E_{\text{min}}$. Для люминесцентных ламп при расчётах берётся равным 1,1; N – число ламп в помещении; η – коэффициент использования светового потока.

Коэффициент использования светового потока показывает, какая часть светового потока ламп попадает на рабочую поверхность. Он зависит от индекса помещения i , типа светильника, высоты светильников над рабочей поверхностью h и коэффициентов отражения стен $\rho_{\text{с}}$ и потолка $\rho_{\text{п}}$.

Индекс помещения определяется по Формуле 27:

$$i = S / h \cdot (A + B) \quad (27)$$

Проведем расчет индекса помещения:

Площадь помещения:

$$S = A \cdot B = 12 \cdot 10 = 120 \text{ м}^2$$

Индекс:

$$i = \frac{S}{h \cdot (A + B)} = \frac{120}{2.35 \cdot (12 + 10)} = 2.32$$

Согласно этим данным коэффициент использования светового потока будет равен 56 % или в долях = 0,56.

Коэффициенты отражения оцениваются субъективно (табл. 4.10) [БЖД Практикум 2009-2020].

Согласно указанной методике выбираем тип источника света.

Наиболее подходящим вариантом является 40 ваттная лампа ЛБ, у которой $\Phi=2800$ лм. Для выбранного типа лампы подходит светильник ОД-2-40 с размерами: длина = 1230 мм, ширина = 266 мм.

Из уравнения 1.5.1 находим количество ламп для помещения

$$N = E_H \cdot S \cdot K_3 \cdot Z / (\Phi \cdot \eta) = 200 \cdot 120 \cdot 1,3 \cdot 1,1 / 2800 \cdot 0,56 = 21,875;$$

Принимаем $N=24$ лампы или 12 светильников..

Размещаем светильники в 3 ряда по 4 светильника в ряду с соблюдением условий: L – расстояние между соседними светильниками или рядами (если по длине (А) и ширине (В) помещения расстояния различны, то они обозначаются L_A и L_B),

L – расстояние между соседними светильниками или рядами (если по длине (А) и ширине (В) помещения расстояния различны, то они обозначаются L_A и L_B),

l – расстояние от крайних светильников или рядов до стены.

Оптимальное расстояние l от крайнего ряда светильников до стены рекомендуется принимать равным $L/3$.

Сначала определим световой поток расчетный.

$$\Phi = E_H \cdot S \cdot K_3 \cdot Z / (N \cdot \eta) = 200 \cdot 120 \cdot 1,3 \cdot 1,1 / (24 \cdot 0,56) = 2554 \text{ лм};$$

Проведем проверку выполнения условия соответствия:

$$- 10\% \leq (\Phi_{\text{расч}} - \Phi_{\text{станд}}) / \Phi_{\text{расч}} \cdot 100\% \leq + 20\%$$

Подставляя численные значения получаем:

$$- 10\% \leq (2800 - 2554) / 2554 \cdot 100 \% \leq + 20\%$$

$$- 10\% \leq +9,6\% \leq + 20\%$$

Результат расчета укладывается в допустимые пределы.

Определим мощность осветительной установки:

$$P = N \cdot P_i = 24 \cdot 40 \text{ Вт} = 960 \text{ Вт}.$$

Теперь определим расстояния между светильниками по длине и ширине помещения.

$$12000 = 3 \cdot L_A + 4 \cdot 1230 + 2 / 3 \cdot L_A;$$

$$L_A = (12000 - 4920) \cdot 3 / 11 = 1930 \text{ мм};$$

$$L_A / 3 = 644 \text{ мм};$$

$$10000 = 2 \cdot L_B + 3 \cdot 266 + 2 / 3 \cdot L_B;$$

$$L_B = (10000 - 798) \cdot 3 / 8 = 3450 \text{ мм};$$

$$L_B / 3 = 1150 \text{ мм}.$$

Рисуем схему размещения светильников на потолке для обеспечения общего равномерного освещения (Рисунок 43).

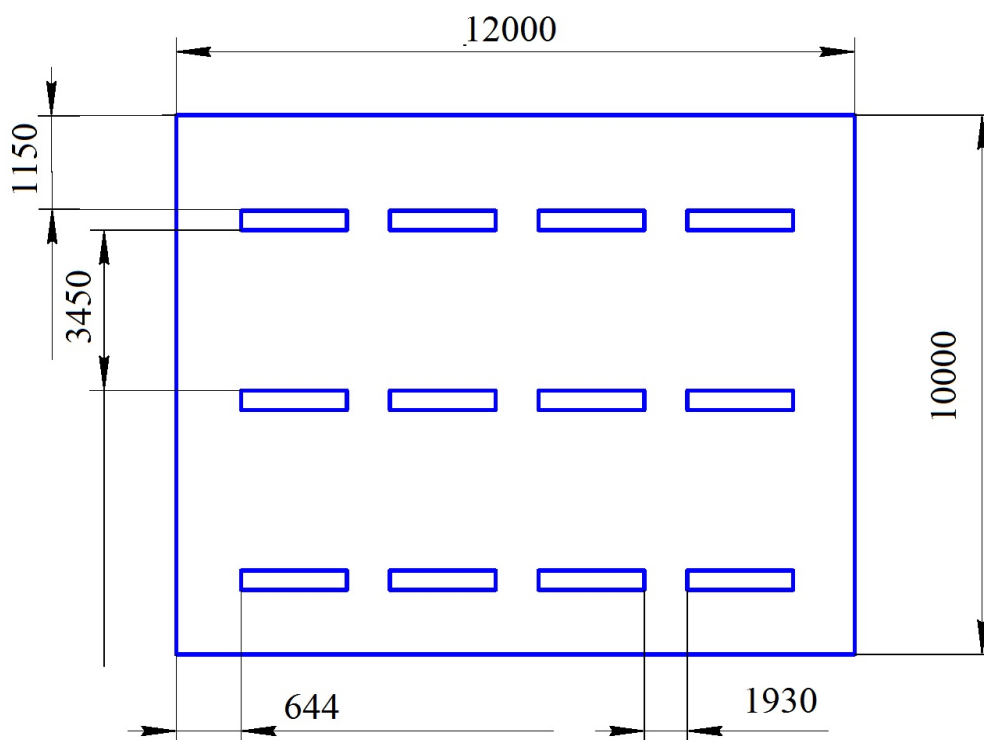


Рисунок 43. План размещения светильников на потолке

6.2.2 Опасные факторы

6.2.2.1 Электроопасность; класс электроопасности помещения, безопасные номиналы I , U , $R_{\text{заземления}}$, СКЗ, СИЗ; Поражение электрическим током

К опасным факторам можно отнести наличие в помещении большого количества аппаратуры, использующей однофазный электрический ток напряжением 220 В и частотой 50 Гц. По опасности электропоражения комната относится к помещениям без повышенной опасности, так как отсутствует повышенная влажность, высокая температура, токопроводящая пыль и возможность одновременного сприкосновения токоведущих элементов с заземленными металлическими корпусами оборудования [6].

Лаборатория относится к помещению без повышенной опасности поражения электрическим током. Безопасными номиналами являются: $I < 0,1$ А; $U < (2-36)$ В; $R_{\text{зазем}} < 4$ Ом.

Для защиты от поражения электрическим током используют СИЗ и СКЗ.

Средства коллективной защиты:

1. защитное заземление, зануление;
2. малое напряжение;
3. электрическое разделение сетей;
4. защитное отключение;
5. изоляция токоведущих частей;
6. оградительные устройства.

7. использование щитов, барьеров, клеток, ширм, а также заземляющих и шунтирующих штанг, специальных знаков и плакатов.

Средства индивидуальной защиты:

1. использование диэлектрических перчаток, изолирующих клещей и штанг, слесарных инструментов с изолированными рукоятками, указатели величины напряжения, калоши, боты, подставки и коврики.

6.2.2.2 Пожароопасность, категория пожароопасности помещения, марки огнетушителей, их назначение и ограничение применения; Приведена схема эвакуации.

По взрывопожарной и пожарной опасности помещения подразделяются на категории А, Б, В1-В4, Г и Д.

Согласно НПБ 105-03 лаборатория относится к категории В – горючие и трудно горючие жидкости, твердые горючие и трудно горючие вещества и материалы, вещества и материалы, способные при взаимодействии с водой, кислородом воздуха или друг с другом только гореть, при условии, что помещения, в которых находится, не относятся к категории наиболее опасных А или Б.

По степени огнестойкости данное помещение относится к 1-й степени огнестойкости по СНиП 2.01.02-85 (выполнено из кирпича, которое относится к трудносгораемым материалам).

Возникновение пожара при работе с электронной аппаратурой может быть по причинам как электрического, так и неэлектрического характера.

Причины возникновения пожара неэлектрического характера:

а) халатное неосторожное обращение с огнем (курение, оставленные без присмотра нагревательные приборы, использование открытого огня);

Причины возникновения пожара электрического характера: короткое замыкание, перегрузки по току, искрение и электрические дуги, статическое электричество и т. п.

Для локализации или ликвидации загорания на начальной стадии используются первичные средства пожаротушения. Первичные средства пожаротушения обычно применяют до прибытия пожарной команды.

Огнетушители водо-пенные (ОХВП-10) используют для тушения очагов пожара без наличия электроэнергии. Углекислотные (ОУ-2) и порошковые огнетушители предназначены для тушения электроустановок, находящихся под напряжением до 1000В. Для тушения токоведущих частей и электроустановок применяется переносной порошковый огнетушитель, например, ОП-5.

В общественных зданиях и сооружениях на каждом этаже должно размещаться не менее двух переносных огнетушителей. Огнетушители следует располагать на видных местах вблизи от выходов из помещений на высоте не более 1,35 м. Размещение первичных средств пожаротушения в коридорах, переходах не должно препятствовать безопасной эвакуации людей.

Для предупреждения пожара и взрыва необходимо предусмотреть:

1. Специальные изолированные помещения для хранения и разлива легковоспламеняющихся жидкостей (ЛВЖ), оборудованные приточно-вытяжной вентиляцией во взрывобезопасном исполнении - соответствии с ГОСТ 12.4.021-75 и СНиП 2.04.05-86;
2. Специальные помещения (для хранения в таре пылеобразной канифоли), изолированные от нагревательных приборов и нагретых частей оборудования;
3. Первичные средства пожаротушения на производственных участках (передвижные углекислые огнетушители ГОСТ 9230-77, пенные огнетушители ТУ 22-4720-80, ящики с песком, войлок, кошма или асбестовое полотно);
4. Автоматические сигнализаторы (типа СВК-3 М 1) для сигнализации о присутствии в воздухе помещений предвзрывных концентраций горючих паров растворителей и их смесей.

Лаборатория полностью соответствует требованиям пожарной безопасности, а именно, наличие охранно-пожарной сигнализации, плана эвакуации, изображенного на рисунке 1, порошковых огнетушителей с поверенным клеймом, табличек с указанием направления к запасному (эвакуационному) выходу (Рисунок 44).

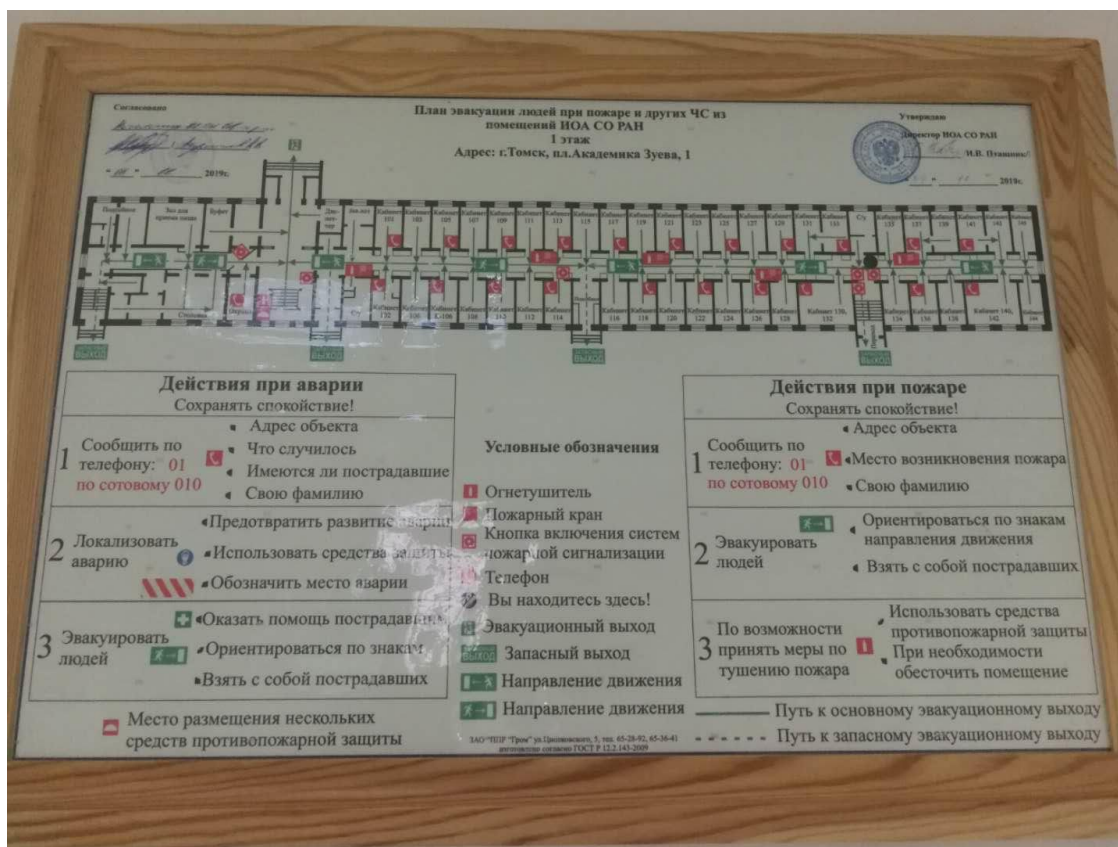


Рисунок 44. План эвакуации

6.3 Экологическая безопасность

Во время выполнения выпускной квалификационной работы вынуждены использовать черновики (предварительная запись информации) на бумажном носителе. Записи несут в себе конфиденциальную, а иногда даже секретную информацию. Чтобы повторно использовать бумагу для записей необходимо бумагу с записями shredded с помощью shredder, спрессовать для уменьшения объема, упаковать в герметичную упаковку и хранить на складе до накопления объема для 1 транспортной единицы, после чего отправить на утилизацию макулатуры в ближайший ее пункт приема.

В компьютерах огромное количество компонентов, которые содержат токсичные вещества и представляют угрозу, как для человека, так и для окружающей среды.

К таким веществам относятся:

- Свинец (накапливается в организме, поражая почки, нервную систему);
- Никель и цинк (могут вызывать дерматит);
- Щелочи (прожигают слизистые оболочки и кожу).

Поэтому компьютер требует специальных комплексных методов утилизации.

Утилизацию компьютера можно провести следующим образом:

- Отделить металлические детали от неметаллов;
- Разделить углеродистые металлы от цветмета;

- Пластмассовые изделия (крупногабаритные) измельчить для уменьшения объема;
- Копир-порошок упаковать в отдельную упаковку, точно также, как и все проклассифицированные и измельченные компоненты оргтехники, и после накопления на складе транспортных количеств отправить предприятиям и фирмам, специализирующимся по переработке отдельных видов материалов.
- Люминесцентные лампы утилизируют следующим образом. Не работающие лампы немедленно после удаления из светильника должны быть упакованы в картонную коробку, бумагу или тонкий мягкий картон, предохраняющий лампы от взаимного соприкосновения и случайного механического повреждения. После накопления ламп объемом в 1 транспортную единицу их сдают на переработку на соответствующее предприятие. Недопустимо выбрасывать отработанные энергосберегающие лампы вместе с обычным мусором, превращая его в ртутьсодержащие отходы, которые загрязняют ртутными парами

6.4 Безопасность в чрезвычайных ситуациях

Природная чрезвычайная ситуация – обстановка на определенной территории или акватории, сложившейся в результате возникновения источника природной чрезвычайной ситуации, который может повлечь или повлек за собой человеческие жертвы, ущерб здоровью людей и (или) окружающей природной среде, значительные материальные потери и нарушение условий жизнедеятельности людей.

Производство находится в городе Томске с континентально-циклоническим климатом. Природные явления (землетрясения, наводнения, засухи, ураганы и т. д.), в данном городе отсутствуют.

Возможными ЧС на объекте в данном случае, могут быть сильные морозы и диверсия.

Для Сибири в зимнее время года характерны морозы. Достижение критически низких температур приводит к авариям систем тепло- и водоснабжения, сантехнических коммуникаций и электроснабжения, приостановке работы. В этом случае при подготовке к зиме следует предусмотреть, а) газобаллонные калориферы (запасные обогреватели), б) дизель или бензоэлектрогенераторы; в) запасы питьевой и технической воды на складе (не менее 30 л на 1 человека); г) теплый транспорт для доставки работников на работу и с работы домой в случае отказа муниципального транспорта. Их количества и мощности должно хватать для того, чтобы работа на производстве не прекратилась.

В производственном помещении, где выполнялось научно-техническое исследование, наиболее вероятно возникновение чрезвычайных ситуаций (ЧС) техногенного характера.

Для предупреждения вероятности осуществления диверсии предприятие необходимо оборудовать системой видеонаблюдения, круглосуточной охраной, пропускной системой, надежной системой связи, а также исключения распространения информации о системе охраны объекта, расположении помещений и оборудования в помещениях, системах охраны, сигнализаторах, их местах установки и количестве. Должностные лица раз в полгода проводят тренировки по отработке действий на случай экстренной эвакуации.

6.5 Вывод

В рамках раздела «Социальная ответственность» процесс выполнения и результаты дипломной работы были рассмотрены с точки зрения социальной ответственности за моральные, общественные, экономические, экологические последствия и ущерб здоровью человека. Дополнительно был выполнен анализ на предмет выявления основных опасных и вредных факторов и оценена степень их воздействия на человека, общество и природную среду. Были предложены методы для защиты и минимизации воздействий выявленных факторов, а также методы предотвращения и устранения возможных чрезвычайных ситуаций.

6.6 Перечень НТД

1. ГОСТ 54 30013-83. Электромагнитные излучения СВЧ. Предельно допустимые уровни облучения. Требования безопасности;
2. ГОСТ 12.4.154-85. «ССБТ. Устройства экранирующие для защиты от электрических полей промышленной частоты»;
2. ГН 2.2.5.1313-03. Предельно допустимые концентрации (ПДК) вредных веществ в воздухе рабочей зоны;
3. СанПиН 2.2.4/2.1.8.055-96. «Электромагнитные излучения радиочастотного диапазона (ЭМИ РЧ)»;
4. СанПиН 2.2.4.548-96. Гигиенические требования к микроклимату производственных помещений.;
5. СН 2.2.4/2.1.8.562-96. Шум на рабочих местах, в помещениях жилых, общественных зданий и на территории жилой застройки.;
6. ГОСТ 12.4.123-83. Средства коллективной защиты от инфракрасных излучений. Общие технические требования.;

7. ГОСТ Р 12.1.019-2009. Электробезопасность. Общие требования и номенклатура видов защиты.;
8. ГОСТ 12.1.030-81. Электробезопасность. Защитное заземление. Зануление.;
9. ГОСТ 12.1.004-91. Пожарная безопасность. Общие требования.;
10. ГОСТ 12.2.037-78. Техника пожарная. Требования безопасности;
11. СанПиН 2.1.6.1032-01. Гигиенические требования к качеству атмосферного воздуха;
12. ГОСТ 30775-2001. Ресурсосбережение. Обращение с отходами. Классификация, идентификация и кодирование отходов. ;
13. СНиП 21-01-97. Противопожарные нормы.;
14. ГОСТ 12.4.154. Система стандартов безопасности труда. Устройства экранирующие для защиты от электрических полей промышленной частоты. Общие технические требования, основные параметры и размеры.

Заключение

В ходе проведения работ в рамках производственной практики получены следующие результаты:

- Разработана собственная версия алгоритма распознавания пола и возраста людей на кадрах в видеопотоке, использующая признаки силуэта и движения. Предложенный конвейер методов обеспечивает качество распознавания пола, равное 0,89 по мере F1, и качество распознавания категории возраста, равное 0,74 по мере F1;
- Разработана собственная версия алгоритмов создания индивидуальных отпечатков объектов по графическим признакам для повторного распознавания, работающая с данными, приближенными к наиболее общим условиям съёмки. Реализованный подход обеспечивает качество решения задачи повторного распознавания объектов, равное 0,98 по метрике СМС с рангом 1;
- Алгоритмы распознавания пола и возраста, а также создания индивидуальных отпечатков объектов по графическим признакам для повторного распознавания внедрены в работу платформы, обеспечивающую детектирование и трекинг объектов;
- Выполнена разработка расширенной обучающей выборки, позволившая создать алгоритмы на основе глубокого обучения, учитывающие особенности данных видеозаписей в общих условиях съёмки;
- Разработана программа и методика испытаний модуля ядра платформы с внедрёнными в его работу новыми функциями. Проведение испытаний позволило как отобрать наилучшие методы и их параметры для функционирования разработанных алгоритмов, так и улучшить общее качество всего конвейера видеоаналитики, реализованного в платформе.

Список источников

1. Apparent Age Estimation Using Ensemble of Deep Learning Models (Malli R.C. Apparent Age Estimation Using Ensemble of Deep Learning Models / R.C. Malli, M. Aygun, H.K. Ekenel // IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW). – 2016. – Pp. 714-721.
2. Hasan M.K. Human Face Detection Techniques: A Comprehensive Review and Future Research Directions / M.K. Hasan, M.S. Ahsan, Abdullah-Al-Mamun [at al.] // Electronics. – 2021. – Vol. 10(19). – Pp. 2354.
3. Naseem I. Linear Regression for Face Recognition / I. Naseem, R. Togneri, M. Bennamoun // IEEE Transactions on Pattern Analysis and Machine Intelligence. – 2010. – Vol. 32 (11). – Pp. 2106–2112.
4. Perlibakas V. Distance measures for PCA-based face recognition / V. Perlibakas // Pattern Recognition Letters. – 2004. – Vol. 25(6). – Pp. 711–724.
5. Liu W. A survey of deep neural network architectures and their applications / W. Liu, Z. Wang, X. Liu [at al.] // Neurocomputing. – 2017. – Vol. 234. – Pp. 11-26.
6. Gu W. Facial expression recognition using radial encoding of local Gabor features and classifier synthesis / W. Gu, C. Xiang, Y. Venkatesh // Pattern Recognition. – 2012. – Vol. 45(1). – Pp. 80-91.
7. Serrano S. Recent advances in face biometrics with Gabor wavelets: A review / S. Serrano, I. Diego, C. Conde // Pattern Recognition Letters. – 2010. – Vol. 31(5). – Pp. 372-381.
8. Chandrappa D.N. Face Detection Using a Boosted Cascade of Features Using OpenCV / D.N. Chandrappa, G. Akshay, M. Ravishankar // Wireless Networks and Computational Intelligence (ICIP 2012). – 2012. Vol. 292. – Pp. 399-404.
9. Serrano S. Analysis of variance of Gabor filter banks parameters for optimal face recognition / S. Serrano, I. Diego, C. Conde [at al.]. // Pattern Recognition Letters. – 2011. Vol. 31(15). – Pp. 1998-2008.
10. Chen J. WLD: a robust local image descriptor / J. Chen, S. Shan, C. He, [at al.]. // IEEE Trans. Pattern Anal. Mach. Intell. – 2010. – Vol. 32(9) – Pp. 1705–1720.
11. Bereta M. Local descriptors and similarity measures for frontal face recognition: a comparative analysis / M. Bereta, W. Pedrycz, M. Reformat // Journal of Visual Communication and Image Representation. – 2013. – Vol. 24(8) – Pp. 1213–1231.
12. Bereta M. Local descriptors in application to the aging problem in face recognition / M. Bereta, P. Karczmarek, W. Pedrycz, [at al.]. // Pattern Recognit. – 2013. – Vol. 46(10) – Pp. 2634–2646.

13. What is a LiDAR scanner, the iPhone 12 Pro's camera upgrade, anyway. // tecradar. 2021. URL: <https://www.techradar.com/news/what-is-a-lidar-scanner-the-iphone-12-pros-rumored-camera-upgrade-anyway> (Дата обращения 21.07.2022).
14. Kevin W. A survey of approaches and challenges in 3D and multi-modal 3D + 2D face recognition / W. Kevin, C. Kyong, F. Patrick // *Computer Vision and Image Understanding*. – 2006. – Vol. 101(1) – Pp. 1-15.
15. Blanz V. Face Recognition Based on Fitting a 3D Morphable Model / V. Blanz, T. Vetter // *IEEE Trans. Pattern Anal. Mach. Intell.* – 2003. – Vol. 25(9) – Pp. 1063–1074.
16. Zhang L. Face recognition from a single training image under arbitrary unknown lighting using spherical harmonics / L. Zhang, D. Samaras // *IEEE Transactions on Pattern Analysis and Machine Intelligence*. – 2006. – Vol. 28(3) – Pp. 351-363.
17. Passalis G. Using Facial Symmetry to Handle Pose Variations in Real-World 3D Face Recognition / G. Passalis, P. Panagiotis, T. Theoharis [at al.]. // *IEEE Transactions on Pattern Analysis and Machine Intelligence*. – 2011. – Vol. 33(10) – Pp. 1938–1951.
18. Zakariya Q. Deep Convolutional Neural Network for Age Estimation based on VGG-Face Model / Q. Zakariya, A. Arafat, D. Buket [at al.] // *ArXiv*. – 2017. – 1709.01664.
19. Chen S. Using Ranking-CNN for Age Estimation / S. Chen, C. Zhang, M. Dong [at al.] // *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. – 2017. – Pp. 742-751.
20. Liu X. AgeNet: Deeply Learned Regressor and Classifier for Robust Apparent Age Estimation / X. Liu, S. Li, M. Kan [at al.] // *015 IEEE International Conference on Computer Vision Workshop (ICCVW)*. – 2015. – Pp. 258-266.
21. Agbo-Ajala O. Deep learning approach for facial age classification: a survey of the state-of-the-art / O. Agbo-Ajala, S. Viriri // *Artificial Intelligence Review*. – 2021. – Pp. 179–213.
22. Antipov G. Apparent Age Estimation from Face Images Combining General and Children-Specialized Deep Learning Models / G. Antipov, M. Baccouche, S. Berrani [at al.] // *2016 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*. – 2016. – Pp. 801-809.
23. Makihara Y. Gait-based age estimation using a whole-generation gait database / Y. Makihara, M. Okumura, H. Iwama [at all] // *Multimedia Tools and Applications*. – 2018. – Vol. 77. – Pp. 28333–28354.
24. Li X. Gait-based human age estimation using age group-dependent manifold learning and regression / X. Li, Y. Makihara, C. Xu [at al.] // *Multimedia Tools and Applications*. – 2018. – Vol. 77. – Pp. 28333–28354.

25. El-Alfy E.M. Silhouette-Based Gender Recognition in Smart Environments Using Fuzzy Local Binary Patterns and Support Vector Machines / E. M. El-Alfy, A. G. Binsaadoon // *Procedia Computer Science*. – 2017. – Vol.109. – Pp. 164-171.
26. Kitchat K. Gender classification from gait silhouette using observation angle-based GEIs / K. Kitcha, N. Khamsemanan, C. Nattee // 2019 IEEE International Conference on Cybernetics and Intelligent Systems (CIS) and IEEE Conference on Robotics, Automation and Mechatronics (RAM). – 2018. – Pp. 485-490.
27. Mansouri N. Gait-based human age classification using silhouette model / N. Mansouri, A.I. Mohammed, B.J. Yousra // *IET Biom*. – 2018. – Vol. 7(2). – Pp. 116-124.
28. Baek N. Multimodal Camera-Based Gender Recognition Using Human-Body Image With Two-Step Reconstruction Network / N. Baek, S. W. Cho, J. H. Koo [at al.] // *IEEE Access*. – 2017. – Vol. 7. – Pp. 104025-104044.
29. Jung C.R. Efficient Background Subtraction and Shadow Removal for Monochromatic Video Sequences / C. R. Jung // *IEEE Transactions on Multimedia*. – 2009. – Vol. 11(3). – Pp. 571-577.
30. Braham M. Deep background subtraction with scene-specific convolutional neural networks / M. Braham, M.V. Droogenbroeck // 2016 International Conference on Systems, Signals and Image Processing (IWSSIP). – 2016. – Pp. 1-4.
31. Wang Y. CDnet 2014: an expanded change detection benchmark dataset / Y. Wang, P. Jodoin, F. Porikli [et al.] // 2014 IEEE Conference on Computer Vision and Pattern Recognition Workshops. – 2014. – Pp. 393-400.
32. Bouwmans T. Deep Neural Network Concepts for Background Subtraction: A Systematic Review and Comparative Evaluation / T. Bouwmans, S. Javed, M. Sultana [at al.] // *Neural Networks*. – 2019. – Vol. 117. – Pp. 8-66.
33. Dedeoğlu Y. Silhouette-Based Method for Object Classification and Human Action Recognition in Video / Y. Dedeoğlu, B.U. Töreyn, U. Güdükbay [at al.] // *Lecture Notes in Computer Science*. – 2006. – Vol. 3979. – Pp. 64-77.
34. Elharrouss O., Gait recognition for person re-identification / O. Elharrouss, N. Almaadeed, S. Al-Maadeed [et al.] // *The Journal of Supercomputing*. – 2021. – Vol. 7. – Pp. 3653-3672.
35. Zheng S. Robust view transformation model for gait recognition / S. Zheng, J. Zhang, K. Huang [at al.] // 2011 18th IEEE International Conference on Image Processing. – 2011. – Pp. 2073-2076.

36. Iwama H. The OU-ISIR Gait Database Comprising the Large Population Dataset and Performance Evaluation of Gait Recognition / H. Iwama, M. Okumura, Y. Makihara [at al.] // IEEE Transactions on Information Forensics and Security. – 2012. – Vol. 7(5). – Pp. 1511-1521.
37. Takemura N. Multi-view large population gait dataset and its performance evaluation for cross-view gait recognition / N. Takemura, Y. Makihara, D. Muramatsu [at al.] // IPSJ Transactions on Computer Vision and Applications. – 2018. – Vol. 10(4). – Pp. 1-14.
38. Sivapalan S. The Backfilled GEI – A Cross-Capture Modality Gait Feature for Frontal and Side-View Gait Recognition / S. Sivapalan, D. Chen, S. Denman [at al.] // 2012 International Conference on Digital Image Computing Techniques and Applications (DICTA). – 2012. – Pp. 1-8.
39. Liu T. Combining Convolutional Neural Network and Support Vector Machine for Gait-based Gender Recognition / T. Liu, X. Ye, B. Sun // 2018 Chinese Automation Congress (CAC). – 2018. – Pp. 3477-3481.
40. Li X. Object Re-Identification Based on Deep Learning / X. Li, Z. Zhou // Visual Object Tracking with Deep Neural Networks. – 2019.
41. What is reidentification? // KLAP – Augmenting fan experience for the sport industry. 2021. URL: <https://www.getklap.com/blog/what-is-reidentification> (Дата обращения 23.07.2022).
42. Ristani E. Features for Multi-Target Multi-Camera Tracking and Re-Identification / E. Ristani, C. Tomasi. // ArXiv – 2018. – 1803.10859.
43. Person Re-Identification // Papers With Code. 2022. URL: <https://paperswithcode.com/task/person-re-identification> (Дата обращения 24.07.2022).
44. Двоеносова Г. Биометрия как наука, метод и способ документирования [Электронный ресурс] / Г. Двоеносова, М. Двоеносова // Управление персоналом. – 2009. – № 11. – URL: <https://www.top-personal.ru/issue.html?2039> (Дата обращения 23.07.2022).
45. Person Re-Identification Shaogang Gong [Electronic resource] / Shaogang Gong [et al.]. – Springer. – 2014. – 444 p.
46. Zhou K. Learning Generalisable Omni-Scale Representations for Person Re-Identification / K. Zhou, Y. Yang, A. Cavallaro [at al.] // IEEE Transactions on Pattern Analysis and Machine Intelligence. – 2021. – Pp. 1-1.
47. Cai Y. Exploring context information for inter-camera multiple target tracking / Y. Cai, G. Medioni. // IEEE Winter Conference on Applications of Computer Vision. – 2014. – Pp. 761-768.
48. Das A. Consistent Re-identification in a Camera Network / A. Das, A. Chakraborty, A.K. Roy-Chowdhury // Lecture Notes in Computer Science. – 2014. – Vol. 8690 – Pp. 330-345).

49. Chen, X. Multitarget Tracking in Nonoverlapping Cameras Using a Reference Set / X. Chen, L. An, B. Bhanu // IEEE Sensors Journal. – 2015. – Pp. 2692-2704.
50. Zhang S. A Camera Network Tracking (CamNeT) Dataset and Performance Baseline / S. Zhang, E. Staudt, E. Faltemier [et al.] // IEEE Winter Conference on Applications of Computer Vision. – 2015. – 365-372.
51. Martinel N. Saliency Weighted Features for Person Re-Identification / N. Martinel, C. Micheloni, G. L. Foresti. // European Conference on Computer Vision. – 2015. – Pp. 191–208.
52. He L. Semi-Supervised Domain Generalizable Person Re-Identification / L. He, W. Liu, J. Liang [et al.] // ArXiv. – 2015. – 2108.05045v2.
53. Ma X. Person Re-Identification by Unsupervised Video Matching / X. Ma, X. Zhu, S. Gong [et al.]. // ArXiv. – 2016. – 1611.08512v3.
54. Chen Y., Person Re-identification using group context / Y. Chen, S. Duffner, A. Stoia [et al.] // HAL. – 2018. – hal-01895373.
55. Zhou K. Omni-Scale Feature Learning for Person Re-Identification // K. Zhou, Y. Yang, A. Cavallaro [et al.] // ArXiv. – 2019. – 1905.00953.
56. Quan R. Auto-ReID: Searching for a Part-Aware ConvNet for Person Re-Identification: materials from ICCV 2019 / R. Quan, X. Dong, Y. Wu [et al.] // ArXiv – 2019. – 1903.09776.
57. Zheng L. Person Re-identification: Past, Present and Future / L. Zheng, Y. Yang, A. G. Hauptmann. // ArXiv. – 2016. – 1610.02984v1.
58. Zhong Z. Unlabeled Samples Generated by GAN Improve the Person Re-identification Baseline in vitro / Z. Zhong, L. Zheng, Y. Yang. // ArXiv – 2017. – 1701.07717.
59. Zhong Z. Random Erasing Data Augmentation / Z. Zhong, L. Zheng, G. Kang [et al.]. // ArXiv – 2017. – 1708.04896v2.
60. Sun Y. SVDNet for Pedestrian Retrieval: materials from CVPR / Y. Sun, L. Zheng, W. Deng [et al.] // ArXiv. – 2017. – 1703.05693v1.
61. Zheng Z. Pedestrian Alignment Network for Large-scale Person Re-identification / Z. Zheng, L. Zheng, Y. Yang. // ArXiv. – 2017. – 1707.00408.
62. Hermans A. In Defense of the Triplet Loss for Person Re-Identification: materials from CVPR / A. Hermans, L. Beyer, B. Leibe. // ArXiv. – 2017. – 1703.07737.
63. Triplet Loss and Online Triplet Mining in TensorFlow // Olivier Moindrot blog. GitHub. 2018. URL: <https://omoindrot.github.io/triplet-loss> (Дата обращения 24.07.2022).
64. He T. Dense Interaction Learning for Video-based Person Re-identification / T. He, X. Jin, X. Shen [et al.]. // ArXiv. – 2021. – 2103.09013v3.

65. Wieczorek M. On the Unreasonable Effectiveness of Centroids in Image Retrieval / M. Wieczorek, B. Rychalska, J. Dabrowski. // ArXiv. – 2021. – 2104.13643v1.
66. Varior R.R. A Siamese Long Short-Term Memory Architecture for Human Re-Identification: materials from CVPR / R.R. Varior, B. Shuai, J. Lu [et al.] // ArXiv. – 2016. – 1607.08381v1.
67. Zheng L. Person Re-identification: Past, Present and Future / L. Zheng, Y. Yang, A. G. Hauptmann. // ArXiv. – 2016. – 1610.02984v1.
68. Deep Learning – 018 The re identification problem in computer vision // Alexey Artemov. 2021. URL: https://www.youtube.com/watch?v=_DqF4233m5Y&list=WL&index=13&t=2s (Дата обращения 24.07.2022).
69. Tan S. Dense invariant feature based support vector ranking for person re-identification / S. Tan, F. Zheng, L. Shao. // IEEE Transactions on Circuits and Systems for Video Technology. – 2015. – Pp. 687-691.
70. Yin J. Unsupervised Person Re-identification via Simultaneous Clustering and Consistency Learning / J. Yin, J. Qiu, S. Zhang [et al.]. // ArXiv. – 2021. – 2104.00202v1.
71. Li M. Unsupervised Person Re-identification by Deep Learning Tracklet Association / M. Li, X. Zhu, S. Gong. // ECCV – 2018. – Pp. 772–788.
72. Munjal B. Query-guided End-to-End Person Search / B. Munjal, S. Amin, F. Tombari [et al.]. // ArXiv. – 2019. – 1905.01203.

ПРИЛОЖЕНИЕ А

(справочное)

Implementation of methods for classifying people by gender and age and their re-identification in a video stream using deep learning technologies

Студент

Группа	ФИО	Подпись	Дата
8ПМ11	Семенюта Антон Вадимович		

Руководитель ВКР

Должность	ФИО	Ученая степень, звание	Подпись	Дата
Доцент	Аксенов Сергей Владимирович	к.т.н.		

Консультант-лингвист отделения иностранных языков ШБИП

Должность	ФИО	Ученая степень, звание	Подпись	Дата
Старший преподаватель	Куркан Наталия Владимировна	к.ф.н.		

1 Literature Review

1.1 Theoretical studies of gender and age recognition methods

1.1.1 Methods for gender and age recognition by facial image

Facial recognition is a popular and accurate approach to determining gender and age from photos or videos. It is due to the presence on the face of all visual features that allow you to determine gender, age, race, mood and many other human features [1]. In the video, facial recognition is preferable for determining gender and age in cases where a person's face is clearly visible in the frames.

There are three stages that can be distinguished in the sequence of actions to solve the described problem (Figure 1):

- Identification of people in the video stream and the construction of detections;
- Face detection;
- Recognition of gender and age by the face.

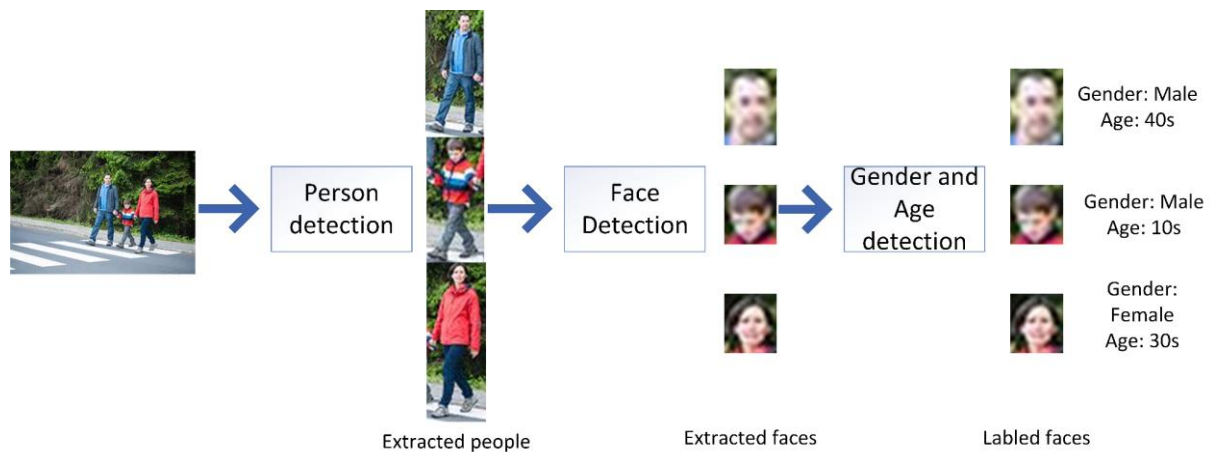


Figure 1. The steps in determining gender and age

1.1.1.1 Face detection

Face detection is a fairly well-known task that has many different approaches to solve [2]. The system requires you to limit the search area of the person. This is necessary for the possibility of unambiguous comparison of the detected face with a certain person. Also, this solution will speed up the development and overall system performance, since there is no need to write and test an additional module for comparing the limiting windows of detected faces and the limiting windows of persons. The approaches to the detecting a face in the frame are described below.

Classical methods for determining faces

The approaches widely used in the early 2000s included classical statistical methods, principal component analysis (PCA), linear and nonlinear classifiers [3]. There were methods that were also used to display the image on a plane and found the degree of proximity of the current

detected object with the face reference vector measuring the Mahalanobis distance [4]. Despite a good theoretical basis for the corresponding methods, they were hardly used in real industrial applications. The problem was the extremely low ability of such approaches to adapt to the invariance of the environment.

Methods based on the use of neural networks

Deep convolutional neural networks (deep CNN) have proven themselves well in the tasks of detecting objects in images. They are able to determine the characteristic features of an object directly from the pixels of the image [5]. The most well-known architectures of convolutional neural networks include AlexNet, GoogLeNet, VGGNet, ResNet. CNN training requires a huge amount of data, so to apply them to narrowly focused tasks, the models pre-trained on known datasets are usually used, and then, using the transfer learning method, weights are calculated on the last layers of the network. CNN-based approaches are the most accurate and at the same time the most demanding for computing capability.

Methods based on Haar and Gabor wavelets

Wavelets helped to obtain the strongest signals from the image, actually highlighting the most important features in the image, discarding the rest. Gabor wavelets have been widely used in research on facial recognition algorithms [6; 7]. Various wavelets were used with the idea of extracting visual features regardless of the task. Furthermore, these visual features were used in the author's algorithms, including for finding a face, as, for example, in the Viola-Jones algorithm [8]. The use of features obtained by the Gabor wavelet is complex and time-consuming [9]. The existing simplified wavelets and algorithms based on them work faster, but they strongly depend on lighting and other external factors [10].

Methods based on finding face descriptors

Instead of using the entire image, the algorithms divide it into subdomains in each of which histograms are calculated and find the most striking and important features for the actual problem [11]. In the future, the obtained features can be used both for the face recognition and for the identification, if there is an example for which it is also possible to obtain features in a similar way. For example, the authors used an approach based on local descriptors, including various local binary templates (LBP) for the face recognition in [12]. At the same time, there were such descriptors that were resistant to age-related changes in facial features. These methods are compact and work quickly, handle well with changes in lighting and various emotions on the face. However, these approaches are very sensitive to image noise.

Methods based on three-dimensional images and depth maps

Many modern cameras and even smartphones are equipped with lidars and have the ability to quickly and accurately obtain a depth map of the image [13]. This method of shooting allows

you to significantly reduce the sensitivity of algorithms to lighting and external factors of shooting. Models working with three-dimensional images are much more accurate and invariant to changes than those working with two-dimensional ones [14]. The methods use image textures [15], compensate for variable lighting [16], and compensate for different human poses by striving for a symmetrical face in the analysis [17]. Such systems require special equipment, careful calibration and data synchronization, which makes them difficult to use in general-purpose systems.

1.1.1.2 Gender and age recognition

Gender determination is a binary classification problem, and the age determination problem can be solved as a multiclass classification by age clusters, and as a regression prediction of the number of years. For example, in order to determine the age, the authors of [18] retrained the VGG-face model by transferring training on a dataset with images of people, for each of whom there was an age label. There were 8 outputs corresponding to the age groups on the output layer of the neural network. In order to determine the age, CNN (Ranking-CNN) was also used in [19], but with a binary output, and the labeling of a person's age occurred by aggregation of the outputs of the “basic” neural networks. Each of the basic convolutional networks was trained on the labels of a person belonging to the group “older” or “younger” than a certain age.

While methods not based on neural networks have sufficiently good accuracy and can compete with CNN for the task of face recognition, convolutional networks are unsurpassed leaders for the task of determining the age by face. So, in 2015, the AgeNet network was presented [20], demonstrating high accuracy in the tasks of determining age by face and winning a prize in the “ChaLearn 2015 apparent age competition” [21]. The approach is a deep convolutional neural network in which both the Euclidean norm and cross-entropy are used as a loss function. Most modern methods for determining gender and age use well-known architectures of deep convolutional neural networks, such as VGG-16 [1; 22] or ImageNet [19].

1.1.2 Methods of gender and age recognition based on silhouette and movement characteristics

The demand for the gender and age recognition based on movement or silhouette arises in situations where the person's face in the frames is poorly distinguishable due to the camera distance, or is not visible at all. Such situations are typical for monitoring systems installed in crowded places (for example, stadiums, airports, train stations) in order to cover as much space as possible on video.

All gender and age recognition algorithms use the results of object detection and tracking stages. The gender recognition is usually presented as a binary classification task, and the age recognition is either a regression task [23] or a multiclass classification task [24] when age groups are distinguished (for example, 0-10, 11-20, 21-30, etc.).

The key feature that is used in approaches to determining gender and age based on features of body shape or gait is GEI (Gait Energy Image). GEI is an image that provides dynamic and static information about a person's gait. This is one of the ways to obtain biometric data at a distance. GEI usually reflects the spatio-temporal characteristics of gait, such as stride length, stride speed, stride width, stride time, mach time, support time (average gait characteristics) and their corresponding variability and asymmetry (dynamic gait characteristics). An example of a GEI image is shown in Figure 2.



Figure 2. Example of a GEI image

GEI-based approaches show high quality metrics both in gender determination [25; 26] and in age determination [27].

Let's describe the main stages of GEI-based approaches. The input data are the images of the bounding boxes of objects during the observation obtained at the detection and tracking stages. The main stages of the approaches are listed below.

- Cropping and scaling bounding windows;
- Object segmentation and image binarization;
- Classification by observation angles (optional);
- GEI Design;
- Determination of gender and age.

1.1.2.1 Cropping and scaling of bounding windows

At this stage objects are cropped according to the coordinates of the corresponding bounding windows.

After receiving the cropped image of the object, it is scaled. In [26] all images are reduced to a format of 224x224 pixels. In [25] – 240x240 pixels. In order not to distort the proportions of the picture, it is accepted to add black paddings on the sides.

In the case of low image quality, additional approaches are used to increase the resolution. A two-stage CNN-based image restoration method was applied in [28]. During the first step, noise and blur are removed from the image using the Image Restoration CNN (IRCNN) model. At the second step, image resolution is improved (Super-Resolution Reconstruction, SR) using Very deep convolution networks SR (VDSR).

1.1.2.2 Segmentation of objects and binarization of images

At this stage the image is segmented by background subtraction. Methods for background subtraction can be divided into two groups: (1) methods based on neural networks and (2) methods that do not use neural networks. The last ones were widely used until the 2010s, before the appearance of modern convolutional neural network architectures. For example, an effective approach for determining the background and removing shadows on a black-and-white video is proposed in [29]. Local spatial coherence is used to determine moving objects. The approach is based on the assumption that the background pixels on a black-and-white video captured by a static camera have a normal distribution. Then the pixels of the moving object will create outliers in the distribution. Therefore, a metrically averaged average is used to determine the distribution parameters. After the initial determination of the background, morphological operators are used to remove shadows and isolated approach errors.

The most effective approaches to background subtraction are approaches based on convolutional neural networks. The first researchers to apply this approach to the task of background selection were Braham and Van Droogenbroeck. The neural network they built is described in [30] and is called ConvNet. To train the model, a data set with CDnet 2014 videos was used as a benchmark for competitions [31; 32].

The authors of the article [33] used the difference between the reference background image (before the object appeared) and the current image and created their own implementation of the background subtraction algorithm. Images from an infrared camera were used to adapt to dynamic changes in the scene.

After separating the object from the background, the image is binarized: the pixels of the image belonging to the object are colored white, and the rest are black.

The result of the stage is a sequence of cropped and scaled binary silhouette images (usually a white silhouette of an object on a black background).

1.1.2.3 Classification by observation angles

The need for this stage depends on the specifics of the place where the surveillance is installed. People tend to change the direction of movement for a short period of time, turning to the camera from different sides in places like shopping malls. To describe this type of movement, there is the term “freestyle walk” – movement in which the observed object changes the direction of movement and, accordingly, the angle of movement relative to the observer during the entire time interval of observation. It is the opposite of walking in one direction (fixed-direction walk) [26].

This stage is designed to increase the accuracy of gender and age recognition during freestyle walk. Instead of combining all the silhouette images extracted from the entire walk into

a GEI characteristic, silhouette images are classified into groups in accordance with the direction of movement of the object relative to the camera (viewing angle) during this stage. The construction of several GEIS (one GEI for one group of silhouette images) effectively shows the unique characteristics of a person. This is because it reduces interference from silhouette images taken from different viewing angles. In addition, some silhouette images may not reflect the entire body of the subject due to the limited field of view of the camera. These partial silhouette images of the body are considered noise. The exclusion of such silhouette images from GEI construction results in higher GEI quality.

The paper [26] presents a classification model to categorize the silhouette of an image into 10 groups of observation angles. Each group represents a range of 36 degrees ($360^\circ/10$) of the viewing angle. A special group is also used for partial body silhouette images that should not be involved in obtaining GEI. VGG16 is used as a classification model. The activation function of the output layer is softmax for 11 classes. The study [34] also uses the definition of the angle before constructing the GEI. For this purpose, the authors use CNN trained on CASIA-(B) [35], OU-ISIR [36], and OU-MVLP [37] datasets.

1.1.2.4 GEI design

After classifying the silhouette images of the walk into groups corresponding to the observation angles, the silhouette images in each group are combined into averaged images (GEI). For each GEI, Formula 1 determines its value at position (x, y).

$$GEI(x, y) = \frac{1}{N} \sum_{t=1}^N I_t(x, y), \quad (1)$$

where N is the number of silhouette images in the group, $I_t(x, y)$ is the binary value written in the image pixel at position (x, y) of the i-th silhouette, where 1 – pixel belongs to the silhouette, 0 – pixel does not belong to the silhouette.

The same formula is used in the case when the step of classification by observation angles was skipped. In this case, the final image will have one channel.

After GEI was compiled, the use of local binary patterns with fuzzy logic (Fuzzification of Local Binary Patterns) helped to improve the quality of the gender recognition model in [25]. The method helped to reduce the noise impact on binary patterns and create more informative features by including fuzzy logic in the procedure for calculating the histogram of variable codes.

There are approaches using GEI modifications. For example, an approach is described in [38] where a characteristic representing gait in three-dimensional space (Gait Energy Volume or GEV) is calculated using depth images. This is a modification of GEI in three-dimensional space, where binary voxels serve as an analogue of binary silhouette pixels. Obtaining such images requires the presence of depth cameras.

The paper [38] also describes an approach where BGEI (Back-filled Gait Energy Image) is calculated – a gait representation using the assumption that the frontal surface of the model contains all information about gait. When using BGEI, there is a chance to still lose some information, since the hind leg is no longer represented in the calculation of the characteristic. However, in some cases such information may be unnecessary or make a minimal contribution to the accuracy of gender and age recognition. BGEI can be built both on the basis of GEI and on the basis of GEV. For the side view silhouettes, the BGEI is constructed using the method of backfilling binary silhouettes. To do this, there is a pixel that characterizes the frontal surface of the object on the profile projection in each line, and from it the line is filled to the end of the image. Filled binary silhouettes are aligned along the centroid of the frontal surface.

1.1.2.5 Gender and age recognition

Before applying classification or regression algorithms directly to recognize gender or age, the N GEI images constructed at the previous stages are combined into one image with N number of channels. If no classification has been made by observation angles, then the number of channels will be equal to one. The authors of [26] used image of 224x224x10 size with 10 GEI.

Classification or regression methods are applied after obtaining the resulting image. These can be both machine learning algorithms and neural networks. The authors of [39] used a convolutional neural network VGG16 for the task of determining gender in. In order to classify gender, the authors of [26] developed their CNN architecture, which has 7 layers. In [25] the method of support vectors with a linear kernel was applied.

The authors of the paper [24] tested several approaches to determine the age: both classification and regression methods. Age groups were allocated for classification (0-5, 6-10, 11-15, 16-20, 21-30, 31-40, 41-50, 51-60, and more than 60 years). Then, by calculating the L2 distance based on GEI between neighboring pairs of groups, and experimentally setting a threshold value for such a distance, the authors combined some age groups into one. There were five groups after merging. To classify images into groups, a modification of the support vector machine called DAGSVM was used. It allows expanding the use of classical SVM for the task of multiclass classification.

1.1.2.6 The approach based on longitudinal and transverse silhouette projections

GEI is a well-established characteristic in relation to determining the gender and age, reflecting the dynamic and figurative features of gait. In addition to GEI-based approaches, there are also other approaches that demonstrate high quality metrics for the age determination task. For example, the problem of binary classification of people into “young” and “elderly” was considered in [27]. Longitudinal and transverse silhouette projections (SLP and STP respectively) were used as characteristics to determine which group a person belongs to. To construct those ones, an

approach was firstly applied to localize the vertical position of body points without using special sensors. Secondly, the length of a person's step was analyzed during an video fragment of his gait.

Then, SLP and STP were built on binary silhouette images, followed by aggregation and obtaining the resulting descriptor.

The authors confirmed that this descriptor reflects the gait features of young and elderly people: a wide step, characteristic of young people, is manifested on the longitudinal projection, and a wider silhouette, usually characteristic of older people, is reflected on the transverse projection.

1.1.2.7 Approach using different types of cameras

Typical images of objects are limited in the ability to recognize gender and age due to information such as lighting, background, shadows, various types of clothing and accessories. Therefore, the study [28] used a combination of images from visible light cameras and infrared cameras based on the temperature difference between the human body and the background to determine gender. In this study, the authors trained two convolutional neural networks ResNet-101 with an input image resolution of 197x447 pixels, taking into account human proportions. To make a decision about the gender, the weights of the models on the last layers were first normalized, then aggregated to obtain the resulting indicator. If the value of the indicator exceeded the experimentally selected threshold value, then the person in the image was classified as a male subject, otherwise as a female.

1.2 Theoretical studies of re-identification methods

1.2.1 Problem statement

The development of individual prints of objects based on graphical features is used to ensure the ability to remember an object and recognize it again both on subsequent frames of the video stream (to improve the quality of the tracking algorithm), and when this object appears in the video stream after a long period of time. So, by creating such prints, the functionality of repeated anonymous recognition of objects is implemented without the need to have a pre-filled database of some features of objects. In most sources, that is called the task of object re-identifying in a video stream [40; 41]. Further, both terms will be used to denote this task.

The task of object re-identification (Person / Target / Object re-identification, ReID) is defined as the task of searching for object visual descriptions in a gallery, according to the requested subset of visual descriptions of the same object taken from other scenes and/or at other points in time on the same scene of the video stream. Gallery is a set of the same descriptions of object sets observed in the video stream. The task also involves ranking descriptions of gallery objects in descending order of similarity with the input/requested object [42; 43].

1.2.2 Limitations in solving the problem

The most accurate methods of repeated recognition of people are based on biometric features that can be extracted from the graphical information of displaying objects in the video stream (face image, gait and silhouette features) [44]. However, these same methods can fail due to the external impacts: mutual overlap of people, different weather conditions and lighting conditions, different distance from the camera to objects, different quality of video material. Such conditions can create significant difficulties for extracting graphic prints of objects based on biometrics. Considering that for the high-quality operation of re-identification methods, graphic features must be sufficiently different in order to successfully distinguish identical-looking objects, and at the same time carry general information in order to be replicated under different conditions, then graphic features determined by the appearance of objects are more preferable for use in re-identification algorithms [45; 46]. Graphical features based on biometric indicators of a person can be used to clarify the results of re-identification only in cases when it is possible to isolate the relevant data on video surveillance cameras. The decision to consider these methods for the formation of individual prints is made in the next stage of work.

Methods of repeated recognition (re-identification) of objects using external visual features of objects without biometric information as a graphic fingerprint are considered below.

1.2.3 The stages of re-identification algorithms

All algorithms of re-identification based on the fingerprint of graphic features work with the results of detection methods, i.e. with bounding windows of objects. The initial platform, which includes the YOLOv4 object set detector, fully meets the initial requirements for the implementation of such algorithms. In general, the pipeline of re-identification algorithms based on the graphical characteristics of the object is divided into 3 main stages [45]:

1. Extraction of graphical features of an object from a bounding box on a video stream frame – obtaining a graphical fingerprint;
2. Calculation of the similarity descriptor of graphic prints – obtaining metrics for searching in the gallery;
3. Ranking the gallery in descending order of similarity of the graphic prints of objects in it and searching for the closest combination to get the result – extracting from the gallery the most suitable for the input graphic description of the object.

Next, the methods used at various stages of the re-recognition algorithms will be discussed in detail.

1.2.4 The stage of extraction of graphic features

At this stage, the calculation of the graphic prints of the object for which it is necessary to search in the gallery is carried out. Usually, such a fingerprint is a certain feature vector obtained from an external graphical representation of an object on a frame or frames of a video stream.

Color features in the form of color histograms [47; 48; 49], as well as texture features (GLCM – Gray Level Co-occurrence Matrices, LBP – Local Binary Patterns, Haralik, Laws, and others) [47; 49; 50] were among the first to be used to extract graphical prints of objects. These methods made it possible to quite successfully determine the similarity of objects during re-identification on a variety of cameras (using ready-marked data sets), as well as distinguish objects in a crowd. However, these methods strongly depended on the environmental conditions, lighting and the quality of the video material – they could not be easily replicated to the general conditions inherent in video surveillance cameras.

One of the developments of the methods mentioned above is the struggle with the limitations of the external environment. For example, this happens by color normalization, which allows you to smooth out the influence of unstable lighting and shading of objects in images from different cameras [47]. In addition, for the tasks of multicameral object re-identification, brightness transfer methods are used between several cameras, which makes it possible to average the ambient lighting conditions between the compared prints of graphic features [48; 50]. They improve the overall quality of the re-identification algorithm and expand the applicability of the proposed methods for extracting graphic features for the task of repeated recognition.

As additional information for prints of graphic features, designed to improve the quality of re-identification of objects of certain classes, the approach of highlighting areas of importance on objects of the human class is used [51]. To highlight such areas, an algorithm is used on graphs, which allows you to select a map of groups of pixels that differ significantly from the other pixels surrounding them. The selected areas are used as weights of graphical features describing the bounding window of the object. Another approach that is widely used to graphically describe people in re-recognition algorithms is the use of shape features [47; 49; 52]: as additional information describing a graphic fingerprint, information about the size of objects or data about the pose taken by a person is used. This feature enrichment makes it easier to compare objects on closely spaced frames of one camera or on frames of multiple cameras. An even more advanced method of describing the graphical features of a human class object on a set of video stream frames is the use of GEI (Gait Energy Image) features that combine dynamic and statistical information about the movement of an object within a tracklet [53].

Another direction of introducing supportive information into the prints of objects based on graphical features is to take into account the environment in which the object is located. Methods

are mentioned in [45; 52], in addition to the features of the object itself, use graphical features of the scene context and rely on them to obtain more accurate information about the location of the object in different time periods or on multiple cameras. A more advanced approach is used in the work [54], where instead of the features of only one person class object, the features of a group of people are considered, next to which this person is located. This approach is based on the assumption that a person moving in a video stream on neighboring frames or different cameras will be part of a group of people for a relatively long period of time. At the same time, at the subsequent stages of the algorithm, the objects are compared in a multi-step manner - first a group of people is re-identified, and then each person in this group.

The most popular and at the same time the most accurate methods of extracting graphic prints for repeated recognition of objects today are neural network algorithms that use deep learning to calculate the most informative features from the bounding boxes. The most successful neural network architectures for such tasks are based on convolutional layers of popular architectures, such as VGG16 or ResNet [55; 56; 42], which, having previously trained weights on large datasets, after additional fine tuning on the marked-up data of the re-identification task, show some of the best results under any external conditions. Thus, the main research at the moment is aimed at improving the quality of the basic architectures of neural networks by fine-tuning them [56; 57], generating realistic images to expand the training sample [57], augmenting training data [58], developing special layers for the task of re-identification [59; 60] and stable error functions such as triplet-loss [61; 62]. These functions allow training models simultaneously on similar and significantly different data of graphic prints of objects.

In modern models showing the best quality of solving the problem of re-identification – SOTA solutions (SOTA - State Of The Art) – approaches of separate extraction of graphic features from various parts of the bounding box are used. Such approaches makes it possible to successfully deal with problems of different scale of objects (objects moving away and approaching in the video stream), different shooting conditions and various the quality of videos [63; 64; 55]. In addition, a generalization of methods for extracting graphical features for video data is applied: graphical prints of objects are extracted not from a single frame, but are aggregated based on information about a set of frames for moving an object [63]. Another case – a set of frames for moving an object from a tracklet is used as input data for training neural networks with recurrent layers [65].

1.2.5 The stage of calculating the similarity descriptor

At this stage of the pipeline of re-identification algorithms, the extracted graphic features of the object are used to calculate some metric of similarity with similar graphic features of the gallery. The calculated metric is the main tool for further searching for similar objects in the gallery.

The most popular approaches for creating a similarity descriptor are metrics for estimating the distance between two vectors: L2-norm (Euclidean distance) or the cosine distance [45]. These metrics are one of the basic mathematical norms for calculating the similarity of vectors, which have proven themselves so well in solving computer vision problems and, in particular, re-identification, that they are still used, including in SOTA solutions [55].

Another approach that is also quite common at the moment is the use of deep machine learning models. For example, Siamese neural networks are used to calculate similarity metrics [67], which allow calculating a similarity descriptor for them based on two vectors of graphical features of objects presented to the input. The development of Siamese neural networks for the task of re-identifying objects in a video stream is the use of recurrent neural network architectures such as Long Short-Term Memory (LSTM) to extract graphical features in time dependence with simultaneous training based on these features of a neural network to calculate a similarity descriptor [66; 68].

One of the most advanced approaches to calculating the similarity metric is the neural network architecture of [64]. It uses the Attention Mechanism to calculate the similarity metric of graphical features of an object. It allows calculating a set of spatiotemporal metrics between a set of input features of an object and a set of features from the gallery when training the network. This makes it possible to significantly improve the quality of repeated recognition of objects in different time periods or on video streams from multiple cameras [64].

1.2.6 The ranking stage

At this final stage of the algorithms for repeated recognition of objects by graphical features, based on the calculated similarity metric, the feature vectors presented in the gallery are ranked and the most similar object (if any) is determined based on the first K vectors of the gallery, where K is the hyperparameter of the algorithm.

Ranking methods are divided into several types according to the method of comparing the input set of features (query of images) with the gallery and according to the method of training the model to compare these sets with the gallery. Algorithms that use only one feature vector for comparison and ranking and also rank by one feature vector from the gallery form a subset of one-shot approaches. Algorithms that compare and rank sets of feature vectors are multi-shot approaches. By the way of training, algorithms are divided into those that compare and rank marked-up examples of a training sample during training (training with a teacher), as well as those that during training only learn to identify strong features that separate different objects well and allow comparing the same objects (training without a teacher). The comparison itself and ranking is carried out using non-neural network approaches in the last case [45].

For the ranking stage, re-identification algorithms offer both the use of more optimal approaches to ordering the gallery based on similarity metrics, and simplification of the ordering algorithm itself by grouping feature vectors in the gallery. Typical methods by which the ranking and comparison of the input vector of features with the features from the gallery is carried out are the algorithms of full search, search for nearest neighbors (for example, k-nearest neighbors) and ranking directly (for example, Ranking SVM) [45]. More advanced approaches include an algorithm [69] where the method of invariant ranking of support vectors (Invariant Ranking SVM) is used, offering training of the SVM algorithm on combined graphical features of objects to ensure invariance of ranking relative to various poses, shapes, types of representation of objects and camera angles. In addition to modifications of the ranking itself, a neural network is used in [70] to cluster objects into semantic groups based on features of objects from individual frames and changes in these features over time in the video stream. Dividing objects into groups allows you to improve the quality and optimize the ranking process – first groups are selected, then objects in these groups.

Modern ranking methods that implement teacher-based learning approaches are most often neural network methods mentioned earlier [64; 65], or approaches that implement all stages of re-identification algorithms in one model [71; 72] - end-to-end approaches. The latter method implements automatic acquisition of object tracklets, extraction of graphical features and comparison of them with the features of tracklets of other objects in the data (multi-shot approach). The method generalizes the approaches of using marked-up data for the task of re-identification both in one video stream and on multiple cameras: the ranking and search for the most similar object tracklets are carried out simultaneously in several video streams. This approach provides the possibility of learning both on marked-up (supervised learning) and on unmarked data (unsupervised learning).

1.3 References

1. Apparent Age Estimation Using Ensemble of Deep Learning Models (Malli R.C. Apparent Age Estimation Using Ensemble of Deep Learning Models / R.C. Malli, M. Aygun, H.K. Ekenel // IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW). – 2016. – Pp. 714-721.
2. Hasan M.K. Human Face Detection Techniques: A Comprehensive Review and Future Research Directions / M.K. Hasan, M.S. Ahsan, Abdullah-Al-Mamun [et al.] // Electronics. – 2021. – Vol. 10(19). – Pp. 2354.
3. Naseem I. Linear Regression for Face Recognition / I. Naseem, R. Togneri, M. Bennamoun // IEEE Transactions on Pattern Analysis and Machine Intelligence. – 2010. – Vol. 32 (11). – Pp. 2106–2112.

4. Perlibakas V. Distance measures for PCA-based face recognition / V. Perlibakas // Pattern Recognition Letters. – 2004. – Vol. 25(6). – Pp. 711–724.
5. Liu W. A survey of deep neural network architectures and their applications / W. Liu, Z. Wang, X. Liu [at al.] // Neurocomputing. – 2017. – Vol. 234. – Pp. 11-26.
6. Gu W. Facial expression recognition using radial encoding of local Gabor features and classifier synthesis / W. Gu, C. Xiang, Y. Venkatesh // Pattern Recognition. – 2012. – Vol. 45(1). – Pp. 80-91.
7. Serrano S. Recent advances in face biometrics with Gabor wavelets: A review / S. Serrano, I. Diego, C. Conde // Pattern Recognition Letters. – 2010. – Vol. 31(5). – Pp. 372-381.
8. Chandrappa D.N. Face Detection Using a Boosted Cascade of Features Using OpenCV / D.N. Chandrappa, G. Akshay, M. Ravishankar // Wireless Networks and Computational Intelligence (ICIP 2012). – 2012. Vol. 292. – Pp. 399-404.
9. Serrano S. Analysis of variance of Gabor filter banks parameters for optimal face recognition / S. Serrano, I. Diego, C. Conde [at al.]. // Pattern Recognition Letters. – 2011. Vol. 31(15). – Pp. 1998-2008.
10. Chen J. WLD: a robust local image descriptor / J. Chen, S. Shan, C. He, [at al.]. // IEEE Trans. Pattern Anal. Mach. Intell. – 2010. – Vol. 32(9) – Pp. 1705–1720.
11. Bereta M. Local descriptors and similarity measures for frontal face recognition: a comparative analysis / M. Bereta, W. Pedrycz, M. Reformat // Journal of Visual Communication and Image Representation. – 2013. – Vol. 24(8) – Pp. 1213–1231.
12. Bereta M. Local descriptors in application to the aging problem in face recognition / M. Bereta, P. Karczmarek, W. Pedrycz, [at al.]. // Pattern Recognit. – 2013. – Vol. 46(10) – Pp. 2634–2646.
13. What is a LiDAR scanner, the iPhone 12 Pro's camera upgrade, anyway. // tecradar. 2021. URL: <https://www.techradar.com/news/what-is-a-lidar-scanner-the-iphone-12-pro-sumored-camera-upgrade-anyway> (Date of application 21.07.2022).
14. Kevin W. A survey of approaches and challenges in 3D and multi-modal 3D + 2D face recognition / W. Kevin, C. Kyong, F. Patrick // Computer Vision and Image Understanding. – 2006. – Vol. 101(1) – Pp. 1-15.
15. Blanz V. Face Recognition Based on Fitting a 3D Morphable Model / V. Blanz, T. Vetter // IEEE Trans. Pattern Anal. Mach. Intell. – 2003. – Vol. 25(9) – Pp. 1063–1074.
16. Zhang L. Face recognition from a single training image under arbitrary unknown lighting using spherical harmonics / L. Zhang, D. Samaras // IEEE Transactions on Pattern Analysis and Machine Intelligence. – 2006. – Vol. 28(3) – Pp. 351-363.

17. Passalis G. Using Facial Symmetry to Handle Pose Variations in Real-World 3D Face Recognition / G. Passalis, P. Panagiotis, T. Theoharis [at al.]. // IEEE Transactions on Pattern Analysis and Machine Intelligence. – 2011. – Vol. 33(10) – Pp. 1938–1951.
18. Zakariya Q. Deep Convolutional Neural Network for Age Estimation based on VGG-Face Model / Q. Zakariya, A. Arafat, D. Buket [at al.] // ArXiv. – 2017. – 1709.01664.
19. Chen S. Using Ranking-CNN for Age Estimation / S. Chen, C. Zhang, M. Dong [at al.] // 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). – 2017. – Pp. 742-751.
20. Liu X. AgeNet: Deeply Learned Regressor and Classifier for Robust Apparent Age Estimation / X. Liu, S. Li, M. Kan [at al.] // 015 IEEE International Conference on Computer Vision Workshop (ICCVW). – 2015. – Pp. 258-266.
21. Agbo-Ajala O. Deep learning approach for facial age classification: a survey of the state-of-the-art / O. Agbo-Ajala, S. Viriri // Artificial Intelligence Review. – 2021. – Pp. 179–213.
22. Antipov G. Apparent Age Estimation from Face Images Combining General and Children-Specialized Deep Learning Models / G. Antipov, M. Baccouche, S. Berrani [at al.] // 2016 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW). – 2016. – Pp. 801-809.
23. Makihara Y. Gait-based age estimation using a whole-generation gait database / Y. Makihara, M. Okumura, H. Iwama [at all] // Multimedia Tools and Applications. – 2018. – Vol. 77. – Pp. 28333–28354.
24. Li X. Gait-based human age estimation using age group-dependent manifold learning and regression / X. Li, Y. Makihara, C. Xu [at al.] // Multimedia Tools and Applications. – 2018. – Vol. 77. – Pp. 28333–28354.
25. El-Alfy E.M. Silhouette-Based Gender Recognition in Smart Environments Using Fuzzy Local Binary Patterns and Support Vector Machines / E. M. El-Alfy, A. G. Binsaadoon // Procedia Computer Science. – 2017. – Vol.109. – Pp. 164-171.
26. Kitchat K. Gender classification from gait silhouette using observation angle-based GEIs / K. Kitcha, N. Khamsemanan, C. Nattee // 2019 IEEE International Conference on Cybernetics and Intelligent Systems (CIS) and IEEE Conference on Robotics, Automation and Mechatronics (RAM). – 2018. – Pp. 485-490.
27. Mansouri N. Gait-based human age classification using silhouette model / N. Mansouri, A.I. Mohammed, B.J. Yousra // IET Biom. – 2018. – Vol. 7(2). – Pp. 116-124.
28. Baek N. Multimodal Camera-Based Gender Recognition Using Human-Body Image With Two-Step Reconstruction Network / N. Baek, S. W. Cho, J. H. Koo [at al.] // IEEE Access. – 2017. – Vol. 7. – Pp. 104025-104044.

29. Jung C.R. Efficient Background Subtraction and Shadow Removal for Monochromatic Video Sequences / C. R. Jung // IEEE Transactions on Multimedia. – 2009. – Vol. 11(3). – Pp. 571-577.
30. Braham M. Deep background subtraction with scene-specific convolutional neural networks / M. Braham, M.V. Droogenbroeck // 2016 International Conference on Systems, Signals and Image Processing (IWSSIP). – 2016. – Pp. 1-4.
31. Wang Y. CDnet 2014: an expanded change detection benchmark dataset / Y. Wang, P. Jodoin, F. Porikli [et al.] // 2014 IEEE Conference on Computer Vision and Pattern Recognition Workshops. – 2014. – Pp. 393-400.
32. Bouwmans T. Deep Neural Network Concepts for Background Subtraction: A Systematic Review and Comparative Evaluation / T. Bouwmans, S. Javed, M. Sultana [at al.] // Neural Networks. – 2019. – Vol. 117. – Pp. 8-66.
33. Dedeoğlu Y. Silhouette-Based Method for Object Classification and Human Action Recognition in Video / Y. Dedeoğlu, B.U. Töreyn, U. Güdükbay [at al.] // Lecture Notes in Computer Science. – 2006. – Vol. 3979. – Pp. 64-77.
34. Elharrouss O., Gait recognition for person re-identification / O. Elharrouss, N. Almaadeed, S. Al-Maadeed [et al.] // The Journal of Supercomputing. – 2021. – Vol. 7. – Pp. 3653-3672.
35. Zheng S. Robust view transformation model for gait recognition / S. Zheng, J. Zhang, K. Huang [at al.] // 2011 18th IEEE International Conference on Image Processing. – 2011. – Pp. 2073-2076.
36. Iwama H. The OU-ISIR Gait Database Comprising the Large Population Dataset and Performance Evaluation of Gait Recognition / H. Iwama, M. Okumura, Y. Makihara [at al.] // IEEE Transactions on Information Forensics and Security. – 2012. – Vol. 7(5). – Pp. 1511-1521.
37. Takemura N. Multi-view large population gait dataset and its performance evaluation for cross-view gait recognition / N. Takemura, Y. Makihara, D. Muramatsu [at al.] // IPSJ Transactions on Computer Vision and Applications. – 2018. – Vol. 10(4). – Pp. 1-14.
38. Sivapalan S. The Backfilled GEI – A Cross-Capture Modality Gait Feature for Frontal and Side-View Gait Recognition / S. Sivapalan, D. Chen, S. Denman [at al.] // 2012 International Conference on Digital Image Computing Techniques and Applications (DICTA). – 2012. – Pp. 1-8.
39. Liu T. Combining Convolutional Neural Network and Support Vector Machine for Gait-based Gender Recognition / T. Liu, X. Ye, B. Sun // 2018 Chinese Automation Congress (CAC). – 2018. – Pp. 3477-3481.

40. Li X. Object Re-Identification Based on Deep Learning / X. Li, Z. Zhou // Visual Object Tracking with Deep Neural Networks. – 2019.
41. What is reidentification? // KLAP – Augmenting fan experience for the sport industry. 2021. URL: <https://www.getklap.com/blog/what-is-reidentification> (Дата обращения 23.07.2022).
42. Ristani E. Features for Multi-Target Multi-Camera Tracking and Re-Identification / E. Ristani, C. Tomasi. // ArXiv – 2018. – 1803.10859.
43. Person Re-Identification // Papers With Code. 2022. URL: <https://paperswithcode.com/task/person-re-identification> (Дата обращения 24.07.2022).
44. Dvoenosova G. Biometry as science, a method and a documentation approach [Electronic resource] / G. Dvoenosova, M. Dvoenosova // Personal Management. – 2009. – № 11. – URL: <https://www.top-personal.ru/issue.html?2039> (Дата обращения 23.07.2022).
45. Person Re-Identification Shaogang Gong [Electronic resource] / Shaogang Gong [et al.]. – Springer. – 2014. – 444 p.
46. Zhou K. Learning Generalisable Omni-Scale Representations for Person Re-Identification / K. Zhou, Y. Yang, A. Cavallaro [et al.] // IEEE Transactions on Pattern Analysis and Machine Intelligence. – 2021. – Pp. 1-1.
47. Cai Y. Exploring context information for inter-camera multiple target tracking / Y. Cai, G. Medioni. // IEEE Winter Conference on Applications of Computer Vision. – 2014. – Pp. 761-768.
48. Das A. Consistent Re-identification in a Camera Network / A. Das, A. Chakraborty, A.K. Roy-Chowdhury // Lecture Notes in Computer Science. – 2014. – Vol. 8690 – Pp. 330-345).
49. Chen, X. Multitarget Tracking in Nonoverlapping Cameras Using a Reference Set / X. Chen, L. An, B. Bhanu // IEEE Sensors Journal. – 2015. – Pp. 2692-2704.
50. Zhang S. A Camera Network Tracking (CamNeT) Dataset and Performance Baseline / S. Zhang, E. Staudt, E. Faltemier [et al.] // IEEE Winter Conference on Applications of Computer Vision. – 2015. – 365-372.
51. Martinel N. Saliency Weighted Features for Person Re-Identification / N. Martinel, C. Micheloni, G. L. Foresti. // European Conference on Computer Vision. – 2015. – Pp. 191–208.
52. He L. Semi-Supervised Domain Generalizable Person Re-Identification / L. He, W. Liu, J. Liang [et al.] // ArXiv. – 2015. – 2108.05045v2.
53. Ma X. Person Re-Identification by Unsupervised Video Matching / X. Ma, X. Zhu, S. Gong [et al.]. // ArXiv. – 2016. – 1611.08512v3.
54. Chen Y., Person Re-identification using group context / Y. Chen, S. Duffner, A. Stoia [et al.] // HAL. – 2018. – hal-01895373.

55. Zhou K. Omni-Scale Feature Learning for Person Re-Identification // K. Zhou, Y. Yang, A. Cavallaro [et al.] // ArXiv. – 2019. – 1905.00953.
56. Quan R. Auto-ReID: Searching for a Part-Aware ConvNet for Person Re-Identification: materials from ICCV 2019 / R. Quan, X. Dong, Y. Wu [et al.] // ArXiv – 2019. – 1903.09776.
57. Zheng L. Person Re-identification: Past, Present and Future / L. Zheng, Y. Yang, A. G. Hauptmann. // ArXiv. – 2016. – 1610.02984v1.
58. Zhong Z. Unlabeled Samples Generated by GAN Improve the Person Re-identification Baseline in vitro / Z. Zhong, L. Zheng, Y. Yang. // ArXiv – 2017. – 1701.07717.
59. Zhong Z. Random Erasing Data Augmentation / Z. Zhong, L. Zheng, G. Kang [et al.]. // ArXiv – 2017. – 1708.04896v2.
60. Sun Y. SVDNet for Pedestrian Retrieval: materials from CVPR / Y. Sun, L. Zheng, W. Deng [et al.] // ArXiv. – 2017. – 1703.05693v1.
61. Zheng Z. Pedestrian Alignment Network for Large-scale Person Re-identification / Z. Zheng, L. Zheng, Y. Yang. // ArXiv. – 2017. – 1707.00408.
62. Hermans A. In Defense of the Triplet Loss for Person Re-Identification: materials from CVPR / A. Hermans, L. Beyer, B. Leibe. // ArXiv. – 2017. – 1703.07737.
63. Triplet Loss and Online Triplet Mining in TensorFlow // Olivier Moindrot blog. GitHub. 2018. URL: <https://omoindrot.github.io/triplet-loss> (Дата обращения 24.07.2022).
64. He T. Dense Interaction Learning for Video-based Person Re-identification / T. He, X. Jin, X. Shen [et al.]. // ArXiv. – 2021. – 2103.09013v3.
65. Wiecezorek M. On the Unreasonable Effectiveness of Centroids in Image Retrieval / M. Wiecezorek, B. Rychalska, J. Dabrowski. // ArXiv. – 2021. – 2104.13643v1.
66. Varior R.R. A Siamese Long Short-Term Memory Architecture for Human Re-Identification: materials from CVPR / R.R. Varior, B. Shuai, J. Lu [et al.] // ArXiv. – 2016. – 1607.08381v1.
67. Zheng L. Person Re-identification: Past, Present and Future / L. Zheng, Y. Yang, A. G. Hauptmann. // ArXiv. – 2016. – 1610.02984v1.
68. Deep Learning – 018 The re identification problem in computer vision // Alexey Artemov. 2021. URL: https://www.youtube.com/watch?v=_DqF4233m5Y&list=WL&index=13&t=2s (Дата обращения 24.07.2022).
69. Tan S. Dense invariant feature based support vector ranking for person re-identification / S. Tan, F. Zheng, L. Shao. // IEEE Transactions on Circuits and Systems for Video Technology. – 2015. – Pp. 687-691.

70. Yin J. Unsupervised Person Re-identification via Simultaneous Clustering and Consistency Learning / J. Yin, J. Qiu, S. Zhang [et al.]. // ArXiv. – 2021. – 2104.00202v1.
71. Li M. Unsupervised Person Re-identification by Deep Learning Tracklet Association / M. Li, X. Zhu, S. Gong. // ECCV – 2018. – Pp. 772–788.
72. Munjal B. Query-guided End-to-End Person Search / B. Munjal, S. Amin, F. Tombari [et al.]. // ArXiv. – 2019. – 1905.01203.

ПРИЛОЖЕНИЕ Б

(справочное)

Скан свидетельства о государственной регистрации программы для ЭВМ No2023614799

«Модуль распознавания пола и возраста людей Visius Demographic Recognition».

РОССИЙСКАЯ ФЕДЕРАЦИЯ



СВИДЕТЕЛЬСТВО
о государственной регистрации программы для ЭВМ
№ 2023614799

**Модуль распознавания пола и возраста людей Visius
Demographic Recognition**

Правообладатель: **Общество с ограниченной
ответственностью «Рубиус Групп» (RU)**

Авторы: **Костин Кирилл Александрович (RU), Кудинов
Антон Викторович (RU), Чурсина Елена Андреевна (RU),
Семенюта Антон Вадимович (RU)**

Заявка № **2023613929**
Дата поступления **06 марта 2023 г.**
Дата государственной регистрации
в Реестре программ для ЭВМ **06 марта 2023 г.**



Руководитель Федеральной службы
по интеллектуальной собственности

документ подписан электронной подписью
Сертификат: 68c8007b14e321b194edbd24145d5c7
Владелец: **Зубов Юрий Сергеевич**
Действителен с 06.03.2022 по 26.05.2023

Ю.С. Зубов

ПРИЛОЖЕНИЕ В

(справочное)

Скан свидетельства о государственной регистрации программы для ЭВМ 2023614800

«Модуль повторного распознавания людей Visius Person Re-Identification».

РОССИЙСКАЯ ФЕДЕРАЦИЯ



СВИДЕТЕЛЬСТВО
о государственной регистрации программы для ЭВМ
№ 2023614800

**Модуль повторного распознавания людей Visius Person
Re-Identification**

Правообладатель: **Общество с ограниченной
ответственностью «Рубиус Групп» (RU)**

Авторы: **Костин Кирилл Александрович (RU), Кудинов
Антон Викторович (RU), Чурсина Елена Андреевна (RU),
Семенюта Антон Вадимович (RU)**

Заявка № **2023613930**
Дата поступления **06 марта 2023 г.**
Дата государственной регистрации
в Реестре программ для ЭВМ **06 марта 2023 г.**



Руководитель Федеральной службы
по интеллектуальной собственности

Ю.С. Зубов

документ подписан электронной подписью
Сертификат 68b80027614e9210294edcd24145d5c7
Владимир Зубов Юсиф Сергеевич
Действителен с 2012 по 26.05.2023

ПРИЛОЖЕНИЕ Г

(справочное)

Скан Акта No 2 внедрения программы для ЭВМ «Модуль распознавания пола и возраста людей Visius Demographic Recognition».

Утверждаю.

Генеральный директор

управляющей организации ООО «Рубиус»

«03» апреля 2023 г.



Акт № 2 внедрения программы для ЭВМ

(изобретения, полезной модели, промышленного образца, программы для ЭВМ, базы данных, секрета производства (ноу-хау))

в ООО «Рубиус Групп» от «03» апреля 2023 г.

Настоящий Акт составлен в том, что нижеуказанное(ая) программа для ЭВМ

(изобретение, полезная модель, промышленный образец, программа для ЭВМ, база данных, секрет производства (ноу-хау))

Модуль распознавания пола и возраста людей Visius Demographic Recognition

(название объекта интеллектуальной собственности)

Автор(ы):

1. Костин Кирилл Александрович

2. Кудинов Антон Викторович

3. Чурсина Елена Андреевна

4. Семенюта Антон Вадимович

(указывается фамилия, имя, отчество каждого автора)

Свидетельство о государственной регистрации № 2023614799 от 06 марта 2023 г. (дата регистрации в Государственном реестре программ для ЭВМ и баз данных)

Патентообладатель(ли)/правообладатель (и): ООО «Рубиус Групп»

внедрено(а) с «03» апреля 2023 г.

при разработке web-сервиса Visius для подсчёта пешеходного трафика по видео, который помогает быстрее подбирать помещения при открытии новых торговых точек.

Операционный директор

управляющей организации «Рубиус»

(по доверенности № 01 от 01.01.2023 г.)

А.Е. Васильева

ПРИЛОЖЕНИЕ Д

(справочное)

Скан Акта No 3 внедрения программы для ЭВМ «Модуль повторного распознавания людей Visius Person Re-Identification».

Утверждаю.

Генеральный директор
управляющей организации ООО «Рубиус»

А.В. Кудинов

«03» апреля 2023 г.



Акт № 3 внедрения программы для ЭВМ

(изобретения, полезной модели, промышленного образца, программы для ЭВМ, базы данных, секрета производства (ноу-хау))
в ООО «Рубиус Групп» от «03» апреля 2023 г.

Настоящий Акт составлен в том, что нижеуказанное(ая) программа для ЭВМ

(изобретение, полезная модель, промышленный образец, программа для ЭВМ, база данных, секрет производства (ноу-хау))

Модуль повторного распознавания людей Visius Person Re-Identification

(название объекта интеллектуальной собственности)

Автор(ы):

1. Костин Кирилл Александрович

2. Кудинов Антон Викторович

3. Чурсина Елена Андреевна

4. Семенюта Антон Вадимович

(указывается фамилия, имя, отчество каждого автора)

Свидетельство о государственной регистрации № 2023614800 от 06 марта 2023 г. (дата регистрации
в Государственном реестре программ для ЭВМ и баз данных)

Патентообладатель(ли)/правообладатель (и): ООО «Рубиус Групп»

внедрено(а) с «03» апреля 2023 г.

при разработке web-сервиса Visius для подсчёта пешеходного трафика по видео, который помогает
быстрее подбирать помещения при открытии новых торговых точек.

Операционный директор
управляющей организации «Рубиус»
(по доверенности № 01 от 01.01.2023 г.)

А.Е. Васильева