Министерство науки и высшего образования Российской Федерации
Федеральное государственное бюджетное образовательное учреждение
высшего образования
«Самарский государственный технический университет»

На правах рукописи

ПАНФИЛОВА Ирина Евгеньевна

МОДЕЛИ И АЛГОРИТМЫ НЕЙРОСЕТЕВОЙ БИОМЕТРИЧЕСКОЙ АУТЕНТИФИКАЦИИ В ЗАЩИЩЕННОМ РЕЖИМЕ ИСПОЛНЕНИЯ

Специальность 2.3.6.
Методы и системы защиты информации, информационная безопасность

Диссертация на соискание ученой степени кандидата технических наук

Научный руководитель:
доктор технических наук, доцент
Ложников Павел Сергеевич

Самара – 2024

Оглавление

Введение

Актуальность темы исследования. Активное развитие искусственного интеллекта (ИИ) обуславливает появление принципиально новых исследовательских задач, направленных на обеспечение его безопасности. Так, по данным MarketsandMarkets, глобальный рынок кибербезопасности для ИИ достигнет 38,2 млрд. долларов к 2026 году, при среднегодовом темпе роста в 26,3% (при 9,3 млрд. долларов в 2020 г.). Теоретические аспекты обеспечения безопасности ИИ отражены в концепции доверенного искусственного интеллекта (ДИИ), предполагающей высокий уровень надёжности, прозрачности и конфиденциальности технологий ИИ. Перечисленные требования особенно актуальны для систем и алгоритмов, принимающих ответственные решения или оперирующих критически важными данными. Таковым в полной мере может считаться искусственный интеллект, лежащий в основе биометрических систем аутентификации. Варианты защищённого исполнения таких систем (при которых невозможны раскрытие логики работы ИИ, извлечение его знаний и управление ими) представлены в рамках исследований *высоконадежной биометрической аутентификации*, в отечественной практике отраженных в серии стандартов ГОСТ Р 52633. Базовыми характеристиками высоконадежной нейросетевой биометрической аутентификации можно считать:

− обеспечение сокрытия решающих правил и конфиденциальности знаний;

− защиту биометрических данных от утечки;

− обеспечение устойчивости к внешним воздействиям (атакам).

Перечисленным критериям, на сегодняшний день, в большей степени соответствуют нейросетевые преобразователи «биометрия-код» (НПБК) − разновидность биометрических криптосистем (БКС), основанных на принципах работы искусственных нейронных сетей. Основной целью НПБК является связывание биометрического образа человека с криптографическим ключом. Чем длиннее криптографический ключ, продуцируемый НПБК, тем ниже возможность компрометации исходного биометрического образа. Разнообразие модификаций

нейросетевых преобразователей, представленное как зарубежной, так и отечественной литературой, наглядно демонстрируют актуальность разработки процедур биометрической аутентификации в защищенном режиме исполнения – *защищенной биометрической аутентификации (ЗБА)*, одновременно с этим, демонстрируя наличие ряда нерешенных задач. Одной из таких задач является необходимость поддержания высокого уровня защищенности процедуры аутентификации от деструктивных воздействий при заданном уровне точности распознавания биометрических образов (аутентификации). Под деструктивными воздействиями понимаются:

− атаки извлечения знаний НПБК;

− компрометация открытых биометрических образов;

− атаки на биометрическое предъявление (спуфинг атаки), направленные на получение несанкционированного доступа к объектам защиты системы ЗБА.

Особенности реализации указанной задачи зависят от конкретной биометрической модальности: отпечатки пальцев, рукописный почерк, голос, лицо и т.д. В этой связи, работа с биометрическими образами, обладающими особенностями сбора и представления, повышает требования к структуре и функциональным возможностям нейросетевых преобразователей «биометрия-код», лежащих в основе ЗБА. Среди таких биометрических модальностей особенно выделяется лицо человека: системы защищенной биометрической аутентификации по лицу в полной мере подвержены всем перечисленным выше деструктивным воздействиям.

С учетом перечисленных особенностей, можно сформулировать общую **научную задачу**, заключающуюся в необходимости изменения логики функционирования и концептуального исполнения НПБК с целью повышения его защищенности по отношению к деструктивным воздействиям при работе с биометрическими образами лиц. Проведенные в ходе работы исследования показали, что предложенная система защищенной биометрической аутентификации является полноценным решением поставленной научной задачи.

Степень проработки темы исследования. Вопросам защищенной биометрической аутентификации посвящены работы таких отечественных и зарубежных исследователей, как: Ахметов Б. С., Безяев А.В., Васильев В.И., Волчихин В.И., Иванов А.И., Малыгина Е.А., Сулавко А.Е., Derakhshani R., Dong X., Liu W., Rathgeb C., Rattani A., Talreja V. и др. Анализ современного состояния области демонстрирует значительные преимущества нейросетевых преобразователей «биометрия-код» (НПБК) перед альтернативными решениями на основе глубоких нейронных сетей и/или нечетких экстракторов. Однако существующие реализации НПБК обладают значительными недостатками либо с точки зрения длины продуцируемого ключа (только 128 бит для классического НПБК), либо с точки зрения точности работы с лицом человека (корреляционные нейроны работают только с сильно коррелированными признаками, что не характерно для образов лица). Кроме того, ЗБА на основе существующих реализаций НПБК оказывается уязвимой по отношению к спуфинг атакам, что способно нивелировать преимущества ее защищенного исполнения.

Перечисленные недостатки демонстрируют необходимость изменения не только логики работы НПБК, но и концептуального исполнения защищенной биометрической аутентификации с его участием. Результатом вносимых изменений должна стать система защищенной биометрической аутентификации по лицу на основе нейросетевого преобразователя «биометрия-код», устойчивая к атакам извлечения знаний НПБК и компрометации открытых биометрических образов лиц, а также к атакам на биометрическое предъявление (спуфинг атак).

Цель диссертационной работы: повысить защищенность процедуры биометрической аутентификации личности на основе нейросетевого преобразователя «биометрия-код», использующего открытые биометрические образы лица человека.

Для достижения поставленной цели необходимо решить **следующие задачи:**

1. Разработать концепцию защищенной биометрической аутентификации по лицу на основе НПБК, устойчивой к внешним воздействиям в виде атак на биометрическое предъявление (спуфинг атак).

2. Разработать модель нейрона и основанную на ней модель нейросетевого преобразователя «биометрия-код», осуществляющих процедуру биометрической аутентификации по лицу с обеспечением защиты знаний и биометрических образов лиц от компрометации.

3. Разработать алгоритмы обучения нейросетевого преобразователя биометрических образов лица в код на малых выборках.

4. Разработать систему биометрической аутентификации по лицу, устойчивую к атакам извлечения знаний НПБК и компрометации открытых биометрических образов лиц, а также к атакам на биометрическое предъявление (спуфинг атак).

Объектом исследования являются биометрические системы аутентификации человека на основе нейросетевых алгоритмов.

Предметом исследования являются нейросетевые модели преобразователей биометрических образов лица в сильный пароль или криптографический ключ.

Методы исследования. Применялись методы классификации и идентификации образов, биометрической аутентификации, глубокого обучения, теории вероятностей и математической статистики, распознавания образов, компьютерного моделирования, кодирования информации, аппарат искусственных нейронных сетей (ИНС), тригонометрические вычисления.

Достоверность и обоснованность работы подтверждается корректной постановкой задач и выбором известных методов, успешно применяемых в других областях, практическим применением системы, построенной в соответствии с разработанными моделями и алгоритмами, а также апробацией на научных конференциях, публикацией результатов в научных изданиях, в том

числе из Перечня ВАК, актами о внедрении результатов работы в образовательную и производственную сферы.

Научная новизна состоит из предложенных в работе:

1. Концепции защищенной биометрической аутентификации по лицу, *отличающейся* применением механизма защищенного нейросетевого контейнера (ЗНК) для безопасного взаимодействия блока аутентификации на основе пользовательского НПБК и блока обнаружения спуфинг атак на основе НПБК, представленного в виде классификатора реальных и поддельных изображений лиц, что позволяет обеспечивать устойчивость процедуры аутентификации к атакам на биометрическое предъявление (спуфинг атакам), а также дополнительную защиту таблиц нейросетевых функционалов пользовательского НПБК.

2. Модели тригонометрического нейрона, а также основанной на ней модели НПБК, *отличающихся* применением новой тригонометрической меры оценки расстояния между образами субъектов в подпространстве пар признаков вместо исходных признаков, что обеспечивает защиту образов лиц от компрометации путем продуцирования длинного криптографического ключа при высокой точности классификации. Предложенные модели не используют параметры распределений и/или характеристики образов легитимных пользователей, что обеспечивает защиту знаний НПБК от компрометации.

3. Алгоритма калибровки нейросетевых преобразователей «биометрия-код» и алгоритма автоматического обучения НПБК на основе тригонометрических нейронов, *отличающихся* использованием дополнительной информации, полученной путем оценки не участвующего в обучении набора биометрических образов лиц, что дает возможность быстрого и робастного обучения пользовательских НПБК на малых выборках образов лиц.

4. Структуры системы защищенной биометрической аутентификации по лицу, *отличающейся* наличием независимых блоков извлечения признаков, обучения нейросетевых преобразователей и аутентификации, а также применением варианта исполнения ЗНК, при котором в режиме обучения ключом

НПБК для обнаружения спуфинг атак осуществляется защита структуры пользовательского НПБК, а обратный описанному процесс происходит при аутентификации. Разработанная структура позволяет повысить защищенность процедуры биометрической аутентификации личности по лицу на основе НПБК в отношении спуфинг атак, а также атак извлечения знаний НПБК и компрометации биометрических образов лиц.

Теоретическая значимость диссертационной работы заключается в новом математическом аппарате для построения защищенной биометрической аутентификации, работающей со слабо коррелированными признаками лица человека. Предложенная математическая модель нейросетевого преобразователя «биометрия-код» повышает устойчивость биометрической аутентификации по лицу к деструктивным воздействиям в виде атак извлечения знаний и компрометации биометрических образов путем продуцирования длинного криптографического ключа (2048 бит), значительно превышающего длину ключа НПБК, обученного в соответствии с ГОСТ Р 52633.5 (128 бит), а также дополняют функционал нейросетевых преобразователей на основе корреляционных нейронов и позволяют работать со слабо колерованными признаками лица. Особенности работы нейронов НПБК позволяют использовать малое число примеров биометрических образов для обучения преобразователя и осуществлять процедуру обучения автоматически, не раскрывая параметров легитимных пользователей, что может быть актуально для иных приложений ИИ, исполняемых в защищенном режиме.

Практическая значимость работы заключается в разработке системы защищенной биометрической аутентификации по лицу и ее программной реализации. Система основана на предложенных в работе концепции, моделях, алгоритмах и структуре. Коэффициент равной вероятности ошибок при работе системы составил EER=2,5%, что говорит о сравнительно низком уровне ошибок распознавания образов при высоком уровне защищенности процедуры биометрической аутентификации по лицу от атак на биометрическое

предъявление, а также атак компрометации знаний НПБК и биометрических данных.

Положения, выносимые на защиту:

1. Концепция защищенной биометрической аутентификации по лицу, обеспечивающей противодействие атакам на биометрическое предъявление.

2. Модель тригонометрического нейрона и основанная на ней модель нейросетевого преобразователя «биометрия-код», осуществляющие процедуру защищенной биометрической аутентификации по лицу.

3. Алгоритм калибровки нейросетевых преобразователей «биометрия-код» и алгоритм автоматического обучения НПБК на основе тригонометрических нейронов, позволяющие производить быстрое и робастное обучение НПБК на малых выборках образов лиц.

4. Система защищенной биометрической аутентификации по лицу и ее программная реализация, обеспечивающая защищенность процедуры биометрической аутентификации личности по лицу на основе НПБК в отношении спуфинг атак и атак компрометации знаний НПБК и открытых биометрических образов лиц.

Личный вклад. Все результаты, изложенные в диссертации, включая программные реализации предложенных в работе алгоритмов, получены автором самостоятельно. Проработка цели и задач, способов их решения и вариантов представления результатов осуществлены автором совместно с научным руководителем.

Апробация результатов работы. Полученные результаты работы докладывались на следующих конференциях: Всероссийская молодежная научно-практическая конференция «Нанотехнологии. Информация. Радиотехника» (г. Омск); V Всероссийская научно-техническая конференция «Безопасность информационных технологий» (г. Омск)**;** IEEE Conference on the Intelligent Methods, Systems, and Applications (Giza, Egypt); III Всероссийская научная школа-семинар «Современные тенденции развития методов и технологий защиты информации» (г. Москва).

Работа выполнена в рамках государственного задания Минобрнауки России на 2023-2025 годы № FSGF-2023-0004. Часть работы по теме диссертации проводилась в рамках гранта ИБ МТУСИ № 40469-18/23-К. Грант выполнялся автором единолично.

Соответствие паспорту специальности. Тема и содержание диссертации соответствуют паспорту специальности 2.3.6. Методы и системы защиты информации, информационная безопасность, пункту 12: «Технологии идентификации и аутентификации пользователей и субъектов информационных процессов. Системы разграничения доступа», а также пункту 15: «Принципы и решения (технические, математические, организационные и др.) по созданию новых и совершенствованию существующих средств защиты информации и обеспечения информационной безопасности».

Публикации. По теме диссертации лично и в соавторстве опубликовано 10 печатных работ, 6 из которых изданы в журналах, рекомендованных ВАК; 1 научная публикация индексирована в международной информационно-аналитической системе научного цитирования Scopus. Получено 2 свидетельства о государственной регистрации программы для ЭВМ.

Объем и структура работы. Диссертация состоит из введения, 4 глав, заключения, списка литературы (195 наименований) и 2 приложений. Общий объем диссертации составляет 166 страниц, включающих в себя 15 таблиц и 28 рисунков.

Глава 1. Анализ современного состояния исследований в области защищенного исполнения искусственного интеллекта в задачах биометрической аутентификации субъектов

1.1. Проблема функциональной безопасности искусственного интеллекта

Искусственный интеллект (ИИ) занимает центральное место в современном технологическом ландшафте, оказывая значительное влияние на различные сферы общества. Его применение охватывает широкий спектр отраслей, включая здравоохранение, финансы, транспорт, образование и др. Статистика использования искусственного интеллекта полностью подтверждает его актуальность: по данным McKinsey [139] мировые инвестиции в прикладной ИИ выросли до 104 миллиардов долларов в 2022 году, а к 2025 году около 75% компаний планируют интегрировать ИИ в свои бизнес-процессы и продукты. Эта статистика открыто свидетельствует о значительном экономическом потенциале искусственного интеллекта и его способности оказывать существенное влияние на различные отрасли человеческой деятельности.

Ввиду растущего влияния искусственного интеллекта на различные аспекты современной жизни и экономики особенно актуальны новые исследовательские направления, ориентированные на разработку методов защиты искусственного интеллекта. Прежде всего, это связано с потенциальными угрозами, которые возникают при разработке и эксплуатации ИИ. Такие угрозы можно разделить на три основные категории (рис. 1.1):

1. Угрозы безопасности данных. Системы ИИ зависят от больших объемов данных, используемых для их обучения и функционирования, в связи чем, их утечка или компрометация может привести к серьезным негативным последствиям. Наиболее распространёнными в отношении данных являются атаки отравления данных [156] (англ. data poisoning) и атаки на конфиденциальность (англ. privacy attacks) [125]. В первом случае атака характеризуется «порчей» данных, используемых для обучения модели ИИ, с целью ухудшения ее

производительности. Для атак на конфиденциальность, в свою очередь, характерно намерение злоумышленника получить конфиденциальную информацию об ИИ или данных, на которых он был обучен, чтобы использовать их не по назначению.

2. *Угрозы эксплуатационной безопасности.* Данная категория угроз направлена непосредственно на модели и алгоритмы ИИ и может включать в себя состязательные атаки [72], атаки инверсии [191], атаки анализа ответов модели [67] и атаки с использованием трансферного обучения [162]. Для состязательных атак, в свою очередь, характерны незаметные изменения в данных, поступающих на вход уже обученной модели, с целью введения ее в заблуждение. В случае атак инверсии злоумышленник пытается восстановить или «обратить» модель ИИ, чтобы получить данные, использованные для ее обучения или входные данные, соответствующие определенным выходам модели. В свою очередь, анализ ответов модели может помочь злоумышленнику получить информацию о данных, которые не являются прямым результатом работы модели, но могут быть выведены из ее поведения или выходных данных (получение синтетических данных). Атаки с использованием трансферного обучения основываются на необходимости внедрения зловредного кода в предварительно обученные открытые модели, которые затем распространяются среди пользователей.

3. *Контекстные угрозы,* направленные на технологии развертывания моделей ИИ и людей, осуществляющих разработку и внедрение таких моделей в производственные и бизнес-процессы. В большинстве случаев подобные угрозы реализуются в виде атак на серверы или среды, на которых функционирует модель. Однако также стоит учитывать человеческий фактор и возможность реализации методов социальной инженерии для введения неправильных данных или получения доступа к системе ИИ.

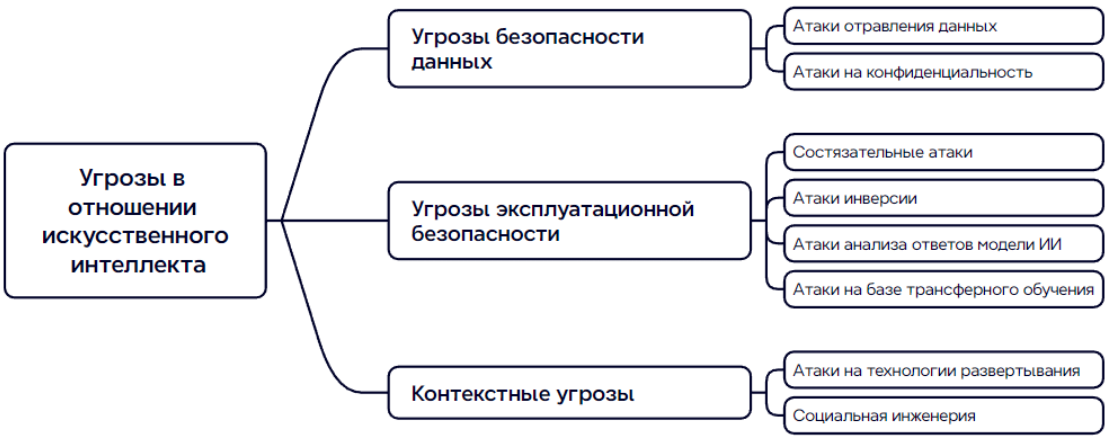


Рисунок 1.1 – Классификация угроз в отношении искусственного интеллекта

Для противодействия перечисленным угрозам необходимо разрабатывать комплексные подходы, включающие как технические, так и методологические решения, направленные на обеспечение безопасности и надежности систем ИИ [55]. Так, конфиденциальность данных, зачастую, достигается с помощью различных криптографических методов, а системы контроля доступа играют ключевую роль в предотвращении несанкционированного доступа к ним. Использование цифровых подписей и контрольных сумм способствует целостности данных, которые помогают обнаруживать в них любые изменения или искажения. Новые технологии анонимизации данных [173], включающие методы деидентификации, защищают персональные данные пользователей, позволяя использовать их для обучения моделей ИИ без риска раскрытия личной информации.

Обучение моделей на защищенных данных с использованием таких методов, как дифференциальная конфиденциальность [192] и федеративное обучение [182], позволяет моделям обучаться, не подвергая конфиденциальную информацию риску утечки. Защита моделей от состязательных атак достигается с помощью генеративно-состязательного обучения [42], в рамках которого модели обучаются распознавать и обрабатывать намеренно искаженные входные данные. Регулярная валидация и верификация моделей, а также методы детекции аномалий (концептуального дрейфа моделей или дрейфа данных [24])

обеспечивает их безопасность, на ранних этапах выявляя потенциальные уязвимости и ошибки.

Перечисленные подходы к обеспечению безопасности искусственного интеллекта представляют собой частные случаи *защищенного исполнения ИИ*. Под «защищенным исполнением» понимается невозможность анализа логики работы ИИ, управления ИИ и извлечения знаний из памяти ИИ (например, персональных данных) любым неавторизованным лицом. Так, вариантом защищенного исполнения ИИ является связывание любого решения ИИ, сформированного на основе анализа данных, с секретным паролем или криптографическим ключом, который известен только пользователю и искусственному интеллекту. Пароль встраивается в структуру модели и сохраняется в виде знаний путем специальной процедуры автоматического обучения (AutoML). Пользователь может в любой момент переобучить ИИ в автоматическом режиме, чтобы поменять пароль, если пароль был скомпрометирован. Таким образом, формируется прямая коммуникация между ИИ и человеком без участия третьих лиц.

Защищенный режим делает затруднительным реализацию следующих сценариев любым неавторизованным лицом: анализ операций, совершаемых ИИ (чтобы понять суть преобразований); управление ИИ (с помощью изменения алгоритма работы, подмены данных ИИ, состязательных атак и т.д.); извлечение и интерпретацию знаний ИИ.

Понятие защищенного исполнения ИИ следует считать составным элементом концепции доверенного искусственного интеллекта (ДИИ) наряду с объяснимостью, робастностью и др. «Доверие» к системам искусственного интеллекта, согласно ГОСТ Р 59276-2020 [7], означает «возможность применения этих систем при решении ответственных задач обработки данных». Исходя из определения, можно сделать вывод, что концепция доверия оказывается особенно актуальной в отношении систем и приложений ИИ, работающих с критически важными данными и/или функционирующими в потенциально враждебной среде. В этой связи, ДИИ играет ключевую роль в системах биометрической

аутентификации на основе искусственного интеллекта из-за специфики их работы с персональными данными. Кроме того, защита от атак и манипуляций также является критическим аспектом для систем биометрической аутентификации: такие системы уязвимы к различным видам атак, включая спуфинг [107] (обман системы с использованием поддельных биометрических данных) и атаки на целостность принимающих решения моделей.

1.2. Технологии искусственного интеллекта для задач биометрической аутентификации по лицу

История применения технологий искусственного интеллекта в системах биометрической аутентификации начиналась с классических алгоритмов распознавания, которые основывались на простых статистических методах и ограниченных объемах данных. С начала 2010-х годов и по настоящее время основное внимание в развитии биометрических систем сосредоточено на использовании методов глубокого обучения [50], в частности, глубоких нейронных сетей (ГНС). Наиболее популярными архитектурами ГНС для различных биометрических модальностей на сегодняшний день являются архитектуры с использованием сверточных слоев, или сверточные нейронные сети (СНС). Они способны эффективно обрабатывать изображения и видео, выявляя их ключевые особенности.

Первой архитектурой СНС, в том числе, применяющейся для задач биометрической аутентификации, стала представленная на соревнованиях ILSVRC-2012 модель AlexNet [77]. Архитектура содержит 5 сверточных слоев, за каждым из которых следуют функции активации ReLU (Rectified Linear Unit) и слои субдискретизации (Max Pooling). После сверточных слоев следуют 3 полносвязных слоя, которые принимают признаки из последнего сверточного слоя и используют их для классификации. Последний полносвязный слой использует функцию softmax. Уже через два года, в 2014 году, в Оксфорде впервые была представлена архитектура сети VGGNet [138], особенностью

которой стало использование крайне малых фильтров свертки (3x3) и удвоенное количество карт признаков после слоев Max Pooling (2x2). Такие «настройки» позволили разработчикам добиться довольно глубокой архитектуры (16 (VGG-16) или 19 (VGG-19) слоев в зависимости от конфигурации). В качестве функций активации также использовались нелинейности ReLU.

В 2015 году в игру вступает компания Google, представившая миру архитектуру 22-слойной сети GoogleNet [148], также известную как Inception. Особенностью архитектуры стали Inception-модули, которые объединяют различные типы операций свертки (1x1, 3x3, 5x5) и субдискретизации в одном слое. В отличие от предыдущих моделей, GoogleNet имеет одновременно глубокую и широкую архитектуру с несколькими Inception-модулями, что значительно повышают уровень возможной абстракции. Кроме того, для предотвращения проблемы затухания градиентов, архитектура GoogLeNet включает в себя вспомогательные классификаторы (auxiliary classifiers), добавляемые после каждого Inception-модуля. На сегодняшний день существуют 4 версии архитектуры Inception.

Следом за GoogleNet компания Microsoft в 2016 году презентует еще одну широко применяемую архитектуру ResNet [63]. Основное отличие ResNet от предыдущих архитектур заключается в использовании так называемых «остаточных связей» (ResNet — Residual Network — «остаточная сеть»). Вместо того чтобы пытаться научить модель напрямую сопоставлять вход и выход каждого слоя, остаточные связи позволяют передавать «остаточную» информацию через блоки, что облегчает обучение глубоких сетей. Такая концепция была реализована с помощью соединений быстрого доступа (shortcut connections). Архитектура ResNet и сегодня используется для разного рода задач из области распознавания образов и имеет, как минимум, 8 версий (например, ResNet-50, ResNet-101 и т.д.) в зависимости от количества слоев в сети и их конфигураций. Кроме того, Inception и ResNet архитектуры имеют объединенные реализации в виде двух моделей Inception-ResNet v1, рассмотренная в данной работе, и Inception-ResNet v2 [121].

Еще две архитектуры, презентованные в 2017 году, представляют интерес для задач биометрической аутентификации. Первая – SENet [66]. Основная идея архитектуры заключается в том, чтобы обучать модель адаптивно взвешивать важность признаков на основе их значимости для конкретной задачи. Основной компонент SENet – это блок Squeeze-and-Excitation, который состоит из двух основных этапов. На этапе «squeeze» используется глобальная субдискретизация для сжатия признаков по каналам, а затем на этапе «excitation» применяются полносвязные слои с нелинейной активацией для вычисления весов, которые адаптивно взвешивают каждый канал признаков. SENet может быть интегрирована в различные архитектуры сверточных нейронных сетей, такие как ResNet, Inception и другие. Кроме того, модель может быть применена как к каждому блоку, так и к конечным слоям сети. Вторая – DenseNet [69] – состоит из нескольких блоков «плотных» слоев, в каждом из которых слои свертки объединены вместе. Использование этих слоев является ключевым отличием DenseNet от других архитектур. В каждом блоке каждый слой получает на вход выходы всех предыдущих слоев и передает свой выход следующим слоям. Это создает прямые пути обратного распространения градиентов, что улучшает градиентный поток и способствует более эффективному обучению.

Несмотря на то, что использование полноценных глубоких нейронных сетей на основе архитектур, рассмотренных выше, является надежным подходом с точки зрения конечной производительности, актуальным остается вопрос ресурсоемкости процедуры их обучения. Особенно остро обозначенная проблема стоит для небольших устройств, таких как мобильные телефоны или периферийные гаджеты [57]. В связи с этим, за последние годы был разработан ряд архитектур, основным свойством которых является малое число параметров оптимизации (весов) и компактная структура организации сети. Часть из них применяются для задач биометрической аутентификации на мобильных устройствах. Наиболее распространенные архитектуры «маловесных» сетей представлены в таблице 1.1.

Таблица 1.1 – Архитектуры маловесных нейронных сетей

Маловесная сеть	Год появления	Количество слоев*	Количество параметров обучения*
SqueezeNet [70]	2016	11	1.25 миллиона
MobileNet [65]	2017	27	4.2 миллиона
ShuffleNet [188]	2017	50	1.87 миллиона
Xception [44]	2018	36	22.8 миллиона
MixNets [151]	2019	20	5 миллионов
ShiftNet [167]	2018	10	1.1 миллиона
VarGNet [185]	2019	10	7.41 миллиона

* – в таблице указаны значения наиболее оптимизированных версий архитектур

Уникальная структура рассмотренных СНС позволяет с их помощью осуществлять работу с самыми разными биометрическими модальностями. Однако наибольшей эффективности описанные архитектуры достигают в работе со сложными признаками лица человека: СНС обладают уникальной структурой, которая позволяет эффективно извлекать и иерархически обрабатывать признаки из изображений лиц. Современные архитектуры СНС обеспечивают высокую точность идентификации, позволяя эффективно работать даже в условиях изменяющегося освещения и при наличии частичных окклюзий лица.

Распознавание лиц [154] является наиболее популярным видом биометрической аутентификации благодаря своей неинвазивной природе и высокой точности. В отличие от других биометрических методов, таких как отпечатки пальцев или радужной оболочки глаза, распознавание лиц не требует физического контакта пользователя с дополнительными устройствами, что делает его более удобным и приемлемым для широкого применения в общественных и частных сферах, от личных гаджетов до общественных мест и транспортных систем.

Основные компоненты системы распознавания лиц на основе нейросетевых технологий представлены на рисунке 1.2 [13]: блок детектирования лиц, блок выравнивания, анти-спуфинг система, блок обогащения данных и

непосредственное распознавание, разделенное на сценарии обучения и тестирования.

Первым этапом работы с лицевой биометрией, как правило, является процедура детекции лица или нескольких лиц на входном изображении или видеопотоке. Указанному этапу может предшествовать предварительная обработка входных изображений с целью улучшения их качества или уменьшения шума (операции фильтрации, изменения контраста или яркости).

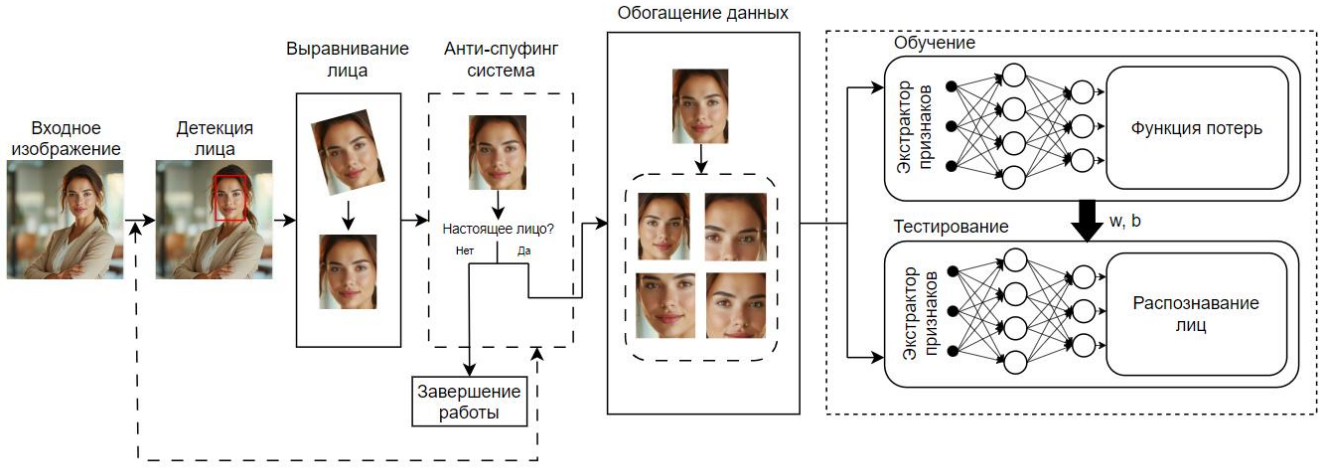


Рисунок 1.2 – Структурная схема системы распознавания лиц на основе СНС

Полученные изображения обрабатываются с помощью различных методов компьютерного зрения, таких как каскады Хаара [140], метод Виолы-Джонса [171] или методы обучения с использованием глубоких нейронных сетей (YOLO [128], SSD [97], MTCNN [183], Retina [87] и др.). Основной задачей этапа детекции является поиск и выделение лица на изображении или в видеопотоке – алгоритм, чаще всего, возвращает координаты точек, «обрезающих» лицо, а также ключевые точки лица (уголки глаз, брови и рот, кончик носа и т.д.). Наличие данного этапа значительно упрощает процедуру распознавания лиц и повышает эффективность работы моделей извлечения признаков.

Полученные после детектирования координаты ключевых точек позволяют определить геометрическую структуру лица и «выровнять» лицо относительно рамок изображения путем переноса (термин из евклидовой геометрии),

масштабирования или вращения. Эффективность описанной процедуры для задачи распознавания лиц была продемонстрирована в 2017 году группой американских исследователей в работе [29], которые смогли доказать, что точность распознавания нейросетевых моделей можно повысить, применяя хорошо подобранную предобработку изображений в виде операции выравнивания по ключевым точкам. На сегодняшний день процедура выравнивания изображения лица приобретает автоматический характер, и, зачастую, встраивается в системы распознавания лиц в качестве дополнительных нейронных сетей предобработки [86, 113].

В качестве одного из дополнительных этапов обработки входного изображения лица может выступать подсистема определения спуфинг (англ. spoofing — подмена) атак. Более узким определением указанной задачи является термин liveness detection, который сводит проблему спуфинга биометрических образов лица к задаче определения подлинности поступающего входного образа. В современном исследовательском сообществе проектирование и разработка подобных систем является самостоятельной задачей, в рамках которой представлены различные методы [14] детектирования атак на биометрическое предъявление (спуфинг атак): атаки распечатанного изображения, атаки воспроизведения, атаки с помощью 3D маски и др. Целью работы модуля является ответ на вопрос: поддельное или реальное изображение лица поступает на вход системы? В связи с этим, анти-спуфинг модуль может быть представлен в виде нейросетевой модели глубокого обучения, решающей задачу бинарной классификации (spoof/live) [172]. В ряде случаев анти-спуфинг система предшествует процедуре детектирования лица, так как алгоритм liveness detection использует дополнительную информацию контекста изображения для принятия решения о живости пользователя [75].

Экспериментально доказанное [109] влияние на производительность нейросетевых моделей глубокого обучения, кроме всего прочего, оказывают условия получения входного изображения: освещение, выражение лиц, окклюзия (загораживание), угол поворота лица и т.д. В связи с этим, в области

распознавания лиц за последние несколько лет стали появляться все больше исследований, связанных с разработкой методов обогащения входных данных и исходных наборов данных [52, 92, 94, 176, 178]. В данном случае под «обогащением данных» понимаются действия, направленные на расширение, повышение разнообразия или качества набора данных, который является обучающим для модели распознавания лиц. При этом обогащение возможно как на уровне входных изображений [174], так и на уровне признакового пространства алгоритма распознавания [80].

Еще более актуальной указанная задача становится в контексте специфики сбора данных для обучения систем распознавания лиц: обычно, для каждого класса (субъекта) собирается только одна обучающая выборка, полученная в определённых условиях. Разнообразие данных в таком случае оказывается слишком затратным как со стороны пользователя, так и со стороны разработчиков. Тогда при обучении моделей можно столкнуться с проблемой «распознавания по одной выборке» [79], при которой производительность модели, в том числе, способность к обобщению, значительно снижается из-за ограниченности обучающего набора данных.

Для преодоления описанной проблемы применяются различные методы, целью которых может являться как пополнение разнообразия набора данных (тренировочного или тестового) [30, 112, 122, 147], так и улучшение качества используемых изображений путем генерации высокоинформативных синтетических образов [157, 193, 195]. Набор самых простых и наиболее часто используемых подходов к обогащению данных объединяется под общим термином аугментации [147]. Аугментация включает себя такие процедуры обработки данных, как изменение размеров изображений, поворот, отражение, добавление шума, изменение контрастности или яркости (рис. 1.3). Аугментация данных не требует дополнительных ресурсов и основывается исключительно на математических преобразованиях входных изображений. Однако даже такие модификации изображений способны повысить точность моделей и их обобщающую способность.

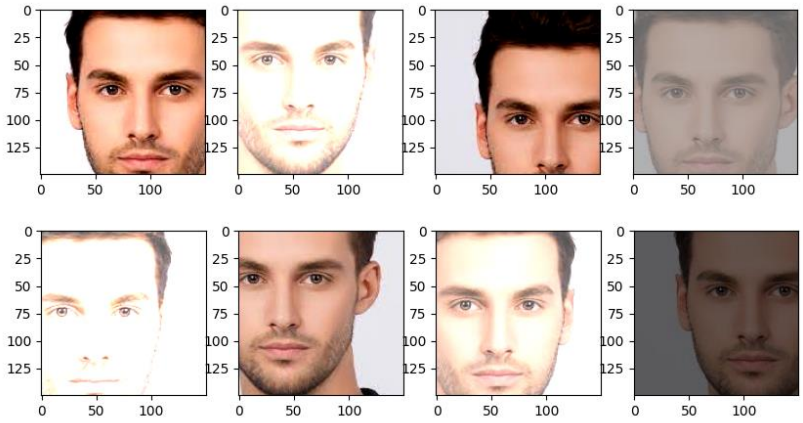


Рис. 1.3 – Пример аугментации изображения лица

Однако значительно более высокой производительности систем распознавания лиц можно достичь, применяя методы обогащения данных, в основе которых лежат технологии глубокого обучения: автокодировщики [122, 195], генеративно-состязательные сети [30, 157] или методы 3D реконструкции лица на основе сверточных нейронных сетей. Последняя группа методов характерна исключительно для задачи распознавания лиц и позволяет создавать трехмерное представление лица человека на основе двухмерных изображений [112]. Для осуществления 3D реконструкции лица могут применяться различные методы, включая структуру общего пространства для моделирования формы лица, методы шаблонного соответствия для оценки параметров лица, методы морфинга для создания трехмерной модели лица и другие.

Процесс непосредственного распознавания лиц в системах, основанных на нейросетевых алгоритмах, сопряжен с предварительным проектированием и обучением нейронных сетей, способных извлекать из входных образов лица информацию для однозначного определения субъекта. Процедуру получения такой информации называют извлечением признаков, а алгоритмы осуществляющие извлечение – экстракторами признаков. Целью обучения экстракторов признаков является приобретение ими «знаний» о некоторых обобщённых представлениях человеческих лиц (получение семантического представления), которые будут использоваться в процессе

идентификации/верификации на тестовом наборе данных. В терминах машинного обучения рассмотренная задача обучения нейросетевого экстрактора признаков определяется как задача, решаемая методом zero-shot обучения: модель способна к обобщению и классификации новых классов без предварительного обучения на них.

Благодаря постоянно совершенствующимся архитектурам и передовым методам обучения, таким как пакетная нормализация (англ. batch-normalization), нейронные сети становится все глубже, а обучение становится все более управляемым. Вслед за этими тенденциями развиваются и специализированные архитектуры, предназначенные для распознавания лиц в разных сценариях функционирования. Одной из первых таких сетей стала разработанная в 2014 году архитектура DeepFace [149], объединившая в себе преимущества ResNet («остаточные слои»), новую для того времени функцию потерь (триплетную) и функции активации ReLU. На датасете LFW [68] модели удалось достичь 97.35% точности.

В 2015 году исследователями Google была разработана еще одна архитектура FaceNet [134] (на основе GoogleNet), основной целью которой является создание уникальных векторных представлений лиц, которые можно использовать для идентификации или верификации личности. Модель включала в себя триплетную функцию потерь, основанную на тройках примерно выровненных совпадающих/несовпадающих участков лица, созданных с помощью нового метода онлайн-майнинга триплетов, и достигла крайне высокой производительности в 99,63%. В том же году архитектура VGGNet была обучена группой исследователей на новом наборе данных изображений лиц VGGface [115], разработанном в рамках проекта Visual Geometry Group (VGG) в Университете Оксфорда. Достигнутая точность алгоритма на указанном датасете составила 98,95%.

В 2017 году специально разработанная для решения задачи распознавания и верификации лиц архитектура SphereFace [96] продемонстрировала принципиально новый подход к понимаю процедуры создания векторных

представления лиц: она создает векторные представления лиц на сфере, а не на плоскости. SphereNet использует свойство сферической гармоники для создания векторных представлений, которые обладают геометрическими инвариантами. Это означает, что представления лиц остаются неизменными при поворотах, сдвигах и масштабировании изображений. Кроме того, в SphereNet применяются сверточные слои, которые работают непосредственно на сферической поверхности. Такое переосмысление подхода к обучению специализированных архитектур для распознавания лиц позволило SphereNet достичь 99,42% на наборе данных LFW. Однако немного позже, в 2019 году, сравнимой с SphereFace производительности удалось достичь уже на маловесных архитектурах: MobiFace [53] на том же наборе данных LFW показала точность 99,7%.

Обучение описанных глубоких сетей отличается от стандартной классификации образов: крайне сложно и нецелесообразно учить модель классифицировать все возможные лица, которые потенциально будут иметь доступ к системе. Чаще всего обучение возможно произвести только на ограниченном наборе данных, хуже или лучше описывающем понятие «лицо». Тогда обучение нейронной сети, направленной на распознавание лиц, должно сводиться не к цели «знать каждого человека в лицо», а к цели «знать, что представляет собой человеческое лицо и уметь классифицировать даже не использованные в обучении лица».

Однако изменение поставленной задачи приводит к следующему вопросу: как добиться максимальной обобщающей способности модели и сформировать у нее надежное представление лица человека, если обучающая выборка невелика и не способна продемонстрировать все возможные вариации строения и особенностей лиц? Ответ на поставленный вопрос может быть найден, если внимательно изучить два основных аспекта обучения любой модели: а) на каких данных проводится обучение; б) какой алгоритм осуществляет обучение. В первом случае единственным вариантом развития событий может быть кратное увеличение обучающего набора так, чтобы он включал в себя максимально возможное многообразие человеческих лиц. Однако как нетрудно догадаться,

такой сценарий является крайне затратным с точки зрения ресурсов (хранение данных и обучение на них моделей) [164].

Второй подход оказался более перспективным в глазах научного сообщества, поскольку в последние годы было проведено значительное количество исследований, сосредоточенных на модификации процедур обучения алгоритмов распознавания лиц, в частности, на разработке новых функций потерь для таких моделей. Функция потерь представляет собой ключевой инструмент, используемый для оптимизации модели с целью достижения требуемого поведения. В связи с этим, ниже будут рассмотрены основные методы, включая те, что относятся к области metric learning [164], нацеленные на создание функций потерь, которые позволяют оценивать расстояние (или сходство) между двумя точками в n-мерном пространстве, представляющими образы лиц.

Функции потерь, основанные на евклидовом расстоянии (metric learning)

Такие функции переводят вектор признаков класса (лица) в евклидово пространство (L2), где вектор признаков есть точка в этом пространстве. Классическими представителями таких функций являются константная функция потерь (contrastive loss) и триплетная функция потерь (triplet loss). Константная функция потерь [133, 146-147] работает с парой образов лица, стараясь как можно ближе «свести» между собой образы одного класса и «развести» как можно дальше образы из разных классов:

$$L = y_{ij} \max\left(0, \left\|f(x_i) - f(x_j)\right\|_2 - \varepsilon^+\right) + (1 - y_{ij})\max(0, \varepsilon^- - \left\|f(x_i) - f(x_j)\right\|_2) \qquad (1.1)$$

где $y_{ij} = 1$ означает что образы x_i и x_j из одного класса, $y_{ij} = 0$ означает принадлежность образов разным классам. $f()$ – сверточная нейронная сеть, генерирующая векторы признаков (англ. embeddings), а ε^+ и ε^- - пороги принятия решения относительно схожести образов (принадлежности одному классу). Для режима верификации указанная функция использовалась в нескольких

реализациях решения DeepID [146]. Однако главной проблемой contrastive loss остается сложная процедура подбора значений ε^+ и ε^-.

Альтернативой константной потере стала триплетная функция потерь [132, 133], на вход понимающая сразу три образа: x_i^a – якорь (anchor), x_i^p – положительный образ (positive), тот же класс, что и якорь, и x_i^n - негативный образ (negative), отличный от якоря класс. Функция потерь приобрела иной вид:

$$L = \sum \left[\left\| f(x_i^a) - f\left(x_i^p\right) \right\|_2^2 - \left\| f(x_i^a) - f(x_i^n) \right\|_2^2 + \alpha \right]_+ \tag{1.2}$$

где α – порог принятия решения. Суть обучения заключается в минимизации расстояния между якорем и положительным примером и максимизации расстояния между якорем и отрицательным примером. Такая функция потерь, в том числе, была опробована при обучении архитектуры FaceNet [134]. Она является одной из самых популярных на практике и иногда используется совестно с softmax [45]: тогда сеть сначала обучают с помощью softmax, а затем повышают ее производительность за счет triplet loss.

Softmax и ее модификации

Константная и триплетная функции потерь не всегда оказываются эффективными, особенно в случае крупного набора данных и большого числа обучающих примеров: обучение становится вычислительно затратным и крайне чувствительным с точки зрения подбора гиперпараметров (ε^+, ε^-, α). В связи с этим, имеют место альтернативные методы, способные избежать зависимости от конкретных образов и расстояний между ними, но при этом хорошо снижать внутриклассовую вариабельность образов. Логика построения альтернативных методов обучения заключается в модификации функции потерь Softmax (cross-entropy) Loss (табл. 2) являющейся одной из самых широко используемых функций в задачах классификации, включая задачу распознавания лиц [168]. Так, функция потерь center loss [165], а также различные варианты ее построения, опираются на идею о том, что можно обучить модель «группировать» точки

одного класса относительного некоторого центра. Обучение будет заключаться в минимизации расстояния между центром и *i*-ым положительным образом: к softmax loss добавляется оценка разброса признаков от центров соответствующих классов. Одним из вариантов функции center loss является представленная в 2017 году в работе [187] функция range loss, главная цель которой заключается в том, чтобы максимизировать различия между распределениями признаков (например, активаций нейронов на выходном слое) для лиц разных людей. Для этого в некоторых работах предлагают измененные вариации center loss, например center-invariant loss [169] или contrastive center loss [120], однако основными недостатками всех модификаций center loss подходов остаются высокая ресурсоемкость со стороны графического процессора и необходимость использования хорошо сбалансированных данных для осуществления классификации.

Одним из прорывов в понимании функционала обучения нейронных сетей для задачи распознавания лиц стали функции потерь, направленные преодоление проблем неверной классификации сложных образов (с низкой межклассовой вариацией). Одной из первых реализаций стала large-margin softmax loss (LM-Softmax), представленная в работе [95]. Основная идея LM-softmax состоит в добавлении «отступа» (margin) к обычному softmax, который представляет собой заданное значение, определяющее минимальный угол между векторами признаков разных классов. Если с помощью отступов корректировать (вычитать/прибавлять/домножать на отступ) этот угол, высчитываемый с помощью скалярного или косинусного расстояния, можно добиться высокой степени разделения классов в заданном пространстве.

Следом за LM-softmax стали появляться ее различные модификации, меняющие и улучшающие способы дифференциации углов между классами в том или ином пространстве. Например, AM-softmax (additive margin softmax) [159] с более наглядным (с геометрической точки зрения) способом введения отступа между классами или A-Softmax (angular softmax) [96], в которой благодаря дополнительной L2 нормализации весов векторы признаков оказываются

лежащими на гиперсфере, где и происходит дополнительная работа с отступами. Адаптированная под задачу распознавания лиц AM-softmax представлена в работе [161] и носит название CosFace. Наиболее популярной реализацией работы с отступами в softmax на практике оказалась функция ArcFace [47], при расчете которой граница принятия решения о принадлежности объекта классу является окружностью фиксированного радиуса на поверхности гиперсферы единичного радиуса в пространстве признаков. Функции FairLoss [91] и AdaptiveFace [93], в свою очередь, вводят адаптивную корректировку отступов для разных классов.

Таблица 1.2 – Softmax и ее модификации

Функция потерь	Формула
Softmax [168]	$L_i = -\log(\frac{e^{f_{y_i}}}{\sum_j e^{f_j}})$
Center loss [165]	$L = L_s + \lambda\frac{1}{2}\sum_{i=1}^{m}\|x_i - c_{y_i}\|_2^2$
Range loss [187]	$L_{R_{intra}} = \sum_{i\subset I}\frac{k}{\sum_{j=1}^{k}\frac{1}{D_{ij}}}$ $L_{R_{inter}} = \max(m - \min_{a,b\in I}\|\bar{x}_a - \bar{x}_b\|_2^2, 0)$
Center-invariant loss [169]	$L = L_{softmax} + \lambda\frac{1}{2}\sum_{i=1}^{m}\|x_i - c_{y_i}\|_2^2 + \gamma\frac{1}{4}(\|c_{y_i}\|_2^2 - \frac{1}{m}\sum_{k=1}^{m}\|c_k\|_2^2)^2$
Contrastive center loss [120]	$L = \frac{1}{2}\sum_{i=1}^{m}\frac{\|f_i - c_{y_i}\|_2^2}{\sum_{j=1,j\neq y_i}^{k}\|f_i - c_j\|_2^2 + \delta}$
LM-Softmax [95]	$L_i = -\log(\frac{e^{\|W_{y_i}\|\|x_i\|\psi(\theta_{y_i})}}{e^{\|W_{y_i}\|\|x_i\|\psi(\theta_{y_i})} + \sum_{j\neq y_i} e^{\|W_i\|\|x_i\|cos(\theta_j)}})$
AM-softmax [159]	$L = -\frac{1}{n}\sum_{i=1}^{n} log\frac{e^{s*(cos\theta_{y_i}-m)}}{e^{s*(cos\theta_{y_i}-m)} + \sum_{j=1,j\neq y_i}^{c} e^{s*cos\theta_j}}$
A-Softmax [96]	$L = \frac{1}{N}\sum_{n=1}^{N} -log\frac{e^{\|x^{(n)}\|\phi(\theta_{y_n}^{(n)})}}{e^{\|x^{(n)}\|\phi(\theta_{y_n}^{(n)})} + \sum_{j\neq y_n} e^{\|x^{(n)}\|cos(\theta_j^{(n)})}}$
ArcFace [47]	$L = -log\frac{e^{s*(cos\theta_{y_i}+m)}}{e^{s*(cos\theta_{y_i}+m)} + \sum_{j=1,j\neq y_i}^{N} e^{s*cos\theta_j}}$

FairLoss [91]	$L = \sum_{i=1}^{N} \omega_i l_{CE}(\hat{y}_i^*, \hat{y}_i')$
AdaptiveFace [93]	$L = -\frac{1}{M}\sum_{n=1}^{M} log \frac{e^{s*(cos(\theta_{y_i^{(j)}})-m_{y^{(j)}})}}{e^{s*(cos(\theta_{y_i^{(j)}})-m_{y^{(j)}})} + \sum_{i=1, i\neq y^{(j)}}^{N} e^{s*cos(\theta_{ij})}} - \frac{1}{N}\sum_{i=1}^{N} m_i$
L2-Softmax [126]	$L = -\frac{1}{M}\sum_{n=1}^{M} log \frac{e^{W_{y_i}^T f(x_i)+b_{y_i}}}{\sum_{j=1}^{C} e^{W_j^T f(x_i)+b_j}}$
CoCo loss [99]	$L = -\sum_{i=1}^{m} log \frac{\exp(\hat{c}_{y_i}^T * \hat{f}_i)}{\sum_{j=1}^{m} \exp(\hat{c}_j^T * \hat{f}_i)}$

Кроме рассмотренных выше подходов по работе с отступами для снижения межклассовой вариации, с 2017 года в некоторых работах также делались попытки нормализовать характеристики и веса в функциях потерь для улучшения производительности модели. Двумя яркими представителями такого направления являются L2-Softmax [126] и CoCo loss [99]. В первом случае, основываясь на наблюдении, что L2-норма признаков, полученных с использованием Softmax Loss, положительно влияет на процедуру классификации лиц, перед применением функции softmax все признаки модели приводятся к одинаковой L2-норме путем нормализации. Во втором случае, после нормализации признаков и весов CoCo loss оптимизирует косинусное расстояние между признаками данных.

Опираясь исключительно на семантическую информацию, полученную от ранее обученного экстрактора признаков, системе распознавания лиц необходимо принять решение об отнесении *i*-ого входного образа к тому или иному классу. Безусловно, принятие такого решения возможно только в случае, если на этапе функционирования модели в реальных условиях добавить к экстрактору признаков блок, осуществляющий распознавание или, иначе, сопоставление признаковых описаний лиц (face matching).

Ситуация, при которой модель «знает» лица, которые будет распознавать является наиболее простой и не требует изменения структуры системы. Однако такой вариант редко встречается на практике и, как правило, для новых классов приходится проводить процедуру регистрации. В самом простом случае

регистрация – это формирование усредненного значения вектора представлений нового класса, с которым, в соответствии с заранее заданным порогом, на этапе распознавания будет осуществляться сравнение с помощью евклидова (L2) или косинусного расстояния.

Сравнение шаблонов может быть осуществлено в виде двух сценариев: верификация и идентификация. В первом случае, сравнение признакового описания зарегистрированного пользователя с вектором признаков *i*-ого входного образа производится в режиме «один к одному». Целью такого сравнения является определение, относятся ли эти два изображения к одному и тому же объекту (пользователю). В приложении к информационной безопасности верификация лиц особенно актуальна для систем контроля доступа. Второй сценарий – идентификация – подразумевает сравнение «один ко многим»: из всех возможных классов (пользователей) выбирается один, наиболее похожий на целевой. Для оценки качества работы нейросетевых алгоритмов верификации и/или идентификации могут использоваться такие метрики, как:

1) Точность распознавания (accuracy), представляющая собой процент правильных классификаций.

2) Метрика FMR (False Match Rate) или False Acceptance Rate (FAR), представляет собой процент случаев, при которых система ошибочно принимает пользователя за легитимного.

3) Метрика FNMR (False Non-Match Rate) или FRR (False Reject Rate), представляет собой процент подлинных образцов, которые были ошибочно отклонены.

4) Метрика GAR (Geniune Accept Rate) или TAR (True Acceptance Rate) – это процент подлинных образов, которые были правильно приняты (т.е. TAR = 1 – FNMR).

5) Анализ ROC-кривой. ROC-кривая – это график, который показывает зависимость между FAR (False Acceptance Rate) и TAR (True Acceptance Rate) при различных пороговых значениях для принятия решения в биометрической системе. На ROC-кривой TAR обычно отображается на оси ординат (y), а FAR на

оси абсцисс (x). Путем изменения порогового значения для принятия решения можно создать разные точки на ROC кривой, каждая из которых представляет собой пару значений FAR и TAR. Чем ближе ROC кривая к верхнему левому углу, тем лучше производительность системы. Идеальная ROC кривая проходит через точку (0,1), что означает FAR равный 0% и TAR равный 100%.

6) Метрика EER (Equal Error Rate), представляющая собой точку на кривой ROC, где вероятность ложного положительного срабатывания (False Positive Rate) равна вероятности ложного отрицательного срабатывания (False Negative Rate). Иначе EER может быть описана как точка пересечения на графике значений FAR и FRR. Чем ниже значение EER, тем лучше производительность системы, поскольку это указывает на меньшее количество ошибок и более точную идентификацию или верификацию.

1.3. Уязвимости процедур биометрической аутентификации по лицу на базе искусственного интеллекта

Согласно последним исследованиям [23], большинство существующих атак на биометрические системы, в том числе, реализованные на базе ИИ и ориентированные на распознавание лиц, можно разделить на две масштабные группы: атаки на биометрическое предъявление [179] (спуфинг-атаки) и атаки на структурные составляющие алгоритма аутентификации с целью извлечения из них знаний или данных.

Спуфинг атаки представляют собой значительную угрозу безопасности систем лицевой биометрической аутентификации, поскольку они могут быть использованы для несанкционированного доступа к системе путем маскировки под авторизованного пользователя. Для эффективной защиты от таких атак необходимо детально изучить разнообразие их реализаций. Например, в зависимости от способа представления поддельных образов в физическом пространстве, выделяются 2D и 3D атаки [54]. К 2D атакам относятся случаи, когда изображение лица подается системе через экран устройства, фотографию на

бумаге или видеоролик, воспроизведенный с экрана. Эти атаки считаются относительно простыми в воспроизведении, так как они не требуют специального оборудования или глубоких знаний о личности жертвы — достаточно нескольких качественных снимков или короткого видеоролика.

3D атаки, в свою очередь, имеют менее четкие границы сложности воспроизведения, так как эта сложность во многом зависит от системы распознавания лиц. Для компрометации системы может использоваться трехмерная маска лица, напечатанная на принтере и повторяющая основные контуры лица жертвы. Хотя данный метод требует качественного изображения лица, он все еще не является чрезмерно сложным и может быть реализован в «домашних условиях». Тем не менее, современные системы распознавания лиц способны обнаруживать атаки такого уровня, поэтому наиболее сложными становятся атаки с использованием высокотехнологичного оборудования, таких как 3D копии лица, театральные или силиконовые маски. Для создания таких масок требуется не только специальное оборудование, но и детальная информация о пропорциях лица жертвы и особенностях его мимики.

Следует отметить, что классификация спуфинг атак (табл. 1.3), основанная на способе представления поддельного образа в физическом пространстве, не является исчерпывающей, поскольку не учитывает возможности цифровых манипуляций. В работе [179] предложена отдельная категория угроз, которая охватывает ряд незаметных манипуляций в цифровой структуре видеопотока или изображения. В дополнение к этому, авторы вводят два дополнительных понятия: «имперсонализация» и «запутывание». Имперсонализация включает реализацию атаки через копирование атрибутов лица подлинного пользователя на различные носители, такие как фотографии, экраны и 3D маски. Таким образом, этот класс атак охватывает как 2D, так и 3D реализации, обобщая подходы независимо от их физического представления. Запутывание же осуществляется с использованием дополнительных элементов (очков, макияжа, татуировок и т.д.) для имитации внешности жертвы.

Таблица 1.3 – Типы спуфинг атак на системы распознавания лиц

Категория	**Тип атаки**	**Описание**	**Пример**	**Сложность реализации**
Цифровые манипуляции	Цифровые манипуляции	Незаметные манипуляции в цифровой структуре видеопотока или изображения.	Изменение пикселей изображения, вставка чужого лица в видеопоток.	Высокая: требует технических навыков и доступа к данным.
Атаки физического представления	Фотография	Использование фотографии лица для обмана системы распознавания лиц.	Показ фото перед камерой.	Низкая: требуется только качественное фото.
	Видео-воспроизведение	Воспроизведение видео с лицом, пытаясь имитировать живое лицо.	Показ видео на экране смартфона перед камерой.	Низкая: требуется видео с лицом.
	3D маска	Использование 3D маски, чтобы подделать лицо реального человека.	Ношение маски, напечатанной на 3D принтере.	Средняя: требуется 3D печать и качественное изображение лица.
	Очки	Использование очков для изменения внешности.	Очки различной формы и цвета.	Низкая: легко реализуемо.
	Макияж	Изменение внешности с помощью макияжа.	Использование косметики для изменения черт лица.	Низкая: требует базовых навыков макияжа.
	Татуировка	Нанесение временных или постоянных татуировок для изменения внешности.	Татуировки на лице.	Средняя: требует оборудования и навыков нанесения татуировки.

Атаки на структурные составляющие алгоритма аутентификации чаще всего выполняются с использованием сложных инструментов и навыков взлома и/или пентестинга [19]. Принято разделять такие атаки по количеству блоков, входящих в состав классической системы биометрической аутентификации: блок извлечения признаков, блок распознавания (принятия решения) и база данных. Для первого блока наиболее характерными являются атаки, основанные на

использовании «знаний» алгоритмов системы для восстановления исходных биометрических данных из шаблонов. В частности, к ним относятся атака инверсии [191], атака подбором [67] (то же, что и атака анализа ответов ИИ) и атака ложного принятия [160]. Последняя атака является наиболее труднореализуемой и заключается в обмане системы аутентификации, путем использования знаний о процессе преобразования образов, ключевых параметров алгоритма и общедоступных наборах данных для создания ложных шаблонов.

В отношении блока принятия решения возможно реализация, как минимум, двух атак: атаки по словарю [40] и атаки на совпадение. В первом случае злоумышленник создает большой набор шаблонов (словарь), которые используются для нахождения подходящего шаблона путем перебора. Во втором случае атака осуществляется непосредственно в отношении блока принятия решения, встроенного в систему.

Наиболее характерными для систем биометрической аутентификации являются атаки в отношении персональных биометрических данных, хранящихся в соответствующей базе. В этой связи, защита биометрических данных от компрометации является одной из ключевых задач при построении безопасных систем биометрической аутентификации. С учетом того, что база данных отвечает за хранение биометрических шаблонов, защита данных от компрометации возможна за счет продуцирования и хранения защищенных шаблонов. Защищенные шаблоны могут считаться устойчивыми к атакам, если они выполняют такие требования, как отменяемость, неинвертируемость и отсутствие очевидных взаимосвязей. Если же шаблоны остаются незащищенными, они могут быть уязвимы к атакам отравления данных [156] и перекрестным атакам [110]. Атака подмены предполагает замену оригинальных данных пользователя иными биометрическими образцами с целью отказа в обслуживании легитимного пользователя. Перекрестные (корреляционные) атаки заключается в извлечении подлинных биометрических данных легитимных пользователей путем анализа сразу нескольких хранилищ и/или шаблонов.

Приведенный перечень характерных для биометрических систем аутентификации (в том числе по лицу) атак имеет ряд схожих черт с атаками в отношении искусственного интеллекта. С одной стороны, это объясняется тем, что категории атак направлены на компрометацию систем, использующих сложные алгоритмы для обработки данных. С другой стороны, существующая корреляция связана с тем, что абсолютное большинство современным систем биометрической аутентификации по лицу функционируют на базе моделей и алгоритмов искусственного интеллекта [26], в частности, глубоких нейронных сетей.

Противодействие атакам, направленным на предъявление системе поддельных образов (спуфинг атакам), а также ориентированным на компрометацию знаний ИИ, лежащего в основе системы, или биометрических образов, является приоритетной задачей в отношении разработки безопасных систем биометрической аутентификации по лицу. Разработка методов и алгоритмов противодействия этим атакам производится в рамках исследований *защищенной биометрической аутентификации (ЗБА)* [20] или биометрической аутентификации в защищенном режиме исполнения.

1.4. Методы и принципы построения процедур биометрической аутентификации в защищенном режиме исполнения

Одной из распространенных практик обеспечения конфиденциальности биометрических данных является хранение таковых в виде биометрических шаблонов (БШ) [104], получаемых путем тех или иных преобразований исходного образа. Однако недавние исследования показывают, что исходные биометрические данные могут быть восстановлены из открыто хранящихся биометрических шаблонов [103], а затем использованы с целью доступа к системе. В связи с этим, особенно актуальными являются задачи защиты биометрических шаблонов от компрометации путем реализации такого процесса преобразования биометрического образа в его шаблон, при котором невозможно

раскрытие параметров работы преобразующей системы, а также самих биометрических данных. Именно такая задача, отдельно от области развития технологий биометрической аутентификации, на протяжении последних десятилетий решается в контексте направления защиты биометрических шаблонов (ЗБШ).

Область исследований ЗБШ представляет собой довольно широкий спектр подходов (рис. 1.4) к сокрытию, шифрованию, изменению и преобразованию биометрических образов людей. В общем случае все перечисленные подходы позволяют обеспечивать защиту биометрических систем от рассмотренных ранее атак и позволяют осуществлять построение процедур биометрической аутентификации в защищенном режиме исполнения.

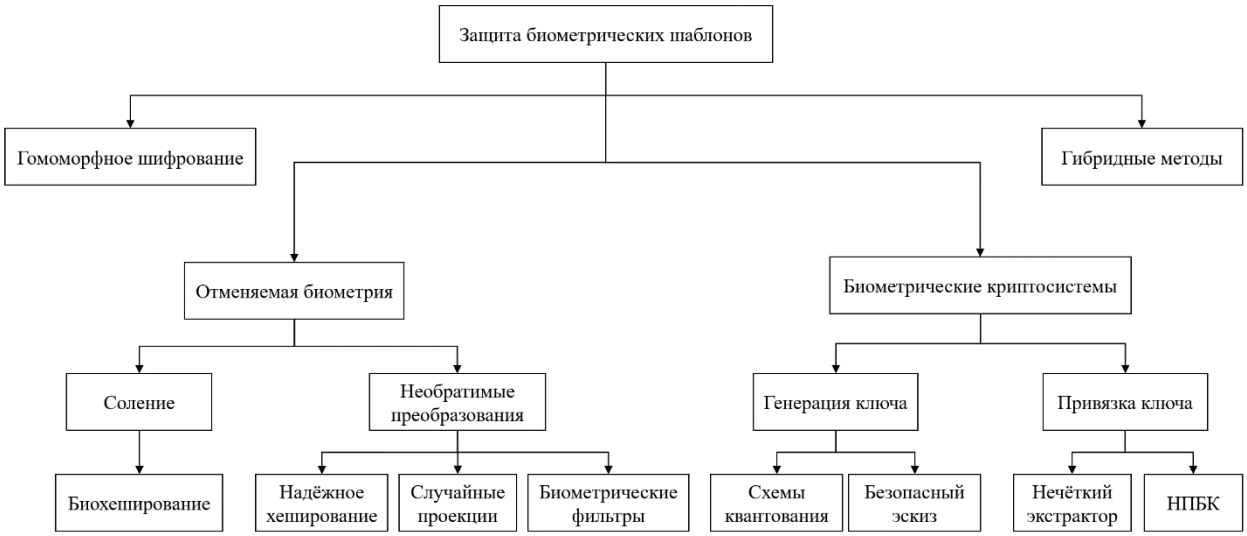


Рисунок 1.4 – Классификация методов защиты биометрических шаблонов

Одним из наиболее простых способов защиты биометрических шаблонов является использование стандартных шифров [141], таких как SHA-3. Однако из-за изменчивости биометрических шаблонов их необходимо расшифровывать перед сравнением, что снижает их эффективность по сравнению с традиционными паролями, которые можно сравнить в зашифрованной (хешированной) форме. На сегодняшний день наиболее перспективным методом шифрования биометрических шаблонов является гомоморфное шифрование

(ГШ). Этот метод позволяет применять шифрование к признакам, извлеченным из изображений лиц, с использованием предварительно обученных моделей глубоких нейронных сетей [155]. Несмотря на очевидные преимущества, вычислительная сложность ГШ ограничивает его широкое применение в задачах защиты биометрических шаблонов. Более того, алгоритм подвержен накоплению ошибок: после выполнения множества операций с зашифрованными данными результаты могут существенно отличаться от результатов операций с незашифрованными данными. Современные исследования в этой области сосредоточены на нахождении оптимального баланса между ускорением операций ГШ и минимизацией потерь в точности распознавания [32].

Отменяемая биометрия заключается в создании уникального шаблона с использованием обратимых (например, рандомизация блоков изображения) или необратимых преобразований (например, хэширования) на основе биометрических данных пользователя с возможностью отзыва и изменения этого шаблона в случае компрометации [106]. В случае компрометации биометрического шаблона (например, если шаблон украден), система может генерировать новый шаблон на основе тех же биометрических данных, но с новым набором преобразований. Таким образом, можно генерировать различные шаблоны при неизменности биометрических данных. Отменяемая биометрия имеет определенные ограничения, в частности, при «солении» шаблон может перестать быть безопасным, если вспомогательные данные скомпрометированы [106].

Альтернативным подходом к решению проблем конфиденциальности биометрических шаблонов, объединяющим в себе преимущества биометрии и криптографии, являются биометрические криптосистемы (БКС) [101]. Биометрические криптосистемы можно рассматривать как частный случай отменяемой биометрии, однако они не всегда предусматривают возможность простой замены шаблонов в случае их компрометации. В рамках обеспечения безопасности биометрических данных в биометрических криптосистемах дополнительно решаются специфические задачи управления криптографическими

ключами (генерация ключей с обеспечением необходимой длины и энтропии, а также обеспечение безопасности хранения ключей). Одной из центральных задач при проектировании биометрических криптосистем становится продуцирование ключа максимально возможной длины, с учетом требований криптографии. В этом отношении важным понятием является энтропия, которая понимается как степень случайности ключа, формируемого на выходе биометрических криптосистем, в случае попытки идентификации/аутентификации нелегитимным пользователем.

Биометрические криптосистемы делятся на два класса в зависимости от принципа функционирования: генерирующие ключи из биометрического шаблона и связывающие ключ с биометрическим шаблоном. В случае генерации ключей в основу работы БКС закладывается принцип квантования биометрических данных: векторы признаков нескольких биометрических образцов раскладываются на интервалы элементов признаков, затем полученные интервалы кодируются и сохраняются в виде вспомогательных данных [16].

Альтернативным подходом является использование схем защиты биометрических шаблонов на основе БКС, которые позволяют связывать исходные биометрические данные с определенным извне криптографическим ключом. Примером такого подхода является схема нечеткого обязательства, при которой случайное кодовое слово, специфичное для пользователя, связывается с биометрическим шаблоном этого пользователя. Другим известным типом привязки является схема нечеткого контейнера, где случайно выбранное, уникальное для пользователя кодовое слово служит набором коэффициентов для секретного полинома, а элементы биометрического шаблона используются как входные данные полинома. Эти методы можно классифицировать как "нечеткие экстракторы".

Также для связывания внешнего кода с биометрическим образом могут применяться глубокие нейронные сети (ГНС). В этом контексте задача сводится к присвоению каждому пользователю системы распознавания лиц случайного двоичного кода с максимальной энтропией, а затем к обучению нейронной сети

(сверточной, глубокой или рекуррентной) для сопоставления изображения лица пользователя с соответствующим кодом. Обученная нейронная сеть должна эффективно обрабатывать естественные вариации биометрических образов от одного и того же пользователя, обеспечивая воспроизведение одного и того же кода как при регистрации, так и при каждой попытке аутентификации. Двоичные коды, полученные при регистрации и аутентификации, криптографически хешируются, и сравнение производится на основе точного совпадения хэшей. Современные улучшения этого подхода сосредоточены на добавлении дополнительной, специфичной для пользователя информации, чтобы усложнить сопоставление между биометрическими признаками и предопределенным кодом. Однако одним из основных недостатков такого подхода является сложность автоматического обучения нейронных сетей, так как процесс обучения обычно основывается на методе градиентного спуска, который трудно автоматизировать. Кроме того, итерационные методы обучения сетей имеют тенденцию к переобучению, что делает их менее надежными в реальных условиях.

Основное преимущество методов рассмотренных методов ЗБШ на основе прямого сопоставления внешнего кода и биометрического образа пользователя путем обучения НС, заключается в том, что результирующие защищенные шаблоны (т.е. криптографические хэши предварительно определенных кодов) принципиально не связаны с исходными (незащищенными) шаблонами лиц. Это означает, что защищенные шаблоны не должны раскрывать информацию о шаблонах лиц, которым они были назначены. Следовательно, если злоумышленник получит доступ к защищенным шаблонам, хранящимся в базе данных системы распознавания лиц, он не сможет восстановить соответствующие изображения или черты лица. Конечно, это основано на предположении, что доступ к параметрам обученной НС не позволит злоумышленнику обнаружить какую-либо связь между шаблоном лица и его хешем. Однако такое предположение может не подтвердиться на практике, особенно если рассматривать наихудший сценарий полностью информированного злоумышленника. Основная проблема с методами ЗБШ, которые полагаются на

заранее определенные выходные данные, заключается в том, что НС необходимо будет переобучать (полностью или частично) каждый раз, когда новый пользователь хочет зарегистрироваться в системе распознавания лиц или когда скомпрометированный пользователь должен быть повторно зарегистрирован с новым защищенным шаблоном. Кроме того, для обучения глубоких нейронных сетей требуются большие наборы данных, хранения которых не является безопасным.

В свою очередь, одной из базовых концепций построения биометрических криптосистем, позволяющих работать с малыми выборками биометрических образов, можно считать концепцию нейросетевого преобразователя «биометрия-код» (НПБК) [144]. Концепция представлена в рамках исследований *высоконадежной биометрической аутентификации,* основные положения которой отражены в серии стандартов ГОСТ Р 52633 [25]. НПБК – это «черный ящик», построенный на основе искусственной нейронной сети, «знающий» своего владельца и надёжно хранящий его пароль или криптографический ключ. Нейросетевой преобразователь обучается формировать и отдавать пользователю его пароль (ключ) при предъявлении соответствующего биометрического образа «Свой» (легитимного образа). При предъявлении образа любого другого субъекта или состязательного примера (образа «Чужой») НПБК должен формировать случайный бинарный код с высокой энтропией. Это свойство позволяет реализовать защиту от состязательных атак. На практике могут предъявляться требования к длине и информационной энтропии ключа.

Спецификой построения НПБК является использование широких нейронных сетей, которые строятся индивидуально для каждого пользователя, при этом формируется нейронная сеть, количество входов которой равно числу признаков биометрического образа, а количество выходов – длине его персонального ключа. Каждый нейрон последнего слоя генерирует один или более бит. Процедура обучения первой реализации НПБК (рис. 1.5), доведенной до реальной практики, представлена в ГОСТ Р 52633.5-2011 [6]. Эти стандарты строятся на базе алгоритма автоматического обучения широких нейронных сетей

с одним или двумя слоями. Из-за малого числа слоев такие сети имеют не очень высокую способность к обобщению и способны работать только с предварительно обработанными данными – вектором признаков с нормальным распределением значений. Недостатком схемы является подверженность некоторым специфическим типам атак [34, 108], недостаточная длина ключа для ряда приложений либо не очень высокая точность классификации по сравнению с передовыми методами машинного обучения.

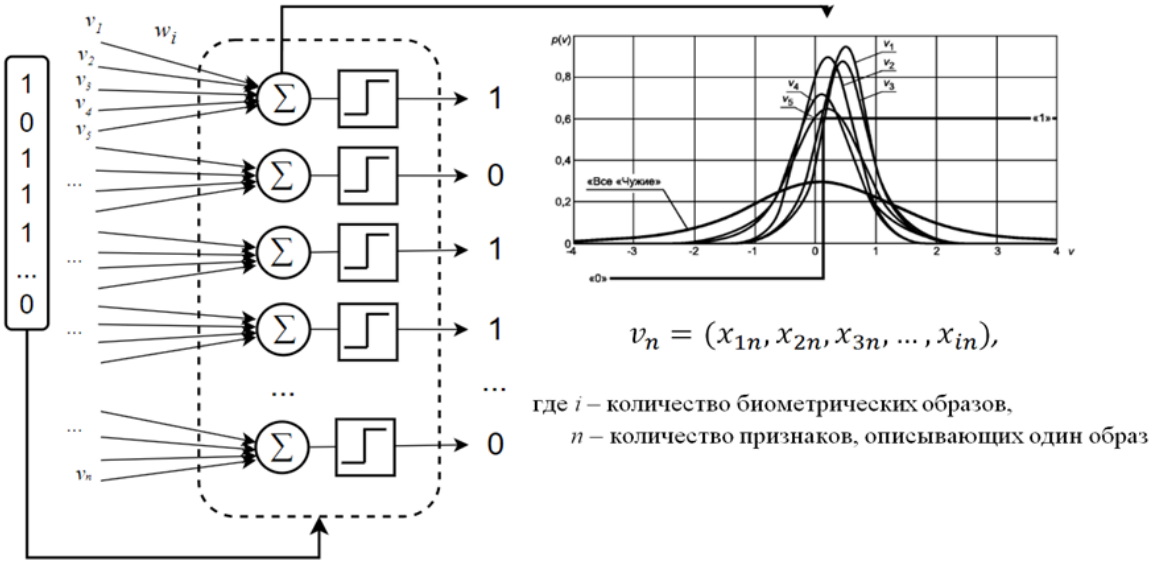


Рисунок 1.5 – Принцип работы НПБК в соответствии с ГОСТ Р 52633.5-2011

Среди наиболее значимых модификаций классической схемы построения НПБК можно выделить преобразователь на основе корреляционных [144] нейронов. Указанная модель работает с парами признаков (вместо исходного вектора признаков биометрического образа) и анализирует степень их коррелированности (положительно коррелированные, независимые или отрицательно коррелированные). Такой подход позволяет эффективно работать с биометрическими модальностями, характеризующимися сильной внутренней корреляцией признаков и создавать НПБК, параметры которого не компрометируют обучающую выборку и криптографические ключи. Однако корреляционные нейроны затруднительно применять, если биометрический образ

пользователя (биометрический шаблон) не имеет достаточное количество пар сильно коррелированных признаков. Такая особенность свойственна образам лиц, что приводит к необходимости создания альтернативного подхода к биометрической аутентификации по лицу (а также многим другим модальностям).

Создание НПБК на основе глубокой нейронной сети, обучающейся с помощью алгоритмов оптимизации на базе градиентного спуска, пока что является сложной задачей, так как, как отмечалось ранее, такие алгоритмы сложно автоматизировать (система должна обучаться на данных пользователя без участия инженера). Так, в работе [158] была предпринята попытка создания НПБК для лицевой биометрии на основе метода обратного распространения ошибки. Авторы отмечают еще один важный недостаток таких реализаций: необходимость полного переобучения всей сети при регистрации нового пользователя. Аналогичные трудности характерны для решения, представленного в работе [119]. Общим недостатком подобных систем является низкая робастность.

Альтернативой НПБК является нечеткий экстрактор [49] (в зависимости от особенностей реализации схемы защиты биометрического шаблона встречаются другие названия – нечеткого обязательства, нечеткого хранилища, нечеткого вложения). Так, в работе [127] авторы объединяют улучшенную схему нечеткого хранилища с глубокой нейронной сетью, позволяющей извлекать информативные признаки из биометрических образов, при этом достигая низкого уровня ошибки первого рода (FMR < 0.01%) и приемлемого уровня ошибки второго рода (FNMR <1%). В работе представлена всесторонняя оценка производительности комбинаций различных методов квантования и бинаризации в экспериментах с перекрестными базами данных на двух общедоступных базах данных лиц (FERET и FRGCv2). Исследование [62], в свою очередь предлагает схему BTP, основанную на схеме нечеткого обязательства, которая использует евклидово расстояние в качестве метрики для расчета оценки соответствия биометрических шаблонов. В работе [51] предлагается криптосистема идентификации лиц, состоящая из двух подсистем: поиска «один ко многим» и схемы «очищенного»

нечеткого хранилища, причем первая подсистема использует хеширование IoM для сжатия и защиты черт лица, а вторая решает проблему равномерного смешивания подлинных и «лишних» наборов в схемах нечеткого хранилища. Несмотря на высокие показатели точности идентификации на тестовых данных, в работе ярко проявлены ограничения, связанные с зависимостью от распознавания черт лица и поиском компромисса в настройке параметров предложенной системы. Исследования показывают [105], что схемы нечеткого экстрактора обычно уступают НПБК как в отношении длины криптографического ключа, так и в точности распознавания. Важно также отметить, что извлечение признаков с помощью глубоких архитектур нейронных сетей не равнозначно процессу генерации ключа на основе биометрических образов. Так, в исследовании [131] авторы осуществляют преобразование векторов признаков в бинарный код (ключ). Однако, как показывает практика, в таком случае выходы экстрактора признаков всегда будут коррелированы с входными биометрическими образами, что способно нарушить требования к энтропии криптографических ключей. Извлечение признаков может выступать исключительно в качестве предварительного этапа обработки биометрического образа.

На основании проведённого анализа можно заключить, что нейросетевые преобразователи «биометрия-код» (НПБК), которые позволяют эффективно работать с малыми выборками данных и обеспечивать высокую степень защиты данных от компрометации, играют важную роль в развитии методов защиты биометрических шаблонов и защищенной биометрической аутентификации. Однако использование глубоких нейронных сетей для создания таких преобразователей остается вызовом из-за сложности автоматизации процесса обучения и рисков переобучения. Таким образом, дальнейшие исследования в области ЗБШ на основе НПБК должны быть направлены на разработку методов, способных обеспечить надежную защиту биометрических данных при минимальных вычислительных затратах и высокой устойчивости к различным видам атак.

Выводы по первой главе

Проблема защиты искусственного интеллекта включает в себя все ключевые аспекты кибербезопасности. Атаки, направленные на данные, которыми оперирует ИИ, или на модели и алгоритмы принятия решения, способны нанести серьезный ущерб различным сервисам и критически важным приложениям на основе ИИ. В этой связи, интеграция принципов доверенного искусственного интеллекта во все этапы жизненного цикла моделей способствует повышению защищённости указанного рода технологий и обеспечивает соответствие базовым нормативным требованиям безопасности, что крайне важно для широкого принятия технологий искусственного интеллекта.

Технологии ИИ в задачах лицевой биометрической аутентификации широко используются благодаря высокой точности и скорости обработки. Современные решения включают в себя использование предобученных моделей глубоких нейронных сетей для детекции лиц и извлечения биометрических признаков. Эти системы могут достигать высокой точности при распознавании лиц в различных условиях функционирования. Основные ограничения текущих решений связаны с необходимостью затраты значительных вычислительных ресурсов и возможностью осуществления таких атак на модели, как спуфинг.

Защищенная биометрическая аутентификация по лицу, в том числе на базе ИИ, строится на методах защиты биометрических шаблонов, среди которых выделяются 3 ключевых направления: биометрические криптосистемы, отменяемая биометрия и гомоморфное шифрование. На фоне принципиальных недостатков отменяемой биометрии и гомоморфного шифрования, БКС способны обеспечивать надежное связывание биометрического образа с криптографическим ключом, заменяющим длинные и ненадежные пароли. Одним из ключевых направлений развития БКС на сегодняшний день являются нейросетевые преобразователи «биометрия-код» (НПБК).

Концепция НПБК относится к высоконадежной биометрической аутентификации, критерии которой определены в серии стандартов ГОСТ Р

52633. Анализ существующих вариантов исполнения НПБК позволил выделить ряд недостатков: низкая длина ключа классического НПБК, обучаемого в соответствии с ГОСТ Р 52633.5-2011, и низкая эффективность в работе с лицом человека НПБК на базе корреляционных нейронов.

Кроме того, проведенный анализ проблемы защищенного исполнения искусственного интеллекта в задачах биометрической аутентификации субъектов продемонстрировал принципиальную необходимость изменения логики функционирования и концептуального исполнения процедуры биометрической аутентификации по лицу на основе НПБК с целью повышения ее защищенности по отношению к деструктивным воздействиям, вроде атак на биометрическое предъявление.

Исходя из вышеизложенного, была сформулирована цель исследования, заключающаяся в необходимости повышения защищенности процедуры биометрической аутентификации личности на основе НПБК, использующего открытые биометрические образы лица человека.

Для достижения поставленной цели необходимо решить **следующие задачи:**

1. Разработать концепцию защищенной биометрической аутентификации по лицу на основе НПБК, устойчивой к внешним воздействиям в виде атак на биометрическое предъявление (спуфинг атак).

2. Разработать модель нейрона и основанную на ней модель нейросетевого преобразователя «биометрия-код», осуществляющих процедуру биометрической аутентификации по лицу с обеспечением защиты знаний и биометрических образов лиц от компрометации.

3. Разработать алгоритмы обучения нейросетевого преобразователя биометрических образов лица в код на малых выборках.

4. Разработать систему биометрической аутентификации по лицу, устойчивую к атакам извлечения знаний НПБК и компрометации открытых биометрических образов лиц, а также к атакам на биометрическое предъявление (спуфинг атак).

Глава 2. Обеспечение защиты высоконадежной лицевой биометрической аутентификации от атак на биометрическое предъявление

2.1. Методы и алгоритмы анализа подлинности изображений лиц

Методы и алгоритмы анализа подлинности изображений лиц представляют собой совокупность технологий, направленных на выявление и предотвращение использования поддельных изображений в системах биометрической аутентификации. Современные подходы к данной задаче основываются на использовании машинного обучения и глубоких нейронных сетей для детектирования аномалий, характерных для синтетических биометрических образов. Однако их применению предшествовало разнообразие альтернативных подходов, основанных на детерминированных алгоритмах [14]. Выделим 5 базовых направлений определения подлинности изображений лиц (liveness detection), поступающих на вход системы биометрической аутентификации:

1. Подходы на основе анализа 2D изображений.
2. Динамические подходы (на основе анализа последовательности изображений видеопотока).
3. Подходы на основе анализа глубины изображений.
4. Подходы на основе физиологических характеристик человека.
5. Подходы на основе глубокого обучения.

Одним из ранних направлений в области определения поддельных биометрических образов стало применение алгоритмов анализа текстуры изображения [43, 74, 102]. Этот подход предполагает, что поддельные изображения имеют отличия от реальных образов в аспектах текстуры и детализации. Для обнаружения таких различий используются различные дескрипторы изображений, которые извлекаются из кадров. В общем случае дескрипторы применяются в компьютерном зрении для описания визуальных особенностей изображения, таких как форма, цвет и текстура.

Наиболее распространенным типом дескрипторов для анализа текстуры являются локальные бинарные паттерны (Local Binary Patterns, LBP). Этот дескриптор используется для описания локальных текстурных шаблонов изображения и вычисляется как двоичное представление разности интенсивностей пикселей в локальной окрестности. В работе [46], где впервые рассматривалась возможность применения LBP-дескрипторов для определения живого присутствия, авторам удалось достичь относительно низкого уровня ошибок первого и второго рода (EER = 2,9%), продемонстрировав преимущества LBP по сравнению с устаревшими методами, такими как LPQ-дескрипторы и фильтры Габора. Подобное исследование представлено в работе [43].

Следующим этапом развития методов обработки текстуры стало использование частотного анализа в сочетании с LBP-дескрипторами, как показано в работах [74] и [46]. В работе [46] изображения сначала подвергались частотному анализу с использованием преобразования Фурье, прежде чем извлекались LBP-дескрипторы.

В задачах определения живого присутствия также используются и другие дескрипторы, такие как SIFT [116], SURF [36], DoG [152] и HOG [75], для извлечения паттернов спуфинга из различных цветовых пространств (RGB, HSV и YCbCr). Например, в работе [75] исследователи применяли гистограмму направленных градиентов (HOG) и анализировали контекст изображения и его границы, что позволило достичь низкого значения EER = 1,1% на наборе данных CASIA Anti-Spoofing.

К подобным методам анализа текстуры также относится исследование [116], посвященное обнаружению муаров – специфических узоров, возникающих при наложении двух периодических сеток. Хотя этот метод ограничен изображениями, полученными при демонстрации экрана цифрового устройства, он демонстрирует высокие показатели HTER = 6% на наборах данных DIAP, CASIA и RAFS, используя LBP и Dense SIFT дескрипторы.

В дополнение к анализу отдельных изображений, в области определения живого присутствия также применяются динамические подходы, основанные на

последовательном анализе видеопотока. Классическими примерами являются методы обнаружения микродвижений [27] и моргания [81, 114]. В работе [114] использовался метод условных случайных полей (CRF) для бинарной классификации кадров (глаза открыты/закрыты) на основе временных зависимостей. Другой динамический подход включает анализ изменений оптического потока [27, 31], предполагая, что оптический поток реальных и поддельных изображений имеет значительные различия, которые можно зарегистрировать.

Более эффективными методами предотвращения 2D и 3D спуфинг атак являются подходы, основанные на оценке глубины изображения или видеокадра [28, 33]. Например, вычисление времени, необходимого свету для прохождения от объекта к камере, позволяет получить карту глубины изображения. Однако этот метод требует специального оборудования, что не всегда возможно для приложений биометрической аутентификации.

Оценка глубины по одному RGB-изображению представляет собой сложную задачу в компьютерном зрении. В последние годы многие исследователи обучают глубокие нейронные сети на больших наборах данных RGB-D для решения этой задачи, включая реконструкцию трехмерной модели лица по двумерным изображениям [71, 130].

Некоторые исследования также используют специфические дескрипторы формы лица [76, 152] или трехмерную реконструкцию лица [163] для отличия реального лица от 3D-маски. Эти методы не требуют специального оборудования, однако их эффективность может снижаться при использовании высококачественных 3D-масок.

Отдельное направление в определении подлинности субъекта связано с измерением показателей, характерных для живого человека, таких как частота сердечных сокращений (ЧСС) [83, 85] или температура лица [111] (с использованием тепловизора или инфракрасного датчика). В работах, посвященных измерению ЧСС, применяются методы фотоплетизмографии, которые могут достигать значения EER близкие к нулю, особенно в сочетании с

сверточными нейронными сетями [85]. Эти методы могут использовать обычную RGB камеру, в отличие от инфракрасных технологий.

С появлением методов глубокого обучения (МГО) задача определения живого присутствия перестала быть трудоемким процессом, сопряжённым с ручным извлечением признаков поддельных изображений. Теперь глубокие нейронные сети (ГНС) могут самостоятельно обнаруживать релевантные признаки, включая текстурные и контекстуальные особенности изображений настоящих и поддельных лиц. Кроме того, ГНС могут быть дополнительно улучшены методами регуляризации и предобучением на больших наборах данных, что увеличивает их способность к обобщению и устойчивости к различным типам атак.

Уже в 2014 году в исследовании [172] было предложено первое полноценное решение для систем определения живого присутствия на основе глубокого обучения, использующее 8-слойную неглубокую свёрточную нейронную сеть (CNN) для извлечения признаков. С тех пор стали регулярно появляться работы [61, 100, 143, 179], в которых применялись предобученные глубокие модели, такие как VGG16 или ResNet18, адаптированные для задач определения живого присутствия с помощью трансферного обучения. Для комплексной оценки различных методов определения живого присутствия, использующих глубокое обучение, представлена следующая классификация таких методов [179]:

1. Гибридные методы: извлечение признаков с помощью классических методов с последующим применением глубоких моделей.

2. Традиционные методы обучения с учителем: так называемые end-to-end решения. Определение живого присутствия осуществляется путем применения методов глубокого обучения, чаще всего одной глубокой нейронной сети.

3. Методы, направленные на повышение обобщающей способности моделей глубокого обучения: подразумевают генерализацию моделей в отношении новых условий работы модели (освещение, качество изображения и др.) или новых типов атак.

4. Методы на основе дополнительной информации: используют специальные сенсоры или дополнительные модели для получения информации о входном изображении в иных диапазонах (инфракрасное излучение, тепловое излучение) или измерениях (карты глубины).

Наиболее простым вариантом обнаружения спуфинг атак с помощью метдов глубокого обучения является подход, при котором сначала извлекаются признаки из входных данных лица с помощью традиционных методов, а затем используется глубокое обучение для их семантического представления. Так, например, в работе [37] авторы используют LBP в качестве локальных дескрипторов, а затем работают с ними с помощью случайного леса. Стоит отметить, что исследователи не применяют CHC и при этом демонстрируют достаточно высокую эффективность на примере эталонного набора данных REPLAY-ATTACK [43]. Однако ключевым недостатком гибридных методов остается вся та же необходимость предварительного ручного извлечения признаков, свойственная традиционным подходам, а значит увеличение времени и ресурсов для обучения и настройки системы.

В большинстве работ, посвященных определению живого присутствия с использованием традиционных методов глубокого обучения, решение проблемы сводится к задаче бинарной классификации [172]. В таком случае модель обучается дифференцировать входные изображения по принципу «фальшивое лицо» (класс 1) и «реальное лицо» (класс 2). Такой подход является одним из наиболее простых и эффективных с точки зрения сходимости модели, однако не лишен недостатков: бинарная модель плохо поддается интерпретации, а выученные ею признаки сложно понять и использовать для улучшения производительности. Однако бинарная классификация способна демонстрировать довольно высокие результаты в случае учета временных характеристик распознавания (потока образов), что было продемонстрировано в работе [170] на примере модели FasTCo.

Отчасти с указанной проблемой способны справляться методы на основе попиксельного контроля (pixel-wise supervision) [59, 124, 180, 189] – подхода к

обучению глубоких моделей, результатом которого становятся попиксельно маркированные изображения обучающего набора данных. Маркировка может осуществляться с целью получения карт псевдоглубины (pseudo depth labels) [118, 180], карт отражений (reflection maps) [180], карт бинарных масок (binary mask label) [59] или карт 3D облака точек (3D point cloud map) [82]. Идея получения таких различных карт при попиксельном контроле основывается на предположении о том, что подавляющее большинство спуфинг атак основаны на предъявлении системе двумерных изображений (распечаток или экранов устройств), в отличие от которых реальное лицо является объемным. Очевидно, что такой подход оказывается ограниченным с точки зрения работы с 3D атаками (масками), в связи с чем извлечение карт иного рода, например карт бинарных масок [59], являются более предпочтительными для получения робастных моделей.

Одной из первых реализаций подхода стала архитектура DepthNet [28], до сих пор активно использующаяся для практических целей и работящая по принципу извлечения карт глубины (псевдоглубины) из входных изображений. Замена стандартных сверточных слоев DepthNet на специально разработанные для задачи антиспуфинга слои центральной разностной свертки (central difference convolution (CDC)) позволили авторам работы [181] разработать новую архитектуру CDCN, демонстрирующую более высокие показатели точности распознавания спуфинг атак на протоколе №1 набора данных OULU-NPU [35] (ACER = 1.3%).

Однако высокие показатели метода попиксельного обучения на небольших наборах данных, вроде OULU-NPU [35] или SiW-M [98], трудно считать показательными, в особенности после появления одного из самых масштабных датасетов для задачи антиспуфинга – CelebA-Spoof [189]. Авторы коллекции утверждают, что одной из ключевых проблем развития моделей для обнаружения спуфинга являются наборы данных, плохо отражающие разнообразие спуфинг атак и их модификаций. С целью исправления ситуации был собран крупномасштабный набор данных и произведена качественная аннотация

биометрических образов и атак. Дополнительно в работе [189] предложена одна из наиболее популярных моделей для задачи антиспуфинга AENet, за счет попиксельного обучения и механизмов внимания достигающая точности 99,6% на том же наборе данных.

Методы, направленные на повышение обобщающей способности моделей глубокого обучения, являются одними из самых перспективных среди современных подходов борьбы со спуфинг атаками. Актуальность направления связана с тем, что ГНС в принципе обладают слабой обобщающей способностью и часто «переучиваются» на специально подобранных наборах данных. Более того, согласно последним исследованиям [179], значительная часть работ по-прежнему опирается на небольшой пул устаревших наборов данных, которые трудно считать репрезентативными и пригодными для обучения моделей, предназначенных для работы в реальных условиях. В этом смысле показательными являются исследования, основанные на методах обучения с нулевым выстрелом (zero-shot learning) или с несколькими выстрелами (few-shot learning) [123], а также на методах обнаружения аномалий, в частности однокласовой классификации. Так, например, в работе [60] авторы используют одноклассовый классификатор на основе многоканальной сверточной нейронной сети. Применение одноклассовой константной потери (one-class constative loss) для обучения классификатора позволило добиться однозначного разделения спуфинг образов и реальных изображений в пространстве векторных представлений. Такой подход обеспечивает возможность определения новых атак в реальных условиях. Аналогичного эффекта добиваются исследователи в работе [84], в качестве функции потерь использующие Hypersphere Loss Function. Применение функции позволяет добиться специфического распределения реальных образов в гиперсфере радиусом r и однозначно определять остальные образы как поддельные.

Наконец рассмотрим мультимодальный подход [58], применяющийся в настоящем исследовании и основанный на использовании дополнительной информации для обучения глубоких моделей. Основным достоинством подхода

является возможность объединения информации из разных модальностей, что позволяет компенсировать недостатки каждой из них и дополнительно использовать различные алгоритмы и модели для анализа каждой из модальностей. Это позволяет оптимально использовать специализированные алгоритмы для обработки конкретного типа данных, улучшая общую производительность и точность системы. Кроме того, разные методы спуфинга могут быть направлены на уязвимости одной модальности, но крайне трудно обмануть систему, способную анализировать несколько модальностей одновременно.

Наиболее простым вариантом реализации подхода является использование специальных сенсоров для получения изображений в альтернативных диапазонах и обучение ГНС на полученных изображениях совместно со стандартными (RGB) [90]. Очевидны недостатки такого решения, связанные с необходимостью использования дополнительного оборудования. В связи с этим, дополнительной информацией для входов нейронной сети могут служить рассмотренные ранее карты, получаемые путем попиксельного контроля, например, карты псевдоглубины [184]. В данном случае неспособность карт глубины работать с 3D атаками частично компенсируется за счет дополнительного входа с классическим RGB изображением. Основной задачей для данного направления, в таком случае, становится поиск эффективных способов слияния информации от нескольких источников.

Из приведенного анализа видно, что наиболее перспективными методами противодействия спуфинг атакам в биометрических системах аутентификации, в частности в системах распознавания лиц, являются методы на основе мультимодальных подходов и попиксельного контроля. Мультимодальные подходы позволяют объединять информацию из различных источников и биометрических модальностей, что компенсирует недостатки каждой из них и повышает общую точность и устойчивость системы к атакам. Методы попиксельного контроля, такие как карты псевдоглубины и отражений, обеспечивают детализированное представление изображений, что позволяет

более эффективно различать реальные лица и подделки. Комбинирование этих методов может привести к созданию более робастных систем, способных противостоять различным видам атак, улучшая их производительность и надежность в реальных условиях.

Таблица 2.1 – Сводная таблица значимых исследований по теме обнаружения спуфинг атак

Наименование исследования	Год	Подход	Тип спуфинг атак (2D/3D)	Метод, лежащий в основе подхода	Метрика оценки эффективности проведенного исследования
Face Liveness Detection Based on Frequency and Micro-Texture Analysis [46]	2014	Анализ текстуры изображения	2D атаки	LBP дескрипторы	EER = 2.7%
Secure Face Unlock: Spoof Detection on Smartphones [116]	2016	Анализ текстуры изображения	2D атаки	SIFT дескрипторы	EER = 3.51%
Context based Face Anti-Spoofing [75]	2013	Анализ текстуры изображения и его контекста	2D атаки	HOG дескрипторы	EER = 1.1%
Motion-based counter-measures to photo attacks in face recognition [27]	2012	Динамический подход (оптический поток)	2D атаки и некоторые разновидности 3D атак	Метод на основе корреляции движений переднего и заднего фонов	HTER = 1.52%
Generalized face anti-spoofing by detecting pulse from face videos [83]	2016	Подход на основе бесконтактного определения пульса (ЧСС)	3D атаки	Метод бесконтактного определения (ЧСС)	EER = 1.58%
Face liveness detection by rPPG features and contextual patch-based CNN [85]	2019	Подход на основе бесконтактного определения пульса (ЧСС)	2D и 3D атаки	Сверточная нейронная сеть	EER = 3.4%

Visible/Infrared face spoofing detection using texture descriptors [111]	2019	Подход на основе инфракрасного излучения	2D атаки	Сверточная нейронная сеть	EER = 1.01%
Face Spoofing Detection Based on Depthmap and Gradient Binary Pattern [33]	2015	Подход на основе глубины изображения	2D атаки	Метод на основе градиентных бинарных паттернов	EER = 31%
Face Anti-Spoofing Using Patch and Depth-Based CNNs [28]	2017	Подход на основе глубины изображения	2D и 3D атаки	Сверточная нейронная сеть	EER = 0.1%
Face anti-spoofing to 3d masks by combining texture and geometry features [163]	2018	Подход на основе реконструкции 3D модели лица	3D атаки	Трехмерная морфологическая модель + сверточная нейронная сеть	EER = 0.9%
3d facial geometric attributes based anti-spoofing approach against mask attacks [153]	2017	Подход на основе оценки глубины изображения с помощью специальных дескрипторов	3D атаки	Дескрипторы на основе meshSIFT	EER = 6.72%
Learn convolutional neural network for face anti-spoofing [172]	2014	Подход на основе глубокого обучения	2D атаки	Сверточная нейронная сеть	HTER < 5%
Deep pixel-wise binary supervision for face presentation attack detection [59]	2019	Подход на основе глубокого обучения	2D атаки	Сверточная нейронная сеть	HTER = 12%
Learning deep forest with multi-scale local binary pattern features for face anti-spoofing [37]	2019	Подход на основе глубокого обучения	2D и 3D атаки	Глубокий лес	EER = 1.56%
Searching central difference convolutional networks for face anti-spoofing [181]	2020	Подход на основе глубокого обучения	2D и 3D атаки	Сверточная нейронная сеть	HTER = 6.5%
Celeba-spoof: Large-scale face anti-spoofing dataset with rich annotations [189]	2020	Подход на основе глубокого обучения	2D и 3D атаки	Сверточная нейронная сеть	HTER = 11.9%
Cross modal focal loss for rgbd face	2021	Подход на основе	2D и 3D	Сверточная нейронная	HTER = 2.6%

anti-spoofing [58]		глубокого обучения	атаки	сеть	

Отметим, что одним из перспективных, но мало изученных, направлений в задачах определения живого присутствия при биометрической аутентификации по лицу сегодня считается применение технологий объяснимости и интерпретируемости искусственного интеллекта [73]. Авторы двух имеющихся на сегодняшний день работ [135], [166] по определению живого присутствия в задачах биометрической аутентификации по лицу с помощью объяснимого ИИ, выбрали подход на основе интерпретируемости и применили в своих работах Grad-CAM, технологию, позволяющую получать объяснения для конкретных классов и предоставлять объяснения для каждого слоя сети.

Однако стоит отметить, что защищенное исполнение и объяснимый искусственный интеллект (ИИ) представляют собой противоположные концепции. Защищенное исполнение основывается на минимизации утечек информации и обеспечении конфиденциальности данных, что достигается через ограничение доступа к внутренним процессам и механизмам принятия решений ИИ. Напротив, объяснимый ИИ направлен на предоставление пользователю прозрачных объяснений относительно того, как и почему была принята та или иная рекомендация или решение. Это требует раскрытия значительной части внутренней логики и процессов ИИ, что повышает информированность и доверие, но потенциально увеличивает риски утечек информации.

2.2. Обзор открытых наборов данных изображений и видеозаписей лиц для определения подлинности лица

Современные исследования в области лицевой биометрической аутентификации активно используют открытые наборы данных изображений и видеозаписей лиц для разработки и оценки алгоритмов определения подлинности лица. Специализированные наборы данных (табл. 2.2), такие как Replay-Attack, MSU-MFSD и CelebA-Spoof, сфокусированы на задачах определения подлинности

лиц и детекции спуфинг-атак. Эти наборы данных включают видеозаписи различных атак, таких как использование фотографий, видео и 3D-масок. Наличие подобных специализированных наборов данных способствует возможности проведения экспериментальных оценок моделей глубокого обучения, направленных на выявление широкого спектра спуфинг атак.

Таблица 2.2 – Сравнение открытых наборов данных изображений и видеозаписей лиц для определения подлинности лица

Набор данных	**Год**	**Соотношение классов Live/Spoof (И – изображение, В – видео)**	**Типы атак**
CASIA-MFSD [190]	2012	150/450(В)	Распечатка (плоская, завернутая, обрезанная), воспроизведение (планшет)
REPLAY-ATTACK [43]	2012	200/1000(В)	Распечатка (плоская), воспроизведение (планшет, телефон)
MSU USSA [117]	2016	1140/9120(И)	Распечатка (плоская), воспроизведение (ноутбук, планшет, телефон)
OULU-NPU [35]	2017	720/2880(В)	Распечатка (плоская), воспроизведение (телефон)
CASIA-SURF [186]	2019	3000/18000(В)	Распечатка (плоская, завернутая, обрезанная)
WMCA [56]	2019	347/1332(В)	Распечатка (плоская), воспроизведение (планшет), частичное изменение (очки), маски (пластик, силикон, бумага, манекен)
CeFA [88]	2020	6300/ 27900(В)	Распечатка (плоская, завернутая), воспроизведение, Маски (3D печать, силикагель)
HQ-WMCA [64]	2020	555/2349(В)	Распечатка (плоская), воспроизведение (планшет, телефон), маски (пластик, силикон, бумага, манекен), макияж, частичное изменение (очки, парик, тату)
CelebA-Spoof [189]	2020	156384/469153(И)	Распечатка (плоская, завернутая), воспроизведение (монитор, планшет, телефон), маска (бумага)
PADISI-Face [129]	2021	1105/924(В)	Распечатка (плоская), воспроизведение (планшет, телефон), частичное изменение (очки), маска (пластик, силикон, манекен)

HiFiMask [89]	2021	13650/40950(B)	Маска (резина)
SiW-M [98]	2022	785/915(B)	Распечатка (плоская, завернутая), воспроизведение (монитор, планшет, телефон), маска (пластик, силикон, бумага, манекен), макияж, частичное изменение (очки, парик, тату)

Для проведения экспериментальных исследований, описанных ниже, на основу был взят один из немногих открытых наборов данных – CelebA-Spoof [189]. Набор данных CelebA-Spoof представляет собой обширную и тщательно аннотированную коллекцию изображений, специально созданную для задач анти-спуфинга в системах распознавания лиц. Набор данных включает 625537 изображений 10177 субъектов, которые охватывают широкий спектр реальных лиц и различных типов атак. Каждое изображение в наборе данных сопровождается подробными аннотациями, включающими метки спуфинга (истинное или поддельное лицо), тип атаки (например, использование фотографий, видео или масок) и другие релевантные атрибуты. Изображения в наборе данных сделаны в различных условиях освещения, с разными позами и выражениями лиц.

2.3. Концепция защищенной биометрической аутентификации по лицу на основе нейросетевых преобразователей «биометрия-код», устойчивая к атакам на биометрическое предъявление

В связи с тем, что процедура обнаружения спуфинг образов является отдельной задачей, отличной от процедуры аутентификации, предложенная концепция (рис. 2.1) состоит из двух параллельно функционирующих блоков. Параллельное исполнение процедур обнаружения спуфинга и распознавания субъектов мотивировано необходимостью минимизации влияния ошибок классификатора спуфинг атак на результирующую процедуру аутентификации

(при последовательном построении блоков точность работы всей системы аутентификации придется свести к блоку обнаружения спуфинг атак).

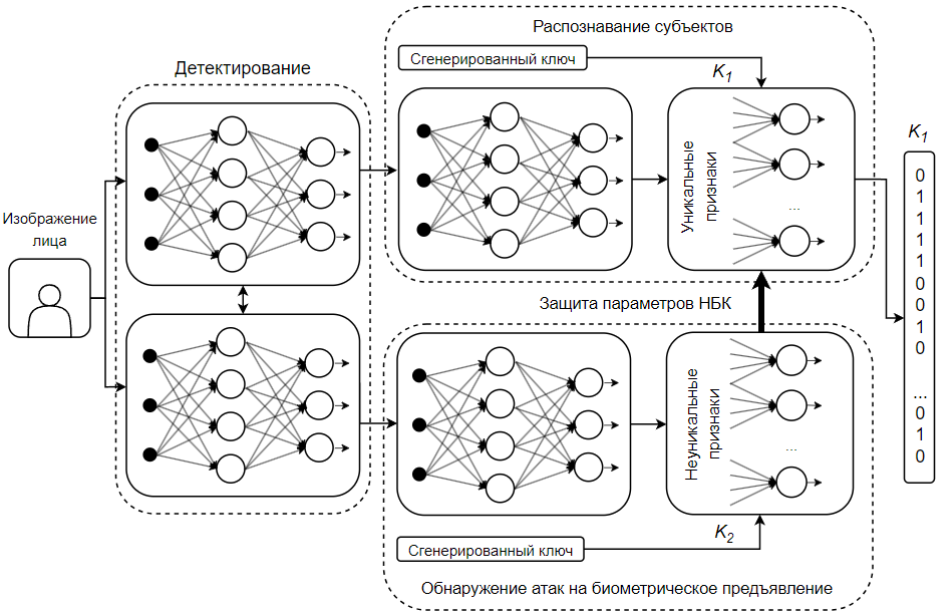


Рисунок 2.1 – Структурная схема концепции защищенной биометрической аутентификации по лицу, устойчивой к атакам на биометрическое предъявление

Оба блока концепции устроены по принципу разделения экстракторов векторного представления входного образа от классификаторов. В качестве таких классификаторов в обоих случаях используются нейросетевые преобразователи «биометрия-код», обладающие принципиальными различиями в логике функционирования и целевом назначении [15].

Нейросетевой преобразователь, отвечающий за распознавание субъектов (аутентификацию), должен обеспечивать связывание биометрического образа лица человека с длинным криптографическим ключом K_1, отвечающим требованиям высоконадежной биометрической аутентификации. Будем называть такой преобразователь *пользовательским*. Модель и алгоритмы обучения пользовательского НПБК представлены в главе 3 настоящей диссертационной работы.

Использование для задачи классификации поддельных и реальных изображений нейросетевого преобразователя, процедура обучения которого

обозначенной задаче представлена в параграфе 2.4, во многом продиктовано необходимостью обеспечения комплексной безопасности конечной процедуры биометрической аутентификации по лицу. Очевидно, что добавление решений анти-спуфинга в системы распознавания лиц только усложняет процедуру аутентификации и влечет за собой формирование потенциальных уязвимостей: на этапе обнаружения поддельных изображений может быть осуществлена кража конфиденциального биометрического образа. В этой связи, применение методов защиты биометрических шаблонов, в частности НПБК, позволяет не только распознавать спуфинг атаки, но и обеспечивать защищенный режим конечной процедуры биометрической аутентификации.

Кроме того, применение ключа, полученного от НПБК для обнаружения спуфинга в качестве дополнительного средства защиты (шифрования) нейросетевых параметров пользовательского НПБК позволяет организовывать дополнительный уровень защиты в рамках рассмотренной концепции и организовывать взаимодействие двух функциональных блоков с учетом четырех возможных векторов атак в отношении предложенной концепции (рис. 2.2).

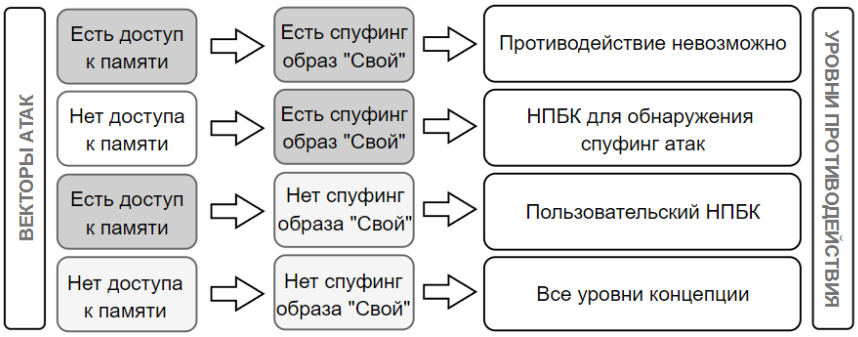


Рисунок 2.2 – Векторы атак в отношении предложенной концепции

Каждый из представленных векторов учитывает одну потенциальную комбинацию атак, исходящую из возможностей злоумышленника. Если злоумышленник планирует провести одну из актуальных для систем нейросетевой биометрической аутентификации атак – на структурные компоненты алгоритма аутентификации или атаку спуфинг – при этом не имея

фактического доступа к памяти устройства или приложения и не обладая спуфинг образом легитимного пользователя («Свой»), то предложенная концепция в целом станет эффективным способом противодействия любым альтернативным угрозам.

Наличие доступа к памяти или спуфинг образу «Свой» по-прежнему не приведет к возможности реализации ни одной из угроз за счет двух уровней обеспечения защиты. Наиболее негативный сценарий, при котором злоумышленник имеет доступ к памяти и обладает спуфинг-образом «Свой», является единственным, при котором невозможно противодействие ни на одном из уровней концепции.

2.4. Классификация атак на биометрическое предъявление с помощью модификации нейросетевого преобразователя «биометрия-код»

Глубокие нейросетевые архитектуры, решающие задачу распознавания атак на биометрическое предъявление, обладают слабой обобщающей способностью по отношению к разнообразию реализации таких угроз и часто «переучиваются» на специально подобранных наборах данных. С учетом того, что значительная часть из них по-прежнему обучается на небольшом пуле устаревших наборов данных, которые не могут обеспечить эффективное решение задачи определения подлинного присутствия, актуальность приобретают новые подходы, сразу направленные на повышение обобщающей способности моделей глубокого обучения. Такие подходы исключают недостатки классического глубокого обучения, связанные с низкой эффективностью моделей при работе на новых данных в реальных условиях: новый тип атак или отличающиеся от обучающих данные приводят к заметному снижению производительности моделей.

Одним из способов нивелирования недостатков применения глубоких нейронных сетей для обнаружения спуфинг атак является логика разделения блока векторного представления образов и блока защиты биометрического шаблона (блока классификации) [150]. Если в качестве последнего для классификации образов использовать широкие нейронные сети, способные

обучаться автоматически (без применения градиентного спуска), то можно избежать переобучения всей сети на специальном наборе данных и несколько повысить обобщающую способность модели за счет классификатора, обученного общему представлению реальных и поддельных изображений. В таком случае, спектр атак, представленный в наборе данных CelebA-Spoof, не оказывает решающего значения при обучении и тестировании предложенных моделей.

Как отмечалось ранее, в качестве описанной широкой нейронной сети может выступать нейросетевой преобразователь «биометрия-код», позволяющий достигать сразу двух ключевых целей в разработке блока обнаружения спуфинг атак: противодействие угрозам компрометации биометрических образов пользователей, проходящим через систему обнаружения, а также высокая точность классификации реальных и поддельных изображений лиц.

В рамках экспериментальной реализации блока обнаружения спуфинг атак с целью векторного представления входных изображений была обучена глубокая нейронная сеть, основанная на архитектуре FeatherNet [184]. FeatherNet — это легковесная архитектура сверточной нейронной сети, разработанная для задачи обнаружения спуфинг атак в системах распознавания лиц. В основе архитектуры лежат идеи минимизации вычислительных затрат и параметров модели без потери точности.

Одной из ключевых особенностей оригинальной архитектуры FeatherNet, предложенной в работе [184], является замена широко применяемого для снижения размерности слоя глобального усредняющего пулинга (Global Average Pooling (GAP)) на так называемый модуль потоковой передачи данных (Streaming Module), основанный на глубинной свертке (depthwise convolution) с шагом > 1. С помощью такой замены удается избегать негативных последствий применения GAP для задач распознавания лиц, связанных с усреднением всех значений карт признаков вне зависимости от степени их «важности» для конкретного примера. Кроме того, FeatherNet является примером мультимодальных архитектур, принимающих на вход не только стандартные изображения (RGB), но и

дополнительную информацию в виде карт глубины [184] и инфракрасных изображений лиц [90].

Для обучения описанной архитектуры и получения блока векторного представления (экстрактора признаков), способного дифференцировать признаки, полученные из изображений реальных и поддельных лиц, были произведены ряд модификаций в структуре сети (табл. 2.3). В первую очередь, для обучения сети кроме стандартных RGB изображений использовались только карты глубины лиц, представленные во вспомогательном наборе данных CelebA-Spoof Depth Image [39]. Исключение изображений в инфракрасном диапазоне из процедуры обучения обосновано отсутствием таковых для экспериментального набора данных CelebA-Spoof.

Таблица 2.3 – Архитектура модифицированной сверточной нейронной сети для извлечения признаков на основе FeatherNet

Назначение	**Слой**	**Описание**	**Размерность выходного вектора**
Усреднение мульти-модальных входных значений	conv1	Блок с Conv2d_cdcn (3 входных канала, 32 выходных, ядро 3x3, шаг 2), BatchNorm2d и Hardswish активацией.	32
	avgpool	AdaptiveAvgPool2d, выполняет адаптивное усреднение по размерам входного изображения для слоя layer1.	32
Извлечение признаков	layer1	Блоки, содержащие InvertedResidual блоки, включая Downsample, Conv2d, BatchNorm2d, Hardswish и SELayer (для некоторых блоков).	16
	layer2		32
	layer3		64
	layer4		**96**
Классифи-каторы («головы»)	fc_face	Блок с Dropout (вероятность отсева 0.15), BatchNorm1d, Hardswish и Linear (входных 96, выходных 40).	40
	fc_attack	Блок с Dropout (вероятность отсева 0.15),	11

		BatchNorm1d, Hardswish и Linear (входных 96, выходных 11).	
	fc_light	Блок с Dropout (вероятность отсева 0.15), BatchNorm1d, Hardswish и Linear (входных 96, выходных 5).	5
	fc_live	Блок с Dropout (вероятность отсева 0.15), BatchNorm1d, Hardswish и AngleSimpleLinear (вычисление косинуса угла между векторами в пространстве признаков).	1
	depth	Блок с Conv2d (входных 96, выходных 1), Upsample (размер 14x14, метод 'bilinear') и Sigmoid активацией.	1

В качестве дополнительных улучшений были произведены замены функций активаций ReLU (Rectified Linear Unit) на относительно новый тип нелинейности для обучения глубоких нейронных сетей HardSwish (h-swish), впервые представленный в рамках исследований архитектуры MobileNetV3 [65]:

$$hardswish(x) = x\frac{ReLU6(x+3)}{6}$$

где ReLU6 – это функция активации ReLU, ограниченная значением 6. HardSwish является аппроксимацией swish, которая умножает входное значение x на сигмовидную функцию от этого же значения. Для современных архитектур, разработанных для мобильных и встраиваемых систем, использование функции активации HardSwish позволяет достигать высокого соотношения производительности и затрат.

Кроме того, для блока усреднения мультимодальных входных значений были изменены процедуры свертки, позволяющие извлекать пространственные признаки из входных данных: классическая 2D свертка (Conv2d) в настоящей работе заменяется на ее модификацию, предназначенную для улучшения способности модели извлекать локальные градиентные признаки из данных –

central difference convolution (CDC) [181]. Центральная разностная свертка работает за счет добавления к процедуре стандартной свертки входного изображения с фильтром *K* дополнительной информации в виде разности каждого элемента окна свертки и центрального элемента этого окна, умноженной на соответствующий элемент фильтра. Добавление описанной модификации в первые слои усреднения позволяют повысить устойчивость модели к небольшим изменениям и искажениям в данных, что является критически важным для возможности распознавания максимально приближенных к реальности спуфинг атак.

Для повышения обобщающей способности модели, а также предотвращения возможного переобучения была произведена аугментация данных тренировочного набора. Для этого входные изображения 7 видам дополнительных преобразований:

1. Добавление шума, имитирующего шум цифровых камер (ISOnoise), с определенным сдвигом цвета и интенсивностью. Применяется с вероятностью 5%.

2. Изменение яркости и контраста в заданных пределах. Применяется с вероятностью 12.5%.

3. Имитация движения, осуществляющее размытие, имитирующее движение, с ограничением размытия до 3 пикселей. Применяется с вероятностью 20%.

4. Имитация сжатия изображения, помогающее снижать его качество. Применяется с вероятностью 25%.

5. Случайное удаление фрагментов изображения (заполнение черным) в виде прямоугольных областей. Применяется с вероятностью 25%.

6. Добавление гауссовского шума. Применяется с вероятностью 20%.

7. Нормализация изображения с заданными средними значениями и стандартными отклонениями для каждого канала (RGB).

Применение указанных видов аугментации способно значительно повысить эффективность антиспуфинг систем распознавания лиц за счет увеличения

разнообразия обучающих данных и повышения устойчивости моделей к различным искажениям. Добавление шума, изменение яркости и контраста, а также имитация движения и сжатия изображений способствуют адаптации моделей к вариативным условиям съемки и различным уровням качества изображений. Случайное удаление фрагментов изображения и добавление гауссовского шума увеличивают способность моделей справляться с частичными потерями информации и шумами. Нормализация изображений обеспечивает стабильность процесса обучения, выравнивая данные и ускоряя сходимость моделей.

В качестве детектора лиц на входных изображениях использовалась предобученная на крупномасштабном наборе данных WIDERFACE [175] модель RetinaFace [48]. Модель разработана для одностадийного обнаружения лиц с высокой точностью и использует стратегии многозадачного обучения с дополнительным обучением.

Обучение полученной архитектуры сети FeatherNet осуществлялось на полном наборе данных CelebA-Spoof в течение 15 эпох с помощью функции потерь Cross Entropy для «головы» fc_live, осуществляющей бинарную классификацию реальных и поддельных изображений (рис. 2.3). Предварительно обучающие данные были разделены на тренировочную и тестовую (валидационную) выборки, каждая из которых случайным образом подвергалась аугментации в соответствии с описанными выше преобразованиями.

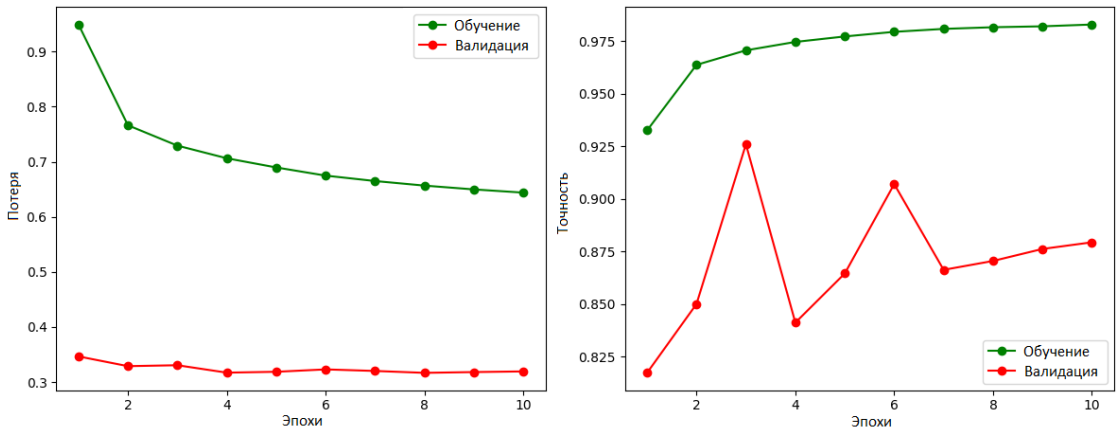


Рисунок 2.3 – Результаты обучения модифицированной архитектуры FeatherNet

На вариационной выборке максимальное значение точности работы сети составило только 93%, несмотря на то, что значения точности на тренировочных данных превышают 97%. Для дальнейшего использования обученной сети в качестве экстрактора признаков, слои классификатора «замораживались», а работа осуществлялась с 96-мерным вектором признаков на выходе предшествующего классификатору слоя.

Классификацию реальных и поддельных изображений в рамках предложенной концепции предлагается осуществлять с помощью классического нейросетевого преобразователя «биометрия-код», обученного соответствующей задаче [17]. В дополнение к преимуществам использования НПБК для задачи защищенного исполнения блока обнаружения спуфинг атак, логика его работы является легко адаптируемой под задачу бинарной классификации, так как в общем случае НПБК умеет разделять входные биометрические образы на два класса: «Свой» и «Чужой». Для этого обучение одного нейрона сводится к задаче разделения распределения откликов нейрона на образы «Свой» и распределения откликов на образы «Чужой» за счет «выталкивания» первого из второго [4]. Выталкивание осуществляется путем вычисления весовых коэффициентов нейронов по формуле:

$$\mu_i = \frac{E_{\text{Чужой}}(v_i) - E_{\text{Свой}}(v_i)}{\sigma_{\text{Чужой}}(v_i)\sigma_{\text{Свой}}(v_i)}$$

где v_i – i-ое значение вектора биометрических признаков $\bar{v}$, $E_{\text{Чужой}}(v_i)$ – математическое ожидание i-ого признака для образа «Чужой», $E_{\text{Свой}}(v_i)$ – математическое ожидание i-ого признака для образа «Свой», $\sigma_{\text{Чужой}}(v_i)$ – квадратичное отклонение i-ого признака для образа «Чужой», $\sigma_{\text{Свой}}(v_i)$ – квадратичное отклонение i-ого признака для образа «Свой». Таким образом, НПБК на уровне отдельного нейрона обучается уникальному представлению «Своего», выраженному в виде вектора признаков обучающего набора, и

однозначно отделяет полученное представление от общего распределения «Чужих» за счет полученных весовых коэффициентов.

Адаптация логики работы НПБК под задачу бинарной классификации осуществляется путем построения преобразователя для реальных изображений лиц (класс C_1). В таком случае, поддельные изображения (класс C_2) расцениваются НПБК как «Чужие», а случайный код на выходе свидетельствует о решении в пользу класса C_2. Принадлежность классу C_i, i = 1, 2, оценивается исходя из получаемого на выходе НПБК бинарного кода (ключ K_2 в случае класса C_1). Особенностью построения НПБК в качестве классификатора является допущение о возможности дублировании входов нейрона в связи с отсутствием необходимости сокрытия структуры преобразователя. Для оценки качества осуществляемой бинарной классификации *i*-ого входного образа применяется следующее правило:

$$\begin{cases} (\bar{a}_i \in C_1 \ \wedge \ h_i < \ threshold) \rightarrow \ TP \\ (\bar{a}_i \in C_1 \ \wedge \ h_i > \ threshold) \rightarrow \ FN \\ (\bar{a}_i \in C_2 \ \wedge \ h_i < \ threshold) \rightarrow \ FP \\ (\bar{a}_i \in C_2 \ \wedge \ h_i > \ threshold) \rightarrow \ TN \end{cases}$$

где h_i – *i*-ое значение расстояния Хэмминга между ожидаемым кодом и выходом НПБК, $threshold$ – порог, определяющий допустимое количество ошибок в коде *i*-ого входного образа для корректного отнесения его к одной из групп классифицированных образов: TP – True Positive, FN – False Negative, FP – False Positive или TN – True Negative. Полученный классификатор не требует итерационного обучения (осуществляется автоматически) и большого числа обучающих примеров. На основе полученных значений указанных метрик (TP, FN, FP, TN) высчитывается точность классификации по формуле:

$$accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$

Дополнительно высчитывается значение ACER (Average Classification Error Rate) – общепринятая метрика, использующаяся для оценки общей производительности антиспуфинговых систем и учитывающая ключевые ошибки – количество ложных принятий (APCER (Attack Presentation Classification Error Rate)) и ложных отказов (BPCER (Bona Fide Presentation Classification Error Rate)) – влияющие на надёжность системы.

$$ACER = \frac{APCER + BPCER}{2}$$

где $APCER = \frac{FP}{FP+TN}$, $BPCER = \frac{FN}{FN+TP}$.

Итоговый эксперимент по оценке точности работы нейросетевого преобразователя «биометрия-код» в качестве классификатора (рис. 2.4) проводился с помощью специально подготовленных для этой задачи выборок (тренировочной и тестовой) из датасета CelebA-Spoof. В случае обучающей выборки случайным образом из классов (субъектов) основного набора данных выбирались 100 изображений реальных лиц, представляющих собой класс C_1, и 100 изображений поддельных лиц, представляющих собой разнообразные атаки исходного набора данных CelebA-Spoof (класс C_2). Аналогичная процедура была произведена для получения тестовой выборки, однако полученные 200 изображений дополнительно случайным образом перемешивались. Лучший результат на тестовых выборках составил 97,2% точности (ACER=2,9%) при значении $threshold$ = 3.

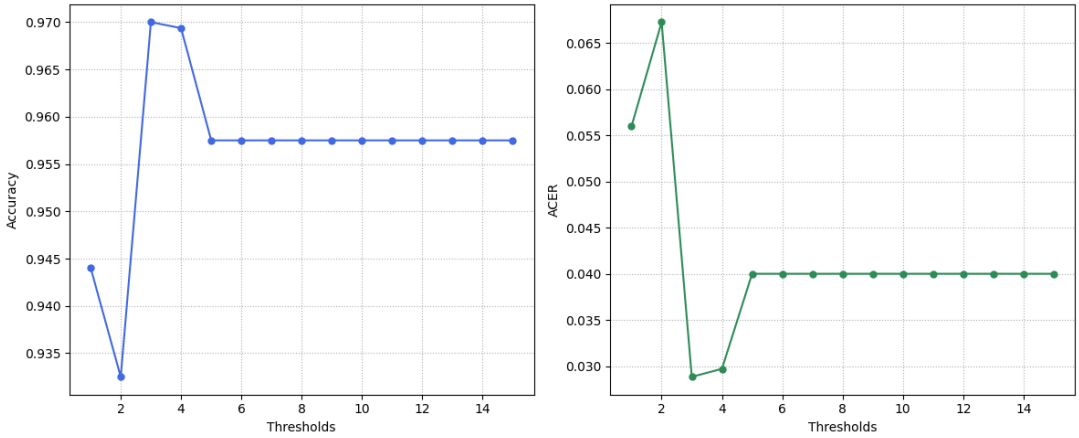


Рисунок 2.4 – Точность (accuracy) работы НПБК при разных значениях *threshold*

Результаты сравнения предложенного решения с классическими моделями глубокого обучения, используемыми для задачи обнаружения спуфинг атак, представлены в таблице 2.4. Несмотря на то, что в части исследований [59, 60, 170, 181] авторы не предоставляют информации о результирующей точности работы сети, оценку моделей можно производить исходя из метрики ACER.

Из таблицы видно, что разработанный модуль (FeatherNet + НПБК) демонстрирует точность 97,2%, что уступает только модели AENet (99,6%), однако по-прежнему показывает высокий уровень распознавания реальных и поддельных лиц. В свою очередь, значение ACER для предложенного решения составляет 2,9%, что ниже, чем у AENet (3,09%) и значительно ниже по сравнению с моделями DeepPixBiS (5,97%) и CDCN++ (1,3%). Такие показатели ACER свидетельствуют о достаточно высокой надежности модели при условии обеспечения безопасного режима работы антиспуфинг модуля.

Таблица 2.4 – Сравнительные результаты работы предложенного решения с альтернативными моделями глубокого обучения для обнаружения спуфинг атак

№	**Модель**	**Датасет**	**Accuracy**	**ACER**
1	DeepPixBiS [59]	OULU-NPU (p.1)	-	5.97%
2	CDCN++ [181]	OULU-NPU (p.1)	-	1.3%
3	AENet [189]	CelebA-Spoof Dataset	99,6%	3.09%
4	MCCNN (BCE+OCCL)-	SiW-M	-	14.9 ± 7.8%

	GMM [60]			
5	FasTCo [170]	SiW-M	-	10.1 ± 5.6%
6	MobileNetv3 [65]	CelebA-Spoof Dataset	99,8%	3.8%
7	**FeatherNet + НПБК**	**CelebA-Spoof Dataset**	**97,2%**	**2,9%**

Показатели ACER для моделей MCCNN (BCE+OCCL)-GMM и FasTCo на наборе данных SiW-M демонстрируют большую вариативность (14.9 ± 7.8% и 10.1 ± 5.6% соответственно), что указывает на их меньшую стабильность по сравнению с FeatherNet + НПБК, а также на тот факт, что глубокие нейронные сети для обнаружения спуфинг атак, зачастую, проучиваются на специализированных наборах данных и показатели их эффективности, полученные в результате обучения, могут существенно отличатся от значений, полученных в реальных условиях функционирования системы.

Несмотря на сравнительно невысокие показатели точности работы НПБК в качестве классификатора, сохраняется ключевое преимущество предложенного решения: безопасная реализация концепции защищенной биометрической аутентификации по лицу, при которой модуль обнаружения спуфинг атак не становится «слабой» точкой потенциальных уязвимостей. Кроме того, стоит учитывать, что представленные в таблице 2.4 глубокие нейронные сети для обнаружения спуфинг атак, зачастую, переучиваются на специализированных наборах данных и показатели их эффективности (accuracy и ACER), полученные в результате обучения, могут существенно отличатся от реальных значений данных показателей в реальных условиях функционирования системы.

2.5. Механизм защиты нейросетевого контейнера пользовательского нейросетевого преобразователя «биометрия-код»

Согласно ГОСТ Р 52633.4-2011 [6], структурированный блок данных, хранящий параметры обученного нейросетевого преобразователя «биометрия-код», носит название нейросетевого биометрического контейнера (НБК) и

представляет собой таблицы связей и/или весов пользовательского НПБК. В рамках предложенной концепции и необходимости использования ключа НПБК для обнаружения спуфинг атак для дополнительной защиты пользовательского НПБК, предлагается защищать нейросетевые контейнеры путем их полного или частичного сокрытия с помощью обратимых и необратимых преобразований. Такой подход лежит в основе идеи защищенного нейросетевого контейнера (ЗНК) [12].

Для реализации конкретного механизма ЗНК после обучения пользовательского НПБК нейросетевые параметры каждого нейрона необходимо шифровать наложением гаммы, представляющей ключ НПБК для обнаружения спуфинг-атак. Описанный механизм не является полным аналогом шифра Вернама, так как в рамках концепции может быть затруднительно выполнить требования по длине гаммы, которая должна соответствовать длине сообщения (НБК). Кроме того, если гамма (ключ) используется более одного раза для шифрования нескольких сообщений (НБК), метод шифрования наложением гаммы перестает быть безопасным. Это происходит потому, что повторное использование той же гаммы позволяет атакующему провести различные виды анализа для восстановления исходных сообщений или, по крайней мере, выявления определенных закономерностей. В этой связи, использование одного ключа НПБК для обнаружения спуфинг для шифрования всех пользовательских НПБК может стать уязвимостью, если ключ является секретным. Однако в рамках концепции предлагается не скрывать указанный ключ, так как его рассекречивание способно обезвредить только первый уровень системы (блок обнаружения спуфинг атак), а для преодоления второго уровня злоумышленнику дополнительно произвести ряд дополнительных действий в отношении блока аутентификации. Кроме того, скомпрометированный ключ можно легко заменить, переобучив НПБК для обнаружения спуфинг атак.

В режиме ЗНК энтропия выходов пользовательского НПБК при поступлении образа «Чужой» или спуфинг-образа повышается [1]. Это объясняется тем, что если часть нейронов НПБК для обнаружения спуфинг атак

ошибается, и на его выходе получается неверный ключ (гамма), таблицы параметров пользовательского НПБК дешифруются неверно (рис. 2.5). Это приводит к тому, что нейроны с неверно восстановленными таблицами связей дают почти случайный отклик.

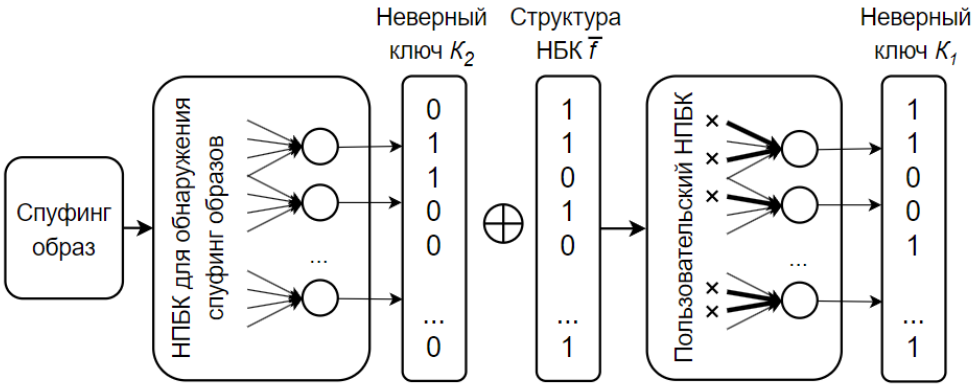


Рисунок 2.5 – Механизм защиты нейросетевого контейнера

Таким образом, ЗНК дополнительно препятствует осуществлению направленного перебора образов «Чужой» для несанкционированного восстановления ключа (или его отдельных бит).

Выводы по второй главе

С целью защиты процедуры биометрической аутентификации по лицу на основе нейросетевого преобразователя «биометрия-код» от спуфинг атак (атак на биометрическое предъявление) разработана концепция защищенной биометрической аутентификации по лицу. Предложенная концепция подразумевает наличие двух отдельно функционирующих блоков на основе разных реализаций НПБК. Нейросетевой преобразователь, отвечающий за аутентификацию, должен обеспечивать связывание биометрического образа лица человека с длинным криптографическим ключом, отвечающим требованиям высоконадежной биометрической аутентификации. Нейросетевые параметры НПБК, осуществляющего аутентификацию, хранятся в нейросетевых

биометрических контейнерах и представляют собой таблицы связей и/или весов НПБК. Предлагается защищать нейросетевые контейнеры путем наложения гаммы, представляющей ключ НПБК для обнаружения спуфинг-атак.

Блок обнаружения спуфинг атак на основе классического НПБК исполняется в режиме, при котором минимизировано влияние ошибок классификации спуфинг атак на результирующую процедуру аутентификации. За счет разделения блока векторного представления образов (глубокой нейронной сети FeatherNet) и блока принятия решения в виде классификатора на основе НПБК, повышается обобщающая способность предложенного решения по отношению к разнообразию реализации спуфинг атак и решается проблема несанкционированного доступа в системах биометрической аутентификации по лицу, осуществляемого через блоки обнаружения спуфинг атак. Лучшее значение точности работы модуля на наборе данных CelebA-Spoof составило 97,2% (ACER = 2,9%), что говорит о приемлемом уровне производительности решения при высоком уровне защищенности процедуры аутентификации.

Глава 3. Разработка процедуры аутентификации пользователей по изображениям лиц в защищенном режиме исполнения

3.1. Сравнительный анализ глубоких нейросетевых моделей для детекции и извлечения признаков из биометрических образов лиц

Исходя из представленного в главе 1 обзора существующих решений защищенного исполнения процедур биометрической аутентификации по лицу, в данном исследовании предлагается такой вариант построения описанной процедуры, при котором модель для детекции лиц (детектор лиц), а также нейросетевая модель для извлечения признаков (экстрактор признаков) отделены от блока защиты биометрического шаблона (рис. 3.1). Блок защиты биометрического шаблона представляет собой нейросетевой преобразователь «биометрия-код» на базе тригонометрических нейронов, модель и алгоритмы обучения которого представлены в настоящей главе.

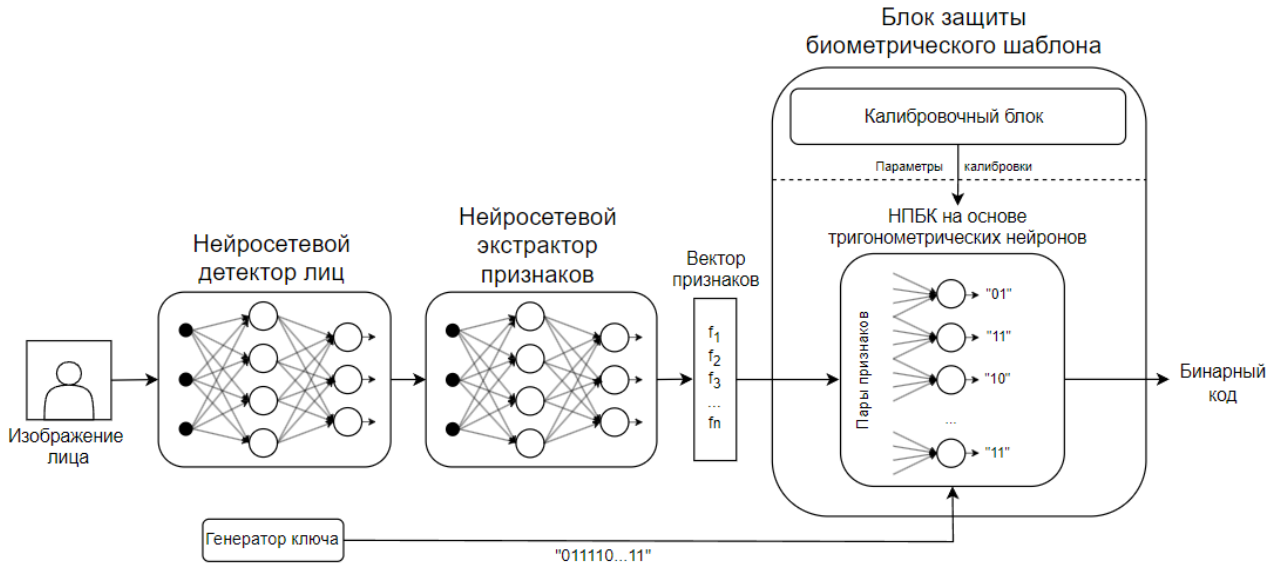


Рисунок 3.1 – Структурная схема построения защищенной системы аутентификации пользователя по лицу

Экстрактор признаков дает возможность получать нормально распределенные признаки [5], однозначно описывающие биометрический образ с

точки зрения их принадлежности субъекту (классу) [3]. Экстрактор признаков не хранит персональные биометрические данные зарегистрированных пользователей в виде знаний и должен обучаться на совокупности обезличенных примеров образов для обеспечения дифференциальной конфиденциальности. В случае добавления в систему нового пользователя или его исключения дообучение этого блока не потребуется.

Так же как и экстрактор признаков, предшествующая ему модель для детектирования лиц строится на основе глубоких нейронных сетей. В рамках данного исследования, такими алгоритмами выступают сверточные нейронные сети (СНС), позволяющие сначала обнаружить лицо на входном изображении, обрезав его по контуру заданных точек, а затем извлечь n-мерный вектор признаков (embedding). Вектор признаков $\bar{a} = (a_1, a_2, \dots, a_n)$, поступает на вход в НПБК, базирующийся на модели тригонометрических нейронов, который осуществляет классификацию биометрических образов пользователей в защищенном режиме. Он позволяет связать заданный извне криптографический ключ (пароль) с откликами своих выходов, при этом каждый отдельный тригонометрический нейрон продуцирует два бита ключа (в отличие от классической модели НПБК).

Для задач детекции лиц и извлечения из них признаков в виде векторов на сегодняшний день используется масса различных архитектур глубоких нейронных сетей [144]. Каждая из них обладает своими особенностями анализа входного изображения и его последующей обработки. С целью проведения сравнительного анализа и выбора одной, подходящей для защищенной процедуры аутентификации, реализации экстрактора признаков, были протестированы 3 открытых библиотеки на языке Python, каждая из которых включает в себя как модель для детекции лица, так и модель для извлечения признаков.

Первой библиотекой с открытым исходным кодом, включающей в себя этапы детекции лиц и извлечения признаков, является facenet-pytorch. Модуль детекции лиц представлен в виде предобученной сверточной нейронной сети типа MTCNN (Multi-Task Convolution Neural Network) [183]. Процесс детекции состоит

из трех подэтапов, каждый из которых реализуется отдельной CNN (P-Net, R-Net и O-Net). Модуль выделения признаков представлен в виде предобученной сверточной нейронной сети типа InceptionResnet [121]. Результатом работы Inception-ResNet является вектор из 512 признаков, выделенных из лица человека. Отметим, что сеть InceptionResNet v.1 имеет две реализации, предобученные на наборе данных VGGFace2 [38] (содержит 3,31 миллиона изображений 9131 субъекта, в среднем 362,6 изображения для каждого субъекта) и на наборе CASIA-Webface [177] (содержит 494414 изображений лиц 10575 реальных субъектов). Для задач данного исследования использовалась реализация, предобученная на наборе данных VGGFace2.

Вторая библиотека, InsightFace [142] — это библиотека Python для 2D- и 3D-анализа и распознавания лиц. InsightFace эффективно реализует широкий спектр современных алгоритмов распознавания, детектирования и выравнивания лиц, которые оптимизированы как для обучения, так и для развертывания. Одна из предобученных моделей для детектирования и распознавания лиц в представленной библиотеке – модель «buffalo_l». Модель состоит из двух частей: детектора лиц RetinaFace [48] и модуля извлечения признаков на основе архитектуры ResNet50 [63].

RetinaFace [48] – это современный одноэтапный детектор лиц, выполняющий попиксельную локализацию лица в различных масштабах, используя преимущества совместного многозадачного обучения с дополнительным и самостоятельным наблюдением. Архитектура ResNet50 (Residual Network) [63], используемая в библиотеке для извлечения 512 информативных признаков из ранее детектированных лиц, основывается на технологиях глубокого обучения для распознавания изображений. ResNet появляется в результате наблюдения о том, что обучение нейронных сетей усложняется с увеличением количества добавляемых слоев, а в некоторых случаях также снижается точность. В связи с этим была предложена архитектура, в которой данная проблема решается путем внедрения структуры глубокого «остаточного» обучения. ResNet имеет множество вариантов, которые работают

согласно одной и той же концепции, но имеют разное количество слоев. Resnet50 используется для обозначения варианта, который может работать с 50 слоями нейронной сети. Представленная модель была предобучена на датасете WebFace600K.

Наконец, третьей библиотекой, используемой для детекции и извлечения признаков является Deepface [136]. Библиотека представляет собой гибридную систему распознавания лиц, объединяющую самые современные модели: VGG-Face, Google FaceNet, OpenFace, Facebook DeepFace, DeepID, ArcFace, Dlib и SFace. Среди представленного списка моделей для задач исследования использовались ранее рассмотренный детектор RetinaFace и глубокая сверточная нейронная сеть для распознавания лиц VGG-Face [115]. VGG-Face представляет собой аналог глубокой архитектуры VGG-16. Обученная на базе данных из 2,6 миллионов изображений лиц и состоящая из 2622 уникальных личностей, используемая база данных включает до тысячи экземпляров каждого субъекта. Модель настроена на получение в качестве входных данных изображения RGB фиксированного размера 224 x 224; в качестве формы предварительной обработки они изначально нормализуют все тренировочные изображения по центру. Сеть способна возвращать эммбендинг размерностью 2622 признака.

Оценка сетей для детектирования лиц и советующих им экстракторов признаков производилась с точки зрения понятия «информативности» признаков, получаемых с помощью них. Под информативностью понимается количество собственной информации *j*-го признака для определенного класса образов [10]:

$$I_j = -log_2(AUC(\Phi_G(a_j), \Phi_I(a_j))) \quad (3.1)$$

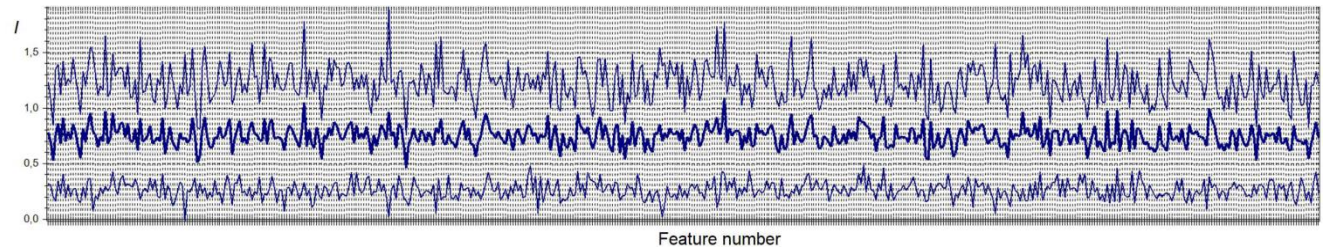
где *AUC* — площадь (area under curve), ограниченная функциями плотности вероятности «Свой» $\Phi_G(a_j)$ и «Чужие» $\Phi_I(a_j)$, а также осью абсцисс. $\Phi_G(a_j)$ характеризует значения признака строго для определенного класса образов, $\Phi_I(a_j)$ характеризует значения этого же признака для всех остальных классов образов.

Чем выше I в среднем, тем дальше разнесены собственные области классов в пространстве признаков и тем признаки информативнее.

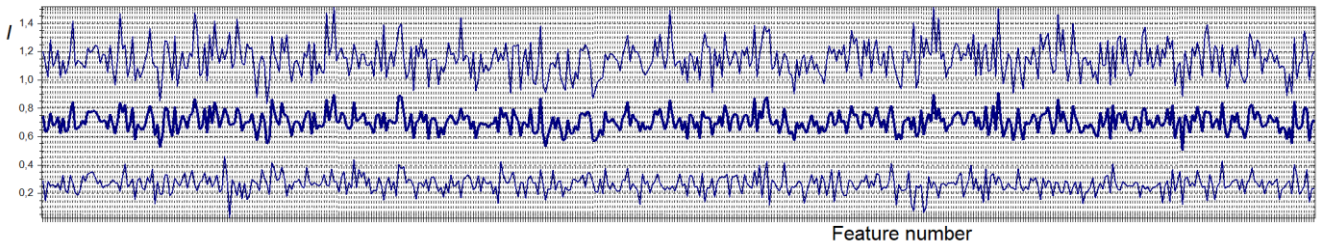
Признаки были извлечены из изображений для 3 комбинаций нейронных моделей для обнаружения лиц и извлечения признаков:

1. MTCNN + Inception-Resnet v1 (512 признаков);
2. RetinaFace + ResNet50 (512 признаков);
3. RetinaFace + VGG-Face (2622 признаков).

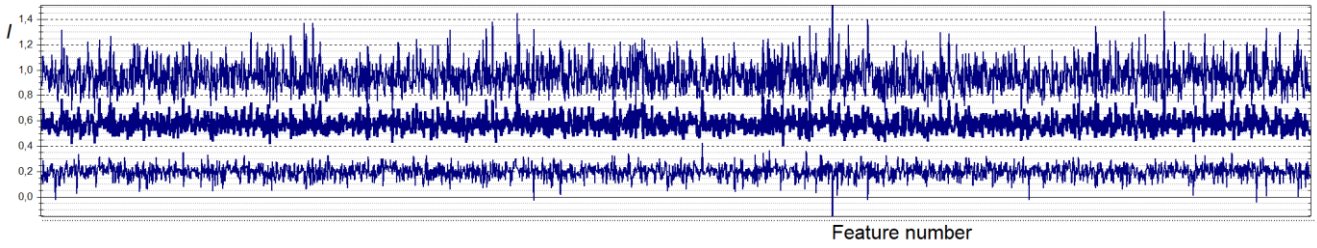
Далее оценивалась информативность признаков (рис. 3.2).



а) MTCNN + Inception-Resnet v1



б) RetinaFace + ResNet50



в) RetinaFace + VGG-Face

Рисунок 3.2 – Информативность признаков (жирная линия – математическое ожидание I, тонкая линия – математическое ожидание ± стандартное отклонение I)

Как видно, распределение информативности признаков имеет практически равномерное распределение. Этот факт объясняется нейросетевой природой

обработки изображений. Преимущество нейросетевых методов извлечения признаков заключается в стремлении нейронной сети равномерно распределять информацию между выходами. Также видно, что первая комбинация моделей дает в среднем наиболее информативные признаки, поэтому сочетание MTCNN и Inception-Resnet v1 было выбрано в качестве основного для проведения дальнейших исследований. Рассмотрим структуру этих сетей подробней.

Как отмечалось ранее, сверточная нейронная сеть MTCNN (табл. 3.1) [183] состоит из трех подсетей - P-Net, R-Net и O-Net. Первая подсеть (P-Net) принимает на вход изображение и использует небольшие сверточные сети для генерации предположений о наличие лица на входном изображении. P-Net генерирует ограничивающие рамки (bounding boxes) и способна производить оценку уверенности каждого из своего предположений. На втором этапе работы MTCNN используется R-Net, которая принимает предположения от P-Net и выполняет более точное обнаружение лиц. Она отбрасывает ложные срабатывания и корректирует позиции ограничивающих рамок лица. Последняя сеть, O-Net, принимает результаты, полученные от R-Net, и выполняет финальное обнаружение лиц. Она дополнительно улучшает точность определения ограничивающих рамок лица и ключевых лицевых ориентиров (глаза, нос, уголки рта). Результатом работы всей сети MTCNN является тензор (матрицу) из координат ограничивающей рамки лица и ключевых лицевых ориентиров. Благодаря полученному тензору изображение, поступившие на вход системы, может быть «обрезано» по заданным координатам для получения изображения лица субъекта, не содержащего дополнительной визуальной информации.

Таблица 3.1 – Архитектура MTCNN

Этап	**Компонент**	**Описание**
Входные данные	Входное изображение	Изображение, подаваемое на вход сети
P-Net	Conv1	Сверточный слой (3x3), 10 фильтров, активация ReLU
	Pool1	Max pooling (2x2), шаг 2
	Conv2	Сверточный слой (3x3), 16 фильтров, активация ReLU
	Conv3	Сверточный слой (3x3), 32 фильтра, активация ReLU
	Output	Карта вероятностей, ограничивающие рамки,

		ключевые точки
R-Net	Conv1	Сверточный слой (3x3), 28 фильтров, активация ReLU
	Pool1	Max pooling (3x3), шаг 2
	Conv2	Сверточный слой (3x3), 48 фильтров, активация ReLU
	Pool2	Max pooling (3x3), шаг 2
	Conv3	Сверточный слой (2x2), 64 фильтра, активация ReLU
	Fc1	Полносвязный слой, 128 нейронов, активация ReLU
	Output	Карта вероятностей, ограничивающие рамки, ключевые точки
O-Net	Conv1	Сверточный слой (3x3), 32 фильтра, активация ReLU
	Pool1	Максимальное объединение (3x3), шаг 2
	Conv2	Сверточный слой (3x3), 64 фильтра, активация ReLU
	Pool2	Max pooling (3x3), шаг 2
	Conv3	Сверточный слой (3x3), 64 фильтра, активация ReLU
	Pool3	Max pooling (2x2), шаг 2
	Conv4	Сверточный слой (2x2), 128 фильтров, активация ReLU
	Fc1	Полносвязный слой, 256 нейронов, активация ReLU
	Output	Карта вероятностей, ограничивающие рамки, ключевые точки

В свою очередь, основные компоненты архитектуры Inception-ResNet v1 (табл. 3.2) включают Inception-блоки и остаточные связи (residual connections). Inception-блоки используются для извлечения признаков на различных масштабах посредством параллельных сверточных операций с разными размерами ядер, такими как 1x1, 3x3 и 5x5, а также операций понижения разрешения изображения со взятием максимума в окне фильтра (max pooling). Эти блоки захватывают локальные и глобальные характеристики изображения, улучшая представление данных. Остаточные связи, в свою очередь, позволяют передавать информацию от предыдущих слоев к последующим, обходя несколько промежуточных слоев, что облегчает обучение глубоких сетей и предотвращает проблему затухания градиентов. Изначально предложенная для задачи классификации архитектура Inception-ResNet v1 в рамках данного исследования подвергается модификации в виде исключения последних полносвязных слои для классификации. После модификации сеть позволяет извлекать векторы из 512 признаков.

Таблица 3.2 – Архитектура Inception-ResNet v1

Этап	**Компонент**	**Описание**
Входные данные	Входное изображение	Изображение, подаваемое на вход сети
Входные слои	Conv2d	Сверточный слой (3x3), 32 фильтра, шаг 2
	Conv2d	Сверточный слой (3x3), 32 фильтра, шаг 1
	Conv2d	Сверточный слой (3x3), 64 фильтра, шаг 1
	MaxPool2d	Максимальное объединение (3x3), шаг 2
	Conv2d	Сверточный слой (1x1), 80 фильтров, шаг 1
	Conv2d	Сверточный слой (3x3), 192 фильтра, шаг 1
	MaxPool2d	Максимальное объединение (3x3), шаг 2
Inception-ResNet-A	Block A	Состоит из параллельных сверточных слоев (1x1, 3x3, 5x5) и остаточных связей
Reduction-A	Block A	Уменьшение размеров карт признаков, комбинация сверток и объединений
Inception-ResNet-B	Block B	Состоит из параллельных сверточных слоев и остаточных связей
Reduction-B	Block B	Уменьшение размеров карт признаков, комбинация сверток и объединений
Inception-ResNet-C	Block C	Состоит из параллельных сверточных слоев и остаточных связей
Финальные слои	Average Pooling	Усреднение (8x8)
	Dropout	Dropout с вероятностью 0.8
	Fully Connected	Полносвязный слой (512 нейронов для эмбеддингов лиц)
	Output	Эмбеддинг размером 512

3.2. Модель тригонометрического нейрона

В задачах биометрической идентификации и аутентификации, поступающий на вход системы вектор признаков $\bar{a} = (a_1, a_2, \dots, a_n)$, описывающий биометрический образ субъекта, сравнивается с некоторым эталонным вектором $\bar{a}'$, однозначно характеризующим этого пользователя. Однако анализируя взаимосвязи между признаками, в том числе, корреляционные связи, можно извлечь дополнительную информацию о различии или сходстве сравниваемых образов [144]. В таком случае, любая пара признаков $(a_i,\ a_j)$ может быть представлена в отдельном подпространстве признаков и изучена на предмет функциональной зависимости (рис. 3.2).

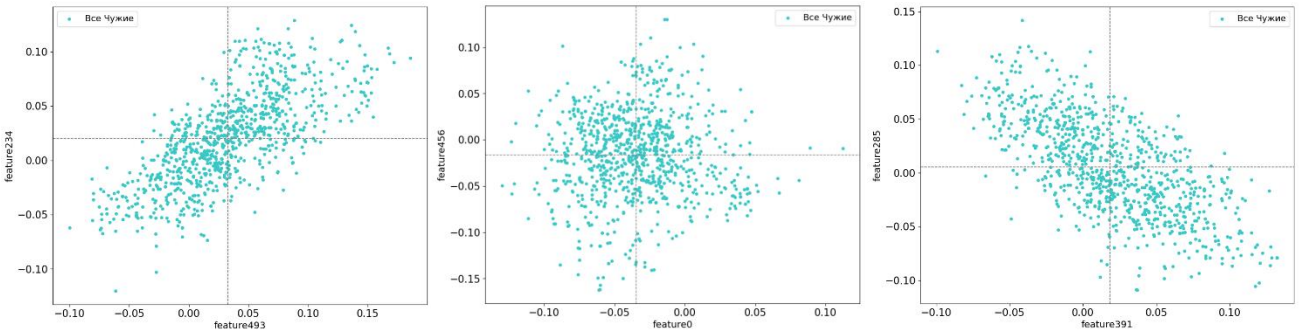


Рисунок 3.2 – Подпространства случайных пар признаков обучающего набора

В зависимости от наличия или отсутствия корреляционных связей между признаками, «облако» точек, описывающих образы в подпространстве пары признаков, распределится от левого нижнего края к верхнему правому (в случае положительной корреляции) или от правого нижнего к верхнему левому краю (в случае отрицательной). Координаты центра собственной области класса «Чужие» («центр массы») равны средним значениям соответствующих признаков $O = (m_i, m_j)$, на рисунке 2 через «центр массы» проходят пунктирные линии.

Любую пару признаков потенциально можно конвертировать в мета-признак [144], а пространство исходных признаков – в спрямляющее пространство мета-признаков с помощью отображения:

$$a'_l = f(a_i, a_j)$$

где i и j – номера признаков исходного вектора признаков ($i \neq j$), a'_l – *мета-признак*, т.е. признак, полученный путем синтеза двух или более исходных признаков с помощью функционального преобразования f, l – номер мета-признака. Число возможных подпространств пар признаков и, соответственно, количество мета-признаков определяется по формуле (3.2) [144]:

$$n' = 0.5(n(n-1)) \tag{3.2}$$

где n – исходное количество признаков.

Такое функциональное преобразование f должно отвечать, как минимум, двум критериям:

1. Не содержать в себе характеристик, компрометирующих образ легитимного пользователя («Своего»). Этот критерий является основополагающим для построения защищенной системы биометрической аутентификации.

2. Позволять с достаточно высокой точностью описывать расположение образа в спрямляющем гиперпространстве мета-признаков с учетом высокой вариативности биометрических образов.

Возьмём в качестве оценки расстояния от точки $O = (m_i,\ m_j)$ до некоторого биометрического образа евклидову метрику:

$$d(a, m) = \sqrt{(a_i - m_i)^2 + (a_j - m_j)^2}$$

Как видно из рисунка 3.3, расстояния от «центра массы» до биометрических образов разных субъектов d_1 и d_2 будут равны. Каждый из этих образов принадлежит разным классам. Некоторые классы образов располагаются достаточно близко друг к другу (образы одного класса - оранжевые треугольники, образы другого – синие треугольники на рис. 3.3), поэтому с помощью евклидовой метрики (или другой подобной) образы будет проблематично соотнести с их классами (провести разделяющую гиперплоскость).

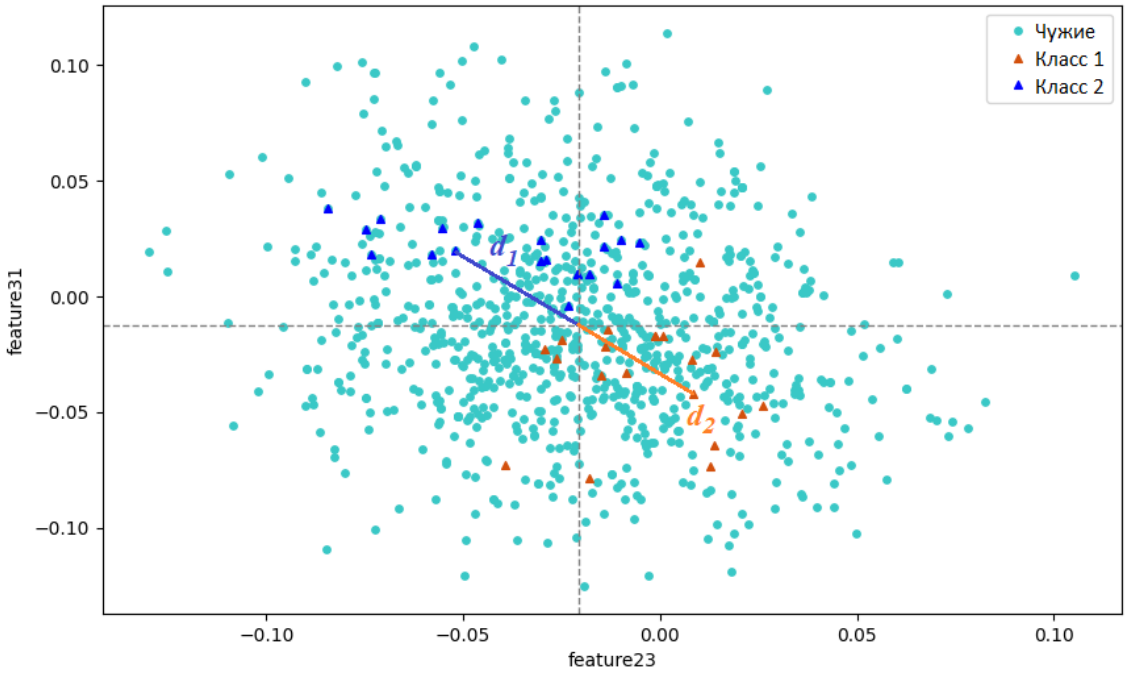


Рисунок 3.3 – Евклидовы расстояния от центра собственной области «Чужие» до образа определенного класса (d_2) и случайного образа (d_1)

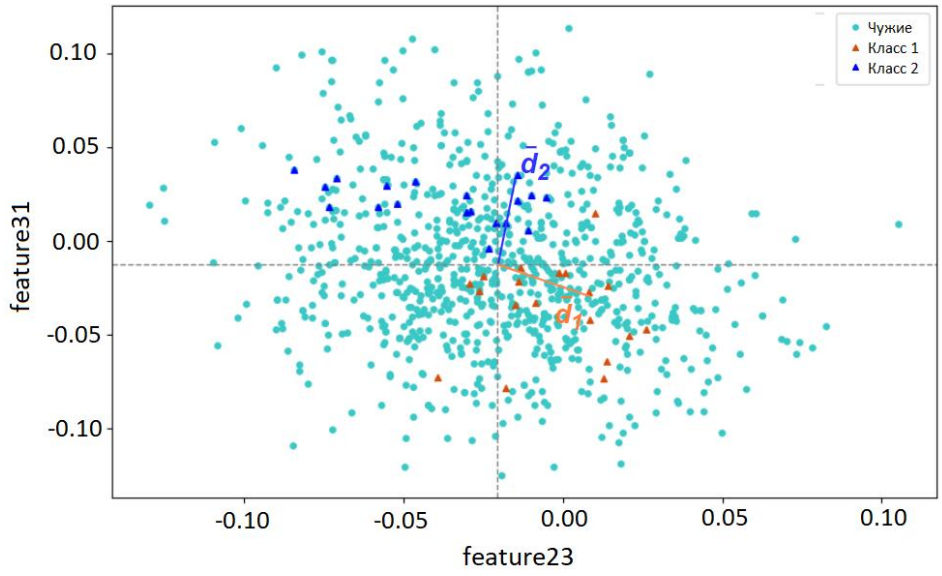
Предлагается две меры близости (3.3) и (3.4) образов в подпространстве пар признаков, учитывающие угол между вектором $\bar{d}$ (отрезок между «центром массы» $O = (m_i,\ m_j)$ и конкретным биометрическим образом), и вектором $\bar{v}$, равным по длине вектору $\bar{d}$, но совпадающим по направлению с осью OX (рис. 3.4).

$$a'_l = \sqrt{(a_i - m_i)^2 + (a_j - m_j)^2} * sin(\widehat{\bar{d}, \bar{v}}), \qquad (3.3)$$

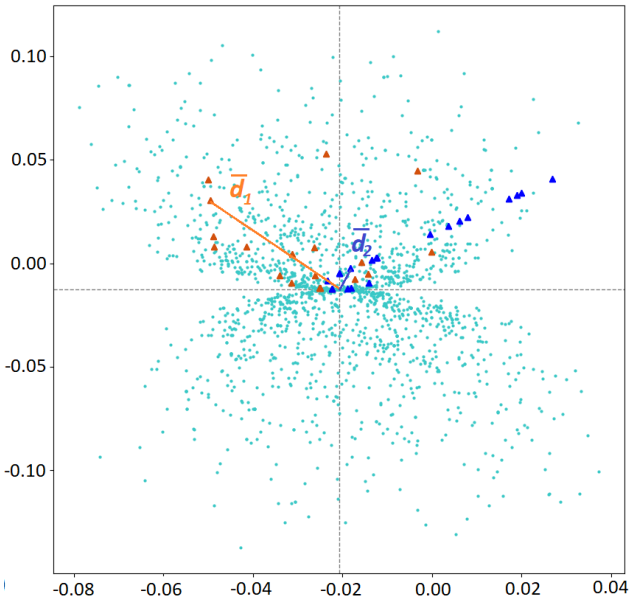
где $sin(\widehat{\bar{d}, \bar{v}})$ – синус угла между векторами $\bar{d}_l$ и $\bar{v}_l$.

Мера (3.3) основана на так называемом тригонометрическом нивелировании расстояния, которое применяется при проведении геодезических работ, а точнее производстве тахеометрических съемок [194]. Несмотря на то, что метрика (3.3) не является полным аналогом тригонометрического нивелирования, она оказывает корректирующее воздействие на исходное расстояние и дает более информативное представление о расположении образов относительно друг друга.

Проиллюстрируем, как визуально изменится подпространство исходной пары признаков (рис. 3.4а), если изменить расположение всех образов относительно «центра массы» в соответствии не с евклидовым расстоянием, а с расстоянием после нивелирования (3.3). Для этого построим производное подпространство признаков с учетом тригонометрического нивелирования (или мета-подпространство, рис. 3.4б).



а)



б)

Рисунок 3.4 – Вычисление и визуализация расстояния от «центра масс» до образов в соответствии с (2): исходное (вверху) и мета-подпространство (внизу)

Альтернативная метрика (3.4) позволяет избегать дополнительных вычислительных затрат для подсчета расстояний:

$$a'_l = sin(\widehat{\bar{d}, \bar{v}}) \quad (3.4)$$

Сравнить метрики по эффективности можно только на основании эксперимента (параграф 3). Величины, вычисляемые по формулам (3.3) и (3.4), также можно рассматривать в качестве мета-признаков, описывающих расположение биометрического образа в подпространстве пары исходных признаков с учетом тригонометрического нивелирования. Тогда формулы (3.3) и (3.4) являются отображением исходного пространства признаков в пространство мета-признаков.

Простейший тригонометрический нейрон строится на метрике (3.5) и принимает на вход мета-признаки, построенные с помощью отображений, учитывающих любой вариант тригонометрического нивелирования, чем и было обусловлено название предложенной модели нейрона. Представленный нейрон суммирует значения мета-признаков, однако возможны и более сложные схемы интегрирования входных значений (например, взвешенное суммирование, перемножение и т.д.), а также иные варианты тригонометрического нивелирования расстояний (формулы (3.3) и (3.4) – это частный случай подобных преобразований). Более сложные конструкции тригонометрических нейронов выходят за рамки настоящей статьи.

$$y = \sum_{l=1}^{k} a'_l, \quad (3.5)$$

где k – количество синапсов (входов) нейрона. Принятие решения по сумме входных значений мета-признаков нейрон осуществляет согласно двухуровневой пороговой функции активации $\varphi(y)$:

$$\varphi(y) = \begin{cases} 1, \ y \geq T_2 \\ 0, \ T_1 < y < T_2 \\ -1, \ y \leq T_1 \end{cases} \quad (3.6)$$

где T_1 и T_2 – пороги принятия решения в пользу одного из трех значений функции.

Пороговая функция необходима для квантования результатов преобразований, так как нейроны должны генерировать на выходах бинарный код.

3.3. Модель нейросетевого преобразователя «биометрия-код» на базе тригонометрического нейрона

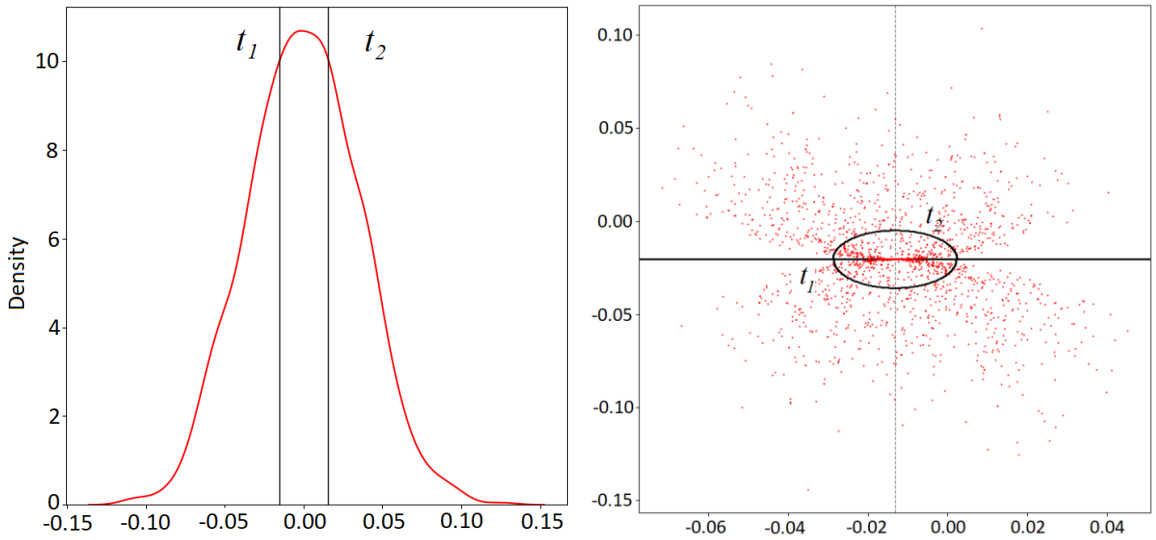
Чтобы НПБК осуществлял связывание ключа пользователя с биометрическим образом, значения на выходе функции активации необходимо преобразовать в двоичные состояния типа {«10», «00», «01»}. Тогда каждый нейрон будет продуцировать 2 бита информации. Для осуществления данной процедуры нужно выбрать номер хеш-преобразования для нейрона (табл. 3.3).

Таблица 3.3. - 24 варианта хеширующих преобразований отклика нейрона в двоичный код

№	-1	0	1	№	-1	0	1
1	11	00	01	**13**	01	00	11
2	11	00	10	**14**	01	00	10
3	11	01	00	**15**	01	10	00
4	11	01	10	**16**	01	10	11
5	11	10	00	**17**	01	11	10
6	11	10	01	**18**	01	11	00
7	00	01	11	**19**	10	00	01
8	00	01	10	**20**	10	00	11
9	00	10	01	**21**	10	01	11

10	00	10	11	**22**	10	01	00
11	00	11	10	**23**	10	11	00
12	00	11	01	**24**	10	11	01

Тригонометрический нейрон является частично связным. Для его настройки требуется подобрать несколько пар признаков и определить для каждой пары пороги t_1 и t_2. Пороги должны разделять соответствующие подпространства на три сектора ($[-\infty; t_1]$, $(t_1; t_2), [t_2; \infty]$) таким образом, чтобы в каждый сектор попадало примерно равное количество обучающих примеров «Чужие». Пары признаков следует выбирать так, чтобы все обучающие примеры «Свой» для каждой пары попадали в один определенный сектор (рис. 3.5).



a) *b)*

Рисунок 3.5 – Визуализация порогов, получаемых на основе метрики (3.3)
a) на графике плотности вероятности мета-признака,
b) в мета-подпространстве пары признаков

Все пороги тригонометрического нейрона зависят только от обучающей выборки образов «Чужие», не являющихся секретными. Выборка «Чужие» должна быть репрезентативной и состоять из данных случайных людей, она собирается разработчиком биометрической системы один раз и должна быть

обезличена. По этой причине она не содержит данных регистрируемых в системе субъектов. Соответственно, пороги нейронов являются открытой информацией. Держать пороги в секрете, как и обучающую выборку «Чужих» бессмысленно, так как злоумышленник может собрать репрезентативную выборку «Чужих» и вычислить пороги, не строго равные системным, но достаточно близкие к ним.

Таким образом, пороги едины для всех пользователей и вычисляются разработчиком биометрической системы заранее по алгоритму, описанному в следующем параграфе.

Следовательно, чтобы обучить нейрон достаточно подобрать подходящие номер хеш-преобразования и пары признаков, т.е. определить таблицу входов, которая тоже является открытой. Хеш-преобразование выбирается случайно, но только среди тех вариантов, которые соответствуют выбранному сектору и паре бит ключа, на которые настраивается нейрон. Например, если нейрон должен генерировать «11» и был выбран второй сектор, то доступные варианты – 11, 12, 17, 18, 23, 24.

Единственным секретным элементом нейрона является номер сектора, который был выбран при обучении и который связан с двумя битами ключа пользователя. Злоумышленник не знает, на какие биты настроен нейрон, и не знает, какой сектор является верным, что определяет безопасность биометрического шаблона.

При подборе ключа следует осуществлять грубый перебор биометрических образов, представленных векторами признаков. Возможность направленного перебора недоступна злоумышленнику, так как у него нет индикации близости генерируемого на выходе НПБК бинарного кода и ключа пользователя (есть только проверка – верно/неверно, как в случае с хешем пароля). Поэтому для взлома НПБК нужно подобрать верные секторы всех нейронов сразу (не по отдельности), так как только в этом случае будет получен верный ключ пользователя целиком. Таким образом, если злоумышленник не знает пользователя «в лицо» (не обладает его биометрическим образом), то он не сможет извлечь ключ (и наоборот).

Таблицу связей нейронов можно рассматривать как знания НПБК конкретного пользователя.

Таким образом, поступив на вход нейрона, пары признаков обрабатываются с помощью функционала (3.3) или (3.4) и попадают в пороговую функцию, относящую полученную сумму значений в одно из трех состояний «-1», «0», «1», которое, в свою очередь, преобразуется в двоичный код.

3.4. Алгоритмы автоматического синтеза и обучения нейросетевого преобразователя биометрических образов лица в код на малых выборках

Для корректной работы НПБК необходимо правильно настроить пороги для каждого подпространства пар признаков. Проще всего разделение на секторы можно продемонстрировать на графике плотности вероятности мета-признаков (рис. 3.5а). В общем случае, пороги, представленные на графике в виде вертикальных линий, рассчитываются следующим образом:

1. Строится эмпирическая функция плотности вероятности $f(.)$ мета-признаков, вычисляемых по формуле (3.3) или (3.4);

2. Функция $f(.)$ интегрируется с целью нахождения функции распределения $F(.)$;

3. Интервал [0, 1] делится на m равных секторов ([0,$\frac{1}{m}$,$\frac{2}{m}$,…,1]);

4. Для каждого значения $z \in [0, \text{m-1}]$ из полученного набора высчитывается соответствующее значение функции распределения t_z, для которого $F(t_z) = z$. Полученные значения t_z и являются искомыми порогами.

Для определения двух порогов и обхода необходимости построения эмпирической функции плотности вероятности (в случае, если закон распределения мета-признаков неизвестен или не соответствует нормальному распределению), предлагается рассчитывать пороги согласно следующему алгоритму:

1. Рассчитываются значения мета-признаков с помощью метрик (3.3) или (3.4):

$$A = \{a'_{l_1}, a'_{l_2}, a'_{l_3}, \dots, a'_{l_q}\},$$

где q – количество биометрических образов в подпространстве пары признаков $(a_i,\ a_j)$.

2. Полученные значения мета-признаков ранжируются по возрастанию:

$$B = sort(A)$$

где *sort(A)* – функция, сортирующая элементы множества A по возрастанию.

3. Полученный возрастающий массив делится на 3 равных сектора, в результате чего получаются пороги t_1 и t_2:

$$t_1 = b_{\frac{q}{3}}, \qquad t_2 = b_{\frac{2q}{3}},$$

где b – значение элемента массива *B*.

Алгоритм на рис. 3.6 стремится найти пороги, при которых $\Phi(t_1) \approx 0{,}333$, $\Phi(t_2) \approx 0.667$, где Ф(.) – функция распределения мета-признака. Чем больше объем обучающей выборки «Чужие», тем ближе значения порогов к искомым.

Для каждой пары признаков значение порогов сохраняются и используются при синтезе и обучении НПБК. После завершения работы алгоритма, нет необходимости хранить данные «Чужих», использовавшиеся в ходе его работы, если пороги не нуждаются в уточнении (донастройке).

Алгоритм потенциально применим при различных законах распределения признаков и мета-признаков, хотя характер распределения может повлиять на эффективность НПБК. Наиболее значимым фактором является несоответствие законов распределения признаков для обучающих выборок «Свой» и «Чужие». Например, если для калибровки использовать образы, полученные при плохом освещении или высокой окклюзии, а для обучения НПБК использовались высококачественные изображения с хорошим освещением без загораживающих объектов, то выборка «Свой» может оказаться смещенной относительно «центра масс», что снизит качество обучения НПБК. Наличие чувствительных ко смещению признаков определяется архитектурой экстрактора признаков. В

настоящей работе распределения исходных признаков были близки к нормальному (иногда наблюдалась незначительная асимметрия распределения), что проверялось критерием хи-квадрат.

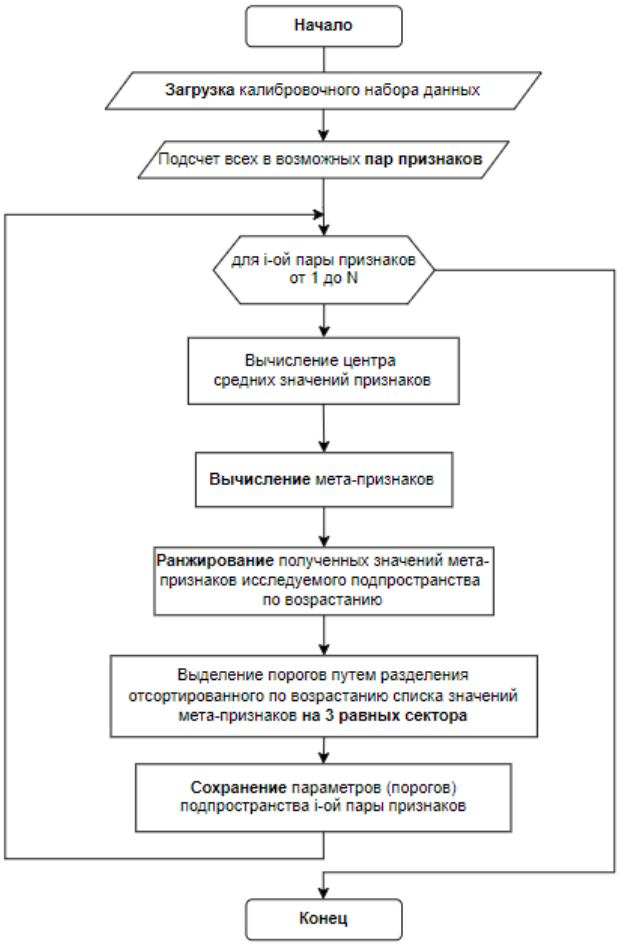


Рисунок 3.6 – Алгоритм калибровки НПБК

НПБК строится отдельно для каждого субъекта, то есть для каждого класса обучающего набора биометрических данных. Процедура синтеза и обучения НПБК осуществляется автоматически без применения итерационного обучения на основе метода обратного распространения ошибки. Алгоритм обучения НПБК приведен на рисунке 3.7.

Первым этапом построения НПБК на базе тригонометрического нейрона будет отбор таких пар признаков обучающего набора данных, в подпространстве которых биометрические образы субъекта будут располагаться строго в одном из трех секторов. Отметим, что «строгость» расположения образов в секторах можно

варьировать, исходя из объема обучающей выборки «Свой» (чем больше обучающих примеров, тем больше отклонений от заданного сектора допустимо). При этом следует учитывать, чем больше обучающих примеров «Свой» лежит вне выбранного сектора, тем выше будет вероятность ошибок «ложного отказа». Слишком строгие правила при больших объемах обучающей выборки «Свой» могут привести к невозможности синтеза НПБК из-за отсутствия достаточного количества подходящих пар признаков.

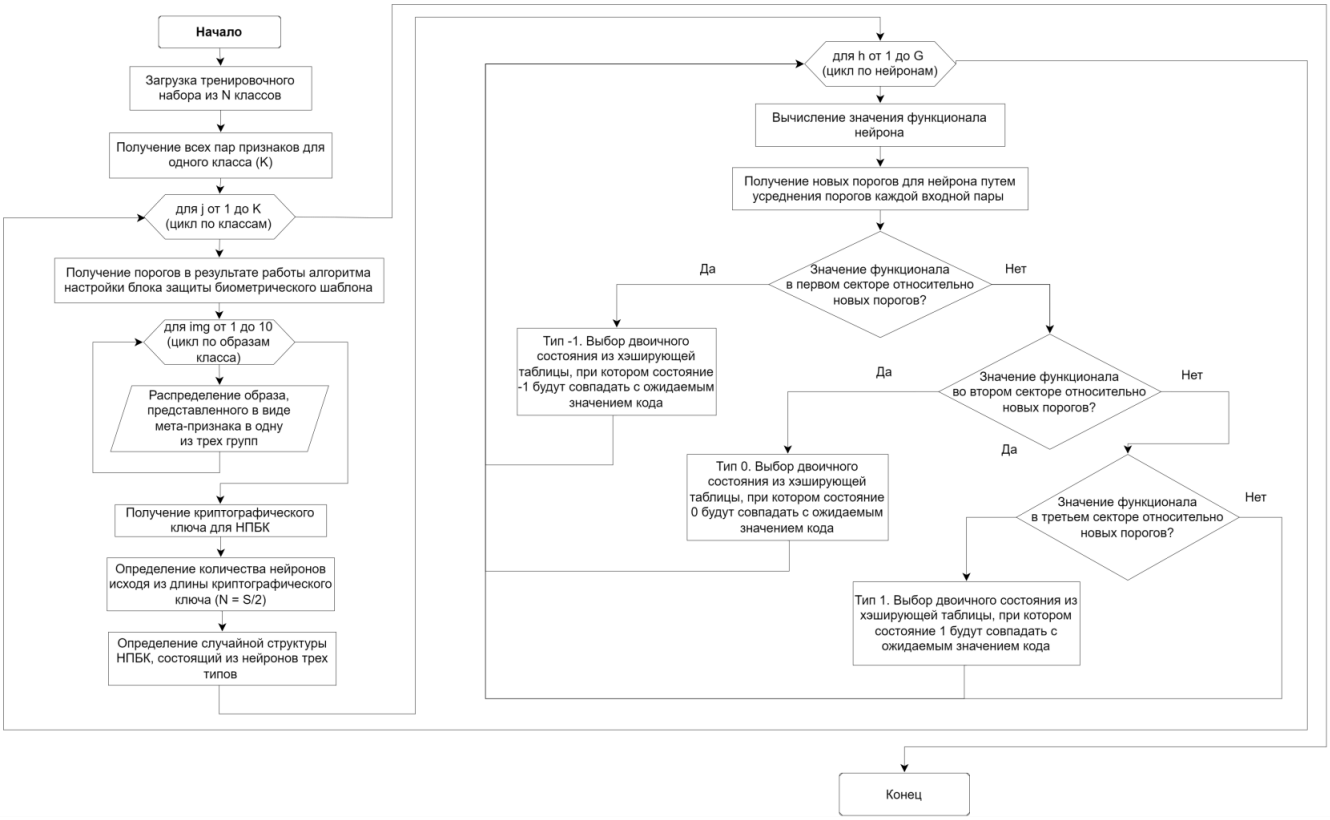


Рисунок 3.7 – Алгоритм обучения нейросетевого преобразователя «биометрия-код» на основе тригонометрического нейрона

Формируется три непересекающиеся группы пар признаков (отдельная группа на каждый сектор). Для каждого нейрона случайным образом выбираются пары признаков из определенной (одной) группы. Количество пар признаков равно числу синапсов нейрона. Входы каждого нейрона должны быть уникальны и ни одна пара не может повторно использоваться в другом нейроне. На базе

каждой группы пар признаков формируется примерно равное количество нейронов, чтобы избежать статистических смещений и снижения информационной энтропии кодов на выходе НПБК при поступлении на его входы образов «Чужой» (допускается незначительное расхождение в 2-3 нейрона). При случайном перемешивании нейронов в структуре НПБК становится невозможно осуществить направленный перебор входных образов для определения ключа пользователя (или его частей).

Пороговые значения для функции активации (3.6) нейрона вычисляются по формулам:

$$T_1 = \frac{1}{k}\sum_{z=1}^{k} t_{1z} \qquad \text{и} \qquad T_2 = \frac{1}{k}\sum_{z=1}^{k} t_{2z},$$

где k – количество синапсов (входов) нейрона; z – номер синапса (входа) нейрона; t_{1z} и t_{2z} – пороги для пары признаков, поступающей на вход нейрона, полученные в результате работы алгоритма калибровки.

Структура НПБК должна строиться из расчета $N = S/q$, где N – количество нейронов, S – желаемая длина криптографического ключа, q – количество бит, продуцируемое одним нейроном и вычисляемое по формуле:

$$q = [\, log_2 h \,],$$

где h – количество порогов, делящее функцию плотности вероятности мета-признаков на равные секторы ($h = m\text{-}1$), [.] – операция округления до большего значения.

После обучения нейронов распределение хеш-преобразований является равномерным, поэтому достигается равновероятное появление состояний «0» и «1» на выходе НПБК при поступлении на входы образов «Чужих», и как следствие высокая энтропия.

3.5. Оценка надежности предложенных моделей и алгоритмов

Данный раздел посвящен комплексной оценке эффективности работы предложенной биометрической системы аутентификации по лицу. В качестве основной метрики использовалась Equal Error Rate (EER), представляющая собой состояние системы при котором False Acceptance Rate (FAR) и False Rejection Rate (FRR) имеют примерно одинаковое значение. Показатели FAR, FRR и EER могут измеряться процентом или вероятностью. EER вычисляется, исходя из ROC-кривых, которые позволяют продемонстрировать способность системы к балансу между безопасностью (минимизация FAR) и удобством использования (минимизация FRR). Чем ниже значение EER, тем выше эффективность биометрической системы. Балансировать показатели FRR и FAR можно путем изменения порога принятия, измеряемого числом допустимых ошибок в бинарном коде, генерируемом с помощью НПБК. На практике это достигается применением кодов, исправляющих ошибки [143].

В отдельных опытах также применялась метрика несбалансированной accuracy, отражающей общий процент или вероятность верных решений.

Для выбора оптимальных параметров НПБК, его первичного тестирования и последующего обучения в рамках данного исследования был сформирован специализированный набор данных SFDv1. Этот набор включал видеозаписи лиц 75 испытуемых. Для каждого участника было собрано три видеозаписи при различных условиях освещения: дневной свет (окно как источник света), искусственный свет в помещении (освещение от ламп сверху) и отсутствие света (тёмный угол помещения, где лица ещё различимы). Средняя продолжительность каждой видеозаписи составила примерно 45 секунд. Испытуемые выполняли круговые движения головой и поворачивались вправо, влево, вверх и вниз, чтобы обеспечить разнообразие позиций лица в кадре. Все участники подписали информированное согласие на участие в эксперименте и на обработку персональных данных. Данные участников хранились и обрабатывались в обезличенном виде. Каждая из трех видеозаписей для одного испытуемого

подвергалась обработке и случайному выделению 20 фреймов (изображений лица, образов). Таким образом, на каждого испытуемого формировался пакет из 60 образов лица, представленных тремя видами освещения.

В качестве дополнительных наборов данных для проведения сравнительного анализа и проверки эффективности работы предложенной системы, в работе использовались:

1. Один из наиболее популярных и широко используемых датасетов в области распознавания лиц LFW (Labeled Faces in the Wild) [68]. Датасет LFW содержит набор изображений лиц знаменитостей (формат JPEG), собранных в различных условиях освещения, ракурсах и эмоциональных проявлениях. Всего LFW содержит более 13000 изображений лиц 5749 различных людей. Однако только несколько классов представлено более чем 20 изображениями, а точнее только 57. В связи с этим, только 57 классов набора данных LFW являются пригодными для проведения обучения и тестирования предложенной системы защищенной системы биометрической аутентификации.

2. Небольшой, но репрезентативный датасет Faces94 [137], представляющий собой набор изображений лиц (формат JPEG) 153 человек. Каждый человек представлен в среднем 20 изображениями. Общее количество изображений составляет 3060. Отметим также, что изображения в Faces94 представлены в низком разрешении, что также добавляет сложности для алгоритмов распознавания.

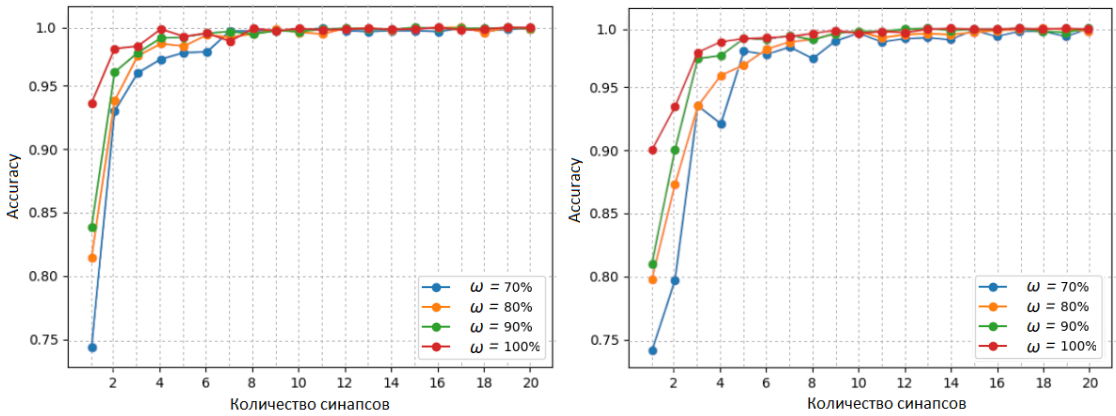
Прежде чем производить оценку эффективности работы предложенной защищенной биометрической системы, необходимо определить граничные значения гиперпараметров, таких как количество синапсов нейрона, в пределах которых имеет смысл осуществлять тестирование системы (из соображений скорости обучения и потенциальной точности работы НПБК). Искомыми гиперпараметрами выступают:

1. Количество входов одного нейрона (количество синапсов) η;
2. ω – доля (процент) образов обучающего набора, попадающих в один из трех секторов, при котором принимается решение отнесения пары

признаков к определенному сектору («строгость» расположения образов в секторах).

С одной стороны, интуитивно понятен тот факт, что чем больше синапсов у каждого нейрона, тем «точнее» будет функционировать НПБК: каждый нейрон будет получать на вход исчерпывающее количество информации, что потенциально снижает вероятность ошибочных решений. Тем не мене, количество пар признаков ограничено. Поэтому необходимо найти баланс между количеством синапсов и длиной ключа. Тестирование, описанное в настоящем разделе, позволяет упростить поиск оптимальной конфигурации НПБК. Оно основано на метрике accuracy, примененной по отношению к отдельно взятому нейрону, как к классификатору.

На рисунке 3.8 представлены графики, полученные в результате вычисления accuracy в зависимости от числа синапсов. Количество синапсов варьировалось от $k = 2$ до $k = 20$. Кроме того, каждая из описанных зависимостей реализовывалась при разных значениях ω ($\omega = 70\%$, $\omega = 80\%$, $\omega = 90\%$ и $\omega = 100\%$). Тестирование нейрона проводилось на основе набора данных SFDv1, для обучения использовались по 10 примеров каждого пользователя. Эксперимент был произведен отдельно для каждого из двух представленных в работе типов тригонометрического нейрона (3.3) и (3.4).



a) *b)*

Рисунок 3.8 – Зависимость точности работы тригонометрического нейрона от количества синапсов при разных значениях ω: a) для нейрона на базе меры (3.3); b) для нейрона на базе меры (3.4)

Следует подбирать такие конфигурации нейрона, при которых возможно получение максимальной точности при наименьших значениях k и ω. Это позволит получить ключ наивысшей длины.

На базе датасета SFDv1 проведен эксперимент по выбору начальной тестовой конфигурации НПБК (табл. 3.4). Для его обучения и тестирования из набора данных были отобраны 10 случайных пользователей, остальные использовались для калибровки. При выборе подходящих конфигураций ключевое внимание уделяется способности НПБК продуцировать ключ максимальной длины для каждого из пользователей при минимальной ошибке распознавания его отдельных нейронов.

Таблица 3.4 – Сравнительная таблица оптимальных конфигураций сборки нейросетевого преобразователя

№	«Строгость» попадания в сектор, ω	Количество синапсов, k	Количество нейронов	Длина ключа, бит	Успешный синтез
Нейрон на базе меры (2)					
1	70%	7	256	512	100%
			512	1024	100%
			1024	2048	100%
2	80%	8	256	512	96%

			512	1024	92%
			1024	2048	86%
3	90%	7	256	512	80%
			512	1024	68%
			1024	2048	60%
4	100%	4	256	512	72%
			512	1024	64%
			1024	2048	58%
Нейрон на базе меры (3)					
1	70%	10	256	512	100%
			512	1024	100%
			1024	2048	100%
2	80%	9	256	512	100%
			512	1024	100%
			1024	2048	100%
3	90%	12	256	512	100%
			512	1024	84%
			1024	2048	74%
4	100%	9	256	512	72%
			512	1024	68%
			1024	2048	52%

Обозначенным требованиям соответствуют только одна конфигурация НПБК на базе меры (3.3): по-видимому, использование в основе евклидового расстояния накладывает шумы. Тем не менее, такой НПБК способен продуцировать 2048 битный ключ, используя всего 70% образов попадания в один из секторов. В случае нейронов на базе меры (3.4) подходящими являются сразу две конфигурации, позволяющие для 100% пользователей успешно продуцировать длинные криптографические ключи (2048 бит). Однако согласно эксперименту, представленному на рисунке 3.8, наиболее высокой точностью распознавания отдельного нейрона обладает конфигурация №1 (99,93% против 99,81% для №2). В связи с этим, указанная конфигурация была выбрана в качестве основной для НПБК на базе меры (3.4).

Рассмотренные конфигурации далее использовались для проведения сравнительного эксперимента для трех экспериментальных датасетов.

Проведен эксперимент по оценке эффективности предложенной системы, который повторялся 3 раза, каждый раз осуществлялось иное разделение наборов данных на непересекающиеся множества:

– калибровочная выборка (одна треть образов из набора данных);
– множество, используемое для обучения и тестирования НПБК (две трети образов из набора данных).

Таким образом, каждый пример был задействован в качестве обучающего, тестового и калибровочного.

Результаты проведенных расчетов представлены на рисунке 3.9 соответственно.

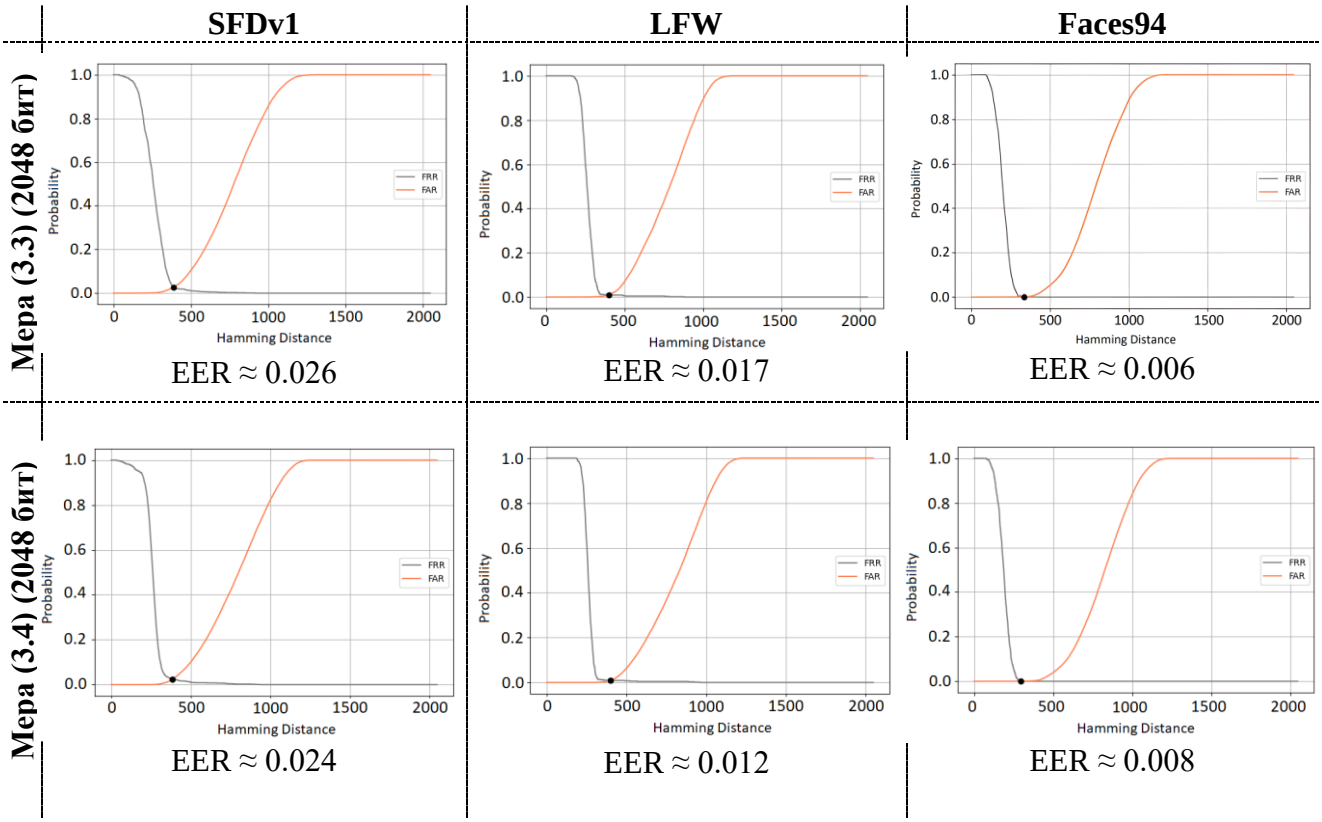


Рисунок 3.9 – Графики средних значений EER для НПБК на базе тригонометрических нейронов при максимальной длине ключа (2048 бит)

Графики показывают зависимость FRR и FAR от допустимого числа неверных бит в бинарном коде, генерируемом НПБК. Число неверных бит измеряется расстоянием Хэмминга между бинарным кодом, формируемым на выходе НПБК, и верным ключом пользователя. Максимальное значение оси расстояния Хэмминга равно длине ключа.

Как видно из рисунков, лучшего показателя удалось достичь при проведении экспериментов на основе набора данных Faces94 (EER ≈ 0.006). При этом длина ключа достигает значения в 2048 бит, что является высоким показателем.

В таблице 3.5 представлено сравнение предложенной модели нейросетевого преобразователя с наиболее актуальными работами за последние годы, предлагающими альтернативные решения по защите биометрических шаблонов лиц. Для корректного сравнения с классической реализацией НПБК в контексте работы с биометрическими образами лиц, был проведен эксперимент по обучению такого преобразователя на наборе данных SFDv1 [21]. Процедура вычисления EER (табл. 3.6) повторялась при различных параметрах I_{min}, k и N (где I_{min} – минимальная информативность признаков, которые учитывались при синтезе НПБК, k – число входов нейронов НПБК, N – количество нейронов НПБК). Обучение преобразователя проводилось при допущении, что входы (синапсы) нейрона могут дублироваться. Хотя такой подход к синтезу НПБК может быть уязвим для атаки Маршалко [11], он позволяет генерировать ключи, сопоставимые по длине с современными решениями (до 512 бит). В ином случае, длина ключа НПБК будет ограничена 128 битами, а минимальное значение ошибки составит EER = 0.5%.

Таблица 3.5. Полученные оценки коэффициента равной вероятности ошибок (EER, %) для классического НПБК

I_{min}	0		0,25			0,5			0,75		1	
N / n	128	256	128	256	512	128	256	512	128	256	128	256
4	0,50	0,36	0,62	0,36	0,36	0,40	0,37	0,37	0,37	0,32	0,27	0,36
6	0,60	0,30	0,41	0,36	0,36	0,32	0,41	0,37	0,60	0,29	0,56	0,47
8	0,47	0,35	0,47	0,33	0,36	0,42	0,37	0,37	0,59	0,42	0,28	0,42
10	0,56	0,35	0,49	0,39	0,37	0,42	0,42	0,32	0,43	0,36	0,59	0,49
12	0,41	0,43	0,49	0,45	0,36	0,57	0,44	0,38	0,46	0,59	0,36	0,38

Все признаки, информативность которых ниже I_{min} не учитываются при синтезе и обучении НПБК. Отметим, что информативность признаков для каждого субъекта различна, поэтому для каждого пользователя набор признаков отличается при $I_{min} > 0$.

Из таблицы видно, что наилучшие результаты EER = 0,27% достигаются при $I_{min} = 1$, $k = 4$, $N = 128$. Сравнимые значения при большей длине ключа (256 бит) достигаются при $I_{min} = 0{,}75$, $k = 6$. При $I_{min} = 0$ используются все признаки, однако многие из них являются шумовыми для отдельных субъектов и не несут полезной информации. Значение $I_{min} = 1$ оптимально, поскольку в этом случае неинформативные признаки отсекаются и не участвуют в обучении и распознавании пользователей. Дальнейшее тестирование показало, что при $I_{min}>1$ вероятность ошибки возрастает, что означает, что отсеиваются уже полезные признаки. Меньшее количество входов нейрона обеспечивает более низкие показатели EER, однако дополнительное снижение k может негативно сказаться на безопасности биометрического шаблона.

Таблица 3.6. – Сравнительный анализ предложенного решения с существующими методами защищенной биометрической аутентификации

Подход к BTP	**Набор данных**	**Метрика**	**Максимальная длина ключа, бит**
Random projection и NN (Peng J. et al. [119])	Faces94	EER = 0.0022	322
Deep Face Fuzzy Vault (Rathgeb C. et al. [127])	FERET + FRGCv2	FRR < 0,01 при FAR < 0.0001	-
Chaff-less Fuzzy Vault (Dong X. et al. [51])	LFW	Accuracy = 98.53% при EER = 0.001	-
	VGGFace2	Accuracy = 98.53% при EER = 0.001	-
	IJB-C	Accuracy = 43.73% при EER = 0.001	-
MEB Encoding (Kumar Pandey R. et al. [79])	PIE	EER = 0,0114	1024
CNN (Kumar Jindal A. et al. [78])	PIE	EER = 0,036	1024
Классическая реализация	SFDv1	EER = 0.0029	256

НПБК			
НПБК (мера (2))	SFDv1	EER = 0.026	2048
	LFW	EER = 0.017	2048
	Faces94	EER = 0.006	2048
НПБК (мера (3))	SFDv1	EER = 0.024	2048
	LFW	EER = 0.012	2048
	Faces94	EER = 0.008	2048

Остальные подходы показывают значительное различие в метриках эффективности защищенной биометрической аутентификации и итоговой длине ключа. Несмотря на то, что в части исследований [51, 127], предложенные системы защищенной аутентификации по лицу демонстрируют высокие показатели точности (accuracy > 95%), превышающие лучшие значения предложенных реализаций НПБК, авторы не дают информации об итоговой длине ключа, что затрудняет сравнения указанных реализаций по степени надежности.

Остальные работы [78, 79, 119], объединяющим свойством которых является применение глубоких нейронных сетей для задачи извлечения информативных признаков из биометрических образов лиц, зачастую демонстрируют сравнительно высокие показатели ошибки работы системы распознавания (EER ≥ 0,0114). Исключением является исследование [119], в рамках которого осуществляется преобразование исходных биометрических признаков в зашифрованный вектор фиксированной длины с помощью метода случайной проекции. Проецированный вектор привязывается к случайному ключу с помощью нейросетевой модели, обеспечивая тем самым защиту биометрических данных и низкие показатели ошибки распознавания (EER = 0.0022). Однако применение самокорректирующихся кодов (БЧХ) для исправления ошибок в кодах предложенного решения накладывает некоторые ограничения на длину результирующего ключа, максимальное значения которого составляет всего 322 бита.

Классическая реализация НПБК, воспроизведенная в рамках представленных экспериментальных исследований и обученная на специально подготовленном наборе данных SFDv1, продемонстрировала один из лучших

результатов с точки зрения точности распознавания (EER = 0.0029), однако сохранила свой ключевой недостаток, на нивелирование которого направлено настоящее исследование: низкая длина ключа даже при достаточно длинном векторе исходных биометрических признаков (512 признаков).

Выводы по третьей главе

Защищенный режим исполнения нейросетевых алгоритмов классификации образов является важной составляющей концепции доверенного ИИ. Защищенный режим строится на базе специальных нейросетевых моделей и архитектур, примером которой является предложенная модель НПБК на базе тригонометрических нейронов. Модель интегрируется в структуру защищенной системы биометрической аутентификации по лицу и представляет собой отдельный блок, осуществляющий классификацию биометрических образов пользователей в защищенном режиме. Математический аппарат НПБК основывается на применении двух альтернативных мер близости биометрических образов в подпространстве пар признаков, что позволяет значительно улучшить точность классификации биометрических образов и длину криптографических ключей, при этом обеспечивая защиту от известных атак в отношении защищенных нейросетевых алгоритмов.

Экспериментальная оценка системы защищенной аутентификации на основе предложенной модели продемонстрировала более высокую эффективность ее работы по сравнению с альтернативными реализациями биометрических криптосистем. Лучшие значения на тестовых наборах данных составило EER ≈ 0.006 на открытом наборе данных Faces94, EER ≈ 0.012 на LFW и EER ≈ 0.024 на специально подготовленном наборе данных SFDv1. Полученная реализация способна продуцировать ключ длиной 2048 бит, что является более высоким показателем надёжности предложенного решения по сравнению с достигнутыми ранее.

Для извлечения признаков использована нейросетевая архитектура Inception-ResNet v1. По результатам экспериментов извлекаемые из изображений лиц признаки имели преимущественно слабую взаимную корреляционную зависимость, а для формирования нейронов, как правило, отбирались признаки с отсутствием корреляции или незначительной корреляцией. Таким образом, разработанная модель хорошо работает со слабо коррелированными признаками. В этом смысле она дополняет модель корреляционных нейронов, которая, наоборот, показывает высокие результаты, когда входные признаки сильно коррелированы и неспособна работать с независимыми признаками.

Глава 4. Разработка системы аутентификации пользователей компьютерных систем по лицу в защищенном режиме исполнения

4.1. Структура системы защищенной биометрической аутентификации по лицу

Определение структуры системы предполагает определение и анализ ее компонентов, их взаимосвязей, а также принципов и правил, в соответствии с которыми осуществляется функционирование обозначенных элементов. В этой связи, процесс разработки структуры системы защищенной биометрической аутентификации по лицу полностью соответствует указанным этапам, однако их описание предлагается осуществлять в рамках трех функциональных блоков (модулей) структуры системы, представленной на рисунке 4.1.

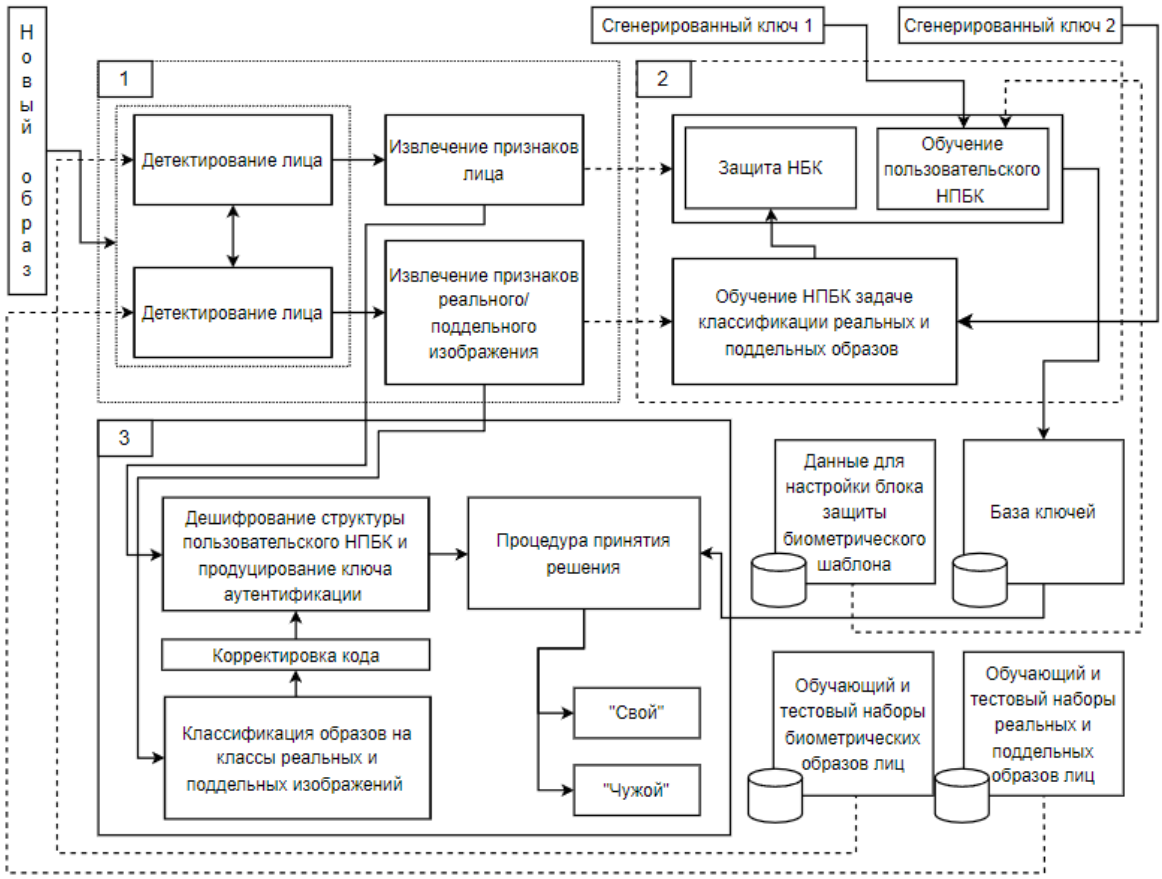


Рисунок 4.1 – Структура системы защищенной биометрической аутентификации по лицу

Предложенная структура включает в себя разработанные в ходе настоящего исследования:

− концепцию защищенной биометрической аутентификации по лицу, обеспечивающую противодействие атакам на биометрическое предъявление;

− модель нейросетевого преобразователя «биометрия-код» на базе тригонометрических нейронов, осуществляющую аутентификацию пользователя в защищенном режиме;

− алгоритмы калибровки и обучения НПБК, позволяющие осуществлять настройку и обучение пользовательских нейросетевых преобразователей.

Целью разработки структуры системы защищенной биометрической аутентификации по лицу является оптимальное объединение предложенных решений в рамках одной функциональной схемы, отвечающей техническим, правовым и организационным требованиям к системам распознавания лиц. Защищенный режим исполнения системы во многом обусловлен опубликованным в рамках приказа Минцифры РФ № 902 от 01.09.2021 г. [18] перечнем угроз безопасности, актуальных при обработке биометрических персональных данных, их проверке и передаче информации о степени их соответствия биометрическим данным, хранящимся в информационных системах.

Ниже представлены три ключевых блока системы, в рамках которых осуществляется работа ее основных функциональных элементов:

1. **Блок детекции лиц и извлечения признаков.** Все глубокие модели, обеспечивающие извлечения признаков из входных изображений, объединяются в один блок. Такое объединение обусловлено применением концепции для обнаружения спуфинг атак и процедуры аутентификации с помощью НПБК, которые подразумевают разделение процессов получения векторного представления лиц и принятия решения. Такие глубокие нейронные сети не требуют повторного переобучения при изменении контекста функционирования системы или компрометации биометрических образов и, при необходимости, могут быть заменены альтернативными архитектурами.

2. **Блок обучения нейросетевых преобразователей «биометрия-код» для задач распознавания лиц и обнаружения спуфинг атак.** НПБК для задачи обнаружения спуфинг атак обучается один раз в соответствии с процедурой, описанной в главе 2 и универсален для всех пользователей.

Для пользовательского НПБК определена процедура шифрования параметров нейросетевого биометрического контейнера (НБК) ключом НПБК для обнаружения спуфинг атак. Для этого предполагается, что НБК можно представить в виде матрицы (таблицы):

$$F = \|f_{ij}\|, \;\; i = \overline{1,n} \;\; j = \overline{1,m}$$

где F – матрица номеров признаков НПБК, n – количество синапсов одного нейрона, m – общее число нейронов НПБК. Каждое значение f_{ij} матриц может быть представлено в двоичной системе счисления и записано в соответствующий двоичный вектор:

$$\bar{f} = (0, 1, 1, 0 \ldots 0)$$

где $\bar{f}$ – двоичный вектор, описывающий структуру (последовательность номеров признаков) НПБК. Полученные в результате рассмотренных преобразований векторы подвергаются гаммированию внешнем ключом НПБК для обнаружения спуфинг атак. Для этого методом скользящего окна осуществляется сложение по модулю 2 N-битного ключа K_2 и равного ему участка соответствующего двоичного вектора. Смещение окна производится на значение длины ключа K_2, а в случае если ключ K_2 «не умещается» в $|\bar{f}|$, остаток вектора складывается с его началом.

3. **Блок аутентификации, осуществляющий распределение новых входных образов по классам «Свой»/«Чужой».** Аутентификация субъекта осуществляется в два этапа: полученный на вход образ определяется как

поддельное или реальное изображение. В случае реального изображения ключ НПБК для обнаружения спуфинг атак успешно дешифровывает параметры НБК, после чего пользовательский НПБК принимает решение об аутентификации. В силу того, что сложение по модулю является обратной по отношению к себе операцией, дешифрование структуры НПБК осуществляется аналогичным шифрованию образом: $\left|\bar{f}\right|' \oplus K_2$, где $\left|\bar{f}\right|'$ – зашифрованный ранее двоичный вектор.

Для нивелирования эффекта повышения энтропии выходов пользовательского НПБК при поступлении образа «Чужой» или спуфинг-образа при режиме работы ЗНК необходимо осуществлять предварительную корректировку кода на выходе НПБК для обнаружения спуфинг атак. Такая корректировка возможна за счет применения самокорректирующихся кодов, призванных выявлять и исправлять ошибки, возникающие при передаче данных по зашумленным каналам связи. Наиболее популярными для этой задачи являются так называемые циклические коды, представляющие собой подмножество линейных блоковых кодов, и обладающие свойством циклической инвариантности, при котором любое циклическое сдвинутое кодовое слово также является кодовым словом. Широко используемыми на практике реализациями циклических кодов являются коды Рида-Соломона и коды Бозе-Чоудхури-Хоквингема (БЧХ). Основным недостатком указанных классических кодов с обнаружением и исправлением ошибок является их крайне высокая избыточность.

Несмотря на то, что ключи классических НПБК содержат значительно меньшее количество ошибок по сравнению с ключами на выходе нечетких экстракторов (за счет предварительного обогащения биометрических данных), и для исправления этих ошибок вполне могут подойти описанные выше коды с избыточностью, лучшей практикой для повышения общей производительности системы защищённой аутентификации является применение самокорректирующихся кодов с нулевой избыточностью. К таким кодам относятся так называемые коды Безяева, хранящие синдромы ошибок в виде фрагментов хэш-функций [2]. В своих работах, Безяев А.В. предлагает

использовать предложенную схему для повышения эффективности работы биометрических приложений на базе ИИ, защищенного криптографическими алгоритмами, в том числе для нейросетевых преобразователей «биометрия-код». Экспериментально подтвержденная эффективность применения схемы рекурсивного формирования эталонных хеш-остатков для задачи защищенной биометрии [2] стала ключевым аргументом в пользу выбора кодов Безяева для корректировки выходов НПБК для обнаружения спуфинг атак в предложенной системе.

Описанные блоки системы функционируют в рамках двух ключевых процессов, характерных для систем биометрической аутентификации на основе ИИ: процесс обучения (на рисунке 4.1 представлен пунктирной линией) и процесс эксплуатации (сплошные линии на рисунке 4.1). Процесс обучения включает в себя этапы, связанные с подготовкой, настройкой, а также обучением глубоких нейросетевых моделей и НПБК:

1. **Этап подготовки**. На данном этапе осуществляется сбор первичных данных образов лиц, необходимых для обучения отдельных функциональных элементов системы, а также их предварительная обработка, направленная на улучшение качества изображений. Для создания надежной системы требуется собрать изображения лиц в различных условиях освещения, с разными выражениями лица, углами съемки и фонами. Существенным преимуществом системы, описанной в виде структуры на рисунке 4.1, является отсутствие необходимости сбора большого числа обучающих примеров биометрических персональных данных субъектов, так как модель НПБК на базе тригонометрических нейронов способна работать с малыми выборками лиц. Кроме того, для обучения глубоких нейросетевых алгоритмов (для детекции лиц или извлечения признаков) сбор специализированных биометрических данных не требуется, так как для обучения подходят крупномасштабные наборы данных общего пользования.

Процесс сбора данных для обучения нейросетевых преобразователей и регистрации пользователей в системе должен осуществляться в соответствии с

Федеральным законом «О персональных данных» от 27.07.2006 №152 [22], согласно которому обработка биометрических персональных данных осуществляется только с письменного согласия гражданина.

2. **Этап настройки.** Настройка подразумевает под собой выбор оптимальных параметров функциональных элементов системы, в частности, гиперпараметров моделей глубокого обучения и конфигурации НПБК.

3. **Этап обучения**. Данный этап включает в себя следующие процедуры:

а. Обучение моделей для детектирования лиц на входном изображении. Допустимо использование предобученных архитектур и использование одного детектора для задач обнаружения спуфинг атак и аутентификации.

б. Обучение экстрактора признаков для обнаружения спуфинг изображений.

в. Обучение экстрактора признаков для задачи аутентификации. Выполняется на специальном наборе данных, представляющем собой образы легитимным пользователей системы.

г. Обучение классического нейросетевого преобразователя «биометрия-код» задаче обнаружения спуфинг атак с помощью признаков, полученных из соответствующего экстрактора.

д. Обучение нейросетевых преобразователей «биометрия-код» на базе тригонометрических нейронов, осуществляющих аутентификацию пользователей.

Процесс эксплуатации связан с использованием системы в реальных условиях. Сбор данных на этом этапе предполагает захват изображений лиц пользователей в реальном времени с помощью веб-камер. Аутентификация включает в себя: выделение ключевых признаков из текущего изображения (нового образа); классификацию с помощью НПБК реальных и поддельных изображений и получение соответствующего его решению ключа; корректировку полученного ключа; успешное или неуспешное дешифрование структуры пользовательского НПБК; продуцирование соответствующего ключа с помощью пользовательского НПБК, который затем сравнивается с эталонным ключом из

базы. На основе результатов сравнения система принимает решение о предоставлении или отказе в доступе.

4.2. Архитектура программного обеспечения, реализующего функционал системы защищенной биометрической аутентификации по лицу

Первым этапом реализации любого решения (системы) является определение ключевых требований, позволяющих описать его функциональные возможности и ограничения. Требования служат основой для проектирования и разработки системы, обеспечивая соответствие конечного продукта ожиданиям пользователей и нормативным требованиям. Кроме того, они помогают выявить и учесть потенциальные угрозы безопасности, особенно в контексте биометрических данных, и разработать меры для их защиты, что минимизирует риски компрометации и увеличивает доверие пользователей к системе. В этой связи, разработка архитектуры программного обеспечения, реализующего функционал системы защищенной биометрической аутентификации по лицу, осуществлялась с учетом следующих технических требований (табл. 4.1):

1. **Точность распознавания**. Система должна демонстрировать высокую точность распознавания, характеризуемую минимальными значениями ошибок первого и второго рода. Для достижения этой цели, средний показатель точности (Accuracy) должен составлять не менее 95%, что подразумевает правильное распознавание лиц в 95% случаев из общего числа попыток. Кроме того, система должна обеспечивать уровень ошибок первого рода (FAR) не выше 0,1%, чтобы минимизировать вероятность ошибочного доступа потенциального злоумышленника к системе.

Таблица 4.1 – Требования к реализации системы

Технические требования	**Описание**
Точность распознавания	Средний показатель точности (Accuracy) ≥ 95%. Ошибка первого рода (FAR) ≤ 0,1%.
Скорость обработки	Время обработки одного изображения ≤ 500 мс.

	Пакетная обработка (10 изображений) ≤ 5 секунд. Время обработки изображения высокого качества ≤ 1 секунда.
Масштабируемость	Поддержка горизонтального и вертикального масштабирования. Модульная и компонентная архитектура. Независимое масштабирование модулей.
Аппаратная и программная совместимость	Поддержка многоядерных процессоров x86 и x64 (≥ 2.0 ГГц), графических процессоров NVIDIA с CUDA, минимальный объем RAM — 8 ГБ (рекомендуется 16 ГБ). Дисковое пространство ≥ 50 ГБ (рекомендуется SSD). Поддержка веб-камер, IP-камер, HD, Full HD, 4K камер.
Стандарты и сертификация	Соответствие ФЗ №152-ФЗ 'О персональных данных. Соблюдение ГОСТ Р 52633.5-2011 и ГОСТ Р ИСО/МЭК 19794-5-2014.
Поддержка различных условий освещения	Алгоритмы предварительной обработки: нормализация, адаптивная коррекция яркости и контраста, фильтрация шумов. Обучение на разнообразных данных: разные условия освещения.
Толерантность к изменениям внешности	Алгоритмы глубокого обучения, распознающие изменения во внешности: причёска, очки, борода, макияж, старение, повреждения кожи. Обучение на репрезентативных данных: разные возрастные группы, аксессуары, изменения внешности.

Балансирование между ошибками первого и второго рода является отдельной задачей и должно быть основано на конкретных требованиях и контексте использования системы. В высоко критичных приложениях, таких как банковские системы, правительственные учреждения или объекты с повышенными требованиями безопасности, приоритет отдается минимизации ошибки первого рода даже при допущении некоторого увеличения ошибки второго рода. В менее критичных приложениях, таких как системы доступа в офисные здания или идентификация клиентов в розничных магазинах, можно допустить более высокий уровень FAR с целью обеспечения удобства пользователей и снижения FRR.

2. **Скорость обработки**. Среднее время обработки одного изображения не должно превышать 500 миллисекунд, чтобы обеспечить оперативность распознавания лиц в реальном времени. Это время включает все этапы обработки, начиная с захвата изображения, его предобработки и детектирования лица, и заканчивая извлечением признаков и идентификацией. Кроме того, система

должна быть способна обрабатывать пакетные запросы, сохраняя производительность на приемлемом уровне. В случае пакетной обработки изображений (например, при обработке 10 изображений одновременно), суммарное время обработки не должно превышать 5 секунд.

Дополнительным аспектом, касающимся скорости обработки кадров, является возможность обеспечения стабильной производительности при обработке изображений различного качества и разрешения. Допустимое увеличение времени обработки изображений высокого качества (например, 1920x1080 пикселей) составляет до 1 секунды на изображение. Для систем, предназначенных для использования в условиях, требующих высокой пропускной способности (например, контроль доступа в местах с большим потоком людей), среднее время обработки должно быть дополнительно оптимизировано.

3. **Масштабируемость**. Горизонтальное масштабирование предусматривает возможность распределения нагрузки между несколькими узлами или экземплярами программы на разных компьютерах. Это может быть достигнуто за счет использования распределенных вычислений и параллельной обработки данных. Вертикальное масштабирование подразумевает возможность увеличения производительности системы за счет использования более мощного аппаратного обеспечения на одном компьютере, такого как многоядерные процессоры и дополнительные объемы оперативной памяти. Архитектура системы должна быть модульной и компонентной, что позволяет легко добавлять или изменять функциональные модули без необходимости полной переработки системы. Каждый модуль должен быть независимо масштабируемым, что обеспечивает гибкость в управлении ресурсами и нагрузкой.

4. **Аппаратная и программная совместимость**. Современные системы поддерживают работу на многоядерных процессорах x86 и x64 архитектур с минимальной тактовой частотой 2.0 ГГц. При этом, для разработанной системы, рекомендуется использование процессоров с поддержкой инструкций SIMD (AVX, SSE) для ускорения операций машинного обучения и обработки изображений. Для улучшения производительности, особенно при обработке

больших объемов данных и в режиме реального времени, система должна поддерживать использование графических процессоров NVIDIA с архитектурой CUDA (Compute Capability 3.0 и выше), а также желательна поддержка других GPU, таких как AMD с ROCm. Минимальный объем оперативной памяти для стабильной работы системы должен составлять 8 ГБ, при этом рекомендуется наличие 16 ГБ или более для обработки больших объемов данных и обеспечения многозадачности. Дисковое пространство должно составлять не менее 50 ГБ свободного места на жестком диске для установки и хранения данных, с рекомендацией использования SSD для ускорения операций ввода-вывода. Система должна поддерживать работу с широким спектром камер, включая веб-камеры, IP-камеры и камеры с поддержкой высоких разрешений (HD, Full HD, 4K), а также протоколы захвата видео, такие как USB Video Class (UVC) и RTSP. Кроме того, система должна быть совместима с основными операционными системами, такими как Windows 10 и выше, и Linux (дистрибутивы Ubuntu 18.04 и выше, CentOS 7 и выше).

5. **Стандарты и сертификация**. Система должна соответствовать требованиям Федерального закона №152-ФЗ "О персональных данных", который регулирует обработку и защиту персональных данных. В соответствии с этим законом, система должна обеспечивать конфиденциальность, целостность и доступность биометрических данных, применяя адекватные меры защиты на всех этапах их обработки и хранения. Это включает использование криптографических методов защиты данных в соответствии с требованиями Федеральной службы по техническому и экспортному контролю (ФСТЭК) и Федеральной службы безопасности (ФСБ). Кроме того, необходимо соблюдать требования ГОСТ Р ИСО/МЭК 19794-1-2008 "Автоматическая идентификация. Идентификация биометрическая. Форматы обмена биометрическими данными" [8] и ГОСТ Р ИСО/МЭК 19794-5-2014 "Информационные технологии. Биометрия. Форматы обмена биометрическими данными. Часть 5. Данные изображения лица" [9]. Первый из стандартов устанавливает основные требования к качеству, точности и надежности биометрических систем идентификации, включая процедуры

тестирования и оценки эффективности. Способы хранения и объема изображений лиц, в свою очередь, отражены в рамках второго стандарта.

6. **Поддержка различных условий освещения**. Необходимо наличие алгоритмов предварительной обработки изображений, такие как нормализация, адаптивная коррекция яркости и контраста, а также фильтрация шумов для минимизации влияния теней и неравномерного освещения. Дополнительно, система должна быть обучена на разнообразных наборах данных, включающих изображения, снятые в широком диапазоне условий освещения — от яркого солнечного света до слабого искусственного освещения.

7. **Толерантность к изменениям внешности**. Находящиеся в структуре системы алгоритмы глубокого обучения должны учитывать и распознавать лица с потенциальными отличиями от исходных (например, измененная причёска, добавление очков, старение и временные повреждения кожи). Для этого используется модели, обученные на разнообразных наборах данных, включающих изображения людей в разных возрастных группах, с различными аксессуарами и изменениями внешности. Алгоритмы должны извлекать стабильные биометрические признаки, которые остаются неизменными при вариациях внешности, обеспечивая надёжное распознавание.

С учетом перечисленных требований, а также разработанной ранее структуры системы можно, описать ее программную архитектуру. Очевидно, что для разработки архитектуры необходимо использовать модульный подход, так как в контексте разработки программного обеспечения (ПО) блоки системы аутентификации посредством декомпозиции легко конвертируются в независимые компоненты, именуемые модулями. Каждый модуль реализует определенную часть функциональности системы, имеет четко определенные интерфейсы для взаимодействия с другими модулями и обладает высокой степенью автономности. Это позволяет разрабатывать, тестировать и изменять модули независимо друг от друга, а также повышает гибкость и устойчивость системы к изменениям.

Разработанная с учетом модульного подхода архитектура системы аутентификации состоит из нескольких базовых компонентов (модулей), представленных на диаграмме 4.2: модуль обработки видеопотока, модуль распознавания лиц, модуль обнаружения спуфинг атак, модуль принятия решений об аутентификации, модуль базы данных и модуль логирования и мониторинга.

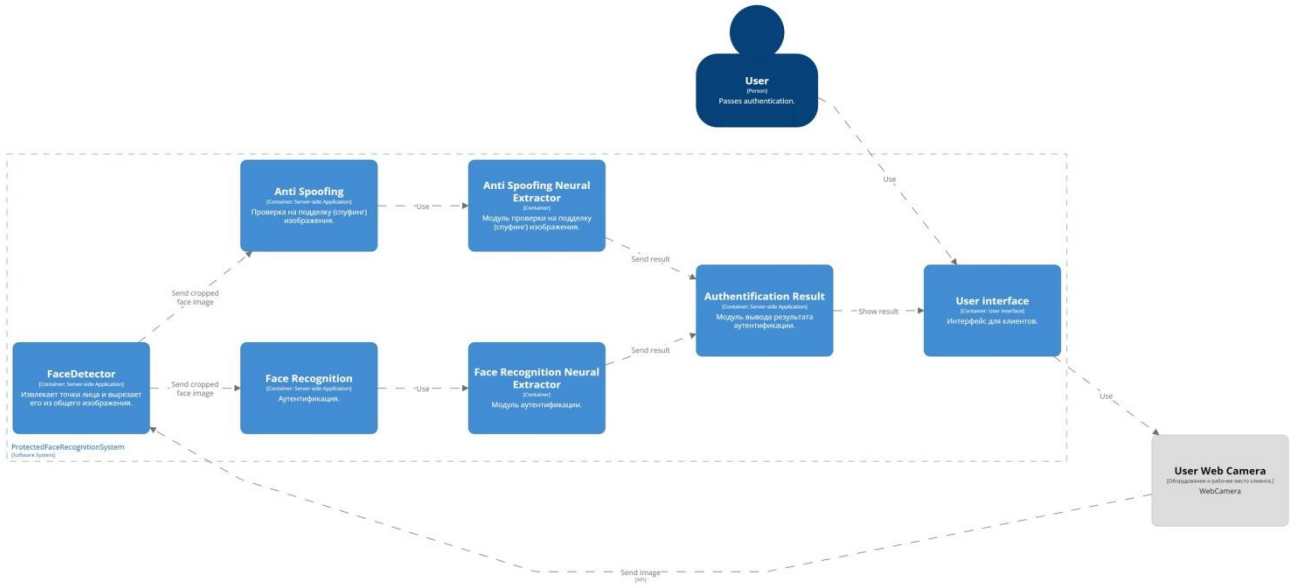


Рисунок 4.2 – Архитектура C4 программной реализации системы

Результатом работы модуля обработки видеопотока являются извлекаемые из видеоизображения кадры (фреймы) с заранее заданной частотой. Частота фреймирования видеопотока зависит от технических характеристик камеры, возможностей обработки данных, а также требований к точности и скорости системы. Дополнительными факторами могут являться сетевые условия или конфигурационные настройки со стороны пользователя. Полученные кадры передаются для дальнейшей обработки в модуль распознавания лиц.

В модуле распознавания лиц нейронные сети, предобученные на соответствующих наборах данных, выполняют детектирование лиц и извлечение из них биометрических признаков. Сверточная нейронная сеть для детекции определяет координаты лица на изображении, а затем сеть для извлечения признаков генерирует вектор, состоящий из 512 элементов. Этот вектор служит

основой для последующей аутентификации и с этой целью отправляется в модуль принятия решений. Параллельно с работой модуля распознавания лиц осуществляется обработка изображения на предмет спуфинг атак. С помощью последовательности операций, включающей детектирование лиц, извлечение признаков и классификацию с помощью классического НПБК, система анализирует изображения на предмет их подлинности. Модуль позволяет сгенерировать криптографический ключ размером > 128 бит, который передается в модуль принятия решений, где используется для проверки легитимности данных.

На следующем этапе система объединяет полученные данные для принятия решения об аутентификации. Модуль, ответственный за этот процесс, принимает вектор признаков и криптографический ключ, после чего пытается расшифровать структуру системы с использованием полученного ключа. Успешная расшифровка указывает на подлинность изображения, и вектор признаков преобразуется в личный ключ пользователя. Сравнение этого ключа с данными, хранящимися в базе данных, позволяет завершить процедуру аутентификации. База данных взаимодействует с другими модулями, предоставляя необходимую информацию для аутентификации и управления учетными записями. Важно отметить, что вся система контролируется модулем логирования и мониторинга, который собирает метрики производительности, фиксирует события и анализирует их для обеспечения стабильности работы. Этот модуль предоставляет отчеты о работе системы, что позволяет оперативно реагировать на любые неполадки, а также поддерживает аудит всех операций.

Каждый из описанных модулей реализован с помощью языка программирования Python. В процессе их программной реализации были использованы такие подходы к написанию кода, как объектно-ориентированное и функциональное программирование. Объектно-ориентированное программирование (ООП) способствует возможности функционального расширения системы в случае добавления новых модулей, а также позволяет инкапсулировать (изолировать) потенциально уязвимую информацию от попыток

внешнего вмешательства. В частности, параметры обучения нейросетевых преобразователей являются приватными полями класса НПБК. Диаграмма класса нейросетевого преобразователя представлена на рисунке 4.3.

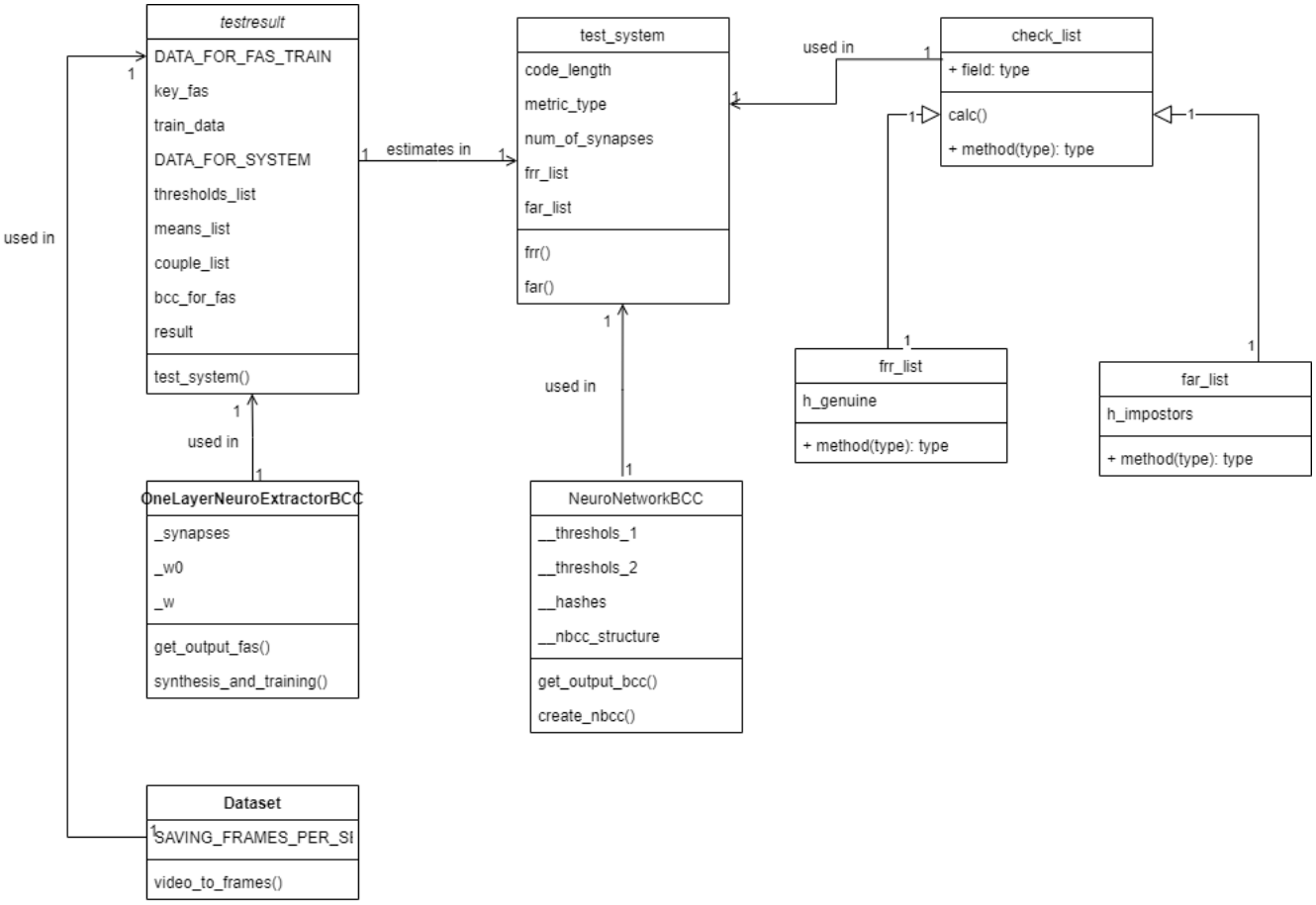


Рисунок 4.3 – UML диаграмма НПБК

В свою очередь, функциональное программирование используется для создания изменяемых и тестируемых функций, обеспечивающих надёжность и простоту верификации отдельных компонентов (модулей) системы (например, функции обработки видеопотока).

Дополнительно, для разработки системы использовался широкий стек технологий, каждая из которых играет независимую роль в создании комплексного решения. Одним из центральных компонентов системы являются библиотеки для обработки изображений и видеопотоков — OpenCV и PIL (Pillow). В частности, OpenCV предоставляет обширный набор инструментов для

обработки изображений и видео, включая функции для захвата, трансформации и анализа визуальных данных в реальном времени. Данная библиотека была применена для захвата входного видеопотока и его «фреймирования» с целью получения отдельных изображений для детектирования и распознавания лиц.

Для создания и обучения моделей глубокого обучения была использована библиотека PyTorch и её расширение torchvision. torchvision ускоряет процесс разработки и тестирования за счёт создания последовательностей преобразования входных изображений и их аугментации. В сочетании с библиотекой facenet_pytorch, которая предоставляет доступ к предобученным моделям MTCNN [183] и InceptionResnet v1 [121], описанные решения позволяют эффективно решать задачи идентификации и верификации лиц. В свою очередь, все процессы обработки и анализа данных для моделей системы осуществляется с помощью библиотек Pandas и NumPy. Начальная разработка и эксперименты над моделями проводились в среде Jupyter Notebook.

Фиксирование логов и отладка системы в случае внештатных ситуаций можно осуществлять с помощью библиотеки Logging, в частности, ее популярного расширения – Loguru. Библиотека существенно облегчает процессы мониторинга и поддержки системы. Для хранения метаданных и результатов работы системы на локальном сервере применяется SQLite. Данная система управления базами данных (СУБД) легко интегрируется с Python через встроенный модуль ‘sqlite3’, который входит в стандартный набор пакетов. SQLite поддерживает реляционные структуры данных и стандартные SQL-запросы, что позволяет выполнить мигрированные уже разработанного решения (системы) на более купные СУБД в случае горизонтального масштабирования. В качестве СУБД для работы с большим количеством данных могут выступать PostgreSQL или MySQL. Организация структуры наиболее простого варианта организации базы данных для разрабатываемой системы, представлена на рисунке 4.4 в виде диаграммы сущностей и связей (Entity-Relationship diagram (ERD)).

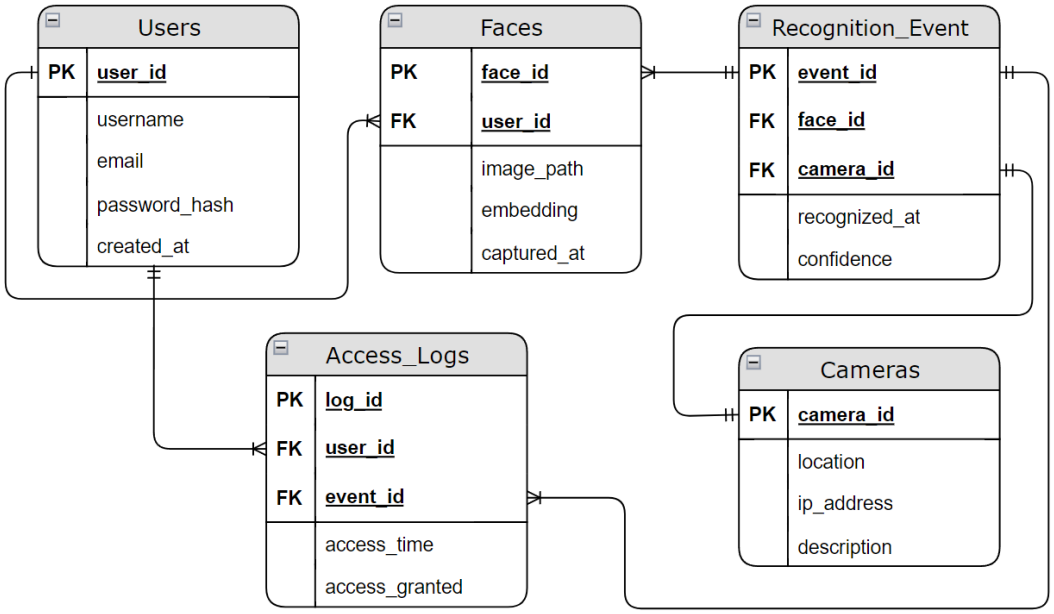


Рисунок 4.4 – Диаграмма сущностей и связей базы данных системы аутентификации

В представленной диаграмме сущностей и связей представлено несколько ключевых таблиц. Таблица пользователей («Users») содержит данные о пользователях, которые могут иметь несколько связанных записей с таблицей лиц («Faces»), что означает, что один пользователь может быть ассоциирован с несколькими лицами. Таблица лиц, в свою очередь, связана с таблицей событий распознавания («Recognition_Events»). Эти события также включают информацию о камерах, задействованных в распознавании, которая хранится в таблице камер («Cameras»). Каждое событие распознавания может быть связано с одной камерой, также как одна камера может участвовать только в одном событии. Таблица журналов доступа («Access_Logs») записывает действия пользователей, регистрируя их взаимодействие с системой и связывая это с соответствующими событиями распознавания.

Границы применимости системы.

Предложенная система защищенной биометрической аутентификации по лицу, в основе которой используются как глубокие нейронные сети, так и широкие нейронные сети автоматического обучения (НПБК), способна эффективно использовать ресурсы благодаря технологиям трасферного обучения

и возможности применения предобученных глубоких моделей. В случае необходимости, однократное переобучение ГНС может потребовать дополнительного количества долговременной памяти в размере от 50 до 150 Гб. В свою очередь, средний объем памяти, занимаемый одной моделью ГНС, составляет около 1 ГБ, что делает их использование крайне экономичным.

С помощью применения нейросетевых преобразователей, обучающихся на малых выборках лиц, достигаются минимальные значения объемов памяти для первичного обучения и настройки системы аутентификации. Выборка для обучения НПБК для обнаружения спуфинг атак собирается один раз из обучающего набора данных крупномасштабного датасета для обучения ГНС. Такая выборка занимает около 50 Мб памяти и может быть удалена сразу после обучения преобразователя. Обучающие данные для аутентификации пользователей также требуют временного хранения только небольшого количества изображений для каждого пользователя (от 7 до 10 примеров), каждое из которых занимает около 200 КБ. Таким образом, для 1000 пользователей потребуется примерно от 1 до 2 ГБ памяти. Хранение ключей для аутентификации в таком объеме используемой памяти практически не влияет на производительность системы, так как современные устройства обладают достаточной емкостью памяти и производительностью процессоров для быстрой обработки данных.

Система оптимизирована для работы на частных устройствах, таких как персональные компьютеры и мобильные устройства, и не требует использования мощных процессоров или большого объема оперативной памяти. Система способна эффективно функционировать на процессорах уровня Intel Core i5 или эквивалентных ему с тактовой частотой 2.5 ГГц и оперативной памятью не менее 16 ГБ. Для ускоренной обработки рекомендуется использование графического процессора, такого как NVIDIA GTX 1650 или выше. Пропускная способность системы позволяет обрабатывать каждый запрос на аутентификацию в течение 200-300 миллисекунд благодаря оптимизации моделей и использованию аппаратного ускорения.

При использовании системы в виде приложения на частных устройствах производительность определяется конфигурацией устройства. Система способна обрабатывать до 10 запросов в секунду на настольных ПК средней мощности и до 2-3 запросов в секунду на современных мобильных устройствах. Система оптимизирована для минимизации задержек и энергопотребления, обеспечивая плавную и эффективную работу даже при интенсивном использовании.

4.3. Построение конвейера обработки данных в системе AIC ModelOps Platform

Разработка систем на основе ИИ подразумевает необходимость эффективного управления жизненным циклом моделей с целью поддержания их работоспособности при различных условиях функционирования. В этой связи крайне актуальным решением указанной задачи являются ModelOps платформы, позволяющие автоматизировать и оптимизировать процессы разработки, тестирования, развертывания и мониторинга моделей искусственного интеллекта. Эти платформы обеспечивают комплексное управление моделями ИИ на всех стадиях их жизненного цикла, что значительно сокращает временные и ресурсные затраты на внедрение моделей в производственные процессы. В условиях возрастающей конкуренции и увеличивающегося объема данных, ModelOps платформы способствуют более быстрой адаптации технологических решений к изменениям.

Важным аспектом использования ModelOps платформ является обеспечение надежности, прозрачности и безопасности процессов работы ИИ. Эти аспекты особенно важны для отраслей с высокой степенью ответственности, таких как здравоохранение, финансы и государственное управление. В этой связи, разработка уникальных ModelOps решений, адаптированных под нужды бизнеса позволяет не только улучшить интеграцию с существующими системами и процессами компании, повышая ее общую эффективность и производительность, но и полностью контролировать доступ к данным и моделям, минимизируя риски

утечек и несанкционированного использования. Кроме того, такая платформа может быть настроена на соответствие внутренним стандартам и нормативным требованиям, что особенно важно в регулированных отраслях.

В ответ на указанные вызовы совместно с автором диссертационного исследования была разработана отечественная ModelOps платформа AIC ModelOps Platform (свидетельство о регистрации программы прил. Б), представляющая собой веб-сервис, входящий в линейку продуктов AIC Constructor и обеспечивающий надежность, прозрачность и безопасность процессов разработки и внедрения искусственного интеллекта. С учетом того, что платформа обладает такими функциональными возможностями, как:

- поддержка широко распространенных библиотек машинного обучения и фреймворков;
- возможность отслеживания эксперимента и расчета базовых метрик оценки производительности;
- интеграция с базами данных;
- возможность импорт внешних наборов данных;
- переиспользование ранее разработанных моделей;
- масштабирование и версионирование моделей, CI/CD модели;
- возможность визуального конструирования бизнес-процессов и систем;
- возможность визуального мониторинга данных и моделей (отображение принимаемых решений и входных данных, на основе которых оно было принято, оповещение при выявлении предвзятости и дрейфующих характеристик);

было принято решение реализовать структуру системы защищенной биометрической аутентификации по лицу в виде цепочки блоков внутри конструктора бизнес-процессов платформы, с целью проведения экспериментальной оценки производительности и эффективности предложенного решения. Для этого предварительно каждая модель глубокого обучения и модели нейросетевых преобразователей, входящие в состав системы и представленная в

виде программного кода на языке программирования Python, реализовывались в виде отдельного «эксперимента».

Эксперимент представляет собой функциональную единицу платформы, позволяющую как таблично, так и графически оценивать уникальную реализацию работы кода модели в виде мета данных. В свою очередь, мета данные представляют собой, так называемые, «логи» и «артефакты» реализации. Лог – это данные эксперимента, которые записываются в серверную базу данных и представляют собой «связку» произвольного количества метрик с произвольным числом гиперпараметров. Метрики можно рассматривать, как многомерную функцию от нескольких аргументов, в качестве которых выступают гиперпараметры. Артефактом может быть как параметры обученной модели (таблицы связей и весовых коэффициентов нейронов), так и результаты ее работы. Артефактом также может являться один из входных параметров эксперимента.

С учетом того, что оптимальные конфигурации моделей глубокого обучения системы были получены при проведении экспериментальных исследований нейросетевых преобразователей, а часть из них является предобученными, реализация каждой модели средствами AIC Platform представляет собой только один эксперимент (рис. 4.5). Однако в случае необходимости (при изменении условий функционирования системы и дрейфа данных) могут быть произведены дополнительные эксперименты в отношении каждой из моделей. Кроме того, с помощью функционала сравнения экспериментов по параметрам и метрикам, можно подобрать новые реализации моделей, при которых достигаются наиболее высокие показатели точности их работы.

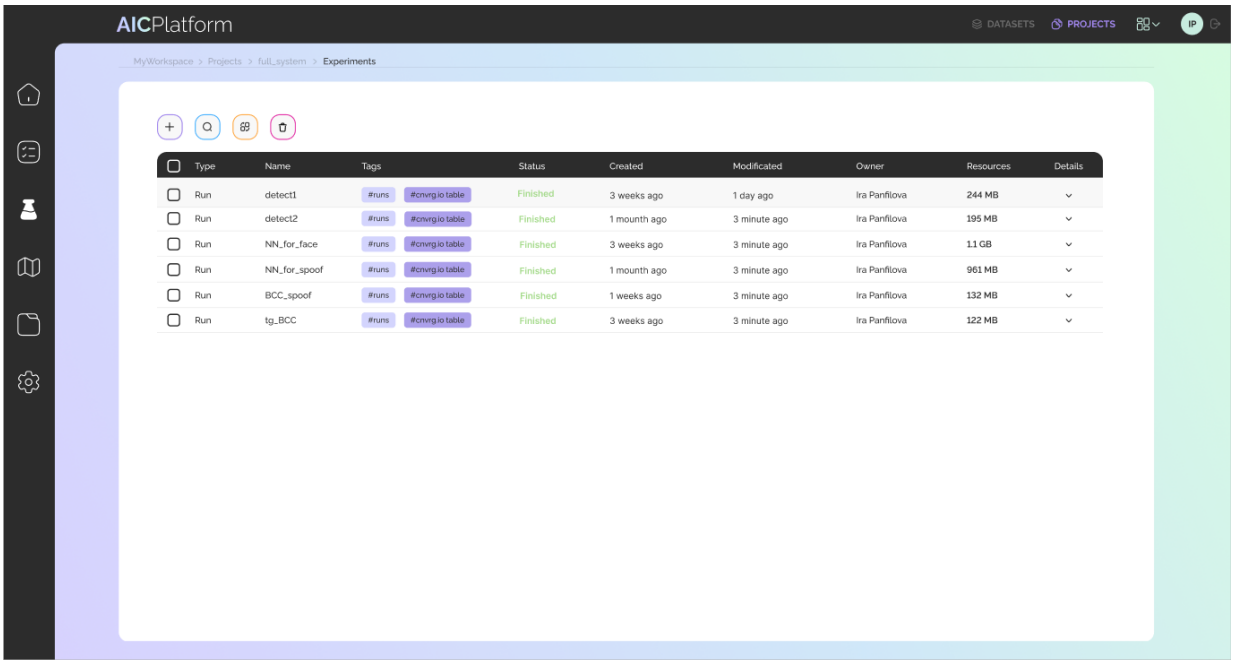


Рисунок 4.5 – Таблица экспериментов разработанной системы аутентификации

Реализация нейросетевых преобразователей представляет собой аналогичный для моделей глубокого обучения процесс создания отдельных экспериментов, однако отличается невозможностью анализа временных характеристик обучения, так как НПБК обучаются автоматически (не итерационно).

Объединение результатов экспериментов в единую структуру осуществляется с помощью конструктора бизнес-процессов, который является эффективным инструментом для удобного проектирования и быстрого запуска конвейеров обработки данных (пайплайнов от англ. pipelines) машинного обучения. Процесс создания нового конвейера начинается с его формирования и переходит к проектированию его составных элементов (задач — исполняемых блоков кода), добавляемых с помощью функционального меню. Каждый новый составной элемент конвейера может быть представлен одним из трех типов (рис. 4.6):

1. Data — блок работы с данными.
2. Exec — блок работы с исполняемыми файлами.
3. Deploy — блок развертывания приложения.

Добавленные в конструктор блоки можно соединять между собой с помощью буквальных связей, что позволяет формировать конвейер. Связи обеспечивают передачу выходных данных одного блока на вход другого, что обеспечивает непрерывный процесс работы приложения. Структура системы, представленная в виде конвейера, изображена на рисунке 4.7. Конструктор бизнес-процессов также поддерживает версионирование конвейеров, позволяя пользователям проектировать системы с использованием различных подходов к их построению.

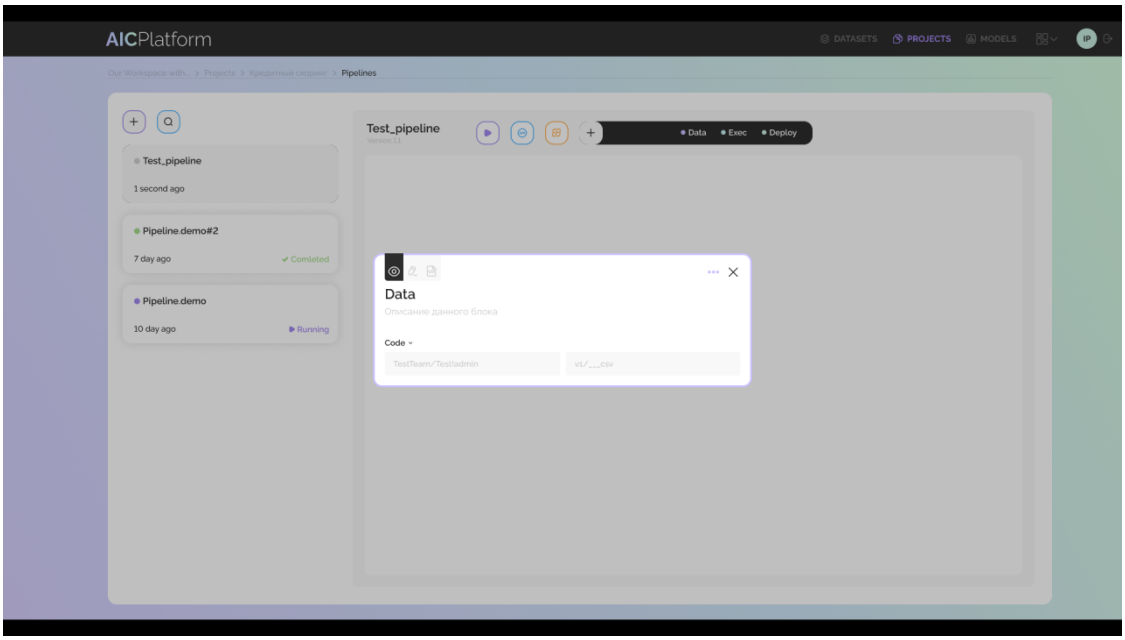


Рисунок 4.6 – Пример построения блока работы с данными

Для проектирования системы с помощью ресурсов конструктора бизнес-процессов потребовалось использовать один блок Data с экспериментальными данными, шесть исполняемых блоков с соответствующими моделями, а также один блок построения приложения и принятия решений об аутентификации.

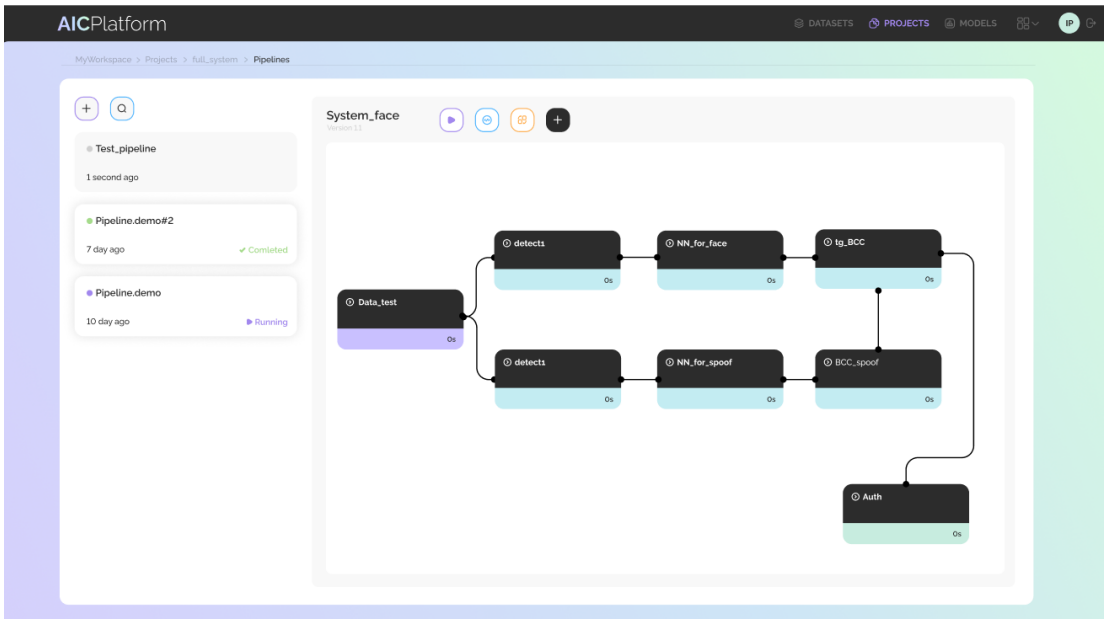


Рисунок 4.7 – Конвейер обработки данных системы аутентификации

Разработка тестового конвейера системы и его экспериментальная оценка являются критически важными этапами, обеспечивающими всестороннюю проверку функциональности, производительности, устойчивости и безопасности предложенной системы биометрической аутентификации по лицу. Кроме того, тестовый конвейер обеспечивает проверку качества данных, интеграции и совместимости всех компонентов системы.

4.4. Экспериментальная оценка надежности системы защищенной биометрической аутентификации по лицу с помощью AIC ModelOps Platform

Целью описанной ниже процедуры тестирования является экспериментальная оценка надежности предложенной системы как комплексного решения для распознавания лиц, исполняемого в защищенном режиме. Под надёжностью, в данном случае, понимается способность системы сохранять приемлемый уровень точности распознавания биометрических образов лиц при заданном уровне защищенности от спуфинг атак, атак извлечения знаний блока принятия решений (НПБК) и компрометации биометрических данных. Экспериментальная реализация осуществляется посредствам запуска конвейера

обработки данных, описанного в предыдущем параграфе и реализованного средствами AIC Platform.

В рамках подготовки данных для тестирования системы защищённой биометрической аутентификации по лицу, был подготовлен набор данных из 120 субъектов, являющийся расширением описанного в главе 3 (параграф 3.5) набора данных SFDv1 и нацеленный на оценку стабильности системы в различных эксплуатационных условиях. Процесс расширения набора данных SFDv1 включал в себя привлечение дополнительных субъектов с целью формирования калибровочного набора в объеме 45 субъектов. Сбор данных происходил по аналогии с исходным набором: каждый дополнительный участник подписывал информированное согласие на участие в эксперименте и обработку данных.

Для усиления набора данных и проверки системы на предмет способности противодействия спуфинг атакам в тренировочный и тестовый блоки полученного набора данных были добавлены изображения, имитирующие попытки обмана системы. Для этого каждый класс был дополнен 10-ью спуфинг образами, включающими изображения фотографий лиц, изображения лиц в бумажных масках в виде лиц легитимных пользователей, а также изображения лиц, полученные с экранов мобильных устройств. Таким образом, итоговый набор данных, названный SFDv2 (таблица 4.2), включает в себя как оригинальные изображения, так и спуфинг-образы.

Таблица 4.2 – Структура набора данных SFDv2

Категория	**Количество изображений на одного субъекта**	**Количество субъектов**	**Общее количество изображений**
Тренировочный/тестовый наборы данных			
Оригинальные изображения (дневной свет)	20	75	1500
Оригинальные изображения (искусственный свет)	20	75	1500
Оригинальные изображения (темное помещение)	20	75	1500
Спуфинг образы (всего)	10	75	750
Распечатанная фотография	4	75	-

Фотография с цифрового устройства (мобильного телефона)	4	75	-
Бумажная маска	2	75	-
Калибровочный набор данных			
Оригинальные изображения (дневной свет)	20	45	900
Оригинальные изображения (искусственный свет)	20	45	900
Оригинальные изображения (темное помещение)	20	45	900
Всего изображений			**7950**

Оценка эффективности и надежности биометрических систем определяется вероятностями ошибок «ложного принятия» (FAR) и «ложного отказа» (FRR), также известными как ошибки первого и второго рода соответственно. Эти показатели взаимосвязаны: изменение порога принятия образа влияет на баланс вероятностей FRR и FAR. Когда FRR и FAR равны, это характеризуется коэффициентом EER (Equal Error Rate), который измеряет вероятность или процент ошибок. В рассматриваемой системе аутентификации порогом служит расстояние Хемминга между бинарным кодом, генерируемым с помощью пользовательского НПБК, и верным ключом пользователя.

Для расчета указанных показателей эффективности (FAR, FRR и EER) набор данных SFDv2 был разделен на две категории: «Зарегистрированные пользователи» (45 субъектов) и «Злоумышленники» (30 субъектов). Для каждого зарегистрированного пользователя был создан пользовательский нейросетевой преобразователь на основе семи случайно выбранных оригинальных образов (изображений), согласно конфигурации, описанной в параграфе 3.5. Остальные образы категории «Свой» (при «искусственном свете» и «отсутствии света», а также спуфинг-образы) использовались для вычисления FRR. Образы из категории «Злоумышленники» (вместе со спуфинг-образами) применялись для оценки FAR. На основе этих данных был вычислен коэффициент EER.

Описанный набор данных необходимо добавить в файловое хранилище AIC Platform (рис. 4.8) для того, что использовать его в блоке Data при построении

конвейера обработки данных. Обладая необходимыми правами доступа, инициировать загрузку набора данных можно через пользовательский интерфейс. Наиболее простым вариантом является загрузка фала в формате.csv непосредственно с локального устройства. Во время загрузки AIC Platform выполняет валидацию файла, проверяя его на соответствие заданным требованиям, таким как корректность структуры, наличие обязательных полей и соответствие формату данных. В случае успешной валидации файл сохраняется в хранилище, присваивается уникальный идентификатор и метаданные, включая дату загрузки, размер файла и информацию о пользователе, выполнившем загрузку.

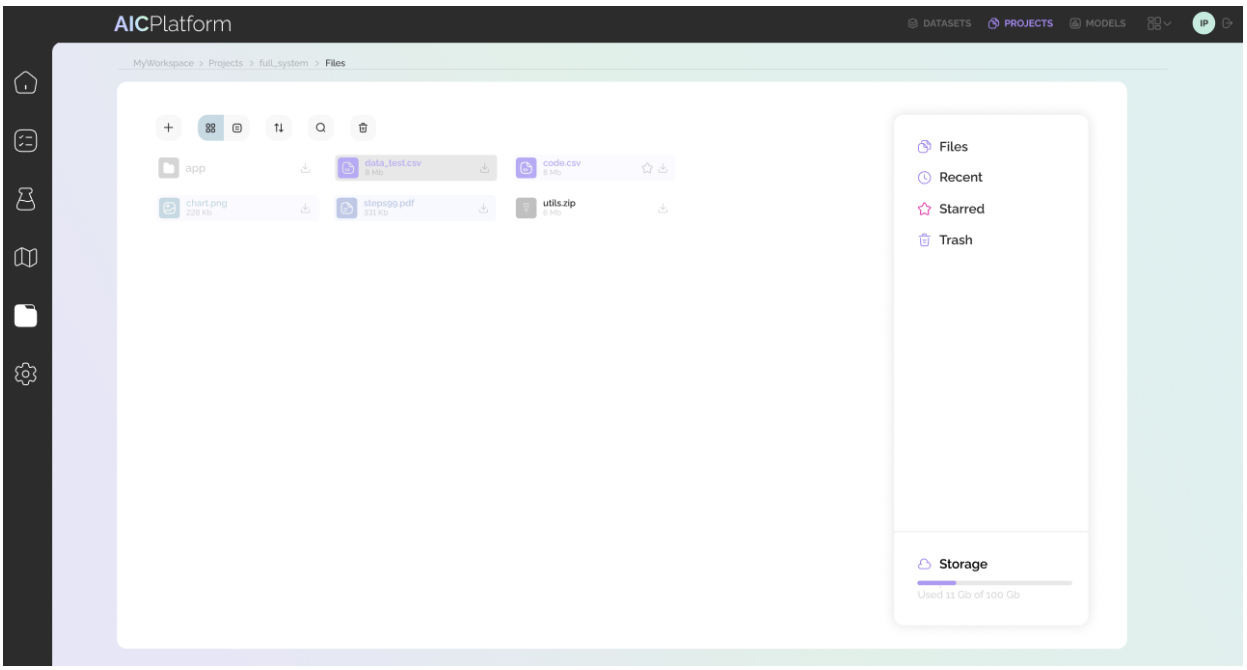


Рисунок 4.8 – Файловое хранилище проекта

При расчете конечной надежности системы, выраженной в виде равной вероятности ошибок (EER), важно учитывать факт использования смешанного набора данных: значения ошибок первого (FRR) и второго рода (FAR) для категорий реальных и поддельных изображений будут разными, так как наличие блока обнаружения спуфинг образов в системе аутентификации будет оказывать следующее корректирующее воздействие на конечную процедуру распознавания:

$$FRR_{\text{аут}} \leq FRR_{\text{общ}} \leq FRR_{\text{аут}} + FRR_{\text{спуф}}$$
$$FAR_{\text{аут}} - FAR_{\text{спуф}} \leq FAR_{\text{общ}} \leq FAR_{\text{аут}}$$

где $FRR_{\text{аут}}$ – ошибка первого рода при работе аутентификации на основе пользовательского НПБК; $FRR_{\text{спуф}}$ – ошибка первого рода блока обнаружения спуфинг атак на основе классического НПБК; $FAR_{\text{аут}}$ – ошибка второго рода при работе аутентификации на основе пользовательского НПБК; $FAR_{\text{спуф}}$ – ошибка второго рода блока обнаружения спуфинг атак на основе классического НПБК; $FRR_{\text{общ}}$ – общее значение ошибки первого рода для системы, работающей на смешанных данных; $FAR_{\text{общ}}$ – общее значение ошибки второго рода для системы, работающей на смешанных данных. Тогда надежность можно оценить по формуле:

$$EER_{\text{общ}} = \frac{FRR_{\text{общ}} + FAR_{\text{общ}}}{2}$$

Для проверки описанной гипотезы 10 раз проводился эксперимент по работе системы на смешанном наборе данных, при этом каждое повторение проводилось независимо с новой выборкой, то сеть с новым случайным выбором данных (7 образов) для построения пользовательских нейросетевых преобразователей. Остальные изображения зарегистрированных пользователей (вместе с их спуфинг образами), использующиеся для вычисления FRR, также перемешивались для того, чтобы избежать систематических ошибок и предвзятости. Калибровка нейросетевых преобразователей выполнялась единожды. После выполнения всех повторений эксперимента, полученные значения EER, представленные в таблицах 4.3 и 4.4, были предварительно усреднены с целью отражения более надежных и устойчивых к случайным выбросам значений.

Таблица 4.3 – Значения EER при 10-ти сценариях работы системы аутентификации

№	1	2	3	4	5	6	7	8	9	10
EER	2,46	2,36	2,57	2,51	2,27	2,6	2,65	2,58	2,73	2,62
Среднее значение EER (%)				≈ 2,5%						
Стандартное отклонение (%)				0.13						
Доверительный интервал (95%)				2.29 - 2.72						

Результаты, представленные в таблицах, позволяют сделать несколько ключевых выводов: среднее значение EER составило ≈ 2.5% (0,0253), что говорит о сравнительно низком уровне ошибок распознавания образов при высоком уровне защищенности процедуры биометрической аутентификации по лицу от атак на биометрическое предъявление, а также атак компрометации знаний НПБК и биометрических данных. Дополнительные значения, в виде стандартного отклонения и доверительного интервала, необходимы для полной интерпретации результатов и обеспечивают дополнительную информацию о надежности и точности системы. Небольшое стандартное отклонение, полученное по результатам эксперимента, указывает на то, что значения EER не сильно варьируются между экспериментами, что свидетельствует о стабильности системы. Полученный «узкий» доверительный интервал в диапазоне (от 2.29% до 2.72%) также демонстрирует низкую вариативность полученных результатов, и, следовательно, высокую достоверность усредненного значения.

4.5. Внедрение результатов исследования

Результаты исследований нашли свое применение как в корпоративной среде, так и в образовательных учреждениях. Разработанные концепция, модели и алгоритмы прошли успешную апробацию в рамках компании ООО «Открытый код» города Самары и использованы при разработке «Системы обеспечения

безопасности и выявления девиантного поведения по видеоизображению на основе базы знаний и искусственных нейронных сетей». Внедрённые решения позволили повысить устойчивость биометрической системы предприятия к спуфинг атакам, а также обеспечить конфиденциальность биометрических образов сотрудников компании.

В учебные процессы результаты диссертации были внедрены в рамках образовательных программ для подготовки магистров по направлению 10.04.01 «Информационная безопасность» ФГБОУ ВО «Самарский государственный технический университет». Предложенные модели и алгоритмы применяются при изучении дисциплин «Интеллектуальные системы и базы данных» и «Информационно-аналитические системы безопасности». Кроме того, предложенные решения используются для подготовки магистров по направлению 09.04.01 «Информатика и вычислительная техника» (направленность «Безопасность и этика искусственного интеллекта») во ФГАОУ ВО «Омский государственный технический университет» при изучении дисциплин «Машинное обучение в приложениях биометрии» и «Защищенное исполнение искусственного интеллекта». Внедрение результатов исследования способствует подготовке высококвалифицированных специалистов, способных эффективно работать с передовыми технологиями в области искусственного интеллекта и биометрии.

Выводы по четвертой главе

По результатам главы разработана система защищенной биометрической аутентификации по лицу, отвечающая современным техническим, правовым и организационным требованиям. Применение современных методов глубокого обучения, трансферного обучения и нейросетевых преобразователей позволило создать эффективную и надежную систему, способную работать на персональных компьютерах и мобильных устройствах. Модульный подход к архитектуре системы позволил разделить её на функциональные блоки, каждый из которых выполняет независимые задачи, такие как детекция лиц, обнаружение спуфинг-

атак и принятие решений об аутентификации. Отдельное внимание в ходе разработки системы уделено вопросам соответствия нормативным требованиям и стандартам, таким как Федеральный закон №152-ФЗ "О персональных данных", ГОСТ Р ИСО/МЭК 19794-1-2008 и ГОСТ Р ИСО/МЭК 19794-5-2014. Реализована поддержка различных аппаратных конфигураций и операционных систем, что позволяет интегрировать систему в широкий спектр IT-инфраструктур.

Результаты экспериментального тестирования системы, проведенного с использованием платформы AIC ModelOps Platform, подтвердили высокую эффективность и надёжность системы при работе в реальных условиях. Средний коэффициент равной вероятности ошибок (EER) составил 2,5%, что указывает на низкий уровень ошибок при аутентификации и высокую устойчивость системы к деструктивным воздействиям. Полученные показатели демонстрируют, что система способна обеспечивать надёжную защиту биометрических данных при высоком уровне производительности, что делает её подходящей для использования в задачах, требующих повышенной безопасности и точности распознавания.

Таким образом, предложенная система биометрической аутентификации по лицу не только демонстрирует высокие технические характеристики, но и полностью соответствует требованиям безопасности и защищенности, что делает её перспективным решением для применения в критически важных секторах, таких как банковское дело, государственные учреждения и объекты с повышенными требованиями к безопасности.

Заключение

В рамках диссертационной работы решена актуальная задача повышения защищенности реализуемой с помощью нейросетевого преобразователя «биометрия-код» процедуры биометрической аутентификации по лицу в отношении деструктивных воздействий (атак извлечения знаний НПБК, компрометации открытых биометрических образов лиц и спуфинг атак). К основным результатам диссертационной работы относятся следующие теоретические и практические результаты:

1. Разработана концепция защищенной биометрической аутентификации по лицу, обеспечивающая устойчивость к атакам на биометрическое предъявление (спуфинг атак). За счет применения дополнительного нейросетевого преобразователя, обученного задаче классификации реальных и поддельных изображений, решается задача противодействия спуфинг атакам, а также обеспечивается дополнительная защита параметров пользовательского НПБК. Защита осуществляется путем применения механизма защиты нейросетевого контейнера (ЗНК). Точность классификации нейросетевого преобразователя, адаптированного для задачи обнаружения спуфинг атак, составляет 97,2% на тестовых выборках, что сравнимо с точностью распознавания спуфинг атак с помощью аналогичных архитектур, обучаемых на основе Cross-Entropy Loss.

2. Разработаны модель тригонометрического нейрона и основанная не ней модель нейросетевого преобразователя биометрия-код (НПБК), позволяющие работать со слабо коррелированными признаками лица человека и продуцировать длинный криптографический ключ на выходе НПБК при высокой точности классификации биометрических образов. Предложенные модели не используют параметры распределений и/или характеристики образов «Свой», что обеспечивает защиту биометрических данных и знаний НПБК от компрометации. В рамках проведенных экспериментов продемонстрированы высокие показатели эффективности итогового исполнения НПБК, продуцирующего ключ длиной в 2048 по сравнению со 128 битами классического НПБК, обученного в

соответствии с ГОСТ Р 52633.5. При этом достигнуты сравнительно высокие показатели точности распознавания EER ≈ 0.024.

3. Разработаны алгоритмы предварительной калибровки нейросетевого преобразователя биометрия-код и алгоритм автоматического обучения НПБК на основе тригонометрических нейронов на малых выборках. Совместная работа алгоритмов позволяет предварительно оценить особенности распределений биометрических образов лиц, не компрометирующих легитимных пользователей, с целью последующей сборки и быстрого обучения НПБК, не требующего большого числа примеров «Свой». Проведенные эксперименты продемонстрировали возможность использования всего 7 образов.

4. Разработана структура системы защищенной биометрической аутентификации по лицу, в которой за счет работы независимых блоков извлечения признаков, обучения нейросетевых преобразователей и аутентификации, а также применения варианта исполнения ЗНК, при котором шифруется структура (порядок расположения синапсов в нейронах) пользовательского НПБК обеспечивается защищенность процедуры биометрической аутентификации личности по лицу на основе НПБК в отношении спуфинг атак и атак компрометации знаний НПБК и открытых биометрических образов лиц. Коэффициент равной вероятности ошибок при работе системы составил EER=0,025, что говорит о сравнительно низком уровне ошибок распознавания образов при высоком уровне защищенности процедуры биометрической аутентификации по лицу от атак на биометрическое предъявление, а также атак компрометации знаний НПБК и биометрических данных.

Результаты диссертационного исследования приняты к внедрению в производственные и бизнес-процессы компаний ООО «Открытый код» г. Самары и ООО «АИ ЗИОН» г. Омска, а также в учебные процессы ФГБОУ ВО «Самарский государственный технический университет» и ФГАОУ ВО «Омский государственный технический университет».

Перспективы дальнейшего развития темы. В рамках дальнейших исследований планируется разработка защищенной биометрической аутентификации, устойчивой к новому типу атак — дипфейкам. Дипфейки представляют собой высококачественные подделки, созданные с использованием методов глубокого обучения, таких как генеративно-состязательные сети (GAN), способные имитировать внешность и голос реального человека с высокой степенью точности. Интеграция методов защиты от дипфейков с алгоритмами анти-спуфинга и специализированными моделями защищенной биометрической аутентификации становится критически важной для обеспечения безопасности данных пользователей и предотвращения мошенничества.

Список сокращений и условных обозначений

БКС – биометрика-криптографическая система

ГНС – глубокая нейронная сеть

ГШ – гомоморфное шифрование

ЗБА – защищенная биометрическая аутентификация

ЗБШ – защита биометрических шаблонов

ЗНК – защита нейросетевого контейнера

ИИ – искусственный интеллект

ИНС – искусственная нейронная сеть

ИС – информационная система

НБК – нейросетевой биометрический контейнер

НПБК – нейросетевой преобразователь «биометрия-код»

ОБ – отменяемая биометрия

СНС – сверточная нейронная сеть

ПО – программное обеспечение

EER – Equal Error Rate (коэффициент равной вероятности ошибок)

FAR – False access rate (ошибка «ложного допуска»)

FRR – False reject rate (ошибка «ложного отказа»)

Список литературы

1. Ахметов Б. Б. и др. Быстрый алгоритм оценки высокоразмерной энтропии биометрических образов на малых выборках // Труды Международного симпозиума «Надежность и качество». – 2015. – Т. 2. – С. 285-287.

2. Безяев А. В., Иванов А. И., Фунтикова Ю. В. Оптимизация структуры самокорректирующегося био-кода, хранящего синдромы ошибок в виде фрагментов хеш-функций //Вестник УрФО. Безопасность в информационной сфере. – 2014. – №. 3 (13). – С. 4-13.

3. Васильев В.И., Панфилова И.Е., Сулавко А.Е., Серикова А.Е. Система верификации личности по изображению лица в защищенном режиме на основе искусственных нейронных сетей // Прикладная информатика. – 2023. – Т. 18. № 5. – С. 33–47.

4. Волчихин В. И. и др. Соотношение мощности нейронов с линейным и квадратичным обогатителями биометрических данных //Известия высших учебных заведений. Поволжский регион. Технические науки. – 2018. – №. 1 (45). – С. 17-25.

5. Волчихин В. И., Иванов А. И., Серикова Ю. И. Компенсация методических погрешностей вычисления стандартных отклонений и коэффициентов корреляции, возникающих из-за малого объема выборок //Известия высших учебных заведений. Поволжский регион. Технические науки. – 2016. – №. 1 (37). – С. 103-110.

6. ГОСТ Р 52633.5-2011. Защита информации. Техника защиты информации. Автоматическое обучение нейросетевых преобразователей биометрия-код доступа. – Москва: Стандартинформ, 2013. 16 с.

7. ГОСТ Р 59276-2020. Системы искусственного интеллекта. Способы обеспечения доверия. Общие положения. – Москва : Стандартинформ, 2021.

8. ГОСТ Р ИСО/МЭК 19794-1-2008. Автоматическая идентификация. Идентификация биометрическая. Форматы обмена биометрическими данными. – Москва : Стандартинформ, 2009.

9. ГОСТ Р ИСО/МЭК 19794-5-2014. Информационные технологии. Биометрия. Форматы обмена биометрическими данными. Часть 5. Данные изображения лица. – Москва : Стандартинформ, 2015.

10. Жумажанова С. С., Панфилова И. Е., Ложников П. С., Сулавко А. Е., Серикова А. Е. Биометрическая аутентификация по тепловым изображениям лица на основе преобразователей "биометрия-код" // Вопросы защиты информации: Науч.-практ. журн. ФГУП «НТЦ оборонного комплекса «Компас». – 2023. – Вып. 1 (140). – С. 8—19.

11. Крохин И. А., Михеев М. Ю. Противодействие атакам Маршалко на сети искусственных нейронов за счет введения ложных связей //Надежность и качество сложных систем. – 2022. – №. 3 (39). – С. 86-94.

12. Майоров А. В. и др. Оценка стойкости защищенных нейросетевых преобразователей биометрия-код с использованием больших баз синтетических биометрических образов //Известия высших учебных заведений. Поволжский регион. Технические науки. – 2018. – №. 4 (48). – С. 65-74.

13. Панфилова И. Е. Глубокие нейронные сети в задачах идентификации и верификации лиц // Вопросы защиты информации: Науч.-практ. журн. ФГУП «НТЦ оборонного комплекса «Компас». – 2024. – Вып. 2 (145). – С. 33—41.

14. Панфилова И. Е., Сулавко А. Е. Методы определения живого присутствия пользователя перед видеокамерой в задачах биометрической аутентификации по лицу / Вопросы защиты информации: Науч.-практ. журн. ФГУП «НТЦ оборонного комплекса «Компас». – 2023. – Вып. 2 (141). – С. 17—26.

15. Панфилова И. Е., Сулавко А. Е., Ложников П. С. Повышение защищенности процедуры биометрической аутентификации по лицу на основе нейросетевых преобразователей «биометрия-код» // Вопросы защиты информации: Науч.-практ. журн. ФГУП «НТЦ оборонного комплекса «Компас». – 2024. – Вып. 3 (146). – С. 3-11

16. Панфилова И.Е., Ниватов Д.П. Обзор методов защиты данных биометрических шаблонов // Сборник научных статей по материалам V

Всероссийской научно-технической конференции «Безопасность информационных технологий». — Том 1.— 2023. – С. 135-145.

17. Панфилова И.Е., Ложников П.С. Исследование применимости нейросетевых преобразователей «биометрия-код» для задачи обнаружения атак на биометрическое предъявление // Вестник УрФО. Безопасность в информационной сфере. – 2024. – Т. 2. – №. 52. – С. 106-121.

18. Приказ Минцифры России № 902 [электронный ресурс]. Режим доступа: https://digital.gov.ru/ru/documents/8062/, свободный (дата обращения: 21.08.2024).

19. Романов Д. Р. Пентест: эффективная методология и инструменты //ББК 1 Н 34. – С. 1224.

20. Сулавко А. Е. Искусственный интеллект в защищенном исполнении //Информационная безопасность: современная теория и практика. – 2020. – С. 112-114.

21. Сулавко А.Е., Панфилова И.Е. Верификация личности субъектов по лицу на основе методов глубокого обучения и нейросетевых преобразователей «биометрия-код» //Нанотехнологии. Информация. Радиотехника (НИР-23). – 2023. – С. 336-340.

22. Федеральный закон "О персональных данных" от 27.07.2006 N 152-ФЗ (последняя редакция) [электронный ресурс]. Режим доступа: https://www.consultant.ru/document/cons_doc_LAW_61801/ , свободный (дата обращения: 21.08.2024).

23. Abdullahi S. M. et al. Biometric template attacks and recent protection mechanisms: A survey //Information Fusion. – 2024. – Т. 103. – С. 102144.

24. Agrahari S., Singh A. K. Concept drift detection in data stream mining: A literature review //Journal of King Saud University-Computer and Information Sciences. – 2022. – Т. 34. – №. 10. – С. 9523-9540.

25. Akhmetov B. S., Ivanov A. I., Alimseitova Z. K. Training of neural network biometry-code converters //News of the National Academy of Sciences of the Republic of Kazakhstan, Series of Geology and Technical Sciences. – 2018. – Т. 1. – №. 427. – С. 61-68.

26. Albalawi S. et al. A comprehensive overview on biometric authentication systems using artificial intelligence techniques //International Journal of Advanced Computer Science and Applications. – 2022. – Т. 13. – №. 4. – С. 1-11.

27. Anjos A., Chakka M. M., Marcel S. Motion-based counter-measures to photo attacks in face recognition //IET biometrics. – 2014. – Т. 3. – №. 3. – С. 147-158.

28. Atoum Y. et al. Face anti-spoofing using patch and depth-based CNNs //2017 IEEE international joint conference on biometrics (IJCB). – IEEE, 2017. – С. 319-328.

29. Bansal A. et al. The do's and don'ts for CNN-based face verification //Proceedings of the IEEE international conference on computer vision workshops. – 2017. – С. 2545-2554.

30. Bao J. et al. Towards open-set identity preserving face synthesis //Proceedings of the IEEE conference on computer vision and pattern recognition. – 2018. – С. 6713-6722.

31. Bao W. et al. A liveness detection method for face recognition based on optical flow field //2009 International Conference on Image Analysis and Signal Processing. – IEEE, 2009. – С. 233-236.

32. Bassit A. et al. Hybrid biometric template protection: Resolving the agony of choice between bloom filters and homomorphic encryption //IET biometrics. – 2022. – Т. 11. – №. 5. – С. 430-444.

33. Beham M. P., Roomi S. M., Dharmalakshmi D. Face spoofing detection based on depthmap and gradient binary pattern //International Journal of Applied Engineering Research. – Т. 9. – №. 21. – С. 2014.

34. Bogdanov D. S., Mironkin V.O. Data recovery for a neural network-based biometric authentication scheme //Математические вопросы криптографии. – 2019. – Т. 10. – №. 2. – С. 61–74.

35. Boulkenafet Z. et al. OULU-NPU: A mobile face presentation attack database with real-world variations //2017 12th IEEE international conference on automatic face & gesture recognition (FG 2017). – IEEE, 2017. – С. 612-618.

36. Boulkenafet Z., Komulainen J., Hadid A. Face anti-spoofing using speeded-up robust features and fisher vector encoding //IEEE Signal Processing Letters. – 2016. – Т. 24. – №. 2. – С. 141-145.

37. Cai R., Chen C. Learning deep forest with multi-scale local binary pattern features for face anti-spoofing //arXiv preprint arXiv:1910.03850. – 2019.

38. Cao Q. et al. Vggface2: A dataset for recognising faces across pose and age //2018 13th IEEE international conference on automatic face & gesture recognition (FG 2018). – IEEE, 2018. – С. 67-74.

39. CelebA Spoof Depth Image [электронный ресурс]. Режим доступа: https://www.kaggle.com/datasets/attentionlayer241/celeba-spoof-depth-image, свободный (дата обращения: 21.08.2024).

40. Chakraborty A., Biswas A., Khan A. K. Artificial intelligence for cybersecurity: Threats, attacks and mitigation //Artificial Intelligence for Societal Issues. – Cham : Springer International Publishing, 2023. – С. 3-25.

41. Chen H. et al. Attention-based two-stream convolutional networks for face spoofing detection //IEEE Transactions on Information Forensics and Security. – 2019. – Т. 15. – С. 578-593.

42. Chen W., Hu H. Generative attention adversarial classification network for unsupervised domain adaptation //Pattern Recognition. – 2020. – Т. 107. – С. 107440.

43. Chingovska I., Anjos A., Marcel S. On the Effectiveness of Local Binary Patterns in Face Anti-spoofing//2012 BIOSIG-proceedings of the International Conference of the Biometrics Special Interest Group (BIOSIG). – IEEE, 2012. – С. 1-7.

44. Chollet F. Xception: Deep learning with depthwise separable convolutions //Proceedings of the IEEE conference on computer vision and pattern recognition. – 2017. – С. 1251-1258.

45. Crosswhite N. et al. Template adaptation for face verification and identification //Image and Vision Computing. – 2018. – Т. 79. – С. 35-48.

46. Das D., Chakraborty S. Face liveness detection based on frequency and micro-texture analysis //2014 International Conference on Advances in Engineering & Technology Research (ICAETR-2014). – IEEE, 2014. – С. 1-4.

47. Deng J. et al. Arcface: Additive angular margin loss for deep face recognition //Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. – 2019. – С. 4690-4699.

48. Deng J. et al. Retinaface: Single-shot multi-level face localisation in the wild //Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. – 2020. – С. 5203-5212.

49. Dodis Y. et al. Fuzzy extractors: How to generate strong keys from biometrics and other noisy data //SIAM journal on computing. – 2008. – Т. 38. – №. 1. – С. 97-139.

50. Dong S., Wang P., Abbas K. A survey on deep learning and its applications //Computer Science Review. – 2021. – Т. 40. – С. 100379.

51. Dong X. et al. Secure chaff-less fuzzy vault for face identification systems //ACM Transactions on Multimidia Computing Communications and Applications. – 2021. – Т. 17. – №. 3. – С. 1-22.

52. Duan Q., Zhang L. Look more into occlusion: Realistic face frontalization and recognition with boostgan //IEEE transactions on neural networks and learning systems. – 2020. – Т. 32. – №. 1. – С. 214-228.

53. Duong C. N. et al. Mobiface: A lightweight deep learning face recognition on mobile devices //2019 IEEE 10th international conference on biometrics theory, applications and systems (BTAS). – IEEE, 2019. – С. 1-6.

54. Galbally J., Satta R. Three-dimensional and two-and-a-half dimensional face recognition spoofing using three-dimensional printed models //IET Biometrics. – 2016. – Т. 5. – №. 2. – С. 83-91.

55. Garcia J. L. C. et al. Securing AI Systems: A Comprehensive Overview of Cryptographic Techniques for Enhanced Confidentiality and Integrity //2024 13th Mediterranean Conference on Embedded Computing (MECO). – IEEE, 2024. – С. 1-8.

56. George A. et al. Biometric face presentation attack detection with multi-channel convolutional neural network //IEEE transactions on information forensics and security. – 2019. – Т. 15. – С. 42-55.

57. George A. et al. EdgeFace: Efficient Face Recognition Model for Edge Devices //arXiv e-prints. – 2023. – C. arXiv: 2307.01838.

58. George A., Marcel S. Cross modal focal loss for rgbd face anti-spoofing //Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. – 2021. – C. 7882-7891.

59. George A., Marcel S. Deep pixel-wise binary supervision for face presentation attack detection //2019 International Conference on Biometrics (ICB). – IEEE, 2019. – C. 1-8.

60. George A., Marcel S. Learning one class representations for face presentation attack detection using multi-channel convolutional neural networks //IEEE Transactions on Information Forensics and Security. – 2020. – T. 16. – C. 361-375.

61. George A., Marcel S. On the effectiveness of vision transformers for zero-shot face anti-spoofing // arXiv preprint, 2020.

62. Gilkalaye B. P., Rattani A., Derakhshani R. Euclidean-distance based fuzzy commitment scheme for biometric template security //2019 7th International Workshop on Biometrics and Forensics (IWBF). – IEEE, 2019. – C. 1-6.

63. He K. et al. Deep residual learning for image recognition //Proceedings of the IEEE conference on computer vision and pattern recognition. – 2016. – C. 770-778.

64. Heusch G. et al. Deep models and shortwave infrared information to detect face presentation attacks //IEEE Transactions on Biometrics, Behavior, and Identity Science. – 2020. – T. 2. – №. 4. – C. 399-409.

65. Howard A. G. Mobilenets: Efficient convolutional neural networks for mobile vision applications //arXiv preprint arXiv:1704.04861. – 2017.

66. Hu J., Shen L., Sun G. Squeeze-and-excitation networks //Proceedings of the IEEE conference on computer vision and pattern recognition. – 2018. – C. 7132-7141.

67. Hu Y. et al. Artificial intelligence security: Threats and countermeasures //ACM Computing Surveys (CSUR). – 2021. – T. 55. – №. 1. – C. 1-36.

68. Huang G. B. et al. Labeled faces in the wild: A database forstudying face recognition in unconstrained environments //Workshop on faces in'Real-Life'Images: detection, alignment, and recognition. – 2008.

69. Huang G. et al. Densely connected convolutional networks //Proceedings of the IEEE conference on computer vision and pattern recognition. – 2017. – С. 4700-4708.

70. Iandola F. N. SqueezeNet: AlexNet-level accuracy with 50x fewer parameters and< 0.5 MB model size //arXiv preprint arXiv:1602.07360. – 2016.

71. Jourabloo A., Liu X. Large-pose face alignment via CNN-based dense 3D model fitting //Proceedings of the IEEE conference on computer vision and pattern recognition. – 2016. – С. 4188-4196.

72. Kaviani S., Han K. J., Sohn I. Adversarial attacks and defenses on AI in medical imaging informatics: A survey //Expert Systems with Applications. – 2022. – Т. 198. – С. 116815.

73. Khairnar S. et al. Face liveness detection using artificial intelligence techniques: A systematic literature review and future directions //Big Data and Cognitive Computing. – 2023. – Т. 7. – №. 1. – С. 37.

74. Kim G. et al. Face liveness detection based on texture and frequency analyses //2012 5th IAPR international conference on biometrics (ICB). – IEEE, 2012. – С. 67-72.

75. Komulainen J., Hadid A., Pietikäinen M. Context based face anti-spoofing //2013 IEEE Sixth International Conference on Biometrics: Theory, Applications and Systems (BTAS). – IEEE, 2013. – С. 1-8.

76. Kose N., Dugelay J. L. On the vulnerability of face recognition systems to spoofing mask attacks //2013 IEEE International Conference on Acoustics, Speech and Signal Processing. – IEEE, 2013. – С. 2357-2361.

77. Krizhevsky A., Sutskever I., Hinton G. E. Imagenet classification with deep convolutional neural networks //Advances in neural information processing systems. – 2012. – Т. 25.

78. Kumar Jindal A., Chalamala S., Kumar Jami S. Face template protection using deep convolutional neural network //Proceedings of the IEEE conference on computer vision and pattern recognition workshops. – 2018. – С. 462-470.

79. Kumar Pandey R. et al. Deep secure encoding for face template protection //Proceedings of the IEEE conference on computer vision and pattern recognition workshops. – 2016. – С. 9-15.

80. Kuo C. W. et al. Featmatch: Feature-based augmentation for semi-supervised learning //Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part XVIII 16. – Springer International Publishing, 2020. – С. 479-495.

81. Li J. W. Eye blink detection based on multiple Gabor response waves //2008 International Conference on Machine Learning and Cybernetics. – IEEE, 2008. – Т. 5. – С. 2852-2856.

82. Li X. et al. 3DPC-Net: 3D point cloud network for face anti-spoofing //2020 IEEE International Joint Conference on Biometrics (IJCB). – IEEE, 2020. – С. 1-8.

83. Li X. et al. Generalized face anti-spoofing by detecting pulse from face videos //2016 23rd International Conference on Pattern Recognition (ICPR). – IEEE, 2016. – С. 4244-4249.

84. Li Z. et al. Unseen face presentation attack detection with hypersphere loss //ICASSP 2020-2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). – IEEE, 2020. – С. 2852-2856.

85. Lin B. et al. Face liveness detection by rppg features and contextual patch-based cnn //Proceedings of the 2019 3rd international conference on biometric engineering and applications. – 2019. – С. 61-68.

86. Lin C. H., Huang W. J., Wu B. F. Deep representation alignment network for pose-invariant face recognition //Neurocomputing. – 2021. – Т. 464. – С. 485-496.

87. Lin T. Y. et al. Focal loss for dense object detection //Proceedings of the IEEE international conference on computer vision. – 2017. – С. 2980-2988.

88. Liu A. et al. Casia-surf cefa: A benchmark for multi-modal cross-ethnicity face anti-spoofing //Proceedings of the IEEE/CVF winter conference on applications of computer vision. – 2021. – С. 1179-1187.

89. Liu A. et al. Contrastive context-aware learning for 3d high-fidelity mask face presentation attack detection //IEEE Transactions on Information Forensics and Security. – 2022. – Т. 17. – С. 2497-2507.

90. Liu A. et al. Face anti-spoofing via adversarial cross-modality translation //IEEE Transactions on Information Forensics and Security. – 2021. – Т. 16. – С. 2759-2772.

91. Liu B. et al. Fair loss: Margin-aware reinforcement learning for deep face recognition //Proceedings of the IEEE/CVF international conference on computer vision. – 2019. – С. 10052-10061.

92. Liu F. et al. Deep learning based single sample face recognition: a survey //Artificial Intelligence Review. – 2023. – Т. 56. – №. 3. – С. 2723-2748.

93. Liu H. et al. Adaptiveface: Adaptive margin and sampling for face recognition //Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. – 2019. – С. 11947-11956.

94. Liu J. et al. Targeting ultimate accuracy: Face recognition via deep embedding //arXiv preprint arXiv:1506.07310. – 2015.

95. Liu W. et al. Large-margin softmax loss for convolutional neural networks //arXiv preprint arXiv:1612.02295. – 2016.

96. Liu W. et al. Sphereface: Deep hypersphere embedding for face recognition //Proceedings of the IEEE conference on computer vision and pattern recognition. – 2017. – С. 212-220.

97. Liu W. et al. Ssd: Single shot multibox detector //Computer Vision–ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, October 11–14, 2016, Proceedings, Part I 14. – Springer International Publishing, 2016. – С. 21-37.

98. Liu Y., Jourabloo A., Liu X. Learning deep models for face anti-spoofing: Binary or auxiliary supervision //Proceedings of the IEEE conference on computer vision and pattern recognition. – 2018. – С. 389-398.

99. Liu Y., Li H., Wang X. Rethinking feature discrimination and polymerization for large-scale recognition //arXiv preprint arXiv:1710.00870. – 2017.

100. Lucena O. et al. Transfer learning using convolutional neural networks for face anti-spoofing //Image Analysis and Recognition: 14th International Conference, ICIAR 2017, Montreal, QC, Canada, July 5–7, 2017, Proceedings 14. – Springer International Publishing, 2017. – С. 27-34.

101. Lutsenko M. et al. Biometric cryptosystems: overview, state-of-the-art and perspective directions //Conference on Mathematical Control Theory. – Cham : Springer International Publishing, 2019. – С. 66-84.

102. Maatta J., Hadid A., Pietikainen M. Face spoofing detection from single images using micro-texture analysis //2011 International Joint Conference on Biometrics (IJCB). – IEEE, 2011. – С. 1-7.

103. Mai G. et al. On the Reconstruction of Face Images from Deep Face Templates //IEEE Transactions on Pattern Analysis and Machine Intelligence. — 2019.

104. Mai G. et al. SecureFace: Face Template Protection //IEEE Transactions on Information Forensics and security. – 2020. – Т. 16. – С. 262-277.

105. Malygin A. et al. Application of artificial neural networks forhandwritten biometric images recognition //Training. – 2017. – Т. 1. – С. 0-1.

106. Manisha, Kumar N. Cancelable biometrics: a comprehensive survey //Artificial Intelligence Review. – 2020. – Т. 53. – №. 5. – С. 3403-3446.

107. Marcel S., Fierrez J., Evans N. (ed.). Handbook of Biometric Anti-Spoofing: Presentation Attack Detection and Vulnerability Assessment. – Berlin, Germany : Springer, 2023. – Т. 1.

108. Marshalko G. B. On the security of a neural network-based biometric authentication scheme //Математические вопросы криптографии. – 2014. – Т. 5. – №. 2. – С. 87–98.

109. Mehdipour Ghazi M., Kemal Ekenel H. A comprehensive analysis of deep learning based representation for face recognition //Proceedings of the IEEE conference on computer vision and pattern recognition workshops. – 2016. – С. 34-41.

110. Mikriukov G. et al. The Anatomy of Adversarial Attacks: Concept-based XAI Dissection //arXiv preprint arXiv:2403.16782. – 2024.

111. Mohamed S., Ghoneim A., Youssif A. Visible/infrared face spoofing detection using texture descriptors //MATEC Web of Conferences. – EDP Sciences, 2019. – T. 292. – C. 04006.

112. Mokhayeri F., Granger E. A paired sparse representation model for robust face recognition from a single sample //Pattern Recognition. – 2020. – T. 100. – C. 107129.

113. Ning X. et al. Real-time 3D face alignment using an encoder-decoder network with an efficient deconvolution layer //IEEE Signal Processing Letters. – 2020. – T. 27. – C. 1944-1948.

114. Pan G. et al. Eyeblink-based anti-spoofing in face recognition from a generic webcamera //2007 IEEE 11th international conference on computer vision. – IEEE, 2007. – C. 1-8.

115. Parkhi O., Vedaldi A., Zisserman A. Deep face recognition //BMVC 2015-Proceedings of the British Machine Vision Conference 2015. – British Machine Vision Association, 2015.

116. Patel K. et al. Live face video vs. spoof face video: Use of moiré patterns to detect replay video attacks //2015 International Conference on Biometrics (ICB). – IEEE, 2015. – C. 98-105.

117. Patel K., Han H., Jain A. K. Secure face unlock: Spoof detection on smartphones //IEEE Transactions on Information Forensics and Security. – 2016. – T. 11. – №. 10. – C. 2268-2283.

118. Peng D. et al. Ts-Fen: Probing feature selection strategy for face anti-spoofing //ICASSP 2020-2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). – IEEE, 2020. – C. 2942-2946.

119. Peng J. et al. A biometric cryptosystem scheme based on random projection and neural network //Soft Computing. – 2021. – T. 25. – C. 7657-7670.

120. Qi C., Su F. Contrastive-center loss for deep neural networks //2017 IEEE international conference on image processing (ICIP). – IEEE, 2017. – C. 2851-2855.

121. Qi X., Zhang L. Face recognition via centralized coordinate learning //arXiv preprint arXiv:1801.05678. – 2018.

122. Qian Y., Deng W., Hu J. Task specific networks for identity and face variation //2018 13th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2018). – IEEE, 2018. – С. 271-277.

123. Qin Y. et al. Learning meta model for zero-and few-shot face anti-spoofing //Proceedings of the AAAI Conference on Artificial Intelligence. – 2020. – Т. 34. – №. 07. – С. 11916-11923.

124. Qin Y. et al. Meta-teacher for face anti-spoofing //IEEE transactions on pattern analysis and machine intelligence. – 2021. – Т. 44. – №. 10. – С. 6311-6326.

125. Rahman M. M. et al. Security risk and attacks in AI: A survey of security and privacy //2023 IEEE 47th Annual Computers, Software, and Applications Conference (COMPSAC). – IEEE, 2023. – С. 1834-1839.

126. Ranjan R., Castillo C. D., Chellappa R. L2-constrained softmax loss for discriminative face verification //arXiv preprint arXiv:1703.09507. – 2017.

127. Rathgeb C. et al. Deep face fuzzy vault: Implementation and performance //Computers & Security. – 2022. – Т. 113. – С. 102539.

128. Redmon J. et al. You only look once: Unified, real-time object detection //Proceedings of the IEEE conference on computer vision and pattern recognition. – 2016. – С. 779-788.

129. Rostami M. et al. Detection and continual learning of novel face presentation attacks //Proceedings of the IEEE/CVF international conference on computer vision. – 2021. – С. 14851-14860.

130. Roth J., Tong Y., Liu X. Adaptive 3D face reconstruction from unconstrained photo collections //Proceedings of the IEEE conference on computer vision and pattern recognition. – 2016. – С. 4197-4206.

131. Roy N. D., Biswas A. Fast and robust retinal biometric key generation using deep neural nets //Multimedia Tools and Applications. – 2020. – Т. 79. – №. 9. – С. 6823-6843.

132. Sankaranarayanan S. et al. Triplet probabilistic embedding for face verification and clustering //2016 IEEE 8th international conference on biometrics theory, applications and systems (BTAS). – IEEE, 2016. – С. 1-8.

133. Sankaranarayanan S., Alavi A., Chellappa R. Triplet similarity embedding for face verification //arXiv preprint arXiv:1602.03418. – 2016.

134. Schroff F., Kalenichenko D., Philbin J. Facenet: A unified embedding for face recognition and clustering //Proceedings of the IEEE conference on computer vision and pattern recognition. – 2015. – С. 815-823.

135. Selvaraju R. R. et al. Grad-cam: Visual explanations from deep networks via gradient-based localization //Proceedings of the IEEE international conference on computer vision. – 2017. – С. 618-626.

136. Serengil S. I., Ozpinar A. Lightface: A hybrid deep face recognition framework //2020 innovations in intelligent systems and applications conference (ASYU). – IEEE, 2020. – С. 1-5.

137. Sikder J. et al. Intelligent face detection and recognition system //2021 International Conference on Intelligent Technologies (CONIT). – IEEE, 2021. – С. 1-5.

138. Simonyan K., Zisserman A. Very deep convolutional networks for large-scale image recognition //arXiv preprint arXiv:1409.1556. – 2014.

139. Smith Z. M., Lostri E. The hidden costs of cybercrime. – McAfee, 2020.

140. Soo S. Object detection using Haar-cascade Classifier //Institute of Computer Science, University of Tartu. – 2014. – Т. 2. – №. 3. – С. 1-12.

141. Stallings W. Cryptography and network security: principles and practice // Pearson. — vol. 7. — 2016.

142. State of the art deep face analysis library [электронный ресурс]. Режим доступа: https://insightface.ai/ , свободный (дата обращения: 21.08.2024).

143. Steane A. M. Simple quantum error-correcting codes //Physical Review A. – 1996. – Т. 54. – №. 6. – С. 4741.

144. Sulavko A. Biometric-Based Key Generation and User Authentication Using Acoustic Characteristics of the Outer Ear and a Network of Correlation Neurons //Sensors. – 2022. – Т. 22. – №. 23. – С. 9551.

145. Sulavko A. et al. Biometric Authentication Using Face Thermal Images Based on Neural Fuzzy Extractor //2023 Intelligent Methods, Systems, and Applications (IMSA). – IEEE, 2023. – С. 80-85.

146. Sun Y. et al. Deep learning face representation by joint identification-verification //Advances in neural information processing systems. – 2014. – T. 27.

147. Sun Y., Wang X., Tang X. Sparsifying neural network connections for face recognition //Proceedings of the IEEE conference on computer vision and pattern recognition. – 2016. – C. 4856-4864.

148. Szegedy C. et al. Going deeper with convolutions //Proceedings of the IEEE conference on computer vision and pattern recognition. – 2015. – C. 1-9.

149. Taigman Y. et al. Deepface: Closing the gap to human-level performance in face verification //Proceedings of the IEEE conference on computer vision and pattern recognition. – 2014. – C. 1701-1708.

150. Talreja V., Valenti M. C., Nasrabadi N. M. Zero-shot deep hashing and neural network based error correction for face template protection //2019 IEEE 10th International Conference on Biometrics Theory, Applications and Systems (BTAS). – IEEE, 2019. – C. 1-10.

151. Tan M., Le Q. V. Mixconv: Mixed depthwise convolutional kernels //arXiv preprint arXiv:1907.09595. – 2019.

152. Tan X. et al. Face liveness detection from a single image with sparse low rank bilinear discriminative model //Computer Vision–ECCV 2010: 11th European Conference on Computer Vision, Heraklion, Crete, Greece, September 5-11, 2010, Proceedings, Part VI 11. – Springer Berlin Heidelberg, 2010. – C. 504-517.

153. Tang Y., Chen L. 3d facial geometric attributes based anti-spoofing approach against mask attacks //2017 12th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2017). – IEEE, 2017. – C. 589-595.

154. Taskiran M., Kahraman N., Erdem C. E. Face recognition: Past, present and future (a review) //Digital Signal Processing. – 2020. – T. 106. – C. 102809.

155. Terhörst P. et al. Beyond identity: What information is stored in biometric face templates? //2020 IEEE international joint conference on biometrics (IJCB). – IEEE, 2020. – C. 1-10.

156. Tolpegin V. et al. Data poisoning attacks against federated learning systems //Computer security–ESORICs 2020: 25th European symposium on research in

computer security, ESORICs 2020, guildford, UK, September 14–18, 2020, proceedings, part i 25. – Springer International Publishing, 2020. – С. 480-501.

157. Tran L., Yin X., Liu X. Disentangled representation learning gan for pose-invariant face recognition //Proceedings of the IEEE conference on computer vision and pattern recognition. – 2017. – С. 1415-1424.

158. Vulfin A. et al. Neural network biometric cryptography system // Proceedings of the Information Technologies and Intelligent Decision Making Systems (ITIDMS2021). CEUR. – 2021. – Т. 2843.

159. Wang F. et al. Additive margin softmax for face verification //IEEE Signal Processing Letters. – 2018. – Т. 25. – №. 7. – С. 926-930.

160. Wang F. et al. Palmprint false acceptance attack with a generative adversarial network (GAN) //Applied Sciences. – 2020. – Т. 10. – №. 23. – С. 8547.

161. Wang H. et al. Cosface: Large margin cosine loss for deep face recognition //Proceedings of the IEEE conference on computer vision and pattern recognition. – 2018. – С. 5265-5274.

162. Wang S. et al. Backdoor attacks against transfer learning with pre-trained deep learning models //IEEE Transactions on Services Computing. – 2020. – Т. 15. – №. 3. – С. 1526-1539.

163. Wang Y. et al. Face anti-spoofing to 3D masks by combining texture and geometry features //Biometric Recognition: 13th Chinese Conference, CCBR 2018, Urumqi, China, August 11-12, 2018, Proceedings 13. – Springer International Publishing, 2018. – С. 399-408.

164. Weinberger K. Q., Saul L. K. Distance metric learning for large margin nearest neighbor classification //Journal of machine learning research. – 2009. – Т. 10. – №. 2.

165. Wen Y. et al. A discriminative feature learning approach for deep face recognition //Computer Vision–ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, October 11–14, 2016, Proceedings, Part VII 14. – Springer International Publishing, 2016. – С. 499-515.

166. Wilson Silva, Tiago Filipe Sousa Gonçalves, Ana Sequeira, João Ribeiro Pinto. Explainable Artificial Intelligence for Face Presentation Attack Detection // 26th Portuguese Conference in Pattern Recognition (RECPAD). – 2020.

167. Wu B. et al. Shift: A zero flop, zero parameter alternative to spatial convolutions //Proceedings of the IEEE conference on computer vision and pattern recognition. – 2018. – С. 9127-9135.

168. Wu Y. et al. Deep convolutional neural network with independent softmax for large scale face recognition //Proceedings of the 24th ACM international conference on Multimedia. – 2016. – С. 1063-1067.

169. Wu Y. et al. Deep face recognition with center invariant loss //Proceedings of the on Thematic Workshops of ACM Multimedia 2017. – 2017. – С. 408-414.

170. Xu X., Xiong Y., Xia W. On improving temporal consistency for online face liveness detection system //Proceedings of the IEEE/CVF International Conference on Computer Vision. – 2021. – С. 824-833.

171. Yang H., Wang X. A. Cascade classifier for face detection //Journal of Algorithms & Computational Technology. – 2016. – Т. 10. – №. 3. – С. 187-197.

172. Yang J., Lei Z., Li S. Z. Learn convolutional neural network for face anti-spoofing //arXiv preprint arXiv:1408.5601. – 2014.

173. Yang L. et al. AI-Driven Anonymization: Protecting Personal Data Privacy While Leveraging Machine Learning //arXiv preprint arXiv:2402.17191. – 2024.

174. Yang S. et al. Image data augmentation for deep learning: A survey //arXiv preprint arXiv:2204.08610. – 2022.

175. Yang S. et al. Wider face: A face detection benchmark //Proceedings of the IEEE conference on computer vision and pattern recognition. – 2016. – С. 5525-5533.

176. Yao G. et al. Mesh guided one-shot face reenactment using graph convolutional networks //Proceedings of the 28th ACM international conference on multimedia. – 2020. – С. 1773-1781.

177. Yi D. et al. Learning face representation from scratch //arXiv preprint arXiv:1411.7923. – 2014.

178. Yu S. et al. Improving face sketch recognition via adversarial sketch-photo transformation //2019 14th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2019). – IEEE, 2019. – С. 1-8.

179. Yu Z. et al. Deep learning for face anti-spoofing: A survey //IEEE transactions on pattern analysis and machine intelligence. – 2022. – Т. 45. – №. 5. – С. 5609-5631.

180. Yu Z. et al. Face anti-spoofing with human material perception //Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part VII 16. – Springer International Publishing, 2020. – С. 557-575.

181. Yu Z. et al. Searching central difference convolutional networks for face anti-spoofing //Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. – 2020. – С. 5295-5305.

182. Zhang C. et al. A survey on federated learning //Knowledge-Based Systems. – 2021. – Т. 216. – С. 106775.

183. Zhang K. et al. Joint face detection and alignment using multitask cascaded convolutional networks //IEEE signal processing letters. – 2016. – Т. 23. – №. 10. – С. 1499-1503.

184. Zhang P. et al. FeatherNets: Convolutional neural networks as light as feather for face anti-spoofing //Proceedings of the IEEE/CVF conference on computer vision and pattern recognition workshops. – 2019. – С. 0-0.

185. Zhang Q. et al. Vargnet: Variable group convolutional neural network for efficient embedded computing //arXiv preprint arXiv:1907.05653. – 2019.

186. Zhang S. et al. Casia-surf: A large-scale multi-modal benchmark for face anti-spoofing //IEEE Transactions on Biometrics, Behavior, and Identity Science. – 2020. – Т. 2. – №. 2. – С. 182-193.

187. Zhang X. et al. Range loss for deep face recognition with long-tailed training data //Proceedings of the IEEE international conference on computer vision. – 2017. – С. 5409-5418.

188. Zhang X. et al. Shufflenet: An extremely efficient convolutional neural network for mobile devices //Proceedings of the IEEE conference on computer vision and pattern recognition. – 2018. – С. 6848-6856.

189. Zhang Y. et al. Celeba-spoof: Large-scale face anti-spoofing dataset with rich annotations //Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part XII 16. – Springer International Publishing, 2020. – С. 70-85.

190. Zhang Z. et al. A face antispoofing database with diverse attacks //2012 5th IAPR international conference on Biometrics (ICB). – IEEE, 2012. – С. 26-31.

191. Zhao X. et al. Exploiting explanations for model inversion attacks //Proceedings of the IEEE/CVF international conference on computer vision. – 2021. – С. 682-692.

192. Zhao Y., Chen J. A survey on differential privacy for unstructured data content //ACM Computing Surveys (CSUR). – 2022. – Т. 54. – №. 10s. – С. 1-28.

193. Zhou E., Cao Z., Sun J. Gridface: Face rectification via learning local homography transformations //Proceedings of the European conference on computer vision (ECCV). – 2018. – С. 3-19.

194. Zhou X. P., Sun M. Study on accuracy measure of trigonometric leveling //Applied Mechanics and Materials. – 2013. – Т. 329. – С. 373-377.

195. Zhu Z. et al. Deep learning identity-preserving face space //Proceedings of the IEEE international conference on computer vision. – 2013. – С. 113-120.

Приложение А. Акты о внедрении научных результатов

УТВЕРЖДАЮ
Первый проректор –
проректор по научной
работе
ФГБОУ ВО «СамГТУ»
_____ М.В. Ненашев
«13» ____06____ 2024 г.

АКТ
о внедрении в учебный процесс университета
результатов кандидатской диссертации старшего преподавателя
кафедры «Электронные системы и информационная безопасность»
Панфиловой Ирины Евгеньевны

Мы, нижеподписавшиеся, исполняющий обязанности заведующего кафедрой «Электронные системы и информационная безопасность» Института автоматики и информационных технологий, к.т.н., доцент Карпова Н.Е., ученый секретарь совета Института автоматики и информационных технологий, к.п.н., доцент Стельмах Я.Г. составили настоящий акт о том, что полученные старшим преподавателем кафедры «Электронные системы и информационная безопасность» Панфиловой И.Е. результаты кандидатской диссертации внедрены в учебный процесс университета.

Предложенные в диссертации модели, алгоритмы, система защищенной биометрической аутентификации, а также описывающий ее программный комплекс используются на кафедре «Электронные системы и информационная безопасность» для подготовки магистров по направлению 10.04.01 «Информационная безопасность» (профиль «Интеллектуальные средства в системах безопасности») при изучении дисциплин «Интеллектуальные системы и базы данных» и «Информационно-аналитические системы безопасности».

И.о. заведующий кафедрой
«Электронные системы и
информационная безопасность»,
к.т.н., доцент — Н.Е. Карпова

Ученый секретарь совета
Института автоматики и
информационных технологий,
к.п.н., доцент — Я.Г. Стельмах

open{code-}

УТВЕРЖДАЮ
Управляющий ООО «Открытый код»
_____ О.Л. Сурнин
«06» _____ 2024 г.

АКТ
использования результатов диссертационной работы
Панфиловой Ирины Евгеньевны, представленной
на соискание ученой степени кандидата технических наук

Мы, нижеподписавшиеся, директор по управлению проектами – исполнительный директор Ситников Павел Владимирович и заместитель руководителя департамента разработки программных средств Крупин Даниил Николаевич, составили настоящий акт о том, что результаты диссертационной работы Панфиловой Ирины Евгеньевны, представленной на соискание ученой степени кандидата технических наук, включающие:

- концепцию защищенной биометрической аутентификации по лицу, устойчивую к спуфинг атакам;
- модель нейросетевого преобразователя «биометрия-код» (НПБК) на основе тригонометрических нейронов, позволяющую осуществлять аутентификацию пользователя по лицу без компрометации его биометрических образов;
- алгоритм предварительной настройки параметров нейросетевых преобразователей «биометрия-код» и алгоритм автоматического обучения НПБК на основе тригонометрических нейронов, позволяющие быстро обучать НПБК на малых выборках образов лиц;
- систему защищенной биометрической аутентификации и способ ее построения, реализованные в виде программного комплекса на основе разработанной концепции, модели и алгоритмов

прошли апробацию и использованы при разработке «Системы обеспечения безопасности и выявления девиантного поведения по видеоизображению на основе базы знаний и искусственных нейронных сетей». Практическое использование результатов работы позволило:

- повысить уровень надежности биометрической аутентификации на предприятии;
- обеспечить защиту системы биометрической аутентификации предприятия от спуфинг атак;
- обеспечить конфиденциальность биометрических образов сотрудников компании.

Результаты диссертационной работы имеют практическое значение и могут быть внедрены на предприятиях, работающих с биометрической аутентификацией по лицу.

Директор по управлению проектами –
Исполнительный директор, д.т.н. П.В. Ситников

Заместитель руководителя
Департамента разработки ПС Д.Н Крупин

УТВЕРЖДАЮ

Проректор по образовательной деятельности
ФГАОУ ВО «Омский государственный
технический университет», к.т.н, доцент
_____ Н.А. Прокудина
«3» 06. 2024 г.

АКТ
о внедрении в учебный процесс университета
результатов кандидатской диссертации младшего научного сотрудника
научно-исследовательской лаборатории «Информационная безопасность»
при кафедре «Комплексная защита информации»
Панфиловой Ирины Евгеньевны

Мы, нижеподписавшиеся, заведующий кафедрой «Комплексная защита информации» Радиотехнического факультета, д.т.н. профессор, Ложников П.С., председатель учебно-методической комиссии Радиотехнического факультета, доцент, к.т.н. Никонов И.В. составили настоящий акт о том, что полученные младшим научным сотрудником научно-исследовательской лаборатории «Информационная безопасность» Панфиловой И.Е. результаты кандидатской диссертации внедрены в учебный процесс университета.

Предложенные в диссертации модели, методы, алгоритмы, и инструментальный комплекс используются на кафедре «Комплексная защита информации» для подготовки магистров по направлению 09.04.01 «Информатика и вычислительная техника» (направленность «Безопасность и этика искусственного интеллекта») при изучении дисциплин «Машинное обучение в приложениях биометрии» и «Защищенное исполнение искусственного интеллекта».

Заведующий кафедрой
«Комплексная защита информации»,
д.т.н., профессор П.С. Ложников

Председатель учебно-методической
комиссии Радиотехнического факультета,
к.т.н., доцент И.В. Никонов

Общество с ограниченной ответственностью
«АИ ЗИОН» (ООО «АИ ЗИОН»)
Телефон: +7 (953) 394-90-54
E-mail: aiconstructor@mail.ru
ОКПО 75754649, ОГРН 1215500030976
ИНН 5507286620, КПП 550701001

АКТ

использования результатов диссертационной работы
Панфиловой Ирины Евгеньевны, представленной
на соискание ученой степени кандидата технических наук

Комиссия в составе: генерального директора, д.т.н. Сулавко Алексея Евгеньевича, ведущего инженера-исследователя, к.т.н. Самотуги А.Е. составили настоящий акт о том, что результаты диссертационной работы Панфиловой И.Е., представленной на соискание ученой степени кандидата технических наук, интегрированы в корпоративную среду управления жизненным циклом искусственного интеллекта «AIC ModelOps Platform», в том числе:

1. Модель тригонометрического нейрона и основанная на ней модель нейросетевого преобразователя образов в код для классификации образов, отличающиеся применением новой тригонометрической меры оценки расстояния между образами, позволяющие повысить защищенность искусственного интеллекта от состязательных атак, а также знаний искусственного интеллекта от компрометации;
2. Алгоритмы предварительной настройки и автоматического обучения нейросетевых преобразователей образов в код на основе тригонометрических нейронов, отличающихся возможностью быстрого и робастного обучения моделей искусственного интеллекта на малых выборках.

Предложенные Панфиловой И.Е. технические решения позволили повысить безопасность моделей и знаний искусственного интеллекта, а также реализовать функционал по автоматическому обучению нейросетевых моделей и интегрировать его в редактор пайплайнов продукта «AIC ModelOps Platform».

Генеральный директор,
доктор технических наук — Сулавко А.Е.

Ведущий инженер-исследователь
кандидат технических наук — Самотуга А.Е.

ОБЩЕСТВО С ОГРАНИЧЕННОЙ ОТВЕТСТВЕННОСТЬЮ «АИ ЗИОН»

Приложение Б. Свидетельства о регистрации программ для ЭВМ и электронных ресурсов

РОССИЙСКАЯ ФЕДЕРАЦИЯ

СВИДЕТЕЛЬСТВО

о государственной регистрации программы для ЭВМ

№ 2022680686

AIC ModelOps Platform

Правообладатель: ***ОБЩЕСТВО С ОГРАНИЧЕННОЙ ОТВЕТСТВЕННОСТЬЮ "АИ ЗИОН" (RU)***

Авторы: ***Сулавко Алексей Евгеньевич (RU), Стадников Денис Геннадьевич (RU), Чобан Адиль Гаврилович (RU), Самотуга Александр Евгеньевич (RU), Панфилова Ирина Евгеньевна (RU)***

Заявка № **2022668418**
Дата поступления **10 октября 2022 г.**
Дата государственной регистрации
в Реестре программ для ЭВМ ***03 ноября 2022 г.***

Руководитель Федеральной службы по интеллектуальной собственности

ДОКУМЕНТ ПОДПИСАН ЭЛЕКТРОННОЙ ПОДПИСЬЮ
Сертификат 68b80077e14e40f0a94edbd24145d5c7
Владелец **Зубов Юрий Сергеевич**
Действителен с 2.03.2022 по 26.05.2023

Ю.С. Зубов

РОССИЙСКАЯ ФЕДЕРАЦИЯ

СВИДЕТЕЛЬСТВО

о государственной регистрации программы для ЭВМ

№ 2023684211

Реализация моделей и алгоритмов обучения нейросетевых преобразователей биометрия-код на основе тригонометрических нейронов, позволяющих симметризовать классы образов относительно пространства признаков

Правообладатель: ***Федеральное государственное автономное образовательное учреждение высшего образования «Омский государственный технический университет» (RU)***

Авторы: ***Панфилова Ирина Евгеньевна (RU), Сулавко Алексей Евгеньевич (RU), Серикова Анастасия Евгеньевна (RU), Дорогов Юрий (KZ)***

Заявка № **2023682402**

Дата поступления **27 октября 2023 г.**

Дата государственной регистрации в Реестре программ для ЭВМ ***14 ноября 2023 г.***

Руководитель Федеральной службы по интеллектуальной собственности

ДОКУМЕНТ ПОДПИСАН ЭЛЕКТРОННОЙ ПОДПИСЬЮ
Сертификат 429b6a0fe3853164baf96f83b73b4aa7
Владелец **Зубов Юрий Сергеевич**
Действителен с 10.05.2023 по 02.08.2024

Ю.С. Зубов