На правах рукописи

Гринчук Олег Валерьевич

МЕТОДЫ ОПРЕДЕЛЕНИЯ ПОДЛИННОСТИ ИЗОБРАЖЕНИЙ ЛИЦ

05.13.17 — Теоретические основы информатики

АВТОРЕФЕРАТ

диссертации на соискание ученой степени
кандидата технических наук

Москва – 2020

Работа выполнена на кафедре интеллектуальных систем федерального государственного автономного образовательного учреждения высшего образования «Московский физико-технический институт (национальный исследовательский университет)».

Научный руководитель: **Цурков Владимир Иванович**
доктор физико-математических наук,
Федеральный исследовательский центр «Информатика и управление» Российской академии наук, руководитель отдела сложных систем.

Официальные оппоненты: **Гостев Иван Михайлович**
доктор технических наук,
ФГБУН Институт проблем передачи информации им. А.А. Харкевича Российской академии наук, ведущий научный сотрудник.

Трёкин Алексей Николаевич
кандидат технических наук,
Автономная некоммерческая организация высшего образования «Сколковский институт науки и технологий», Космический центр, научный сотрудник.

Ведущая организация: Федеральное государственное бюджетное образовательное учреждение высшего образования «Владимирский государственный университет имени Александра Григорьевича и Николая Григорьевича Столетовых».

Защита состоится «24» декабря 2020 года в 14:00 на заседании диссертационного совета Д 002.073.05 при Федеральном исследовательском центре «Информатика и управление» Российской академии наук по адресу: 119333, г. Москва, ул. Вавилова, д. 40.

С диссертацией можно ознакомиться в библиотеке ФИЦ ИУ РАН и на сайте http://frccsc.ru.
Автореферат разослан « » ноября 2020 года.

Ученый секретарь
диссертационного совета Д 002.073.05,
кандидат технических наук Рейер И.А.

Общая характеристика работы

Актуальность темы. С развитием технологий глубокого обучения и увеличением вычислительных мощностей системы биометрической идентификации получили широкое практическое применение. Один из основных разделов данного направления – биометрия по лицу. Аутентификация по изображению лица постепенно вытесняет физические ключи обеспечения доступа на закрытые объекты, используется вместо пароля для удаленного подтверждения операций в финансовых учреждениях и заменяет пин-код или отпечатки пальца для авторизации в мобильных телефонах. С широким внедрением технологии распознавания лиц возникла задача защиты систем от попыток взлома и подлога чужого биометрического шаблона. Модели распознавания лиц проверяют только схожесть образца с биометрическим шаблоном и не могут оценить, зарегистрированный пользователь проходит контроль доступа или злоумышленник пытается обмануть систему. Самые распространенные попытки подлога осуществляются с помощью физических артефактов с изображением нужного человека, таких как бумажные распечатки фотографий, записи с экрана мобильного устройства или силиконовые маски, повторяющие трехмерную геометрию лица жертвы. Чтобы обеспечить надежную работу систем идентификации, требуются комплементарные подходы, проверяющие подлинность пользователя на изображении. Методы, решающие данную задачу, называются методами определения **живости.**

Одни из первых алгоритмов определения живости были основаны на извлечении текстурно-частотных признаков изображения (Li: 2004; Maatta: 2011) с помощью дискретного преобразования Фурье и локальных бинарных шаблонов, и последующей классификацией методом опорных векторов.

Альтернативное семейство алгоритмов строилось на построении полезных признаков по последовательности изображений лиц. В работе (Kim: 2013) изменяется фокусное расстояние камеры в момент съемки и анализируется отличие размытых и четких областей изображения для поддельных и подлинных объектов. В работе (Bao: 2007) считается карта оптического потока между соседними кадрами, на базе которой строится классификатор оценки подлинности. В ряд работ (Jee: 2006, Sun: 2006) рассматривается моргание как одна из характеристик живости.

Общей проблемой рассматриваемых методов является недостаточная для практического применения точность работы в неизвестных сценариях, что обусловлено ограниченностью доступных для обучения обучающих выборок. В отличие от задачи распознавания лиц, выборки для обучения алгоритмов определения живости могут

быть собраны только вручную в лабораториях с приглашенными участниками и, следовательно, сильно ограничены в количестве и разнообразии.

Однако, высокий спрос индустрии на решения определения живости стимулировал активное развитие исследований в данной области. В последние годы в открытом доступе появляется все больше соответственных обучающих выборок (Zhang: 2019; A. Liu: 2020; Y. Liu: 2020), что позволяет строить более устойчивые и точные алгоритмы.

Цели и задачи диссертационной работы.

В работе были поставлены следующие **цели**:

1. Исследовать кооперативные методы определения живости для стационарных и мобильных устройств с высоким уровнем защиты от всех видов взлома.
2. Разработать практически применимые методы определения живости для систем контроля и управления доступом (СКУД) с защитой от самых распространенных попыток взлома.
3. Предложить алгоритм определения живости для мультимодальных данных.
4. Предложить некооперативный метод определения живости для стационарных и мобильных устройств с защитой от неизвестных попыток взлома.

Для достижения поставленных целей были решены следующие **задачи**:

1. Разработка кооперативного метода определения живости для стационарных и мобильных устройств и анализ его работоспособности для различных типов атак.
2. Разработка удобного в применении кооперативного метода защиты от наиболее распространенных видов взлома.
3. Сбор обучающей и тестовой выборки, построение алгоритмов генерации новых данных для увеличения вариативности обучающей выборки.
4. Создание практически применимого алгоритма определения живости, эксплуатирующего особенности сценария.
5. Исследование полезных признаков модальностей глубины и ИК, разработка метода определения живости для мультимодальных данных.
6. Разработка устойчивого к неизвестным видам атак алгоритма определения живости по видеопоследовательности для стационарных и мобильных устройств.

Научная новизна. В диссертации предложена концепция искусственных модальностей как промежуточного представления данных в полезном признаковом пространстве, на основе которой предложены новые практически применимые алгоритмы определения живости для разных сценариев. Разработан новый кооперативный метод определения живости для мобильных и стационарных устройств, защищающий от известных видов атак. Предложен новый высокопроизводительный метод определения живости для систем контроля и управления доступом. Предложен масштабируемый метод определения живости по видеопоследовательности. Также предложена новая архитектура нейронной сети для решения задачи определения живости по мультимодальным данным, основанная на тесной связи промежуточных признаков разных модальностей.

Научная значимость работы заключается в том, что предложенные методы показывают более высокую точность и производительность для разных сценариев применения в сравнении с существующими методами. В рамках разработанных методов предлагаются новые архитектуры нейронных сетей, процедуры генерации данных и искусственные модальности, которые могут быть полезны не только для решения задачи определения подлинности, но и для других задач компьютерного зрения.

Практическая значимость результатов диссертации заключается в том, что разработанные алгоритмы используются в основе продуктов, которые были внедрены в многие компании по всему миру. Программная реализация части предложенных методов выложена в открытый доступ, что дает возможность другим исследовательским группам использовать наработки в своих целях. Выложенные в открытый доступ алгоритмы показали лучший в мире результат на двух самых больших выборках по определению живости изображений лиц.

Методы исследования. Для большей части предложенных алгоритмов широко применяется аппарат глубокого обучения нейронных сетей. Для предобработки и подготовки данных к алгоритмам определения подлинности использовались модели детектирования лица и его ключевых точек (Zhang: 2017) и извлечения оптического потока (Sun: 2018). Для сбора данных и демонстрации результатов применялись методы разработки клиенто-серверных мобильных приложений.

Основные положения, выносимые на защиту:

1. Кооперативный метод определения живости для работы в кооперативных сценариях для мобильных и стационарных устройств.
2. Метод определения живости по движению головы для мобильного сценария.
3. Метод определения живости для систем контроля и управления доступом, включающий в себя три независимых алгоритма: по одному изображению, по картам границ и по динамическим временным признакам лица.
4. Алгоритм определения живости по мультимодальным данным.
5. Алгоритм определения живости по видеопоследовательности для мобильных и стационарных устройств.

Степень достоверности и апробация работы. Достоверность результатов подтверждена экспериментальной проверкой, в том числе сторонними организациями; публикациями результатов исследования в рецензируемых научных изданиях и конференциях по машинному обучению. Результаты работы докладывались и обсуждались на следующих научных конференциях.

1. “Recognizing Multi-Modal Face Spoofing with Face Recognition Networks”, Международная конференция “Computer Vision and Pattern Recognition Workshops”. – Long Beach, CA. - 2019.
2. “Creating Artificial Modalities to Solve RGB Liveness”, Международная конференция “Computer Vision and Pattern Recognition Workshops”. – Virtual. – 2020.

Публикации по теме диссертации. Основные результаты по теме диссертации изложены в 6 печатных изданиях, рекомендованных ВАК (включенных в международную систему цитирования Scopus).

Личный вклад. Все результаты, выносимые на защиту, получены автором лично при научном руководстве д.ф.-м.н. Цуркова В. И. Разработка алгоритмов предобработки данных и проведение экспериментов, описанных в главах 4 и 5, проводились совместно с Паркиным А.Н. [2], вклад автора был решающим.

Структура и объем работы. Диссертация состоит из оглавления, введения, пяти разделов, заключения и списка литературы. Основной текст занимает 113 страниц, включая 32 иллюстрации и 14 таблиц. Библиография включает 77 наименований.

Содержание работы

Во **введении** обоснована актуальность диссертационной работы, сформулированы цели и методы исследования, поставлены основные задачи, обоснована научная новизна, показаны теоретическая и практическая значимость полученных результатов.

В **первой главе** приводится формальная постановка задачи, вводятся основные термины и определения. Описывается процедура получения оптимального алгоритма по заданной выборке и формулируется понятие искусственной модальности.

Определение 1. Пусть $\mathbb{X}$ – множество изображений, на которых присутствует центрированное лицо человека. *Треком* длины T_N назовем последовательность изображений $T = \{X_j\},\ j = 1, \dots, T_N$ таких, что $X_j \in \mathbb{X}$.

Определение 2. Будем рассматривать *живость трека* T как бинарную переменную $l \in \mathbb{L}$, равную 1, если на записи живой человек, и 0, если перед камерой демонстрируется его "копия" – распечатанная фотография, экран электронного устройства с записью лица этого человека, трехмерная силиконовая маска и т.д.

Определение 3. *Моделью алгоритмов определения живости* назовем параметрическое семейство отображений $\{\mathbf{f}\ (T, \mathbf{w})\}$, где $\mathbf{f}\ (T, \mathbf{w})$ – в общем случае не дифференцируемая по параметрам $\mathbf{w} \in \mathbb{W}$ функция из множества треков в множество меток живости:

$$\mathbf{f}: \mathbb{X}^{\mathrm{N}} \times \mathbb{W} \to \mathbb{L}.$$

Пусть дана обучающая выборка

$$\boldsymbol{\mathcal{D}} = \{(T_i,\ l_i)\}, i = 1, \dots, m, \tag{1}$$

где $T_i = \{X_j\}, j = 1, \dots, N_{T_i},\ \ l_i \in \{0,1\}$ – множество треков и соответствующих им меток живости.

Алгоритм определения живости – отображение из множества треков в множество меток живости, чьи параметры оптимизированы по заданной обучающей выборке.

Метод определения подлинности изображений лиц – получение алгоритма определения живости для произвольной обучающей выборки.

Определение 4. *Оценкой живости* для трека T_i по некоторому алгоритму будем считать предсказание $s_i = \mathbf{f}(T_i, \mathbf{w}) \in [0,1]$, т.е. вероятность принадлежности трека классу 1.

Задача поиска оптимального алгоритма сводится к минимизации эмпирического риска по обучающей выборке $\boldsymbol{\mathcal{D}}$:

$$\frac{1}{m}\sum_1^m \mathcal{L}\left(\mathbf{f}(T_i,\ \mathbf{w}),\ l_i\right) \to \min, \tag{2}$$

где $\mathcal{L}$ – некоторая функция потерь.

В **разделе 1.2** вводится классификация сценариев применения алгоритмов определения живости S, кооперативности поведения пользователя I и модальностей изображений M.

Раздел 1.3 посвящен определению устойчивости алгоритмов определения живости, вводится понятие устойчивости обучения и определение внутренних и внешних условий распределения данных.

В **разделе 1.4** описываются виды фальсификаций $\{\boldsymbol{PA}\}$, вводится классификация атак по уровням сложности.

Раздел 1.5 содержит процедуру получения оптимального алгоритма по заданной выборке.

В начале, если возможно, генерируются дополнительные данные, которые разнообразят обучающую выборку. После чего выборка фиксируется и разбивается на подвыборки для кросс-валидации. Разбиение заданной выборки на обучающую и валидационную по подгруппам эмулирует существующее разделение данных на заданную и контрольную выборки:

$$\boldsymbol{\mathcal{D}} = \boldsymbol{\mathcal{D}}_{\mathbf{train}} \cup \boldsymbol{\mathcal{D}}_{\mathbf{val}} \sim \tag{3}$$

$$p(T|\boldsymbol{\theta}_{\mathbf{11}}, S, I, M)p(T|\{\boldsymbol{PA}\}_{\mathbf{11}}) + p(T|\boldsymbol{\theta}_{\mathbf{12}}, S, I, M)p(T|\{\boldsymbol{PA}\}_{\mathbf{12}})$$

Далее выбирается семейство моделей $\mathbf{F} = \{\mathbf{f}\ (T, \mathbf{w})\}$, среди которых будет выбрана лучшая для заданной выборки. Рассматриваются преимущественно рассматриваются сверточные нейронные сети, хорошо работающие для задач компьютерного зрения.

Сначала определяются модели $\hat{\mathbf{F}}$, удовлетворяющие заданному бюджету времени работы T. Пусть $\tau_{\mathbf{f}}$ – время работы прямого прохода нейронной сети для одного трека $\mathbf{f}\ (T_i, \mathbf{w})$. Тогда множество рассматриваемых моделей:

$$\hat{\mathbf{F}} = \{\mathbf{f}(T, \mathbf{w}) \; \forall \, \mathbf{f} \in \mathbf{F}: \tau_{\mathbf{f}} \leq \mathrm{T}\}.$$

Каждая из рассматриваемых архитектур обучается с несколькими базовыми наборами гиперпараметров $\boldsymbol{\gamma}$, после чего выбирается лучшая по кросс-валидации архитектура:

$$\mathbf{f}^* = \arg\min_{\mathbf{f}\in\hat{\boldsymbol{F}},\boldsymbol{\gamma}} \mathcal{L}_{\mathrm{val}}(l\,, \mathbf{f}(T, \mathbf{w} \mid \boldsymbol{\gamma})). \tag{4}$$

Поочередно меняя гиперпараметры от допустимого минимального до максимального, обучается 20-30 моделей $\mathbf{f}^*$ и потом из них выбирается лучшая по точности на валидационной выборке:

$$\mathbf{w}^* = \arg\min_{\mathbf{w},\boldsymbol{\gamma}} \mathcal{L}_{\mathrm{val}}\big(l\,, \mathbf{f}^*(T, \mathbf{w} \mid \boldsymbol{\gamma})\big). \tag{5}$$

Полученный таким образом алгоритм $\mathbf{f}^*(T, \mathbf{w}^*)$ назовем *выбранным по процедуре*. В дальнейшем, если не указано обратное, алгоритмы определения живости строятся по описанной выше процедуре.

В **разделе 1.6** формулируется понятие меры качества Q и рассматривается зависимость качества от размера обучающей выборки. Согласно (Figureoa 2012; Cho: 2015; Lei: 2019), зависимость качества алгоритма $\mathbf{f}$ от размера обучающей выборки подчиняется экспоненциальному закону:

$$Q(\mathbf{f}, \boldsymbol{\mathcal{D}}_{\mathbf{test}}) = a|\boldsymbol{\mathcal{D}}|^{-b} + c, \tag{6}$$

Определение 5. Параметр c из (6) назовем *пределом потенциала* алгоритма определения живости $\mathbf{f}$ на выборке $\boldsymbol{\mathcal{D}}$ по контрольной выборке $\boldsymbol{\mathcal{D}}_{\mathbf{test}}$.

Предел потенциала показывает максимальное качество, которое можно получить на неограниченной по размеру обучающей выборке. В идеальном случае $c = 0$ при $|\boldsymbol{\mathcal{D}}| \to \infty$, но если задано ограничение сверху на количество параметров нейронной сети либо обучающая выборка покрывает не все множество внутренних условий $\boldsymbol{\theta}$, достичь 0 не всегда возможно (рис. 1.В).

Определение 6. Параметр b из (6) назовем *степенью эффективности* алгоритма определения живости $\mathbf{f}$ на выборке $\boldsymbol{\mathcal{D}}$ по контрольной выборке $\boldsymbol{\mathcal{D}}_{\mathbf{test}}$.

Степень эффективности чаще всего принимает значения из диапазона $[0, 1]$. Чем выше степень эффективности, тем меньше данных нужно алгоритму, чтобы достичь хорошего значения качества (рис 1.А).

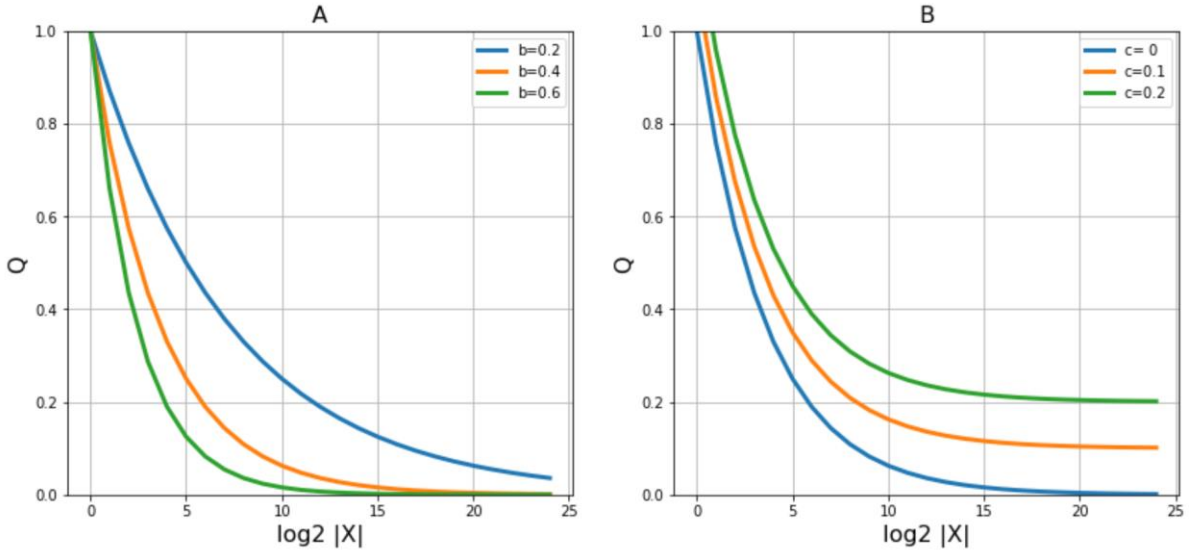

Рис 1. Зависимость меры качества от размера обучающей выборки для разных значений параметров b (слева) и c (справа).

В **разделе 1.7** формулируется понятие искусственной модальности и рассматриваются ее свойства.

Обычное цветное изображение лица X выбирается из множества изображений лиц $\mathbb{X}$, при этом

$$X \in \mathbb{X} \subset \mathbb{Z}_{[0,255]}^{3WH},$$

где $\mathbb{Z}_{[0,255]}^{3\mathrm{WH}}$ – пространство матриц размера $3 \times W \times H$, состоящих из интенсивностей пикселей с значениями от 0 до 255. Аналогично, для трека длины N

$$T \in \mathbb{X}^{\mathrm{N}} \subset \mathbb{Z}_{[0,255]}^{3\mathrm{NWH}},$$

$\mathbb{X}^{\mathrm{N}}$ – богатое пространство, объекты которого содержат множество мелких деталей, в том числе черты лица человека и элементы заднего плана. Поэтому, при обучении алгоритма определения живости на выборке небольшого размера либо собранной при очень ограниченных внутренних условиях $\boldsymbol{\theta}$, возможно переобучение на не имеющие отношения к живости признаки.

Рассмотрим некоторую функцию $\phi: \mathbb{X}^N \rightarrow \mathbb{M}$, где $\mathbb{M}$ – пространство изображений размера $C \times W \times H$, т.е.

$$\phi(T) \in \mathbb{M} \subset \mathbb{R}^{CWH} \tag{7}$$

Пусть заданы обучающая и контрольная выборки обычных цветных изображений $\boldsymbol{\mathcal{D}}$ и $\boldsymbol{\mathcal{D}}_{\mathbf{test}}$, а также процедура выбора алгоритма $\mathbf{f}^*(\boldsymbol{\mathcal{D}})$. Переведем выборки в пространство $\mathbb{M}$, т.е.

$$\widetilde{\boldsymbol{\mathcal{D}}} = \{(\phi(T_i), l_i)\ \forall\ (T_i, l_i) \in\ \boldsymbol{\mathcal{D}}\}$$

$$\widetilde{\boldsymbol{\mathcal{D}}}_{\boldsymbol{test}} = \{(\phi(T_i), l_i)\ \forall\ \ (T_i, l_i) \in \boldsymbol{\mathcal{D}}_{\mathbf{test}}\}$$

Построим алгоритмы $\mathbf{f}^*(\boldsymbol{\mathcal{D}})$ и $\mathbf{f}^*(\widetilde{\boldsymbol{\mathcal{D}}})$ по процедуре и посчитаем степени эффективности алгоритмов $\tilde{b}$ и b по контрольным выборкам $\boldsymbol{\mathcal{D}}_{\mathbf{test}}$ и $\widetilde{\boldsymbol{\mathcal{D}}}_{\mathbf{test}}$ соответственно.

Определение 7. Пространство $\mathbb{M}$ назовем *искусственной модальностью,* а функцию ϕ – функцией *преобразования модальности*, если $\tilde{b} > b$.

Если размер выборки небольшой, то алгоритм, построенный на изображениях искусственной модальности, будет работать лучше на контрольной выборке, чем алгоритм, обученный на оригинальных изображениях. Для больших и разнообразных выборок предел потенциала алгоритмов на обычных изображениях выше, чем на изображениях из искусственной модальности, но в прикладных задачах собрать выборку такого размера чаще всего не представляется возможным.

Вторая глава посвящена кооперативным методам определения живости для мобильных и стационарных сценариев, в условиях отсутствия или малого количества обучающих данных.

В **разделе 2.1** предлагается атомарный алгоритм, который состоит из комбинации простых кооперативных проверок.

Назовем *алгоритмом атрибута* некоторый алгоритм $\mathcal{K}$ компьютерного зрения, который по заданному кадру X_i возвращает некоторое действительное число, вектор или метку класса k_i:

$$\mathcal{K}(X_i) = k_i \tag{8}$$

Назовем *атомом* $\mathcal{A}$ алгоритм определения живости , который по последовательности $\{k_i\}$ и фиксированным гиперпараметрам $\boldsymbol{\gamma}$ определяет бинарный ответ живости l, при этом

$$\mathcal{A}(\{\mathcal{K}(X_i)\}, \boldsymbol{\gamma}) = l, \tag{9}$$

Функции агрегации для последовательности $\{k_i\}, i = 1, \dots, n$:

1. $\boldsymbol{\psi}_{\mathbf{max}}(\{k_i\}) = \max(\{k_i\})$
2. $\boldsymbol{\psi}_{\mathbf{avg}}(\{k_i\}) = \frac{1}{n}\sum_i^n k_i$
3. $\boldsymbol{\psi}^{>}_{\mathbf{max/avg}}(\{k_i\}, x) = [\boldsymbol{\psi}_{\mathbf{max/avg}}(\{k_i\}) > x]$
4. $\boldsymbol{\psi}^{\leq}_{\mathbf{max/avg}}(\{k_i\}, x) = [\boldsymbol{\psi}_{\mathbf{max/avg}}(\{k_i\}) \leq x]$

где [] – оператор, равный 1, если условие в скобках выполняется и 0 в противном случае. Для оценки бинарного действия разделим исходную последовательность на две части:

$$K_1 = \{k_i\}, i = 1, \dots, \lceil dN \rceil,$$
$$K_2 = \{k_i\}, i = \lceil dN \rceil + 1, \dots, N,$$

где $\lceil\ \rceil$ – целая часть числа, d – гиперпараметр, показывающий, какую долю трека отнести к первой части. Вводятся следующие виды атомов:

1. По улыбке
$$\mathcal{A}(\{k_i\}) = \boldsymbol{\psi}^{\leq}_{\mathbf{avg}}(K_1, t) * \boldsymbol{\psi}^{>}_{\mathbf{avg}}(K_2, t)$$
2. По открытому рту
$$\mathcal{A}(\{k_i\}) = \boldsymbol{\psi}^{\leq}_{\mathbf{avg}}(K_1, t) * \boldsymbol{\psi}^{>}_{\mathbf{avg}}(K_2, t)$$
3. По поднятым бровям
$$\mathcal{A}(\{k_i\}) = \boldsymbol{\psi}^{>}_{\mathbf{max}}(K_2 - \boldsymbol{\psi}_{\mathbf{avg}}(K_1), t)$$

4–7. По повороту головы
$$\mathcal{A}_{\text{влево}}(\{k_i\}) = \boldsymbol{\psi}^{>}_{\mathbf{max}}(K_2^{yaw} - \boldsymbol{\psi}_{\mathbf{avg}}(K_1^{yaw}), t)$$
$$\mathcal{A}_{\text{вправо}}(\{k_i\}) = \boldsymbol{\psi}^{>}_{\mathbf{max}}(\boldsymbol{\psi}_{\mathbf{avg}}(K_1^{yaw}) - K_2^{yaw}, t)$$
$$\mathcal{A}_{\text{вниз}}(\{k_i\}) = \boldsymbol{\psi}^{>}_{\mathbf{max}}(K_2^{pitch} - \boldsymbol{\psi}_{\mathbf{avg}}(K_1^{pitch}), t)$$
$$\mathcal{A}_{\text{вверх}}(\{k_i\}) = \boldsymbol{\psi}^{>}_{\mathbf{max}}(\boldsymbol{\psi}_{\mathbf{avg}}(K_1^{pitch}) - K_2^{pitch}, t)$$

8. По морганию
$$\forall d = 1, \dots, N - 2n_o - n_c\colon \mathrm{K}_1, \mathrm{K}_2, K_3 = \{k_i\}, \{k_j\}, \{k_v\},$$
$$i = d, \dots, d + n_o, \qquad j = d + n_o + 1, d + n_o + n_c,$$
$$v = d + n_o + n_c + 1, \dots, d + 2n_o + n_c$$
$$\mathcal{A}(\{k_i\}) = \boldsymbol{\psi}^{>}_{\mathbf{avg}}(K_1, 1 - t) * \boldsymbol{\psi}^{\leq}_{\mathbf{avg}}(K_2, t) * \boldsymbol{\psi}^{>}_{\mathbf{avg}}(K_3, 1 - t)$$

где t, n_o, n_c – гиперпараметры алгоритмов.

Схема предлагаемого мультиатомарного алгоритма показана на рис 2. Агрегация отдельных атомов в связанную последовательность и расставление компонент в

произвольном порядке существенно улучшают защиту от различных видов атак. Так, отдельные атомы не обеспечивают защиту от динамических фальсификаций, но случайная последовательность, подкрепленная требованием непрерывности, такую защиту обеспечить может.

Вход: набор атомов $\{\mathcal{A}_i\}$, $i = 1, \dots, n$; количество проверок m.
Выход: l
1: **выбрать** $i_1, i_2, \dots, i_m$ случайных чисел из $1, \dots, n$.
2: **запросить у пользователя** треки по действиям $\left\{\mathcal{A}_{i_j}\right\}, j = 1, \dots, m : \{T_{i_j}\}$
3: $a_{i_j} = \mathcal{A}_{i_j}(\mathcal{K}_{i_j}\left(T_{i_j}\right))$
4: $l = \prod_{j=1}^{m} a_{i_j}$

Рис 2. Мультиатомарный алгоритм определения живости.

В **разделах 2.2, 2.3** рассматривается кооперативный алгоритм определения живости, требующий меньшей вовлеченности пользователя по сравнению с мультиатомарным алгоритмом. Вводится оптический поток как искусственная модальность, по определению выделяющая полезные признаки для оценки подлинности (рис 3).

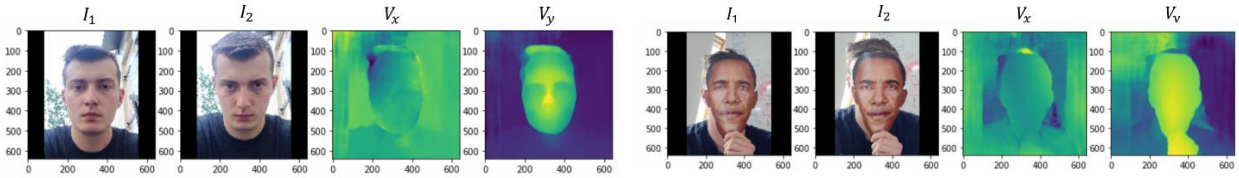

Рис 3. Пример карт оптического потока для реального и поддельного треков.

Для сбора данных реализован программный комплекс в виде приложения на телефон, собирающий данные пользователей. На собранных данных проведена процедура получения алгоритма определения живости, описанная в первой главе.

Алгоритм определения живости, обученный на картах оптического потока, работает значительно лучше алгоритма по обычным изображениям (рис 4).

В **разделе 2.4** приведены заключения к главе.

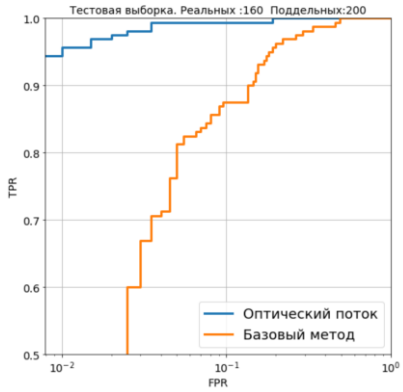

Рис 4. Сравнение меры качества алгоритмов, обученных на модальности оптического потока и на исходных изображениях.

В **третьей главе** предлагается метод определения подлинности для систем контроля и управления доступом (СКУД). Собирается обучающая и тестовые выборки, а также описываются методы синтеза новых данных, эффективно увеличивающие размер и вариативность выборки для обучения. Предлагаются три алгоритма определения живости в описанном сценарии. Все алгоритмы оптимизированы под скорость работы для возможности внедрения в промышленные объекты. Первый алгоритм базируется на идее генерации синтетических данных, эмулирующих атаку масками. Второй алгоритм основывается на идее различия границ на подлинных и поддельных изображениях. Метод устойчив к полноразмерным и экранным видам атак, но пасует перед вырезанными масками. Метод хорошо работает на этом типе атак, но неустойчив к полноразмерным артефактам. Третий метод эксплуатирует идею динамического изменения лицевой мимики и углов наклона головы при подходе человека к турникету. Ансамбль из трех алгоритмов показывает высокую точность на целевой выборке.

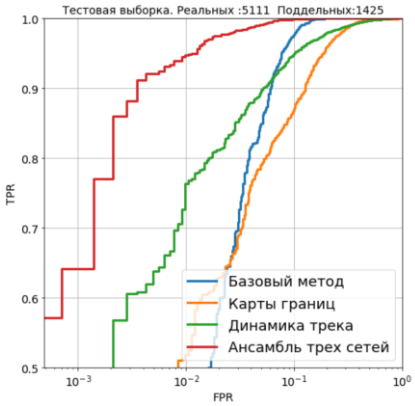

Рис. 5. ROC-кривая предложенных алгоритмов на тестовой выборке СКУД.

В **разделе 3.1** описывается сбор обучающей выборки. В **разделе 3.2** строится алгоритм определения живости по обычным изображениям, предлагается быстрая архитектура нейронной сети SimpleNet, показанная на рис. 6.

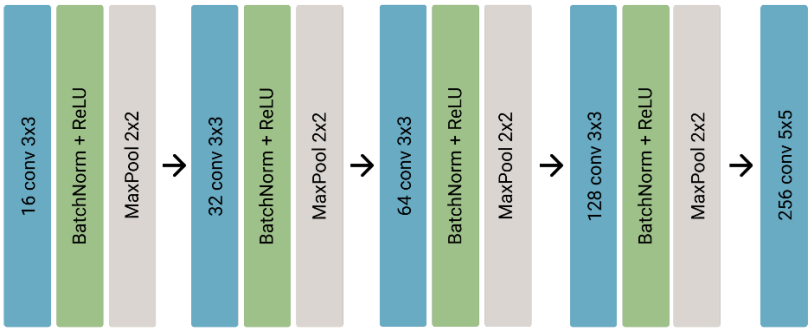

Рис. 6. Архитектура SimpleNet.

Проверяется, что качество алгоритма зависит от размера обучающей выборки (рис. 7).

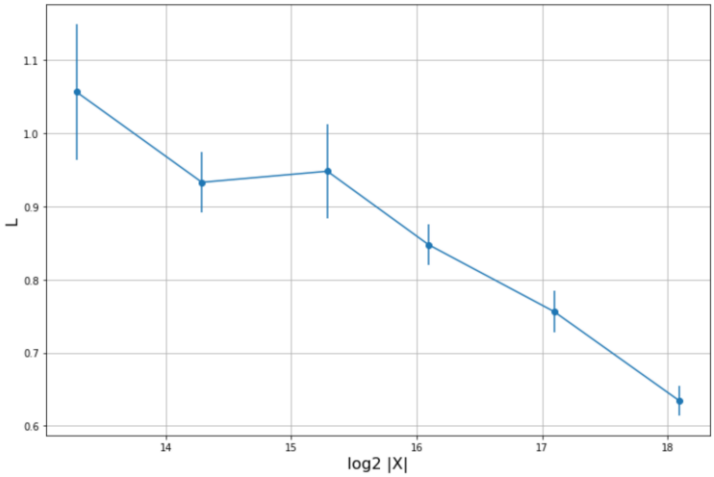

Рис. 7. Зависимость значения функции потерь на контрольной выборке от размера обучающей выборки.

Далее, в **разделе 3.3**, предлагается использование искусственной модальности границ, так как согласно рис. 7, качество нейронной сети по исходным изображениям не выходит на плато при полном размере выборки. Обучение алгоритма по картам границ аналогично обучению по обычному изображению. Семейство моделей

$\mathbf{f}_X(\phi(X_i), \mathbf{w})$, где ϕ – функция преобразования искусственной модальности, выбирается исходя из ограничений по времени работы – рассматривается SimpleNet,. Результат на треках считается, как усреднение результатов по кадрам:

$$\mathbf{f}(T, \mathbf{w}) = \boldsymbol{\psi}_{\mathbf{avg}}(\{\mathbf{f}_X(\phi(X_i), \mathbf{w})\}).$$

В **разделе 3.4** предлагается альтернативный подход, оценивающий весь трек целиком. Предлагается архитектура агрегации динамических признаков (рис. 8). Идея алгоритма состоит в том, что базовая сеть SimpleNet учит дескриптор, совпадающий для лиц с одинаковой мимикой/поворотом головы и различный для лиц с изменением мимики. Это свойство потом ловится сверточными слоями, которые смотрят на дескрипторы всех кадров одновременно и в конце обрабатывается полносвязным для перевода в одно число.

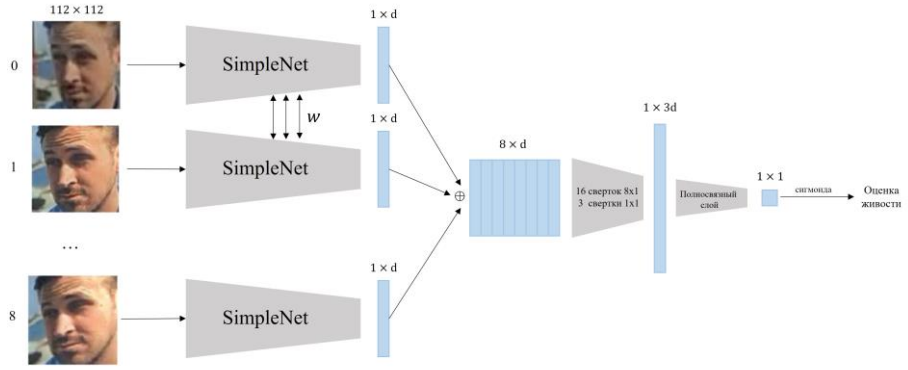

Рис. 8. Архитектура нейронной сети определения живости по динамике трека. $\oplus$ – оператор конкатенации.

В **разделе 3.5** содержатся выводы к главе и демонстрируется итоговую точность ансамбля алгоритмов на целевой выборке (рис. 5).

В **четвертой главе** предлагается алгоритм, работающий с мультимодальными изображениями (RGB, ИК, Глубина). Представляется универсальное улучшение мультимодальных архитектур нейронных сетей (рис. 6), позволяющее лучше агрегировать признаки на всех уровнях детализации. Эксперименты показали, что новые модальности добавляют полезные признаки и улучшают точность на целевой метрике.

Предложенное решение заняло первое место на самом крупном на момент разработки алгоритма мультимодальном датасете CASIA-SURF (**раздел 4.2**) в 2019 году.

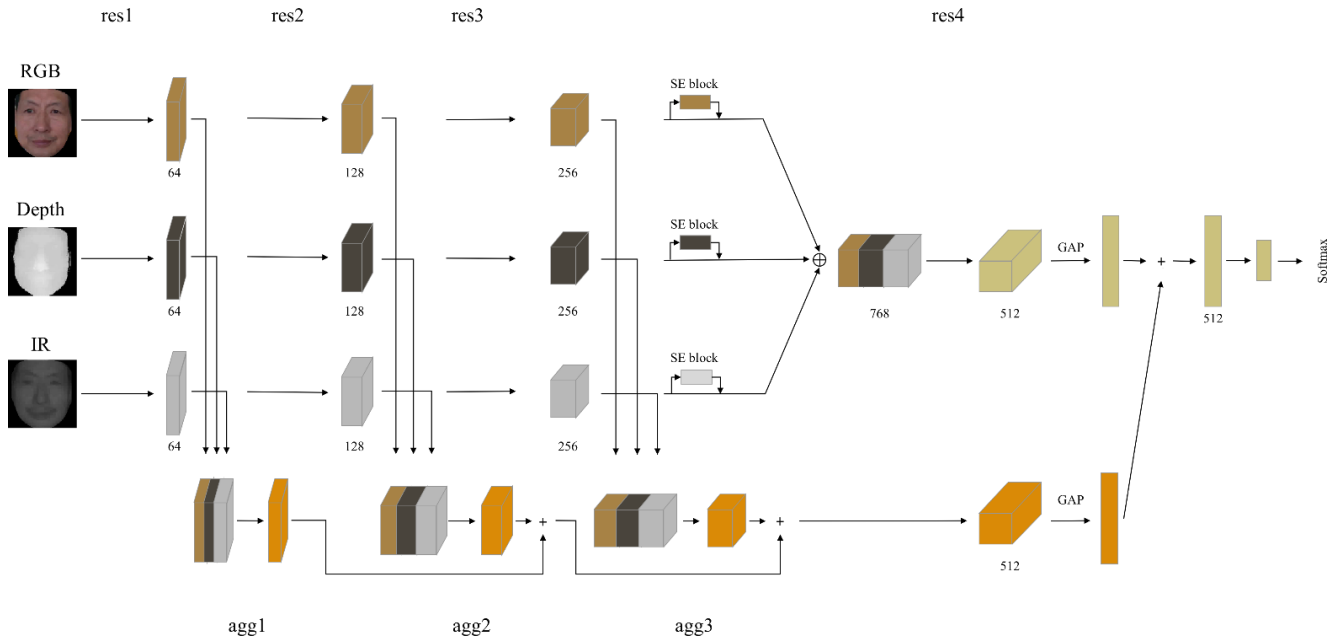

Рис. 9. Предлагаемая архитектура. GAP - общий слой усреднения; ⊕ - оператор объединения; + - оператор почленного суммирования.

Предлагаются три направления работы: данные, архитектура нейронной сети и инициализация весов (**разделы 4.3**, **4.4**). Комплексный подход выявил существенные улучшения точности по сравнению с базовым методом. Тщательный выбор обучающей подвыборки по типам атак позволяет модели лучше противостоять незнакомым попыткам взлома. Предлагается новая архитектура сети с модулем мультиуровневой агрегации признаков, что улучшает обмен полезными признаками между подсетями разных модальностей как на поверхностных, так и на глубоких слоях модели. Используется метод переноса признаков с обученных моделей распознавания лиц, что улучшило стабильность модели и увеличило точность на целевой выборке.

Результаты экспериментов и показатели точности на контрольной выборке показаны в **разделе 4.5**.

Когда есть возможность установить в систему вместо обычных камер специализированные, например, с сенсорами ИК и глубины, сделать надежный алгоритм определения живости становится в разы проще, так как дополнительные модальности обеспечивают модель очень информативными признаками Карта глубины позволяет

показывает трехмерную структуру демонстрируемого объекта, тем самым значительно упрощая отсечение двумерных артефактов, а инфракрасный диапазон помогает с трехмерными масками, так как изображение глаз у живых людей в ИК отличается от статических изображений глаз.

В **разделе 4.6** анализируется влияние каждой из модальностей на итоговую меру каества. Табл. 1 показывает, что наибольший вклад вносит глубина, но лучшее решение достигается при использовании всех трех модальностей.

В **разделе 4.7** приводятся выводы к главе.

Таблица 1. Влияние дополнительных модальностей на целевую метрику.

Модальность	TPR в точке FPR =		
	$\mathbf{10^{-2}}$	$\mathbf{10^{-3}}$	$\mathbf{10^{-4}}$
RGB	71.74	22.34	7.85
ИК	91.82	72.25	57.41
Глубина	100.00	99.77	98.40
RGB+ИК+Глубина	100.00	100.00	99.87

В **пятой главе** предлагается метод решения задачи определения живости для видеопоследовательности на заданной обучающей выборке. Показывается, что аккуратный выбор искусственных модальностей, как ранк-пулинг или оптический поток, уменьшают риск переобучения и повышают итоговую точность модели по сравнению с наивным использованием исходных изображений.

Также предлагается быстрая и масштабируемая архитектура нейронной сети (рис 7), применимая в прикладных задачах. Наконец, показывается простой трюк по обогащению поддельных данных, что всегда является узким местом для подавляющего большинства задач определения живости. В результате, предложенное решение заняло первое место в соревновании Chalearn Singlemodal Face Anti-spoofing Attack Detection на конференции CVPR2020 по выборке CASIA-SURF CeFa (**раздел 5.1**).

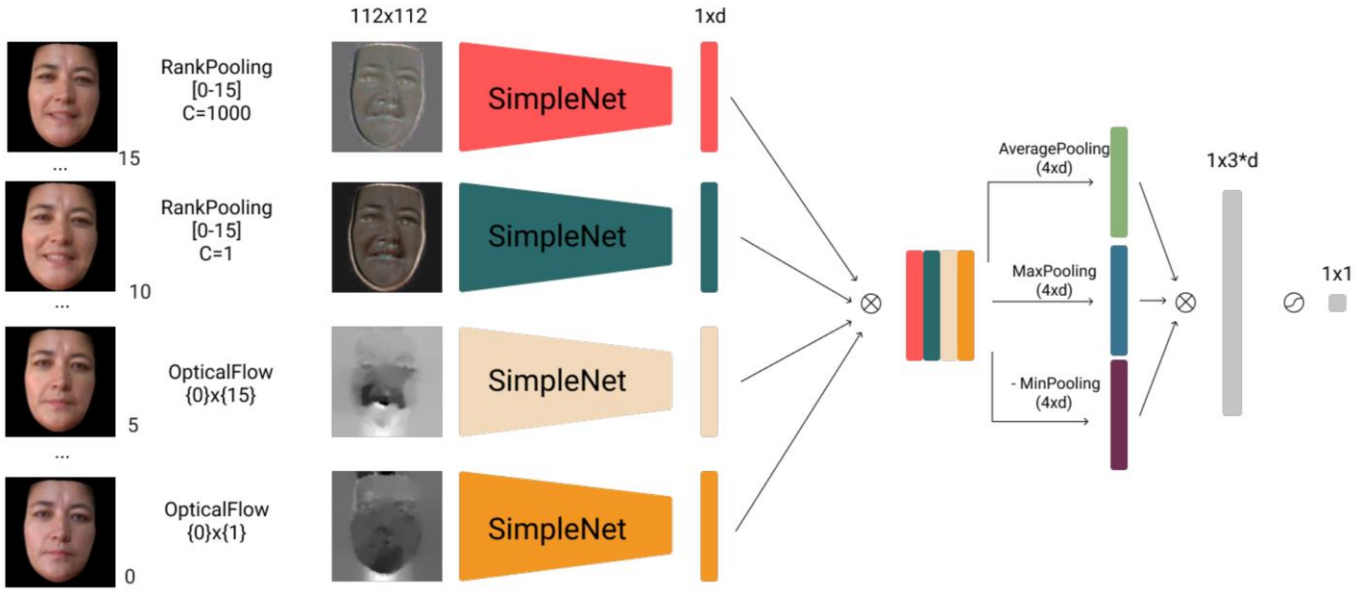

Рис. 10. Предлагаемая архитектура для обработки видеопоследовательности. 16 равномерно выбранных изображений из трека используются для получения четырех модальностей: 2 ранк-пулинга и 2 оптических потока. Модальности обрабатываются различными SimpleNet, после чего агрегируются полносвязным слоем.

В **разделе 5.2** описывается предлагаемый алгоритм. Модальность ранк-пулинга кодирует видеопоследовательность в вектор признаков с помощью процесса оптимизации, который может быть сформулирован как метод опорных признаков для задачи регрессии SVR (Fernando: 2017). После решения оптимизационной задачи вектор признаков можно визуализировать, получая динамическое изображение, которое отображает временную эволюцию покадровых признаков. В данной задаче выбираются гиперпараметры $C = 1$ и $C = 1000$, т.е. низкий и высокий уровень регуляризации для SVR и получаются два визуально различных представления (рис. 11). $C = 1$ сохраняет больше информации об объекте, в то время как $C = 1000$ показывает изменение черт лица со временем.

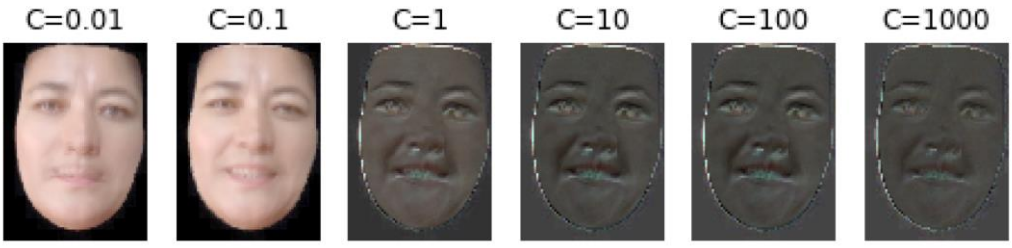

Рис. 11. Ранг-пулинг для разных значений параметра регуляризации C.

Помимо предоставленных модальностей, для увеличения вариативности выборки используется *аугментация последовательности* – преобразование реальных последовательностей в синтетические. Для этого в процессе обучения выбирается один кадр из реального трека и дублируется нужное число раз, после все кадры новой последовательности индивидуально аугментируются поворотами, сдвигом и цветовой коррекцией. Новое семейство поддельных треков больше похоже на распечатанные атаки, присутствующие в тестовой выборке.

В **разделе 5.3** предлагается новая архитектура, показанная на рис. 10. Используемые базовые нейросети SimpleNet достаточно глубокие для извлечения полезных признаков из изображений модальностей, но достаточно узкие, чтобы избежать переобучения. Каждый из четырех полученных тензоров обрабатывается отдельной сетью SimpleNet, которые возвращают дескрипторы размера $1 \times d$. Дескрипторы конкатенируются, после чего к полученной $4 \times d$ матрице применяются операторы Max, Min и Avg пулинга, получая $3 \times d$ матрицу. Обработка завершается полносвязным слоем с сигмоидой.

Таблица 2. Результаты на тестовой выборке CASIA-SURF CeFa.

Метод	APCER, %	BPCER, %	ACER, %
Базовый	23.83 ± 1.70	25.20 ± 22.00	23.42 ± 12.14
Ранк-пулинг(C=1000)	14.11 ± 13.52	11.25 ± 12.75	12.68 ± 4.39
+аугментация последовательности	0.68 ± 0.21	13.91 ± 10.03	7.30 ± 5.00
+Ранк-пулинг(C=1)	1.07 ± 0.53	13.00 ± 10.75	7.03 ± 5.20
+Оптический поток	0.11 ± 0.11	5.33 ± 2.37	2.72 ± 1.21

Раздел 5.4 посвящен экспериментам. Изменение точности на контрольной выборке от введения искусственных модальностей и аугментации последовательности, и итоговый результат показаны в табл. 2. В р**азделе 5.5** описаны выводы к главе.

В **заключении** описаны основные результаты диссертационной работы.

1. Предложены кооперативные методы определения подлинности, основанные на интерактивном взаимодействии с пользователем. Представлен атомарный алгоритм, построенный без обучающей выборки, но неудобный для пользователя. Показаны улучшенные алгоритмы, требующие меньшего уровня кооперативности, основанные на оптическом потоке.
2. Рассмотрена задача определения подлинности для систем контроля и управления доступом. Предложены три алгоритма определения живости в описанном сценарии. Все алгоритмы оптимизированы под скорость работы для возможности внедрения в промышленные объекты.
3. Собраны и обработаны тестовые и обучающие данные для сценария СКУД. Описаны методы синтеза новых данных, эффективно увеличивающие размер и вариативность выборки для обучения.
4. Предложен алгоритм по различию контуров подлинных и поддельных изображениях. Исследована идея временной зависимости изображений лиц на треке, предложен алгоритм, эксплуатирующий различия в динамическом поведении реальных и поддельных примеров.
5. Разработан алгоритм, работающий с мультимодальными изображениями. Предложено универсальное улучшение мультимодальных архитектур нейронных сетей, позволяющее лучше агрегировать признаки на всех уровнях детализации. Получен лучший результат по итогам открытого конкурса на целевой выборке.
6. Предложен алгоритм определения подлинности для мобильных и настольных устройств по видеопоследовательности. Предложена масштабируемая архитектура на основе использования искусственных модальностей. Алгоритм продемонстрировал лучший результат по итогам открытого конкурса на целевой выборке.

Публикации соискателя по теме диссертации

В изданиях, рекомендованных ВАК:

1. *O. Grinchuk, V.I. Tsurkov.* Training a Multimodal Neural Network to Determine the Authenticity of Images. // Journal of Computer and Systems Sciences International. - 2020. – V. 59 (4). – P. 575-582.
2. *A. Parkin, O. Grinchuk*. Recognizing Multi-Modal Face Spoofing With Face Recognition Networks. // Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition Workshops. – 2019.
3. *O. Grinchuk, V.I. Tsurkov, L.P. Wang*. Neural Network Training System for Marker Encoding. // Journal of Computer and Systems Sciences International. – 2019. – V. 58 (3). – P.434-440.
4. *A. Gladkov, O. Grinchuk, Y. Pigareva, I. Mukhina, V. Kazantsev, A. Pimashkin.* Theta rhythm-like bidirectional cycling dynamics of living neuronal networks in vitro. // PloS one. – 2018. – V. 13 (2).
5. *O. Grinchuk, V.I. Tsurkov*. Cyclic Generative Neural Networks for Improved Face Recognition in Nonstardard Domains. // Journal of Computer and Systems Sciences International. – 2018. – V. 57 (4). – P.620-625.
6. *O. Grinchuk, V. Lebedev, V. Lempitsky*. Learnable Visual Markers. // Advances in Neural Information Processing Systems – 2016. – P. 4143-4151.