

На правах рукописи

Свитов Давид Вячеславович

**Оптимизация производительности свёрточных
нейронных сетей в системе распознавания лиц**

Специальность 1.2.2 —
«Математическое моделирование, численные методы и комплексы
программ»

Автореферат
диссертации на соискание учёной степени
кандидата технических наук

Новосибирск — 2023

Работа выполнена в федеральном государственном бюджетном учреждении науки Институте автоматики и электрометрии сибирского отделения Российской академии наук.

Научный руководитель: доктор технических наук, ведущий научный сотрудник ИАиЭ СО РАН

Нежевенко Евгений Семёнович

Официальные оппоненты: **Спектор Александр Аншелевич**,
доктор технических наук, профессор,
Федеральное государственное бюджетное образовательное учреждение высшего образования Новосибирский государственный технический университет,

Андриянов Никита Андреевич,

кандидат технических наук,
доцент Департамента анализа данных и машинного обучения Финансового университета при Правительстве Российской Федерации,

Ведущая организация: Федеральное государственное автономное образовательное учреждение высшего образования «Новосибирский национальный исследовательский государственный университет»

Защита состоится 31 мая 2023 г. на заседании диссертационного совета 24.1.028.01 (Д 003.005.02) при ИАиЭ СО РАН по адресу: 630090, Новосибирская область, г. Новосибирск, пр-кт Академика Коптюга, д.1.

Отзывы на автореферат в двух экземплярах, заверенные печатью учреждения, просьба направлять по адресу: 630090, Новосибирская область, г. Новосибирск, пр-кт Академика Коптюга, д.1, ученому секретарю диссертационного совета 24.1.028.01 (Д 003.005.02).

Телефон для справок: (383) 330-79-69.

Ученый секретарь
диссертационного совета
24.1.028.01 (Д 003.005.02),
академик РАН, профессор

Шалагин Анатолий Михайлович

Общая характеристика работы

Актуальность темы. Нейросетевые модели позволяют решать широкий спектр задач по анализу изображений, что позволяет повсеместно применять их на практике. Развитие технологий обучения нейронных сетей и разработка новых, более глубоких архитектур позволили для большинства прикладных задач автоматизации получить точности, сравнимые с точностью человека-оператора.

Так современные нейросетевые модели позволяют производить идентификацию и верификацию лиц с точностью, достаточной для задач автоматизации. Системы биометрической идентификации по лицу, как правило, состоят из трёх элементов, каждый из которых требует выполнения нейронной сети: обнаружение лица в кадре, выравнивание лица, построение биометрического вектора. Практическая применимость таких систем ограничена требованиями к вычислительным ресурсам. Для обработки изображений такими нейросетевыми моделями требуются вычислительные сервера, оборудованные графическими или тензорными процессорами.

Для повышения практической применимости нейронных сетей необходимо избавиться от ограничений, вызванных требуемыми вычислительными ресурсами. Для решения этой проблемы были предложены мобильные архитектуры нейронных сетей с меньшим числом параметров. Однако точность таких моделей существенно ниже их серверных аналогов, что может быть критично для задач биометрического доступа.

Целью диссертационной работы является разработка методов, позволяющих сократить разницу в точности между серверными и мобильными нейросетевыми архитектурами для задачи распознавания человека по лицу. Предлагаемые методы позволяют улучшить соотношение скорости и точности для нейронных сетей в системе распознавания лиц.

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

1. Исследовать эффективность современных методов ускорения нейросетевых моделей и их применимость к нейронным сетям, используемым в задаче распознавания лиц;
2. Разработать метод ускорения обработки потокового видео нейросетевым детектором лиц для встраивания детектора в устройства с маломощными мобильными вычислителями;
3. Разработать метод максимизации скорости работы и минимизации объёма занимаемой памяти для нейронных сетей, выполняющих построение биометрического вектора по изображению лица;
4. Исследовать возможность комбинирования предложенных методов оптимизации и их практическую применимость в задаче идентификации на встраиваемых устройствах, таких как умные домофоны.

Научная новизна:

1. Предложен и реализован новый алгоритм дистилляции для моделей, обученных с функцией Софтмакс с отступом, для задачи построения биометрического вектора по изображению лица. Софтмакс с отступом — это модификация функции Софтмакс с добавлением константы для большего разнесения векторов в пространстве. Предложенный алгоритм впервые использует адаптивное вычисление отступов в функции Софтмакс на основе расстояния до центра кластера;
2. Предложен подход, позволяющий эффективнее разносить на гиперсфере биометрические вектора, полученные нейронной сетью с малым числом параметров. Предложенный подход впервые использует модель с большим числом параметров для нахождения центров кластеров векторов для повышения эффективности обучения малой нейронной сети;
3. Разработан и впервые применён метод ранней остановки исполнения нейросетевого детектора объектов на основе значения признаков промежуточных слоёв сети. Продемонстрировано повышение средней скорости обработки кадров за счёт использования предложенного метода;

4. Предложен и реализован устойчивый к шуму алгоритм локализации движения в видеопотоке. Впервые для решения этой задачи применены глубокие признаки предобученного детектора объектов, что не влечёт дополнительных вычислительных затрат.

Теоретическая значимость полученных результатов заключается в предложенном в диссертации новом подходе к передаче знаний от сети-учителя к сети-ученику для задач нейросетевой биоидентификации. Копирование последнего слоя сети-учителя в сеть-ученик позволяет последнему использовать более хорошее пространство векторов при обучении. Предлагаемый метод позволяет расширить пространство возможных каналов передачи знаний между нейронными сетями при дистилляции.

Практическая значимость полученных результатов заключается в реализации разработанного подхода в виде программного комплекса для обучения свёрточных нейронных сетей на ЭВМ. Получаемые таким образом нейросетевые модели могут использоваться для распознавания лиц на маломощных вычислителях. Описанный в данной диссертации подход позволил реализовать алгоритмы обнаружения и идентификации лица в видеопотоке на конечном устройстве пользователя с мобильным ARM процессором. Технология распознавания лиц, основывающаяся на разработанном подходе, внедрена в продуктах компаний:

1. “ООО Новотелеком”. Нейронные сети, обученные с помощью разработанного подхода, применяются в Проекте «Свободные руки»: системе идентификации жильцов по видео с камеры домофона;
2. “ООО Рубетек РУС”. Разработанный подход применялся для обучения нейронной сети для распознавания людей, имеющих доступ в помещение, по камере, установленной на входе. Данная технология применяется в жилых комплексах Группы Компаний ПИК;
3. “ООО Открытая мобильная платформа”. Обученные с использованием разработанного метода нейронные сети применяются для распознавания лиц в операционной системе “Аврора”.

Основные положения, выносимые на защиту:

1. Использование обучаемых весов последнего слоя сети-учителя для инициализации весов сети-ученика в задаче дистилляции СНС для построения биометрических векторов лиц позволяет получать более компактные кластеры биометрических векторов на гиперсфере;
2. Инициализация весов копированием последнего слоя из большей модели вынуждает сеть-ученик генерировать вектора с сохранением пространственных взаимоотношений как в сети-учителе, что позволяет использовать модели разной ёмкости в одной системе и открывает новые возможности для разработки методов дистилляции нейронных сетей;
3. Реализация предложенного подхода инициализации весов последнего слоя нейронной сети продемонстрировала повышение процента верно верифицированных изображений лиц на открытом наборе данных;
4. При использовании Софтмакс с отступом, вычисление отступа на основе угла между биометрическим вектором и центром класса, вычисленных нейронной сетью с большим числом параметров, позволяет увеличивать точность моделей для построения биометрических векторов лиц. Софтмакс с отступом — это модификация функции Софтмакс с добавлением константы для большего разнесения векторов в пространстве;
5. Реализация предложенного подхода дистилляции на основе адаптивного отступа в Софтмакс продемонстрировала повышение процентов верно идентифицированных и верифицированных людей на открытых наборах данных;
6. Использование карт активаций на промежуточных слоях извлекателя признаков в задаче детектирования позволяет производить устойчивое к зашумлённости входного видеопотока обнаружение движения. Применение такой техники обнаружения движения в кадре позволяет с высокой надёжностью отфильтровывать статичные кадры из зашумлённого видеопотока и не требует дополнительных вычислительных ресурсов;

7. Реализация предложенного подхода была экспериментально проверена на открытых наборах данных и продемонстрирована его эффективность в сравнении с подходом на основе попиксельного сравнения кадров;
8. Разработанный комплекс программных средств для обучения СНС через дистилляцию позволяет создавать программные решения на основе нейросетевых технологий для задачи распознавания лиц, работающие на борту встраиваемого устройства или маломощном сервере;

Достоверность полученных результатов обеспечивается корректным проведением большого числа тестов на реальных данных, в том числе независимым институтом стандартов. Для замера точности системы использовались различные объективные метрики, продемонстрировавшие результаты непротиворечивые друг-другу и теоретическим выкладкам.

Апробация работы. Основные положения и результаты диссертационной работы докладывались на конференциях: “ICDLFR 2020: 22nd International Conference on Deep Learning and Face Recognition”, Амстердам, 2020 г.; “ММРО 2021: Математические методы распознавания образов”, Москва 2021 г. Доклад на конференции ICDFLR 2021 был отмечен дипломом за лучший доклад.

Личный вклад. Автор разработал и реализовал подходы к оптимизации производительности свёрточных нейронных сетей в системе распознавания лиц. Разработанные подходы описаны в главах: "Глава 2 Оптимизация детектора лиц" и "Глава 3 Оптимизация вычисления биометрического вектора". Автор провёл полное тестирование разработанной системы как описано в "Глава 4 Разработка программного модуля для встраиваемых систем".

Публикации. Материалы диссертации опубликованы в 5 печатных работах в рецензируемых журналах [1–5]. Работы [1; 2] в индексируемых Scopus журналах, относящихся к квартили Q1. Работы [4; 5] изданы в рецензируемых журналах, индексируемых ВАК РФ и относящихся к квартилям Q3 и Q2.

Содержание работы

Во **введении** обосновывается актуальность исследований, проводимых в рамках диссертационной работы, формулируются цель и задачи, решению которых посвящена работа.

В **первой главе** приводится обзор на основе литературных источников методов оптимизации скорости работы нейросетевых моделей и применимость этих методов к технологии распознавания лиц. Рассматриваются принципы работы и основные характеристики таких методов как прунинг, дистилляция, квантование и нейросетевой архитектурный поиск. Глава состоит из 6 разделов.

Раздел **1.1 Классификация методов оптимизации скорости нейросетевых моделей** начинается с описания разницы в вычислительной мощности серверных и мобильных CPU. Приводится перечисление категорий методов оптимизации производительности нейронных сетей для работы на мобильных CPU. Приводится краткое описание следующих подходов: прунинг, дистилляция, квантование, нейросетевой архитектурный поиск. Приводится сценарий комбинирования данных подходов.

Раздел **1.2 Прунинг** посвящён описанию алгоритмов прунинга весов нейронной сети. В основе прунинга лежит идея об удалении связей между нейронами, вносящих наименьший вклад в итоговый результат. В разделе описывается общий подход к задаче прунинга весов и детально анализируются методы, дающие наилучший результат.

В частности, обсуждается решение задачи прунинга основываясь на "гипотезе о лотерейных билетах". Приводится разбор алгоритма прунинга, основанного на данной гипотезе, его ограничений и недостатков. Производится разбор модификаций данного алгоритма и подходов к устранению его недостатков. После формулирования проблемы коллапса слоя, описывается метод SynFlow, позволяющий избегать данной проблемы.

В конце раздела производится классификация существующих алгоритмов прунинга по их структуре, способу выбора параметров и частоте применения к обучаемой модели.

Раздел **1.3 Дистилляция** посвящён подходу, позволяющему более эффективно обучить модель с малым числом параметров, используя информацию, получаемую от обученной модели с большим числом параметров. Описывается понятие “dark knowledge” (тайное знание), впервые введённое Хинтоном, и методы позволяющие осуществлять его передачу между моделями. В начале раздела содержится описание предложенного Хинтоном метода и его наиболее распространённых модификаций. Рассматриваются техники передачи знаний с промежуточных слоёв сети и от ансамбля сетей.

Затем приводится детальное описание метод дистилляции для нейронных сетей в задаче распознавания по лицу, а именно: триплетная дистилляция, дистилляция по углу и дистилляция на основе отступа.

Раздел **1.4 Квантование** посвящён методу оптимизации производительности и потребляемой памяти нейронной сети за счёт снижения дискретности значений её весов. Раздел описывает задачу квантования и приводит общую формулу для её решения. Далее в разделе производится классификация методов квантования по следующим признакам: симметричное или ассиметричное, статичное или динамичное, поканальное или послойное, во время или после обучения.

В конце раздела обсуждается реализация подходов к квантованию в наиболее популярных библиотеках машинного обучения и применимость данных методов к мобильным архитектурам нейронных сетей.

Раздел **1.5 Нейросетевой архитектурный поиск** посвящён описанию задачи автоматического нейросетевого поиска архитектур сетей для конкретного домена. В данном случае под доменом работы понимаются данные, подаваемые на вход нейронной сети и значения, которые ей необходимо предсказать. В разделе приводится разбор ключевых работ по архитектурному поиску и приводится классификация методов по способу задания пространства поиска и по методу оптимизации сети.

Раздел **1.6 Выводы по первой главе** описывает комбинирование приведённых в главе подходов и анализирует преимущества и недостатки их совместного применения. Делается аргументированный выбор дистилляции в качестве основного исследуемого в диссертации метода.

Во **второй главе** описан, предложенный в диссертации, метод для оптимизации детектора лиц, позволяющий снизить число ложных срабатываний за счёт обработки только движущихся объектов. Предлагаемый метод заключается в модификации уже обученной на детекцию свёрточной нейронной сети и может быть применён для повышения точности некоторой имеющейся системы путём небольших изменений в ней.

Раздел **2.1 Задача детектирования лиц** описывает постановку задачи детектирования лица в кадре. В разделе освещается актуальность решения поставленной задачи, заключающаяся в критичности снижения числа ложных положительных срабатываний детектора для систем распознавания лиц, так как они могут приводить к недетерминированному поведению.

В разделе **2.2 Методы детектирования лиц** освещаются наиболее распространённые модели для детектирования, такие как: Faster RCNN, YOLO (You only look once) и SSD: Single Shot MultiBox Detector. Обсуждаются разновидности и версии данных архитектур. Далее рассматриваются методы отделения движущихся объектов от фона на основе построения модели фона и сравнение с ней новых кадров. В заключение раздела приводится анализ недостатков существующих методов, реализующих объединения детекторов объектов и детекторов движения.

Раздел **2.3 Разработка устойчивого подхода к обнаружению движения** посвящён описанию, предлагаемого в диссертации, метода к устойчивому обнаружению движения целевого объекта в видео потоке.

Входное изображение обрабатывается первыми слоями извлекателя признаков, входящим в состав модели детектирования объектов (Рисунок 0.1). Полученные на этом этапе карты признаков используются для построения модели фона. Карта признаков для текущего кадра сравнивается с моделью фона по косинусной близости. Полученная после этого карта движений используется для проверки необходимости заканчивать выполнение детектора. Карта движений также используется для отфильтровывания ограничивающих рамок, чтобы отбраковать ложные срабатывания, соответствующие статичным объектам. Этот метод позволяет сократить среднее время обработки кадра

видеопотока и снизить число ложных срабатываний. Ключевая идея предлагаемого метода - использование промежуточных карт признаков из нейронной сети детектора для определения регионов движения в кадре.

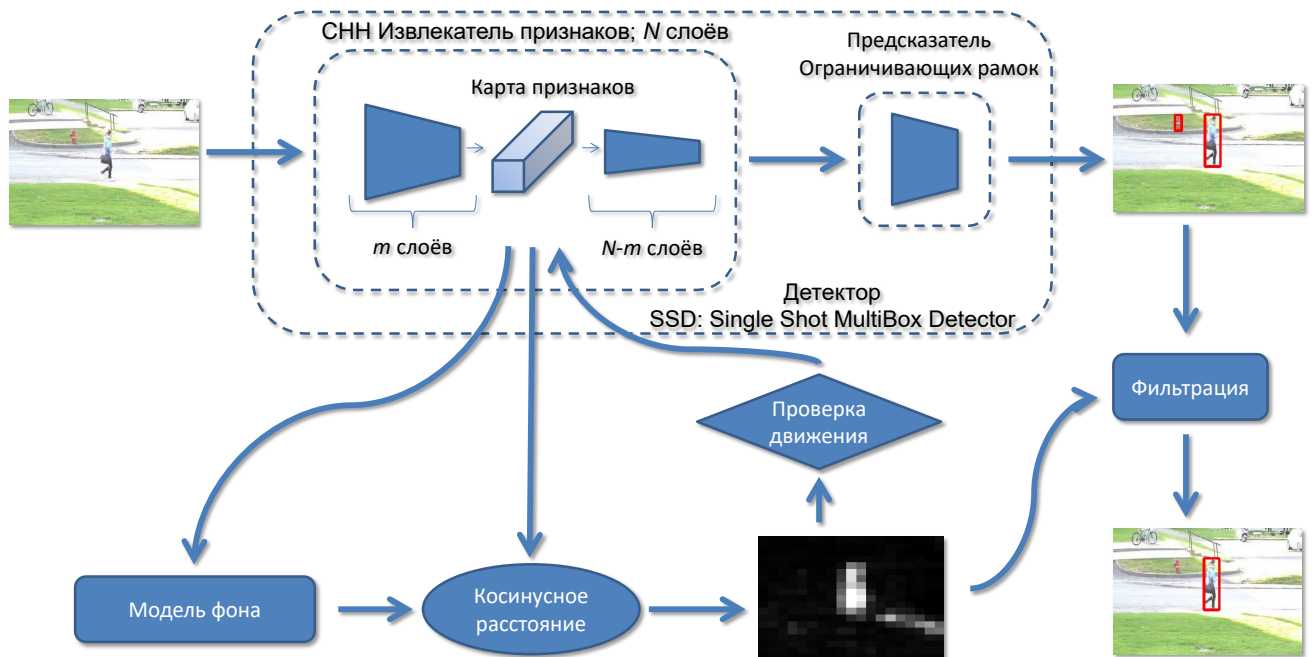


Рис. 0.1 — Схема устойчивого к зашумлённости входных данных метода для обнаружения движения в видеопотоке на основе промежуточных карт признаков СНС.

В разделе **2.4 Экспериментальная апробация предложенного подхода** в свою очередь приводятся результаты тестирования скорости и точности разработанного и реализованного¹ подхода на открытом наборе данных CDNet2014.

В разделе **2.5 Заключение по второй главе** подытоживаются полученные в данной главе результаты.

В **третьей главе** описывается разработка и тестирование метода оптимизации вычисления биометрического вектора. Использование СНС в сочетании с функцией Софтмакс (анг. softmax) с отступами позволяет достичь наибольшей точности в задаче распознавания лиц. В данной диссертации предлагается метод дистилляции для моделей, обученных с функцией Софтмакс с отступами, который позволяет получить большую точность, чем другие методы для задачи

¹<https://github.com/david-svitov/AmphibianDetector>

распознавания лиц на наборах данных LFW, AgeDB-30 и Megaface. Основная идея предлагаемого подхода заключается в использовании центров классов сети-учителя для инициализации сети-ученика. Затем сеть-ученик обучается производить биометрические вектора, углы от которых до центров классов равны углам в сети-учителе (Рисунок 0.2).

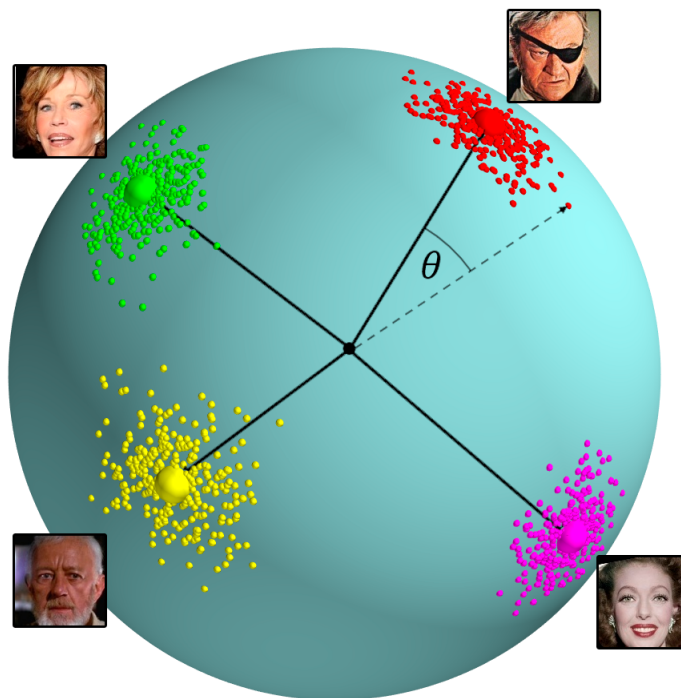


Рис. 0.2 — Гиперсфера нормированных биометрических векторов. Различными цветами обозначены разные классы из обучающего набора данных: большими кругами обозначены центры классов, меньшими - изображения, относящиеся к классу. Для вычисления близости изображения к классу необходимо вычислить угол θ между векторами.

В разделе **3.1 Задача распознавания лиц** описывается постановка и актуальность задачи, принципы работы Софтмакс с отступами для обучения моделей и современные мобильные архитектуры для распознавания лиц. В конце раздела приводится интуитивное описание предлагаемого в диссертации метода дистилляции.

Идея предложенного подхода заключается в копировании последнего слоя сети, содержащего обученные центры классов обучающего набора данных, из сети-учителя в сеть-ученика и заморозке этого слоя во время всей процедуры

дистилляции. Дистилляция заключается в обучении сети-ученика воспроизводить углы между центрами классов и биометрическими векторами для лиц, равные углам между соответствующими векторами и центрами классов в сети-учителе. Такой подход позволяет сети-ученику лучше воспроизводить результат работы сети-учителя.

В следующем разделе **3.2 Существующие подходы к построению биометрических векторов** приводится подробное описание функции Софт-макс с отступами, а также частные её случаи, такие как: Cosface, Sphreface и Arcface. Также в главе приводится подробный обзор существующих методов дистилляции как для общих случаев, так и для задачи распознавания лиц.

Раздел **3.3 Разработка специализированного метода дистилляции для задачи распознавания лиц** описывает предлагаемый в диссертации метод дистилляции для моделей, обученных с функцией Софтмакс с отступом. В начале раздела приводится описание сети-ученика и сети-учителя с обоснованием выбора их архитектур. Приводится описание числа параметров архитектур, времени работы и объёма занимаемой памяти.

Далее приводится формальное описание предлагаемого метода с подробным его изложением в виде формул. В разделе описывается процедура дистилляции на примере модели обученной с Arcface.

Затем в разделе **3.4 Экспериментальная апробация предлагаемого метода** приводятся детали его реализации², подробное описание процедуры обучения моделей и подхода к процедуре тестирования. Далее в разделе приводятся результаты тестирования полученных моделей на открытых наборах данных LFW, AgeDB-30 и MegaFace.

Раздел **3.5 Исследование предложенного метода** анализирует как различные аспекты предлагаемого алгоритма влияют на его точность. Для этого из предлагаемого подхода поочерёдно исключаются компоненты алгоритма и замеряется их влияние на итоговую точность.

Раздел **3.6 Заключение по третьей главе** подытоживает представленные в ней результаты.

²<https://github.com/david-svitov/margindistillation>

Глава 4 Разработка программного модуля распознавания лиц для встраиваемых систем представляет представлено описание разработки и реализации системы распознавания лиц в виде набора нейронных сетей и программного модуля для работы на борту маломощных встраиваемых устройств. К сценариям применения данной системы могут быть отнесены: домофон с функцией распознавания жильцов, система биометрической идентификации смартфона, пропускная система предприятия.

В разделе **4.1 Обучение нейронных сетей** обосновывается выбор используемых в системе архитектур нейронных сетей. Обсуждается выбор фреймворка для их обучения.

В разделе **4.2 Конвертация обученных параметров** приводится описание процедуры переноса весов обученной модели в конечный фреймворк, используемый на устройстве.

Раздел **4.3 Разработка ключевых узлов программного модуля** посвящён описанию узлов системы, использующих нейросетевые модели. Как видно из схемы на рисунке 0.3, система содержит два ключевых программных узла: "выравнивание лица на изображении" и "сравнение с базой данных биометрических векторов".

Раздел **4.4 Тестирование системы институтом стандартов NIST** описывает процедуру тестирования, используемые метрики и тестовые данные, использованные институтом. Приводятся результаты тестирования на данных, полученных в лабораторных условиях и на реальных данных. Приводятся замеры скорости разработанной системы.

В разделе **4.5 Тестирование системы на данных с камеры домофона** приводятся точности разработанной системы на реальных данных непосредственного заказчика системы. Замеры проводятся на данных с видеодомофонов компаний ООО "Новотелеком" и ООО "Рубитек".

Оценка точности работы системы производилась в два этапа: сначала оценивалась точность работы детектора лиц, затем точность идентификации на основе биометрических векторов. Подробно описываются используемые метрики и тестовые данные. Приводится сравнение точности для различных

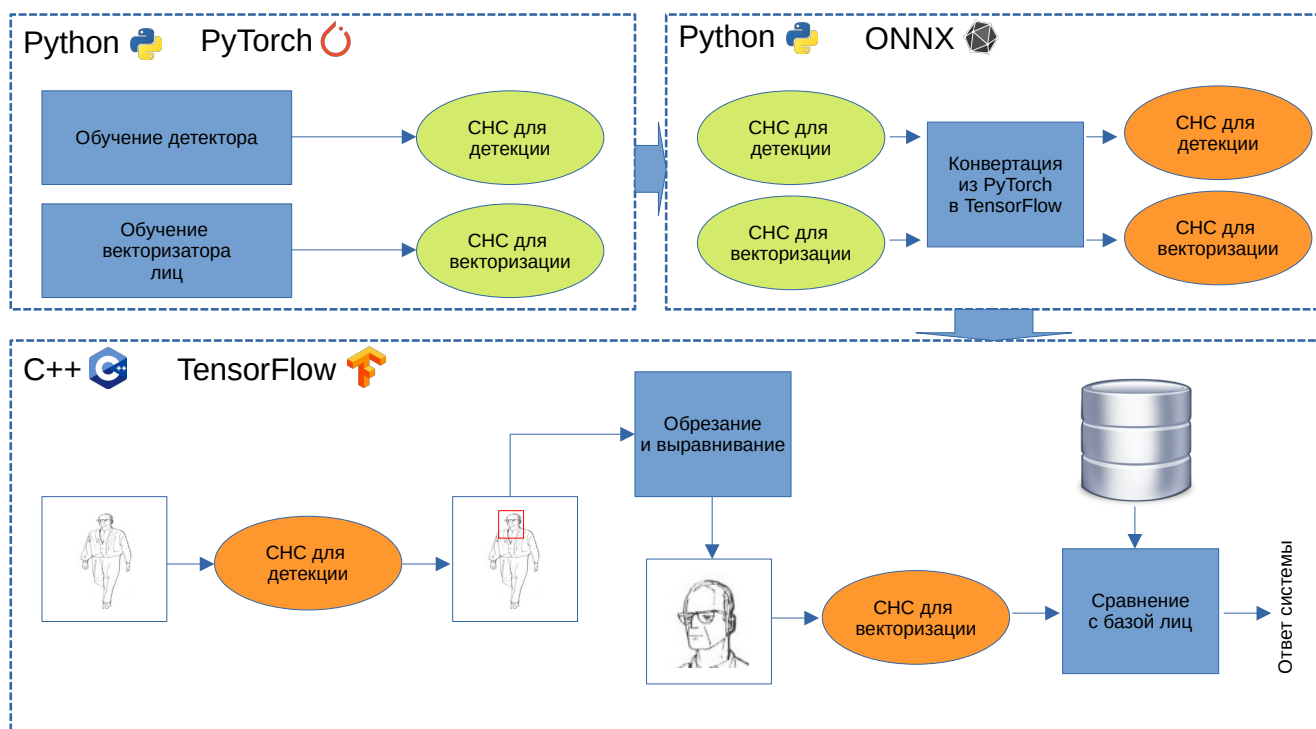


Рис. 0.3 — Схематическое представление трёх этапов разработки системы распознавания лиц: обучение нейронных сетей, конвертация обученных параметров, разработка ключевых узлов программного модуля.

Таблица 0.1 — Сравнение времени на 1th Gen Intel(R) Core(TM) i7-11700K @ 3.60GHz.

Модель	Время, мс
Resnet100 (учитель)	509.4
MibiFace_x3_swish_se (ученик)	36.7

сценариев использования системы (Таблица 0.2). И скорости работы системы (Таблица 0.1).

В разделе **4.6 Заключение по четвёртой главе** приводятся выводы по произведённым замерам. Из представленных экспериментов видно, что предлагаемые методы позволили ускорить построение биометрических векторов **в 13 раз**, при получении сравнимой или большей (на некоторых категориях данных) точности.

Таблица 0.2 — Сравнение точности моделей на реальных данных для домофонов с различной степенью сжатия для видео.

Модель	Сильное сжатие TAR@FAR 10%	Слабое сжатие TAR@FAR 10%
Resnet100 (учитель)	91.84%	100%
MibiFace_x3_swish_se (ученик)	89.8%	93.55%
MibiFace_x3_swish_se (дистиллированная)	93.88%	96.77%
MibiFace_x3_swish_se (дистиллированная; квантованная)	93.88%	96.77%

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

1. Предложен метод инициализации весов СНС с малым числом параметров для обучения сети на задачи распознавания лиц. Предложенный метод демонстрирует эффективность использования весов последнего слоя сети с большим числом параметров в извлекателе признаков для инициализации сети с малым числом параметров.
2. В работе продемонстрировано повышение точности СНС для распознавания лиц на примере модели с архитектурой MobileFaceNet на наборе данных LFW с 99.51% до 99.60% за счёт применения разработанного метода инициализации последнего слоя сети-ученика значениями весов сети сети-учителя.
3. Предложен метод дистилляции для повышения точности СНС с малым числом параметров, обученных для задачи распознавания лиц. Предложенный метод дистилляции учитывает специфику работы нейронных сетей, обученных с функцией Софтмакс с отступами.
4. На примере нейронной сети с архитектурой MobileFaceNet был продемонстрирован прирост точности на наборе данных MegaFace с 90.62% до 91.70% за счёт применения разработанного подхода. Предложенный

- подход к дистилляции позволяет получить большую точность модели, чем применение прочих существующих методов дистилляции, позволявших получить точность до 90.77% на этом наборе данных.
5. Предложен метод для снижения числа ложноположительных срабатываний детектора людей в видеопоток. Это позволило повысить надёжность всей системы распознавания лиц, за счёт исключения из дальнейшей обработки случаев, приводящих к недетерминированному поведению системы. Метод использует промежуточные карты признаков из СНС, извлекающей признаки в детекторе для устойчивого к шуму обнаружения движения. Также предложенный подход позволил ускорить этап детектирования человека за счёт ранней остановки работы детектора на кадрах, не содержащих движущихся объектов.
 6. Реализация разработанного метода обнаружения движения на основе промежуточных карт признаков позволяет снизить среднее время работы нейронной сети с 0.135 сек до 0.112 сек для модели с архитектурой SSD+MobileNetV2. При применении реализации разработанного подхода метрика Mean Average Precision возросла с 0.326 до 0.548 на наборе данных CDNet2014 pedestrian.
 7. Предложенные в данной работе методы реализованы и внедрены в виде ключевых модулей в системы распознавания лиц следующих партнёров ООО Экспасофт: ООО Новотелеком, ООО Рубитек РУС, ООО Открытая мобильная платформа, о чём имеется соответствующий акт о внедрении, прикреплённый к данной диссертации.
 8. Для данных с видеодомофона разработанный программный комплекс, реализующий представленные подходы, позволил произвести ускорение системы в 13 раз без просадки точности для некоторых сценариев использования. Предложенные методы и все приведённые эксперименты реализованы на языке Python с использованием библиотеки Pytorch, и их исходный код выложен в открытый доступ на платформе GitHub для того, чтобы на них могли опираться дальнейшие исследования.

Список литературы

1. Low-power computer vision: Status, challenges, and opportunities [Текст] / S. Alyamkin, ..., D. Svitov, G. K. Thiruvathukal, B. Zhang, J. Zhang, X. Zhang, S. Zhuo [и др.] // IEEE Journal on Emerging and Selected Topics in Circuits and Systems. — 2019. — Т. 9, № 2. — С. 411—421.
2. NTIRE 2021 challenge on image deblurring [Текст] / S. Nah, ..., D. Svitov, D. Pakulich, J. Kim, J. Jeong // Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. — 2021. — С. 149—165.
3. *Svitov, D.* MarginDistillation: Distillation for Face Recognition Neural Networks with Margin-Based Softmax [Текст] / D. Svitov, S. Alyamkin // International Journal of Computer and Information Engineering. — 2021. — Т. 15, № 3. — С. 206—210.
4. *Свитов, Д. В.* ОПТИМИЗАЦИЯ НЕЙРОСЕТЕВОГО ДЕТЕКТОРА ДВИЖУЩИХСЯ ОБЪЕКТОВ [Текст] / Д. В. СВИТОВ, С. А. АЛЯМКИН // Автометрия. — 2021. — Т. 57, № 1. — С. 21—30.
5. *Свитов, Д. В.* Дистилляция моделей для распознавания лиц, обученных с применением функции Софтмакс с отступами [Текст] / Д. В. СВИТОВ, С. А. АЛЯМКИН // Автоматика и телемеханика. — 2022. — № 10. — С. 35—46.