

# Разработка и исследование биометрической системы распознавания лиц на основе применения метода глубокого обучения

Е.Е. Истратова, Д.А. Пустовских

**Аннотация**—Для увеличения надежности и эффективности работы в большинстве информационных систем применяются технологии биометрической идентификации личности, где в качестве объекта исследования используется изображение лица человека. Цель исследования заключалась в разработке модели валидации биометрических данных для обеспечения быстрой и эффективной верификации и идентификации пользователей, а также контроля их доступа в режиме реального времени. В качестве методов исследования были использованы технология компьютерного зрения и нейросетевые технологии. Для реализации модели валидации биометрических данных был разработан и исследован собственный ансамбль алгоритмов, спроектированный на основе базовой модели для детектирования лица BlazeFace. В блоке обнаружения и распознавания лица были выделены 478 лицевых точек, на базе которых была построена трехмерная модель лица. Найденные маркеры позволили вычислить параметры для валидации: наклон головы по разным осям; направление взгляда и открытость глаз; размытость анализируемого кадра. Отличительной особенностью предложенной модели является акцент на семантически значимые области лица за счет выполнения дополнительных вычислений, что позволяет более точно прогнозировать ориентиры вокруг губ, глаз и радужной оболочки, повышая точность распознавания. Результаты сравнительного анализа с базовой моделью показали увеличение точности распознавания на 14%, и сокращение общей продолжительности распознавания лиц в 3,6 раза. Разработанное программное обеспечение может быть использовано для видеонаблюдения, контроля за ходом испытаний в системах прокторинга, а также для обеспечения безопасности в составе модуля систем безопасности зданий.

**Ключевые слова**—нейронные сети, валидация, биометрические данные, программное обеспечение, распознавание лиц.

## I. ВВЕДЕНИЕ

Биометрические системы постоянно развиваются и предлагают технологии, которые можно использовать в автоматических системах для однозначной и эффективной идентификации и аутентификации личности человека без необходимости пользователю что-либо носить с собой или запоминать, в отличие от

традиционных методов, таких как пароли, идентификаторы.

Согласно статье [1], развитие биометрических технологий используется предприятиями для проверки личности пользователей с целью ограничения и предоставления доступа к технологическим исследовательским центрам или различным производственным объектам. Так, в статье [2] предлагается улучшенная архитектура на основе глубокой сверточной нейронной сети для системы управления повседневным доступом. Однако в системах безопасности, основанных на биометрии, частота отказов зависит не от самой конфигурации системы, а от извлечения признаков и связанных с ними статистических данных. В исследовании [3] представлен подход к минимизации частоты отказов и поддержанию высокой точности и единообразия распознавания характерных точек. В результате предложенная методика подтвердила эффективность применения на наборе данных при его непредсказуемых вариациях.

Ряд исследований посвящены повышению скорости распознавания и сокращению времени распознавания в системах валидации биометрических данных. Как правило, указанные методы основаны на статистическом или нейросетевом выделении признаков. Наряду с классическими методами распознавания лиц также исследуются новейшие направления исследований: глубокое обучение, разреженные модели и теория нечетких множеств. Кроме того, в статьях представлены будущие аспекты технологий распознавания лиц и их потенциальное значение в грядущем цифровом обществе [4].

## II. АНАЛИЗ КЛАССИЧЕСКИХ МОДЕЛЕЙ И ПОДХОДОВ ДЛЯ ВАЛИДАЦИИ ДАННЫХ

Исходя из проведенного предварительного исследования в основе процесса валидации данных лежат методики обнаружения и распознавания человеческих лиц. Причем именно обнаружение лиц на изображениях является одной из актуальных задач в области автоматической обработки визуальной информации [5].

Обнаружение лица — это первый и важный шаг для его распознавания, который используется для определения лиц на изображениях и применяется в области контроля доступа [6].

В системах видеонаблюдения автоматическое определение личности человека является приоритетной

Истратова Евгения Евгеньевна, Новосибирский государственный технический университет, [istratova@mail.ru](mailto:istratova@mail.ru)

Пустовских Дмитрий Александрович, Новосибирский государственный технический университет, [dimon.pustovskih@mail.ru](mailto:dimon.pustovskih@mail.ru)

задачей, так как значительно увеличивает пропускную способность при входе в здание. Методы обнаружения позволяют определять несколько лиц на изображении. В случае множественного обнаружения лиц их конечное местоположение просто определяется как множество их местоположений [7].

Все методы обнаружения лиц делятся на два основных вида: эмпирические и связанные с моделированием изображения лица. Эмпирические методы основываются на опыте человека в решении задачи распознавания лиц и пытаются формализовать и алгоритмизировать этот опыт, построив на его основе автоматическую систему распознавания. При этом формируется ряд эвристик, на базе наличия и взаимного соответствия которых информационная система может определить факт присутствия лица на изображении. Среди данной группы методов можно выделить два направления: методы распознавания «сверху-вниз», основанные на знаниях, и методы распознавания «снизу-вверх», опирающиеся на особенности применения. Распознавание «сверху-вниз» означает построение некоторого набора правил, которым должен отвечать фрагмент изображения, для того, чтобы быть признанным человеческим лицом. В качестве ключевых правил могут выступать следующие правила:

1. Лицо зачастую симметрично относительно вертикальной оси.
2. На лице присутствуют отдельные черты (глаза, нос, рот).
3. Черты лица (глаза, нос, рот) отличаются от кожи по яркости (обычно им соответствуют области резкого изменения яркости).
4. Черты лица имеют строго определенные отношения размеров и взаимное положение.

Распознавание «снизу-вверх» использует инвариантные признаки изображений лиц, связанные с условиями съемки, к ним относятся следующие методы обнаружения:

1. Обнаружение элементов и особенностей, характерных для изображения лица.
2. Анализ обнаруженных особенностей, вынесение решения о количестве и расположении лиц. При этом производится комплексная проверка выделенных областей с характерными для человеческого лица свойствами для выявления областей, действительно являющихся лицами.

Вторая категория методов, связанная с моделированием изображения лица, не использует нейронные сети для обработки информации, а акцентирует внимание на нахождении закономерностей и свойств лиц на изображении, применяя при этом методы математической статистики и машинного обучения. Основываясь на методах распознавания образов и обнаружения лиц, задачи валидации можно рассматривать как частный случай.

Причем задача обнаружения лица обычно решается в два этапа:

1. Поиск фрагментов для классификации.
2. Разделение найденных фрагментов на классы.

Изображению или его фрагменту ставится в соответствие вычисленный вектор признаков, который используется для классификации изображений на два класса – «лицо» и «не лицо». Каждый пиксель на изображении становится частью вектора признаков, в итоге изображение представляется в виде вектора  $n \times m$  в пространстве  $R_n \times m$ .

Распознавание лиц — одна из наиболее изученных задач, решаемых при помощи технологии компьютерного зрения; оно служит фундаментальным шагом для многих приложений для анализа лиц, который состоит из четырех этапов: распознавание лиц, выравнивание лиц, извлечение объектов и сопоставление лиц [8].

В прошлом, большинство исследований по распознаванию лиц были сосредоточены на программных алгоритмах, которые зачастую было трудно превратить в реальные продукты из-за сложности этих алгоритмов и низкой производительности в реальном времени [9].

Существует множество методов распознавания по геометрии лица, основанных на том, что черты каждого человека являются уникальными. Данная область биометрии является перспективной, так как автоматическое распознавание человека исключает множество ошибок со стороны другого человека.

При распознавании лиц по 2D-изображению основным преимуществом является отсутствие необходимости использования дорогостоящего оборудования, требуется лишь наличие видеоприбора и дополнительных условий для повышения точности определения. Ключевые этапы подобного распознавания включают в себя обработку и анализ изображений с последующим распознаванием образов. Первым этапом является низкоуровневая обработка изображений без детального изучения содержимого изображения. Данный этап применяется для подготовки изображения к анализу и заключается в следующих действиях:

1. Восстановление.
2. Очистка от шумов.
3. Сжатие данных.
4. Улучшение характеристик (четкость, контраст).

На входе и выходе этого этапа всегда находятся 2D-изображения.

Следующим этапом является анализ изображений, представляющий собой выделение на изображениях объектов и измерение их параметров, то есть извлечение информации из изображений. Входящей информацией для данного этапа выступают одно или несколько 2D-изображений, к ним могут относиться изображения с цветной или черно-белой камеры, данные с инфракрасной камеры. Карта глубины тоже относится к 2D-изображениям, хотя и характеризует трехмерное пространство. На выходе получается вектор признаков с их значениями и вероятностными оценками.

Этап распознавания образов представляет собой классификацию или кластеризацию вектора признаков, полученного на этапе анализа изображений. На выходе может быть либо номер образа (класса, кластера), либо вектор вероятностей принадлежности исходного изображения к образам.

К основным недостаткам систем по распознаванию лиц можно отнести низкую статистическую достоверность, повышенные требования к окружающей среде (свет, влажность и т.д.), требования к положению человека на изображении. Несмотря на это, системы распознавания лиц по 2D-изображениям получили широкое распространение в связи с их низкой стоимостью. В настоящее время предпочтение отдают данным методам, так как большинство систем видеонаблюдения являются системами, работающими в реальном времени и требующими быстрой обработки информации. Также в связи с развитием техники в современных системах видеонаблюдения все чаще используются камеры высокого разрешения, а это накладывает жесткие требования на время работы [10].

При распознавании лиц по 3D-изображению выделяют два основных подхода, основанные на технологиях стерео и освещения лица структурированным светом. В первом случае 3D-система стерео создает трехмерные изображения, синтезируя две или более двухмерные фотографии. Этот подход на основе компьютерной обработки в некоторой степени осложняет процесс трехмерного сопоставления изображения лица, в результате чего, надежность результата напрямую зависит от условий освещенности. При другом подходе лицо освещается структурированным светом невидимого, почти инфракрасного спектра от трех источников, отраженное от лица излучение улавливается цифровой видеокамерой, при этом собирается около 40 000 измерительных точек.

Поскольку в 3D-сканерах лица на основе структурированного света используются свои источники света, надежность результата применения этого оборудования в плохих условиях освещенности гарантирована.

Детекторы лиц обычно предназначены для фронтальной конфигурации лица. Поэтому их производительность чувствительна как к вращениям в плоскости, так и вне ее. Некритичным является и поворот головы человека, проходящего идентификацию. Конечные пользователи, как правило, не испытывают трудностей при обращении с трехмерным считывателем. Пользователь подходит к считывателю, останавливается перед ним, далее полученное изображение сопоставляется с образами в базе данных.

Еще одним подходом к обнаружению и распознаванию лиц является методика детектирования лица с использованием нейронных сетей. Данная технология использует каскадную архитектуру, построенную на сверточной нейронной сети и

работающую с различными масштабами. Ключевым преимуществом данного подхода является высокая скорость работы, которая позволяет применять сверточные нейронные сети в системах реального времени. Скорость обработки одного кадра может достигать примерно 1-2 мс. К основным недостаткам подобного подхода относятся повышенные требования к вычислительной мощности и необходимость обучения каждой модели [11].

Обучаются нейронные сети на наборе примеров. Суть обучения сводится к настройке весов межнейронных связей в процессе решения оптимизационной задачи методом градиентного спуска. В процессе обучения нейронной сети происходит автоматическое извлечение ключевых признаков, определение их важности и построение взаимосвязей между ними. Предполагается, что обученная нейронная сеть сможет применить опыт, полученный в процессе обучения, на неизвестные образы за счет обобщающих способностей.

Таким образом, для решения задачи обнаружения и распознавания лиц наиболее эффективным является использование искусственных нейронных сетей, поскольку они обеспечивают возможность получения классификатора, хорошо моделирующего сложную функцию распределения изображений лиц, тем самым, увеличивая точность по сравнению с другими методами.

### III. РАЗРАБОТКА МОДЕЛИ ВАЛИДАЦИИ БИОМЕТРИЧЕСКИХ ДАННЫХ ЧЕЛОВЕКА

В качестве основы для проектирования модели для валидации биометрических данных была использована модель BlazeFace для детектирования лица.

BlazeFace — это модель машинного обучения, разработанная Google для быстрого определения местоположения и ключевых точек лиц. Оба эти вида параметров могут быть получены одновременно. Модель BlazeFace выдает шесть координат ключевых точек лица: для центров глаз, ушных раковин, центра рта и кончика носа. Также с ее помощью возможно обнаружение нескольких людей одновременно.

В модели используются два основных блока для организации архитектуры нейронной сети, представленные на рис. 1. В данных блоках используются глубокие сверточные слои, что повышает скорость работы модели. Ядро 5x5 используется для преобразования DW, а ядро 1x1 — для слоев преобразования. В модели используется пять одинарных блоков BlazeBlock и шесть двойных блоков BlazeBlock. Полная архитектура показана в табл. 1.

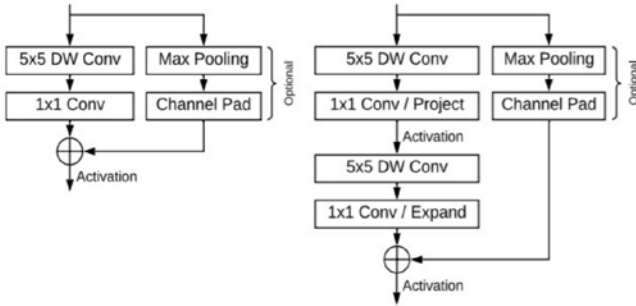


Рисунок 1 — Блоки модели BlazeFace: одиночный BlazeBlock (слева) и двойной BlazeBlock (справа)

Для расчета привязки используются два сверточных слоя Conv с размером ядра 16x16 и размером ядра 8x8 с количеством фильтров два и восемь, соответственно.

Вывод модели требует постобработки, так как количество перекрывающихся привязок значительно увеличивается при увеличении размеров объекта. Этот прогноз сначала обрезается, чтобы соответствовать границам размеров изображения, затем применяется NMS, чтобы получить окончательные блоки. Размер данных полей можно изменить до исходного размера изображения, поскольку оно было изменено на 128x128.

Обзор классических моделей и подходов к обнаружению и распознаванию лиц позволил разработать собственную модель для валидации биометрических данных человека на основе модели BlazeFace. Цель ее проектирования и исследования заключалась в обнаружении лица на изображении и нахождении его ключевых точек.

Слой	Размер входного вектора	Размер ядра сверточного слоя
Сверточный слой	128x128x3	5x5x3x24(stride 2)
Одиночный BlazeBlock	64x64x24	5x5x24x1 1x1x24x24
Одиночный BlazeBlock	64x64x24	5x5x24x1 1x1x24x24
Одиночный BlazeBlock	64x64x24	5x5x24x1(stride 2) 1x1x24x48
Одиночный BlazeBlock	32x32x48	5x5x48x1 1x1x48x48
Одиночный BlazeBlock	32x32x48	5x5x48x1 1x1x48x48
Двойной BlazeBlock	32x32x48	5x5x48x1(stride 2) 1x1x48x24 5x5x24x1 1x1x24x96
Двойной BlazeBlock	16x16x96	5x5x96x1 1x1x96x24 5x5x24x1 1x1x24x96
Двойной BlazeBlock	16x16x96	5x5x96x1 1x1x96x24 5x5x24x1 1x1x24x96
Двойной BlazeBlock	16x16x96	5x5x96x1(stride 2) 1x1x96x24 5x5x24x1 1x1x24x96
Двойной BlazeBlock	8x8x96	5x5x96x1 1x1x96x24 5x5x24x1 1x1x24x96
Двойной BlazeBlock	8x8x96	5x5x96x1 1x1x96x24 5x5x24x1 1x1x24x96

Таблица 1. Архитектура нейронной сети

Результатами работы модели стали формирование ограничивающей рамки вокруг лица и определение нескольких ориентиров, включая центры глаз, кончик

носа и т. д. Схема с последовательностью реализации основных этапов модели представлена на рис. 2.

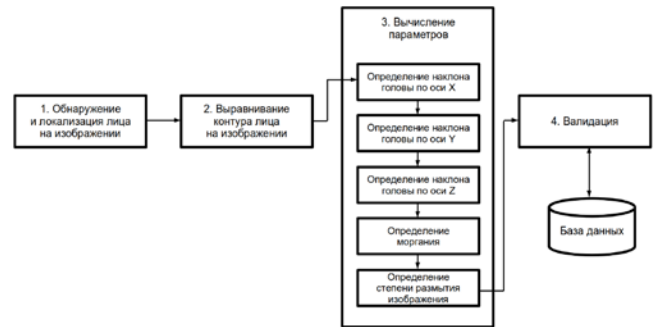


Рисунок 2 — Схема ансамбля алгоритмов

Отличительной особенностью предложенной модели является акцент на семантически значимые области лица за счет выполнения дополнительных вычислений. В результате это позволяет более точно прогнозировать ориентиры вокруг губ, глаз и радужной оболочки, что, в свою очередь, повышает точность распознавания.

Для реализации модели валидации биометрических данных был предложен и исследован разработанный ансамбль алгоритмов. В блоке обнаружения и распознавания лица находятся 478 лицевых точек, на основе которых можно построить трехмерную модель лица. На базе найденных маркеров вычисляются параметры для валидации, в них входят: наклон головы по разным осям, параметры глаз (направление взгляда и их открытость), также определяется, размыт ли анализируемый кадр. На основе найденных параметров можно принимать решение о соответствии изображения стандарту или заданным параметрам.

#### IV. РЕАЛИЗАЦИЯ МОДЕЛИ ВАЛИДАЦИИ БИОМЕТРИЧЕСКИХ ДАННЫХ ЧЕЛОВЕКА

Для реализации разработанной модели валидации биометрических данных человека необходимо предварительно определить требования к инструментальным средствам, чтобы максимально оптимизировать процесс выбора подходящих программных продуктов.

В наше время наблюдается стремительный прогресс в сфере разработки и использования программных решений на базе технологий машинного обучения. Для эффективного и удобного использования результатов работы нейронных сетей предполагается использование значительных по масштабам вычислительных мощностей, которые могут быть недоступны для большинства разработчиков в этой сфере. Для подключения к этим кластерам, как правило, разрабатываются удобные и понятные интерфейсы, которые помогают найти подходы к корректному запуску, управлению и отладке в процессе работы, а также получению и анализу промежуточных результатов. Следовательно, удобное и функциональное средство разработки программ на основе машинного

обучения должно учитывать все возможные способы проектирования, подходы к реализации и банальное удобство взаимодействия с нейронными сетями с возможностью изменения топологии в процессе обучения или разработки [12].

Таким образом, средство проектирования, прежде всего, должно удовлетворять следующим параметрам:

1. Система должна являться полноценным, но не обязательно самодостаточным (для процесса обучения может быть реализована возможность использования сторонних сервисов, предоставляющих удаленные вычислительные мощности) продуктом, который предоставляет возможность обучения моделей, отладки самого процесса обучения и модернизации математического аппарата.
2. Система должна иметь дружелюбный интерфейс для упрощения взаимодействия не ознакомленного разработчика с функционалом системы.
3. Требуется использовать современные средства и библиотеки разработки программных модулей нейронной сети, поскольку основная задача системы разработки заключается в предоставлении возможности обучения нейронной сети посредством выбранного математического аппарата.
4. Взаимозаменяемость программных модулей, которая может быть реализована через наличие прозрачных и описанных в документации протоколов взаимодействия системы, использования API. Наличие документации в принципе является важной составляющей любого программного средства, в особенности, если это средство проектирования.

Машинное обучение описывает вычислительную модель, которая использует алгоритмы для обучения на огромных объемах данных. Благодаря этому, методы машинного обучения позволяют с успехом решать некоторые проблемы, которые трудно решить с помощью традиционных программных систем, например, распознавание изображений, медицинская диагностика, предсказание цен на рыночные активы. В отличие от какого-либо конкретного алгоритма или модели машинного обучения, система машинного обучения в практических приложениях относится к полным системам, включающим:

- 1) одну или несколько моделей машинного обучения;
- 2) данные, используемые для обучения модели;
- 3) интерфейсы и документы для взаимодействия пользователя с моделью.

Традиционные программные системы обычно используют явные правила взаимодействия человека с компьютером для управления поведением системы, в то время как системы машинного обучения больше полагаются на характеристики данных и неявные

модели принятия решений. Из-за сложных и изменчивых атрибутов данных, трудностей для измерения и плохой интерпретируемости традиционных моделей машинного обучения данный вид систем ставит новые вызовы перед проектированием требований к ним.

К основным характеристикам, которые необходимо учитывать при формировании требований к инструментальным средствам для реализации программного обеспечения для валидации биометрических данных человека, относятся следующие: скорость обработки; информационная безопасность; конфиденциальность; эффективность; надежность и интерпретируемость.

Исходя из того, что программное обеспечение для валидации биометрических данных относится к группе систем, основанных на применении технологии компьютерного зрения, то в качестве дополнительных требований к нему могут выступать следующие:

1. Скорость обработки и передачи информации.
2. Дополнительные требования к качеству фото и видео данных.
3. Объем данных для обучения.
4. Точность анализа данных с учетом вероятности ошибки по сравнению с человеком и другими устройствами.
5. Оптические характеристики применяемых веб- или видеокамер по определению цветов, расстояний и размеров, поиску по шаблонам.
6. Дополнительные требования к аппаратному обеспечению, связанные с разрешением видеосенсоров, динамическим диапазоном и объемом вычислительной мощности для обработки.

Таким образом, смысл определения требований к инструментальным средствам заключается в анализе компонентов процесса машинного обучения, таких как: наборы данных и алгоритмы обучения, а также в определении стратегии реализации для удовлетворения всех функциональных и нефункциональных требований, предъявляемых к методам машинного обучения.

Исходя из описанных требований, можно выделить следующие рекомендуемые группы инструментальных средств для проектирования программного обеспечения:

1. Языки программирования.
2. Системы управления базами данных.
3. Библиотеки алгоритмов компьютерного зрения.
4. Фреймворки для разработки интерфейса.
5. Фреймворки для обработки изображений.

Наиболее распространенными языками программирования, применяемыми для разработки систем по обнаружению и распознаванию лиц, являются следующие: C++, Python, Java. Первый вариант применяется, если необходимо соединить несколько сервисов в единую платформу или интегрировать готовый программный продукт в качестве встраиваемого модуля в более сложную

многокомпонентную информационную систему. Python применяется, когда акцент разработки смещается в сторону применения устройств и технологий Интернета вещей, так как в данном случае требуется минимизировать нагрузку на все компоненты системы. Кроме того, этот вариант успешно зарекомендовал себя при использовании в составе систем с небольшой вычислительной мощностью. Язык программирования Java чаще всего используется при проектировании информационных систем и их модулей, которые будут работать с большими объемами «сырых» данных, как правило, подобные системы активно проектируются для экономического сектора.

Системы управления базами данных применяются для хранения цифровых изображений, поступающих с видеокамер или с веб-камер. Основным требованием к данной группе инструментальных средств является высокая производительность, обеспечивающая в большей степени быстрое действие процесса обработки изображений, влияя, тем самым, на производительность процесса валидации в целом. Наиболее распространенными программными продуктами, относящимися к этой группе, являются следующие: MySQL, PostgreSQL. Широкое распространение данных систем управления базами данных связано со стоимостью их лицензий (распространяются бесплатно) и с высокой производительностью. Дополнительным преимуществом выступает наличие полного пакета достаточно подробной документации, регламентирующей процесс работы с этими программными продуктами.

Для работы с изображениями при реализации проекта предполагается использовать библиотеку OpenCV. Выбор в данном случае ограничен отсутствием достаточно производительных и корректных по взаимодействию с изображениями аналогов.

OpenCV – это библиотека, содержащая в себе набор функций для реализации технологии компьютерного зрения в режиме реального времени. Одним из ключевых достоинств библиотеки OpenCV является то, что ее функции реализованы как для Python, так и для языка программирования C++. Это позволяет адаптировать реализации модифицированного подхода обнаружения движения на C++ для его дальнейшего развития и внедрения в системы прокторинга и видеонаблюдения.

Для работы с цифровыми изображениями и для удобства пользователей программного обеспечения необходимо разработать соответствующие интерфейсы. Наиболее рациональным вариантом при этом является применение фреймворков.

Для разработки интерфейса предполагается использовать фреймворк Qt5, представляющий собой кроссплатформенную среду разработки приложений для настольных компьютеров, встроенных и мобильных устройств. К поддерживаемым платформам относятся следующие: Linux, OS X, Windows, VxWorks, QNX,

Android, iOS, BlackBerry, OS Sailfish и другие.

Qt — это фреймворк, написанный на языке программирования C++. Препроцессор MOC (компилятор метаобъектов) используется для расширения языка C++ такими функциями, как сигналы и слоты. Перед этапом компиляции MOC анализирует исходные файлы, написанные на расширенном Qt C++ и генерирует из них стандартные исходные коды C++. Таким образом, сам фреймворк, а также все приложения и библиотеки, использующие его, могут быть скомпилированы любым стандартным компилятором C++, например, при помощи Clang, GCC, ICC, MinGW или MSVC.

Данный фреймворк также выполняет функции графической библиотеки для языка программирования C++, которая содержит множество компонентов для поддержки не только графики, но и сети, базы данных и т. д., то есть целый набор инструментов для быстрого и удобного проектирования графического интерфейса будущего программного обеспечения.

Для создания пользовательского интерфейса в Qt применяются два вида элементов. Первым из них является Designer. С помощью этого средства можно с относительной легкостью создавать окна программ. Полученный интерфейс сохраняется в файл с расширением .ui, который подключается к создаваемой программе с помощью специальных методов.

Ко второму виду элементов относятся сигналы и слоты. В этом заключается отличительная особенность фреймворка Qt по сравнению с другими аналогами. С помощью данного функционала осуществляется связь между объектами. Сигнал отправляется в слот, если происходит запрограммированное событие. Слот же обязан ответить на все подключенные к нему сигналы. Данная схема представляет собой функции-члены объектов, у которых основное направление задач — это оперативное реагирование на действия пользователя.

Для обработки изображений необходимо выбрать специальный программный продукт, отличающийся высокой производительностью и наличием специального функционала для работы с графикой. Рынок программного обеспечения, применяемого в данной сфере, не велик, поэтому наиболее оптимальным вариантом, не имеющим достойных аналогов, является фреймворк Mediapipe.

Благодаря программным особенностям данного фреймворка, работа приложения выполняется через граф, узлами которого являются калькуляторы, выполняющие какое-либо действие или вычисления. Все вершины имеют выход и вход.

Mediapipe — это кроссплатформенный фреймворк для запуска моделей машинного обучения. Сама модель задается в форме графа, граф состоит из следующих элементов:

1. Вершины графа (Calculators) — это некоторые преобразования для данных. У каждой вершины должен быть как минимум один входящий и как

минимум один исходящий поток. Вершина представляет собой C++ класс, реализующий интерфейс CalculatorBase.

2. Ребра графа (Streams) задают связи между вершинами. С помощью потоков по графу перемещаются пакеты с данными. Потоки могут быть внутренними, входящими и исходящими. Внутренний поток соединяет две вершины, через входящий поток из внешнего кода в граф попадают данные, а с помощью исходящего потока граф отправляет данные в вызывающий код.
3. Пакет (Packet) — это единица данных, перемещаемая по потокам и обрабатываемая в вершинах графа. Каждый пакет несет в себе данные определенного типа — это может быть строка, целое число, массив чисел с плавающей запятой или пользовательский тип, описанный и сериализуемый в protobuf. Каждый пакет содержит в себе timestamp, то есть отметку времени, ассоциированную с пакетом, она необходима для того, чтобы отличать, какой пакет был отправлен раньше, какой — позже, причем данная метка с реальным временем напрямую не связана.

Программный интерфейс для реализации функционала в графе содержит следующие методы:

1. Static Status GetContract (CalculatorContract\*) — статический метод, в котором узел описывает форматы данных, которые ждет на вход и готов отдать на выход.
2. Status Open (CalculatorContext\*) — инициализация вершины при создании графа. Здесь, например, может быть загрузка данных, требуемых для работы.
3. Status Process (CalculatorContext\*) — обработка поступившего пакета.
4. Status Close (CalculatorContext\*) — закрытие вершины.

Графы описываются в формате protobuf.

Таким образом, для реализации предложенной модели валидации данных на основании сравнительного анализа и с учетом предъявляемых требований были определены группы инструментальных средств и выбраны следующие наиболее оптимальные решения для реализации программного обеспечения:

1. Язык программирования — C++.
2. Библиотека алгоритмов компьютерного зрения — OpenCV.
3. Система управления базами данных — PostgreSQL.
4. Фреймворк для разработки интерфейсов — Qt5.
5. Фреймворк для обработки изображений — MediaPipe.

Для обучения модели нейронной сети был собран и размечен набор данных. Часть набиралась из сети Интернет, другая — была собрана вручную. Нужно

было обеспечить разнообразие данных, чтобы модель могла работать в различных условиях (зашумленность, разные ракурсы и т.д.), а после для улучшения точности модели необходимо было собрать такой датасет, где один и тот же человек будет сфотографирован с хорошим разрешением и с разных ракурсов. После сбора данных происходила их разметка.

На сегодняшний день разметка данных активно используется в сферах машинного обучения и искусственного интеллекта. Без размеченных данных нельзя было бы обучить алгоритмы машинного обучения. Разметка данных представляет собой процесс выявления людьми объектов в сырых данных, например, в видео или на изображениях, и добавления к ним меток. С помощью размеченных данных модель нейронной сети может обучаться, так как она узнает правильный ответ. Обученная модель может распознавать и классифицировать данные уже на совершенно незнакомом для нее наборе данных. Разметка данных происходила с помощью фреймворка для обработки изображений MediaPipe.

Данное программное обеспечение представляет собой открытый бесплатный инструмент с графическим интерфейсом, который был написан на языке C++. Он используется для аннотирования изображений и имеет несколько функций, адаптированных для использования с библиотекой алгоритмов компьютерного зрения OpenCV.. Для контроля качества предусмотрен быстрый обзор всех аннотаций, с помощью которого можно заметить ошибки в разметке данных.

Для обучения модели нейронной сети необходимо предоставить данные в формате, содержащем один текстовый файл для каждого изображения (содержащий аннотации и числовое представление метки) и карту меток, которая сопоставляет числовые идентификаторы с удобочитаемыми строками. Аннотации нормализованы в диапазоне [0, 1], что упрощает работу с ними даже после масштабирования или растяжения изображений.

Обучение модели нейронной сети происходило в среде google colab в течении 72 часов при количестве итераций (эпох) 46000. Для оценки качества обучения в той же среде google colab был сгенерирован график точности mAP и средней ошибки avg Loss от количества итераций. Средняя точность обнаружения mAP составила 98 %, что является довольно хорошим показателем при условии наличия большого числа классов, необходимых для распознавания. Чтобы убедиться в корректности модели, было проведено сравнение с базовой моделью. Результаты представлены в табл. 2.

Таблица 2. Результаты сравнения разработанной модели с базовой

Критерии сравнения	Модель BlazeFace	Разработанная модель
Скорость распознавания, Fps	39	140
Время распознавания, мс	25	7
Точность распознавания, %	86	98

Исходя из результатов сравнительного анализа можно сделать вывод о том, что применение разработанной модели является целесообразным. Она может быть использована для валидации биометрических данных человека в режиме реального времени, показывая хорошую точность и полноту, а также отличную скорость обработки кадров.

#### В.ЗАКЛЮЧЕНИЕ

Таким образом, в результате исследования была разработана модель для валидации биометрических данных человека. Полученная модель была реализована в виде программного обеспечения и протестирована на наборе данных. Готовый программный продукт направлен на автоматизацию рабочего процесса распознавания лиц, способствует увеличению скорости обработки информации, что приводит к сокращению затрат времени и повышению качества работы персонала, отвечающего за информационную безопасность.

Программа предназначена для анализа входящей информации в системах видеонаблюдения с возможностью интеллектуального анализа данных. Программное решение может применяться как автономно, так и подключаться к системам безопасности зданий в качестве встраиваемого модуля. Отличительной особенностью программы является использование собственных алгоритмов и моделей для решения задачи распознавания объектов в системах видеонаблюдения и прокторинга. Помимо этого, программный продукт может быть использован в научно-исследовательских целях, а также для обеспечения безопасности в составе модуля систем безопасности зданий.

#### БИБЛИОГРАФИЯ

- [1] Begum N., Mustafa A.S. A novel approach for multimodal facial expression recognition using deep learning techniques. *Multimed Tools Appl* 81, 18521–18529 (2022). <https://doi.org/10.1007/s11042-022-12238-y>.
- [2] Rusia M.K., Singh D.K. A comprehensive survey on techniques to handle face identity threats: challenges and opportunities. *Multimed Tools Appl* (2022). <https://doi.org/10.1007/s11042-022-13248-6>.
- [3] Wu C., Nabil S., Zhou S. et al. Gram Matrix-Based Convolutional Neural Network for Biometric Identification Using Photoplethysmography Signal. *J. Shanghai Jiaotong Univ. (Sci.)* (2022). <https://doi.org/10.1007/s12204-022-2426-5>.
- [4] Srivastava S., Kumar A., Singh A. et al. An improved approach towards biometric face recognition using artificial neural network. *Multimed Tools Appl* 81, 8471–8497 (2022). <https://doi.org/10.1007/s11042-021-11721-2>.
- [5] Jayavardivel R., Prabakaran P. Investigation on automated surveillance monitoring for human identification and recognition using face and iris biometric. *J Ambient Intell Human Comput* 12, 10197–10208 (2021). <https://doi.org/10.1007/s12652-020-02787-1>.

- [6] Begum N., Mustafa A.S. A novel approach for multimodal facial expression recognition using deep learning techniques. *Multimed Tools Appl* 81, 18521–18529 (2022). <https://doi.org/10.1007/s11042-022-12238-y>.
- [7] Arora S., Mittal R., Kukreja H. et al. An evaluation of denoising techniques and classification of biometric images based on deep learning. *Multimed Tools Appl* (2022). <https://doi.org/10.1007/s11042-021-11573-w>.
- [8] Senthilkumar R., Gnanamurthy K. A proposed method for the improvement in biometric facial image recognition using document-based classification. *J Supercomput* 76, 4476–4494 (2020). <https://doi.org/10.1007/s11227-018-2408-4>.
- [9] Ali W., Tian W., Din S.U. et al. Classical and modern face recognition approaches: a complete review. *Multimed Tools Appl* 80, 4825–4880 (2021). <https://doi.org/10.1007/s11042-020-09850-1>.
- [10] Jayavardivel R., Prabakaran P. Investigation on automated surveillance monitoring for human identification and recognition using face and iris biometric. *J Ambient Intell Human Comput* 12, 10197–10208 (2021). <https://doi.org/10.1007/s12652-020-02787-1>.
- [11] Tiong L.C.O., Kim S.T. & Ro Y.M. Implementation of multimodal biometric recognition via multi-feature deep learning networks and feature fusion. *Multimed Tools Appl* 78, 22743–22772 (2019). <https://doi.org/10.1007/s11042-019-7618-0>.
- [12] Arora S., Mittal R., Kukreja H. et al. An evaluation of denoising techniques and classification of biometric images based on deep learning. *Multimed Tools Appl* (2022). <https://doi.org/10.1007/s11042-021-11573-w>.

**Истратова Евгения Евгеньевна.** Новосибирский государственный технический университет, г. Новосибирск, Россия. Кандидат технических наук, доцент кафедры автоматизированных систем управления. Количество печатных работ: 97. Область научных интересов: информационные технологии, информационные системы, системы компьютерного зрения. e-mail: [istratova@mail.ru](mailto:istratova@mail.ru) (Ответственная за переписку).

**Пустовских Дмитрий Александрович.** Новосибирский государственный технический университет, г. Новосибирск, Россия. Студент факультета автоматизации и вычислительной техники. Количество печатных работ: 2. Область научных интересов: системы компьютерного зрения, искусственный интеллект, информационные технологии.



# Development and research of the biometric face recognition system based on the application of the deep learning method

E.E. Istratova, D.A. Pustovskih

**Abstract**—To increase the reliability and efficiency of work in most information systems, biometric identification technologies are used, where the image of a person's face is used as an object of study. The aim of the study was to develop a biometric data validation model to provide fast and efficient verification and identification of users, as well as real-time control of their access. Computer vision technology and neural network technologies were used as research methods. To implement the biometric data validation model, we developed and studied our own ensemble of algorithms, designed on the basis of the BlazeFace basic model for face detection. In the face detection and recognition block, 478 facial points were identified, on the basis of which a three-dimensional face model was built. The found markers made it possible to calculate the parameters for validation: head tilt along different axes; gaze direction and eye openness; blurring of the analyzed frame. A distinctive feature of the proposed model is the emphasis on semantically significant areas of the face by performing additional calculations, which makes it possible to more accurately predict landmarks around the lips, eyes, and iris, increasing recognition accuracy. The results of a comparative analysis with the base model showed an increase in recognition accuracy by 14%, and a reduction in the total duration of face recognition by 3.6 times. The developed software can be used for video surveillance, monitoring the progress of tests in proctoring systems, as well as for ensuring security as part of the building security systems module.

**Keywords**—neural networks, validation, biometric data, software, face recognition.

## REFERENCES

- [1] Begum N., Mustafa A.S. A novel approach for multimodal facial expression recognition using deep learning techniques. *Multimed Tools Appl* 81, 18521–18529 (2022). <https://doi.org/10.1007/s11042-022-12238-y>.
- [2] Rusia M.K., Singh D.K. A comprehensive survey on techniques to handle face identity threats: challenges and opportunities. *Multimed Tools Appl* (2022). <https://doi.org/10.1007/s11042-022-13248-6>.
- [3] Wu C., Nabil S., Zhou S. et al. Gram Matrix-Based Convolutional Neural Network for Biometric Identification Using Photoplethysmography Signal. *J. Shanghai Jiaotong Univ. (Sci.)* (2022). <https://doi.org/10.1007/s12204-022-2426-5>.
- [4] Srivastava S., Kumar A., Singh A. et al. An improved approach towards biometric face recognition using artificial neural network. *Multimed Tools Appl* 81, 8471–8497 (2022). <https://doi.org/10.1007/s11042-021-11721-2>.
- [5] Jayavadivel R., Prabakaran P. Investigation on automated surveillance monitoring for human identification and recognition using face and iris biometric. *J Ambient Intell Human Comput* 12, 10197–10208 (2021). <https://doi.org/10.1007/s12652-020-02787-1>.
- [6] Begum N., Mustafa A.S. A novel approach for multimodal facial expression recognition using deep learning techniques. *Multimed Tools Appl* 81, 18521–18529 (2022). <https://doi.org/10.1007/s11042-022-12238-y>.
- [7] Arora S., Mittal R., Kukreja H. et al. An evaluation of denoising techniques and classification of biometric images based on deep learning. *Multimed Tools Appl* (2022). <https://doi.org/10.1007/s11042-021-11573-w>.
- [8] Senthilkumar R., Gnanamurthy .K. A proposed method for the improvement in biometric facial image recognition using document-based classification. *J Supercomput* 76, 4476–4494 (2020). <https://doi.org/10.1007/s11227-018-2408-4>.
- [9] Ali W., Tian W., Din S.U. et al. Classical and modern face recognition approaches: a complete review. *Multimed Tools Appl* 80, 4825–4880 (2021). <https://doi.org/10.1007/s11042-020-09850-1>.
- [10] Jayavadivel R., Prabakaran P. Investigation on automated surveillance monitoring for human identification and recognition using face and iris biometric. *J Ambient Intell Human Comput* 12, 10197–10208 (2021). <https://doi.org/10.1007/s12652-020-02787-1>.
- [11] Tiong L.C.O., Kim S.T. & Ro Y.M. Implementation of multimodal biometric recognition via multi-feature deep learning networks and feature fusion. *Multimed Tools Appl* 78, 22743–22772 (2019). <https://doi.org/10.1007/s11042-019-7618-0>.
- [12] Arora S., Mittal R., Kukreja H. et al. An evaluation of denoising techniques and classification of biometric images based on deep learning. *Multimed Tools Appl* (2022). <https://doi.org/10.1007/s11042-021-11573-w>.