

УДК 004.89

АНАЛИЗ ВОЗМОЖНОСТЕЙ ИСПОЛЬЗОВАНИЯ ГЛУБОКОЙ НЕЙРОННОЙ СЕТИ В КАЧЕСТВЕ ТЕХНОЛОГИИ РАСПОЗНАВАНИЯ ОБРАЗОВ ДЛЯ РЕШЕНИЯ ЗАДАЧИ АВТОМАТИЗАЦИИ УЧЕТА ПОСЕЩАЕМОСТИ

Ильина Анна Владимировна¹, Кореньков Владимир Васильевич²

¹Студент;

ГБОУ ВО МО «Университет «Дубна»,
Институт системного анализа и управления;
141980, Московская обл., г. Дубна, ул. Университетская, 19;
e-mail: anna.ilina.1307@yandex.ru.

²Директор, доктор технических наук, профессор;

Объединенный институт ядерных исследований,
Лаборатория информационных технологий;
141980, Московская обл., г. Дубна, ул. Жолио-Кюри, 6;
ГБОУ ВО МО «Университет «Дубна»,
Институт системного анализа и управления;
141980, Московская обл., г. Дубна, ул. Университетская, 19;
e-mail: korenkov@cv.jinr.ru.

Задача подсчета количества людей актуальна при проведении разного рода мероприятий, к которым могут относиться семинары, лекции, конференции, собрания и пр. Взамен монотонного ручного подсчета участников гораздо эффективнее применить технологию распознавания лиц, предоставляющую возможность не только быстро проводить подсчет присутствующих, но также распознать каждого из них, благодаря чему становится возможным проведение дальнейшей аналитики этих данных, выявление в них закономерностей и прогнозирование. Проведенное в работе исследование определяет оценку качества использования технологии распознавания лиц на изображениях и видеопотоке, основанной на использовании глубокой нейронной сети, для решения задачи автоматизации учета посещаемости.

Ключевые слова: глубокие нейронные сети, распознавание образов, искусственный интеллект.

ANALYSIS OF THE POSSIBILITIES OF USING A DEEP NEURAL NETWORKS AS AN IMAGE RECOGNITION TECHNOLOGY TO SOLVE THE PROBLEM OF AUTOMATING ATTENDANCE TRACKING

Irina Anna¹, Korenkov Vladimir²

¹Student;

Dubna State University;
Institute of the system analysis and management;
141980, Dubna, Moscow reg., Universitetskaya str., 19;
e-mail: anna.ilina.1307@yandex.ru.

²Director, Doctor of Technical Science, professor;

Joint institute for nuclear researches,
Laboratory of Information Technologies;
141980, Moscow reg., Dubna, Joliot-Curie, 6;
Dubna State University;
Institute of the system analysis and management;
141980, Dubna, Moscow reg., Universitetskaya str., 19;
e-mail: korenkov@cv.jinr.ru.

The task of counting the number of people is relevant when conducting various types of events, which may include seminars, lectures, conferences, meetings, etc. Instead of monotonous manual counting of participants, it is much more effective to use facial recognition technology, which makes it possible not only to quickly count those present, but also to recognize each of them, which makes it possible to conduct further

analysis of this data, identify patterns in them and predict. The research conducted in this paper determines the quality assessment of the use of facial recognition technology in images and video streams, based on the use of a deep neural network, to solve the problem of automating attendance tracking.

Keywords: deep neural networks, image recognition, artificial intelligence.

Введение

Задача подсчета количества людей актуальна при проведении разного рода мероприятий, к которым могут относиться семинары, лекции, конференции, собрания, концерты и пр. Взамен монотонного ручного подсчета участников гораздо эффективнее использовать некоторое оборудование с необходимым программным обеспечением, которое позволяло бы детектировать изображения лиц участников и выдавать общее количество всех участников. Таким оборудованием может быть стационарный персональный компьютер, ноутбук или одноплатный¹ микрокомпьютер с предустановленной системой детектирования и распознавания лиц.

Нередко наличие функции обнаружения и идентификации лиц представляет одно из ключевых условий при выборе системы видеонаблюдения, поскольку сфера применения этой технологии достаточно объемна: установление личности людей, поиск преступников и предупреждение преступных действий с помощью применения поведенческого анализа, отслеживание нарушителей правил дорожного движения, поиск человека в толпе, организация контроля доступа на территорию, контроль посещаемости офиса сотрудниками, маркетинг и многое другое.

Одним из главных преимуществ использования системы распознавания лиц в качестве основной технологии для проведения интеллектуальной видеоаналитики является гарантия регистрации того или иного события только в момент появления человека в кадре. Это обеспечивает отсутствие ложных срабатываний системы в моменты случайного появления животных или, например, шелеста листьев на ветру.

Применение технологии распознавания лиц для задачи учета посещаемости предоставляет возможность не только проводить подсчет присутствующих, но также распознать каждого из них, выявив таким образом тех, кто отсутствует на мероприятии. Такое свойство может быть очень полезным при проведении лекций, семинаров и собраний, а наличие дополнительной функции журналирования — для подведения итоговых проверок посещаемости. При этом предпочтительнее использовать быстро внедряемые программно-аппаратные комплексы, которые позволят решить задачу в максимально кратчайшие сроки.

Данная работа носит преимущественно исследовательский характер, целью которой является определение оценки качества использования технологии распознавания лиц на изображениях и видеопотоке для решения задачи автоматизации учета посещаемости. Тем не менее, не исключена возможность использования полученного программного продукта в практических целях.

1. Обзор технологий распознавания образов на изображениях

На сегодняшний день существует две технологии распознавания лиц: двумерная и трехмерная.

Двумерная технология

Двумерная технология опирается на распознавание плоских изображений. Ввиду своего более раннего развития она не предполагает учета нескольких важных аспектов, чем объясняется невысокая точность распознавания при ее использовании. К таким аспектам относятся размер, выпуклость изображения или отдельных его частей (например, лба или подбородка), зависимость от яркости видимого света, что объясняет невысокую точность распознавания лиц в темноте или наоборот при слишком

¹ Одноплатный компьютер — компьютер, имеющий на одной плате процессор, оперативную и постоянную память, устройства ввода-вывода; это универсальное стандартизированное микропроцессорное решение, легко адаптируемое (тем или иным способом) для решения конкретной задачи [1].

ярком свете, и контрастности изображения. Кроме того, большое влияние на качество распознавания имеет положение головы относительно камеры (например, человека, стоящего к камере в профиль, распознать практически невозможно) и наличие макияжа или дополнительных предметов аксессуара таких как очки.

Трехмерная технология

Трехмерная технология, получившая свое развитие гораздо позже и относящаяся к биометрическим, учитывает данные важные аспекты и представляет собой гораздо более точную модель. Это объясняется тем, что признаки, содержащиеся в объемных образах, более устойчивы к изменениям внешности, поскольку используют индивидуальные биологические показатели человека, не меняющиеся со временем или с рядом других условий. Помимо этого, исключен риск ошибок распознавания при некорректном положении головы человека посредством возможности вращения объемного образа лица до положения строго анфас.

Существуют два подхода к трехмерному распознаванию лица: технология стерео, в идее которой лежит синтез двух или более плоских фотографий (все так же присутствует зависимость от освещенности фотографий), и технология освещения лица специальным невидимым структурированным светом. При этом собирается около 40 000 измерительных точек [2]. Поскольку в таких сканерах используются свои источники света, надежность результата применения этого оборудования в плохих условиях освещенности гарантирована.

2. Подход к распознаванию лиц, основанный на использовании нейронных сетей

Существует обширный ряд методов распознавания лиц, к которому относятся метод опорных векторов, линейный дискриминантный анализ, сравнение эластичных графов, скрытые Марковские модели и др. Однако одним из самых популярных является подход, основанный на использовании глубоких нейронных сетей.

Нейронная сеть представляет из себя математическую модель в виде программного и аппаратного воплощения, строящаяся на принципах функционирования биологических нейронных сетей. Сегодня такие сети активно используют в практических целях за счет возможности не только разработки, но и обучения. Их применяют для прогнозирования, распознавания образов, машинного перевода, распознавания аудио и т. д. [3]. Архитектура нейронной сети состоит из нейронов (вычислительных узлов), имеющих между собой связи их веса. Нейроны объединены в слои, которые в свою очередь бывают входными, выходными и «скрытыми» (находящимися между входным и выходным слоями).

Сумма входных сигналов, умноженных на соответствующие веса добавленным порогом активации, передается непосредственно на выход нейрона и рассматривается как выходной сигнал [4]. Каждый нейрон имеет порог активации, который устанавливает одно из двух возможных выходных значений нейрона: 0 или 1. На вход нейронной сети подается вектор числовых данных. После прохождения его по всем нейронам сеть выдает результирующее значение.

Глубокие нейронные сети представляют подвид нейронных сетей, в которых содержится один или более «скрытых» слоев (см. рис. 1).

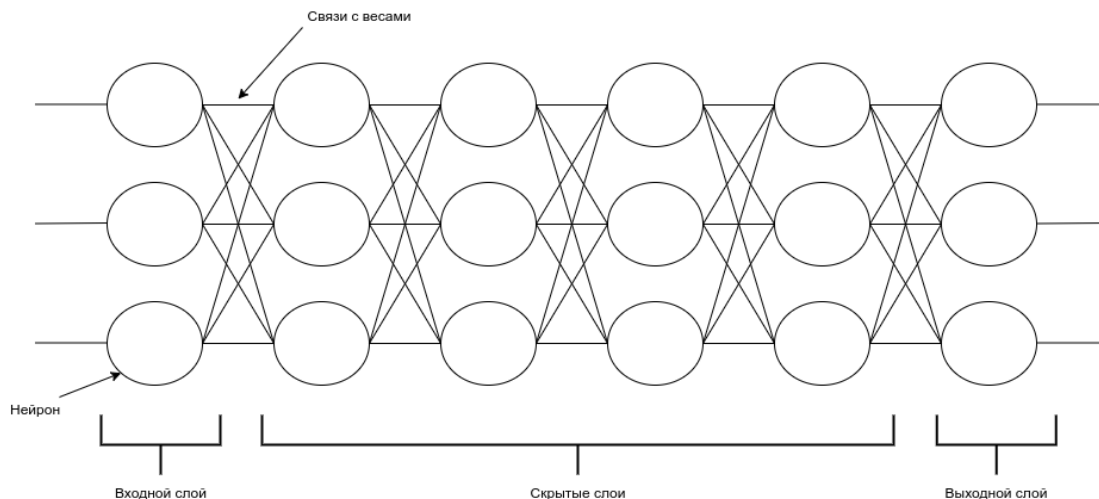


Рис. 1. Пример простой глубокой нейронной сети

Оценить вычислительную сложность алгоритмов вне вычислительных машин с параллельной архитектурой затруднительно. Тем не менее сообщается о достаточно эффективном использовании нейронных сетей в области анализа изображений лиц по трем направлениям: классификация людей по полу, непосредственно распознавание и определение эмоциональных выражений лиц. Однако в применении к распознаванию человека по изображению лица такой подход хорош для задач контроля доступа лишь небольшой группы лиц. Он обеспечивает непосредственное сравнение сетью самих образов, однако с увеличением числа классов время обучения и работы сети возрастает экспоненциально [5]. Поэтому для задачи поиска похожего человека в большой базе данных необходимо извлекать крайне компактный набор ключевых характеристик, на основе которых можно производить этот поиск.

Наиболее успешное применение в решении задач распознавания образов получили сверточные нейронные сети — *Convolutional Neural Networks*. Как показывает практика, сети, имеющие именно такую архитектуру, максимально эффективно справляются с данной задачей. Сверточная нейронная сеть получила свое название от операции свертки, которая предполагает, что каждый фрагмент изображения будет умножен на ядро свертки поэлементно, при этом полученный результат должен суммироваться и записаться в похожую позицию выходного изображения. Такая архитектура обеспечивает инвариантность распознавания относительно сдвига объекта, постепенно укрупняя «окно», на которое «смотрит» свёртка, выявляя всё более и более крупные структуры и паттерны в изображении [3] (см. Рис. 2. Архитектура сверточной нейронной сети [6]. 2).

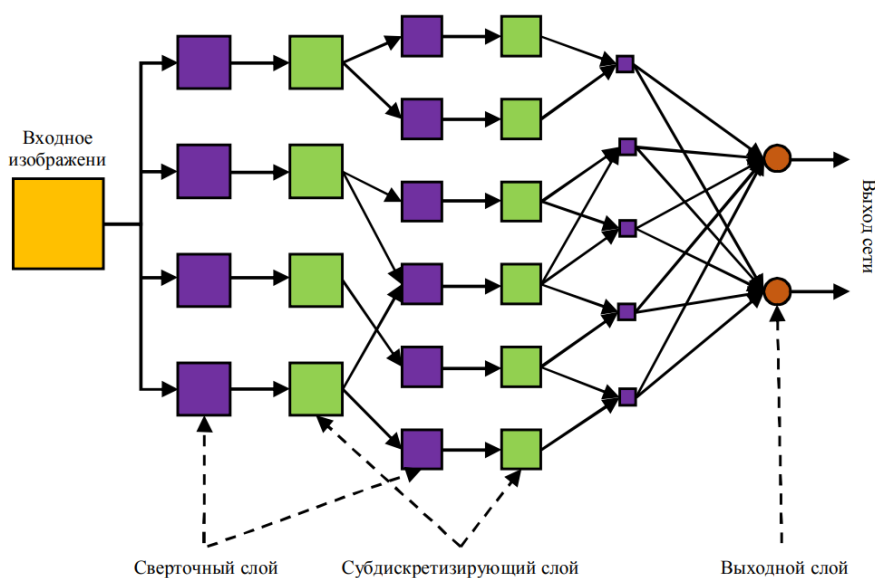


Рис. 2. Архитектура сверточной нейронной сети [6]. 2)

3. Обзор идей и существующих программных технологий для решения задачи распознавания лиц

Две идеи решения задачи

Существует две идеи решения задачи:

1. Разработка собственной нейронной сети с помощью библиотек программного обеспечения по машинному обучению.
2. Применение уже готовых нейронных сетей, имеющих специальную архитектуру и гарантирующих распознавание с некоторой точностью.

Первый способ является весьма трудоемким, поскольку подразумевает нетривиальные подзадачи итеративного подбора архитектуры, размера, количества нейронов, функции активации и многих других параметров нейронной сети, фильтрации и приведения входных данных к определенному виду и пр. Следует добавить, что точность распознавания разработанной нейронной сети может не иметь таких впечатляющих показателей, как уже готовых, и зачастую добиться высокой точности представляется достаточно ресурсоемкой задачей как в отношении времени, так и в отношении аппаратных возможностей оборудования.

Поскольку данная работа не имеет цели исследования возможностей различных нейронных сетей для решения задачи, гораздо целесообразнее и эффективнее использовать уже готовую нейронную сеть со специальной архитектурой, гарантирующей решение данной задачи с некоторой точностью.

В настоящее время имеется некоторый ряд программных библиотек с уже заранее подготовленными реализациями нейронных сетей в том числе и для задач распознавания лиц. К ним относятся такие как *OpenCV*, *FaceNet* и *dlib*. Далее кратко рассмотрены возможности каждой из них.

Описание библиотеки *OpenCV*

OpenCV – кроссплатформенная высоко оптимизированная библиотека компьютерного зрения с открытым исходным кодом, ориентированная на использование приложениями реального времени. Имеет интерфейсы для языков *C++*, *Python* и *Java*. Далее перечислены лишь некоторые программные возможности *OpenCV*:

1. Модуль обработки изображений, который включает в себя линейную и нелинейную фильтрацию изображений, геометрические преобразования изображений (изменение размера, аффинное и перспективное деформирование, общее переназначение на основе таблиц), преобразование цветового пространства, гистограммы и так далее [7].
2. Модуль анализа видео, который включает в себя оценку движения, вычитание фона и алгоритмы отслеживания объекта [7].
3. Обнаружение объектов и экземпляров предопределенных классов (например, лиц, глаз, кружек, людей, автомобилей и т. д.) [7].
4. Простой в использовании интерфейс для захвата видео, изображений и видеокодеков, а также простые возможности пользовательского интерфейса [7].

Кроме того, *OpenCV* имеет собственные реализации нейронных сетей, решающие задачу распознавания лиц на изображениях.

Описание системы распознавания лиц *FaceNet*

FaceNet – система распознавания лиц от сотрудников компании *Google*, основанная на использовании глубокой сверточной нейронной сети, которая предлагает решение трех основных задач:

1. Проверка на соответствие конкретного изображения лица одному из ранее обнаруженных (один и тот же ли человек на изображении) [8].
2. Непосредственно распознавание (установление личности по изображению лица) [8].
4. Поиск лица конкретного человека в базе лиц [8].

Мерой, определяющей степень схожести объектов, является расстояние в Евклидовом пространстве. При этом чем меньше значение этого расстояния, тем более похожими считаются объекты.

Несмотря на то, что *FaceNet* является самостоятельной системой, существует открытая библиотека для решения задач распознавания лиц под названием *OpenFace*, которая использует глубокую сверточную нейронную сеть, основанную на технологии *FaceNet*.

Описание библиотеки *dlib*

Dlib – программная библиотека с открытым исходным кодом для создания приложений машинного обучения и анализа данных на языке *C++*. Содержит в себе следующие алгоритмы, а также огромное множество других:

1. Алгоритмы машинного обучения.
2. Алгоритмы численных вычислений.
3. Алгоритмы для работы с графами.
5. Алгоритмы для работы с изображениями, в том числе обнаружение объектов и распознавание лиц.

Библиотека имеет полную подробную документацию и множество примеров кода на языках *C++* и *Python* и регулярно тестируется в системах *MS Windows*, *Linux* и *Mac OS X*. Соотношение строк кода модульного тестирования к строкам кода библиотеки составляет от 1 до 4 [9]. Все это вкуче с высокой скоростью работы алгоритмов является несомненными преимуществами перед другими подобными библиотеками и подталкивает разработчиков использовать именно этот интерфейс. Таким образом появились так называемые программные «обертки» над *dlib* – например, проект с открытым исходным кодом под названием *face_recognition*.

Обзор прикладного программного интерфейса *face_recognition*

Face_recognition появился на свет в марте 2017 года под лицензией *MIT*, автором проекта является *Adam Geitgey* [10]. Проект представляет собой набор простых функций на языке *Python*, позволяющих легко произвести кодирование и распознавание необходимых изображений как с использованием командной строки, так и в собственном программном обеспечении, написанном на языке *Python*. Интерфейс базируется на двумерной технологии распознавания и использует модель глубокой сверточной нейронной сети из библиотеки *dlib* (*cnn_face_detection_model_v1*) [11] и позволяет обнаружить и распознать сразу нескольких людей на изображении (см. рис. 3).

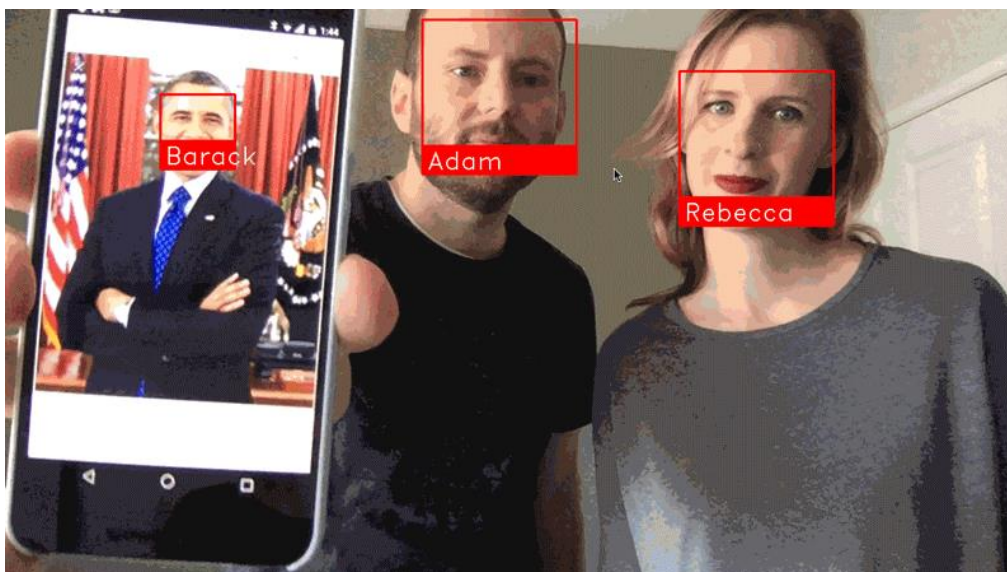


Рис. 3. Демонстрация распознавания нескольких людей на изображении [10]

Подключение функции захвата веб-камеры из библиотеки *OpenCV* позволяет использовать *face_recognition* даже в видеопотоке. Идея распознавания основана на кодировании входящего изобра-

жения или кадра видеопотока с помощью вызова функции *face_encodings* и сравнении его с остальными кодировками файлов, имеющимися в хранилище, с помощью вызова функции *compare_faces*, что можно описать всего несколькими строчками кода на языке *Python* [10].

В целом логику работы программы при использовании функций данного интерфейса можно описать простыми шагами:

1. Чтение входного изображения (или кадра видеопотока).
2. Поиск лица на изображении.
3. Кодирование изображения (в результате получается *encoding*, или кодировка – некоторый набор чисел, кодирующих каждый из отличительных признаков лица (например, ширина лба, расстояние между глазами и пр.)).
4. Сравнение данной кодировки с другими кодировками, уже имеющимися в системе (если сеть была предварительно обучена) и вычисление метрик (расстояний, или числовых мер близости между двумя кодировками). Соответственно, чем метрика меньше, тем более похожими считаются изображения, и наоборот. Как только наименьшая из всех метрик была найдена, запоминается индекс файла, который оказался наиболее похожим на входное изображение, и по нему возвращается имя этого файла.

По заявлению автора, данная модель прошла тестирование на наборе из 13000 изображений лиц под названием «*Labeled Faces in the Wild*» [12], по результатам которого точность распознавания достигла 99,38% [10].

Помимо обычного распознавания, *face_recognition* имеет функцию, позволяющую выделить на лице очертания глаз, носа, рта и подбородка человека, что может быть полезным при разработке приложений для создания цифрового макияжа на лице [10].

4. Обзор и обоснование выбора используемого программного и аппаратного обеспечения

Для реализации прототипа системы детектирования и распознавания лиц были применены следующие программные и аппаратные средства:

1. Высокоуровневый язык программирования общего назначения *Python*, не зависящий от какой-либо операционной системы, а также имеющий множество полезных библиотек, исчерпывающую документацию и легко читаемый код. Используемая версия языка: 3.6.
2. Ввиду возможности легкого интегрирования в *Python*-приложение и заявленной высокой точности распознавания прикладной программный интерфейс с готовой нейронной сетью *face_recognition*, рассмотренный в п. 3.4.6.
3. *PyQt4* — набор расширений графического фреймворка *Qt* для языка *Python*, позволяющий создавать кроссплатформенные интерфейсы приложений, работающие в системах *Linux* и других *UNIX*-подобных, *Mac OS X*, *Windows*.
4. *Qt Designer* — кроссплатформенная свободная среда для разработки интерфейсов с помощью графического фреймворка *Qt*.
5. Доступный на момент разработки одноплатный микрокомпьютер *Raspberry Pi 3 Model B+*. Несравненными преимуществами использования такого микрокомпьютера являются достаточно невысокая стоимость его покупки (около 35\$), при этом технические характеристики устройства незначительно уступают смартфону среднего ценового сегмента (64-битный процессор *Broadcom BCM2837B0* с четырьмя ядрами *Cortex A53*, графическим процессором *Videocore IV* и тактовой частотой 1.4 ГГц, 1 Гб оперативной и 14 Гб основной памяти, популярные стандарты беспроводной связи такие как *Wireless LAN* и *Bluetooth*) [13], миниатюрный размер: плата микрокомпьютера чуть больше кредитной карты и имеет размер 85.6*56*21 мм и вес в 45 гр, в связи с чем устройство легко помещается в ладони, а также поддержка различных операционных систем, в том числе и *Linux*, простота установки где бы то ни было и возможность подключения таких устройств как веб-камера, *USB*-накопитель, клавиатура, мышь,

монитор с *HDMI*-выходом и др. К недостаткам можно отнести невысокое быстродействие и возможный перегрев процессора.

6. Также в процессе разработки и тестирования времени работы системы использовался доступный ноутбук *ASUS X751L*.

5. Тестирование алгоритмов и анализ результатов

Были произведены тестирования работы системы с видеопотоком (при этом использовались: встроенная веб-камера ноутбука, камера от *Raspberry Pi*, веб-камера *CBR CW 555M*), а также на наборе квадратных изображений «*Face Research Lab London Set*» [14]. Набор представляет 1020 цветных фотографий 102 людей (по 10 фотографий на каждого человека) размером 1350*1350 пикселей (далее – *px*). Каждая из 10 фотографий представляет человека в определенном расположении к камере, из которых 4 фотографии отображают профиль (см. рис. 4). Как показало тестирование, процент обнаружения лица человека, находящегося на фотографии в профиль, равен 0. Исходя из этого результата было принято решение дальнейшие замеры времени работы и точности распознавания проводить на всех фотографиях, исключая профили, то есть фотографии, названия которых удовлетворяют маскам *_01.jpg, *_05.jpg, *_06.jpg, *_10.jpg.



Рис. 4. Первые 10 фотографий из набора «*Face Research Lab London Set*»

Следующим шагом в подготовке к тестированию было изменение размера каждой фотографии на нужный для конкретного теста. В результате были произведены замеры времени работы и точности распознавания на фотографиях размером 300*300 *px*, 200*200 *px*, 100*100 *px* и 90*90 *px* (границы изменения определены также результатами тестов, которые будут рассмотрены далее). Для решения этой подзадачи был написан скрипт на языке *bash*, который принимает на вход числовое значение необходимого размера (количество пикселей) и затем вызывает нужную функцию из дополнительной библиотеки для работы с изображениями *Imagemagick* [15].

Следующей подзадачей являлась необходимость разделения всего набора получившихся фотографий на три части: фотографии для обучения нейронной сети, фотографии для тестирования качества обучения и фотографии для оценки процента допуска ошибок при появлении в камере незнакомых людей, если нейронная сеть уже обучена. Смысл тестирования на последней части дает возможность понять, в каких случаях система распознает человека, фотографию которого нейронная сеть никогда не видела (то есть однозначно некорректно). Такая ошибка называется ошибкой I рода или *false positive*, в то время как ошибка, при которой известный системе человек распознается как неизвестный, называется ошибкой II рода или *false negative*. Для решения этой подзадачи вся выборка была разделена в соотношении 70%-30%. 30% были помещены в папку «*Unknown*», а оставшиеся 70% фотографий были использованы в качестве обучения и тестирования нейронной сети. Для обучения были взяты фотографии, названия которых удовлетворяют маске *_03.jpg, то есть те, на которых человек представлен строго анфас и без улыбки (одна фотография на человека). Все остальные фотографии (*_02.jpg, *_04.jpg, *_07.jpg, *_08.jpg, *_09.jpg) были помещены в папку для тестирования качества обучения сети. Таким образом в папке «*Train*» содержится 71 фотография, в папке «*Test*» — 355, а в папке «*Unknown*» — 186. Для проведения тестов на каждой из трех частей были написаны три соответствующих скрипта на языке *Python*.

В результате были произведены замеры количества корректно распознанных изображений для каждого из размеров и подсчитаны допущенные ошибки распознавания. Основные результаты сведены в отдельную таблицу (см. рис. 5) и построены соответствующие диаграммы, отображающие динамику изменения времени и количества ошибок распознавания известных и неизвестных системе лиц в зависимости от размера изображений (см. рис. 6, рис. 7, рис. 8).

ASUS X751L			
	Точность корректного распознавания при тестировании	Процент отсутствия ошибки распознавания неизвестного системе человека как известного	Общее время работы алгоритмов (обучение + тестирование)
90 px	92,9178470254957	38,7096774193548	2,07256572643916
100 px	96,8838526912181	46,7741935483871	2,1446165005366
200 px	98,5875706214689	74,7311827956989	2,71279185612996
300 px	98,0281690140845	67,741935483871	3,02157008250554

Raspberry Pi 3 Model B+			
	Точность корректного распознавания при тестировании	Процент отсутствия ошибки распознавания неизвестного системе человека как известного	Общее время работы алгоритмов (обучение + тестирование)
200 px	98,5875706214689	71,505376344086	8,89115488529205

Рис. 5. Сводная таблица, отображающая основные полученные в ходе тестов результаты

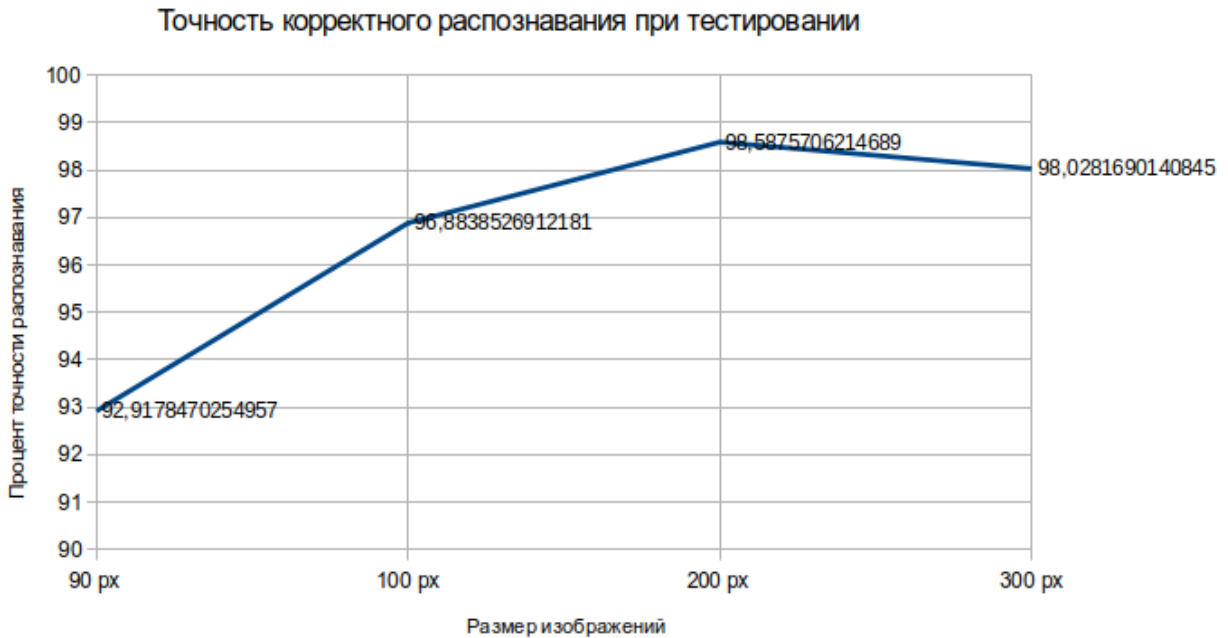


Рис. 6. График зависимости точности корректного распознавания лиц от размера входного изображения (на тестовых данных)



Рис. 7. График зависимости точности определения неизвестных системе лиц от размера входного изображения (на данных из папки «Unknown»)

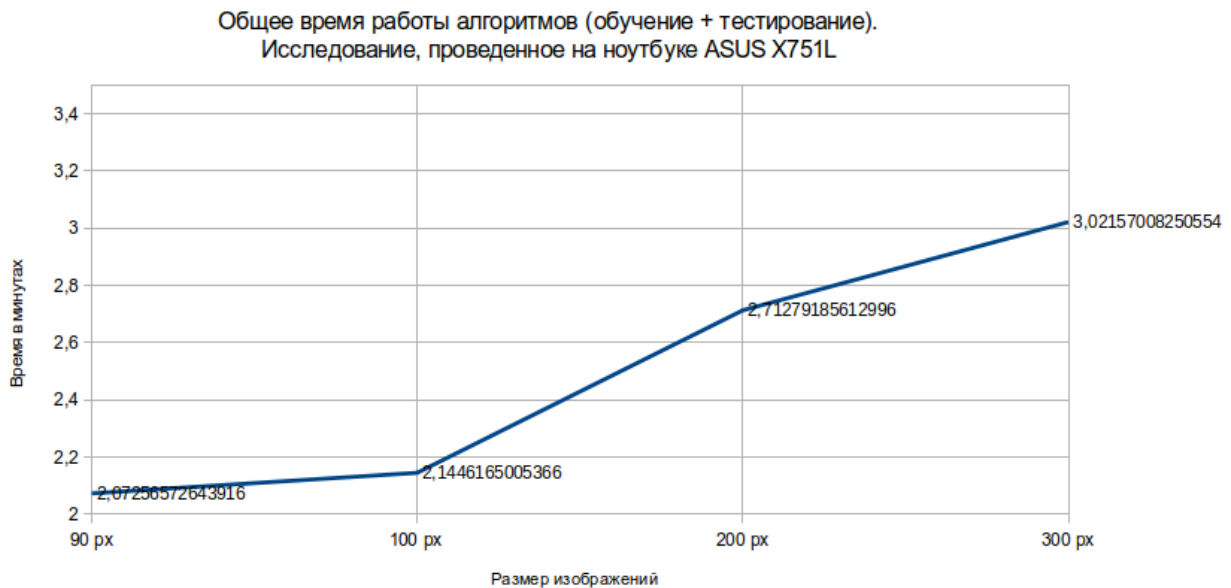


Рис. 8. График динамики времени работы алгоритмов обучения и проверки на тестовых данных в зависимости от размера входных изображений (в минутах)

Тестирование системы в целях определения оптимального размера изображений производилось на ноутбуке *ASUS X751L*. Как показали результаты тестирования, оптимальный размер изображений составляет 200×200 px, что определяет наименьшее количество ошибок распознавания при наименьшем времени работы системы (примерно 2,71 мин на ноутбуке). По результатам замера на *Raspberry Pi* время работы алгоритмов на изображениях такого размера занимает примерно 8,89 минут, что примерно в три раза медленнее, чем на ноутбуке. При этом:

1. Среднее время кодирования изображения составляет:
 - ~0,34 сек на ноутбуке.
 - ~0,93 сек на микрокомпьютере.
2. Среднее время поиска лица на изображении составляет:
 - ~0,04 сек на ноутбуке.

- ~0,38 сек на микрокомпьютере.
- 3. Среднее время поиска сходства обнаруженного лица с имеющимися в базе составляет:
 - ~0,0002 сек на ноутбуке.
 - ~0,008 сек на микрокомпьютере.
- 4. Общее время работы программы при кодировании изображений (= обучение сети) составляет:
 - ~0,58 мин на ноутбуке.
 - ~1,10 мин на микрокомпьютере.
- 5. Общее время работы программы при распознавании изображений (= тестирование обученной сети) составляет:
 - ~2,13 мин на ноутбуке.
 - ~7,78 мин на микрокомпьютере.

Из сказанного следует, что в работе будущей системы оптимально использовать изображения размером 200×200 пикс. Также было замечено, что при тестировании на изображениях размером 300×300 пикс нейронная сеть принимает во внимание слишком мелкие детали, в результате чего может обнаружить лицо даже в тех местах, где его нет (см. рис. 9). В связи с этим было принято решение не проводить исследования на изображениях большего размера. Также стоит отметить, что отдельно проведенный тест на определение нижней границы размера входного изображения показал, что при размере 70×70 пикс процент обнаружения лица на изображении равен 0.

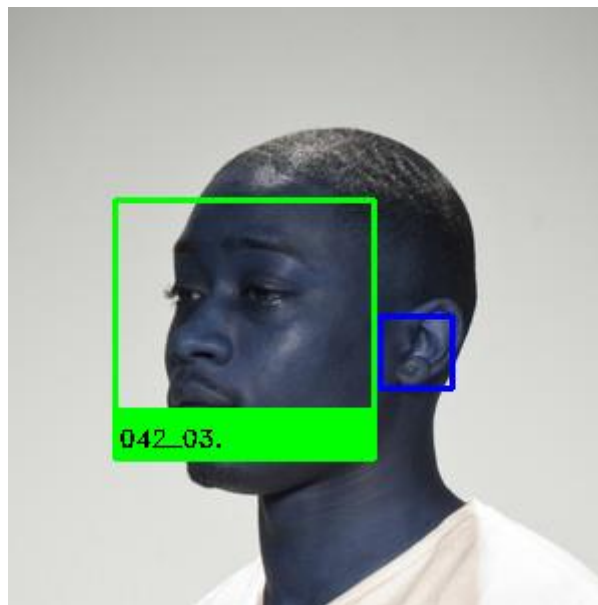


Рис. 9. Пример некорректного обнаружения лица на фотографии размером 300×300 пикс

Из всего вышеизложенного следует вывод, что разрабатываемая система имеет точность корректного распознавания лиц, равную 98,6%, и может быть пригодна для применения в практических целях, например, для автоматизации процесса учета посещаемости семинаров и лекций в университетах. Следует заметить, что при организации архитектуры разрабатываемой системы в виде модели «клиент-сервер» можно добиться распределения нагрузки в процессе обмена информацией между участниками, а также разделить части программного кода системы, отвечающие за распознавание и вывод информации, что значительно сократит время работы каждой из частей.

На данный момент разработан прототип системы с графическим интерфейсом, работающий как на ноутбуке ASUS X751L, так и на одноплатном микрокомпьютере Raspberry Pi 3 Model B+ и имеющий в себе следующие функции:

1. Загрузка и чтение фотографий лиц людей для последующего их сравнения.
2. Сравнение текущего изображения лица в видеопотоке с имеющимися фотографиями лиц в хранилище.

3. Отображение информации о распознанном человеке на интерфейсе системы. В случае обнаружения неизвестного лица система должна сохранять его изображение в хранилище для последующей регистрации и отображать соответствующую информацию на интерфейсе.
4. Ведение журналирования, полезного для обеспечения возможности проведения аналитики, как в окне интерфейса системы, так и с возможностью выгрузки информации в файл.

В будущем планируется организовать архитектуру системы в виде модели «клиент-сервер», а также получить ответы на дополнительный ряд вопросов, связанный с ускорением работы системы на других процессорах, оптимизацией алгоритма, а также с процессами внедрения данной системы на предприятиях.

Заключение

В рамках данного исследования была исследована оценка возможности применения технологии распознавания лиц на изображениях *face_recognition*, основанной на использовании глубокой нейронной сети, для решения задачи автоматизации учета посещаемости. Эксперименты показали эффективность работы данного подхода. Был проведен анализ существующих технологий и программных методов распознавания лиц на изображениях и в видеопотоке, анализ входных данных для тестирования, проведено тестирование, а также произведен анализ полученных результатов, из которого определено, что данный подход является эффективным при использовании входных изображений, имеющих размер 200*200 px. При этом точность распознавания составляет 98,6%.

Все разработанные нами алгоритмы были реализованы на языке *Python*. Весь исходный код доступен под лицензией *GNU General Public License v3.0* по ссылке [16].

Список литературы

1. Кругляк К. Одноплатные компьютеры для встраиваемых систем. // Журнал «Современные технологии автоматизации». 2003. №4. — С. 6-17.
2. Тест-эквайринг Иванов. 3D технология распознавания лиц. — 2016. — [Электронный ресурс]. URL: <https://sec-group.ru/blog/tryohmernye-tehnologii-v-raspoznovanii-lic/>.
3. «Центр 2М». Нейронные сети: распознавание образов и изображений с помощью ИИ. — 2018-2020. — [Электронный ресурс]. URL: <https://center2m.ru/ai-recognition>.
4. Элементарное введение в технологию нейронных сетей с примерами программ / Р. Тадеусевич, Б. Боровик, Т. Гончаж, Б. Леппер. — М.: ООО «Научно-техническое издательство «Горячая линия-Телеком», 2011. — С. 403.
5. Гришин А. И. Исследование эффективности методов обработки изображений в системах распознавания лиц: Выпускная квалификационная работа магистра направления 080500.68 «Бизнес-информатика» / Южно-Уральский государственный университет. — Челябинск, 2016. — С. 75.
6. Федорова А. А. Распознавание английского текста сверточной нейронной сетью: Выпускная квалификационная работа бакалавра направления 010400.62 «Прикладная математика и информатика» / Санкт-Петербургский Государственный Университет. — СПб., 2016. — С. 30.
7. OpenCV dev team. Introduction – OpenCV 2.4.13.7 documentation. — 2018. — [Электронный ресурс]. URL: <https://docs.opencv.org/2.4.13.7/modules/core/doc/intro.html>.
8. F. Schroff, D. Kalenichenko, J. Philbin. Facenet: A unified embedding for face recognition and clustering // Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition. 2015.
9. Dlib C++ Library. — 2019. — [Электронный ресурс]. URL: <http://dlib.net/>.
10. Ageitgey. Face_recognition: The world's simplest facial recognition api for Python and the command line. — 2020. [Электронный ресурс]. URL: https://github.com/ageitgey/face_recognition.
11. Ageitgey. Face_recognition: api.py at master. — 2020. — [Электронный ресурс]. URL: https://github.com/ageitgey/face_recognition/blob/master/face_recognition/api.py.

12. LFW Face Database : Main. — 2018. — [Электронный ресурс]. URL: <http://vis-www.cs.umass.edu/lfw/>.
13. Raspberry Pi. Buy a Raspberry Pi 3 Model B+. — 2018. — [Электронный ресурс]. URL: <https://www.raspberrypi.org/products/raspberry-pi-3-model-b-plus/>.
14. Face Research Lab London Set. — 2017. [Электронный ресурс]. URL: https://figshare.com/articles/Face_Research_Lab_London_Set/5047666.
15. Mike. Batch Resize Images using Linux Command Line and Imagemagick. — 2017. — [Электронный ресурс]. URL: <https://guides.wp-bullet.com/batch-resize-images-using-linux-command-line-and-image-magick/>.
16. Репозиторий GitHub. — [Электронный ресурс]. URL: <https://github.com/trnkv/facerec>.