

Раздел 4 АВТОМАТИЗАЦИЯ, АНАЛИЗ И ОБРАБОТКА ИНФОРМАЦИИ, УПРАВЛЕНИЕ ТЕХНОЛОГИЧЕСКИМИ ПРОЦЕССАМИ В СОЦИАЛЬНЫХ И ЭКОНОМИЧЕСКИХ СИСТЕМАХ

УДК 303.732.4

DOI: 10.34046/aumsuomt96/21

ИССЛЕДОВАНИЕ МЕТОДОВ И АЛГОРИТМОВ КОМПЬЮТЕРНОГО ЗРЕНИЯ НА ОСНОВЕ СВЁРТОЧНЫХ И РЕКУРРЕНТНЫХ НЕЙРОННЫХ СЕТЕЙ

Н.А. Полковникова, кандидат технических наук, доцент

В статье произведено исследование методов и алгоритмов компьютерного зрения на основе глубоких свёрточных и рекуррентных нейронных сетей. Обработка изображений с целью их распознавания является одной из центральных и практически важных задач при создании систем искусственного интеллекта. Распознавание лиц на изображениях является классической задачей в компьютерном зрении и самым современным трендом в интеллектуальных системах безопасности и контроля доступа, в авторизации пользователей, организации видеоконференций, робототехнике и биометрии. Рассмотрена эволюция подходов и алгоритмов к решению задач обработки, распознавания и локализации лиц на изображениях с использованием алгоритма Виола-Джонса, свёрточных и рекуррентных нейронных сетей. Рассмотрены технологии восстановления трёхмерного изображения лица по исходному 2D-изображению нейросетевыми методами. Приведены результаты компьютерных экспериментов работы программы на основе свёрточной и рекуррентной нейронной сети для обнаружения лица, определения возраста, эмоционального состояния человека и наличия маски на лице в видеопотоке реального времени в среде Matlab R2020a. Практическое применение полученных результатов может быть использовано в дистанционном обучении при идентификации личности курсантов и студентов.

Ключевые слова: распознавание образов, компьютерное зрение, глубокое машинное обучение, алгоритм Виола-Джонса, признаки Хаара, свёрточные и рекуррентные нейросети, эмоциональный искусственный интеллект.

The article made a study of methods and computer vision algorithms based on deep convolutional and recurrent neural networks. Image processing for the purpose of their recognition is one of the central and practically important tasks in creation of artificial intelligence systems. Face detection and recognition in images is a classic task in computer vision and the most modern trend in intelligent security and access control systems, in user authorization, in organizing video conferencing, in robotics and biometrics. Considered evolution of approaches and algorithms to problems solving of processing, recognition and localization of faces in images using Viola-Jones object detection algorithm, convolutional and recurrent neural networks. Technologies for the restoration of a three-dimensional image of a face from an initial 2D image by neural methods are considered. The results of computer experiments based on a convolutional and recurrent neural network for face detection, age determination, face emotional state and mask detection in real-time video stream are implemented in Matlab R2020a. Practical application of the obtained results can be used in distant education in identifying the personality of cadets and students.

Keywords: pattern recognition, computer vision, deep machine learning, Viola-Jones object detection, Haar-like features, convolutional and recurrent neural networks; emotional artificial intelligence.

За последние годы всё более широкое распространение получают биометрические системы идентификации человека. Традиционные системы идентификации требуют знание пароля, наличия ключа, идентификационной карточки либо иного идентифицирующего документа, который можно забыть, потерять или подделать. В отличие от них биометрические системы основываются на уникальных биологических характеристиках человека, которые трудно подделать и которые однозначно определяют конкретного человека. К таким характеристикам относятся отпечатки пальцев, форма ладони, узор радужной оболочки, изображение сетчатки глаза, голос и лицо

человека. Обработка изображений с целью их распознавания является одной из центральных и практически важных задач при создании систем искусственного интеллекта (СИИ). Компьютерное зрение стало ключевой технологией искусственного интеллекта во многих областях и разрабатывается для:

- беспилотных автомобилей (Tesla, Volvo, Audi, BMW, Mercedes-Benz, Google, Yandex, Alphabet Waymo, Baidu Apollo, Uber);
- безэкипажных морских судов (Kongsberg, Rolls-Royce, Hyundai, Transas);
- интеллектуальных диалоговых систем (Apple Siri, Amazon Alexa, Google Assistant, Яндекс)

Алиса, Samsung Bixby, Microsoft Cortana, Mail Group Маруся) и многих других направлений [1-5].

Распознавание лиц является классической задачей в компьютерном зрении и самым современным трендом в интеллектуальных системах безопасности и контроля доступа, в авторизации пользователей, организации видеоконференций, робототехнике, биометрии и т. п. Распознавание лиц является частью биометрического распознавания. Биометрия – это набор технических приёмов распознавания человека путём измерения и анализа его уникальных физических и поведенческих характеристик.

Эволюция методов и алгоритмов распознавания лиц. Первая биометрическая система для описания и идентификации лиц разработана английским учёным Francis Galton в 1888 г. и использовалась в криминалистике. Затем в 1964 г. американский учёный W.W. Bledsoe разработал первую систему автоматического распознавания лиц на базе данных снимков по методу кортежей. В 1973 году японский учёный-программист Т. Kanade на основе базы данных из 850 оцифрованных фотографий со Всемирной выставки в Суите (Япония) разработал программу, которая могла извлекать черты лица (нос, рот и глаза) без участия человека.

За последние годы было предложено множество алгоритмов обработки, распознавания и локализации лиц: цепи Маркова, фильтры Габора («Gabor wavelets», 1968 г, Dennis Gabor), метод главных компонент («principle component analysis»), метод собственных лиц («Eigenfaces»), метод «Fisherfaces», алгоритм Виолы-Джонса, алгоритм Лукаса-Канаде, свёрточные нейронные сети и т. д. В настоящее время биометрия используется в смартфонах, в банковском деле (например, удалённо подтверждать личность клиента для выдачи кредита), маркетинге (программы лояльности), социальном управлении (контроль сотрудников на рабочем месте, учёт рабочего времени) и др. Например, Apple использует технологию Face ID, а компания OnePlus – технологию Face Unlock. В 2020 году компания Apple запатентовала технологию распознавания лиц по рисунку вен под кожей человека, что сделает невозможной разблокировку устройства людьми с похожей на владельца внешностью, манекеном и даже близнецами. Данная технология будет реализована путём просвечивания кожи пользователя инфракрасным датчиком устройства, таким образом технология Apple Face ID может перейти на совершенно новый уровень безопасности. В компа-

нии Huawei разработана технология, которая позволяет по лицу человека довольно точно измерять кровяное давление с использованием фронтальной камеры смартфона. В компании Baidu распознавание лиц используется вместо ID-карт для обеспечения доступа в офис, а при повторном пересечении границы в ОАЭ достаточно только посмотреть в веб-камеру. В медицинском научном журнале Nature Medicine сообщается, что в 2019 году учёными разработана система искусственного интеллекта, способная по фотографии человека диагностировать 216 редких наследственных заболеваний с высокой точностью.

Распознавание человека по изображению лица выделяется среди биометрических систем тем, что, во-первых, не требует специального дорогостоящего оборудования (достаточно наличие видеокamеры), во-вторых, отсутствует физический контакт человека с устройствами (нет необходимости ни к чему прикасаться или ждать срабатывания системы). В большинстве случаев достаточно пройти перед камерой несколько секунд.

Существуют множество признаков, по которым можно распознать человека, т.н. «биометрические модальности» («biometric modality») – биометрические характеристики, применяемые в биометрическом процессе. Самые точные биометрические модальности: ДНК, радужная оболочка глаза, отпечатки пальцев, отпечаток ладони и рисунок вен на руке. Менее точными биометрическими модальностями являются голос, лицо, походка, манера набора текста на клавиатуре (клавиатурный почерк) и т.д. На рис. 1 представлены сравнительные оценки основных ошибок аутентификации для различных биометрических модальностей.

На рис. 1а показано, что отпечатки пальцев, радужная оболочка глаза и лицо являются наиболее надёжными биометрическими параметрами и обладают высокой уникальностью. К недостаткам распознавания человека по изображению лица является отсутствие 100% надёжности идентификации. При возникновении ошибок ложного доступа не зарегистрированная в базе данных личность получает доступ к частной информации, что снижает безопасность системы. В случае ложного отказа в доступе возникает проблема с удобством использования системы, когда легитимным пользователям отказывается в доступе к приложению или предписывается пройти дополнительную проверку. Это не влияет на безопасность, но причиняет беспокойство пользова-

телу и может иметь другие последствия, препятствуя человеку заниматься своими делами. Поэтому нужен определенный компромисс при выборе уровня ошибок ложного доступа и ложного

отказа в доступе. В системах, где требуется высокая надёжность, применяют комбинирование нескольких биометрических методов.

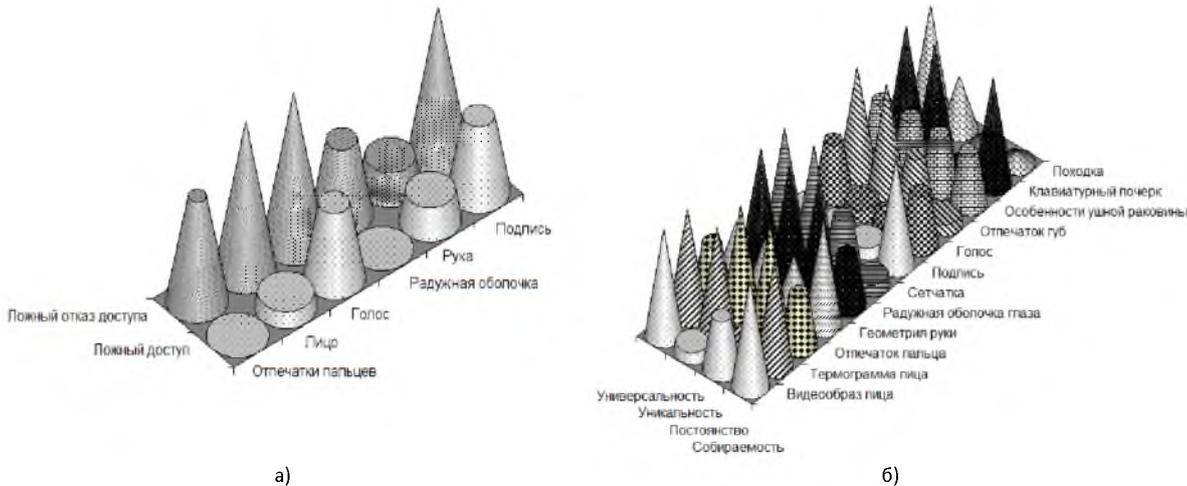


Рисунок 1 – Сравнительные оценки основных ошибок аутентификации для различных биометрических модальностей: а) оценка ошибок аутентификации для различных биометрических модальностей; б) оценка качественных свойств биометрических модальностей

Техника идентификации личности, основанная на распознавании лица, отличается от распознавания по другим биометрическим показателям (радужная оболочка глаза, отпечаток пальца). Преимуществом идентификации личности по лицу является то, что физический контакт с устройством не производится. Это делает данную технологию наиболее приемлемой для массового применения. Сложность задачи распознавания лиц можно обосновать следующими причинами:

1) лицо человека – это динамический объект, имеющий высокую степень вариативности (изменяемости) во внешнем виде (например, по форме, мимике и эмоциональному состоянию, цвету кожи, усов, бороды, возрасту, причёске и т.п.);

2) различные параметры освещённости при съёмке, определенные типом и направлением источника света;

3) фрагментарное перекрытие лиц другими объектами (очки, маска и т.д.);

4) необходимость локализации и распознавания лиц, имеющих произвольные положения в пространстве (различные ракурсы).

При выборе биометрической системы нужно руководствоваться требованиями, которые часто несовместимы друг с другом. Например, система должна быть безопасной, т.е. обеспечивать высокую точность идентификации и в то же время должна быть удобной для пользователей. Точность аутентификации оценивается с помощью количественных параметров таких, как уровень ошибок, которые совершает система при работе. Основными ошибками биометрических систем яв-

ляются принятие «чужого» пользователя за «своего» (ошибка ложного доступа) и отказ «своему» пользователю (ошибка ложного отказа доступа).

Задачи распознавания лиц и классы решаемых задач. В общем виде все задачи распознавания делятся на следующие категории:

- начальная идентификация («initial identification»);
- выделение сущностей («extraction»);
- отслеживание («tracking»);
- повторное распознавание («reidentification»).

Задачи распознавания лиц можно разделить на две категории: идентификация и верификация. Идентификация – это когда есть некоторая база, в которой множество лиц (классов) и каждое лицо относится к определённому человеку (объекту). Когда поступает новая фотография, система должна выбрать из базы человека, к которому эта новая фотография относится либо выдать заключение, что этот объект не относится к известным классам. Таким образом, идентификация – это выбор конкретного человека из множества людей.

В задаче верификации объект заявляет, что он принадлежит некоторому известному системе классу, а система подтверждает или опровергает это заявление. Другими словами, верификация – это когда в базе находятся фотографии только одного человека, а при поступлении новой фотографии, система должна определить: на фото этот же человек или другой. В частности, верификация используется в смартфонах, когда устройству нужно распознать владельца для разблокировки.

Многие компании активно внедряют liveness-технологии для определения живой ли человек перед камерой устройства (глаза открыты или закрыты, система просит человека улыбнуться, моргнуть или поднести камеру, или смартфон ближе к лицу), что требует разработки всё более совершенных алгоритмов распознавания в реальном времени.

При решении задачи распознавания лиц возникают две категории ошибок:

1) ошибки первого рода (type I error, «misdetection») – ситуации, когда человек есть в базе, но система его не опознала и в доступе отказано (в биометрии FRR – «false reject rate»);

2) ошибки второго рода (type II error, «false alarm») – ситуации, когда человека нет в базе, но он опознаётся системой как человек присутствующий в базе и ему разрешается доступ (в биометрии FAR – «false access rate»).

Перед тем, как решать задачу распознавания лиц с помощью нейронных сетей необходимо определить, к какому классу относится решаемая задача. Класс задач «один к одному» означает, что есть один объект на входе и один объект на выходе. На вход поступает изображение, а на выходе определяется класс изображения (рис. 2а).

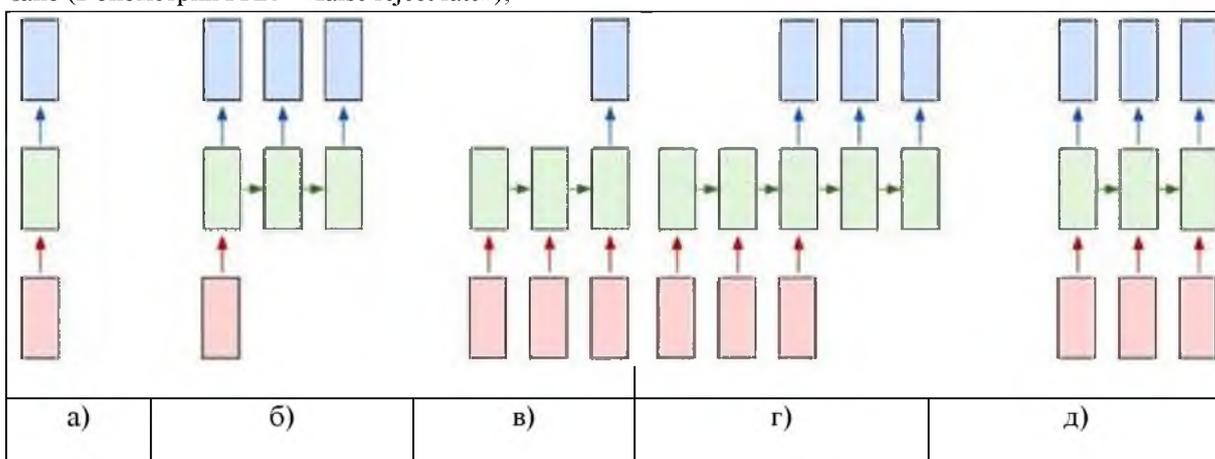


Рисунок 2 – Классы решаемых задач с помощью нейросетей: а) один к одному; б) один ко многим; в) многие к одному; г) многие ко многим; д) многие ко многим

Например, сравнение лица с фотографией в документах. Данный класс задач наиболее сложен, поскольку система никогда не сталкивалась с изображением лица данного человека, имеются возрастные или другие изменения лица и качество фотографии, как правило, хуже, чем изображение лица, снятого камерой. Класс задач «один ко многим», например, на вход поступает изображение, а на выходе словесное описание, что на изображении (рис. 2б). Класс задач «многие к одному», например, на вход поступает видео, а на выходе определяется человек, кто на видео (рис. 2в). Класс задач «многие ко многим», например, на вход поступают изображения, а на выходе определяются объекты на изображениях (рис. 2г). Другой пример класса задач «многие ко многим» – на вход поступает видео, каждый кадр которого классифицируется для дальнейшего анализа (рис. 2д).

Обычно в большой базе данных ($10^4 - 10^7$ изображений) требуется найти изображения, наиболее похожие на заданное (класс задач «один ко многим»). Поиск должен быть произведён за приемлемое время. Одно из решений состоит в хранении в базе данных небольших

наборов заранее извлечённых ключевых признаков, максимально характеризующих изображение. К решению класса задач «один ко многим», прежде всего, относятся метод главных компонент («principle component analysis») и метод собственных лиц («Eigenfaces»). Коэффициенты, полученные разложением входного изображения на главные компоненты, в ранних методах использовались для сравнения изображений путём вычисления Евклидова расстояния, а в более совершенных методах – на основе метрики Махаланобиса с использованием Гауссовского распределения.

Распознавание лиц на основе метода Виолы-Джонса. Задачу распознавания лиц вполне успешно решали алгоритмом Виола-Джонса (Viola-Jones) – основополагающим алгоритмом для поиска объектов на изображении в реальном времени. Алгоритм разработан Паулом Виолой (Paul Viola) и Майклом Джонсом (Michael Jones) в 2001 году и основан на использовании техники «скользящего окна» и каскадов Хаара в качестве признаков. Техника «скользящего окна» («sliding window») заключается в том, что рамка, размером меньшим, чем исходное изображение, двигается с некоторым шагом по изображению, и с помощью

каскадов Хаара определяет, есть ли в рассматриваемом окне лицо. Обученный каскад Хаара, принимая на вход изображение, определяет, есть ли на нем искомый объект, т. е. выполняет задачу классификации, разделяя входные данные на два класса (есть искомый объект, нет искомого объекта) [7, 8]. Метод Виолы-Джонса в настоящее время является одним из лучших по соотношению показателей: эффективность распознавания/скорость работы. Также детектор имеет очень низкую вероятность ложного срабатывания. Перед идентификацией лица алгоритм преобразует изображение в градации серого цвета для сокращения данных и упрощения обработки [9]. Алгоритм Виолы-Джонса работает следующим образом:

Шаг 1. Исходное цифровое цветное трёхканальное (RGB) изображение преобразуется в градации серого цвета (одноканальное изображение с $2^8 = 256$ уровнями яркости) и превращается в матрицу яркости, т.е. интегральное представление изображения. Интегральное представление позволяет быстро рассчитывать суммарную яркость произвольного прямоугольника на данном изображении, причём какой бы прямоугольник не был, время расчёта неизменно. Интегральное представление изображения – это матрица, совпадающая по размерам с исходным изображением. В каждом элементе её хранится сумма интенсивностей всех пикселей, находящихся левее и выше данного элемента. Элементы матрицы рассчитываются по следующей формуле:

$$L(x, y) = \sum_{i=0, j=0}^{i \leq x, j \leq y} I(i, j), \quad (1)$$

где $I(i, j)$ – яркость пикселя исходного изображения. Каждый элемент матрицы $L(x, y)$ представляет собой сумму пикселей в прямоугольнике от $(0, 0)$ до (x, y) , т.е. значение каждого пикселя (x, y) равно сумме значений всех пикселей левее и выше данного пикселя (x, y) . Расчёт матрицы занимает линейное время, пропорциональное числу пикселей в изображении, поэтому интегральное изображение просчитывается за один проход.

Шаг 2. На полученную матрицу яркости накладывается одна из квадратных масок (признаки Хаара). Признак Хаара является набором прямоугольных областей изображения, примыкающих друг к другу и разделенных на две группы. Возможных признаков Хаара множество: разнообразные комбинации областей разной ширины и высоты с разными позициями на изображении. Маска «сканирует» изображение, меняя положение и размер. Признак f – это отображение $f: X \rightarrow D_f$, где X – множество исследуемых объектов, D_f – множество допустимых значений признака f . Если заданы признаки f_1, \dots, f_n , то вектор признаков $x = (f_1(x), \dots, f_n(x))$ называется признаковым описанием объекта $x \in X$. Признаковые описания допустимо отождествлять с самими объектами, при этом множество $X = D_{f_1} * \dots * D_{f_n}$ называют признаковым пространством. Примеры признаков Хаара и их расположение на изображении показаны на рис. 3.

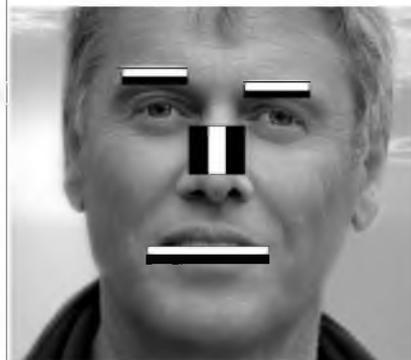
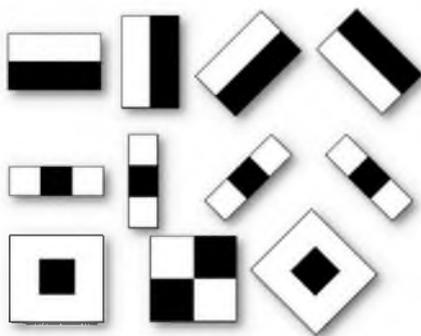


Рисунок 3 – Признаки Хаара и их расположение на изображении

Признаки Хаара используют факт того, что на изображениях лиц для бровей характерны тёмные горизонтальные линии, носовая перегородка – более светлая вертикальная линия и т.д.

Шаг 3. Значение признака равняется разности суммы пикселей в белом прямоугольнике и в чёрном прямоугольнике. Для их вычисления используется понятие интегрального изображения, рассмотренное выше. Таким образом, цифровые

значения яркости из тех ячеек матрицы, которые попали под белую часть маски суммируются, а те значения, что попали под чёрную часть вычитаются. Если хотя бы в одном из случаев разность белых и чёрных областей оказалась выше определенного порога, то эта область изображения подлежит дальнейшей обработке. Если указанное условие не выполняется, то данная область изоб-

ражения не содержит лица. Но для задачи обнаружения лица, многие признаки бесполезны, поэтому для того, чтобы выбрать полезные признаки применяется алгоритм AdaBoost («Adaptive Boosting») – алгоритм усиления классификаторов, путём их объединения, разработанный Й. Фройндом (Yoav Freund) и Р.Е. Шапире (R.E. Schapire). Алгоритм AdaBoost комбинирует большое число «слабых» классификаторов («weak classifiers») с целью создания одного «сильного» классификатора («strong classifier»), что позволяет решить задачу обнаружения лица намного эффективнее. «Слабый» классификатор – это функция $h_j(x)$, которая принимает на вход изображение, вычисляет значение соответствующего ей признака Хаара для этого изображения и сравнивает это значение с порогом, возвращая либо 0, либо 1:

$$h_j(x) = \begin{cases} 1, & \text{если } p_j f_j(x) < p_j \theta_j, \\ 0, & \text{иначе} \end{cases} \quad (2)$$

где $h_j(x)$ – слабый классификатор; x – входное изображение, $f_j(x)$ – значение соответствующего признака Хаара для изображения x ; θ_j – порог (вещественное число); p_j – изменяет направление знака неравенства, если значение признака будет отрицательным и указывается для каждого признака отдельно (направление знака неравенства); j – признак объекта на изображении, имеет структуру вида: $j = \{\text{маска, положение, размер}\}$. Задача «слабого» классификатора – определять присутствие объекта в более 50% случаев.

Для алгоритма AdaBoost необходимо заранее подготовить тестовую выборку из I изображений, содержащих искомым объект и n изображений не содержащих искомым объект. Тогда количество всех тестовых изображений будет составлять $X = [x_1, x_2, \dots, x_n]$, где X – множество всех тестовых изображений. При этом для каждого изображения заранее известно присутствует искомым объект или нет, что отражено во множестве Y , $Y = [y_1, y_2, \dots, y_n]$, где

$$y_i = \begin{cases} 1, & \text{объект присутствует на изображении } x_i, \\ 0, & \text{иначе} \end{cases} \quad (3)$$

В процессе обучения алгоритма AdaBoost создаётся «сильный» классификатор $H(x)$, состоящий из T «слабых» классификаторов:

$$H(x) = \begin{cases} 1, & \text{если } \sum_{t=1}^T a(t) \times h_j(t) \geq \frac{1}{2} \sum_{t=1}^T a(t), \\ 0, & \text{иначе} \end{cases} \quad (4)$$

где $t = [1 \dots T]$ – значения «слабых» классификаторов, вычисляемых на этапе обучения из базы изображений.

Алгоритм AdaBoost перебирает все возможные «слабые» классификаторы и выбирает те, которые допускают меньше всего ошибок. После

выбора очередного «слабого» классификатора весовые коэффициенты перераспределяются так, что неверно классифицированные изображения начинают сильнее влиять на значение ошибки. После применения алгоритма AdaBoost количество признаков значительно сокращается (с 16000 до 6000). Но в большинстве случаев на фотографиях лицо занимает малую область, и чтобы не выполнять прогон по всем признакам ненужные области, ввели понятие «каскад классификаторов». После этого признаки группируются по стадиям: первые стадии содержат малое количество признаков, в каждой последующей стадии их количество будет увеличиваться.

Шаг 4. Повторяем с шага 2 уже с новой маской – но только в той области изображения, которая прошла первое испытание. Область, которая прошла все стадии, является лицом.

На рис. 4 представлен пример работы алгоритма Виолы-Джонса в среде Matlab. В качестве исходного изображения загружена цветная фотография, которая на первом шаге работы алгоритма была преобразована в градации серого цвета, а затем выполнено распознавание лиц на изображении.

В данном примере в среде Matlab с помощью пакета Computer Vision Toolbox использован предобученный детектор FrontalFaceCART, вызов которого осуществляется по умолчанию, либо следующим образом:
faceDetector=vision.CascadeObjectDetector('FrontalFaceCART');

При этом размер изображений, использованных при обучении детектора FrontalFaceCART, составляет 20×20 пикселей:

$$\text{faceDetector.MinSize} = [20 \ 20];$$

Помимо детектора FrontalFaceCART пакет Computer Vision Toolbox содержит следующие предобученные детекторы по алгоритму Виолы-Джонса: FrontalFaceLBP, UpperBody, EyePairBig, EyePairSmall, LeftEye, RightEye, LeftEyeCART, RightEyeCART, ProfileFace, Mouth, Nose.

Однако, алгоритм Виола-Джонса изначально разработан для идентификации лиц в положении анфас или под небольшим углом (примерно до 30°). Если угол наклона больше 30° или есть лица, смотрящие вверх, вниз или в сторону (положение профиль), то вероятность обнаружения резко падает. Указанная особенность метода не позволяет обнаружить лицо человека, повернутое под произвольным углом (на рис. 4 лица без рамки), что в значительной мере затрудняет (требуются новые классификаторы) или делает невозможным использование алгоритма Виола-Джонса

в современных производственных системах с учётом растущих потребностей. Другой недостаток

алгоритма заключается в том, что заранее подготовить размеченную выборку изображений не представляется возможным.

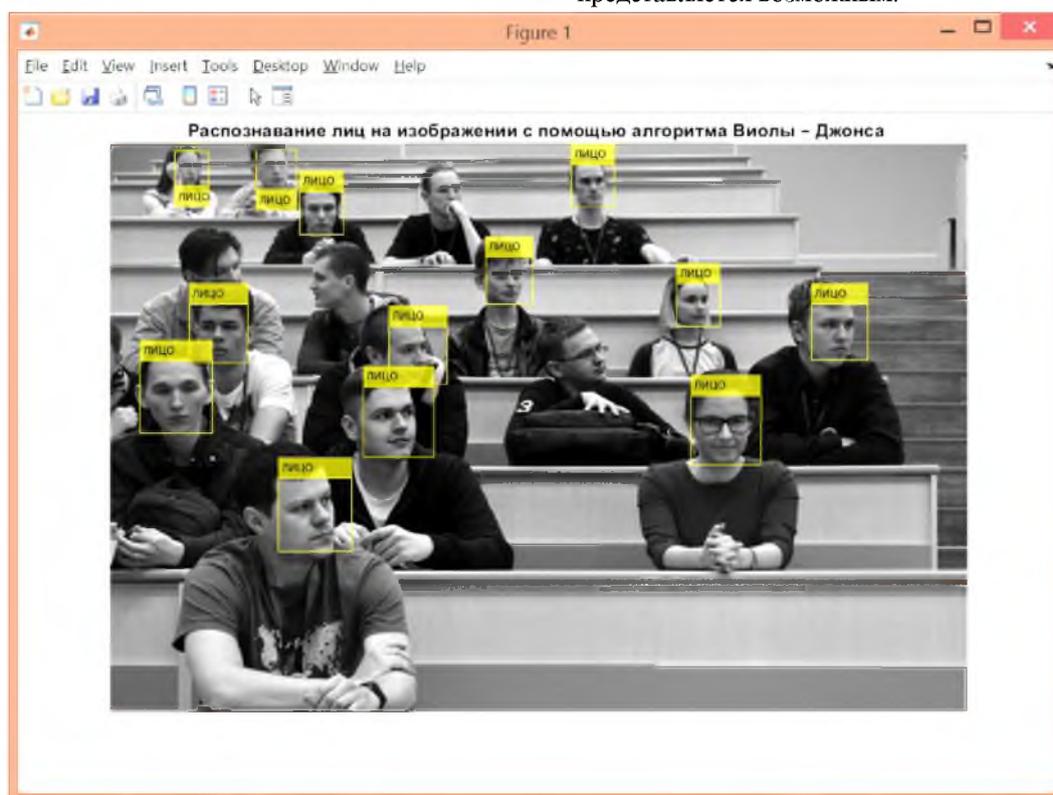


Рисунок 4 – Результат распознавания лиц с помощью алгоритма Виолы-Джонса

Применение другого метода – собственных лиц («Eigenfaces») – требует идеализированных условий съёмки и получения изображений: единые параметры освещённости, нейтральное выражение лица, отсутствие помех и перекрытие лица другими объектами. При несоблюдении этих условий главные компоненты не будут отражать межклассовые вариации. Например, при различных условиях освещённости метод собственных лиц практически неприменим, поскольку первые главные компоненты преимущественно отражают изменения освещения. При соблюдении идеализированных условий точность распознавания с использованием данного метода может достигать значения выше 90%. Поскольку в реальности идеализированных условий достигнуть затруднительно, в настоящее время задача распознавания лиц решается с помощью технологий глубокого обучения: свёрточных и рекуррентных нейронных сетей [10, 11].

Обобщённый алгоритм распознавания лиц. Распознавание изображений пересекается с распознаванием образов. Такие задачи не имеют точного аналитического решения, при этом требуется выделение ключевых признаков, характеризующих зрительный образ, определение относительной важности признаков путём выбора их

весовых коэффициентов и учёт взаимосвязей между признаками. Изначально эти задачи выполнялись человеком-экспертом вручную, путём экспериментов, что занимало много времени и не гарантировало качества. В новых методах выделение ключевых признаков осуществляется путём автоматического анализа обучающей выборки, но часть информации о признаках задаётся вручную. Для автоматического применения таких анализаторов выборка должна быть достаточно большой и охватывать все возможные ситуации, например, изменение ракурса, внешности, условий освещённости и т.п.

Почти все алгоритмы распознавания лиц построены на машинном обучении с использованием многослойных нейронных сетей (МНС). Нейросети обучаются на больших наборах данных (датасетах) с размеченными изображениями. При этом, чем качественнее исходные данные, тем лучше и точнее алгоритм будет справляться с поставленной задачей. Обучаются МНС при помощи алгоритма обратного распространения ошибки («backpropagation»), являющегося разновидностью градиентного спуска в пространстве весовых коэффициентов с целью минимизации суммарной ошибки:

$$\Delta W = -\eta \frac{\partial E}{\partial W}, \quad (5)$$

где η – скорость обучения («learning rate»). Правильный выбор скорости обучения очень важен и влияет на качество обучения всей нейросети. При этом функция суммарной ошибки E для нейросети составляет:

$$E = \frac{1}{2} \sum_j (y_j - t_j)^2 \quad (6)$$

где t_j – эталонное значение выходов МНС, y_j – фактическое значение выходов МНС. При этом ошибки (величины коррекции весовых коэффициентов) распространяются в обратном направлении от входов к выходам.

Алгоритм обратного распространения ошибки является NP-трудным, поэтому время обучения сети увеличивается экспоненциально с ростом размерности данных. В процессе обучения нейросети происходит автоматическое извлечение ключевых признаков, определение их важности и построение взаимосвязей между ними. Обученная нейронная сеть может успешно применять опыт, полученный в процессе обучения, на неизвестные образы за счёт хороших обобщающих способностей.

Главные проблемы МНС заключаются в следующем:

1. Проблема локального минимума – заключается в том, что при итерационном спуске может наступить момент, когда решение нейросети заходит в локальный минимум, вследствие малой величины шага. Локальный минимум не всегда обеспечивает приемлемое решение, поэтому решение проблемы заключается в применении стохастических методов.

2. Выбор архитектуры нейросети (количество нейронов, слоёв, характер связей, функции активации). С этим связана проблема переобучения, которая заключается в том, что нейросеть с избыточным числом элементов теряет обобщающую способность и хорошо работает только на

тренировочной выборке. В настоящее время разработаны различные априорные оценки выбора архитектуры, методы прореживания обученных нейросетей, эволюционирующие нейросети возрастающих топологий.

3. Выбор шага (скорости) обучения. Данная проблема связана с тем, что при малом шаге время обучения будет большим и нейросеть может «застревать» в локальных минимумах, а при больших шагах возможно расхождение процесса обучения. Проблема эффективно решается адаптивным шагом, который на каждой итерации позволяет сделать шаг, минимизирующий ошибку нейросети на данной итерации. Существуют методы, которые на каждом тренировочном цикле (эпохе) анализируют всю тренировочную выборку и выбирают оптимальное значение и направление шага.

В общем виде алгоритм распознавания лиц заключается в следующем: сначала на фотографии детектируется лицо, потом на лице выделяются несколько ключевых точек, затем лицо выравнивается по этим ключевым точкам так, чтобы подбородок был внизу. Затем производится проверка: возможно ли распознавание или нет? На это влияют такие факторы как освещение, перекрытие лица другими предметами (очки, маска и т.д.). После чего производится финальная предобработка, и затем верификация или идентификация в зависимости от поставленной задачи.

Свёрточные и рекуррентные нейронные сети. После определения класса решаемой задачи производится выбор архитектуры нейронной сети. Современное глубокое обучение основывается на двух типах нейронных сетей: свёрточные и рекуррентные. В основе рекуррентных нейросетей – рекуррентный LSTM-нейрон, а в основе свёрточных – операция свёртки (рис. 5).

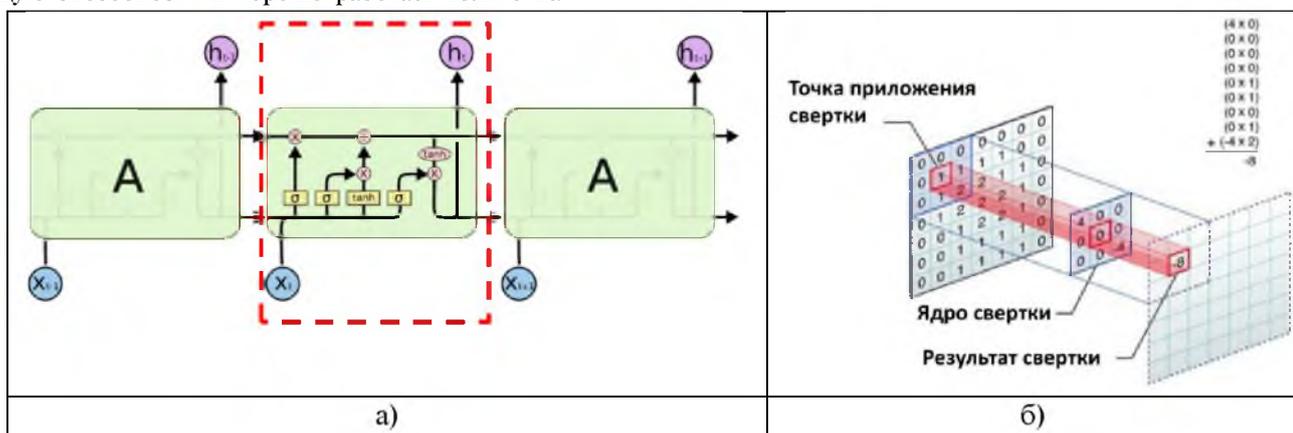


Рисунок 5 – Базовые элементы рекуррентных и свёрточных нейронных сетей:
а) рекуррентный нейрон (LSTM-ячейка); б) операция свёртки

На рис. 5а рекуррентный LSTM-нейрон (выделен пунктирной линией) содержит один слой гиперболического тангенса (\tanh) и три сигмоидальных слоя (σ). На вход LSTM-нейрон получает некоторые данные X_t и подаёт на выход некоторое значение h_t . Сигмоидальный слой («input layer gate» – «слой входного фильтра») возвращает числа в диапазоне $[0...1]$, которые обозначают, какую долю каждого блока информации следует пропустить дальше по нейросети. Циклическая связь позволяет передавать информацию от текущего шага нейросети к следующему. Рекуррентная нейронная сеть (РНС) с «долгой краткосрочной памятью» («Long short-term memory», LSTM) – это специфическая модификация классической РНС, которая способна учитывать долгосрочные зависимости между подаваемыми на вход данными. Отличие РНС от традиционных искусственных нейронных сетей заключается в наличии обратных связей (как внутри, так и между слоями), что позволяет при анализе текущего набора входных значений учитывать результаты предыдущих итераций. Данная особенность позволяет говорить о наличии «эффекта памяти», позволяющего анализировать последовательности подаваемых на вход данных.

На рис. 5б представлена операция свёртки для изображения размера 7×7 пикселей, при этом ядро свёртки размером 3×3 пикселя. Операция свёртки заключается в том, что ядро перемещается с верхнего левого угла изображения, производя операцию умножения пикселей. Например, в результате свёртки число – 8 получено следующим образом: $4 \times 0 + 0 \times 0 + 0 \times 0 + 0 \times 0 + 0 \times 1 + 0 \times 1 + 0 \times 0 + 0 \times 1 + (-4) \times 2 = -8$. Свёрточные нейросети обеспечивают высокую скорость обучения и обработки изображений, частичную устойчивость к изменениям масштаба, смещениям, поворотам, смене ракурса и прочим искажениям изображения.

Таким образом, РНС предназначены для обработки последовательностей (зависимостей), например, для анализа видео-потока, а свёрточные нейросети – в основном для обработки изображений. Ниже рассмотрено применение свёрточных и рекуррентных нейросетей для распознавания лиц на изображениях и в видеопотоке.

Алгоритм распознавания лиц свёрточной нейросетью. Свёрточные нейросети являются самой мощной технологией глубокого обучения для распознавания изображений и являются сложными многоуровневыми структурами,

каждый уровень которых представляет собой свёртку изображений со своим обучаемым набором фильтров. В отличие от признаков Хаара, которые используются в алгоритме Виолы-Джонса, нейросети не требуют участия человека и самостоятельно находят признаки в процессе обучения. Для этого необходимо обучить нейросеть на большой базе изображений (обучающей выборке), в которой имелись бы снимки самых разных лиц на различном фоне. Свёрточные нейросети обеспечивают частичную устойчивость к изменениям масштаба, смещениям, поворотам, смене ракурса и прочим искажениям изображения.

Каждый слой свёрточной нейронной сети представляет собой набор плоскостей, состоящих из нейронов. Нейроны одной плоскости имеют одинаковые синаптические (весовые) коэффициенты, ведущие ко всем локальным участкам предыдущего слоя. Каждый нейрон слоя получает входы от некоторой области предыдущего слоя (локальное рецептивное поле), т. е. входное изображение предыдущего слоя сканируется небольшим окном и пропускается сквозь набор синаптических коэффициентов, а результат отображается на соответствующий нейрон текущего слоя. Таким образом, набор плоскостей представляет собой карты характеристик, и каждая плоскость находит «свои» участки изображения в любом месте предыдущего слоя. Размер локального рецептивного поля выбирается самостоятельно в процессе разработки нейронной сети.

Особенностью таких нейросетей является наличие т.н. «свёрточных слоёв». Основное отличие свёрточных слоёв от полносвязных состоит в том, что каждый нейрон в нем соединен только с ограниченным числом нейронов из предыдущего слоя. Такой слой рассматривают как совокупность фильтров. Каждому фильтру ставят в соответствие несколько параметров: размер, шаг, набор весовых коэффициентов. Главным назначением одного конкретного свёрточного слоя является выделение простых шаблонов на входе с помощью обучаемых фильтров. Располагая свёрточные слои один за другим и в комбинации с другими типами слоёв, получаем, что с ростом глубины нейросети растёт абстрактность и сложность выделяемых признаков. На рис. 6 представлена визуализация признаков на различных уровнях свёрточной нейронной сети при распознавании лиц на изображениях.

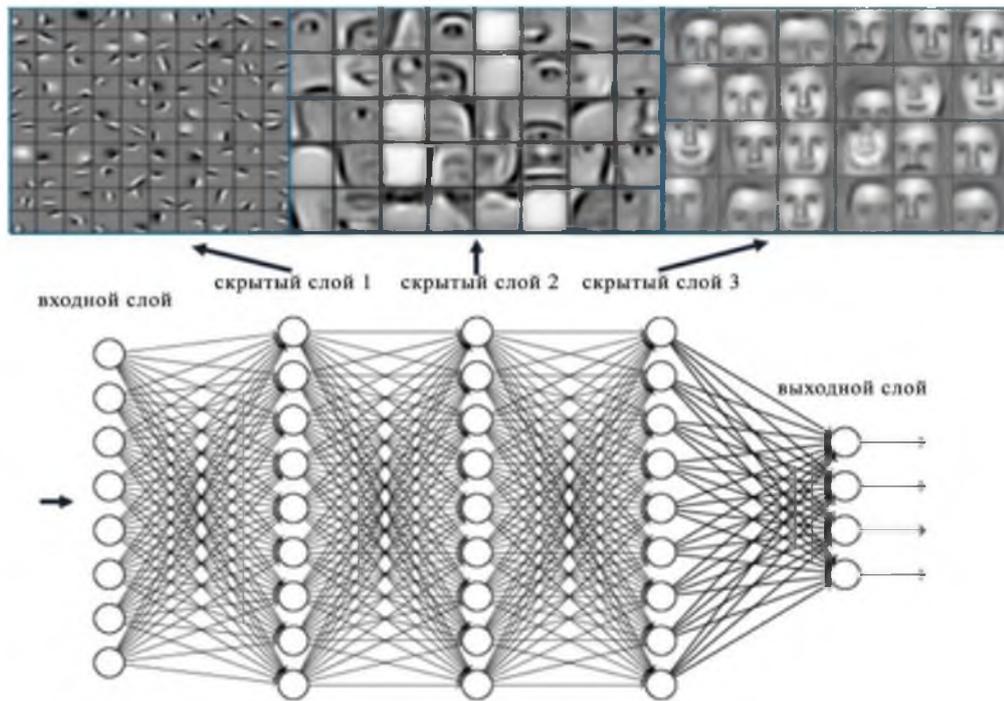


Рисунок 6 – Визуализация признаков на различных уровнях нейронной сети

Рассмотрим первый шаг работы алгоритма – детекцию лица на фотографии. По изображению перемещается окно определённого размера и для каждого фрагмента изображения вычисляется вероятность того, что в этом фрагменте находится лицо. Затем выбирается фрагмент изображения с наибольшей вероятностью. Второй шаг работы алгоритма – детекция особых точек, с помощью которых можно построить форму лица. Следующий шаг – верификация лица. На данном этапе необходимо построить функцию, которая по части преобработанной фотографии с лицом строит набор числовых признаков (вектор признаков), которые достаточно хорошо разделяются по

кластерам. В каждом кластере оказывается фотография одного человека. Чтобы сделать распознавание лиц более точным, глубокие нейросети часто поворачивают лицо в анфас с помощью опорных точек на лице, чтобы экстраполировать новое положение.

Архитектура свёрточной нейросети состоит из многих слоёв: несколько свёрточных слоёв («convolutional»), следующих друг за другом, затем, как правило, нелинейная функция активации и слой подвыборки (субдискретизации, «subsampling»). Такая последовательность может повторяться несколько раз. На рис. 7 представлена обобщённая схема распознавания лиц с помощью свёрточной нейросети.



Рисунок 7 – Схема распознавания лица свёрточной нейронной сетью

Для распознавания лица система наблюдения делает фотографию человека, затем нейронная сеть выполняет поиск области лица. Полученная область выделяется, оптимизируется масштаб, корректируется яркость и контраст изображения. При распознавании лица нейросетью происходит поиск из базы данных похожих лиц, детальное сравнение с выборкой похожих на него лиц по ключевым признакам: определяется пол, возраст и эмоциональное состояние человека на изображении. После этого нейронная сеть выводит полученные результаты на изображение [12]. Технология аутентификации СППР на основе свёрточной нейронной сети при предоставлении доступа состоит из трёх этапов:

- 1) измерение параметров объекта и получение текущего образца;
- 2) распознавание текущего образца;

3) принятие решения.

Все системы распознавания лиц можно условно разделить на две широкие категории: использующие (2D) двумерные изображения и (3D) трёхмерные изображения лиц. При использовании баз данных 2D лиц на качество распознавания влияют положение лиц на изображении и условия освещения, при этом 3D изображения разработаны для снятия этих ограничений. Современные свёрточные нейронные сети способны по исходному 2D-изображению восстановить трёхмерное изображение лица, менять ориентацию головы и источник света (рис. 8).

Данные возможности позволяют более точно идентифицировать человека по одной фотографии. В подобном направлении развиваются технологии получения трёхмерного отпечатка пальца (рис. 9).

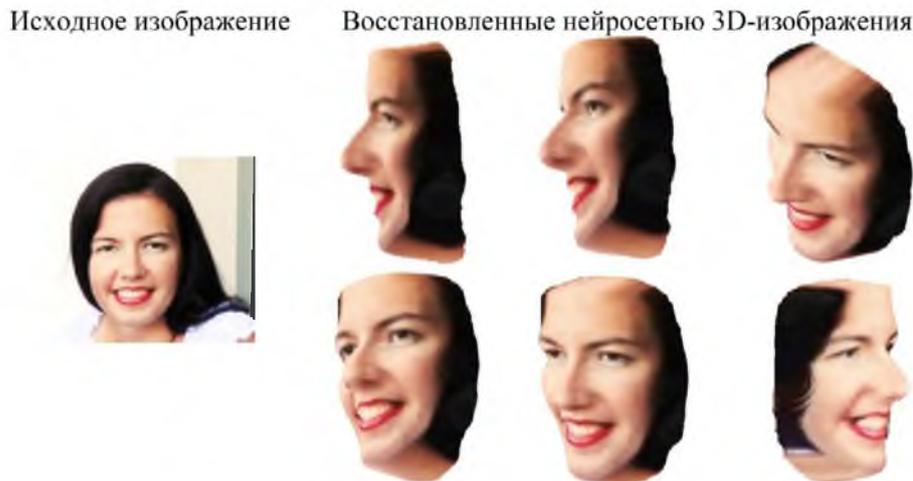


Рисунок 8 – Результат восстановления трёхмерного изображения нейросетью

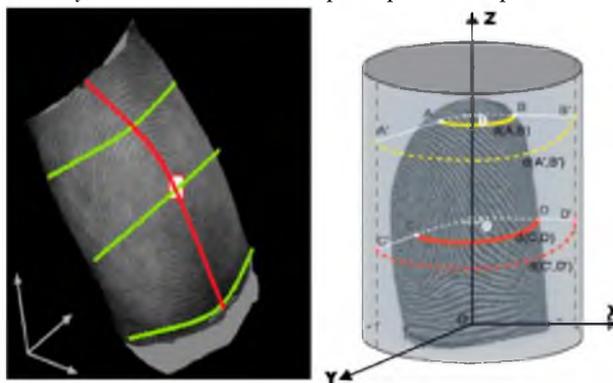


Рисунок 9 – Трёхмерное представление отпечатка пальца

Преимущество трёхмерной модели отпечатка заключается в том, что при соприкосновении с сенсором человек может по-разному прикладывать палец, что не повлияет на качество распознавания, но значительно повысит точность.

Нейросетевые детекторы. В настоящее время разработано множество архитектур специализированных нейронных сетей для распознавания человеческих лиц. Например, сиамская

нейронная сеть Google FaceNet, свёрточные нейросети Facebook DeepFace и VGG-Face (VGG-16) достигли точности распознавания лиц сравнимой с человеком [13, 14]. Современные реализации обнаружения объектов на основе свёрточных нейронных сетей строятся поверх предварительно обученных нейросетевых архитектур, таких как AlexNet, Google Inception, VGG. У данных

нейросетевых архитектур натренированы миллионы параметров и они могут конвертировать входное изображение в набор признаков, которые могут далее использоваться для таких задач, как классификация изображений, обнаружение объектов и другие задачи, связанные с компьютерным зрением. В 2014 году был предложен детектор объектов под названием R-CNN («Regions with CNN features» – участки с CNN-признаками), в котором нейросеть AlexNet была вторично натренирована набором помеченных данных для обнаружения объектов. Нейросетевой детектор R-CNN показал большое улучшение точности, по сравнению с традиционными методами обнаружения. С тех пор как был предложен R-CNN-детектор, один за другим стали выходить более эффективные алгоритмы обнаружения объектов: быстрый R-CNN детектор (Fast R-CNN), более быстрый R-CNN-детектор (Faster R-CNN), детектор Mask R-CNN, детектор YOLO («You Only Look Once»), SSD-детектор («Single Shot Multibox Detector» – однократный многорамочный детектор), детектор YOLO v2 и др. В настоящее время существует большое количество открытых библиотек для распознавания лиц, например, OpenCV и OpenFace [15-17].

Определение возраста и эмоционального состояния с помощью нейросети. Эмоциональный искусственный интеллект («emotion artificial intelligence») и эмоциональные вычисления («affective computing») – это технология, которая позволяет распознавать и интерпретировать человеческие эмоции и реагировать на них. Камера формирует видео-поток, затем из него выделяются отдельные изображения и обрабатываются. Видеопоток представляет собой последовательность изображений, сопровождаемых аудио-

рожкой. Однако в настоящем исследовании работа велась только с визуальными данными, а аудиопотоки не анализировались.

Современные нейронные сети способны идентифицировать пол людей на видео с 90% точностью, но при определении возраста задача усложняется. Из-за условий съёмки или даже при небольшом повороте головы определение возраста одного и того же человека на разных кадрах колеблется в диапазоне ± 5 лет. Традиционно для определения возраста человека берётся промежуток от 1 до 100 лет: нейросеть по каждому видеокадру оценивает степень уверенности в том, что человек на изображении соответствует определённому возрасту. Например, если на 30% кадров нейросеть определяет, что человеку 21 год, а на 10% – что 60 лет, то вывод будет таким: с вероятностью 30% возраст составляет 21 год, а с вероятностью 10% – 60 лет. Черты лица (брови, глаза, губы и т. д.) размечаются маркерами (точками).

По результатам анализа алгоритмов идентификации лиц в видеопотоке был предложен подход, использующий в качестве классификатора рекуррентную LSTM-нейросеть, одним из преимуществ которой является способность обработки длинных последовательностей информации. Нейросеть определяет положение маркеров, сопоставляет их с признаками эмоций из шаблона и делает вывод о том, какая эмоция отражена: злость, страх, удивление, грусть, радость или спокойствие. Получены результаты компьютерных экспериментов работы программы в среде Matlab R2020a на основе свёрточной и рекуррентной нейронной сети для обнаружения лица, определения возраста, эмоционального состояния человека и наличия маски на лице в видеопотоке в режиме реального времени (рис. 10).

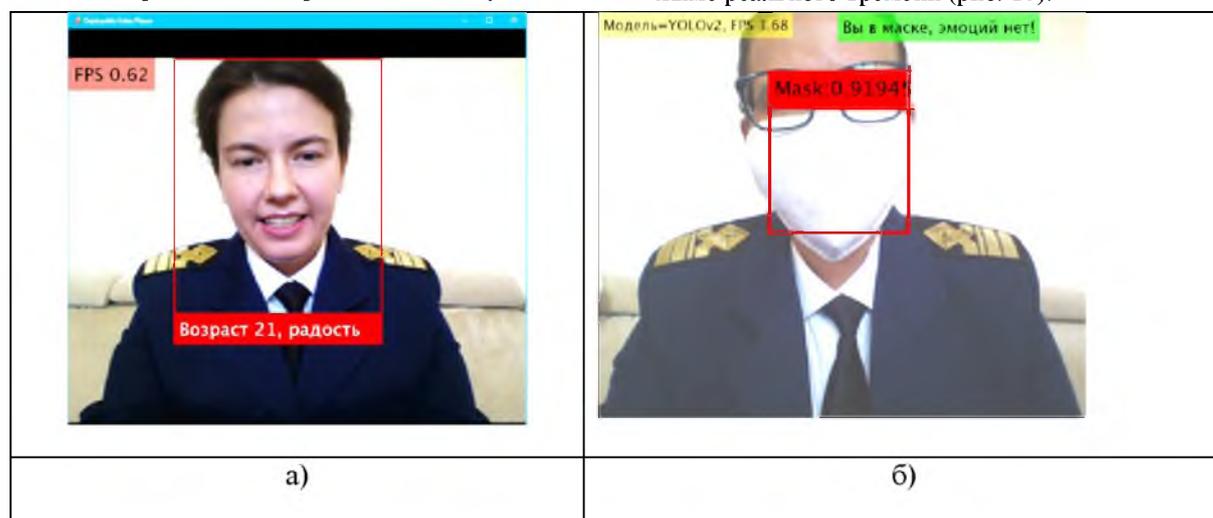


Рисунок 10 – Результаты работы нейронной сети в видеопотоке:
а) результат определения возраста и эмоции; б) результат распознавания маски на лице

Распознавание производится в видеопотоке: на рис. 10а нейросеть определила возраст 21 год и эмоцию «радость» при кадровой частоте 0,62 FPS («frames per second» – кадров в секунду), а на рис. 10б нейросеть на базе модели YOLOv2 определила наличие маски на лице с коэффициентом уверенности $\approx 0,92$ при кадровой частоте 3,68 FPS. Поскольку на лице обнаружена маска, определить эмоцию на лице невозможно. Для обучения использовался набор заранее размеченных изображений лиц, каждому из которых поставлена в соответствие изображённая на нём эмоция. Ниже приведён фрагмент листинга программы с комментариями на языке Matlab:

```

1. % подключение к веб-камере
2. wcam = webcam(1);
3. wcam.Resolution = wcam.AvailableResolutions{end};
4. player = vision.DeployableVideoPlayer();
5. fps = 0; % начальная кадровая частота
6. ages = 0:100; % диапазон возраста
7. % вектор-строка из распознаваемых эмоций
8. emotions = {'злость', 'отвращение', 'страх', 'радость', 'спокойствие', 'грусть', 'удивление'};
9. origSize = cellfun(@str2num, strsplit(wcam.Resolution, 'x'));
10. size1 = [448,448];
11. size2 = [224,224];
12. cont = true;
13. color = [150,0,0];
14. % загрузка моделей для обнаружения лица,
    возраста и эмоции
15. modelsFolder = 'models';
16. matfileFace = [modelsFolder, '/', 'net_face_yolo.mat'];
17. matfileAge = [modelsFolder, '/', 'net_age.mat'];
18. matfileEmotion = [modelsFolder, '/', 'net_emotion.mat'];

```

В процессе обучения РНС выявляет признаки, определяющие эмоцию человека. Чем точнее удаётся интерпретировать эмоции, намерения и мотивацию, тем выше уровень эмоционального интеллекта и эмпатии. Например, интеграция эмоционального интеллекта в голосовые помощники повысит способность определять эмоциональное состояние человека и скорректировать стратегию коммуникации, исходя из эмоционального состояния собеседника. Интеграция эмоционального искусственного интеллекта в автоматизированные системы управления (например, бес-

пилотные автомобили) необходима для обеспечения безопасности и корректировки стратегии управления.

Заключение. Рассмотрена эволюция подходов и алгоритмов к решению задач обработки, распознавания и локализации лиц с использованием алгоритма Виола-Джонса, свёрточных и рекуррентных нейронных сетей. Свёрточные нейросети обеспечивают частичную устойчивость к изменениям масштаба, смещениям, поворотам, смене ракурса и прочим искажениям изображения. В свою очередь, рекуррентные нейросети более эффективны для обработки последовательностей (зависимостей), в частности, для анализа видеопотока. Использование свёрточной нейронной сети обладает рядом преимуществ перед другими методами распознавания лиц:

- 1) способность к обучению для выделения ключевых характеристик лица и высокий результат распознавания даже при искажении данных;
- 2) возможность получения классификатора, который хорошо моделирует сложную функцию распределения изображения лица, что позволяет увеличить точность распознавания.

Рассмотрены технологии восстановления трёхмерного изображения лица по исходному 2D-изображению с помощью свёрточных нейронных сетей. По результатам анализа алгоритмов идентификации лиц в видеопотоке был предложен подход, использующий в качестве классификатора рекуррентную LSTM-нейросеть, одним из преимуществ которой является способность обработки длинных последовательностей информации в видео-потоке. Получены результаты компьютерных экспериментов работы программы в среде Matlab R2020a на основе свёрточной и рекуррентной нейронной сети для обнаружения лица, определения возраста, эмоционального состояния человека и наличия маски на лице в видеопотоке в режиме реального времени. Анализ результатов компьютерных экспериментов показал высокую эффективность работы нейросетевого алгоритма и программы на основе свёрточной и рекуррентной нейронной сети при распознавании лиц в видеопотоке.

Абстрактные и формальные задачи, являющиеся одними из самых сложных умственных задач для человека, успешно решаются алгоритмами глубокого машинного обучения. Сложности, с которыми сталкиваются системы, основанные на строго формализованных знаниях, свидетельствуют о том, что системы искусственного интеллекта должны обладать способностью приобретать собственные знания путём извлечения

шаблонов из исходных данных, что решается с помощью искусственных нейронных сетей.

Развитие техники и технологий в ближайшем будущем, возможно, будет развиваться по пути мульти-биометрии, т.е. совмещения нескольких биометрических модальностей для более точной идентификации и верификации. В свою очередь технология распознавания лиц граничит с более продвинутой технологией на базе генеративно-состязательных нейронных сетей, которая используется для создания дипфейков («deepfakes»), генерации лиц и различных объектов.

Практическое применение полученных результатов также может быть использовано в дистанционном обучении при идентификации личности курсантов и студентов.

Литература

1. Студеникин Д.Е., Бован С.Д., Хекерт Е.В., Модина М.А. Использование нейронных сетей для организации визуального наблюдения // Морские интеллектуальные технологии. – 2019. – № 4-3 (46). – С. 91-95.
2. Брилок Д.В., Старовойтов В.В. Распознавание человека по изображению лица нейросетевыми методами. – Минск, 2002. – 54 с.
3. LeCun Y. Deep Learning Hardware: Past, Present, and Future. IEEE International Solid-State Circuits Conference-(ISSCC), 2019. – 12-19 pp.
4. Клетте Р. Компьютерное зрение. Теория и алгоритмы / пер. с англ. А.А. Слинкин. – М.: ДМК Пресс, 2019. – 506 с.
5. Полковникова Н.А., Полковников А.К. Система поддержки принятия решений для выбора режима и прогнозирования отказов главного судового двигателя // Эксплуатация морского транспорта. – 2019. – №3(92). – С. 170-180.
6. Полковникова Н.А. Нейросетевые системы автоматического распознавания морских объектов // Эксплуатация морского транспорта. – 2020. – №1(94). – С. 207-219.
7. Белых Е. А. Обучение каскадов Хаара // Вестник Сыктывкарского университета. Сер. 1: Математика. Механика. Информатика. – 2017. – Вып. 1 (22). – С. 41-53.
8. Чанг Буй Тхи Тху, Хоанг Фан Нгок, Спицын В.Г. Распознавание лиц на основе применения метода Виолы–Джонса, вейвлет-преобразования и метода главных компонент // Известия Томского политехнического университета. – 2012. – т. 320. – № 5. – С. 54-59.
9. Шешкус А.В., Маталов Д.П., Арлазаров В.В., Николаев Д.П. Исследование композиции алгоритмов компьютерного зрения, базирующихся на машинном обучении, для решения задачи локализации и классификации объектов // Труды Института системного анализа Российской академии наук. – 2019. – Т. 69. – № 1. – С. 29-36.
10. Schroff F., Kalenichenko D., Philbin J. FaceNet: A Unified Embedding for Face Recognition and Clustering <https://arxiv.org/abs/1503.03832> (Дата обращения 05.08.2020)
11. Слитченко О.О., Клименко С.В., Степанова А.Ю. Использование аппарата нейросетей при распознавании лица // Актуальні проблеми автоматизації та інформаційних технологій. – 2017. – том 21. – С. 143-151.
12. Полковникова Н.А. Система поддержки принятия решений на основе свёрточных нейронных сетей по распознаванию объектов на изображениях // Материалы III международного научно-технического форума «Современные технологии в науке и образовании» (4-6 марта 2020, Рязань). – Рязань: Рязанский государственный радиотехнический университет имени В.Ф. Уткина, т.5. – с. 108-112.
13. Recent advances on memetic algorithms and its applications in image processing. Editors: Hemanth D.J., Kumar B.V., Karpagam Manavalan G.R. – Springer, 2020. – 199 p.
14. Leskovec J., Rajaraman A., Ullman J.D. Mining of Massive Datasets. – Cambridge University Press, 3rd edition, 2019. – 583 p.
15. Redmon J., Farhadi A. YOLO9000: Better, Faster, Stronger. Proceedings of the IEEE conference on computer vision and pattern recognition, 2017. – 7263-7271 pp.
16. LeCun Y., Bengio Y., Hinton G. Deep learning. Nature, vol. 521 (7553), 2015. – 436-444 pp.
17. Krizhevsky A., Sutskever I., Hinton G.E. Imagenet classification with deep convolutional neural networks. Advances in neural information processing systems, 2012. – 1097-1105 pp.

REFERENCES

1. Studenikin D.E., Bovan S.D., Khekert E.V., Modina M.A. Ispol'zovanie neyronnykh setey dlya organizatsii vizual'nogo nablyudeniya // Morskie intellektual'nye tekhnologii. 2019. № 4-3 (46). – S. 91-95.
2. Brilyuk D.V., Starovoytov V.V. Raspoznavanie cheloveka po izobrazheniyu litsa neyrosetevymi metodami. – Minsk, 2002. – 54 s.
3. LeCun Y. Deep Learning Hardware: Past, Present, and Future. IEEE International Solid-State Circuits Conference-(ISSCC), 2019. – 12-19 pp.
4. Klette R. Komp'yuternoe zrenie. Teoriya i algoritmy / per. s angl. A.A. Slinkin. – M.: DMK Press, 2019. – 506 s.
5. Polkovnikova N.A., Polkovnikov A.K. Sistema podderzhki prinyatiya resheniy dlya vybora rezhima i prognozirovaniya otkazov glavnogo sudovogo dvigatelya // Ekspluatatsiya morskogo transporta. Gos. morskoy universitet im. admirala F.F. Ushakova, Novorossiysk. – 2019, №3(92). – S. 170-180.
6. Polkovnikova N.A. Neyrosetevye sistemy avtomaticheskogo raspoznavaniya morskikh ob"ektov

- // Eksploatatsiya morskogo transporta. Gos. morskoy universitet im. admirala F.F. Ushakova, Novorossiysk. – 2020, №1(94). – С. 207-219.
7. Belykh E. A. Obuchenie kaskadov Khaara // Vestnik Syktyvkarskogo universiteta. Ser. 1: Matematika. Mekhanika. Informatika. 2017. Vyp. 1 (22). С. 41–53.
 8. Chang Buy Tkhi Tkhu, Khoang Fan Ngok, Spitsyn V.G. Raspoznavanie lits na osnove primeniya metoda Violy–Dzhonsa, veyvlet preobrazovaniya i metoda glavnykh komponent // Izvestiya Tomskogo politekhnicheskogo universiteta, t. 320, № 5, 2012. – S. 54-59.
 9. Sheshkus A.V., Matalov D.P., Arlazarov V.V., Nikolaev D.P. Issledovanie kompozitsii algoritmov komp'yuternogo zreniya, baziruyushchikhsya na mashinnom obuchenii, dlya resheniya zadachi lokalizatsii i klassifikatsii ob"ektov // Trudy Instituta sistemnogo analiza Rossiyskoy akademii nauk, 2019, T. 69, № 1. – S. 29-36.
 10. Schroff F., Kalenichenko D., Philbin J. FaceNet: A Unified Embedding for Face Recognition and Clustering <https://arxiv.org/abs/1503.03832> (Data obrashcheniya 05.08.2020)
 11. Slipchenko O.O., Klimenko S.V., Stepanova A.Yu. Ispol'zovanie apparata neyrosetey pri raspoznavanii litsa // Aktual'ni problemi avtomatizatsii ta informatsiynikh tekhnologiy. tom 21, 2017. – S. 143-151.
 12. Polkovnikova N.A. Sistema podderzhki prinyatiya resheniy na osnove svertochnykh neyronnykh setey po raspoznavaniyu ob"ektov na izobrazheniyakh // Materialy III mezhdunarodnogo nauchno-tekhnicheskogo foruma «Sovremennye tekhnologii v nauke i obrazovanii» (4-6 marta 2020, Ryazan'). – Ryazan': Ryazanskiy gosudarstvennyy radiotekhnicheskii universitet imeni V.F. Utkina, t.5.– s. 108-112.
 13. Recent advances on memetic algorithms and its applications in image processing. Editors: Hemanth D.J., Kumar B.V., Karpagam Manavalan G.R. – Springer, 2020. – 199 p.
 14. Leskovec J., Rajaraman A., Ullman J.D. Mining of Massive Datasets. – Cambridge University Press, 3rd edition, 2019. – 583 p.
 15. Redmon J., Farhadi A. YOLO9000: Better, Faster, Stronger. Proceedings of the IEEE conference on computer vision and pattern recognition, 2017. – 7263-7271 pp.
 16. LeCun Y., Bengio Y., Hinton G. Deep learning. Nature, vol. 521 (7553), 2015. – 436-444 pp.
 17. Krizhevsky A., Sutskever I., Hinton G.E. Imagenet classification with deep convolutional neural networks. Advances in neural information processing systems, 2012. – 1097-1105 pp.

УДК 629.125

DOI: 10.34046/aumsuomt96/22

РАЗРАБОТКА И ИСПЫТАНИЕ ТЕСТОВЫХ ЭКРАНОПЛАНОВ РАЗЛИЧНЫХ КОМПОНОВОК

*И. М. Данцевич, кандидат технических наук
В.Н. Прасолов*

В данной статье описываются эксперименты, проведённые с использованием радиоуправляемых и виртуальных моделей экранопланов. Описаны основные проблемы, выявленные в ходе проведения экспериментов, а также обозначены дальнейшие перспективы развития проекта.

Ключевые слова: Экранный эффект, экраноплан, высокоскоростной водный транспорт.

This article describes the experiments carried out using the constructed radio-controlled and virtual models of WISE-craft. The main problems identified during the experiments are described, and further prospects for the development of the project are indicated.

Keywords: Wing in surface effect (WISE), WISE-craft, high-speed water transport.

В настоящее время всё чаще встречается информация об успешных испытаниях таких видов транспорта, идею о создании которых ещё пару десятков лет назад не восприняли бы даже в серьёз. Активно строятся и испытываются мультироторные аппараты [1] перевозящие, как большие грузы, так и людей; персональные летательные ранцы с турбореактивной тягой с возможностью вертикального взлёта и посадки [2] и многие другие.

Тенденция развития нестандартных видов транспорта объясняется большим скачком в развитии области электронных, материальных и интеллектуальных технологий, а также относительной доступностью всех необходимых материалов, инструментов и информации, позволяющих даже в кустарных условиях осуществлять разработку, постройку и испытание достаточно серьёзных прототипов инновационных транспортных средств.