

СРАВНИТЕЛЬНЫЙ АНАЛИЗ АЛГОРИТМОВ РАСПОЗНАВАНИЯ ЧЕЛОВЕКА ПО БИОМЕТРИИ ЛИЦА

Белорусский государственный университет, Минск, Республика Беларусь

В настоящее время всё более широкое распространение получают биометрические системы идентификации человека. Они 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. Лицо и голос каждого человека также индивидуальны. Распознавание человека по изображению лица выделяется среди биометрических систем тем, что, во-первых, не требует специального дорогостоящего оборудования. Для большинства приложений персонального компьютера достаточно обычной видеокамеры, а почти все современные смартфоны имеют фронтальную камеру достаточно хорошего качества. Во-вторых, быстрое и удобное использование. В большинстве случаев достаточно просто пройти мимо или задержать лицо перед камерой на несколько секунд. В данной работе сравниваются готовые и доступные решения для распознавания человека по изображению лица с целью выбора наиболее оптимальной для использования в будущем проекте по разработке системы аутентификации пользователя по биометрии лица.

К основным классам решаемых задач распознавания человека по лицу на изображении или в видеопотоке можно отнести: поиск в больших базах данных, контроль доступа, получение информации о человеке. Они различаются как по требованиям, предоставляемым к системам распознавания, так и по способам решения. Различны и требования, предъявляемые к ошибкам первого и второго рода для таких классов. Ошибкой первого рода в данном случае называется ситуация, когда объект заданного класса не распознаётся системой. Ошибка второго рода происходит, когда объект заданного класса принимается за объект другого класса. Следует также отметить различие понятий верификации и идентификации. В задаче верификации неизвестный объект заявляет, что он принадлежит к некоторому известному системе классу. Система подтверждает или опровергает это заявление. В системах верификации ошибкой первого рода является ситуация, когда объект, принадлежащий к известным системе классам, принимается за объект, относящийся к неизвестным системе классам, и в доступе ему отказывают. Ошибка второго рода совершается, когда объект неизвестного класса принимается за объект, относящийся к известным системе классам, и ему разрешается доступ. При идентификации требуется отнести объект к одному из n известных классов или выдать заключение о том, что этот объект не относится к известным классам. При верификации критическими являются требования к ошибкам второго рода. Система не должна определять незнакомых людей как знакомых, возможно даже за счёт увеличения ошибок первого рода (отказов в доступе знакомым людям).

Все известные методы сходятся в том, что имеется набор изображений лиц заданной группы людей, к которому алгоритм обращается в процессе распознавания или изначально настраивается на него в процессе обучения. Одним из распространенных подходов к решению такой задачи является использование нейронных сетей, которые после обучения обладают хорошей обобщающей способностью. числовое значение.

Процесс распознавания лица на изображении (в видеопотоке) можно разделить на этапы [1]: 1) Детектирование и локализация лица на изображении; 2) Выравнивание изображения лица; 3) Вычисление вектора признаков; 4) Распознавание.

На каждом этапе используются различные методы их решения. Они могут различаться по затраченному времени и точности выполнения. по входным данным.

В работе сравнения подходов к распознаванию делались с использованием языка программирования Python. Тесты проводились на различных датасетах изображений и видео.

Для детектирования лиц на изображениях и видео сравнивались следующие методы: Виолы-Джонса [2], Histogram of Oriented Gradients (HOG), нейронные сверточные сети (Single Shot MultiBox Detector (SSD), MTCNN, RetinaFace, MediaPipe)

Датасет по подсчету числа лиц на фотографии.

Данный датасет содержит более 8000 фотографий с разным количеством людей на них. Также на фотографиях лица людей могут находиться под разными углами. Такой датасет применяется для обучения моделей по поиску лиц на фотографиях. Суммарно лиц на всех фотографиях — 24533. Результаты работы с этим набором данных представлены в таблице 1.

Таблица 1

Результаты работы методов по подсчету лиц на фото

Метод	Найдено лиц всего, %	Найдено лиц на одном фото в среднем, %	Времени затрачено на весь датасет, с	Времени на обработку одного фото в среднем, с
Виолы-Джонса	41.66	44.13	388.48	0.047
HOG	63.70	67.95	1275.25	0.16
SSD	46.23	51.59	279.40	0.034
MTCNN	93.15	94.51	4203.33	0.51
RetinaFace	98.92	99.23	1485.12	0.181
Mediapipe(RGB)	35.47	44.55	70.76	0.008

Скорость обработки изображений у метода Mediapipe самая высокая, что очень важно при обработке потокового видео, хотя результаты обнаружения лиц не выглядят впечатляюще по сравнению с, например, методом RetinaFace. Из-за слишком низкой скорости MTCNN по сравнению с RetinaFace, которая имеет большую точность, было принято решение не использовать этот метод в сравнениях на других датасетах.

Датасет фотографий приближенных лиц под разными углами.

Содержит 6000 фотографий лиц небольшого числа людей, которые повернуты под разными углами к объективу камеры. Разрешение фотографий 200×200 пикселей. На каждой фотографии присутствует только одно лицо, а сама съемка сделана на фронтальную камеру телефона. В ходе исследований установлено, что некоторые методы отработали быстрее, что связано с малым разрешением самих фотографий. Точность Mediapipe на данном датасете лучше по сравнению с предыдущим.

Видео лиц, поворачивающихся на разные углы.

Видео сняты на фронтальную камеру телефона. На каждом присутствует только один человек, который поворачивает голову в разных направлениях. Разрешение видео — 750×630, частота кадров в секунду — 30.54, общая продолжительность всех видео — 229 секунд. Результаты показали большую точность Mediapipe, а также его высокую скорость.

Таким образом, для решения задачи детектирования мной был выбран метод детектирования из фреймворка Mediapipe. Он очень быстр и детектирует лицо на 99% кадров на видео с интервью в разных разрешениях (144p, 240p, 360p, 480p, 720p, 1080p) и видео, снятых на фронтальную камеру. Нагружает процессор среднего сегмента AMD Ryzen 5 2600 всего на 20% в пике. А также сами веса нейронной сети занимают в памяти компьютера менее 90 МБ. видео.

Получение вектора признаков и распознавание.

Для получения вектора признаков использовались следующие нейронные сети: ArcFace, Facebook DeepFace, DeepID, Dlib Face Recognition, Google Facenet, Facenet512, OpenFace, VGG-Face. Выбор обусловлен открытостью использования нейронных сетей, разнообразием архитектур выбранных сетей, разнообразием датасетов, на которых обучались сети. Для полученных векторов вычислялось как косинусное, так и евклидово расстояние [3]. Сравнение нейронных сетей производилось на датасете из 2000 фотографий лиц 16 людей при разных ракурсах и разных углах поворота головы. Анализ результатов исследования по идентификации лица человека, что сети DeepFace и Dlib работают медленно, что неприемлемо при работе с потоковым видео. Сеть SFace показывает высокую производительность, но имеет слишком большую вероятность ошибок второго рода, что неприемлемо для систем верификации. Далее для проверки выбран небольшой датасет из видео с интервью разных людей. Для детектирования лица использовалась нейросеть Mediapipe. По результатам эксперимента хорошо себя показала Facenet512. Она также, как и на датасете фотографий лиц, имела практически не имела ошибок второго рода. А процент ошибок первого рода был минимален. Скорость обработки видео у системы с данной сетью позволяет эффективно обрабатывать более 15 изображений лиц в секунду.

Выводы

В качестве метода для детектирования лица на изображении наилучшим образом себя зарекомендовала мобильная нейронная сеть Mediapipe. Она обладает высокой скоростью обработки изображений, с высокой вероятностью детектирует лицо, расположенное под небольшими углами к объективу камеры, минимально нагружает процессор среднего класса и требует небольшого количества памяти, которое занимают веса данной сети. Для вычисления вектора признаков наилучшим решением оказалась сверточная нейронная сеть Facenet512, поскольку длина её выходного вектора не слишком велика по сравнению с DeepFace и VGG-Face, что позволит быстро вычислять расстояние между сравниваемыми векторами. Также у данной сети низкие вероятности ошибок первого и второго рода, что особенно важно в контексте систем верификации, где требуется минимальность ошибки второго рода.

Список литературы

1. Татаренков Д. А., Анализ методов обнаружения лиц на изображении (2015) [Электронный ресурс] – Режим доступа: <https://moluch.ru/archive/84/15524/> – Дата доступа: 10.04.2023.
2. P. Viola and M. J. Jones, «Rapid Object Detection using a Boosted Cascade of Simple Features», proceedings IEEE Conf. on Computer Vision and Pattern Recognition (CVPR 2001), 2001 [Electronic resource] – Mode of access: <https://ieeexplore.ieee.org/document/990517> – Date of access: 10.04.2023.
3. M., Deza, E. Encyclopedia of Distances — Fourth Edition. — Springer, 2016. [Electronic resource] – Mode of access: <https://link.springer.com/book/10.1007/978-3-662-52844-0> – Date of access: 10.04.2022.