

УДК 004.93'12

**ИСПОЛЬЗОВАНИЕ МАШИННОГО ОБУЧЕНИЯ ДЛЯ ГЛУБОКОГО
РАСПОЗНАВАНИЯ ЛИЦ***В. А. Филиппенко, А. В. Зотов*

Донской государственной технической университет (г. Ростов-на-Дону, Российская Федерация)

В данной работе рассматриваются возможности машинного обучения при создании программного обеспечения для распознавания лиц в видеопотоке.

Ключевые слова: машинное обучение, распознавание лиц, сверточная нейронная сеть, модель, классификация, набор данных, Python, видеопоток.

UDC 004.93'12

DEEP FACE RECOGNITION USING MACHINE LEARNING*V. A. Filippenko, A. V. Zotov*

Don State Technical University (Rostov-on-Don, Russian Federation)

In this paper, the authors consider the process of creating software for facial recognition in a video stream using machine learning.

Keywords: machine learning, face recognition, convolutional neural network, model, classification, data set, Python, video stream.

Введение. Распознавание лиц — одно из наиболее перспективных направлений машинного обучения. Системы распознавания могут применяться для биометрической бесконтактной идентификации по лицу, что значительно упростит процедуру подтверждения личности без предъявления документов и оптимизирует работу соответствующих служб. Системы, о которых идет речь, могут за доли секунды идентифицировать человека в толпе. Использование глубокого машинного обучения для распознавания лиц демонстрирует отличные результаты. Уже несколько лет данная технология активно применяется для идентификации не только лиц, но и практически всех предметов на изображении. Благодаря высокой производительности некоторых обученных моделей нейронных сетей, они применимы для процессоров смартфонов.

В данной статье рассмотрена одна из самых производительных технологий систем распознавания лиц, основанная на глубоком машинном обучении. Работа созданного программного средства состоит из трех основных этапов: обнаружение лица в потоке видеоданных, его нормализация и классификация. Для программирования модели использован язык Python [1, 2].

Обнаружение лиц в видеопотоке. Прежде чем обработать поток видео, его нужно представить покадрово. Затем на каждом кадре необходимо определить местоположение лица. Для этого используют один из двух способов:

— задействуют специальный алгоритм, который находит опорные точки или области лица, а после выделяет всю область лица;

— применяют модели нейронной сети, обученной на огромном объеме данных [3].

Подобных методов довольно много [4], и их можно разбить на две категории:

— основанные на выявлении особых черт человека, которые затем пытаются обнаружить программно;

— основанные на внешних признаках, при которых необходимо провести этап обучения системы путем обработки тестовых изображений.

Методом обнаружения лиц по внешним признакам стараются найти и выявить закономерности и свойства лица неявно, применяя методы математической статистики и машинного обучения. Один из вариантов обнаружения лиц в данном случае — создание и обучение модели сверточной нейронной сети. Для обучения требуется довольно большой набор данных, содержащий фотографии с лицами и без. Один из наиболее распространенных и крупных наборов таких данных — VGGFace2. Содержащиеся в нем изображения (более 3 млн) собираются из базы картинок Google и довольно существенно различаются по позиции, освещенности и т. д. Набор свободно распространяется в коммерческих и исследовательских целях. В представленной работе используется DlibResNet [5]. Эта модель представляет собой сеть ResNet с 29 сверточными слоями. Основой для обучения послужила коллекция лиц VGGFace2. Обучение сети начиналось со случайно инициализированных весов и использовало структурированную функцию потерь, которая пытается спроецировать все тождества в непересекающихся окружностях радиусом 0,6. После модель была протестирована с помощью общедоступного теста на нахождение лиц LFW (англ. Labeled Faces in the Wild можно перевести как «лица, отмеченные в естественной среде»). Обученная модель получила среднюю ошибку 0,993833 со стандартным отклонением 0,00272732, что является очень хорошим показателем.

После нахождения лица на кадре функция `get_frontal_face_detector()`, которая содержится в библиотеке Dlib, возвращает две координаты: левая верхняя и правая нижняя точка обнаруженного лица.

Нахождение антропометрических точек лица. Метод базируется на учете статистических связей между расположением антропометрических точек лица. Всего выделяется 68 точек: подбородок, внешние края глаз, края бровей, кончик носа и так далее (рис. 1).



Рис. 1. Расположение антропометрических точек на лице

После обнаружения точек вычисляется дескриптор лица — набор характеристик, описывающий лицо независимо от посторонних факторов. Сопоставление таких дескрипторов позволяет идентифицировать личность человека, например, с помощью расчета Евклидовой метрики между этими точками, обозначаемыми $P(x_1, x_2, \dots, x_n)$ и $Q(y_1, y_2, \dots, y_n)$ на плоскости (x, y) [6].

$$d(x, y) = \sqrt{\sum_{i=1}^n (x_i - y_i)^2}.$$

В данной работе использована модель Dlib, которая возвращает массив с 68 координатами нужных точек.

Нормализация обнаруженных лиц. После того как лицо выделено на изображении, программное средство выполняет следующий этап работы — нормализацию изображения. Различия между изображениями одного и того же лица из-за освещения и направления угла просмотра почти всегда больше, чем вариации изображения из-за изменения идентичности лица. Замечено, что различные ракурсы вызывают серьезные проблемы в реальных системах распознавания лиц. В

непринужденной обстановке, как правило, положение головы меняется по вертикали и горизонтали. Чтобы добиться лучших результатов, следует предварительно обработать изображения лица для устранения погрешности.

На распознавание лица больше всего влияет:

- ориентация на экране,
- освещение,
- перекрытие другими объектами.

Кусочно-аффинная деформация позволяет нормализовать изменения позы. Деформация применяется к треугольным элементам, определяемым описанными ниже моделями улучшенного активного внешнего вида (рис. 2).

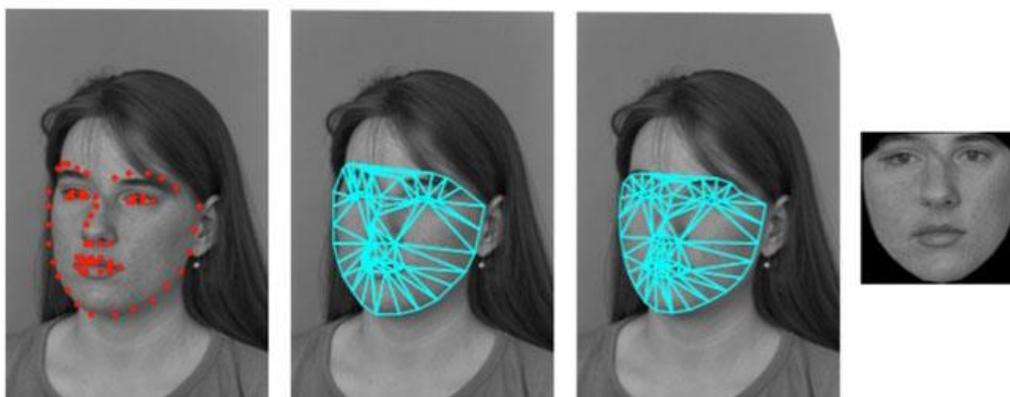


Рис. 2. Процесс выравнивания (или «поворота») лица с помощью аффинной деформации

Такое преобразование помогает повысить точность распознавания на 5–7 %.

Классификация обнаруженных лиц. Основной составляющей для классификации обнаруженных лиц будет извлечение дескриптора лица с помощью функции обученной модели. При разработке программного обеспечения авторы статьи применили модель, которая извлекает дескриптор лица, используя обнаруженные ранее антропометрические точки. Но для классификации лиц нужна база данных (БД), содержащая дескрипторы лиц, которые будут сравниваться с дескриптором лица, обнаруженного на кадре потока. Для создания такой базы авторы считали дескриптор каждого лица, а затем сохранили их в набор.

При наличии набора данных с дескрипторами лиц и дескриптора лица с видеопотока можно определить принадлежность данного лица к лицам, дескрипторы которых хранятся в БД. Для этого рассчитывается Евклидово расстояние между двумя дескрипторами. Согласно документации Dlib, если Евклидово расстояние меньше 0,6, то на фотографиях — один и тот же человек. Однако в ходе тестирования приложения было замечено, что при данном пороге лицо человека на кадре иногда классифицируется неправильно. Поэтому было решено снизить указанный порог до 0,587. При таком значении не отмечены погрешности в определении лица.

После выполнения вышеперечисленных действий лицо выделяется на кадре из потока с помощью функции `cv2.rectangle()`, принимающей две координаты обнаруженного на изображении лица, а также определяющей толщину и цвет рамки в формате RGB. Затем с помощью функции `cv2.putText()` выводится имя и на лицо наносится маска с антропометрическими точками. Для этого используется цикл, в котором обнаруженные точки ставятся на изображение с помощью функции `cv2.circle()`. Наконец, с помощью функции `cv2.imshow()` изображение выводится в видеопоток, и перечисленные выше действия повторяются для следующего кадра. Результат работы приложения показан на рис. 3.

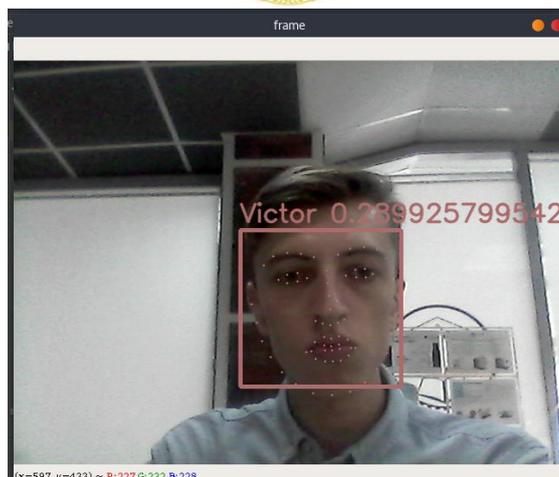


Рис. 3. Результат работы приложения для распознавания лица в видеопотоке

Заключение. В результате проведенной работы с помощью языка программирования Python создано программное обеспечение, которое получает на вход видеопоток, рассматривает его покадрово и выполняет действия, необходимые для идентификации лица в видеопотоке.

Программное обеспечение отвечает следующим требованиям:

- 1) полная работоспособность;
- 2) вывод всей необходимой информации;
- 3) поддержка графического интерфейса, в котором транслируется видеопоток;
- 4) выделение обнаруженного лица и вывод информации о человеке;
- 5) возможность демонстрации антропометрических точек обнаруживаемого лица.

Библиографический список

1. Мюллер, А. Введение в машинное обучение с помощью Python / А. Мюллер, С. Гвидо. — Москва : Вильямс, 2017. — 480 с.
2. Rosebrock, A. Deep Learning for Computer Vision with Python [Электронный ресурс] / A. Rosebrock. — Режим доступа: <https://www.pyimagesearch.com/deep-learning-computer-visionpython-book/> (дата обращения 22.10.19).
3. Грас, Дж. Data Science. Наука о данных с нуля / Дж. Грас. — Санкт-Петербург : БХВ-Петербург, 2017. — 336 с.
4. Форсайт, Д. Компьютерное зрение. Современный подход / Д. Форсайт, Ж. Понс. — Москва : Вильямс, 2004. — 928 с.
5. Dlib C++ Library [Электронный ресурс] / Boost Software License. — Режим доступа: <http://dlib.net/> (дата обращения: 20.10.19).
6. Лутц, А. Изучаем Python / А. Лутц. — Санкт-Петербург : Символ-Плюс, 2011. — 1280 с.

Об авторах:

Филиппенко Виктор Александрович, студент Донского государственного технического университета (344000, РФ, г. Ростов-на-Дону, пл. Гагарина, 1), philippenko@inbox.ru

Зотов Алексей Вячеславович, студент Донского государственного технического университета (344000, РФ, г. Ростов-на-Дону, пл. Гагарина, 1), azotzot@gmail.com