

## Исследование алгоритма Виолы-Джонса для разработки системы распознавания лиц с помощью нейронных сетей

*Д.Ю. Романюта<sup>1</sup>, А.В. Коваленко<sup>1</sup>, М.В. Шарпан<sup>2</sup>*

<sup>1</sup>*Кубанский государственный университет, Краснодар*

<sup>2</sup>*Краснодарский университет МВД России*

**Аннотация:** В современном мире распознавание лиц становится все более важной и актуальной задачей. С развитием технологий и возрастанием объема данных, увеличивается потребность в надежных, точных и эффективных системах распознавания лиц. Нейронные сети демонстрируют высокую эффективность в решении задач компьютерного зрения и имеют большой потенциал для улучшения существующих математических моделей распознавания лиц. Данная статья посвящена исследованию методов для распознавания лиц человека, будет подробно рассмотрен алгоритм Виолы-Джонса, который можно применять в задаче по распознаванию лиц с применением нейронных сетей. Также будут проанализированы методики обучения моделей глубокого обучения с помощью библиотек, использующие также алгоритм Виолы-Джонса и будет описан алгоритм использования обученной модели в API, которая будет использована в десктопных и мобильных приложениях.

**Ключевые слова:** биометрическая идентификация, распознавание лица человека, математические модели, методы распознавания лиц, глубокое обучение, сверточные нейронные сети, tensorflow.

### Введение

В последние годы распознавание лиц на изображении стало широко применяемой технологией в различных сферах, таких, как безопасность, аутентификация, видеонаблюдение и социальные сети. Однако, существует ряд сложностей, которые затрудняют достижение высокой точности и надежности распознавания лиц. К таким сложностям относят влияние освещённости на изображения, угол съёмки, различные аксессуары, которые влияют на точность распознавания.

Одной из важных проблем является вариабельность лицевых черт. Лица людей имеют уникальные особенности, такие как форма лица, расположение глаз, носа и рта, а также текстура кожи. Однако, эти черты могут изменяться в зависимости от возраста, эмоций, причёски, наличия или отсутствия грима и других факторов. Это приводит к тому, что одно и то же лицо может выглядеть совершенно по-разному на разных изображениях.

Поэтому в данной работе будет проведено исследование алгоритма распознавания лиц Виолы-Джонса, а также рассмотрено ее использование в различных библиотеках для глубокого обучения.

Актуальность темы данной статьи обусловлена несколькими факторами:

1. Возрастающая потребность в системах распознавания лиц: с развитием технологий и увеличением объема данных, потребность в надежных и точных системах распознавания лиц растет.

2. Необходимость в улучшении математических моделей: существующие модели распознавания лиц не всегда обеспечивают достаточную точность и требуют улучшения, разработка новых математических моделей с использованием нейронных сетей может повысить эффективность систем распознавания лиц и снизить количество ложных срабатываний.

3. Важность обучения и тестирования математических моделей: обучение и тестирование моделей распознавания лиц требует большого объема данных и вычислительных ресурсов.

Технологии для распознаванию объектов все чаще внедряются в разные сферы деятельности жизни человека, например, в медицине [1] и в строительстве [2]. В данных работах рассматриваются различные особенности использования нейронных моделей и детекторов для решения задач распознавания объектов. В правоохранительной сфере или в банкинге также появляются новые возможности для обеспечения безопасности пользователей, реализованных с помощью технологий распознавания объектов, в частности биометрическая идентификация.

Системы искусственного интеллекта показали свою эффективность в различных областях науки и техники. В данный момент проводится цифровизация различных направлений науки, например, электрохимии [3] и

---

робототехники [4]. В данных работах рассматриваются методы использования методов применения нейронных сетей для анализа и обучения.

### **Постановка задачи**

Объектом исследования является биометрическая идентификация личности с помощью распознавания лиц, а предметом исследования алгоритмы для реализации идентификации личности с помощью лица человека. Областью исследования являются математические модели и инструментальные средства для построения моделей сверточных нейронных сетей для распознавания лиц.

В данной работе поставлены следующие задачи:

1. Рассмотреть основные методы применяемые для распознавания лиц.
2. Исследовать математическую модель алгоритма Виолы-Джонса, применяемого для решения задачи распознавания лиц.
3. Обучить нейросетевую модель для распознавания лиц с помощью библиотеки Tensorflow и OpenCV, в которых отражен алгоритм Виолы-Джонса.

Для обучения и проведения экспериментов с нейросетевой моделью использовался язык программирования Python. Данный язык позволяет реализовать огромное количество алгоритмов для распознавания объектов. Также для данного языка было реализовано огромное количество полезных библиотек (библиотека NumPy, OpenCV, Tensorflow и т.п.).

### **Методы распознавания лиц на изображении**

Рассмотрим некоторые методы обнаружения и распознавания лица. Данные методы делятся на эмпирические и методы моделирования изображения лица. К методам первой группы относят, например, метод шаблонов (основан на сохранении шаблонов лиц в базе данных и сравнении их с изображением, полученным в режиме реального времени). К методам

---

второй группы относят метод распределения опорных точек (основан на отслеживании геометрических параметров лица) [5].

С развитием моделей глубокого обучения сверточные нейронные сети (Convolutional Neural Networks, CNN) становятся одним из самых используемых методов для распознавания лиц. CNN для распознавания лиц на изображениях является очень эффективным методом и широко используется в различных областях, таких как безопасность, медицина, развлечения и другие. Однако, этот метод также имеет некоторые ограничения, такие как возможность ошибок в распознавании лиц при наличии сильного освещения, наклонов головы и т.д. [6].

### Математическая модель алгоритма Виолы-Джонса

Опишем способ распознавания лица человека с помощью математической модели. В основе такой модели могут быть использованы различные алгоритмы для распознавания объектов. Для распознавания лиц обычно применяют алгоритм Виолы-Джонса. Определим начальные переменные для работы метода. Разделим набор изображений лиц  $X_1, X_2, X_3, \dots, X_L$  сохраненных в базе данных на  $L$  классов, где классы соответствуют зарегистрированным лицам. Каждое изображение представим в виде вектора значений  $K$ :

$$\mu = (\mu_1, \mu_2, \mu_3, \dots, \mu_K)^\tau, \quad (1)$$

где  $\tau$  – оператор транспонирования. Далее определим функцию расстояния для каждого изображения (1). Данная функция  $d(\mu_l, \mu_s)$  определяет для вектора изображения признаки на наименьшем расстоянии  $\mu$  во входной форме принадлежащие классу  $X_L$ :

$$d(\mu_l, \mu_s) < d(\mu_j, \mu_s), l \neq j, i, j = 0, 1, \dots, L-1, \quad (2)$$

Определим входные и выходные данные, которые используются для работы алгоритма. В качестве входных данных будут использованы

изображения из базы данных, а на выходе будут получена последовательность координат лица на кадре изображения. Обычно выходной кадр координат лица представляет изображение лица, ограниченное квадратом [6].

Теперь после определения начальных условий опишем математическую модель алгоритма Виолы-Джонса. Данный алгоритм использует интегральные изображения  $i(N*N)$ , представляющую из себя матрицу. В ней каждый элемент является суммой значений яркости, соответствующих пикселям на начальном изображении, расположенных слева и сверху от него:

$$I_i(x, y) = \sum_{x' \leq x, y' \leq y} i(x', y'), \quad (3)$$

Элементы матрицы (3) могут рассчитываться независимо от размера и местоположения исходного изображения. В дальнейшем полученная матрица используется для эффективного расчета яркости необходимых фрагментов изображения [6].

Алгоритм Виолы-Джонса построен на обучение методом Adaboost, который в свою очередь работает с помощью классификаторов, образующие признаки Хаара. Комбинация данных признаков позволяет получить сложный классификатор. В работе алгоритма выбираются наиболее подходящие признаки для обнаружения необходимого объекта. Одно из преимуществ метода Adaboost заключается в том, что при его работе можно напрямую определить частоту ошибок простого классификатора [7]. Данный подход позволяет не допустить итеративного обучения.

Далее происходит процесс обучения классификатора для работы алгоритма. Перед обучением происходит этап разделения изображений, хранящихся в базе на положительные и отрицательные образцы, которые будут выступать в качестве обучающей выборки. Далее происходит обучение с помощью классификатора. Классификатор применяет итерационный

подход для обучения, что позволяет увеличить эффективность работы классификатора, повысив точность распознавания и уменьшив число ошибок.

Во время работы алгоритма были освещены участки на которых могло быть лицо. Поэтому проводится этап сравнения входных и выходных изображений для сбора статистики работы классификатора.

Рассмотренный алгоритм используется в различных системах распознавания лиц и был реализован во многих библиотеках. Адаптивность метода позволяет использовать его в обучения нейросетевых моделей для распознавания лиц [8].

### **Обучение модели с помощью Tensorflow**

Идея CNN основана на работах по распознаванию визуальных образов, в том числе на алгоритме Виолы-Джонса, где было показано, что извлечение и объединение локальных признаков в более абстрактные признаки высшего порядка полезно. Для создания моделей сверточных нейронных сетей были разработаны библиотеки которые упрощают процесс настройки и обучения.

Рассмотренный ранее алгоритм используется во многих библиотеках для компьютерного зрения, одной из самых популярных таких библиотек является Tensorflow. Данная библиотека широко распространена, поскольку предоставляет интерфейс для выражения общих алгоритмов машинного обучения и исполняемого кода моделей. Модели, созданные в Tensorflow, можно переносить в системы с небольшими изменениями или вообще без них, используя самые разные устройства — от мобильных телефонов до распределенных серверов. Tensorflow был создан и поддерживается Google, а также используется внутри компании для целей машинного обучения [9].

Поскольку для хорошего обучения модели CNN требуется большой объем данных, который может привести к длительному сроку обучения на кластерах графических процессоров, предложили обучать модели CNN на

заранее обученной модели. Обычной практикой является использование предварительно обученной сети в качестве инициализации собственной модели или для импортирования функций для реализации задач. Такой тип обучения, при котором используется предварительно обученная сеть был назван трансферным обучением и подразумевает передачу знаний из одной модели в другую [9].

Предварительно обученную сеть также можно использовать, но не так часто, для инициализации всех весов в сети, где все веса корректируются во время обучения. При применении этого подхода в трансферном обучении некоторые из более ранних слоев обычно оставляют фиксированными, поскольку они представляют собой общие функции.

Для решения задачи обучения нейросетевой модели для обнаружения лиц в качестве предварительно обученной сети была выбрана модель MobileNet v2. Поскольку в будущем данная модель будет использована в мобильном приложении выбор данной модели обусловлен тем, что она предварительно оптимизирована для мобильных устройств, в которой для построения CNN используются отделимые по глубине свертки [10]. Это приводит к тому, что облегченная нейронная сеть ограничивает не только размер модели, но и, прежде всего, задержку модели. Данная модель позволит оптимизировать гиперпараметры для настройки соотношения задержки и точности в зависимости от решаемой проблемы [9].

Для создания нейросетевой модели для идентификации личности, которая будет встроена в мобильное приложение использовался набор данных DigFace 1M. Датасет содержащий 72 тысяч изображений для 10 тысяч сгенерированных человек.

Для обучения модели с целью распознавания лиц необходимо использовать изображения, соответствующие следующим требованиям: формат jpeg, png, gif (первый кадр) или bmp; размер файла составляет от 1 КБ

---

до 6 МБ, а размер изображения варьируется от  $36 \times 36$  пикселей до  $4096 \times 4096$  пикселей [5]. Подбор подходящих изображений для обучения является важной задачей, от которой зависит точность правильного распознавания объекта на изображении. Особенно при подборе изображений нужно обратить внимание на следующие факторы: разрешение (размер); яркость; осветительные приборы; угол захвата. Полный процесс обучения модели глубокого обучения представлен на рис.1.



Рис. 1. – Блок-схема обучения модели глубокого обучения

После настройки параметров и обработки датасета было начато обучение. На каждом шаге обучения сообщается значение функции потерь. Так первоначальное значение модели была равна 88,1216 и по прошествии 300 итераций стала равна 0,1366, что является хорошим результатом. На переобучение модели потребовалось около 3 часов. Переобученная модель сохраняется в файле вывода (типа pb). После обучения сеть можно использовать для распознавания лиц.

Обученную нейросетевую модель можно использовать в различных системах для распознавания лиц. Так как в качестве предварительно обученной модели была выбрана MobileNet v2, предпочтительным вариантом для использования модели являются мобильные устройства.

После обучения модели для распознавания лиц, следующий шаг состоит в реализации алгоритма для ее использования. Было реализовано API

для использования данной модели с целью идентификации личности. Данная программа будет импортировать библиотеки Tensorflow и OpenCV, а также написанный ранее API для распознавания лиц. После определения начальных параметров (источник кадров, длина и ширина кадра и т.п.) запускается цикл, который извлекает каждый кадр и запускает API для распознавания на нем лица. Если на кадре будет идентифицирован зарегистрированный пользователь, лицо будет обведено квадратом и над ним выведется надпись с именем пользователя. Алгоритм изображен на рис. 2.



Рис.2. – Алгоритм работы системы для идентификации лиц

Подробнее рассмотрим алгоритм кода, который осуществляет цикл сканирования кадров из источника для идентификации. Важным методом в данном алгоритме является проверка расстояния между объектом и камерой для наилучшей идентификации. Далее идет метод для загрузки API, которой загружает модель нейронной сети, используемой алгоритме распознавания лиц.

Важным элементом, который необходимо реализовать, является метод для определения неизвестных лиц и сохранение их как новых для дальнейшей идентификации. Когда в цикл попадает неизвестное лицо, программа должна предлагать вписать имя пользователя, после чего кадр с

лицом человека сохраняется, и оно будет определено при последующем попадании в цикл. Пример работы алгоритма показан на рис. 3.

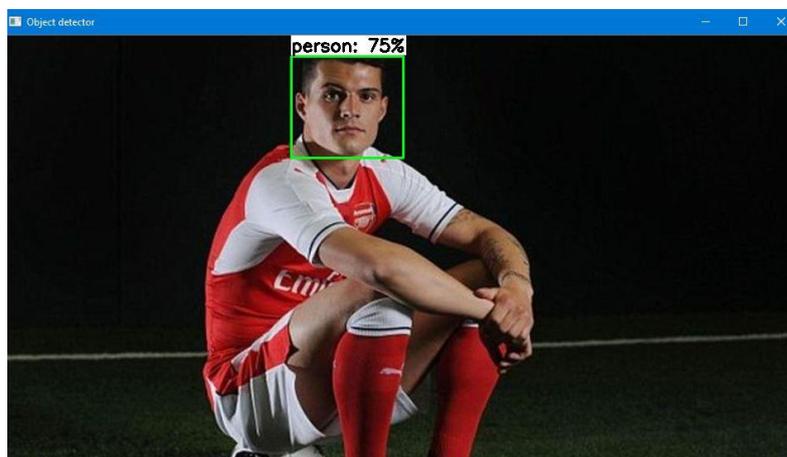


Рис. 3. – Пример обнаружения лица

### Заключение

С развитием математических моделей и нейронных сетей, возможности распознавания лица могут значительно усовершенствоваться. Нейронные сети смогут обрабатывать и анализировать более сложные и разнообразные данные, что позволит им точнее и быстрее распознавать лица различных людей.

В представленной работе были рассмотрены методы распознавания лиц на изображении, математическая модель в основе которой лежит алгоритм Виолы-Джонса, библиотеки, позволяющие строить модели сверточных нейронных сетей. Была обучена модель для распознавания лиц с помощью трансферного обучения. Данная модель легла в основу API, которое может использоваться в десктоп и мобильных приложениях, а также была написана программа для десктоп обнаружения лиц, с настраиваемыми источниками кадров. В дальнейшем будет собрана статистика по использованию данной программы в условиях разной видимости, а также эффектами и углами положения кадра для распознавания лиц. Данная статистика поможет улучшить алгоритм распознавания лиц на основе CNN.

## Литература

1. Евсина В.А., Жжонов В.А., Широбокова С.Н. Выявление границ органов воздушного дыхания на снимках компьютерной томографии с помощью сверточных нейронных сетей архитектуры U-NET // Инженерный вестник Дона, 2023, №1. URL: [ivdon.ru/ru/magazine/archive/n1y2023/8151/](http://ivdon.ru/ru/magazine/archive/n1y2023/8151/).

2. Егунов В.А., Кравченя П.Д., Большакова Е.И., Суменова З.Ш. Использование нейросетевых детекторов для предотвращения несчастных случаев на производстве // Инженерный вестник Дона, 2023, №10. URL: [ivdon.ru/ru/magazine/archive/n10y2023/8757/](http://ivdon.ru/ru/magazine/archive/n10y2023/8757/).

3. Sun, X., J. Li, A.V. Kovalenko, W. Feng and Y. Ou Integrating Reinforcement Learning and Learning From Demonstrations to Learn Nonprehensile Manipulation. IEEE, 2023, 20(3). URL: [ieeexplore.ieee.org/document/9810199/citations?tabFilter=papers#citations](https://ieeexplore.ieee.org/document/9810199/citations?tabFilter=papers#citations) Дата обращения 14.11.2023.

4. Gudza, I.V., A.Kh. Kovalenko, M.Kh. Urtenov and A.V. Pismenskiy Artificial intelligence system for the analysis of theoretical and experimental current-voltage characteristics. Journal of Physics, 2021, 2131. URL: [iopscience.iop.org/article/10.1088/1742-6596/2131/2/022089/pdf](https://iopscience.iop.org/article/10.1088/1742-6596/2131/2/022089/pdf) Дата обращения 13.11.2023.

5. Кузнецов Д.А., Никольский П.Г., Рачков Д.С., Кузнецов А.В., Хахамов А.П. Классификация методов обнаружения и распознавания лица на изображении // Научный результат. Информационные технологии, 2019, №1. URL: [rinformation.ru/journal/annotation/1642/](http://rinformation.ru/journal/annotation/1642/).

6. The Study of Mathematical Models and Algorithms for Face Recognition in Images Using Python in Proctoring System. URL: [mdpi.com/2079-3197/10/8/136](https://mdpi.com/2079-3197/10/8/136) Дата обращения 10.11.2023.

7. Друки А.А., Каковкин П.А., Чернета Д.С. Применение алгоритма Виолы-Джонса, преобразования Хафа и искусственных нейронных сетей для

---

распознавания символьной информации на сложном фоне // Scientific Visualization, 2016, №3 URL: [sv-journal.org/2016-3/05/index.php?lang=ru](http://sv-journal.org/2016-3/05/index.php?lang=ru).

8. Deep learning in bioinformatics. URL: [academic.oup.com/bib/article/18/5/851/2562808](http://academic.oup.com/bib/article/18/5/851/2562808) Дата обращения 10.11.2023.

9. Mobile Object Detection using TensorFlow Lite and Transfer Learning. URL: [diva-portal.org/smash/get/diva2:1242627/FULLTEXT01.pdf](http://diva-portal.org/smash/get/diva2:1242627/FULLTEXT01.pdf) Дата обращения 12.11.2023.

10. MobileNetV2: Inverted Residuals and Linear Bottlenecks. URL: [arxiv.org/abs/1801.04381](http://arxiv.org/abs/1801.04381). Дата обращения 12.12.2023.

### References

1. Evsina V.A., Zhzhonov V.A., Shirobokova S.N. Inzhenernyj vestnik Dona, 2023, №1. URL: [ivdon.ru/ru/magazine/archive/n1y2023/8151/](http://ivdon.ru/ru/magazine/archive/n1y2023/8151/).

2. Egunov V.A., Kravchenja P.D., Bol'shakova E.I., Sutenova Z.Sh. Inzhenernyj vestnik Dona, 2023, №10. URL: [ivdon.ru/ru/magazine/archive/n10y2023/8757/](http://ivdon.ru/ru/magazine/archive/n10y2023/8757/).

3. Sun, X., J. Li, A.V. Kovalenko, W. Feng and Y. Ou Integrating Reinforcement Learning and Learning From Demonstrations to Learn Nonprehensile Manipulation. IEEE, 2023, 20(3). URL: [ieeexplore.ieee.org/document/9810199/citations?tabFilter=papers#citations](http://ieeexplore.ieee.org/document/9810199/citations?tabFilter=papers#citations) Date accessed 14.11.2023.

4. Gudza, I.V., A.Kh. Kovalenko, M.Kh. Urtenov and A.V. Pismenskiy Journal of Physics, 2021, 2131. URL: [iopscience.iop.org/article/10.1088/1742-6596/2131/2/022089/pdf](http://iopscience.iop.org/article/10.1088/1742-6596/2131/2/022089/pdf) Date accessed 13.11.2023.

5. Kuznecov D.A., Nikol'skij P.G., Rachkov D.S., Kuznecov A.V., Nahamov A.P. Nauchnyj rezul'tat. Informacionnye tehnologii, 2019, №1 URL: [rrinformation.ru/journal/annotation/1642/](http://rrinformation.ru/journal/annotation/1642/).



6. The Study of Mathematical Models and Algorithms for Face Recognition in Images Using Python in Proctoring System. URL: [mdpi.com/2079-3197/10/8/136](https://mdpi.com/2079-3197/10/8/136) Date accessed 10.11.2023.

7. Druki A.A., Kakovkin P.A., Cherneta D.S. Scientific Visualization, 2016, №3 URL: [sv-journal.org/2016-3/05/index.php?lang=ru](https://sv-journal.org/2016-3/05/index.php?lang=ru).

8. Deep learning in bioinformatics. URL: [academic.oup.com/bib/article/18/5/851/2562808](https://academic.oup.com/bib/article/18/5/851/2562808) Date accessed 10.11.2023.

9. Mobile Object Detection using TensorFlow Lite and Transfer Learning. URL: [diva-portal.org/smash/get/diva2:1242627/FULLTEXT01.pdf](https://diva-portal.org/smash/get/diva2:1242627/FULLTEXT01.pdf) Date accessed 12.11.2023.

10. MobileNetV2: Inverted Residuals and Linear Bottlenecks. URL: [arxiv.org/abs/1801.04381](https://arxiv.org/abs/1801.04381) Date accessed 12.12.2023.

**Дата поступления: 19.11.2023**

**Дата публикации: 6.01.2024**