

Разработка и интеграция модели нейронной сети для распознавания лиц

Д.Н. Веклич, Э.С. Анисимова

Елабужский институт Казанского федерального университета, Елабуга

Аннотация: В статье представлено исследование, посвященное реализации нейросетевого подхода к задаче распознавания лиц. Проведён обзор существующих методов распознавания лиц. Разработана модель нейронной сети, обученная на наборе данных DigiFace-1M. В работе подробно описана архитектура разработанной модели нейронной сети, а также поэтапный процесс её обучения. Достигнутая точность модели на валидационной выборке составила 78%, а на обучающей - 92%. В статье также освещены вопросы интеграции разработанной модели в российский сервис Amvera Cloud. В результате создано веб-приложение, позволяющее проводить идентификацию пользователей по загруженным изображениям лиц. Данное исследование демонстрирует потенциал нейронных сетей для задач распознавания лиц и предлагает практическое решение для реализации подобных систем в различных сферах.

Ключевые слова: распознавание лиц, глубокое обучение, нейронные сети, идентификация пользователей, архитектура модели, обучение модели, интеграция модели, облачные сервисы, безопасность, биометрические технологии.

Введение

В современном мире, где информационные технологии стремительно развиваются, автоматическое распознавание лиц становится всё более важным и востребованным [1]. Оно используется в различных сферах человеческой деятельности, но основное своё применение нашло в сфере безопасности и охране правопорядка. Распознавание лиц относится к процессу идентификации или верификации людей путем анализа и сравнения уникальных черт лица. Среди разнообразных методов распознавания человека по биометрии, распознавание лиц выделяется своей популярностью и доступностью для массового применения, поскольку оно не требует какого-либо физического контакта и может быть выполнено на расстоянии, например с помощью камер видеонаблюдения.

С активным развитием технологий и увеличением количества данных, нейронные сети, в частности глубокие нейронные сети, стали эффективным и важным инструментом для решения различных задач компьютерного зрения. Они способны выявлять уникальные черты лица, такие, как форма глаз, носа

и рта, и использовать их для идентификации или верификации личности. Распознавание лиц с помощью нейронных сетей также имеет свои преимущества перед другими методами идентификации и верификации, такими как отпечатки пальцев или сменные пароли. Это быстрый и удобный способ проверки личности, который может быть реализован в реальном времени.

Методы распознавания лиц

Методы можно разделить на две категории – методы, основанные на распознавании по локальным признакам и методы, основанные на распознавании по глобальным признакам. Первая категория методов фокусируется на определённых частях лица (глаза, рот, нос и др.), вторая же рассматривает человеческое лицо как единое целое [2]. Существует большое количество методов распознавания лиц [3], рассмотрим некоторые из них.

- 1) Нейронные сети являются гибридным методом, это подразумевает, что в нейронных сетях может использоваться распознавание как по локальным признакам, так и по глобальным признакам. Нейронные сети могут использоваться не только для распознавания лиц, но и для распознавания эмоций и гендера по лицу [4, 5].
 - 2) Скрытые Марковские модели являются методом распознавания лиц по локальным признакам [6]. Они представляют собой статистическую модель, которая используется для характеристики статистических свойств сигнала.
 - 3) Метод главных компонент является статистическим методом распознавания лиц по глобальным признакам [7], который используется для сокращения многомерных данных до данных меньшей размерности при сохранении большинства информации путем преобразования исходных переменных в новые, некоррелированные переменные, известные как главные компоненты.
-

- 4) Метод гибкого сравнения на графах – это биологически вдохновленный алгоритм, где распознавание лиц происходит по локальным признакам [8].
- 5) Активные модели внешнего вида – это статистический метод, в котором распознавание лиц происходит по локальным признакам [9].
Для решения задачи распознавания лиц в данной работе был выбран нейросетевой подход.

Разработка и обучение модели нейронной сети по распознаванию лиц

Средой для разработки нейронной сети была выбрана Google Colaboratory. Одно из главных преимуществ Google Colaboratory - возможность использовать графический процессор или процессор для машинного обучения Tensor Processing Unit для выполнения вычислительно интенсивных задач, что делает Google Colab очень удобным инструментом для работы с искусственным интеллектом.

Для обучения нейронной сети был выбран набор данных под названием DigiFace-1M. Этот набор данных поделен на 2 части и содержит синтезированные изображения людей. Первая часть содержит в себе 720 тысяч изображений для 10 тысяч людей, для каждого человека имеется 72 изображения.

Выборка с изображениями была разделена на обучающую, валидационную и тестовую выборки, в соотношении 80% изображений для обучающей и по 10% для валидационной и тестовой. В обучающую выборку попало 227 943 изображения, в валидационную – 27 993 изображения, а в тестовую выборку 31 992 изображения.

Разработанная архитектура модели состоит из двух скрытых слоёв, двух слоёв Dropout, слоя Flatten, полносвязного слоя и выходного слоя (рис.1).

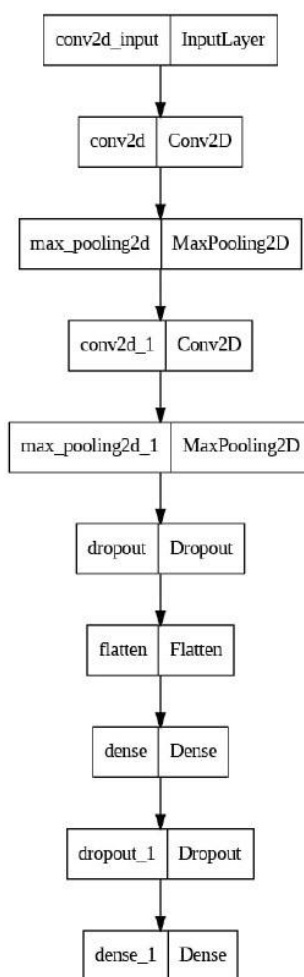


Рис. 1. – Архитектура модели

К скрытым слоям относятся слои Conv2D и MaxPooling2D. Conv2D является свёрточным слоем [10], в данной модели имеется два таких слоя. В первом слое число фильтров было задано равным 64, размер ядра (5,5). Во втором слое число фильтров равно 128, размер ядра остался таким же. Данные значения были подобраны экспериментально. Также после каждого слоя Conv2D идёт слой MaxPooling2D, он снижает размерность изображения со взятием максимального значения пикселя, попадающего в сетку размером pool_size. Как наиболее оптимальное значение, pool_size был взят размером (2,2).

Данная модель также содержит два слоя Dropout. Dropout нужен для предотвращения переобучения нейронной сети, путём исключения определённого процента случайных нейронов, данный процент задаётся в

слое. В ходе анализа точности и ошибок обучения нейросети для первого Dropout-слоя было выбрано 30 процентов, для второго - 40 процентов.

После создания модели свёрточной нейронной сети её необходимо обучить.

Обучение нейронной сети проводилось в несколько этапов. На первом этапе обучения точность нейронной сети на обучающей выборке составила 81 процент, точность на валидационной (проверочной) выборке составила 72 процента. График изменения потерь и точности в ходе обучения на первом этапе изображен на рисунке 2.

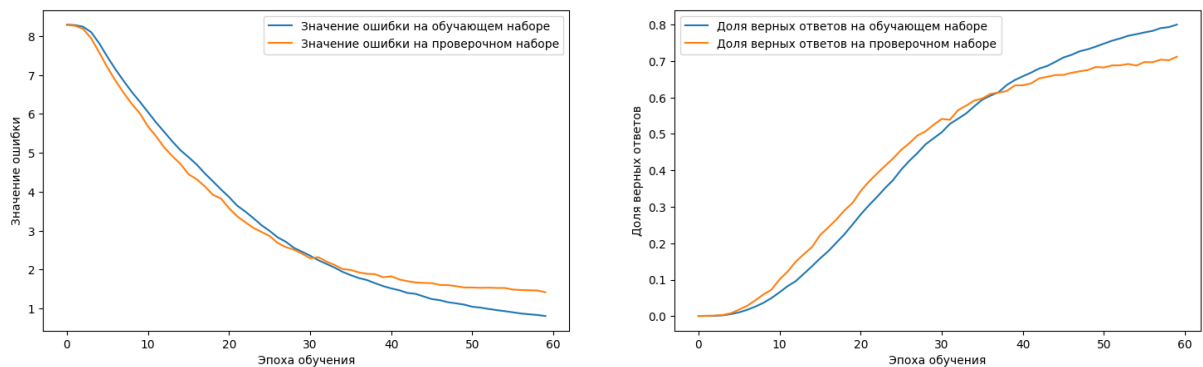


Рис. 2. – График потерь и точностей на первом этапе обучения

На втором этапе обучения точность нейронной сети на обучающей выборке составила 90 процентов, точность на валидационной (проверочной) выборке составила 77 процентов (рис. 3).

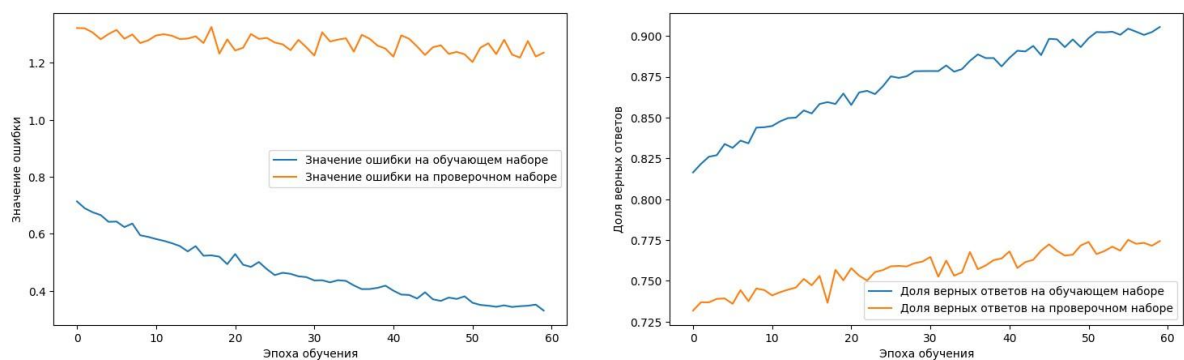


Рис. 3. – График потерь и точностей на втором этапе обучения

На третьем этапе обучения точность нейронной сети на обучающей выборке составила 92 процента, точность на валидационной (проверочной) выборке составила 78 процентов (рис. 4).

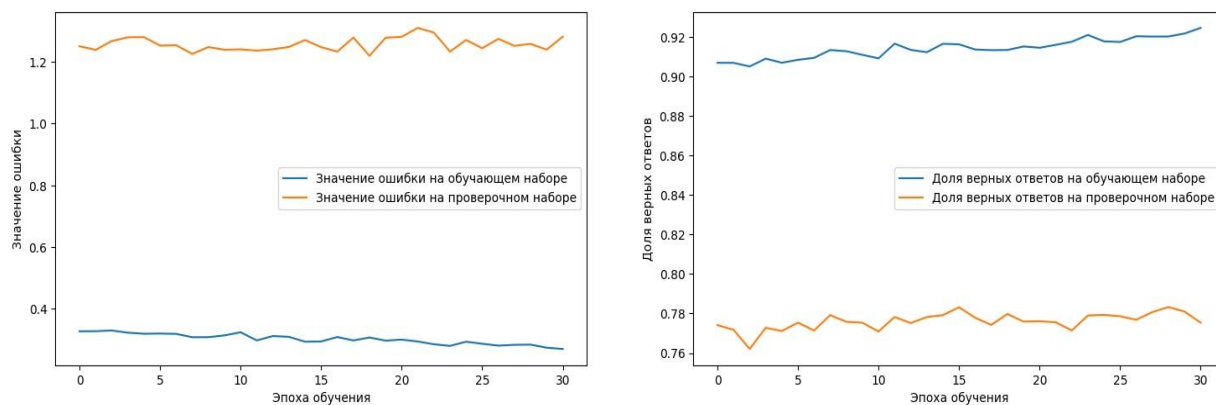


Рис. 4. – График потерь и точностей на третьем этапе обучения

После обучения, итоговая точность данной модели свёрточной нейронной сети на валидационной выборке составила 78 процентов, точность на обучающей выборке составила 92 процента.

Интеграция модели нейронной сети в веб-приложение

В качестве сервиса для онлайн хостинга был выбран российский сервис Amvera Cloud. Amvera Cloud представляет собой облако для развертывания ботов, сайтов и других приложений.

Для начала была выполнена загрузка и распаковка архива с некоторыми изображениями людей из набора данных, словаря ResultsMap и сохранённой модели нейросети.

После этого необходимо создать следующие файлы:

– prediction.py. В данном файле производится загрузка обученной модели нейросети и загрузка словаря ResultsMap в переменную ResultMap. Также в файле имеются функции image_path() и predict(). В функции image_path() происходит создание папки и сохранение в неё загруженного изображения, на выходе функция выдаёт путь к этому изображению. В

функции `predict()` производится обработка изображения и распознавание человека с помощью нейросети.

- `app.py`, который является файлом приложения. В нём задано название приложения, добавлен загрузчик для файлов с расширением `.png`.

- `setup.sh`, который является файлом конфигурации `Streamlit`, и `requirements.txt`, представляющий собой список необходимых пакетов. Они, совместно со скриптом `app.py`, составляют установочный набор, на который будет ориентироваться хостинг при развертывании приложения.

После создания скриптов и файлов был создан проект в `Amvera Cloud`.

Для развертывания приложения в `Amvera Cloud` была проведена настройка и заполнение репозитория `Git`.

Далее был создан файл `amvera.yml`, который является файлом конфигурации и используется для сборки и запуска приложения.

После этого были созданы переменные с именем пользователя, паролем, названием созданного репозитория в `Amvera Cloud` и проведено клонирование пустого удаленного репозитория.

Далее ранее созданные скрипты и файлы, а также загруженные файлы были перемещены в локальный репозиторий.

Перед развертыванием приложения необходимо создать коммит. Для этого проводится инициализация локального репозитория, индексирование и добавление всех файлов в промежуточное хранилище `Git`, а также сохранение изменений в локальном репозитории. Последним шагом работы с `Git` является отправка файлов в удаленный репозиторий. После отправки файлов начинается автоматическая сборка и развертывание проекта в `Amvera Cloud`.

В результате было получено веб-приложение, пример его работы изображен на рис. 5.

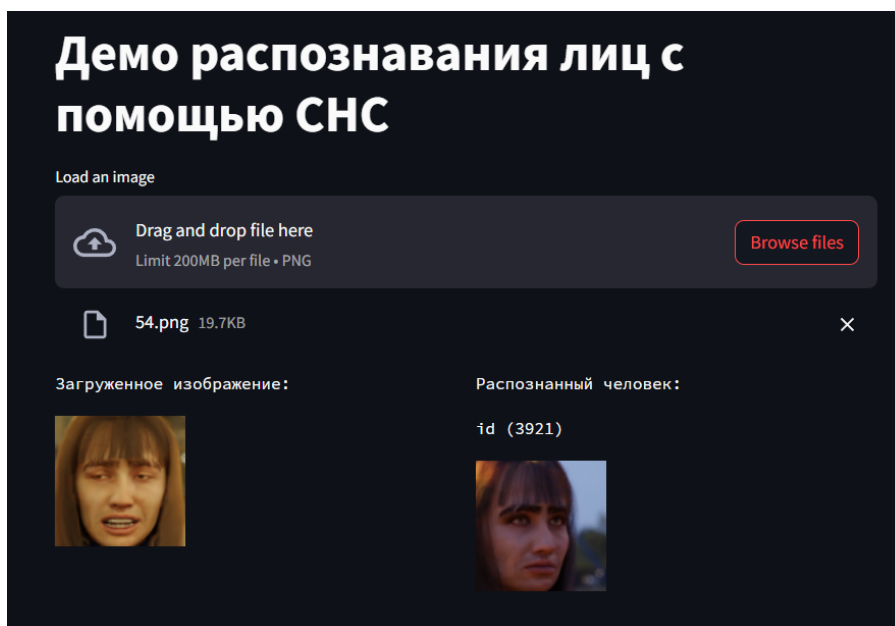


Рис. 5. – Пример работы веб-приложения

Заключение

В данной работе:

1. Был проведен обзор методов распознавания лиц: нейронные сети, Скрытые марковские модели, метод главных компонент, метод гибкого сравнений на графах, активные модели внешнего вида.
2. Была разработана и обучена модель свёрточной нейронной сети для решения задачи распознавания лиц.
3. Проведено исследование эффективности работы разработанной свёрточной нейронной сети. Точность распознавания лиц разработанной нейронной сети на наборе данных DigiFace-1M составила 78%.
4. Было создано веб-приложение, в которое была интегрирована свёрточная нейронная сеть. Приложение было развернуто и запущено на облачном сервисе Amvera Cloud.

Литература

1. Дергачев В.В., Александров А.А. Методы анализа и структурированного распознавания лиц в естественных условиях //



Инженерный вестник Дона. - 2017. - №4. URL:
ivdon.ru/ru/magazine/archive/n4y2017/4549.

2. Кузнецов Д.А., Дамм В.А., Кузнецов А.В., Трегубов Р.Б. Сравнительный анализ популярных методов распознавания лиц на изображениях // Научный результат. Информационные технологии. - 2019. - №4(4). - С. 3-9.

3. Романюта Д.Ю., Коваленко А.В., Шарпан М.В. Исследование алгоритма Виолы-Джонса для разработки системы распознавания лиц с помощью нейронных сетей // Инженерный вестник Дона. - 2024. - №1. URL:
ivdon.ru/ru/magazine/archive/n1y2024/8941.

4. Лекун Я. Как учится машина: революция в области нейронных сетей и глубокого обучения. - М.: Альпина PRO, 2021. - 335 с.

5. Шевцов Н.И., Шиянов Д.А. Аспекты использования различных методов распознавания лиц в современных системах безопасности // Молодой учёный. - 2021. - №47. - С. 8-9.

6. Мухамадиева К.Б. Сравнительный анализ алгоритмов распознавания лиц // Современные материалы, техника и технологии. - 2017. - №7(15). - С. 58-63.

7. 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. - №4(1). - С. 38-46.

8. Левчук С.А., Якименко А.А. Исследование характеристик алгоритмов распознавания лиц // Сборник научных трудов НГТУ. - 2018. - №3-4(93). - С. 40-58.

9. Alabort-i-Medina, J. and S. Zafeiriou, A Unified Framework for Compositional Fitting of Active Appearance Models. International Journal of Computer Vision, 2017, 121(1). pp. 26-64.

10. Tareen, S. and K.F. Tareen, Convolutional Neural Networks for Beginners. SSRN Electronic Journal, 2023. № 3. URL: researchgate.net/publication/373813288_Convolutional_Neural_Networks_for_Begginsers. (Дата обращения 03.11.2024)

References

1. Dergachev V.V., Aleksandrov A.A. Inzhenernyj vestnik Dona, 2017, №4. URL: ivdon.ru/ru/magazine/archive/n4y2017/4549.
2. Kuznecov D.A., Damm V.A., Kuznecov A.V., Tregubov R.B. Nauchnyj rezul'tat. Informacionnye tehnologii. 2019. №4(4). pp. 3-9.
3. Romanjuta D.Ju., Kovalenko A.V., Sharpan M.V. Inzhenernyj vestnik Dona, 2024, №1. URL: ivdon.ru/ru/magazine/archive/n1y2024/8941.
4. Lekun Ja. Kak uchitsja mashina: revoljucija v oblasti nejronnyh setej i glubokogo obuchenija [How Machines Learn: The Revolution in Neural Networks and Deep Learning]. M.: Al'pina PRO, 2021. 335 p.
5. Shevcov N.I., Shijanov D.A. Molodoj uchjonyj. 2021. №47. pp. 8-9.
6. Muhamadieva K.B. Sovremennye materialy, tehnika i tehnologii. 2017. №7(15). pp. 58-63.
7. Kuznecov D.A., Nikol'skij P.G., Rachkov D.S., Kuznecov A.V., Hahamov A.P. Nauchnyj rezul'tat. Informacionnye tehnologii. 2019. №4(1). pp. 38-46.
8. Levchuk S.A., Jakimenko A.A. Sbornik nauchnyh trudov NGTU. 2018. №3-4(93). pp. 40-58.
9. Alabort-i-Medina, J. and S. Zafeiriou, A International Journal of Computer Vision, 2017, 121(1). P. 26-64.
10. Tareen, S. and K.F. Tareen. SSRN Electronic Journal, 2023. No 3. URL: researchgate.net/publication/373813288_Convolutional_Neural_Networks_for_Begginsers. (Date accessed 03.11.2024)

Дата поступления: 3.10.2024 Дата публикации: 17.11.2024
