

**Жаров Эльдар Сергеевич**, студент магистратуры,  
Поволжский государственный университет  
телекоммуникаций и информатики  
Zharov Eldar Sergeevich, Master's Student,  
Povolga state university of telecommunications and informatics

**Пальмов Сергей Вадимович**, доц. каф. ИСТ, к.т.н.,  
Поволжский государственный университет  
телекоммуникаций и информатики  
Palmov Sergey Vadimovich,  
Povolga state university of telecommunications and informatics

**ИССЛЕДОВАНИЕ ПРИМЕНИМОСТИ ГЛУБОКИХ НЕЙРОННЫХ СЕТЕЙ  
ПРИ РЕШЕНИИ ЗАДАЧИ КЛАССИФИКАЦИИ ЧЕЛОВЕЧЕСКИХ ЛИЦ  
A STUDY ON THE APPLICABILITY OF DEEP NEURAL  
NETWORKS FOR HUMAN FACE CLASSIFICATION**

**Аннотация.** В работе рассматриваются современные методы распознавания и классификации человеческих лиц с использованием глубоких нейронных сетей. Проведён анализ ключевых архитектур, включая AlexNet, VGG, ResNet, а также специализированных решений, таких как MTCNN и FaceNet. Описаны принципы работы свёрточных нейронных сетей, их преимущества перед классическими алгоритмами и методы оптимизации для работы на мобильных устройствах. Отдельное внимание уделено вопросам быстродействия и точности, а также перспективам развития технологий компьютерного зрения.

**Abstract.** This paper reviews modern methods of human face recognition and classification using deep neural networks. It analyzes key architectures, including AlexNet, VGG, ResNet, as well as specialized solutions such as MTCNN and FaceNet. The principles of convolutional neural networks, their advantages over classical algorithms, and optimization methods for mobile devices are described. Special attention is paid to performance and accuracy issues, as well as to the prospects for the development of computer vision technologies.

**Ключевые слова:** Глубокие нейронные сети, классификация лиц, распознавание лиц, свёрточные нейронные сети, компьютерное зрение, извлечение признаков, обучение с учителем, датасет лиц.

**Keywords:** Deep neural networks, face classification, face recognition, convolutional neural networks, computer vision, feature extraction, supervised learning, face dataset.

**Введение**

В последние годы проблема автоматического распознавания лиц вызывает всё больший интерес как у учёных, так и у разработчиков коммерческого ПО. Объясняется это просто: количество фотографий и видеозаписей в интернете растёт лавинообразно, и обрабатывать их вручную уже невозможно [1]. Люди – одни из самых частых объектов на фото, поэтому умение быстро и точно определять, кто именно изображён, становится важной задачей для поисковых систем, социальных сетей и систем безопасности [1].

Помимо чисто идентификационных задач, существует потребность в классификации лиц по различным признакам, например по полу, возрасту или расовой принадлежности. Эти характеристики относятся к разным типам данных: пол – это простой бинарный выбор, раса – выбор из нескольких вариантов, а возраст – числовая величина, которую нужно угадать максимально точно [1]. Самым сложным считается именно определение возраста, потому что



на внешность человека влияет не только количество прожитых лет, но и образ жизни, наследственность и даже настроение в момент съёмки [1, 2].

Цель нашей работы – разобраться, насколько хорошо современные нейросети справляются с классификацией лиц, и какие методы для этого используются. В обзоре мы рассмотрим как классические алгоритмы, заложившие основу, так и новейшие разработки, позволяющие добиваться высокой точности даже на обычных смартфонах.

### **1. Классические методы распознавания лиц**

До того, как нейросети стали главным инструментом компьютерного зрения, учёные придумали немало способов сравнивать лица между собой. Один из самых известных – метод главных компонент, который часто называют «собственными лицами». Его суть в том, чтобы сжать изображение до набора главных признаков, которые и используются для сравнения [5]. Если фотографии сделаны в хороших условиях, точность может достигать 90 %, но стоит измениться освещению или человеку чуть повернуть голову – результат сильно падает [4, 5].

Чуть более продвинутый подход – линейный дискриминантный анализ, или метод Фишера. Он работает не просто со сжатием, а старается найти такие признаки, которые лучше всего разделяют разные классы [5]. Благодаря этому он лучше справляется с помехами вроде очков или бороды и показывает точность около 94 %. Правда, как он поведёт себя при смене ракурса, изучено недостаточно [5].

Есть и совсем другой способ – сравнение лиц с помощью графов. Представьте, что на лице расставлены точки в ключевых местах (уголки глаз, кончик носа и так далее), и эти точки соединены линиями. Получается своеобразная сетка. Один граф остаётся неизменным, а второй пытаются на него наложить, слегка деформируя [4, 5]. Точность тут высокая – до 97 %, и метод неплохо переносит мимику и небольшие повороты головы. Но есть огромный минус: считать всё это приходится очень долго, и чем больше лиц в базе, тем медленнее работает система [4, 5].

Отдельно стоит упомянуть алгоритм Виолы-Джонса, который в своё время стал настоящим прорывом в области детектирования объектов. Он использует специальные признаки (их называли признаками Хаара) и каскады классификаторов, обученных с помощью AdaBoost [5]. Главное его достоинство – скорость: он работает в реальном времени и редко ошибается. Но если лицо повернуто больше, чем на 30 градусов, алгоритм его уже не видит, что для современных систем часто неприемлемо [5].

Все эти методы, несмотря на свои плюсы, имеют серьёзные ограничения, которые удалось преодолеть только с приходом глубокого обучения.

### **2. Как устроены свёрточные нейронные сети**

Сегодня главный инструмент для работы с изображениями – это свёрточные нейронные сети (их часто обозначают аббревиатурой CNN). Идею таких сетей предложили ещё в конце 80-х, вдохновившись устройством зрительной коры мозга. Там есть простые клетки, реагирующие на линии и края, и сложные, которые собирают из этих линий целостные образы [7, 10]. Название «свёрточные» они получили из-за операции свёртки – способа выделять на картинке локальные признаки, такие как границы предметов, углы или текстуры, причём делать это независимо от того, где именно на фото находится объект [9, 10].

Свёрточная сеть собирается из нескольких типов слоёв, как конструктор. Основной тип – свёрточный слой. В нём есть набор обучаемых фильтров (их ещё называют ядрами). Каждый фильтр сканирует изображение и определяет, насколько сильно его «любимый» признак присутствует в каждом участке картинки. Результат работы фильтра – карта признаков [8, 9]. После свёртки обычно идёт слой подвыборки, или пулинга. Чаще всего используется MaxPooling, который оставляет только самые яркие значения, отбрасывая всё лишнее. Это уменьшает размер картинки и помогает сети не обращать внимания на мелкие дефекты [7, 9].



Очень важно, что свёрточные сети умеют выстраивать иерархию признаков. Первые слои учатся распознавать простые вещи: линии, перепады яркости, цветовые пятна. Следующие слои комбинируют их в более сложные формы – круги, прямоугольники, части объектов. А самые глубокие слои уже оперируют абстрактными понятиями: «это похоже на глаз», «это похоже на колесо» [8, 9]. Чем глубже сеть, тем более сложные концепции она может усвоить, но с глубиной возникают проблемы с обучением – градиенты затухают, и сеть перестаёт учиться. Чтобы это исправить, придумали специальные архитектурные решения, например остаточные связи [10].

В конце сети обычно ставят один или несколько полносвязных слоёв – они уже и выносят финальный вердикт, что же изображено на картинке [7, 10]. Чтобы сеть не переобучалась (то есть не запоминала примеры, а училась обобщать), используют разные приёмы: например, Dropout, когда часть нейронов случайно отключается во время обучения, или раннюю остановку, если точность на проверочных данных перестала расти [7].

### 3. Современные архитектуры для распознавания лиц

За последние годы придумано множество вариантов свёрточных сетей, которые значительно обошли классические методы по точности и скорости. Одной из первых ласточек стала AlexNet – она победила в конкурсе ImageNet в 2012 году и показала, что глубокие сети с большим числом фильтров способны на то, что не под силу традиционным алгоритмам [10].

Потом появилась VGG-16, где большие фильтры заменили на связки из маленьких (3×3). Это позволило делать сети глубже, не увеличивая слишком сильно количество параметров [10]. Следующим важным шагом стала ResNet – разработка Microsoft Research. Там придумали остаточные связи, которые позволяют информации проходить в обход нескольких слоёв, что решает проблему затухания градиента. Благодаря этому удалось создавать сети с десятками и даже сотнями слоёв, которые при этом хорошо обучаются [9, 10].

Для задачи распознавания лиц разрабатывают и специальные архитектуры. Например, сеть MTCNN состоит из трёх каскадов. Первый (P-net) быстро пробегает по картинке и находит области, где могут быть лица. Второй (R-net) отсеивает лишнее, а третий (O-net) уже точно определяет границы лица и находит ключевые точки – глаза, нос, уголки рта [4]. Благодаря такому каскадному подходу удаётся сохранить высокую скорость работы и при этом получать отличное качество [4].

Другая известная сеть – FaceNet от Google. Она относится к классу сиамских сетей и учится особым образом: создаёт такое пространство, где фотографии одного человека лежат рядом, а разных – далеко друг от друга [4, 9]. Для этого на этапе обучения используют тройки снимков: опорное изображение, ещё одно фото того же человека и фото кого-то другого. Сеть старается уменьшить расстояние между первыми двумя и увеличить – между первым и третьим [4, 9].

Ещё один важный приём – использование функций потерь с отступами, например ArcFace или CosFace. Они немного меняют стандартную формулу обучения так, чтобы векторы одного класса не просто были близки, а имели дополнительный зазор от чужих векторов [2]. Это помогает сети лучше различать похожих людей и делает кластеры каждого человека более компактными [2].

### 4. Как ускорить нейросети и запустить их на телефоне

Высокая точность часто достигается ценой огромных вычислительных затрат. Например, сеть ResNet100, которая отлично показывает себя на тестах, обрабатывает одно изображение почти полсекунды на обычном процессоре. Для видеопотока это слишком медленно [2]. Чтобы использовать распознавание лиц в телефонах, домофонах или камерах, нужны более лёгкие и быстрые модели.



Один из способов – дистилляция. Суть в том, что большая и точная сеть (учитель) обучает маленькую сеть (ученика). Ученик старается подражать не только ответам учителя, но и его внутренним представлениям, и в итоге показывает результаты гораздо лучше, чем если бы его учили просто на примерах [2, 9]. В одной из работ предложили для сетей с ArcFace особый метод дистилляции: веса последнего слоя (центры классов) просто копируют из большой сети в маленькую и замораживают, а потом учат ученика так, чтобы его векторы были направлены под теми же углами к этим центрам [2]. Это позволило поднять точность маленькой сети MobileFaceNet на сложном тесте MegaFace с 90,6 % до 91,7 %, что совсем немного уступает большой сети [2].

Другой способ – квантование. Обычно веса хранятся как 32-битные числа с плавающей точкой. Если перевести их в 8-битные целые числа, модель занимает меньше места и считает быстрее, хотя точность может чуть упасть [2, 3]. Иногда после квантования и дообучения сеть даже лучше работает на реальных данных, особенно если они зашумлены [2].

Ещё можно сэкономить ресурсы ещё до того, как запускать сеть. Например, в системе AmphibianDetector предложили использовать промежуточные карты признаков самого детектора, чтобы определять, есть ли на кадре движение. Если ничего не движется, обработку можно остановить досрочно [2]. Это не только ускоряет работу, но и снижает количество ложных срабатываний, когда система пытается найти лицо там, где его нет [2].

### **5. Где уже применяют распознавание лиц**

Технологии распознавания лиц давно вышли из лабораторий и используются в самых разных сферах. В маркетинге, например, с их помощью анализируют аудиторию и подбирают рекламу. Сервис YouScan умеет находить упоминания брендов не только в тексте, но и на фотографиях в соцсетях – так выяснили, что коты чаще фотографируются с техникой Apple, а собаки – с одеждой Adidas [6].

В системах видеонаблюдения распознавание лиц помогает искать преступников. В Москве такая система работает с 2017 года и уже помогла задержать десятки правонарушителей [6]. В Китае пошли дальше: там камеры распознают не только лица, но и марки машин, элементы одежды – эту информацию тоже можно использовать для разных целей [6].

Медицина – ещё одна область, где нейросети приносят реальную пользу. Они анализируют снимки КТ, МРТ и даже биопсийные препараты, находя раковые клетки с точностью до 95–97 %, что часто выше, чем у человека [6, 8]. Например, сеть Inception от Google показывает чувствительность 92 % при поиске метастазов, в то время как у врачей этот показатель около 72 % [6].

В автомобилях распознавание объектов стало частью систем помощи водителю. Одной камеры достаточно, чтобы видеть пешеходов, знаки и светофоры в реальном времени [6]. Специальные процессоры вроде Visconti от Toshiba строят трёхмерную модель дороги и могут заметить даже то, чему сеть специально не учили, например упавшее дерево [6].

Дроны с системами распознавания используют для инспекции линий электропередач, поиска поломок в труднодоступных местах и даже для съёмки с автослежением [6].

Есть и более специфические задачи, например биометрическая криптография, где по лицу нужно получить уникальный ключ для шифрования. В России эта область регулируется ГОСТами [3]. Оказалось, что нейросетевые преобразователи справляются с этим не хуже классических методов, а за счёт встроенной регуляризации даже меньше переобучаются [3].

### **6. Проблемы и ограничения**

При всех успехах у нейросетей хватает и недостатков. Главный – им нужно очень много размеченных данных. Чтобы сеть научилась хорошо обобщать, нужны тысячи, а то и миллионы примеров [9]. Если данных мало, приходится брать готовую модель, обученную на чём-то другом, и доучивать её под свою задачу, что получается не всегда [9].



Вторая проблема – вычислительная сложность. Даже небольшие сети для мобильных устройств содержат миллионы параметров, и без специальных процессоров они работают медленно [2, 3]. Квантование и дистилляция помогают, но не решают проблему полностью.

Не стоит забывать и про этическую сторону. Массовое видеонаблюдение с функцией распознавания лиц может нарушать приватность и создавать угрозу тотального контроля [6]. В разных странах идут споры о том, где проходит граница допустимого, и пока единого мнения нет.

К тому же нейросети могут ошибаться, особенно если условия съёмки отличаются от тех, на которых их учили. Например, сеть, обученная на лицах европеоидов, будет хуже работать на других расах [1]. Поэтому при разработке важно учитывать разнообразие данных и тестировать систему в реальных условиях.

### **Заключение**

Мы рассмотрели, как менялись методы распознавания лиц: от простых сравнений и графов до сложных нейросетевых архитектур. Понятно, что свёрточные сети сегодня – это основной инструмент, и благодаря им точность классификации достигла очень высоких значений.

Особенно интересны разработки, которые позволяют запускать сложные модели на мобильных устройствах: дистилляция, квантование, эффективные архитектуры вроде MobileFaceNet. Это открывает дорогу для массового внедрения технологии в повседневную жизнь – от разблокировки телефонов до умных домофонов.

Но проблемы остаются: нужны большие данные, мощные процессоры, и важно не переступить этическую грань. Дальнейшее развитие, скорее всего, пойдёт по пути создания ещё более компактных моделей, улучшения обобщения на разнородных данных и выработки правил использования.

В любом случае, можно уверенно сказать, что глубокие нейронные сети доказали свою состоятельность в задаче классификации лиц, и сфера их применения будет только расширяться.

### *Список литературы:*

1. Рыбинцев А.В. Исследование, модификация и разработка методов компьютерного зрения для задач определения атрибутов личности по изображению лица: дис. ... канд. техн. наук. – М., 2018.
2. Свитов Д.В. Оптимизация производительности свёрточных нейронных сетей в системе распознавания лиц: дис. ... канд. техн. наук. – Новосибирск, 2023.
3. Куликов С.В., Захаров О.С., Андреев Д.Ю. Исследование возможности совместного применения нейросетевого преобразователя биометрия-код и глубокой свёрточной нейронной сети в распознавании лиц // Информационные технологии. – 2018. – № 2.
4. Безгачев Ф.В. Распознавание лиц на основе нейронных сетей: современные технологии // Научная электронная библиотека «КиберЛенинка». – 2021. – № 4.
5. Левчук С.А., Якименко А.А. Исследование характеристик алгоритмов распознавания лиц // Сборник научных трудов НГТУ. – 2018. – № 3–4 (93).
6. Топ-5 сфер применения систем распознавания объектов [Электронный ресурс] / Habr.com. – Режим доступа: <https://habr.com/ru/company/toshibarus/blog/433544/> (дата обращения: 13.11.2025).
7. Комкова Е.С., Унжакова С.В. Актуальность применения сверточной нейронной сети для решения задачи классификации эмоций на изображениях // Научные высказывания. – 2024. – № 17 (64).



8. Как свёрточные нейросети имитируют работу мозга [Электронный ресурс] / Яндекс Практикум. – Режим доступа: <https://practicum.yandex.ru/blog/svertochnye-neyronnye-seti/> (дата обращения: 12.11.2025).

9. Шолле Ф. Глубокое обучение на Python. – СПб.: Питер, 2018.

10. Маршалко Д.А., Кубанских О.В. Архитектура свёрточных нейронных сетей // Ученые записки Брянского государственного университета. – 2019. – № 4 (16).

