

МИНОБРНАУКИ РОССИИ
Федеральное государственное бюджетное образовательное учреждение
высшего образования

**«САРАТОВСКИЙ НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ
ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ
ИМЕНИ Н. Г. ЧЕРНЫШЕВСКОГО»**

Кафедра дискретной математики и информационных технологий

**ИСПОЛЬЗОВАНИЕ СВЕРТОЧНЫХ НЕЙРОННЫХ СЕТЕЙ ДЛЯ
ОЦЕНКИ ВОЗРАСТА**

АВТОРЕФЕРАТ МАГИСТЕРСКОЙ РАБОТЫ

студентки 2 курса 271 группы
направления 09.04.01 — Информатика и вычислительная техника
факультета КНиИТ
Ситниковой Русланы Эдуардовны

Научный руководитель
доцент, к. ф.-м. н.

И. Д. Сагаева

Заведующий кафедрой
доцент, к. ф.-м. н.

Л. Б. Тяпаев

Саратов 2025

ВВЕДЕНИЕ

В настоящее время нейронные сети имеют высокий спрос и используются в большом спектре различных задач. Существует несколько видов нейронных сетей, каждый из которых имеет свои определенные особенности. Для работы с изображениями обычно используются сверточные нейронные сети.

Из изображения лица человека можно получить различную информацию. Например, по фотографии можно узнать пол, примерный возраст, расу, эмоциональное состояние.

Оценка возраста человека по фотографии может быть использована в разных сферах:

- в коммерции для подбора подходящих рекомендаций пользователю;
- для поддержания возрастных ограничений, например, для ограничения доступа к различным ресурсам;
- при расследовании преступлений или при поиске людей по фотографии из базы данных;
- в здравоохранении;
- в сфере развлечений.

Поэтому актуальными являются исследования, направленные на оценку возраста человека по различным видам изображений с помощью сверточных нейронных сетей.

Целью данной работы является разработка, обучение и анализ работы нейронной сети, способной по изображению лица человека определить его примерный возраст.

Для реализации данной цели необходимо решить следующие задачи.

1. Изучить общие сведения о сверточных нейронных сетях.
2. Рассмотреть существующие аналоги решений для оценки возраста и их особенности.
3. Найти датасеты, содержащие изображения лиц людей в подходящем формате.
4. Подобрать архитектуру сверточной нейронной сети и обучить ее.
5. Реализовать аугментацию изображений найденного датасета.
6. Проверить результаты работы нейронной сети на аугментированных изображениях.
7. Проверить результаты работы нейронной сети на изображениях, качество

которых было улучшено в сравнении с оригиналом.

Для сверточной нейронной сети, использованной в данной работе, в ходе экспериментов была подобрана архитектура, показывающая наиболее высокий результат. Также были проведены исследования предсказаний данной модели на различных изображениях: неаугментированных, аугментированных и улучшенных с помощью онлайн-сервиса, использующего для работы технологии искусственного интеллекта.

Структура и объем работ. Для решения поставленных задач выполнена выпускная квалификационная работа, включающая в себя введение, 5 основных глав, заключение, список использованных источников из 37 наименований и 6 приложений. Работа изложена на 97 страницах, содержит 32 рисунка и 12 таблиц.

Первый раздел имеет название «Общие сведения о сверточных нейронных сетях» и содержит информацию, необходимую для построения сверточной нейронной сети, а также для оценки результатов ее работы.

Второй раздел имеет название «Обзор аналогичных исследований» и содержит примеры исследований, в которых были разработаны модели для оценки возраста по изображению.

Третий раздел имеет название «Реализация сверточной нейронной сети». В данной главе описан процесс создания и обучения сверточной нейронной сети на оригинальных изображениях датасетов UTKFace и MORPH.

Четвертый раздел имеет название «Аугментация изображений и анализ работы нейронной сети» и содержит описание реализованного процесса аугментации изображений датасета MORPH, а также результаты предсказаний модели на аугментированных изображениях.

Пятый раздел имеет название «Проверка работы нейронной сети на улучшенных изображениях» и содержит результаты предсказаний на 100 изображениях датасета MORPH, которые были улучшены с помощью стороннего сервиса Imglarger Online.

Выпускная квалификационная работа заканчивается заключением, списком использованных источников, а также приложениями с кодом А–Е.

КРАТКОЕ СОДЕРЖАНИЕ РАБОТЫ

Общие сведения о сверточных нейронных сетях. Нейронные сети в общем смысле — это сети запрограммированных нейронов, имитирующие работу человеческого мозга. Такие сети содержат первый входной слой, последний выходной и неограниченное количество так называемых скрытых слоев, расположенных между ними. Нейроны в искусственных сетях представляют собой наборы входных данных, веса и функцию активации. Таким образом, каждый нейрон является математической функцией, представляющей собой функционирование биологического нейрона.

Главной особенностью нейронных сетей является то, что при их использовании не нужно создавать набор правил для решения задачи — данный набор формируется в процессе обучения. Устройство нейронных сетей похоже на устройство человеческого мозга тем, что информация в них поступает из «окружающей среды», а показателем накопления знаний служат веса нейронов.

Сверточные нейронные сети или CNN (Convolutional Neural Network) используют математическую операцию свертки для работы с изображениями. В сверточных нейронных сетях применяются различные типы слоев:

- входной слой (Input layer);
- сверточный слой (Convolution layer);
- субдискретизационный слой (Pooling layer);
- слой с операцией выравнивания (Flatten);
- полносвязный или выходной слой.

При работе с моделью может возникнуть эффект переобучения — состояние модели, при котором она «запоминает» обучающий набор данных, на котором достигаются хорошие результаты, но оценка работы на тестовом наборе при этом низкая. Один из вариантов избежания проблемы переобучения — применение метода аугментации существующих изображений. Аугментация — это искусственное увеличение на основе преобразования с сохранением выходных меток. Способами преобразования изображения являются добавление различных шумов, повороты и отображение, масштабирование и смещение, изменение цвета, различные варианты размытия.

Для оценки работы модели регрессии могут быть использованы следующие метрики:

- среднеквадратичная ошибка или Mean Squared Error (MSE), представлен-

ная формулой 1;

- средняя абсолютная ошибка или Mean Absolute Error (MAE), представленная формулой 2.

$$MSE = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2, \quad (1)$$

где n — количество объектов, y_i — истинное значение, $\hat{y}_i$ — предсказанное значение.

$$MAE = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i|, \quad (2)$$

где n — количество объектов, y_i — истинное значение, $\hat{y}_i$ — предсказанное значение.

Обзор аналогичных исследований. В данном разделе приведены ссылки и краткие описания иных исследований, посвященных оценке возраста по изображению.

Реализация сверточной нейронной сети. Для обработки датасетов и реализации нейронной сети использован язык программирования Python 3.9.6. Обучение нейронной сети производилось с помощью Google Collab — сервиса для исполнения файлов Jupyter Notebook. Для обучения использовалась среда выполнения с ускорителями T4.

Предварительная обработка изображений состояла из нескольких шагов.

1. Выделение лица на изображении с помощью каскадов Хаара;
2. Изменение размеров изображения до 224 на 244 пикселя;
3. Извлечение возраста из названия файла;
4. Добавление в датасет новой строки, соответствующей обработанному изображению.

Для обучения модели были использованы датасеты UTKFace и MORPH, приведенные к возрастному диапазону от 10 до 60 лет.

Итоговая нейронная сеть состоит из пятнадцати слоев. Ядро свертки для всех сверточных слоев имеет одинаковый размер 3 на 3 пикселя, на каждом сверточном слое используется функция активации ReLU, изображение не теряет размерность после свертки. Каждый из субдискретизационных слоев имеет ядро 2 на 2 пикселя, на слоях данного типа не используется функция акти-

вации. Перед каждым полносвязным слоем применяется операция дропаут с вероятностью исключения нейронов равной 0.5. На вход нейронная сеть принимает трехканальные изображения 224 на 224 пикселя, выход нейронной сети является числом, соответствующим предсказанному по изображению возрасту. Архитектура используемой сверточной нейронной сети представлена в таблице 1.

Таблица 1 – Архитектура используемой в работе сверточной нейронной сети

Слой	Размерность выхода	Количество параметров
Conv2D	(None, 224, 224, 32)	896
MaxPooling2D	(None, 112, 112, 32)	0
Conv2D	(None, 112, 112, 64)	18,496
MaxPooling2D	(None, 56, 56, 64)	0
Conv2D	(None, 56, 56, 128)	73,856
MaxPooling2D	(None, 28, 28, 128)	0
Conv2D	(None, 28, 28, 256)	295,168
MaxPooling2D	(None, 14, 14, 256)	0
Flatten	(None, 50176)	0
Dropout	(None, 50176)	0
Dense	(None, 512)	25,690,624
Dropout	(None, 512)	0
Dense	(None, 256)	131,328
Dropout	(None, 256)	0
Dense	(None, 1)	257

При обучении в качестве функции потерь использовалась средняя абсолютная ошибка, оптимизатор Adam со скоростью обучения, равной $5e - 4$, параметр `batch_size` был задан значением 4. При обучении было указано максимальное количество эпох — 30, а также в качестве меры предупреждения переобучения была использована функция ранней остановки, принудительно останавливающая обучение, если минимальное значение функции потерь не обновляется в течение трех эпох.

Нейросеть получает на вход датасет, состоящий из двух столбцов — первый столбец `Faces` хранит относительный путь до изображения, второй столбец `Ages` хранит значения возраста, соответствующего изображению. Пример датасета приведен на рисунке 1.

Нейронная сеть с одной и той же архитектурой и параметрами была обучена с нуля несколько раз на разных наборах данных. Для тестирования исполь-

	Faces	Ages
0	Train/052229_1M43.JPG	43
1	Train/336861_01F35.JPG	35
2	Train/127390_05F36.JPG	36

Рисунок 1 – Формат датасета, который принимает на вход нейронная сеть

зовались выборки из 13391 тысяч изображений датасета UTKFace, 13391 тысяч изображений датасета MORPH и 13391 изображений из смешанного датасета, который содержал изображения как из датасета UTKFace, так и из датасета MORPH.

Первоначально модель была обучена на обработанном датасете UTKFace. Данный датасет содержал 27317 изображений, 80% которых использовались для обучения, оставшиеся — для валидации. Обучение длилось 14 эпох. Модель показала значение MAE, равное 7.11 на изначальной валидационной выборке. Лучший результат по двум метрикам модель показала на выборке из датасета UTKFace (значение MAE равно 5.2313), хуже — на смешанном датасете (значение MAE равно 6.6183), самый худший результат — на датасете MORPH (значение MAE равно 7.0497).

При обучении на датасете MORPH модель показала следующие результаты. Датасет содержал 37845 изображений, разделенных в соотношении 8 к 2 на обучающую и валидационную выборки. Обучение длилось 8 эпох, значение MAE на исходной валидационной выборке равно 5. При тестировании лучший результат по двум метрикам модель показала на выборке из датасета MORPH (значение MAE равно 4.4323), хуже — на смешанном датасете (значение MAE равно 6.3674), самый худший результат — на датасете UTKFace (значение MAE равно 7.3989).

Модель также была с нуля обучена на смешанном датасете, содержащем изображения как из датасета UTKFace, так и изображения датасета MORPH. Датасет, на котором происходило обучение, содержал 66954 изображений, 20% которого использовались для валидации. Обучение данной модели длилось 15 эпох, после чего была произведена его ранняя остановка. Метрика MAE для данной модели на исходной валидационной выборке составила 5, как и при обучении на датасете MORPH. При тестировании лучший результат модель показала на выборке из датасета MORPH (значение MAE равно 3.9045), хуже — на смешанном датасете (значение MAE равно 4.8311), самый худший результат —

на датасете UTKFace (значение MAE равно 4.9674).

Модель также была проверена на датасете MegaAge, выборка из которого содержала 28821 изображение. Лучшую оценку показала модель, обученная на датасете MORPH (значение MAE составило 9.8967), худшую — на смешанном датасете (значение MAE составило 11.8934), а на датасете UTKFace значение MAE составило 11.1244.

Аугментация изображений и анализ работы нейронной сети. Для аугментации была использована часть датасета MORPH, состоящая из 4964 исходных изображений. В качестве модели выбрана сверточная нейронная сеть, которая была обучена на датасете, составленном из изображений датасета MORPH и UTKFace.

Результаты работы модели с добавлением различных размеров участка шума представлены в таблице 2, аугментация проводилась несколько раз для одних и тех же значений области шума. При возрастании размера участка шума увеличивается и ошибка предсказания. При добавлении одних только участков шума средняя абсолютная ошибка модели находится в пределах 4–5 лет, а среднеквадратическая ошибка колеблется в интервале значений от 28 до 37.

Таблица 2 – Оценка работы модели на 5000 изображениях с добавлением шума

Размер обл. шума в процентах от ориг. размера	MAE	MSE
25%	4.2667	30.0553
25%	4.2465	30.0434
25%	4.1757	28.8312
35%	4.3039	30.3342
35%	4.4270	32.5847
35%	4.3402	31.6061
45%	4.2766	29.9809
45%	4.2828	29.8899
45%	4.4277	33.0741
55%	4.7513	37.2019
55%	4.5757	34.5171
55%	4.5741	34.4539

Затем шум на данных изображениях был заменен однотонным белым цветом. При этом метрика MAE на выборке изображений снизилась с 0.87 до 0.40.

Для еще одного эксперимента были использованы следующие возможные преобразования изображений:

- добавление шума областью 0.4 от оригинального размера изображения с вероятностью 75% применения данной аугментации;
- искажение изображения с вероятностью 75% применения данной аугментации;
- искажение перспективы с вероятностью 75% применения данной аугментации.

При добавлении одних только участков шума на изображения средняя абсолютная ошибка оставалась в пределах 4 лет, но при добавлении возможности искажений самого изображения или искажения перспективы средняя абсолютная возросла до 6-ти лет. Это может быть связано с тем, что при искажении изображений изменяются и специфические возрастные черты — возможные морщины у носогубной складки, на лбу, в уголках глаз.

Проверка работы нейронной сети на улучшенных изображениях. Искажать результат предсказаний нейронной сети могут не только дефекты изображений или неправильный ракурс. На определяемый возраст также могут влиять улучшения фотографий — например, применение коррекции изображений или фильтров.

Для проверки работы модели на улучшенных фотографиях были выделены 100 изображений из датасета MORPH с наименьшей ошибкой предсказания.

Для улучшения изображений был использован сервис Imglarger Online, использующий технологии искусственного интеллекта для улучшения изображений, предоставляющий фильтр AI Portrait Retouch, который позволяет ретушировать портреты. Качество изображений после улучшения повысилось без изменения размера изображения, кожа стала выглядеть более гладкой, также повысилась четкость изображения.

Наименьшая ошибка на выборке до применения улучшений составляла 0.006107 лет, наибольшая — 0.376980 лет, в среднем ошибка предсказаний достигала 0.17725. После улучшения наименьшая ошибка составила 0.052406 лет, наибольшая — 26.515064 лет, в отличие от предсказаний на оригинальных изображениях, где ошибка не выходила за рамки одного года. В среднем ошибка предсказаний по улучшенным изображениям составляет 10.116.

ЗАКЛЮЧЕНИЕ

Оценка возраста человека по фотографии является достаточно сложной задачей в силу того, что старение — это индивидуальный процесс, зависящий от таких факторов, как генетика, состояние здоровья, питание, образ жизни, социальные условия. Не менее важным фактором является качество непосредственного изображения человека и качество датасетов, которые могут иметь ошибки в разметке, неправильные метки возраста, плохие изображения.

При решении данной задачи исследователи уделяют особое внимание предобработке изображений, их аугментации, структуре нейронной сети и функции потерь.

Таким образом, следующие задачи становятся особенно важными при разработке приложения для оценки возраста:

- предобработка датасета для обучения;
- предобработка изображения для передачи его на вход нейронной сети;
- разработка структуры нейронной сети;
- извлечение особенностей для оценки возраста;
- подбор функции потерь, которая будет использована для обучения нейронной сети.

В ходе работы была реализована и обучена модель сверточной нейронной сети. Для обучения использовались датасеты UTKFace, MORPH и их комбинация, для тестирования — датасет MegaAge. Перед обучением и тестированием датасеты были обработаны: из изображений были извлечены лица, датасет UTKFace также был очищен от выбросов и аугментирован.

На выборке изображений, принадлежащих известным датасетам, лучший результат показала модель после обучения на смешанном датасете — ее средняя оценка MAE составила 4.57, MSE — 34.38. На изображениях, принадлежащих датасету MegaAge лучший результат показала модель при обучении на датасете MORPH: оценка MAE составила 9.8967, MSE — 142.3571.

Также была произведена оценка работы модели на аугментированных изображениях датасета MORPH, на котором модель показывала лучшие результаты. В качестве преобразований были использованы добавление шума и участков белого цвета, а также искажение самого изображения и перспективы.

При обучении модели средняя абсолютная ошибка составила 3.9045, среднеквадратическая — 25.8131. При добавлении областей шума различной пло-

щади средняя абсолютная ошибка модели возросла до интервала в 4–5 лет, а среднеквадратическая ошибка показала результат в интервале значений от 28 до 37. При замене участков шума на участки однотонного белого цвета значение ошибок уменьшалось. При применении искажений изображения средняя абсолютная ошибка возросла до 6 лет, а среднеквадратическая ошибка — до значения 63.

Улучшение качества изображений тоже может значительно влиять на результат работы модели. При улучшении качества ста изображений датасета MORPH ошибка возросла со среднего значения, равного 0.17725, до значения в 10.116 лет.

Отдельные части магистерской работы были опубликованы в сборнике, индексируемом в РИНЦ:

- 1 Ситникова, Р. Использование сверточных нейронных сетей для определения возраста по изображению лица человека / Р. Ситникова, И. Сагаева // ЕСТЕСТВЕННЫЕ И ТЕХНИЧЕСКИЕ ИССЛЕДОВАНИЯ В РОССИИ И ЗА РУБЕЖОМ: ПРОБЛЕМЫ, ПУТИ СОВЕРШЕНСТВОВАНИЯ. — Краснодар: ООО «КОНТУР СП», 2025. — С. 200–204.

Основные источники информации:

- 1 Толмачев, С. ИССЛЕДОВАНИЕ ВЛИЯНИЯ ФУНКЦИИ АКТИВАЦИИ ИСКУССТВЕННОЙ НЕЙРОННОЙ СЕТИ ПРИ РЕШЕНИИ ЗАДАЧИ РАСПОЗНАВАНИЯ ОБРАЗОВ / С. Толмачев, Л. Гришина // Шаг в науку. — 2021. — Т. 4, № 23. — С. 77–82.
- 2 Юрьевич, А. Д. ПРИМЕНЕНИЕ АУГМЕНТАЦИИ ДЛЯ УЛУЧШЕНИЯ КАЧЕСТВА КЛАССИФИКАЦИИ / А. Д. Юрьевич // StudNet. — 2022. — № 4. — С. 2397–2411.
- 3 Ricanek, K. MORPH: a longitudinal image database of normal adult age-progression / K. Ricanek, T. Tesafaye // 7th International Conference on Automatic Face and Gesture Recognition (FGR06). — 2006. — Pp. 341–345.
- 4 Saleem, M. Application of deep convolution neural network for image / M. Saleem // International Journal of Advanced Trends in Computer Science and Engineering. — 2023. — Vol. 11. — Pp. 42–46.
- 5 Zhifei, Z. Age progression/regression by conditional adversarial autoencoder / Z. Zhifei, S. Yang, H. Qi // IEEE Conference on Computer Vision and Pattern Recognition (CVPR). — 2017. — Pp. 5810–5818.