

МИНОБРНАУКИ РОССИИ
Федеральное государственное бюджетное образовательное учреждение
высшего образования
«САРАТОВСКИЙ НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ
ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ
ИМЕНИ Н. Г. ЧЕРНЫШЕВСКОГО»

Кафедра системного анализа и
автоматического управления

**РЕАЛИЗАЦИЯ АЛГОРИТМА РАСПОЗНАВАНИЯ ВВОДА НА
ОСНОВЕ ИСКУССТВЕННОЙ НЕЙРОННОЙ СЕТИ**
АВТОРЕФЕРАТ БАКАЛАВРСКОЙ РАБОТЫ

студента 4 курса 481 группы
направления 27.03.03 — Системный анализ и управление
факультета КНиИТ
Чугунова Дениса Алексеевича

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

М. В. Корнилов

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

И. Е. Тананко

Саратов 2025

ВВЕДЕНИЕ

Актуальность темы обусловлена стремительной цифровизацией всех сфер человеческой деятельности проблема автоматического распознавания текста (OCR) приобретает особую значимость. Современный мир генерирует колоссальные объемы текстовой информации в самых разных форматах - от печатных документов и рукописных заметок до изображений с текстовыми элементами. При этом значительная часть этой информации по-прежнему существует в аналоговом виде, что создает серьезные препятствия для ее автоматизированной обработки и анализа. Именно в этом контексте разработка эффективных алгоритмов оптического распознавания символов становится критически важной задачей, имеющей как научное, так и прикладное значение.

Особую актуальность данная тема приобретает в свете развития технологий искусственного интеллекта и машинного обучения. Современные нейронные сети, особенно сверточные архитектуры, демонстрируют впечатляющие результаты в задачах компьютерного зрения, включая распознавание слов. Однако, несмотря на существование коммерческих OCR-решений, потребность в разработке специализированных алгоритмов остается чрезвычайно высокой. Это связано с рядом фундаментальных ограничений существующих систем.

Универсальные OCR-системы часто демонстрируют недостаточную точность при работе с нестандартными шрифтами, низкокачественными изображениями или специфическими типами документов. Многие коммерческие решения требуют подключения к облачным сервисам, что создает проблемы с конфиденциальностью данных и делает невозможной работу в условиях отсутствия интернет-соединения. В-третьих, существующие системы часто оказываются избыточными и ресурсоемкими для решения узкоспециализированных задач.

В данном исследовании акцент делается на разработке алгоритма распознавания печатных символов (цифр и заглавных латинских букв) с использованием сверточных нейронных сетей. Этот выбор обусловлен несколькими факторами. С одной стороны, распознавание печатных символов является базовой задачей, на основе которой можно строить более сложные системы. С другой стороны, даже в этой, казалось бы, простой области остаются нере-

шенные проблемы, связанные с обеспечением устойчивости алгоритма к различного рода искажениям, шумам и вариациям в изображениях.

Таким образом, исследование и разработка алгоритмов OCR на основе нейронных сетей представляет собой актуальную научно-практическую задачу, решение которой может внести существенный вклад как в развитие методов компьютерного зрения, так и в создание конкретных программных решений для автоматизации обработки текстовой информации. Особое значение эта работа приобретает в условиях, когда потребность в простых, эффективных и доступных инструментах распознавания слов продолжает неуклонно расти.

Целью данной бакалаврской работы является разработка алгоритма распознавания слов, основанного на сверточных нейронных сетях, который сочетает в себе высокую точность, устойчивость к искажениям и практическую применимость в реальных сценариях. Для достижения этой цели предстоит решить комплекс взаимосвязанных задач, включая анализ современных методов OCR, подготовку и augmentation данных, проектирование архитектуры нейронной сети, а также реализацию полного пайплайна обработки изображений от предварительной бинаризации до финального распознавания символов. Особое внимание уделяется оптимизации модели для работы с печатными символами, включая цифры и заглавные латинские буквы, что является базовым, но критически важным этапом для любых систем автоматического ввода данных.

Теоретическая значимость исследования заключается в углубленном изучении возможностей сверточных нейронных сетей для задач классификации символов. В ходе работы анализируются различные подходы к построению архитектуры сети, такие как применение слоев Batch Normalization и Dropout для борьбы с переобучением, а также исследуется влияние методов аугментации данных на обобщающую способность модели. Полученные результаты могут служить основой для дальнейших исследований, например, в области распознавания рукописного текста или адаптации алгоритма для работы с кириллицей и сложными шрифтами.

Практическая ценность работы проявляется в широком спектре потенциальных применений разработанного алгоритма. В сфере документооборота он может быть использован для автоматизации ввода данных из счетов,

накладных или анкет, значительно сокращая временные затраты и минимизируя человеческие ошибки. В образовательных проектах на его основе возможно создание систем проверки заданий или инструментов для оцифровки учебных материалов. Отдельного внимания заслуживает возможность интеграции алгоритма в мобильные приложения или embedded-системы, например, для распознавания номеров товаров в логистике или обработки почтовых индексов. Важным преимуществом предложенного решения является его автономность – возможность работы без подключения к облачным сервисам, что особенно актуально для задач, связанных с обработкой конфиденциальной информации.

Бакалаврская работа состоит из введения, 4 разделов, заключения, списка использованных источников и цифрового носителя в качестве приложения. Общий объем работы — 48 страниц, из них 42 страницы — основное содержание, включая 6 рисунков и 2 таблицу, список использованных источников информации — 23 наименования.

КРАТКОЕ СОДЕРЖАНИЕ РАБОТЫ

Первый раздел «Искусственные нейронные сети» посвящен теоретическим основам искусственных нейронных сетей и их применению в задачах оптического распознавания символов.

В подразделе 1.1 приведены основные понятия и архитектура искусственных нейронных сетей, включая математическую модель отдельного нейрона:

$$y = f \left(\sum_{i=1}^n w_i x_i + b \right)$$

где x_i - входные сигналы, w_i - весовые коэффициенты, b - смещение, f - функция активации [1].

Подраздел 1.2 описывает принципы работы сверточных нейронных сетей (CNN), включая:

- Сверточные слои с операцией двумерной свертки
- Слои подвыборки (Max Pooling и Average Pooling)
- Полносвязные слои для классификации [2]

Подраздел 1.3 содержит обзор функций активации (ReLU, Leaky ReLU, ELU) и функции Softmax для задач классификации [3].

Подраздел 1.4 посвящен методам регуляризации в нейронных сетях:

- Dropout для предотвращения коадаптации нейронов
- Batch Normalization для ускорения обучения
- L2-регуляризация для контроля весовых коэффициентов [4]

Второй раздел «Основы оптического распознавания символов (OCR)» описывает ключевые этапы обработки текстовых изображений.

В подразделе 2.1 представлена общая схема работы OCR-систем, включающая:

- Предварительную обработку изображений
- Сегментацию текстовых блоков
- Распознавание и классификацию символов [5]

Подраздел 2.2 посвящен методам предварительной обработки:

- Бинаризация (метод Оцу и адаптивные методы)
- Морфологические операции (эрозия, дилатация)
- Анализ контуров [6]

Подраздел 2.3 описывает методы сегментации текста:

- Выделение текстовых блоков
- Сегментация строк и символов
- Особенности обработки печатного и рукописного текста [7]

Третий раздел «Используемые технологии и библиотеки» содержит описание инструментария для реализации системы OCR.

Подраздел 3.1 описывает применение языка Python для машинного обучения [8].

Подраздел 3.2 посвящен библиотекам:

- TensorFlow/Keras для реализации нейронных сетей
- OpenCV для обработки изображений
- Pandas/NumPy для управления данными [9]

Четвертый раздел «Практическая реализация системы OCR» описывает процесс создания и тестирования моделей.

В подразделе 4.1 представлены используемые датасеты:

- EMNIST (Extended MNIST) для печатных символов
- Stanford для рукописных цифр [10]

Подраздел 4.2 описывает архитектурные особенности моделей CNN для каждого датасета.

Подраздел 4.3 содержит процедуру обучения моделей с применением:

- Аугментации данных
- Оптимизатора Adam
- Методов регуляризации [11]

Подраздел 4.4 представляет результаты сравнения моделей. Результаты представлены в таблице 1.

Таблица 1 – Сравнение характеристик моделей

Параметр	EMNIST	Stanford
Размер изображения	28x28	32x32
Количество классов	47	36
Точность	0.84	0.88

Результаты работы системы демонстрируют высокую точность распознавания (до 88%) и возможность практического применения в задачах автоматизации документооборота [12].

ЗАКЛЮЧЕНИЕ

В данной выпускной квалификационной работе проведено исследование, посвященное разработке и сравнительному анализу двух моделей оптического распознавания символов на основе сверточных нейронных сетей. Основной целью работы являлось создание эффективной системы распознавания текста с использованием современных методов глубокого обучения и компьютерного зрения. Анализ современных подходов подтвердил, что сверточные нейронные сети обеспечивают высокую точность даже при наличии шумов, искажений и вариаций шрифтов. В ходе исследования были разработаны две модели на основе датасетов EMNIST и Stanford, достигшие точности распознавания порядка 90

Качество предварительной обработки изображений оказало значительное влияние на итоговую точность моделей. Для датасета EMNIST оптимальным методом бинаризации стал метод Оцу, в то время как для Stanford более эффективными оказались адаптивные методы. Применение морфологических операций позволило улучшить качество изображений за счет устранения шумов и восстановления структуры символов. Сравнительный анализ показал, что EMNIST-модель поддерживает большее количество классов, что

делает её универсальной, тогда как Stanford-модель демонстрирует повышенную устойчивость к вариациям качества входных данных благодаря сложным методам регуляризации и аугментации. Обе модели обладают сопоставимой скоростью обработки (около 4 мс на изображение), что позволяет использовать их в системах реального времени.

Практическим результатом работы стало создание полного конвейера обработки изображений, включающего предобработку, распознавание и визуализацию. Система учитывает возможные исключительные ситуации, такие как низкое качество входных данных или отсутствие текста, и включает механизмы обработки ошибок и логирования, что делает её пригодной для промышленного внедрения. Разработанное решение может применяться в автоматизации документооборота, банковских технологиях и других прикладных областях.

Перспективы дальнейшего развития включают исследование современных архитектур нейронных сетей (ResNet, Transformer), улучшение методов аугментации данных, повышение точности распознавания рукописного текста, а также разработку эффективных механизмов постобработки с использованием языковых моделей. Особый потенциал имеет создание мультимодальных систем, объединяющих визуальное распознавание с семантическим анализом, что особенно актуально для работы со специализированными текстами.

Таким образом, проведенное исследование подтвердило эффективность глубокого обучения в задачах оптического распознавания символов и позволило разработать практическую систему, готовую к реальному применению. Полученные результаты открывают возможности для дальнейшего совершенствования алгоритмов и расширения функциональности, что делает их перспективными для использования в различных сферах экономики и социальной деятельности.

Основные источники информации:

1. Гудфеллоу, И. Глубокое обучение / И. Гудфеллоу, И. Бенджио, А. Курвилль ; пер. с англ. А. А. Слинкина. — М. : ДМК Пресс, 2022. — 652 с.
2. Russakovsky, O. ImageNet Large Scale Visual Recognition Challenge / O. Russakovsky, J. Deng, H. Su // International Journal of Computer Vision.

- 2015. — Vol. 115, No. 3. — P. 211–252.
3. ЛеКун, Я. Градиентное обучение применяется к распознаванию документов / Я. ЛеКун, Л. Ботту, Й. Хаффнер // Proceedings of the IEEE. — 1998. — Т. 86, № 11. — С. 2278–2324.
 4. Simonyan, K. Very Deep Convolutional Networks for Large-Scale Image Recognition / K. Simonyan, A. Zisserman // International Conference on Learning Representations. — 2015. — 14 p.
 5. He, K. Deep Residual Learning for Image Recognition / K. He, X. Zhang, S. Ren // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. — 2016. — P. 770–778.
 6. Оцу, Н. Пороговый выбор методом градации серого / Н. Оцу // IEEE Transactions on Systems, Man, and Cybernetics. — 1979. — Т. 9, № 1. — С. 62–66.
 7. Саувола, Й. Адаптивный метод бинаризации документов / Й. Саувола, М. Пиетаксинен // Pattern Recognition. — 2000. — Т. 33, № 2. — С. 225–236.
 8. Брадски, Г. Изучение OpenCV 4 с помощью Python / Г. Брадски, А. Кэлер ; пер. с англ. А. А. Слинкина. — М. : ДМК Пресс, 2020. — 524 с.
 9. Шолле, Ф. Глубокое обучение на Python / Ф. Шолле ; пер. с англ. А. А. Слинкина. — СПб. : Питер, 2021. — 400 с.
 10. Cohen, G. EMNIST: An Extension of MNIST to Handwritten Letters / G. Cohen, S. Afshar, J. Tapson // Proceedings of the International Joint Conference on Neural Networks. — 2017. — P. 2921–2926.
 11. Смит, Р. Современная система оптического распознавания символов / Р. Смит // IEEE Transactions on Pattern Analysis and Machine Intelligence. — 2007. — Т. 29, № 4. — С. 530–553.
 12. Abadi, M. TensorFlow: A System for Large-Scale Machine Learning / M. Abadi, P. Barham, J. Chen // Proceedings of the 12th USENIX Symposium on Operating Systems Design and Implementation. — 2016. — P. 265–283.