

МИНОБРНАУКИ РОССИИ

Федеральное государственное бюджетное образовательное учреждение  
высшего образования

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

Кафедра теории функций и стохастического анализа

**ПРИМЕНЕНИЕ МЕТОДОВ МАШИННОГО ОБУЧЕНИЯ К  
ЗАДАЧАМ РАСПОЗНАВАНИЯ ОБРАЗОВ**

**АВТОРЕФЕРАТ БАКАЛАВРСКОЙ РАБОТЫ**

студентки 4 курса 451 группы  
направления 38.03.05 — Бизнес-информатика

механико-математического факультета

Шевченко Алины Игоревны

Научный руководитель

д. ф.-м. н., доцент

\_\_\_\_\_

С. П. Сидоров

Заведующий кафедрой

д. ф.-м. н., доцент

\_\_\_\_\_

С. П. Сидоров

Саратов 2024

## ВВЕДЕНИЕ

**Актуальность темы.** Распознавание образов, являясь ключевой составляющей искусственного интеллекта, играет центральную роль в исследованиях, направленных на автоматизацию и оптимизацию процессов обработки и анализа данных. Особое внимание уделяется использованию искусственных нейронных сетей и сверточных нейронных сетей для решения задач распознавания образов, что подчеркивает актуальность и перспективность развития технологий глубокого обучения (deep learning).

Актуальность работы также обусловлена широким спектром приложений методов распознавания образов во многих сферах науки, техники, медицины, биологии и т.д.

**Целью бакалаврской работы** является решение одной из задач компьютерного зрения, состоящей в распознавании рукописных математических выражений при помощи аппарата искусственных нейронных сетей с дальнейшей конвертацией полученной формулы в LaTeX-формулу на основе использования аппарата искусственных нейронных сетей.

**Объектом исследования** является задача распознавания рукописных математических выражений. Исследование фокусируется на разработке и применении методов и алгоритмов, которые могут анализировать изображения или сканы рукописных математических выражений и определять их содержание.

**Предметом исследования** является процесс использования искусственных нейронных сетей для решения этой задачи. Искусственные нейронные сети — это модели машинного обучения, которые имитируют работу биологических нейронных сетей и способны обучаться на больших объемах данных, автоматически определяя сложные зависимости и шаблоны. В данном случае, нейронные сети используются для распознавания и интерпретации рукописных математических выражений.

Для достижения цели были поставлены следующие задачи:

- изучить основные определения, применяемые в искусственных нейронных сетях;
- исследовать современные методы решения задач компьютерного зре-

ния, а также недостатки этих методов;

- познакомиться с библиотекой машинного обучения pytorch;
- рассмотреть архитектуры искусственных нейронных сетей;
- разработать алгоритм для распознавания рукописных математических выражений с применением аппарата искусственных нейронных сетей;
- научиться применять многослойные перцептроны и предобученные сверточные нейронные сети для решения задач.

**Практическая значимость** данной темы заключается в её применении в различных областях науки, техники, медицины и биологии, где методы машинного обучения и искусственного интеллекта находят всё более широкое применение. Особенно актуальна тема распознавания образов, поскольку она лежит в основе многих современных технологий, таких как компьютерное зрение, распознавание речи, обработка естественного языка и многие другие.

Применение искусственных нейронных сетей и сверточных нейронных сетей для решения задач распознавания образов открывает новые возможности для автоматизации и оптимизации процессов обработки и анализа данных. Это позволяет значительно ускорить и упростить многие процессы, связанные с обработкой больших объемов информации, что в свою очередь ведёт к повышению эффективности работы в различных отраслях.

**Структура и содержание работы.** Работа состоит из введения, 3 разделов, заключения и списка используемых источников, содержащего 20 наименований.

## ОСНОВНОЕ СОДЕРЖАНИЕ РАБОТЫ

Во **введении** раскрывается актуальности темы работы, формулируется цель работы, задачи, которые необходимо решить, и отмечается практическая значимость результатов

В **первом** разделе приводится теоретическая часть работы. Рассматриваются основы машинного обучения, подчеркивая его роль в искусственном интеллекте как области, посвященной разработке методов решения задач через обучение на основе данных. Основная идея машинного обучения заключается в том, чтобы использовать обучающую выборку, состоящую из пар объект-ответ, для обучения модели, которая затем может генерировать приближенные ответы на новые объекты. Объекты могут быть представлены различными типами признаков, что делает машинное обучение гибким инструментом для решения широкого спектра задач.

Задачи машинного обучения классифицируются на несколько категорий, включая регрессию, классификацию и кластеризацию. Классификация, в частности, фокусируется на определении принадлежности объекта к одному из заранее определенных классов на основе его признаков. Математически это выражается через поиск функции, которая отображает объекты на метки классов, и может быть решена с помощью функции решающего правила, которая определяет класс объекта на основе его признаков. Примеры задач классификации включают медицинскую диагностику, геологоразведку, распознавание текстов, кредитный скоринг и синтез химических соединений. Эти примеры демонстрируют широкий спектр применения машинного обучения в различных областях, от здравоохранения до финансов и науки.

Далее подробно описываются различные подходы к решению задач классификации в машинном обучении. Эти методы включают:

- Логистическая регрессия – метод, который используется для предсказания вероятности принадлежности объекта к определенному классу.
- Метод ближайших соседей (k-NN) – метод, который основывается на том, что объект относится к тому же классу, что и ближайшие к нему объекты.
- Деревья решений – метод, который используется для построения дерева, которое разделяет объекты на классы на основе признаков.

- Случайный лес – метод, который основан на построении множества деревьев решений и предсказывает класс объекта на основе голосования всех деревьев.
- Наивный Байесовский классификатор – метод, который использует статистические методы для предсказания класса объекта на основе вероятностей появления каждого признака.
- Метод опорных векторов (SVM) – метод, который использует математические методы для построения гиперплоскости, которая разделяет объекты на классы.
- Нейронные сети – метод, который использует алгоритмы машинного обучения для построения модели, которая способна предсказывать класс объекта на основе множества признаков.

Выбор метода зависит от характеристик данных и задачи, а иногда комбинация методов используется для улучшения результатов.

Во **втором** разделе обсуждаются две ключевые проблемы в машинном обучении: выбор признаков и влияние факторов вариации на исходные данные. Выбор признаков критически важен для успешности алгоритмов машинного обучения, особенно в задачах, где нет очевидного набора признаков, например, в распознавании изображений или речи. Исследователи обычно выбирают признаки через аналитическую работу, что может быть трудоемким и не всегда эффективным, поскольку может пропустить важные признаки, не очевидные для человека.

Автоматическое определение признаков, которое позволяет алгоритмам машинного обучения работать более эффективно, становится возможным благодаря использованию deep learning. Deep learning использует многослойные нейронные сети, которые самообучаются на больших наборах данных, формируя иерархическую структуру признаков. Это позволяет алгоритмам автоматически выявлять и использовать признаки различных уровней абстракции, что особенно полезно для обработки сложных данных, таких как изображения или звуки.

Deep learning достигло популярности благодаря нескольким причинам: появлению новых архитектур нейронных сетей, развитию вычислительных технологий и доступности графических процессоров, а также распростране-

нию подхода послойного обучения, который позволяет обучать сети более эффективно и использовать неразмеченные данные для обучения. Этот подход значительно упрощает процесс обучения и делает его более доступным для исследователей, позволяя использовать одну обученную сеть для решения различных задач классификации с помощью разных наборов данных.

Далее в разделе описываются нейронные сети как математическая модель, имитирующая работу нервной системы, состоящая из связанных между собой элементов - нейронов. Ключевые компоненты нейронной сети включают:

- Нейрон - основной элемент, принимающий входные данные и выдающий результат, состоящий из входных и выходных связей и функции активации.
- Связи - передают информацию от одного нейрона к другому.
- Функция активации - определяет активационный потенциал нейрона.
- Слой - группа нейронов, выполняющих одну функцию.
- Обучение - процесс настройки нейронной сети на задачу.

Искусственный нейрон состоит из векторов входных значений, весов, смещения и функции активации, которые вместе определяют выход нейрона. Нейроны организованы в слои, включая входной, скрытый и выходной, каждый из которых может иметь произвольное количество нейронов и входных/выходных сигналов.

Преимущества нейронных сетей включают самообучаемость, эффективную фильтрацию шумов, адаптацию к изменениям во входных данных, отказоустойчивость и высокую скорость работы.

Далее описываются ключевые аспекты работы с нейронными сетями, сосредоточиваясь на определении функции потерь, архитектуре нейронных сетей, включая многослойные перцептроны (MLP) и сверточные нейронные сети (CNN), а также на методе transfer learning.

Многослойный перцептрон (MLP) представляет собой нейронную сеть прямого распространения сигнала, где входной сигнал преобразуется в выходной, проходя через несколько слоев. MLP обучается с помощью метода обратного распространения ошибки, который корректирует веса нейронов для минимизации функции ошибок.

Сверточная нейронная сеть (CNN) используется для анализа визуальных образов и состоит из чередующихся слоев свертки и подвыборки, за которыми следуют полносвязные слои. CNN применяет операцию свертки для обработки изображений, что позволяет значительно сократить количество параметров и ускорить обучение.

Transfer learning предполагает использование предобученных нейронных сетей для решения новых задач. Например, сверточные нейронные сети могут быть обучены на большом наборе данных, таком как ImageNet, и затем адаптированы для решения других задач, изменяя последний полносвязный слой и дообучая модель на новом наборе данных. Эти концепции являются фундаментальными для понимания и применения нейронных сетей в различных областях, включая распознавание образов, обработку естественного языка и многое другое.

В **третьем** разделе показано, как на практике решать задачу распознавания рукописных математических выражений при помощи аппарата искусственных нейронных сетей. Для обучения модели нейронной сети распознавать числа, математические знаки и буквы, необходимо подготовить обучающий набор данных, который включает в себя тренировочные и проверочные данные. Тренировочные данные используются для корректировки параметров модели, а проверочные данные - для оценки её точности.

Выбранный датасет включает в себя цифры от 0 до 9, буквы латинского алфавита («h», «t», «w», «x», «y») и наиболее часто встречающиеся математические символы, такие как знаки минуса, плюса, скобок, запятой и звездочки.

Из обучающих наборов данных, содержащих изображения этих объектов, была произведена выборка. Из-за неравномерного количества изображений для каждого класса (от 1000 до 30000), была выполнена равномерная выборка тестовой коллекции, что привело к созданию множества из 20205 файлов-изображений. Большая часть этих изображений (1000) была добавлена в тренировочный набор данных, а 200 изображений - в валидационный.

Этот процесс подготовки данных обеспечивает базу для обучения модели распознаванию указанных символов, что является ключевым шагом в разработке нейронных сетей для обработки и интерпретации изображений.

В процессе обучения моделей нейронных сетей используются ключевые параметры, такие как эпоха, батч и итерация.

- Эпоха - это процесс прохождения всего тренировочного набора данных через нейронную сеть, завершается, когда все тренировочные экземпляры были обработаны.
- Батч - это небольшой набор тренировочных данных, используемый для корректировки параметров модели.
- Итерация - количество батчей, которые должны пройти через модель, чтобы закончить одну эпоху.

Обучение модели можно остановить, когда достигнута требуемая точность или если ошибка перестала уменьшаться. Модель считается хорошо обученной, если она точно определяет объекты как на обучающих, так и на новых данных. Переобучение, когда модель хорошо работает только на обучающих данных, указывает на необходимость корректировки архитектуры или количества обучающих данных. Для решения проблемы переобучения можно использовать раннюю остановку, сохраняя весовые коэффициенты модели до начала переобучения.

В качестве первого метода распознавания элементов математического выражения был выбран многослойный персептрон (MLP) с архитектурой, включающей входной слой, два скрытых слоя и выходной слой, соответствующий количеству классов. Обучение сети проходило с использованием алгоритма обратного распространения ошибки и функции потерь кросс-энтропии.

Обучение модели осуществлялось в Jupyter Notebook, где данные делились на батчи для эффективного обучения. Валидационный набор данных использовался для оценки обобщающей способности модели, не влияя на корректировку параметров.

Оценка точности модели производилась с помощью метрики Accuracy, которая рассчитывается как отношение количества правильных прогнозов к общему количеству. Однако, несмотря на достаточно высокую эффективность, при дальнейшем анализе тестового изображения модель проявила недостаточную точность в детектировании границ объектов, а также в различении цифр и букв, у которых могут быть схожие очертания. Для улучшения точности модели в будущем планируется дополнительное обучение на изобра-



жениях с индивидуальными особенностями почерка. Можно сделать вывод, что MLP может столкнуться с проблемами при работе с изображениями из-за своей линейной природы, что делает его менее эффективным для задач, требующих учета пространственной структуры данных, такой как распознавание образов.

В качестве второго метода распознавания элементов математического выражения была выбрана предобученная нейронная сеть, в частности, архитектура EfficientNet, которая показывает высокую эффективность в классификации изображений на наборе данных ImageNet. EfficientNet использует сверточные слои, а также предлагает различные размеры сетей, что обеспечивает гибкость в выборе подходящего количества весов. Обучение сети проходило с использованием алгоритма обратного распространения ошибки и функции потерь кросс-энтропии, с количеством эпох обучения 20. Данные разделялись на батчи для эффективного обучения.

Далее при помощи библиотеки компьютерного зрения и машинного обучения OpenCV изображение в градациях серого было преобразовано в бинарное изображение. Это сделано для упрощения последующих этапов обработки, так как бинарное изображение содержит только два цвета (черный и белый), что упрощает выделение контуров и объектов на изображении.

Затем выделялись контуры объектов, после этого изображения объектов центрировались, и далее подавались на классификацию с помощью обученной модели.

Результаты классификации интерпретировались, и для каждого изображения определялась метка класса, соответствующая наиболее вероятному символу или цифре. Полученные метки сопоставлялись с названиями папок, содержащих изображения математических элементов, и результат записывался в строку.

Также была выполнена задача конвертации полученной формулы в LaTeX-формулу, что позволяет представлять результаты в удобном для чтения виде.

**В заключении** приведены результаты бакалаврской работы.

## ОСНОВНЫЕ РЕЗУЛЬТАТЫ

1. Изучены основные определения, применяемые в искусственных нейронных сетях.
2. Исследованы современные методы решения задач компьютерного зрения, а также недостатки этих методов.
3. Разработан алгоритм для распознавания рукописных математических выражений с применением аппарата искусственных нейронных сетей.
4. Рассмотрены архитектуры полносвязных и сверточных нейронных сетей, а также рассмотрены недостатки и преимущества этих архитектур.
5. Применены многослойные перцептроны и предобученные нейронные сети для решения задачи распознавания рукописных математических выражений.