

МИНОБРНАУКИ РОССИИ

Федеральное государственное бюджетное образовательное учреждение

высшего образования

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

Кафедра теории функций и стохастического анализа

**СВЕРТОЧНЫЕ НЕЙРОННЫЕ СЕТИ И ИХ ПРИЛОЖЕНИЯ К
ЗАДАЧАМ РАСПОЗНАВАНИЯ ОБРАЗОВ**

АВТОРЕФЕРАТ МАГИСТЕРСКОЙ РАБОТЫ

студента 2 курса 248 группы

направления 09.04.03 — Прикладная информатика

механико-математического факультета

Власенко Дмитрия Ильича

Научный руководитель

доцент, к. ф.-м. н.

Д. В. Мельничук

Заведующий кафедрой

д. ф.-м. н., доцент

С. П. Сидоров

Саратов 2024

Введение

В последние годы сверточные нейронные сети (CNN) получили широкое распространение в задачах компьютерного зрения благодаря способности автоматически извлекать признаки из изображений. Однако на практике эффективность таких моделей снижается при наличии искажений входных данных, таких как шум, размытие или сжатие. Это делает проблему устойчивости нейросетей актуальной в контексте реального применения.

Данная проблема особенно актуальна для прикладных задач, в которых надёжность систем компьютерного зрения критична: агротехнические комплексы, системы мониторинга окружающей среды, автономные транспортные средства, медицинская диагностика и др. В таких условиях данные, поступающие на вход модели, часто оказываются неидеальными — загрязнёнными, частично закрытыми или сжатыми, что делает анализ устойчивости нейросетей не только научно интересной, но и практически значимой задачей.

Цель магистерской работы — исследовать устойчивость предобученной модели ResNet50 к различным искажениям изображений при решении задачи классификации растений (датасет Plant Seedlings).

Задачи работы включали:

1. изучение архитектур CNN и программных средств для их реализации;
2. реализацию функций искажений различной природы и интенсивности;
3. проведение экспериментов по оценке точности модели на искажённых данных;
4. построение графиков робастности и тепловых карт;
5. анализ результатов и выявление факторов, влияющих на устойчивость.

В работе исследуется архитектура ResNet50, используется подход трансферного обучения и проводится эмпирическая оценка устойчивости модели при воздействии одиночных и парных искажений.

В качестве основы для оценки выбрана модель ResNet50, обладающая хорошим соотношением точности и сложности, широким распространением в реальных системах и доступными предобученными весами. Используется подход трансферного обучения, при котором базовые слои модели замораживаются, а выходная часть адаптируется под конкретную задачу.

В качестве набора данных применяется датасет Plant Seedlings, содер-

жащий более 4 500 цветных изображений 12 видов молодых растений в условиях открытого грунта. Изображения имеют высокое разрешение и различаются по углам съёмки, освещению и структуре фона, что делает датасет репрезентативным и подходящим для оценки робастности.

Практическая часть выполнена с использованием языка Python и библиотек TensorFlow, Keras, OpenCV и NumPy. Реализован полный цикл: от предобработки изображений и обучения модели до генерации искажений и визуализации результатов в виде графиков и тепловых карт. Это позволило комплексно оценить поведение модели в условиях деградации входных данных и предложить направления для повышения устойчивости.

Структура работы:

- Глава 1 — архитектура и характеристики свёрточных нейросетей;
- Глава 2 — устойчивость CNN к искажениям входных данных;
- Глава 3 — обзор современных архитектур CNN;
- Глава 4 — программные средства для реализации нейросетей;
- Глава 5 — практическая часть: оценка устойчивости ResNet50 на датасете Plant Seedlings.

Цель магистерской работы — исследовать устойчивость предобученной модели ResNet50 к различным искажениям изображений при решении задачи классификации растений (датасет Plant Seedlings).

Работа прошла апробацию на различных конференциях, в частности, на ежегодной студенческой конференции "Актуальные проблемы математики и механики" которую проводил механико-математический факультет СГУ в апреле 2025 года, в секции "Анализ данных" в XIII Международной молодежной научно-практической конференции «Математическое и компьютерное моделирование в экономике, страховании и управлении рисками», ноябрь 2024 года.

Основное содержание работы

В первой главе рассматриваются теоретические основы функционирования нейронных сетей, включая понятие искусственного нейрона, многослойной архитектуры и активационных функций. Описываются основные этапы обработки информации в нейросетях: от подачи входных данных до получения выходных предсказаний. Подробно рассматриваются принципы работы нейрона, включающие взвешенное суммирование входных сигналов, применение функции активации и передача результата далее по сети. Основное внимание уделяется топологии свёрточных нейронных сетей (CNN), включая описание входного, свёрточного, подвыборочного (пулингового), полносвязного и выходного слоёв. Для каждого слоя поясняется его назначение, математическая модель и влияние на размерность данных.

Также рассмотрены функции активации: сигмоида, $\tanh$ и наиболее популярная ReLU, её модификации (Leaky ReLU, ELU) и влияние на обучение сети. Подчёркивается важность выбора активации для избежания проблемы исчезающего градиента.

Рассматриваются методы нормализации данных, включая batch normalization, и их влияние на стабильность обучения. Также отдельно обсуждаются архитектурные различия между полносвязными и свёрточными слоями: свёртки обеспечивают локальность восприятия и меньшую параметризацию, в то время как fully connected слои чаще используются на финальных этапах для принятия решений.

В главе приводится краткий исторический обзор развития CNN — от LeNet до современных моделей, а также обсуждаются ключевые идеи, повлиявшие на широкое распространение свёрточных сетей в задачах обработки изображений, включая backpropagation, большие датасеты и рост вычислительных мощностей.

Во второй главе подробно рассматриваются причины потери точности CNN при искажении входных данных. Начинается с анализа природы искажений, которые могут возникать в реальных условиях, включая шумы, размытость, фотометрические искажения (изменение яркости, контрастности), геометрические трансформации (вращения, масштабирование, сдвиги), сжатие изображений и частичное перекрытие объектов (окклюзия). Дана класси-

фикация искажений по характеру воздействия: случайные (например, шум), систематические (например, JPEG-артефакты) и целенаправленные (например, adversarial атаки).

Каждый тип искажения сопровождается кратким описанием механизма его возникновения и последствий для активаций в свёрточных слоях. Например, гауссов шум приводит к разрушению локальных текстур, особенно на начальных слоях, а сильное размытие влияет на способность сети распознавать границы объектов. JPEG-сжатие нарушает регулярность текстур и приводит к искажению пространственных зависимостей. Геометрические трансформации могут привести к потере инвариантности модели к поворотам и масштабам.

Показано, что свёрточные сети чаще опираются на локальные текстурные признаки, нежели на глобальные формы и семантику. Это делает модели уязвимыми к искажениям, которые не изменяют форму объекта, но разрушают текстуру. Подчёркивается важность использования методов объяснимого ИИ (ХАИ), таких как визуализация карт активаций, чтобы выявить, какие именно признаки используются моделью при классификации.

Рассмотрены существующие методы повышения устойчивости: — data augmentation (искусственное увеличение обучающей выборки путём добавления искажений); — регуляризация (dropout, weight decay); — архитектурные модификации (использование residual connections, attention-механизмов); — использование adversarial training (обучение на специально созданных искажениях).

Кроме того, освещены подходы на стадии предобработки (фильтрация, нормализация), а также методы post-hoc повышения устойчивости, включая ансамблирование моделей и адаптивную настройку порогов. Приведён краткий обзор литературы по robust deep learning, включая основные направления: adversarial robustness, corruptions robustness, out-of-distribution detection. Подчёркиваются открытые проблемы, такие как невозможность обеспечить устойчивость ко всем типам искажений одновременно и нехватка универсальных метрик для оценки робастности моделей.

Третья глава содержит систематический обзор архитектур CNN: от классической LeNet до современных моделей VGG, GoogleNet, ResNet и EfficientNet.

Каждая из этих архитектур сыграла важную роль в развитии глубокого обучения и нашла широкое применение в прикладных задачах компьютерного зрения. Рассматриваются ключевые идеи, заложенные в архитектуру LeNet — одной из первых сверточных сетей, использовавшей слои свёртки и субдискретизации. VGG характеризуется простой, но глубокой архитектурой, в которой используются свёртки с ядрами 3×3 и одинаковая структура блоков. GoogleNet (Inception) предложила концепцию инцепшн-блоков с параллельными путями обработки, позволяющими извлекать признаки разных масштабов.

Отдельное внимание уделяется архитектуре ResNet, в частности, ResNet50. Принципиальное новшество — остаточные соединения (skip-connections), которые позволяют передавать градиенты на большие глубины, избегая эффекта затухающего градиента. Благодаря этому, ResNet показала превосходные результаты на множестве задач и стала одной из самых используемых моделей в промышленности.

Обоснован выбор ResNet50 как компромисса между глубиной, стабильностью и скоростью обучения. Модель обеспечивает высокую точность при сравнительно умеренном количестве параметров, легко поддаётся модификации и эффективно обучается при помощи transfer learning. В работе также рассматривается влияние глубины, ширины и структуры соединений на производительность и устойчивость модели.

Подчёркивается значение архитектурных улучшений, влияющих на устойчивость модели, таких как batch normalization, позволяющая ускорить сходимость и стабилизировать обучение, dropout, снижающий риск переобучения, и global average pooling, уменьшающий количество параметров в финальных слоях.

Особое внимание уделено модели EfficientNet, предлагающей систематический подход к масштабированию глубины, ширины и разрешения входных данных. Описана функция активации SiLU (Swish), применяемая в EfficientNet, обеспечивающая более плавную нелинейность и улучшающая качество обобщения.

В завершение главы проводится сравнение архитектур по ряду метрик: количество параметров, точность на стандартных тестовых наборах, устой-

чивость к шумам и трансформациям, скорость инференса. Делается вывод о том, что при ограниченных вычислительных ресурсах и наличии искажённых данных ResNet50 представляет собой сбалансированный выбор, обеспечивающий надёжную производительность и устойчивость.

Четвёртая глава посвящена современным программным инструментам: TensorFlow, Keras и PyTorch. Подробно рассматриваются их архитектурные особенности, поддерживаемые модели программирования, а также ключевые компоненты, обеспечивающие гибкость и масштабируемость. TensorFlow, разработанный Google, предоставляет поддержку как для статических, так и для динамических графов, обладает мощным API и возможностью экспорта обученных моделей для продакшн-сред. Keras, интегрированный в TensorFlow, используется как высокоуровневая оболочка для быстрой и удобной сборки моделей. PyTorch, напротив, предлагает гибкий подход с динамической моделью вычислений, что делает его особенно удобным для исследований и быстрой отладки.

Рассматриваются возможности графов вычислений: в TensorFlow основной упор делается на построение статического графа с последующей компиляцией и оптимизацией, тогда как в PyTorch все операции выполняются в режиме реального времени, что упрощает дебаг и ускоряет экспериментирование. Описаны подходы к инициализации моделей, настройке оптимизаторов и управлению обучающими циклами. Освещается поддержка работы на GPU и TPU, включая механизмы распределённого обучения, автоматическое управление памятью, а также интеграция с такими библиотеками, как CUDA и cuDNN.

В разделе представлены основные методы оптимизации: стохастический градиентный спуск (SGD), Adam, RMSProp. Показано влияние гиперпараметров (скорости обучения, коэффициентов регуляризации) на процесс сходимости. Приведены типовые функции потерь (categorical cross-entropy, MSE) и метрики оценки качества (accuracy, precision, recall, F1-score).

Приводится краткое сравнение библиотек по удобству использования, скорости, доступности документации, сообществу пользователей и доступу к готовым предобученным моделям (Model Zoo, TensorFlow Hub, Hugging Face).

Рассматриваются современные платформы для деплоя моделей: TensorFlow

Serving, TorchServe, ONNX. Обсуждаются преимущества и недостатки универсальных форматов (ONNX) для переноса моделей между фреймворками. Также рассмотрены облачные решения: Google Cloud AI Platform, Amazon SageMaker и Microsoft Azure ML, обеспечивающие масштабируемость, мониторинг и безопасность.

Показана роль современных фреймворков в быстром прототипировании, тестировании гипотез, отладке и последующем масштабировании моделей машинного обучения. Подчёркивается значение доступности open source-инструментов для развития исследовательской среды и коммерческих решений.

Пятая глава представляет собой практическую реализацию. На основе ResNet50, предобученной на ImageNet, построена модель для классификации растений по изображениям. Слой классификации был заменён, добавлены слои Dropout и Dense. Обучение проводилось на 80% выборки, остальные 20% использовались для валидации. Для проверки устойчивости были реализованы пять видов искажений: гауссов шум, размытие, JPEG-сжатие, изменение яркости и окклюзия. Каждое искажение применялось с пятью уровнями тяжести. Искажения реализованы с помощью библиотеки OpenCV, а выбор параметров основан на принятых в литературе шкалах интенсивности. Для каждого искажения строилась зависимость точности от уровня тяжести, что позволяло наглядно сравнивать чувствительность модели к различным типам деградации.

Эксперименты показали:

- значительное снижение точности при шуме и JPEG-сжатии (до 10%);
- устойчивость к изменению яркости;
- умеренное снижение при размытии и окклюзии;
- усиление негативного эффекта при парных искажениях, выявлено с помощью тепловых карт.

Результаты были визуализированы в виде графиков точности и потерь, кривых робастности, тепловых карт для всех сочетаний искажений. Визуализация обеспечила качественную интерпретацию поведения модели и позволила выделить наиболее уязвимые комбинации искажений. Также был реализован код автоматической генерации искажений и построения матриц точно-

сти, пригодный для повторного использования и расширения. Анализ позволил оценить, насколько ResNet50 чувствительна к разным типам искажений, и подтвердить гипотезу о зависимости от текстурных признаков. Наиболее критичными оказались шумы и сжатие, в то время как яркость практически не повлияла на предсказания. Дополнительно были выделены особенности поведения модели при воздействии различных искажений: например, при наличии размытия модель демонстрировала склонность к переобобщению, при этом наблюдалось снижение уверенности выходного слоя. JPEG-сжатие, особенно на высоких уровнях, вносило артефакты, которые маскировали ключевые признаки объектов, затрудняя распознавание. Одной из сильных сторон ResNet50 является наличие остаточных связей, позволяющих сохранять стабильные градиенты при обучении, что благоприятно сказалось на адаптации модели к задаче классификации растений. Важным преимуществом архитектуры является также наличие хорошо масштабируемой структуры, позволяющей её адаптацию под задачи различной сложности. Среди плюсов ResNet50 можно отметить:

- устойчивость к переобучению при использовании transfer learning;
- возможность тонкой настройки отдельных слоёв под целевой датасет;
- хорошая интерпретируемость при использовании методов визуализации активаций и вниманий (например, Grad-CAM);
- высокая скорость сходимости по сравнению с более глубокими архитектурами.

Таким образом, проведённый анализ позволил не только количественно оценить устойчивость модели, но и выявить качественные аспекты её поведения при различных внешних воздействиях, что делает исследование практически полезным для построения надёжных систем компьютерного зрения в условиях ограниченного качества входных данных.

Заключение

В ходе выполнения магистерской работы была проведена комплексная оценка устойчивости сверточной нейронной сети ResNet50 к искажениям входных данных на примере задачи классификации изображений растений. Исследование включало как теоретический анализ современных архитектур CNN и причин их чувствительности к деградации изображений, так и реализацию практического эксперимента с использованием датасета Plant Seedlings.

Разработанная методика позволила выявить конкретные виды искажений, наиболее критично влияющие на точность модели. Установлено, что такие искажения, как аддитивный шум и JPEG-сжатие, вызывают наиболее существенное снижение точности. В то же время изменение яркости и частичная окклюзия оказывают умеренное влияние. Комбинированные искажения приводят к значительной деградации качества распознавания, что подчёркивает важность многомерного анализа устойчивости.

Практическая часть показала, что использование подхода transfer learning на базе ResNet50 даёт хорошие результаты при обучении на ограниченном по объёму датасете. Модель демонстрирует высокую точность на чистых данных и позволяет провести оценку устойчивости без необходимости полного переобучения. Визуализация результатов в виде кривых робастности и тепловых карт обеспечила наглядное представление о поведении модели в различных условиях и дала возможность выделить уязвимые зоны.

Результаты работы подтверждают актуальность и значимость оценки устойчивости CNN в прикладных задачах компьютерного зрения. Полученные данные могут использоваться при разработке и тестировании систем, функционирующих в реальных условиях с высокой вероятностью искажений изображений. Также работа может служить основой для дальнейших исследований в области повышения робастности моделей, включая аугментацию, архитектурные улучшения и методы объяснимого ИИ.

Таким образом, поставленные цели и задачи магистерской работы были достигнуты, а полученные результаты представляют практическую и исследовательскую ценность для области машинного обучения и компьютерного зрения.

СПИСОК ИСПОЛЬЗОВАННЫХ ИСТОЧНИКОВ

- 1 *Чжан, А.* Погружение в глубокое обучение / А. Чжан, З. Липтон, М. Ли, А. Смола. — 2-е изд. — М.: ООО «ДМК Пресс», 2021. — 428 с.
- 2 *Лекун, Я.* Как учится машина: революция в области нейронных сетей и глубокого обучения / Я. Лекун. — Москва: Альпина ПРО, 2021. — 320 с.
- 3 *Картер, Дж.* Свёрточные нейросети / Дж. Картер. — Москва: ЛитРес, 2024. — 200 с.
- 4 *Дайзенрот, М. П.* Математика в машинном обучении / М. П. Дайзенрот, А. Фейзал, Ч. С. Он. — Санкт-Петербург: Питер, 2023. — 480 с.
- 5 *Калин, О.* Архитектуры глубокого обучения: математический подход / О. Калин; пер. с англ. А. Киселева. — Москва: ДМК Пресс, 2024. — 700 с.
- 6 *Поляков, Д.* Глубокое обучение на Python 3 / Д. Поляков. — Санкт-Петербург: БХВ-Петербург, 2022. — 320 с.
- 7 *Воронцов, К. В.* Искусственный интеллект и машинное обучение / К. В. Воронцов. — Москва: Бином. Лаборатория знаний, 2022. — 480 с.
- 8 *Цуканова, Н. И.* Программирование глубоких нейронных сетей на Python / Н. И. Цуканова. — Москва: КУРС, 2023. — 368 с.
- 9 *Шолле, Ф.* Машинное обучение и TensorFlow: искусственный интеллект на языке Python / Ф. Шолле. — Санкт-Петербург: Питер, 2023. — 400 с.
- 10 *Стивенс, Э.* Глубокое обучение с PyTorch / Э. Стивенс, Л. Антига, Т. Виманн. — Москва: Вильямс, 2021. — 456 с.

Приложение. Список сокращений

- CNN** Convolutional Neural Network — свёрточная нейронная сеть
- ReLU** Rectified Linear Unit — функция активации с выпрямлением
- Grad-CAM** Gradient-weighted Class Activation Mapping — визуализация значимых областей изображения
- PCA** Principal Component Analysis — метод главных компонент
- PSNR** Peak Signal-to-Noise Ratio — пиковое отношение сигнал/шум
- SSIM** Structural Similarity Index Measure — индекс структурного сходства
- IoU** Intersection over Union — мера пересечения над объединением
- MAE** Mean Absolute Error — средняя абсолютная ошибка
- MSE** Mean Squared Error — среднеквадратичная ошибка
- GPU** Graphics Processing Unit — графический процессор
- CPU** Central Processing Unit — центральный процессор
- ResNet** Residual Network — остаточная нейронная сеть
- TL** Transfer Learning — перенос обучения
- XAI** Explainable Artificial Intelligence — объяснимый искусственный интеллект
- SGD** Stochastic Gradient Descent — стохастический градиентный спуск
- TPU** Tensor Processing Unit — тензорный процессор Google