

Ордена Трудового Красного Знамени  
федеральное государственное бюджетное  
образовательное учреждение высшего образования  
«Московский технический университет связи и информатики»

На правах рукописи

Зиядинов Вадим Валерьевич

## **Оптимизация помехоустойчивости и точности нейросетевого распознавания изображений**

Специальность 2.2.13

Радиотехника, в том числе системы и устройства телевидения

Диссертация на соискание ученой степени

кандидата технических наук

Научный руководитель  
доктор технических наук, доцент  
Терешонок Максим Валерьевич

Научный консультант  
доктор технических наук, доцент  
Гладышев Анатолий Иванович

Москва 2024

## ОГЛАВЛЕНИЕ

ВВЕДЕНИЕ.....	2
1. АНАЛИЗ МЕТОДОВ АУГМЕНТАЦИИ ОБУЧАЮЩИХ ДАННЫХ ДЛЯ РАСПОЗНАВАНИЯ ИЗОБРАЖЕНИЙ .....	13
1.1 Проблемы практического применения свёрточных нейронных сетей.....	13
1.2 Аугментация данных.....	19
1.3 Виды и способы аугментации данных .....	22
1.3.1 Матричная фильтрация.....	24
1.3.2 Геометрические преобразования.....	24
1.3.3 Изменения цветового пространства.....	25
1.3.4 Дропаут.....	26
1.3.5 Наложение изображений .....	27
1.3.6 Состязательное обучение .....	29
1.3.7 Перенос стиля.....	31
1.3.8 Генеративно-состязательные сети.....	32
1.4 Естественные состязательные примеры .....	34
1.5 Оценка помехоустойчивости нейронных сетей.....	37
1.6 Постановка и формализация научной задачи исследования .....	38
Выводы по разделу 1.....	42
2. ИССЛЕДОВАНИЕ ПОМЕХОУСТОЙЧИВОСТИ СВЁРТОЧНОЙ НЕЙРОННОЙ СЕТИ.....	43
2.1 Метод оценки помехоустойчивости СНС .....	43
2.2 План исследования.....	47

2.3 Модель формирования и искажения изображений с низкой плотностью точек .....	49
2.4 Структура нейронной сети.....	52
2.5 Оценка зависимости качества распознавания от величины неопределенности в тестовых наборах данных .....	53
2.6 Исследование помехоустойчивости сети, обученной на наборах данных с искажениями.....	54
Выводы по разделу 2.....	57
3. ИССЛЕДОВАНИЕ ВЛИЯНИЯ НЕОПРЕДЕЛЁННОСТИ В ОБУЧАЮЩИХ ДАННЫХ НА ПОМЕХОУСТОЙЧИВОСТЬ СВЁРТОЧНОЙ НЕЙРОННОЙ СЕТИ.....	59
3.1 Зависимость точности распознавания от неопределённости в тестовых и обучающих данных ( $U_{TR}$ и $U_{TS}$ ).....	59
3.2 Интегральная точность распознавания изображений при различных пороговых значениях требуемой минимальной точности распознавания..	61
3.3 Распознавание зашумленных естественных изображений и обучение свёрточных нейронных сетей на зашумленных естественных изображениях.....	66
3.4 Результаты анализа работы свёрточной сети при прочих видах искажений естественных изображений.....	70
Выводы по разделу 3.....	73
4. МЕТОД ОПТИМАЛЬНОЙ АУГМЕНТАЦИИ ОБУЧАЮЩИХ ДАННЫХ БЕЗ УВЕЛИЧЕНИЯ ИХ ОБЪЁМА .....	75
4.1 Проблема распознавания естественных изображений.....	75
4.2 Наборы данных, структура свёрточной нейронной сети, типы искажений и моделирование.....	77

4.3 Зависимости точности распознавания изображений от интенсивности размытия.....	81
Выводы по разделу 4.....	84
5. НИЗКОЧАСТОТНАЯ ФИЛЬТРАЦИЯ ИЗОБРАЖЕНИЙ ДЛЯ ПРОТИВОДЕЙСТВИЯ СОСТЯЗАТЕЛЬНЫМ ИСКАЖЕНИЯМ .....	86
5.1 Методы противодействия состязательным искажениям.....	86
5.2 Инструменты и методы.....	92
5.2.1 Наборы данных.....	92
5.2.2 Свёрточные нейронные сети.....	95
5.2.3 Состязательные атаки .....	97
5.3 Разработанный метод противодействия высокочастотным искажениям .....	99
5.4 Постановка эксперимента.....	103
5.5 Результаты работы предложенного метода .....	105
Выводы по разделу 5.....	110
ЗАКЛЮЧЕНИЕ .....	112
СПИСОК СОКРАЩЕНИЙ И УСЛОВНЫХ ОБОЗНАЧЕНИЙ .....	114
СПИСОК ТЕРМИНОВ .....	115
СПИСОК ЛИТЕРАТУРЫ.....	116
Приложение 1. Акты о внедрении и использовании результатов диссертационной работы.....	140

## ВВЕДЕНИЕ

### **Актуальность темы исследования**

Глубокое обучение и аналитика больших данных на сегодняшний день являются важными областями вычислительных наук. Различные организации сталкиваются с необходимостью внедрения этих направлений в свои рабочие процессы, чтобы не отставать от современных тенденций [1], [2], [3], [4], [5], [6]. Нейронные сети глубокого обучения могут быстро и эффективно выявлять самые сложные закономерности в данных на высоких уровнях абстракции, в то время как эти закономерности в первом приближении не наблюдаются. Использование машинного обучения может решить проблемы прогнозирования и автоматизации во многих областях жизни, например, таких как распознавание речи [7], [8], компьютерное зрение [9] [10], [11] и визуализация данных [12].

Технологии автоматического распознавания находят самое широкое применение в обработке изображений. Свёрточные нейронные сети (СНС, англ. CNN – «Convolutional neural network») все более успешно применяются для обработки изображений, распознавания символов и рукописного текста [13], распознавания номерных знаков [14], обнаружения патологий человека, растений и животных [15], [16], [17], [18], распознавания лиц и эмоций [19], [20], выделения объектов интереса в видеопотоке [21], [22] и т.д.

Большинство современных публикаций, рассматривающих свёрточные нейронные сети, посвящено применению известных нейронных сетей к новым наборам данных из различных проблемных областей [23]. Многие публикации посвящены совершенствованию топологий нейронных сетей и методов обучения [24]. Однако в задачах распознавания образов остается много нерешенных проблем. Во-первых, точность распознавания классификаторами бывает низкой или недостаточной. Ложные диагнозы, поставленные автоматами с использованием нейронных сетей, хотя и не являются большой проблемой в настоящее время (поскольку данные, полученные от сети, проверяются оператором), могут стать препятствием для расширения применения алгоритмов

автоматического распознавания в будущем. То же самое можно сказать и, например, о системах автоматического вождения, таких как автомобильные автопилоты.

Во-вторых, на результаты работы нейронных сетей влияют искажения данных, такие как атаки состязательного характера [25], [26]. На рисунке 1 показан пример такой атаки: изображение собаки с внесёнными искажениями малой интенсивности распознается сетью как изображение улитки с большим коэффициентом достоверности.

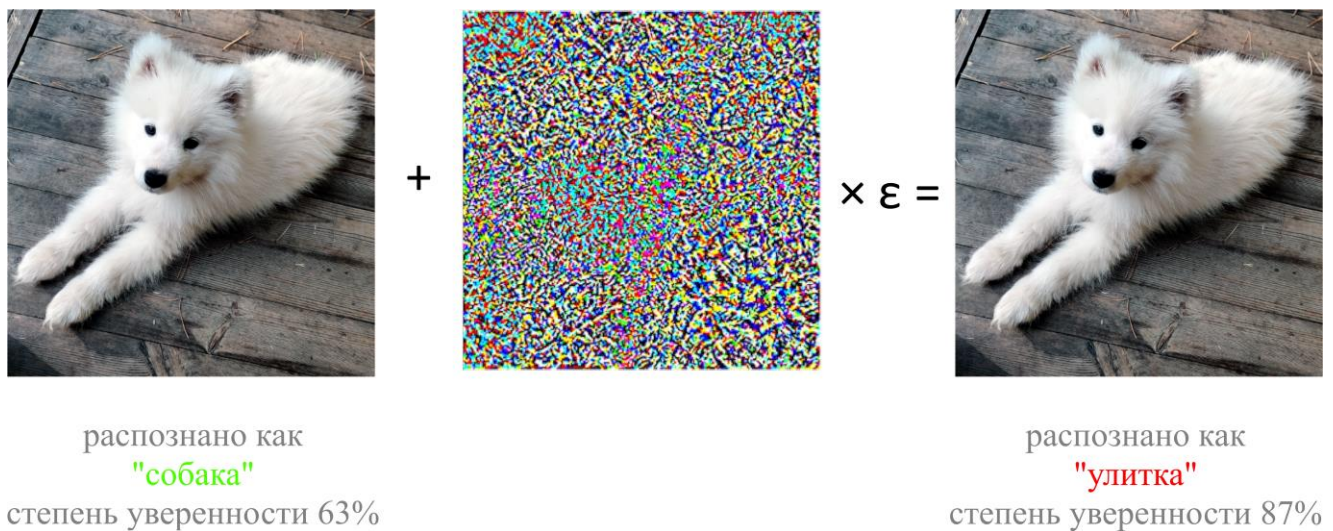


Рисунок 1 – Пример состязательной атаки на свёрточную нейронную сеть

В-третьих, не существует универсального подхода к оценке оптимальности и робастности обученной нейронной сети. Нельзя заранее предсказать, как поведет себя обученная нейронная сеть при получении новых данных, и нельзя быть однозначно уверенным, что сеть правильно распознает новые данные, особенно если статистические характеристики новых данных отличаются от характеристик данных, использованных для обучения.

Анализ недавних публикаций показал, что исследования робастности проводятся в основном с точки зрения построения кривой "точность-полнота" [27]. Некоторые публикации посвящены оценке успешности состязательных атак [28]. Недавние работы, касающиеся количественной оценки влияния

неопределенности в данных, не дали решений для повышения помехоустойчивости нейронных сетей [29]. Исследование помехоустойчивости нейронных сетей на данный момент находится на начальной стадии. Однако такое исследование представляется ключом к решению проблем, связанных с состязательными атаками, и повышению надежности и корректности распознавания различных данных нейронными сетями. В связи с вышеизложенным, тема исследования вопросов, касающихся повышения помехоустойчивости и точности распознавания изображений с помощью свёрточных нейронных сетей, является актуальной.

### **Степень разработанности темы**

Важный вклад в развитие тематики помехоустойчивости нейронных сетей внесли С. А. Доленко [30], [31], [32], И. И. Соловьёв [33], [34], [35], И. В. Оселедец [36], М. В. Терешонок [33], [37], Ian J. Goodfellow [38], [39], [40], Geoffrey Hinton [41], [42], [43], Alex Krizhevsky [44], Yann LeCun [45], [46], Christian Szegedy [47], [48], [49], Ilya Sutskever [50], [51], Yoshua Bengio [52], [53], [54].

Свёрточные нейронные сети обеспечивают высокие результаты регрессии и классификации во многих задачах [55], показывают высокую эффективность классификации наборов данных с тысячами классов [56], вытесняют другие методы в таких задачах, как распознавание речи [57] и рукописного текста [58], в биомедицинских приложениях [59] и т.д. При этом нейронные сети уязвимы к атакам, основанным на изменениях входных данных таким образом, чтобы незаметно для человека исказить различаемые нейронной сетью признаки, или создать примеры, хорошо различаемые нейронной сетью, но являющиеся шумом для восприятия человека – так называемые состязательные атаки. Такие атаки могут быть опасны для систем с применением нейронных сетей, например, авторы статьи [60] демонстрируют атаку на системы распознавания голосовых команд, создающую аудиосигнал, распознаваемый системами как голосовая команда, но воспринимаемый человеком как шум. Множество исследований [61],

[62], [63] показало, что такие искажения возникают и по естественным причинам. Существует несколько методов борьбы с чрезмерным снижением качества распознавания искажённых изображений, в частности, оптимизация структуры нейронной сети [64], [65], [66]. Другим перспективным методом является так называемое «состязательное обучение» [67], [68], [69]. Важным методом устойчивого обучения также является аугментация данных [70], [71]. В большом числе исследований не продемонстрировано значительных успехов при использовании аугментации обучающих данных, однако, в статье [72] авторы показали, что аугментация обучающих данных совместно с усреднением весов модели по ансамблю нейронных сетей [73] может значительно повысить робастность (помехоустойчивость) сети. На данный момент не разработан системный подход к изучению влияния методов искажения данных на качество обучения и помехоустойчивость нейронной сети.

**Объектом исследования** являются автоматы распознавания изображений на основе свёрточных нейронных сетей.

**Предметом исследования** являются характеристики помехоустойчивости (робастности) автоматов распознавания изображений на основе свёрточных нейронных сетей.

**Цель диссертационного исследования** – обеспечить повышение точности распознавания свёрточной нейронной сетью изображений при наличии в них искажений различной природы, описываемых разнообразными математическими моделями.

#### **Задача диссертационного исследования**

Научная задача диссертационного исследования состоит в разработке метода оптимальной аугментации обучающих изображений, обеспечивающего повышение точности распознавания тестовых изображений при наличии в них искажений различной природы.

Научная задача разделена на три частные научные задачи:



1. Нахождение оптимального значения неопределённости в обучающих изображениях.

2. Выбор оптимальных пропорций аугментированных и исходных изображений в обучающем наборе.

3. Нахождение оптимального метода предварительной обработки тестовых данных.

### **Научная новизна**

Доказательство существования оптимального значения неопределённости в *обучающих* изображениях, позволяющего достичь максимальной интегральной точности распознавания *тестовых* изображений с различными искажениями при заданном пороге минимальной точности распознавания получено автором впервые [67], [71].

Подход к повышению точности распознавания изображений, подвергнутых состязательным атакам, на основе низкочастотной фильтрации изображений в совокупности с предварительным обучением нейронной сети размытыми изображениями, предложен автором впервые [74].

### **Теоретическая значимость работы**

Теоретическая значимость результатов диссертационного исследования обусловлена вкладом в развитие исследований робастности и устойчивости методов искусственного интеллекта к внешним воздействиям, в том числе:

1) разработкой метода нахождения оптимума количества искажений в обучающих данных;

2) разработкой метода противостояния высокочастотным искажениям;

3) доказательством методом статистического моделирования существования оптимального значения неопределённости в *обучающих* изображениях, позволяющего достичь максимальной интегральной точности распознавания *тестовых* изображений с различными искажениями при заданном пороге минимальной точности распознавания [67];

4) разработкой подхода к повышению точности распознавания изображений на основе низкочастотной фильтрации изображений.

### **Практическая значимость работы**

Предложенный метод аугментации *обучающих* изображений позволяет повысить точность распознавания *тестовых* изображений, что может быть использовано в различных практических приложениях. Практическая значимость подтверждена актом использования результатов диссертационной работы.

### **Личный вклад**

Все основные научные положения, а также промежуточные выводы, представленные в диссертации, получены автором лично. Из публикаций, написанных в соавторстве, в диссертации использованы только части, подготовленные автором лично.

### **Методология и методы исследования**

В работе использованы следующие методы: численное моделирование, теория вероятностей, математическая статистика, статистическое моделирование, теория машинного обучения, методы искусственного интеллекта, методы цифровой обработки изображений.

### **Апробация и публикация результатов**

По материалам исследования всего опубликовано 19 научных трудов. Основные результаты диссертационной работы изложены в 9 печатных публикациях в рецензируемых изданиях, входящих в список ВАК или индексируемых в международных базах данных Web of Science и Scopus [65], [66], [67], [71], [74], [75], [76], [77], [78].

Материалы диссертационной работы были доложены и одобрены на четырёх научно-технических конференциях:

1. XIV Международная отраслевая научно-техническая конференция «ТЕХНОЛОГИИ ИНФОРМАЦИОННОГО ОБЩЕСТВА», Москва, МТУСИ, 18-19 марта 2020 г. [75];

2. 2020 Systems of Signal Synchronization, Generating and Processing in Telecommunications (SYNCHROINFO-2020), Светлогорск, Россия, 01 — 03 июля 2020 г. [78];

3. 2022 Systems of Signal Synchronization, Generating and Processing in Telecommunications (SYNCHROINFO-2022), Архангельск, Россия, 29 июня — 01 июля 2022 г. [71];

4. Объединенный семинар лаборатории "Физика наноструктур" МГУ и Лаборатории сверхпроводящих и квантовых технологий ВНИИА.

### **Реализация и внедрение результатов**

Полученные в ходе диссертационного исследования алгоритмы, программы и методики их применения реализованы в НИР «Шеренга-2020», НИР «Интеллект-В» и СЧ ОКР «5P17K302-МТУСИ», выполненных по Государственному заказу в МТУСИ в 2018 — 2023 годах. Акт об использовании результатов приведён в приложении.

Получено 9 свидетельств об официальной регистрации программы для ЭВМ [79], [80], [81], [82], [83], [84], [85], [86], [87].

**Работа соответствует паспорту специальности 2.2.13 «Радиотехника, в том числе системы и устройства телевидения» по направлению исследований в части пункта № 11 «Разработка информационных технологий, в том числе цифровых, а также с использованием нейронных сетей для распознавания сигналов, изображений и речи в интеллектуальных радиотехнических, робототехнических системах технического зрения».**

### **Положения, выносимые на защиту:**

1. Существует оптимальное значение неопределённости в *обучающих* изображениях, позволяющее достичь максимальной интегральной точности распознавания *тестовых* изображений с различными искажениями.

2. Оптимальное значение неопределённости в *обучающих* изображениях может быть оценено методом статистического моделирования. Использование обучающего набора данных с оптимальным значением неопределённости позволяет снизить вероятность ошибки распознавания в среднем в 20 раз по сравнению с использованием исходного набора изображений без дополнительных искажений.

3. Существует оптимальный способ аугментации *обучающих* изображений, позволяющий повысить интегральную точность распознавания *тестовых* изображений с различными искажениями при заданном пороге минимальной точности распознавания, без увеличения объёма обучающей выборки. Использование оптимального способа аугментации позволяет снизить вероятность ошибки распознавания в среднем на 60 % по сравнению с использованием исходного набора изображений без дополнительных искажений.

4. Низкочастотная фильтрация изображений в совокупности с предварительным обучением нейронной сети размытыми изображениями позволяет в среднем в 8,8 раз снизить вероятность ошибки распознавания изображений, подвергнутых состязательным атакам, по сравнению с использованием исходного набора изображений без дополнительных искажений.

#### **Объем и структура работы**

Текст диссертации изложен на 141 странице и включает введение, пять разделов, заключение, список сокращений и условных обозначений, список терминов, список литературы и приложение. Список литературы содержит 247 наименований. В работе представлены 56 рисунков и 2 таблицы.

# 1. АНАЛИЗ МЕТОДОВ АУГМЕНТАЦИИ ОБУЧАЮЩИХ ДАННЫХ ДЛЯ РАСПОЗНАВАНИЯ ИЗОБРАЖЕНИЙ

## 1.1 Проблемы практического применения свёрточных нейронных сетей

Модели глубокого машинного обучения являются наиболее успешными и прогрессивными методами решения задач, формализация способов решения которых затруднительна. Отличительной особенностью глубоких нейронных сетей является наличие более чем одного слоя обработки между входным и выходным слоями [88], [89] (рисунок 1.1).

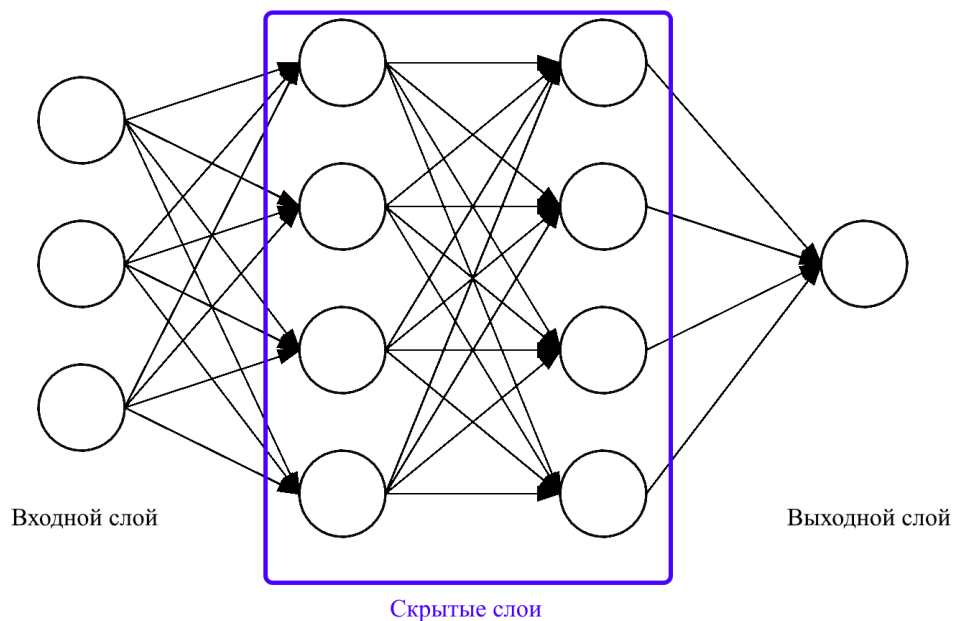


Рисунок 1.1 – Глубокая нейронная сеть

Как видно из рисунка 1.1, скрытые слои, включающие искусственные нейроны, принимают набор взвешенных входов от предыдущих слоёв и вычисляют выходные значения через внутреннюю функцию активации.

Сети, в которых используется множество слоёв, характеризуются высоким разнообразием преобразований. Методы оптимизации и поиска правил для

решаемой нейронной сетью задачи работают с многомерными функциями признаков с использованием структуры сети (рисунок 1.2).

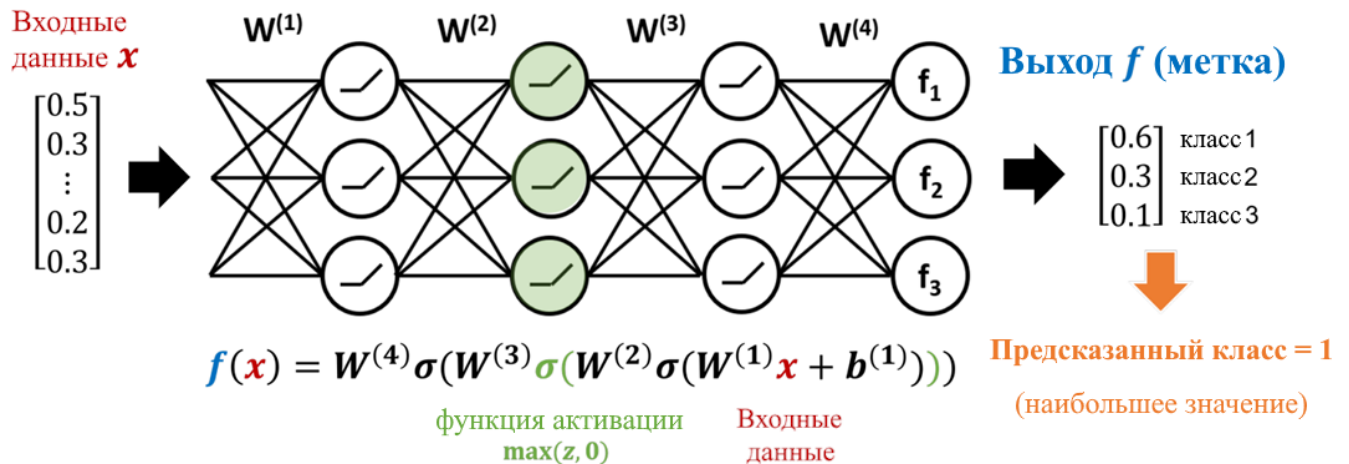


Рисунок 1.2 – Представление нейронной сети в виде функции [90]

На рисунке 1.2 показано представление нейронной сети в виде функции, принимающей в качестве аргументов входные данные, и выдающей значения в виде выходной стохастической матрицы вероятностей соответствия входных данных соответствующему классу. Таким образом, любая архитектура нейронной сети сводится к некоторой функции  $f(x)$ . Обучение нейронной сети сводится к минимизации функции ошибки, представляющей собой расстояние (в заранее выбранной метрике) между истинной (целевой) и фактической выходной матрицей вероятностей [91], [39].

Так как СНС благодаря своей способности к выявлению неявных признаков быстро и качественно справляются со многими задачами, автоматизация решения которых была ранее затруднительна, существует множество сфер применения СНС [92], [93]. Различные архитектуры СНС используются в обработке речи [94], текста и видеопотока (рекуррентные нейронные сети) [95], [96], [97], для классификации изображений (свёрточные нейронные сети), для поиска изменений и верификации (сиамские архитектуры) [76], [98], локализации [99], [100] и т.д. (рисунок 1.3).

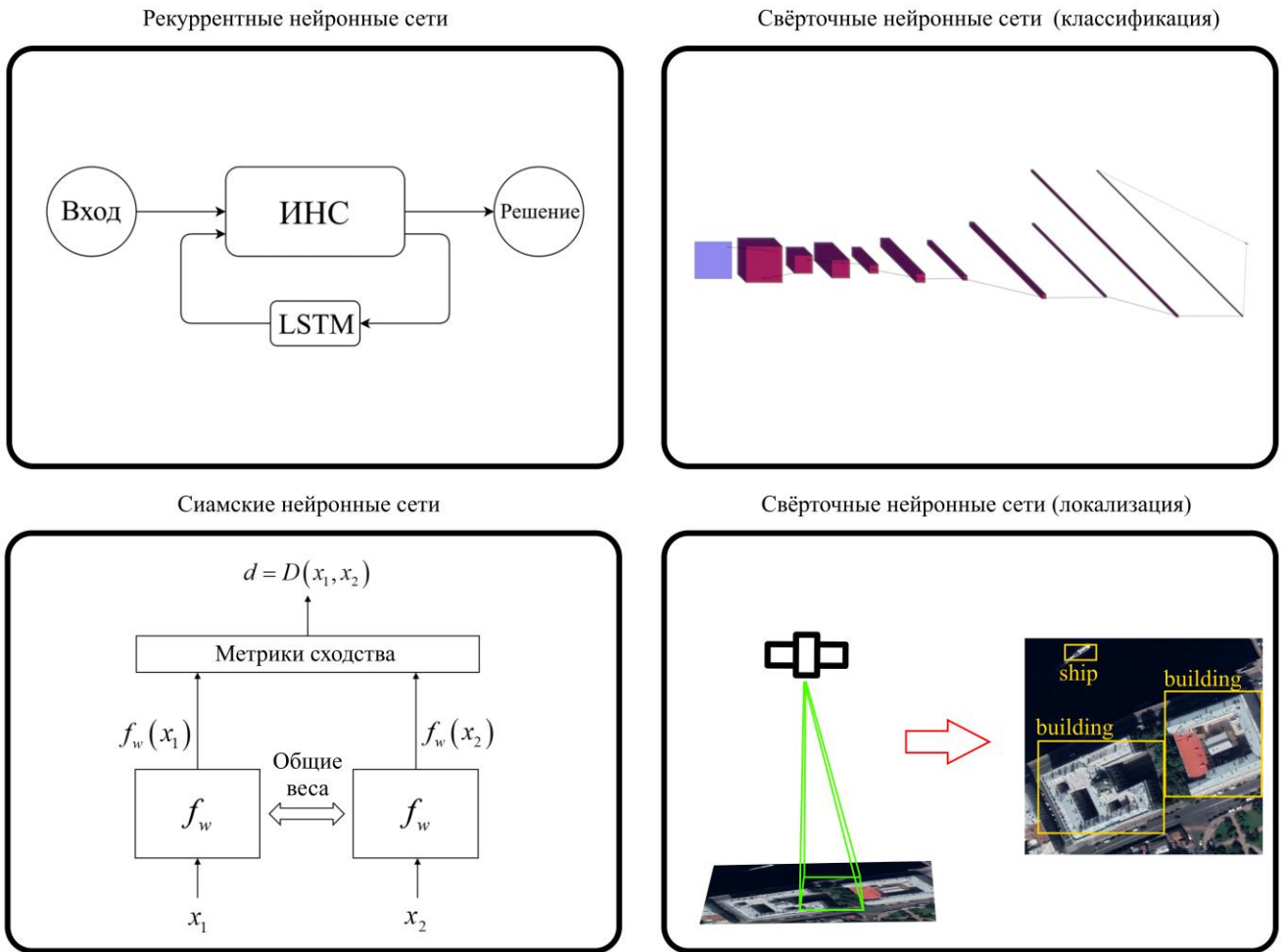


Рисунок 1.3 – Некоторые применения глубоких нейронных сетей [101], [66], [76], [102]

Сети, показанные на рисунке 1.3, по сути являются адаптацией простейшей нейронной сети для решения задач разного рода.

Точность распознавания в задачах классификации является ключевой метрикой качества модели. С ограниченной обучающей выборкой для повышения точности классификации разрабатываются и адаптируются структуры сетей [103], [104], [105], проводится работа по поиску оптимальных гиперпараметров обучения сети [106], [107], [108], а также расширение доступных обучающих данных – аугментация [109], [110].

Рост доступных вычислительных мощностей позволяет использовать сети с тысячами слоёв [111], [112]. Например, модель сети VGG16 содержит около 140 миллионов параметров [113], задействованных в обучении, а глубина модели

InceptionResNetV2 достигает 572 слоёв [114]. Для решения задач различного рода работа с архитектурами оправдана, и большая часть исследований в области СНС направлена на решение проблем с архитектурами сетей и с особенностями обучения этих архитектур [115], [116], [117].

При этом в пределах одной задачи для повышения качества работы нейронных сетей наибольшее значение имеет разнообразие и репрезентативность информации, используемой для обучения. Однако основная проблема обучения СНС заключается в том, что обучающих данных, как правило, недостаточно для полного обобщения, либо доступные данные значительно отличаются от необходимых для решения конкретной задачи. СНС эффективно определяют уникальные признаки изображений [118], но в реальных условиях не все возможные признаки обязательно присутствуют на изображениях, признаки не очевидны, взаимозависимы и вариативны. Описанные проблемы обучающих наборов данных ведут к снижению обобщающей способности сетей, что эквивалентно общему снижению качества классификации, локализации и регрессии. При решении задач классификации с помощью СНС проблема недостаточности данных проявляется явно – количество подходящих естественных изображений (фотографий) ограничено, описание математических моделей генерации этих изображений затруднительно и зачастую невозможно (при недостаточности или неточности описания использование генерируемых данных неэффективно), накопление достаточного набора данных обучения – весьма трудоёмкий процесс, требующий соблюдения множества условий – выдерживания баланса классов и репрезентативности набора данных, исключения признакового дисбаланса. Помимо этого, для использования набора данных (изображений) в качестве обучающего набора требуется предварительная ручная разметка каждого из уникальных образцов. Решение этих проблем на этапе формирования обучающей выборки даст значительный прирост качества распознавания при обработке реальных данных, однако отдача от трудозатрат постоянно снижается.



Снижение обобщающей способности сети в процессе обучения с использованием доступных данных проявляется, помимо прочего, эффектом переобучения [119]. Переобучение представляет собой "запоминание" параметров обучающего набора данных, но не выявление основных информативных признаков. Точность классификации нового набора данных при этом, как правило, снижается. На рисунке 1.4 наблюдается повышение значения ошибки тестового набора данных при снижении значения ошибки обучающего набора данных с каждой последующей итерацией обучения, что свидетельствует о переобучении СНС (рисунок 1.4).

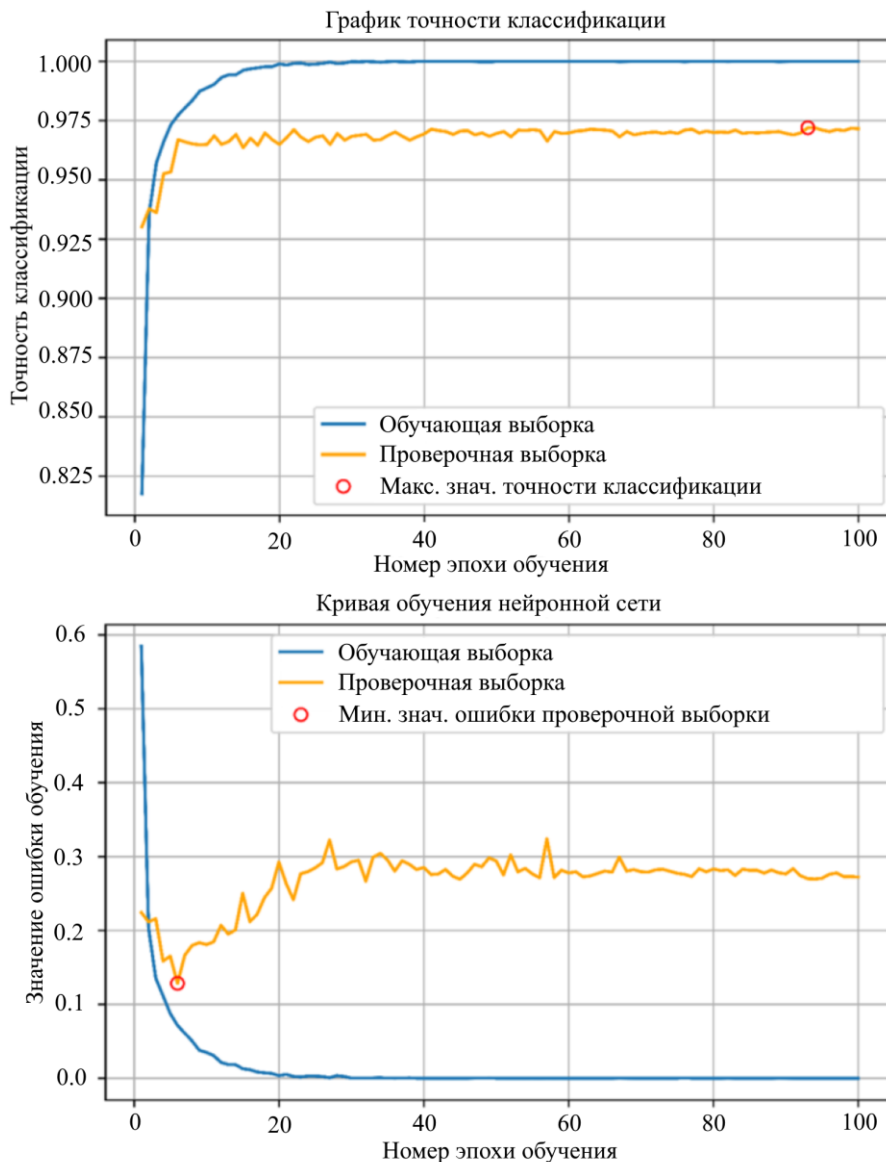


Рисунок 1.4 – Переобучение нейронной сети [75]

Таким образом, как видно из рисунка 1.4, с каждой последующей эпохой обучения функция, реализуемая СНС, позволяет всё лучше находить оптимум в признаковом пространстве обучающих данных, но не данных в целом. Нейронная сеть «запоминает» обучающий набор данных, и ее дальнейшее обучение не обеспечит улучшения работы [120], [121], [122]. Это может свидетельствовать как о проблемах с процессом обучения, так и с архитектурой СНС или обучающими данными, однако необходимо заметить, что отсутствие тенденции некоей модели к переобучению не означает высокого качества работы этой модели.

Наличие в свободном доступе большого числа открытых наборов данных не является гарантией повышения качества работы системы распознавания. Состав наиболее подходящего набора обучающих данных во многом зависит от условий, в которых планируется использование обученной системы и от источника данных для распознавания. Например, в работе [123] авторы указывают, что обобщение полученной в процессе обучения сети информации на новые наборы данных происходит плохо, в особенности для систем классификации (сети, обученные на картинах или набросках, не справляются с распознаванием естественных изображений). Такие виды изменений входных данных обобщены в работе [124]. Авторы классифицируют различия в наборах данных на «сдвиг предметной области» (англ. Domain shift) [125] и «сдвиг субпопуляции» (англ. Subpopulation shift) [126]. Под «обобщением предметной области» авторы понимают разницу в данных, полученных различными способами (например, данные, полученные от пациентов из разных больниц [127], изображения, сделанные разными камерами [128], биологические пробы из разных типов клеток [129] или спутниковые снимки из разных стран или временных периодов [130]). Сдвиг субпопуляции – различие пропорций классов в обучающем и распознаваемом наборе данных [131]. Авторы доказывают, что качество работы нейронных сетей, обученных стандартным подходом, значительно ниже на данных из другого набора, чем на том же наборе данных. Это свойственно и для моделей, обученных с помощью существующих методов

борьбы с искажениями данных, и авторы [124], [132] подчеркивают необходимость новых методов обучения моделей, более устойчивых к сдвигам распределения, возникающих на практике. Следует заметить также, что результаты, полученные авторами, подтверждают, что параметры данных в обучающем наборе должны быть приближены к параметрам данных для распознавания. Для решения данных проблем важным методом является аугментация данных.

## 1.2 Аугментация данных

Аугментация данных – один из способов уменьшить переобучение моделей машинного обучения, повысить устойчивость модели к искажениям и увеличить разнообразие и количество обучающих данных, при этом используя только данные, доступные изначально [133]. Аугментация данных также может использоваться для повышения робастности моделей классификации (за счет увеличения доли важных для классификации признаков – инвариантов).

Термин "аугментация данных" относится к итеративным алгоритмам оптимизации [134], в общей теории статистики для детерминированных алгоритмов этот метод был введен Демпстером, Лэрдом и Рубиным в 1977 году [135]. В своей статье для максимизации функции правдоподобия авторы использовали алгоритм EM («ожидаемое-максимальное», или «максимизация ожидаемого», англ. «Expectation-maximization»). Модификацию алгоритма EM для расчета апостериорных вероятностей при использовании стохастических алгоритмов предложили Таннер и Вонг в 1985 [136]. Авторы указывают, что использование аугментации данных позволяет производить моделирование при недостаточном количестве исходных данных.

Первые наработки, демонстрирующие эффективность дополнения данных для обучения СНС, были получены на основе простых преобразований изображений, таких как горизонтальное отражение, расширение цветового пространства и произвольная обрезка изображения. Одно из первых

практических применений аугментации обучающих данных в моделях многослойных нейронных сетей, использующихся для распознавания рукописного текста, была использована в 1998 году [137]. Авторы использовали 3 набора обучающих данных, состоящих из 15000, 30000 и 60000 уникальных изображений. Проверка обученных экземпляров нейронных сетей на этих наборах данных показала снижение значения ошибки распознавания с ростом числа уникальных изображений в наборах обучающих данных (рисунок 1.5). Для проверки этой гипотезы авторы из исходных изображений путём добавления аффинных преобразований получили набор данных обучения, состоящий из 600000 изображений и достигли ещё меньшего значения ошибки распознавания тестового набора данных.

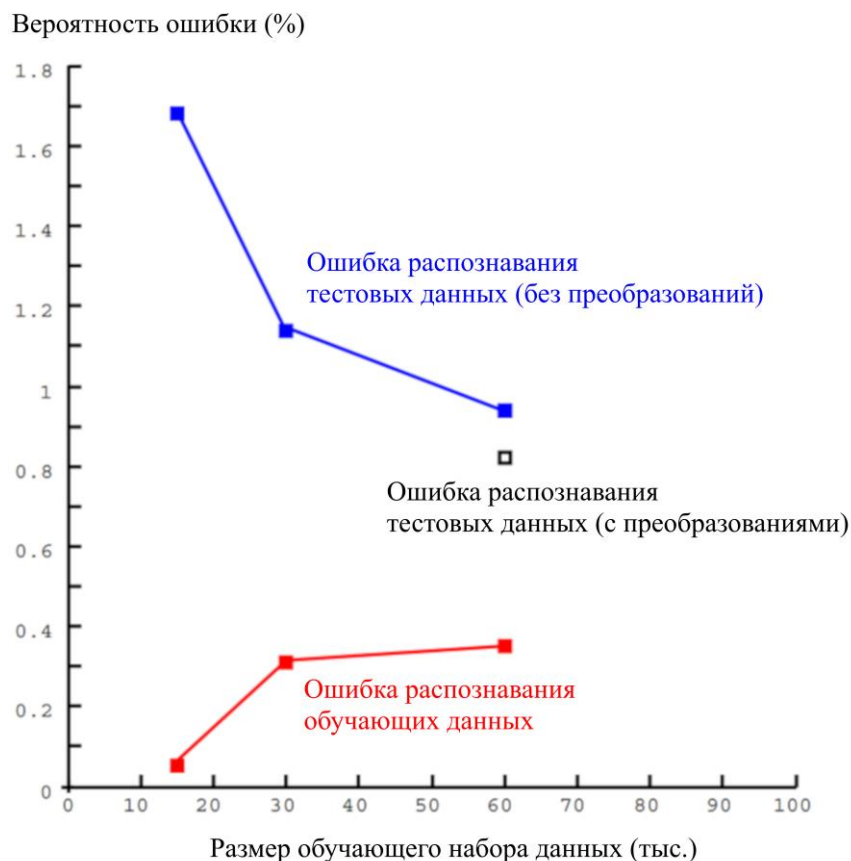


Рисунок 1.5 – Минимальное значение ошибки обучающих и тестовых данных при использовании наборов данных, включающих 15000, 30000 и 60000 уникальных изображений, а также набора, включающего 600000 изображений, полученных с использованием аффинных преобразований [137]

Искажения представляли собой комбинации следующих плоскостных аффинных преобразований: горизонтальные и вертикальные отражения, масштабирование, сжатие (одновременное горизонтальное сжатие и вертикальное растяжение, или наоборот) и горизонтальный сдвиг. Как видно из рисунка 1.5, когда изменённые данные использовались для обучения, коэффициент ошибок теста снизился до 0,8% (вместо 0,95% при обучении с изображениями без искажений). Были использованы те же параметры обучения, что и для данных без аугментации. Общая продолжительность сеанса обучения осталась неизменной (20 эпох по 60000 экземпляров в каждой).

СНС благодаря своей архитектуре инвариантны к некоторой степени таких искажений, как сдвиг и масштабирование, однако заранее неизвестно, приведут ли другие виды искажений к существенному изменению признаков, используемых экземпляром нейронной сети для идентификации класса изображения. Использование аугментированных данных может решить эту проблему.

Стандартной практикой аугментации является изменение различными способами определённой части изображений в наборе данных без расширения этого набора данных. Однако недавние исследования показали результативность расширения исходных наборов данных. Авторы [138] показали, что использование множества экземпляров для каждого из изображений повышает точность классификации тестовых данных как при обучении на малых, так и на больших наборах данных.

Известны исследования, использующие аугментацию данных для балансировки классов. Дисбаланс классов в машинном обучении представляет из себя неравномерное распределение данных различных классов. Это явление свойственно для большинства приложений машинного обучения [139]. При обучении с использованием несбалансированных наборов данных модель будет ошибочно склонна к классу с большим числом экземпляров в обучающем наборе данных. В работе [140] авторы доказывают, что данная проблема часто проявляется в реальных приложениях машинного обучения, и требует поиска её

решений. Два наиболее простых и используемых метода решения проблемы дисбаланса классов – oversampling и undersampling (рисунок 1.6) – в анализе данных – методы, используемые для корректировки распределения классов в наборе данных.

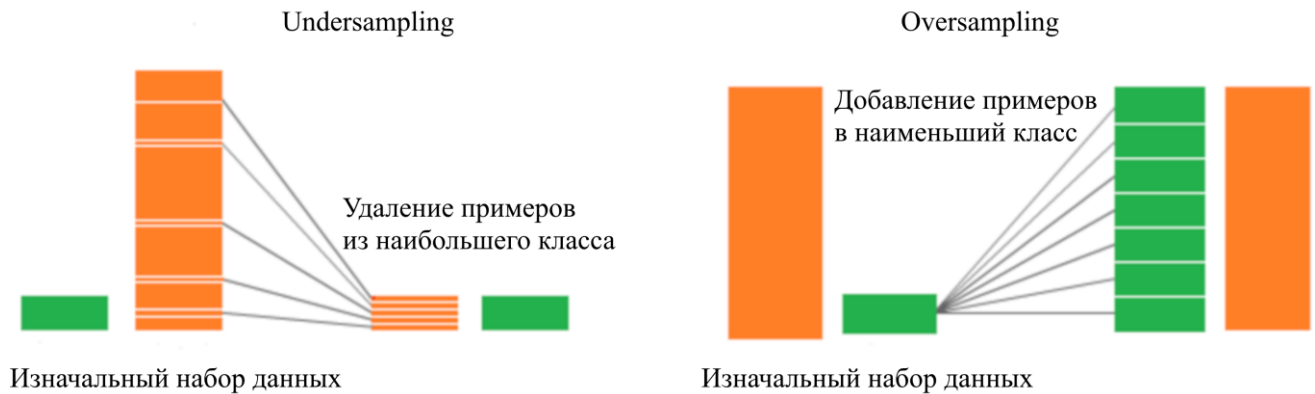


Рисунок 1.6 – Oversampling и undersampling [139]

Методами oversampling и undersampling, таким образом, происходит выравнивание количества экземпляров различных классов. В работе [140] показано, что использование метода Oversampling даёт наилучшие результаты при обучении классификаторов, что требует использования аугментации для классов с наименьшим числом уникальных изображений.

### 1.3 Виды и способы аугментации данных

Как было указано ранее, первые методы аугментации для обучения СНС, предназначенных для решения задач классификации, включали горизонтальные и вертикальные отражения, масштабирование, сжатие изображений и горизонтальный сдвиг. Так как использование аугментации для обучения моделей нейронных сетей оказало положительное влияние на качество работы модели, к настоящему времени исследователями успешно применено множество различных способов преобразования изображений для аугментации, при этом используемые способы часто специфичны для решаемой исследователями

практической задачи. В работе [141] авторы используют искажения изображений символов, и доказывают, что использование методов аугментации специфично для решаемой задачи. Основные способы аугментации изображений, используемые в современных исследованиях, показаны на рисунке 1.7 [142]. В то же время все эти методы и их комбинации позволяют как преодолеть проблемы различия данных типа обобщения предметной области, так и сдвига субпопуляции.

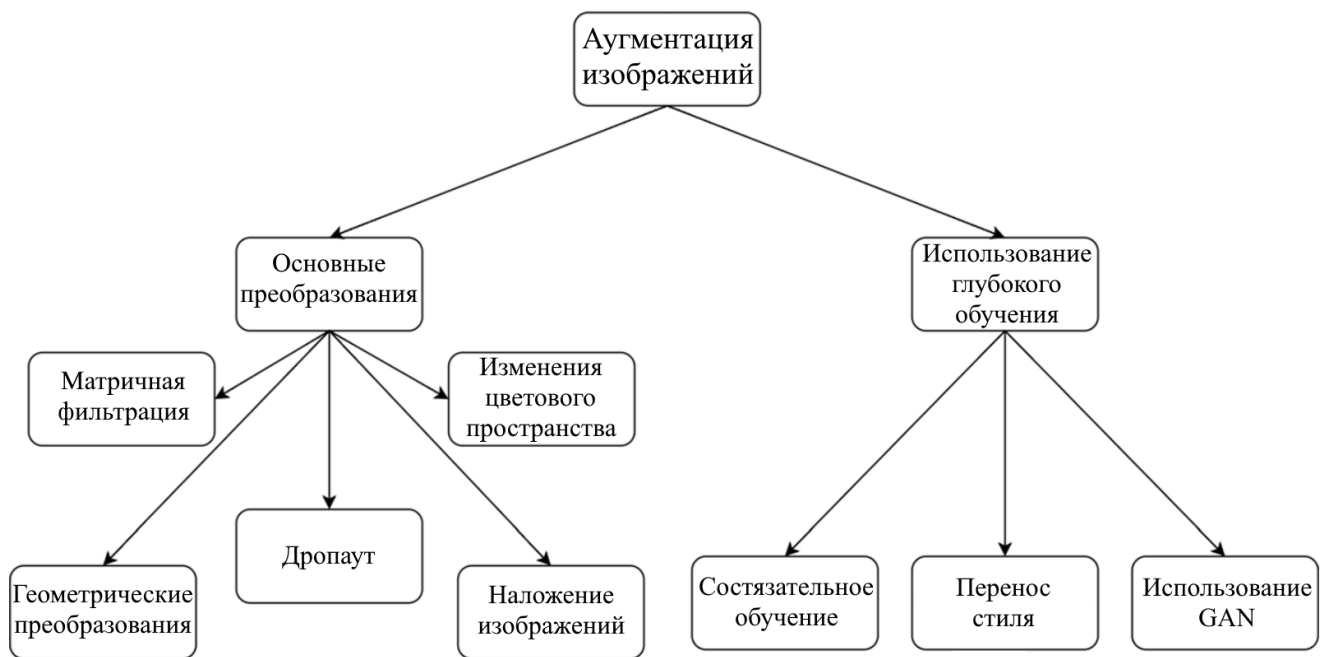


Рисунок 1.7 – Основные способы аугментации изображений

На рисунке 1.7 показано, что все виды преобразований изображений для аугментации делятся в основном на два типа: преобразования изображений различными математическими методами и преобразования с помощью подходов глубокого обучения. Преобразования математическими методами, в свою очередь, можно условно разделить на 5 видов, преобразования с использованием подходов глубокого обучения – на 3.

### 1.3.1 Матричная фильтрация

Фильтрация изображений с применением матриц – изменение изображений с помощью ядер свертки (матриц свертки). С помощью различных матриц можно, например, изменять резкость изображения, добавлять размытие и выделять границы на изображениях (рисунок 1.8).

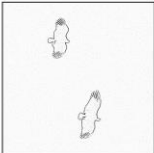





	Выделение границ	Размытие в движении	Размытие по Гауссу	Повышение резкости	Раздвоение	Тождественное отображение
Ядро свёртки	$\begin{bmatrix} -1 & -1 & -1 \\ -1 & 8 & -1 \\ -1 & -1 & -1 \end{bmatrix}$	$\frac{1}{3} \begin{bmatrix} 0 & 0 & 0 \\ 1 & 1 & 1 \\ 0 & 0 & 0 \end{bmatrix}$	$\frac{1}{16} \begin{bmatrix} 1 & 2 & 1 \\ 2 & 4 & 2 \\ 1 & 2 & 1 \end{bmatrix}$	$\begin{bmatrix} 0 & -1 & 0 \\ -1 & 5 & -1 \\ 0 & -1 & 0 \end{bmatrix}$	$\frac{1}{2} \begin{bmatrix} 0 & 0 & 0 \\ 1 & 0 & 1 \\ 0 & 0 & 0 \end{bmatrix}$	$\begin{bmatrix} 0 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix}$
Результат						

Рисунок 1.8 – Применение различных матриц свёртки

Преимуществом матричной фильтрации является возможность её включения в структуру СНС, так как принцип применения матричной фильтрации идентичен работе этих сетей. Таким образом можно осуществлять аугментацию «на лету» за счёт повышения потребляемых сетью вычислительных мощностей.

### 1.3.2 Геометрические преобразования

Геометрические преобразования – класс простейших в реализации методов аугментации, основанный на сдвигах, поворотах изображений, сжатиях-растяжениях, а также локально-аффинных преобразованиях. Преимуществом метода является простота реализации и низкие требования к вычислительным ресурсам. Значительным недостатком можно назвать ненадёжность для некоторых применений аппаратов распознавания, при которых геометрические



преобразования могут изменить класс изображения (например, поворот буквы «Е» может привести к распознаванию как «Ш», цифры «9» – как цифру «6», отражение буквы «Р» – как «Ь»).

### 1.3.3 Изменения цветового пространства

Поскольку естественные изображения, полученные различными моделями матриц и фотокамер, в различных условиях освещенности и обработанные различными моделями процессоров значительно отличаются с точки зрения цветового пространства и динамического диапазона, различные способы изменения цветового пространства для аугментации могут сделать свёрточную сеть более устойчивой к различиям характеристик источников изображений (сдвигам предметной области). Существует множество способов изменения цветового пространства – такие как исключение определённых цветовых каналов, расширение, ограничение или сужение динамического диапазона, преобразование изображения в градации серого или изменение количества тонов (рисунок 1.9).

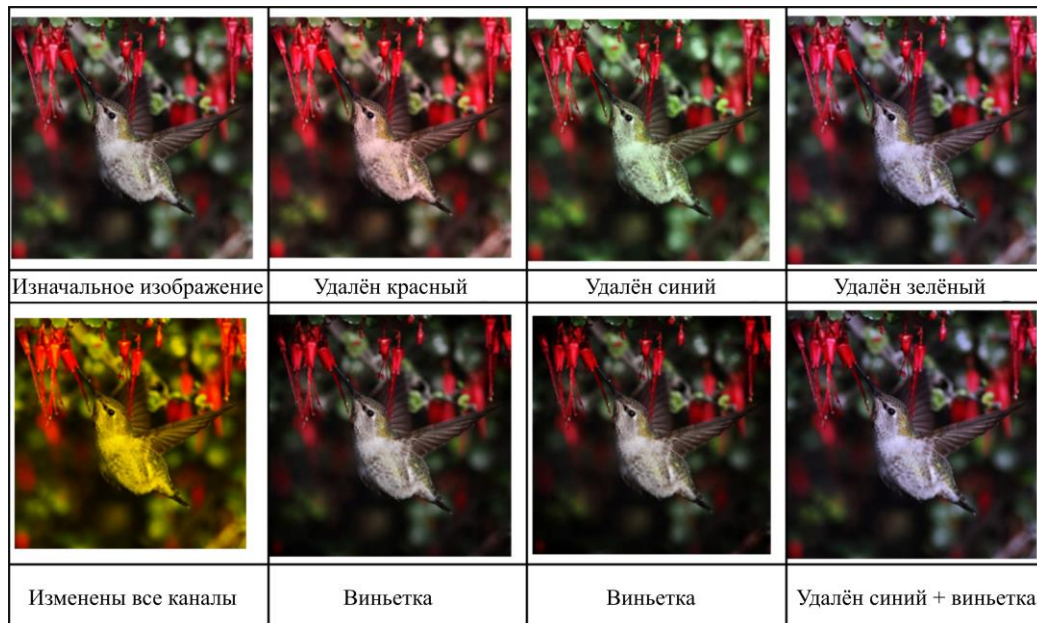


Рисунок 1.9 – Примеры аугментации методом изменения цветового пространства

Стоит заметить, что изменения цветового пространства может также изменить важные для классификации признаки и снизить точность распознавания естественных изображений нейронной сетью.

### 1.3.4 Дропаут

Дропаут (англ. dropout – исключение, выпадение) как метод аугментации – по аналогии с методом регуляризации в СНС – затирание случайных частей изображения с целью снижения влияния определённых признаков на решение сети [144]. Как правило, алгоритм, осуществляющий затирание, случайным образом выбирает область (в данном контексте эту область принято называть регионом) изображения или объекта на изображении и заменяет значения пикселей в этой части на 0 (черный), 255 (белый), средним значением пикселей, случайными значениями или другими произвольными значениями (рисунок 1.10).

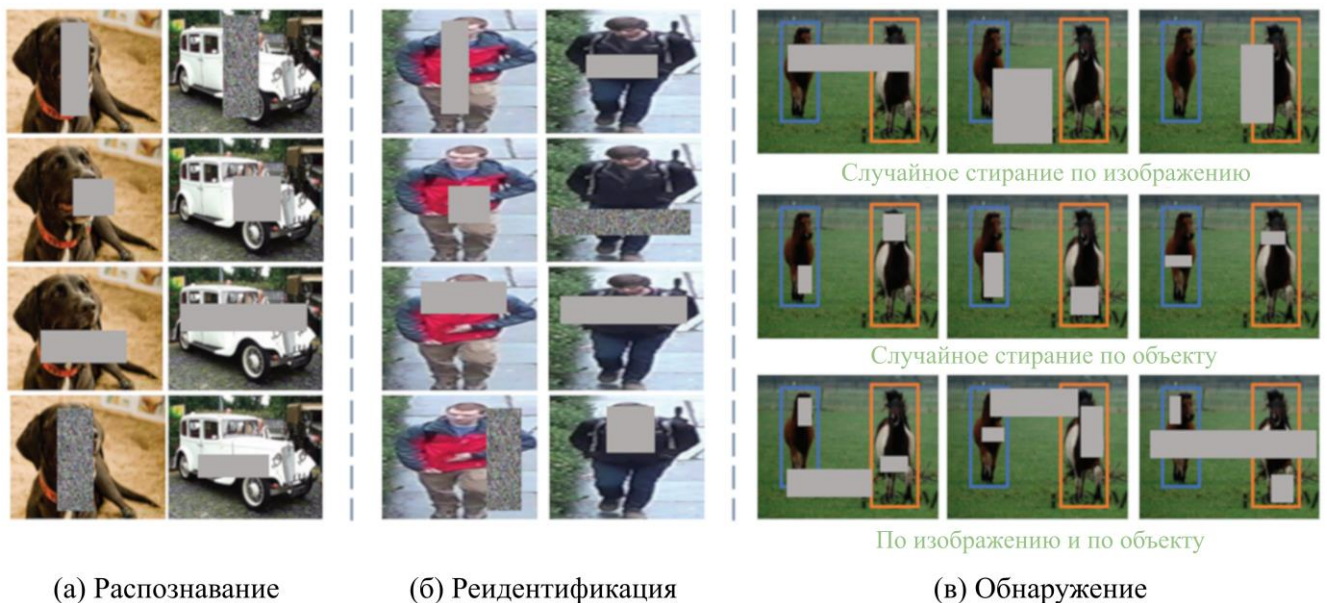


Рисунок 1.10 – Примеры аугментированных изображений методом дропаут для различных применений СНС: (а) распознавание изображений, (б) реидентификация, (в) обнаружение объектов [144]

Исследования [144], [145] продемонстрировали значительное снижение ошибки распознавания при использовании аугментации методом Дропаут. Однако недостатком является вероятное изменение класса изображения (затирание части буквы «О» может изменить на «С») и, соответственно, необходимость ручной верификации аугментированных изображений для некоторых применений и задач.

### 1.3.5 Наложение изображений

Наложение изображений заключается в смешении двух и более изображений различными методами: усреднение значений пикселей [146], различные способы сшивания [147] и комбинации этих методов (рисунок 1.11).

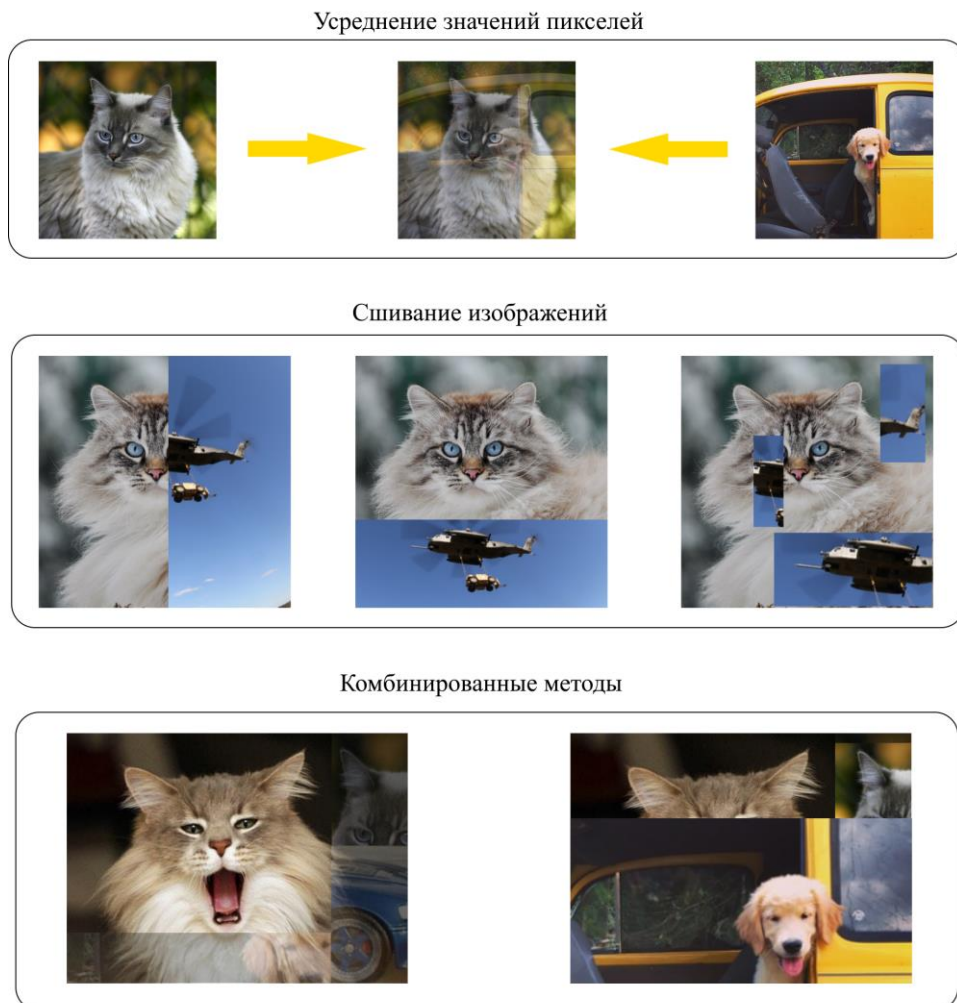


Рисунок 1.11 – Методы аугментации путем наложения изображений

Метод аугментации путем усреднения значений пикселей оказался особенно эффективен на небольших наборах данных. В работе [146] на наборе данных CIFAR-10, состоящем из 1000 уникальных экземпляров (по 100 в каждом классе) автор добился снижения коэффициента ошибок с 43,1% до 31,0%. При использовании метода возможно расширение любых наборов данных до  $N^2 \cdot N$  (где  $N$  – количество уникальных экземпляров в изначальном наборе).

Использование сшивания изображений и комбинированных методов в значительной степени усложняет предварительную подготовку данных, поскольку требует задания более сложных меток класса (в виде набора вероятностей соответствия классам пропорционально представлению на изображении, либо пропорции классов на изображении и расположение соответствующих изображений) (рисунок 1.12).

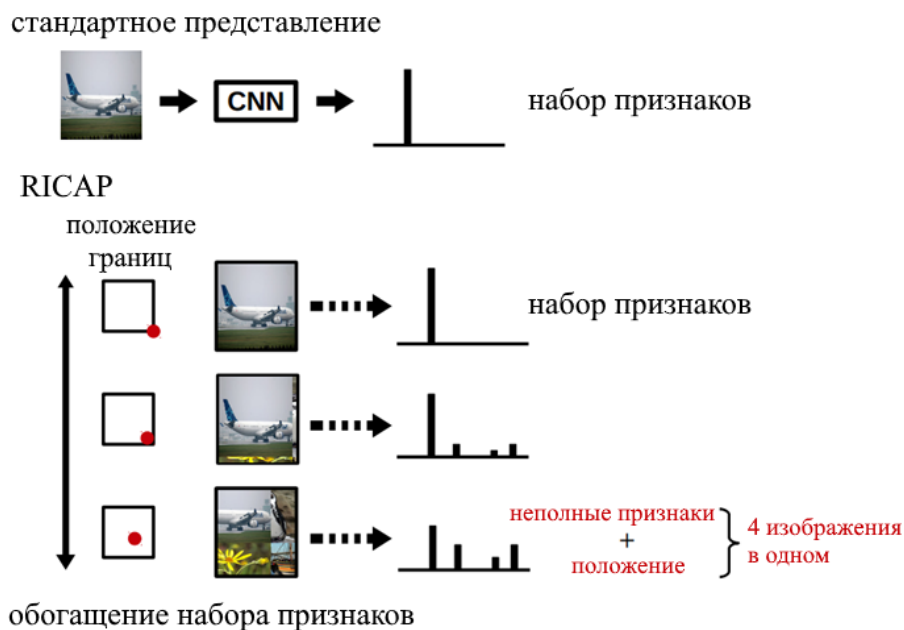


Рисунок 1.12 – Сравнение представления меток классов при использовании изначальных изображений и аугментированных методом сшивания [148]

Авторы [148] заявляют повышение точности классификации для таких применений СНС, как классификация, обнаружение и локализация объектов.

### 1.3.6 Состязательное обучение

В работе [149] авторы показали, что СНС крайне неустойчивы к небольшим искажениям (особенно в виде небольших изменений значений яркости пикселей) и указывают, что не было предложено эффективных методов расчета устойчивости современных классификаторов на основе СНС на различных наборах данных. Незначительные и незаметные для человека изменения полностью искажают работу систем классификации. Более того, в статье [150] продемонстрировано, что большинство изображений (более 68,36% на наборах данных CIFAR-10 и 41,22% ImageNet) могут быть неправильно классифицированы при изменении всего одного пикселя на изображении (рисунок 1.13).

С использованием алгоритма построения состязательной атаки Fast Gradient Sign Method (FGSM), сеть maxout [151] (изначально достигавшая вероятности ошибки 0,45%) неправильно классифицировала 89,4% состязательных примеров со средней уверенностью 97,6%. Более того, с ростом разрешения используемых изображений ошибка распознавания состязательных примеров растёт. Описанная авторами проблема оказала значительное влияние на развитие подходов к обучению и оценке устойчивости систем на основе СНС, поскольку описанный ими факт полностью изменил представление научного сообщества о том, каким образом нейронные сети выявляют признаки из данных.

Одним из первых упоминаний проблемы незаметных искажений является [152]. Авторы также обнаружили, что состязательные искажения относительно устойчивы для нейронных сетей с различным количеством слоев, архитектурой или обученных на различных подмножествах обучающих данных. То есть, состязательные примеры изображений являются переносимыми на различные нейронные сети даже если они обучены с другими гиперпараметрами или на другом наборе данных.

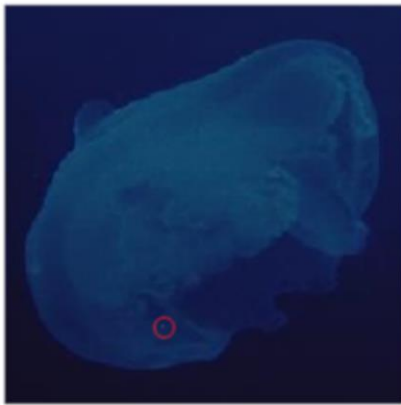




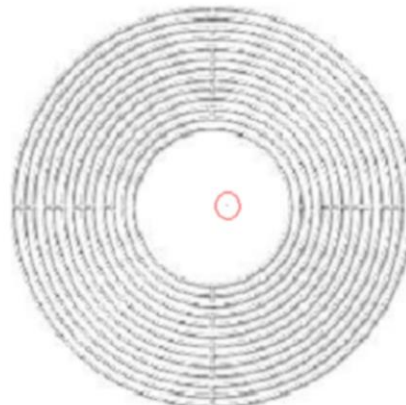
Планетарий  
Мечеть(7.81%)



Покрывало  
Подушка(6.83%)



Медуза  
Ванна(21.18%)



Моток  
Вентилятор(37.00%)

Рисунок 1.13 – Однопиксельные атаки на изображения из наборов данных ImageNet. Измененные пиксели выделены красными кружками. Оригинальные метки классов показаны черным цветом, а итоговые результаты распознавания и соответствующий им коэффициент уверенности приведены ниже [150]

Позже были предложены методы для создания состязательных примеров [25], [149]. С точки зрения аугментации данных и взаимодействия нейронных сетей подход к глубокому устойчивому обучению заложен в использовании двух или более сетей с противоположными целями. Конкурирующая (состязательная) сеть обучается генерации состязательных искажений в изображениях с целью обеспечить неправильную классификацию изображений классификатором, а классификатор обучается противостоять таким состязательным примерам. Состязательные примеры могут являться хорошим источником аугментации,

поскольку такой способ аугментации является эффективным для повышения устойчивости СНС к неочевидным и незаметным для человека искажениям, однако, как показано в [149], часто не повышает точности распознавания данных.

### 1.3.7 Перенос стиля

Перенос стиля (англ. Neural Style Transfer) с помощью нейронных сетей в настоящее время широко известен благодаря своему художественному применению [153], [154], [155], [156]. Однако методы переноса стиля также служат инструментом для аугментации данных. Общая идея алгоритмов в том, чтобы использовать СНС таким образом, что стиль одного изображения может быть перенесен на другое изображение с сохранением его оригинального содержания (рисунок 1.14). Этот эффект достигается путем манипулирования представлениями данных в различных слоях СНС.



Рисунок 1.14 – Примеры изображений, полученных переносом стиля с применением свёрточных нейронных сетей [157]

Одной из самых важных проблем современных систем классификации с применением СНС является их тенденция к извлечению признаков из комбинаций соседних пикселей, а не объекта в целом, что объясняет неустойчивость сетей к шумам и импульсным искажениям (в том числе и к состоятельным искажениям на изображениях). Развитие подхода к обучению сетей с применением методов, выделяющих именно важные признаки объектов на изображениях, вероятно, в дальнейшем позволят решить проблемы восприимчивости сетей к незначительным изменениям изображений. Наибольшим потенциалом в данном направлении обладают современные методы переноса стиля. При использовании этих методов, как правило, меняются условия освещенности объекта на изображении, либо текстура этого объекта. Соответственно, при аугментации набора обучающих данных важен выбор методов переноса стиля, опорных стилей и степени изменения исходного изображения. Самое очевидное применение переноса стиля – при обучении систем распознавания в робототехнике – включает изменение освещенности, погодных условий или времени года. Авторы статьи [158] показывают, что применение аугментированного с использованием методов переноса стиля набора обучающих данных для задач робототехники оказывается более эффективным, чем простой сбор более широкого набора данных.

У такого метода аугментации также есть одно неявное, но важное преимущество – при недостаточном количестве исходных изображений некоего класса становится возможным правильное обучение сети, поскольку упрощается применение метода oversampling из-за расширения вариативности изменений изображения.

### **1.3.8 Генеративно-состязательные сети**

В основе генеративно-состязательных методов лежит использование сети – генератора как источника обучающих данных. Сеть, обученная признакам на исходном малом наборе данных используется в качестве генератора новых



данных, на которых и обучается новая сеть (рисунок 1.15). Такой способ аугментации является эффективным при недостаточном размере изначального набора данных [159].

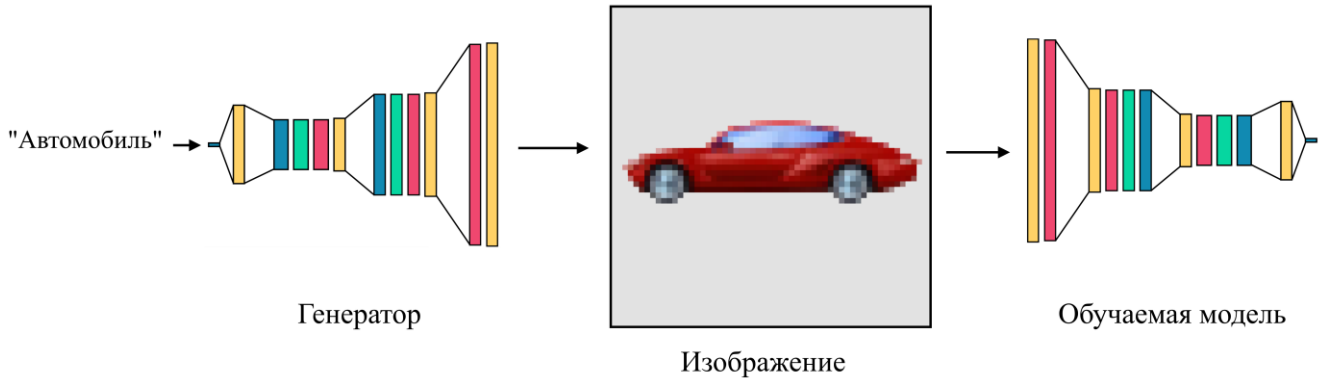


Рисунок 1.15 – Принцип работы GAN

Использование генеративно-сопоставительных сетей (англ. Generative Adversarial Networks, GAN) для аугментации данных – достаточно новый способ повышения качества обучения, представленный в 2014 году [38]. Начиная с 2017 года количество исследований стало заметно расти.

Известны несколько успешных исследований, включающих использование сопоставительных генерирующих сетей. Авторам статьи [160] удалось добиться повышения точности классификации до 13 п.п. на наборах данных Omniglot, до 12 п.п. на наборе данных VGG Face и до 2 п.п. на наборе данных EMNIST. В статье [133] оценивается эффективность различных методов аугментации. Авторы заключают, что GAN и другие варианты аугментации с применением нейронных сетей работают ненамного лучше традиционных аугментаций, но потребляют почти в 3 раза больше вычислительного времени, а ключевой потенциал заложен в возможности комбинации нейросетевых и основных методов аугментации данных. Сочетание традиционной аугментации с последующей нейронной аугментацией еще больше повышает эффективность классификации. Авторы [161] представили новую систему RenderGAN, которая может генерировать большое количество реалистичных размеченных изображений путем объединения 3D-модели объекта и сопоставительной

генерирующей сети. Дополнения, применяемые к изображению (освещение, фон и прочие детали) генерируются отдельной сетью, обученной на основе немаркированных данных таким образом, чтобы создаваемые изображения были наиболее реалистичными, сохраняя при этом метки класса, известные из исходной 3D-модели.

Также известны исследования, направленные на балансировку классов с применением состязательных генерирующих сетей. Авторы [162] представили структуру, использующую модель СНС в качестве классификатора и циклически согласованные состязательные сети (CycleGAN) в качестве генератора обучающего набора данных (изначальные наборы данных для классификации эмоций несбалансированы). Авторы заявляют 5% – 10% увеличение точности классификации после применения методов дополнения данных на основе GAN.

Важной проблемой применения генеративно-состязательных сетей для обучения является значительное усложнение процесса обучения, необходимость предварительного обучения генератора, а также потенциальные проблемы с репрезентативностью данных, представленных аппаратом генерации данных [163].

#### **1.4 Естественные состязательные примеры**

Одной из проблем устойчивости обученных нейронных сетей является несоответствие реальных данных тем, на которых эта сеть обучалась. Это может быть связано с отличающимися свойствами источников для обучающих данных и реальных (например, различающееся разрешение или апертура объектива). Эти виды искажений принято называть естественными состязательными примерами. Одной из первых работ, в которой были рассмотрены естественные состязательные примеры, является [164]. На основе набора данных ImageNet, включающего десятки миллионов изображений, авторы создали свои наборы данных (ImageNet-A и ImageNet-O), содержащие изображения, которые наименее качественно классифицируются с помощью современных моделей машинного

обучения. При этом изображения, включаемые авторами в эти наборы, содержат ограниченное число ложных признаков (рисунок 1.16).



Рисунок 1.16 – Примеры естественных состязательных изображений из наборов данных ImageNet-A. Черным показан действительный класс изображения, красным – результат распознавания с помощью сети ResNet-50

Модель нейронной сети DenseNet-121 на наборе данных ImageNet-A достигает точности распознавания около 2% (что на 90 п. п. меньше точности распознавания набора данных ImageNet той же сетью). На рисунке 1.17 показаны результаты экспериментов по распознаванию изображений из набора ImageNet-A с применением современных архитектур свёрточных сетей [164].

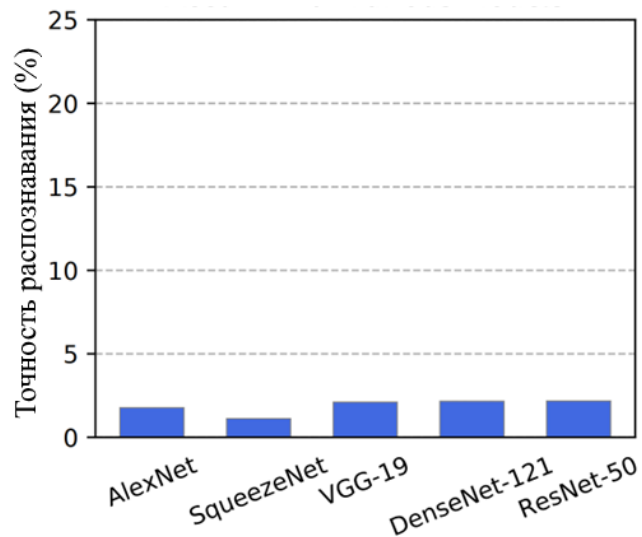


Рисунок 1.17 – Результаты экспериментов по распознаванию изображений из набора ImageNet-A с применением современных архитектур свёрточных сетей [164]

Из рисунка 1.17 видно, что все классификаторы различных архитектур, используемые для ImageNet, не обобщают данные в наборе ImageNet-A. Общая точность распознавания данных из набора на современных архитектурах нейронных сетей не превышает 2,2%. Авторы показали, что методы аугментации, включающие основные преобразования, данных практически не повышают производительность СНС, а использование других публичных наборов данных для обучения дает ограниченное улучшение.

Вместе с тем способность СНС к выявлению неявных признаков приводит к тому, что незначительное изменение информации приводит к непредсказуемым изменениям работоспособности всей системы [152]. Более того, изучив особенности определённого экземпляра нейронной сети и используя состязательные атаки, возможно «заставить» сеть ошибочно классифицировать данные таким образом, как задумано злоумышленником [165]. Существуют исследования, ориентированные на обеспечение устойчивости классификаторов на основе СНС от атак, основанных на изменении входных данных, однако большинство существующих методов специфично для той атаки, которая рассматривалась в исследовании.

Также известны исследования, ориентированные на разработку более широкого класса методов борьбы с искажениями и шумами в изображениях, распознаваемых с применением СНС. В части этих методов используются различные сложные алгоритмы (denoiser), т.е. предобработка изображений, использование состязательных сетей и обучение с зашумленными данными. Значительная часть разработанных систем предварительной обработки изображений специфична к определённым видам искажений, поэтому достаточно быстро преодолевается в новых алгоритмах состязательных искажений [166], [167].

### **1.5 Оценка помехоустойчивости нейронных сетей**

Значимой проблемой в задачах распознавания с применением СНС является наличие деградаций, искажений и шумов в изображениях. Впервые оценка робастности (англ. Robustness – «устойчивость») распознавания изображений с использованием машин опорных векторов наименьших квадратов (англ. LS-SVM – “Least-squares support-vector machine”) была исследована в [168]. Исследование робастности было проведено применительно к проверке подписи и в результате были получены графики зависимости вероятности ложной тревоги от вероятности промаха на готовых наборах данных подписей. Известны исследования, ориентированные на распознавание искаженных изображений. Распознавание изображений рукописных цифр при наличии в распознаваемых изображениях шумов также исследовалось в [169]. В качестве классификатора использовалась сеть RCN – «Reservoir Computing Network», обученная на изображениях без искажений. В обеих работах не исследовалось влияние неопределенности обучающего и тестового наборов данных на результаты распознавания.

В статье [170] авторы проанализировали влияние таких искажений, возникающих в реальных задачах, как размытие в движении, шумы, артефакты сжатия, цветовые искажения на производительность СНС по распознаванию лиц, используя закрытый набор данных LFW (англ. Labeled Faces in the Wild).

Множество статей описывают разработку денойзеров (от англ. «denoise» – «шумоподавление») [171], [172], [173], однако большинство денойзеров могут быть эффективны для улучшения качества изображения с точки зрения субъективного восприятия, но не повышают точность распознавания. Также денойзеры специфичны для решаемой задачи, и часто невозможно достоверно априорно оценить качество обработки изображения денойзером.

Известны исследования способов обнаружения искажений на изображениях [174] и избавления от такого рода искажений [175]. В статье [68] авторы указывают на наименьшую устойчивость различных архитектур СНС (VGG16, GoogleNet, VGG-CNN-S) перед такими искажениями, как шумы и размытия, и указывают, что обучение зашумленными или размытыми данными решением проблемы не является, т.к. при обучении такими данными снижается точность распознавания неискаженных изображений. Комплексной оценки влияния параметров обучающих данных на качество распознавания авторы [68] не проводят. Несмотря на расширяющееся применение СНС в задачах классификации изображений, влияние деградации изображений изучено недостаточно полно, и унифицированная количественная оценка помехоустойчивости отсутствует.

## **1.6 Постановка и формализация научной задачи исследования**

На данный момент большинство исследований, посвященных практическим применениям СНС, рассматривают разработку новых наборов данных, применение различных архитектур, модификации этих архитектур и методы предварительной обработки обучающих данных. Возможность применения методов аугментации в исследованиях до сих пор учитывается редко. При этом исследования не рассматривают проблему низкой устойчивости СНС к искажениям различного рода, а также способов снижения влияния этих искажений на качество распознавания изображений. Игнорирование таких проблем при практическом применении СНС может вызвать полную

неэффективность работы системы распознавания, например, при воздействии злоумышленника или в присутствии искажений рода сдвига предметной области.

Для решения такого рода проблем необходимо:

1. Проведение количественной оценки влияния искажений в данных на качество распознавания изображений свёрточной нейронной сетью;
2. Разработка способа нахождения оптимальных характеристик обучающих данных (с точки зрения помехоустойчивости обучаемой системы распознавания);
3. Разработка способа предварительной обработки классифицируемых изображений для повышения точности классификации искаженных данных.

Таким образом, научная задача данного диссертационного исследования состоит в выборе оптимального метода аугментации *обучающих* изображений, нахождении оптимального значения неопределённости в *обучающих* изображениях, а также поиске оптимального метода предварительной обработки, позволяющего достичь максимальной интегральной точности распознавания *тестовых* изображений с различными искажениями при заданном пороге минимальной точности распознавания.

В работе не рассматривается:

- 1) реализация методов оптимизации обучения и предварительной обработки данных применительно к системам локализации объектов с использованием СНС;
- 2) разработка новых архитектур искусственных нейронных сетей;
- 3) оптимизация гиперпараметров обучения;
- 4) аппаратная реализация системы распознавания изображений на основе СНС;
- 5) сложные системы предварительной обработки изображений, такие как GAN или автоэнкодеры.

Математическая постановка задачи диссертационного исследования:

1) Нахождение оптимального значения неопределённости в обучающих изображениях  $U_{TRopt}$ :

$$U_{TRopt} = \arg \left( \max \left( \sum_{U_{TS}=0}^{U_{TSmax}} P(U_{TS}, U_{TR}) \right) \right), \quad (1)$$

где  $U_{TR}$  – значение неопределённости в обучающем наборе данных,  $U_{TS}$  – значение неопределённости в тестовом наборе данных.

Величина неопределенности в обучающем наборе данных  $U_{TR}$  существенно влияет на точность распознавания и зависимость точности распознавания от неопределенности в тестовом наборе данных  $U_{TS}$ . Предположено существование оптимальной (с точки зрения точности распознавания) неопределённости  $U_{TRopt}$  в обучающем наборе изображений (для нейронных сетей, распознающих изображения с неизвестной заранее неопределенностью). Оптимальным значением неопределённости в обучающих изображениях  $U_{TRopt}$  является значение, позволяющее получить максимальное значение интегральной точности классификации при сохранении монотонности функции зависимости точности классификации от неопределённости в тестовых данных (1).

2) Выбор оптимального метода аугментации обучающих изображений, состоящего в выборе оптимального закона распределения значений неопределённости в обучающем наборе данных:

$$p(U_{TR})_{opt} = \arg \left( \max \left( \sum_{U_{TS}=0}^{U_{TSmax}} P(U_{TS}, p(U_{TR})) \right) \right), \quad (2)$$

где  $p(U_{TR})$  – закон распределения значений неопределённости в обучающем наборе данных.

Способ аугментации обучающего набора изображений существенно влияет на точность распознавания изображений с различной неопределённостью.



Предположено существование оптимального метода аугментации обучающих изображений, максимизирующего интегральную точность распознавания изображений с различными значениями неопределённости. Оптимальным методом аугментации обучающих изображений является метод, позволяющий получить максимальное значение интегральной точности классификации при сохранении монотонности функции зависимости точности классификации от значения неопределённости в тестовых данных  $p(U_{TS})$  (2).

3) Нахождение оптимального метода предварительной обработки тестовых данных:

$$F_{TSopt} = \arg \left( \max \left( \sum_{U_{TS}=0}^{U_{TSmax}} P(F_{TS}, U_{TS}) \right) \right), \quad (3)$$

где  $F_{TS}$  – функция обработки тестовых данных.

Предположено существование оптимального метода предварительной обработки тестовых данных, максимизирующего интегральную точность распознавания изображений с различной неопределённостью (интенсивностью состязательного искажения Fast Gradient Sign Method). Оптимальным методом предварительной обработки тестовых изображений является метод, позволяющий получить максимальное значение интегральной точности классификации при различной интенсивности состязательного искажения (3).

Нахождение  $U_{TROpt}$ ,  $p(U_{TR})_{opt}$ ,  $F_{TSopt}$  проводилось в ходе диссертационного исследования методом статистического моделирования. Результаты моделирования приведены в разделах 2 – 5 диссертации.

## Выводы по разделу 1

1. Недостаточность, неполнота или несоответствие тестовым доступных для обучения СНС данных зачастую обуславливает снижение качества работы системы распознавания. В практических условиях сбор широкого (достаточного, полного) набора обучающих данных затруднителен ввиду высокой требовательности к временным ресурсам, необходимости анализа полноты признаков, а также соответствия неявных параметров обучающих данных тестовым.

2. Для решения вышеописанных проблем важным методом является аугментация данных. К настоящему времени исследователями успешно применено множество различных способов преобразования изображений для аугментации, при этом используемые способы часто специфичны для решаемой исследователями практической задачи.

3. Существуют исследования, ориентированные на обеспечение устойчивости классификаторов на основе СНС от атак, основанных на изменении входных данных, однако большинство существующих методов специфично для той атаки, которая рассматривалась в исследовании, соответственно, достаточно быстро преодолевается в новых алгоритмах состязательных искажений.

4. Несмотря на расширяющееся применение СНС в задачах классификации изображений, влияние деградации изображений изучено недостаточно полно, и унифицированная количественная оценка помехоустойчивости отсутствует.

5. Необходимо разработать метод оптимальной аугментации обучающих изображений, разработать метод нахождения оптимального значения неопределённости в обучающих изображениях, а также разработать метод предварительной обработки изображений, позволяющий достичь максимальной интегральной точности распознавания тестовых изображений с различными искажениями при заданном пороге минимальной точности распознавания.

## 2. ИССЛЕДОВАНИЕ ПОМЕХОУСТОЙЧИВОСТИ СВЁРТОЧНОЙ НЕЙРОННОЙ СЕТИ

### 2.1 Метод оценки помехоустойчивости СНС

Наиболее важным аспектом в применении нейронных сетей является обучение, и успех обучения в основном зависит от правильного представления обучающих данных. Качество работы сложных и больших нейронных сетей, обученных на плохо представленных данных, на практике оказывается гораздо ниже качества работы простых сетей, обученных на правильных и хорошо организованных данных [176], [177]. Помимо того, сложные нейронные сети, состоящие из десятков миллионов нейронов, характеризуются высоким потреблением вычислительных ресурсов. Таким образом, Конструирование признаков (англ. Feature Engineering) - построение репрезентативной обучающей выборки и правильный процесс генерации признаков в обучающих данных - является наиболее важной задачей машинного обучения.

Корректное и продуманное конструирование признаков позволяет задействовать потенциал способности СНС к обобщению, который следует (и даже необходимо) использовать для снижения влияния шума и неопределённостей, вносимых в распознаваемые данные, и, соответственно, противостояния искажениям. Устойчивая к искажениям в распознаваемых данных СНС использует информацию из своих слоев, сформированных при обучении таким образом, что влияние этих искажений на распознаваемые признаки достаточно мало. Например, для борьбы с высокочастотным шумом СНС должна использовать признаки, выявленные в глубоких свёрточных слоях, где входные данные усредняются по большим областям исходного изображения. В ходе диссертационной работы предположено и методом статистического моделирования доказано, что обучение СНС на данных с внесёнными искажениями может помочь улучшить обобщающие способности сети и

возможность сети противостоять этим искажениям, тем самым снизив влияние искажений на аппараты распознавания изображений.

Для СНС общая структура системы контролируемого обучения и распознавания выглядит так, как показано на рисунке 2.1.



Рисунок 2.1 – Общая структура системы обучения и распознавания с учителем

Использование реального цифрового изображения для проведения оценки устойчивости свёрточной нейронной сети к искажениям на практике невозможно, поскольку нет возможности оценить долю полезной информации в этих изображениях (отношение сигнал/шум на изображениях). Для проведения такой оценки требуется разработка модели контролируемой генерации обучающих и тестовых изображений (контролируемое конструирование признаков). В данной работе для исследования поведения нейронной сети без потери общности была использована модель генерации изображений с низкой плотностью точек. Изображения, получаемые математической моделью генерации, являются, помимо прочего, наглядными: количество шума, представленное как неопределенность расположения точек на изображении, можно оценить визуально. Изображения с низкой плотностью точек (рисунок 2.2) удобны для внесения искажений в формы объектов.

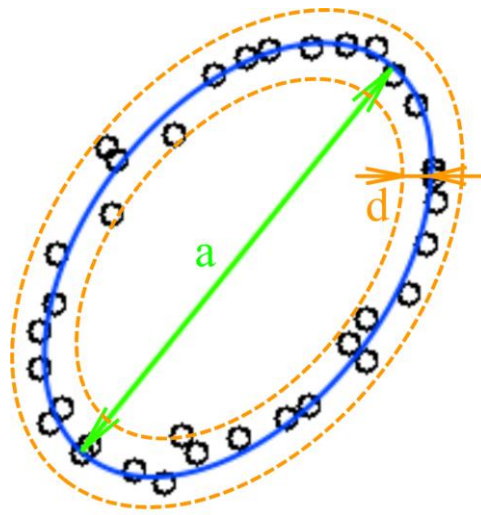


Рисунок 2.2 – Пример изображения с низкой плотностью точек

На рисунке 2.2 показан пример изображения с низкой плотностью точек. Идеальная фигура представляет собой эллипс. Видны отклонения расположения точек от идеального эллипса. Эти отклонения следует рассматривать как шум или неопределенность. В данном разделе анализируется точность распознавания изображений с низкой плотностью точек в зависимости от величины неопределенности расположения точек (далее — неопределенность) в тестовом (в разделе 3 – в обучающем) наборе данных.

Метод исследования заключается в генерации наборов данных с псевдослучайными изображениями различных классов. Генерируемые наборы данных содержат большое число изображений с низкой плотностью точек и отличаются различной неопределенностью. Далее проводится анализ точности распознавания этих наборов данных свёрточной нейронной сетью, обученной на изображениях без внесённых искажений.

Параметр, определяющий меру неопределенности, может быть описан как

$$U = \frac{d}{a}, \quad (4)$$

где  $d$  - дисперсия,  $a$  - линейный размер фигуры (рисунок 2.2). Далее в работе неопределенность обучающего набора данных будет обозначаться как  $U_{TR}$ , а неопределенность тестового набора данных как  $U_{TS}$ . В других задачах, таких как

распознавание зашумленных изображений, показанных на рисунке 2.2,  $U$  можно описать как

$$U = \frac{I_{noise}}{I_{info}}, \quad (5)$$

где  $I_{noise}$  - средняя интенсивность шума, а  $I_{info}$  - средняя интенсивность значимой части распознаваемого изображения [28]. В обоих случаях неопределенность  $U$  описывает отношение динамического диапазона (интенсивности) шумовой составляющей к динамическому диапазону информативной составляющей в изображении. Добавление шума любой природы и интенсивности может резко снизить точность распознавания изображений нейронной сетью, обученной на идеальном наборе данных, поэтому необходимо провести исследование робастности, чтобы избежать этого эффекта путем изменения свойств обучающего набора данных.

Задачи распознавания изображений с низкой плотностью точек исследовались в серии работ автора данного диссертационного исследования, посвященных оценке и предсказанию поведения групп абонентов сети мобильной связи и сложных кластеров путем анализа телетрафика и геолокационных данных [75], [178], [179]. Информацию о местоположениях абонентов в группах можно представить в виде изображений с низкой плотностью точек. СНС показали свою эффективность в решении этой задачи [88], однако анализ и обоснование устойчивости применяемого метода к неопределенности исходных данных не проводились. В работе [75] была описана реализованная математическая модель, генерирующая формы кластеров абонентов, а также показано, что типичные формы кластеров, представляющие изображения с низкой плотностью точек, могут быть автоматически классифицированы с применением СНС.

Исчерпывающие характеристики устойчивости СНС пока не получены, но есть все основания предполагать, что различные значения неопределенности в наборах данных могут существенно влиять на качество распознавания этих форм. Таким образом, оценка и оптимизация робастности СНС при решении задач

распознавания требуют специального статистического моделирования. Эта задача представляет большой теоретический и практический интерес, поскольку результаты данной работы могут быть в дальнейшем применены во всех областях машинного обучения. В основе данной работы описана в значительной степени упрощенная модель, что позволяет обобщить ее выводы для большинства случаев, решаемых нейронными сетями и другими моделями машинного обучения.

## 2.2 План исследования

Как уже говорилось ранее, оптимизация исключительно архитектуры нейронной сети без учета влияния характеристик обучающего набора данных, как правило, не дает исчерпывающего результата и позволяет улучшить поведение системы лишь в некоторых случаях. Хотя углубление нейронной сети (в общем случае, но не применимо к распознаванию изображений с низкой плотностью точек из-за простоты изображений, что будет показано далее) часто приводит к улучшению способности этой сети обнаруживать и обобщать скрытые признаки [88], [180], [181], но также создает множество проблем, таких как увеличение потребления вычислительных ресурсов сетью, проблема «исчезающих» или «взрывающихся» градиентов [182] и др. Оптимизация должна проводиться с учетом свойств всех компонентов системы "обучающий набор данных - модуль обработки - тестовый набор данных" (рисунок 2.1).

Широко распространенный традиционный подход предполагает получение фиксированной точности распознавания тестового набора данных сетью, обученной на фиксированном обучающем наборе данных с заданной неопределенностью  $U_0$ . Эта точность может быть описана одним числом, скаляром  $P_0$ . Точность  $P_0$  описывается следующим образом:

$$P_0 = \frac{M_{correct}}{M_{total}}, \quad (6)$$

где  $M_{correct}$  – количество правильно распознанных элементов в тестовом наборе данных, а  $M_{total}$  – общее количество элементов в тестовом наборе данных.

Этот скалярный подход позволяет оценить только локальные свойства системы обучения-распознавания, но не позволяет оценить поведение этой системы при различных значениях неопределенности в данных. В разделах 2 и 3 предлагается более глубокий векторно-матричный подход для оценки устойчивости и робастности сети, который включает следующие последовательные шаги:

1. Получение массива точности распознавания тестовых наборов данных  $P$  при различных неопределенностях тестовых наборов данных  $U_{TS}$  с фиксированной неопределенностью обучающего набора данных  $U_{TR}$  – вектор  $P(U_{TS})$  (раздел 2 диссертации).

2. Получение двумерного массива точностей распознавания  $P$  наборов данных в зависимости от неопределенностей тестового ( $U_{TS}$ ) и обучающего набора данных  $U_{TR}$  - матрица  $P(U_{TR}; U_{TS})$  (раздел 3 диссертации).

Таким образом, на каждом следующем шаге происходит увеличение информативности относительно оценки робастности и оптимальности системы обучения-распознавания. Имея известные  $P(U_{TR}; U_{TS})$  возможно получить любое  $P(U_{TS})$  и  $P_0$ :

$$P(U_{TS}) = \frac{1}{N_{TR}} \cdot \sum_{U_{TR}} P(U_{TR}; U_{TS}); \quad (7)$$

$$P_0 = \frac{1}{N_{TS}} \cdot \sum_{U_{TS}} P(U_{TS}), \quad (8)$$

где  $N_{TR}$  – количество наборов данных с различными значениями неопределенности для обучения  $U_{TR}$ ,  $N_{TS}$  – количество наборов данных с различными значениями неопределенности тестовых данных  $U_{TS}$ .



### 2.3 Модель формирования и искажения изображений с низкой плотностью точек

Для оценки внешних характеристик системы обучения-распознавания была выбрана удобная модель экспериментов с изображениями. Математическая модель, описанная в работе [75], позволяет автоматически генерировать наборы данных, используемые для обучения и тестирования СНС, а также задавать различные параметры неопределенности (например, дисперсия смещения положения точек относительно векторной модели фигуры (рисунок 2.3). Это позволяет оценить устойчивость СНС к изменению параметров неопределенности набора данных и оценить характеристики нейронной сети в условиях действия факторов, увеличивающих интенсивность искажения во входных данных.

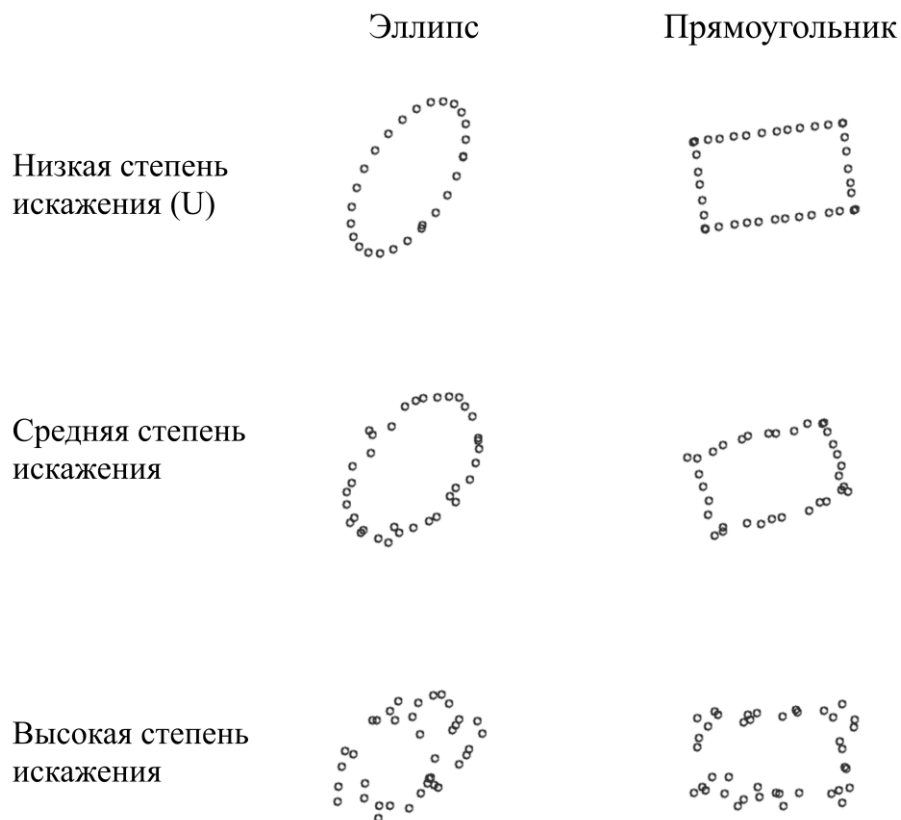


Рисунок 2.3 – Модель изображений с низкой плотностью точек с различными значениями неопределенности

Для оценки характеристик помехоустойчивости обученной нейронной сети в работе сгенерировано 250 наборов данных с различными значениями неопределенности. Примеры изображений с различной интенсивностью искажений, входящие в состав наборов данных, приведены на рисунке 2.4.

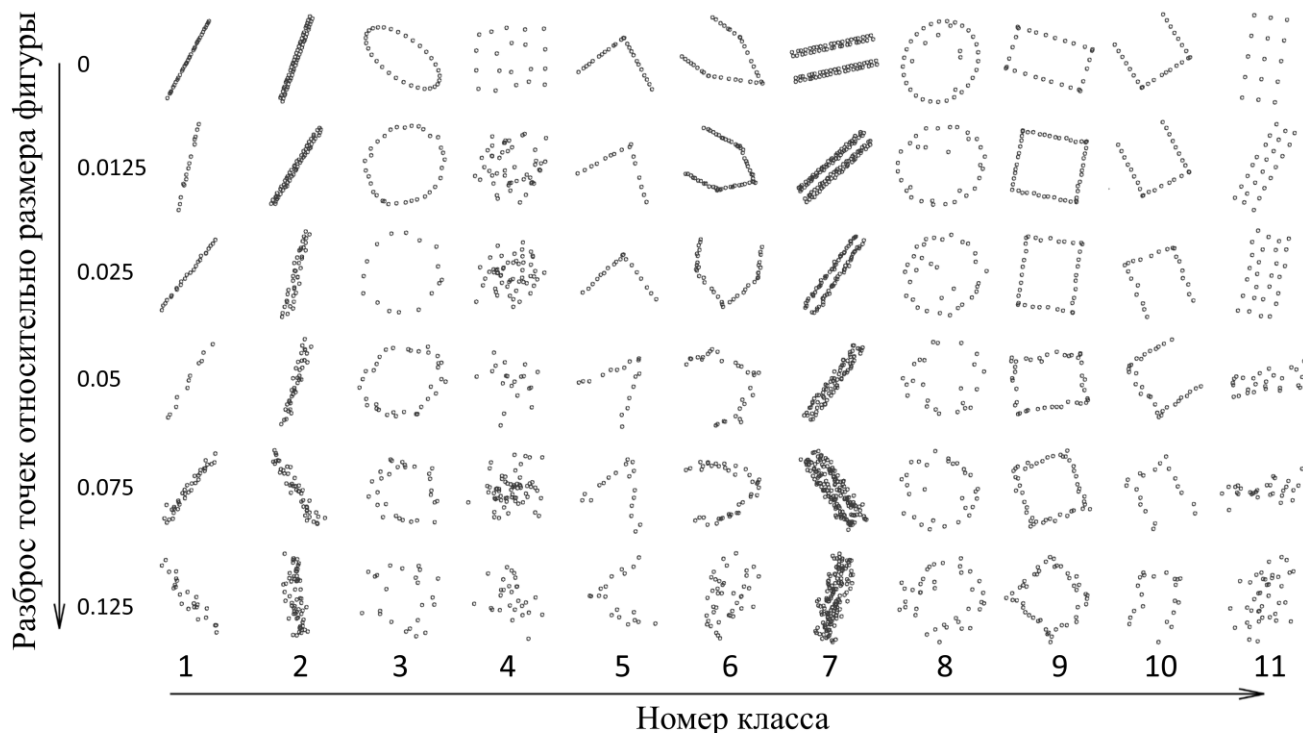


Рисунок 2.4 – Примеры изображений с различной интенсивностью искажений

Как видно из рисунка 2.4, неопределенность координат отдельных точек искажает изображение, но характерные черты фигур сохраняются.

На рисунке 2.5 показан используемый в данной работе способ реализации смещения — по мере увеличения значения неопределенности изображения искажаются сильнее. Результат моделирования изображений показан на рисунке 2.4. Результатом генерации является набор данных, включающий изображения с разрешением  $256 \times 256$  пикселей. Выбранное разрешение достаточно для обеспечения необходимой точности представления искаженных изображений без ущерба для скорости обработки этих изображений и требовательности СНС к вычислительным ресурсам, а также для размера массива данных. Полученная модель генерации и искажения изображений может быть описана следующим образом (рисунок 2.5):

1. Создание векторной модели фигуры с равномерно распределенными точками (количество точек является случайным и выбирается в некотором диапазоне).

2. Добавление к координатам каждой точки индивидуального смещения, описываемого гауссовским случайным распределением (дисперсия распределения задает величину неопределенности). Гауссовское случайное распределение хорошо подходит для описания неопределенности, возникающей по ряду различных причин физической природы. Неопределенность измеряется относительно модуля вектора  $a$ , определяющего линейный размер фигуры. Например, значение неопределенности 0,1 относительных единиц говорит о том, что  $\sigma$  положения точек определяется как 0,1 от максимального линейного размера фигуры.

3. Полученная фигура поворачивается на произвольный угол.

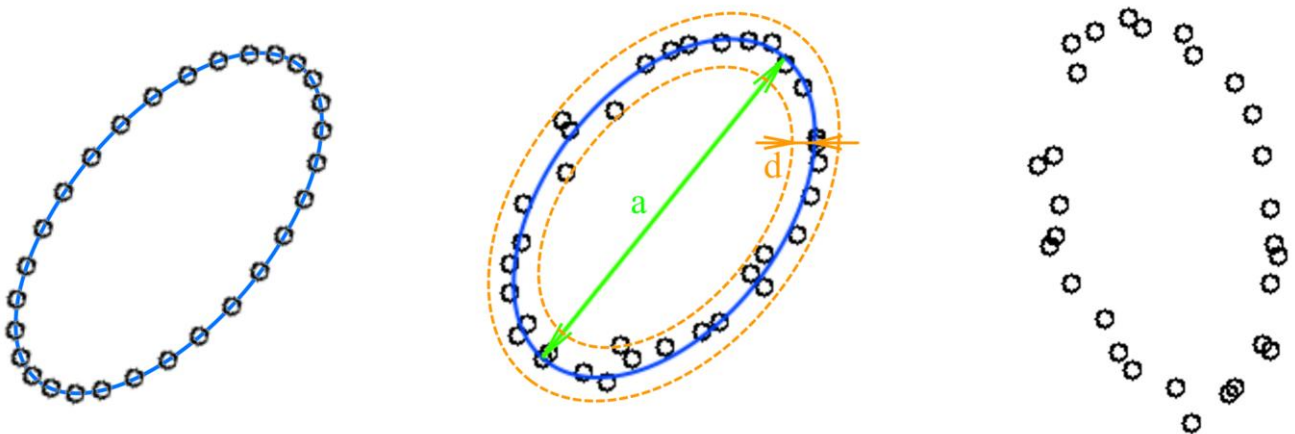


Рисунок 2.5 – Модель генерации изображений

Изображения, полученные с помощью модуля генерации изображений, объединяются в единый набор данных для создания обучающего набора данных. Преимуществом данной модели генерации изображений является простота их интерпретации; результаты исследований могут быть обобщены на более широкий класс задач.

## 2.4 Структура нейронной сети

Свёрточная нейронная сеть – алгоритм глубокого обучения, благодаря своей структуре являющийся эффективным для классификации изображений (рисунок 2.6). Свёрточная структура позволяет оценивать значимость различных объектов (признаков) в изображении. Предварительная обработка данных, требуемая для работы СНС, значительно менее затратна по времени и вычислительным ресурсам по сравнению с другими алгоритмами классификации. В то время как в методах классификации без использования обучения ключевые признаки могут быть заданы вручную, при достаточном обучении СНС имеют возможность «изучать» эти характеристики автоматически. Каждый экземпляр данных (применительно к СНС в данной работе – изображения) представляет из себя набор (вектор) признаков. В обучении с учителем для каждого из экземпляров задана метка класса. Задача нейронной сети при обучении – вычисление с использованием определённого набора пар «вектор признаков – метка класса» классификационной функции. Количественную оценку различия полученной и целевой функций при обучении называют ошибкой обучения.

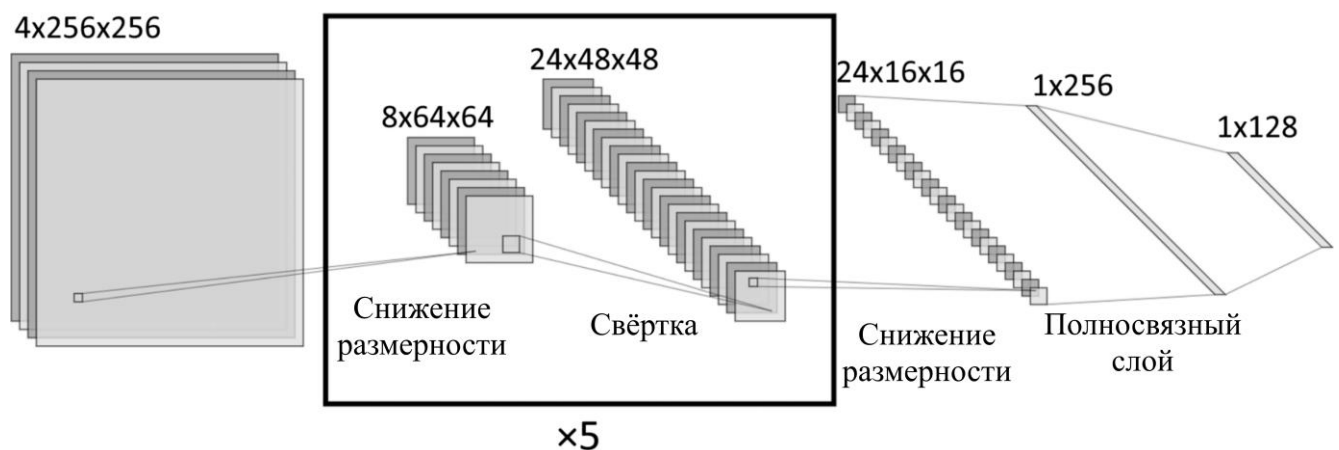


Рисунок 2.6 – Структура СНС

Используемая в данной работе архитектура является одной из самых простых и эффективных для относительно простых наборов данных [91], что

позволяет провести обобщенные эксперименты и в дальнейшем экстраполировать результаты на более широкий класс архитектур. Сверточная сеть состоит из чередующихся слоёв свертки и подвыборки (рисунок 2.6), а процесс обучения заключается в многократном предъявлении сети обучающего набора данных (каждая итерация называется эпохой) и коррекции синаптических весов сети на каждой итерации. Когда синаптические веса стабилизируются, а средняя ошибка на всем обучающем множестве минимизируется, сеть можно считать обученной [91].

## 2.5 Оценка зависимости качества распознавания от величины неопределенности в тестовых наборах данных

В настоящем исследовании с использованием нейронной сети, обученной на данных без внесённых искажений, было проведено распознавание каждой из множества сгенерированных тестовых выборок с разной мерой неопределённости  $U_{TS}$ , что позволило выявить зависимость точности распознавания от меры неопределённости. В результате получен массив со множеством значений точности распознавания в зависимости от неопределённости в тестовых данных  $P(U_{TS})$ . Результаты анализа сведены в график, представленный на рисунке 2.7.

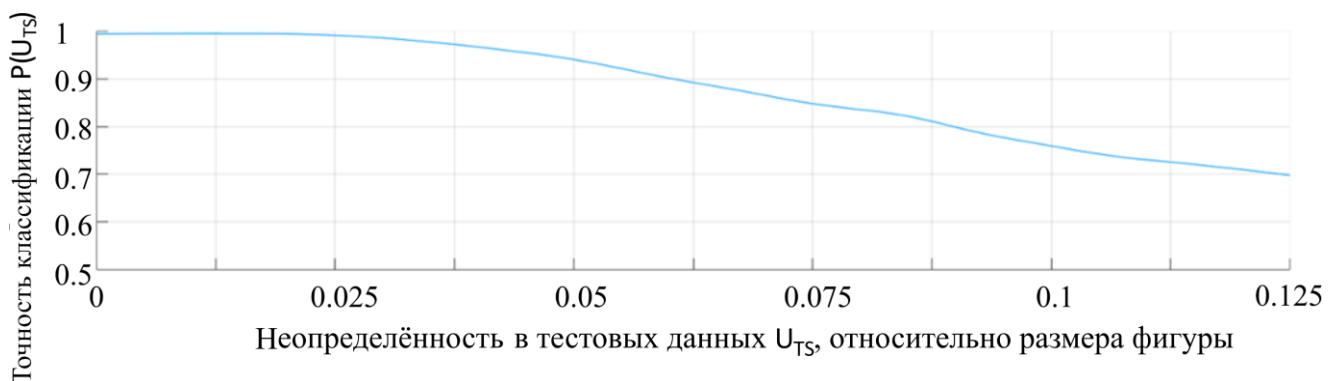


Рисунок 2.7 – График зависимости точности распознавания изображений от неопределённости в тестовых наборах данных при использовании СНС, обученной на наборах данных без искажений

Как видно из графика, наибольшая точность распознавания достигается при значении неопределённости в тестовых данных (среднеквадратического отклонения) от 0 до 0,025. Точность распознавания в этом диапазоне достигает 0,99, а при увеличении неопределённости в тестовых данных монотонно убывает, что указывает на корректность работы модели.

Качество работы сети (точность распознавания) на различных данных зависит от степени сходства тестовых данных с данными, полученными сетью при обучении. Соответствие характеристик и признаков обучающих данных тестовым – репрезентативность обучающей выборки – важнейшее условие корректного обучения нейронных сетей [183]. Различия классификационных функций обучающего и тестового наборов данных проявляется при оценке значения ошибки классификации и, соответственно, в снижении точности классификации (в снижении уверенности ответа сети). Для подтверждения данного тезиса проведено исследование помехоустойчивости сети, обученной на наборах данных с искажениями.

## **2.6 Исследование помехоустойчивости сети, обученной на наборах данных с искажениями**

При внесении неопределённости в положения точек в обучающей выборке характеристики качества распознавания тестовых выборок значительно меняется. Для получения дополнительной информации были обучены две независимые СНС с идентичной структурой на двух обучающих наборах данных с двумя различными значениями неопределённости:  $U_{TR} = 0$  и  $U_{TR}$ , выбранное случайным образом для каждого изображения в диапазоне от 0 до 0,025. Гиперпараметры нейронной сети оставались неизменными для проведения анализа именно влияния обучающей выборки. Для каждого эксперимента были созданы отдельные наборы данных с произвольными последовательностями изображений (что важно для обеспечения стохастичности поиска целевой классификационной функции). Правила создания наборов данных описаны в подразделе 2.3 и в [75].

Изначальные веса СНС задавались случайно в процессе выполнения эксперимента. Обученные нейронные сети использовались для распознавания отдельно сгенерированных наборов данных, содержащих изображения с различными неопределенностями  $U_{TS}$ . Все вероятности распознавания  $P$ , полученные в сериях экспериментов, усреднялись по всем сериям экспериментов с фиксированными  $U_{TS}$ .

В результате были получены два массива значений точности распознавания от неопределенности данных тестового набора  $P(U_{TS})$ . Результаты этого эксперимента сведены в график, представленный на рисунке 2.8.

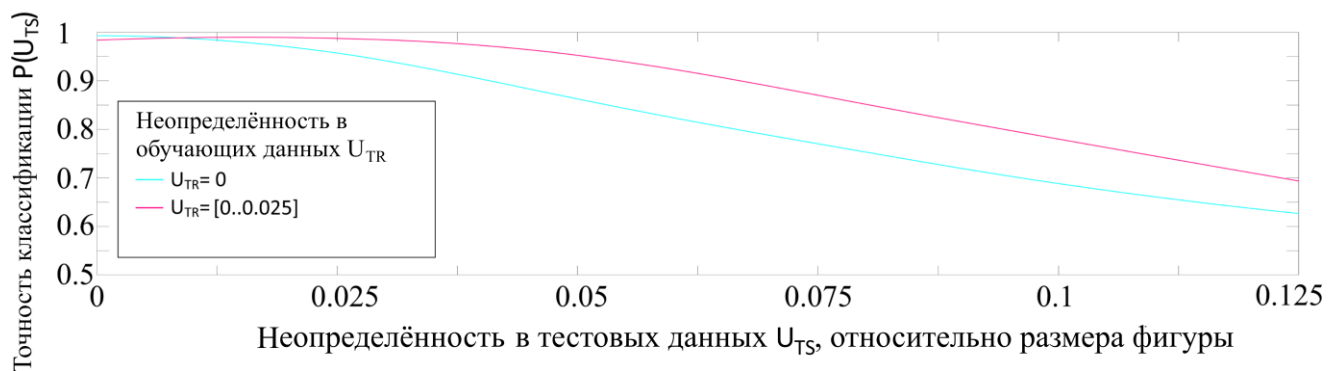


Рисунок 2.8 – Графики зависимости точности распознавания изображений от неопределённости в тестовых наборах данных, полученные сетями, обученными на наборах данных со значениями неопределённости, равной 0 и задаваемой случайно для каждого экземпляра в диапазоне от 0 до 0,025 от размера фигуры

Сравнение двух графиков на рисунке 2.8 позволяет сделать три основных вывода:

1. При  $U_{TR} = 0$  кривая точности имеет монотонный характер. Это подтверждает согласованность и устойчивость выбранной модели.

2. При  $U_{TR}$ , выбранном случайным образом для каждого изображения в диапазоне от 0 до 0,025, кривая точности немонотонна; она показывает небольшое падение точности (относительно графика точности  $U_{TR} = 0$ ) при значениях  $U_{TS}$  ниже 0,01. Это явление доказывает, что изменение пропорций

неопределенности в обучающем наборе данных может повлиять на качество распознавания идеальных изображений (без внесённых искажений).

3. Максимальная точность достигается при  $U_{TR} = 0$  и  $U_{TS} = 0$ , но интегральная (общая) точность при всех рассмотренных значениях  $U_{TS}$  достигается сетью, обученной при  $U_{TR}$ , выбранном случайным образом для каждого изображения в диапазоне от 0 до 0,025. Это явление можно объяснить ограниченной способностью нейронной сети, обученной только на идеальных изображениях (без искажений), к обобщению признаков, представленных в изображениях с искаженными данными.

Как говорилось выше, характеристики обучающего набора данных значительно влияют на характеристики обученной нейронной сети и на точность выполнения ими будущих задач распознавания на новых наборах данных [184]. Для оптимизации обучения СНС (в отношении поиска оптимальных параметров обучающего набора данных для улучшения качества распознавания изображений с различными значениями неопределенности) в данной работе проведены эксперименты, включающие обучение СНС на наборах данных с множеством различных значений неопределенности. Графиков, представленных на рисунках 2.7 и 2.8, недостаточно для формирования однозначных выводов об оптимальности обучения. В данном исследовании проведены дальнейшие эксперименты по обучению нейронных сетей, тем самым произведено "развёртывание" результатов в новом измерении (количество неопределенности в наборах данных для обучения). На рисунке 2.9 показаны зависимости точности распознавания от величины неопределенности  $U_{TS}$ , полученные сетями, обученными на наборах данных с  $U_{TR}$ , изменяющимся от 0 до 0,125 с шагом 0,025, показанные на одном графике для наглядности.



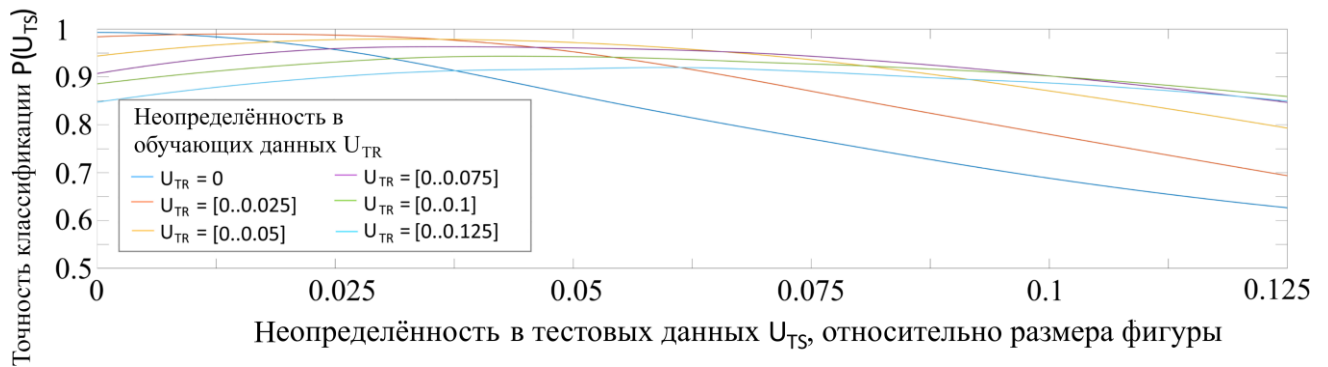


Рисунок 2.9 – Графики зависимости точности распознавания от неопределённости в тестовых наборах данных, полученные сетями, обученными на наборах данных различными значениями неопределённости

Из рисунка 2.9 можно заключить, что максимальная точность распознавания обученной нейронной сетью достигается при  $U_{TS} \approx U_{TR}$ . Более того, анализ графиков на рисунке 2.9 показывает, что при использовании сетей, обученных на данных с  $U_{TR} \geq 0,025$ , графики точности классификации меняют свою форму с монотонной на немонотонную, что указывает на неоптимальность обучения и, соответственно, ненадежность модели распознавания. Зависимость  $P(U_{TS})$  должна быть монотонной ( $dP/dU_{TS} \leq 0$ ), так как в наборах данных с более высоким значением неопределённости снижается доля значимой информации. Это правило можно использовать в качестве критерия правильности и робастности обучения.

## Выводы по разделу 2

1. Разработанная математическая модель, описанная в данном разделе, позволяет точно оценить характеристики устойчивости СНС к искажениям, и генерирует простые для понимания примеры для исследования поведения нейронной сети.

2. С использованием нейронной сети, обученной на сгенерированных наборах данных с разной интенсивностью внесённых искажений, проведено

распознавание каждой из множества сгенерированных тестовых выборок (с разной мерой неопределённости тестового набора), что позволило выявить зависимость точности распознавания от меры неопределённости в тестовом наборе данных.

3. При обучении сети на данных без внесённых искажений наибольшая точность распознавания достигается при значении неопределённости в тестовых данных (среднеквадратического отклонения) от 0 до 0,025. Точность распознавания в этом диапазоне достигает 0,99, а при увеличении неопределённости в тестовых данных монотонно убывает, что указывает на корректность работы модели.

4. При внесении избыточных искажений в обучающий набор данных кривая точности приобретает немонотонный характер; максимальная точность распознавания достигается при  $U_{TS} \approx U_{TR}$ , что указывает на неоптимальность обучения. При этом интегральная точность распознавания тестовых наборов данных растёт.

### 3. ИССЛЕДОВАНИЕ ВЛИЯНИЯ НЕОПРЕДЕЛЁННОСТИ В ОБУЧАЮЩИХ ДАННЫХ НА ПОМЕХОУСТОЙЧИВОСТЬ СВЁРТОЧНОЙ НЕЙРОННОЙ СЕТИ

#### 3.1 Зависимость точности распознавания от неопределённости в тестовых и обучающих данных ( $U_{TR}$ и $U_{TS}$ )

Результаты, полученные на предыдущем этапе (представленные в виде графиков на рисунке 2.9), привели к проведению комплексного исследования поведения СНС при изменении неопределенности данных в обучающем наборе  $U_{TR}$ . Для более детального анализа точности распознавания изображений было сгенерировано множество обучающих наборов данных с различной неопределенностью  $U_{TR\ i} = d_i/a$  (от 0 до 0,125 от линейного размера фигуры с шагом 0,0005). Отдельные копии СНС, показанной на рисунке 2.6, были обучены, и их веса были получены для каждого из обучающих наборов данных. Затем каждая обученная сеть распознавала каждый из тестовых наборов данных с различными неопределенностями  $U_{TS\ j} = d_j/a$ . Это позволило получить двумерный массив (матрицу) точности распознавания в зависимости от неопределенностей обучающего и тестового наборов данных  $P = P(U_{TR}; U_{TS})$ . Полученная матрица точности распознавания содержит полную информацию о внешних характеристиках системы обучения-распознавания и может быть использована для оценки согласованности и устойчивости системы (рисунок 3.1).

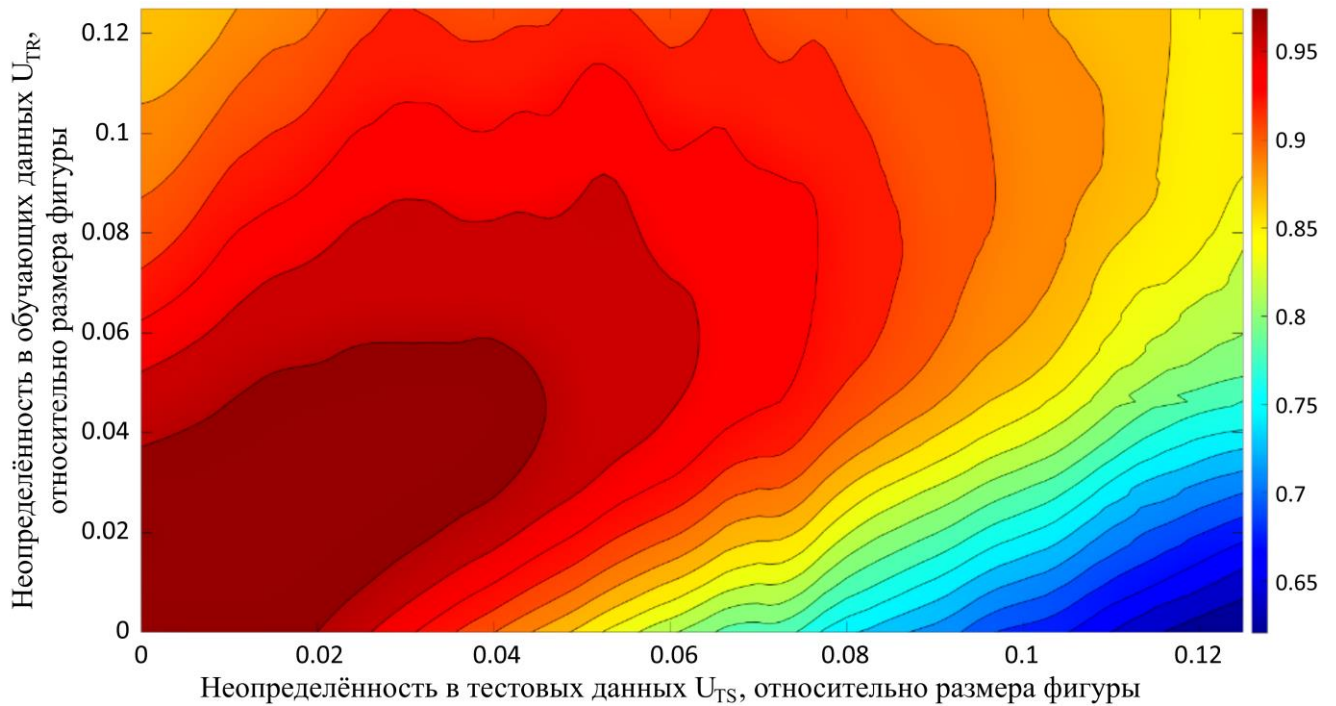


Рисунок 3.1 – График зависимости точности распознавания изображений от неопределённости в тестовых и обучающих данных  $U_{TR}$  и  $U_{TS}$

Как видно из рисунка 3.1, сеть лучше распознает данные с такой же или меньшей неопределенностью, чем та, которая использовалась в наборе данных для обучения. Этот факт подтверждает правильность работы системы обучения-тестирования, поскольку сеть лучше распознает данные со статистическими характеристиками, схожими с теми, которые использовались во время обучения (либо с большей долей значащих данных).

Точность распознавания, достигнутая сетью, обученной на наборе данных с высоким значением  $U_{TR}$ , заметно снижается на данных с низким значением  $U_{TS}$ ; это вызвано изменением доли значимых данных и шума в процессе обучения. Неравномерность, которую можно увидеть на графике, является следствием ограниченности наборов данных и должна рассматриваться как статистическая погрешность.

Данные, полученные в результате данного эксперимента, позволят произвести определение оптимальных параметров обучающих наборов данных для СНС с целью получения наилучших результатов точности распознавания

тестовых наборов данных. Рассмотрим способ определения оптимальных параметров обучения.

### **3.2 Интегральная точность распознавания изображений при различных пороговых значениях требуемой минимальной точности распознавания**

Поскольку большинство систем распознавания в практическом применении имеет свои минимальные требования к точности [66], необходимо проанализировать точность распознавания при различных значениях минимальных порогов точности распознавания.

В практических задачах, решаемых с помощью нейронных сетей, часто нет необходимости распознавать данные с чрезвычайно высокими искажениями. Более того, к разработанным решениям, использующим нейронные сети, часто предъявляются минимальные требования по точности классификации/распознавания. Часто на практике возникает необходимость получить достаточно высокую уверенность системы, что часто является более важным критерием качества системы. Для получения более практически значимых результатов в данной работе из полученных данных были выбраны области, включающие значения неопределенности тестового набора данных с точностью распознавания выше  $P_{thr}$ , (в которых точность распознавания выше выбранных пороговых значений), что позволило оценить допустимые значения неопределенности тестовых наборов данных для обеспечения необходимой точности распознавания. На рисунке 3.2 показана область, включающая значения неопределенности тестового набора данных  $U_{TS}$ , в которых обеспечена точность распознавания  $P_{thr}$  выше 90%.

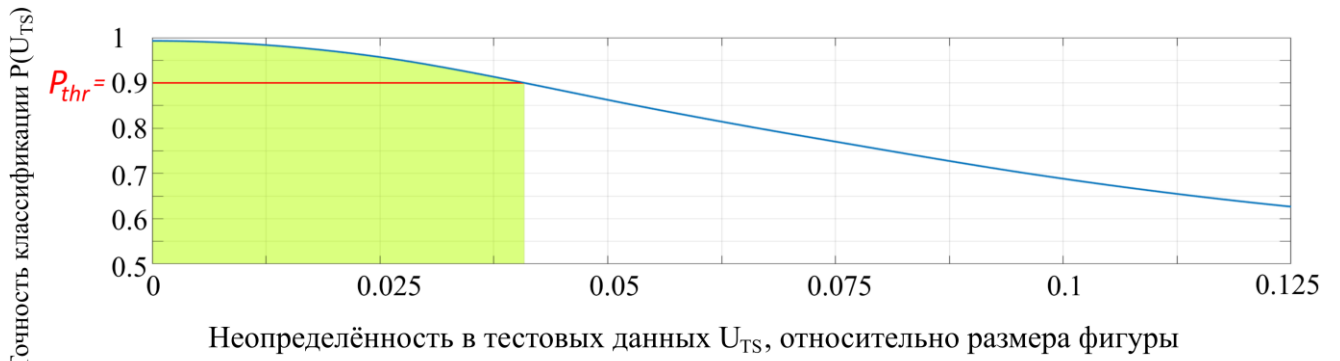


Рисунок 3.2 – Область, включающая значения неопределённости тестового набора данных с точностью распознавания выше 90%

Выделенная область на рисунке 3.2 рассчитывается как

$$Q(P_{thr}) = \sum_{U_{TS}=U_{TS}^{\min}, P \geq P_{thr}}^{U_{TS}=U_{TS}^{\max}, P \geq P_{thr}} P(U_{TS}). \quad (9)$$

Такая оценка позволяет рассчитать качество работы системы не только с точки зрения максимально возможной точности классификации, но и с точки зрения устойчивости системы к неопределённости, возникающей в тестовых данных.

После оценки области, включающей значения неопределённости тестового набора данных с точностью распознавания не ниже определённого порога  $P_{thr}$  возможно обоснование выбора оптимального значения неопределённости в обучающем наборе данных. Возникает задача определения методом статистического анализа оптимальных параметров обучающего набора данных для получения требуемой точности распознавания выше порога  $P_{thr}$ . Для каждой сети, обученной ранее на наборах данных с различными неопределённостями, были получены интегральные значения точности распознавания  $Q$  при различных пороговых значениях:

$$Q(U_{TR}; P_{thr}) = \sum_{U_{TS}=U_{TS}^{\min}, P \geq P_{thr}}^{U_{TS}=U_{TS}^{\max}, P \geq P_{thr}} P(U_{TR}; U_{TS}) \quad (10)$$

В данном случае  $Q$  — это интегральное значение точности классификации для всех тестовых наборов данных (со всеми значениями неопределенности), для которых точность распознавания превысила порог  $P \geq P_{thr}$ . Полученные данные обобщены на графике, представленном на рисунке 3.3 ниже.

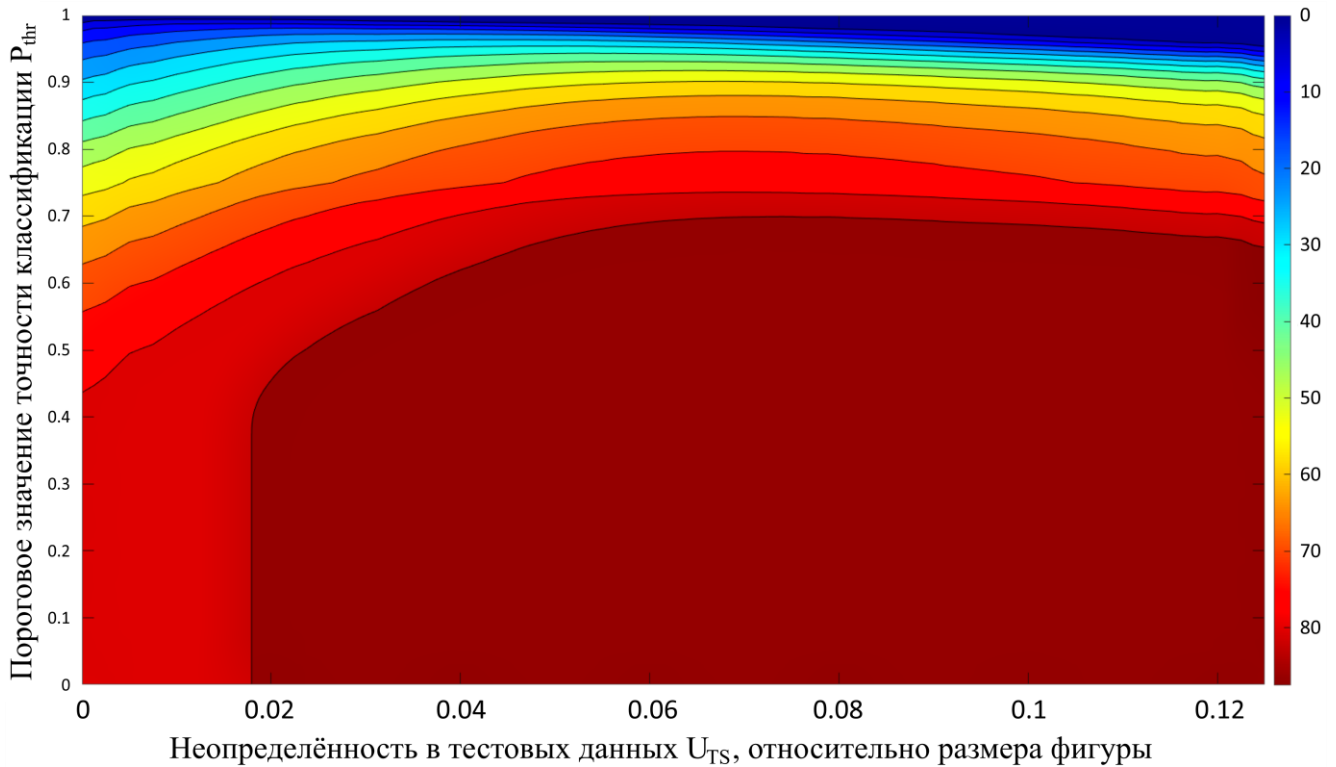


Рисунок 3.3 – График зависимости интегральной точности распознавания изображений  $Q$  для всех значений  $P > P_{thr}$  от  $P_{thr}$  и неопределённости в обучающих данных  $U_{TR}$

Данный график отображает общий показатель вероятности правильного распознавания для данных, дающих вероятность распознавания больше  $P_{thr}$ , в зависимости от  $P_{thr}$  и неопределенности в обучающем наборе данных  $U_{TR}$ . Результаты для тестовых данных, распознанных с точностью ниже требуемого  $P_{thr}$ , не учитывались. Цвета на графике показывают области с одинаковой средней точностью распознавания для всех  $P > P_{thr}$ . Из графика можно заключить, что всегда существует оптимальная неопределенность обучающего набора данных  $U_{TR}$ , зависящая от нижнего порога требуемой точности распознавания.

Используя рисунок 3.3, можно определить оптимальную неопределенность в обучающем наборе данных  $U_{TR}$ , необходимую для обучения сети и достижения максимальной интегральной точности распознавания  $Q$  для всех данных с локальной вероятностью распознавания, превышающей порог  $P \geq P_{thr}$ . График зависимости интегральной точности распознавания изображений  $Q$  от неопределённости в обучающих данных  $U_{TR}$  для разных значений  $P_{thr}$  проиллюстрирован на рисунке 3.4.

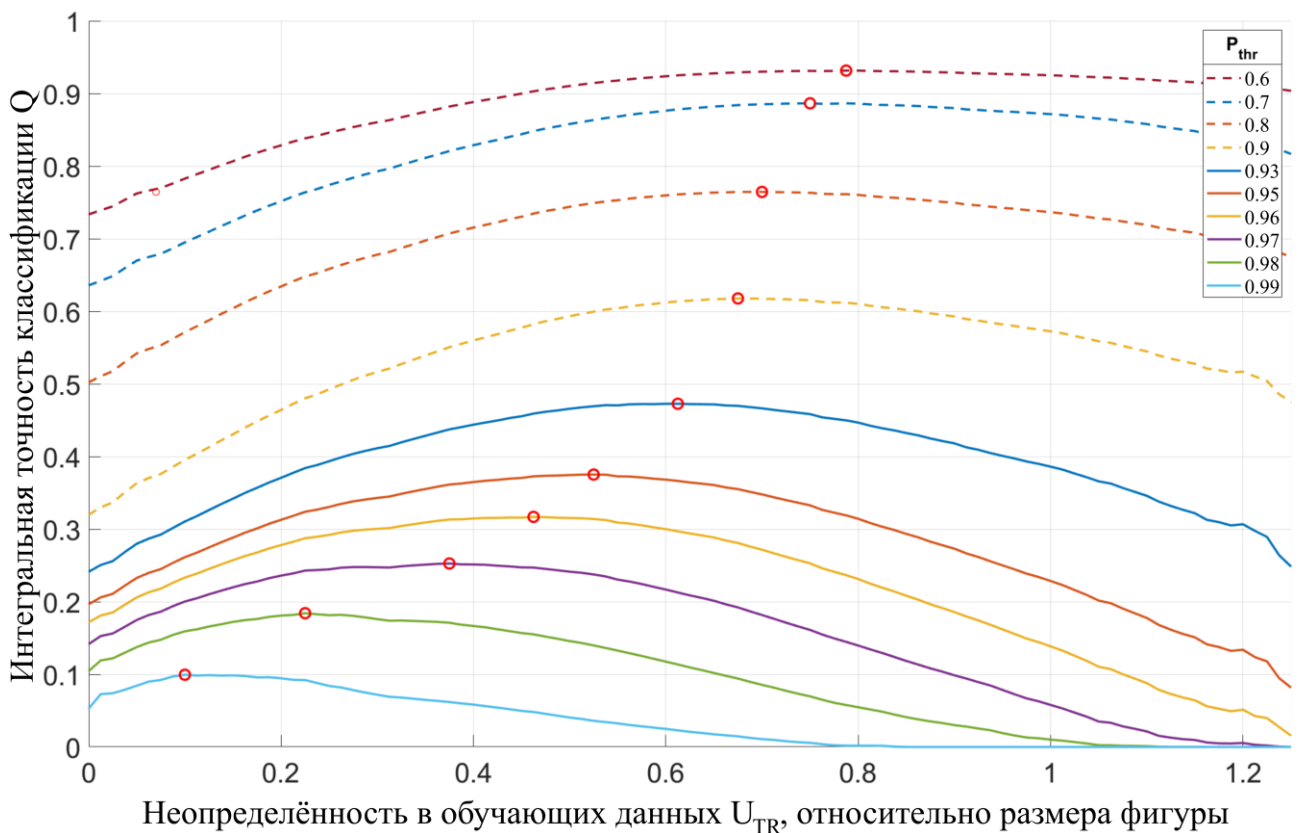


Рисунок 3.4 – График зависимости интегральной точности распознавания изображений  $Q$  от неопределённости в обучающих данных  $U_{TR}$  для разных значений  $P_{thr}$  и оптимальные значения неопределённости в обучающих данных  $U_{TR}$  для различных значений порога  $P_{thr}$

Если считать оптимальным набором данных для обучения при требуемом пороге минимальной точности классификации тот набор данных, который дает наибольшее значение интегральной точности классификации  $Q$ , то график на рисунке 3.4 удобен для определения оптимальной неопределенности в наборе



обучающих данных  $U_{TR}$ . При анализе зависимости интегральной точности распознавания  $Q$  от неопределенности обучающего набора данных  $U_{TR}$  при фиксированном пороге  $P_{thr}$ , выявляется четкий максимум кривой, положение этого максимума будет указывать на оптимальное значение неопределенности обучающего набора данных  $U_{TR}$  (рисунок 3.4,  $Q_{max}$  для различных  $P_{thr}$ ). Красные точки на графике отображают максимумы кривых зависимости интегральной точности распознавания изображений  $Q$  от неопределённости в обучающих данных  $U_{TR}$  для каждого из значений  $P_{thr}$ . Эти максимумы формируют монотонный тренд зависимости оптимального значения  $U_{TR}$  от требуемого  $P_{thr}$  (рисунок 3.5).

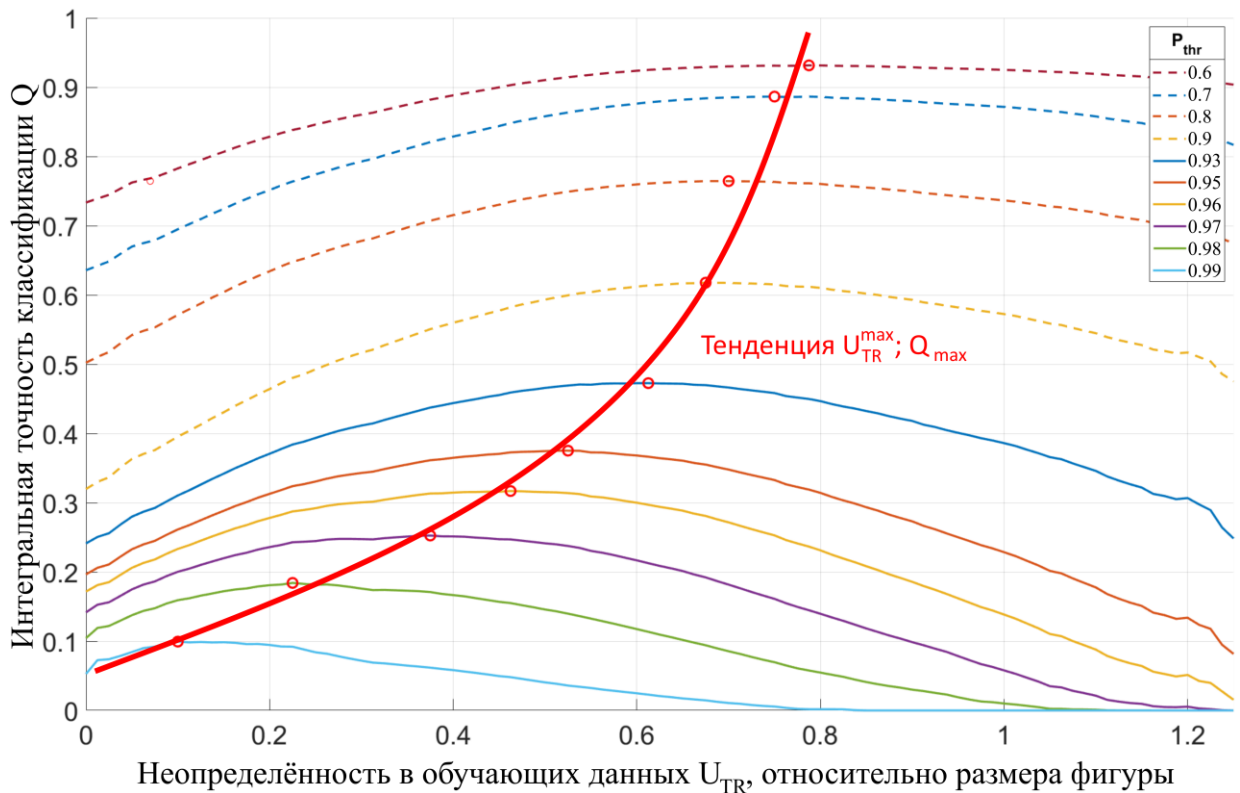


Рисунок 3.5 – Тенденция оптимальных значений неопределённости в обучающих данных  $U_{TR}$  для различных значений порога  $P_{thr}$

Анализ рисунка 3.5 позволяет сделать вывод, что обучение сети с оптимальным значением  $U_{TR}$  для фиксированного значения  $P_{thr}$  значительно повышает интегральную точность распознавания по сравнению с обучением сети

на идеальном наборе данных ( $U_{TR} = 0$ ). Например, для  $P_{thr} = 0,9$  значение  $Q_{max}$  превышает  $Q_0$  на 94% ( $Q_{max} = 0,62$  получено при  $U_{TR} = 0,068$ , а  $Q_0 = 0,32$  - при  $U_{TR} = 0$ ).

### 3.3 Распознавание зашумленных естественных изображений и обучение свёрточных нейронных сетей на зашумленных естественных изображениях

Для обобщения результатов данного исследования проведена серия экспериментов с различными типами изображений и шумов с использованием той же структуры СНС (подраздел 2.4) и того же подхода, включающего анализ зависимости  $P(U_{TR}; U_{TS})$  (подраздел 3.1). Примеры использовавшихся естественных изображений показаны на рисунке 3.6. К изображениям добавлен белый гауссовский шум с математическим ожиданием  $\mu = 0$  и различными стандартными отклонениями  $\sigma$ , чтобы создать отдельные наборы данных для обучения и распознавания. Таким образом, неопределённость  $U$  определяется через формулу (5), где  $I_{noise} = \sigma$ .

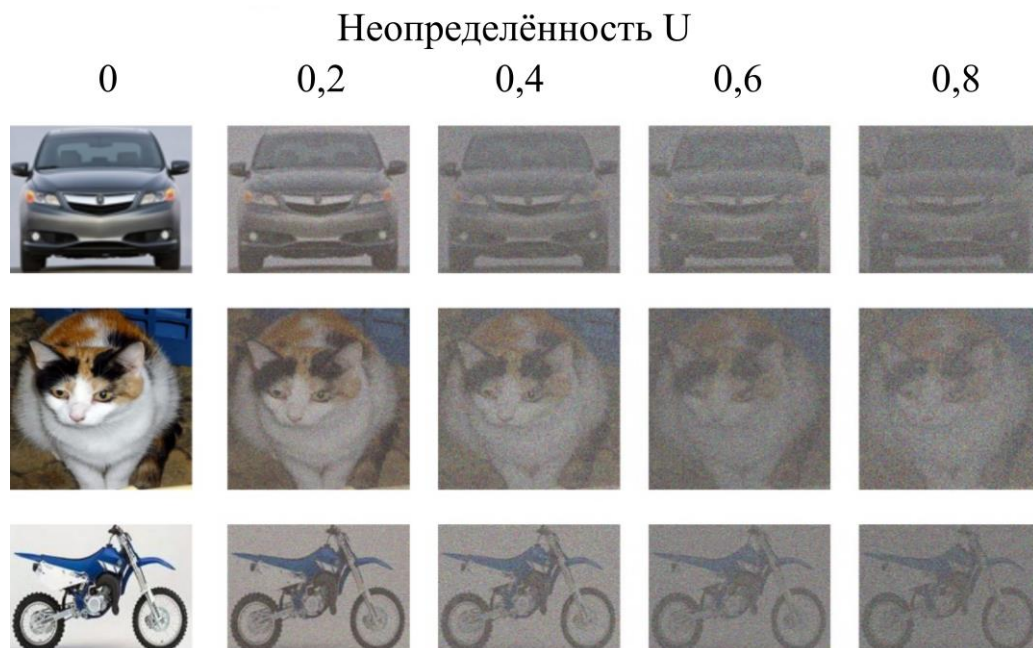


Рисунок 3.6 – Примеры изображений с различной интенсивностью шума

Для решения задачи классификации зашумленных изображений использована СНС, идентичная описанной ранее. Было сгенерировано пять наборов данных для обучения. Параметры наборов данных для обучения следующие:

1. В первом наборе данных  $U_{TR} = 0$  для всех изображений (шум не добавлялся).

2. Второй набор данных был разделен на три части, содержащие равное количество изображений; в первой части  $U_{TR} = 0$ , во второй части  $U_{TR} = 0,04$ , в третьей части  $U_{TR} = 0,08$ .

3. Третий набор данных был разделен на три части, содержащие равное количество изображений; в первой части  $U_{TR} = 0$ , во второй части  $U_{TR} = 0,12$ , в третьей части  $U_{TR} = 0,16$ .

4. Четвертый набор данных был разделен на три части, содержащие равное количество изображений; в первой части  $U_{TR} = 0$ , во второй части  $U_{TR} = 0,2$ , в третьей части  $U_{TR} = 0,4$ .

5. Пятый набор данных был разделен на три части, содержащие равное количество изображений; в первой части  $U_{TR} = 0$ , во второй части -  $U_{TR} = 0,4$ , в третьей части -  $U_{TR} = 0,8$ .

Таким образом, в четырех из пяти наборов данных исходные изображения были искажены шумами различной интенсивности.

Пять независимых СНС с идентичной структурой были обучены на пяти описанных выше наборах данных. Гиперпараметры сетей оставались неизменными. Для каждого эксперимента были созданы отдельные наборы данных со случайной последовательностью классов изображений для обеспечения стохастичности поиска целевой классификационной функции свёрточной нейронной сетью, при этом последовательность представления классов изображений нейронной сети (в разных наборах данных) оставалась неизменной. Начальные веса СНС задавались случайно для каждого уникального экземпляра нейронной сети. Обученные СНС использовались для распознавания

отдельно сгенерированных тестовых наборов, содержащих изображения с различными значениями интенсивности шума  $U_{TS}$ . Тестовые наборы данных имели однородную структуру: все изображения в каждом наборе данных имели одинаковое количество дополнительного шума, что предоставляет в результате фиксированное значение  $U_{TS}$  для всего набора данных. Тестовые и обучающие наборы содержали уникальные экземпляры изображений для избежания их пересечения и нарушения точности экспериментов. Все вероятности распознавания  $P$ , полученные в этих сериях симуляций, были усреднены по всем сериям экспериментов с фиксированным значением  $U_{TS}$ .

Результаты моделирования показаны на рисунках 3.7 и 3.8.

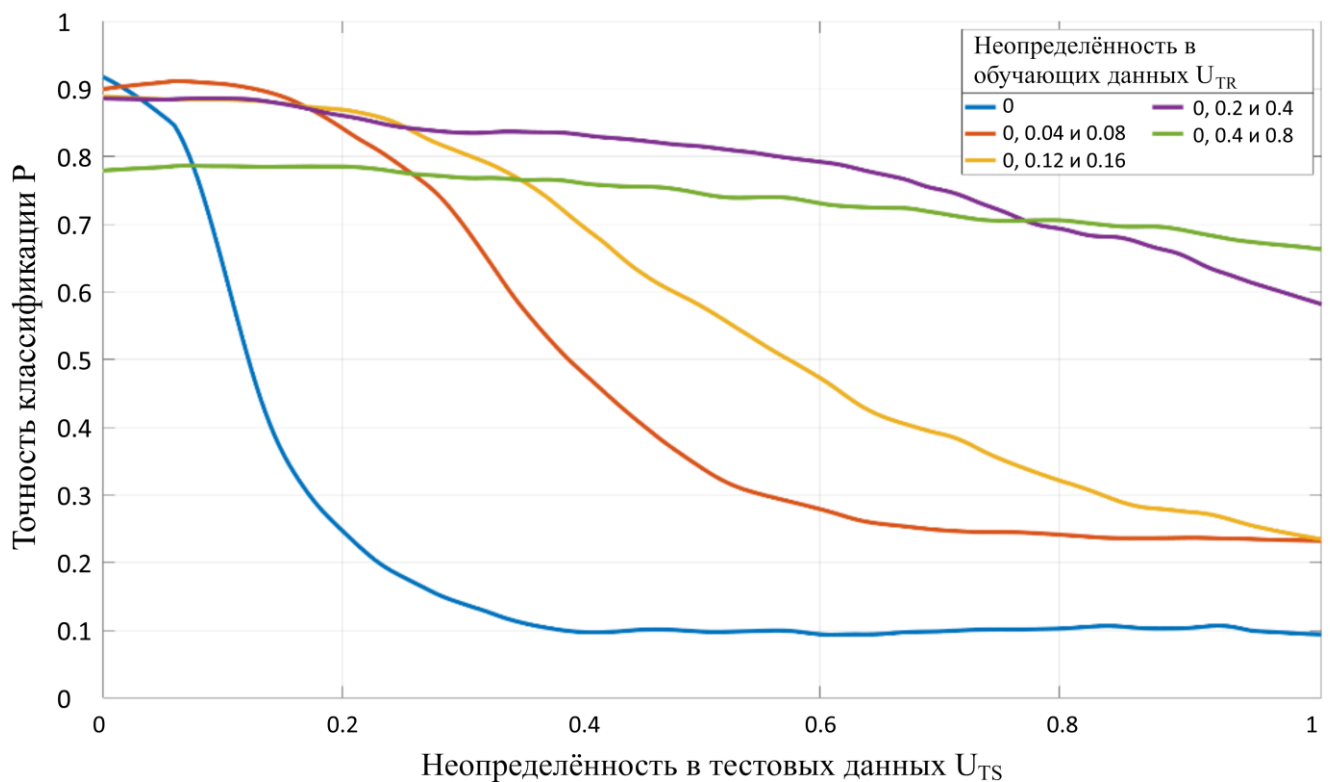


Рисунок 3.7 – Графики зависимости точности распознавания от неопределённости тестовых данных  $U_{TS}$ , полученной пятью реализациями CNN, обученными с различными значениями неопределённостями  $U_{TR}$

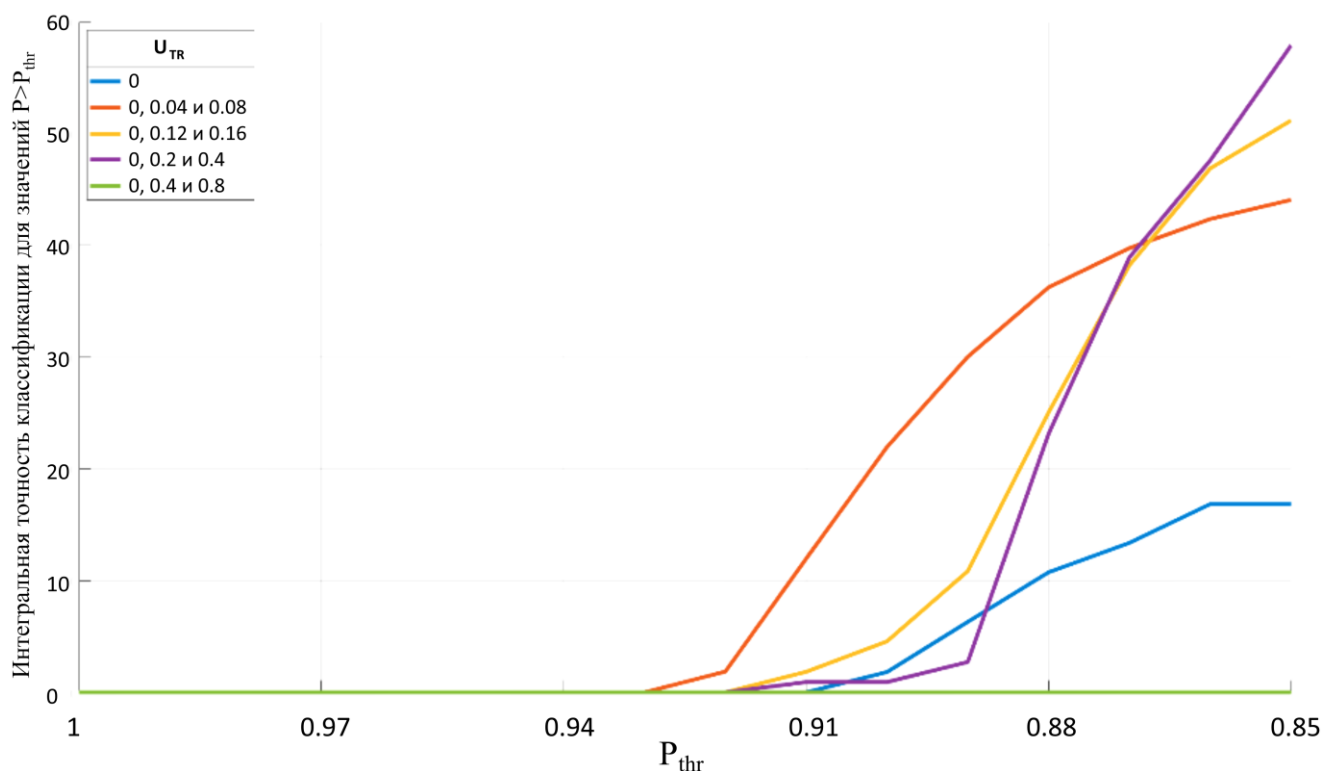


Рисунок 3.8 – Графики зависимости интегрального значения точности распознавания  $Q$  для всех  $P > P_{thr}$  от  $P_{thr}$  для пяти реализаций CNN, обученных с различными значениями неопределённостями  $U_{TR}$

Результаты, показанные на рисунках 3.7 и 3.8, позволяют утверждать, что умеренная интенсивность искажений в наборах обучающих данных  $U_{TR}$  является оптимальной для обучения нейронных сетей и для последующего распознавания зашумленных изображений с высокой пороговой вероятностью распознавания  $P_{thr}$ . Для рассматриваемого примера оптимальным набором данных является второй, имеющий  $U_{TR} = \{0; 0,04; 0,08\}$  для  $P_{thr} > 0,87$ . Дальнейшее увеличение значения  $U_{TR}$  приводит к падению интегральной точности классификации для высоких  $P_{thr}$ . Этот результат позволяет обобщить подтверждение существования оптимальной неопределенности обучающего набора данных для этих типов данных.

### 3.4 Результаты анализа работы свёрточной сети при прочих видах искажений естественных изображений

Чтобы подтвердить правильность и характерность выводов для разных видов искажений, была проведена серия экспериментов с различными типами искажений естественных изображений. Эти серии симуляций с различными типами изображений и шумов/искажений проведена с использованием одной и той же структуры СНС и одного подхода, включающего анализ зависимости  $P(U_{TR}; U_{TS})$ . Примеры изображений показаны на рисунке 3.9.

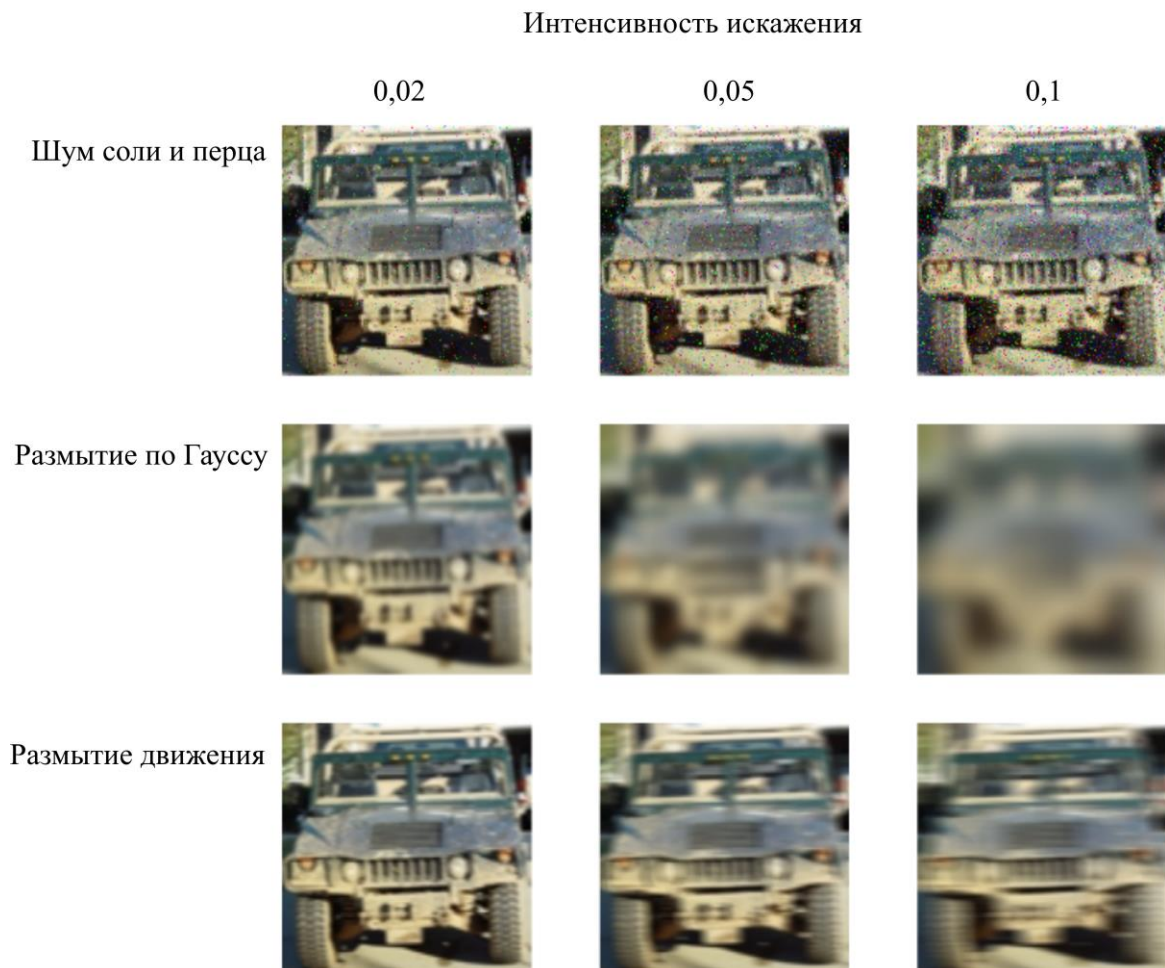


Рисунок 3.9 – Примеры изображений с различной интенсивностью шума соли и перца, гауссовым размытием и размытием движения

Результаты моделирования показаны на рисунках 3.10-3.12. Для шумов "соль и перец" неопределенность рассчитывалась по формуле (5), где  $I_{noise}$  - количество "шумовых" пикселей в изображении, а  $I_{info}$  - общее количество пикселей. Для размытия по Гауссу и размытия движения неопределенность рассчитывалась следующим образом:

$$U = S_{kernel} / S_{image}, \quad (11)$$

где  $S_{kernel}$  - размер ядер гауссовского фильтра, а  $S_{image}$  - размер изображения.

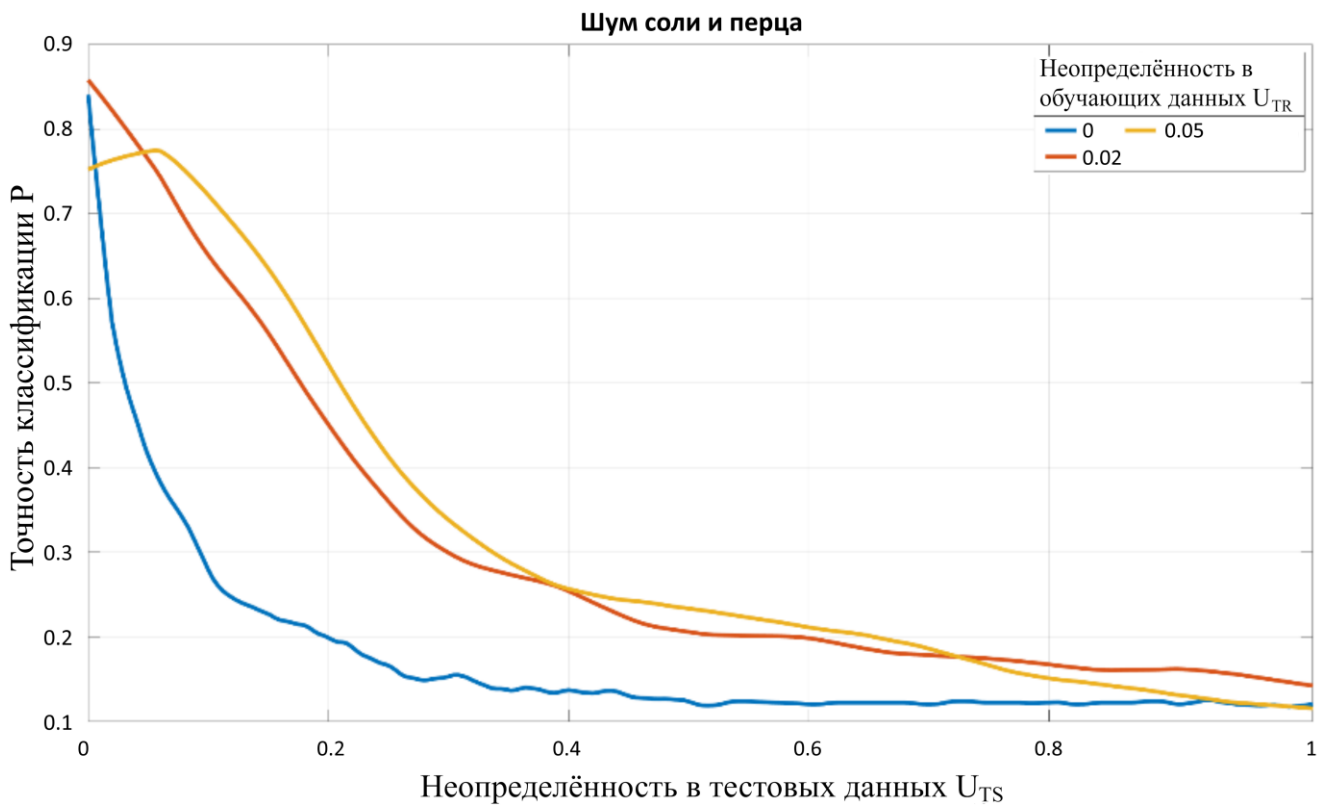


Рисунок 3.10 – Графики зависимости точности распознавания от неопределённости тестовых данных  $U_{TS}$ , полученной пятью реализациями CNN, обученными с различными значениями неопределённостями  $U_{TR}$ .

Неопределенность создается путем добавления шума в виде соли и перца и рассчитывается по (5)



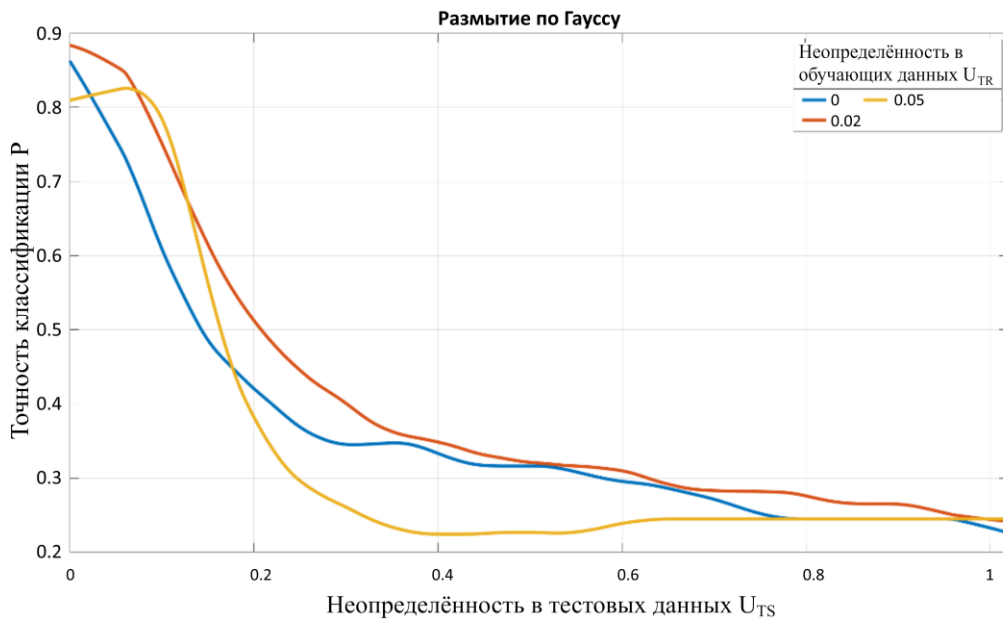


Рисунок 3.11 – Графики зависимости точности распознавания от неопределённости тестовых данных  $U_{TS}$ , полученной пятью реализациями CNN, обученными с различными значениями неопределённостями  $U_{TR}$ . Неопределенность получена путем добавления гауссовского размытия и рассчитана с помощью (11)

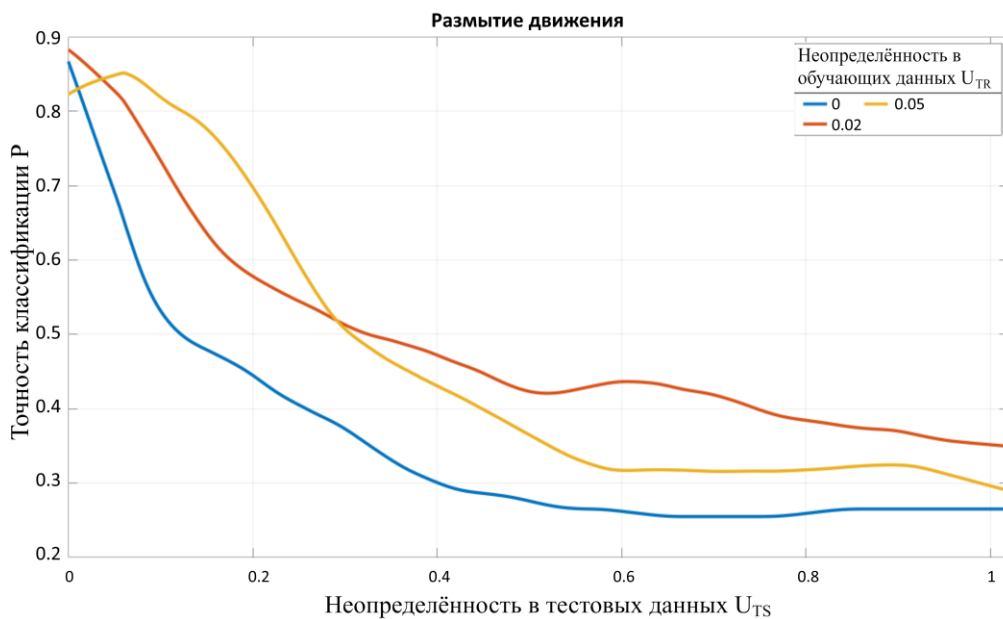


Рисунок 3.12 – Графики зависимости точности распознавания от неопределённости тестовых данных  $U_{TS}$ , полученной пятью реализациями CNN, обученными с различными значениями неопределённостями  $U_{TR}$ . Неопределенность создается путем добавления размытия движения и рассчитывается с помощью (11)



Анализ рисунков 3.10-3.12 показывает одну тенденцию: независимо от типа шума/искажения при обучении, его величина одинаково влияет на точность распознавания изображений. Использование для обучения изображений без дополнительного искажения ( $U_{TR} = 0$ , синие кривые) приводит к быстрому снижению точности распознавания с ростом количества искажений в тестовых данных  $U_{TS}$ . Этот факт говорит о том, что СНС, обученная таким образом, будет уязвима для состязательных атак и шумов. Результаты также показывают, что умеренная неопределенность  $U_{TR}$  обучающего набора данных (красные кривые) в случае шума "соль и перец", а также размытия по Гауссу и размытия движения является оптимальной для распознавания зашумленных/искаженных изображений без потери точности распознавания оригинального изображения. Этот результат позволяет нам обобщить подтверждение существования оптимального значения неопределенности в обучающем наборе данных для этих типов изображений с различными видами шума и искажений.

Таким образом, проведено моделирование в различных условиях. Все результаты показали одну и ту же закономерность: существует оптимальное количество искажений различной физической природы, которые, дополняя обучающие наборы данных, приводят к значительному улучшению помехоустойчивости обученной СНС и повышению общего качества распознавания. Чрезмерное количество искажений обучающего набора данных (желтые кривые на рис. 3.10-3.12) делает обученную нейронную сеть неустойчивой. Это видно по немонотонному характеру соответствующих (желтых) кривых помехоустойчивости (рост точности распознавания с увеличением  $U_{TS}$ ).

### **Выводы по разделу 3**

1. Величина неопределенности в обучающем наборе данных  $U_{TR}$  существенно влияет на точность распознавания и зависимость точности распознавания от неопределенности в тестовом наборе данных  $U_{TS}$ . В данном

разделе проанализирована точность распознавания множества наборов данных с различными значениями неопределенности и получена зависимость точности распознавания от интенсивности искажений в обучающем наборе данных. Существование оптимальной (с точки зрения точности распознавания) интенсивности искажений в обучающем наборе данных (для нейронных сетей, работающих с данными с неизвестным заранее значением неопределенности) было предположено и доказано для различных типов изображений и шумов.

2. Определение оптимума интенсивности искажений в обучающем наборе данных может быть выполнено с помощью статистического моделирования. Обучение сети с использованием набора данных с оптимальным значением неопределенности  $U_{TR}$  обеспечивает значительное увеличение точности распознавания по сравнению с обучением на идеальном наборе данных. СНС учится не использовать шум/искажения в качестве признаков во время обучения, поскольку шум/искажения не способствуют нахождению целевой классификационной функции при обучении. Это повышает обобщающую способность нейронной сети и ее устойчивость к состязательным атакам и шумам. В то же время чрезмерное количество шума/искажений разрушают процесс обучения, что приводит к общему снижению точности распознавания. Этот вывод можно использовать для повышения качества распознавания изображений аппаратами классификации изображений на основе СНС путём добавления некоторого (оптимального) количества шума в набор обучающих данных.

3. Полученные результаты применимы к СНС с распространёнными структурами и различным типам искажений в данных (гауссовский шум, искажение расположения точек, шум соли и перца, размытие по Гауссу, размытие в движении и т.д.).

## 4. МЕТОД ОПТИМАЛЬНОЙ АУГМЕНТАЦИИ ОБУЧАЮЩИХ ДАННЫХ БЕЗ УВЕЛИЧЕНИЯ ИХ ОБЪЁМА

### 4.1 Проблема распознавания естественных изображений

Проблема робастности обучения глубоких СНС для распознавания естественных изображений в последнее время всё чаще привлекает внимание исследователей по всему миру. Классификация естественных изображений свёрточными нейронными сетями, обученными на качественных образцах, хорошо выполняется для большинства изображений, но в значительной степени зависит от наличия искажений во входных изображениях, как было показано ранее. При наличии искажений во входных изображениях возникает эффект деградации качества распознавания [64], [185]. Искажения в естественных изображениях, вызывающие деградацию качества распознавания, появляются из-за множества причин, таких как сжатие с потерями, тепловые шумы матриц фотокамер, абберации оптических элементов и т. д., и чаще всего непредсказуемы, однако также могут быть и малозаметными для человека. На рисунке 4.1 показано снижение визуального качества изображений с увеличением степени сжатия, а также значение схожести сжатых изображений с оригиналом, полученные методом SSIM (Structural Similarity Index Measure, «мера индекса структурного сходства») [186]. Чтобы проиллюстрировать эффект снижения значения индекса структурного сходства при сжатии изображений, в данной работе также проведено моделирование, результаты которого приведены на рисунке 4.1. При моделировании набор исходных изображений подвергался сжатию с применением алгоритма JPEG [187] с различными коэффициентами, после чего оценивалась точность распознавания этих изображений нейронной сетью со структурой, описанной в [67], [66] (для обучения нейронной сети были

использованы другие примеры изображений без искажений и не подвергавшиеся сжатию).

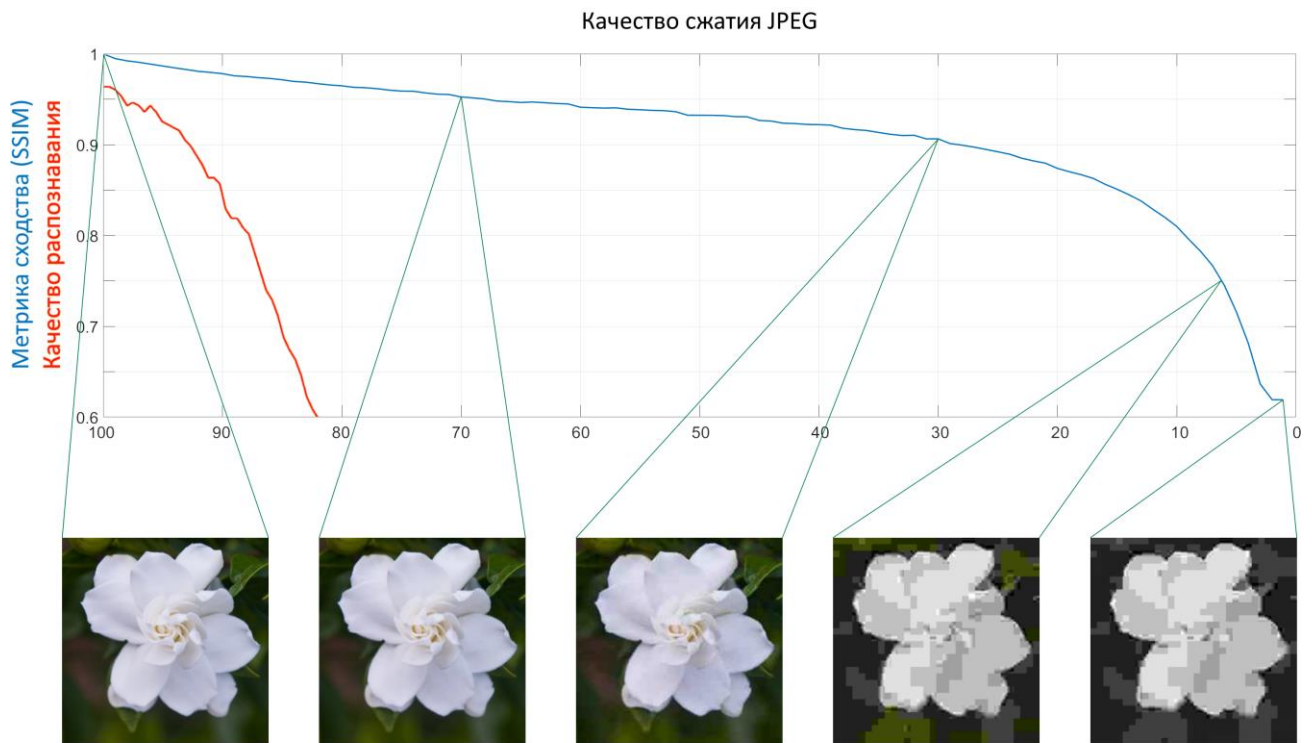


Рисунок 4.1 – Качество изображения и значения индекса структурного сходства изображения с оригиналом при использовании сжатия JPEG

Как показано на рисунке 4.1, заметные искажения изображения, вызванные сжатием, появляются при коэффициенте сжатия менее 30. Однако точность распознавания нейронной сетью снижается быстрее, чем визуально снижается качество изображения относительно исходного.

Для решения этой проблемы был ранее предложен метод повышения робастности обучения путём добавления в обучающие выборки зашумленных изображений [185], [67], [188]. В данном разделе проведён анализ различных способов аугментации обучающих данных для достижения наилучшего качества распознавания изображений, слабо деградирующего по мере увеличения интенсивности искажений во входных изображениях. Проведён анализ снижения качества распознавания изображений отдельных классов и показаны

преимущества сетей, обученных на изображениях, аугментированных оптимальным методом.

#### **4.2 Наборы данных, структура свёрточной нейронной сети, типы искажений и моделирование**

На качество обучения нейронных сетей, помимо прочего, влияет количество и разнообразие обучающих данных [189]. Принято характеризовать обучающие данные параметром «репрезентативность», дающим представление о том, насколько полно обучающие данные отражают закономерности, присущие всему практически возможному их многообразию. Увеличение количества обучающих данных приводит к избыточным затратам на обучение нейронных сетей, поэтому предпочтение следует по возможности отдавать методам, обеспечивающим наибольшее разнообразие данных при их относительно небольшом количестве.

В данном разделе изучено влияние способов аугментации обучающих данных на примере одного варианта искажения изображений – размытие по Гауссу. Из-за своего широкого распространения этот вид искажения часто встречается на практике и может чаще других видов искажений попадать в наборы изображений, подвергающихся распознаванию системами с применением СНС. Размытие по Гауссу используется для уменьшения цифрового шума изображений, теплового шума матриц, часто используется в системах компьютерного зрения и для сокрытия определённых частей изображений. Помимо снижения влияния высокочастотного шума в изображении, размытие по Гауссу уменьшает количество деталей в изображении и может негативно влиять на способность сети к выявлению важных признаков (рисунок 4.2).



Рисунок 4.2 – Изображения, размытые по Гауссу (с размером окна размытия)

В данном исследовании проведено сравнение несколько различных способов аугментации обучающих данных. Многие исследователи, использующие аугментацию для повышения качества работы нейронных сетей и систем компьютерного зрения, используют искажение малой части присутствующих в обучающих выборках изображений, однако неизвестно, является ли данный способ наиболее эффективным. В данном исследовании использовано 5 различных способов аугментации, показанных на рисунке 4.3. Зелёным цветом обозначены части обучающего набора, не подвергавшиеся искажению. Красным цветом обозначены части обучающего набора с максимальным искажением изображений. Части обучающего набора, искажённые промежуточным образом, изображены цветами, переходными от зелёного к красному.

Общий объём обучающей выборки при каждом способе аугментации одинаков, что позволяет оценить влияние на качество обучения именно многообразия изображений, а не их количества.



Рисунок 4.3 – Способы аугментации обучающих выборок

На рисунке 4.3 показаны 5 различных делений обучающего набора изображений.

1. Классический набор изображений для обучения – без использования искажённых изображений.

2. В обучающем наборе все изображения искажены с размером окна = 25.

3. Для каждого из изображений в обучающем наборе размер окна искажения задавался случайно в диапазоне от 0 до 25.

4. Использовалась половина изображений из исходного набора (обозначенная штриховкой) - без искажений, а также другая половина (обозначенная сплошным цветом) - с искажением при размере окна = 25.

5. Использовалась половина изображений из исходного набора (обозначенная штриховкой) дважды – без искажений и с искажением при размере окна = 25.

Во всех наборах, в которых использовались искажённые изображения, использовался размер окна размытия до 25 пикселей. Данное значение выбрано исходя из результатов предыдущих исследований. Ранее в разделе 3 показано, что независимо от типа искажения, закономерность влияния его интенсивности в

обучающих изображениях на точность распознавания тестовых изображений однотипная. В каждом случае наблюдается оптимальное значение интенсивности искажения обучающих изображений (в случае присутствия искажений различной физической природы существует некоторое оптимальное значение неопределенности обучающего набора изображений для распознавания искаженных изображений без потери точности распознавания неискаженных изображений).

Для исследования был использован открытый набор изображений для классификации, включающий в себя 8 классов изображений [64], [190]. Современные глубокие СНС, такие как Inception [114], ResNet [191] и VGG [192], успешно используются для распознавания набора изображений ImageNet [193], содержащего 1000 классов изображений. Однако проведение исследования на базовой СНС с большими наборами изображений затруднено ввиду ограниченности обобщающей способности выбранной сети. Использование упрощённой структуры сети и сокращённого набора изображений обусловлено соображениями ресурсоёмкости вычислительных экспериментов, необходимых для получения кривых помехоустойчивости. Общий принцип при изменении структуры сети и расширении набора изображений остаётся неизменным.

С использованием наборов изображений, схематически представленных на рисунке 4.3, были обучены 5 экземпляров нейронной сети с идентичной структурой. Архитектура нейронной сети показана в таблице 4.1.

Таблица 4.1 – Архитектура сети

<b>Слой (тип)</b>	<b>Выходная размерность</b>	<b>Число параметров</b>
Zero Padding	(260, 260, 3)	0
Convolution	(256, 256, 8)	608
Batch Norm	(256, 256, 8)	32
Activation	(256, 256, 8)	0



Max Pooling	(128, 128, 8)	0
Zero Padding	(132, 132, 8)	0
Convolution	(128, 128, 128)	25728
Batch Norm	(128, 128, 128)	512
Activation	(128, 128, 128)	0
Max Pooling	(64, 64, 128)	0
Zero Padding	(68, 68, 128)	0
Convolution	(64, 64, 256)	819456
Batch Norm	(64, 64, 256)	1024
Activation	(64, 64, 256)	0
Max Pooling	(32, 32, 256)	0
Zero Padding	(36, 36, 256)	0
Convolution	(32, 32, 512)	3277312
Batch Norm	(32, 32, 512)	2048
Activation	(32, 32, 512)	0
Max Pooling	(16, 16, 512)	0
Convolution	(16, 16, 8)	4104
Flatten	(2048)	0
Dense	(8)	16392
Activation	(8)	0

Суммарное число параметров: 4,147,216

Суммарное число обучаемых параметров: 4,145,408

Суммарное число необучаемых параметров: 1,808

### **4.3 Зависимости точности распознавания изображений от интенсивности размытия**

Для обученных нейронных сетей были получены кривые помехоустойчивости (зависимости точности распознавания от интенсивности искажений в тестовых изображениях). Множества тестовых наборов генерировались путём размытия исходных изображений по Гауссу с различными

размерами окна. Мерой интенсивности искажения в данном случае является размер окна размытия. Графики зависимости точности распознавания изображений от интенсивности размытия (размера окна размытия) представлены на рисунке 4.4. Также были получены графики зависимостей точности распознавания худших четырех классов для демонстрации скорости спада качества распознавания нейронной сетью, обученной на наборах изображений, аугментированных разными способами (рисунок 4.5).

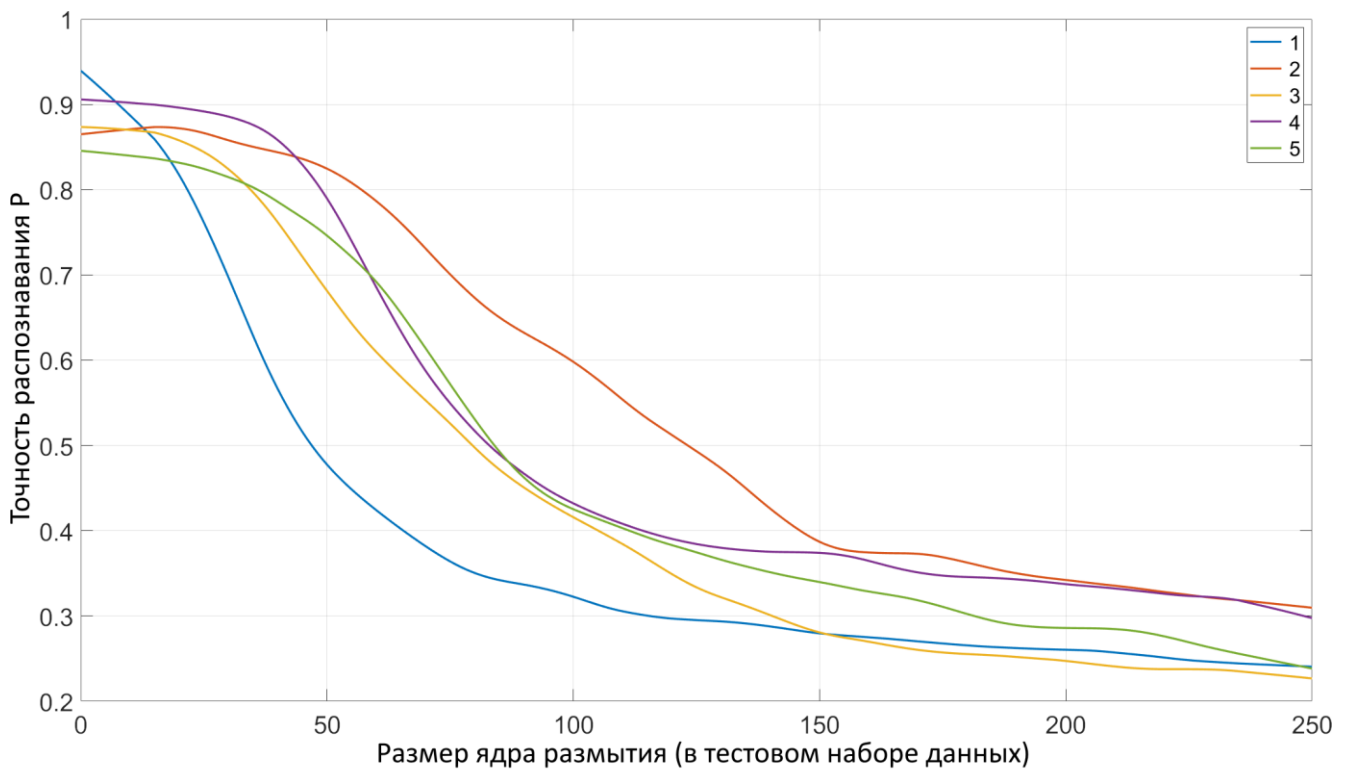


Рисунок 4.4 – Графики зависимости точности распознавания изображений от интенсивности размытия (размера окна размытия)

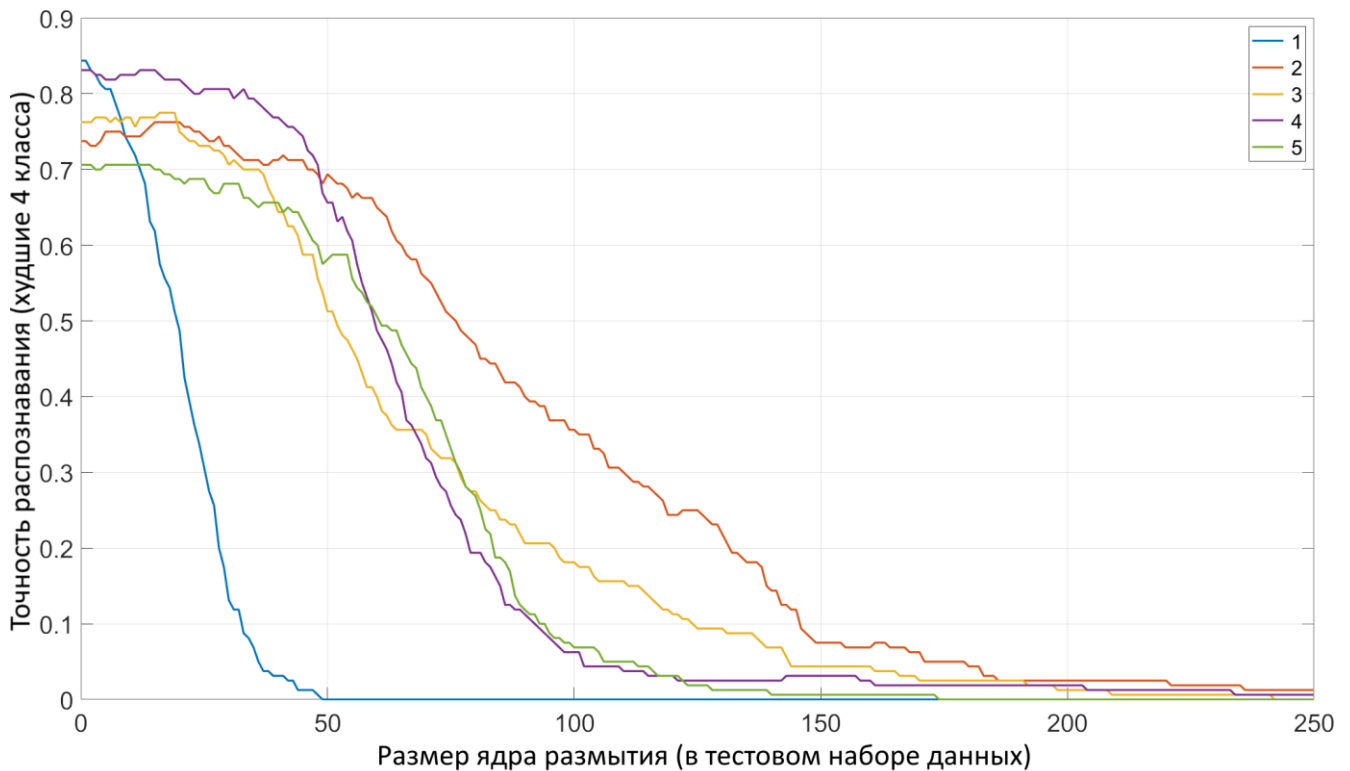


Рисунок 4.5 – Графики зависимости точности распознавания худших 4 классов от интенсивности размытия (размера окна размытия)

Из полученных графиков видно, что аугментация методом 4 (использование половины изображений из исходного набора без искажений и другой половины с искажением при размере окна = 25) обеспечивает наилучший практический результат (высокую точность распознавания слабо искажённых изображений и медленный спад качества распознавания при росте интенсивности искажения). Использование в качестве обучающих данных изображений без аугментации (дополнения обучающих данных) приводит к низкому качеству распознавания изображений, искаженных размытием по Гауссу и быстрому спаду точности классификации изображений с ростом интенсивности размытия. Как видно из рисунка 4.5, при обучении на неискаженных изображениях нейронная сеть теряет способность распознавания половины классов из набора при размере окна размытия, равном 48. При этом обучение с искажением половины изображений из набора (вариант 4) незначительно снижает точность

классификации неискажённых изображений, но существенно повышает точность при размытии распознаваемых изображений в важном диапазоне размеров окна размытия от 7 до 50. При этом робастность распознавания изображений по отдельным классам при обучении с аугментацией методом 4 (использование половины изображений из исходного набора без искажений и другой половины с искажением при размере окна, равном 25) также повышается по сравнению с обучением без использования аугментации данных.

#### **Выводы по разделу 4**

1. Произведён анализ точности распознавания множества наборов изображений с различными способами аугментации, получена зависимость точности распознавания тестовых изображений от интенсивности размытия по Гауссу тестовых изображений для нейронных сетей, обученных с использованием различных методов аугментации. Данное исследование показывает, что способ аугментации обучающего набора изображений существенно влияет на точность распознавания и зависимость точности распознавания от интенсивности размытия по Гауссу в распознаваемых изображениях. Проведённое исследование показало, что способ аугментации, заключающийся в искажении с определённой интенсивностью определённой доли изображений из исходного обучающего набора, позволяет обеспечить более устойчивую работу нейронной сети в режиме распознавания искажённых изображений и позволяет добиться значительного увеличения точности их распознавания. Методом статистического моделирования показано, что разработанный способ аугментации обеспечивает наивысшую интегральную точность распознавания тестовых изображений с различной интенсивностью искажений при сохранении монотонности функции зависимости точности

распознавания от неопределённости тестовых изображений и, таким образом, является оптимальным.

2. Предложенный способ аугментации целесообразно использовать в системах технического зрения различного назначения, в которых возможно размытие изображений по Гауссу, например, в системах распознавания фотографий, системах космической съемки [78] и т.п.

## **5. НИЗКОЧАСТОТНАЯ ФИЛЬТРАЦИЯ ИЗОБРАЖЕНИЙ ДЛЯ ПРОТИВОДЕЙСТВИЯ СОСТЯЗАТЕЛЬНЫМ ИСКАЖЕНИЯМ**

### **5.1 Методы противодействия состязательным искажениям**

Свёрточные нейронные сети нашли широкое применение при решении современных вычислительных задач, поскольку позволяют автоматизировать решение широкого класса проблем, таких как классификация и сегментация изображений [194], обнаружение и отслеживание объектов в видеопотоке [195], а также генерацию изображений [196], [197]. Также СНС – наиболее эффективный инструмент машинного обучения при обработке аудио [198], [199]. В последние годы всё более значительная доля вычислительных мощностей задействована в обработке мультимедиа, а рост вычислительных мощностей позволяют применять всё более сложные и требовательные алгоритмы машинного обучения [200], [201]. СНС способны эффективно извлекать признаки из мультимедиа и работать с большими объемами информации, поэтому все чаще используются для решения задач, формализация которых затруднительна или невозможна.

Вместе с тем, значимой нерешённой проблемой для СНС является их чувствительность к присутствию искажений или шумов в данных. Нейронные сети, обученные на изображениях без шума и искажений, не обеспечивают обобщающей способности, достаточной для классификации искаженных или зашумленных изображений. До сих пор достоверно не известны точные характеристики устойчивости СНС к шумам и искажениям в изображениях, известны лишь отдельные исследования в данном направлении [202], [203], [204]. Наиболее сильно снижают точность распознавания изображений т.н. состязательные искажения, поскольку являются целенаправленными и используют некоторые свойства моделей нейронных сетей, на которые проводится атака. Одним из первых упоминаний данной проблемы является

исследование [152], продемонстрировавшее, помимо прочего, недостатки обобщающей способности нейронных сетей. Авторы также обнаружили, что состязательные искажения относительно эффективны для нейронных сетей с различным количеством слоев, архитектурой или обученных на различных подмножествах обучающих данных. Состязательные примеры изображений являются переносимыми на различные нейронные сети, даже если эти сети обучены с другими гиперпараметрами или на другом наборе данных. Позже было предложено большое число методов для создания состязательных примеров, в том числе Fast Gradient Sign Method (FGSM) [25], Deepfool [149], One-pixel attack [150], [205] и прочие. При использовании FGSM, сеть maxout [151], изначально достигавшая вероятности ошибки 0,45%, неправильно классифицировала 89,4% состязательных примеров, а средняя степень уверенности составила 97,6%. Более того, с ростом разрешения используемых изображений ошибка распознавания состязательных примеров растёт. На данный момент актуально «противостояние» различных состязательных атак и методов борьбы с ними [206], [207], [208]. До сих пор не существует эффективных методов противодействия высокочастотным состязательным атакам. Методы, использующие автокодировщики, помогают обнаружить высокочастотные атаки противника, но не предотвратить их [209].

Большое число естественных изображений, представленных в цифровом виде, также имеют искажения. Большинство таких искажений внесены в процессе получения этих изображений. Искажения такого рода вносятся в наборы данных без участия злоумышленника (необычные ракурсы и углы фотосъемки, тепловые шумы матриц фотокамер и особенности объективов, искажения атмосферы, артефакты оцифровки изображения, и т.д.). Естественные состязательные примеры часто являются непредсказуемыми, поэтому методы борьбы с ними часто неочевидны.

Также искажения классифицируются как domain shift [210], могут быть использованы злоумышленниками [211]. Одной из первых работ, в которой были рассмотрены естественные состязательные примеры, является [164]. На основе набора данных ImageNet, включающего десятки миллионов изображений, авторы создали свои наборы данных (ImageNet-A и ImageNet-O), содержащие изображения, которые наименее качественно классифицируются с помощью современных моделей машинного обучения. При этом изображения, включаемые авторами в эти наборы, содержат ограниченное число ложных признаков.

Современные архитектуры СНС, такие как AlexNet [11], DenseNet-121 [56], ResNet-50 [212], SqueezeNet [213], VGG-19 [113] на наборе данных ImageNet-A достигает точности распознавания не выше 2,2% (что примерно на 90 п. п. меньше точности распознавания набора данных ImageNet теми же сетями). Авторы [164] показали, что существующие методы аугментации данных практически не повышают производительность, а использование других публичных наборов данных для обучения дает ограниченное улучшение. При этом авторы [164] не предлагают эффективных способов преодоления эффекта состязательных искажений. Все вышеописанные проблемы должны учитываться при разработке современных систем распознавания изображений на основе СНС.

Известны исследования, ориентированные на разработку методов борьбы с искажениями и шумами в изображениях, распознаваемых свёрточными нейронными сетями [214], [215], [216], [217], [218], [219]. В части этих методов используются различные сложные фильтры (денойзеры), т.е. предобработка изображений, использование состязательных сетей и обучение с зашумленными данными. Значительная часть разработанных систем предварительной обработки изображений специфична к определённым видам искажений и способам построения состязательных атак, поэтому достаточно быстро преодолевается в новых алгоритмах состязательных искажений [220], [221]. В частности, важные



требования к денойзерам, такие как сохранение чётких границ и текстур объектов, не позволяют успешно противостоять состязательным атакам.

Другой известный способ обеспечения устойчивости к состязательным атакам заложен в использовании двух или более сетей с противоположными целями. В данном случае конкурирующая (состязательная) сеть обучается генерации состязательных искажений в изображениях с целью обеспечить неправильную классификацию изображений классификатором, а классификатор обучается противостоять таким состязательным примерам [222], [223]. Соответственно, состязательные примеры могут являться хорошим источником аугментации. Такой способ аугментации является эффективным для повышения устойчивости СНС к неочевидным и незаметным для человека искажениям. Однако такой подход заметно усложняет процесс разработки системы, обучения нейронной сети, а также требует постоянного контроля процесса обучения, и не всегда является надёжным [224]. Важнейшим способом противостояния наличию шумов и искажений в тестовых данных является обучение нейронной сети с использованием аугментированных данных [225], [226]. Для дополнения данных могут использоваться различные методы, специфичные для решаемой задачи. Однако значительная доля исследований, связанных с применением СНС, до сих пор не рассматривает данной проблемы.

Можно обобщить известные методы борьбы с состязательными искажениями следующим образом:

1. защитная дистилляция [216] предполагает использование двух или более сетей; метод эффективен против некоторых неопределённых угроз, но уязвим против тонкой подстройки высокочастотных атак;

2. градиентная регуляризация [227], [228] трудно реализуема; количественная оценка устойчивости к атакам на основе градиента отсутствует;

3. денойзеры используются в основном для улучшения визуального качества изображения или повышения качества изображения, не доказана их

эффективность против атак на основе градиента; количественных оценок мало [218];

4. существует работа, реализующая генератор для синтеза изображений [229], авторы которой используют сложные алгоритмы генерации изображений, модель СНС и наборы данных;

5. генеративные состязательные сети [219] эффективны для обнаружения состязательного шума, но не противостояния; дискриминатор (важная часть GAN) также уязвим для тех же состязательных атак;

6. низкоуровневые преобразования изображений [230] еще один эффективный метод. Тем не менее, реализация метода требует временных затрат, а имеющиеся результаты несопоставимы (разные модели CNN и наборы данных).

В данной работе предложен метод противостояния высокочастотным шумам, основанный на принципах, заимствованных из радиотехнических систем – фильтрация зашумлённых изображений с помощью низкочастотного (low-pass) фильтра Гаусса [231], [232]. Фильтрация изображений позволяет достаточно эффективно подавить высокочастотный шум, но одновременно размывает изображение, снижает его чёткость. Это в свою очередь приводит к снижению точности распознавания размытого изображения нейронными сетями, исходно обученными распознаванию чётких изображений (рисунок 5.1). Таким образом, фильтрация изображений фильтром Гаусса позволяет свести проблему противостояния высокочастотным состязательным атакам к проблеме распознавания размытых изображений, уже рассмотренной в предыдущей работе [71]. В данном разделе проводится большой набор тестов для различных интенсивностей FGSM и различных размеров гауссова фильтра. Это позволяет определить оптимальный размер фильтра Гаусса для предлагаемого метода. Суть предлагаемого метода показана на рисунке 5.1. В работе не рассматриваются сложные системы предварительной обработки изображений, такие как GAN или

автоэнкодеры, поскольку они неэффективны против состязательных атак на основе градиента. Предложенный метод прост в реализации и эффективен. Метод может быть использован в различных системах распознавания образов, реализованных на различных аппаратных платформах, в том числе с крайне ограниченными вычислительными ресурсами.



Рисунок 5.1 – Схема предложенного метода

Для повышения эффективности системы распознавания необходимо обучить её эффективно классифицировать размытые данные. Для этого следует проводить обучение на данных, дополненных размытыми изображениями [67]. В данной работе доказано, что такой способ обучения системы распознавания и предварительной обработки классифицируемых изображений является эффективным для повышения точности классификации зашумлённых данных без существенного снижения качества распознавания, вызванного размытием. В данном разделе доказано наличие оптимума интенсивность размытия зашумлённых данных, подлежащих классификации, и предлагается метод нахождения этого оптимума. Анализируется и сравнивается поведение двух нейронных сетей: простой СНС, показывающей высокие результаты на наборах данных с небольшим количеством классов, а также сети EfficientNetV3 на

наборах данных ImageNet, CIFAR-10, Natural Images и Rock-Paper-Scissors. Простая CNN тестируется на наборах данных CIFAR-10, Natural Images и Rock-Paper-Scissors. EfficientNetB3 тестируется на Natural Images и ImageNet-1k.

## **5.2 Инструменты и методы**

### **5.2.1 Наборы данных**

Для оценки эффективности предложенного подхода в различных условиях и подтвердить его переносимость, проведены эксперименты на общедоступных наборах данных. В исследовании были использованы 4 набора данных, на которых проводилось обучение сетей и анализ результатов, в том числе CIFAR-10, ImageNet, Rock-Paper-Scissors и набор данных Natural Images. В данном подразделе представлено описание этих наборов данных и кратко описано обоснование выбора.

CIFAR-10 – один из самых широко используемых для обучения и тестирования классификаторов наборов изображений. Набор включает 60000 изображений в 10 классах, размерность изображений приведена к  $32 \times 32 \times 3$  [233]. Такая размерность является сравнительно низкой, что, с одной стороны, позволяет затрачивать значительно меньше временных и вычислительных ресурсов на обучение, с другой – значительно снижает точность классификации при наличии на изображениях искажений или шумов даже небольшой интенсивности (рисунок 5.2).

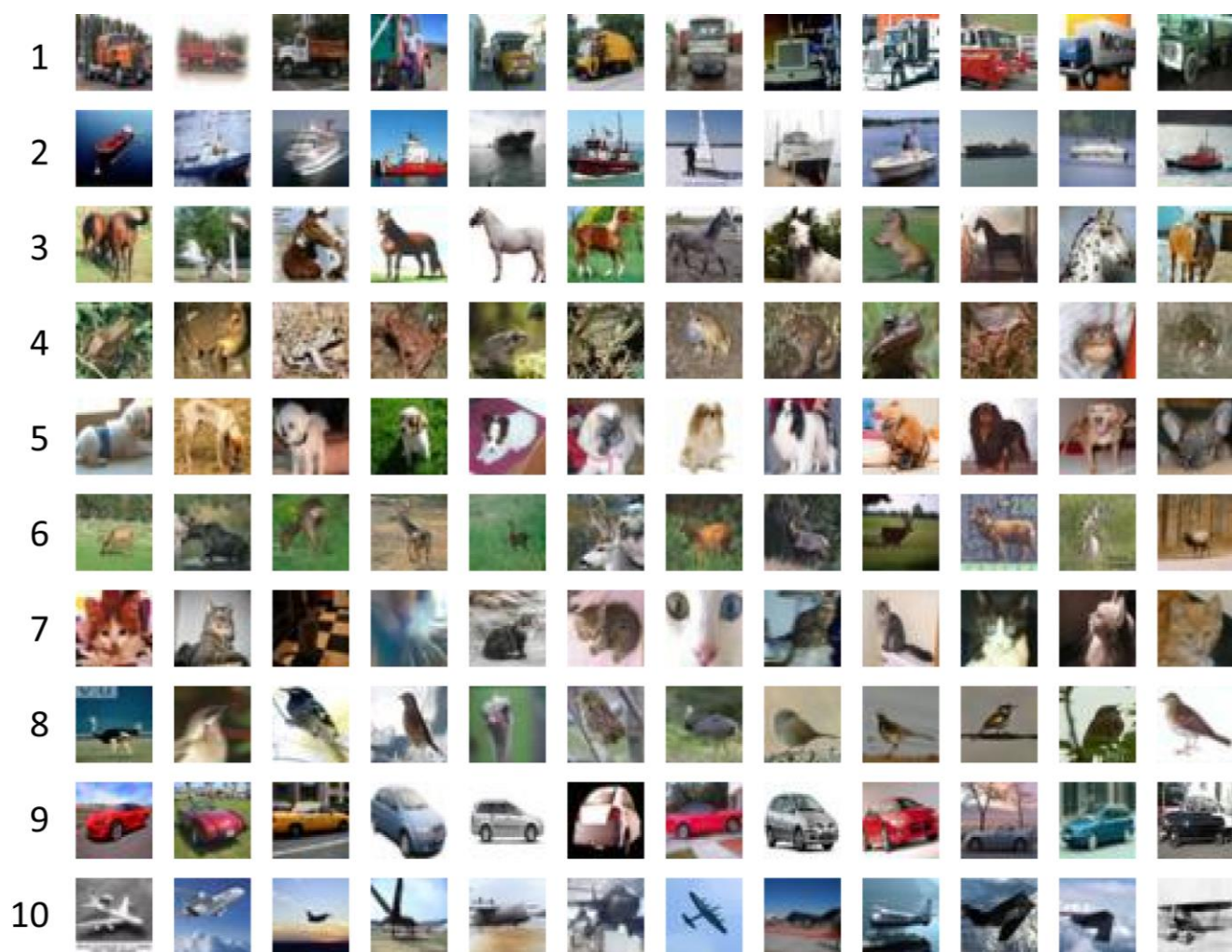


Рисунок 5.2 – Примеры изображений из набора CIFAR-10 (классы: 1 – грузовик, 2 – корабль, 3 – лошадь, 4 – лягушка, 5 – собака, 6 – олень, 7 – кошка, 8 – птица, 9 – автомобиль, 10 – самолёт)

Natural Images – небольшой набор данных естественных изображений [64], состоящий из 6899 изображений, включенных в 8 различных классов (самолёт, автомобиль, кошка, собака, цветок, фрукт, мотоцикл, человек). Поскольку обучение нейронных сетей на больших наборах данных, таких как ImageNet-1k, затруднительно, для сокращения временных и вычислительных затрат набор Natural Images использовался для проведения широкого класса тестов (рисунок 5.3).



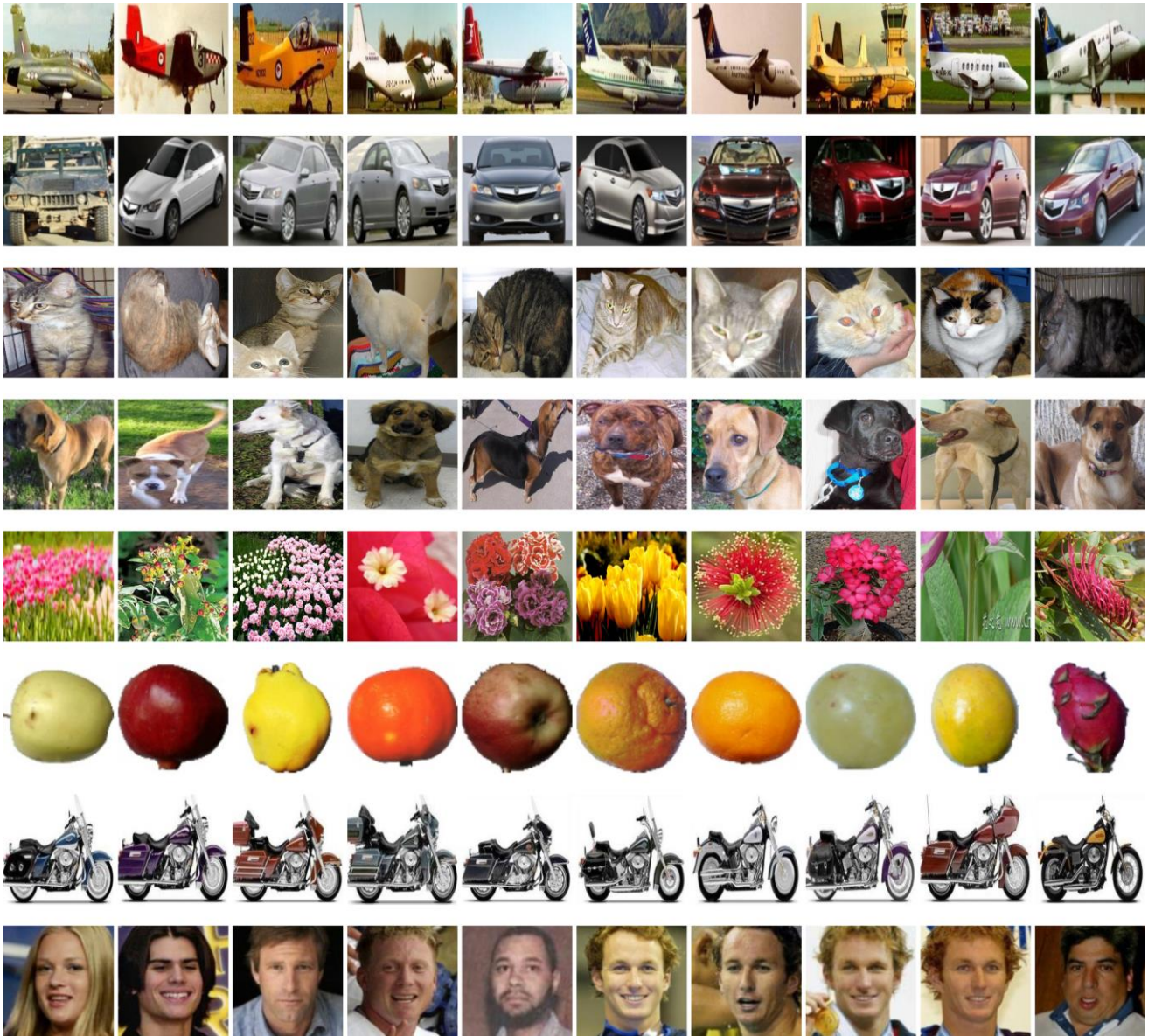


Рисунок 5.3 – Примеры изображений из набора Natural Images

Для расширения и проверки результатов исследования на сложном наборе данных был использован набор ImageNet-1k. ImageNet-1k [234] – большой набор данных (изображений), состоящий из ~1400000 изображений, маркированных в 1000 классов. Является подмножеством набора изображений ImageNet. Размерность изображений в наборе не стандартизирована, но все изображения представлены в 3 каналах. ImageNet-1k – широко используемый набор данных для тестирования систем автоматизированной локализации и классификации

изображений, поскольку является достаточно сложным набором с точки зрения наборов признаков и разнообразия классов.

Набор данных Rock-Paper-Scissors (RPS) Images [235] содержит изображения жестов рук из игры "Камень-ножницы-бумага". Изображения были получены в рамках разрабатываемого проекта по созданию игры "Камень-ножницы-бумага" с использованием компьютерного зрения и машинного обучения. Набор данных содержит 2188 изображений, соответствующих жестам "Камень" (726 изображений), "Бумага" (710 изображений) и "Ножницы" (752 изображения). Все изображения сделаны на зеленом фоне с относительно одинаковой освещенностью и балансом белого. Все изображения представляют собой RGB-изображения размером 300 пикселей в ширину на 200 пикселей в высоту в формате .png.

### 5.2.2 Свёрточные нейронные сети

В данном исследовании были использованы две основные архитектуры СНС:

1. упрощённая сеть высокого быстродействия, описанная ниже в тексте;
2. широко известная EfficientNetB3 [103].

Результаты первых экспериментов были получены нами на упрощённой сети высокого быстродействия. Сеть содержит всего 914,960 параметров, что довольно мало по сравнению с современными СНС. Это позволит проводить относительно короткие тесты в ущерб общей точности классификации (рисунок 5.4). Простая СНС тестируется на нескольких небольших наборах данных, поскольку ее обобщающая способность крайне ограничена. Эту простую СНС используется для подтверждения переносимости результатов на различные наборы данных.

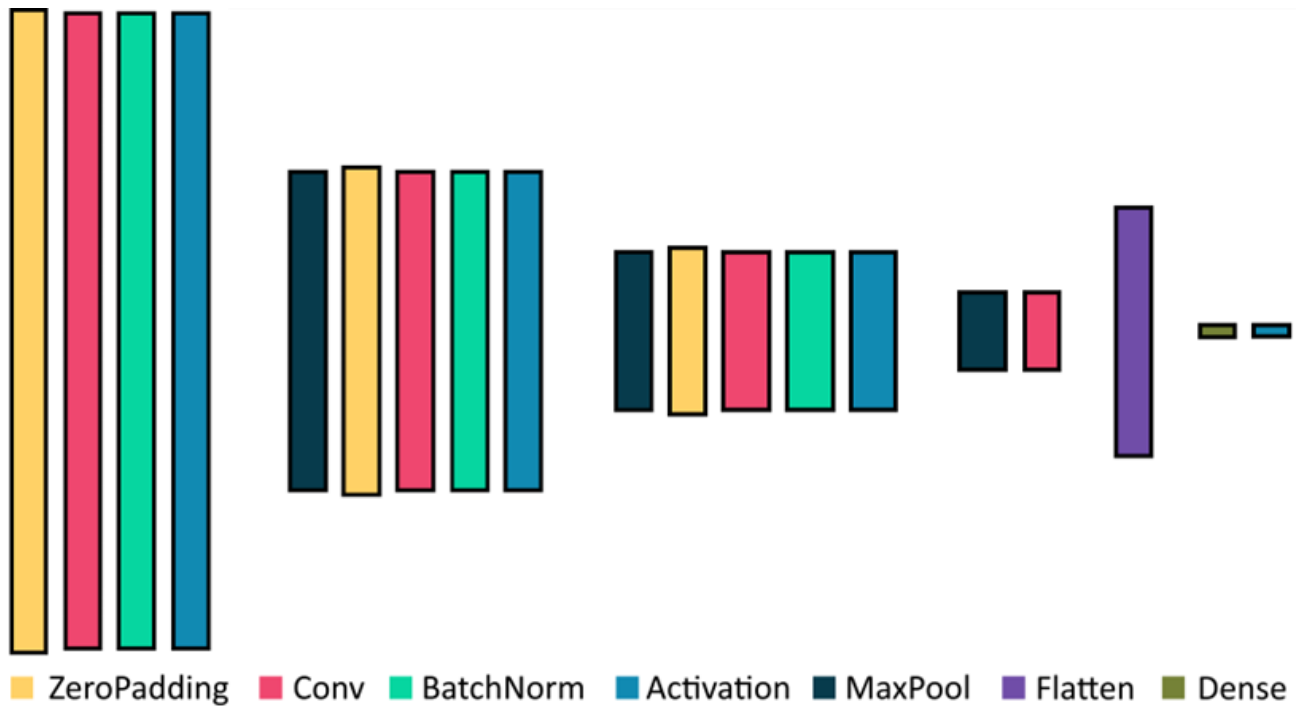


Рисунок 5.4 – Архитектура использованной свёрточной нейронной сети

Для расширения исследования и получения более репрезентативных данных использована СНС, называемая EfficientNet [103]. Авторы обратили внимание, что при разработке новых архитектур СНС уделяется недостаточное внимание балансировке разрешения, ширины и глубины, а также указали на важность такой балансировки. В своём исследовании авторы предложили эффективный метод комбинированного масштабирования СНС до любого размера. При на порядки меньшем количестве параметров и времени обучения по сравнению с многими современными архитектурами сетей, архитектура EfficientNetB3 достигает более высоких результатов точности классификации Top-1 на различных наборах данных. Так как в данном исследовании используются результаты большого числа тестов, для ограничения времени и вычислительных ресурсов, затраченных на эксперименты, было принято решение использовать архитектуру EfficientNetB3, что позволит анализировать сложные наборы изображений с приемлемым уровнем точности.



### 5.2.3 Состязательные атаки

FGSM (Fast Gradient Sign Method) – на данный момент один из самых распространённых методов построения состязательной атаки [25]. Суть метода заключается в добавлении к исходному изображению некоторого неслучайного вектора, направление которого совпадает с градиентом функции потерь. Добавочный вектор FGSM можно представить как:

$$\eta = \varepsilon \cdot \text{sgn}(\nabla_x J(\theta, x, y)), \quad (12)$$

где  $\theta$  – параметры атакуемой модели нейронной сети;

$x$  – входной вектор (изображение);

$y$  – истинный класс вектора  $x$  (если есть);

$J(\theta, x, y)$  – функция потерь модели нейронной сети;

$\varepsilon$  – коэффициент, выбираемый эмпирически;

$\nabla_x$  – градиент в пространстве изображения;

$\text{sgn}$  - функция знака;

$\eta$  - вектор состязательной атаки.

Такой состязательный вектор выглядит для восприятия человека как высокочастотный шум малой интенсивности, не влияющий на возможность распознавания объекта на изображении. В то же время этот шум крайне эффективно снижает точность классификации примеров нейронными сетями. Интенсивность атаки устанавливается множителем и выбирается таким образом, чтобы изменение распознаваемого изображения было минимальным, но достаточным для осуществления эффективной атаки. На некоторых современных моделях СНС атаку получается провести таким образом, что изменения на искаженном изображении незаметны для человека (рисунок 5.5).

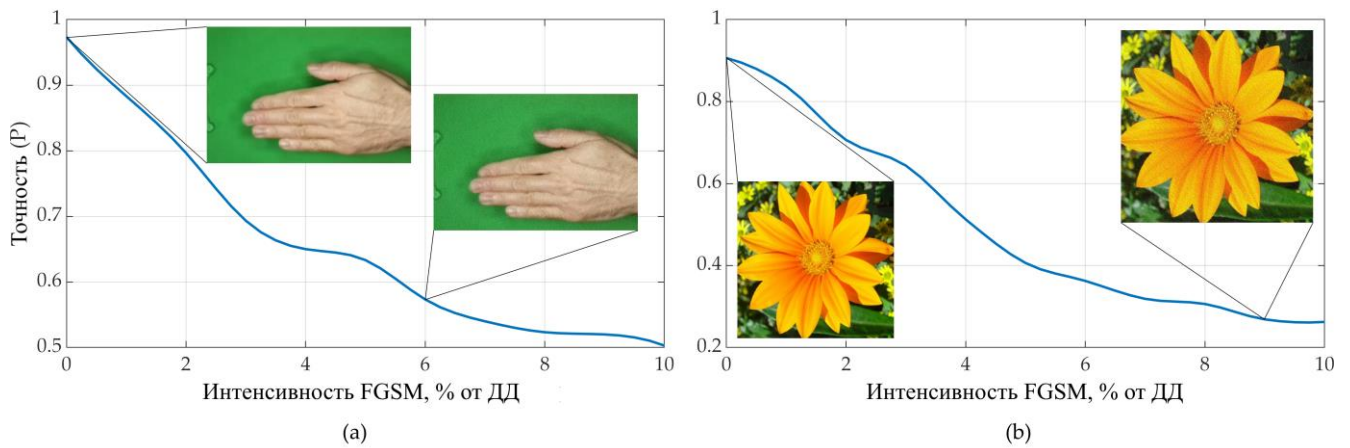


Рисунок 5.5 – Влияние FGSM на качество распознавания наборов изображений  
(a) набора Rock-Paper-Scissors Images и (b) Natural Images

Несмотря на то, что FGSM является одним из первых алгоритмов построения состязательных атак на нейронные сети, он до сих пор считается одним из наиболее эффективных, являясь при этом простым с точки зрения реализации и временных затрат. Более сложным вариантом FGSM является алгоритм PGD (projected gradient descent). Суть алгоритма PGD сводится к итерационному применению алгоритма FGSM для повышения эффективности атаки [236]. Многие другие алгоритмы состязательных атак также основаны на FGSM [237]. Необходимо заметить, что предлагаемые в данной работе способы противодействия высокочастотным шумам могут быть малоэффективны против таких видов низкочастотных состязательных искажений, как атаки в физическом пространстве [238], Square attack [239], однако достаточно эффективны против таких высокочастотных искажений, как PGD [236], C&W attack [240], Zeroth Order Optimization (ZOO) [241], HopSkipJumpAttack (HSJA) [242] и DeepFool [149].

### **5.3 Разработанный метод противодействия высокочастотным искажениям**

Важной особенностью СНС для распознавания изображений является слабая восприимчивость сетей к масштабу (размеру) объектов на изображении. Это обстоятельство обуславливает практически одинаковое влияние на распознавание как низкочастотных, так и высокочастотных компонент изображения. В этом состоит принципиальное отличие работы современных искусственных СНС от человеческого восприятия. Авторы [243] исследовали влияние различных компонент частотного спектра изображений на работу СНС и показали, что высокочастотные компоненты изображений являются причиной многих уязвимостей в работе СНС, в том числе и состязательных атак. При этом человеческое зрение невосприимчиво для высокочастотных компонент изображений [244]. При этом авторы [243] показывают, что некоторые фильтры, часто используемые при обучении и работе сетей, могут усугублять данную проблему. Также авторы [243] показывают, что более состязательно-устойчивые нейронные сети, как правило, имеют менее резкие градиенты в свёрточных ядрах (фильтрах), составляющих эти сети.

Большинство алгоритмов построения состязательных атак используют именно эту уязвимость СНС [245]. Известны исследования, направленные на обнаружение наличия состязательных атак, основанные на анализе спектра изображений [246], [247]. Для противодействия и защиты системы распознавания изображений от состязательных атак, основанных на высокочастотных искажениях, эффективным является применение фильтров нижних частот, таких как фильтр Гаусса. При применении подобной фильтрации высокочастотные компоненты изображений будут утеряны, однако общая структура изображения, положение объектов интереса и их формы остаются различимы (рисунок 5.6).

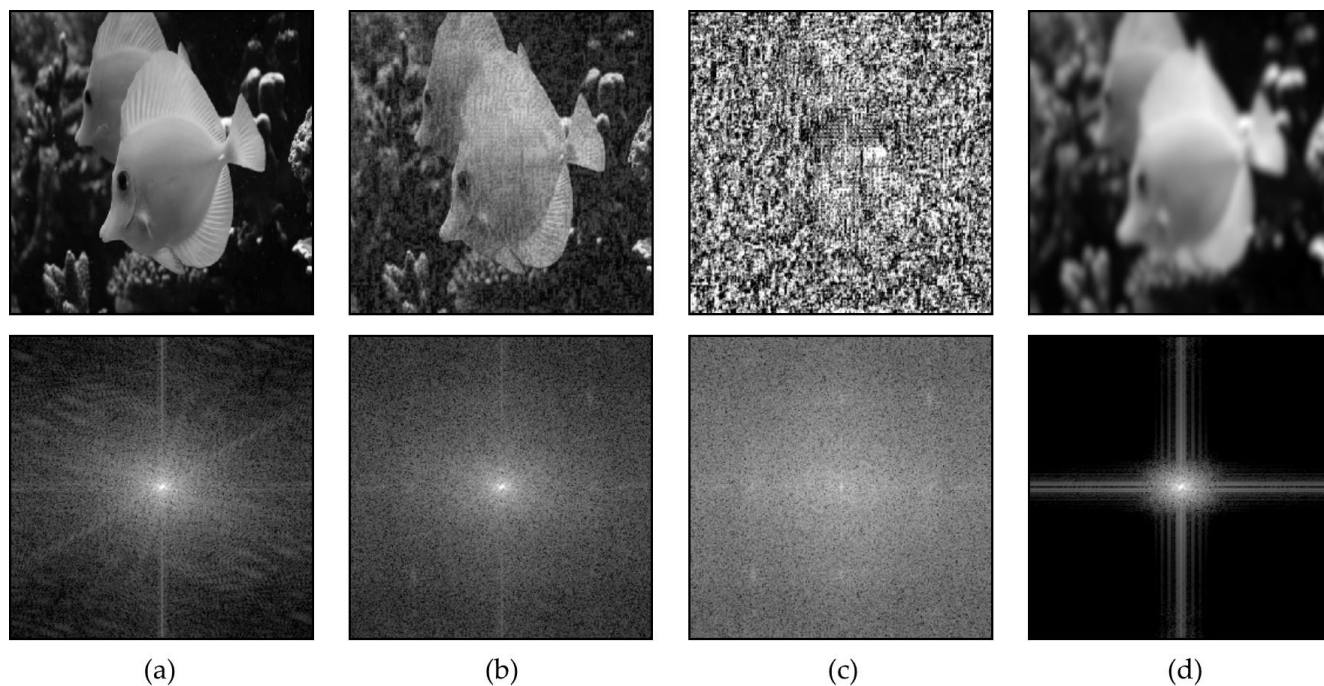


Рисунок 5.6 – Двумерное преобразование Фурье изображения (a) без FGSM; (b) с FGSM (10%); (c) Только FGSM; (d) с FGSM, с применением гауссовского фильтра нижних частот

На рисунке 5.6 представлены двумерные спектры (декартово преобразование Фурье) изображения. Из рисунка видно, что атака FGSM размывает спектр изображения, а фильтр низких частот ограничивает этот спектр, приближая его к исходному. В качестве другого примера уменьшения влияния атак противника на изображение возможно рассмотреть его с точки зрения профиля яркости. На рисунке 5.7 представлены одномерные профили яркости изображения, FGSM с 10 % динамического диапазона изображения, изображения противника и размытого изображения противника.

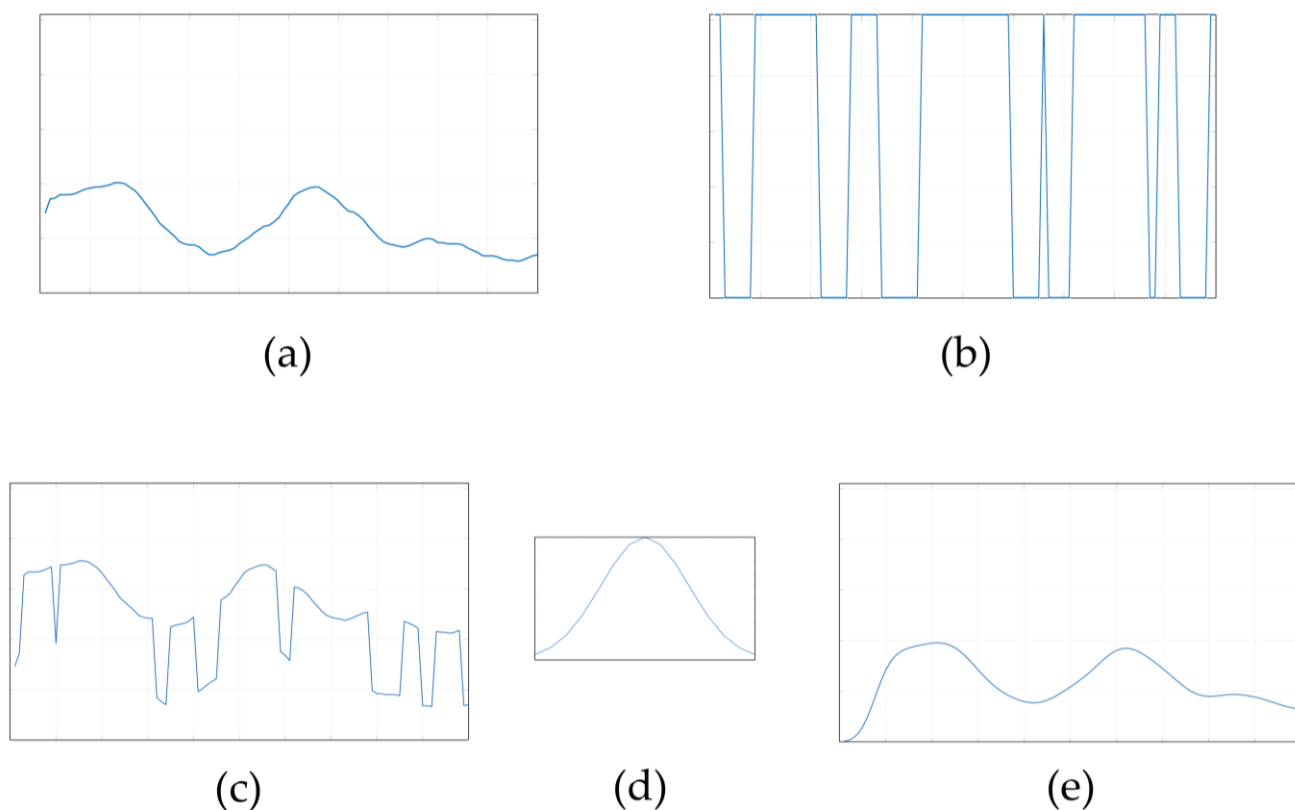


Рисунок 5.7 – Влияние размытия по Гауссу на содержание изображения:  
 (a) профиль яркости исходного изображения, выровненный по одной линии,  
 (b) профиль яркости FGSM той же размерности, (c) состязательное изображение  
 (изображение + 0,1 FGSM), (d) импульсная характеристика фильтра Гаусса, (e)  
 свертка состязательного изображения и импульсной характеристики фильтра  
 Гаусса

Как видно из рисунка 5.7, состязательные атаки сильно влияют на профиль яркости, делая его неузнаваемым. В то же время фильтрация по Гауссу, выполненная поверх состязательного изображения, восстанавливает профиль яркости изображения, приближая его к исходному. Чтобы подтвердить гипотезу об эффективности низкочастотной фильтрации для преодоления состязательной атаки, произведён анализ влияния размытия по Гауссу на изображение и структуру матрицы атаки. Красная кривая на рисунке 5.8 показывает зависимость скалярного произведения размытого и исходного изображения от размера фильтра Гаусса. Синяя кривая на рисунке 5.8 показывает зависимость скалярного

произведения размытой и исходной матрицы атаки от размера фильтра Гаусса. Скалярное произведение двух изображений (представленных в виде векторов) можно рассматривать как меру сходства или корреляцию. Векторы со схожими направлениями и величинами будут давать большее скалярное произведение, а меньшее скалярное произведение указывает на ортогональность векторов.

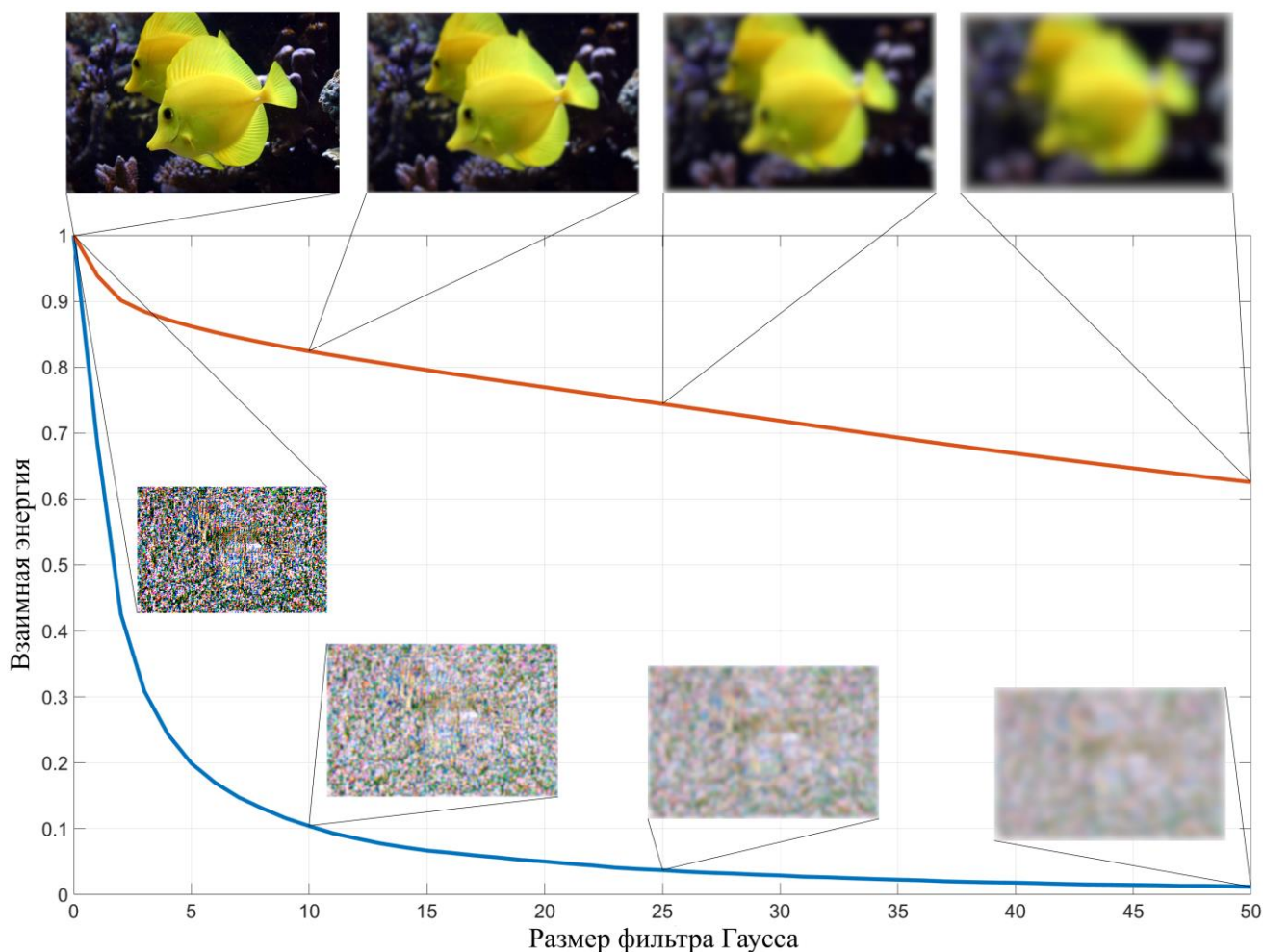


Рисунок 5.8 – Значения взаимной энергии изображения и вектора атаки

Как видно из рисунка 5.8, значения функции взаимной корреляции исходной и размытой матрицы атаки при применении фильтра Гаусса снижается быстрее, чем значения функции взаимной корреляции размытого и исходного изображения. При размере окна размытия (среднеквадратического отклонения фильтра Гаусса) более 10 пикселей, размытая и исходная матрицы атаки

становятся практически некоррелированы. Поскольку матрица атаки является целевой функцией (каждый пиксель не является случайным), результативность атаки на нейронную сеть с ростом ядра фильтра Гаусса будет снижаться значительно, чем качество распознавания изображения.

#### **5.4 Постановка эксперимента**

Блок-схема предложенного алгоритма обработки изображений для противостояния состязательным атакам приведена на рисунке 5.9. Поскольку метод предполагает размытие тестовых изображений, важное значение приобретает обучение нейронной сети на размытых изображениях. СНС предварительно обучается на аугментированных данных [67], [75]. Этот подход эффективен, поскольку реализация фильтра Гаусса не требует больших вычислительных затрат. Процедура аугментации использует только один вычислительно простой фильтр Гаусса. Обучение не требует сложных с вычислительной точки зрения алгоритмов состязательных атак для дополнения данных. Обучение нейронной сети выполняется за один подход. Исходный обучающий набор данных разбивается на две части. Одна часть осталась неизменной, вторая была размыта с помощью фильтра Гаусса, размер которого был выбран случайным образом в диапазоне от 0 до 0,1 от размера изображения.

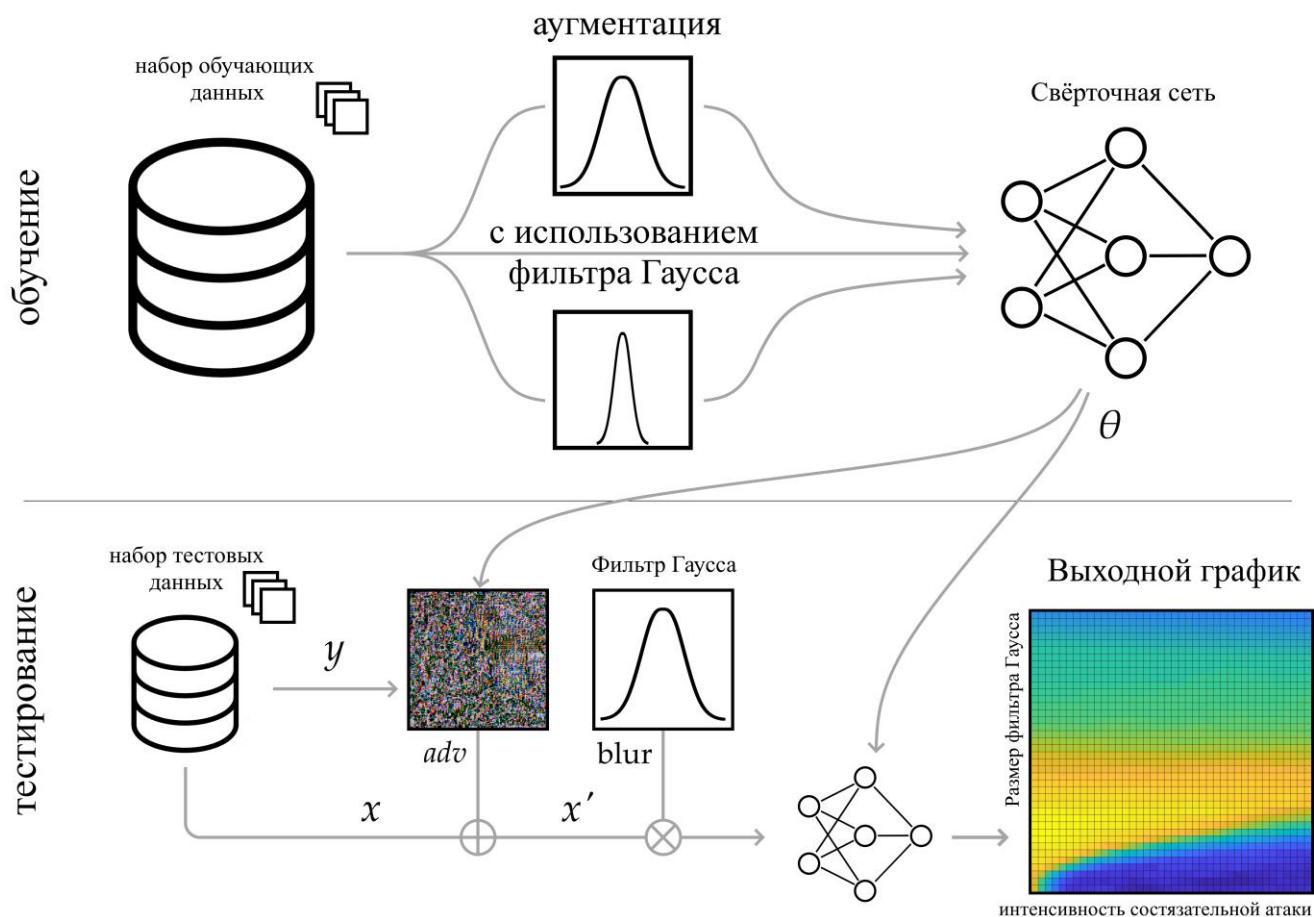


Рисунок 5.9 – Схема разработанного метода

На этапе тестирования к изображениям добавлены векторы FGSM различной интенсивности. После этого состязательные изображения были отфильтрованы с помощью фильтра Гаусса. Обученная нейронная сеть использовалась для распознавания размытых состязательных изображений. Высокочастотный компонент изображения включает в себя состязательную атаку, другие высокочастотные шумы (например, импульсный или тепловой шум для естественных изображений) и мелкие паттерны изображения. Гауссовский фильтр значительно снижает влияние высокочастотной составляющей изображения. При этом общая структура изображения ухудшается гораздо менее значительно. Эта техника представляет собой компромисс между общей точностью распознавания и точностью распознавания состязательных изображений. Первый показатель снижается незначительно, а второй



значительно возрастает. В данном разделе проводится большое количество тестов по распознаванию изображений с широким диапазоном интенсивностей FGSM и размеров фильтра Гаусса. Это позволяет получить трехмерные графики зависимости точности распознавания изображений от интенсивности FGSM и размера фильтра Гаусса, а также определить оптимальный размер фильтра Гаусса для распознаваемых изображений.

## 5.5 Результаты работы предложенного метода

С использованием метода, схема которого показана на рисунке 5.9, получены результаты распознавания тестовых данных для разных нейронных сетей. На следующих графиках (рисунок 5.10) показана зависимость точности распознавания изображений от интенсивности атаки FGSM и от размера фильтра Гаусса. Интенсивность атаки FGSM далее измеряется в процентах от динамического диапазона изображения (DR). Размеры фильтра Гаусса измеряются в процентах от размера исходного изображения.

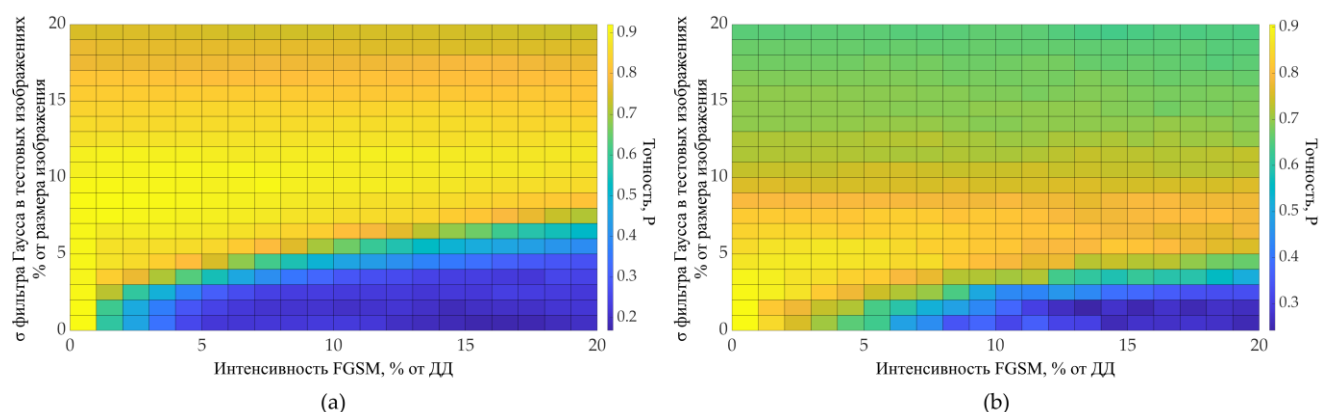


Рисунок 5.10 – Точность распознавания набора изображений Natural Dataset с применением простой СНС: (а) СНС обучена с применением аугментированного фильтром Гаусса набора данных; (б) обучение проводилось без аугментации

Для получения данных графиков проведено по 441 (1681 для некоторых тестов с использованием простой архитектуры СНС) независимых экспериментов по распознаванию тестового набора данных с внесением состязательных искажений с различной интенсивностью и последующей фильтрацией по Гауссу. Общее число независимых экспериментов по распознаванию тестовых наборов данных, представленных на рисунках 5.10-5.12, равняется 2646. Как видно из рисунка 5.10, точность распознавания изображений при наличии состязательных искажений быстро падает с ростом интенсивности этих искажений. При интенсивности состязательных искажений, равной 4-5% от динамического диапазона изображения (в случае, показанном на рисунке 5.10, (а)) точность распознавания падает до уровня случайного ответа. Однако при обработке состязательных тестовых изображений с применением фильтра Гаусса точность возрастает. При дальнейшем увеличении размеров фильтра важные для распознавания признаки изображения утрачиваются, и точность распознавания падает. На рисунке 5.10 заметно, что с повышением интенсивности состязательных искажений требуется больший размер фильтра Гаусса, но точность распознавания изображений не достигает максимально возможных значений (полученной при отсутствии состязательных искажений), однако приближается к нему. При дальнейшем увеличении интенсивности состязательных искажений (снижении отношения сигнал-шум) применение фильтра Гаусса становится малоэффективным. Также из рисунка 5.10 видно, что оптимальное значение размеров фильтра Гаусса зависит от интенсивности состязательных искажений, а также от характеристик используемых данных и нейронной сети, что показано на рисунках 5.11-5.12. Например, обучение СНС для распознавания набора данных Rock-Paper-Scissors с применением аугментации (размытыми изображениями) показало высокую эффективность при низких значениях интенсивности состязательных искажений в тестовых данных (< 3 % от динамического диапазона изображения). При дальнейшем увеличении

интенсивности состязательных искажений большой выигрыш был получен сетью, обученной без применения аугментации (рисунок 5.11).

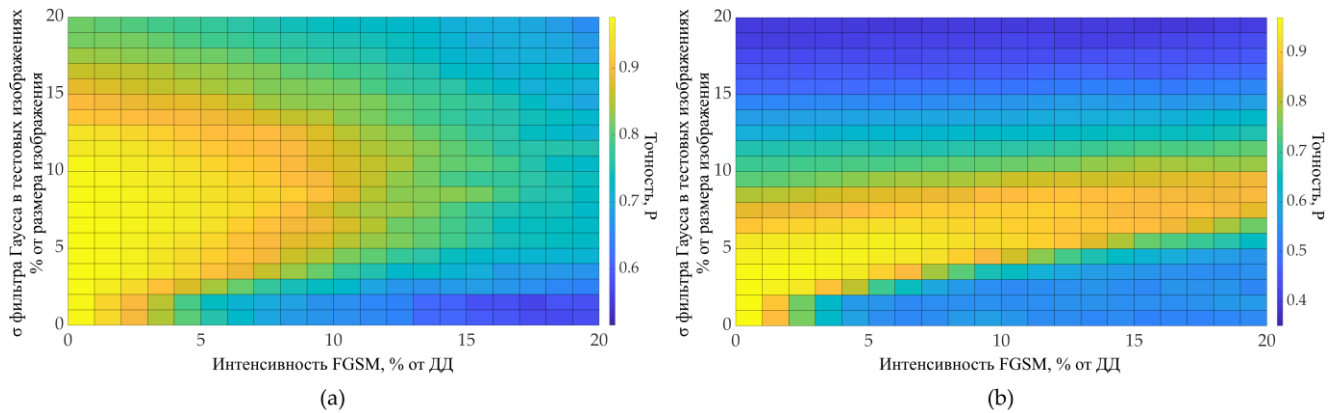


Рисунок 5.11 – Точность распознавания набора изображений Rock-Paper-Scissors Dataset с применением простой СНС: (а) СНС обучена с применением аугментированного фильтром Гаусса набора данных; (б) обучение проводилось без аугментации

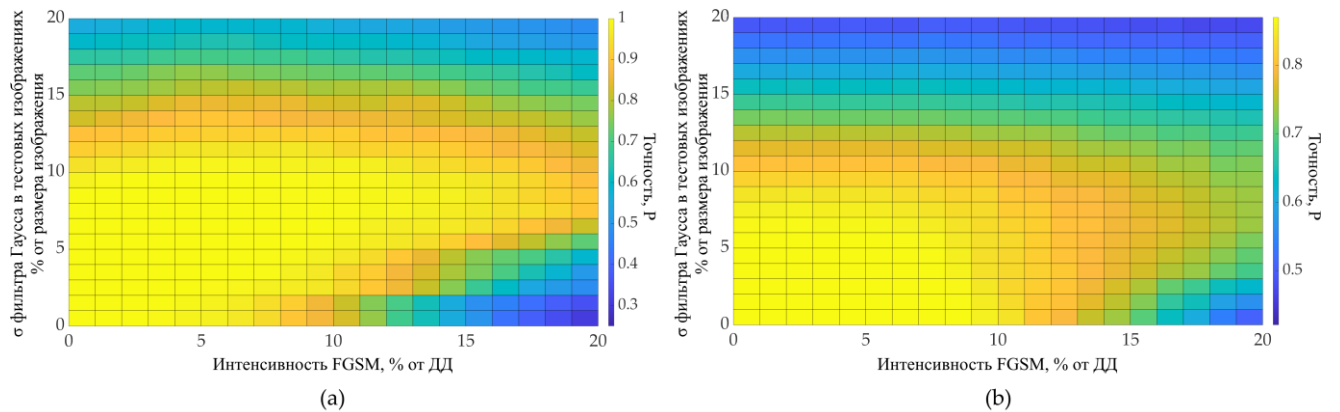


Рисунок 5.12 – Точность распознавания сетью EfficientNetB3: (а) Natural Dataset; (б) ImageNet

Поскольку в практически важных случаях используется интенсивность состязательной атаки, не превышающая 5% от динамического диапазона исходных изображений, использование аугментации изображений при обучении нейронных сетей даёт преимущество в точности распознавания при более

широком разбросе значений размера фильтра Гаусса. Полученные результаты являются репрезентативными для более сложных архитектур СНС. В данном разделе проведены эксперименты с использованием алгоритма, показанного на рисунке 5.9 для сети EfficientNetV3 с применением наборов изображений Natural и ImageNet. Для обучения сети EfficientNet V3 использовался аугментированный набор данных ImageNet (аугментация с применением фильтра Гаусса), без использования Transfer Learning.

В таблице 5.1 показано значение точности классификации при различной интенсивности состязательных искажений и возможный выигрыш при применении фильтрации. Оптимальный размер фильтра выбирался, исходя из максимизации точности распознавания при различных значениях интенсивности состязательной атаки.

$$\sigma_{opt} = \arg \left( \max \left( \sum_{I_{FGSM}=0}^{I_{FGSM}^{max}} P_{LPF}(\sigma, I_{FGSM}) \right) \right) \quad (13)$$

, где  $\sigma_{opt}$  – оптимальный размер фильтра,

$P_{LPF}$  – точность распознавания, полученная с применением низкочастотного фильтра Гаусса,

$I_{FGSM}$  – интенсивность состязательной атаки,

$I_{FGSM}^{max}$  – максимальное значение интенсивности состязательной атаки.

Выигрыш в точности  $G$  рассчитан с использованием следующей формулы:

$$G = \frac{(1 - P_{no LPF})}{(1 - P_{LPF})} \quad (14)$$

, где  $G$  – выигрыш в точности,

$P_{no LPF}$  – точность распознавания, полученная без применения низкочастотного фильтра Гаусса,

$P_{LPF}$  – точность распознавания, полученная с применением низкочастотного фильтра Гаусса оптимального размера.

Таблица 5.1 – Точность классификации при различных интенсивностях искажений и возможный выигрыш в точности при использовании фильтра.

Нейронная сеть и Набор данных	Интенсивность FGSM	Точность распознавания с FGSM и без LPF $P_{no\ LPF}$	Точность распознавания с FGSM и LPF $P_{LPF}$	Оптимальный размер НЧ фильтра	Выигрыш в точности G
Сеть высокого быстродействия (Natural Dataset)	5	0,206	<b>0,913</b>	10	9,1
	10	0,206	<b>0,900</b>		7,9
	20	0,188	<b>0,894</b>		6,7
Сеть высокого быстродействия (RPS)	5	0,738	<b>0,947</b>	8	4,9
	10	0,660	<b>0,879</b>		2,8
	20	0,576	<b>0,738</b>		1,6
EfficientNet (ImageNet)	15	0,699	<b>0,781</b>	7	1,4
	20	0,481	<b>0,720</b>		1,9
EfficientNetB3 (Natural Dataset)	5	0,977	<b>1,000</b>	7	$\infty$
	10	0,814	<b>0,996</b>		46,5
	20	0,250	<b>0,881</b>		6,3

Выигрыш в точности  $G$  показывает относительное снижение частоты ошибок распознавания при использовании низкочастотной фильтрации по сравнению с использованием СНС без фильтра.

Предложенный подход является вычислительно эффективным, поскольку требует лишь простой аугментации обучающего набора изображений, выполняемой один раз перед обучением, а также простой фильтрации изображений перед распознаванием. Время, затрачиваемое на фильтрацию,

зависит от размеров изображения. При использовании простых СНС, таких как описанная выше сеть высокого быстродействия, время, затрачиваемое на фильтрацию, составляет менее 0,4% от общего времени распознавания изображений. Для сложных сетей, таких как EfficientNetB3, относительные затраты времени на фильтрацию изображений составляют 0,25%.

При выборе размеров окна размытия стоит учитывать несколько параметров, таких как разрешение изображения, характеристики нейронной сети. Слишком высокое значение размера окна размытия может исказить важные для классификации характеристики объекта, тем самым снизив общее качество работы нейросетевого алгоритма. В настоящей работе показан метод выбора оптимального размера фильтра.

## **Выводы по разделу 5**

1. Предложен простой в реализации метод повышения устойчивости глубоких свёрточных нейронных сетей к высокочастотным искажениям, в том числе высокочастотным состязательным атакам. До сих пор не существует комплексного исследования эффективности фильтрации низких частот для противодействия высокочастотным атакам. Предлагаемый метод позволяет повысить устойчивость глубоких свёрточных нейронных сетей к неблагоприятным воздействиям. Метод основан на низкочастотной фильтрации изображений и их распознавании сетью, предварительно обученной распознаванию размытых изображений. Показан метод выбора оптимального размера фильтра.

2. Показано, что влияние состязательной атаки с увеличением граничной частоты фильтра Гаусса снижается быстрее, чем качество исходного изображения. Таким образом, размен влияния состязательной атаки на размытие изображения оказывается эффективным. Предварительное обучение нейронной

сети распознаванию размытых изображений является важной частью предложенного метода, поскольку позволяет снизить влияние размытия изображений на качество их распознавания.

3. Выигрыш в точности  $G$ , достигаемый при использовании предложенного метода, в любом случае составляет не менее 1,4, средний выигрыш в точности составляет  $G=8,8$  раз (без учета EfficientNetB3, оцененного на ImageNet, и интенсивности FGSM  $I_{FGSM} = 5$ , где выигрыш принимает бесконечное значение за счет отсутствия ошибок распознавания при использовании низкочастотной фильтрации).

4. Предложенный метод ввиду его высокой эффективности и низкой сложности может быть использован в различных системах распознавания изображений и технического зрения, реализованных на различных аппаратных платформах, в том числе и с крайне ограниченными вычислительными ресурсами. В то же время следует отметить, что предложенная методика может оказаться неэффективной против низкочастотных состязательных атак. В дальнейших исследованиях целесообразно расширить изучение поведения свёрточной нейронной сети с точки зрения предварительной обработки изображений. Это исследование будет включать более широкие наборы современных архитектур свёрточных нейронных сетей, в том числе сети локализации. Кроме того, будут проведены тесты для различных видов состязательных атак (BIM, PGD, CW, низкочастотные атаки и т. д.). Хорошим направлением для будущих исследований может стать изучение эффективности предложенного метода против доменных сдвигов. Также будут рассмотрены более широкие наборы фильтров, включая медианные фильтры, режекторные фильтры и т. д.

## ЗАКЛЮЧЕНИЕ

Основные результаты диссертационной работы сведены к следующему.

1. Разработанные математические модели генерации изображений и обучения-тестирования свёрточной сети позволяют точно оценить характеристики устойчивости СНС к искажениям. Выявлена зависимость точности распознавания от меры неопределённости в тестовом наборе данных. Для корректно работающей модели при увеличении неопределённости в тестовых данных точность распознавания монотонно убывает. При внесении чрезмерных искажений в обучающий набор проявляется неоптимальность обучения.

2. Проанализирована точность распознавания множества наборов данных с различными значениями неопределённости и получена зависимость точности распознавания от интенсивности искажений в обучающем наборе данных. Существование оптимальной интенсивности искажений в обучающем наборе данных было предположено и доказано для различных типов изображений и шумов. Показано, что определение этого оптимума может быть выполнено с помощью статистического моделирования. Полученные результаты применимы к СНС с распространёнными структурами и различным типам искажений в данных. Использование обучающего набора данных с оптимальным значением неопределённости позволяет снизить вероятность ошибки распознавания в среднем в 20 раз по сравнению с использованием исходного набора изображений без дополнительных искажений.

3. Получена зависимость точности распознавания от интенсивности размытия по Гауссу для нейронных сетей, обученных с использованием различных методов аугментации. Доказано, что существует оптимальный способ аугментации обучающего набора данных, позволяющий снизить вероятность ошибки в среднем в 2,5 раза по сравнению с использованием исходного набора изображений без дополнительных искажений.

4. Предложен метод повышения устойчивости глубоких свёрточных нейронных сетей к высокочастотным атакам. Показано, что влияние



сопоставительной атаки с увеличением граничной частоты фильтра Гаусса снижается быстрее, чем качество исходного изображения. Выигрыш в точности, достигаемый при использовании предложенного метода, в любом случае составляет не менее 1,4, средний выигрыш в точности составляет 8,8 раз.

Таким образом, цель диссертационной работы достигнута, научная задача разработки метода оптимальной аугментации обучающих изображений, обеспечивающего повышение точности распознавания тестовых изображений при наличии в них искажений различной физической природы, решена.

Дальнейшая работа будет посвящена расширению результатов на различные структуры нейронных сетей и различные задачи (например, обнаружение объектов). Также возможны исследования по поиску аналитического решения для задачи определения оптимального значения неопределенности обучающего набора данных без массивного статистического моделирования.

**СПИСОК СОКРАЩЕНИЙ И УСЛОВНЫХ ОБОЗНАЧЕНИЙ**

ИНС – Искусственная Нейронная Сеть;

НИР – Научно-Исследовательские Работы;

СНС – Свёрточная Нейронная Сеть;

CNN – «Convolutional neural network»;

DR – Dynamic Range – Динамический диапазон;

EM – «Expectation-maximization» – «ожидаемое-максимальное»,  
«максимизация ожидаемого»;

FGSM – Fast Gradient Sign Method;

GAN – Generative Adversarial Networks – Генеративно-сопоставительных  
сетей;

LPF – Low-Pass Filter – Низкочастотный фильтр;

LS-SVM – Least-Squares Support-Vector Machine – Машина опорных  
векторов наименьших квадратов;

LSTM – Long Short-Term Memory – Длинная цепь элементов краткосрочной  
памяти;

SSIM – Structural Similarity Index Measure – Мера индекса структурного  
сходства.

## СПИСОК ТЕРМИНОВ

**Денойзер** – математическая модель, обеспечивающая повышение отношения сигнал/шум на изображениях.

**Дропаут** – удаление случайных частей изображения с целью снижения влияния определённых признаков на решение сети.

**Неопределённость ( $U$ )** – отношение интенсивности искажённой составляющей к интенсивности информативной составляющей в изображении. При этом искажённая составляющая может иметь различную физическую природу (шум, размытие, случайный разброс расположения точек и т. д.).

**Оптимальность** – решение, удовлетворяющее одновременно двум следующим условиям:

1) обеспечение максимального значения интегральной точности распознавания;

2) сохранение монотонности функции зависимости точности распознавания от неопределённости в тестовых данных.

**Помехоустойчивость** – свойство модели нейронной сети, характеризующее зависимость точности распознавания изображений от интенсивности искажений различной природы.

**Точность распознавания** – отношение количества правильно распознанных изображений к общему количеству изображений в некотором наборе.

**Оценка вероятности ошибки распознавания** – единица минус точность распознавания.

**Робастность** – то же, что и **Помехоустойчивость**.

**СПИСОК ЛИТЕРАТУРЫ**

1. Deng L. et al. Recent advances in deep learning for speech research at Microsoft // 2013 IEEE International Conference on Acoustics, Speech and Signal Processing. 2013. P. 8604–8608.
2. Jang J.-W. et al. Sparsity-Aware and Re-configurable NPU Architecture for Samsung Flagship Mobile SoC // 2021 ACM/IEEE 48th Annual International Symposium on Computer Architecture (ISCA). 2021. P. 15–28.
3. Jacob B. et al. Quantization and Training of Neural Networks for Efficient Integer-Arithmetic-Only Inference. 2018. P. 2704–2713.
4. Siddegowda S. et al. Neural Network Quantization with AI Model Efficiency Toolkit (AIMET): arXiv:2201.08442. arXiv, 2022.
5. Rueckauer B. et al. NxTF: An API and Compiler for Deep Spiking Neural Networks on Intel Loihi // J. Emerg. Technol. Comput. Syst. 2022. Vol. 18, № 3. P. 48:1-48:22.
6. Dias F.M., Antunes A., Mota A.M. Artificial neural networks: a review of commercial hardware // Engineering Applications of Artificial Intelligence. 2004. Vol. 17, № 8. P. 945–952.
7. Hinton G. et al. Deep Neural Networks for Acoustic Modeling in Speech Recognition: The Shared Views of Four Research Groups // IEEE Signal Processing Magazine. 2012. Vol. 29, № 6. P. 82–97.
8. Zhang Z. et al. Deep Learning for Environmentally Robust Speech Recognition: An Overview of Recent Developments: arXiv:1705.10874. arXiv, 2018.
9. Khan S. et al. A Guide to Convolutional Neural Networks for Computer Vision. Cham: Springer International Publishing, 2018.
10. Bishop C.M. Neural networks for pattern recognition. Reprinted. Oxford: Oxford University Press, 2010. 482 p.
11. Krizhevsky A., Sutskever I., Hinton G.E. ImageNet classification with deep convolutional neural networks // Commun. ACM. 2017. Vol. 60, № 6. P. 84–90.

12. Hong P., Wen Z., Huang T.S. Real-time speech-driven face animation with expressions using neural networks // *IEEE Transactions on Neural Networks*. 2002. Vol. 13, № 4. P. 916–927.
13. Kavitha B.R., Srimathi C. Benchmarking on offline Handwritten Tamil Character Recognition using convolutional neural networks // *Journal of King Saud University - Computer and Information Sciences*. 2022. Vol. 34, № 4. P. 1183–1190.
14. Du S. et al. Automatic License Plate Recognition (ALPR): A State-of-the-Art Review // *IEEE Transactions on Circuits and Systems for Video Technology*. 2013. Vol. 23, № 2. P. 311–325.
15. Liu B. et al. Identification of Apple Leaf Diseases Based on Deep Convolutional Neural Networks // *Symmetry*. 2017. Vol. 10, № 1. P. 11.
16. Lu Y. et al. Identification of rice diseases using deep convolutional neural networks // *Neurocomputing*. 2017. Vol. 267. P. 378–384.
17. Beritelli F. et al. Automatic heart activity diagnosis based on Gram polynomials and probabilistic neural networks // *Biomed. Eng. Lett*. 2018. Vol. 8, № 1. P. 77–85.
18. Deepak S., Ameer P.M. Brain tumor classification using deep CNN features via transfer learning // *Computers in Biology and Medicine*. 2019. Vol. 111. P. 103345.
19. Jain N. et al. Hybrid deep neural networks for face emotion recognition // *Pattern Recognition Letters*. 2018. Vol. 115. P. 101–106.
20. Li K. et al. Facial expression recognition with convolutional neural networks via a new face cropping and rotation strategy // *Vis Comput*. 2020. Vol. 36, № 2. P. 391–404.
21. Ciaparrone G. et al. Deep learning in video multi-object tracking: A survey // *Neurocomputing*. 2020. Vol. 381. P. 61–88.
22. Egmont-Petersen M., de Ridder D., Handels H. Image processing with neural networks—a review // *Pattern Recognition*. 2002. Vol. 35, № 10. P. 2279–2301.
23. Singh D. et al. Classification of COVID-19 patients from chest CT images using multi-objective differential evolution-based convolutional neural networks // *Eur J Clin Microbiol Infect Dis*. 2020. Vol. 39, № 7. P. 1379–1389.

24. Khan A. et al. A survey of the recent architectures of deep convolutional neural networks // *Artif Intell Rev.* 2020. Vol. 53, № 8. P. 5455–5516.
25. Goodfellow I.J., Shlens J., Szegedy C. Explaining and Harnessing Adversarial Examples. arXiv, 2014.
26. Chaturvedi A., Garain U. Mimic and Fool: A Task-Agnostic Adversarial Attack // *IEEE Transactions on Neural Networks and Learning Systems.* 2021. Vol. 32, № 4. P. 1801–1808.
27. Girshick R. et al. Region-Based Convolutional Networks for Accurate Object Detection and Segmentation // *IEEE Transactions on Pattern Analysis and Machine Intelligence.* 2016. Vol. 38, № 1. P. 142–158.
28. Xiao Y., Pun C.-M., Liu B. Fooling deep neural detection networks with adaptive object-oriented adversarial perturbation // *Pattern Recognition.* 2021. Vol. 115. P. 107903.
29. Abdar M. et al. A review of uncertainty quantification in deep learning: Techniques, applications and challenges // *Information Fusion.* 2021. Vol. 76. P. 243–297.
30. Dolenko T.A. et al. Fluorescence diagnostics of oil pollution in coastal marine waters by use of artificial neural networks // *Appl. Opt.* 2002. Vol. 41, № 24. P. 5155.
31. Guzhva A., Dolenko S., Persiantsev I. Multifold Acceleration of Neural Network Computations Using GPU // *Artificial Neural Networks – ICANN 2009* / ed. Alippi C. et al. Berlin, Heidelberg: Springer Berlin Heidelberg, 2009. Vol. 5768. P. 373–380.
32. Burikov S.A. et al. Application of artificial neural networks to solve problems of identification and determination of concentration of salts in multi-component water solutions by Raman spectra // *Opt. Mem. Neural Networks.* 2010. Vol. 19, № 2. P. 140–148.
33. Schegolev A.E. et al. Superconducting Neural Networks: from an Idea to Fundamentals and, Further, to Application // *Nanotechnol Russia.* 2021. Vol. 16, № 6. P. 811–820.

34. Schegolev A.E. et al. Bio-Inspired Design of Superconducting Spiking Neuron and Synapse // *Nanomaterials*. 2023. Vol. 13, № 14. P. 2101.
35. Sidorenko A. et al. Base Elements for Artificial Neural Network: Structure Modeling, Production, Properties // *International Journal of Circuits, Systems and Signal Processing*. 2023. Vol. 17. P. 177–183.
36. Lebedev V. et al. Speeding-up Convolutional Neural Networks Using Fine-tuned CP-Decomposition: arXiv:1412.6553. arXiv, 2015.
37. Schegolev A. et al. Learning cell for superconducting neural networks // *Supercond. Sci. Technol.* 2021. Vol. 34, № 1. P. 015006.
38. Goodfellow I. et al. Generative Adversarial Nets // *Advances in Neural Information Processing Systems*. Curran Associates, Inc., 2014. Vol. 27.
39. Goodfellow I., Bengio Y., Courville A. *Deep learning*. Cambridge, Massachusetts: The MIT Press, 2016. 775 p.
40. Abadi M. et al. TensorFlow: Large-Scale Machine Learning on Heterogeneous Distributed Systems: arXiv:1603.04467. arXiv, 2016.
41. Srivastava N. et al. Dropout: A Simple Way to Prevent Neural Networks from Overfitting // *Journal of Machine Learning Research*. 2014. Vol. 15, № 56. P. 1929–1958.
42. Rumelhart D.E., Hinton G.E., Williams R.J. Learning representations by back-propagating errors // *Neurocomputing: foundations of research*. Cambridge, MA, USA: MIT Press, 1988. P. 696–699.
43. Graves A., Mohamed A., Hinton G. Speech Recognition with Deep Recurrent Neural Networks: arXiv:1303.5778. arXiv, 2013.
44. Hinton G.E. et al. Improving neural networks by preventing co-adaptation of feature detectors: arXiv:1207.0580. arXiv, 2012.
45. LeCun Y., Bengio Y. Convolutional networks for images, speech, and time series // *The handbook of brain theory and neural networks*. Cambridge, MA, USA: MIT Press, 1998. P. 255–258.
46. LeCun Y. et al. Efficient BackProp // *Neural Networks: Tricks of the Trade* / ed. Orr G.B., Müller K.-R. Berlin, Heidelberg: Springer, 1998. P. 9–50.

47. Ioffe S., Szegedy C. Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift: arXiv:1502.03167. arXiv, 2015.
48. Szegedy C. et al. Going deeper with convolutions // 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Boston, MA, USA: IEEE, 2015. P. 1–9.
49. Erhan D. et al. Scalable Object Detection using Deep Neural Networks. 2014. P. 2147–2154.
50. Sutskever I., Vinyals O., Le Q.V. Sequence to Sequence Learning with Neural Networks // Advances in Neural Information Processing Systems. Curran Associates, Inc., 2014. Vol. 27.
51. Zaremba W., Sutskever I., Vinyals O. Recurrent Neural Network Regularization: arXiv:1409.2329. arXiv, 2015.
52. Glorot X., Bengio Y. Understanding the difficulty of training deep feedforward neural networks // Proceedings of the Thirteenth International Conference on Artificial Intelligence and Statistics. JMLR Workshop and Conference Proceedings, 2010. P. 249–256.
53. Bengio Y. Deep Learning of Representations: Looking Forward: arXiv:1305.0445. arXiv, 2013.
54. Bengio Y. Practical recommendations for gradient-based training of deep architectures: arXiv:1206.5533. arXiv, 2012.
55. LeCun Y., Bengio Y., Hinton G. Deep learning: 7553 // Nature. Nature Publishing Group, 2015. Vol. 521, № 7553. P. 436–444.
56. Huang G. et al. Densely Connected Convolutional Networks // 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Honolulu, HI: IEEE, 2017. P. 2261–2269.
57. Nassif A.B. et al. Speech Recognition Using Deep Neural Networks: A Systematic Review // IEEE Access. 2019. Vol. 7. P. 19143–19165.
58. Ahlawat S. et al. Improved Handwritten Digit Recognition Using Convolutional Neural Networks (CNN) // Sensors. 2020. Vol. 20, № 12. P. 3344.



59. Zhang Y. et al. Neural network-based approaches for biomedical relation classification: A review // *Journal of Biomedical Informatics*. 2019. Vol. 99. P. 103294.
60. Carlini N. et al. Hidden voice commands // *Proceedings of the 25th USENIX Conference on Security Symposium*. USA: USENIX Association, 2016. P. 513–530.
61. Smith D.F., Wiliem A., Lovell B.C. Face Recognition on Consumer Devices: Reflections on Replay Attacks // *IEEE Transactions on Information Forensics and Security*. 2015. Vol. 10, № 4. P. 736–745.
62. Sharif M. et al. Accessorize to a Crime: Real and Stealthy Attacks on State-of-the-Art Face Recognition // *Proceedings of the 2016 ACM SIGSAC Conference on Computer and Communications Security*. New York, NY, USA: Association for Computing Machinery, 2016. P. 1528–1540.
63. Kurakin A., Goodfellow I., Bengio S. Adversarial examples in the physical world: arXiv:1607.02533. arXiv, 2017.
64. Roy P. et al. Effects of Degradations on Deep Neural Network Architectures: arXiv:1807.10108. arXiv, 2023.
65. Зиядинов В.В., Курочкин П.С., Терешонок М.В. Оптимизация обучения сверточных нейронных сетей при распознавании изображений с низкой плотностью точек // *Радиотехн. и электрон.* 2021. Vol. 66, № 12. P. 1207–1215.
66. Ziyadinov V.V., Kurochkin P.S., Tereshonok M.V. Convolutional Neural Network Training Optimization for Low Point Density Image Recognition // *J. Commun. Technol. Electron.* 2021. Vol. 66, № 12. P. 1363–1369.
67. Ziyadinov V., Tereshonok M. Noise Immunity and Robustness Study of Image Recognition Using a Convolutional Neural Network // *Sensors*. 2022. Vol. 22, № 3. P. 1241.
68. Dodge S., Karam L. Understanding how image quality affects deep neural networks // *2016 Eighth International Conference on Quality of Multimedia Experience (QoMEX)*. Lisbon, Portugal: IEEE, 2016. P. 1–6.

69. Zheng S. et al. Improving the Robustness of Deep Neural Networks via Stability Training // 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Las Vegas, NV, USA: IEEE, 2016. P. 4480–4488.
70. Shafahi A. et al. Adversarial training for free! // Proceedings of the 33rd International Conference on Neural Information Processing Systems. Red Hook, NY, USA: Curran Associates Inc., 2019. P. 3358–3369.
71. Ziyadinov V.V., Tereshonok M.V. Neural Network Image Recognition Robustness with Different Augmentation Methods // 2022 Systems of Signal Synchronization, Generating and Processing in Telecommunications (SYNCHROINFO). Arkhangelsk, Russian Federation: IEEE, 2022. P. 1–4.
72. Rebuffi S.-A. et al. Data Augmentation Can Improve Robustness // Advances in Neural Information Processing Systems. Curran Associates, Inc., 2021. Vol. 34. P. 29935–29948.
73. Izmailov P. et al. Averaging Weights Leads to Wider Optima and Better Generalization: arXiv:1803.05407. arXiv, 2019.
74. Ziyadinov V., Tereshonok M. Low-Pass Image Filtering to Achieve Adversarial Robustness // Sensors. 2023. Vol. 23, № 22. P. 9032.
75. Ziyadinov V.V., Tereshonok M.V., Moscow Technical University of Communications and Informatics. MATHEMATICAL MODELS AND RECOGNITION METHODS FOR MOBILE SUBSCRIBERS MUTUAL PLACEMENT // T-Comm. 2021. Vol. 15, № 4. P. 49–56.
76. Ilina O. et al. A Survey on Symmetrical Neural Network Architectures and Applications // Symmetry. 2022. Vol. 14, № 7. P. 1391.
77. Зиядинов В.В., Талалаев А.Б., Терешонок М.В. TRAFFIC JAM DETECTION USING CLUSTER ANALYSIS OF GEOLOCATION DATA // Труды НИИР. 2022. № 2(9).
78. Ziyadinov V.V., Tereshonok M.V. Analytical Survey on MANET and VANET Clusterisation Algorithms // 2020 Systems of Signal Synchronization, Generating and Processing in Telecommunications (SYNCHROINFO). Svetlogorsk, Russia: IEEE, 2020. P. 1–5.

79. Свидетельство о государственной регистрации программы для ЭВМ № 2022660463 Российская Федерация. Программа сравнительного анализа и визуализации результатов работы свёрточных нейронных сетей: № 2022619579: заявл. 23.05.2022: опублик. 03.06.2022 / В. В. Зиядинов, О. В. Ильина, М. В. Терешонок; заявитель Ордена Трудового Красного Знамени Федеральное государственное бюджетное образовательное учреждение высшего образования «Московский технический университет связи и информатики». – EDN LEUEVA.
80. Свидетельство о государственной регистрации программы для ЭВМ № 2020660537 Российская Федерация. Моделирование типов взаимного расположения абонентов сетей мобильной связи: № 2020619728: заявл. 27.08.2020: опублик. 04.09.2020 / В. В. Зиядинов, С. С. Аджемов; заявитель Ордена Трудового Красного Знамени федеральное государственное бюджетное образовательное учреждение высшего образования «Московский технический университет связи и информатики» (МТУСИ). – EDN HFPMMJ.
81. Свидетельство о государственной регистрации программы для ЭВМ № 2022660552 Российская Федерация. Программа моделирования шума в реальных изображениях и генерации обучающих выборок для систем распознавания: № 2022619577: заявл. 23.05.2022: опублик. 06.06.2022 / В. В. Зиядинов, О. В. Ильина; заявитель Ордена Трудового Красного Знамени Федеральное государственное бюджетное образовательное учреждение высшего образования «Московский технический университет связи и информатики». – EDN JBCIOD.
82. Свидетельство о государственной регистрации программы для ЭВМ № 2022660553 Российская Федерация. Программа оценки результативности работы алгоритмов кластеризации: № 2022619578: заявл. 23.05.2022: опублик. 06.06.2022 / В. В. Зиядинов; заявитель Ордена Трудового Красного Знамени Федеральное государственное бюджетное образовательное учреждение высшего образования «Московский технический университет связи и информатики». – EDN DDTMAW.

83. Свидетельство о государственной регистрации программы для ЭВМ № 2022660554 Российская Федерация. Программа визуализации характеристик обучения свёрточных нейронных сетей для определения оптимальных параметров обучающих выборок при требуемой минимальной точности классификации: № 2022619580: заявл. 23.05.2022: опубл. 06.06.2022 / В. В. Зиядинов, М. В. Терешонок; заявитель Ордена Трудового Красного Знамени Федеральное государственное бюджетное образовательное учреждение высшего образования «Московский технический университет связи и информатики». – EDN UDOZJP.
84. Свидетельство о государственной регистрации программы для ЭВМ № 2021619356 Российская Федерация. Программа для оптимизации работы свёрточных нейронных сетей: № 2021618615: заявл. 02.06.2021: опубл. 08.06.2021 / В. В. Зиядинов, В. И. Иванов, М. В. Терешонок; заявитель Ордена Трудового Красного Знамени Федеральное государственное бюджетное образовательное учреждение высшего образования «Московский технический университет связи и информатики». – EDN SLLDLB.
85. Свидетельство о государственной регистрации программы для ЭВМ № 2021619626 Российская Федерация. Программа генерации обучающих выборок для систем распознавания изображений с низкой плотностью точек: № 2021618601: заявл. 02.06.2021: опубл. 15.06.2021 / В. В. Зиядинов, Е. В. Алтухов; заявитель Ордена Трудового Красного Знамени Федеральное государственное бюджетное образовательное учреждение высшего образования «Московский технический университет связи и информатики». – EDN KBVHTP.
86. Свидетельство о государственной регистрации программы для ЭВМ № 2024611339 Российская Федерация. Программный комплекс расчета внешних характеристик точности и устойчивости свёрточной нейронной сети к высокочастотным искажениям и оптимизации параметров предварительной обработки изображений: № 2023689369: заявл. 27.12.2023: опубл. 19.01.2024 / В. В. Зиядинов, О. В. Ильина, М. В. Терешонок; заявитель Ордена Трудового

Красного Знамени Федеральное государственное бюджетное образовательное учреждение высшего образования «Московский технический университет связи и информатики».

87. Свидетельство о государственной регистрации программы для ЭВМ № 2024612396 Российская Федерация. Программный комплекс для демонстрации работы свёрточной нейронной сети, решающей задачу распознавания состязательных изображений: № 2023689454: заявл. 27.12.2023: опубл. 31.01.2024 / О. В. Ильина, М. В. Терешонок, В. В. Зиядинов; заявитель Ордена Трудового Красного Знамени федеральное государственное бюджетное образовательное учреждение высшего образования «Московский технический университет связи и информатики» (МТУСИ).
88. Bengio Y. Learning Deep Architectures for AI // FNT in Machine Learning. 2009. Vol. 2, № 1. P. 1–127.
89. Schmidhuber J. Deep learning in neural networks: An overview // Neural Networks. 2015. Vol. 61. P. 85–117.
90. Weng T.-W.-W.L. Evaluating robustness of neural networks: Thesis. Massachusetts Institute of Technology, 2020.
91. Haykin S. Neural networks and learning machines. 3rd ed. Upper Saddle River: Pearson, 2009.
92. Voulodimos A. et al. Deep Learning for Computer Vision: A Brief Review // Computational Intelligence and Neuroscience. 2018. Vol. 2018. P. 1–13.
93. Guo Y. et al. Deep learning for visual understanding: A review // Neurocomputing. 2016. Vol. 187. P. 27–48.
94. Kumar A., Verma S., Mangla H. A Survey of Deep Learning Techniques in Speech Recognition // 2018 International Conference on Advances in Computing, Communication Control and Networking (ICACCCN). Greater Noida (UP), India: IEEE, 2018. P. 179–185.
95. Mikolov T. et al. Recurrent neural network based language model // Interspeech 2010. ISCA, 2010. P. 1045–1048.

96. Recurrent Neural Networks: Design and Applications / ed. Jain L.M. Lakhmi C. Boca Raton: CRC Press, 1999. 416 p.
97. Ebrahimi Kahou S. et al. Recurrent Neural Networks for Emotion Recognition in Video // Proceedings of the 2015 ACM on International Conference on Multimodal Interaction. Seattle Washington USA: ACM, 2015. P. 467–474.
98. Chicco D. Siamese Neural Networks: An Overview // Artificial Neural Networks / ed. Cartwright H. New York, NY: Springer US, 2021. Vol. 2190. P. 73–94.
99. Khan A. et al. Crowd Monitoring and Localization Using Deep Convolutional Neural Network: A Review // Applied Sciences. 2020. Vol. 10, № 14. P. 4781.
100. Maggiori E. et al. High-Resolution Aerial Image Labeling With Convolutional Neural Networks // IEEE Trans. Geosci. Remote Sensing. 2017. Vol. 55, № 12. P. 7092–7103.
101. Yu Y. et al. A Review of Recurrent Neural Networks: LSTM Cells and Network Architectures // Neural Computation. 2019. Vol. 31, № 7. P. 1235–1270.
102. Ильина О.В., Терешонок М.В. Исследование помехоустойчивости глубокой сверточной нейронной сети при обнаружении транспортных средств на аэрофотоснимках Земли // Радиотехн. и электрон. 2022. Vol. 67, № 2. P. 166–173.
103. Tan M., Le Q. EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks // Proceedings of the 36th International Conference on Machine Learning / ed. Chaudhuri K., Salakhutdinov R. PMLR, 2019. Vol. 97. P. 6105–6114.
104. Shah S.A.R. et al. AmoebaNet: An SDN-enabled network service for big data science // Journal of Network and Computer Applications. 2018. Vol. 119. P. 70–82.
105. Sze V. et al. Efficient Processing of Deep Neural Networks: A Tutorial and Survey // Proc. IEEE. 2017. Vol. 105, № 12. P. 2295–2329.
106. Claesens M., De Moor B. Hyperparameter Search in Machine Learning: arXiv:1502.02127. arXiv, 2015.
107. Yang L., Shami A. On hyperparameter optimization of machine learning algorithms: Theory and practice // Neurocomputing. 2020. Vol. 415. P. 295–316.

108. Feurer M., Hutter F. Hyperparameter Optimization // Automated Machine Learning / ed. Hutter F., Kotthoff L., Vanschoren J. Cham: Springer International Publishing, 2019. P. 3–33.
109. Kostrikov I., Yarats D., Fergus R. Image Augmentation Is All You Need: Regularizing Deep Reinforcement Learning from Pixels: arXiv:2004.13649. arXiv, 2021.
110. Bloice M.D., Stocker C., Holzinger A. Augmentor: An Image Augmentation Library for Machine Learning: arXiv:1708.04680. arXiv, 2017.
111. Hu J., Shen L., Sun G. Squeeze-and-Excitation Networks // 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Salt Lake City, UT: IEEE, 2018. P. 7132–7141.
112. Real E. et al. Regularized evolution for image classifier architecture search // Proceedings of the Thirty-Third AAAI Conference on Artificial Intelligence and Thirty-First Innovative Applications of Artificial Intelligence Conference and Ninth AAAI Symposium on Educational Advances in Artificial Intelligence. Honolulu, Hawaii, USA: AAAI Press, 2019. P. 4780–4789.
113. Simonyan K., Zisserman A. Very Deep Convolutional Networks for Large-Scale Image Recognition: arXiv:1409.1556. arXiv, 2015.
114. Szegedy C. et al. Inception-v4, Inception-ResNet and the Impact of Residual Connections on Learning // AAAI. 2017. Vol. 31, № 1.
115. Sladojevic S. et al. Deep Neural Networks Based Recognition of Plant Diseases by Leaf Image Classification // Computational Intelligence and Neuroscience. 2016. Vol. 2016. P. 1–11.
116. Masood S.Z. et al. License Plate Detection and Recognition Using Deeply Learned Convolutional Neural Networks: arXiv:1703.07330. arXiv, 2017.
117. Ptucha R. et al. Intelligent character recognition using fully convolutional neural networks // Pattern Recognition. 2019. Vol. 88. P. 604–613.
118. Vougioukas K., Petridis S., Pantic M. Realistic Speech-Driven Facial Animation with GANs // Int J Comput Vis. 2020. Vol. 128, № 5. P. 1398–1413.

119. Lawrence S., Giles C.L., Tsoi A.C. Lessons in Neural Network Training: Overfitting May be Harder than Expected // Proceedings of the Fourteenth National Conference on Artificial Intelligence and Ninth Innovative Applications of Artificial Intelligence Conference, AAAI 97, IAAI 97, July 27-31, 1997, Providence, Rhode Island, USA / ed. Kuipers B., Webber B.L. AAAI Press / The MIT Press, 1997. P. 540–545.
120. Ying X. An Overview of Overfitting and its Solutions // J. Phys.: Conf. Ser. 2019. Vol. 1168. P. 022022.
121. Tetko I.V., Livingstone D.J., Luik A.I. Neural network studies. 1. Comparison of overfitting and overtraining // J. Chem. Inf. Comput. Sci. 1995. Vol. 35, № 5. P. 826–833.
122. Encyclopedia of Machine Learning / ed. Sammut C., Webb G.I. Boston, MA: Springer US, 2010.
123. Li D. et al. Deeper, Broader and Artier Domain Generalization // 2017 IEEE International Conference on Computer Vision (ICCV). Venice: IEEE, 2017. P. 5543–5551.
124. Koh P.W. et al. WILDS: A Benchmark of in-the-Wild Distribution Shifts // Proceedings of the 38th International Conference on Machine Learning / ed. Meila M., Zhang T. PMLR, 2021. Vol. 139. P. 5637–5664.
125. Zhou K. et al. Domain Generalization: A Survey // IEEE Trans. Pattern Anal. Mach. Intell. 2022. P. 1–20.
126. Subbaswamy A., Adams R., Saria S. Evaluating Model Robustness and Stability to Dataset Shift // Proceedings of The 24th International Conference on Artificial Intelligence and Statistics. PMLR, 2021. P. 2611–2619.
127. Zech J.R. et al. Variable generalization performance of a deep learning model to detect pneumonia in chest radiographs: A cross-sectional study // PLoS Med / ed. Sheikh A. 2018. Vol. 15, № 11. P. e1002683.
128. Beery S., Van Horn G., Perona P. Recognition in Terra Incognita // Computer Vision – ECCV 2018 / ed. Ferrari V. et al. Cham: Springer International Publishing, 2018. Vol. 11220. P. 472–489.



129. Pyo J. et al. Cyanobacteria cell prediction using interpretable deep learning model with observed, numerical, and sensing data assemblage // *Water Research*. 2021. Vol. 203. P. 117483.
130. Jean N. et al. Combining satellite imagery and machine learning to predict poverty // *Science*. 2016. Vol. 353, № 6301. P. 790–794.
131. Santurkar S., Tsipras D., Madry A. BREEDS: Benchmarks for Subpopulation Shift: arXiv:2008.04859. arXiv, 2020.
132. Гайер А.В., Чернышова Ю.С., Шешкус А.В. Генерация искусственной обучающей выборки для задачи распознавания символов полей паспорта РФ // *Сенсорные системы*. 2018. Vol. 32, № 3. P. 230–235.
133. Perez L., Wang J. The Effectiveness of Data Augmentation in Image Classification using Deep Learning: arXiv:1712.04621. arXiv, 2017.
134. Van Dyk D.A., Meng X.-L. The Art of Data Augmentation // *Journal of Computational and Graphical Statistics*. 2001. Vol. 10, № 1. P. 1–50.
135. Dempster A.P., Laird N.M., Rubin D.B. Maximum Likelihood from Incomplete Data Via the *EM* Algorithm // *Journal of the Royal Statistical Society: Series B (Methodological)*. 1977. Vol. 39, № 1. P. 1–22.
136. Tanner M.A., Wong W.H. The Calculation of Posterior Distributions by Data Augmentation // *Journal of the American Statistical Association*. 1987. Vol. 82, № 398. P. 528–540.
137. Lecun Y. et al. Gradient-based learning applied to document recognition // *Proc. IEEE*. 1998. Vol. 86, № 11. P. 2278–2324.
138. Fort S. et al. Drawing Multiple Augmentation Samples Per Image During Training Efficiently Decreases Test Error: arXiv:2105.13343. arXiv, 2022.
139. Mohammed R., Rawashdeh J., Abdullah M. Machine Learning with Oversampling and Undersampling Techniques: Overview Study and Experimental Results // 2020 11th International Conference on Information and Communication Systems (ICICS). Irbid, Jordan: IEEE, 2020. P. 243–248.
140. Haibo He, Garcia E.A. Learning from Imbalanced Data // *IEEE Trans. Knowl. Data Eng.* 2009. Vol. 21, № 9. P. 1263–1284.

141. Gayer A., Chernyshova Y., Sheshkus A. Effective real-time augmentation of training dataset for the neural networks learning // Eleventh International Conference on Machine Vision (ICMV 2018) / ed. Nikolaev D.P. et al. Munich, Germany: SPIE, 2019. P. 64.
142. Shorten C., Khoshgoftaar T.M. A survey on Image Data Augmentation for Deep Learning // J Big Data. 2019. Vol. 6, № 1. P. 60.
143. Wu R. et al. Deep Image: Scaling up Image Recognition: arXiv:1501.02876. arXiv, 2015.
144. Zhong Z. et al. Random Erasing Data Augmentation // AAAI. 2020. Vol. 34, № 07. P. 13001–13008.
145. DeVries T., Taylor G.W. Improved Regularization of Convolutional Neural Networks with Cutout: arXiv:1708.04552. arXiv, 2017.
146. Inoue H. Data Augmentation by Pairing Samples for Images Classification: arXiv:1801.02929. arXiv, 2018.
147. Liang D. et al. Understanding Mixup Training Methods // IEEE Access. 2018. Vol. 6. P. 58774–58783.
148. Takahashi R., Matsubara T., Uehara K. Data Augmentation Using Random Image Cropping and Patching for Deep CNNs // IEEE Trans. Circuits Syst. Video Technol. 2020. Vol. 30, № 9. P. 2917–2931.
149. Moosavi-Dezfooli S.-M., Fawzi A., Frossard P. DeepFool: A Simple and Accurate Method to Fool Deep Neural Networks // 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Las Vegas, NV, USA: IEEE, 2016. P. 2574–2582.
150. Su J., Vargas D.V., Sakurai K. One Pixel Attack for Fooling Deep Neural Networks // IEEE Trans. Evol. Computat. 2019. Vol. 23, № 5. P. 828–841.
151. Goodfellow I. et al. Maxout Networks // Proceedings of the 30th International Conference on Machine Learning. PMLR, 2013. P. 1319–1327.
152. Szegedy C. et al. Intriguing properties of neural networks: arXiv:1312.6199. arXiv, 2014.

153. Gatys L., Ecker A., Bethge M. A Neural Algorithm of Artistic Style // *Journal of Vision*. 2016. Vol. 16, № 12. P. 326.
154. Luan F. et al. Deep Photo Style Transfer // *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. Honolulu, HI: IEEE, 2017. P. 6997–7005.
155. Mechrez R., Shechtman E., Zelnik-Manor L. Photorealistic Style Transfer with Screened Poisson Equation: arXiv:1709.09828. arXiv, 2017.
156. Jing Y. et al. Neural Style Transfer: A Review // *IEEE Trans. Visual. Comput. Graphics*. 2020. Vol. 26, № 11. P. 3365–3385.
157. Li Y. et al. Demystifying Neural Style Transfer // *Proceedings of the Twenty-Sixth International Joint Conference on Artificial Intelligence*. Melbourne, Australia: International Joint Conferences on Artificial Intelligence Organization, 2017. P. 2230–2236.
158. Tobin J. et al. Domain randomization for transferring deep neural networks from simulation to the real world // *2017 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. Vancouver, BC: IEEE, 2017. P. 23–30.
159. Bowles C. et al. GAN Augmentation: Augmenting Training Data using Generative Adversarial Networks: arXiv:1810.10863. arXiv, 2018.
160. Antoniou A., Storkey A., Edwards H. Data Augmentation Generative Adversarial Networks: arXiv:1711.04340. arXiv, 2018.
161. Sixt L., Wild B., Landgraf T. RenderGAN: Generating Realistic Labeled Data // *Front. Robot. AI*. 2018. Vol. 5. P. 66.
162. Zhu X. et al. Emotion Classification with Data Augmentation Using Generative Adversarial Networks // *Advances in Knowledge Discovery and Data Mining* / ed. Phung D. et al. Cham: Springer International Publishing, 2018. Vol. 10939. P. 349–360.
163. Salimans T. et al. Improved techniques for training GANs // *Proceedings of the 30th International Conference on Neural Information Processing Systems*. Red Hook, NY, USA: Curran Associates Inc., 2016. P. 2234–2242.
164. Hendrycks D. et al. Natural Adversarial Examples. arXiv, 2019.

165. Shafahi A. et al. Poison frogs! targeted clean-label poisoning attacks on neural networks // Proceedings of the 32nd International Conference on Neural Information Processing Systems. Red Hook, NY, USA: Curran Associates Inc., 2018. P. 6106–6116.
166. Alatwi H.A., Morisset C. Adversarial Machine Learning In Network Intrusion Detection Domain: A Systematic Review: arXiv:2112.03315. arXiv, 2021.
167. Mohammadpour L. et al. A Survey of CNN-Based Network Intrusion Detection // Applied Sciences. 2022. Vol. 12, № 16. P. 8162.
168. Ferrer M.A. et al. Robustness of Offline Signature Verification Based on Gray Level Features // IEEE Trans.Inform.Forensic Secur. 2012. Vol. 7, № 3. P. 966–977.
169. Jalalvand A. et al. On the application of reservoir computing networks for noisy image recognition // Neurocomputing. 2018. Vol. 277. P. 237–248.
170. Karahan S. et al. How Image Degradations Affect Deep CNN-Based Face Recognition? // 2016 International Conference of the Biometrics Special Interest Group (BIOSIG). Darmstadt, Germany: IEEE, 2016. P. 1–5.
171. Zhang K. et al. Learning Deep CNN Denoiser Prior for Image Restoration // 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Honolulu, HI: IEEE, 2017. P. 2808–2817.
172. Ghosh S. et al. Robustness of Deep Convolutional Neural Networks for Image Degradations // 2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). Calgary, AB: IEEE, 2018. P. 2916–2920.
173. Endo K., Tanaka M., Okutomi M. Classifying Degraded Images Over Various Levels Of Degradation // 2020 IEEE International Conference on Image Processing (ICIP). Abu Dhabi, United Arab Emirates: IEEE, 2020. P. 1691–1695.
174. Kalalembang E., Usman K., Gunawan I.P. DCT-based local motion blur detection // International Conference on Instrumentation, Communication, Information Technology, and Biomedical Engineering 2009. Bandung, Indonesia: IEEE, 2009. P. 1–6.

175. Ramakrishnan S. et al. Deep Generative Filter for Motion Deblurring // 2017 IEEE International Conference on Computer Vision Workshops (ICCVW). Venice: IEEE, 2017. P. 2993–3000.
176. Domingos P. A few useful things to know about machine learning // Commun. ACM. 2012. Vol. 55, № 10. P. 78–87.
177. Woods W.A. Important issues in knowledge representation // Proc. IEEE. 1986. Vol. 74, № 10. P. 1322–1334.
178. Studer S. et al. Towards CRISP-ML(Q): A Machine Learning Process Model with Quality Assurance Methodology // MAKE. 2021. Vol. 3, № 2. P. 392–413.
179. Limberg C., Wersing H., Ritter H. Beyond Cross-Validation—Accuracy Estimation for Incremental and Active Learning Models // MAKE. 2020. Vol. 2, № 3. P. 327–346.
180. Szegedy C., Toshev A., Erhan D. Deep Neural Networks for Object Detection // Advances in Neural Information Processing Systems. Curran Associates, Inc., 2013. Vol. 26.
181. Goodfellow I.J. et al. Multi-digit Number Recognition from Street View Imagery using Deep Convolutional Neural Networks: arXiv:1312.6082. arXiv, 2014.
182. Roodschild M., Gotay Sardiñas J., Will A. A new approach for the vanishing gradient problem on sigmoid activation // Prog Artif Intell. 2020. Vol. 9, № 4. P. 351–360.
183. Borovicka T. et al. Selecting Representative Data Sets // Advances in Data Mining Knowledge Discovery and Applications / ed. Karahoca A. InTech, 2012.
184. Shin H.-C. et al. Deep Convolutional Neural Networks for Computer-Aided Detection: CNN Architectures, Dataset Characteristics and Transfer Learning // IEEE Trans. Med. Imaging. 2016. Vol. 35, № 5. P. 1285–1298.
185. Ilina O.V., Tereshonok M.V. Robustness Study of a Deep Convolutional Neural Network for Vehicle Detection in Aerial Imagery // J. Commun. Technol. Electron. 2022. Vol. 67, № 2. P. 164–170.
186. Wang Z. et al. Image Quality Assessment: From Error Visibility to Structural Similarity // IEEE Trans. on Image Process. 2004. Vol. 13, № 4. P. 600–612.

187. Wallace G.K. The JPEG still picture compression standard // IEEE Trans. Consumer Electron. 1992. Vol. 38, № 1. P. xviii–xxxiv.
188. Menshih P.G., Erokhin S.D., Gorodnichev M.G. Efficiency Analysis of Neural Networks Ensembles Using Synthetic Data // 2020 Wave Electronics and its Application in Information and Telecommunication Systems (WECONF). St. Petersburg, Russia: IEEE, 2020. P. 1–3.
189. Емельянов С.О. et al. Методы аугментации обучающих выборок в задачах классификации изображений // Сенсорные системы. 2018. Vol. 32, № 3. P. 236–245.
190. Natural Images [Electronic resource]. URL: <https://www.kaggle.com/datasets/prasunroy/natural-images> (accessed: 14.09.2023).
191. Wen L., Li X., Gao L. A transfer convolutional neural network for fault diagnosis based on ResNet-50 // Neural Comput & Applic. 2020. Vol. 32, № 10. P. 6111–6124.
192. Sengupta A. et al. Going Deeper in Spiking Neural Networks: VGG and Residual Architectures // Front. Neurosci. 2019. Vol. 13. P. 95.
193. Deng J. et al. ImageNet: A large-scale hierarchical image database // 2009 IEEE Conference on Computer Vision and Pattern Recognition. Miami, FL: IEEE, 2009. P. 248–255.
194. Liu F., Lin G., Shen C. CRF learning with CNN features for image segmentation // Pattern Recognition. 2015. Vol. 48, № 10. P. 2983–2992.
195. Yang L. et al. Deep Location-Specific Tracking // Proceedings of the 25th ACM international conference on Multimedia. Mountain View California USA: ACM, 2017. P. 1309–1317.
196. Ren Y. et al. Deep Image Spatial Transformation for Person Image Generation // 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Seattle, WA, USA: IEEE, 2020. P. 7687–7696.
197. Borji A. Generated Faces in the Wild: Quantitative Comparison of Stable Diffusion, Midjourney and DALL-E 2. arXiv, 2022.

198. Jasim H.A. et al. Classify Bird Species Audio by Augment Convolutional Neural Network // 2022 International Congress on Human-Computer Interaction, Optimization and Robotic Applications (HORA). Ankara, Turkey: IEEE, 2022. P. 1–6.
199. Mustaqeem, Kwon S. A CNN-Assisted Enhanced Audio Signal Processing for Speech Emotion Recognition // Sensors. 2019. Vol. 20, № 1. P. 183.
200. Zaharia M. et al. Accelerating the machine learning lifecycle with MLflow. // IEEE Data Eng. Bull. 2018. Vol. 41, № 4. P. 39–45.
201. Baylor D. et al. Continuous Training for Production {ML} in the {TensorFlow} Extended ({{{{{{TFX}}}}}) Platform. 2019. P. 51–53.
202. Huang H. et al. Exploring Architectural Ingredients of Adversarially Robust Deep Neural Networks. arXiv, 2021.
203. Wu B. et al. Do Wider Neural Networks Really Help Adversarial Robustness? arXiv, 2020.
204. Akrouf M. On the Adversarial Robustness of Neural Networks without Weight Transport. arXiv, 2019.
205. Papernot N. et al. The Limitations of Deep Learning in Adversarial Settings // 2016 IEEE European Symposium on Security and Privacy (EuroS&P). Saarbrücken: IEEE, 2016. P. 372–387.
206. Hu Y. et al. Artificial Intelligence Security: Threats and Countermeasures // ACM Comput. Surv. 2023. Vol. 55, № 1. P. 1–36.
207. Chakraborty A. et al. A survey on adversarial attacks and defences // CAAI Trans on Intel Tech. 2021. Vol. 6, № 1. P. 25–45.
208. Xu H. et al. Adversarial Attacks and Defenses in Images, Graphs and Text: A Review // Int. J. Autom. Comput. 2020. Vol. 17, № 2. P. 151–178.
209. Han C. et al. Adversarial Example Detection and Restoration Defensive Framework for Signal Intelligent Recognition Networks // Applied Sciences. 2023. Vol. 13, № 21. P. 11880.

210. Ben-David S. et al. Analysis of Representations for Domain Adaptation // Advances in Neural Information Processing Systems / ed. Schölkopf B., Platt J., Hoffman T. MIT Press, 2006. Vol. 19.
211. Athalye A. et al. Synthesizing Robust Adversarial Examples // Proceedings of the 35th International Conference on Machine Learning. PMLR, 2018. P. 284–293.
212. He K. et al. Deep Residual Learning for Image Recognition. arXiv, 2015.
213. Iandola F.N. et al. SqueezeNet: AlexNet-level accuracy with 50x fewer parameters and <math>0.5\text{MB}</math> model size. arXiv, 2016.
214. Shaham U., Yamada Y., Negahban S. Understanding adversarial training: Increasing local stability of supervised models through robust optimization // Neurocomputing. 2018. Vol. 307. P. 195–204.
215. Samangouei P., Kabkab M., Chellappa R. Defense-GAN: Protecting Classifiers Against Adversarial Attacks Using Generative Models: arXiv:1805.06605. arXiv, 2018.
216. Hinton G., Vinyals O., Dean J. Distilling the Knowledge in a Neural Network. arXiv, 2015.
217. Xu W., Evans D., Qi Y. Feature Squeezing: Detecting Adversarial Examples in Deep Neural Networks // Proceedings 2018 Network and Distributed System Security Symposium. 2018.
218. Liao F. et al. Defense Against Adversarial Attacks Using High-Level Representation Guided Denoiser // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR). 2018.
219. Creswell A., Bharath A.A. Denoising Adversarial Autoencoders // IEEE Trans. Neural Netw. Learning Syst. 2019. Vol. 30, № 4. P. 968–984.
220. Rahimi N., Maynor J., Gupta B. Adversarial Machine Learning: Difficulties in Applying Machine Learning to Existing Cybersecurity Systems. P. 40–31.
221. Xu H. et al. Adversarial Attacks and Defenses: Frontiers, Advances and Practice // Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining. Virtual Event CA USA: ACM, 2020. P. 3541–3542.



222. Rebuffi S.-A. et al. Fixing Data Augmentation to Improve Adversarial Robustness: arXiv:2103.01946. arXiv, 2021.
223. Wang D. et al. Improving Global Adversarial Robustness Generalization With Adversarially Trained GAN: arXiv:2103.04513. arXiv, 2021.
224. Zhang H. et al. The Limitations of Adversarial Training and the Blind-Spot Attack: arXiv:1901.04684. arXiv, 2019.
225. Lee H., Kang S., Chung K. Robust Data Augmentation Generative Adversarial Network for Object Detection // Sensors. 2022. Vol. 23, № 1. P. 157.
226. Xiao L. et al. Adversarial and Random Transformations for Robust Domain Adaptation and Generalization // Sensors. 2023. Vol. 23, № 11. P. 5273.
227. Ross A., Doshi-Velez F. Improving the Adversarial Robustness and Interpretability of Deep Neural Networks by Regularizing Their Input Gradients // AAAI. 2018. Vol. 32, № 1.
228. Ross A.S., Hughes M.C., Doshi-Velez F. Right for the Right Reasons: Training Differentiable Models by Constraining their Explanations. arXiv, 2017.
229. Li H. et al. Online Alternate Generator Against Adversarial Attacks // IEEE Trans. on Image Process. 2020. Vol. 29. P. 9305–9315.
230. Yin Z. et al. Defense against adversarial attacks by low-level image transformations // Int J Intell Syst. 2020. Vol. 35, № 10. P. 1453–1466.
231. Ito K., Xiong K. Gaussian filters for nonlinear filtering problems // IEEE Trans. Automat. Contr. 2000. Vol. 45, № 5. P. 910–927.
232. Blinchikoff H.J., Zverev A.I. Filtering in the Time and Frequency Domains. Institution of Engineering and Technology, 2001.
233. Krizhevsky A. Learning Multiple Layers of Features from Tiny Images. 2009.
234. Russakovsky O. et al. ImageNet Large Scale Visual Recognition Challenge // Int J Comput Vis. 2015. Vol. 115, № 3. P. 211–252.
235. Rock-Paper-Scissors Images [Electronic resource]. URL: <https://www.kaggle.com/datasets/drgfreeman/rockpaperscissors> (accessed: 11.09.2023).

236. Madry A. et al. Towards Deep Learning Models Resistant to Adversarial Attacks. arXiv, 2017.
237. Tramèr F. et al. The Space of Transferable Adversarial Examples: arXiv:1704.03453. arXiv, 2017.
238. Wang J. et al. Defensive Patches for Robust Recognition in the Physical World // 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). New Orleans, LA, USA: IEEE, 2022. P. 2446–2455.
239. Andriushchenko M. et al. Square Attack: A Query-Efficient Black-Box Adversarial Attack via Random Search // Computer Vision – ECCV 2020 / ed. Vedaldi A. et al. Cham: Springer International Publishing, 2020. Vol. 12368. P. 484–501.
240. Carlini N., Wagner D. Towards Evaluating the Robustness of Neural Networks // 2017 IEEE Symposium on Security and Privacy (SP). San Jose, CA, USA: IEEE, 2017. P. 39–57.
241. Chen P.-Y. et al. ZOO: Zeroth Order Optimization Based Black-box Attacks to Deep Neural Networks without Training Substitute Models // Proceedings of the 10th ACM Workshop on Artificial Intelligence and Security. Dallas Texas USA: ACM, 2017. P. 15–26.
242. Chen J., Jordan M.I., Wainwright M.J. HopSkipJumpAttack: A Query-Efficient Decision-Based Attack // 2020 IEEE Symposium on Security and Privacy (SP). San Francisco, CA, USA: IEEE, 2020. P. 1277–1294.
243. Wang H. et al. High-Frequency Component Helps Explain the Generalization of Convolutional Neural Networks // 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Seattle, WA, USA: IEEE, 2020. P. 8681–8691.
244. Bradley A. et al. Visual orientation and spatial frequency discrimination: a comparison of single neurons and behavior // Journal of Neurophysiology. 1987. Vol. 57, № 3. P. 755–772.
245. Zhou Y. et al. High frequency patterns play a key role in the generation of adversarial examples // Neurocomputing. 2021. Vol. 459. P. 131–141.

246. Zhang Z., Jung C., Liang X. Adversarial Defense by Suppressing High-frequency Components: arXiv:1908.06566. arXiv, 2019.
247. Thang D.D., Matsui T. Automated Detection System for Adversarial Examples with High-Frequency Noises Sieve // *Cyberspace Safety and Security* / ed. Vaidya J., Zhang X., Li J. Cham: Springer International Publishing, 2019. Vol. 11982. P. 348–362.

# Приложение 1. Акты о внедрении и использовании результатов диссертационной работы

МИНИСТЕРСТВО ЦИФРОВОГО  
РАЗВИТИЯ, СВЯЗИ И МАССОВЫХ  
КОММУНИКАЦИЙ  
РОССИЙСКОЙ ФЕДЕРАЦИИ

Ордена Трудового Красного Знамени  
федеральное государственное  
бюджетное образовательное  
учреждение высшего образования

«МОСКОВСКИЙ ТЕХНИЧЕСКИЙ  
УНИВЕРСИТЕТ СВЯЗИ И  
ИНФОРМАТИКИ»  
(МТУСИ)



MINISTRY OF DIGITAL  
DEVELOPMENT,  
COMMUNICATIONS  
AND MASS MEDIA OF  
THE RUSSIAN FEDERATION

MOSCOW TECHNICAL  
UNIVERSITY  
OF COMMUNICATIONS  
AND INFORMATICS  
(MTUCI)

ул. Авиамоторная, д. 8а, Москва, 111024,  
www.mtuci.ru; mtuci.pf; e-mail: kanc@mtuci.ru  
Телефон (495) 957-77-31; факс (495) 957-77-36  
ОГРН 1027700117191; ИНН/КПП 7722000820/772201001; ОКПО 01179952;  
ОКВЭД 85.22, 46.19, 58.19, 61.10, 68.32, 72.19, 85.21, 85.23, 85.42.9, 71.20, 33.13, 26.60 ; ОКТМО 45388000

\_\_\_\_\_ 20 \_\_\_\_\_ № \_\_\_\_\_  
На № \_\_\_\_\_ от \_\_\_\_\_

УТВЕРЖДАЮ  
Ректор МТУСИ

С.Д. Ерохин



\_\_\_\_\_ 2024 г.

## АКТ

### об использовании результатов диссертационной работы Зядинова Вадима Валерьевича на тему «Оптимизация помехоустойчивости и точности нейросетевого распознавания изображений»


Комиссия в составе:


начальник НИО, д.т.н., доцент М. В. Терешонок;  
главный научный сотрудник, д.т.н., доцент Н. В. Кленов;  
заведующий НИЛ, к.ф.-м.н. А. Е. Щеголев


составила настоящий акт о том, что **результаты** диссертационной работы «Оптимизация помехоустойчивости и точности нейросетевого распознавания изображений», представленной на соискание учёной степени кандидата технических наук, **использованы** в МТУСИ при выполнении НИОКР в соответствии со следующей таблицей.

№ п.п.	Результат диссертационной работы	Шифр НИОКР, в которых использован результат диссертационной работы	Итог использования результата диссертационной работы в НИОКР
1	Математические модели генерации изображений; метод оценки оптимальной степени неопределённости в обучающих данных; рациональный метод аугментации обучающих изображений.	НИР «Шеренга-2020», Государственный контракт от 12.08.2020 № 2022187150262452655002912	Реализация программных комплексов и алгоритмов функционирования; повышение качества их работы.
2	Математические модели генерации изображений; метод оценки оптимальной степени неопределённости в обучающих данных; рациональный метод аугментации обучающих изображений.	СЧ ОКР «5P17K302-МТУСИ», договор от 07.12.2018 № 18/201/ОКР/4806/18	Реализация программных комплексов и алгоритмов функционирования; повышение качества их работы.
3	Метод фильтрации изображений для противодействия высокочастотным искажениям.	НИР «Интеллект-В», Государственный контракт от 20.07.2023 № 2325187150062452655003606	Повышение эффективности и надёжности работы разрабатываемых методов.

Использование предложенных в диссертации Зиядинова Вадима Валерьевича на тему «Оптимизация помехоустойчивости и точности нейросетевого распознавания изображений» научных методов и технических решений позволило повысить качество работы моделей машинного обучения, задействованных в программных комплексах, разрабатываемых МТУСИ в рамках Государственных контрактов.

 д.т.н., доцент М.В. Терешонок

 д.т.н., доцент Н.В. Кленов

 к.ф.-м.н. А.Е. Щеголев