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

АВТОРЕФЕРАТ

диссертации на соискание ученой степени

кандидата технических наук

Москва 2024

Работа выполнена в ордена Трудового Красного Знамени федеральном государственном бюджетном образовательном учреждении высшего образования «Московский технический университет связи и информатики» (МТУСИ).

Научный руководитель:	**Терешонок Максим Валерьевич** – доктор технических наук, доцент, начальник НИО-49 МТУСИ.
Научный консультант:	**Гладышев Анатолий Иванович** – доктор технических наук, доцент, профессор РосНОУ.
Официальные оппоненты:	**Соловьев Игорь Игоревич**, доктор физико-математических наук, ведущий научный сотрудник отдела фундаментальных исследований Научно-инжинирингового центра специальной радиосвязи и радиомониторинга Федерального государственного бюджетного образовательного учреждения высшего образования «МИРЭА – Российский технологический университет». **Доленко Сергей Анатольевич**, кандидат физико-математических наук, заведующий лабораторией адаптивных методов обработки данных научно-исследовательского института ядерной физики имени Д. В. Скобельцына Московского государственного университета имени М. В. Ломоносова.
Ведущая организация:	Акционерное общество «Научно-исследовательский институт точных приборов» (АО «НИИ ТП»).

Защита диссертации состоится «13» июня 2024 года в 13:00 на заседании диссертационного совета по защите докторских и кандидатских диссертаций 55.2.002.01 при ордена Трудового Красного Знамени федеральном государственном бюджетном образовательном учреждении высшего образования «Московский технический университет связи и информатики» по адресу: 111024, г. Москва, ул. Авиамоторная, д. 8-а, МТУСИ, ауд. А-211.

С диссертацией можно ознакомиться в библиотеке и на сайте МТУСИ: http://srd-mtuci.ru/images/Dis-Ziyadinov/dis-Ziyadinov.pdf

Автореферат разослан «___» _____ 2024 г.

Учёный секретарь
диссертационного совета 55.2.002.01
доктор технических наук, доцент М.В. Терешонок

ОБЩАЯ ХАРАКТЕРИСТИКА РАБОТЫ

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. 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, компьютерное зрение и визуализация данных.

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, распознавания номерных знаков, обнаружения патологий человека, растений и животных, распознавания лиц и эмоций, выделения объектов интереса в видеопотоке и т.д.

Большинство современных публикаций, рассматривающих свёрточные нейронные сети, посвящено применению известных нейронных сетей к новым наборам данных из различных проблемных областей. Многие публикации посвящены совершенствованию топологий нейронных сетей и методов обучения. 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.

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". Некоторые публикации посвящены оценке успешности состязательных атак. Недавние работы, касающиеся количественной оценки влияния неопределенности в данных, не дали решений для повышения помехоустойчивости нейронных сетей. 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, И. И. Соловьёв, И. В. Оселедец, М. В. Терешонок, Ian J. Goodfellow, Geoffrey Hinton, Alex Krizhevsky, Yann LeCun, Christian Szegedy, Ilya Sutskever, Yoshua Bengio.

Свёрточные нейронные сети обеспечивают высокие результаты регрессии и классификации во многих задачах, показывают высокую эффективность классификации наборов данных с тысячами классов, вытесняют другие методы в

таких задачах, как распознавание речи и рукописного текста, в биомедицинских приложениях 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, известны атаки на системы распознавания голосовых команд, создающие аудиосигнал, распознаваемый системами как голосовая команда, но воспринимаемый человеком как шум. Множество исследований 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. Другим перспективным методом является так называемое «состязательное обучение». Важным методом устойчивого обучения также является аугментация данных. В большом числе исследований не продемонстрировано значительных успехов при использовании аугментации обучающих данных, однако, в статье авторы показали, что аугментация обучающих данных совместно с усреднением весов модели по ансамблю нейронных сетей 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.

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.

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;

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.

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.;

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

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

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 «5Р17К302-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.

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.

СОДЕРЖАНИЕ РАБОТЫ

Первая глава посвящена детальному обзору предмета исследований. Рассмотрены проблемы практического применения свёрточных нейронных сетей, приведены теоретические основы аугментации данных, виды и способы аугментации данных, рассмотрена проблема естественных состязательных искажений, приведён обзор исследований, посвященных оценке помехоустойчивости нейронных сетей. Проведена математическая постановка и формализация научной задачи исследования, приведены ограничения исследования. В результате исследования выявлены следующие проблемы:

Возможность применения методов аугментации в исследованиях редко учитывается исследователями. Исследования не рассматривают проблему низкой устойчивости СНС к искажениям различного рода. Способы снижения влияния искажений на качество распознавания изображений также рассматривается редко. 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.

В **Главе 2** описана математическая модель контролируемой генерации изображений, проведена оценка зависимости качества распознавания от неопределенности в тестовых наборах данных, а также проведено исследование помехоустойчивости сети, обученной на наборах данных с искажениями.

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 оценки устойчивости 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. 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.

Параметр, определяющий меру неопределенности, может быть описан как

$$U = \frac{d}{a}, \tag{1}$$

где d - дисперсия, a - линейный размер фигуры. 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, U можно описать как

$$U = \frac{I_{noise}}{I_{info}}, \tag{2}$$

где I_{noise} - средняя интенсивность шума, а I_{info} - средняя интенсивность значимой части распознаваемого изображения. 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.

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:

$$P_0 = \frac{M_{correct}}{M_{total}}, \quad (3)$$

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 **главах 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:

$$P(U_{TS}) = \frac{1}{N_{TR}} \cdot \sum_{U_{TR}} P(U_{TR}; U_{TS}); \quad (4)$$

$$P_0 = \frac{1}{N_{TS}} \cdot \sum_{U_{TS}} P(U_{TS}), \quad (5)$$

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}.

Характеристики обучающего набора данных значительно влияют на характеристики обученной нейронной сети и на точность выполнения ими будущих задач распознавания на новых наборах данных. Для оптимизации обучения СНС (в отношении поиска оптимальных параметров обучающего набора данных для улучшения качества распознавания изображений с различными значениями неопределенности) проведены эксперименты, включающие обучение СНС на наборах данных с множеством различных значений неопределенности. На рисунке 1 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.

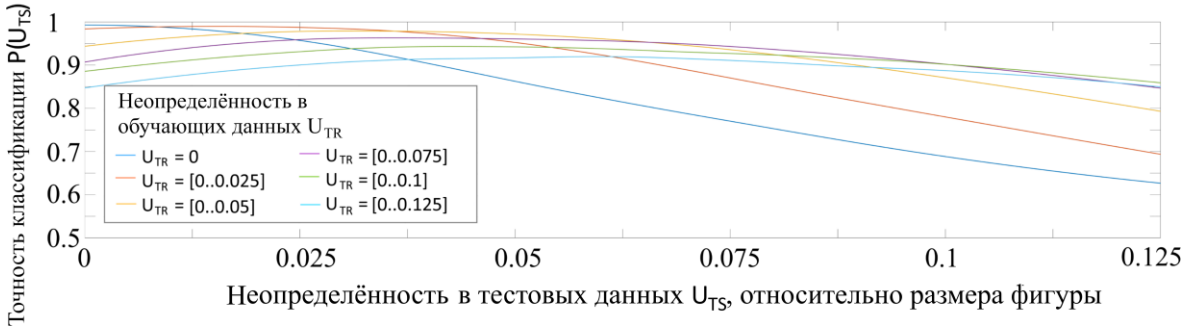

Рисунок 1 – Графики зависимости точности распознавания от неопределённости в тестовых наборах данных, полученные сетями, обученными на наборах данных различными значениями неопределённости.

Из рисунка 1 можно заключить, что максимальная точность распознавания обученной нейронной сетью достигается при $U_{TS} \approx U_{TR}$. Более того, анализ

графиков на рисунке 1 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.

В **Главе 3** проведено комплексное исследование поведения СНС при изменении неопределенности данных в обучающем наборе U_{TR}. Сгенерировано множество обучающих и тестовых наборов данных с различной неопределенностью $U_{TR\ i} = d_i/a$ (от 0 до 0,125 от линейного размера фигуры с шагом 0,0005). Это позволило получить матрицу точности распознавания в зависимости от неопределенностей обучающего и тестового наборов данных $P = P\ (U_{TR};\ U_{TS})$ (рисунок 2).

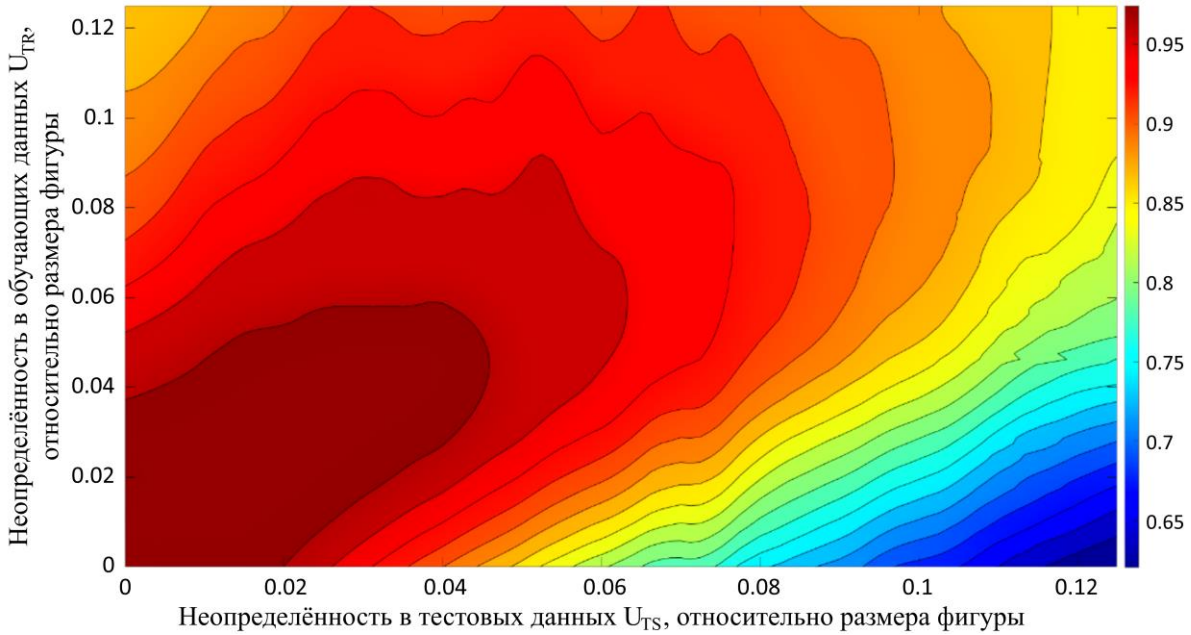

Рисунок 2 – График зависимости точности распознавания изображений от неопределённости в тестовых и обучающих данных U_{TR} и U_{TS}

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.

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 показана область, включающая значения неопределенности тестового набора данных U_{TS}, в которых обеспечена точность распознавания P_{thr} выше 90%.

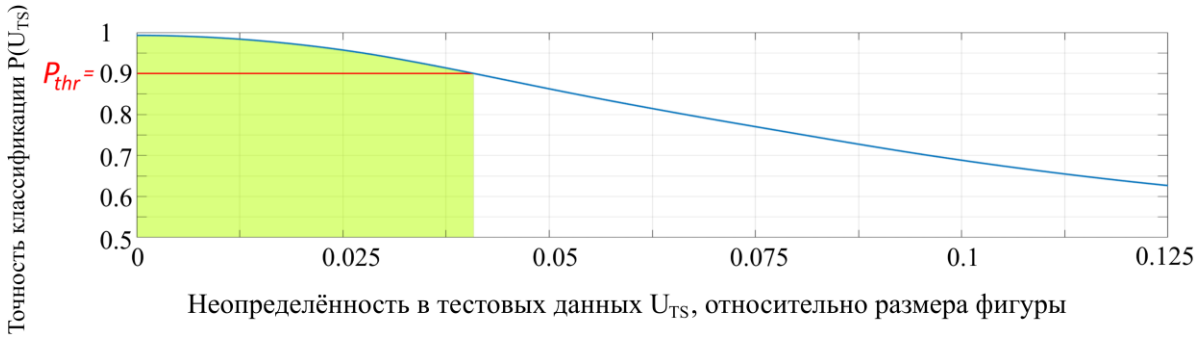

Рисунок 3 – Область, включающая значения неопределенности тестового набора данных с точностью распознавания выше 90%

Выделенная область на рисунке 3 рассчитывается как

$$Q(P_{thr}) = \sum_{U_{TS}=U_{TS}^{\min},P\geq P_{thr}}^{U_{TS}=U_{TS}^{\max},P\geq P_{thr}} P(U_{TS}). \quad (6)$$

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

После оценки области, включающей значения неопределенности тестового набора данных с точностью распознавания не ниже определённого порога P_{thr} возможно обоснование выбора оптимального значения неопределённости в обучающем наборе данных. Для каждой сети, обученной ранее на наборах данных с различными неопределенностями, были получены интегральные значения точности распознавания Q при различных пороговых значениях:

$$Q(U_{TR};P_{thr}) = \sum_{U_{TS}=U_{TS}^{\min},P\geq P_{thr}}^{U_{TS}=U_{TS}^{\max},P\geq P_{thr}} P(U_{TR};U_{TS}) \quad (7)$$

Q —интегральное значение точности классификации для всех тестовых наборов данных, для которых точность распознавания превысила порог $P \geq P_{thr}$. Полученные данные обобщены на графике, представленном на рисунке 4.

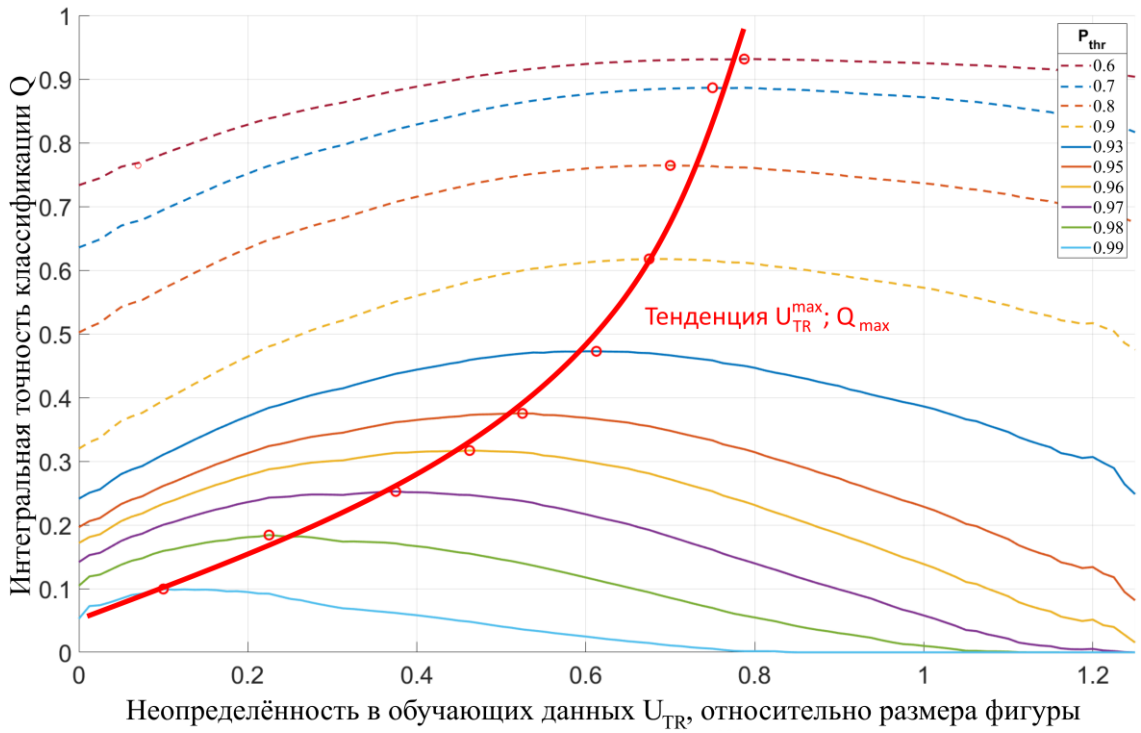

Рисунок 4 – График зависимости интегральной точности распознавания изображений Q от неопределённости в обучающих данных U_{TR} для разных значений P_{thr} и оптимальные значения неопределённости в обучающих данных U_{TR} для различных значений порога P_{thr}

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}. График на рисунке 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 максимумом кривой, положение этого максимума будет указывать на оптимальное значение неопределенности обучающего набора данных U_{TR}. 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 $U_{TR} = 0$).

Чтобы подтвердить правильность и характерность выводов для разных видов искажений, проведена серия экспериментов с различными типами искажений естественных изображений. Независимо от типа шума/искажения при обучении, его величина одинаково влияет на точность распознавания изображений. Существует оптимальное количество искажений в обучающих данных, приводящее к значительному улучшению помехоустойчивости обученной СНС и повышению общего качества распознавания.

В **главе 4** проведён анализ различных способов аугментации обучающих данных для достижения наилучшего качества распознавания изображений. Получены графики зависимости точности распознавания от интенсивности искажений в тестовых изображениях для пяти различных методов аугментации обучающих изображений. 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 приводит к низкому качеству распознавания изображений, искаженных размытием по Гауссу и быстрому спаду точности классификации изображений с ростом интенсивности размытия.

В **Главе 5** предложен метод противостояния высокочастотным шумам, основанный на фильтрации зашумлённых изображений с помощью низкочастотного фильтра Гаусса. 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. Таким образом, фильтрация изображений фильтром Гаусса позволяет свести проблему противостояния высокочастотным состязательным атакам к проблеме распознавания размытых изображений, уже рассмотренной ранее. Доказано наличие оптимума интенсивности размытия зашумлённых данных и предлагается метод нахождения этого оптимума. Рассмотрены 4 набора изображений и 2 архитектуры СНС. Блок-схема предложенного алгоритма обработки изображений для противостояния состязательным атакам приведена на рисунке 5.

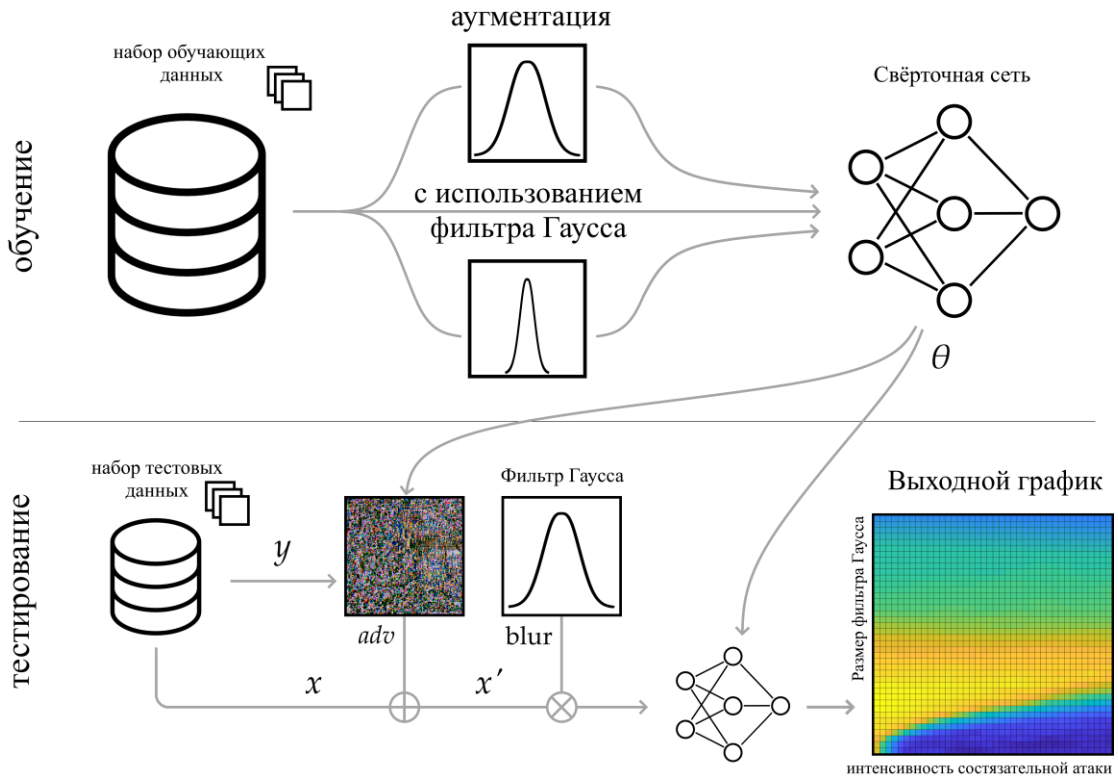

Рисунок 5 – Схема разработанного метода

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 и мелкие паттерны изображения.

Гауссовский фильтр значительно снижает влияние высокочастотной составляющей изображения, при этом общая структура изображения ухудшается гораздо менее значительно. На следующих графиках (рисунок 6) показана зависимость точности распознавания изображений от интенсивности атаки FGSM и от размера фильтра Гаусса. Интенсивность атаки FGSM измеряется в процентах от динамического диапазона изображения (DR). Размеры фильтра Гаусса измеряются в процентах от размера исходного изображения.

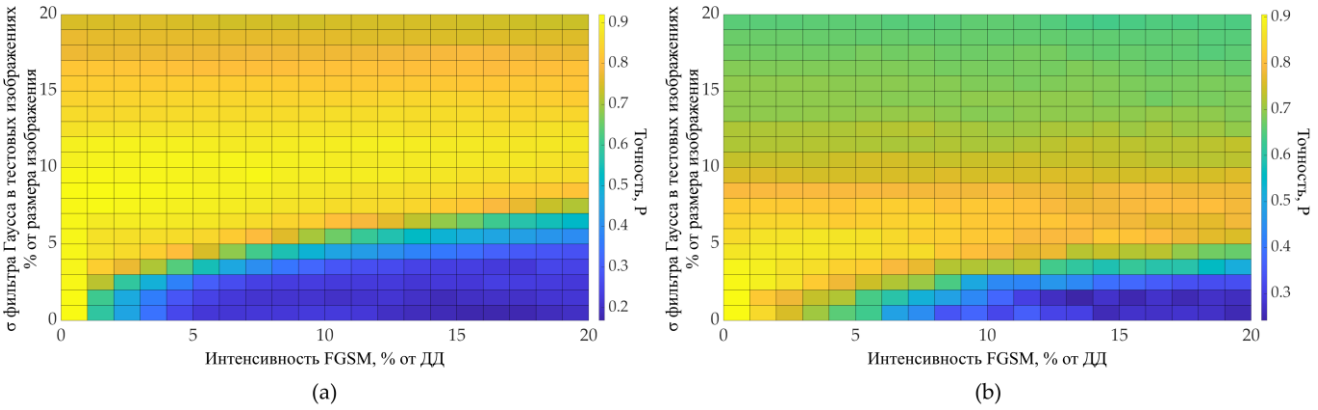

Рисунок 6 – Точность распознавания набора изображений Natural Dataset с применением простой СНС: (а) СНС обучена с применением аугментированного фильтром Гаусса набора данных; (б) обучение проводилось без аугментации

Точность распознавания изображений при наличии состязательных искажений быстро падает с ростом интенсивности этих искажений. При обработке состязательных тестовых изображений с применением фильтра Гаусса точность возрастает. 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 применение фильтра Гаусса становится малоэффективным. Оптимальное значение размеров фильтра Гаусса зависит от интенсивности состязательных искажений, а также от характеристик используемых данных и нейронной сети.

В таблице 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.

$$\sigma_{opt} = \arg\left(max\left(\sum_{I_{FGSM}=0}^{I_{FGSM}^{max}} P_{LPF}(\sigma, I_{FGSM})\right)\right) \quad (8)$$

, где σ_{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:

$$G = \frac{(1 - P_{no\ LPF})}{(1 - P_{LPF})} \quad (9)$$

, 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 1 – Точность классификации при различных интенсивностях искажений и возможный выигрыш в точности при использовании фильтра.

Нейронная сеть и Набор данных	**Интен-сив-ность 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	**0,913**	10	9,1
	10	0,206	**0,900**		7,9
	20	0,188	**0,894**		6,7

Сеть высокого быстродействия (RPS)	5	0,738	**0,947**	8	4,9
	10	0,660	**0,879**		2,8
	20	0,576	**0,738**		1,6
EfficientNet (ImageNet)	15	0,699	**0,781**	7	1,4
	20	0,481	**0,720**		1,9
EfficientNetB3 (Natural Dataset)	5	0,977	**1,000**	7	∞
	10	0,814	**0,996**		46,5
	20	0,250	**0,881**		6,3

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

В **Заключении** перечислены основные результаты и выводы данной диссертационной работы:

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 неопределенности 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 состязательной 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 публикаций автора по теме диссертации:

Статьи в журналах, индексируемых в базах данных Web of Science и Scopus:

1. Ziyadinov, V. V. Convolutional Neural Network Training Optimization for Low Point Density Image Recognition / V. V. Ziyadinov, P. S. Kurochkin, M. V. Tereshonok // Journal of Communications Technology and Electronics. – 2021. – Vol. 66, No. 12. – P. 1363-1369. – DOI 10.1134/S1064226921120202.

2. Ziyadinov, V. и др. A Survey on Symmetrical Neural Network Architectures and Applications // Symmetry. 2022. Т. 14. № 7. С. 1391.

3. Ziyadinov, V. Noise Immunity and Robustness Study of Image Recognition Using a Convolutional Neural Network / V. Ziyadinov, M. Tereshonok // Sensors. – 2022. – Vol. 22, No. 3. – DOI 10.3390/s22031241.

4. Ziyadinov V., Tereshonok M. Low-Pass Image Filtering to Achieve Adversarial Robustness // Sensors. 2023. Т. 23. № 22. С. 9032. DOI: 10.3390/s23229032.

Статьи в журналах из списка ВАК:

1. Зиядинов, В. В. Оптимизация обучения сверточных нейронных сетей при распознавании изображений с низкой плотностью точек / В. В. Зиядинов, П. С. Курочкин, М. В. Терешонок // Радиотехника и электроника. – 2021. – Т. 66, № 12. – С. 1207-1215. – DOI 10.31857/S0033849421120202.

2. 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. DOI: 10.36724/2072-8735-2021-15-4-49-56.

3. Зиядинов, В. В. Обнаружение автомобильных заторов с использованием кластерного анализа данных геолокации / В. В. Зиядинов, А. Б. Талалаев, М. В. Терешонок // Труды Научно-исследовательского института радио. – 2022. – № 2. – С. 28-39. – DOI 10.34832/NIIR.2022.9.2.003.

Материалы конференций, индексируемые в базах данных Web of Science и Scopus:

1. Ziyadinov, V. V. Analytical Survey on MANET and VANET Clusterisation Algorithms / V. V. Ziyadinov, M. V. Tereshonok // 2020 Systems of Signal

Synchronization, Generating and Processing in Telecommunications, SYNCHROINFO 2020, Svetlogorsk, 01–03 июля 2020 года. – Svetlogorsk, 2020. – P. 9166120. – DOI 10.1109/SYNCHROINFO49631.2020.9166120.

2. Ziyadinov, V. V. Neural Network Image Recognition Robustness with Different Augmentation Methods / V. V. Ziyadinov, M. V. Tereshonok // Systems of Signal Synchronization, Generating and Processing in Telecommunications. – 2022. – Vol. 5, No. 1. – P. 441-444. – DOI 10.1109/SYNCHROINFO55067.2022.9840987.

Свидетельства о государственной регистрации программ для ЭВМ:

1. Свидетельство о государственной регистрации программы для ЭВМ № 2024612396 РФ. Программный комплекс для демонстрации работы свёрточной нейронной сети, решающей задачу распознавания состязательных изображений: опубл. 31.01.2024.

2. Свидетельство о государственной регистрации программы для ЭВМ № 2024611339 РФ. Программный комплекс расчета внешних характеристик точности и устойчивости свёрточной нейронной сети к высокочастотным искажениям и оптимизации параметров предварительной обработки изображений : опубл. 19.01.2024.

3. Свидетельство о государственной регистрации программы для ЭВМ № 2022660463 РФ. Программа сравнительного анализа и визуализации результатов работы свёрточных нейронных сетей : опубл. 03.06.2022.

4. Свидетельство о государственной регистрации программы для ЭВМ № 2022660552 РФ. Программа моделирования шума в реальных изображениях и генерации обучающих выборок для систем распознавания : опубл. 06.06.2022.

5. Свидетельство о государственной регистрации программы для ЭВМ № 2022660553 РФ. Программа оценки результативности работы алгоритмов кластеризации : опубл. 06.06.2022.

6. Свидетельство о государственной регистрации программы для ЭВМ № 2022660554 РФ. Программа визуализации характеристик обучения свёрточных нейронных сетей для определения оптимальных параметров обучающих выборок при требуемой минимальной точности классификации : опубл. 06.06.2022.

7. Свидетельство о государственной регистрации программы для ЭВМ № 2021619356 РФ. Программа для оптимизации работы свёрточных нейронных сетей : опубл. 08.06.2021.

8. Свидетельство о государственной регистрации программы для ЭВМ № 2021619626 РФ. Программа генерации обучающих выборок для систем распознавания изображений с низкой плотностью точек : опубл. 15.06.2021.

9. Свидетельство о государственной регистрации программы для ЭВМ № 2020660537 РФ. Моделирование типов взаимного расположения абонентов сетей мобильной связи : опубл. 04.09.2020.

Прочие публикации:

1. Зиядинов, В. В. Математические модели и методы распознавания взаимного расположения мобильных абонентов / В. В. Зиядинов, М. В. Терешонок // Технологии информационного общества : Сборник трудов XIV Международной отраслевой научно-технической конференции, Москва, 18–19 марта 2020 года. – Москва: ООО "Издательский дом Медиа паблишер", 2020. – С. 157-159.