

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

На правах рукописи

Ильина Ольга Викторовна

**Повышение устойчивости и точности нейросетевого обнаружения
искажённых объектов при видовом мониторинге**

Специальность 2.2.13

Радиотехника, в том числе системы и устройства телевидения

Диссертация на соискание учёной степени
кандидата технических наук

Научный руководитель:
доктор технических наук, профессор
Терешонок Максим Валерьевич

Москва 2026

ОГЛАВЛЕНИЕ

ВВЕДЕНИЕ.....	5
1. АНАЛИЗ СУЩЕСТВУЮЩИХ СПОСОБОВ НЕЙРОСЕТЕВОГО ОБНАРУЖЕНИЯ ОБЪЕКТОВ ПРИ НАЛИЧИИ ЕСТЕСТВЕННЫХ И ИСКУССТВЕННЫХ ИСКАЖЕНИЙ.....	13
1.1 Анализ существующих методов обнаружения и распознавания объектов на изображениях при видовом мониторинге	13
1.2 Анализ существующих метрик оценки качества и робастности модели обнаружения объектов на изображениях.....	15
1.2.1 Метрики оценки качества.....	15
1.2.2 Метрики оценки робастности	19
1.3 Анализ существующих способов обучения нейросетевых систем, решающих задачу компьютерного зрения для устранения влияния естественных искажений	20
1.4 Анализ существующих методов устранения влияния состязательных патч- атак.....	24
1.5 Границы исследования.....	27
1.6 Постановка научной задачи исследования	29
Выводы по разделу 1.....	32
2. РАЗРАБОТКА СПОСОБА ФОРМИРОВАНИЯ ОБУЧАЮЩЕГО НАБОРА ДАННЫХ ДЛЯ ПОВЫШЕНИЯ УСТОЙЧИВОСТИ МОДЕЛИ ОБНАРУЖЕНИЯ ОБЪЕКТОВ ПРИ ВИДОВОМ МОНИТОРИНГЕ	34
2.1 Исследование влияние введения АБГШ в обучающие изображения нейросетевой модели обнаружения объектов.....	34
2.2 Исследование влияния введения гауссова размытия в обучающие изображения нейросетевой модели обнаружения объектов.....	39

2.2.1 Применение гауссова размытия к изображениям.....	39
2.2.2 Разработка способа формирования обучающего набора данных для повышения устойчивости анализируемой модели обнаружения объектов	41
2.2.3 Оценка устойчивости модели обнаружения объектов при обучении с использованием разработанного способа формирования обучающего набора данных	44
Выводы по разделу 2.....	50
3..РАЗРАБОТКА МЕТОДА АДАПТИВНОЙ ПРЕДВАРИТЕЛЬНОЙ ОБРАБОТКИ ИЗОБРАЖЕНИЙ ПРИ ВИДОВОМ МОНИТОРИНГЕ	52
3.1 Исследование влияния моделей естественных искажений на нейросетевую модель обнаружения объектов, обученную разработанным способом.....	52
3.2 Разработка адаптивной схемы для робастного обнаружения и распознавания объектов различного качества при видовом мониторинге	55
3.3 Оценка эффективности разработанной адаптивной схемы	58
Выводы по разделу 3.....	65
4. РАЗРАБОТКА МЕТОДА ЛОКАЛИЗАЦИИ И УДАЛЕНИЯ СОСЯЗАТЕЛЬНЫХ ПАТЧ-АТАК НА ИЗОБРАЖЕНИЯХ ПРИ ВИДОВОМ МОНИТОРИНГЕ.....	67
4.1 Описание разработанного метода локализации и удаления состязательных патч-атак на изображениях	67
4.1.1 Автокодировщик	68
4.1.2 Вычисление максимальной ошибки между исходным и сгенерированным изображениями	71
4.1.3 Алгоритм локализации аномальных фрагментов	73
4.1.4 Алгоритм кластеризации и обработки карты аномальных фрагментов.....	75

4.1.5 Алгоритм применения маски аномалий к изображению	79
4.2 Оценка эффективности разработанного метода.....	80
Выводы по разделу 4.....	85
ЗАКЛЮЧЕНИЕ	87
СПИСОК СОКРАЩЕНИЙ И УСЛОВНЫХ ОБОЗНАЧЕНИЙ.....	90
СЛОВАРЬ ТЕРМИНОВ.....	91
СПИСОК ЛИТЕРАТУРЫ.....	92
ПРИЛОЖЕНИЕ А	105
ПРИЛОЖЕНИЕ Б.....	111

ВВЕДЕНИЕ

Актуальность темы исследования. В последние годы активно развиваются технологии видового мониторинга, основанные на дистанционном зондировании Земли, что обеспечивает получение больших объемов изображений земной поверхности с высоким пространственным разрешением [1]. Визуальный анализ таких данных востребован при мониторинге транспортных средств, объектов инфраструктуры и природных территорий. Объем и сложность поступающих данных обуславливают необходимость автоматизации их обработки и применения нейросетевых методов обнаружения объектов. Современные исследования подтверждают широкое применение глубокого обучения и нейросетевых моделей обнаружения объектов для анализа аэрофото- и спутниковых изображений. При этом для изображений дистанционного зондирования характерны сложный фон, высокая вариативность масштаба, ориентации и формы объектов, а также наличие большого числа мелких объектов на одном изображении [2, 3].

Обнаружение объектов является одной из ключевых задач компьютерного зрения и заключается в одновременном определении пространственного положения объектов на изображении, класса этих объектов и оценки показателя уверенности своего предсказания. Для задач видового мониторинга особенно важна возможность автоматического обнаружения множества объектов различного размера и ориентации. Развитие методов нейросетевого обнаружения объектов позволило существенно повысить эффективность автоматического анализа подобных данных, однако перенос достигнутых результатов в реальные условия эксплуатации остается сложной задачей.

Эффективность нейросетевых моделей обнаружения объектов может существенно снижаться при ухудшении качества входных изображений. При формировании аэрофотоснимков на качество данных влияют шумы, размытие, изменения освещенности, атмосферные явления, особенности оптической системы и другие факторы. В исследованиях по робастности моделей обнаружения

объектов показано, что современные модели обнаружения объектов могут значительно терять эффективность при воздействии естественных искажений, причем снижение качества зависит как от типа, так и от интенсивности искажения [4].

Дополнительную проблему представляет сдвиг распределения данных между условиями обучения и эксплуатации нейросетевой модели. Изменение характеристик входных изображений относительно обучающего распределения может приводить к снижению обобщающей способности модели и ухудшению качества обнаружения объектов. В связи с этим повышение робастности нейросетевых моделей обнаружения объектов к изменениям условий формирования изображений является важным направлением современных исследований [5].

Наряду с естественными искажениями нейросетевые системы видового мониторинга подвержены преднамеренным состязательным воздействиям. В частности, состязательные патч-атаки позволяют локально изменять изображение таким образом, чтобы нарушать работу модели обнаружения объектов. Для аэрофотоснимков показана практическая реализуемость физических состязательных патч-атак: специально оптимизированные и напечатанные патчи могут быть размещены на реальных объектах и приводить к существенному снижению эффективности их обнаружения на снимках с воздуха [6, 7]. Аналогичная уязвимость продемонстрирована для систем обнаружения людей в задачах видеомониторинга [8].

Таким образом, современные системы видового мониторинга должны обеспечивать не только высокую точность обнаружения объектов в стандартных условиях, но и устойчивость к изменению характеристик входных данных и преднамеренным воздействиям. Разработка методов повышения точности и робастности нейросетевого обнаружения объектов при естественных искажениях и состязательных атаках является актуальной научной задачей, имеющей практическое значение для систем дистанционного зондирования Земли, транспортного мониторинга и других приложений компьютерного зрения.

Степень разработанности темы исследования. Важный вклад в развитие тематики робастности искусственных нейронных сетей внесли С.А. Доленко [9, 10, 11], А.И. Гладышев [12, 13], М.В. Терешонок [14, 15, 16, 17], В.В. Зиядинов [18, 19], И.В. Оселедец [20, 21, 22], I.J. Goodfellow [23, 24], С. Szegedy [25], D. Hendrycks [26, 27].

Существенный прогресс в области обнаружения объектов на изображениях связан с развитием глубоких сверточных нейронных сетей и появлением современных архитектур обнаружения объектов, среди которых наибольшее распространение получили двухэтапные обнаружители объектов семейства R-CNN [28] и одноэтапные архитектуры семейства YOLO [29, 30, 31]. Эти системы заложили основу современных систем обнаружения объектов, обеспечивающих высокую точность и скорость обработки изображений. В последние годы данные архитектуры успешно применяются в задачах дистанционного зондирования Земли, интеллектуального транспорта, автономной робототехники и видового мониторинга, однако их эффективность существенно снижается при отклонении условий эксплуатации от условий обучения вследствие сдвига домена, возникающего при изменении распределения входных данных [32].

Большое количество исследований посвящено повышению робастности моделей компьютерного зрения к естественным искажениям изображений. Наиболее распространёнными направлениями являются аугментация обучающих данных, обучение на изображениях с различными видами деградации, предварительное восстановление изображений, а также разработка специализированных архитектур, менее чувствительных к шуму и размытию [33, 34, 35]. Существенный вклад в развитие данного направления внесли исследования [26], в которых сформированы современные подходы к оценке робастности моделей глубокого обучения. Вместе с тем большинство существующих работ посвящено задачам классификации изображений, тогда как вопросам повышения устойчивости нейросетевых моделей обнаружения объектов к совместному воздействию различных естественных искажений уделено значительно меньше внимания.

Отдельным направлением исследований является защита нейросетевых моделей от состязательных воздействий. После фундаментальных работ Ian Goodfellow [23] и Aleksander Madry [36] были предложены многочисленные методы повышения устойчивости моделей, основанные на состязательном обучении, локализации патчей, использовании сегментационных и генеративных моделей, а также сертифицированных алгоритмах защиты [37].

Несмотря на большое количество исследований в области повышения робастности нейросетевых моделей компьютерного зрения, большинство существующих работ рассматривает отдельные виды естественных искажений либо ориентировано на задачи классификации изображений. Недостаточно исследованы вопросы формирования оптимального распределения уровней искажений, вносимых в процесс обучения моделей обнаружения объектов, адаптивной предварительной обработки изображений, согласованной со статистическими характеристиками обучающих изображений, а также повышения устойчивости моделей одновременно к преднамеренным состязательным атакам без модификации нейросетевой архитектуры.

Целью диссертационной работы является повышение точности и робастности (устойчивости) функционирования нейросетевых моделей обнаружения объектов в условиях воздействия естественных искажений и преднамеренных состязательных атак при видовом мониторинге.

Научная задача исследования состоит в разработке комплекса методов повышения точности и робастности нейросетевого обнаружения объектов в условиях естественных и преднамеренных искажений входных изображений при видовом мониторинге.

Для достижения поставленной цели и решения научной задачи в работе были решены следующие **частные задачи**:

1. Разработка способа формирования обучающего набора данных для повышения устойчивости модели обнаружения объектов при видовом мониторинге.

2. Разработка метода адаптивной предварительной обработки изображений при видовом мониторинге.

3. Разработка метода локализации и удаления состязательных патч-атак на изображениях при видовом мониторинге.

Научная новизна работы. В диссертационной работе исследование проводится для видового мониторинга, где изображения характеризуются наличием объектов, существенно различающихся по масштабу и пространственному расположению на изображении. В отличие от классификации, где оценивается корректность определения класса изображения, задача обнаружения одновременно требует определить наличие объекта, корректно классифицировать его и локализовать его местоположение. Перенос результатов, полученных для классификации, на задачу обнаружения объектов не является непосредственным, поскольку эти задачи имеют разные критерии качества и по-разному реагируют на искажения входных изображений.

Научная новизна диссертационной работы заключается в разработке методов повышения точности и робастности нейросетевого обнаружения объектов при естественных и искусственных искажениях при сохранении высокой точности обнаружения объектов на неискажённых изображениях. Впервые для рассматриваемой задачи предложен способ формирования обучающего набора изображений, заключающийся в целенаправленном добавлении искажённых изображений с заданным распределением уровней искажений; разработан адаптивный метод предварительной обработки изображений на основе оценки их статистической неоднородности; а также разработан метод локализации и удаления состязательных патч-атак на основе реконструкции изображения и анализа карты ошибок.

Теоретическая значимость работы заключается в развитии математического аппарата глубокого обучения и теории робастного компьютерного зрения в условиях искажающих воздействий на входные изображения. Результаты диссертационного исследования углубляют

теоретическую базу систем искусственного интеллекта для дальнейшего проектирования защищенных и помехоустойчивых систем компьютерного зрения.

Практическая значимость работы заключается в повышении точности обнаружения объектов на тестовых изображениях, что может быть использовано в различных практических приложениях. Практическая значимость подтверждена актом использования результатов диссертационной работы в МТУСИ при выполнении НИОКР.

Методология и методы исследования: методы численного моделирования, теории вероятностей, математической статистики, теории машинного обучения, теории искусственного интеллекта, цифровой обработки изображений.

Объектом исследования являются глубокие свёрточные нейронные сети, решающие задачу обнаружения объектов на изображении.

Предметом исследования являются характеристики устойчивости и робастности глубоких свёрточных нейронных сетей, решающих задачу обнаружения объектов.

Положения, выносимые на защиту:

1. Разработанный способ обучения модели обнаружения объектов, основанный на формировании обучающего набора изображений с использованием искаженных изображений, обеспечивает увеличение метрики полноты обнаружения на тестовых данных более чем в 2 раза при воздействии аддитивного белого гауссова шума (отношение сигнал/шум от -8 до 0 дБ) и при воздействии гауссова размытия с радиусом от 6 до 12 пикселей по сравнению с моделью, обученной на неискаженных изображениях.

2. Разработанный адаптивный метод предварительной обработки изображений, обеспечивающий автоматический выбор режима фильтрации на основе оценки статистической неоднородности входного изображения, совместно с разработанным методом робастного обучения позволяет повысить интегральную оценку помехоустойчивости модели в 1,4 раза по сравнению с исходной моделью.

3. Разработанный метод локализации и удаления состязательных патч-атак, основанный на реконструкции изображения и анализе локальных статистических характеристик карты ошибок реконструкции, увеличивает значение метрики качества обнаружения mAP с 46,8% до 81,0% при сохранении высокой точности обнаружения на изображениях без атак (mAP = 93,5%).

Степень достоверности. Достоверность полученных результатов подтверждается обоснованием используемого математического аппарата, а также верификацией разработанных методов на общедоступных наборах тестовых изображений.

Апробация результатов. Материалы диссертационной работы были доложены и одобрены на пяти научно-технических конференциях:

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

- 2023 Intelligent Technologies and Electronic Devices in Vehicle and Road Transport Complex (TIRVED), Москва, Россия, 15 – 17 ноября 2023 г.

- Научно-исследовательский семинар по физике наноструктур, атомной и молекулярной физике, Москва, Россия, 8 сентября 2022 год;

- Научно-исследовательский семинар по физике наноструктур и квантовым технологиям, Москва, Россия, 27 августа 2026 год;

- Международная конференция «Квантовые вычисления на сверхпроводниках», Сколково, Россия, 3 сентября 2026 года.

Публикации результатов. По теме диссертации было опубликовано 13 работ [38 - 50]. Из них 3 работы опубликованы в журналах из белого списка и проиндексированы в международных базах данных Web of Science и Scopus [38, 39, 40], 2 работы опубликованы в журналах из Перечня ВАК [41, 42], ещё 2 работы проиндексированы в международной базе данных Scopus [43, 44], получено 6 свидетельств о государственной регистрации программ для ЭВМ (Приложение А) [45 - 50].

Реализация и внедрение результатов. Разработанные в ходе диссертационного исследования методы и способы использованы в НИР «Шеренга-2020», СЧ ОКР «5P17K302-МТУСИ» и НИР «Лакмус-СПП», выполненных и выполняемых в рамках Государственных контрактов в МТУСИ с 2018 года по настоящее время (Приложение Б).

Личный вклад. Все выносимые на защиту научные результаты получены соискателем лично. Из работ, опубликованных в соавторстве, в диссертацию включена только та их часть, которая получена лично соискателем.

Соответствие паспорту научной специальности. Диссертация соответствует пункту 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» паспорта научной специальности 2.2.13 «Радиотехника, в том числе системы и устройства телевидения».

Объем и структура работы. Текст диссертационной работы изложен на 112 печатных страницах и включает введение, четыре раздела, заключение, список сокращений и условных обозначений, список терминов, список литературы и приложения (на 8 печатных страницах). Список литературы содержит 97 наименований. В работе представлены 23 рисунка и 1 таблица.

1. АНАЛИЗ СУЩЕСТВУЮЩИХ СПОСОБОВ НЕЙРОСЕТЕВОГО ОБНАРУЖЕНИЯ ОБЪЕКТОВ ПРИ НАЛИЧИИ ЕСТЕСТВЕННЫХ И ИСКУССТВЕННЫХ ИСКАЖЕНИЙ

1.1 Анализ существующих методов обнаружения и распознавания объектов на изображениях при видовом мониторинге

Обнаружение объектов представляет собой одну из фундаментальных задач компьютерного зрения и обработки изображений, заключающуюся в определении класса объектов, присутствующих на изображении, а также их местоположения. При этом обнаруживаемые объекты могут наблюдаться в произвольном количестве, иметь произвольное расположение и ориентацию, а также различные размеры и формы.

В настоящее время разработано большое количество архитектур глубоких нейронных сетей, различающихся структурой, способом организации вычислительных блоков и характером взаимодействия между слоями, что позволяет применять их для решения широкого спектра задач компьютерного зрения [38]. Среди них значительное распространение получили методы, основанные на глубоких сверточных нейронных сетях (ГСНС), обладающих способностью автоматически извлекать высокоуровневые признаки из исходных данных без необходимости их ручного проектирования [51].

Современные алгоритмы нейросетевого обнаружения объектов на изображениях можно разделить на две категории: одноэтапные и двухэтапные. В одноэтапных детекторах локализация объектов и их классификация выполняются одновременно за один проход изображения через нейронную сеть. Наиболее широкое распространение среди одноэтапных обнаружителей объектов получило семейство архитектур YOLO [52]. Наиболее распространенным представителем двухэтапных обнаружителей объектов является архитектура Faster R-CNN [28], в которой обработка изображения включает два последовательных

этапа: генерацию предложений регионов с помощью сети Region Proposal Network (RPN) и последующую классификацию объектов, выделенных на предыдущем этапе. Благодаря такому подходу обеспечивается высокая точность как локализации, так и распознавания объектов.

Выбор архитектуры определяется требованиями конкретной задачи. Двухэтапные обнаружители объектов обеспечивают высокую точность локализации и распознавания объектов, однако отличаются большей вычислительной сложностью. В то же время современные одноэтапные архитектуры семейства YOLO сочетают высокую скорость обработки с точностью, сопоставимой с двухэтапными решениями во многих практических сценариях, что делает их эффективным выбором для систем реального времени и широкого круга задач компьютерного зрения.

В настоящей работе под данными, поступающими на вход нейросетевой модели обнаружения понимается совокупность цифровых изображений, используемых для обучения, настройки и проверки нейросетевых моделей. Основным типом данных являются цветные изображения, полученные в условиях видового мониторинга. Цифровое изображение в рамках настоящего исследования рассматривается как двумерная матрица значений яркости с несколькими цветовыми каналами. Для цветного изображения в формате RGB каждому пикселю соответствуют три значения, характеризующие интенсивность красной, зеленой и синей составляющих цвета. При использовании стандартного 8-битного представления каждое из значений принимает целочисленное значение в диапазоне от 0 до 255, где 0 соответствует минимальной, а 255 – максимальной интенсивности соответствующей цветовой компоненты. В ряде вычислительных процедур диссертации значения пикселей предварительно нормируются к диапазону от 0 до 1.

Обучающий набор изображений представляет собой подмножество исходных данных, предназначенное для настройки весовых параметров нейросетевой модели обнаружения объектов в процессе обучения. В настоящей работе обучающий набор изображений может содержать как исходные

неискаженные изображения, так и изображения, полученные в результате контролируемого внесения естественных искажений.

Тестовый набор изображений представляет собой совокупность данных, не использовавшихся непосредственно для настройки весовых параметров нейросетевой модели обнаружения объектов, и предназначенную для оценки качества и робастности ее работы. В настоящем исследовании тестовые изображения формировались таким образом, чтобы обеспечить контролируемое изменение типа и интенсивности искажений. Для этого к тестовым изображениям применялись различные уровни искажений.

При видовом мониторинге обнаружение объектов является первым этапом автоматизированной обработки аэрофотоснимков. Особенностью данной задачи является необходимость обнаружения объектов малого размера, наблюдаемых в различных масштабах, при наличии естественных искажений, обусловленных условиями съемки, характеристиками сенсора и особенностями окружающей среды.

1.2 Анализ существующих метрик оценки качества и робастности модели обнаружения объектов на изображениях

1.2.1 Метрики оценки качества

Системы обнаружения и распознавания объектов на изображении решают задачи локализации объектов (предсказание координат ограничивающих рамок вокруг объекта, когда объект присутствует на изображении) и распознавание объектов (определение класса объектов). Величина полноты обнаружения позволяет оценить способность модели обнаруживать все релевантные объекты и показывает долю правильных предсказаний от общего числа истинных объектов. Величина точности обнаружения показывает способность модели идентифицировать только релевантные объекты и равна отношению количества правильных обнаружений к общему числу предсказаний [53]. Для вычисления

полноты и точности обнаружения каждый предсказанный системой обнаружения ограничивающий прямоугольник сначала должен быть классифицирован как:

- истинно-положительный (true positive, TP) при корректной локализации объекта и правильном определении его класса;
- ложноположительный (false positive, FP) при ошибочной локализации объекта либо неверной классификации обнаруженного объекта;
- ложноотрицательный (false negative, FN) при отсутствии обнаружения истинного объекта.

Формулы для вычисления метрик полноты R и точности P обнаружения [53]:

$$R = \frac{\sum_{n=1}^S TP_n}{\sum_{n=1}^S TP_n + \sum_{n=1}^{G-S} FN_n} \quad (1)$$

$$P = \frac{\sum_{n=1}^S TP_n}{\sum_{n=1}^S TP_n + \sum_{n=1}^{N-S} FP_n} \quad (2)$$

где G – количество всех истинных объектов, N – количество предсказанных системой обнаружения объектов, из которых S объектов предсказаны верно ($S \leq G$), n – номер предсказанного ограничивающего потенциальный объект прямоугольника.

В случае обнаружения объектов на изображении необходимо вычислить метрику отношения площади пересечения к площади объединения истинного и предсказанного прямоугольника (Intersection Over Union, IoU), которая показывает, насколько близки обнаруженные сетью ограничивающие рамки к ограничивающим рамкам истинных объектов. Это измерение выполняется независимо для каждого класса объектов путем оценки степени перекрытия предсказанной и достоверной областей [53]. Полное совпадение двух ограничивающих рамок происходит, когда метрика IoU между ними становится равной 1. В случае, когда оба ограничивающих прямоугольника не пересекают друг друга, метрика IoU равна 0. Чем ближе к единице становится IoU, тем лучше

считается обнаружение. Поскольку обнаружители объектов также выполняют классификацию каждого ограничивающего прямоугольника, только истинные и обнаруженные ограничивающие рамки одного и того же класса сравниваются метрикой IoU.

После вычисления метрики IoU для всех истинных и всех предсказанных ограничивающих рамок необходимо установить пороговое значение метрики IoU для определения правильности предсказанных ограничивающих рамок. В случае, когда IoU больше или равно заданному пороговому значению, предсказание является истинно-положительным (TP), но в случае, когда IoU меньше заданного порогового значения, то предсказание является ложно положительным (FP).

Для комплексной оценки качества алгоритмов обнаружения объектов широко используется метрика средней точности Mean Average Precision (mAP). Данная метрика учитывает одновременно способность модели корректно обнаруживать объекты и точность локализации ограничивающих рамок. В основе вычисления mAP лежит метрика усредненной точности Average Precision (AP), которая рассчитывается отдельно для каждого класса объектов. Для её получения строится кривая зависимости точности P от полноты R (PR-кривая), отражающая изменение точности и полноты при варьировании порога уверенности обнаружителя объектов. Метрика AP определяется как площадь под PR-кривой:

$$AP = \int_0^1 P(R) dR. \quad (3)$$

На практике интеграл вычисляется численно по дискретным точкам PR-кривой. После вычисления AP для каждого класса рассчитывается метрика mAP, представляющая собой среднее значение AP по всем классам:

$$mAP = \frac{1}{K} \sum_{k=1}^K AP_k, \quad (4)$$

где K – количество классов объектов;

AP_k – значение метрики AP k -го класса.

В то же время для оценки качества локализации объектов широко используется метрика усредненной полноты (от англ. Average Recall, AR). В отличие от AP, которая учитывает как качество локализации, так и точность классификации объектов, AR характеризует способность обнаружителя находить объекты независимо от достоверности присвоенного класса. По этой причине AR часто применяется для оценки качества алгоритмов генерации регионов интереса RPN, а также для анализа полноты обнаружения объектов.

Величина AR представляет собой усредненное значение полноты R , вычисленное по набору порогов IoU. В общем случае данная метрика определяется выражением

$$AR = \frac{1}{T} \sum_t^T R(t), \quad (5)$$

где T – множество пороговых значений IoU;

$R(t)$ – значение полноты обнаружения при пороге t .

Усредненное значение полноты вычисляется по десяти порогам IoU в диапазоне от 0,50 до 0,95 с шагом 0,05. Высокое значение усредненной полноты свидетельствует о том, что детектор обнаруживает большую часть присутствующих на изображении объектов и обеспечивает высокое качество локализации.

Таким образом, метрика mAP позволяет получить интегральную оценку качества обнаружителя объектов, учитывая одновременно точность классификации, полноту обнаружения и качество локализации объектов на изображении. В то же время метрика AR позволяет более полно оценить эффективность алгоритма обнаружения объектов находить релевантные объекты и при анализе качества генерации ограничивающих прямоугольников.

1.2.2 Метрики оценки робастности

Несмотря на широкое применение перечисленных метрик качества обнаружения объектов, они позволяют оценивать эффективность обнаружителя объектов только в условиях, близких к распределению обучающих данных. Однако в практических сценариях эксплуатации системы компьютерного зрения часто функционируют в условиях, отличающихся от условий обучения. К таким условиям относятся изменения освещенности, атмосферные явления, размытие вследствие движения или расфокусировки, цифровой шум, сжатие изображения, а также другие естественные искажения, возникающие в процессе формирования изображения. В связи с этим наряду с оценкой точности детектирования важной характеристикой современных моделей является их робастность.

Под робастностью (от англ. *robustness*) понимается способность модели сохранять работоспособность и обеспечивать приемлемое качество обнаружения объектов при наличии изменений входных данных, не изменяющих их семантическое содержание. Другими словами, робастная система должна корректно обнаруживать объекты не только на исходных изображениях, но и в условиях естественных искажений, возникающих при эксплуатации системы. В литературе робастность рассматривается как одна из ключевых характеристик надежности моделей компьютерного зрения, отражающая их способность к обобщению на изображения, отличающиеся от обучающего распределения [4, 54, 55].

Оценка робастности заключается в сравнении качества работы модели на исходных изображениях и на изображениях, содержащих контролируемые искажения различного типа и интенсивности. Для этого используется набор тестовых данных, к которому последовательно применяются различные виды естественных искажений. После этого для каждого уровня искажения вычисляются стандартные метрики обнаружения объектов.

В качестве естественных искажений обычно рассматриваются аддитивный белый гауссов шум (АБГШ), импульсный шум, размытие вследствие

дефокусировки или движения камеры, изменение яркости и контрастности, атмосферные эффекты (дождь, туман, снег), Joint Photographic Experts Group (JPEG)-сжатие, а также другие виды деградации качества изображения [56, 57]. Такие искажения моделируют реальные условия эксплуатации систем компьютерного зрения и позволяют исследовать устойчивость алгоритма к наиболее распространенным внешним воздействиям.

Наиболее распространенным способом оценки робастности является анализ зависимости значения метрики качества системы от уровня интенсивности искажения. Чем меньше снижение метрики качества относительно исходных изображений, тем более робастной считается модель.

Таким образом, оценка робастности дополняет традиционные метрики качества обнаружения объектов и позволяет определить способность модели сохранять высокую эффективность при наличии искажений различной природы и интенсивности, возникающих в реальных условиях эксплуатации.

1.3 Анализ существующих способов обучения нейросетевых систем, решающих задачу компьютерного зрения для устранения влияния естественных искажений

В настоящее время существует большой интерес в области применения нейронных сетей для устранения размытия на входном изображении, но не для определения класса объекта на этом изображении [58 - 62]. Также существуют работы, направленные на усложнение архитектур ГСНС для повышения точности классификации объектов на изображении. Так, авторы [63] предложили усовершенствовать ГСНС для задач обнаружения лиц путем добавления ансамбля ветвей к изначальной ГСНС, а также путем расширения обучающего набора изображений её размытой копией. В статье [64] предложено несколько вариантов улучшенных моделей ГСНС для распознавания изображений путем робастного предварительного обучения. Для этого авторы предлагают

обучить сверточную часть модели реконструировать исходное изображение из искаженного, после чего полная модель обучается уже распознавать образы на искаженных изображениях.

В ряде источников представлены исследования влияния различного типа и степени искажений изображений на качество работы ГСНС. Так, исследование влияния внешних объектов (линейка, волосы) и качества изображения (размытие, шум, контраст) на классификацию меланомы с использованием современных моделей ГСНС проведено в [65], где авторы эмпирическим путем продемонстрировали способности различных ГСНС справляться с тем или иным видом искажения входных данных. В работе [66] показано, что тонкая настройка первых нескольких слоев ГСНС и обучение всей сети с нуля на искаженных данных положительно влияет на качество классификации искаженных тестовых наборов изображений. Однако, авторы предполагают, что обучение ГСНС с нуля на искаженных данных может привести к адаптации модели распознавать исключительно искаженные изображения. Практически одновременно с предыдущей работой вышла статья [67], в которой авторы сравнивают качество распознавания ГСНС (в том числе, полученных после тонкой настройки последнего полносвязного слоя на частично неискаженных и частично искаженных изображениях одновременно) с наблюдениями людей на искаженных изображениях. Проведенные в данной работе исследования показали, что, хотя ГСНС работают лучше или наравне с людьми на изображениях хорошего качества, на искаженных изображениях качество работы ГСНС все еще намного ниже, чем у человека. Также в работе показано, что качество распознавания человеком снижается быстрее для размытых изображений по сравнению с зашумленными изображениями, тогда как для ГСНС наблюдается противоположная тенденция. Проведенный в данной работе анализ матриц ошибок показал, что существует низкая корреляция между ошибками, совершенных ГСНС и людьми. Это может быть признаком того, что внутреннее представление изображений отличается между ГСНС и зрительной системой человека.

Авторы работ [57], [68] исследовали качество классификации ГСНС на изображениях с добавленными разнообразными видами искажений, где показали, что при обучении ГСНС на обучающем наборе данных, включающем несколько уровней искажений, точность классификации на тестовом наборе данных с разными уровнями искажений выше, чем при обучении на фиксированном уровне искажений. Однако, данный вывод не распространяется на точность классификации на неискажённых данных. В [68] продемонстрировано, что точность классификации будет наивысшей при обучении и тестирования ГСНС на наборах данных с одинаковым фиксированным уровнем искажений, а в случае, когда разница между фиксированным уровнем искажений обучающих данных и фиксированным уровнем искажений тестовых данных велика, то точность классификации на тестовом наборе данных может быть очень низкой. Также авторы показали, что предварительная очистка тестовых изображений от искажений современными алгоритмами (а именно, удаление тумана и размытия движения) не улучшает качество классификации ГСНС по сравнению с методом добавления искаженных данных в процесс обучения сети. Для повышения робастности авторы [57] использовали предварительно обученную современную архитектуру ГСНС и дообучили ее за три эпохи на неискаженных и искаженных изображениях с различной степенью искажений одновременно, что привело к значительному увеличению робастности такой ГСНС. Также авторы исследовали аналогичным образом стабильность системы сегментации объектов на изображении и обнаружили, что дообучение такой системы на неискаженных и искаженных изображениях хоть и увеличивает качество работы системы сегментации объектов на изображении, однако существует большой разрыв между качеством сегментации неискаженных и размытых изображений. Отмечается большее влияние искажений на качество сегментации по сравнению с качеством распознавания. Авторы считают, что это связано с тем, что система сегментации объектов решает задачу идентификации и разделения объектов, а размытие границ объектов сильно затрудняет эту задачу. Повышение качества распознавания изображений путём добавления искажений в обучающий набор изображений

подробно исследовано в работе [5]. Для большого разнообразия искажающих воздействий в [5] показано, что существует оптимальное значение интенсивности (или уровня) добавляемого искажения в обучающий набор данных.

В работе [18] исследуется влияние различных способов аугментации обучающего набора изображений на устойчивость нейросетевой модели к искажению изображений. Авторами проведено численное моделирование зависимости точности распознавания от размера ядра гауссова размытия для различных вариантов формирования обучающего набора. Рассматривались, в частности, обучение на исходных данных и обучение на данных, часть которых подвергалась гауссову размытию. Показано, что способ формирования обучающего набора изображений существенно влияет как на абсолютную точность распознавания, так и на характер зависимости качества модели от степени размытия тестовых изображений. Наиболее устойчивое поведение было получено при аугментации, при которой часть обучающих изображений подвергалась искажению, что позволило повысить качество распознавания на искаженных данных. Работа [19] посвящена повышению устойчивости сверточных нейронных сетей за счет предварительной фильтрации входных изображений. Авторами исследуется предположение о том, что высокая чувствительность нейросетей к локальным состязательным воздействиям связана с наличием высокочастотных компонент в искажении. Для подавления таких составляющих предлагается использовать низкочастотную фильтрацию изображения, реализованную с помощью гауссова фильтра. Данные исследования рассматривают преимущественно задачу классификации изображений, что ограничивает непосредственный перенос полученных результатов на задачу обнаружения объектов, где помимо классификации требуется точная локализация объекта.

Таким образом, анализ существующих исследований показывает, что большинство методов повышения устойчивости разработано для задач классификации изображений, в которых объект занимает центральную часть изображения и отсутствует необходимость его локализации. Для задачи обнаружения объектов подобные исследования представлены значительно меньше

и преимущественно ограничиваются изучением отдельных видов искажений. При этом практически отсутствуют работы, посвященные повышению устойчивости современных обнаружителей объектов к одновременному воздействию нескольких естественных искажений различного происхождения, что определяет актуальность настоящего исследования.

1.4 Анализ существующих методов устранения влияния состязательных патч-атак

Несмотря на значительный прогресс в разработке и применении моделей нейросетевого обнаружения объектов, они сталкиваются с серьезной угрозой в виде состязательных атак [23], которые представляют собой специально оптимизированные преднамеренные искажения, вводящие в заблуждение целевые ИНС. Состязательные атаки можно разделить на атаки с известным доступом к модели (white-box attacks), которые являются более эффективными, чем аналоги с ограниченным доступом (black-box attacks) [69]. Также состязательные атаки реализуются в цифровом и физическом формате. Цифровые состязательные атаки напрямую манипулируют значениями пикселей входных изображений, что предполагает полный доступ к изображениям, и реализуются исключительно в цифровом виде. Физические состязательные атаки добавляют в окружающую среду реальные объекты, которые могут повлиять на выходные изображения нейронных сетей. Как правило, это небольшие фрагменты изображений (патчи), которые оптимизируются в цифровой среде перед печатью [70].

Подобные атаки, направленные на обман нейронных сетей, могут привести к критическим ошибкам в различных сферах. Например, в автономном вождении ошибки распознавания объектов могут иметь катастрофические последствия. Атаки с использованием патчей показали свою эффективность в обмане систем классификации дорожных знаков [71], а также в сокрытии пешеходов [8] или даже автомобилей [6].

Наиболее распространённым подходом к повышению устойчивости нейросетевых моделей является обучение с использованием состязательных примеров (adversarial training) [72, 73]. Этот метод повышает надежность модели за счет добавления в обучающий набор изображений состязательных изображений. В работе [74] предложен другой подход: авторы добавили новый класс (состязательных патчи) и повторно обучили обнаружитель объектов на наборе данных, дополненном состязательными изображениями. Таким образом, обученная нейросетевая модель обнаружения объектов обнаруживает как целевые объекты, так и состязательные патчи. Несмотря на эффективность обучения с состязательными примерами для повышения устойчивости модели, этот метод требует значительных временных затрат. Его использование для крупномасштабных наборов данных и сложных глубоких архитектур становится дорогостоящим, а появление новых типов атак приводит к необходимости повторного обучения нейронных сетей. С целью снижения вычислительных затрат в последующих исследованиях основной акцент был сделан на разработке внешних модулей защиты, функционирующих без повторного обучения обнаружителей объектов и направленных на удаление состязательных патчей до обработки входных изображений обнаружителем объектов.

Существует множество работ, посвященных обучению внешних сегментаторов-защитников. Метод PatchZero [75] использует модель сегментации для обнаружения и удаления враждебных патчей. Метод Segment and Complete (SAC) [76] включает архитектуру U-Net [77], проверенную на практике, для сегментации и удаления патчей на входных изображениях. В работе [78] предложен метод Adversarial Pixel Masking (APM), где модель U-Net встроена в общую архитектуру обнаружителя объектов. Однако все эти методы включают в себя внешние сегментаторы, обученные на враждебных данных. Несмотря на это, обнаружители объектов остаются уязвимыми для новых атак, а внешние модели защиты требуют обучения на большом количестве новых враждебных примеров, что снова увеличивает затраты на обучение. DIFfusion-based DeFender [79] основан на диффузионных моделях, которые используют способность диффузионных

моделей восстанавливать исходную структуру изображения посредством итерационного удаления шумовой составляющей для противодействия атакам с использованием патчей. Авторы подчеркивают потенциал генеративных моделей в повышении устойчивости к состязательным атакам. Universal Defense Filter (UDFilter) [80] состоит из двух модулей: модуля генерации враждебных патчей и модуля генерации фильтра защиты. Этот подход ослабляет влияние патчей, применяя фильтр к входному изображению. Авторы также предложили стратегию Joint Detection Strategy, которая поддерживает эффективность модели. Patch-Agnostic Defense (PAD) [81] предназначен для локализации и удаления патчей без предварительного знания об атаке. В PAD используется модель Segment Anything Model [82] для точной сегментации патчей. Метод Local Gradient Smoothing (LGS) [83] также не зависит от атаки с патчами; он рассчитывает градиент входного изображения для локализации враждебного патча, поскольку градиент позволяет определить области изображения с высокими частотами.

В последние годы появились сертифицированные методы защиты от патчей. Например, метод Detectorguard [84] обнаруживает атаки без их удаления, что может привести к снижению функциональности модели во время атаки. Метод Objectseeker [85] удаляет части входного изображения и предлагает хотя бы одно разбиение, при котором патч не влияет на обнаружение объектов. Однако в сценариях атак с множественными патчами этот метод может оказаться неэффективным.

Таким образом, анализ существующих методов показывает, что современные подходы к защите от состязательных патч-атак преимущественно основаны на обучении на состязательных примерах, использовании внешних сегментаторов, восстановлении изображений либо применении сертифицированных методов защиты. Несмотря на достигнутые результаты, большинство существующих решений ориентировано на определённые типы атак, характеризуется высокой вычислительной сложностью и требует дополнительного обучения при появлении новых алгоритмов формирования состязательных воздействий.

Следовательно, разработка универсальных методов повышения устойчивости нейросетевых моделей обнаружения объектов, способных эффективно противодействовать широкому спектру искусственных воздействий без существенного увеличения вычислительных затрат и без необходимости повторного обучения модели, остаётся актуальной научной задачей.

Следует отметить, что в отличие от естественных искажений, возникающих в процессе формирования изображения, состязательные патч-атаки являются преднамеренно сформированными локальными воздействиями. Несмотря на различную природу данных факторов, обе группы воздействий приводят к снижению качества обнаружения объектов, что обуславливает необходимость разработки методов повышения робастности современных нейросетевых моделей обнаружения объектов.

1.5 Границы исследования

В данной диссертационной работе исследование ограничено разработкой методов повышения точности и робастности нейросетевого обнаружения объектов за счет процедуры подготовки обучающих данных и внешней обработки изображений. В данной работе не исследуются:

- способы модификации архитектуры нейронных сетей для повышения качества их работы;
- обработка мультиспектральных и гиперспектральных изображений;
- функции потерь, используемые для обучения нейронных сетей;
- способы оптимизации весовых параметров нейронных сетей;
- способы интерпретации выходных данных;
- вопросы достаточности объёмов обучающих выборок.

Основной предмет исследования заключается в исследовании устойчивости уже выбранной модели обнаружения объектов к искаженным данным при сохранении ее базовой архитектуры. В связи с этим качество модели

рассматривается одновременно в двух аспектах: как точность обнаружения на неискаженных изображениях и как способность сохранять это качество при увеличении интенсивности внешнего воздействия.

Основное внимание в работе уделено исследованию общего характера изменения поведения нейросетевых моделей обнаружения объектов при различных искажениях входных данных. При этом повышение устойчивости достигается не за счет изменения внутреннего устройства модели обнаружения объектов, а путем целенаправленного воздействия на входные изображения и внешние этапы обработки изображений. К таким воздействиям относятся формирование обучающего набора изображений с искаженными изображениями, адаптивная предварительная обработка входных данных, а также использование внешнего модуля обнаружения и удаления состязательных патч-атак. Такой подход позволяет использовать разработанные методы независимо от вида нейросетевой архитектуры и используемых наборов данных.

В настоящей работе естественные искажения рассматриваются на примере аддитивного белого гауссова шума и гауссова размытия, для которых существуют многочисленные исследования влияния на качество работы нейросетевых моделей. При исследовании состязательных воздействий рассматривается отдельный класс преднамеренных локальных искажений – состязательных патч-атак. В отличие от естественных искажений, целью которых является моделирование деградации качества изображения, состязательный патч формируется целенаправленно с учетом особенностей работы атакуемой нейросетевой модели обнаружения объектов.

Таким образом, границы настоящего исследования можно определить как исследование поведения фиксированных нейросетевых архитектур обнаружения объектов на цифровых изображениях при контролируемом искажении входных данных. Повышение устойчивости достигается за счет формирования обучающего набора изображений с оптимальным распределением уровней естественных искажений, адаптивной предварительной обработки входных изображений и применения внешнего метода локализации и удаления состязательных патч-атак.

Такой подход позволяет отделить влияние свойств самих данных и процедур их обработки от влияния архитектурных изменений нейросетевой модели и тем самым оценить возможности повышения робастности без усложнения модели обнаружения объектов.

1.6 Постановка научной задачи исследования

Проведенный в предыдущих подразделах анализ показал, что современные нейросетевые модели обнаружения объектов обеспечивают высокие показатели качества на данных со статистическими распределениями, соответствующими распределениям обучающего набора изображений. Вместе с тем, при изменении условий формирования тестовых изображений эффективность нейросетевых моделей может существенно снижаться. Для задач видового мониторинга данная проблема обусловлена, в частности, воздействием естественных искажений, возникающих при фотосъемке, передаче и обработке изображений, а также возможностью преднамеренного формирования состязательных воздействий, направленных на нарушение функционирования модели обнаружения объектов.

В связи с этим научная задача настоящего исследования состоит в разработке метода повышения точности и робастности (устойчивости) функционирования нейросетевых моделей обнаружения объектов в условиях воздействия естественных искажений и преднамеренных состязательных атак при видовом мониторинге.

В рамках поставленной научной задачи рассматривается нейросетевая модель обнаружения объектов F_θ , параметризованная набором параметров θ , на вход которой поступает изображение x . Результатом работы модели является множество обнаруженных объектов:

$$F_\theta(x) = \{(\mathbf{b}_i, c_i, s_i)\}, \quad (6)$$

где $\mathbf{b}_i$ – вектор предсказанных координат ограничивающей рамки i -го объекта;
 c_i – идентификатор класса предсказанного объекта;
 s_i – оценка уверенности обнаружения.

Пусть x_0 представляет собой исходное изображение, не содержащее искусственно внесенных искажений, а $D_\varphi^l(x_0)$ – изображение, полученное в результате воздействия искажения типа φ и интенсивности l . К естественным искажениям в настоящей работе относятся, прежде всего, аддитивный белый гауссов шум и гауссово размытие. Тогда задача повышения робастности заключается в нахождении такого F_θ , при котором выполняется следующее условие:

$$\sum_l Q\left(F_\theta\left(D_\varphi^l(x_0)\right)\right) \rightarrow \max, \quad (7)$$

где Q – метрика оценки качества обнаружения объектов. Конкретная метрика качества (точность, полнота, усредненная точность и т.п.) может выбираться в зависимости от задачи, это не меняет общего подхода, описываемого выражением (7).

При этом повышение робастности не должно сопровождаться существенным снижением качества обнаружения объектов на неискаженных изображениях x_0 . Поэтому задача рассматривается как поиск такого способа формирования обучающих данных и обработки входных изображений, при котором одновременно обеспечивается высокая точность обнаружения объектов на исходных изображениях и высокая устойчивость к искажениям различной интенсивности.

Весовые параметры модели F_θ зависят в общем случае от выбора метода формирования обучающего набора изображений, так и от метода предварительной обработки тестовых изображений. Таким образом, сформулированная постановка

задачи подразумевает работу как с предварительной обработкой изображений, так и с обучением модели. Это обстоятельство определяет общую структуру последующего исследования:

- разработка способа формирования обучающего набора данных для повышения устойчивости модели обнаружения объектов при видовом мониторинге;

- разработка метода адаптивной предварительной обработки изображений при видовом мониторинге;

- разработка метода локализации и удаления состязательных патч-атак на изображениях при видовом мониторинге.

Полученные решения должны обеспечить повышение устойчивости и точности нейросетевого обнаружения объектов без существенного усложнения архитектуры модели обнаружения объектов и с сохранением возможности практического применения разработанных методов в системах видового мониторинга.

Выводы по разделу 1

В результате проведенного анализа существующих методов обнаружения объектов, способов оценки качества и робастности, а также методов повышения устойчивости нейросетевых моделей к естественным и искусственным искажениям получены следующие выводы.

1. По результатам анализа существующих методов обнаружения и распознавания объектов на изображениях при видовом мониторинге установлено, что современные системы обнаружения объектов преимущественно основаны на глубоком обучении и представлены одноэтапными и двухэтапными архитектурами. Двухэтапные обнаружители объектов обеспечивают высокую точность локализации и классификации объектов, однако характеризуются высокой вычислительной сложностью. Современные одноэтапные архитектуры семейства YOLO обеспечивают высокую скорость обработки изображений при сохранении точности, сопоставимой с двухэтапными решениями, что делает их наиболее перспективными для задач видового мониторинга, связанных с обработкой аэрофотоснимков в условиях ограниченного времени обработки.

2. Анализ существующих метрик оценки качества и робастности моделей обнаружения объектов показал, что для оценки качества работы обнаружителей объектов наиболее широко используются метрики Precision, Recall, Average Precision (AP), Mean Average Precision (mAP) и Average Recall (AR), позволяющие комплексно оценивать точность классификации, полноту обнаружения и качество локализации объектов. В большинстве современных исследований робастность оценивается по изменению традиционных метрик качества (прежде всего mAP) при воздействии различных естественных искажений.

3. Анализ существующих способов обучения нейросетевых систем для устранения влияния естественных искажений показал, что большинство существующих исследований посвящено задачам классификации изображений и основано на использовании предварительного восстановления изображений,

модификации архитектуры нейронной сети либо обучении на смеси неискаженных и искаженных изображений.

4. Анализ существующих методов устранения влияния состязательных патч-атак показал, что современные методы защиты основаны преимущественно на обучении с использованием состязательных примеров, применении внешних моделей сегментации патчей, использовании генеративных моделей восстановления изображений и сертифицированных методов защиты. Несмотря на достигнутые результаты, большинство существующих решений ориентировано на ограниченный класс атак, требует значительных вычислительных затрат и дополнительного обучения при появлении новых методов формирования состязательных воздействий, что ограничивает их практическое применение.

5. На основании результатов проведенного анализа сформулирована научная задача исследования, состоящая в разработке метода повышения точности и робастности (устойчивости) функционирования нейросетевых моделей обнаружения объектов в условиях воздействия естественных искажений и преднамеренных состязательных атак при видовом мониторинге. В то же время установлены границы исследования, заключающиеся в исследовании поведения фиксированных нейросетевых архитектур обнаружения объектов на цифровых изображениях при контролируемом искажении входных данных. Повышение устойчивости достигается за счет формирования обучающего набора изображений с оптимальным распределением уровней естественных искажений, адаптивной предварительной обработки входных изображений и применения внешнего метода локализации и удаления состязательных патч-атак. Такой подход позволяет отделить влияние свойств самих данных и процедур их обработки от влияния архитектурных изменений нейросетевой модели и тем самым оценить возможности повышения робастности без усложнения модели обнаружения объектов.

2. РАЗРАБОТКА СПОСОБА ФОРМИРОВАНИЯ ОБУЧАЮЩЕГО НАБОРА ДАННЫХ ДЛЯ ПОВЫШЕНИЯ УСТОЙЧИВОСТИ МОДЕЛИ ОБНАРУЖЕНИЯ ОБЪЕКТОВ ПРИ ВИДОВОМ МОНИТОРИНГЕ

2.1 Исследование влияния введения аддитивного белого гауссова шума в обучающие изображения нейросетевой модели обнаружения объектов

В настоящей работе для исследования влияния введения АБГШ нейросетевой модели обнаружения объектов предложено несколько подходов к обучению такой модели. Первый подход основан на обучении модели обнаружения объектов на исходных (не зашумленных) данных. Два других подхода заключаются в обучении на зашумленных изображениях с разной степенью искажения [39]. Зашумленные изображения создаются из исходного обучающего набора по следующим правилам:

- 1) исходный обучающий набор изображений случайным образом перемешивается и делится на три части с равным количеством изображений;
- 2) одна часть остается без изменений;
- 3) к изображениям из второй части добавляется белый гауссов шум по правилу (8);
- 4) к изображениям из третьей части добавляется белый гауссов шум по правилу (9);

$$\mathbf{Y}_2 = \mathbf{X}_2 + \mathbf{W}_2, \quad (8)$$

$$\mathbf{Y}_3 = \mathbf{X}_3 + \mathbf{W}_3, \quad (9)$$

где $\mathbf{X}_2, \mathbf{X}_3$ – матрицы исходных изображений, используемые для второй и третьей подвыборок соответственно, представленные в виде трехмерных матриц пикселей, имеющих высоту h пикселей, ширину w пикселей и глубину $d = 3$

пикселя, что соответствует количеству каналов цветного (RGB) изображения (диапазон значений яркости в каналах RGB от 0 до 255); $\mathbf{W}_2, \mathbf{W}_3$ – матрицы случайных величин с нормальным распределением и нулевой средней величиной, имеющие такие же размеры, что и матрицы $\mathbf{X}_2$ и $\mathbf{X}_3$ соответственно, причем случайные величины матрицы $\mathbf{W}_2$ имеют стандартное отклонение, равное 30 для второго подхода обучения и 100 для третьего подхода обучения (в этих случаях отношения сигнал/шум составляют 4.55 и -5.9 дБ соответственно), а случайные величины матрицы $\mathbf{W}_3$ имеют стандартное отклонение, равное 40 для второго подхода обучения и 200 для третьего подхода обучения (в этих случаях отношения сигнал/шум составляют 2.05 и -11.93 дБ соответственно); $\mathbf{Y}_2, \mathbf{Y}_3$ – изображения второй и третьей подвыборок соответственно.

Шумовое искажение изображений является следствием множества различных процессов. Центральная предельная теорема (ЦПТ) утверждает, что сумма достаточно большого количества некоррелированных случайных процессов с примерно одинаковыми масштабами имеет распределение, близкое к нормальному [86]. Кроме того, множество различных случайных процессов, влияющих на итоговое искажение, имеет различную физическую природу и потому независимо, а, следовательно, не коррелировано. Таким образом, автокорреляционная функция итогового искажения имеет вид дельта-функции. Исходя из изложенного выше в качестве шумового воздействия в данной работе используется белый гауссов шум.

Описанные выше варианты обучения (обучение на исходном и частично зашумленном наборах данных соответственно) проводятся на одной и той же нейросетевой архитектуре модели обнаружения объектов и с одинаковыми гиперпараметрами обучения.

Для получения кривых помехоустойчивости на вход сети подаются изображения из тестового набора изображений с разным уровнем шума и оценивается величина полноты обнаружения для каждого случая. Величина полноты обнаружения вычисляется для всех предсказанных ограничивающих рамок, имеющих наивысшую вероятность нахождения в данной рамке объекта, по

формуле (1) при метрике пересечения предсказанных рамок с истинными (IoU) больше 0,5. Для вычисления степени зашумленности в данном исследовании используется величина отношения сигнал/шум (ОСШ), формула которого имеет вид:

$$\text{ОСШ} = 20 \times \log_{10} \left(\frac{\sigma_{\text{и}}}{\sigma_{\text{АБГШ}}^{\text{тест}}} \right), \quad (10)$$

где $\sigma_{\text{и}}$ – стандартное отклонение тестового набора данных, характеризующее средний контраст изображения, $\sigma_{\text{АБГШ}}^{\text{тест}}$ – стандартное отклонение аддитивного белого гауссова шума.

Для визуализации на рисунке 1 представлены образцы изображений с различными значениями ОСШ, используемые для данной оценки.

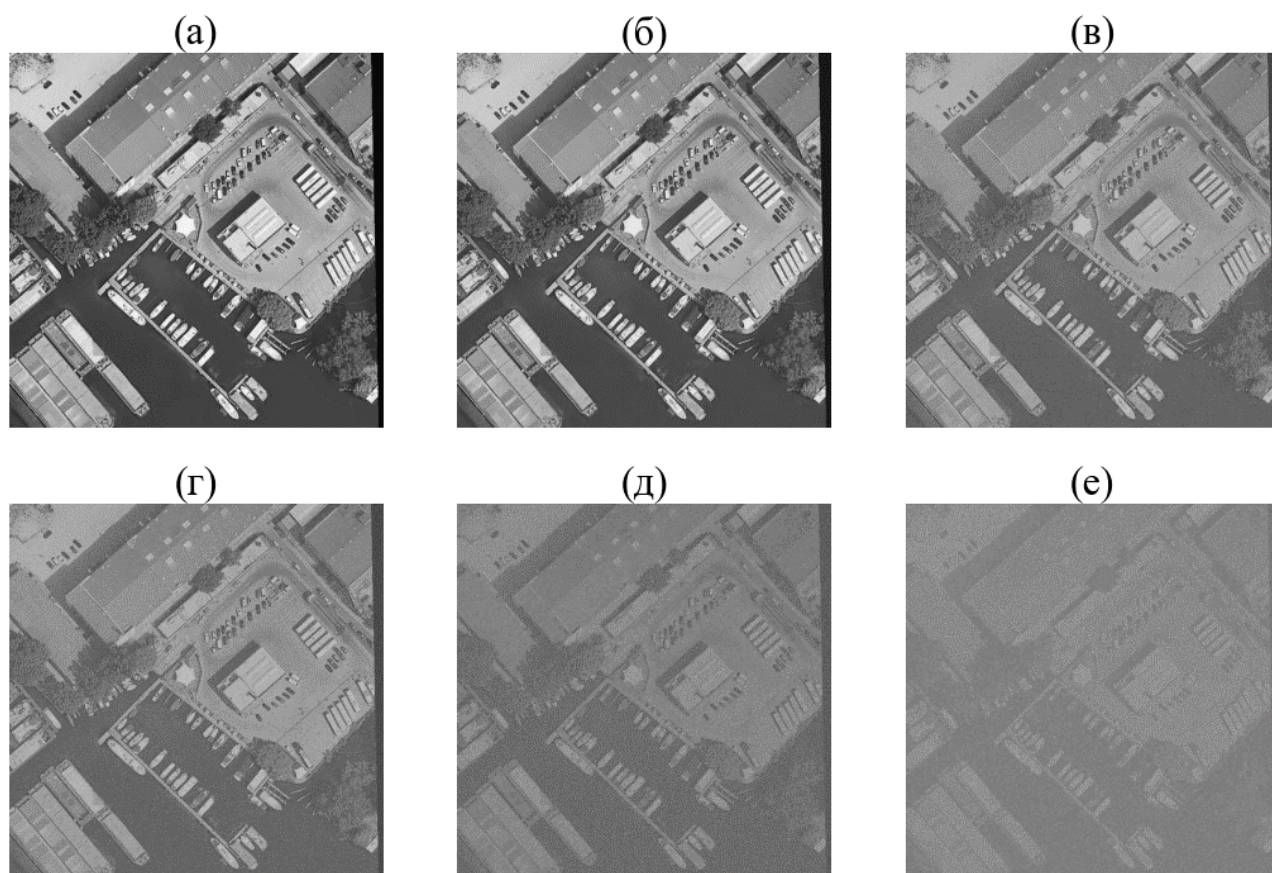


Рисунок 1 – Примеры изображений из тестового набора с различными значениями ОСШ: 94,1 (а); 20,1 (б); 4,55 (в); 2 (г); -5,9 (д); -11,9 дБ (е)

На рисунке 2 представлены графики зависимости величины полноты обнаружения объектов от ОСШ для трех различных моделей обученной на исходном обучающем наборе изображения (кривая 1); на умеренно зашумленном обучающем наборе изображение, полученной добавлением зашумленных изображений по правилу (8)-(9) с ОСШ = 4,55 и 2,05 дБ (кривая 2); и на избыточно зашумленном обучающем наборе изображение, полученной добавлением зашумленных изображений по правилу (8)-(9) с ОСШ = -5,9 дБ и ОСШ = -11,93 дБ (кривая 3).

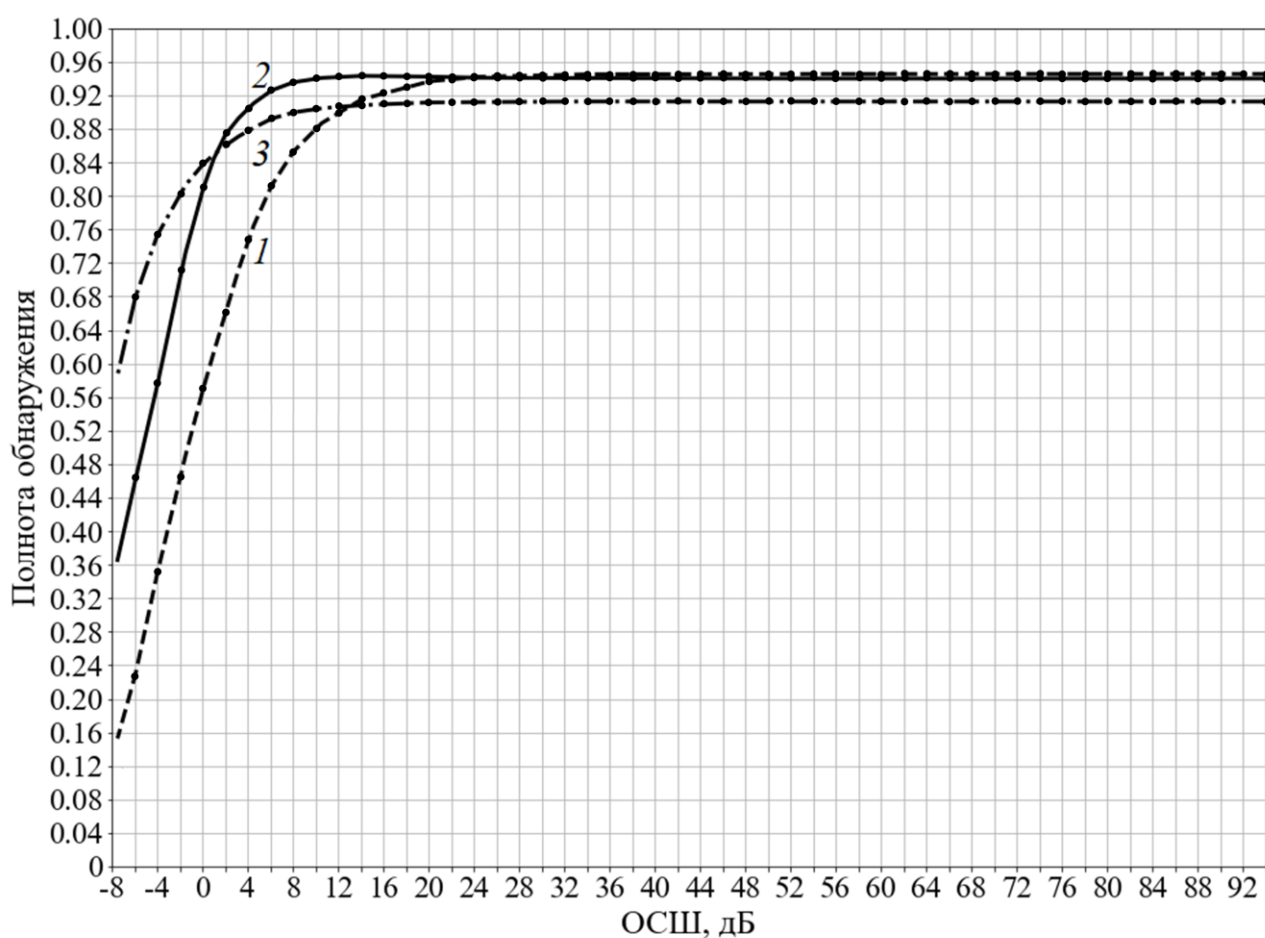


Рисунок 2 – Графики зависимости величины полноты обнаружения от ОСШ

Графики помехоустойчивости модели обнаружения, представленные на рисунке 2, позволяют выделить две области отношений сигнал/шум:

- «переходная область», в которой наблюдается заметное (на десятки процентов) повышение полноты обнаружения с ростом ОСШ (от -8 до 8 дБ);

- «область высоких ОСШ», в которой полнота обнаружения высока и мало меняется, объекты на изображении визуальны четко различимы ($\text{ОСШ} > 8 \text{ дБ}$).

Из графиков, представленных на рисунке 2, видно, что при увеличении уровня шума модель, обученная на умеренно зашумленном наборе данных (кривая 2), показывает способность обнаруживать больше релевантных объектов, чем та же модель, но обученная на исходном наборе данных (кривая 1). Стоит отметить, что в «переходной области» ОСШ кривая 2 показывает прирост метрики полноты обнаружения относительно кривой 1 более чем в 2 раза в диапазоне значений ОСШ от -8 до 0 дБ. В то же время при ОСШ, превышающем 20 дБ в проведенном в данной работе эксперименте, модель, обученная на исходном наборе данных, показывает величину полноты обнаружения на 0,4% выше, что является незначительно более высоким значением. Из этого можно сделать вывод, что такая модель лучше работает на неискаженных данных, однако хуже способна противостоять шумовому воздействию (переходная область). Добавление в обучающий набор изображения с чрезмерно низким ОСШ (кривая 3), т.е. с таким уровнем шума, при котором визуальны объекты на изображениях аэрофотоснимках сложно различить, приводит к существенному снижению качества распознавания изображений в области ОСШ выше 0 дБ (в данном эксперименте). При ОСШ ниже 0 дБ наблюдается более высокая помехоустойчивость данной модели по сравнению с моделью, обученной на умеренно зашумленных данных (кривая 2). В то же время добавление в обучающий набор изображения с чрезмерно низким ОСШ снижает интегральную (суммарную по всем значениям ОСШ) полноту обнаружения из-за внесения слишком большой неопределенности в весовые коэффициенты нейронной сети. То есть обучение модели обнаружения на наборе данных с чрезмерно низким ОСШ приводит к сдвигу влево и вниз кривой помехоустойчивости. Данное наблюдение свидетельствует о необходимости поиска оптимального уровня неопределенности, которая вводится в обучающий набор изображений, для получения наивысшей помехоустойчивости модели.

Таким образом, исследование влияния шумового воздействия на качество работы современной архитектуры глубокой сверточной нейронной сети,

решающей задачу распознавания объектов на аэрофотоснимках Земли, показало, что интегральная оценка качества обнаружения при различных ОСШ может служить индикатором истинной обобщающей способности модели и корректности её выбора [39].

Кроме того, предложен способ повышения помехоустойчивости нейросети путем ее обучения на частично искаженном наборе данных. Показано, что избыточное зашумление обучающего набора изображений приводит к снижению интегральной оценки полноты обнаружения объектов.

Также из работы следует, что обучение и оценка модели обнаружения объектов исключительно на неискаженных данных не гарантируют, что она покажет аналогичный результат при обработке зашумленных входных изображений. В то же время добавление в обучающий набор кадры с умеренным уровнем шума обеспечивает прирост качества обнаружения более чем в 2 раза в области низких ОСШ.

В результате проведенный анализ зависимости эффективности сверточных нейронных сетей от уровня искажений позволяет получить более полные сведения о качестве работы модели обнаружения объектов на аэрофотоснимках в реальных условиях.

2.2 Исследование влияния введения гауссова размытия в обучающие изображения нейросетевой модели обнаружения объектов

2.2.1 Применение гауссова размытия к изображениям

Эффект размытия представляет собой ухудшение качества изображения, вызванное усреднением яркостей ближайших пикселей. Размытое изображение может быть смоделировано с помощью применения операции свертки исходного неискаженного изображения и ядра размытия, которое представляет собой фильтр нижних частот, которое приводит к потере или ослаблению высокочастотных деталей изображения (т.е. уменьшению резкости изображения).

В данном исследовании работе в качестве искажений используется гауссово размытие, которое реализовано путем применения операции двумерной свертки матрицы изображения с ядром гауссовых значений размером $2k + 1$ при заданном радиусе ядра размытия k :

$$I_{\text{размыт.}}(x, y) = \sum_{m=-k}^k \sum_{n=-k}^k G_{\text{норм.}}(m, n) \cdot I(x - m, y - n), \quad (11)$$

где $I(x, y)$ – матрица неискаженного цветного изображения, значения которого находятся в диапазоне от 0 до 255 и представляют собой значения яркости пикселей в каждом RGB-канале;

$I_{\text{размыт.}}(x, y)$ – матрица размытого изображения, значения которого также находятся в диапазоне от 0 до 255;

$G_{\text{норм.}}(x, y)$ – двумерное гауссово ядро с нулевым средним и стандартным отклонением d , которое описывается следующей формулой:

$$G_{\text{норм.}}(x, y) = \frac{G(x, y)}{\sum_{m=-k}^k \sum_{n=-k}^k G(m, n)}, \quad x \in [-k; k], y \in [-k; k], \quad (12)$$

где

$$G(x, y) = \frac{1}{2\pi d^2} \cdot e^{-\frac{x^2 + y^2}{2d^2}}, \quad x \in [-k; k], y \in [-k; k]. \quad (13)$$

В данной работе используются гауссовы ядра с варьируемым параметром стандартного отклонения d пикселей и радиуса ядра $k = \max(1, [d])$ пикселей, где $[\cdot]$ – операция округления в большую сторону. Так, в данном исследовании степень искажений в дальнейшем будет описываться единственным параметром – стандартным отклонением гауссовых значений d . На рисунке 3 представлены проекции ядер, полученных по формулам (12)-(13) (при $y=0$ в выражении (12)), для трех различных значений стандартного отклонения d .

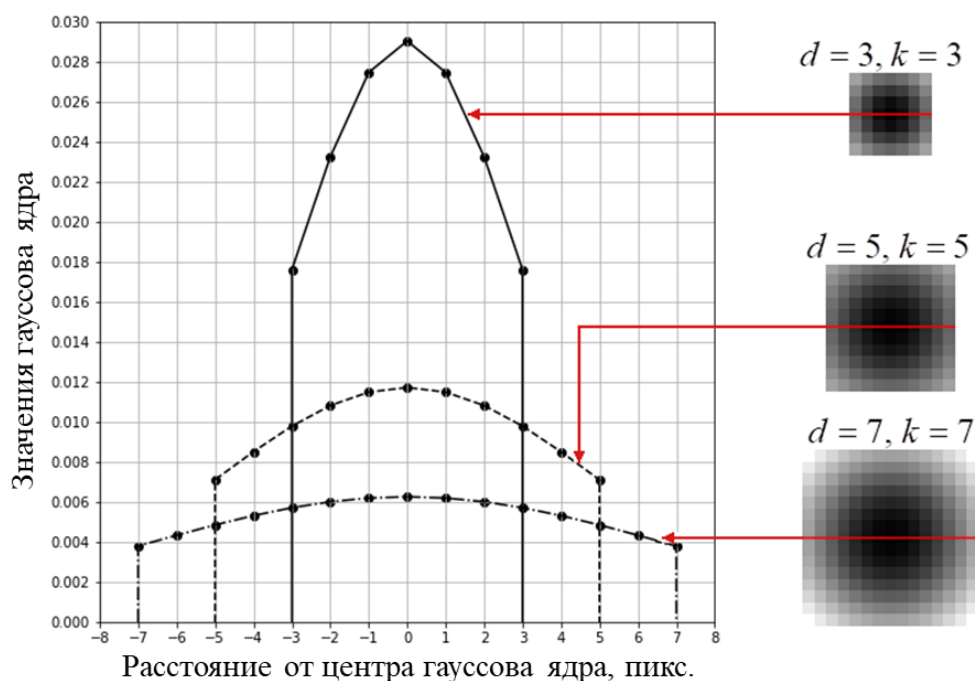


Рисунок 3 – Двумерные изображения ядер размытия (справа) и их проекции (слева) для трех различных значений степеней искажений

Из рисунка 3 видно, что с увеличением стандартного отклонения гауссовых величин увеличивается количество соседних пикселей, которые участвуют в размытии изображений.

2.2.2 Разработка способа формирования обучающего набора данных для повышения устойчивости анализируемой модели обнаружения объектов

Опыт предыдущих исследований [5, 66 - 68] показывает, что добавление неискаженных и искаженных изображений в обучающий набор изображений показывает более высокое качество предсказаний ГСНС. Однако нет исследований, определяющих долю неискаженных и искаженных изображений в обучающем наборе изображений, а также диапазон изменения уровня искажений, позволяющий добиться повышения качества распознавания, стабильности и робастности нейронной сети. В настоящем исследовании в качестве уровня искажения изображений d в обучающем наборе изображений предложено

использовать модуль случайной величины, подчиняющейся нормальному распределению с нулевым средним значением и стандартным отклонением $\sigma_{\text{обуч.}}^{\text{размыт.}}$ [43]:

$$d = |\hat{d}|, \quad (14)$$

где

$$\hat{d} \sim N\left(0, \left(\sigma_{\text{обуч.}}^{\text{размыт.}}\right)^2\right), \quad (15)$$

где $N\left(0, \left(\sigma_{\text{обуч.}}^{\text{размыт.}}\right)^2\right)$ – нормальное распределение с нулевым математическим ожиданием и стандартным отклонением $\sigma_{\text{обуч.}}^{\text{размыт.}}$.

На рисунке 4 изображены столбчатые диаграммы для различных обучающих выборок, обозначенных как $\sigma_{\text{обуч.}}^{\text{размыт.}}$. Каждая столбчатая диаграмма описывает долю обучающего набора изображений, которая размывается в соответствии с параметром искажения.

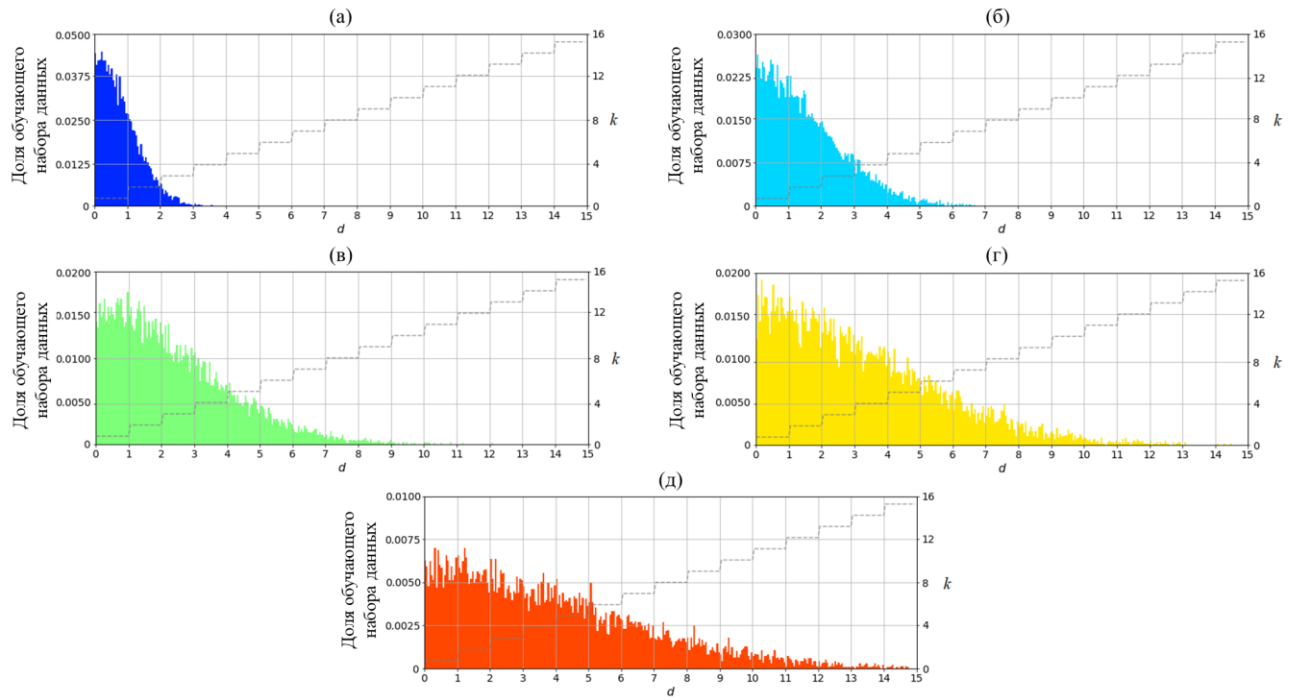


Рисунок 4 – Гистограммы величины искажений обучающих выборок d при $\sigma_{\text{обуч.размыт.}} = 1$ (а), $\sigma_{\text{обуч.размыт.}} = 2$ (б), $\sigma_{\text{обуч.размыт.}} = 3$ (в), $\sigma_{\text{обуч.размыт.}} = 4$ (г), $\sigma_{\text{обуч.размыт.}} = 5$ (д); зависимость радиуса ядра размытия k от величины искажения d (серая пунктирная кривая)

Из рисунка 4 видно, что при $\sigma_{\text{обуч.размыт.}} = 1$ большая доля обучающего набора данных имеет степень искажений d в диапазоне от 0 до 1 пикселей, при $\sigma_{\text{обуч.размыт.}} = 5$ обучающий набор данных содержит изображения с уровнями искажений d , изменяющимися в широком диапазоне (примерно от 0 до 10 пикселей). В ходе настоящего исследования была поставлена цель определить влияние распределения уровня искажений в обучающем наборе изображений на качество распознавания тестовых изображений.

При тестировании систем локализации транспортных средств, обученных на наборах данных с различным значениями $\sigma_{\text{обуч.размыт.}}$, использовался один и тот же набор данных. Однако, для исследования робастности обученных систем в данной работе изображения в тестовом наборе данных размываются, как описано в (11)-(13), при фиксированных значениях размера ядра размытия, связанных с $\sigma_{\text{тест.размыт.}}$.

2.2.3 Оценка устойчивости модели обнаружения объектов при обучении с использованием разработанного способа формирования обучающего набора данных

Для выбора способа размытия обучающего набора изображений было исследовано семь моделей: модель, обученная на неискажённом обучающем наборе изображений (т.е. без добавления размытия $\sigma_{\text{обуч.}}^{\text{размыт.}} = 0$); модели, обученные на наборах данных, у которых все изображения в обучающей наборе подверглись искажениям с постоянным параметром размытия $d_{\text{обуч.}}^{\text{размыт.}} = 1$, $d_{\text{обуч.}}^{\text{размыт.}} = 2$ и $d_{\text{обуч.}}^{\text{размыт.}} = 3$, соответственно; модели, обученные на наборах данных, изображения которых искажены в соответствии с правилом, описанном в подразделе 2.2.2 с параметром $\sigma_{\text{обуч.}}^{\text{размыт.}} = 1$, $\sigma_{\text{обуч.}}^{\text{размыт.}} = 2$ и $\sigma_{\text{обуч.}}^{\text{размыт.}} = 3$ соответственно. На рисунке 5 представлены графики зависимости величины усредненной полноты (AR) (5) перечисленных выше моделей от различной степени искажений тестового набора данных $\sigma_{\text{тест}}^{\text{размыт.}}$.

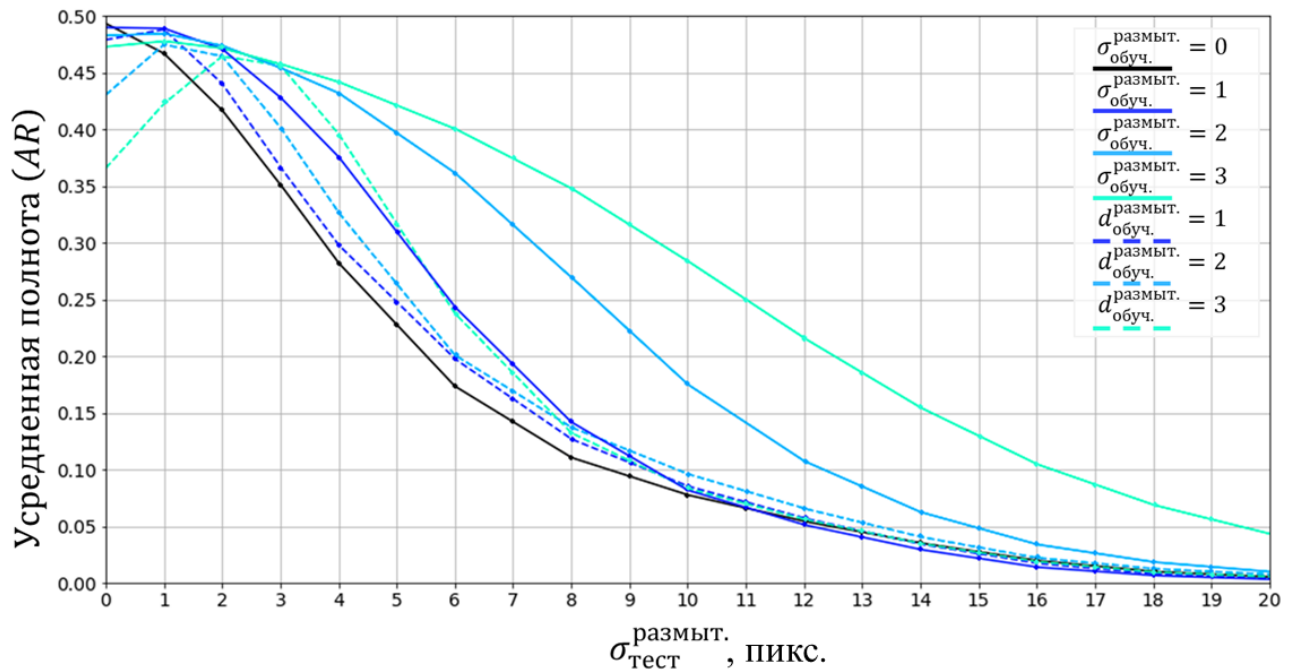


Рисунок 5 – Графики зависимости метрики усредненной полноты (AR) от $\sigma_{\text{размыт.тест}}$ для моделей с параметром размытия обучающего набора изображений $\sigma_{\text{размыт.обуч.}} = 0$ (черная сплошная кривая), $d_{\text{размыт.обуч.}} = 1$, $d_{\text{размыт.обуч.}} = 2$ и $d_{\text{размыт.обуч.}} = 3$ (синяя, голубая и зеленая пунктирные кривые, соответственно), $\sigma_{\text{размыт.обуч.}} = 1$, $\sigma_{\text{размыт.обуч.}} = 2$ и $\sigma_{\text{размыт.обуч.}} = 3$ (синяя, голубая и зеленая сплошные кривые соответственно)

Из рисунка 5 видно, при обучении моделей на изображениях с фиксированным уровнем искажений (случаи $d_{\text{размыт.обуч.}} = 1$, $d_{\text{размыт.обуч.}} = 2$ и $d_{\text{размыт.обуч.}} = 3$) максимум метрики качества перемещается вдоль оси абсцисс, при этом сильно снижается качество локализации объектов на неискаженных данных. Также из рисунка 5 видно, что способ обучения модели, обученной на наборе изображений, описанной в подразделе 2.2.2 значительно увеличивает робастность такой модели. Однако видно, что на неискаженных данных качество работы моделей $\sigma_{\text{размыт.обуч.}} = 1$, $\sigma_{\text{размыт.обуч.}} = 2$ и $\sigma_{\text{размыт.обуч.}} = 3$ незначительно уменьшилось относительно модели, обученной на данных без искажений (при $\sigma_{\text{размыт.обуч.}} = 0$), однако при больших значениях $\sigma_{\text{размыт.тест}}$ такие модели показывают более высокую метрику усредненной полноты. Таким образом, можно сделать вывод, что существует проблема запоминания уровня размытия моделями обнаружения

объектов на изображении. Также существует степень и способ размытия обучающих данных, благодаря которым можно достичь высокой обобщающей способности модели, обеспечивающей высокое качество распознавания данных как без размытия, так и с различным его уровнем.

Для более полного поиска и анализа способности анализируемой модели локализации транспортных средств на аэрофотоснимках Земли были подготовлены следующие модели: модель, обученная на исходном (без искажений) обучающем наборе изображений ($\sigma_{\text{обуч.}}^{\text{размыт.}} = 0$); модели, обученные на наборах данных, изображения которых искажены в соответствии с правилом, описанном в подразделе 2.2.2 с параметром $\sigma_{\text{обуч.}}^{\text{размыт.}} = 1$, $\sigma_{\text{обуч.}}^{\text{размыт.}} = 2$, $\sigma_{\text{обуч.}}^{\text{размыт.}} = 3$, $\sigma_{\text{обуч.}}^{\text{размыт.}} = 4$ и $\sigma_{\text{обуч.}}^{\text{размыт.}} = 5$ соответственно. На рисунке 6 представлены кривые зависимости усредненной полноты AR (5) перечисленных выше моделей от различной степени искажений тестового набора данных $\sigma_{\text{тест}}^{\text{размыт.}}$.

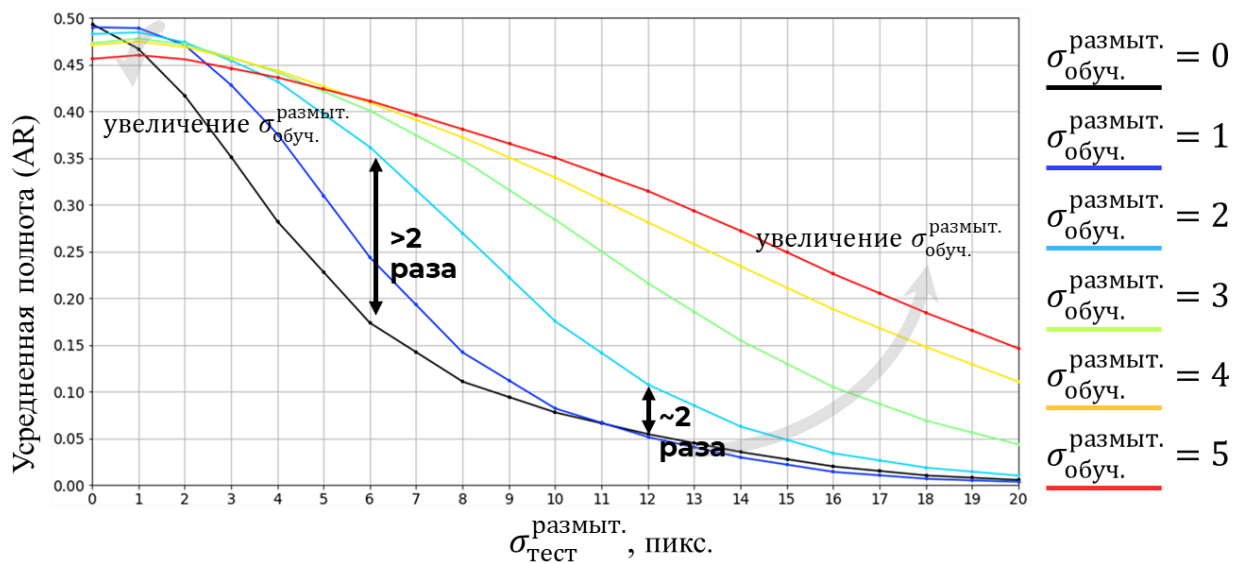


Рисунок 6 – Графики зависимости усредненной полноты AR от $\sigma_{\text{тест}}^{\text{размыт.}}$ для моделей с параметром $\sigma_{\text{обуч.}}^{\text{размыт.}} = 0$ (черная кривая), $\sigma_{\text{обуч.}}^{\text{размыт.}} = 1$, $\sigma_{\text{обуч.}}^{\text{размыт.}} = 2$, $\sigma_{\text{обуч.}}^{\text{размыт.}} = 3$, $\sigma_{\text{обуч.}}^{\text{размыт.}} = 4$ и $\sigma_{\text{обуч.}}^{\text{размыт.}} = 5$ (синяя, голубая, зеленая, желтая и красная кривые соответственно)

Из рисунка 6 видно, что при обучении и тестировании ГСНС на неискаженных данных ($\sigma_{\text{обуч.}}^{\text{размыт.}} = 0$) показывает наивысшую метрику усредненной полноты, но с увеличением степени размытия тестовых данных $\sigma_{\text{тест}}^{\text{размыт.}}$ качество обнаружения быстро падает. В случае увеличения разброса степени искажений в обучающем наборе изображений (т.е. с увеличением $(\sigma_{\text{обуч.}}^{\text{размыт.}})$ видно, что уровень усредненной полноты анализируемой модели уменьшается медленнее при больших значениях $\sigma_{\text{тест}}^{\text{размыт.}}$, однако, на исходных (без искажений) данных оценка качества обнаружения заметно падает. Таким образом, можно сделать вывод, что существует оптимальное значение $\sigma_{\text{обуч.}}^{\text{размыт.}}$, при котором обученная с данным будет иметь довольно высокий охват истинных объектов на неискаженных или слабо размытых входных данных и медленно ухудшаться с ростом степени искажений входных изображений.

Для оценки полноты обнаружения моделью локализации объектов используется метрика перекрытия истинного объекта с лучшей предсказанной ограничивающей рамкой – IoU. На рисунке 7 представлены контурные графики для моделей с параметром $\sigma_{\text{обуч.}}^{\text{размыт.}} = 0$ (черная кривая), $\sigma_{\text{обуч.}}^{\text{размыт.}} = 1$, $\sigma_{\text{обуч.}}^{\text{размыт.}} = 2$, $\sigma_{\text{обуч.}}^{\text{размыт.}} = 3$, $\sigma_{\text{обуч.}}^{\text{размыт.}} = 4$ и $\sigma_{\text{обуч.}}^{\text{размыт.}} = 5$, на которых интенсивностью красного цвета показана доля правильно обнаруженных и локализованных транспортных средств на аэрофотоснимках Земли от метрики перекрытия IoU. Данные карты показывают, что чем точнее предсказаны ограничивающие рамки, тем (т.е. чем выше их метрика перекрытия с истинными объектами IoU), тем выше на контурных графиках будет отображаться интенсивнее красный цвет.

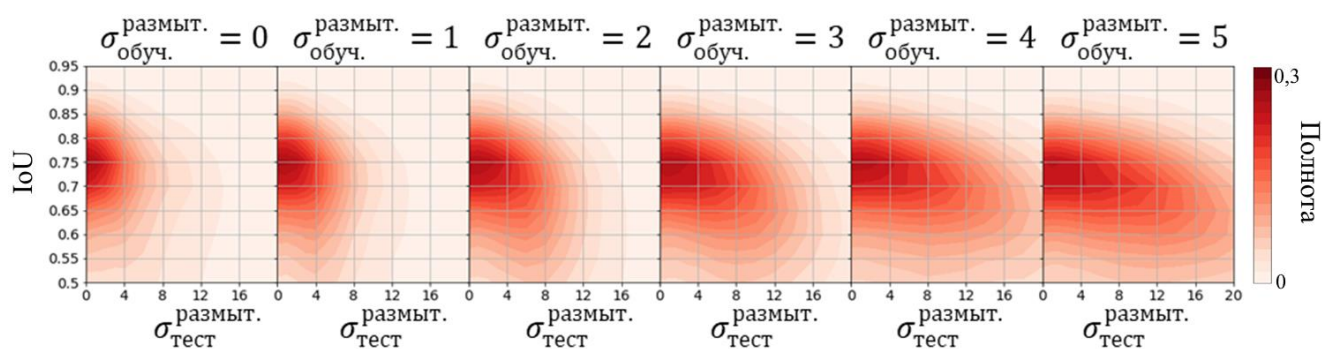


Рисунок 7 – Контурные графики зависимости доли правильно обнаруженных и локализованных транспортных средств на аэрофотоснимках от метрики их перекрытия с истинными объектами для моделей с параметром $\sigma_{\text{обуч.}}^{\text{размыт.}} = 0$, $\sigma_{\text{обуч.}}^{\text{размыт.}} = 1$, $\sigma_{\text{обуч.}}^{\text{размыт.}} = 2$, $\sigma_{\text{обуч.}}^{\text{размыт.}} = 3$, $\sigma_{\text{обуч.}}^{\text{размыт.}} = 4$ и $\sigma_{\text{обуч.}}^{\text{размыт.}} = 5$ при различных уровнях размытия тестовых изображений

На рисунке 7 видно, что контур «растекается» на большие значения $\sigma_{\text{тест}}^{\text{размыт.}}$ с увеличением разброса искажений в обучающем наборе изображений $\sigma_{\text{обуч.}}^{\text{размыт.}}$, что подтверждает вышеописанное наблюдение. При этом, хотя доля правильно обнаруженных объектов растет с увеличением $\sigma_{\text{обуч.}}^{\text{размыт.}}$ при высоких уровнях размытия в тестовом наборе изображений, расширение вниз ярких пятен на «тепловых картах» рисунка 7 с ростом $\sigma_{\text{обуч.}}^{\text{размыт.}}$ и $\sigma_{\text{тест}}^{\text{размыт.}}$ свидетельствует о соответствующем снижении меры перекрытия обнаруженных объектов с истинными. Причина данной закономерности может заключаться в размытии границ обнаруживаемых объектов.

Таким образом, можно заключить, что избыточное размытие обучающих изображений, несмотря на рост усредненной полноты AR, приводит к нежелательным результатам в виде снижения меры перекрытия границ предложенных нейронной сетью регионов интереса с истинными объектами на изображении.

В ходе решения проблемы влияния размытия аэрофотоизображений Земли на качество обнаружения объектов при мониторинге был предложен способ формирования обучающего набора изображений. В результате проведенного моделирования удалось экспериментально показать, что эффективной аугментацией обучающих изображений может быть их размытие части

с интенсивностью, описываемой положительной ветвью распределения Гаусса. Выбор дисперсии этого распределения (и, соответственно, доли искажённых изображений и уровня их искажений в обучающем наборе) обусловлен двумя конкурирующими процессами – ростом усредненной полноты AR и снижением меры перекрытия границ предложенных нейронной сетью регионов интереса с истинными объектами на изображении по мере увеличения дисперсии уровня размытия.

В частности, достаточно эффективная модель, обученная на наборе изображений с параметром разброса уровней размытия $\sigma_{\text{обуч.}}^{\text{размыт.}} = 2$, демонстрирует прирост полноты обнаружения более, чем в 2 раза при уровнях искажения тестовых изображений в диапазоне $6 \leq \sigma_{\text{тест}}^{\text{размыт.}} \leq 12$, по сравнению с обученной на неискажённых изображениях.

Таким образом, выбор оптимального значения параметром разброса уровней размытия обучающих изображений должен в каждой задаче определяться особенностями обнаруживаемых объектов и взаимодействием между блоками локализации и классификации в обобщённой структуре нейросетевого обнаружителя.

Выводы по разделу 2

В результате проведенных исследований, направленных на разработку способа формирования обучающего набора изображений для повышения устойчивости модели обнаружения объектов при видовом мониторинге, исследовано влияние различных видов естественных искажений на качество работы нейросетевых моделей и разработаны способы формирования обучающих выборок, обеспечивающие повышение устойчивости моделей обнаружения объектов. Получены следующие результаты.

1. Исследование влияния шумового воздействия на глубокую сверточную нейронную сеть для распознавания объектов на аэрофотоснимках Земли показало, что оценка качества при различных ОСШ служит индикатором истинной обобщающей способности модели, а включение в обучающий набор аэрофотоизображений с умеренным уровнем шума обеспечивает прирост качества распознавания более чем в 2 раза в области низких ОСШ (от -8 до 0 дБ). Таким образом, комплексный анализ зависимости эффективности нейросети от уровня искажений позволяет достоверно оценить корректность выбранной архитектуры и качество её работы в реальных условиях эксплуатации.

2. По результатам исследования влияния введения гауссова размытия в обучающие изображения нейросетевой модели обнаружения объектов установлено, что аугментация обучающего набора изображениями с различными уровнями гауссова размытия позволяет существенно повысить устойчивость модели к размытию тестовых изображений. Показано, что модель, обученная с использованием распределения уровней размытия, описываемого положительной ветвью распределения Гаусса с параметром $\sigma_{\text{обуч.}}^{\text{размыт.}} = 2$ пикселя, обеспечивает увеличение полноты обнаружения более чем в 2 раза при уровнях размытия тестовых изображений в диапазоне $6 \leq \sigma_{\text{тест}}^{\text{размыт.}} \leq 12$ пикселей по сравнению с моделью, обученной на неискажённых изображениях. При этом избыточное размытие обучающих изображений приводит к снижению качества локализации объектов, что определяет необходимость выбора оптимального параметра разброса

уровней размытия как компромисса между повышением полноты обнаружения и сохранением точности локализации.

3. В результате проведенных исследований разработан и экспериментально обоснован способ обучения модели обнаружения объектов, основанный на формировании обучающего набора изображений с использованием искаженных изображений, интенсивность которых задается полунормальным распределением. Полученные результаты, включая увеличение полноты обнаружения более чем в 2 раза при воздействии гауссова размытия при радиусе размытия в диапазоне от 6 до 12 пикселей, а также повышение качества распознавания более чем в 2 раза в области низких ОСШ (от -8 до 0 дБ), подтверждают эффективность разработанного метода. Таким образом, первое положение, выносимое на защиту, экспериментально подтверждено.

3. РАЗРАБОТКА МЕТОДА АДАПТИВНОЙ ПРЕДВАРИТЕЛЬНОЙ ОБРАБОТКИ ИЗОБРАЖЕНИЙ ПРИ ВИДОВОМ МОНИТОРИНГЕ

3.1 Исследование влияния моделей естественных искажений на нейросетевую модель обнаружения объектов, обученную разработанным способом

В данном исследовании в качестве модели нейросетевого обнаружения объектов используется модель Faster R-CNN, обученная пятью разными способами, упомянутыми в подразделе 2.2 настоящей диссертационной работы.

На рисунке 8 показаны кривые зависимости метрики качества обнаружения mean Average Precision (mAP) от степени искажения гауссова размытия $\sigma_{\text{тест}}^{\text{размыт.}}$ в тестовом наборе изображений для моделей с параметром $\sigma_{\text{обуч.}}^{\text{размыт.}} = 0$, $\sigma_{\text{обуч.}}^{\text{размыт.}} = 1$, $\sigma_{\text{обуч.}}^{\text{размыт.}} = 2$, $\sigma_{\text{обуч.}}^{\text{размыт.}} = 3$ и $\sigma_{\text{обуч.}}^{\text{размыт.}} = 4$ и $\sigma_{\text{обуч.}}^{\text{размыт.}} = 5$.

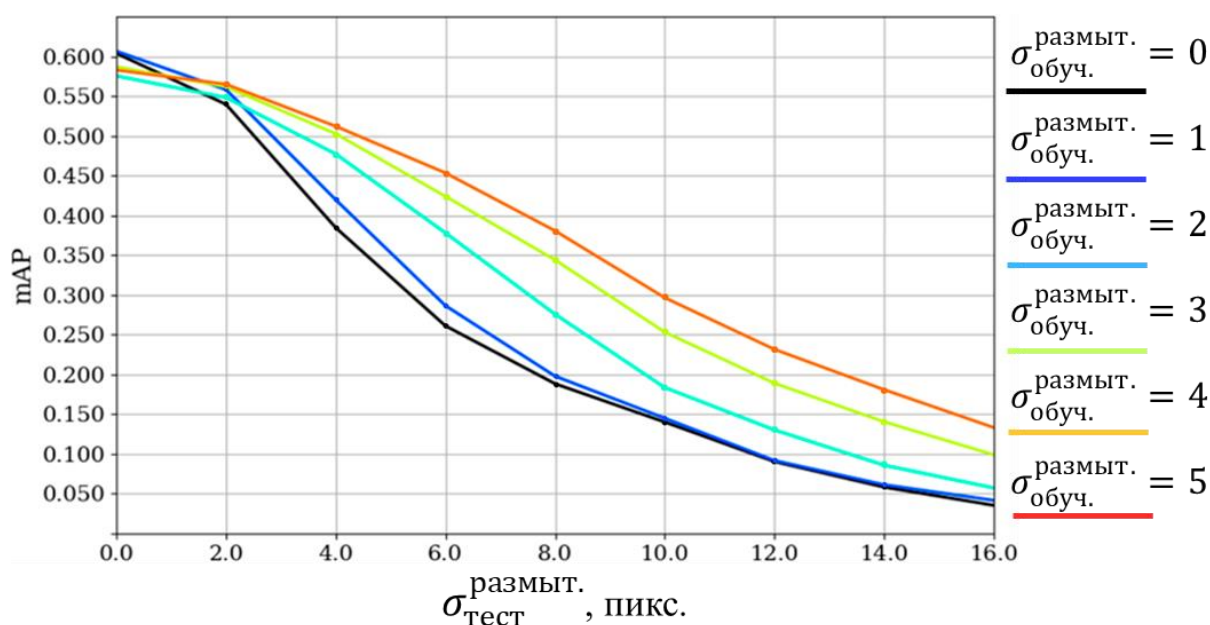


Рисунок 8 – Кривые зависимости mAP от степени искажения гауссова размытия $\sigma_{\text{тест}}^{\text{размыт.}}$ для моделей с параметром $\sigma_{\text{обуч.}}^{\text{размыт.}} = 0$, $\sigma_{\text{обуч.}}^{\text{размыт.}} = 1$, $\sigma_{\text{обуч.}}^{\text{размыт.}} = 2$, $\sigma_{\text{обуч.}}^{\text{размыт.}} = 3$ и $\sigma_{\text{обуч.}}^{\text{размыт.}} = 4$ (черная, синяя, голубая, зеленая и оранжевая кривые соответственно)

Из рисунка 8 видно, что с увеличением степени размытия тестовых изображений качество обнаружения моделей с параметром $\sigma_{\text{обуч.}}^{\text{размыт.}} = 0$ и $\sigma_{\text{обуч.}}^{\text{размыт.}} = 1$ быстро падает. В случае же увеличения разброса степени размытия в обучающем наборе изображений (т.е. с увеличением параметра $\sigma_{\text{обуч.}}^{\text{размыт.}}$) видно, что величина mAP уменьшается медленнее при больших значениях степени размытия тестовых изображений. Однако на исходных (без искажений) изображениях оценка качества обнаружения заметно падает. Полученный результат подтверждает выводы, полученные в разделе 2 настоящей диссертационной работы, а именно, что существует оптимальное значение разброса степени размытия в обучающем наборе изображений, при котором обученная нейросетевая модель будет показывать достаточно высокую интегральную оценку качества обнаружения объектов на неискаженных или слабо искаженных входных изображениях и медленно ухудшаться с ростом степени размытия входных изображений.

Полученные зависимости свидетельствуют о том, что увеличение диапазона уровней размытия в обучающем наборе изображений способствует повышению инвариантности модели к данному виду искажения. Вместе с тем чрезмерное увеличение разброса уровней размытия приводит к ухудшению качества обнаружения на неискаженных изображениях, что обусловлено смещением обучающего распределения в сторону сильно деградированных изображений.

На рисунке 9 показаны кривые зависимости метрики качества обнаружения mAP (4) от различных значений стандартного отклонения аддитивного белого гауссова шума $\sigma_{\text{тест}}^{\text{АБГШ}}$ в тестовом наборе изображений для моделей с параметром $\sigma_{\text{обуч.}}^{\text{размыт.}} = 0$, $\sigma_{\text{обуч.}}^{\text{размыт.}} = 1$, $\sigma_{\text{обуч.}}^{\text{размыт.}} = 2$, $\sigma_{\text{обуч.}}^{\text{размыт.}} = 3$ и $\sigma_{\text{обуч.}}^{\text{размыт.}} = 4$.

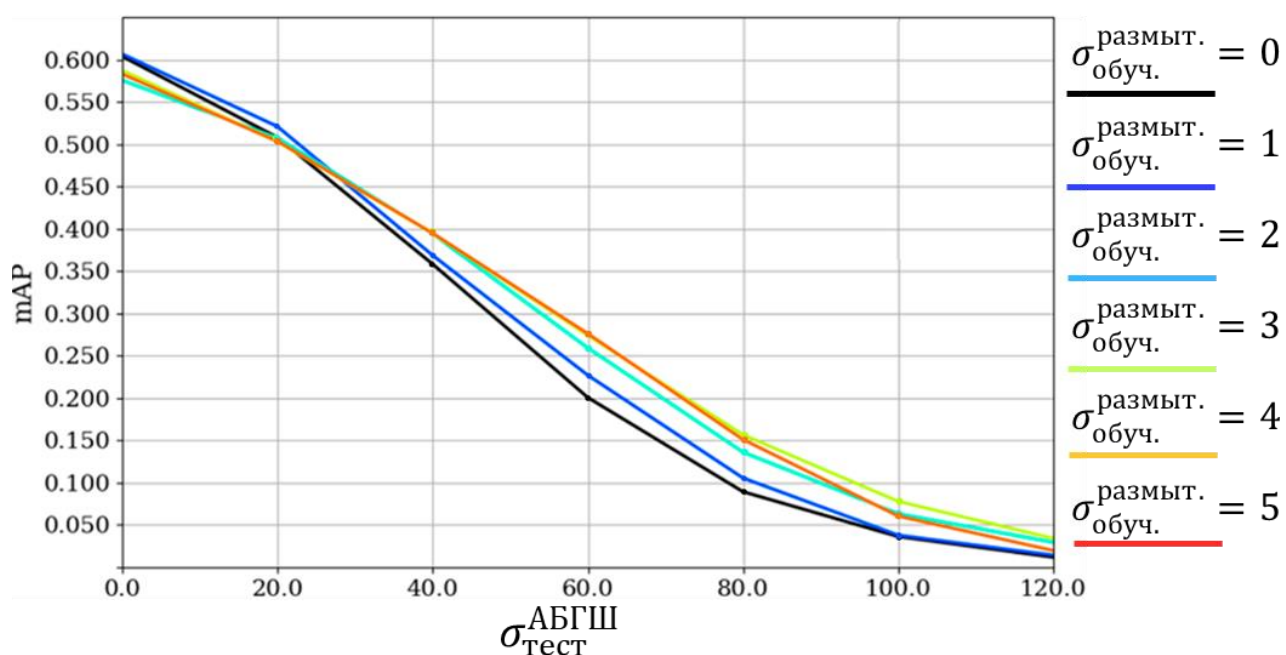


Рисунок 9 – Кривые зависимости mAP от величины стандартного отклонения АБГШ $\sigma_{\text{тест}}^{\text{АБГШ}}$ для с параметром $\sigma_{\text{обуч.}}^{\text{размыт.}} = 0$, $\sigma_{\text{обуч.}}^{\text{размыт.}} = 1$, $\sigma_{\text{обуч.}}^{\text{размыт.}} = 2$, $\sigma_{\text{обуч.}}^{\text{размыт.}} = 3$ и $\sigma_{\text{обуч.}}^{\text{размыт.}} = 4$ (черная, синяя, голубая, зеленая и оранжевая кривые соответственно)

Как видно на рисунке 9, добавление в обучение размытых изображений не приводит к значительному росту качества обнаружения объектов при наличии на распознаваемых изображениях АБГШ.

Полученный результат показывает, что робастность, сформированная при обучении на изображениях с гауссовым размытием, не переносится на искажения иной природы. Это объясняется тем, что гауссово размытие приводит преимущественно к подавлению высокочастотных компонент изображения, тогда как аддитивный белый гауссов шум, напротив, увеличивает уровень высокочастотных составляющих. Следовательно, механизмы компенсации данных видов искажений существенно различаются.

Следовательно, повышение робастности нейросетевой модели к одному виду естественных искажений не обеспечивает аналогичного повышения устойчивости к другим видам деградации изображений. Это свидетельствует о необходимости разработки адаптивной системы предварительной обработки, способной учитывать преобладающий тип и уровень искажений входного изображения перед его

подачей в модель обнаружения объектов. Адаптивный подход к повышению качества функционирования интеллектуальных систем также рассматривается в телекоммуникационных задачах, где изменение параметров системы в соответствии с условиями её работы позволяет повысить эффективность функционирования [42]. В настоящей работе аналогичный принцип адаптации реализуется применительно к предварительной обработке изображений перед их подачей в модель обнаружения объектов.

3.2 Разработка адаптивной схемы для робастного обнаружения и распознавания объектов различного качества при видовом мониторинге

Поскольку влияние размытия на распознавание зашумлённых и размытых изображений различно, необходима предварительная оценка уровня шума на изображении. На данный момент известны методы оценки отношения сигнал/шум по одному изображению, основанные на анализе двумерной автокорреляционной функции изображения [87, 88]. Данные методы применимы к изображениям с высоким разрешением и малым числом значимых крупных объектов. На изображениях с малым разрешением и большим числом мелких объектов данные методы не дают результата ввиду близости автокорреляционных функций таких изображений к шумовым (дельта-функциям). Для решения данной проблемы предложен метод оценки степени статистической неоднородности изображения с малым разрешением и большим числом объектов различной формы, цвета и размера.

Для вычислительно не затратной оценки коэффициента статистической неоднородности, входное изображение разбивается на несколько небольших фрагментов (в данной работе фрагменты имеют размеры 20×20 пикс.). Для входного изображения и всех его фрагментов отдельно рассчитываются величины стандартного отклонения, которые используются для расчета нового

условного коэффициента статистической неоднородности (Statistical Heterogeneity Factor, SHF), формула получения которого выглядит следующим образом [44]:

$$SHF = \frac{\sqrt{\frac{1}{N_{patches}} \sum_{p=1}^{N_{patches}} (std_{image} - std_p)^2}}{std_{image}}, \quad (16)$$

где std_{image} – величина стандартного отклонения яркости изображения;

std_p – величина стандартного отклонения яркости фрагмента с индексом p ,

$N_{patches}$ – количество фрагментов.

Данный параметр характеризует степень различия между глобальной дисперсией изображения и дисперсией его локальных областей.

На рисунке 10 представлены кривые зависимостей величины коэффициента статистической неоднородности для различных искажений, добавленные к изображениям. Полученные данные были рассчитаны для случайной части обучающего набора изображений и усреднены.

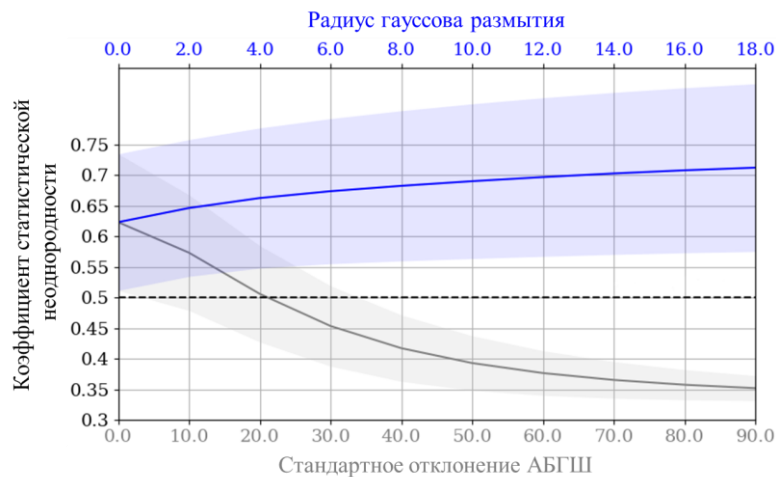


Рисунок 10 – Среднее значение коэффициента статистической неоднородности в зависимости от радиуса гауссова размытия (синяя кривая) и стандартного отклонения аддитивного белого гауссова шума (серая кривая). Полупрозрачная полоса показывает разброс значений коэффициента статистической неоднородности для каждого искажения

Из полученных на рисунке 10 результатов видно, что величина коэффициента статистической неоднородности падает с увеличением значения стандартного отклонения АБГШ, что показывает на отсутствие сильных статистических различий между статистиками целого изображения и статистиками фрагментов этих изображений. Также из графиков на рисунке 10 видно, что величина коэффициента статистической неоднородности растет с увеличением радиуса гауссова размытия, из чего следует, что значения стандартных отклонений размытого изображения и его фрагментов уже значительно отличаются. Данные статистические свойства позволяют без серьезных и вычислительно затратных преобразований входных изображений идентифицировать наличие шума в них. Предложенный метод оценки статистической неоднородности на сложном изображении позволяет без априорного знания о степени и виде искажения, опираясь только на статистические характеристики входных изображений, определить наличие искажения.

Полученная оценка статистической неоднородности используется в разработанной адаптивной системе обработки, упрощенная схема которой представлена на рисунке 11. В соответствии с предложенной схемой, перед входом в модель обнаружения объектов изображение поступает в блок расчета условного коэффициента статистической неоднородности. В случае, если полученный коэффициент меньше определенного порогового значения, изображение считается зашумленным и его необходимо передать на вход гауссова фильтра с предопределённым радиусом размытия, который выполняет роль фильтра нижних частот и снижает влияние высокочастотных составляющих на изображении. В ином случае, считается, что на изображении нет значительного шума и нет необходимости выполнять предварительное размытие. Использование гауссова фильтра обусловлено тем, что он эффективно подавляет высокочастотную составляющую аддитивного белого гауссова шума при низкой вычислительной сложности, что особенно важно для систем видового мониторинга, функционирующих в режиме, близком к реальному времени [44].

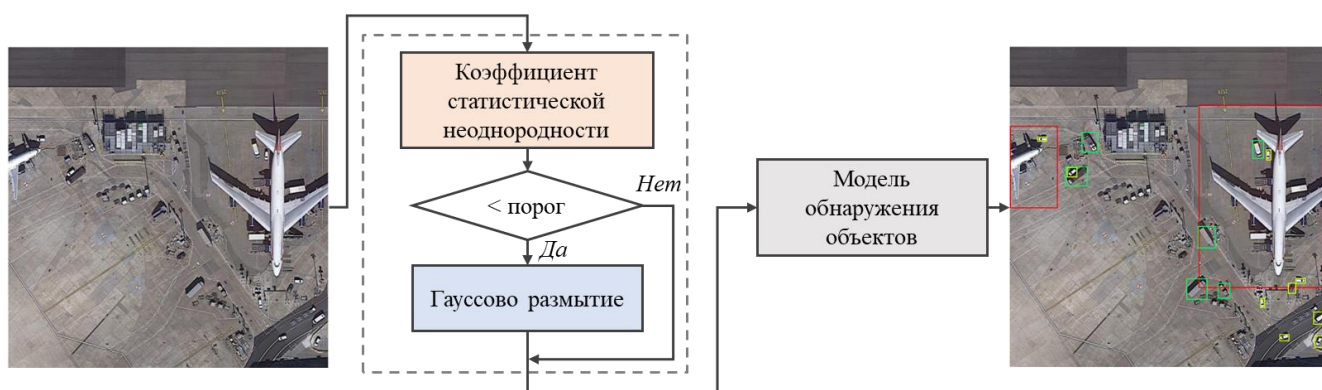


Рисунок 11 – Упрощенная схема разработанного адаптивного метода повышения устойчивости модели обнаружения объектов на искаженных изображениях

Как видно из зависимостей на рисунке 10, если в адаптивной схеме задать слишком низкий порог для решения о пропуске входного изображения через гауссов фильтр, то некоторые изображения с АБГШ не будут проходить через него для снижения влияния высокочастотных составляющих. В случае, если задать слишком высокое пороговое значение, то сильно размытые входные изображения будут дополнительно размываться, что приведет к еще большему искажению изображений. В настоящем исследовании в качестве порога коэффициента статистической неоднородности решено установить значение равное 0,5, которое позволяет не пропускать изображения, подверженные размытию. Данный подход предусматривает выбор порога коэффициента статистической неоднородности на основе анализа статистических характеристик входных изображений, без использования показателей качества модели обнаружения объектов.

Таким образом, разработанная схема реализует адаптивный выбор режима предварительной обработки в зависимости от статистических характеристик входного изображения, обеспечивая вычислительно эффективное определение необходимости предварительной фильтрации перед обнаружением объектов.

3.3 Оценка эффективности разработанной адаптивной схемы

Для полного поиска и анализа параметров, позволяющих получить наиболее устойчивую к искажениям различной природы модель обнаружения объектов,

в данном исследовании был проведен ряд статистических экспериментов, состоящих из измерения метрик качества обнаружения объектов mAP от величины радиуса гауссова размытия, применяемого в адаптивной системе, описанной в подразделе 3.1 настоящей диссертационной работы при различных значениях уровнях гауссова размытия и АБГШ в тестовом наборе изображений.

Для более наглядного сравнения метрик качества модели обнаружения на искаженных различными видами искажений, в данной работе используется интегральная оценка кривой помехоустойчивости, которая представляет собой среднее значение площадей под кривыми mAP для дискретного множества значений стандартных отклонений гауссова размытия и АБГШ и значений радиуса гауссова ядра в тестовом наборе изображений. Такой подход к оценке использован исключительно для получения экспериментального критерия устойчивости при наличии двух разных типов искажений одновременно. Такое усреднение в данной работе используется как условный показатель при равном учёте двух исследуемых типов искажений.

Для расчета площади под кривой помехоустойчивости в настоящем исследовании используется следующая формула:

$$AUC = \sum_{i=0}^N mAP_i \cdot \Delta\sigma_i \quad (17)$$

где N – количество интервалов, на которые делится анализируемая кривая;
 $\Delta\sigma_i$ – ширина отрезка с индексом i по оси абсцисс анализируемой кривой;
 mAP_i – величина mAP в середине отрезка с индексом i .

Тогда, используя формулу (17), формула интегральной оценки кривой помехоустойчивости:

$$S = \frac{AUC_{\text{размыт.}} + AUC_{\text{АБГШ}}}{2} \quad (18)$$

где $AUC_{\text{размыт.}}$ – площадь под кривой величины качества обнаружения mAP от степени размытых изображений в тестовом наборе изображений $mAP(\sigma_{\text{тест}}^{\text{размыт.}})$; $AUC_{\text{АБГШ}}$ – площадь под кривой величины качества обнаружения mAP от степени зашумления изображений в тестовом наборе изображений $mAP(\sigma_{\text{тест}}^{\text{АБГШ}})$.

На рисунке 12 представлены значения интегральной оценки кривой помехоустойчивости, рассчитанной по формуле (18), полученные при определенном радиусе гауссова размытия $\sigma_{\text{адаптивн.}}^{\text{размытие}}$, применяемой в адаптивной системе для моделей с параметром $\sigma_{\text{обуч.}}^{\text{размыт.}} = 0$, $\sigma_{\text{обуч.}}^{\text{размыт.}} = 1$, $\sigma_{\text{обуч.}}^{\text{размыт.}} = 2$, $\sigma_{\text{обуч.}}^{\text{размыт.}} = 3$ и $\sigma_{\text{обуч.}}^{\text{размыт.}} = 4$.

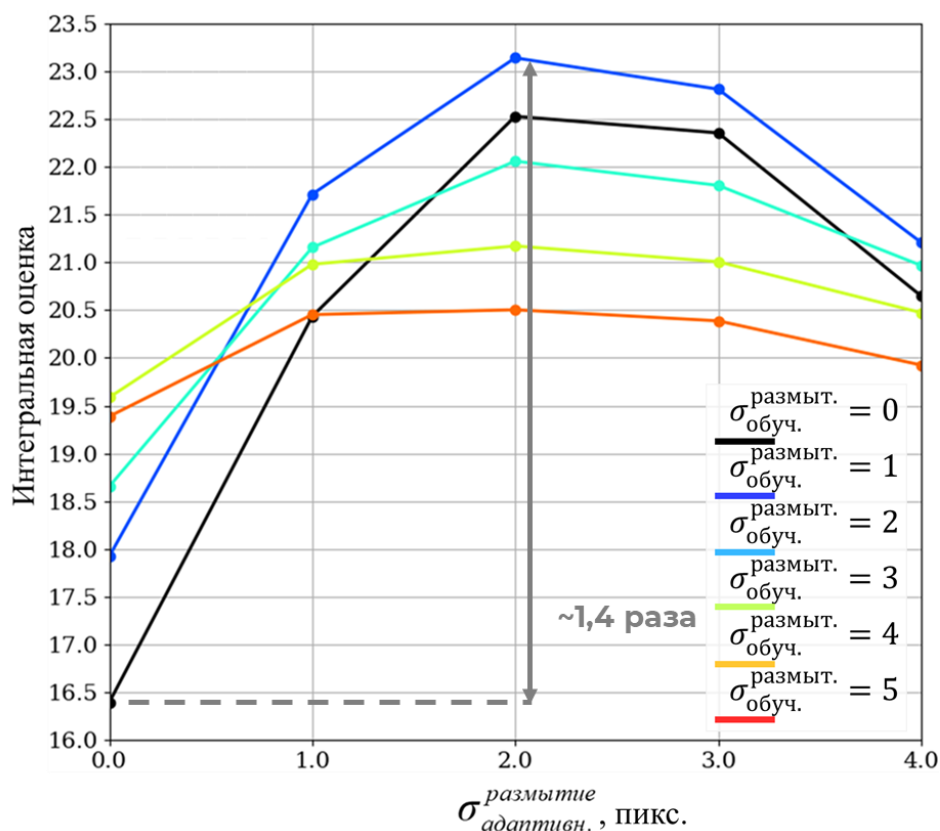


Рисунок 12 – Интегральная оценка кривой помехоустойчивости в зависимости от радиуса гауссова размытия, используемого в адаптивной схеме для моделей с параметром $\sigma_{\text{обуч.}}^{\text{размыт.}} = 0$, $\sigma_{\text{обуч.}}^{\text{размыт.}} = 1$, $\sigma_{\text{обуч.}}^{\text{размыт.}} = 2$, $\sigma_{\text{обуч.}}^{\text{размыт.}} = 3$ и $\sigma_{\text{обуч.}}^{\text{размыт.}} = 4$

Из рисунка 12 видно, что использование предложенной адаптивной схеме с предопределенным радиусом размытия в этой адаптивной схеме приводит к росту величины интегральной оценке кривой помехоустойчивости. Однако существует некоторое критическое значение данного радиуса размытия, выше которого интегральная оценка кривой помехоустойчивости начинает падать. Также из данного рисунка видно, что применение только адаптивной схемы (черная кривая для модели с параметром $\sigma_{\text{обуч.}}^{\text{размыт.}} = 0$) или при использовании только обучения на сформированном обучающем наборе изображение ($\sigma_{\text{адаптивн.}}^{\text{размытие}} = 0$) не приводит к тем результатам, которые получаются при их сочетании.

Также из полученных результатов видно, что модель, обученная на наборе изображений с параметром искажения $\sigma_{\text{обуч.}}^{\text{размыт.}} = 1$, в сочетании с предложенной адаптивной системой с параметром $\sigma_{\text{адаптивн.}}^{\text{размытие}} = 2$, дает наивысшую оценку интегральной величины кривых помехоустойчивости. На рисунке 13 представлена визуализация результатов работы модели обнаружения объектов на нескольких примерах из набора изображений DOTA [89], к которым добавлен аддитивный белый гауссов шум. В левой колонке показаны результаты обнаружения объектов без использования разработанного метода, в то время как в правой колонке показаны результаты работы модели обнаружения объектов с использованием разработанного метода. Из анализа рисунка следует, что использование разработанного метода позволяет повысить полноту обнаружения объектов на зашумленных изображениях, обеспечивая обнаружение объектов, не выявленных исходной моделью.

Результаты обнаружения объектов
без использования разработанного метода



Результаты обнаружения объектов
с использованием разработанного метода



Рисунок 13 – Визуализация результатов работы модели обнаружения объектов на нескольких зашумленных примерах из набора аэрофотоизображений

На рисунках 14-15 представлены кривые, показывающие зависимости метрики качества обнаружения от радиуса гауссова размытия $\sigma_{\text{тест}}^{\text{размыт.}}$ и различных значений стандартного отклонения аддитивного белого гауссова шума $\sigma_{\text{тест}}^{\text{АБГШ}}$ в тестовом наборе изображений, соответственно. Данные зависимости представлены для моделей с параметром $\sigma_{\text{обуч.}}^{\text{размыт.}} = 0$, $\sigma_{\text{обуч.}}^{\text{размыт.}} = 1$ и $\sigma_{\text{обуч.}}^{\text{размыт.}} = 2$ (черная, синяя и голубая кривые соответственно) без добавления адаптивной системы (пунктирные кривые) и с применением предложенной адаптивной системой (сплошные кривые).

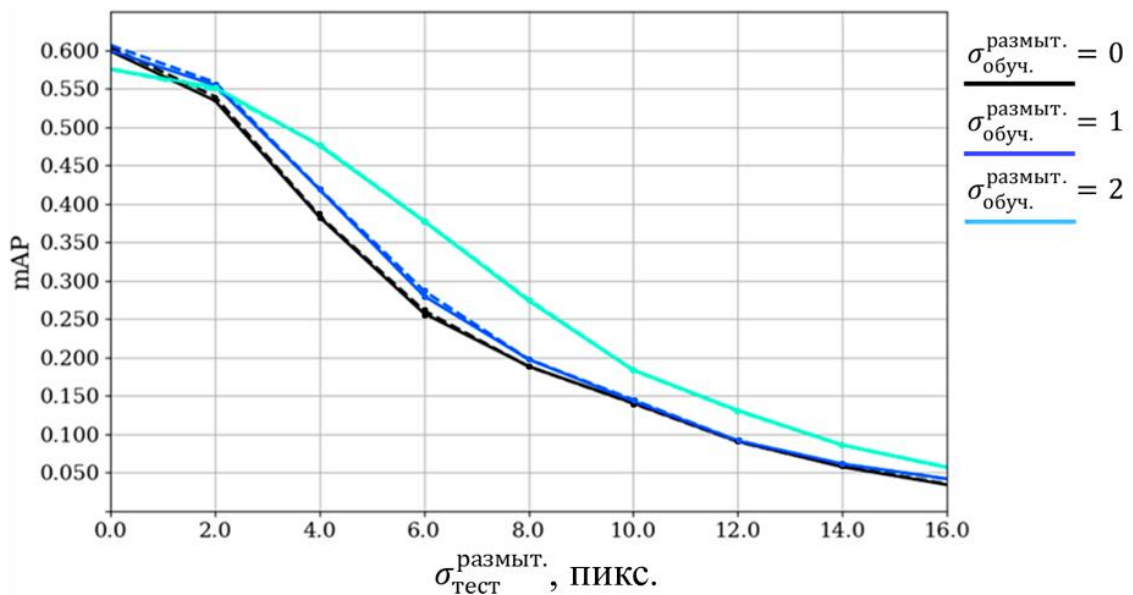


Рисунок 14 – Кривые зависимости метрики качества mAP от радиуса гауссова размытия $\sigma_{\text{тест}}^{\text{размыт.}}$ для моделей с параметром $\sigma_{\text{обуч.}}^{\text{размыт.}} = 0$, $\sigma_{\text{обуч.}}^{\text{размыт.}} = 1$ и $\sigma_{\text{обуч.}}^{\text{размыт.}} = 2$ (черная, синяя и голубая кривые соответственно) без предлагаемой адаптивной системы (пунктирные кривые) и с ней (сплошные кривые)

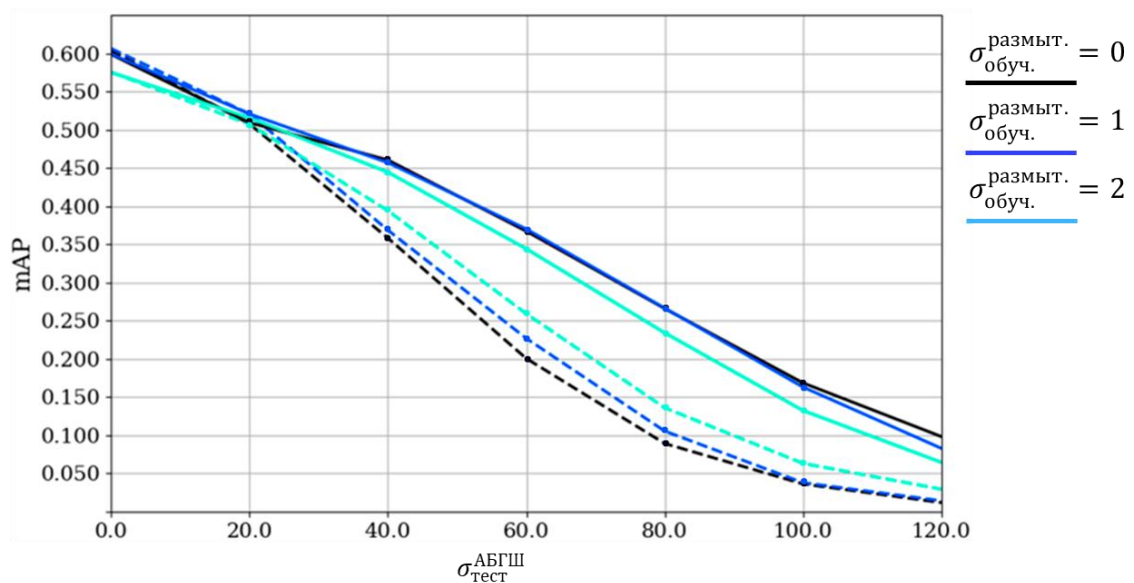


Рисунок 15 – Кривые зависимости метрики качества mAP от величины стандартного отклонения АБГШ $\sigma_{\text{тест}}^{\text{АБГШ}}$ для моделей для моделей с параметром $\sigma_{\text{обуч.}}^{\text{размыт.}} = 0$, $\sigma_{\text{обуч.}}^{\text{размыт.}} = 1$ и $\sigma_{\text{обуч.}}^{\text{размыт.}} = 2$ (черная, синяя и голубая кривые соответственно) без предлагаемой адаптивной системы (пунктирные кривые) и с ней (сплошные кривые)

Как видно на рисунке 14, кривые зависимостей не изменились при добавлении адаптивной системы на размытых тестовых изображениях, что

означает правильное применение коэффициента наличия шума, который не пропускает уже размытые изображения через гауссов фильтр. Из результатов, показанных на рисунке 15 видно, что применение адаптивной системы для изображений с шумом приводит к росту метрики качества работы модели обнаружения объектов. Полученные выводы подтверждают результаты, полученные с использованием интегральной оценки кривой помехоустойчивости.

В результате данного исследования предложен метод повышения робастности системы нейросетевого обнаружения объектов на аэрофотоизображениях. Данный метод состоит из двух этапов. Первый этап включает в себя аугментацию обучающих изображений, заключающуюся в добавлении размытых изображений, степень размытия которых задается генератором случайных величин с полунормальной функцией распределения с параметром стандартного отклонения $\sigma_{\text{обуч.}}^{\text{размыт.}} = 1$. Второй этап представляет собой адаптивную схему, состоящую из блока статистической оценки статистической неоднородности на входном изображении и гауссова фильтра с радиусом размытия $\sigma_{\text{адаптивн.}}^{\text{размытие}} = 2$, позволяющего подавить высокочастотные шумовые составляющие изображения. Разработанный метод приводит к повышению интегральной оценки кривой помехоустойчивости в 1,4 раза по сравнению с исходной моделью обнаружения объектов без использования разработанного способа формирования обучающего набора изображений и адаптивной предварительной обработки.

Таким образом, разработанный метод позволяет существенно повысить устойчивость нейросетевой модели обнаружения транспортных средств одновременно к гауссову размытию и аддитивному белому гауссову шуму без модификации архитектуры модели обнаружения объектов. Это подтверждает эффективность совместного применения разработанного способа формирования обучающего набора изображений и адаптивной предварительной обработки изображений как универсального подхода к повышению качества обнаружения объектов в условиях естественных искажений, характерных для задач видового мониторинга.

Выводы по разделу 3

В результате проведенных исследований, направленных на разработку метода адаптивной предварительной обработки изображений при видовом мониторинге, получены следующие результаты.

1. Исследование влияния различных моделей естественных искажений на модель обнаружения объектов, обученную с использованием разработанного способа формирования обучающего набора изображений, показало, что повышение устойчивости нейросетевой модели к гауссову размытию не обеспечивает аналогичного повышения устойчивости к аддитивному белому гауссову шуму. Установлено, что механизмы компенсации данных видов искажений существенно различаются, поскольку гауссово размытие подавляет высокочастотные составляющие изображения, тогда как аддитивный белый гауссов шум, напротив, увеличивает их уровень. Полученный результат свидетельствует о необходимости разработки адаптивной предварительной обработки, учитывающей преобладающий тип естественного искажения входного изображения.

2. Разработана адаптивная схема предварительной обработки изображений, основанная на оценке статистической неоднородности входного изображения. Предложенная схема обеспечивает вычислительно эффективное определение необходимости предварительной фильтрации и реализует адаптивный выбор режима обработки без априорной информации о типе и уровне естественных искажений.

3. Разработан метод повышения робастности нейросетевого обнаружения объектов на аэрофотоизображениях, объединяющий разработанный способ формирования обучающего набора изображений и адаптивную предварительную обработку изображений. Первый этап метода заключается в аугментации обучающих изображений размытыми изображениями, степень размытия которых задается полунормальным распределением с параметром $\sigma_{\text{обуч.}}^{\text{размыт.}} = 1$, а второй этап включает адаптивную предварительную обработку с использованием оценки

статистической неоднородности изображения и гауссова фильтра с радиусом $\sigma_{\text{адаптивн. размытие}} = 2$. Экспериментально показано, что совместное применение разработанных решений обеспечивает увеличение интегральной оценки кривой помехоустойчивости в 1,4 раза по сравнению с исходной моделью обнаружения объектов.

4. В результате проведенных исследований разработан адаптивный метод предварительной обработки изображений, основанный на оценке статистической неоднородности входного изображения и автоматическом выборе режима фильтрации, а также обоснованы параметры его совместного применения с разработанным способом формирования обучающего набора изображений. Экспериментально показано, что совместное использование разработанных методов обеспечивает повышение интегральной оценки помехоустойчивости модели в 1,4 раза одновременно к гауссову размытию и аддитивному белому гауссову шуму по сравнению с исходной моделью без изменения архитектуры нейросетевой модели обнаружения объектов. Полученные результаты подтверждают эффективность разработанного метода и позволяют сделать вывод о подтверждении второго положения, выносимого на защиту.

4. РАЗРАБОТКА МЕТОДА ЛОКАЛИЗАЦИИ И УДАЛЕНИЯ СОСТЯЗАТЕЛЬНЫХ ПАТЧ-АТАК НА ИЗОБРАЖЕНИЯХ ПРИ ВИДОВОМ МОНИТОРИНГЕ

4.1 Описание разработанного метода локализации и удаления состязательных патч-атак на изображениях

Наряду с естественными деградациями современные системы компьютерного зрения подвержены искусственным воздействиям, целенаправленно создаваемым для нарушения работы нейросетевых моделей. Наиболее опасным видом таких воздействий являются состязательные патч-атаки, представляющие собой локальные области изображения, специально оптимизированные для снижения качества обнаружения объектов.

В отличие от естественных искажений, состязательные патчи формируются с учетом особенностей функционирования нейросетевой модели и способны существенно снижать эффективность обнаружения объектов даже при незначительной занимаемой площади изображения. Это требует разработки специализированных методов защиты, ориентированных на локализацию таких патчей и устранение их влияния до выполнения обнаружения объектов.

Разработанный метод локализации и удаления состязательных патч-атак на изображениях при видовом мониторинге позволяет повысить устойчивость нейросетевой модели обнаружения объектов к искусственным воздействиям без модификации ее архитектуры [40].

Разработанный метод основан на идее обнаружения и устранения аномалии на входных изображениях. В качестве аномалий в данной работе рассматриваются локализованные в пространстве изображения фрагменты, которые отличаются от тех же фрагментов на доброкачественных изображениях, т.е. изображениях без добавления состязательных атак.

Упрощенная схема метода локализации и удаления состязательных патч-атак на изображениях представлена на рисунке 16 [41].

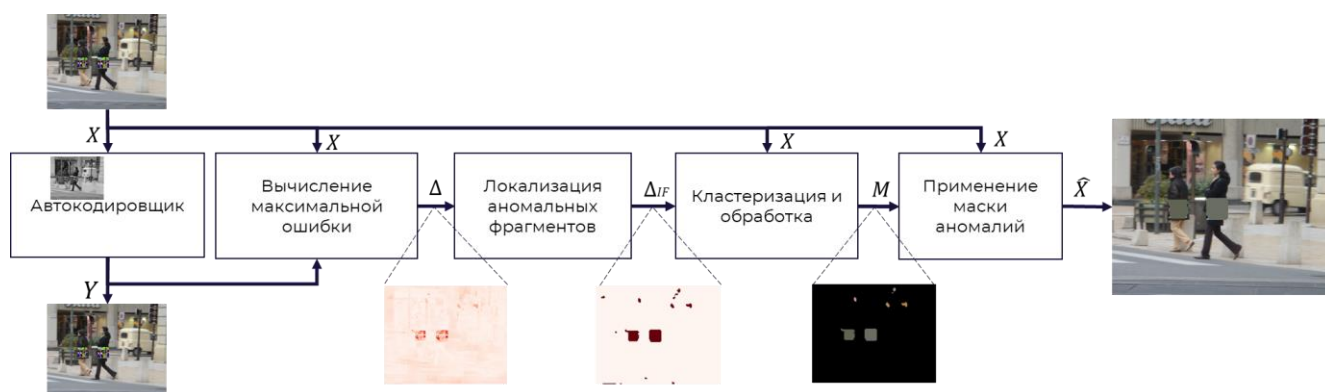


Рисунок 16 – Упрощенная схема метода локализации и удаления состязательных патч-атак на изображениях

На рисунке 16 нейросетевой автокодировщик стремится сгенерировать изображение на основе входного изображения. Блок вычисления максимальной ошибки вычисляет разницу между входным и сгенерированным изображениями. Блок локализации аномальных фрагментов нацелен на выделение аномальных фрагментов из входных изображениях, а блок кластеризация и обработка группирует полученные на предыдущем шаге аномалии в кластеры и обрабатывает их. Полученная в результате этих блоков карта аномалий, на которой выделены участки с аномалиями, применяется к исходному изображению для удаления аномальных участков из изображения. Результат применения карты аномалий поступает на вход обнаружителя объектов.

Ниже представлено более подробное описание каждого блока из рисунка 16.

4.1.1 Автокодировщик

Для решения проблемы отсутствия априорно известных доброкачественных изображений, в данной работе предлагается сгенерировать приближенно доброкачественное изображения на основе входного.

В настоящее время существует метод AnoGAN [90], решающий задачу нейросетевого восстановления изображений. Однако нейросетевая архитектура AnoGAN довольно проста и способна генерировать изображения, содержащие, например, простые текстурные паттерны и изображения объектов, занимающих большую площадь на изображении. Стоит отметить, что AnoGAN показал высокое качество работы при решении задачи поиска аномальных участков на медицинских снимках.

В настоящем же исследовании используются сложные данные, т.е. изображения, имеющих сложный фон и содержащих различные объекты (люди, дорожные знаки, транспортные средства). А также не всегда эти изображения имеют высокое качество. Таким образом, для работы со сложными изображениями необходимо использовать довольно глубокую нейросетевую архитектуру, что в свою очередь приводит к тяжелому обучению в условиях ограниченных наборов данных.

Для решения данной проблемы предложено использовать автокодировщик, представляющий собой глубокую сверточную нейронную сеть (ГСНС), которая также, как и в [90], обучается без учителя исключительно на чистых изображениях. Такой подход позволяет получить систему, у которой отсутствует знание об атаке, что позволяет повысить ее применимость к различным вариациям атак. Автокодировщик состоит из двух частей: кодировщика, решающего задачу сжатия входных изображений; и декодировщика, который стремится воссоздать исходное изображение из их сжатых версий [91]. Данный аспект позволяет сжать информацию о состязательной атаке на этапе кодировщика и восстановить изображение из сжатой информации со значительным ослаблением атаки путём снижения информации о ней.

Также для предварительного ослабления атаки путём снижения информации о ней в данной работе предложено обучать нейронную сеть таким образом, что на ее вход подается черно-белое изображение, а на выходе – цветной вариант входного изображения. Т.е. автокодировщик стремится восстановить исходное цветное изображение из ее черно-белого варианта.

На рисунке 17 представлена упрощенная схема используемого автокодировщика. В качестве кодировщика используется полностью сверточная часть глубокой нейронной архитектуры ResNet50 [92]. Такая глубокая сверточная нейронная сеть состоит из пяти блоков, каждый из которых уменьшает пространственные размеры входных изображений в два раза. Как и в работе [92], блок *conv64* представляет собой последовательно соединенные слои свертки с 64 выходными фильтрами, слой *BatchNormalization*, слой активации *ReLU* и слой уменьшения пространственного разрешения в два раза *MaxPooling*. Блоки *resconvU* представляют собой остаточные блоки с количеством выходных фильтров U .

В качестве декодировщика предложено использовать пять последовательно соединенных блоков, каждый из которых увеличивает пространственные размеры входных изображений. В данной работе в качестве блока *deconvU* предложено использовать последовательно соединенные транспонированный слой свертки слои (*ConvTranspose*) с количеством выходных фильтров U , слой *BatchNormalization*, слой активации *ReLU*. В качестве функции потерь для обучения данного автокодировщика используется функция вычисления среднеквадратической ошибки (СКО) между исходным цветным и сгенерированным изображениями.

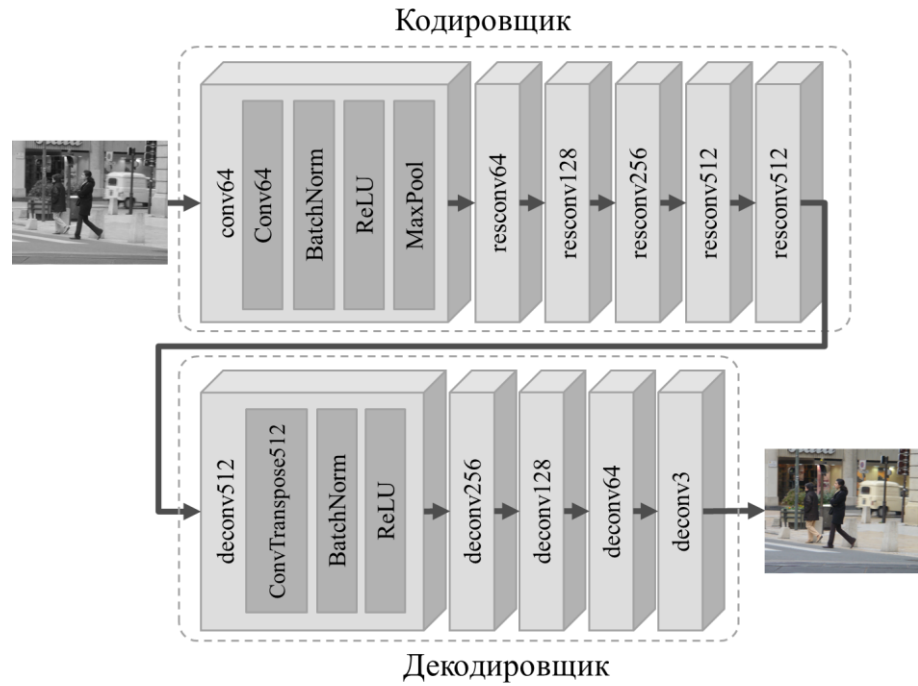


Рисунок 17 – Упрощенная схема используемого автокодировщика

Таким образом, подобный метод не требует разметки изображений, что делает данный подход более гибким для решения многих задач. Также видно, что предлагаемая сеть довольно проста, что позволяет довольно легко повторить данный вариант. Однако стоит отметить, что от качества работы данной архитектуры зависит качество работы алгоритма в целом [40].

4.1.2 Вычисление максимальной ошибки между исходным и сгенерированным изображениями

После обучения автокодировщика необходимо получить разницу между исходным и сгенерированным изображениями. Так, обозначим исходное цветное изображение как $X \in \mathbb{R}^{H \times W \times 3}$, а сгенерированное цветное изображение как $Y \in \mathbb{R}^{H \times W \times 3}$. Пространственные размеры изображений составляют H и W пикселей по высоте и ширине соответственно; количество цветовых каналов равно 3, что соответствует цветовому RGB-формату. Величина яркости пикселей изображений нормирована относительно максимально возможной яркости

пикселей, т.е. значения величин X и Y лежат в диапазоне от 0 до 1. В данной работе в качестве вычисления величины ошибки между исходным и сгенерированным изображениями используется следующая формула:

$$\Delta[i, j] = \max_{color} (|X[i, j, color] - Y[i, j, color]|), \quad (19)$$

где i, j – индексы строки и столбца изображений, соответственно;

$color$ – индекс цветового канала изображения, который может принимать значения 0, 1 и 2 для красного, зеленого и синего каналов соответственно;

$\Delta \in \mathbb{R}^{H \times W}$ – карта ошибок между исходным и сгенерированным изображениями.

Максимизация величины ошибки по цветовой составляющей позволяет уменьшить количество измерений для последующей обработки, а также уменьшить информацию об ошибке восстановления цвета. На рисунке 18 представлен пример полученной по формуле (19) карты ошибок для чистого и состязательного изображений.

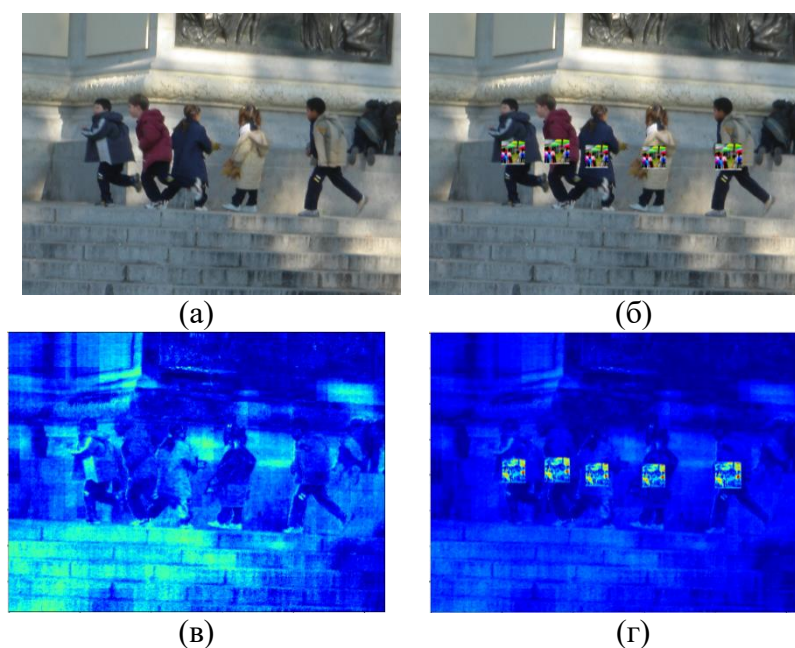


Рисунок 18 – Пример доброкачественного изображения (а) и соответствующая ему карта ошибок (в); пример состязательного изображения (б) и соответствующая ему карта ошибок (г)

На рисунке 18 (г) визуально выделяются области с патч-атакой, которые можно рассматривать как аномалии, т.е. значения карты ошибок велики. Также стоит отметить, что на рисунке 18 видны погрешности восстановления изображений, которые не должны быть выделены как аномалии. Для этих целей разработан алгоритм обнаружения аномалий, который описан ниже.

4.1.3 Алгоритм локализации аномальных фрагментов

Упрощенная схема последовательности операций в блоке локализации аномальных фрагментов представлена на рисунке 19.

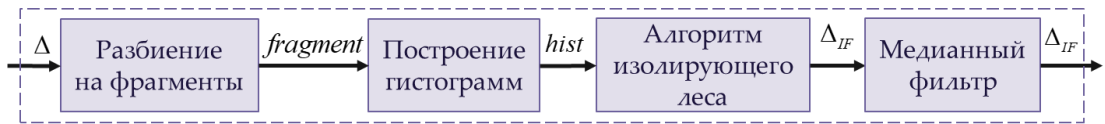


Рисунок 19 – Упрощенная схема алгоритма блока локализации аномальных фрагментов

В данной работе для выявления аномальных участков карта ошибок разбивается на фрагменты скользящим окном размером $F \times F$ с шагом $S \times S$ в соответствии со следующей формулой:

$$fragment[r, c, r_F, c_F] = \Delta[r \cdot S + r_F, c \cdot S + c_F], \quad (20)$$

где $fragment \in \mathbb{R}^{\frac{H}{S} \times \frac{W}{S} \times F \times F}$ – результат разбиения карты ошибок на фрагменты;

r и c – индексы фрагментов по высоте и ширине соответственно;

r_F и c_F – индексы по высоте и ширине каждого фрагмента соответственно.

Из рисунка 18 (г), видно, что на месте патч-атаки величина ошибок резко меняется относительно фрагментов без патч-атаки. Информацию об изменении ошибок внутри фрагментов можно получить из гистограммы. Таким образом, для дальнейшей обработки карты ошибок используются гистограммы яркости

выделенных фрагментов карты ошибок, получение которых можно описать следующей формулой:

$$hist[r, c] = histogram(fragment[r, c, r_F, c_F], B), \quad (21)$$

где $hist \in \mathbb{R}^{\frac{H}{s} \times \frac{W}{s} \times B}$ – нормированный результат вычисления гистограмм всех фрагментов;

B – количество интервалов для построения гистограммы;

$histogram(\cdot)$, – операция вычисления нормированной гистограммы.

Полученные гистограммы поступают на вход алгоритма Isolation Forest [93], который выделяет фрагменты с аномальными гистограммами. Алгоритм Isolation Forest представляет собой мощный инструмент выделения аномальных данных. Таким образом, резкое изменение гистограмм фрагментов, где находится патч-атака, относительно остальных гистограмм можно расценивать как аномальное поведение, что и детектируется алгоритмом Isolation Forest. Данный этап можно представить следующей формулой:

$$\Delta_{IF} = IsolationForest(hist), \quad (22)$$

где $\Delta_{IF} \in \{1, 0\}^{\frac{H}{s} \times \frac{W}{s}}$ – полученная карта аномальных фрагментов, которая представляет собой бинарную матрицу с пространственными размерами $\frac{H}{s} \times \frac{W}{s}$, такую, что

$$\Delta_{IF}[r, c] = \begin{cases} 1, & \text{если } fragment[r, c] \text{ аномален} \\ 0, & \text{в ином случае} \end{cases}. \quad (23)$$

Последний этап алгоритма представляет собой медианный фильтр, позволяющий удалить разрозненные фрагменты, отмеченные как аномальные и, в то же время, сгруппировать плотные аномальные фрагменты.

На рисунке 6 представлен пример работы алгоритма локализации аномальных фрагментов для чистого и состязательного изображений, упомянутых на рисунке 20 (а) и (б) соответственно.

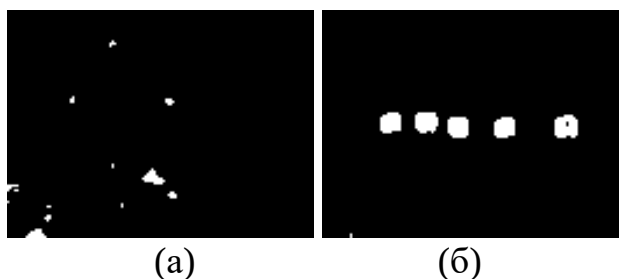


Рисунок 20 – Примеры карт аномальных фрагментов для доброкачественного изображения (а) и для состязательного изображения (б)

Как видно из рисунка 20, описанный выше алгоритм выделяет аномальные фрагменты. Однако из-за отсутствия априорной информации об атаках (их наличие, количество, размеры и формы) необходимо произвести пост-обработку полученных данных информации, алгоритм которой описан в следующем разделе.

4.1.4 Алгоритм кластеризации и обработки карты аномальных фрагментов

Упрощенная схема предложенного алгоритма кластеризации и обработки карты аномальных фрагментов представлена на рисунке 21.

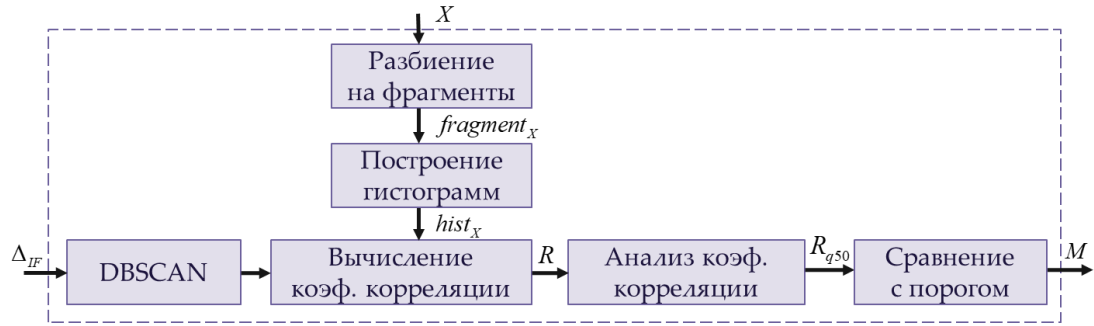


Рисунок 21 – Упрощенная схема алгоритма блока кластеризации и обработки

Полученные на предыдущем этапе координаты аномальных объектов поступают в качестве входных данных в алгоритм кластеризации DBSCAN [94], который решает задачу объединения близкорасположенных аномальных фрагментов и удаления случайно отмеченных аномальных фрагментов. Алгоритм кластеризации DBSCAN позволяет находить группы плотно расположенных точек в условиях отсутствия знаний о количестве и формах этих групп, и в то же время алгоритм DBSCAN устойчив к шумовым выбросам, которые могут быть в случае недостаточно точной генерации изображения в блоке автокодировщика, например. В результате этапа кластеризации получим L кластеров, каждое из которых содержит различное количество фрагментов, входящих в определенный кластер.

Следующим этапом является этап вычисления коэффициента корреляции Пирсона [95] между гистограммами фрагментов, входящих в кластер с индексом l , и гистограммами фрагментов, входящих в окрестность вокруг рассматриваемого кластера с индексом l . Коэффициент корреляции представляет собой скаляр, указывающий на степень взаимосвязи между двумя переменными. Вычисление коэффициента корреляции удобно использовать из-за его однозначной трактовки результатов: коэффициент корреляции Пирсона стремится к 1, если сравниваемые гистограммы очень похожи; или стремится к 0, если сравниваемые гистограммы абсолютно не похожи.

На данном этапе анализируются гистограммы фрагментов исходного изображения X . Т.е. предварительно, исходное изображение X разбивается на фрагменты по тому же правилу, что и в блоке локализации аномальных фрагментов:

$$fragment_X[r, c, r_F, c_F, color] = X[r \cdot S + r_F, c \cdot S + c_F, color], \quad (24)$$

где $fragment_X \in \mathbb{R}^{\frac{H}{S} \times \frac{W}{S} \times F \times F \times 3}$ – результат разбиения исходного изображения X на фрагменты скользящим окном размером $F \times F$ с шагом $S \times S$;
 $color$ – индекс цветового канала.

Далее также, как и в блоке локализации аномальных фрагментов, происходит построение гистограммы для каждого фрагмента по следующему правилу:

$$hists_X[r, c] = \text{histogram}(fragment_X[r, c, r_d, c_d, color], B_X), \quad (25)$$

где $hists_X \in \mathbb{R}^{\frac{H}{S} \times \frac{W}{S} \times B_X}$ – нормированный результат вычисления гистограммы каждого фрагмента;

B_X – количество ячеек для построения гистограммы.

Результат работы алгоритма вычисления коэффициента корреляции для каждого кластера с индексом l представляет собой двумерную матрицу, каждый элемент которой вычисляется следующим образом:

$$\begin{aligned} R[l, a_l, n_l] = \\ = \text{PearsonCorrelation}(hists_{Xabnormal}[a_l], hists_{Xnormal}[n_l]), \end{aligned} \quad (26)$$

где $hists_{Xabnormal} \in \mathbb{R}^{A \times B_X}$ – гистограммы фрагментов исходного изображения X , классифицированные как аномальные в блоке локализации аномальных фрагментов;

$hists_{Xnormal} \in \mathbb{R}^{N \times B_X}$ – гистограммы фрагментов исходного изображения X , классифицированные как нормальные в блоке локализации аномальных фрагментов;

A – количество аномальных фрагментов;

N – количество нормальных фрагментов;

a_l – индекс аномального фрагмента в кластере с индексом l ;

n_l – индекс нормального фрагмента, находящегося в окрестности кластера с индексом l .

В результате выполнения данного алгоритма получаем множество двумерных матриц коэффициентов корреляции Пирсона для каждого кластера.

Ввиду того аспекта, что состязательная патч-атака представляет собой локальное в пространстве возмущение, логически приходим к выводу, что изображение вокруг состязательной патч-атаки должно семантически сильно отличаться (не коррелировать) от изображения самой патч-атаки. Исходя из таких логических рассуждений получаем, что гистограммы состязательных патч-атак должны сильно отличаться от гистограмм окрестности вокруг этой патч-атаки. Так как коэффициент корреляции указывает на уровень схожести, то коэффициент корреляции между состязательной патч-атакой и окрестностью вокруг данной атаки должен быть низким из-за семантической независимости патч-атаки от окружающего изображения. Таким образом, для каждого кластера вычисляется медианное значение коэффициента корреляции Пирсона по следующему правилу:

$$R_{q50}[l] = \underset{a_l}{\text{median}} \left(\underset{n_l}{\text{median}}(R[l, a_l, n_l]) \right), \quad (27)$$

где операции $\underset{n_l}{\text{median}}(\cdot)$ и $\underset{a_l}{\text{median}}(\cdot)$ обозначают вычисление медианного значения вдоль определенной переменной.

Полученное значение медианного значения коэффициента корреляции Пирсона сравнивается с порогом, установленным эмпирическим путем. В случае, если данное значение ниже порога, кластер считается аномальным.

В данной работе для формирования карты аномалий M (рисунок 16) в области, где обнаружен аномальный кластер, извлекаются значения яркости пикселей из исходного изображения X , вычисляется их среднее значение, и это

значение присваивается пикселям в карте аномалий M в месте нахождения данного кластера.

На рисунке 22 представлен пример карт аномальных фрагментов для чистого и состязательного изображений, упомянутых на рисунке 20 (а) и (б) соответственно. Карты получены в результате работы алгоритма кластеризации и обработки.

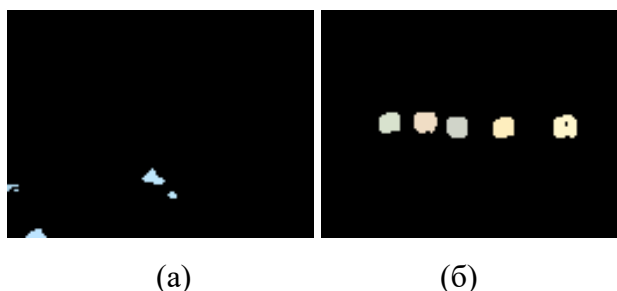


Рисунок 22 – Примеры карт аномалий для доброкачественного изображения (а) и для состязательного изображения (б)

Как видно из рисунка 22, описанный выше алгоритм обрабатывает выделенные в блоке локализации аномальных фрагментов изображения и производит решение по отнесению выделенных фрагментов к аномалиям.

4.1.5 Алгоритм применения маски аномалий к изображению

Благодаря вышеописанному алгоритму получаем маску аномалий, которую необходимо применить к исходному изображению X следующим образом:

$$\hat{X}[i, j] = \begin{cases} M[i, j], & \text{если } M[i, j] > 0, \\ X[i, j], & \text{в противном случае} \end{cases} \quad (28)$$

Данная операция позволяет сохранить исходное изображение и в то же время скрыть аномальные участки на изображении. Полученное изображение $\hat{X}$ в дальнейшем поступает на вход атакуемой системы обнаружения объектов для последующего обнаружения интересующих объектов. Таким образом,

обнаруженная патч-атака на изображении заменяется однородно окрашенной фигурой, что позволяет снизить содержание атаки до минимума, тем самым минимизируя её влияние на систему распознавания объектов.

4.2 Оценка эффективности разработанного метода

В качестве модели обнаружения объектов используется предварительно обученная на обучающем наборе изображений MS COCO модель YOLOv3 [31].

Разработанная в данном исследовании автокодировщик обучался только на той части обучающего набора изображений MS COCO, на который есть изображения людей. Назовем данный набор изображений как MS COCO Person. Объем обучающего набора MS COCO Person составляет 64115 изображений. Автокодировщик обучался 48 эпох, размер батча равен 4. Скорость обучения сначала повышалась до значения $1e-4$ за 500 итераций, после чего снижалась до величины $1e-5$ по косинусоидальному графику. В качестве оптимизатора выбран Adam с параметрами $\beta_1 = 0,9$ и $\beta_2 = 0,99$.

В алгоритме обнаружения аномальных фрагментов используется окно разбиения карты ошибок на фрагменты, окно разбиения имеет размер $F=8$ и шаг $S=8$.

Используемые алгоритмы Isolation Forest и DBSCAN реализованы на языке программирования Python с использованием пакета scikit-learn [96]. Алгоритм DBSCAN требует задания двух параметров: минимального числа точек $minPts=9$, которые должны образовывать плотную область, и размер окрестности $\varepsilon=2\sqrt{2}$ для формирования плотной области. Остальные параметры, используемые для задания алгоритмов Isolation Forest и DBSCAN берутся по умолчанию из пакета scikit-learn. Порог медианного коэффициента Пирсона задается равный 0,1.

Для сравнения в данной работе используются методы, направленные на защиту нейросетевых моделей обнаружения объектов от состязательных патч-атак, а именно метод PAD [81] и метод SAC [76]. Методы, основанные

на состязательном обучении, в сравнение не включались, поскольку они реализуют иной принцип повышения устойчивости, предполагающий модификацию обучающего процесса и, как правило, повторное обучение самой модели обнаружения на состязательных примерах. В настоящем исследовании основной целью является разработка внешнего метода защиты модели обнаружения объектов на этапе обработки входного изображения, который позволяет повысить устойчивость уже обученного обнаружителя без изменения его архитектуры и без его повторного обучения. Также стоит отметить, что реализации сравниваемых методов лежат в открытом доступе, что делает возможным сравнить их при использовании одного и того же тестового набора изображений и обнаружителя объектов.

Изначально метод SAC применялся для обнаружения цифровой патч-атаки projected gradient descent (PGD) [36], используя для этого самообучение. Однако проведенный анализ работоспособности метода SAC на изображениях с патч-атакой из работы [8] не привел к положительным результатам. Для данного эксперимента метод SAC был обучен на обучающем наборе изображений MS COCO Person, к которому применены состязательные патч-атаки [8]. Остальные настройки для этого метода остались без изменений.

В таблице 1 представлен сравнительный анализ качества существующих методов защиты обнаружителя объектов от состязательных патч-атак. Для экспериментальной оценки разработанного метода использован тестовый набор изображений INRIA-Person [97], являющийся одним из эталонных наборов изображений для разработки и сравнительной оценки алгоритмов обнаружения пешеходов на статических изображениях. Тестовый набор изображений содержит 288 изображений с пешеходами. Особенностью INRIA-Person является разнообразие поз, ракурсов и фонов при относительно высоком качестве исходных изображений, благодаря чему набор изображений получил широкое распространение при исследовании алгоритмов обнаружения пешеходов, что в свою очередь позволяет провести сравнительный анализ полученных результатов с существующими методами защиты от состязательных атак.

Таблица 1 – Метрика качества обнаружения объектов mAP(%) в режиме состязательной патч-атаки [8] для различных методов защиты. Наивысшие показатели выделены жирным, ближайшие к наивысшим показателям – подчеркнуты

Методы защиты	mAP без добавления атак, %	mAP с добавлением атак, %
-	<u>93,3</u>	46,8
SAC	<u>93,3</u>	53,7
PAD	92,7	<u>76,9</u>
Разработанный	93,5	81,0

Результаты, представленные в таблице 1, показывают, что разработанный метод обеспечивает наилучшее качество обнаружения объектов среди рассмотренных методов защиты. В частности, на наборе изображений INRIA-Person при отсутствии атаки значение метрики mAP составляет 93,5%, а при наличии состязательной патч-атаки – 81,0%, что превышает результаты методов SAC и PAD.

Визуализация результата работы модели обнаружения объектов на нескольких примерах из набора INRIA-Person, подвергшихся состязательной патч-атаке, представлены на рисунке 23. Зеленые рамки представляют обнаруженные объекты категории «человек», при этом в верхнем левом углу каждой рамки отображается оценка объектности. В левой колонке показаны результаты обнаружения без использования Разработан метод локализации и удаления состязательных патч-атак, в то время как в правой колонке показан результат работы модели обнаружения объектов с использованием разработанного метода.



Рисунок 23 –Визуализация результата работы модели обнаружения объектов на нескольких примерах из набора изображений INRIA-Person, подвергшихся состязательной патч-атаке

В результате исследования разработан метод снижения влияния состязательных патч-атак на систему нейросетевого обнаружения объектов. Предложенный метод включает глубокую сверточную нейронную сеть, алгоритм локализации аномальных областей и алгоритм их кластеризации и последующей обработки.

Использование глубокой сверточной нейронной сети позволяет формировать реконструированное изображение с ослабленным влиянием состязательного патча без необходимости предварительного знания его структуры. Последующая локализация аномалий основана на анализе статистических характеристик локальных гистограмм карты ошибок и входного изображения, что обеспечивает выделение областей, содержащих состязательные возмущения, и их последующую обработку.

В отличие от большинства существующих подходов, предложенный метод не требует повторного обучения модели обнаружения объектов при появлении новых типов состязательных патч-атак и не зависит от конкретного метода их формирования. Совместное использование легкой глубокой сверточной нейронной сети и классических методов обработки данных обеспечивает высокую эффективность защиты при сравнительно низких вычислительных затратах.

Таким образом, разработанный метод обеспечивает эффективное снижение влияния состязательных патч-атак на качество обнаружения объектов, сохраняя высокую точность работы нейросетевой модели без модификации архитектуры модели обнаружения объектов и без повторного обучения при появлении новых типов атак. Полученные результаты подтверждают перспективность применения предложенного подхода для повышения устойчивости систем видового мониторинга к искусственным искажениям.

Выводы по разделу 4

В результате исследований, направленных на разработку метода локализации и удаления состязательных патч-атак на изображениях при видовом мониторинге, получены следующие результаты.

1. Разработан метод защиты нейросетевой системы обнаружения объектов от состязательных патч-атак, включающий глубокую сверточную нейронную сеть для реконструкции входного изображения, алгоритм локализации аномальных областей на основе анализа локальных статистических характеристик карты ошибок и входного изображения, а также алгоритм кластеризации и последующей обработки обнаруженных аномалий. Предложенный подход позволяет локализовать области состязательных возмущений и уменьшить их влияние на качество обнаружения объектов.

2. Показано, что использование реконструированного изображения совместно с алгоритмом локализации аномалий позволяет эффективно снижать влияние состязательных патч-атак без предварительного знания их структуры и без модификации архитектуры нейросетевой модели обнаружения объектов. В отличие от большинства существующих методов защиты, предложенный метод не требует повторного обучения модели при появлении новых типов состязательных патч-атак и не зависит от метода их формирования.

3. Экспериментальная оценка на наборе изображений INRIA-Person показала, что применение разработанного метода совместно с моделью обнаружения объектов YOLOv3 повышает значение метрики mAP в условиях состязательной патч-атаки с 46,8% до 81,0%, при этом на изображениях без атак значение mAP сохраняется на уровне 93,5%.

4. Таким образом, разработанный метод обеспечивает локализацию и удаление состязательных патч-атак без модификации архитектуры модели обнаружения объектов и без повторного обучения при появлении новых типов состязательных воздействий. Полученные экспериментальные результаты

подтверждают его эффективность и позволяют сделать вывод о подтверждении третьего положения, выносимого на защиту.

ЗАКЛЮЧЕНИЕ

В диссертационной работе решена актуальная научно-техническая задача повышения устойчивости и точности нейросетевого обнаружения искаженных объектов при видовом мониторинге за счёт разработки способа формирования обучающего набора изображений, адаптивной предварительной обработки изображений и защиты от состязательных патч-атак.

В ходе выполнения диссертационной работы получены следующие основные результаты.

1. Проведен анализ современных методов нейросетевого обнаружения объектов, существующих метрик оценки качества и робастности, а также способов повышения устойчивости моделей компьютерного зрения к естественным и искусственным искажениям. Установлено, что существующие методы не обеспечивают достаточной устойчивости моделей обнаружения объектов к совокупному воздействию различных видов естественных искажений и состязательных патч-атак без существенного увеличения вычислительных затрат или модификации архитектуры нейронной сети, что определило направление дальнейших исследований.

2. Разработан способ формирования обучающего набора изображений для обучения модели обнаружения объектов, суть которого заключается в использовании изображений, содержащих естественные искажения различной интенсивности. Экспериментально установлено, что повышение устойчивости определяется не только наличием искажённых изображений в обучающем наборе изображений, но и законом распределения уровней искажений. Показано, что использование полунормального распределения уровней искажений позволяет увеличить полноту обнаружения более чем в 2 раза при воздействии гауссова размытия при радиусе размытия в диапазоне от 6 до 12 пикселей, а также повышение качества распознавания более чем в 2 раза в области низких ОСШ (от -8 до 0 дБ), подтверждают эффективность разработанного метода.

3. Разработан метод адаптивной предварительной обработки изображений, основанный на статистической оценке неоднородности входного изображения и автоматическом выборе режима фильтрации. Предложенный метод объединяет разработанный способ формирования обучающего набора изображений и адаптивную предварительную обработку, что обеспечивает повышение интегральной оценки помехоустойчивости системы обнаружения объектов в 1,4 раза одновременно к гауссову размытию и аддитивному белому гауссову шуму по сравнению с исходной моделью без изменения архитектуры нейросетевой модели обнаружения объектов.

4. Разработан метод локализации и удаления состязательных патч-атак, основанный на реконструкции входного изображения и анализе локальных статистических характеристик карты ошибок реконструкции. Предложенный метод не требует модификации архитектуры модели обнаружения объектов и повторного обучения при появлении новых типов состязательных воздействий. Экспериментально показано, что применение разработанного метода увеличивает значение метрики mAP в условиях состязательной патч-атаки с 46,8% до 81,0%, при этом на изображениях без атак значение mAP сохраняется на уровне 93,5%, что превосходит результаты существующих методов защиты.

Полученные результаты доказывают обоснованность положений, выносимых на защиту, а также подтверждают, что цель диссертационной работы достигнута, а поставленная научная задача решена.

Разработанные методы могут быть использованы при создании интеллектуальных систем видового мониторинга, функционирующих в условиях естественных и искусственных искажений изображений, включая системы дистанционного зондирования Земли, автономные транспортные средства и другие системы компьютерного зрения, предъявляющие повышенные требования к надёжности обнаружения объектов.

Перспективы дальнейших исследований связаны с разработкой универсальных методов повышения робастности нейросетевых моделей к более широкому спектру естественных и искусственных искажений, исследованием

совместного воздействия нескольких типов деградации изображений при решении задач компьютерного зрения.

СПИСОК СОКРАЩЕНИЙ И УСЛОВНЫХ ОБОЗНАЧЕНИЙ

АБГШ	– аддитивный белый гауссов шум;
ГСНС	– глубокая свёрточная нейронная сеть;
ДЗЗ	– дистанционное зондирование Земли;
ИНС	– искусственная нейронная сеть;
НИОКР	– научно-исследовательская и опытно конструкторская работа;
НИР	– научно-исследовательская работа;
ОСШ	– отношение сигнал/шум;
СЧ	– составная часть;
ЭВМ	– электронная вычислительная машина;
AP	– Average Precision – усредненная точность;
AR	– Average Recall – усредненная полнота;
DBSCAN	– Density-based spatial clustering of applications with noise;
IoU	– Intersection Over Union;
mAP	– Mean Average Precision;
PAD	– Patch-Agnostic Defense
R-CNN	– Region-Based Convolutional Neural Network;
RGB	– Red, Green, Blue;
SAC	– Segment and Complete;
YOLO	– You Only Look Once.

СЛОВАРЬ ТЕРМИНОВ

Автокодировщик (Autoencoder): разновидность искусственной нейронной сети, предназначенная для обучения эффективному скрытому представлению входных данных путем их кодирования в пространство меньшей размерности и последующего восстановления.

Аугментация изображений: метод увеличения разнообразия обучающего набора изображений путём искусственного преобразования исходных изображений без изменения их семантического содержания.

Видовой мониторинг: наблюдение за объектами земной поверхности по данным дистанционного зондирования Земли или видеонаблюдения с целью обнаружения, распознавания и анализа объектов.

Робастность модели: способность нейросетевой модели сохранять высокое качество обнаружения объектов при наличии искажений входных изображений или изменении условий эксплуатации.

Состязательная атака (Adversarial Attack): преднамеренное изменение входных изображений, вызывающее ошибочную работу нейросетевой модели.

Состязательный патч (Adversarial Patch): локальная область изображения, специально оптимизированная для нарушения работы модели обнаружения объектов независимо от содержания остальной части изображения.

СПИСОК ЛИТЕРАТУРЫ

1. Cheng, B. A Review of Remote Sensing Image Object Detection: Advances and Challenges in Data, Techniques, and Applications / B. Cheng, Q. Deng, W. Gan, Y. Li, C. Wu, Q. Wang. // *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*. – 2026. – Т. 19. – С. 10438–10472.
2. Cao, Z. Real-time object detection based on UAV remote sensing: A systematic literature review / Z. Cao, L. Kooistra, W. Wang, L. Guo, J. Valente. // *Drones*. – 2023. – Т. 7, № 10. – С. 620.
3. Hua, W. A survey of small object detection based on deep learning in aerial images / W. Hua, Q. Chen. // *Artificial Intelligence Review*. – 2025. – Т. 58, № 6. – С. 162.
4. Chen, Z. Object detection for autonomous vehicles under adverse weather conditions / Z. Chen, Z. Zhang, Q. Su, K. Yang, Y. Wu, L. He, X. Tang. // *Expert Systems with Applications*. – 2026. – Т. 296. – С. 128994.
5. Ziyadinov, V. Noise immunity and robustness study of image recognition using a convolutional neural network / V. Ziyadinov, M. Tereshonok. // *Sensors*. – 2022. – Т. 22, № 3. – С. 1241.
6. Du, A. Physical adversarial attacks on an aerial imagery object detector / A. Du, B. Chen, T. J. Chin, Y. W. Law, M. Sasdelli, R. Rajasegaran, D. Campbell. // *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)*. – Waikoloa, HI, USA, 2022. – С. 1796–1806.
7. Zhang, Y. Adversarial patch attack on multi-scale object detection for UAV remote sensing images / Y. Zhang, Y. Zhang, J. Qi, K. Bin, H. Wen, X. Tong, P. Zhong. // *Remote Sensing*. – 2022. – Т. 14, № 20. – С. 5298.
8. Thys, S. Fooling automated surveillance cameras: Adversarial patches to attack person detection / S. Thys, W. Van Ranst, T. Goedemé. // *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*. – Long Beach, CA, USA, 2019. – С. 49–55.

9. Isaev, I. Neural network solution of an inverse problem with integration of geophysical methods on recovered data: Training with noise addition / I. Isaev, I. Osbornev, E. Osbornev, E. Rodionov, M. Shimelevich, S. Dolenko. // *Biologically Inspired Cognitive Architectures Meeting*. – Cham : Springer Nature Switzerland, 2023. – С. 406–413.
10. Isaev, I. V. Training with noise as a method to increase noise resilience of neural network solution of inverse problems / I. V. Isaev, S. A. Dolenko. // *Optical Memory and Neural Networks*. – 2016. – Т. 25, № 3. – С. 142–148.
11. Isaev, I. Training with noise addition in neural network solution of inverse problems: Procedures for selection of the optimal network / I. Isaev, S. Dolenko. // *Procedia Computer Science*. – 2018. – Т. 123. – С. 171–176.
12. Власов, С. О. Решение задачи обнаружения объекта с помощью нейросетевых технологий / С. О. Власов, А. И. Гладышев, А. А. Богуславский, С. М. Соколов. // *Препринты Института прикладной математики им. М. В. Келдыша РАН*. – 2023. – № 0. – С. 16–27.
13. Гладышев, А. И. Применение Big Data для анализа околоземного космического пространства / А. И. Гладышев, А. И. Зимовец, А. А. Нечай, А. В. Обухов. // *Вестник Российского нового университета. Серия: Сложные системы: модели, анализ и управление*. – 2020. – Т. 4, № 4. – С. 127–134.
14. 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.
15. Schegolev, A. Learning cell for superconducting neural networks / A. Schegolev, N. Klenov, I. Soloviev, M. Tereshonok. // *Superconductor Science and Technology*. – 2021. – Т. 34, № 1. – С. 015006.
16. Klenov, N. V. Energy efficient superconducting neural networks for high-speed intellectual data processing systems / N. V. Klenov, A. E. Schegolev, I. I. Soloviev, S. V. Bakurskiy, M. V. Tereshonok. // *IEEE Transactions on Applied Superconductivity*. – 2018. – Т. 28, № 7. – С. 1–6.

17. Adjemov, S. S. A neural-network method for the synthesis of informative features for the classification of signal sources in cognitive radio systems / S. S. Adjemov, N. V. Klenov, M. V. Tereshonok, D. S. Chirov. // Moscow University Physics Bulletin. – 2016. – Т. 71, № 2. – С. 174–179.
18. Ziyadinov, V. V. Neural network image recognition robustness with different augmentation methods / V. V. Ziyadinov, M. V. Tereshonok. // В сборнике: 2022 Systems of Signal Synchronization, Generating and Processing in Telecommunications (SYNCHROINFO). – Архангельск, Россия, 2022. – С. 1–4.
19. Ziyadinov, V. Low-pass image filtering to achieve adversarial robustness / V. Ziyadinov, M. Tereshonok. // Sensors. – 2023. – Т. 23, № 22. – С. 9032.
20. Pautov, M. Cc-cert: A probabilistic approach to certify general robustness of neural networks / M. Pautov, N. Tursynbek, M. Munkhoeva, N. Muravev, A. Petiushko, I. Oseledets. // Proceedings of the AAAI Conference on Artificial Intelligence. – 2022. – Т. 36, № 7. – С. 7975–7983.
21. Pautov, M. Smoothed embeddings for certified few-shot learning / M. Pautov, O. Kuznetsova, N. Tursynbek, A. Petiushko, I. Oseledets. // В сборнике: Advances in Neural Information Processing Systems 35. – New Orleans, Louisiana, USA, 2022. – С. 24367–24379.
22. Korzh, D. General Lipschitz: Certified robustness against resolvable semantic transformations via transformation-dependent randomized smoothing / D. Korzh, M. Pautov, O. Tsymboi, I. Oseledets. // ECAI 2024: 27th European Conference on Artificial Intelligence, 19–24 October 2024, Santiago de Compostela, Spain, including 13th Conference on Prestigious Applications of Intelligent Systems (PAIS 2024). – London : SAGE Publications Pvt. Ltd, 2024. – С. 1591–1598.
23. Goodfellow, I. J. Explaining and harnessing adversarial examples / I. J. Goodfellow, J. Shlens, C. Szegedy. // В сборнике: 3rd International Conference on Learning Representations, ICLR 2015, Conference Track Proceedings. – San Diego, CA, USA, 2015. – С. 1–11.

24. Goodfellow, I. J. Generative adversarial nets / I. J. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, Y. Bengio. // *Advances in Neural Information Processing Systems*. – 2014. – Т. 27. – С. 1–9.
25. Szegedy, C. Intriguing properties of neural networks / C. Szegedy, W. Zaremba, I. Sutskever, J. Bruna, D. Erhan, I. J. Goodfellow, R. Fergus. // В сборнике: 2nd International Conference on Learning Representations (ICLR 2014), Conference Track Proceedings. – Banff, Alberta, Canada, 2014. – С. 1–10.
26. Hendrycks, D. Benchmarking neural network robustness to common corruptions and perturbations / D. Hendrycks, T. Dietterich. // В сборнике: 7th International Conference on Learning Representations (ICLR 2019). – New Orleans, Louisiana, USA, 2019. – С. 1–17.
27. Hendrycks, D. Natural adversarial examples / D. Hendrycks, K. Zhao, S. Basart, J. Steinhardt, D. Song. // В сборнике: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). – 2021. – С. 15262–15271.
28. Ren, S. Faster R-CNN: Towards real-time object detection with region proposal networks / S. Ren, K. He, R. Girshick, J. Sun. // *IEEE Transactions on Pattern Analysis and Machine Intelligence*. – 2016. – Т. 39, № 6. – С. 1137–1149.
29. Redmon, J. YOLO9000: Better, faster, stronger / J. Redmon, A. Farhadi. // В сборнике: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR). – 2017. – С. 7263–7271.
30. Redmon, J. YOLO9000: Better, faster, stronger / J. Redmon, A. Farhadi. // 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). – IEEE, 2017. – С. 6517–6525.
31. Redmon, J. YOLOv3: An incremental improvement [Электронный ресурс] / J. Redmon, A. Farhadi. // arXiv preprint. – 2018. – Режим доступа: <https://arxiv.org/pdf/1804.02767> (дата обращения: 13.01.2025).
32. Wang, M. Deep visual domain adaptation: A survey / M. Wang, W. Deng. // *Neurocomputing*. – 2018. – Т. 312. – С. 135–153.
33. Shorten, C. A survey on image data augmentation for deep learning / C. Shorten, T. M. Khoshgoftaar. // *Journal of Big Data*. – 2019. – Т. 6, № 1. – С. 60.

34. Zhang, K. Plug-and-play image restoration with deep denoiser prior / K. Zhang, Y. Li, W. Zuo, L. Zhang, L. Van Gool, R. Timofte. // *IEEE Transactions on Pattern Analysis and Machine Intelligence*. – 2021. – Т. 44, № 10. – С. 6360–6376.
35. Arsenos, A. Common Corruptions for Evaluating and Enhancing Robustness in Air-to-Air Visual Object Detection / A. Arsenos, V. Karampinis, E. Petrongonas, C. Skliros, D. Kollias, S. D. Kollias, A. Voulodimos. // *IEEE Robotics and Automation Letters*. – 2024. – Т. 9, № 7. – С. 6688–6695.
36. Madry, A. Towards deep learning models resistant to adversarial attacks / A. Madry, A. Makelov, L. Schmidt, D. Tsipras, A. Vladu. // В сборнике: *The Sixth International Conference on Learning Representations (ICLR 2018)*. – Vancouver, Canada, 2018. – С. 1–28.
37. Liang, H. Adversarial attack and defense: A survey / H. Liang, E. He, Y. Zhao, Z. Jia, H. Li. // *Electronics*. – 2022. – Т. 11, № 8. – С. 1283.
38. Ilina, O. A survey on symmetrical neural network architectures and applications / O. Ilina, V. Ziyadinov, N. Klenov, M. Tereshonok. // *Symmetry*. – 2022. – Т. 14, № 7. – С. 1391.
39. Ilina, O. V. Robustness study of a deep convolutional neural network for vehicle detection in aerial imagery / O. V. Ilina, M. V. Tereshonok. // *Journal of Communications Technology and Electronics*. – 2022. – Т. 67, № 2. – С. 164–170.
40. Ilina, O. Increasing neural-based pedestrian detectors' robustness to adversarial patch attacks using anomaly localization / O. Ilina, M. Tereshonok, V. Ziyadinov. // *Journal of Imaging*. – 2025. – Т. 11, № 1. – С. 26.
41. Ильина, О. В. Способ локализации состязательных патч-атак / О. В. Ильина. // *Вестник связи*. – 2026. – № 9. – С. 26–33.
42. Демин, А. Д. Исследование влияния адаптации диаграмм направленности антенн на эффективность сетевого уровня AD НОС сети радиосвязи / А. Д. Демин, М. В. Терешонок, О. В. Ильина. // *T-Comm: Телекоммуникации и транспорт*. – 2025. – Т. 19, № 4. – С. 4–11.
43. Ilina, O. V. Study of deep convolutional neural network for vehicle localization on blurred aerial imagery / O. V. Ilina, M. V. Tereshonok. // В сборнике:

2022 Systems of Signal Synchronization, Generating and Processing in Telecommunications (SYNCHROINFO). – Архангельск, Россия, 2022. – С. 1–7.

44. Ilina, O. V. Adaptive method to increase vehicle detection system robustness on blurred and noisy aerial images / O. V. Ilina, M. V. Tereshonok. // В сборнике: 2023 Intelligent Technologies and Electronic Devices in Vehicle and Road Transport Complex (TIRVED). – Москва, Россия, 2023. – С. 1–6.

45. Свидетельство о государственной регистрации программы для ЭВМ № 2024689957 Российская Федерация. Программный модуль для локализации и удаления состязательных патч-атак на цветных изображениях для решения задач компьютерного зрения / О. В. Ильина, М. В. Терешонок; заявитель и правообладатель МТУСИ. – № 2024689563; заявл. 28.11.2024; опубл. 11.12.2024.

46. Свидетельство о государственной регистрации программы для ЭВМ № 2024612396 Российская Федерация. Программный комплекс для демонстрации работы свёрточной нейронной сети, решающей задачу распознавания состязательных изображений / О. В. Ильина, М. В. Терешонок, В. В. Зиядинов; заявитель и правообладатель МТУСИ. – № 2023689454; заявл. 27.12.2023; опубл. 31.01.2024.

47. Свидетельство о государственной регистрации программы для ЭВМ № 2024610705 Российская Федерация. Программа для моделирования работы нейросетевой модели обнаружения объектов с использованием адаптивной системы для повышения устойчивости на размытых и зашумленных цветных изображениях / О. В. Ильина, М. В. Терешонок; заявитель и правообладатель МТУСИ. – № 2023689457; заявл. 27.12.2023; опубл. 12.01.2024.

48. 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 / В. В. Зиядинов, О. В. Ильина, М. В. Терешонок; заявитель и правообладатель МТУСИ. – № 2023689369; заявл. 27.12.2023; опубл. 19.01.2024.

49. Свидетельство о государственной регистрации программы для ЭВМ № 2022660463 Российская Федерация. Программа сравнительного анализа и визуализации результатов работы свёрточных нейронных сетей / В. В. Зиядинов, О. В. Ильина, М. В. Терешонок; заявитель и правообладатель МТУСИ. – № 2022619579; заявл. 23.05.2022; опубл. 03.06.2022.

50. Свидетельство о государственной регистрации программы для ЭВМ № 2022660552 Российская Федерация. Программа моделирования шума в реальных изображениях и генерации обучающих выборок для систем распознавания / В. В. Зиядинов, О. В. Ильина; заявитель и правообладатель МТУСИ. – № 2022619577; заявл. 23.05.2022; опубл. 06.06.2022.

51. 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. – Т. 66, № 12. – С. 1363–1369.

52. Sapkota, R. YOLO advances to its genesis: A decadal and comprehensive review of the You Only Look Once (YOLO) series / R. Sapkota, M. Flores-Calero, R. Qureshi, C. Badgujar, U. Nepal, A. Poullose, P. Zeno, U. B. P. Vaddevolu, S. Khan, M. Shoman, H. Yan, M. Karkee. // Artificial Intelligence Review. – 2025. – Т. 58, № 9. – С. 274.

53. Padilla, R. A comparative analysis of object detection metrics with a companion open-source toolkit / R. Padilla, W. L. Passos, T. L. Dias, S. L. Netto, E. A. Da Silva. // Electronics. – 2021. – Т. 10, № 3. – С. 279.

54. Geirhos, R. ImageNet-trained CNNs are biased towards texture; increasing shape bias improves accuracy and robustness / R. Geirhos, P. Rubisch, C. Michaelis, M. Bethge, F. A. Wichmann, W. Brendel // В сборнике: 7th International Conference on Learning Representations (ICLR 2019). – New Orleans, Louisiana, USA, 2019. – С. 1–13.

55. Taori, R. Measuring robustness to natural distribution shifts in image classification / R. Taori, A. Dave, V. Shankar, N. Carlini, B. Recht, L. Schmidt. // В сборнике: Advances in Neural Information Processing Systems 34 (NIPS 2020):

Proceedings of the 34th International Conference on Neural Information Processing Systems. – Vancouver, BC, Canada, 2020. – C. 18583–18599.

56. Dodge, S. Understanding how image quality affects deep neural networks / S. Dodge, L. Karam. // 2016 Eighth International Conference on Quality of Multimedia Experience (QoMEX). – IEEE, 2016. – C. 1–6.

57. Vasiljevic, I. Examining the impact of blur on recognition by convolutional networks [Электронный ресурс] / I. Vasiljevic, A. Chakrabarti, G. Shakhnarovich. // arXiv preprint. – 2016. – Режим доступа: <https://arxiv.org/pdf/1611.05760> (дата обращения: 20.05.2022).

58. Li, L. Learning a discriminative prior for blind image deblurring / L. Li, J. Pan, W. S. Lai, C. Gao, N. Sang, M. H. Yang. // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. – 2018. – C. 6616–6625.

59. Chakrabarti, A. A neural approach to blind motion deblurring / A. Chakrabarti. // European Conference on Computer Vision. – Cham : Springer International Publishing, 2016. – C. 221–235.

60. Kupyn, O. DeblurGAN: Blind motion deblurring using conditional adversarial networks / O. Kupyn, V. Budzan, M. Mykhailych, D. Mishkin, J. Matas. // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. – 2018. – C. 8183–8192.

61. Sun, J. Learning a convolutional neural network for non-uniform motion blur removal / J. Sun, W. Cao, Z. Xu, J. Ponce. // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. – 2015. – C. 769–777.

62. Zhang, K. Deep plug-and-play super-resolution for arbitrary blur kernels / K. Zhang, W. Zuo, L. Zhang. // Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. – 2019. – C. 1671–1681.

63. Ding, C. Trunk-branch ensemble convolutional neural networks for video-based face recognition / C. Ding, D. Tao. // IEEE Transactions on Pattern Analysis and Machine Intelligence. – 2017. – Т. 40, № 4. – C. 1002–1014.

64. Liu, D. Enhance visual recognition under adverse conditions via deep networks / D. Liu, B. Cheng, Z. Wang, H. Zhang, T. S. Huang. // IEEE Transactions on Image Processing. – 2019. – T. 28, № 9. – C. 4401–4412.
65. Gazioğlu, B. S. A. Effects of objects and image quality on melanoma classification using deep neural networks / B. S. A. Gazioğlu, M. E. Kamaşak. // Biomedical Signal Processing and Control. – 2021. – T. 67. – C. 102530.
66. Zhou, Y. On classification of distorted images with deep convolutional neural networks / Y. Zhou, S. Song, N. M. Cheung. // 2017 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). – IEEE, 2017. – C. 1213–1217.
67. Dodge, S. A study and comparison of human and deep learning recognition performance under visual distortions / S. Dodge, L. Karam. // 2017 26th International Conference on Computer Communication and Networks (ICCCN). – IEEE, 2017. – C. 1–7.
68. Pei, Y. Effects of image degradation and degradation removal to CNN-based image classification / Y. Pei, Y. Huang, Q. Zou, X. Zhang, S. Wang. // IEEE Transactions on Pattern Analysis and Machine Intelligence. – 2019. – T. 43, № 4. – C. 1239–1253.
69. Wang, Y. Di-AA: An interpretable white-box attack for fooling deep neural networks / Y. Wang, J. Liu, X. Chang, R. J. Rodríguez, J. Wang. // Information Sciences. – 2022. – T. 610. – C. 14–32.
70. Wei, H. Physical Adversarial Attack Meets Computer Vision: A Decade Survey / H. Wei, H. Tang, X. Jia, Z. Wang, H. Yu, Z. Li, S. Satoh, L. Van Gool, Z. Wang. // IEEE Transactions on Pattern Analysis and Machine Intelligence. – 2024. – T. 46, № 12. – C. 9797–9817.
71. Wei, X. Adversarial sticker: A stealthy attack method in the physical world / X. Wei, Y. Guo, J. Yu. // IEEE Transactions on Pattern Analysis and Machine Intelligence. – 2023. – T. 45, № 3. – C. 2711–2725.
72. Zhao, W. Adversarial training methods for deep learning: A systematic review / W. Zhao, S. Alwidian, Q. H. Mahmoud. // Algorithms. – 2022. – T. 15, № 8. – C. 283.

73. Rao, S. Adversarial training against location-optimized adversarial patches / S. Rao, D. Stutz, B. Schiele. // Computer Vision – ECCV 2020: 16th European Conference, Glasgow, UK, 23–28 August 2020 : Proceedings. – Cham : Springer, 2020. – С. 429–448.
74. Ji, N. Adversarial YOLO: Defense human detection patch attacks via detecting adversarial patches [Электронный ресурс] / N. Ji, Y. Feng, H. Xie, X. Xiang, N. Liu. // arXiv preprint. – 2021. – Режим доступа: <https://arxiv.org/pdf/2103.08860> (дата обращения: 13.01.2025).
75. Xu, K. PatchZero: Defending against adversarial patch attacks by detecting and zeroing the patch / K. Xu, Y. Xiao, Z. Zheng, K. Cai, R. Nevatia. // Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision (WACV). – Waikoloa, HI, USA, 2023. – С. 4632–4641.
76. Liu, J. Segment and complete: Defending object detectors against adversarial patch attacks with robust patch detection / J. Liu, A. Levine, C. P. Lau, R. Chellappa, S. Feizi. // Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). – New Orleans, LA, USA, 2022. – С. 14973–14982.
77. Ronneberger, O. U-Net: Convolutional networks for biomedical image segmentation / O. Ronneberger, P. Fischer, T. Brox. // Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015: 18th International Conference, Munich, Germany, 5–9 October 2015 : Proceedings. – Cham : Springer, 2015. – С. 234–241.
78. Chiang, P. H. Adversarial pixel masking: A defense against physical attacks for pre-trained object detectors / P. H. Chiang, C. S. Chan, S. H. Wu. // Proceedings of the 29th ACM International Conference on Multimedia. – Virtual Event, China, 2021. – С. 1856–1865.
79. Wei, X. Real-world adversarial defense against patch attacks based on diffusion model / X. Wei, C. Kang, Y. Dong, Z. Wang, S. Ruan, Y. Chen, H. Su. // IEEE Transactions on Pattern Analysis and Machine Intelligence. – 2025. – Т. 47, № 12. – С. 11124–11140.

80. Mao, Z. Enhancing robustness of person detection: A universal defense filter against adversarial patch attacks / Z. Mao, S. Chen, Z. Miao, H. Li, B. Xia, J. Cai, W. Yuan, X. You. // *Computers & Security*. – 2024. – T. 146. – C. 104066.
81. Jing, L. PAD: Patch-Agnostic Defense against Adversarial Patch Attacks / L. Jing, R. Wang, W. Ren, X. Dong, C. Zou. // *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. – Seattle, WA, USA, 2024. – C. 24472–24481.
82. Kirillov, A. Segment Anything / A. Kirillov, E. Mintun, N. Ravi, H. Mao, C. Rolland, L. Gustafson, T. Xiao, S. Whitehead, A. C. Berg, W. Y. Lo, P. Dollár, R. Girshick. // *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*. – 2023. – C. 4015–4026.
83. Naseer, M. Local gradients smoothing: Defense against localized adversarial attacks / M. Naseer, S. Khan, F. Porikli. // *Proceedings of the 2019 IEEE Winter Conference on Applications of Computer Vision (WACV)*. – Waikoloa, HI, USA : IEEE, 2019. – C. 1300–1307.
84. Xiang, C. DetectorGuard: Provably securing object detectors against localized patch hiding attacks / C. Xiang, P. Mittal. // *Proceedings of the 2021 ACM SIGSAC Conference on Computer and Communications Security (CCS)*. – Virtual Event, Republic of Korea, 2021. – C. 3177–3196.
85. Xiang, C. ObjectSeeker: Certifiably robust object detection against patch hiding attacks via patch-agnostic masking / C. Xiang, A. Valtchanov, S. Mahloujifar, P. Mittal. // *Proceedings of the 2023 IEEE Symposium on Security and Privacy (SP)*. – San Francisco, CA, USA : IEEE, 2023. – C. 1329–1347.
86. Durrett, R. *Probability: Theory and examples*. – Cambridge : Cambridge University Press, 2019. – 430 c.
87. Thong, J. T. L. Single-image signal-to-noise ratio estimation / J. T. L. Thong, K. S. Sim, J. C. H. Phang. // *Scanning*. – 2001. – T. 23, № 5. – C. 328–336.
88. Sim, K. S. Image signal-to-noise ratio estimation using the autoregressive model / K. S. Sim, N. S. Kamel. // *Scanning*. – 2004. – T. 26, № 3. – C. 135–139.

89. Xia, G. S. DOTA: A large-scale dataset for object detection in aerial images / G. S. Xia, X. Bai, J. Ding, Z. Zhu, S. Belongie, J. Luo, M. Datcu, M. Pelillo, L. Zhang. // 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. – IEEE, 2018. – C. 3974–3983.
90. Schlegl, T. Unsupervised anomaly detection with generative adversarial networks to guide marker discovery / T. Schlegl, P. Seeböck, S. M. Waldstein, U. Schmidt-Erfurth, G. Langs. // Information Processing in Medical Imaging: 26th International Conference, IPMI 2017, Boone, NC, USA, 25–30 June 2017 : Proceedings. – Cham : Springer, 2017. – C. 146–157.
91. Li, P. A comprehensive survey on design and application of autoencoder in deep learning / P. Li, Y. Pei, J. Li. // Applied Soft Computing. – 2023. – T. 138. – C. 110176.
92. He, K. Deep residual learning for image recognition / K. He, X. Zhang, S. Ren, J. Sun. // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR). – Las Vegas, NV, USA, 2016. – C. 770–778.
93. Liu, F. T. Isolation forest / F. T. Liu, K. M. Ting, Z. H. Zhou. // Proceedings of the 2008 Eighth IEEE International Conference on Data Mining (ICDM). – Pisa, Italy : IEEE, 2008. – C. 413–422.
94. Ester, M. A density-based algorithm for discovering clusters in large spatial databases with noise / M. Ester, H. P. Kriegel, J. Sander, X. Xu. // Proceedings of the Second International Conference on Knowledge Discovery and Data Mining (KDD). – Portland, OR, USA, 1996. – C. 226–231.
95. Cohen, I. Pearson correlation coefficient / I. Cohen, Y. Huang, J. Chen, J. Benesty. // Noise Reduction in Speech Processing. – Berlin ; Heidelberg : Springer, 2009. – C. 1–4.
96. Pedregosa, F. Scikit-learn: Machine learning in Python / F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, M. Duchesnay. // Journal of Machine Learning Research. – 2011. – T. 12. – C. 2825–2830.

97. Dalal, N. Histograms of oriented gradients for human detection / N. Dalal, B. Triggs. // В сборнике: 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05). – San Diego, CA, USA, 2005. – Т. 1. – С. 886–893.

ПРИЛОЖЕНИЕ А

РОССИЙСКАЯ ФЕДЕРАЦИЯ



СВИДЕТЕЛЬСТВО

о государственной регистрации программы для ЭВМ

№ 2022660463

**«Программа сравнительного анализа и визуализации
результатов работы свёрточных нейронных сетей»**

Правообладатель: **Ордена Трудового Красного Знамени
Федеральное государственное бюджетное
образовательное учреждение высшего образования
«Московский технический университет связи и
информатики» (МТУСИ) (RU)**

Авторы: **Зиядинов Вадим Валерьевич (RU), Ильина Ольга
Викторовна (RU), Терешонок Максим Валерьевич (RU)**

Заявка № 2022619579

Дата поступления 23 мая 2022 г.

Дата государственной регистрации

в Реестре программ для ЭВМ 03 июня 2022 г.



Руководитель Федеральной службы
по интеллектуальной собственности

Ю.С. Зубов

РОССИЙСКАЯ ФЕДЕРАЦИЯ



СВИДЕТЕЛЬСТВО

о государственной регистрации программы для ЭВМ

№ 2022660552

**«Программа моделирования шума в реальных
изображениях и генерации обучающих выборок для
систем распознавания»**

Правообладатель: *Ордена Трудового Красного Знамени
Федеральное государственное бюджетное образовательное
учреждение высшего образования «Московский технический
университет связи и информатики» (МТУСИ) (RU)*

Авторы: *Зиядинов Вадим Валерьевич (RU), Ильина Ольга
Викторовна (RU)*



Заявка № 2022619577

Дата поступления 23 мая 2022 г.

Дата государственной регистрации

в Реестре программ для ЭВМ 06 июня 2022 г.

Руководитель Федеральной службы
по интеллектуальной собственности

Ю.С. Зубов

РОССИЙСКАЯ ФЕДЕРАЦИЯ



СВИДЕТЕЛЬСТВО

о государственной регистрации программы для ЭВМ

№ 2024610705

Программа для моделирования работы нейросетевой модели обнаружения объектов с использованием адаптивной системы для повышения устойчивости на размытых и зашумленных цветных изображениях

Правообладатель: *Ордена Трудового Красного Знамени федеральное государственное бюджетное образовательное учреждение высшего образования «Московский технический университет связи и информатики» (RU)*

Авторы: *Ильина Ольга Викторовна (RU), Терешонок Максим Валерьевич (RU)*



Заявка № 2023689457

Дата поступления 27 декабря 2023 г.

Дата государственной регистрации

в Реестре программ для ЭВМ 12 января 2024 г.

Руководитель Федеральной службы
по интеллектуальной собственности

ДОКУМЕНТ ПОДПИСАН ЭЛЕКТРОННОЙ ПОДПИСЬЮ
Сертификат 429b6a0fe3853164ba96f83b73b4aa7
Владелец **Зубов Юрий Сергеевич**
Действителен с 10.05.2023 по 02.08.2024

Ю.С. Зубов

РОССИЙСКАЯ ФЕДЕРАЦИЯ



СВИДЕТЕЛЬСТВО

о государственной регистрации программы для ЭВМ

№ 2024611339

«Программный комплекс расчета внешних характеристик точности и устойчивости свёрточной нейронной сети к высокочастотным искажениям и оптимизации параметров предварительной обработки изображений»

Правообладатель: *Ордена Трудового Красного Знамени федеральное государственное бюджетное образовательное учреждение высшего образования «Московский технический университет связи и информатики» (RU)*

Авторы: *Зиядинов Вадим Валерьевич (RU), Ильина Ольга Викторовна (RU), Терешонок Максим Валерьевич (RU)*



Заявка № 2023689369

Дата поступления 27 декабря 2023 г.

Дата государственной регистрации

в Реестре программ для ЭВМ 19 января 2024 г.

Руководитель Федеральной службы
по интеллектуальной собственности

ДОКУМЕНТ ПОДПИСАН ЭЛЕКТРОННОЙ ПОДПИСЬЮ
Сертификат 429b6a0fe3853164ba96f83b73b4aa7
Владелец **Зубов Юрий Сергеевич**
Действителен с 10.05.2023 по 02.08.2024

Ю.С. Зубов

РОССИЙСКАЯ ФЕДЕРАЦИЯ



СВИДЕТЕЛЬСТВО

о государственной регистрации программы для ЭВМ

№ 2024612396

**«Программный комплекс для демонстрации работы
свёрточной нейронной сети, решающей задачу
распознавания состязательных изображений»**

Правообладатель: *Ордена Трудового Красного Знамени
федеральное государственное бюджетное образовательное
учреждение высшего образования «Московский технический
университет связи и информатики» (RU)*

Авторы: *Ильина Ольга Викторовна (RU), Терешонок Максим
Валерьевич (RU), Зиядинов Вадим Валерьевич (RU)*

Заявка № 2023689454

Дата поступления 27 декабря 2023 г.

Дата государственной регистрации

в Реестре программ для ЭВМ 31 января 2024 г.



*Руководитель Федеральной службы
по интеллектуальной собственности*

ДОКУМЕНТ ПОДПИСАН ЭЛЕКТРОННОЙ ПОДПИСЬЮ
Сертификат 429b6a0fe3853164baf96f83b73b4aa7
Владелец **Зубов Юрий Сергеевич**
Действителен с 10.05.2023 по 02.08.2024

Ю.С. Зубов

РОССИЙСКАЯ ФЕДЕРАЦИЯ



СВИДЕТЕЛЬСТВО

о государственной регистрации программы для ЭВМ

№ 2024689957

Программный модуль для локализации и удаления
 состязательных патч-атак на цветных изображениях
 для решения задач компьютерного зрения

Правообладатель: *Ордена Трудового Красного Знамени
 федеральное государственное бюджетное образовательное
 учреждение высшего образования «Московский технический
 университет связи и информатики» (RU)*

Авторы: *Ильина Ольга Викторовна (RU), Терешонок Максим
 Валерьевич (RU)*



Заявка № 2024689563

Дата поступления 28 ноября 2024 г.

Дата государственной регистрации

в Реестре программ для ЭВМ 11 декабря 2024 г.

Руководитель Федеральной службы
 по интеллектуальной собственности

ДОКУМЕНТ ПОДПИСАН ЭЛЕКТРОННОЙ ПОДПИСЬЮ
 Сертификат 0692e7c1a6300b154f2401670bca2026
 Владелец **Зубов Юрий Сергеевич**
 Действителен с 10.07.2024 по 03.10.2025

Ю.С. Зубов

ПРИЛОЖЕНИЕ Б

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; мтуси.рф; 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

Проректор по научной работе МТУСИ



Н.В. Кленов

_____ 2026 г.

АКТ

об использовании результатов диссертационной работы

Ильиной Ольги Викторовны на тему

«Повышение устойчивости и точности нейросетевого обнаружения
искажённых объектов при видовом мониторинге»

Комиссия в составе:

начальник НИЧ, д.т.н., профессор М. В. Терешонок;
заведующий НИЛ, к.ф.-м.н. А. Е. Щеголев;
старший научный сотрудник, к.т.н. В. В. Зиядинов

составила настоящий акт о том, что **результаты** диссертационной работы «Повышение устойчивости и точности нейросетевого обнаружения искажённых объектов при видовом мониторинге», 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	Метод адаптивной предварительной обработки изображений при видовом мониторинге	НИР «Шеренга-2020», Государственный контракт от 12.08.2020 № 2022187150262452655002912	Реализация программных комплексов и алгоритмов функционирования; повышение качества их работы.
2	Способ формирования обучающего набора данных для повышения устойчивости модели обнаружения объектов при видовом мониторинге	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 локализации и удаления состязательных патч-атак на изображениях при видовом мониторинге	НИР «Лакмус-СПП», Государственный контракт от 28.07.2025 г. № 2527187150122452655004173	Повышение эффективности и надёжности работы разрабатываемых методов.

Использование предложенных в диссертации Ильиной Ольги Викторовны на тему «Повышение устойчивости и точности нейросетевого обнаружения искажённых объектов при видовом мониторинге» 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., профессор М.В. Терешонок
 к.ф.-м.н. А.Е. Щеголев
 к.т.н. В.В. Зиядинов