

Евдокимова Виктория Витальевна

**МЕТОД КРУПНОШАГОВОГО МЕТАОБУЧЕНИЯ В СКВОЗНОЙ
НЕЙРОСЕТЕВОЙ РЕКОНСТРУКЦИИ ОДНОГО КЛАССА ИЗОБРАЖЕНИЙ**

1.2.1. Искусственный интеллект и машинное обучение

АВТОРЕФЕРАТ
диссертации на соискание ученой степени
кандидата технических наук

Самара – 2024

Работа выполнена на кафедре суперкомпьютеров и общей информатики федерального государственного автономного образовательного учреждения высшего образования «Самарский национальный исследовательский университет имени академика С.П. Королева» и в лаборатории интеллектуального анализа видеоданных отделения «Институт систем обработки изображений – Самара» Курчатовского комплекса кристаллографии и фотоники федерального государственного бюджетного учреждения «Национальный исследовательский центр «Курчатовский институт».

Научный руководитель: доктор технических наук
Никоноров Артем Владимирович

Официальные оппоненты: **Николаев Дмитрий Петрович**,
доктор технических наук, федеральный
исследовательский центр «Информатика и
управление» Российской академии наук,
руководитель отдела 95 «Зрительных систем»;
Конушин Антон Сергеевич,
кандидат физико-математических наук,
АНО «Институт искусственного
интеллекта», старший научный сотрудник,
руководитель группы.

Ведущая организация: автономная некоммерческая образовательная
организация высшего образования «Сколковский
институт науки и технологий», г. Москва.

Защита состоится «24» декабря 2024 г. в 10 часов 00 минут на заседании диссертационного совета 24.2.379.08, созданного на базе ФГАОУ ВО «Самарский национальный исследовательский университет имени академика С.П. Королева», по адресу: 443086, Россия, г. Самара, ул. Московское шоссе, д. 34.

С диссертацией можно ознакомиться в библиотеке и на сайте ФГАОУ ВО «Самарский национальный исследовательский университет имени академика С.П. Королева»: https://ssau.ru/storage/pages/6553/file_67060231736d38.23146147.pdf

Автореферат разослан «_____» _____ 2024 г.

Ученый секретарь диссертационного совета,
доктор физико-математических наук, доцент

Дорошин А.В.

Общая характеристика работы

Актуальность. В настоящее время все больше исследовательских проектов посвящены разработке систем формирования изображений на основе изображающей дифракционной оптики, цель которой – сократить количество оптических элементов в оптической системе, сократить ее вес и стоимость при массовом производстве (Banerji 2019, Nikonov 2018, Peng 2015, Sahin 2017). Однако по сравнению с рефракционной оптикой, для изображающей дифракционной оптики характерны сильные искажения формируемых изображений. На рисунке 1 представлены примеры изображений, полученных с помощью многоуровневой дифракционной линзы (МДЛ) и рефракционного объектива. Основная задача данной работы – формирование изображений потребительского качества, то есть близких к эталонному изображению (рисунок 1.б), путем реконструкции изображений, сформированных оптическими системами на основе МДЛ (рисунок 1.а). Рисунок 1 демонстрирует пример, когда на исходном изображении, сформированном МДЛ, текст не различим, а после нейросетевой реконструкции (рисунок 1.в) часть текста (первый абзац) становится читаемой.



Рисунок 1 – Примеры изображений реальной сцены, полученные: а) изображающей дифракционной оптикой, б) рефракционной оптикой, в) в результате нейросетевой реконструкции изображения а)

В рамках данной диссертационной работы используются наборы изображений, сформированные различными изображающими дифракционно-оптическими системами (ИДОС): одноапертурной (Nikonov 2018, Nikonov 2019) и мультиапертурной системами (Nikonov 2019, Степаненко 2023) на основе МДЛ с разной высотой микрорельефа, а также гибридной рефракционно-дифракционной системой (Evdokimova 2023). Разработанные в данной диссертационной работе алгоритмы реконструкции опираются на модель искажений, которая описана в ранних работах (Nikonov 2015, Nikonov 2019). Согласно данной модели, причиной искажений в ИДОС являются хроматические aberrации, которые проявляются на изображениях в виде хроматического размытия и цветовой засветки (Nikonov 2015, Nikonov 2019). Диаметр цветовой засветки на изображениях, сформированных в ИДОС, составляет больше 200 пикселей (Nikonov 2019).

Для компенсации искажений разрабатываются различные методы вычислительной постобработки изображений от алгоритмов на основе обратной свертки (Peng 2015, Nikonov 2015, Nikonov 2018) до нейросетевой реконструкции (Nikonov 2018, Nikonov 2019, Peng 2019, Евдокимова 2020, Dup 2020, Евдокимова

2021, Ivliev 2022). Ранние подходы к компенсации искажений используют алгоритмы на основе обратной свертки и не обеспечивают желаемого потребительского качества формируемого изображения.

Большой вклад в развитие методов реконструкции изображений в дифракционно-оптических системах внесли работы (Nikonorov 2015, Peng 2015, Nayag 2015, Sun 2017). В ранних работах впервые представлено использование дифракционной линзы Френеля в изображающих оптических системах (Nayag 2015). В 2015 году представлены модель искажений в ИДОС, 4-хуровневая дифракционная линза и алгоритм постобработки регистрируемых изображений, открывшие перспективы использования дифракционно-оптических элементов для получения изображений потребительского качества (Nikonorov 2015). В работах (Nikonorov 2015, Peng 2015, Nikonorov 2018) описываются алгоритмы реконструкции на основе обратной свертки, а в работе (Sun 2017) предлагается учитывать межканальную связь для коррекции хроматических аберраций. В работе (Харитонов 2023) предложена компьютерная технология моделирования процесса формирования оптических изображений дифракционной линзой.

Комбинация алгоритма обратной свертки и нейросетевой реконструкции изображений на основе сети VDSR (Nikonorov 2018) показала многообещающие результаты (средний прирост пикового отношения сигнала к шуму (PSNR) составил 4,72 дБ). Исследования задачи реконструкции изображений в ИДОС в настоящее время мало. В работах коллектива Стэнфордского университета предлагается нейросетевая реконструкция на основе U-Net-подобных архитектур (Peng 2019, Dun 2020). Прирост среднего значения PSNR при использовании комбинации МДЛ и генеративно-состязательной нейронной сети (Peng 2019), разработанных командой Стэнфордского университета, составил 4,61 дБ (средний PSNR изображений до реконструкции – 21,28 дБ, а после реконструкции – 25,89 дБ). Ранние результаты не отражают весь потенциал нейросетевой реконструкции изображений в ИДОС.

Задача реконструкции изображений в ИДОС похожа на задачу сверхразрешения по одному изображению (SISR), но для задачи реконструкции изображений в ИДОС характерен дополнительный тип искажений – цветовая засветка, которая имеет нелокальный характер. Благодаря сходству задач критерии оценки качества реконструкции, архитектуры нейронных сетей и подходы к обучению в данной диссертационной работе берут вдохновение в исследованиях задачи сверхразрешения.

Для оценки качества реконструкции в ИДОС используются классические критерии задачи сверхразрешения – пиковое отношение сигнала к шуму (PSNR) и индекс структурного сходства (SSIM), но оценка качества реконструкции реальных изображений, не имеющих эталона, проводится, как правило, на основе психовизуального критерия (например, mean opinion score, MOS). Поскольку для реальных изображений отсутствует эталон, то такие изображения нельзя включить в процесс обучения нейронной сети. Для обучения нейронных сетей используются наборы данных, полученные путем формирования изображений, выведенных на экран монитора. Лабораторный стенд для формирования обучающей выборки используется в работах (Nikonorov 2015, Nikonorov 2019, Peng 2019, Dun 2020, Евдокимова 2020, Евдокимова 2021, Ivliev 2022). Однако из-за различий в условиях

формирования реальных данных и обучающей выборки при реконструкции реальных изображений появляются артефакты. Высокие значения PSNR или SSIM, полученные при обучении на лабораторных данных, не обеспечивают отсутствие артефактов реконструкции на реальных изображениях. Появляется необходимость в разработке некоторой процедуры, которая позволит оценить количество артефактов реконструкции без наличия эталонного изображения и, как следствие, откроет возможность использования реальных данных в процессе обучения сети. В рамках данной диссертационной работы впервые предложен критерий оценки количества артефактов реконструкции на изображениях реальных сцен, называемый уровнем ложных контуров (False edge level, FEL).

При обработке изображений из тестовой выборки, сформированной в тех же лабораторных условиях, что и обучающий набор данных, нейросетевая реконструкция обеспечивает значительный прирост качества с точки зрения среднего значения PSNR. Однако прикладные задачи требуют использования ИДОС в условиях, отличных от лабораторных (разные ИДОС на основе МДЛ с разной высотой микрорельефа, разные конфигурации линз, разные условия формирования изображений, разные камеры), что приводит к различиям в распределении обучающего и тестового набора данных. Обучающий набор данных не содержит домены, которые характерны для реальных условий формирования изображений. Сбор данных из всех возможных доменов для обучения сети требует большого количества ресурсов и не оправдывает его реализации из-за периодически меняющейся технологии изготовления ИДОС. Поэтому задача реконструкции реальных изображений становится похожа на задачу сверхразрешения с неизвестной моделью искажений (Wang 2021) или на задачу доменной адаптации (Wang 2023), где требуется обработка изображений из неизвестных новых доменов, не используемых в процессе обучения сети. Для решения данной задачи развиваются такие подходы, как метаобучение (Finn 2017), трансферное обучение (Pan 2010), обучение по нулевой выборке (Wang 2019). В рамках данной диссертационной работы разработанный критерий FEL и алгоритмы на основе крупношагового метаобучения позволяют сократить количество артефактов реконструкции на изображениях реальных сцен и адаптировать обученную на лабораторных данных модель к новому домену по одному изображению. В данной диссертационной работе приводится теоретическое обоснование метода крупношагового метаобучения, который является модификацией мелкошагового метаобучения (Finn 2017).

Таким образом тема данной диссертационной работы является актуальной и покрывает важные пробелы области реконструкции изображений в ИДОС.

Целью работы является разработка алгоритмов сквозной нейросетевой реконструкции изображений, полученных в изображающих дифракционно-оптических системах (ИДОС), с возможностью адаптации модели под различные оптические схемы и условия формирования изображений.

В соответствии с поставленной целью сформулированы следующие **основные задачи диссертации**:

1. Разработка алгоритмов сквозной нейросетевой реконструкции изображений, сформированных в ИДОС, с целью повышения качества изображений.

2. Разработка процедуры оценки количества артефактов реконструкции на изображениях реальных сцен, не требующей наличия эталонного изображения.
3. Разработка алгоритмов формирования обучающих выборок и обучения нейросетевых моделей реконструкции изображений в ИДОС с целью уменьшения артефактов реконструкции.
4. Разработка метода и алгоритмов адаптации предложенных нейросетевых моделей реконструкции для различных оптических схем и условий формирования изображений.

Научная новизна работы состоит в следующих впервые полученных результатах.

1. Разработаны алгоритмы сквозной нейросетевой реконструкции, обеспечивающие повышение качества изображений в ИДОС, на основе полнокадровой и локальной нейросетевой обработки.
2. Теоретически обоснован метод крупношагового метаобучения, позволяющий провести адаптацию нейросетевых моделей реконструкции по нулевой выборке под различные условия формирования изображений.
3. Впервые предложен критерий оценки уровня ложных контуров (FEL), позволяющий учитывать реальные данные в процессе обучения сети с целью уменьшения артефактов нейросетевой реконструкции.
4. Разработаны новые алгоритмы аугментации данных с учетом особенностей видеoinформационного тракта, приводящих к артефактам реконструкции.

Практическая значимость работы состоит в возможности создания объективов на основе сверхлегких дифракционных оптических элементов, которые в свою очередь открывают перспективы создания сверхкомпактных беспилотных летательных аппаратов и атмосферных спутников с возможностью получения изображений на больших расстояниях.

Работы по теме диссертации выполнялись в рамках: гранта РФФИ 18-07-01390 «Разработка многообъективных систем технического зрения на основе изображающей дифракционной оптики и методов глубокого обучения» (Самарский университет); гранта РФФИ 19-29-09054 «Разработка и исследование архитектуры мультимодальной цифровой платформы видеосенсорики на основе дифракционно-оптических систем видимого, инфракрасного и гиперспектрального режимов съемки» (ФНИЦ "Кристаллография и фотоника" РАН); гранта РФФИ 22-19-00364 «Разработка прикладных систем технического зрения видимого, инфракрасного и гиперспектрального режима съемки на основе одно- и мультиапертурных дифракционно-оптических систем и методов глубокого обучения» (ФНИЦ "Кристаллография и фотоника" РАН); программы РАН 0022-2019-0001 «Создание цифровой платформы сенсорики на основе дифракционно-оптических систем и глубокого обучения» (ФНИЦ "Кристаллография и фотоника" РАН).

Результаты работы были внедрены в ООО «Локус» и использованы в рамках научно-исследовательских и опытно-конструкторских работ по теме «Конструктор системы дистанционного зондирования Земли для космического аппарата типа Кубсат с дифракционной оптикой и обучающими методиками для формирования

инженерных знаний, навыков и ключевых компетенций в сфере дистанционного зондирования Земли, машинного зрения и искусственного интеллекта».

Методология и методы исследования. В диссертации используются: методы распознавания образов, математические методы обработки изображений, методы обучения искусственных нейронных сетей, метод трансферного обучения, метод обучения по нулевой выборке, метод метаобучения. В ходе выполнения диссертационной работы разработаны алгоритмы решения поставленных задач, а также программные средства, реализующие данные алгоритмы. Обучение нейронных сетей проводилось с использованием открытой программной библиотеки для машинного обучения Tensorflow.

Соответствие паспорту специальности. Диссертация соответствует паспорту научной специальности 1.2.1 Искусственный интеллект и машинное обучение и охватывает следующие области исследования, входящие в данную специальность:

2. Исследования в области оценки качества и эффективности алгоритмических и программных решений для систем искусственного интеллекта и машинного обучения. Методики сравнения и выбора алгоритмических и программных решений при многих критериях.
4. Разработка методов, алгоритмов и создание систем искусственного интеллекта и машинного обучения для обработки и анализа текстов на естественном языке, для изображений, речи, биомедицины и других специальных видов данных.
12. Исследования в области «доверенных» систем класса ИИ, включая проблемы формирования тестовых выборок прецедентов, надежности, устойчивости, переобучения и т.д.
14. Методы и средства формирования массивов условно-реальных данных и прецедентов, необходимых для решения задач искусственного интеллекта и машинного обучения.

Основные положения, выносимые на защиту:

1. Разработанные алгоритмы сквозной нейросетевой реконструкции изображений в ИДОС обеспечивают на 44 % больший прирост среднего значения PSNR по сравнению с подходом на основе обратной свертки.
2. Разработанный критерий оценки уровня ложных контуров (FEL) позволяет оценивать количество артефактов реконструкции в реальных условиях формирования изображений при отсутствии эталона, позволяя выполнять обучение моделей реконструкции с существенно меньшим количеством артефактов и обеспечивая повышение средней экспертной оценки (MOS) на 0,3 балла.
3. Предложенные алгоритмы аугментации данных позволяют уменьшить уровень артефактов реконструкции на изображениях реальных сцен по сравнению с подходом без аугментации данных.
4. Полученное теоретическое обоснование метода крупношагового метаобучения позволяет реализовать независимое последовательное обучение нейросетевых моделей в новых прикладных доменах, в отличие

от мелкошагового метода, в котором при появлении нового домена требуется обучение всех нейросетевых моделей во всех доменах.

5. Алгоритм адаптации на основе метода крупношагового метаобучения позволяет адаптировать модель под различные условия формирования изображений по одному изображению реальной сцены и обеспечивает меньший уровень артефактов реконструкции по сравнению с алгоритмом без крупношагового метаобучения.

Достоверность полученных результатов подтверждена посредством воспроизводимых вычислительных экспериментов на больших наборах изображений, сформированных с использованием различных одно- и мультиапертурных ИДОС, МДЛ с разной высотой микрорельефа, а также гибридной дифракционно-рефракционной системой. Часть собранных наборов изображений и разработанных нейросетевых моделей выложена в общий доступ. Описания предложенных алгоритмов допускают их воспроизводимость и опубликованы в ведущих научных журналах и трудах конференций по машинному зрению и обработке изображений.

Апробация работы. Результаты работы были представлены на 6 международных конференциях:

1. International Conference on Computer Vision Workshop (ICCVW-2019), Seoul, Korea, 2019.
2. V International conference on information technology and nanotechnology (ITNT – 2019), Samara, 2019.
3. VI International conference on information technology and nanotechnology (ITNT – 2020), Samara, 2020.
4. 8th international workshop “Image Mining. Theory and applications” on International Conference of Pattern Recognition, Mailan, Italy, 2022.
5. VIII International conference on information technology and nanotechnology (ITNT – 2022), Samara, 2022.
6. IX International conference on information technology and nanotechnology (ITNT – 2023), Samara, 2023.

Авторский вклад в работу является определяющим. Все изложенные в работе и выносимые на защиту результаты получены автором лично. Постановка задач и формализация результатов проводились совместно с научным руководителем. В диссертации использованы результаты совместных исследований: формирование и подготовка наборов изображений выполнена М.В. Петровым совместно с Н.А. Ивлевым; разработка алгоритмов реконструкции на основе обратной свертки, с которыми проводилось сравнение, выполнена М.В. Петровым; все изображающие оптические системы, исследованные в работе, разработаны Р.В. Скидановым, Н.А. Ивлевым, В.В. Подлипиновым, С.В. Ганчевской; ручная обработка изображений реальных сцен проводилась А. П. Алексеевым. Из работ в соавторстве на защиту выносятся результаты, в получении которых автор принимал непосредственное участие.

Публикации. Основные результаты по теме диссертации изложены в 16 печатных работах: 9 из которых изданы в журналах, индексируемых Scopus и WoS; 3 – в журналах, рекомендованных ВАК; 4 – в тезисах докладов.

Объем и структура работы. Диссертация состоит из введения, 4 глав, заключения и 5 приложений. Объем диссертации составляет 147 страниц, включая 79 изображений и 8 таблиц. Список литературы содержит 110 наименований.

Основное содержание работы

Во **введении** описано состояние рассматриваемой проблемы и обоснована актуальность темы диссертации. Сформулированы цели и задачи исследования, новизна полученных результатов и защищаемые положения.

В первой главе описываются преимущества и недостатки ИДОС, история совершенствования ИДОС, приведены модель искажений в ИДОС и обзор существующих методов реконструкции изображений, описана процедура формирования набора данных и введены критерии оценки качества.

Модель искажений, характерных для ИДОС (Nikonov 2015), включает два типа искажений: хроматическое размытие, которое имеет локальный характер на формируемом изображении, и цветовую засветку (нелокальный характер искажения), диаметр которой составляет больше 200 пикселей (Nikonov 2019), что накладывает ограничение на размер рецептивного поля нейронной сети для компенсации двух типов искажений в ИДОС.

Для оценки качества формируемых изображений используется пиковое отношение сигнала к шуму (англ. peak signal-to-noise ratio, PSNR):

$$PSNR = 10 \log_{10} \left(\frac{MAX_I^2}{RMSE^2} \right) = 20 \log_{10} \left(\frac{MAX_I}{RMSE} \right),$$

где MAX_I – это максимальное значение, принимаемое пикселем изображения, в рамках данной работы MAX_I принимается равным 255,

$$RMSE = \sqrt{\frac{1}{mn} \sum_{i=0}^{m-1} \sum_{j=0}^{n-1} |I(i, j) - K(i, j)|^2},$$

I, K – полутоновые изображения размера $m \times n$.

Также в первой главе описан лабораторный стенд и различные дифракционно-оптические системы, используемые для формирования обучающих данных: одноапертурные системы на основе МДЛ с разной высотой микрорельефа (Nikonov 2018, Nikonov 2019), двухапертурная система на основе МДЛ (Nikonov 2019), трехапертурная система на основе МДЛ (Степаненко 2023), гибридная система (Evdokimova 2023).

В прикладных задачах наибольший интерес представляет реконструкция реальных изображений, которые получены в условиях, отличных от лабораторных. Реальные изображения не имеют эталона, который требуется для расчета PSNR. В таком случае, как правило, оценка качества изображений, не имеющих эталона, проводится на основе психовизуального критерия. В исследованиях задачи сверхразрешения (Ledig, 2017) используется средняя экспертная оценка (mean opinion score, MOS), которая определяется следующим образом:

$$MOS = \frac{\sum_{n=1}^N R_n}{N},$$

где R_n – оценка по шкале от 1 до 5, где 1 – очень высокое количество артефактов реконструкции, а 5 – артефакты реконструкции незаметны, N – количество участников опроса.

Во второй главе описываются алгоритмы сквозной нейросетевой реконструкции изображений, обеспечивающие повышение качества изображений в ИДОС, на основе полнокадровой и локальной нейросетевой обработки.

На рисунке 2 изображена двухэтапная схема реконструкции, которая применима как для одноапертурных, так и для многоапертурных ИДОС на основе МДЛ.

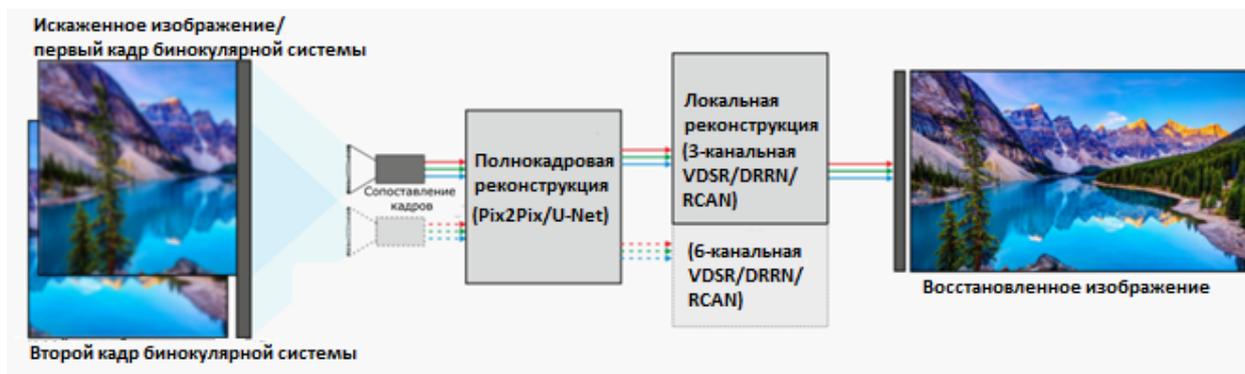


Рисунок 2 – Двухэтапная схема сквозной нейросетевой реконструкции изображений, полученных в монокулярной и бинокулярной системах на основе МДЛ

Для полнокадровой нейросетевой реконструкции используются U-Net-подобные архитектуры (Pix2Pix, модификация U-Net), рецептивное поле которых составляет больше 200×200 пикселей, что позволяет компенсировать цветовую засветку, диаметр которой в ИДОС составляет больше 200 пикселей (Nikonov 2019).

Все используемые архитектуры подробно описаны в главе 2. Модификацию архитектуры U-Net можно записать в следующем виде:

Encoder: CB64-CB128-CB256-CB512-CB512-CB512-CB512-CB512,

Decoder: TcBD512- TcBD512- TcBD512-TcB512-TcB256-TcB128-TcB64,

где CB_k – это комбинация сверточного слоя, пакетной (батч) нормализации и функции активации с количеством фильтров k , TcBD k означает, что перед активационным слоем добавляется dropout-слой.

Локальная реконструкция изображений проводится на основе классических для задачи сверхразрешения (с известной моделью искажений) архитектур – VDSR (Kim 2016), DRRN (Tai 2017) с небольшим рецептивным полем (41×41 пикселей) и сетью RCAN (Zhang 2018) с механизмом внимания. В работе используются трехканальная и шестиканальная модификации сети VDSR (Nikonov 2019), которые описаны в главе 2.

В третьей главе рассматриваются методы и алгоритмы адаптации предложенных нейросетевых моделей реконструкции для различных оптических схем и условий формирования изображений.

В третьей главе предложена новая процедура оценки количества артефактов реконструкции на изображениях реальных сцен. Критерий, называемый уровнем ложных контуров (FEL), позволяет оценить качество реальных изображений без наличия эталона, который требуется для расчета PSNR, а также использовать реальные данные в процессе обучения сети. Критерий ранней остановки обучения на

основе FEL предложен в качестве альтернативы критерию на основе среднего значения PSNR, рассчитанного на лабораторных данных, с целью уменьшения артефактов реконструкции на изображениях реальных сцен. Также с целью уменьшения артефактов реконструкции в третьей главе предложены алгоритмы аугментации обучающего набора данных. Кроме того, в третьей главе теоретически обоснован метод крупношагового метаобучения.

Поскольку артефакты реконструкции на изображении визуально имеют вид контуров (рисунок 3.а и рисунок 3.б), то и количественная оценка таких артефактов основана на обнаружении контуров для фрагмента изображения, на котором визуально контуров быть не должно. Уровень ложных контуров (False edge level, FEL) определяется как процент пикселей, принадлежащих контуру, для фрагмента изображения, на котором визуально не должно быть контуров:

$$FEL = \frac{1}{NM} \sum_{\substack{0 \leq i \leq N-1 \\ 0 \leq j \leq M-1}} E_{ij} \cdot 100, \quad (3.1)$$

где E – это бинарная маска контуров размера $N \times M$, полученная в результате работы алгоритма Canny (эвристически параметры порога установлены, равными 0 и 70). Значения маски контуров E должны быть нормализованы, то есть приведены к диапазону $[0, 1]$.

В главе 3 выделяются следующие требования к выбору фрагмента: ширина или высота фрагмента больше 200 пикселей (что соответствует диаметру цветовой засветки в ИДОС) и на аналогичном фрагменте изображения, сформированным рефракционной камерой, значение FEL должно быть минимально. Выбор алгоритма обнаружения контуров не ограничен, но в рамках данной работы используется алгоритм Canny.

На рисунке 3 приведены результаты работы алгоритма Canny на фрагментах восстановленных изображений реальной сцены. Для каждого фрагмента указаны значения FEL. Подтверждение эффективности критерия FEL проводилось путем расчета средней экспертной оценки MOS по данным 26 участников опроса (Ledig, 2017). Для определения статистической значимости различий средних величин использовался t-критерий Стьюдента.

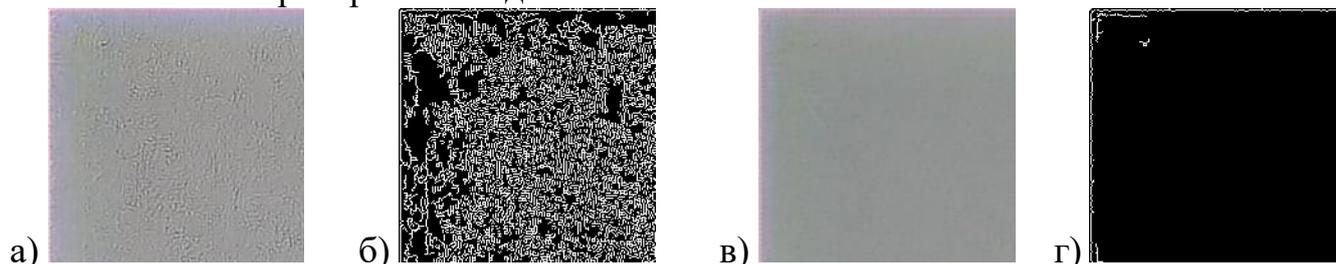


Рисунок 3 – Примеры работы алгоритма Canny: а) фрагмент восстановленного изображения реальной сцены ($FEL = 30,69\%$, $MOS = 1$, $p < 0,05$), б) результат работы алгоритма Canny для а), в) фрагмент восстановленного изображения реальной сцены ($FEL = 1,29\%$, $MOS = 3$, $p < 0,05$), г) результат работы алгоритма Canny для в)

Классический критерий ранней остановки обучения сети (Nikonov 2018, Nikonov 2019, Евдокимова 2021, Ivliev 2022, Степаненко 2023, Евдокимова 2023) основан на среднем значении PSNR, рассчитанным по валидационной выборке.

В главе 3 предлагается новый критерий ранней остановки обучения, основанный на значении FEL для фрагмента изображения реальной сцены. Критерием выбора лучшей точки в пространстве параметров является **минимальное значение FEL** на фрагменте реального изображения (Evdokimova 2023):

$$best_epoch = \underset{e}{\operatorname{argmin}} FEL(T_e^{real}),$$

где T_e^{real} – фрагмент реального изображения, восстановленный моделью с параметрами, полученными на эпохе обучения e .

Также для уменьшения уровня артефактов реконструкции в главе 3 предлагается аугментации обучающего набора данных путем моделирования двух видов искажения: шум, причиной которого является усиление или светочувствительность камеры (ISO шум), и изменение экспозиции изображения.

Изменение экспозиции изображений обучающей выборки проводилось с использованием алгоритма повышения качества изображения, полученного в условиях низкой освещенности (Ying 2017).

Для моделирования ISO-шума в данной работе используется реализация Пуассона-Гауссовского шума (Foi 2008) из библиотеки albuminations, в рамках которой RGB изображение переводится в цветовое пространство HLS (hue – тон, saturation – насыщенность, lightness – яркость), затем шум Пуассона применяется к каналу яркости (L), а шум Гаусса – к каналу тон (H).

В главе 3 представлено теоретическое обоснование метода крупношагового метаобучения, который является модификацией метода мелкошагового метаобучения (Finn 2017).

Трехэтапный алгоритм на основе метода крупношагового метаобучения используется для адаптации модели к новым условиям формирования изображений по нулевой выборке. Схема трехэтапного алгоритма адаптации представлена на рисунке 4.

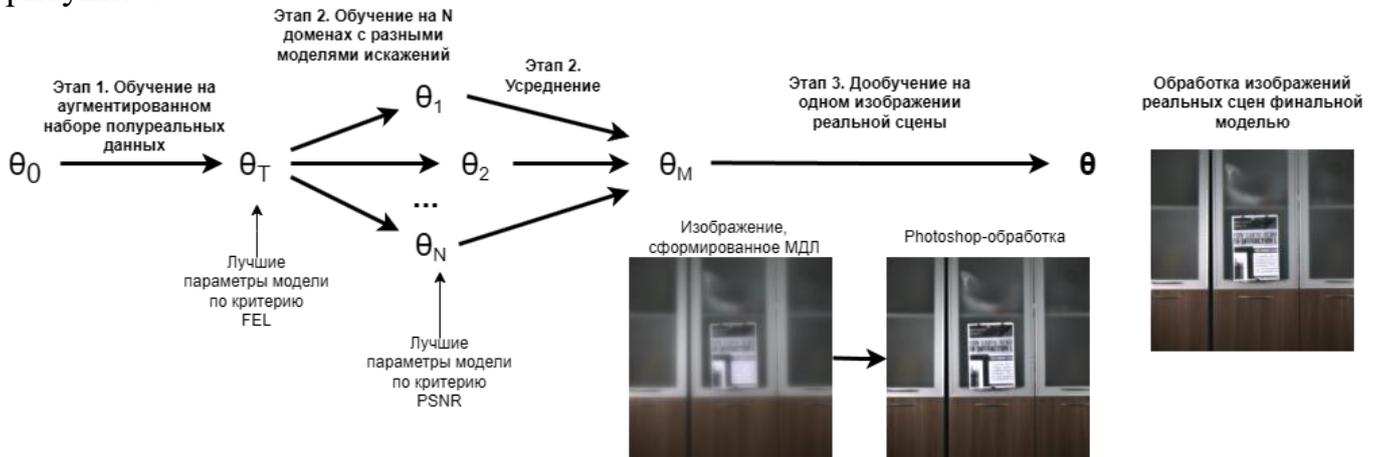


Рисунок 4 – Трехэтапный алгоритм адаптации нейросетевой модели к различным условиям формирования реальных изображений на основе крупношагового метаобучения

Формула расчета параметров финальной модели согласно методу крупношагового метаобучения имеет следующий вид:

$$\theta_M = (1 - \gamma N)\theta_T + \gamma \sum_{i=1}^N \theta^{iK},$$

где θ_T – параметры модели, полученные после 1 этапа обучения, θ_M – параметры модели, полученные после 2 этапа обучения, N – количество доменов, α, β – параметры скорости обучения, $\gamma = \frac{\beta}{\alpha}$, K – номер эпохи ранней остановки.

В четвертой главе приведены результаты экспериментальных исследований алгоритмов сквозной нейросетевой реконструкции изображений, критерия FEL, аугментации обучающего набора данных и алгоритма адаптации нейросетевой модели на основе метода крупношагового метаобучения.

Оценка качества реконструкции изображений проведена в рамках трех экспериментов (таблицы 1–3) для разных конфигураций дифракционно-оптических систем и многоуровневых дифракционных линз с разной высотой микрорельефа.

В таблице 1 представлены результаты реконструкции на основе алгоритмов сквозной нейросетевой реконструкции для монокулярной и бинокулярной систем на основе МДЛ ($m=13$), а также приведено сравнение с подходом на основе обратной свертки. Объемы обучающей, валидационной и тестовой выборок для монокулярной системы составляли 321, 29, 29 изображений соответственно, а для бинокулярной системы – 119, 29 и 29 изображений. По сравнению с подходом на основе обратной свертки алгоритмы сквозной нейросетевой реконструкции изображений обеспечивают лучшее качество реконструкции (на 7,55 дБ больше с точки зрения PSNR).

Таблица 1 – Монокулярная и бинокулярная системы на основе МДЛ ($m=13$)

Метод реконструкции	Входное качество (PSNR, дБ)	Полнокадровая реконструкция (PSNR, дБ)	Локальная реконструкция (PSNR, дБ; прирост)	Общий прирост (PSNR, дБ)
Обратная свертка (Монокулярная система) []	17,21	18,41		1,2
Pix2pix+ трехканальная DRRN (Монокулярная система)	17,21	25,51	25,96 (0,45)	8,75
Pix2Pix + шестиканальная VDSR (Бинокулярная система)	17,21	25,41	26,07 (0,66)	8,86

В таблице 2 приведены результаты исследования влияния технологии изготовления МДЛ на качество реконструкции. В эксперименте использовались наборы данных, сформированные на основе МДЛ с разной высотой микрорельефа ($m=7$ и $m=13$). Объемы обучающей, валидационной и тестовой выборок составили 1226, 15 и 605 изображений соответственно.

Таблица 2 – Монокулярная система на основе МДЛ ($m=13$ и $m=7$)

m	Входное качество (PSNR, дБ)	Полнокадровая реконструкция, U-Net (PSNR, дБ)	Локальная реконструкция, VDSR (PSNR, дБ; прирост)	Общий прирост (PSNR, дБ)
13	19,68	27,02	27,26 (0,24)	7,58
7	18,13	26,65	26,82 (0,17)	8,69

В таблице 3 представлены результаты реконструкции для гибридной и трехапертурных систем. Объемы обучающей, валидационной и тестовой выборок для двух систем составили 1244, 15 и 605 изображений соответственно.

Таблица 3 – Гибридная, трехапертурная системы на основе МДЛ ($m=1$)

ИДОС	Входное качество (PSNR, дБ)	Полнокадровая реконструкция, U-Net (PSNR, дБ)	Локальная реконструкция, RCAN (PSNR, дБ; прирост)	Общий прирост (PSNR, дБ)
Гибридная	17,85	28,09	28,32 (0,23)	10,47
Трехапертурная	12,77	25,67	25,8 (0,13)	13,03

На рисунке 5 представлены результаты экспериментальных исследований комбинации критерия ранней остановки обучения на основе FEL и аугментации обучающего набора данных, которые показали, что данная комбинация позволяет значительно уменьшить уровень артефактов реконструкции на изображениях реальных сцен по сравнению с подходом без аугментации данных и классическим критерием ранней остановки обучения на основе PSNR.

На рисунке 6 представлены результаты исследования алгоритма адаптации нейросетевой модели реконструкции на основе крупношагового метаобучения. Показано сравнение реконструкции изображения реальной сцены, не используемое в процессе обучения сети, для двух случаев: обучение в рамках этапов 1 и 3 с пропуском второго этапа крупношагового метаобучения и обучение в рамках трех этапов алгоритма адаптации. На изображениях выделены фрагменты, для которых рассчитаны значения FEL. Видно, что использование предлагаемого алгоритма адаптации обеспечивает меньший уровень артефактов реконструкции на изображениях реальных сцен по сравнению с подходом без крупношагового метаобучения.

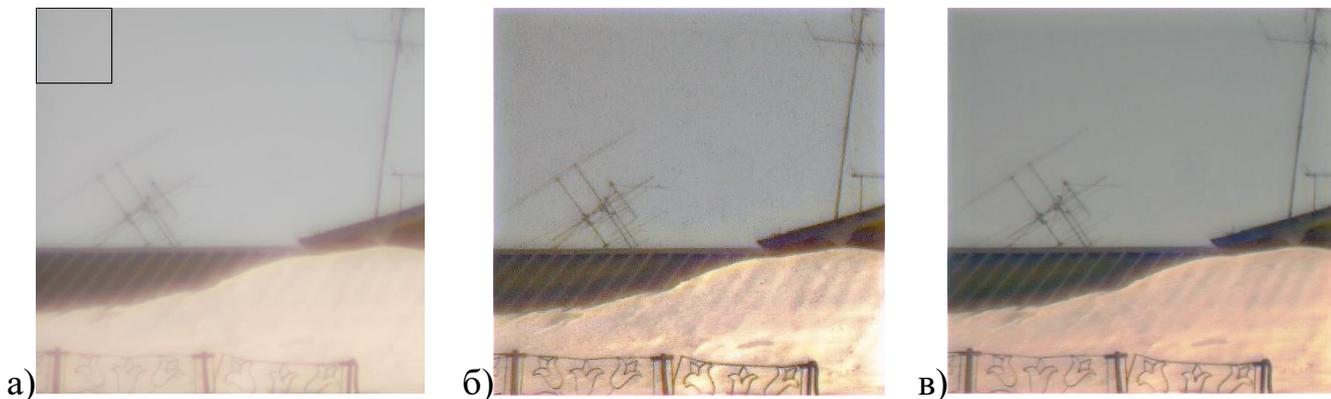
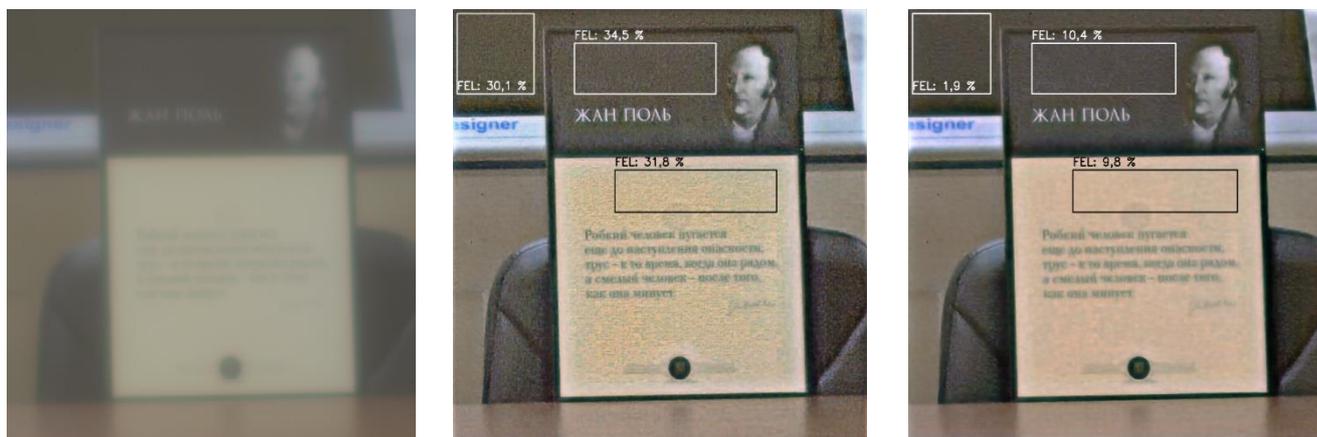
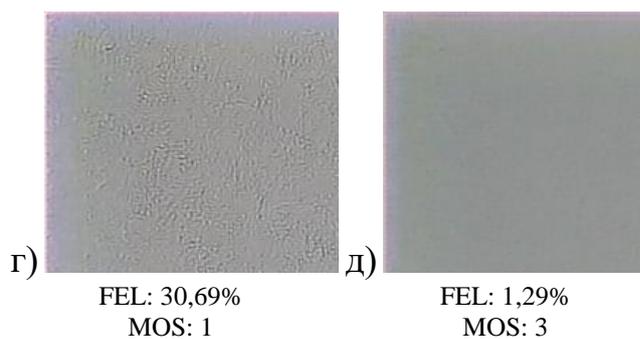


Рисунок 5 – Пример реконструкции изображения реальной сцены: а) изображение, сформированное гибридной системой, б) восстановленное в случае обучения без аугментации данных и с использованием критерия ранней остановки обучения на основе PSNR, в) восстановленное в случае обучения с аугментацией данных и с использованием критерия ранней остановки обучения на основе FEL, г) фрагмент восстановленного изображения б, д) фрагмент восстановленного изображения в



а) б) в)

Рисунок 6 – Результаты адаптации модели к новым условиям формирования реальных изображений по нулевой выборке: а) изображение реальной сцены, полученное МДЛ, б) восстановленное моделью, обученной в рамках 1 и 3 этапов алгоритма (пропущен этап крупношагового метаобучения), в) восстановленное моделью, обученной в рамках трех этапов алгоритма на основе крупношагового метаобучения

Заключение

В данной работе разработаны алгоритмы сквозной нейросетевой реконструкции изображений, полученных в изображающих дифракционно-оптических системах (ДОС), с возможностью адаптации модели под различные оптические схемы и условия формирования изображений. Основные результаты работы заключаются в следующем.

1. Разработанные алгоритмы сквозной нейросетевой реконструкции изображений в ИДОС обеспечивают прирост среднего значения PSNR на 7 дБ больше по сравнению с алгоритмом на основе обратной свертки и не зависят от технологии изготовления МДЛ.
2. Получено теоретическое обоснование метода крупношагового метаобучения, позволяющего провести адаптацию нейросетевой модели реконструкции изображений по нулевой выборке под различные условия формирования изображений.
3. Разработан критерий оценки уровня ложных контуров (FEL), позволяющий оценить количество артефактов реконструкции при отсутствии эталона, который требуется для расчета PSNR и SSIM.
4. Использование критерия ранней остановки обучения на основе FEL обеспечивает уменьшение количества артефактов реконструкции по сравнению с классическим критерием на основе PSNR, для дифракционно-рефракционной системы количество артефактов уменьшается на 16,21 % по критерию FEL и на 0,76 балла по критерию MOS ($p < 0,05$), а в случае трехапертурной системы – на 8,4% по критерию FEL и на 0,3 балла по критерию MOS ($p < 0,05$).
5. Предложенные алгоритмы аугментации обучающих данных в комбинации с критерием ранней остановки на основе FEL обеспечивают наименьший уровень артефактов на изображениях реальных сцен по сравнению с подходом без аугментации данных и с использованием критерия ранней остановки на основе PSNR, в случае дифракционно-рефракционной системы получено сокращение количества артефактов на 29,4% по критерию FEL и на 2 балла по критерию MOS ($p < 0,05$), в случае трехапертурной системы на 17 % по критерию FEL и на 1,5 балла по критерию MOS ($p < 0,05$).
6. Использование разработанного алгоритма адаптации на основе крупношагового метаобучения позволяет уменьшить уровень артефактов реконструкции от 3 до 7 раз.

Публикации автора по теме диссертации

В изданиях, входящих в международные базы цитирования Web of Science и Scopus:

Опубликовано 3 работы в изданиях Q1.

1. Nikonorov, A. Toward Ultralightweight Remote Sensing With Harmonic Lenses and Convolutional Neural Networks / A. Nikonorov, M. Petrov, S. Bibikov, P. Yakimov, V. Kutikova, Y. Yuzifovich, A. Morozov, R. Skidanov, N. Kazanskiy // IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing. – 2018. – Vol. 11, no. 9. – P. 3338-3348 (Q1).

2. Ivliev, N. First Earth-Imaging CubeSat with Harmonic Diffractive Lens / V. Evdokimova, V. Podlipnov, M. Petrov, S. Ganchevskaya, L. Tkachenko, D. Abrameshin, Y. Yuzifovich, A. Nikonorov, R. Skidanov, N. Kazanskiy, V. Soifer // *Remote Sensing*. – 2022. – V. 14, no. 2230 (Q1).
3. Evdokimova, V.V. Hybrid Refractive-Diffractive Lens with Reduced Chromatic and Geometric Aberrations and Learned Image Reconstruction / V.V. Evdokimova, V.V. Podlipnov, N.A. Ivliev, M.V. Petrov, S.V. Ganchevskaya, V.A. Fursov, Y. Yuzifovich, S.O. Stepanenko, N.L. Kazanskiy, A.V. Nikonorov, R.V. Skidanov // *Sensors*. – 2023. – Vol. 23. – p. 415 (Q1).
4. Nikonorov, A. Deep learning-based enhancement of hyperspectral images using simulated ground truth / A. Nikonorov, M. Petrov, S. Bibikov, V. Kutikova, P. Yakimov, A. Morozov, R. Skidanov, N. Kazanskiy // *10th IAPR Workshop on Pattern Recognition in Remote Sensing*. – 2018. – no. 8486408.
5. Nikonorov, A. Towards multi-aperture imaging using diffractive lens / A. Nikonorov, R. Skidanov, V. Evdokimova, M. Petrov, A. Alekseev, S. Bibikov, N. Kazanskiy // *Proceedings of SPIE - The International Society for Optical Engineering*. – 2019. – Vol. 11146.
6. Nikonorov, A. Deep learning-based imaging using single-lens and multi-aperture diffractive optical systems / A. Nikonorov, V. Evdokimova, M. Petrov, P. Yakimov, S. Bibikov, Y. Yuzifovich, R. Skidanov, N. Kazanskiy // *International Conference on Computer Vision Workshop (ICCVW 2019)*. – 2019. – V. 9022384. – P. 3969 – 3977.
7. Nikonorov, A. Deep learning-based image reconstruction for multi-aperture diffractive lens / A. Nikonorov, R. Skidanov, V. Evdokimova, M. Petrov, A. Alekseev, S. Bibikov, N. Kazanskiy // *Journal of Physics: Conference Series*. – 2019. – Vol. 1368, no 5.
8. Evdokimova, V. Study of GAN-based image reconstruction for diffractive optical systems / V. Evdokimova, M. Petrov, M. Kluyeva, N. Firsov, S. Bibikov, R. Skidanov, S. Popov, A. Nikonorov // *Proceedings of ITNT 2020 - 6th IEEE International Conference on Information Technology and Nanotechnology*. – 2020.
9. Evdokimova, V.V. Meta-Learning Approach in Diffractive Lens Computational Imaging. / V.V. Evdokimova, S.A. Bibikov, A.V. Nikonorov // *Pattern Recognit. Image Anal.* – 2022. – Vol. 32. – p. 466–468.

В изданиях из списка ВАК РФ:

10. Nikonorov, A.V. Реконструкция изображений в дифракционно-оптических системах на основе сверточных нейронных сетей и обратной свертки / A.V. Nikonorov, M.V. Petrov, S.A. Bibikov, V.V. Kutikova, A.A. Morozov, N.L. Kazanskiy // *Компьютерная оптика*. – 2017. – Т. 41, № 6. – С. 875-887.
11. Евдокимова, В.В. Нейросетевая реконструкция видеопотока в дифракционных оптических системах массового производства / В.В. Евдокимова, М.В. Петров, М.А. Ключева, Е.Ю. Зыбин, В.В. Косьянчук, И.Б. Мищенко, В.М. Новиков, Н.И. Сельвесюк, Е.И. Ершов, Н.А. Ивлиев, Р.В.

Скиданов, Н.Л. Казанский, А.В. Никоноров // Компьютерная оптика. – 2021. – Т. 45, № 1. – С. 130-141.

12. Степаненко С.О. Получение цветных изображений системой на основе трех дифракционных линз / С.О. Степаненко, В.В. Евдокимова, М.В. Петров, Р.В. Скиданов, А.В. Никоноров // Компьютерная оптика. – 2023. – Т. 47, № 5. – С. 716-724.

В сборниках трудов конференций:

13. Никоноров, А.В. Исследование мультиапертурных систем на основе дифракционной оптики и глубокого обучения/ А.В. Никоноров, Р.В. Скиданов, В.В. Евдокимова, М.В. Петров, А.П. Алексеев, С.А. Бибииков, Н.Л. Казанский // V Международная конференция и молодёжная школа «Информационные технологии и нанотехнологии» (ИТНТ-2019). – 2019. – С. 993-1000.
14. Евдокимова, В. Исследование генеративно-состязательных нейронных сетей для решения задачи реконструкции изображений в дифракционно-оптических системах / В.В. Евдокимова, М.В. Петров, М.А. Ключева, А.П. Алексеев, С.А. Бибииков, Р.В. Скиданов, А.В. Никоноров // VI Международная конференция и молодёжная школа «Информационные технологии и нанотехнологии» (ИТНТ-2020). – 2020. – Т. 4. – С. 305-311.
15. Евдокимова, В. В. Адаптация нейросетевых алгоритмов реконструкции изображений для различных условий экспозиции // VIII Международная конференция и молодёжная школа «Информационные технологии и нанотехнологии» (ИТНТ-2022). – 2022. – Т. 4. – С. 043602.
16. Евдокимова, В.В. Исследование метаобучения для нейросетевой реконструкции изображений в дифракционно-оптических системах / В.В. Евдокимова, В.Д. Рябов, А.П. Алексеев, А.В. Никоноров // Информационные процессы. – 2023. – Т. 23, № 3. – С. 335–342.