

МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ РОССИЙСКОЙ ФЕДЕРАЦИИ  
ФЕДЕРАЛЬНОЕ ГОСУДАРСТВЕННОЕ АВТОНОМНОЕ  
ОБРАЗОВАТЕЛЬНОЕ УЧРЕЖДЕНИЕ ВЫСШЕГО ОБРАЗОВАНИЯ  
«САМАРСКИЙ НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ УНИВЕРСИТЕТ ИМЕНИ  
АКАДЕМИКА С.П. КОРОЛЕВА»  
(САМАРСКИЙ УНИВЕРСИТЕТ)

ОТДЕЛЕНИЕ «ИНСТИТУТ СИСТЕМ ОБРАБОТКИ ИЗОБРАЖЕНИЙ – САМАРА»  
КУРЧАТОВСКОГО КОМПЛЕКСА КРИСТАЛЛОГРАФИИ И ФОТОНИКИ  
ФЕДЕРАЛЬНОГО ГОСУДАРСТВЕННОГО БЮДЖЕТНОГО УЧРЕЖДЕНИЯ  
«НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ ЦЕНТР «КУРЧАТОВСКИЙ ИНСТИТУТ»

На правах рукописи

**Евдокимова Виктория Витальевна**

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

**1.2.1 Искусственный интеллект и машинное обучение**

Диссертация на соискание учёной степени  
кандидата технических наук

Научный руководитель:  
**Никоноров Артем Владимирович,**  
доктор технических наук

Самара – 2024

## ОГЛАВЛЕНИЕ

Введение .....	5
Глава 1. Постановка задачи оценки и повышения качества изображений в дифракционно-оптических системах .....	15
1.1 Преимущества и недостатки ИДОС .....	15
1.2 Формирование изображений в дифракционно-оптических системах .....	18
1.3 Компенсация искажений изображения за счет совершенствования оптической схемы и различные схемы получения изображений.....	20
1.3.1 Гармоническая линза .....	20
1.3.2 Изображающая система из трех дифракционных линз .....	24
1.3.3 Дифракционно-рефракционная оптическая система .....	25
1.4 Вычислительный способ компенсации искажений .....	27
1.5 Оценка качества изображений в ИДОС .....	32
1.5.1 Пиковое отношение сигнала к шуму .....	33
1.5.2 Лабораторный стенд для формирования обучающего набора данных.....	33
1.5.3 Психовизуальный критерий оценки качества реконструкции изображений реальных сцен .....	37
1.5.4 Качество формируемых изображений для разных ИДОС .....	38
Выводы к главе 1 .....	42
Глава 2. Сквозная нейросетевая реконструкция изображений в ИДОС.....	43
2.1. Нейросетевой подход к реконструкции изображений в ИДОС.....	43
2.2. Полнокадровая нейросетевая реконструкция.....	45
2.3 Локальная нейросетевая реконструкция .....	50
Выводы к главе 2 .....	54
Глава 3. Алгоритмы обучения и адаптации нейросетевых моделей реконструкции изображений в ИДОС.....	55
3.1 Ограничения при формировании набора данных .....	55

3.2 Доменная генерализация нейросетевых моделей .....	59
3.3 Критерий ложных контуров .....	61
3.4 Аугментация обучающего набора данных.....	63
3.4.1 Изменение экспозиции обучающих данных.....	63
3.4.2 Моделирование ISO-шума на обучающих данных.....	66
3.5 Метод крупношагового метаобучения .....	67
3.5.1 Мелкошаговое метаобучение для задачи сверхразрешения .....	68
3.5.2 Адаптация модели к различным условиям формирования изображений на основе крупношагового метаобучения .....	70
3.5.3 Теоретическое обоснование метода крупношагового метаобучения .....	75
Выводы к главе 3 .....	78
Глава 4. Экспериментальные исследования .....	79
4.1 Экспериментальное исследование алгоритмов сквозной нейросетевой реконструкции изображений.....	79
4.1.1 Оценка качества сквозной нейросетевой реконструкции для монокулярной и бинокулярной ОС на основе МДЛ ( $m=13$ ).....	79
4.1.3 Исследование влияния технологии изготовления МДЛ на качество реконструкции .....	84
4.2 Исследование аугментации обучающего набора данных и критерия FEL для уменьшения артефактов реконструкции.....	91
4.2.1 Обучение полнокадровой сети U-Net.....	91
4.2.2 Критерий выбора лучшей точки в пространстве параметров модели .....	92
4.2.3 Аугментация обучающих данных моделированием ISO-шума и изменением экспозиции .....	95
4.3 Исследование алгоритма адаптации нейросетевой модели к новым условиям формирования изображений на основе метода крупношагового метаобучения....	103
4.3.1 Обучение на исходном наборе полуреальных лабораторных данных.....	103

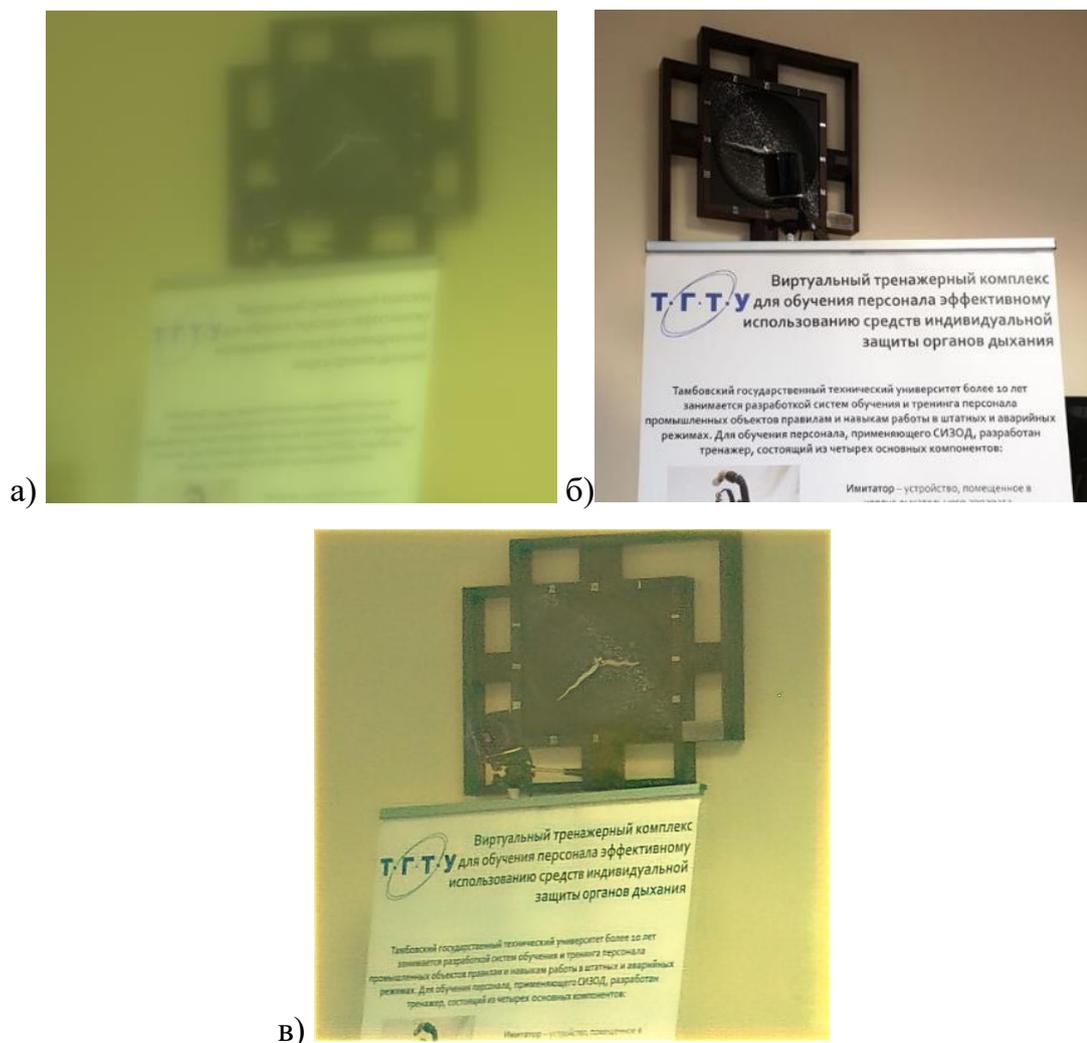
4.3.2 Обучение на N доменах с разными моделями искажений .....	108
4.3.3 Дообучение на одном изображении реальной сцены .....	110
4.3.4 Добавление нового домена .....	116
4.4 Сравнение качества реконструкции изображений для разных ИДОС .....	119
Выводы к главе 4 .....	122
Заключение.....	123
Список сокращений и условных обозначений .....	125
Список литературы.....	126
Приложение А – Акт о внедрении .....	137
Приложение Б – Результаты реконструкции изображений для разных ИДОС .....	138
Приложение В – Примеры реконструкции реальных изображений для рефракционно-дифракционной ОС .....	139
Приложение Г – Дополнительные примеры реконструкции реальных изображений для метода крупношагового метаобучения .....	142
Приложение Д – Свидетельства о регистрации программы для эвм .....	146

## ВВЕДЕНИЕ

### **Актуальность темы и степень ее разработанности.**

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 [1-4]. 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 классического 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.

Существует три подхода к уменьшению размеров изображающих оптических систем по сравнению с рефракционной оптикой: металинзы (метаповерхности) [5], безлинзовые системы [6] и многоуровневые дифракционные линзы (МДЛ) [2, 7-11, 12-16]. Преимущество металинз заключается в возможности создания оптической системы с субволновой толщиной. Однако металинзы, как и безлинзовые системы, сложны в изготовлении при массовом производстве и пока уступают по качеству формируемых изображений многоуровневым дифракционным линзам. МДЛ используются для получения цветных изображений высокого разрешения с 2015 года [12], но дифракционная линза известна как оптический элемент с 19 века.



*Рисунок 1 – Примеры изображений реальной сцены, сформированные: 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): одноапертурными [2, 7] и мультиапертурными системами [7-11] на основе МДЛ с разной высотой микрорельефа, а также гибридной рефракционно-дифракционной системой [17]. Предлагаемые в данной диссертационной работе алгоритмы реконструкции опираются на модель искажений, которая представлена в работах [7, 12]. Согласно данной модели, причиной искажений в ИДОС являются хроматические aberrации, которые проявляются на изображениях в виде локального хроматического размытия и нелокальной цветовой засветки [7, 12].

Для компенсации искажений с 2015 года разрабатываются различные методы вычислительной постобработки изображений от алгоритмов на основе обратной свертки [2, 3, 12, 18-21] до нейросетевой реконструкции [2, 7, 10, 11, 13-16, 22-27]. Ранние подходы к компенсации искажений используют классические методы на основе обратной свертки и не обеспечивают потребительского качества формируемого изображения.

Большой вклад в развитие методов реконструкции изображений в дифракционно-оптических системах внесли работы [2, 3, 12, 28-30]. В [28] впервые описано использование дифракционной линзы Френеля в изображающих оптических системах. В [12] представлены модель искажений в ИДОС, 4-хуровневая дифракционная линза и алгоритм постобработки регистрируемых изображений, открывшие перспективы использования дифракционно-оптических элементов (ДОЭ) для получения цветных изображений потребительского качества. В [30] предложена компьютерная технология моделирования процесса формирования оптических изображений дифракционной линзой. В работах [2, 3, 12] описываются алгоритмы реконструкции на основе обратной свертки, а в работе [29] предлагается учитывать межканальную связь для коррекции хроматических aberrаций. В работе [2] впервые представлена комбинация алгоритма обратной свертки и нейросетевой реконструкции изображений на основе сети VDSR, которая показала многообещающие результаты (средний прирост пикового отношения сигнала к шуму (PSNR) составил 4,72 дБ). Развитие данного подхода к реконструкции изображений в дифракционно-оптических системах описывается в рамках данной диссертационной работы, где предлагаются алгоритмы сквозной нейросетевой реконструкции, позволяющие компенсировать искажения, характерные для ИДОС.

Исследований задачи реконструкции изображений в ИДОС в настоящее время мало. В работах [31, 32] коллектива Стэнфордского университета предлагается нейросетевая реконструкция на основе U-Net-подобных архитектур. В работе [32] используются синтетические данные, а в работе [31] прирост среднего значения PSNR составил 4,61 дБ (качество изображений до реконструкции – 21,28 дБ, а после реконструкции – 25,89 дБ), что не отражает весь потенциал нейросетевой реконструкции.

Задача реконструкции изображений в ИДОС похожа на задачу сверхразрешения по одному изображению (SISR), но для задачи реконструкции изображений в ИДОС характерен дополнительный нелокальный тип искажений – цветовая засветка. Благодаря сходству задач критерии оценки качества реконструкции, архитектуры нейронных сетей и подходы к обучению в данной 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, визуально. Поскольку для реальных изображений отсутствует эталон, то такие изображения нельзя включить в процесс обучения нейронной сети. Для обучения нейронных сетей используются наборы данных, сформированные в лабораторных условиях на основе выведенных на экран монитора изображений. Лабораторный стенд для формирования обучающей выборки используется в работах [2, 7-11, 12-19, 22, 23, 31, 33, 34]. 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 некоторого критерия, который позволит оценить уровень 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 уровня 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 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 распределениях обучающего и тестового набора данных [35]. 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 [36] или на задачу доменной генерализации [37], где требуется обработка изображений из неизвестных новых доменов, не используемых в процессе обучения.

Для задачи сверхразрешения с неизвестной моделью искажений используются различные подходы, заключающиеся в обучении сети для получения точного представления деградации для конкретных изображений [38], адаптации универсальных нейросетевых моделей для конкретного типа искажений изображения [39], а также в моделировании искажений изображений реального мира на основе простых моделей, таких как размытие изображения, изменение размера, добавление шума [40]. Последний подход [40] послужил основой для предлагаемой в данной работе аугментации обучающего набора данных путем моделирования искажений, характерных для реальных условий формирования изображений, что позволило уменьшить уровень артефактов реконструкции на реальных изображениях.

Задаче доменной генерализации нейросетевых моделей за последние годы уделяется много внимания: доменная адаптация [41, 42], метаобучение [39, 43-44], трансферное обучение [45-47], обучение по нулевой выборке [48, 49]. Разработка критерия, позволяющего оценивать уровень артефактов реконструкции без эталона и включать реальные данные в процесс обучения сети, способна упростить задачу доменной генерализации, которая предполагает отсутствие какой-либо информации о новых доменах в обучающей выборке, до задачи доменной адаптации, которая использует информацию о новых доменах при обучении сети.

При обработке изображений реальных сцен также интерес представляют подходы, позволяющие адаптировать уже существующую обученную модель под новые условия формирования изображений по небольшому набору изображений из нового домена. К таким подходам относятся мелкошаговое метаобучение [39, 43, 44]

и обучение по нулевой выборке [48-49]. Впервые метод мелкошагового метаобучения предложен в работе [43]. В данной диссертационной работе представлено теоретическое обоснование модификации метода мелкошагового метаобучения, называемой крупношаговым метаобучением. Использование алгоритма адаптации модели на основе метода крупношагового метаобучения позволяет адаптировать модель к новым условиям формирования изображений по одному реальному изображению и упрощает процесс обучения сети.

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 видеоинформационного 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 РНФ 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». (Приложение А).

**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. Подлипновым, С.В. Ганчевской; ручная обработка изображений реальных сцен проводилась А. П. Алексеевым.

**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. Объём диссертации составляет 147 страниц, включая 79 изображений и 8 таблиц. Список литературы содержит 110 наименований.

# ГЛАВА 1. ПОСТАНОВКА ЗАДАЧИ ОЦЕНКИ И ПОВЫШЕНИЯ КАЧЕСТВА ИЗОБРАЖЕНИЙ В ДИФРАКЦИОННО-ОПТИЧЕСКИХ СИСТЕМАХ

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

## 1.1 Преимущества и недостатки ИДОС

В настоящее время все больший интерес у исследователей и потребителей вызывают миниатюрные камеры [1-4]. В случае с традиционной оптикой повышение числовой апертуры (разрешения) сопровождается увеличением числа оптических элементов объектива для компенсации присущих рефракционной оптике оптических aberrаций [50]. Для упрощения конструкции объективов в последнее время разрабатываются изображающие системы, в основе которых лежит одна плоская линза. Наиболее популярным подходом к проектированию плоской оптики является использование т.н. металинз [5], которые позволяют создать оптическую систему с субволновой толщиной [5]. Другим примером плоской оптики являются многоуровневые дифракционные линзы (МДЛ), которые стали использоваться для получения цветных изображений высокого разрешения только с 2015 года [12], в то время, как дифракционная линза известна как оптический элемент с 19 века. Несмотря на то, что металинзы могут быть тоньше, чем МДЛ, данная характеристика является незначительной перед преимуществами дифракционной оптики. МДЛ являются более дешевым вариантом плоской оптики при их массовом производстве, они просты в изготовлении и эксплуатации, позволяют создавать уникальные мультиапертурные конфигурации, иногда позволяя сократить габариты оптической системы. Кроме того, по сравнению с металинзами МДЛ обеспечивают лучшее качество изображения.

Использование плоской оптики позволяет значительно сократить вес и упростить конструкцию оптической системы, что является большим преимуществом перед классическими рефракционными аналогами. Особенно

заметно такое преимущество перед рефракционной оптикой в многолинзовых системах [104, 105].

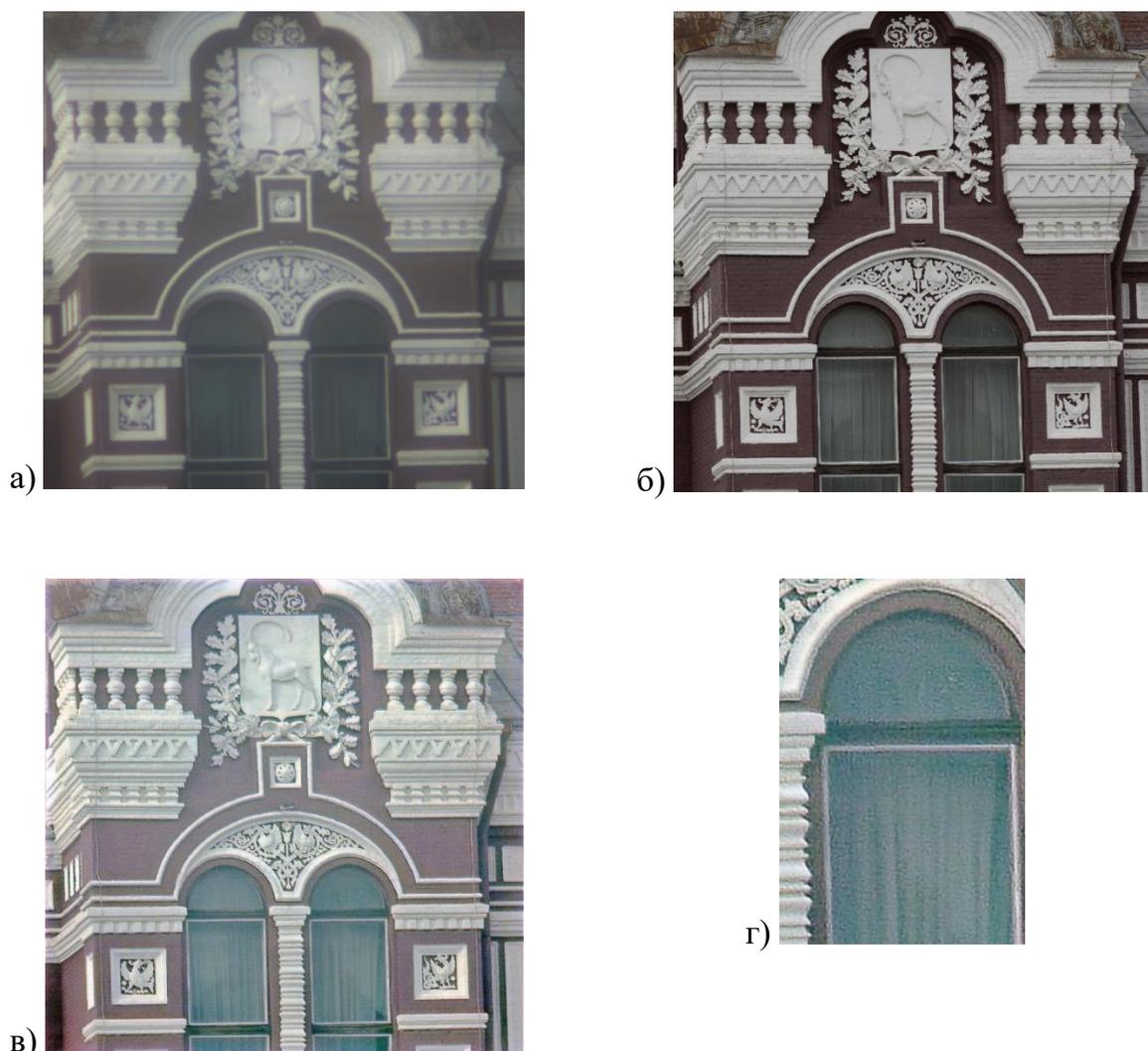
Кроме того, возможно использование гибридных рефракционно-дифракционных элементов, которые применяются в качестве искусственных хрусталиков глаза [51]. Если использовать в качестве дифракционной структуры бинарную линзу, то она выступает как линза и с положительной, и с отрицательной оптической силой, что приводит к возникновению на оптической оси двух разделенных фокусов, между которыми резкость изображения примерно одинакова. Таким образом, исключается необходимость настройки фокусного расстояния искусственного хрусталика.

Основной проблемой как дифракционной оптики, так и металинз являются сильные искажения регистрируемых изображений. В отличие от традиционных оптических систем для плоской оптики характерно большое количество продольных хроматических аберраций. В работах [5, 52] авторы предлагают устранять хроматические искажения при помощи дополнительных оптических элементов. В работе [53] описывается создание на поверхности дифракционной линзы комбинации зон, которые компенсируют хроматические искажения, за счет чего достигается высокий уровень PSNR (27,1 дБ), но только для синтезированных изображений. В работе [12] впервые представлена комбинация многоуровневой линзы Френеля и вычислительной постобработки. Вычислительная реконструкция помогает компенсировать сильные хроматические аберрации и другие искажения, характерные для дифракционных линз и плоских оптических элементов.

В рамках модели искажений, представленной в работах [2, 12], хроматические аберрации вызывают на формируемых искажениях два типа искажений. Первый тип искажений – хроматическое размытие, в результате которого пятно рассеяния самого резкого цветового канала изображения существенно меньше относительно двух других каналов [2, 12]. Как правило, самым резким каналом является зеленый и большинство алгоритмов реконструкции используют межканальную связь для повышения резкости контуров красного и синего каналов на основе информации о зелёном канале [2, 3]. Вторым типом искажений является хроматический сдвиг или цветовая засветка, причиной которой является перераспределение энергии между дифракционными порядками [7].

Дополнительным источником искажений являются артефакты реконструкции, возникающие в результате постобработки изображения из-за различий в условиях формирования обучающего набора данных и обрабатываемого изображения [33], таких как разные уровни экспозиции и значений усиления камеры, приводящих к разному уровню шума на изображении.

На рисунке 1.1 представлен пример изображения (рисунок 1.1.а), полученного на основе дифракционно-рефракционной линзы, в сравнении с изображением, сформированным рефракционным объективом (рисунок 1.1.б).



*Рисунок 1.1 – Пример изображения реальной сцены: а) снятого дифракционно-рефракционным объективом, б) снятого рефракционным объективом, в) результат реконструкции изображения, снятого дифракционно-рефракционным объективом, г) фрагмент восстановленного изображения в), демонстрирующий появление артефактов реконструкции*

Также приведен пример реконструкции изображения (рисунок 1.1.в), на фрагменте которого (рисунок 1.1.г) наблюдаются артефакты реконструкции.

## 1.2 Формирование изображений в дифракционно-оптических системах

Многоуровневая дифракционная линза (МДЛ или гармоническая линза) – дифракционная линза, оптический путь которой равен целому количеству  $m$  базовых длин волн  $\lambda_0$  [106]. Высота МДЛ, изготовленной из материала с показателем преломления  $n$ , вычисляется по формуле:

$$h = \frac{m}{n-1} \lambda_0, \quad (1.1)$$

где  $\lambda_0$  – длина волны освещающего МДЛ источника (базовая длина волны),  $m = 2, 3, 4, \dots$  – мультипликатор высоты микрорельефа [106]. Высота обычной дифракционной линзы рассчитывается по формуле (1.1) при  $m = 1$ .

МДЛ обладают свойствами как дифракционных, так и рефракционных линз. Чем больше высота рельефа гармонической линзы, тем ярче выражены рефракционные свойства.

При использовании одной гармонической линзы для регистрации изображения расстояние до фокальной плоскости зависит от длины волны света [106, 107]:

$$f(\lambda, k) = \frac{m\lambda_0 f_0}{\lambda k}, \quad (1.2)$$

где  $m$  – множитель для высоты микрорельефа гармонической линзы,  $\lambda_0$  – базовая длина волны,  $\lambda$  – текущая длина волны,  $k$  – дифракционный порядок,  $f_0$  – расстояние до фокальной плоскости для длины волны  $\lambda_0$ .

В таком случае эффективность формирования фокального пятна описывается следующим уравнением:

$$e(\lambda, k) = \text{sinc}^2 \left( \frac{m\lambda_0 f_0}{\lambda} - k \right). \quad (1.3)$$

Согласно описанным свойствам гармонической линзы, всегда можно выбрать несколько длин волн с одинаковым фокусным расстоянием. На рисунке 1.2 представлено распределение интенсивности в фокусе на оптической оси  $z$  для гармонической линзы с параметрами:  $m=10$ ,  $\lambda_0=0,65$  мкм,  $f_0=100$  мм. Вычисления были выполнены в рамках работы [12] для интервала от  $\lambda_{min} = 0,38$  мкм до  $\lambda_{max}=0,74$

мкм, в результате определены три длины волны с одинаковым фокусным расстоянием. На рисунке 1.2 представлен график для  $\lambda_1=0,406$  мкм (синий),  $\lambda_2=0,500$  мкм (зеленый),  $\lambda_3=0,650$  мкм (черный). Для данных длин волн фокусное расстояние оказалось равным  $f=100,15$  мм.

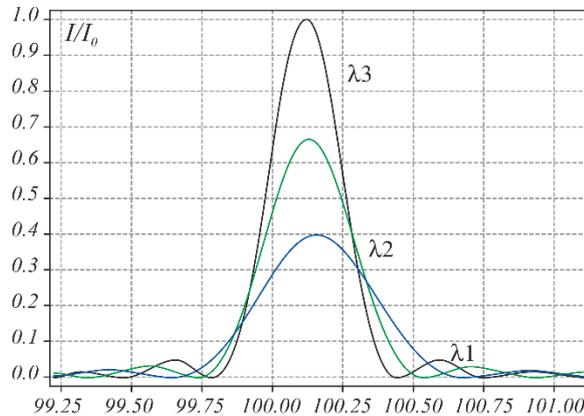


Рисунок 1.2 – График интенсивности фокусировки на оси  $z$  для гармонической линзы, где  $m=10$ ,  $\lambda_0=0,65$  мкм,  $f_0=100$  мм,  $\lambda_1=0,406$  мкм – синий,  $\lambda_2=0,500$  мкм – зеленый,  $\lambda_3=0,650$  мкм – черный

Согласно (1.2), для разных длин волн фокусное расстояние изменяется практически линейно, что приводит к возникновению хроматической aberrации. В каждом канале хроматическая aberrация вызывает размытие, которое может быть описано при помощи функции рассеяния точки (ФРТ), иногда называемой ядром размытия [12]. Для некоторых длин волн данная проблема проявляется в виде хроматического размытия на контурах изображения. Однако для других длин волн видимого света фокус будет сдвигаться сильнее, что приводит к цветовой засветке на значительной части изображения, которая в работе [12] называется хроматическим сдвигом. Изменение  $\Delta D$  оптической силы  $D$  гармонической линзы определяется следующим уравнением:

$$\frac{\Delta D}{D} \approx \frac{1}{m}. \quad (1.4)$$

Площадь искажений определяется шириной ФРТ:

$$W = \frac{\Delta f}{2k}. \quad (1.5)$$

где  $k$  – порядок дифракции, а интенсивность определяется соотношением (1.3).

На рисунке 1.3 приведены примеры распределения интенсивности в фокусе для различных длин волн. На рисунке 1.3 показаны результаты расчетов для относительно узкого диапазона длин волн. Гладкие синие и красные кривые являются огибающими распределения интенсивности в фокусе для порядков дифракции 10 и 9 соответственно. На рисунке 1.3 видно, что расположение фокусов достаточно сильно изменяется в диапазоне от 90 до 110 нм. Более того некоторые длины волн имеют два фокуса. Длина волны 0,65 мкм попадает на огибающую для  $k = 10$  (синяя линия). Величина продольной хроматической aberrации для такой линзы составляет около 10 мм. Для линзы с апертурой 10 мм смещение фокуса  $\Delta f$  составляет 10 мм, при этом ширина пятна в фокальной плоскости составляет порядка 1 мм или 200 пикселей при размере пиксела 5 мкм. Размер засветки пятна фокусировки в 200 пикселей является довольно большим, поэтому пятно будет иметь низкую интенсивность, что приведет к образованию фоновой засветки.

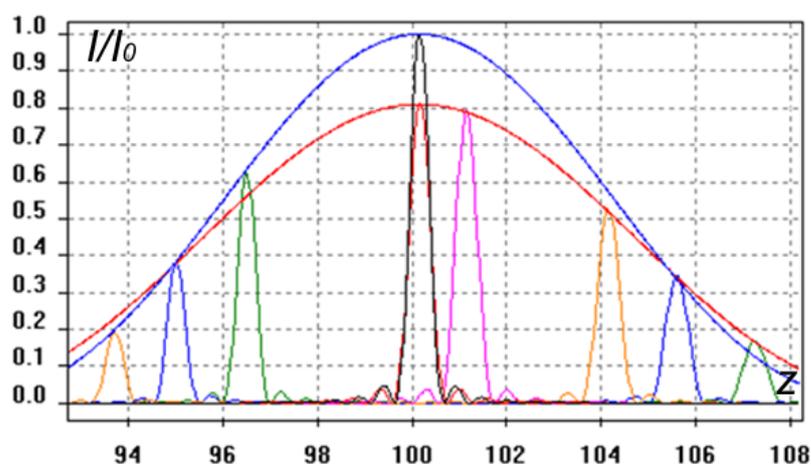


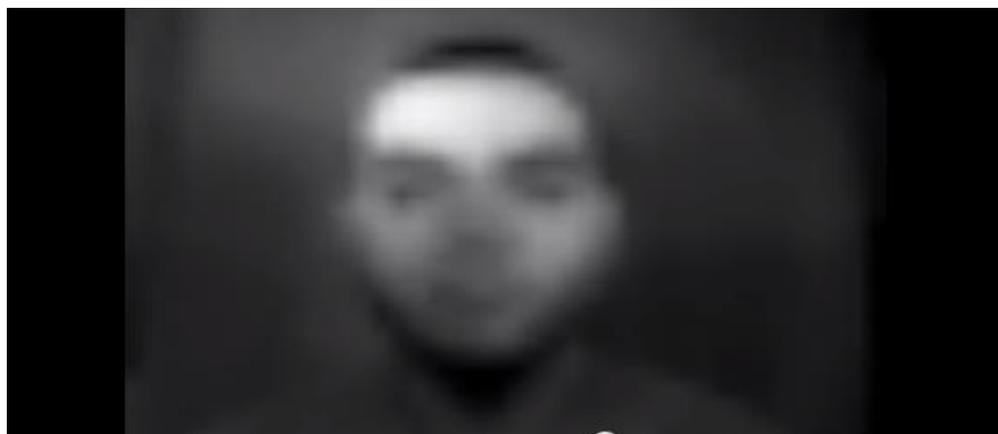
Рисунок 1.3 – График распределения интенсивности в фокусе на оси  $z$  для гармонической линзы с параметрами  $m = 10$ ,  $\lambda_0 = 0,65$  мкм,  $f_0 = 100$  мм,  $\lambda_1 = 0,65$  мкм – черный,  $\lambda_2 = 0,675$  мкм – зеленый,  $\lambda_3 = 0,685$  мкм – синий,  $\lambda_4 = 0,695$  мкм – желтый,  $\lambda_5 = 0,715$  мкм – фиолетовый

### 1.3 Компенсация искажений изображения за счет совершенствования оптической схемы и различные схемы получения изображений

#### 1.3.1 Гармоническая линза

Впервые использование дифракционной линзы Френеля в изображающих оптических системах описано в работе 2015 года [28]. Конструкция

энергонезависимой камеры, предложенной в [28], включает бинарную линзу Френеля. На рисунке 1.4 представлен пример изображения, полученного с помощью предложенного прототипа камеры и бинарной линзы Френеля. Как видно из рисунка 1.4, качество изображения практически не пригодно для прикладного использования.

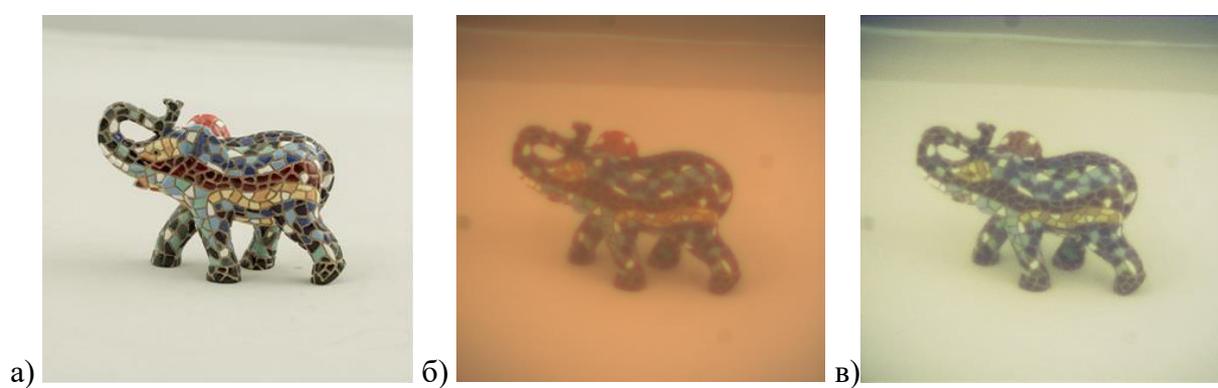


*Рисунок 1.4 – Изображение, снятое энергонезависимой камерой в работе [27]*

Позднее в 2015 году была опубликована работа [12], где предложены 4-уровневая дифракционная линза (или обычная дифракционная линза) и алгоритм постобработки регистрируемых изображений, открывшие перспективы использования дифракционно-оптических элементов (ДОЭ) для получения изображений потребительского качества. Алгоритм постобработки используется для компенсации искажений, характерных для дифракционной оптики. На рисунке 1.5 приведены изображения одной сцены, полученные с помощью рефракционного объектива и обычной дифракционной линзы, а также результат постобработки второго изображения на основе алгоритма, предложенного в работе [12]. Результат использования комбинации четырехуровневой дифракционной линзы и алгоритма реконструкции все еще уступает по качеству изображениям, снятым рефракционным объективом, но по сравнению с бинарной линзой Френеля качество снимков существенно улучшилось.

В 2017 году авторами работ [2, 13] была предложена 256-уровневая гармоническая линза, которая обеспечивала лучшее качество изображения, чем ранее предложенная 4-х уровневая линза. Гармоническая линза позволила уменьшить хроматические aberrации как за счет увеличения числа уровней квантования (256 уровней), так и за счет использования дифракционного рельефа

повышенной высоты. Если в обычной дифракционной линзе, рассчитанной на длину волны 550 нм, высота микрорельефа составляет около 1 мкм, то в рамках работы [2] удалось получить микрорельеф высотой до 8 мкм. На рисунке 1.6 представлено сравнение обычной и гармонической дифракционных линз. Микрорельеф гармонической линзы был изготовлен с помощью двухлазерной круговой записывающей системы (TLCWS) [54], которая позволяет записывать многоуровневый микрорельеф с 256 уровнями квантования на поверхности фоторезиста. Ранее такая система использовалась только для записи ДОО в неизображающих системах или изображающих системах низкого разрешения, предназначенных для работы с монохроматическим когерентным светом.



*Рисунок 1.5 – Формирование изображения рефракционным объективом (а), формирование изображения четырехуровневой дифракционной линзой (б), восстановленное изображение (в)*

Гармонические линзы также называют многоуровневыми дифракционными линзами или МДЛ (MDL, multilevel diffractive lenses) [7, 14]. При совершенствовании рельефа линзы в работах [7, 14] указывается мультипликатор высоты микрорельефа линзы вместо количества уровней квантования на поверхности фоторезиста (4 или 256). Так, в работе [14] используются МДЛ при  $m=7$  и  $m=13$ . На рисунке 1.7 приведено сравнение изображений, полученных с использованием МДЛ ( $m=13$ ) и МДЛ ( $m=7$ ). Чем меньше значение  $m$ , тем меньше высота микрорельефа и тем ниже качество формируемых изображений.

С целью улучшить качество формируемого изображения в работе [7] впервые использовалась бинокулярная оптическая система на основе двух МДЛ. Ранее использовалась только монокулярная система на основе одной МДЛ. На рисунке 1.8.a приведена оптическая схема объектива на основе МДЛ для

монокулярной дифракционной оптической системы, а на рисунке 1.8.б – для бинокулярной системы.

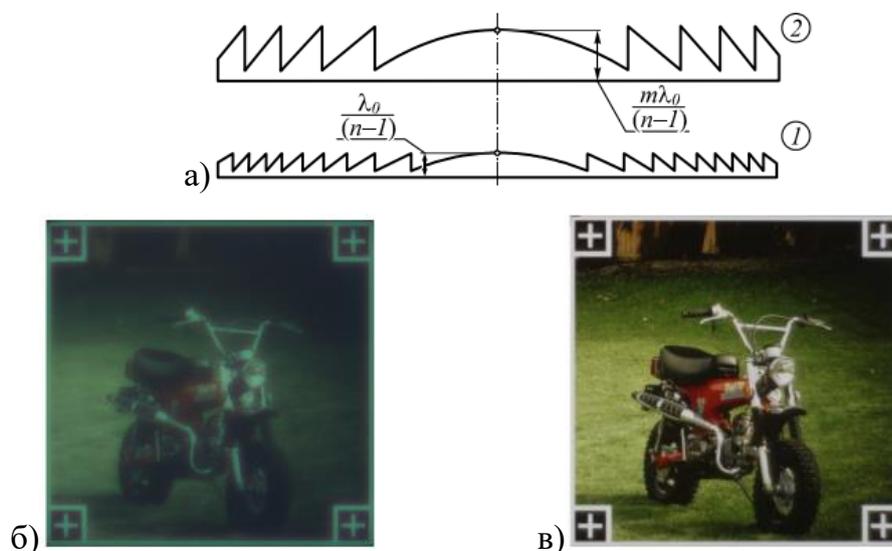


Рисунок 1.6 – Сравнение высоты микрорельефа обычной и дифракционной линз (а), результат формирования изображения с использованием обычной дифракционной линзы (б), результат формирования изображения с использованием гармонической линзы (256 уровневой) (в)



Рисунок 1.7 – Изображение, снятое МДЛ13 (а), изображение, снятое МДЛ 7 (б), эталонное изображение (в)

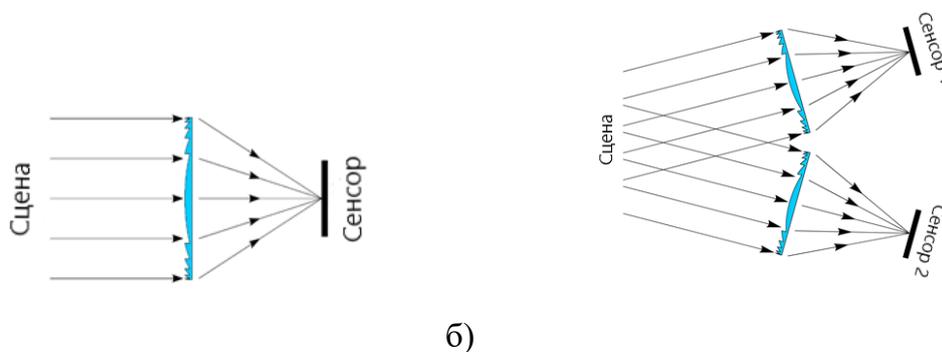


Рисунок 1.8 – Схема монокулярного объектива на основе МДЛ (а), оптическая схема бинокулярной системы на основе МДЛ (б)

Монокулярная оптическая система построена на базе одной МДЛ с фокусным расстоянием 150 мм и диаметром линзы 50 мм. Изготовление основано на технологии прямой лазерной записи [2, 13]. Бинокулярная система объединяет в себе две линзы, которые зафиксированы на расстоянии 40 см друг от друга для формирования изображения одной сцены с разных углов зрения. В обоих случаях фокусировка объектива настраивается вручную.

### 1.3.2 Изображающая система из трех дифракционных линз

В работе [8] предложена новая изображающая система из трех дифракционных линз, каждая из которых формирует один канал цветного RGB изображения. Такой подход позволяет существенно сузить спектральный диапазон каждой линзы, и таким образом существенно снизить искажения изображения, вызванные хроматической аберрацией, присущей дифракционной оптике. На рисунке 1.9 представлена оптическая схема такой мультиапертурной системы, которая состоит из трех дифракционных линз с фокусным расстоянием 100 мм, каждая из которых рассчитана на свой спектральный диапазон.

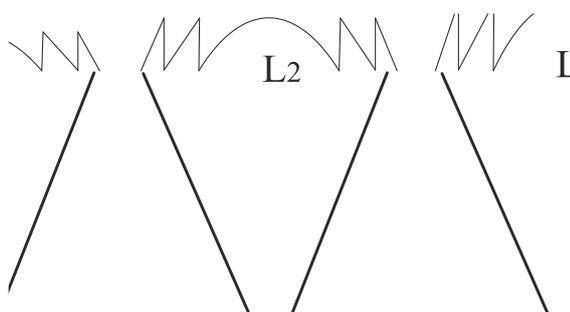


Рисунок 1.9 – Оптическая схема мультиапертурной системы

Рабочая длина волны каждой дифракционной линзы задается высотой микрорельефа. Для формирования изображения на длине волны  $\lambda$  высота микрорельефа определяется по формуле

$$h = \frac{\lambda}{n - 1}, \quad (1.6)$$

где  $n$  – показатель преломления материала линзы.

Для нормального режима работы были выбраны длины волн 450 нм, 550 нм и 650 нм, равномерно покрывающие видимый диапазон. При расчете были использованы значения показателя преломления резиста, полученные в результате измерения на эллипсометре.

Для длины волны 450 нм в расчете  $n=1,68$ , расчетная высота микрорельефа составила 660 нм, для длины волны 550 нм в расчете  $n = 1,64$ , расчетная высота микрорельефа составила 860 нм, для длины волны 650 нм в расчете  $n = 1,63$ , расчетная высота микрорельефа составила 1030 нм. Радиальные сечения микрорельефа для линз на эти длины волн были рассчитаны для диаметра 10 мм. По расчетным микрорельефам методом прямой лазерной записи на фоторезисте ФП-40 были изготовлены дифракционные линзы.

Прототип трехлинзовой системы представляет собой три объектива и три USB 3.0 камеры Basler acA1920-40uc, закрепленные на единой площадке (Рисунок 1.10.а). Юстировка каждого объектива после ручной фокусировки осуществляется с помощью трех подпружиненных регулировочных винтов таким образом, чтобы каждый из объективов смотрел приблизительно в одну и ту же точку. Пример сформированного изображения, представлен на рисунке 1.10.б.

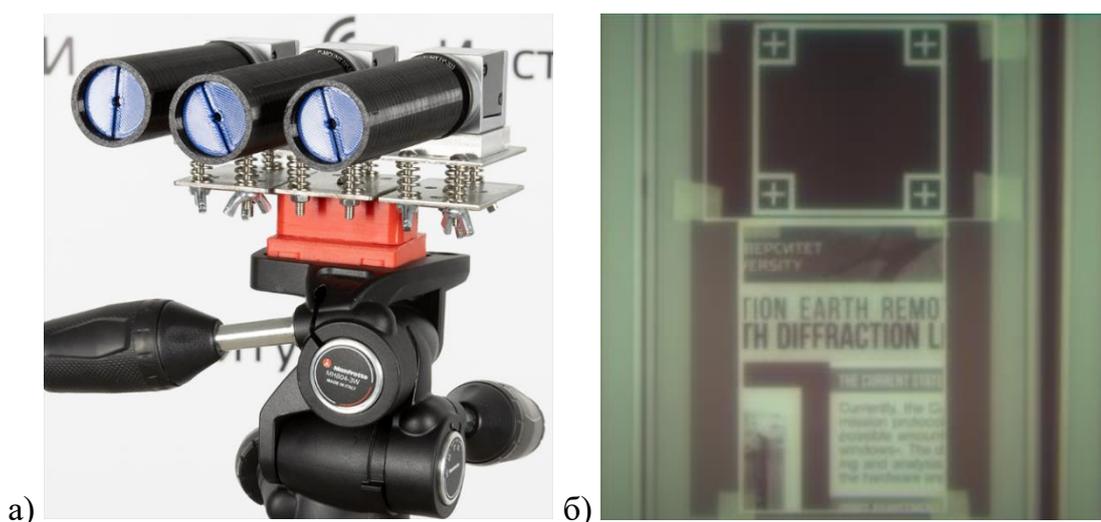
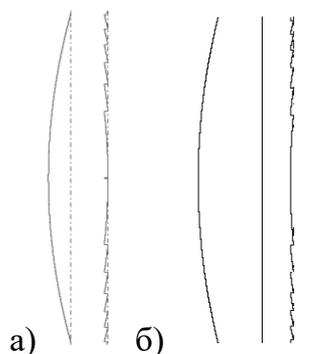


Рисунок 1.10 – Прототип оптической системы на основе трех дифракционных линз (а), пример сформированного изображения (б)

### 1.3.3 Дифракционно-рефракционная оптическая система

В работе [33] разработана гибридная, дифракционно-рефракционная оптическая схема. Для создания такой схемы использовался подход к расчету дифракционно-рефракционной системы, позволяющий одновременно компенсировать как геометрические, так и хроматические aberrации. Использование дублета (рисунок 1.11.а) из рефракционной и дифракционной линзы позволяет при правильном подборе параметров полностью исключить

хроматическую aberrацию на двух длинах волн и сильно уменьшить ее в интервале между этими длинами волн [55]. Поскольку изготовление линзы на рисунке 1.11.а является слишком сложным и дорогостоящим, изготовлена система, конструкция которой представлена на рисунке 1.11.б. Расстояние между оптическими элементами составляет 1 мм.



*Рисунок 1.11 – Оптическая схема дублета из рефракционной и дифракционной линз: а) одна линза с рефракционной и дифракционной поверхностями, б) система из плоско-выпуклой линзы и дифракционной линзы на плоской подложке*

На рисунке 1.12 представлен внешний вид изготовленного экспериментального образца изображающей системы на основе рефракционной и дифракционной линз.

Исторически гибридная оптическая система рассматривается исключительно как непригодная к практическому использованию модель, поскольку при использовании стандартной сферической рефракционной линзы и стандартной дифракционной линзы, аппроксимирующей подобную же сферическую линзу, получается система с существенными геометрическими aberrациями. Поэтому дифракционная оптика рассматривается в составе достаточно сложных оптических систем [56, 57] лишь как элемент компенсации хроматической aberrации, или хроматической aberrации и сферической aberrации [58]. Однако таким использованием возможности дифракционной оптики не ограничиваются. Работа [59], где дифракционный элемент включен в оптическую конструкцию, интересна тем, что в данном случае пришлось использовать сложный преломляющий элемент произвольной формы для компенсации геометрических aberrаций. Работа [59] является примером того, как предвзятость знаний, в данном случае экспертных знаний авторов в области рефракционной оптики, может привести к упущению из виду богатых возможностей хорошо спроектированного

дифракционного элемента. Благодаря опыту авторов работы [33] в области дифракционной оптики известно, что один дифракционный элемент способен аппроксимировать произвольную асферическую поверхность, какой бы сложной она ни была. Теоретически, поскольку каждая отдельная зона дифракции может рассматриваться как параметр оптимизации, дифракционный элемент может быть сконструирован таким образом, чтобы компенсировать как геометрические, так и хроматические аберрации.



Рисунок 1.12 – Внешний вид рефракционно-дифракционного дублета

#### 1.4 Вычислительный способ компенсации искажений

В рамках данной диссертационной работы исследуется повышение качества изображений за счет вычислительной постобработки. В зависимости от качества изображений, формируемых оптической системой, используются разные алгоритмы компенсации искажений.

Впервые в работе [12] для восстановления изображений, регистрируемых обычной дифракционной линзой, применяются алгоритмы процедурной коррекции. Процедурная коррекция основана на модели искажений [12], которая описывается как комбинация двух типов искажений: хроматической аберрации, которая преобразует исходное изображение  $p_{RGB}^0(\mathbf{x})$  в размытое  $p_{RGB}^B(\mathbf{x})$  (хроматическое размытие), и хроматического сдвига, который в свою очередь преобразует размытое изображение  $p_{RGB}^B(\mathbf{x})$  в  $p_{RGB}^{B,S}(\mathbf{x})$ . Модель искажений принимает вид:

$$p_{RGB}^B(\mathbf{x}) = \mathbf{B}_{RGB} \otimes p_{RGB}^0(\mathbf{x}) + \xi, \quad (1.7)$$

$$p_{RGB}^{B,S}(\mathbf{x}) = S_{RGB}(p_R^B(\mathbf{x}), p_G^B(\mathbf{x}), p_B^B(\mathbf{x})), \quad (1.8)$$

где  $p_{RGB}(\mathbf{x})$  – один из цветовых каналов исходного изображения,  $p_{RGB}^{B,S}(\mathbf{x})$  – канал изображения, полученного при помощи ДОЭ,  $\mathbf{V}_{RGB}$  – функция размытия точки для каждого канала,  $S_{RGB}(p_R^B(\mathbf{x}), p_G^B(\mathbf{x}), p_B^B(\mathbf{x}))$  – хроматический сдвиг,  $\xi$  – аддитивный шум.

Модель искажений (1.7)-(1.8) предполагает, что размытие происходит перед хроматическим сдвигом, таким образом нелинейная цветовая коррекция применяется перед сверткой с ФРТ.

Процедурная коррекция, предложенная в работе [12] включает 3 этапа:

1. Применение алгоритма обратной свертки к зеленому каналу для устранения хроматического размытия изображения. Зеленый канал, как правило, является более четким.
2. Устранение хроматического размытия в синем и красном каналах на основе анализа контуров в зеленом канале. Используется гипотеза об ахроматичности контуров [60].
3. Цветовая коррекция изображения для устранения хроматического сдвига.

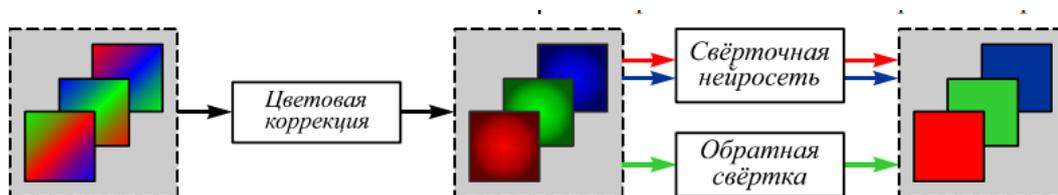
Подходы к реконструкции изображений, используемые в работах [3, 53, 61], схожи между собой и основаны на методах обратной свертки с регуляризацией [62, 63]. Подход из работы [53] основан на регуляризованной обратной свертке с межканальной связью, которая учитывает линейную модель искажений. Для дифракционной линзы один из цветовых каналов зарегистрированного изображения (обычно зелёный канал) менее подвержен размытию, чем другие каналы. В работе [61] показано, что обратная свертка эффективна только для более четких каналов, в то время как другие методы нелинейного преобразования, например, повышение резкости (*sharpening*), могут применяться для более размытых цветовых каналов.

Альтернативным подходом к решению задачи реконструкции изображений являются методы машинного обучения. Так, в работе [3] используется предложенный в [53] подход компенсации присущих обратной свертке артефактов на основе многослойного перцептрона. В работе [64] однако отмечено, что решить задачу восстановления расфокусированного изображения непосредственно полносвязной нейронной сетью без использования обратной свертки не удастся.

Начиная с 2014 года, методы глубокого обучения применяются для решения задачи сверхразрешения по одному изображению, которая является смежной задачей реконструкции. Так, в работе [65] описана архитектура SRCNN, по результатам превосходящая традиционные подходы к задаче сверхразрешения. Предложенная в 2016 году [66] нейронная сеть VDSR позволила достичь лучшего качества изображения и оказалась более стабильной при обучении. Развитие такого подхода для получения изображений различного масштаба описано в [67].

В работе [2] впервые предложено использование сверточной нейронной сети (СНС), которая была встроена в схему реконструкции, описанную в работе [12]. Реконструкция изображения, снятого гармонической линзой, осуществлялась в три этапа (рисунок 1.13):

1. Цветовая коррекция всех каналов RGB изображения.
2. Применение обратной свертки для устранения хроматического размытия в зеленом канале.
3. Устранение хроматического размытия красного и синего каналов с помощью СНС.



*Рисунок 1.13 – Схема реконструкции изображения с использованием СНС, предложенная в работе [2]*

В качестве СНС использовалась модификация сети VDSR [66]. Использование в работе [12] межканальной связи, основанной на гипотезе об ахроматичности контуров [60], вдохновило авторов работы [2] на использование межканальной связи в качестве штрафа за отклонение контура от ахроматичности (grey-edge penalty) как дополнительное слагаемое в функции потерь СНС. Стоит отметить, что в описанных выше работах каналы изображения обрабатываются отдельно, в том числе на вход СНС подается только один канал изображения.

В работе [2] показан потенциал СНС для задачи реконструкции изображений. Использование сети VDSR позволило улучшить качество изображений по сравнению с предыдущими методами на основе обратной свертки.

Работы [31, 32] также опираются на преимущества нейросетевой реконструкции перед алгоритмами обратной свертки для изображений с нелокальными искажениями. В работах [31, 32] исследуется комбинация ДОС и нейросетевой реконструкции на основе U-Net-подобной архитектуре. В работе [31] прирост среднего значения PSNR составил 4,61 дБ (входной PSNR – 21,28 дБ, выходной PSNR – 25,89 дБ). В работе [32] обучение нейронной сети проводилось для подбора параметров ДОС на синтезированных данных.

В ранних исследованиях при реконструкции изображений в дифракционно-оптических системах использовалось упрощение модели (1.7)-(1.8). В работах [29, 66] учитывалось только хроматическое размытие (1.7). В работах [13, 2] решалась дополнительно задача цветовой коррекции засветки на основе трехмерных таблиц поиска. Позднее авторами работы [7] была предложена модификация модели (1.6)-(1.7), учитывающая контекстно-зависимые (нелокальные) искажения. Согласно модели [7] цветовая засветка является нелокальным искажением и зависит от контекста изображения, поэтому модель сформированного изображения принимает следующий вид:

$$p_{RGB}(\mathbf{x}) = \mathbf{B}_{RGB} \otimes p_{RGB}^0(\mathbf{x}) + H(P, p_{RGB}^0(\mathbf{x}), \mathbf{x}) + \xi, \quad (1.9)$$

где  $p_{RGB}(\mathbf{x})$  – красный, зеленый или синий канал изображения, полученного с помощью МДЛ,  $p_{RGB}^0(\mathbf{x})$  – исходное эталонное изображение,  $\mathbf{B}_{RGB}$  – ФРТ для каждого цветового канала,  $H(P, p_{RGB}^0(\mathbf{x}), \mathbf{x})$  – цветовая засветка,  $\xi$  – аддитивный шум.

Необходимо отметить, что цветовая засветка зависит от большей части изображения, то есть контекст изображения влияет на эффект засветки в любой заданной точке. Таким образом,  $H(P, p_{RGB}^0(\mathbf{x}), \mathbf{x})$  – контекстно-зависимые искажения, где зависимость от всего изображения указывает на нелокальный характер таких искажений.

В рамках данной диссертационной работы предлагается полностью нейросетевой подход к реконструкции изображений в виде двухэтапной схемы сквозной нейросетевой реконструкции для дифракционных оптических систем, что является первым пунктом новизны данной работы. Результаты исследования двухэтапной схемы опубликованы в работе [7], где вводятся понятия локальной (patch-wise) и полнокадровой (image-wise) реконструкции. Для компенсации

хроматического сдвига применяется полнокадровая нейронная сеть, при обучении которой используются полноразмерные изображения. Устранение хроматического размытия проводится так называемой локальной сетью, при обучении которой использовались фрагменты изображений из обучающей выборки размера  $41 \times 41$  пикселей. Полнокадровая и локальные сети различаются размером рецептивного поля. Как будет показано в главе 2, полнокадровая сеть имеет большее рецептивное поле, чем локальная сеть. Предложенная схема реконструкции применима для всех оптических систем, описанных в разделе 1.3. Более детальное описание двухэтапной схемы реконструкции будет представлено в главе 2.

Двухэтапная схема сквозной нейросетевой реконструкции показала значительный прирост качества реконструкции согласно критерию PSNR на тестовых полуреальных лабораторных данных. Однако при обработке изображений реальных сцен, полученных как при искусственном, так и при естественном освещении появляются артефакты реконструкции. Для минимизации артефактов реконструкции в данной диссертационной работе предлагается включить фрагмент реального изображения в процесс обучения сети. В процессе обучения проводится оценка доли ложных контуров на данном фрагменте на основе критерия FEL (false-edge level), которая является третьим пунктом новизны данной работы. Как показали результаты исследований, опубликованные в [33], использование FEL компенсирует недостатки обучения сети на полуреальных лабораторных данных. Детально критерий FEL будет описан в главе 3.

Артефакты, возникающие при реконструкции реальных изображений в дифракционно-оптических системах связаны с различиями в условиях формирования обучающего набора данных и изображений реальных сцен. Экспозиция и динамический диапазон на изображениях реальных сцен значительно отличаются от результата формирования изображения с экрана, так как средняя освещенность во втором случае составляет 200 люкс. Кроме того, причинами артефактов реконструкции могут быть усиление (gain) или светочувствительность (ISO) камеры, исследуемые в работах [8, 14, 33, 68]. В данной диссертационной работе для устранения артефактов реконструкции предлагается аугментация обучающей выборки путем моделирования двух типов искажений, характерных для изображений реальных сцен: моделирование шума, причиной которого является

усиление или светочувствительность камеры (ISO шум), и изменение экспозиции изображения. Результаты исследования алгоритмов аугментации обучающего набора данных, относящихся к четвертому пункту новизны данной работы, опубликованы в [33, 14]. Подробнее алгоритмы аугментации обучающего набора будут описаны в главе 3.

В данной диссертационной работе также проводится адаптация универсальных нейросетевых моделей для конкретного типа искажений изображения, основанная на крупношаговом метаобучении. Подход на основе мелкошагового метаобучения использовался для задачи сверхразрешения по одному изображению (SISR) с неизвестной моделью искажений. Теоретическое обоснование метода крупношагового метаобучения для адаптации нейросетевой модели реконструкции по нулевой выборке является вторым пунктом новизны данной работы. Результаты исследования описаны в работах [34, 69]. Описание данного подхода представлено в главе 3.

Важно отметить, что разные технологии изготовления дифракционной оптики приводят к разному качеству изображений и требуют разных подходов к реконструкции изображений как процедурного, так и нейросетевого. В работе [15] показано, что при реконструкции изображения, полученного дифракционным объективом на борту наноспутника CubeSat, используются и нейросетевые, и процедурные методы реконструкции. Авторы работы [70] считают, что трудно получить нейросетевую модель с хорошей обобщающей способностью при наличии механических погрешностей при производстве камер, поэтому используют постобработку на основе расширенной многомерной динамической свертки, которая адаптирована к отклонениям камеры от модели, возникающим при производстве.

### **1.5 Оценка качества изображений в ИДОС**

В разделе 1.3 показано многообразие оптических схем изображающих систем на основе дифракционной оптики, а в разделе 1.4 представлены подходы к вычислительной реконструкции с целью повышения качества формируемых изображений. Таким образом, введение критерия и процедур оценки качества формируемых изображений является ключевым для настоящей работы. Как было сказано ранее предлагаемый подход к компенсации искажений в ДОС близок к

задаче повышения разрешения по одному изображению, поэтому предлагается использовать принятые в данной области критерии PSNR (пиковое отношение сигнала к шуму) и SSIM (индекс структурного сходства) для оценки качества изображений.

### 1.5.1 Пиковое отношение сигнала к шуму

Поскольку оценка качества реконструкции изображений в дифракционных оптических системах на основе PSNR хорошо зарекомендовала себя в работах [2, 7, 12, 13, 61], в данной работе критерий PSNR считается базовым.

PSNR – пиковое отношение сигнала к шуму (англ. peak signal-to-noise ratio). Данный показатель определяет соотношение между максимумом возможного значения сигнала и мощностью шума, искажающего значения сигнала. Многие сигналы имеют широкий динамический диапазон, PSNR обычно измеряется в логарифмической шкале в децибелах (дБ).

PSNR можно определить через среднеквадратичное отклонение (СКО, англ. root-mean-square error):

$$RMSE = \sqrt{\frac{1}{mn} \sum_{i=0}^{m-1} \sum_{j=0}^{n-1} |I(i, j) - K(i, j)|^2}, \quad (1.10)$$

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

Тогда PSNR определяется следующим соотношением:

$$PSNR = 10 \log_{10} \left( \frac{MAX_I^2}{RMSE^2} \right) = 20 \log_{10} \left( \frac{MAX_I}{RMSE} \right), \quad (1.11)$$

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

Для цветных изображений с тремя компонентами RGB на пиксель применяется такое же определение PSNR, но RMSE считается по всем трем компонентам (и делится на утроенный размер изображения).

### 1.5.2 Лабораторный стенд для формирования обучающего набора данных

Расчет PSNR требует сопоставления изображения с эталоном. Для формирования обучающих наборов данных, используемых в данной работе, использовался лабораторный стенд, представленный на рисунке 1.14. Подобный

лабораторный стенд также применяется командой Стэнфордского университета [31].

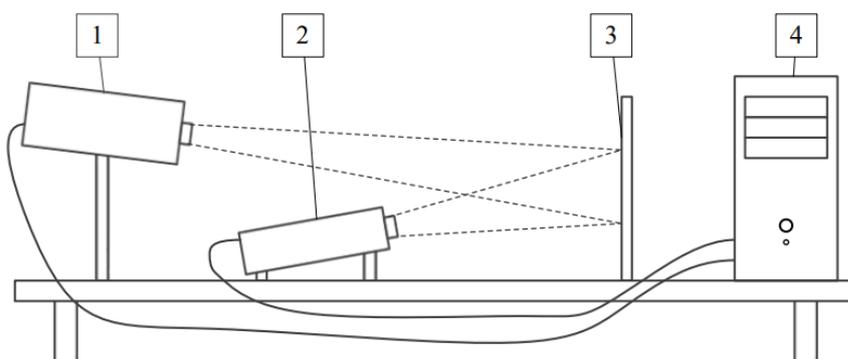


Рисунок 1.14 – Схема лабораторного стенда для формирования набора данных: 1 – камера, 2 – проектор, 3 – экран, 4 – компьютер

Изображения из заранее подготовленного набора выводятся на экран с помощью проектора, расположенного на расстоянии 6 м от экрана. Изображение, выводимое на экран, считается эталонным. Далее осуществляется формирование эталонных изображений камерой на основе МДЛ. После формирования изображений выполняется процедура сопоставления сформированного с использованием ДОС изображения с эталоном, выводимым на экран. Сформированные МДЛ изображения обладают проективными искажениями. Согласно подходу, описанному в [61], коррекция таких искажений осуществляется обратным проективным преобразованием. Для этого на исходное изображение необходимо добавить служебную информацию – маркеры (рисунок 1.15).

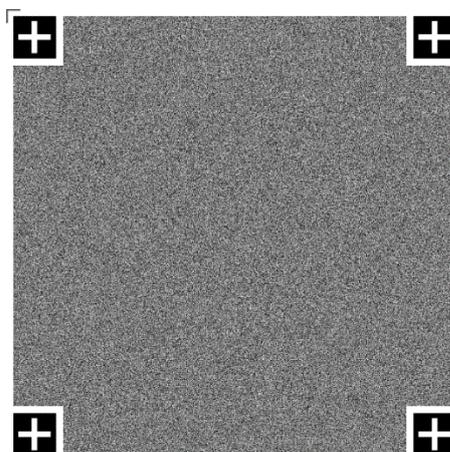
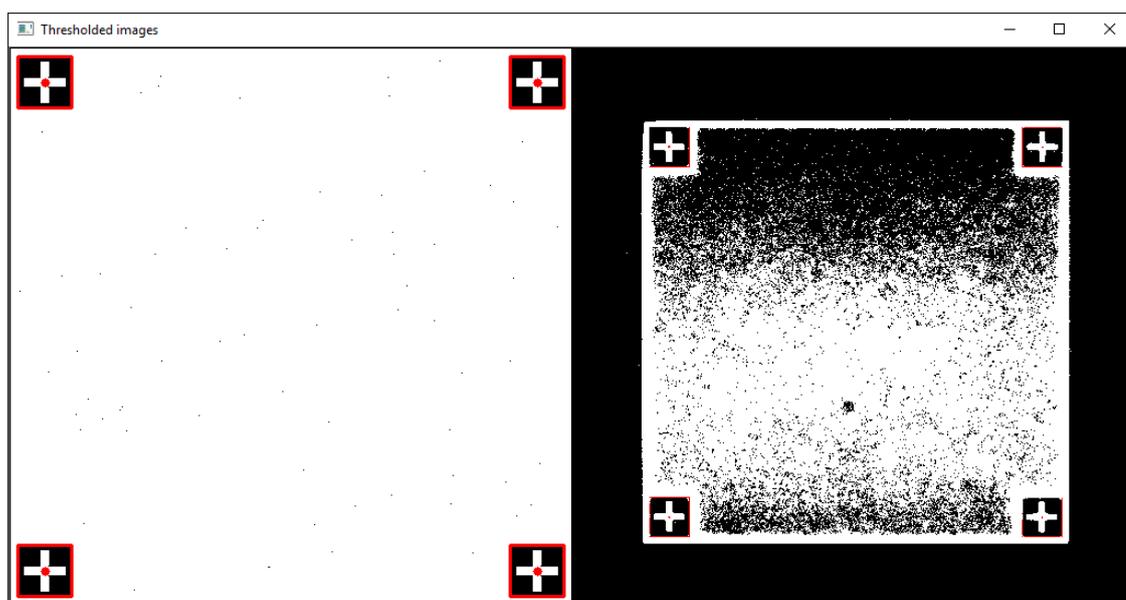
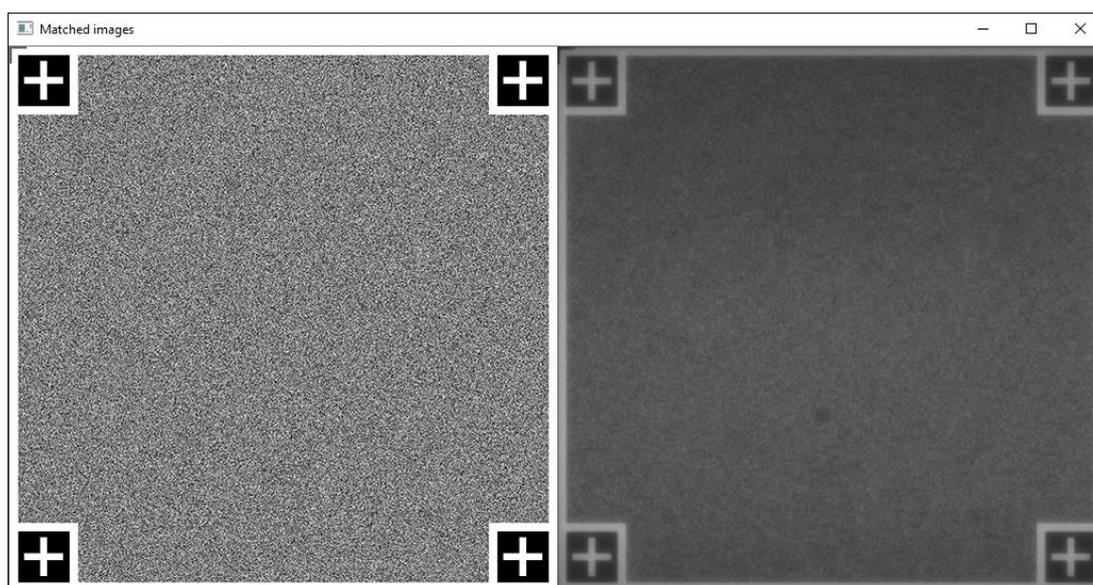


Рисунок 1.15 – Пример регистрируемого изображения с добавленными маркерами

Затем регистрируемое исходное изображение сопоставляется со снимком с экрана проектора, полученным с использованием дифракционного объектива. Сначала осуществляется пороговая обработка сопоставляемых изображений и нахождение координат маркеров (рисунок 1.16). Последним шагом сопоставления изображений является применение проективного преобразования к снимку, полученному с использованием дифракционной линзы (рисунок 1.17).



*Рисунок 1.16 – Обнаружение маркеров на сопоставляемых изображениях*



*Рисунок 1.17 – Результат сопоставления исходного изображения (слева) со снимком, полученным с использованием дифракционной линзы (справа)*

Поскольку PSNR возможно рассчитать, если существует соответствующее эталонное изображение, то PSNR используется только для оценки качества реконструкции тестовой выборки, а также валидационной выборки в процессе обучения сети для выбора наилучшей точки в пространстве параметров модели.

В прикладных задачах наиболее важным является качество реконструкции изображений сцен реального мира. В этом случае используется визуальная оценка качества реконструкции, а также предложенный в главе 3 критерий FEL, позволяющий учитывать фрагменты реальных изображений при обучении сети.

При использовании ИДОС на основе трех линз, алгоритм сопоставления полученных кадров и эталонных изображения модифицирован [8]. Поскольку формирование набора данных осуществляется последовательно для каждого из трех объективов, то для каждого эталонного изображения формируются три файла, соответствующие трем разным каналам формируемого изображения.

Сопоставление калибровочного изображения и снятого кадра осуществляется для каждого из трех наборов полученных данных. Сначала осуществляется бинаризация изображения с автоматическим подбором значения порога, локализация угловых маркеров на бинарном изображении, затем вычисление параметров проективного преобразования. Полученные проективные преобразования применяются к каждому набору данных. После этого осуществляется склейка полутоновых изображений в единое цветное RGB изображение.

Формирование реального изображения проводится по аналогичной схеме. Рядом с интересующим объектом размещается калибровочное изображение с маркерами (рисунок 1.18.а). Последовательно осуществляются формирование изображения каждым из объективов и сопоставление сформированного изображения с исходным. Далее полученные координаты маркеров используются для вычисления проективного преобразования красного и синего каналов изображения к зеленому. Полученный результат изображен на рисунке 1.18.б.

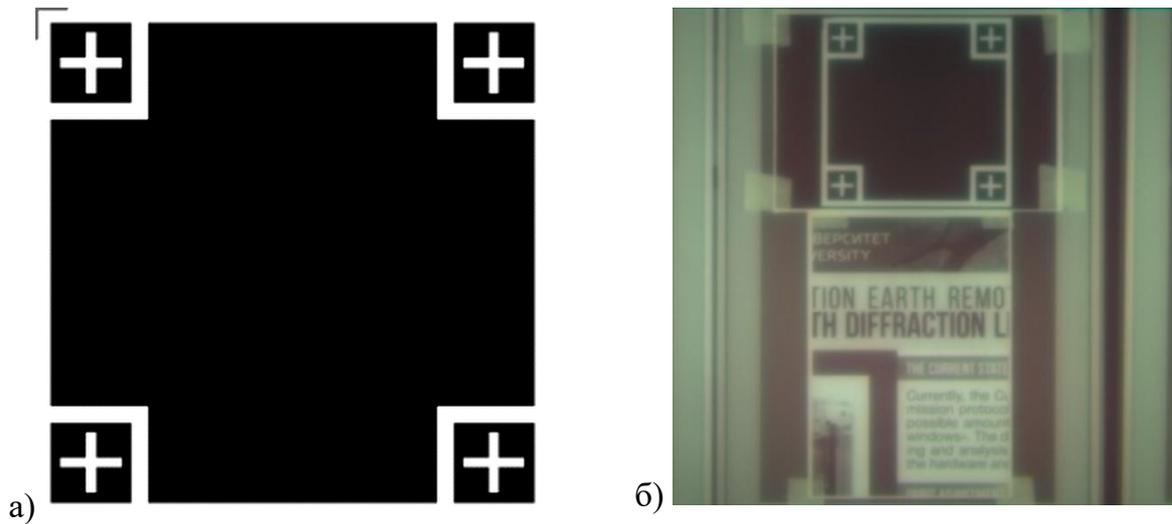


Рисунок 1.18 – Калибровочное изображение для сопоставления изображений (а) и снятое изображение реальной сцены после сопоставления кадров с трех линз (б)

### 1.5.3 Психовизуальный критерий оценки качества реконструкции изображений реальных сцен

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 визуально.

В исследованиях задачи сверхразрешения [71] используется средняя экспертная оценка (mean opinion score, MOS) – субъективное измерение, определенное в рекомендации P.800 МСЭ-T для оценки качества передачи в телефонных сетях [72]. MOS определяется следующим образом:

$$MOS = \frac{\sum_{n=1}^N R_n}{N}, \quad (1.12)$$

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.

В рамках данной диссертационной работы MOS используется для подтверждения эффективности предлагаемого в главе 3 критерия оценки качества реконструкции изображений, не имеющих эталона.

### 1.5.4 Качество формируемых изображений для разных ИДОС

В таблице 1.1 приведены различные оптические системы на основе плоской оптики и указано среднее значение PSNR на наборе формируемых данными системами изображений. Для МДЛ указан мультипликатор высоты микрорельефа  $m$ . Формула высоты микрорельефа (1.1) МДЛ представлена в разделе 1.2. Чем больше высота микрорельефа линзы, тем больше проявляются рефракционные свойства и, как результат, лучше качество формируемого изображения. На рисунках 1.19 – 1.21 представлены примеры изображений, сформированные разными типами плоской оптики.

Предлагаемые в данной работе алгоритмы реконструкции актуальны для всех типов плоской оптики, но в рамках данной работы используются наборы данных, сформированные различными дифракционно-оптическими системами, указанными в таблице: одноапертурная система на основе МДЛ с разной высотой микрорельефа ( $m=7$  и  $m=13$ ), двухапертурная и трехапертурная системы, а также гибридная рефракционно-дифракционная система.

По качеству формируемых входных изображений ДОС можно разделить на системы со средним значением PSNR:

1. меньше 12 Дб (четырёхуровневая линза, представленная в работе [12]);
2. больше 12 дБ, но меньше 17 дБ (трехапертурная система на низком рельефе, металинзы [73, 74]);
3. больше 17 дБ (все оставшиеся системы).

Таблица 1.1 – Качество формируемых изображений для разных видов плоской оптики

№	Тип ОС	Входной PSNR, дБ
1	Металинзы [73] и [74]	14,568; 9,7 (синий канал), 12,8 (зеленый канал), 13,1 (красный канал)
2	МДЛ (m=1) [12]	<12
3	МДЛ (m неизвестно) [31] и [32]	21,28 и 19,9
4	Одноапертурная система на основе МДЛ (m=7) [14]	18,13
5	Одноапертурная система на основе МДЛ (m=13) [2, 7]	19,68
6	Двухапертурная система на основе МДЛ (m=13) [7]	17,21
7	Трехапертурная система на основе МДЛ (m=1) [8]	12,77
8	Гибридная система (m=1) [33]	17,85

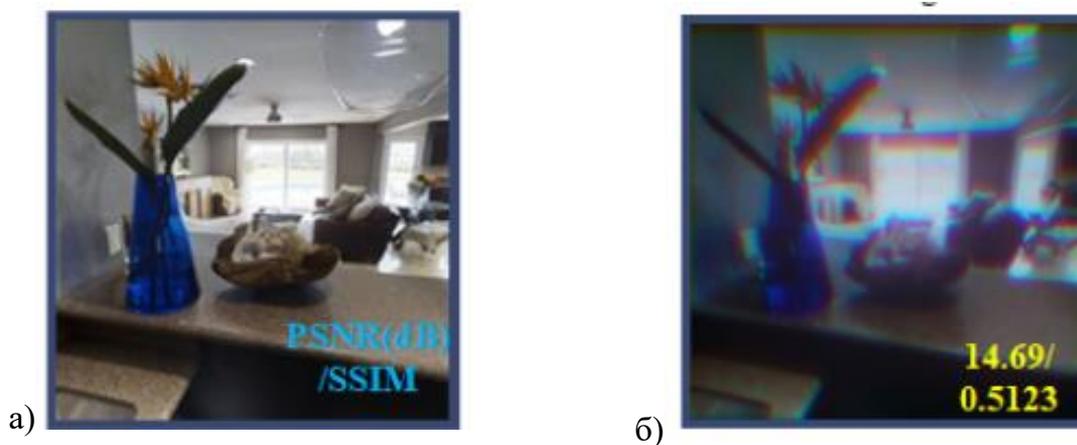
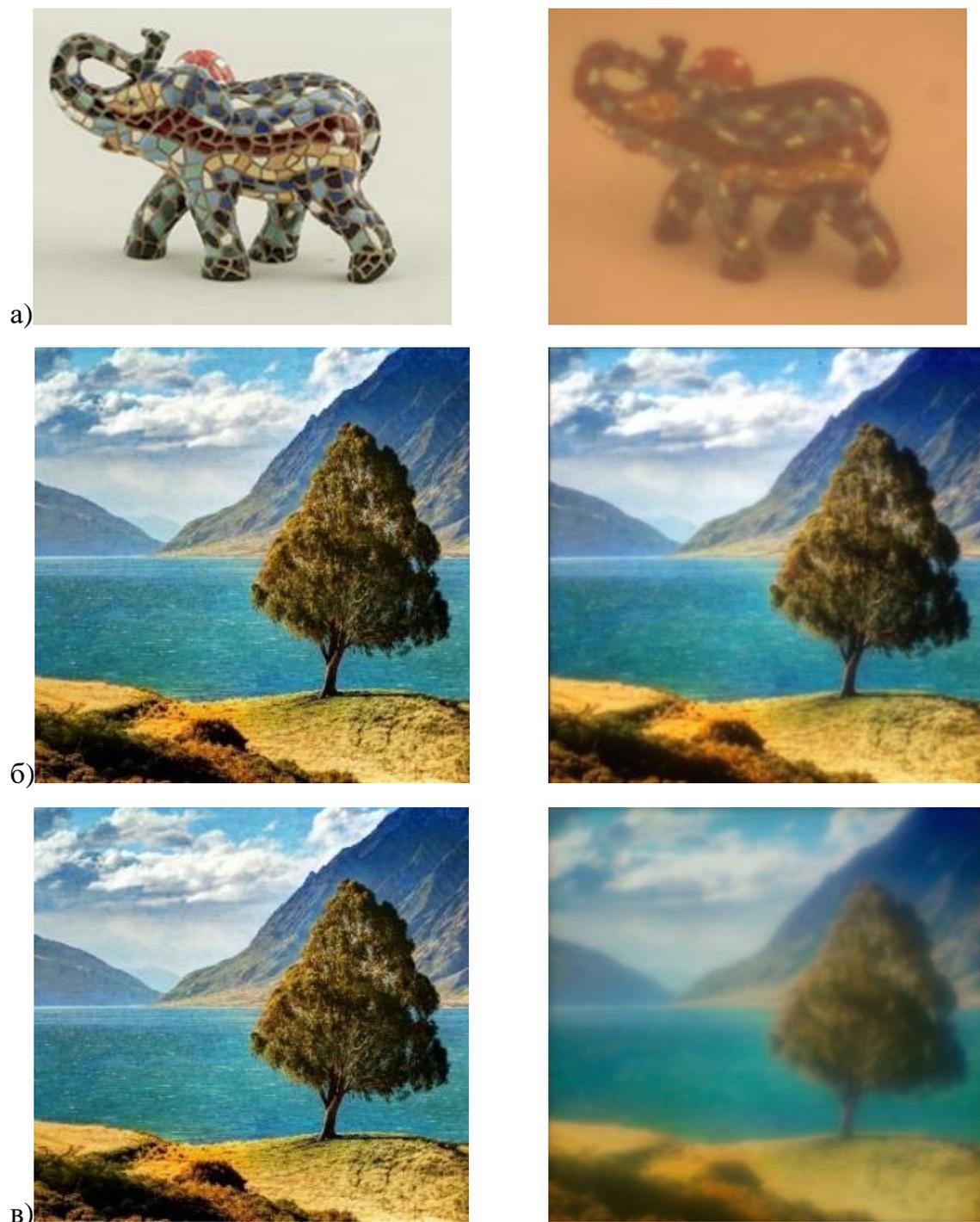


Рисунок 1.19 – Пример формирования изображения реальной сцены (а) с помощью металинзы[73] (б)



*Рисунок 1.20 – Примеры изображений, полученных разными конфигурациями ИДОС (справа) и соответствующие эталоны (слева): а) четырехуровневая линза [12], б) одноапертурная система на основе МДЛ ( $m=13$ ) [2, 7], в) одноапертурная система на основе МДЛ ( $m=7$ ) [14]*

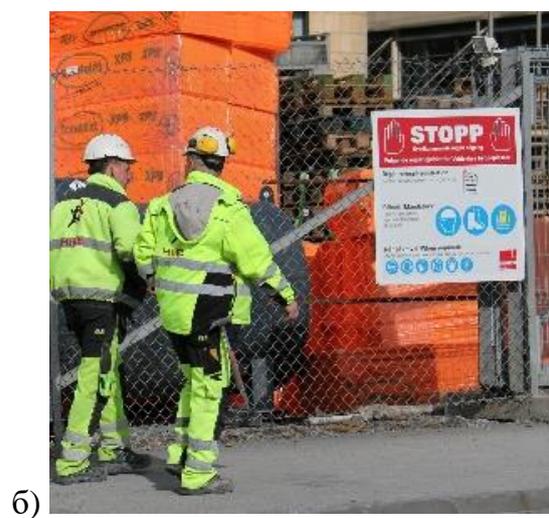


Рисунок 1.21 – Примеры изображений, полученных разными конфигурациями ИДОС (справа) и соответствующие эталоны (слева): а) гибридная система ( $m=1$ ) [33], б) трехапертурная система на основе МДЛ ( $m=1$ ) [8]

## Выводы к главе 1

1. Основной проблемой систем на основе плоской оптики являются сильные искажения формируемых изображений, проявляемые в виде локального хроматического размытия и нелокальной цветовой засветке из-за хроматических aberrаций, артефактов реконструкции, появляющихся из-за различий в условиях формирования обучающего набора данных и реальных изображений.
2. Результирующее качество реконструкции сильно зависит от исходного качества данных. Разнообразие ИДОС приводят к разному исходному качеству изображений и требуют разных подходов к реконструкции изображений как процедурного, так и нейросетевого.
3. Диаметр цветовой засветки на изображениях, формируемых в ИДОС, составляет больше 200 пикселей, поэтому размер рецептивного поля полнокадровой сети должен быть больше  $200 \times 200$  пикселей.
4. Для расчета PSNR требуется наличие эталонного изображения. Поскольку для изображений реальных сцен не имеется эталонных изображений, то появляется необходимость в разработке критерия, не требующего эталона, для оценки их качества.

## **ГЛАВА 2. СКВОЗНАЯ НЕЙРОСЕТЕВАЯ РЕКОНСТРУКЦИЯ ИЗОБРАЖЕНИЙ В ИДОС**

В настоящей главе предлагается двухэтапная схема сквозной нейросетевой реконструкции с описанием каждого шага предложенной схемы: полнокадровой и локальной реконструкции.

### **2.1. Нейросетевой подход к реконструкции изображений в ИДОС**

Задача реконструкции изображений в дифракционно-оптических системах схожа с задачей сверхразрешения по одному изображению (single image super resolution, SISR) [38, 66, 75, 76], которая заключается в восстановлении изображения высокого разрешения из изображения низкого разрешения. Сходство заключается в наличии локальных искажений – хроматического размытия, а также в том, что обе задачи (сверхразрешения и реконструкции изображений в ДОС) являются некорректно поставленными, так как существует множество изображений высокого разрешения, соответствующих входному изображению низкого разрешения. Однако в отличие от задачи сверхразрешения для задачи реконструкции изображений в ИДОС характерны нелокальные искажения в виде цветовой засветки.

Согласно обзору [36], современные нейросетевые решения задачи сверхразрешения обеспечивают лучшее визуальное качество изображений высокого разрешения, а также высокие значения пикового отношения сигнала к шуму (PSNR) и индекса структурного сходства (SSIM) [36, 75], по сравнению с классическими подходами. Разнообразные решения, основанные на сверточных (SRCNN [65, 77] и генеративно-состязательных архитектурах (SRGAN [71]) достигают улучшения критериев качества за счет модификации архитектуры [66, 78, 79] и функции потерь [80, 81], а также разработке новых стратегий обучения сетей [81-83]. С точки зрения наличия модели искажений в обзоре [36] нейросетевые подходы классифицируют на задачу с известной моделью искажений (supervised) и с неизвестной моделью искажений (unsupervised). В основу двухэтапной схемы реконструкции легла классификация архитектур с точки зрения размера рецептивного поля – области на входном изображении, вносящей вклад в выходной пиксел. Архитектуры с рецептивным полем, большим 200 пикселей, в рамках данной работы называются

полнокадровыми (image-wise), а реконструкция изображений на основе такой сети называется полнокадровой реконструкцией. Архитектуры с размером рецептивного поля, меньшим 200 пикселей, называются локальными (patch-wise), а реконструкция изображения на основе такой сети называется локальной реконструкцией. Порог 200 пикселей выбран на основе размера пятна фокусировки, который в разделе 1.2 составил 200 пикселей и говорит о том, что диаметр искажений на изображениях, формируемых в ДОС, составляет больше 200 пикселей.

В работах [2, 7] показано, что искажения изображений, характерные для дифракционной оптики, включают локальные искажения, вызванные хроматическими аберрациями, и нелокальные (контекстно-зависимые искажения, хроматический сдвиг), вызванные перераспределением энергии между дифракционными порядками линзы, диаметр которых составляет больше 200 пикселей.

В данном разделе предлагается двухэтапный нейросетевой подход к реконструкции изображений в ИДОС. Комбинация полнокадровой и локальной нейронных сетей позволяет компенсировать локальные и нелокальные искажения изображений. На рисунке 2.1 изображена предлагаемая двухэтапная схема реконструкции, которая применима для монокулярной и бинокулярной ИДОС на основе МДЛ ( $m=13$ ) [7], а также для ИДОС на основе МДЛ ( $m=7$ ) [16].

Использование локальных нейронных сетей характерно для задачи сверхразрешения [65, 66, 77, 84]. В данной работе локальные сети используются для дополнительного повышения качества локальных фрагментов изображения. В рамках предлагаемой двухэтапной схемы возможна комбинация с различными локальными сетями. В данной работе используется сеть VDSR [66] и DRRN [84], а также исследована возможность использования более современной архитектуры с механизмом внимания RCAN [85].

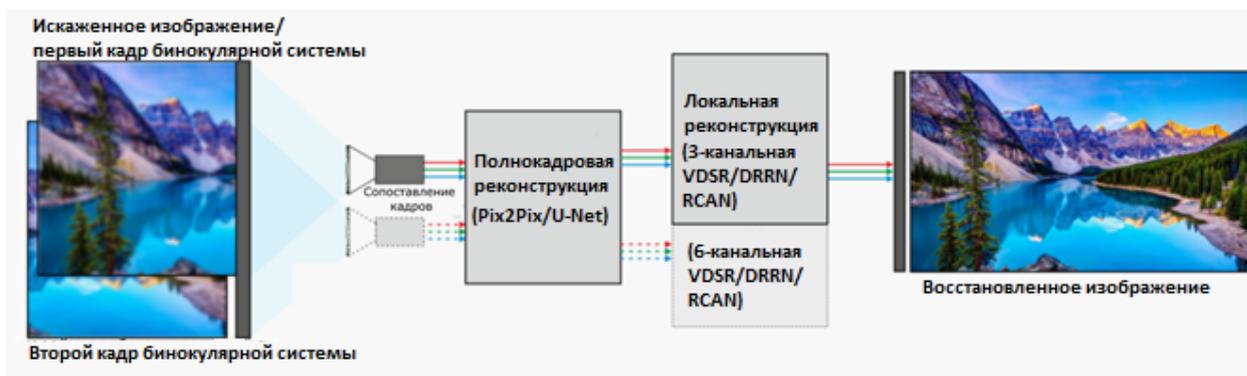


Рисунок 2.1 – Двухэтапная схема сквозной нейросетевой реконструкции изображений, полученных в монокулярной и бинокулярной системах на основе МДЛ

## 2.2. Полнокадровая нейросетевая реконструкция

В данной работе полнокадровая нейросетевая реконструкция выполняется на основе U-Net-подобных архитектур. В частности, исследуется как генеративно-сопоставительная нейронная сеть Pix2Pix [86], генератор которой является модификацией сети U-Net [87], так и сверточная нейронная сеть на основе архитектуры U-Net.

Как правило, архитектура генеративно-сопоставительных сетей состоит из двух сверточных нейронных сетей (СНС, англ. Convolutional neural network или сокращенно CNN), называемых генератором  $G$  и дискриминатором  $D$ . На рисунке 2.2 показан принцип работы генеративно-сопоставительной сети для полнокадровой реконструкции изображения в ИДОС. На вход генератору подается искаженное изображение в цветовом пространстве RGB, а на выходе генератора формируется восстановленное RGB изображение  $Y$ . Затем дискриминатор  $D$  определяет, является ли  $Y$  эталонным изображением из обучающей выборки или сгенерированным. Цель генератора – формирование изображений, которые дискриминатор не сможет отличить от эталонных. Генератор используемой в данной работе архитектуры Pix2Pix основан на сети U-Net, а дискриминатор – это архитектура PatchGAN [109].

Функция потерь для Pix2Pix имеет следующий вид:

$$L_{GAN}(G, D) = E_x [\log D(x)] + E_z [1 - \log D(G(z))], \quad (2.1)$$

где  $z$  – изображение, полученное с помощью МДЛ,  $x$  – оригинальное изображение.

В работе [87] в качестве функции потерь предлагается сумма  $L_{GAN}$  и функции расстояния  $L_1$ , которая отвечает за генерацию изображения, близкого к эталонному:

$$L_1(G) = E_{x,z} \left[ \|x - G(z)\|_1 \right]. \quad (2.2)$$

Таким образом задача обучения Pix2Pix принимает следующий вид:

$$G^* = \arg \min_G \max_D L_{GAN}(G, D) + \lambda L_1(G). \quad (2.3)$$

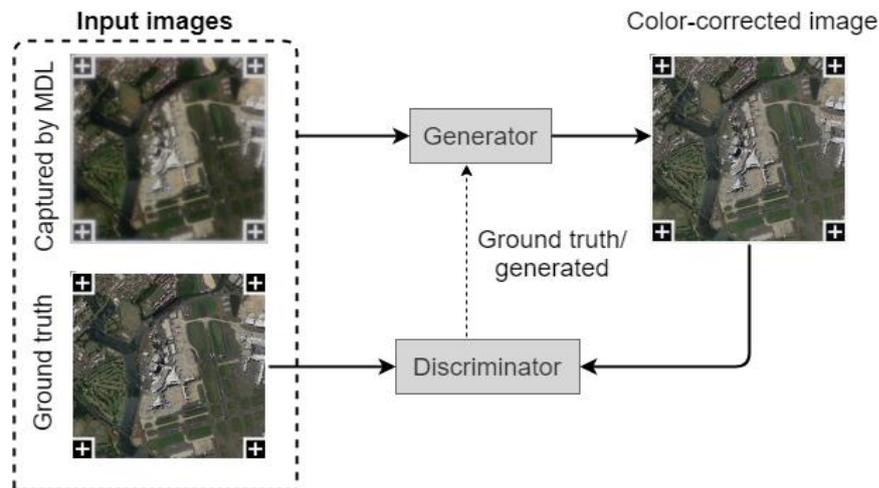


Рисунок 2.2 – Полнокадровая реконструкция на основе генеративно-сопоставительной нейронной сети

Классическая архитектура сети U-Net, разработанная для задачи сегментации медицинских изображений, представлена на рисунке 2.3. Сеть U-Net состоит из кодировщика (encoder) и декодировщика (decoder). Кодировщик включает блоки слоев, понижающих размерность изображения (понижающий блок), а декодировщик – блоки слоев, повышающих размерность изображения (повышающий блок). Понижающий размерность блок состоит из двух последовательных сверточных слоев с ядром свертки размера  $3 \times 3$  и функцией активации ReLU и последующим слоем подвыборки (max-pooling), предназначенным для понижения размерности с окном размера  $2 \times 2$  и шагом, равным 2. В каждом понижающем блоке количество каналов карт свойств увеличивается в два раза. Повышающий размерность блок состоит из транспонированного сверточного слоя с ядром свертки размера  $2 \times 2$ , операции объединения результата выполнения транспонированной свертки с картами свойств соответствующей размерности, полученными на выходе одного из понижающих блоков, а также двух

сверточных слоев с ядром свертки размера  $3 \times 3$  и функцией активации ReLU. В каждом повышающем блоке количество каналов карт свойств уменьшается в два раза. В качестве последнего слоя используется сверточный слой с ядром свертки размера  $1 \times 1$ .

В оригинальной работе [87] архитектура U-Net состоит из 4 понижающих блоков и 4 повышающих блоков. Между понижающим и повышающим блоками находятся два дополнительных сверточных слоя с ядром свертки размера  $3 \times 3$  и функцией активации ReLU.

Оригинальную архитектуру сети U-Net можно записать в следующем виде:

**Encoder:** C64-C64-P-C128-C128-P-C256-C256-P-C512-C512-P-C1024-C1024,

**Decoder:** Tc1024- C512- C512- Tc512- C256- C256- Tc256- C128- C128- Tc128- C64- C64-C2,

где  $C_k$  – это комбинация сверточного слоя и функции активации с количеством фильтров  $k$ , P – слой подвыборки, Tc $k$  - это комбинация транспонированного сверточного слоя и функции активации с количеством фильтров  $k$ .

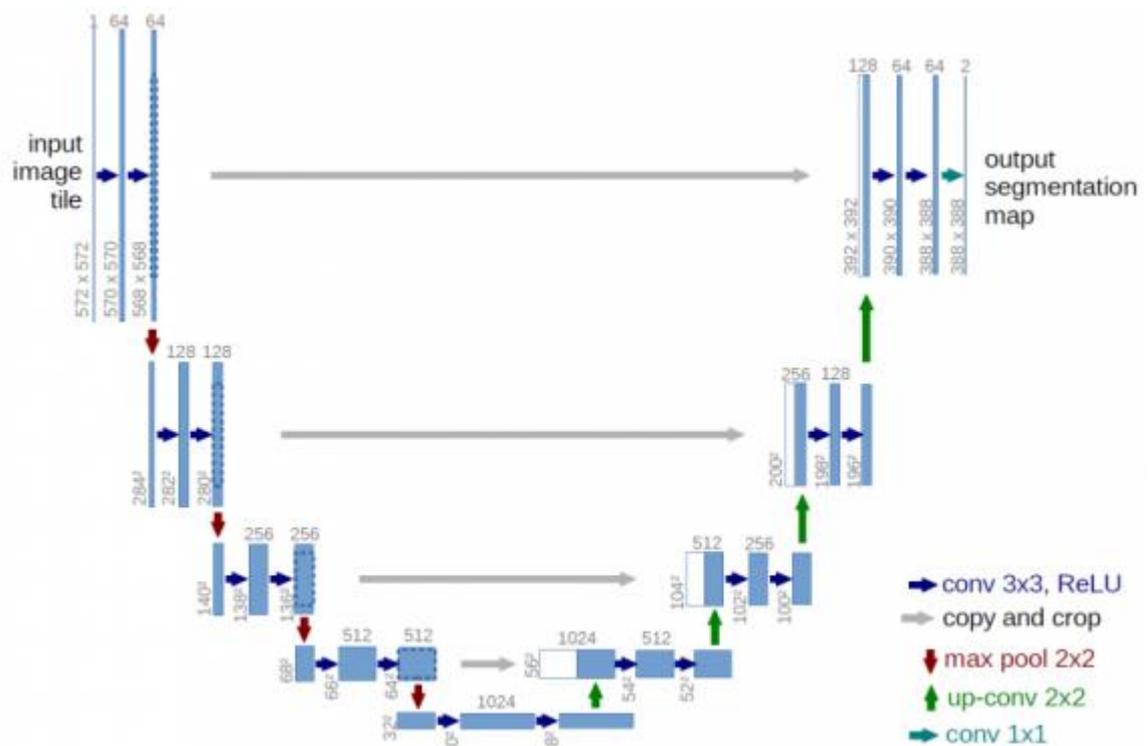


Рисунок 2.3 – Классическая архитектура сети U-Net [87]

В данной работе используется следующая модификация классической архитектуры U-Net:

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

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

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

Размер окна свертки во всех случаях –  $4 \times 4$ . На каждом сверточном слое кодировщика (encoder) ширина и высота карт признаков уменьшаются в 2 раза, а слои декодировщика за счет транспонированной свертки напротив увеличивают размеры карт признаков, что позволяет получить на выходе сети восстановленное изображение такого же разрешения, как у входного изображения. В качестве функции активации в кодировщике используется Leaky ReLU, а в декодировщике - ReLU. Для понижения разрешения вместо слоя подвыборки, который использовался в оригинальной архитектуре, используется шаг, равный 2, в сверточном слое.

Основными отличиями предложенной архитектуры от классической архитектуры U-Net [87] являются меньшее количество сверточных слоев (15 в предлагаемой архитектуре и 23 в оригинальной архитектуре), а также наличие слоев пакетной нормализации и dropout, которые используются для борьбы с эффектом переобучения. Кроме того, размеры ядер сверток и шага (stride) обусловлены тем, что при переходе из кодировщика в декодировщик в оригинальной архитектуре проводится процедура сопоставления размерностей, которая сопровождается обрезанием изображения, что в случае задачи сегментации не является критичным, но при реконструкции изображений в ИДОС приводит к потере информации по краям изображения.

Как было сказано ранее, для полнокадровой реконструкции важно, чтобы размер рецептивного поля сети был больше  $200 \times 200$  пикселей. Оценка рецептивного поля кодировщика используемой U-Net-подобной архитектуры составила  $778 \times 778$  пикселей, то есть рецептивное поле всей сети больше  $778 \times 778$  пикселей (оценка рецептивного поля всей сети снизу), что удовлетворяет требованию к размеру рецептивного поля для полнокадровой реконструкции. Расчет рецептивного поля кодировщика проводился по стандартной формуле расчета рецептивного поля для сверточных нейронных сетей, представленной в работе [88].

Как правило, полнокадровая реконструкция изображений эффективна для ИДОС с качеством входных изображений, большим 17 дБ. Главная проблема

полнокадровой реконструкции – артефакты реконструкции на изображениях реальных сцен. Для локальной реконструкции проблема артефактов не столь актуальна, так как она обеспечивает более низкий прирост качества в целом. Артефакты реконструкции и подходы к их уменьшению рассматриваются в главе 3.

Для выбора функции потерь, используемой при обучении сети U-Net, проведено исследование стандартных функций потерь для задачи сверхразрешения, описанных в работе [89], а также дополнительно исследуется функция Хьюбера [90]:

1. Средняя абсолютная ошибка

$$L^{l_1}(P) = \frac{1}{N} \sum_{p \in P} |x(p) - y(p)|,$$

где  $p$  – индекс пиксела,  $P$  – фрагмент изображения,  $x(p)$  и  $y(p)$  – значения пикселей, восстановленного и эталонного изображений соответственно,  $N$  – количество пикселей в  $P$ .

2. Средняя квадратичная ошибка:

$$L^{l_2}(P) = \frac{1}{N} \sum_{p \in P} (x(p) - y(p))^2.$$

3. Индекс структурного сходства SSIM [92]:

$$L^{SSIM}(P) = \frac{1}{N} \sum_{p \in P} 1 - SSIM(p),$$

где  $SSIM(p)$  – значение индекса структурного сходства пиксела  $p$ .

4. Многомасштабный индекс структурного сходства [91]:

$$L^{MS-SSIM}(P) = \frac{1}{N} \sum_{p \in P} 1 - MS\_SSIM(p),$$

где  $MS\_SSIM(p)$  – значение многомасштабного индекса структурного сходства пиксела  $p$ .

5. Комбинация функций  $L^{MS-SSIM}$  и  $L^{l_1}$ :

$$L^{Mix_1} = \alpha L^{MS-SSIM} + (1 - \alpha) L^{l_1},$$

где параметр  $\alpha$  выбран на основе исследований, проведённых в работе [89].

6. Комбинация функций  $L^{MS-SSIM}$  и  $L^{l_2}$ :

$$L^{Mix_2} = \alpha L^{MS-SSIM} + (1 - \alpha) L^{l_2},$$

где параметр  $\alpha = 0,84$ .

7. Функция Хьюбера [90]:

$$L^{Huber} = \begin{cases} \text{если } |x(p) - y(p)| \leq \delta, \frac{1}{2}(x(p) - y(p))^2 \\ \text{иначе, } \frac{1}{2}|x(p) - y(p)| - \frac{1}{2}\delta^2 \end{cases},$$

где  $\delta$  – параметр функции Хьюбера.

### 2.3 Локальная нейросетевая реконструкция

Локальная реконструкция изображений проводится на основе архитектур VDSR [66] и DRRN [84]. В данной работе предлагается трехканальная и шестиканальная модификации сети VDSR. В работе [2] сеть VDSR успешно применялась для устранения хроматического размытия, однако каждый канал изображения обрабатывался сетью отдельно. На рисунке 2.4.в представлена используемая архитектура СНС, основанная на сети VDSR.

Трехканальная VDSR принимает на вход одновременно все три канала RGB изображения и на выходе формируется реконструированное RGB изображение. На рисунке 2.4.а представлена архитектура сети VDSR для монокулярной ДОС. Архитектура сети VDSR состоит из повторяющихся блоков сверточного слоя и слоя активации. В качестве последнего слоя используется остаточный слой (residual layer или Addition слой на рисунке 2.4), на котором проводится попиксельное сложение выхода предыдущего слоя и входного изображения. Также в отличие от оригинальной архитектуры VDSR в качестве функции активации используется функция PReLU (parametric rectified linear unit) и 15 сверточных слоев вместо 20. Размер рецептивного поля сети VDSR, включающей 15 сверточных слоев, составляет  $31 \times 31$  пикселей.

Локальная реконструкция изображений, полученных в бинокулярной системе, выполняется на основе 6-канальной модификации сети VDSR, которая принимает на вход два RGB изображения после полнокадровой реконструкции и формирует на выходе восстановленное RGB изображение. На рисунке 2.4.б представлена предложенная архитектура 6-канальной VDSR.

На последнем слое VDSR (Addition слой) предлагаемой модификации выполняется сложение выходного изображения предыдущего слоя и линейной комбинации входных изображений:

$$\begin{aligned}
Addition &= Y_{prev} + X_{in} \\
X_{in} &= (1 - \gamma)X_{left} + \gamma X_{right}
\end{aligned}
\tag{2.4}$$

где  $Y_{prev}$  – выход предыдущего слоя,  $X_{left}$  – входное RGB изображение, снятое МДЛ, расположенной слева,  $X_{right}$  – входное RGB изображение, снятое МДЛ, расположенной справа,  $\gamma$  – обучаемый параметр сети.

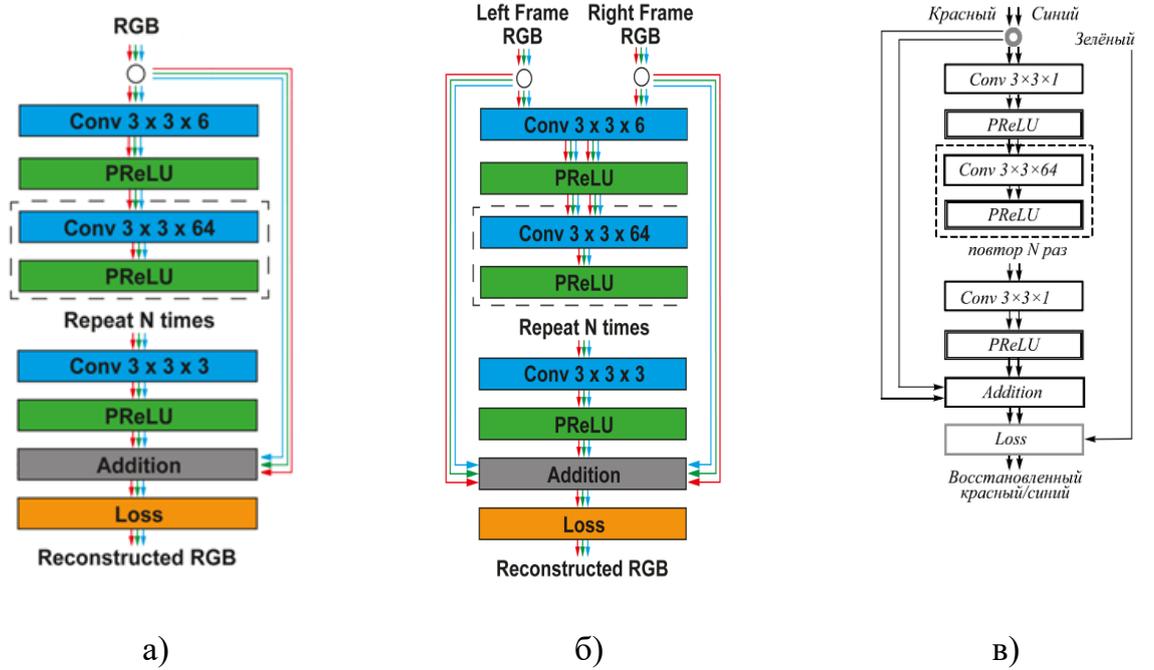


Рисунок 2.4 – Архитектура сети VDSR: а) для монокулярной ДОС, б) для бинокулярной ДОС, в) для коррекции хроматического размытия в работе [2]

Учитывая гипотезу об ахроматичности контуров [60] и утверждение, что зелёный канал является более чётким в ДОС [2, 12], к функции потерь для сети VDSR добавляется штраф за отклонение контура в красном и синем каналах от контура в зеленом канале [2]:

$$\begin{aligned}
L(\hat{p}_i, p_i^0, p_G^0, w) &= \frac{1}{N} \sum_{j=1}^N d(\hat{p}^{(j)} - p^{0(j)}) + \\
&+ \lambda \frac{1}{N} \sum_{j=1}^N \|p_G^{0(j)} \nabla \hat{p}_i^{(j)} - \hat{p}_i^{(j)} \nabla p_G^{0(j)}\|_2 + \eta \|w\|_2, i \in \{R, B\},
\end{aligned}
\tag{2.5}$$

где  $N$  – количество изображений в одном мини-пакете обучения,  $j$  – индекс изображения в выборке,  $p_i^{0(j)}$  –  $i$ -й цветовой канал эталонного изображения,  $\hat{p}_i^{(j)}$  –  $i$ -й цветовой канал восстановленного изображения. Первое слагаемое функции потерь обеспечивает минимизацию расстояния  $d()$  между эталонным изображением

и изображением, полученным на выходе сети. Второе слагаемое реализует штраф за отклонение градиента от зелёного канала  $p_G^{0(j)}$  эталонного изображения. Последнее слагаемое накладывает ограничения на норму весовых коэффициентов сети (weight decay). В качестве  $d()$  может использоваться  $L_2$ -норма или дифференцируемый аналог  $L_1$ -нормы – функция потерь Чарбоньера (Charbonnier) [68].

Архитектура сети DRRN, используемая в данной работе представлена на рисунке 2.5 и состоит из трех вложенных друг в друга остаточных блоков и 6 сверточных слоев. На вход сети (input) подается трехканальное RGB изображение. Выход сети (output) – также трехканальное RGB изображение. Для обучения сети использовалась функция потерь, учитывающая гипотезу об ахроматичности контуров (2.5).

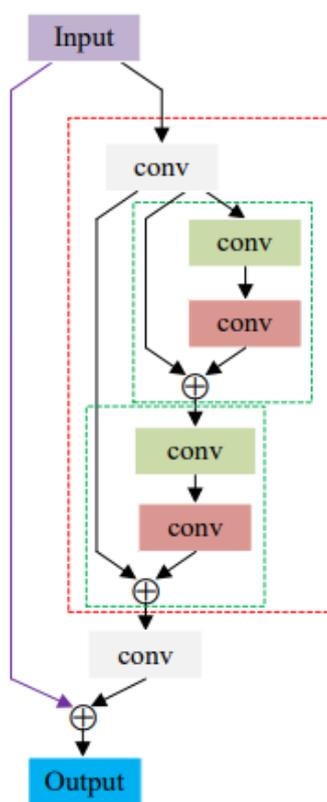


Рисунок 2.5 – Архитектура сети DRRN

Среди архитектур, используемых для задачи суперразрешения с известной моделью искажений, многообещающие результаты, согласно обзору [36] обеспечивает сеть с механизмом внимания RCAN [85], поэтому в данной работе проведена попытка использования сети RCAN для локальной реконструкции.

Архитектура сети RCAN, представленная на рисунке 2.6, учитывает различия в каналах изображения, выделяя в них разные признаки. RCAN состоит из 4 основных частей: сверточный слой для предварительного извлечения признаков (conv), основной остаточный блок, состоящий из  $G$  остаточных блоков (residual in residual), слои повышения разрешения и сверточный слой реконструкции. Каждый остаточный блок состоит из  $V$  блоков с вниманием (RCAB), которые позволяют масштабировать выделяемые признаки для каждого канала отдельно. Предлагаемый в [85] механизм с вниманием позволяет извлекать признаки из более информативных каналов изображения. Для задачи локальной реконструкции блок повышения разрешения (upscale module) удаляется из архитектуры.

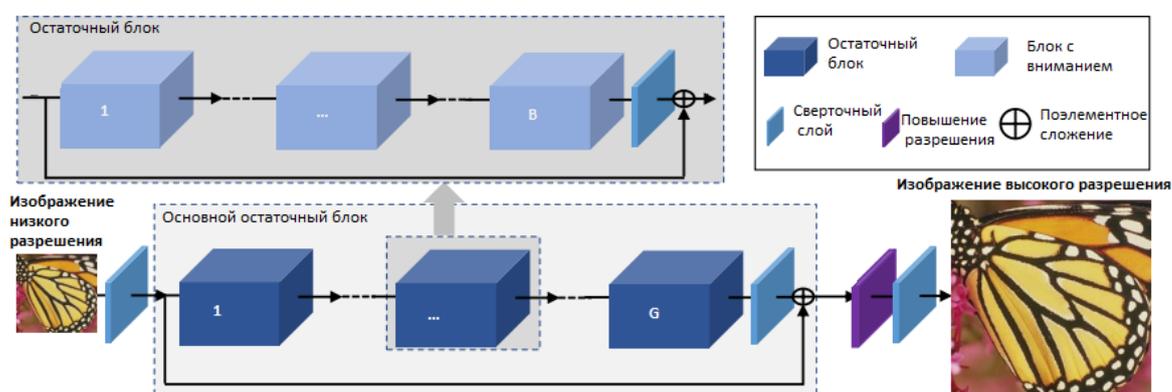


Рисунок 2.6 – Архитектура сети RCAN

## Выводы к главе 2

1. Двухэтапная схема сквозной нейросетевой реконструкции включает этапы полнокадровой и локальной реконструкций. Архитектуры с рецептивным полем, большим 200 пикселей, в рамках данной работы называются полнокадровыми, а реконструкция изображений на основе такой сети называется полнокадровой реконструкцией. Архитектуры с размером рецептивного поля, меньшим 200 пикселей, называются локальными, а коррекция изображения на основе такой сети называется локальной реконструкцией.
2. Главная проблема полнокадровой реконструкции – артефакты реконструкции на изображениях реальных сцен. Для локальной реконструкции проблема артефактов не столь актуальна, так как она обеспечивает более низкий прирост качества в целом.
3. Полнокадровая реконструкция осуществляется на основе U-Net – подобной архитектуры. Локальные сети используются для дополнительного повышения качества локальных фрагментов изображения. В рамках предлагаемой двухэтапной схемы возможна комбинация с различными локальными сетями. В данной работе используется сеть VDSR и DRRN, а также исследована возможность использования более современной архитектуры с механизмом внимания RCAN.

## ГЛАВА 3. АЛГОРИТМЫ ОБУЧЕНИЯ И АДАПТАЦИИ НЕЙРОСЕТЕВЫХ МОДЕЛЕЙ РЕКОНСТРУКЦИИ ИЗОБРАЖЕНИЙ В ИДОС

В настоящей главе рассматриваются методы и алгоритмы адаптации нейросетевой модели реконструкции к различиям в условиях формирования обучающего набора данных и изображений реальных сцен. В данной главе введено такое понятие, как уровень ложных контуров (FEL) и предложен критерий выбора лучшей точки в пространстве параметров модели на основе FEL, а также представлены алгоритмы аугментации обучающего набора данных для уменьшения артефактов реконструкции. Кроме того, в данной главе теоретически обоснован метод крупношагового метаобучения.

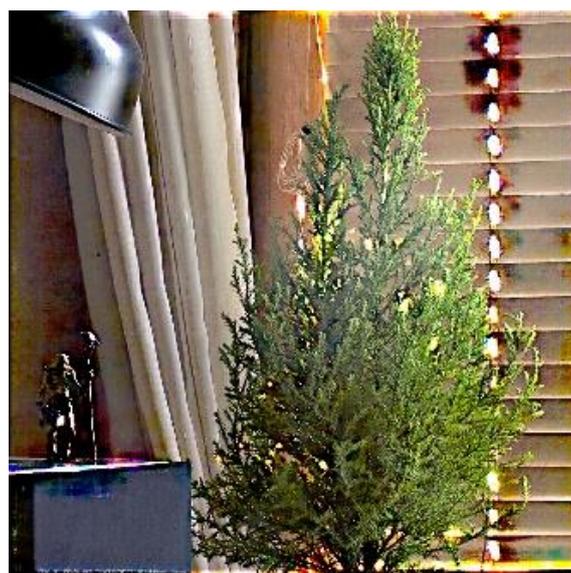
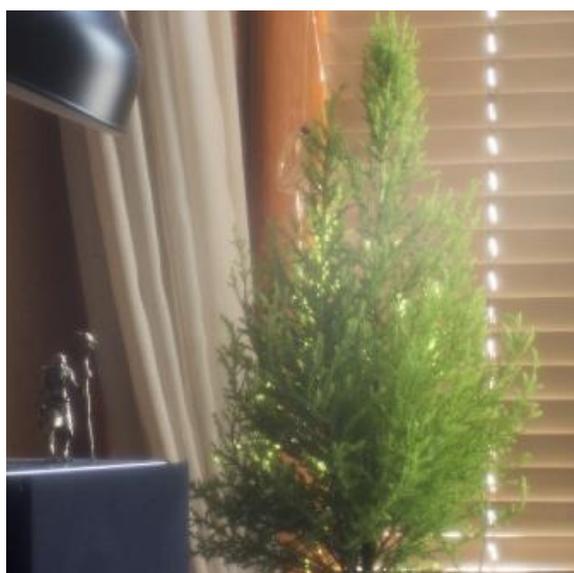
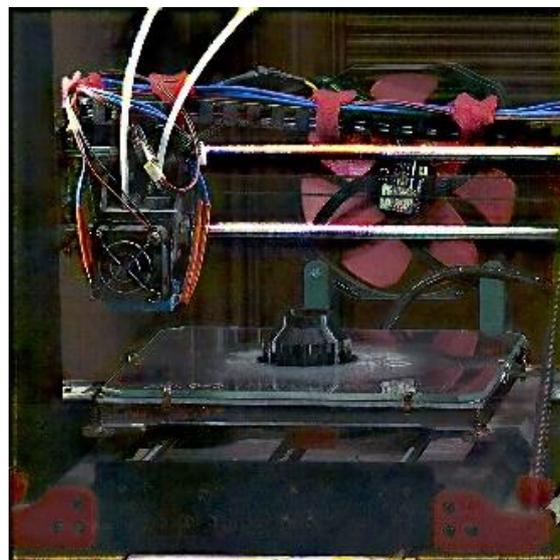
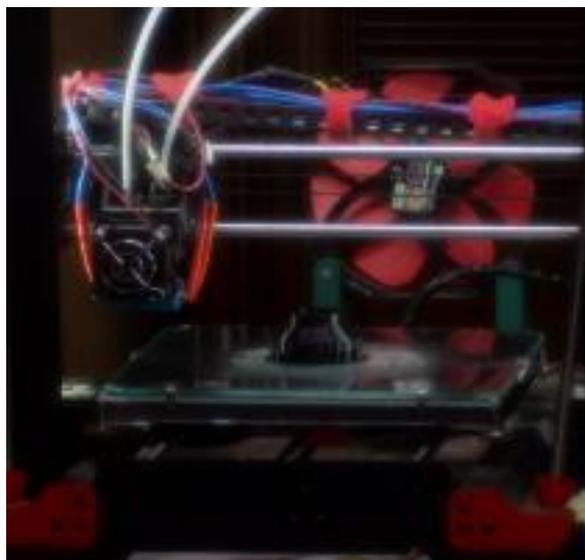
### 3.1 Ограничения при формировании набора данных

Успешное использование двухэтапного сквозного нейросетевого подхода к реконструкции изображений позволяет получить систему компьютерного зрения, имеющую практическое применение. Работа [15] посвящена практическому применению двухэтапной схемы сквозной нейросетевой реконструкции изображений, полученных на спутнике CubeSat с камерой на основе МДЛ. Однако, как показано в [15], при практическом применении ИДОС не обходится без проблем, связанных с артефактами реконструкции, возникающими из-за различий в условиях формирования реальных изображений и обучающей выборки.

Экспозиция и динамический диапазон на изображениях реальных сцен значительно отличаются от результата формирования обучающей выборки, которое проводилось в условиях средней освещенности в 200 люкс. Кроме того, причинами артефактов реконструкции могут быть усиление (gain) или светочувствительность (ISO) камеры.

На рисунке 3.1 представлены примеры реальных изображений, на которых появились артефакты при реконструкции полнокадровой сетью. Артефакты реконструкции наблюдаются в наиболее ярких областях изображения, что приводит к предположению о влиянии уровня экспозиции и динамического диапазона входного изображения. На рисунке 3.2 приведены примеры полнокадровой реконструкции реального изображения с разными значениями экспозиции.

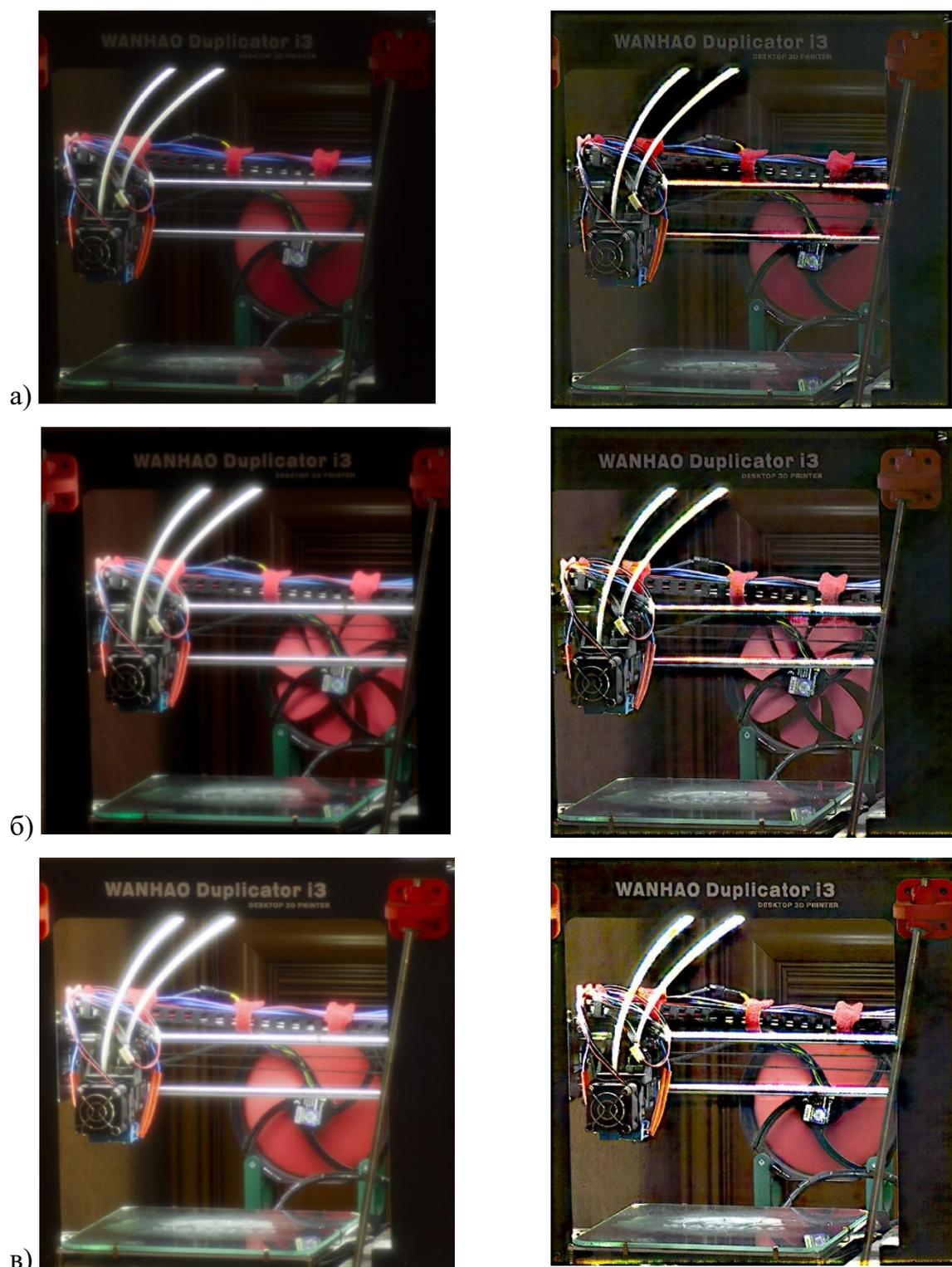
Увеличение и уменьшение значения экспозиции на 1 стоп проводилось в программе Adobe Photoshop. Видно, что уменьшение экспозиции на 1 стоп позволило уменьшить артефакты реконструкции в наиболее ярких (засвеченных) областях.



а)

б)

*Рисунок 3.1 – Примеры артефактов на реальных изображениях, обработанных сетью Pix2Pix на этапе полнокадровой реконструкции: а) изображения, полученные МДЛ, б) результат обработки полнокадровой сетью Pix2Pix*

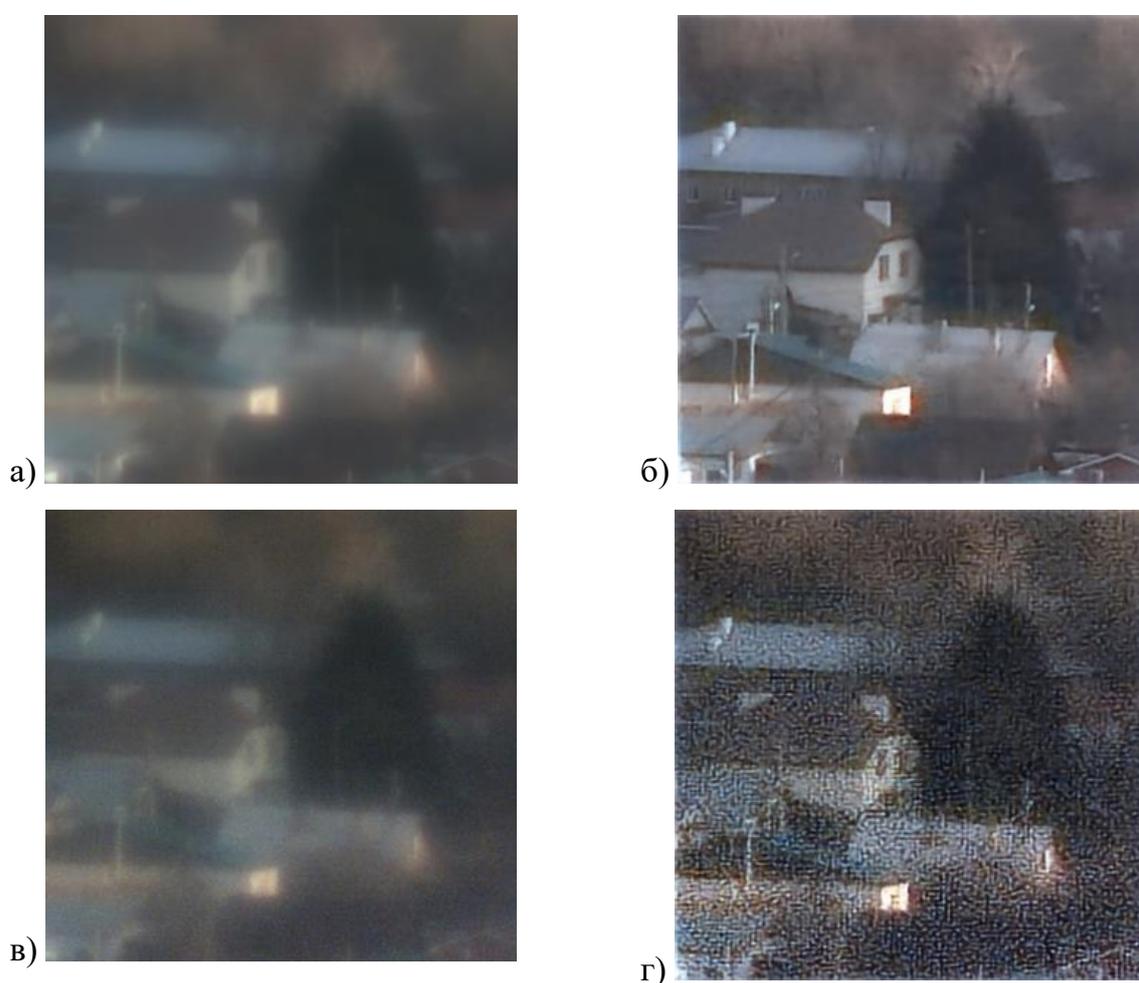


*Рисунок 3.2 – Полнокадровая реконструкция реального изображения с разными значениями экспозиции, слева – входное, справа – восстановленное: а) уменьшение экспозиции на 1 стоп, б) исходное изображение, в) увеличение экспозиции на 1 стоп*

Кроме того, на рисунке 3.3 показано влияние коэффициента усиления камеры на появление артефактов реконструкции в виде ложных контуров на

изображении. Видно, что при увеличении коэффициента усиления камеры количество артефактов тоже увеличивается, поэтому было принято решение при формировании изображений использовать коэффициент усиления камеры (gain), равный 0. В этом случае количество артефактов реконструкции будет варьироваться в зависимости от используемой камеры и сенсора, так как одно и то же значение светочувствительности матрицы (ISO) на разных камерах может приводить к разному уровню ISO-шума.

Поскольку артефакты реконструкции возникают только после обработки полнокадровой сетью, дальнейшие исследования с целью уменьшения количества артефактов будут проводиться только для этапа полнокадровой реконструкции.



*Рисунок 3.3 – Примеры реконструкции реальных изображений, полученных при разных значениях коэффициента усиления камеры (gain): а) изображение, полученное МДЛ при gain = 2, б) результат полнокадровой реконструкции а), в) изображение, полученное МДЛ при gain = 16, г) результат его полнокадровой реконструкции в)*

Стоит отметить, что в проводимых экспериментах не брались в расчет эффекты различий в пиксельной сетке между экраном и камерой, оснащенной дифракционным объективом, а также эффекты, связанные с конечной величиной глубины резко изображаемого пространства (ГРИП). Первое исключено из рассмотрения, т.к. при формировании изображений в лаборатории разрешение экрана и камеры приблизительно равны, а пятно размытия ИДОС составляет не менее 2-3 пикселей. Отметим также, что диафрагменное число исследуемых систем не менее 16, что позволяет не учитывать эффекты конечной ГРИП. При формировании изображения сцена сосредоточена вдоль плоскости, перпендикулярной оптической оси, либо размещена в гиперфокальном пространстве оптической системы.

### 3.2 Доменная генерализация нейросетевых моделей

Учитывая ограничения формирования обучающего набора данных лабораторными условиями и важность реконструкции изображений реальных сцен для прикладного использования ИДОС, появляется необходимость в доменной генерализации полнокадровой нейронной сети. В классификации подходов по наличию модели искажений, реконструкция реальных изображений в ИДОС также относится к задаче с неизвестной моделью искажений, поскольку такие изображения не имеют эталона и появление артефактов реконструкции связано с различиями в условиях формирования обучающего набора данных и реальных кадров.

Согласно обзору [48], под доменом понимается выборка данных из генеральной совокупности:

$$S = \{(x_i, y_i)\}_{i=1}^n \sim P_{XY},$$

где  $x \in X \subset \mathbb{R}^d$  и  $y \in Y \subset \mathbb{R}$ ,  $X$  и  $Y$  – случайные вектора, обозначающие вход и выход сети с совместным распределением  $P_{XY}$ . Обучающий набор данных  $S_{train}$  в задаче доменной генерализации включает  $M$  обучающих доменов:

$$S_{train} = \{S^i | i = 1, \dots, M\},$$

где  $S^i = j\{(x_i^j, y_i^j)\}_{i=1}^{n_i}$  –  $i$ -й домен. Совместное распределение каждой пары доменов обучающего набора данных различно:  $P_{XY}^i \neq P_{XY}^j, 1 \leq i \neq j \leq M$ .

Целью доменной генерализации является обучение сети на обучающем наборе доменов для минимизации потерь на неизвестном тестовом домене  $S_{test}$  с неизвестным распределением  $P_{XY}^{test} \neq P_{XY}^i, i = \overline{1, M}$ :

$$\min_h E_{(x,y) \in S_{test}} [l(h(x), y)],$$

где  $E$  – математическое ожидание,  $l$  – функция потерь,  $h: X \rightarrow Y$  – нейросетевая модель. На рисунке 3.4 приведена иллюстрация доменной генерализации.

Близкими к решению задачи доменной генерализации являются рассматриваемые в разделе 3.4 методы трансферного обучения, метаобучения и обучения по нулевой выборке. Сложностью задачи доменной генерализации является полное отсутствие информации о распределении тестового домена. При включении информации о новом домене в процесс обучения задача упрощается до доменной адаптации модели. В разделе 3.3 вводится новый критерий FEL, позволяющий учитывать реальные данные в процессе обучения сети с целью уменьшения артефактов нейросетевой реконструкции. Поскольку при использовании FEL информация о неизвестном домене реального изображения учитывается в процессе обучения, то задачу обучения модели с целью уменьшения артефактов реконструкции на реальных кадрах можно отнести к задаче доменной адаптации [37].

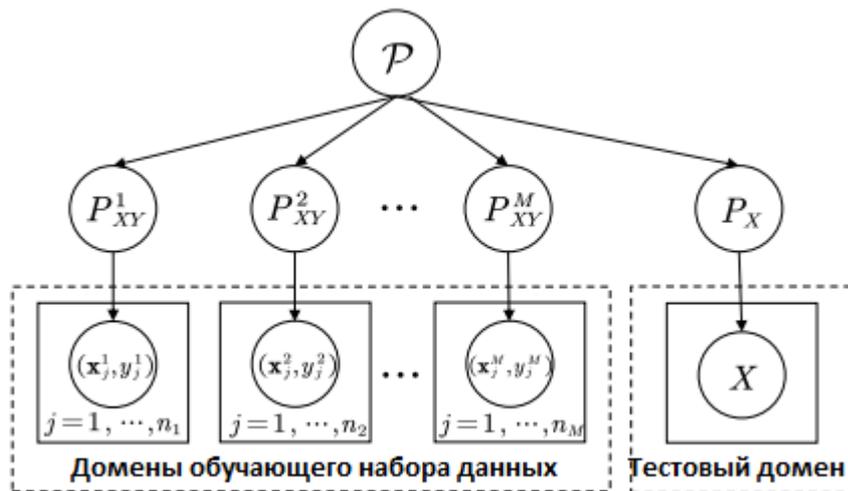


Рисунок 3.4 – Иллюстрация задачи доменной генерализации [37]

### 3.3 Критерий ложных контуров

С целью уменьшения артефактов реконструкции на этапе полнокадровой коррекции в данной работе предлагается измерять уровень артефактов на реальном изображении в процессе обучения сети для поиска лучшей точки в пространстве параметров модели. Для измерения уровня артефактов вводится уровень ложных контуров (False edge level, FEL) [33], который позволяет использовать изображение реальной сцены в процессе обучения и не требует наличия соответствующего эталонного (ground truth) изображения.

Поскольку артефакты реконструкции на изображении визуально имеют вид контуров (рисунок 3.5.а и рисунок 3.5.б), то и количественная оценка таких артефактов основана на обнаружении контуров для фрагмента изображения, на котором визуально контуров быть не должно, то есть фрагменты без перепадов яркости. На выбранном фрагменте нет перепадов яркости и визуально контуры не наблюдаются, поэтому все обнаруженные контуры на данном фрагменте относятся к артефактам реконструкции. 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  $[0, 1]$ .

Можно выделить следующие требования к выбору фрагмента: ширина или высота фрагмента больше 200 пикселей (данный порог выбран на основе диаметра цветовой засветки в ИДОС) и на аналогичном фрагменте изображения, сформированным рефракционной камерой, значение FEL должно быть минимально. Выбор алгоритма детекции контуров не ограничен.

На рисунке 3.5 приведены результаты работы алгоритма Canny на фрагментах восстановленных изображений реальной сцены. Для каждого фрагмента указаны значения FEL.

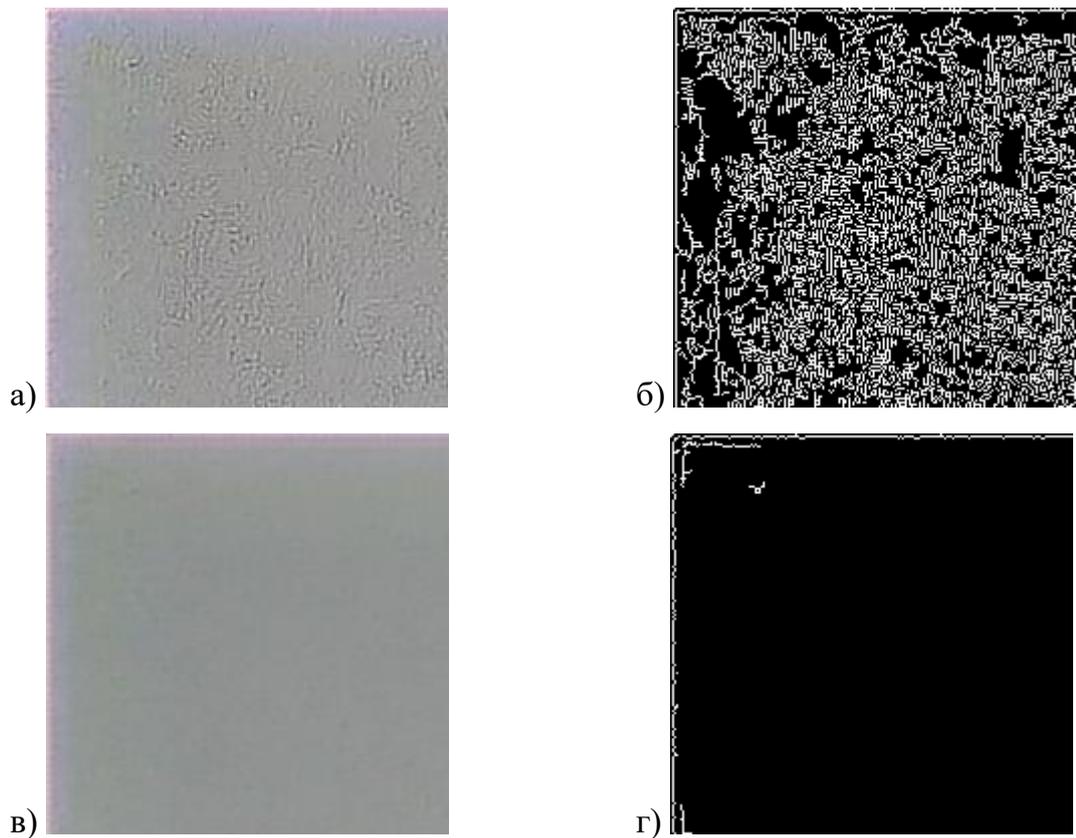


Рисунок 3.5 – Примеры работы алгоритма Sanny: а) фрагмент восстановленного изображения реальной сцены ( $FEL = 30,69 \%$ ), б) результат работы алгоритма Sanny для а), в) фрагмент восстановленного изображения реальной сцены ( $FEL = 1,28 \%$ ), г) результат работы алгоритма Sanny для в)

Классический критерий ранней остановки обучения сети [14, 15] основан на среднем значении PSNR, рассчитанный по валидационной выборке. Согласно данному подходу, критерием выбора лучшей точки в пространстве параметров модели является максимальное значение среднего PSNR для валидационной выборки:

$$best\_epoch = \operatorname{argmax}_e \operatorname{meanPSNR}(T_e^{val}),$$

где  $T_e^{val}$  – валидационная выборка, восстановленная моделью с параметрами, полученными на эпохе обучения  $e$ ,  $\operatorname{meanPSNR}$  – среднее значение PSNR на валидационной выборке. Параметры модели, полученные на эпохе  $best\_epoch$  считаются лучшей точкой в пространстве параметров сети согласно критерию PSNR.

В данной работе предлагается альтернативный критерий ранней остановки обучения, основанный на расчете FEL для фрагмента изображения реальной сцены.

Критерием выбора лучшей точки в пространстве параметров является **минимальное значение FEL** на фрагменте реального изображения [33]:

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

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

### **3.4 Аугментация обучающего набора данных**

Как было сказано в разделе 3.1, причинами артефактов реконструкции на реальных изображениях являются различия уровней экспозиции реального изображения и изображений обучающей выборки, а также светочувствительность сенсора ISO, приводящая к шуму на входном изображении.

В данной работе предлагается решение для уменьшения артефактов реконструкции на основе аугментации обучающего набора данных. Стоит отметить, что аугментация проводится с целью увеличения репрезентативности выборки, а не с целью увеличения ее объема. Искажение данных применяется с некоторой вероятностью к уже существующему набору данных, не повышая его объем.

Для улучшения качества реконструкции изображений реальных сцен аугментация данных проводится путем моделирования двух видов искажения: шум, причиной которого является усиление или светочувствительность камеры (ISO шум), и изменение экспозиции изображения.

Уровень артефактов на реальном изображении в процессе обучения сети для поиска лучшей точки в пространстве параметров модели измеряется с помощью FEL.

#### **3.4.1 Изменение экспозиции обучающих данных**

Изменение экспозиции изображений обучающей выборки проводилось с использованием алгоритма повышения качества изображения, сформированного в условиях низкой освещенности, предложенного в работе [92]. Алгоритм основан на функции отклика камеры, связывающей освещенность сенсора со значениями пикселей изображения. Для повышения качества изображения алгоритм формирует оценку модели отклика камеры на основе гистограммных характеристик двух изображений с разной экспозицией и карту значений экспозиции. Эксперименты в

работе [92] показали, что алгоритм позволяет варьировать экспозицию изображения, не внося дополнительные искажения цвета по сравнению с другими современными подходами.

Как правило, современные камеры выполняют различные нелинейные преобразования, например, баланс белого [93-95], устранение шума [96], цветовую коррекцию [97] и др. Нелинейная функция, связывающая освещенность сенсора и значение пиксела формируемого изображения называется функцией отклика камеры (camera response function, CRF).

Как показано на рисунке 3.6, модель такого нелинейного процесса можно представить в следующем виде:

$$P = f(E),$$

где  $E$  – освещенность сенсора,  $P$  – формируемое изображение,  $f$  – функция отклика камеры. Оценка функции  $f$  может быть проведена с помощью моделирования функции преобразования яркости (brightness transform function, BTF). Как показано на рисунке 3.6, функция преобразования яркости сопоставляет два изображения  $P_0$  и  $P_1$  одной сцены, снятой с разной экспозицией:

$$P_1 = g(P_0, k),$$

где  $g$  – функция преобразования яркости,  $k$  – коэффициент экспозиции. Оценка функции отклика камеры может быть получена следующим образом:

$$g(f(E), k) = f(kE).$$

Среди различных подходов к моделированию функции преобразования яркости часто используют Гамма-коррекцию и аффинные преобразования [92]. Однако предложенный в работе [92] подход обеспечивает более реалистичное улучшение качества изображения.

Изображения, полученные при недостаточной освещенности могут быть глобально недоэкспонированными, но могут быть и локально недоэкспонированными, поэтому для того, чтобы улучшить качество такого изображения, повышая его экспозицию, недостаточно использование постоянного коэффициента экспозиции  $k$ . Авторы работы [92] предлагают использовать карту коэффициентов экспозиции (exposure ratio map)  $K$ :

$$P' = g(P, K),$$

где  $P$  и  $P'$  – входное и улучшенное выходное изображения. Карта коэффициентов экспозиции содержит подходящий для каждого пиксела коэффициент экспозиции. Таким образом, задача улучшения качества изображения, снятого в условиях недостаточной освещенности заключается в оценке функции преобразования яркости  $g$  и карты коэффициентов экспозиции  $K$ .

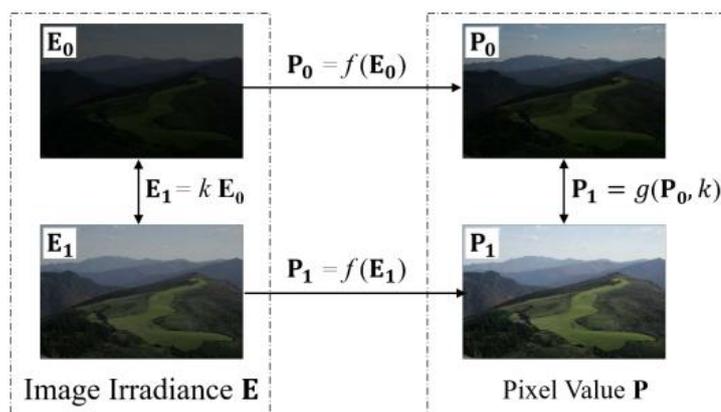


Рисунок 3.6 – Связь освещенности и пикселов изображений одной сцены, полученных при разных значениях экспозиции

В рамках данной работы алгоритм, предложенный в работе [92], используется для увеличения и уменьшения уровня экспозиции изображения. На рисунке 3.7 представлен пример изменения экспозиции изображения, полученного МДЛ.

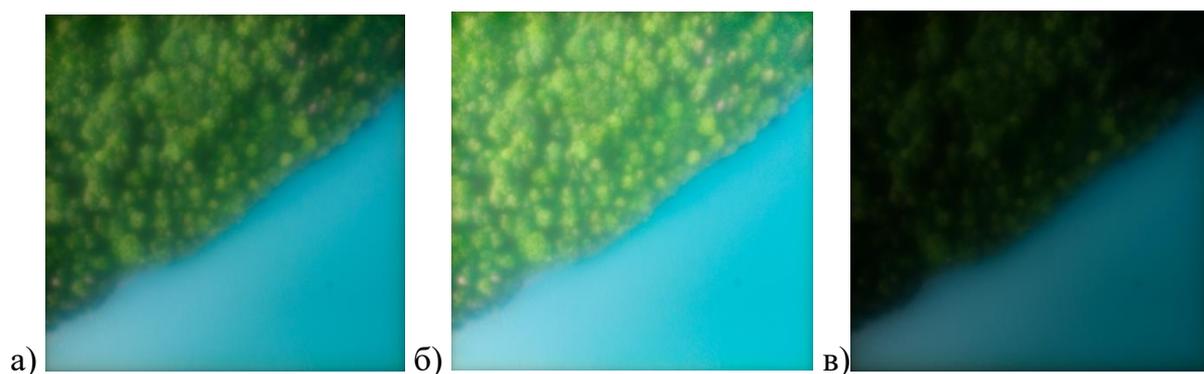


Рисунок 3.7 – Пример изменения экспозиции изображения обучающей выборки на основе алгоритма, предложенного в [92]: а) изображение, полученное МДЛ ( $m=13$ ), б) результат увеличения экспозиции, в) результат уменьшения экспозиции

### 3.4.2 Моделирование ISO-шума на обучающих данных

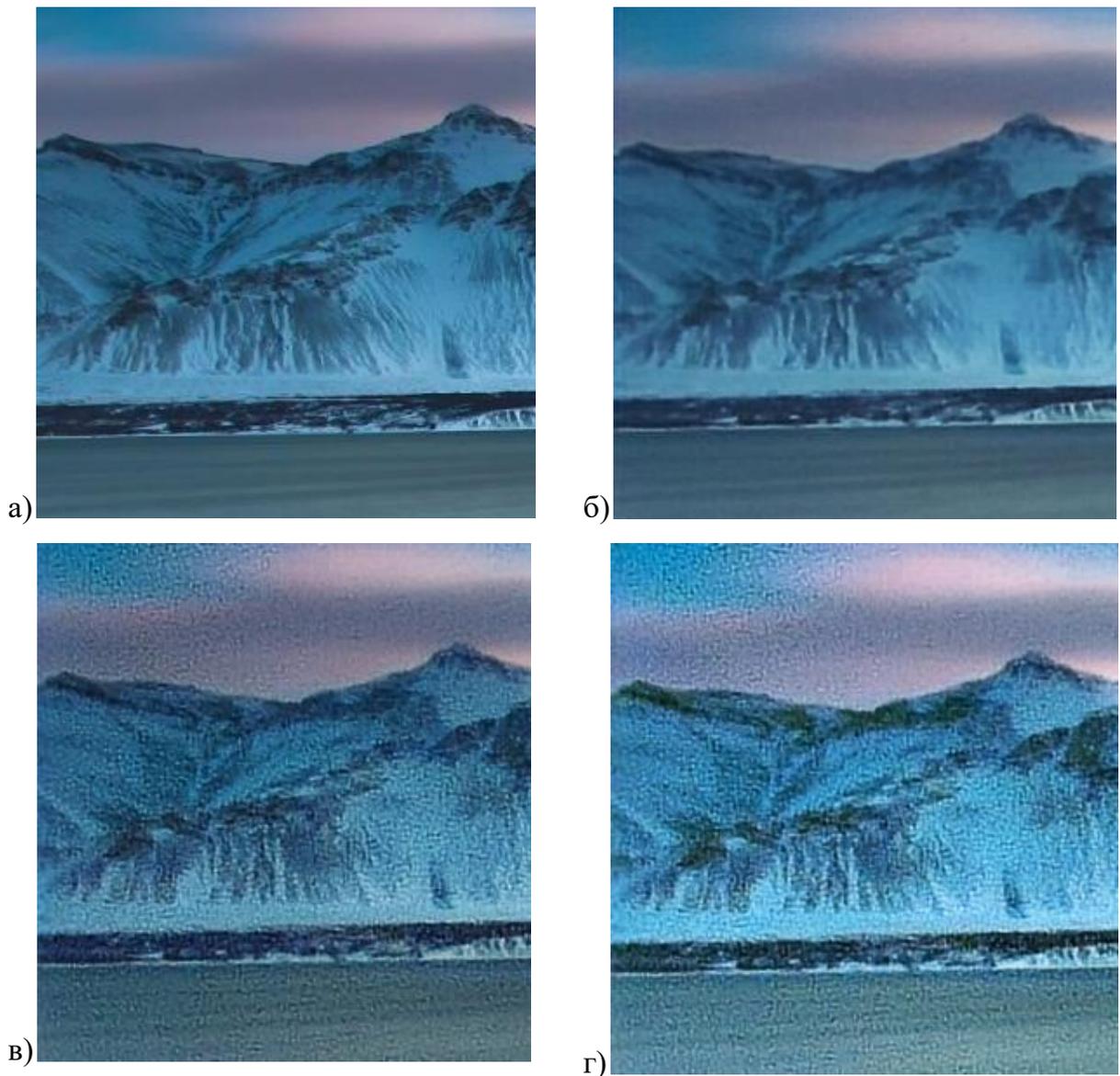
Электрические и тепловые характеристики сенсоров в современных цифровых камерах заметно улучшаются [98], при этом размер пиксела становится все меньше и меньше, что приводит к увеличению чувствительности выходного сигнала сенсора к фотонному шуму. Как утверждают авторы работы [98], на данный момент фотонный шум вносит наиболее значимый вклад в шумовую составляющую выходного сигнала сенсора, что делает ее заметно зависимой от сигнала. В таком случае классический подход моделирования аддитивного гауссовского шума, который используется при обработке изображений в целом и, в частности, авторами работы [32] при моделировании искажений в ИДОС, к данной работе не применим. Согласно работе [98], наиболее подходящая к исходным данным модель шума включает два слагаемых: зависимый от сигнала фотонный шум  $\eta_p$  (распределение Пуассона) и независимый от сигнала стационарный шум  $\eta_g$  (распределение Гаусса). Шум Пуассона-Гаусса имеет следующий вид [98, 99]:

$$y = \eta_p + \eta_g = \frac{1}{a} \alpha + \beta,$$

где  $y$  – наблюдаемый зашумленный сигнал,  $\alpha \sim \Pi(ax)$ ,  $\beta \sim N(0, b^2)$ ,  $x$  – исходный сигнал без шума,  $a$  и  $b$  – параметры распределений.

Для моделирования ISO-шума в данной работе используется реализация Пуассона-Гауссовского шума из библиотеки `albuminations` [100], 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).

На рисунке 3.8 представлен пример наложения ISO-шума на изображение, полученное дифракционно-рефракционной ОС, а также демонстрация артефактов реконструкции, возникающих после применения предложенной аугментации ISO-шумом (рисунок 3.8.в) и ISO-шумом совместно с увеличением экспозиции (рисунок 3.8.г). Артефакты реконструкции на рисунках 3.8.в и 3.8.г визуально похожи на артефакты, возникающие при обработке изображений реальных сцен (рисунок 4.10.а и 4.10.в), что показывает целесообразность использования предложенных алгоритмов аугментации.



*Рисунок 3.8 – Демонстрация артефактов реконструкции на изображении из тестовой выборки для дифракционно-рефракционной ОС: а) эталонное изображение, б) восстановленное входное изображение без дополнительной искажения, в) восстановленное входное изображение с предварительно наложенным ISO-шумом, г) восстановленное входное изображение с предварительно наложенным ISO-шумом и увеличенной экспозицией*

### **3.5 Метод крупношагового метаобучения**

Решение, основанное на аугментации обучающих данных и критерии ложных контуров, решает проблему артефактов реконструкции, позволяя использовать реальные данные в процессе обучения. Однако в реальных задачах для формирования новых реальных данных могут использоваться другие виды ИДОС,

другие условия формирования, а также другие камеры. Например, в случае получения всего нескольких снимков от спутника, оборудованного ИДОС [15], нет возможности обучить новую модель реконструкции для данного домена из-за отсутствия обучающей выборки. Появляется необходимость в некотором методе, который позволил бы на основе небольшого количества новых данных из новых доменов дообучить имеющуюся модель, тем самым адаптировав ее к новым условиям формирования изображений.

Вдохновением для такого подхода послужила работа [39], содержание которой описано в пункте 3.5.1. В пункте 3.5.2 приведена модификация данного подхода, основанная на крупношаговом метаобучении, а в пункте 3.5.3 представлено теоретическое обоснование метода крупношагового метаобучения.

### **3.5.1 Мелкошаговое метаобучение для задачи сверхразрешения**

Все большую популярность набирает разработка новых подходов к решению задачи сверхразрешения с неизвестной моделью искажений. Одним из решений является комбинация трансферного обучения (transfer learning) [45-47], мелкошагового метаобучения, независимого от модели искажений (model-agnostic metalearning, MAML) [43] и обучения по нулевой выборке (zero-shot superresolution, ZSSR) [48, 49]. Такая трехэтапная комбинация, названная метаобучением для сверхразрешения по нулевой выборке (MZSR, meta-transfer learning for zero-shot super-resolution), успешно применялась в работе [39].

При трансферном обучении предварительно обученная модель используется в качестве «начальной точки» для обучения на целевом наборе данных при решении похожей задачи. Цель метаобучения – обучить модель на наборах данных, соответствующих различным задачам или доменам, и последующее использование обученной модели для быстрой адаптации к новой задаче или домену с небольшим объемом обучающей выборки. Метаобучение использует отдельные домены (задачи) в качестве обучающих образцов. Обучение по нулевой выборке предполагает отсутствие обучающей выборки и обучение по одному тестовому изображению, не имеющему эталона.

Подход, описанный в работе [39], включает 3 этапа (рисунок 3.9):

1. Обучение на достаточно большом наборе данных (large scale training) высокого качества с известной моделью искажений. На данном этапе определяются значения весов начальной модели  $\theta_T$ .
2. Мелкошаговое метаобучение (meta-transfer learning). Дообучение начальной модели  $\theta_T$  на  $N$  наборах данных, сформированных с использованием модели искажений с различными параметрами. При одновременном дообучении на  $N$  наборах данных (доменах) проводится усреднение  $N$  траекторий градиентного спуска, что позволяет определить «среднюю» точку  $\theta_M$  в пространстве параметров модели для  $N$  задач (доменов).
3. Этап метатестирования (meta test) с использованием подхода к сверхразрешению по нулевой выборке (ZSSR), в рамках которого модель  $\theta_M$  дообучается на одном изображении низкого разрешения (input LR) с неизвестной моделью искажений. Поскольку для данного изображения нет эталона, то дообучение проводится на следующей паре: применяется понижение разрешения исходного изображения низкого разрешения (input LR) и сформированное изображение подается на вход сети, а исходное изображение низкого разрешения (input LR) считается эталонным и используется при расчете функции потерь. Спустя несколько шагов градиентного спуска обучение останавливается. Количество шагов определяется экспериментальным путем. В результате дообучения сформирована финальная модель  $\hat{\theta}_k$  на вход которой подается исходное изображение низкого разрешения (input LR). Финальная модель генерирует изображение высокого разрешения для изображения низкого разрешения с неизвестной моделью искажений. Подробнее данный процесс описан в работе [39].

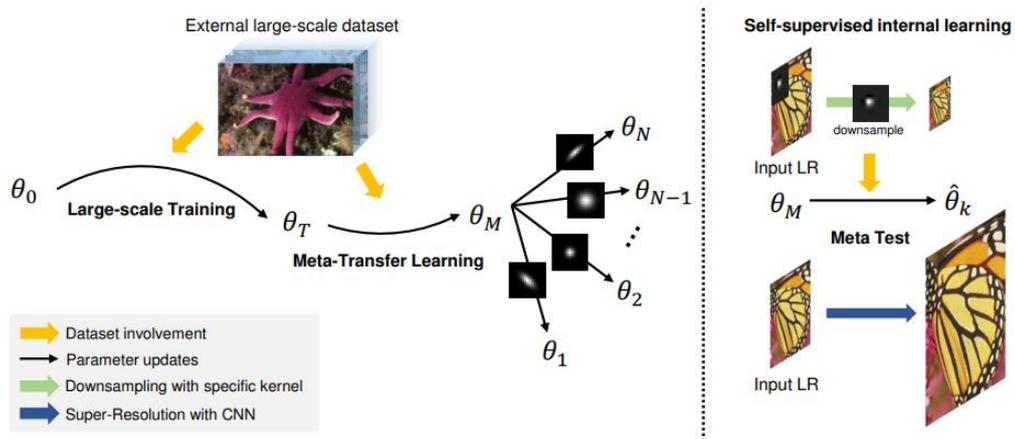


Рисунок 3.9 – Схема метаобучения для задачи сверхразрешения по нулевой выборке (MZSR), представленная в работе [39]

### 3.5.2 Адаптация модели к различным условиям формирования изображений на основе крупношагового метаобучения

В данной работе используется алгоритм крупношагового метаобучения, в рамках которого усреднение траекторий градиентного спуска заменяется усреднением значений параметров моделей, сформированных на втором этапе алгоритма – этапе мелкошагового метаобучения. Кроме того, на этапе адаптации полученной усредненной модели к новому домену для расчета функции потерь используется улучшенное экспертом Photoshop изображение вместо эталона. Обоснование алгоритма крупношагового обучения с усреднением моделей, а не траекторий градиентного спуска представлено в разделе 3.4.3.

Предлагаемый в данной работе алгоритм включает 3 этапа (рисунок 3.10):

1. Обучение на исходном наборе полуреальных данных. Перед обучением проводится аугментация обучающей выборки  $D_{train}$  моделированием ISO-шума и изменением экспозиции. Процедура аугментации описана в разделе 3.4. Выбор начальной точки данного подхода  $\theta_T$  в пространстве параметров модели проводится на основе критерия FEL, предложенного в разделе 3.3.
2. Крупношаговое метаобучение. Последовательное дообучение начальной модели  $\theta_T$  на  $N$  наборах данных (доменах), сформированных с

использованием различных типов искажений, характерных для ДОС: увеличение экспозиции, уменьшение экспозиции, моделирование ISO-шума с различными параметрами. В отличие от оригинальной работы, где проводилось усреднение траекторий градиентного спуска, в данной работе предлагается усреднять значения параметров  $N$  моделей, полученных в результате дообучения начальной модели  $\theta_T$  на  $N$  наборах данных (доменах):

$$\theta_M = \frac{1}{N} \sum_{i=1}^N \theta_i. \quad (3.2)$$

Выбор лучшей точки в пространстве параметров модели при дообучении начальной модели  $\theta_T$  на  $N$  наборах данных (доменах) осуществляется на основе значения PSNR, рассчитанного по валидационной выборке.

3. Обучение на одном реальном изображении или дообучение на нулевой выборке. На данном этапе модель  $\theta_M$  дообучается на одном реальном изображении с неизвестной моделью искажений. Для расчета функции потерь используется обработанное экспертом Photoshop изображение. На рисунке 3.11 представлен пример реального изображения и его результат обработки в Photoshop. Количество эпох обучения определяется экспериментальным путем на основе значения FEL. Модель  $\hat{\theta}_k$ , сформированная в результате дообучения на изображении реальной сцены, используется для обработки новых изображений реальных сцен.

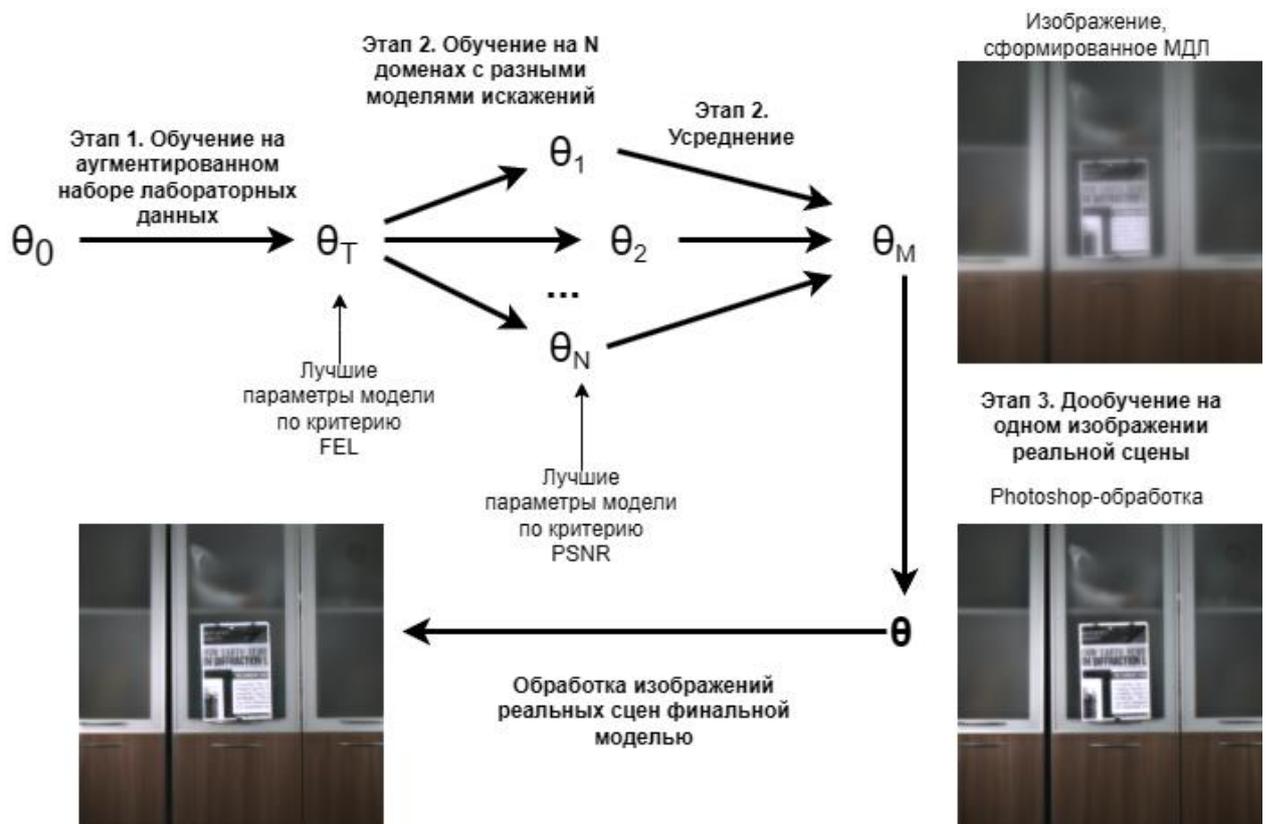


Рисунок 3.10 – Трехэтапный алгоритм адаптации нейросетевой модели к различным условиям формирования реального изображения на основе крупношагового метаобучения

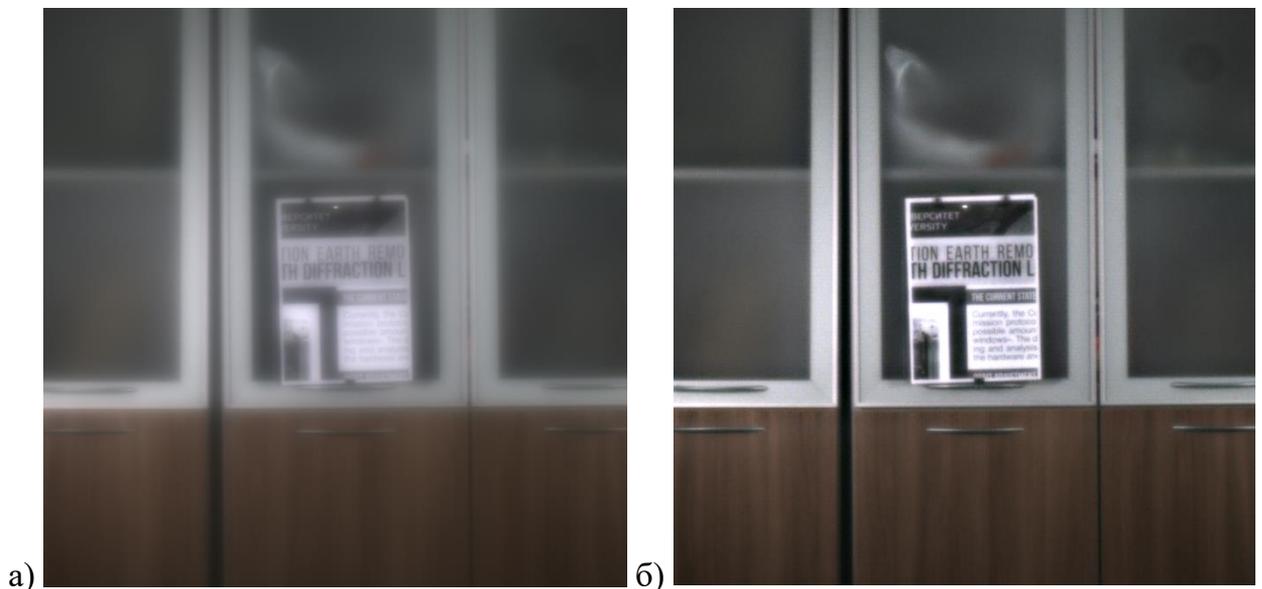


Рисунок 3.11 – Пример реального изображения (а) и результата его обработки в Photoshop (б)

Результаты экспериментальных исследований данного подхода представлены в главе 4. Преимуществом усреднения моделей, а не траекторий градиентного спуска является то, что при появлении нового домена «средняя»

модель  $\theta_M$  формируется с минимальными вычислительными затратами. Достаточно дообучить модель  $\theta_T$  на новом домене, и добавить новую модель  $\theta_{N+1}$  в уже существующий стек из  $N$  моделей. При усреднении траекторий градиентного спуска второй этап необходимо повторить полностью, дообучая модель  $\theta_T$  на всем стеке из  $N+1$  моделей. Ниже представлен псевдокод описанного трехэтапного алгоритма.

---

**Псевдокод алгоритма адаптации модели на основе крупношагового метаобучения: этапы 1–2**

---

**Input:**  $D_{train} = \{x^{(j)}, y^{(j)}\}$  и  $D_{val} = \{x^{(j)}, y^{(j)}\}$  – обучающий и валидационный наборы изображений;  
 $\alpha$  – скорость обучения;  
 batch\_size – размер пакета;  
 real\_patch – фрагмент реального изображения.

**Output:**  $\theta = (\theta_1, \dots, \theta_M)$  – значения параметров усредненной модели.

- 1 Аугментация данных  $D_{train}^{aug}$ .
- 2 Инициализация весов  $\theta$  случайными значениями
- 3 **for** all epochs **do**
- 4     Получение пакета изображений из  $D_{train}^{aug}$
- 5     Вычисление функции потерь MSE и градиента в точке  $\theta$
- 6
- 7     Шаг градиентного спуска. Обновление весов  $\theta$
- 8 **end**
- 9 выбор лучшей точки  $\theta$  в пространстве весов по критерию FEL на фрагменте real\_patch
- 10 Генерация  $N$  пар наборов данных  $T_i^{train}$  и  $T_i^{val}$ ,  $i = \overline{1, N}$  для обучения и валидации, путем применения различных типов искажений, характерных для ДОС, к наборам  $D_{train}$  и  $D_{val}$ .
- 11 **for** all  $(T_i^{train}, T_i^{val})_i$  **do**
- 12     Инициализация весов  $\theta_i = \theta$
- 13     **for** all epochs **do**
- 14         Вычисление функции потерь MSE и ее градиента в точке  $\theta_i$  для  $T_i^{train}$

```

15         Шаг градиентного спуска. Обновление весов
            $\theta_i \leftarrow \theta_i - \alpha \nabla_{\theta} L_{T_i^{train}}(\theta_i)$ 
16     end
17     выбор лучшей точки  $\theta_i$  в пространстве весов по
       критерию PSNR для  $T_i^{val}$ :
            $best\_epoch = \underset{epoch}{\operatorname{argmax}} \operatorname{meanPSNR}(T_i^{val})$ 
18 end
19 Усреднение  $\theta_i$ :

```

$$\theta_{Mean} = \frac{1}{N} \sum_{i=1}^N \theta_i^{best\_epoch}$$

---

### Этап 3. Обучение на одном реальном изображении

---

```

Input: real_image - реальное изображение; ph_image улучшенное
экспертом real_image;  $\theta_{Mean}$  - значения параметров
усредненной модели; steps - количество итераций;
real_patch - фрагмент реального изображения
Output:  $\theta_{real} = (\theta_1, \dots, \theta_M)$  - значения параметров финальной модели
20 Инициализация параметров модели  $\theta$  значениями параметров
     $\theta_{Mean}$ 
21 for all steps do
22     Вычисление функции потерь MSE и ее градиента в точке
23     Шаг градиентного спуска. Обновление весов:
            $\theta \leftarrow \theta - \alpha \nabla_{\theta} Loss(\theta)$ 
24 End
25 Выбор лучшей точки  $\theta$  в пространстве весов по критерию
    FET на фрагменте real_patch
26 Обработка реальных изображений

```

Псевдокод оригинального алгоритма мелкошагового метаобучения с усреднением траекторий градиентного спуска выглядит следующим образом:

---

Псевдокод оригинального алгоритма мелкошагового метаобучения с усреднением траекторий градиентного спуска [39]

**Строки 1-10 повторяются**

---

---

```

11 for all steps do
12     for all  $T_i^{train}$  do
13         Вычисление функции потерь MSE и ее градиента в
           точке  $\theta$  для  $T_i^{train}$ 
14         Шаг градиентного спуска. Обновление весов:
           
$$\theta_i \leftarrow \theta - \alpha \nabla_{\theta} L_{T_i^{train}}(\theta)$$

15     end
16 Обновление весов на мета-шаге:
           
$$\theta \leftarrow \theta - \beta \nabla_{\theta} \sum_{i=1, \overline{N}} L_{T_i^{val}}(\theta_i)$$

17 end

```

### 3.5.3 Теоретическое обоснование метода крупношагового метаобучения

Согласно [43], мелкошаговое метаобучение состоит из следующей последовательности шагов. Расчет оценки параметров для каждого набора данных (домена)  $T_i$  при  $i = \overline{1, N}$ :

$$\theta'_i = \theta - \alpha \nabla_{\theta} L_{T_i}(f_{\theta}), \quad (3.3)$$

где  $\theta$  – вектор параметров модели  $f_{\theta}$ ,  $L_{T_i}$  – функция потерь, рассчитанная по набору данных  $T_i$ ,  $\alpha$  – параметр скорости обучения.

Расчет градиентного шага метаобучения для оценки параметров выглядит следующим образом:

$$\theta \leftarrow \theta - \beta \nabla_{\theta} \sum_{i=1}^N L_{T_i}(f_{\theta'_i}), \quad (3.4)$$

где  $\beta$  – параметр скорости метаобучения.

Согласно [43] последовательность (3.3) – (3.4) приводит к вычислению гессиана. Также в [43] экспериментально показано, что аппроксимация гессиана первым порядком, возможно, приводит к незначительному снижению точности. Согласно [43], для сетей с ReLU активацией это может обосновано тем, что для ReLU активаций вторая производная почти всегда близка к нулю [102]. При таком упрощении (3.4) примет вид:

$$\theta \leftarrow \theta - \beta \nabla_{\theta} \sum_{i=1}^N L_{T_i}(f_{\theta}). \quad (3.5)$$

В предположении, что градиентные шаги (3.5) выполняются в течение  $K$  итераций финальная оценка параметров примет вид:

$$\boldsymbol{\theta}_K = \boldsymbol{\theta} - \beta \sum_{k=1}^K \left( \nabla_{\boldsymbol{\theta}} \sum_{i=1}^N L_{T_i}(f_{\boldsymbol{\theta}_k}) \right) = \boldsymbol{\theta} - \beta \sum_{i=1}^N \sum_{k=1}^K \nabla_{\boldsymbol{\theta}} L_{T_i}(f_{\boldsymbol{\theta}_k}). \quad (3.6)$$

Следующие рассуждения будут также верны в том случае, когда для разных наборов данных будет использоваться разное число итераций. С другой стороны, для каждого из наборов данных финальная оценка параметров, полученная за  $K$  шагов примет вид:

$$\boldsymbol{\theta}_{i,K} = \boldsymbol{\theta} - \alpha \sum_{k=1}^K \nabla_{\boldsymbol{\theta}} L_{T_i}(f_{\boldsymbol{\theta}_k}). \quad (3.7)$$

Представляет интерес выражение оценки (3.6) через среднее значение оценок (3.7).

Выражение для средних оценок (3.7) примет вид:

$$\frac{1}{N} \sum_{i=1}^N \boldsymbol{\theta}_{i,K} = \boldsymbol{\theta} - \alpha \frac{1}{N} \sum_{i=1}^N \sum_{k=1}^K \nabla_{\boldsymbol{\theta}} L_{T_i}(f_{\boldsymbol{\theta}_k}) \quad (3.8)$$

Тогда:

$$\alpha \frac{1}{N} \sum_{i=1}^N \sum_{k=1}^K \nabla_{\boldsymbol{\theta}} L_{T_i}(f_{\boldsymbol{\theta}_k}) = \boldsymbol{\theta} - \frac{1}{N} \sum_{i=1}^N \boldsymbol{\theta}_{i,K}. \quad (3.9)$$

Используя (3.9), второе слагаемое в (3.6) представляется как:

$$\beta \sum_{i=1}^N \sum_{k=1}^K \nabla_{\boldsymbol{\theta}} L_{T_i}(f_{\boldsymbol{\theta}_k}) = N \frac{\beta}{\alpha} \boldsymbol{\theta} - \frac{\beta}{\alpha} \sum_{i=1}^N \boldsymbol{\theta}_{i,K}. \quad (3.10)$$

Используя выражение (3.10), можно записать (3.7) в следующем виде:

$$\boldsymbol{\theta}_k = \left( 1 - \frac{\beta}{\alpha} N \right) \boldsymbol{\theta} + \frac{\beta}{\alpha} \sum_{i=1}^N \boldsymbol{\theta}_{i,K}. \quad (3.11)$$

При введении параметра  $\gamma = \frac{\beta}{\alpha}$  выражение (3.11) принимает следующий вид:

$$\boldsymbol{\theta}_k = (1 - \gamma N) \boldsymbol{\theta} + \gamma \sum_{i=1}^N \boldsymbol{\theta}_{i,K}. \quad (3.12)$$

Случай при  $\gamma = \frac{1}{N}$ , приводит (3.12) к виду (3.2), используемому в настоящей работе в качестве основной формулы для оценки крупношагового метаобучения.

Отметим, что  $\alpha$  и  $\beta$  в (3.11) малые положительные числа, меньшие единицы и представляющие собой параметр скорости обучения.

Альтернативный способ обоснования крупношагового метаобучения предлагается, выражая (3.6) через (3.7), преобразуя (3.6) к следующему виду:

$$\frac{\alpha}{\beta} \boldsymbol{\theta}_K = \frac{\alpha}{\beta} \boldsymbol{\theta} - \sum_{i=1}^N \left( \alpha \nabla_{\boldsymbol{\theta}} \sum_{k=1}^K L_{T_i}(f_{\boldsymbol{\theta}_k}) \right). \quad (3.13)$$

Далее выполняется подстановка выражения (3.7) в (3.13):

$$\frac{\alpha}{\beta} \boldsymbol{\theta}_K = \frac{\alpha}{\beta} \boldsymbol{\theta} - \sum_{i=1}^N (\boldsymbol{\theta} - \boldsymbol{\theta}_{i,K}), \quad (3.14)$$

$$\frac{\alpha}{\beta} \boldsymbol{\theta}_K = \frac{\alpha}{\beta} \boldsymbol{\theta} - N\boldsymbol{\theta} + \sum_{i=1}^N \boldsymbol{\theta}_{i,K}, \quad (3.15)$$

$$\boldsymbol{\theta}_K = \boldsymbol{\theta} \left( \mathbf{1} - \frac{\beta}{\alpha} N \right) + \frac{\beta}{\alpha} \sum_{i=1}^N \boldsymbol{\theta}_{i,K}. \quad (3.16)$$

После введения параметра  $\gamma = \frac{\beta}{\alpha}$  выражение (3.16) принимает вид:

$$\boldsymbol{\theta}_K = \boldsymbol{\theta}(\mathbf{1} - \gamma N) + \gamma \sum_{i=1}^N \boldsymbol{\theta}_{i,K}. \quad (3.17)$$

Финально выражение (3.17) можно представить в виде:

$$\boldsymbol{\theta}_K = \boldsymbol{\theta}(\mathbf{1} - \gamma N) + \gamma N \left( \frac{1}{N} \sum_{i=1}^N \boldsymbol{\theta}_{i,K} \right). \quad (3.18)$$

### Выводы к главе 3

1. Артефакты реконструкции возникают из-за различий в условиях формирования реальных изображений и обучающей выборки. Причинами артефактов реконструкции могут быть разные уровни экспозиции и усиление (gain) или светочувствительность (ISO) камеры. Поскольку артефакты реконструкции возникают только после обработки полнокадровой сетью, дальнейшие исследования с целью уменьшения количества артефактов будут проводиться только для этапа полнокадровой реконструкции.
2. Реконструкция реальных изображений в ДОС относится к задаче с неизвестной моделью искажений и задаче доменной адаптации, поскольку использование FEL позволяет учитывать информацию о новом домене в процессе обучения сети.
3. Предлагаемая аугментация обучающего набора данных проводится с целью увеличения репрезентативности выборки, уменьшая различия доменов обучающего набора данных и изображений реальных сцен.
4. Метод крупношагового метаобучения для задачи реконструкции изображений в ИДОС требует меньших вычислительных затрат на обучение сети при появлении нового домена за счет усреднения параметров моделей, а не траекторий градиентного спуска.

## ГЛАВА 4. ЭКСПЕРИМЕНТАЛЬНЫЕ ИССЛЕДОВАНИЯ

В настоящей главе приведены результаты экспериментальных исследований двухэтапной схемы нейросетевой реконструкции, аугментации обучающего набора данных и алгоритма адаптации нейросетевой модели на основе крупношагового метаобучения.

### 4.1 Экспериментальное исследование алгоритмов сквозной нейросетевой реконструкции изображений

#### 4.1.1 Оценка качества сквозной нейросетевой реконструкции для монокулярной и бинокулярной ОС на основе МДЛ ( $m=13$ )

В данном разделе сравниваются результаты реконструкции на основе метода сквозной нейросетевой реконструкции для монокулярной и бинокулярной ИДОС, а также приведено сравнение результатами классического метода, основанного на обратной свертке [61].

В ранних работах авторы использовали синтезированные изображения [29] или проводили оценку качества реконструкции только для центральной части изображений [7]. Данное упрощение было сделано из-за сложной природы искажений, которые служили причиной низких значений PSNR для полноразмерного изображения. За счет использования полнокадровой сети предложенный в данной работе двухэтапный подход эффективно обрабатывает нелокальные искажения, позволяя оценивать PSNR для целого изображения.

Для проведения экспериментов использовались два набора данных, сформированных в рамках работы [7] с использованием монокулярной и бинокулярной систем на основе МДЛ ( $m=13$ ). Набор данных для монокулярной системы составил 350 пар искаженных изображений, полученных с помощью МДЛ ( $m=13$ ), и эталонных изображений размера  $1024 \times 1024$ . Второй набор данных, сформированный на основе бинокулярной системы, включает 148 искаженных изображений размера  $1024 \times 1024$  для каждой из двух линз и соответствующие им эталонные изображения. Процедура формирования набора данных описана в разделе 1.5.

Наборы данных случайным образом разбиты на обучающую и тестовую выборки. Обучающие выборки для монокулярной и бинокулярной систем составили 321 и 119 изображений соответственно. Тестовый набор данных в обоих случаях состоял из 29 изображений. В качестве полнокадровой сети использовалась архитектура Pix2Pix, которая в случае бинокулярной системы обучалась на 238 изображениях.

Для обучения локальных сетей VDSR и DRRN изображения разбиты на фрагменты размера  $31 \times 31$  с шагом 31 (размер рецептивного поля). Кроме того, обучающая выборка была расширена путем поворота изображений на  $90^\circ$ ,  $180^\circ$ ,  $360^\circ$ , операции «отражения» (flip) по горизонтали и добавления Гауссовского шума с нулевым средним и дисперсией, равной 0,01. В результате для каждого фрагмента изображения из обучающей выборки было сформировано 15 дополнительных фрагментов.

Согласно работе [7], для набора данных, полученных с помощью бинокулярной системы, в качестве предварительной обработки данных были проведены сопоставление изображений, сформированных разными линзами, и коррекция перспективных искажений. Первое изображение, полученное на основе МДЛ, расположенной в оптической системе слева, было использовано в качестве базового, а второе изображение, снятое МДЛ, расположенной справа, приводилось к системе координат первого.

Обучение полнокадровой сети Pix2Pix проводилось на полноразмерных изображениях обучающей выборки с помощью алгоритма оптимизации Adam [103]. Параметр скорости обучения (learning rate) составил 0,0002, а параметр момента (momentum) оптимизатора был установлен, равным 0,5. Функция потерь генеративно-сопоставительной сети представлена в разделе 2.2.

Для обучения сетей VDSR и DRRN использовался также алгоритм оптимизации Adam. Размер мини-пакета составил 64 фрагмента. Глубина сети VDSR, начальное значение скорости обучения, значение коэффициентов  $\eta$  и  $\lambda$  в функции потерь (2.5) были установлены, равными 18,  $10^{-4}$ ,  $10^{-4}$  и  $10^{-2}$  соответственно. Функция потерь (2.5) для обучения сети VDSR учитывает гипотезу об ахроматичности контуров и описана в разделе 2.3.

В таблице 4.1 указаны значения PSNR для описанного в данной работе подхода и для метода, основанного на обратной свертке из [3, 61]. В работах [5, 61] представлена более подробная информация и сравнение методов реконструкции на основе обратной свертки. Также в таблице указаны средние значения PSNR для входных искаженных изображений, изображений после полнокадровой реконструкции и восстановленного изображения. Средние значения PSNR рассчитывались по тестовому набору, состоящему из 29 изображений.

Из таблицы видно, что для монокулярной системы после полнокадровой реконструкции прирост среднего значения PSNR составил 8,3 дБ. Прирост среднего значения PSNR после локальной реконструкции сетью DRRN составил 0,45 дБ. Сеть VDSR не справилась с локальной реконструкцией в данной задаче. Общий прирост среднего значения PSNR для монокулярной системы составил 8,75 дБ, что на 7,55 дБ больше, чем при использовании классического подхода на основе обратной свертки.

Для бинокулярной системы общий прирост среднего значения PSNR составил 8,86 дБ: 8,2 дБ после полнокадровой сети и 0,66 дБ после локальной сети.

Значимость небольшого прироста PSNR (0,45 дБ и 0,66 дБ) при локальной реконструкции по сравнению с полнокадровой подтверждается при исследовании публикаций по теме задачи сверхразрешения. В работах [84, 85] даже прирост среднего значения PSNR около 0,1 дБ учитывается для обоснования значимости улучшений в архитектуре сети. Поэтому в данной работе на этапе локальной реконструкции значимым считается прирост PSNR, превышающий 0,1 дБ.

На рисунке 4.1 представлены графики среднего значения PSNR, рассчитанного на валидационной выборке в процессе обучения 3-х и 6-ти канальных VDSR. Как видно из графика, 6-канальная VDSR достигает почти такого же качества реконструкции, что и трехканальная, уже на 12 эпохе, тем самым показывая более быструю сходимость.

На рисунках 4.2-4.3 представлены примеры восстановленных изображений из тестовой выборки для монокулярной и бинокулярной систем, которые демонстрируют качество восстановленных изображений, сопоставимое с эталонами. Результаты экспериментов демонстрирует эффективность предлагаемого

двухэтапного подхода к реконструкции изображений в монокулярных и бинокулярных ДОС по сравнению с ранними методами на основе обратной свертки.

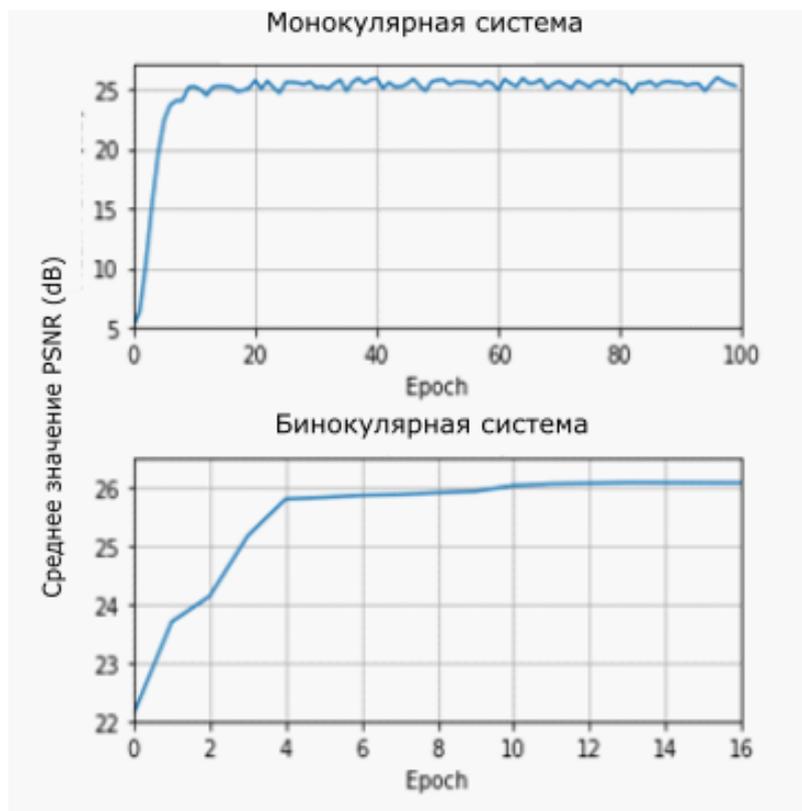


Рисунок 4.1 – График среднего значения PSNR на валидационной выборке в процессе обучения локальной сети

Таблица 4.1 – Результаты реконструкции тестового набора изображений

Метод реконструкции	Входное изображение (PSNR, дБ)	Изображение после полнокадровой реконструкции (PSNR, дБ)	Восстановленное изображение (PSNR, дБ)
Обратная свертка (Монокулярная система)	17,21	–	18,41
Pix2pix+ трехканальная VDSR (Монокулярная система)	17,21	25,51	25,59
Pix2pix+ трехканальная DRRN (Монокулярная система)	17,21	25,51	25,96
Pix2Pix + шестиканальная VDSR (Бинокулярная система)	17,21 (Слева), 17,19 (Справа)	25,41 (Слева), 25,08 (Справа)	26,07

В рамках данного эксперимента видно, что сеть DRRN хорошо сработала в комбинации с сетью Pix2Pix. Однако в дальнейших экспериментах при замене полнокадровой сети на U-Net в качестве локальной сети лучшие результаты показала архитектура VDSR, поэтому она используется далее в разделе 4.1.4. В приложении Б приведена таблица с результатами исследований различных комбинаций локальных и полнокадровых архитектур.



Рисунок 4.2 – Пример изображения из тестовой выборки: а) сформированное МДЛ ( $t=13$ ), б) восстановленное для монокулярное системы, в) восстановленное для бинокулярной системы, г) эталонное

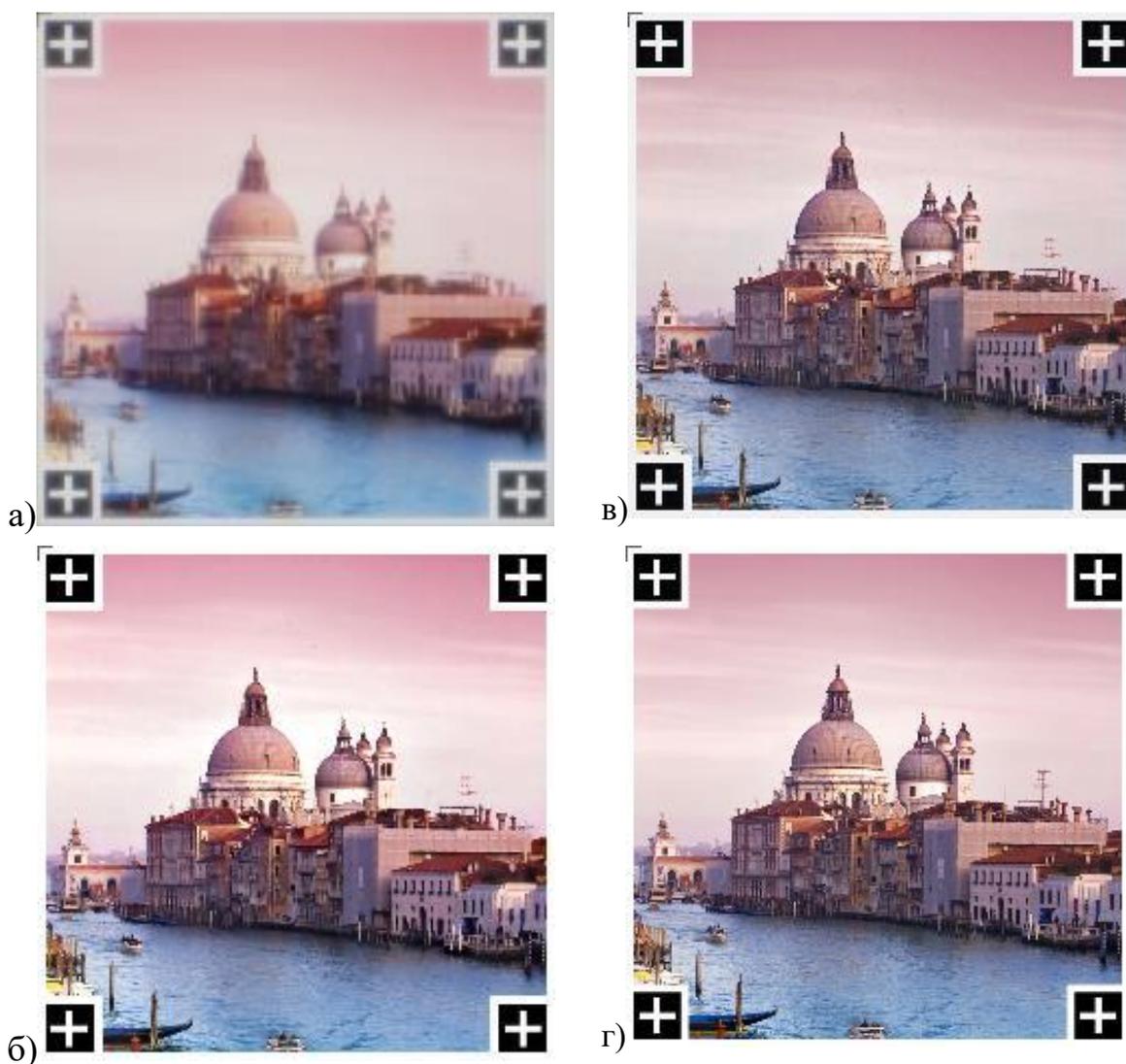


Рисунок 4.3 – Пример изображения из тестовой выборки: а) сформированное МДЛ ( $m=13$ ), б) восстановленное для монокулярное системы, в) восстановленное для бинокулярной системы, г) эталонное

#### 4.1.3 Исследование влияния технологии изготовления МДЛ на качество реконструкции

Ранее исследования двухэтапной схемы реконструкции проводились на небольшом объеме данных (350 уникальных изображений), сформированном только для МДЛ ( $m=13$ ). В данном разделе предлагается увеличить объем обучающей выборки и провести обучение двухэтапной схемы реконструкции для двух разных ИДОС на основе МДЛ ( $m=7$ ) и МДЛ ( $m=13$ ) с разной высотой микрорельефа.

Формирование данных для обучения осуществлялось в рамках работы [14] с использованием стенда в помещении. В состав стенда входят: откалиброванный монитор, разрешение которого составляет 1920x1080 пикселей; камера Basler

асА1920-40u, интерфейс USB 3.1 Gen 1, разрешение 1920x1200 пикселей; ноутбук; объективы с резьбовым креплением C-mount, в которые установлены МДЛ. Штатив с камерой установлен таким образом, чтобы в область кадра камеры целиком умещался экран монитора. Внешние источники света были выключены, помещение затемнено. Выдержка и баланс белого установлены по выведенному на экран монитора изображению с заливкой белого цвета. Фокусировка объектива осуществлена вручную с использованием миры «Siemens star», установленной в плоскости монитора на таком же расстоянии.

Набор изображений для формирования обучающей выборки включает 1831 эталонных изображений и одно калибровочное изображение с маркерами. Формирование изображений осуществлялось в один проход для каждой МДЛ. Сопоставление эталонных изображений с полученными снимками осуществлялось с использованием процедуры, описанной в разделе 1.5.

В результате сформированы два набора изображений, полученных на основе МДЛ ( $m=7$ ) и МДЛ ( $m=13$ ). Каждый набор данных включает 1831 пару изображений (эталонное и сформированное МДЛ) размера  $1024 \times 1024$  и случайным образом разбит на обучающую (1226 изображений) и тестовую (605 изображений) выборки. Валидационная выборка включает 15 тестовых изображений.

В рамках данного эксперимента в качестве полнокадровой сети используется модификация архитектуры U-Net, описанная в разделе 2.2. Обучение U-Net проводилось на полноразмерных изображениях ( $1024 \times 1024$ ) обучающей выборки с помощью мини-пакетного стохастического градиентного спуска и алгоритма оптимизации Adam [103]. Параметр скорости обучения (learning rate) составил 0,0002,  $\beta_1 = 0,5$ ,  $\beta_2 = 0,999$ ,  $\epsilon = 10^{-8}$ . Выбор функции потерь для обучения сети U-Net проводился на основе результатов исследования качества реконструкции при использовании стандартных для задачи сверхразрешения функций потерь [89], а также функции Хьюбера [90]. Результаты оценки среднего значения PSNR для разных функций потерь представлены в таблице 4.2.

Для оценки качества реконструкции использовалось среднее значение PSNR на тестовой выборке. Видно, что для МДЛ ( $m=13$ ) наилучшее качество реконструкции обеспечивает среднеквадратическое отклонение восстановленного изображения от эталона  $L^2$ , а для МДЛ ( $m=7$ ) лучший результат достигнут благодаря

использованию функции Хьюбера. Таким образом функция потерь  $L^{Huber}$  и  $L^{l_2}$  выбраны для обучения полнокадровой сети U-Net. Процедура обучения локальной сети описана в разделе 4.1.1.

Таблица 4.2 – Оценка качества реконструкции для разных функций потерь

Функция потерь	PSNR для МДЛ (m=7 13) (дБ)	PSNR для МДЛ (m=7) (дБ)
$L^{l_1}$	24,67	26,54
$L^{l_2}$	27,02	26,44
$L^{SSIM}$	25,6	24,94
$L^{MS-SSIM}$	25,9	25,03
$L^{Mix_1}$	26,28	26,57
$L^{Mix_2}$	26,37	26,55
$L^{Huber}$	25,86	26,65

Оценка качества метода сквозной нейросетевой реконструкции на основе сети U-Net представлена в таблице 4.3, где указаны средние значения PSNR, рассчитанные на тестовой выборке (605 изображений) для исходных изображений, изображений после полнокадровой и локальной реконструкции. Кроме того, в таблице 4.3 указаны средние значения PSNR, рассчитанные для 25 лучших по критерию PSNR изображений (ТОП-25).

Для МДЛ (m=13) прирост среднего значения PSNR составил 7,58 дБ: 7,34 дБ после полнокадровой сети и 0,24 дБ после локальной сети. Для МДЛ (m=7) прирост среднего значения PSNR составил 8,69 дБ (8,52 дБ после полнокадровой сети и 0,17 дБ после локальной сети), что на 1,11 дБ больше, чем в случае с МДЛ (m=13). Несмотря на более низкое качество входных изображений в случае МДЛ (m=7) двухэтапная схема сквозной нейросетевой реконструкции обеспечивает высокий прирост среднего значения PSNR на тестовой выборке, что показывает ее эффективность для разных технологий изготовления ИДОЭ.

Из таблицы 4.3 видно, что предложенная схема реконструкции не только обеспечивает высокий уровень PSNR, но и не зависит от качества входного набора данных, сокращая разницу между средними значениями PSNR для изображений, снятых двумя разными МДЛ, с 1,55 дБ до 0,4 дБ. Визуально качество входных изображений, полученных с помощью МДЛ (m=7) и МДЛ (m=13), можно оценить

по рисунку 4.5, где заметно ухудшение входного качества при использовании МДЛ ( $m=7$ ).

Таблица 4.3 – Среднее значение PSNR на тестовой выборке

m	Исходное	Результат полнокадровой реконструкции (U- Net)	Результат локальной реконструкции (VDSR)	ТОП-25 (VDSR)
13	19,68	27,02	27,26	31,94
7	18,13	26,65	26,82	31,28

На рисунке 4.4 показаны графики изменения функций потерь (функция Хьюбера для МДЛ ( $m=7$ ) и  $l_2$  для МДЛ ( $m=13$ )) и среднего значения PSNR на валидационной выборке (15 изображений из тестовой выборки) в процессе обучения сети U-Net.

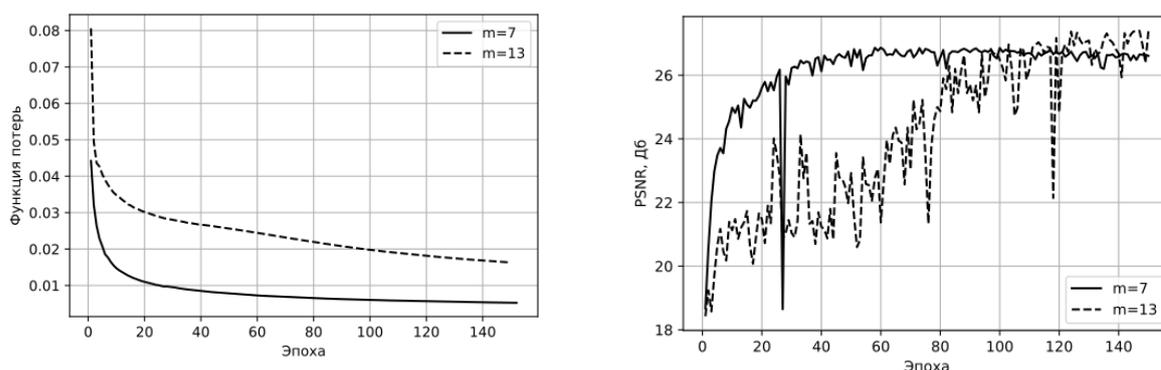
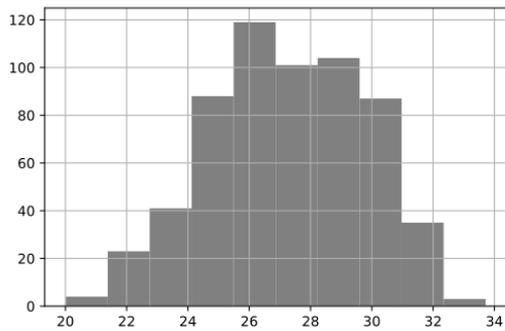


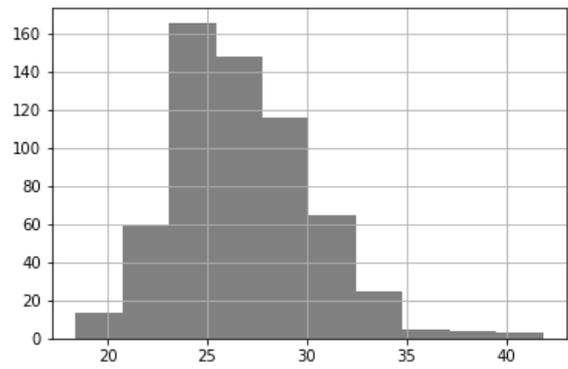
Рисунок 4.4 – Графики обучения сети U-Net: а) функции потерь, б) средний PSNR на валидационной выборке

На рисунке 4.5 представлены гистограммы распределения среднего значения PSNR на тестовой выборке, построенные после двухэтапной реконструкции.

На рисунке 4.6 приведены примеры восстановленных изображений из тестовой выборки. Видно, что визуально качество реконструкции изображений из тестового набора сопоставимо с эталонными изображениями.



а)



б)

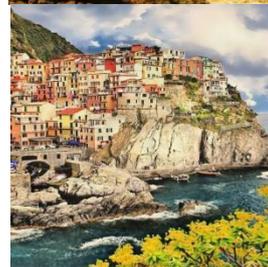
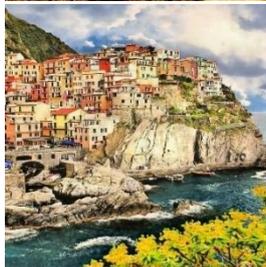
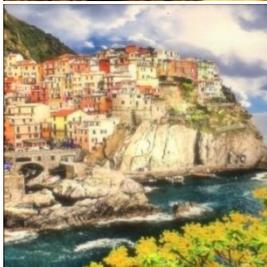
Рисунок 4.5 – Гистограммы распределения PSNR на тестовой выборке для МДЛ ( $m=13$ ) (а) и МДЛ ( $m=7$ ) (б)

Сформированное МДЛ

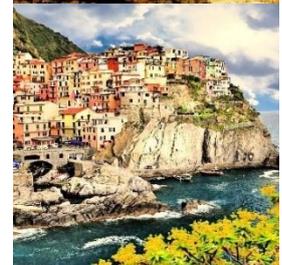
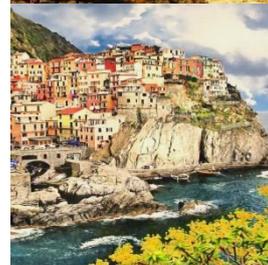
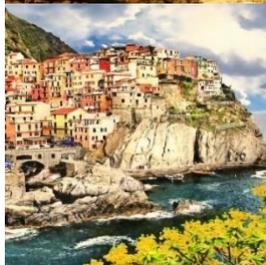
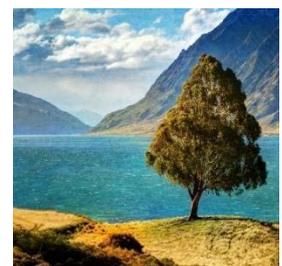
U-Net реконструкция

VDSR реконструкция

Эталонное



а)



б)

Рисунок 4.6 – Примеры изображений из тестовой выборки для МДЛ ( $m=13$ ) (а) и МДЛ ( $m=7$ ) (б)

На рисунках 4.7 и 4.8 приведены результаты полнокадровой реконструкции изображений сцен реального мира с использованием архитектур U-Net и Pix2Pix. Входные изображения (рисунок 4.7.a и рисунок 4.8.a) получены с помощью МДЛ ( $m=13$ ). Видно, что артефакты, характерные при реконструкции сетью Pix2Pix, исчезли в результате обработки сетью U-Net. Таким образом упрощение архитектуры полнокадровой сети (замена Pix2Pix на U-Net) без уменьшения размера рецептивного поля оказала положительное влияние на результирующее качество реконструкции, поэтому в дальнейших экспериментах при полнокадровой реконструкции будет использоваться архитектура U-Net.

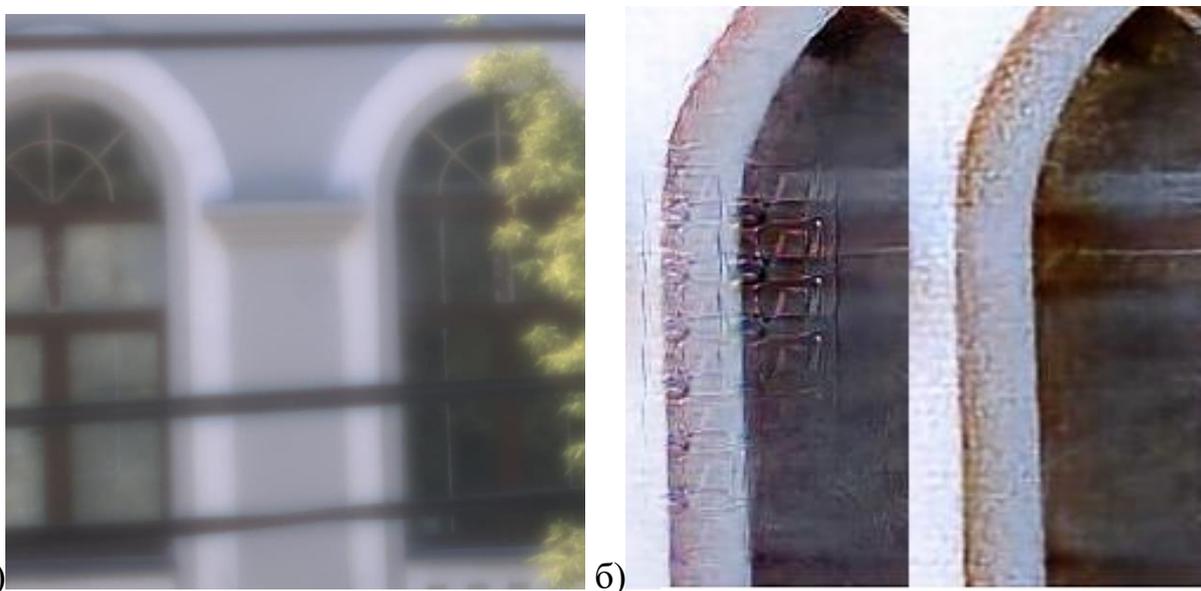


*Рисунок 4.7 – Результат полнокадровой реконструкции изображения сцены реального мира, полученного с помощью МДЛ ( $m=13$ ): а) входное изображение, б) сравнение фрагментов обработанных изображений сетями Pix2Pix (сверху) и U-Net (снизу)*



а) б)  
 Рисунок 4.8 – Результаты полнокадровой реконструкции изображения сцены реального мира, полученного с помощью МДЛ ( $m=13$ ) при естественном освещении: а) входное изображение, б) сравнение фрагментов обработанных изображений сетями Pix2Pix (сверху) и U-Net (снизу).

На рисунке 4.9 приведены результаты полнокадровой реконструкции изображения, сформированного МДЛ ( $m=7$ ).



а) б)  
 Рисунок 4.9 – Результат полнокадровой реконструкции изображения сцены реального мира, полученного с помощью МДЛ ( $m=7$ ) при естественном освещении: а) входное изображение, б) сравнение фрагментов обработанных изображений сетями Pix2Pix (слева) и U-Net (справа)

По сравнению с результатами реконструкции для МДЛ ( $m=13$ ), в случае МДЛ ( $m=7$ ) оба восстановленных сетями Pix2Pix и U-Net изображения (рисунок

4.9.б) имеют артефакты реконструкции. Реконструкция сетью Pix2Pix приводит к более грубым артефактам, продемонстрированным на фрагменте изображения (рисунок 4.9.б). На изображении, восстановленном сетью U-Net, наблюдаются менее заметные артефакты в виде ложных контуров, исследование которых проводится в разделе 4.2.

## **4.2 Исследование аугментации обучающего набора данных и критерия FEL для уменьшения артефактов реконструкции**

Исследование алгоритмов аугментации проводилось на наборах данных, сформированных с помощью двух разных изображающих систем: дифракционно-рефракционной ОС и ДОС на основе трех линз, схемы которых описаны в разделе 1.3. Как было сказано в главе 2, для локальной реконструкции проблема артефактов не столь актуальна, так как локальная сеть обеспечивает более низкий прирост качества в целом. Поэтому исследование с целью уменьшения артефактов реконструкции проводится только для полнокадровой сети.

### **4.2.1 Обучение полнокадровой сети U-Net**

Формирование набора данных на основе дифракционно-рефракционной ОС и ДОС на основе трех линз осуществлялось с использованием камеры Basler acA1920-40uc USB 3.0 и монитора с диагональю 27 дюймов и разрешением 3840×2160 пикселей.

Набор данных состоит из 1857 RGB изображений размера 1024×1024 и нескольких калибровочных изображений размера 1024×1024, содержащих маркеры, для сопоставления снятых кадров с цифровыми оригиналами (рисунок 1.18.а). Разработанное в рамках работ [9, 17] ПО с использованием SDK камеры автоматизирует процесс вывода эталонного изображения на экран монитора и его формирования с выставляемой задержкой захвата изображения.

В случае ДОС на основе трех линз подбор коэффициента баланса белого для красного и синего каналов изображения осуществляется в два этапа. Сначала определяется минимум из максимальных значений выдержки каждой камеры, чтобы исключить возможность переэкспозиции («пересвета»), но и не допустить сужения динамического диапазона регистрируемого изображения. Для этого на экран

монитора выводится RGB изображение белого цвета с интенсивностью (255, 255, 255). Полученное значение выдержки фиксируется для каждой камеры, производится формирование изображения белой заливки экрана, затем вычисляются коэффициенты.

Набор данных случайным образом разбит на обучающую (1244 изображения) и тестовую (613 изображений) выборки. В качестве валидационной выборки используются 21 изображение из тестовой выборки.

Для обучения сети U-Net, архитектура которой описана в разделе 2.2, использовался алгоритм оптимизации ADAM с параметрами  $\beta_1 = 0,5$  и  $\beta_2 = 0,999$  и коэффициентом скорости обучения, равным 0,0002. В качестве функции потерь использовалось среднеквадратическое отклонение восстановленного изображения от эталона.

#### **4.2.2 Критерий выбора лучшей точки в пространстве параметров модели**

В рамках проведенных экспериментов нейронная сеть обучалась на двух наборах данных (для дифракционно-рефракционной ОС и ДОС на основе трех линз) в течение 200 эпох, поскольку при дальнейшем обучении наблюдается стагнация среднего значения PSNR на валидационной выборке. После каждой эпохи рассчитывались среднее значение PSNR на валидационной выборке и значение FEL на фрагментах (200×200 пикселей) реальных изображений, представленных на рисунке 4.10.а и 4.10.б для разных ДОС. Используемый фрагмент выделен рамкой черного цвета. На выбранном фрагменте нет перепадов яркости и визуально не наблюдается контуров. Поэтому все обнаруженные контура на данном фрагменте относятся к артефактам реконструкции. В качестве критерия ранней остановки обучения исследовались критерии PSNR и FEL, описанные в разделе 3.2.

На рисунках 4.11 и 4.12 представлены результаты реконструкции реальных изображений (рисунок 4.10), полученных разными ИДОС, при разных критериях ранней остановки обучения. Приведены как полноразмерные восстановленные изображения, так и фрагменты, по которым рассчитывалось значение FEL.

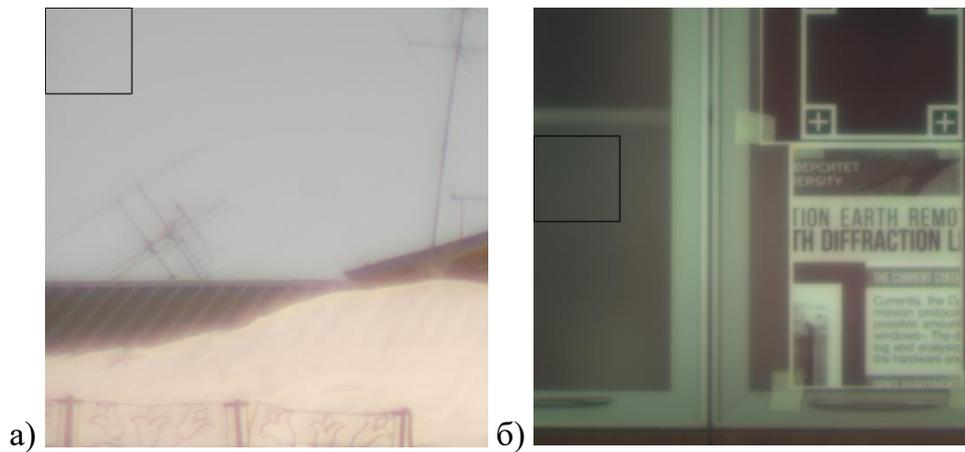


Рисунок 4.10 – Примеры изображений сцен реального мира с выделенным фрагментом  $200 \times 200$  пикселей ограничивающими рамками черного цвета: а) , дифракционно-рефракционная ОС, б) ДОС на основе трех линз

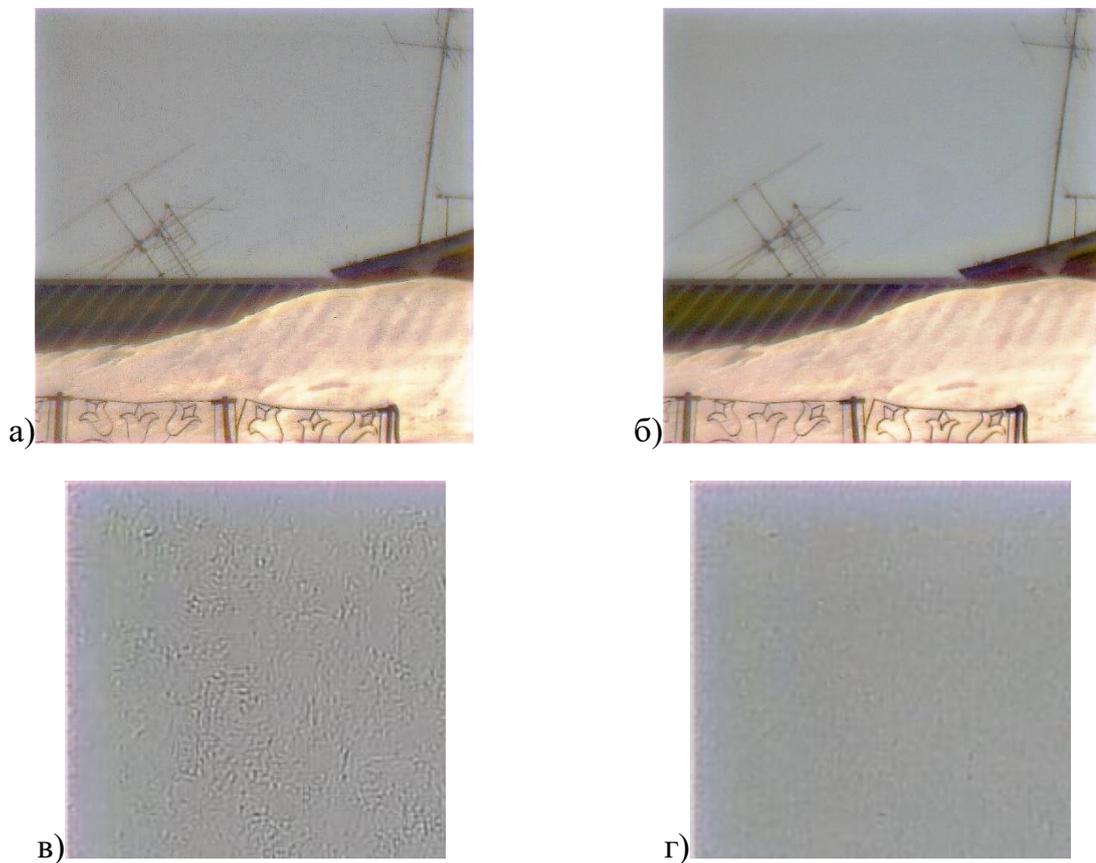


Рисунок 4.11 – Результаты реконструкции реального изображения, сформированного дифракционно-рефракционной ОС (рисунок 4.10.а), для разных критериев выбора лучшей точки в пространстве параметров модели: а) в) критерий PSNR (FEL: 30,69%), б) г) критерий FEL (FEL: 14,48% MOS: 1,75,  $p < 0,05$ )

В случае дифракционно-рефракционной ОС при использовании критерия ранней остановки обучения PSNR значение FEL на фрагменте восстановленного изображения (рисунок 4.11.в) составило 30,69 %, а при использовании критерия FEL (рисунок 4.11.г) – 14,48 %.

Для ДОС на основе трех линз при использовании критерия ранней остановки обучения PSNR значение FEL на фрагменте восстановленного изображения (рисунок 4.12.в) составило 17,51 %, а при использовании критерия FEL (рисунок 4.12.г) – 9,11 %. И визуально, и на основе значения FEL видно, что использование критерия FEL при выборе лучшей точки в пространстве параметров модели обеспечивает меньший уровень артефактов реконструкции по сравнению с результатами для критерия PSNR.

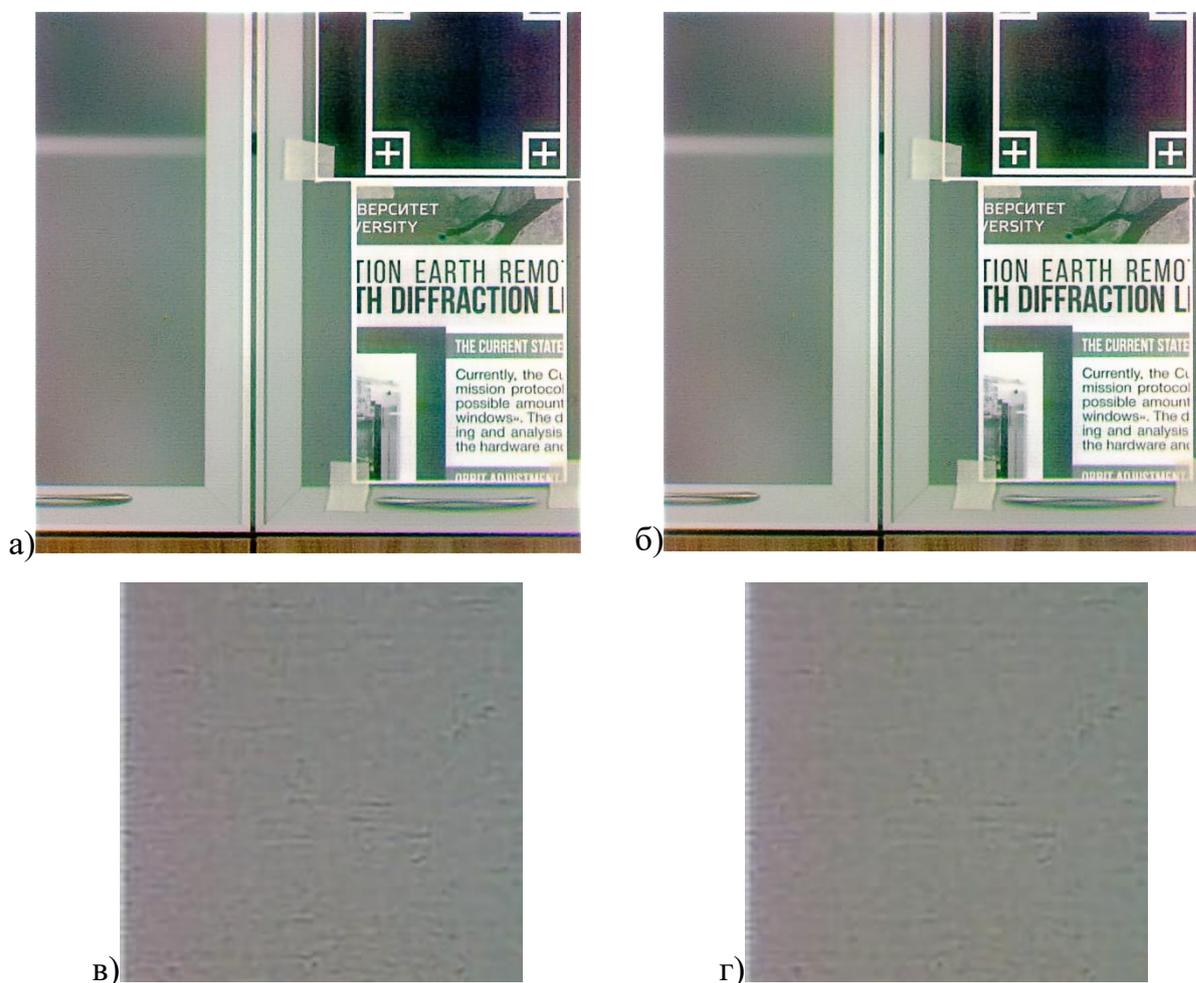


Рисунок 4.12 – Результаты реконструкции реального изображения, сформированного ДОС на основе трех линз (рисунок 4.9.б) для разных критериев выбора лучшей точки в пространстве параметров модели: а) в) критерий PSNR (FEL: 17,51%), б) г) критерий FEL (FEL: 9,11%%)

### 4.2.3 Аугментация обучающих данных моделированием ISO-шума и изменением экспозиции

В рамках данного исследования рассматриваются 3 случая обучения сети: без аугментации данных, аугментация обучающей выборки путем моделирования ISO-шума, а также аугментация моделированием ISO-шума с последующим изменением экспозиции. Зашумление данных проводилось с вероятностью 0,5 и значением параметра интенсивности, выбранным случайным образом из множества {0,1; 0,2; 0,3}. При увеличении значения параметра интенсивности больше 0,3 на входном изображении наблюдается визуально заметный шум, который, как правило, не заметен при визуальной оценке реальных изображений в практических задачах. Изменение экспозиции проводилось с вероятностью 0,1 после аугментации ISO-шумом. Подбор параметров был сделан за счет визуального анализа сходства полученных изображений с реальными данными.

В таблицах 4.4-4.5 представлены результаты расчета среднего значения PSNR на тестовой выборке и FEL, рассчитанного на фрагменте восстановленного изображения реальной сцены для случаев без предобработки обучающей выборки и с предложенными выше видами предобработки, а также указаны значения как при использовании критерия выбора лучшей точки в пространстве параметров модели FEL, так и при использовании критерия PSNR.

При использовании критерия ранней остановки обучения PSNR достигается максимальное значение среднего PSNR на тестовой выборке: 28,09 дБ в случае дифракционно-рефракционной ОС и 25,67 дБ для ДОС на основе трех линз. При использовании критерия ранней остановки обучения FEL в случае дифракционно-рефракционной ОС среднее значение PSNR составило 26,91 дБ, а для ДОС на основе трех линз – 25,32 дБ. Поскольку целью данного исследования является уменьшение артефактов реконструкции на изображениях реальных сцен, то за основу берется критерий FEL и уменьшение среднего значения PSNR на 1,18 дБ (дифракционно-рефракционная ОС) и на 0,35 (ДОС на основе трех линз) по сравнению с уменьшением уровня артефактов реконструкции в данном случае считается незначительным.

Таблица 4.4 – Сравнение критериев качества для разных видов аугментации для дифракционно-рефракционной ОС

№	Аугментация	Критерий ранней остановки обучения	FEL (%)	PSNR на тестовой выборке (дБ)	Восстановленный фрагмент
1	Отсутствует	PSNR	30,69	27,68	Рисунок 4.11.в
		FEL	14,48	27,01	Рисунок 4.11.г
2	Наложение ISO ( $p=0,5$ ) и интенсивностью 0,1	PSNR	3,32	28,09	–
		FEL	2,83	27,65	Рисунок 4.13.а
3	Наложение ISO ( $p=0,5$ ) и интенсивностью, случайно выбираемой из множества $\{0,1; 0,2; 0,3\}$	PSNR	1,47	27,4	–
		FEL	1,35	27,37	Рисунок 4.13.б
4	Наложение ISO ( $p=0,5$ ) с интенсивностью, случайно выбираемой из множества $\{0,1; 0,2; 0,3\}$ и изменение экспозиции	PSNR	1,69	27,08	–
		FEL	1,29	26,91	Рисунок 4.13.в

**В случае дифракционно-рефракционной ОС** минимальное значение FEL (1,29 %) достигается при использовании аугментации ISO-шумом и варьированием экспозиции. На рисунке 4.13 представлены восстановленные фрагменты реального изображения (рисунок 4.10.а) для различных видов аугментации при использовании критерия FEL. Рисунки 4.13.б и 4.13.обеспечивают FEL практически на одном уровне с разницей в 0,06%. Для выбора подходящей аугментации в данном случае проводилось визуальное сравнение полноразмерных восстановленных изображений (рисунок 4.14). Учитывая качество цветов и визуальный уровень шума на полноразмерном изображении, аугментация ISO-шумом и изменением экспозиции обеспечивает лучшее качество реконструкции реального изображения. В приложении В представлены примеры обработанных реальных изображений, не

участвовавших в процессе обучения сети, а также приведены результаты формирования изображения такой же сцены рефракционным объективом.

Таблица 4.5 – Сравнение критериев качества для разных видов аугментации для ДОС на основе трех линз

№	Аугментация	Критерий ранней остановки обучения	FEL (%)	PSNR на тестовой выборке (дБ)	Восстановленный фрагмент
5	Отсутствует	PSNR	17,51	25,54	Рисунок 4.12.в
		FEL	9,11	25,52	Рисунок 4.12.г
6	Наложение ISO ( $p=0,5$ ) и интенсивностью, случайно выбираемой из множества $\{0,1; 0,2; 0,3\}$	PSNR	1,20	25,67	Рисунок 4.15.а
		FEL	0,51	25,32	Рисунок 4.15.б
7	Наложение ISO ( $p=0,5$ ) с интенсивностью, случайно выбираемой из множества $\{0,1; 0,2; 0,3\}$ и изменение экспозиции	PSNR	4,70	25,23	Рисунок 4.15.в
		FEL	0,8	25,27	Рисунок 4.15.г

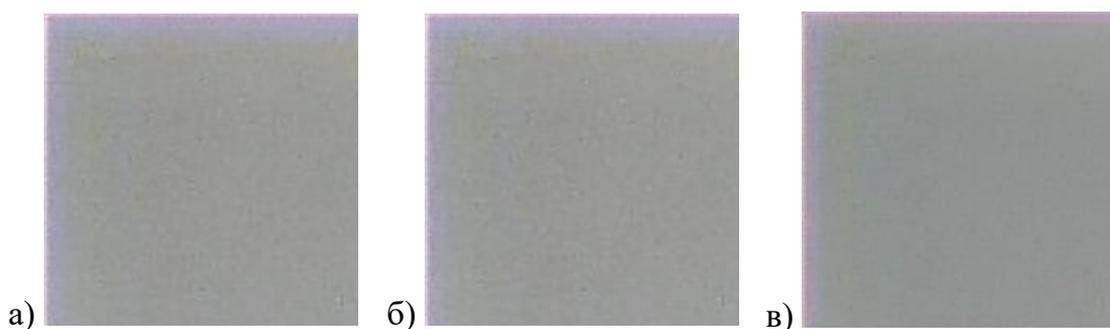
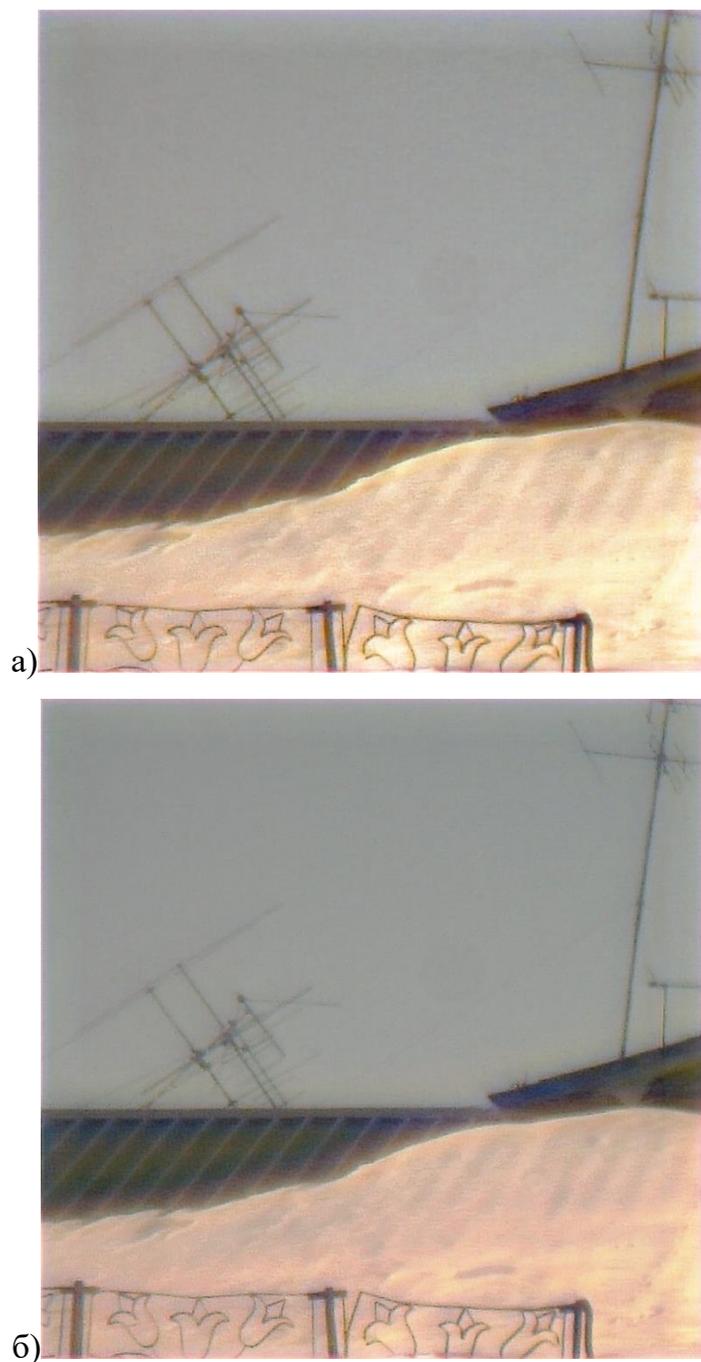


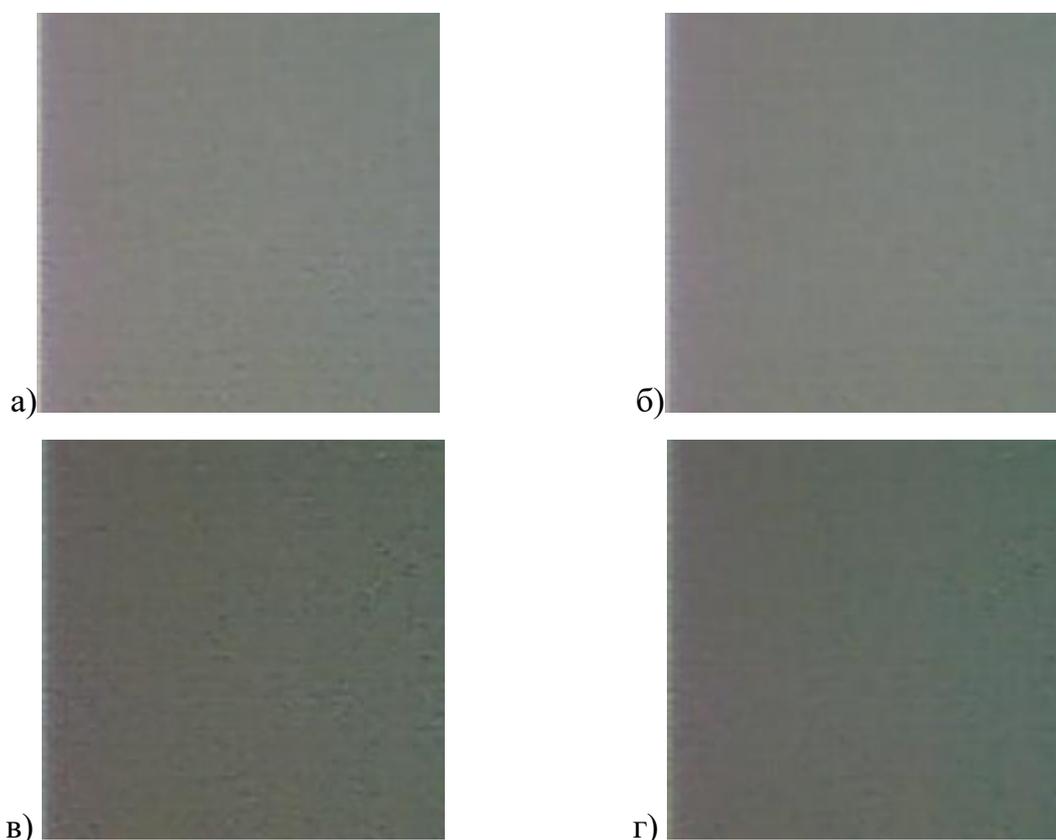
Рисунок 4.13 – Восстановленные фрагменты реального изображения (рисунок 4.10.а), полученного дифракционно-рефракционной ОС для разных видов аугментации: а) наложение шума ISO ( $p=0,5$ ) с интенсивностью 0,1, б) наложение шума ISO ( $p=0,5$ ) с интенсивностью, случайно выбираемой из множества  $\{0,1; 0,2; 0,3\}$ , в) наложение шума ISO ( $p=0,5$ ) с интенсивностью, случайно выбираемой из множества  $\{0,1; 0,2; 0,3\}$  и изменением экспозиции



*Рисунок 4.14 – Результат реконструкции реального изображения (рисунок 4.10.а), полученного дифракционно-рефракционной ОС для разных аугментаций: а) наложение шума ISO ( $p=0,5$ ) с интенсивностью, случайно выбираемой из множества  $\{0,1; 0,2; 0,3\}$ , б) наложение шума ISO ( $p=0,5$ ) с интенсивностью, случайно выбираемой из множества  $\{0,1; 0,2; 0,3\}$  и изменением экспозиции*

Для ДОС на основе трех линз минимальное значение FEL (0,51 %) достигается при использовании аугментации ISO-шумом. Данный результат соответствует визуально меньшему числу артефактов реконструкции на рисунке 4.15. На рисунке 4.15 указаны восстановленные фрагменты реального изображения,

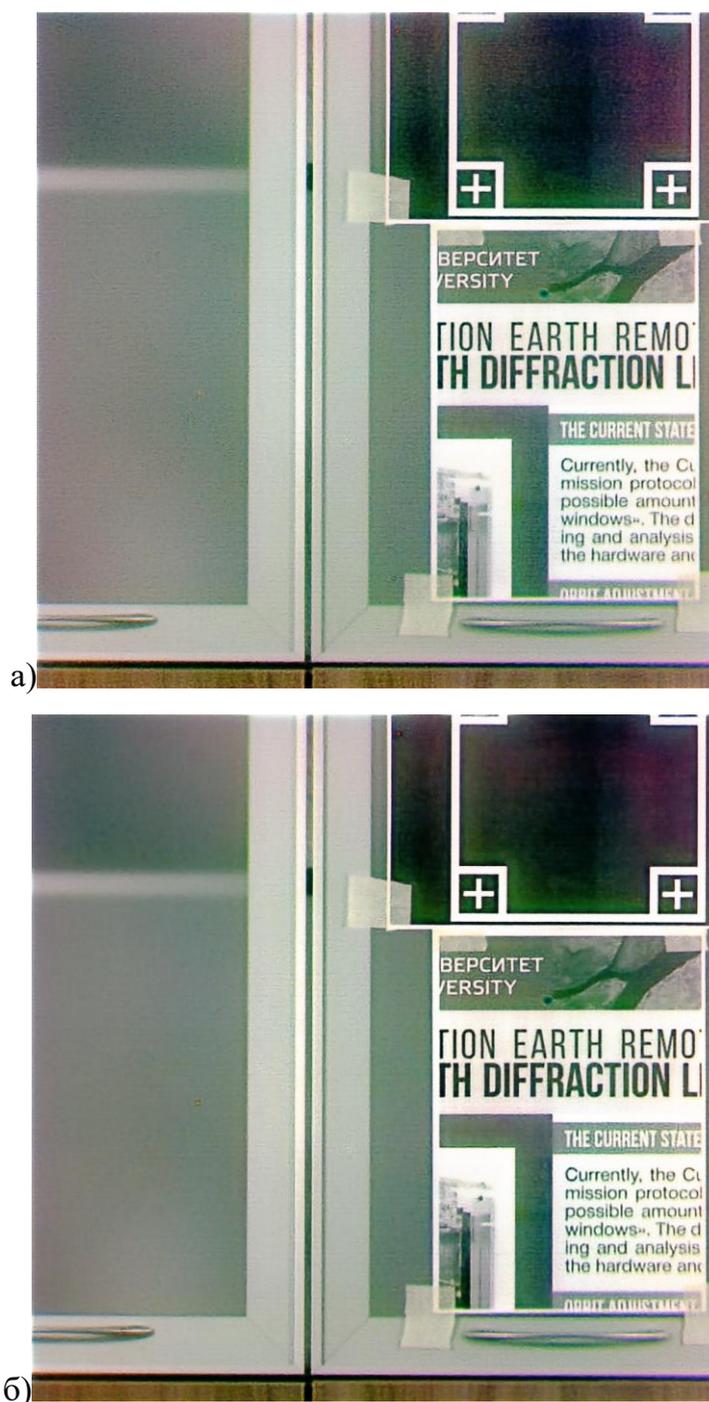
представленного на рисунке 4.10.б. Результаты в таблице 4.4 и на рисунке 4.15 еще раз подтверждают эффективность критерия FEL для выбора точки в пространстве параметров модели, обеспечивающей снижение количества артефактов реконструкции на изображениях реальных сцен.



*Рисунок 4.15 – Результат реконструкции фрагмента изображения реальной сцены для ДООС на основе трех линз (рисунок 4.10.б): а) аугментация обучающей выборки моделированием ISO-шума, использование критерия PSNR, б) аугментация обучающей выборки моделированием ISO-шума, использование критерия FEL, в) аугментация обучающей выборки моделированием ISO-шума и изменением экспозиции, использование критерия PSNR, г) аугментация обучающей выборки моделированием ISO-шума и изменением экспозиции, использование критерия FEL*

На рисунке 4.16 представлено сравнение результатов реконструкции изображения реальной сцены (рисунок 4.10.б) при использовании критерия FEL для случая аугментации ISO-шумом и при отсутствии аугментации. Видно, что в случае аугментации обучающей выборки ISO-шумом наблюдается меньшее количество артефактов и визуально они становятся менее заметными. Стоит отметить, что при визуальной оценке фрагментов, содержащих текст, на рисунках 4.10.б и 4.16 видно,

что текст нечитаемый на изображении 4.10.б становится читаемым после проведенной нейросетевой реконструкции.



*Рисунок 4.16 – Результаты реконструкции полноразмерного изображения реальной сцены, представленного на рисунке 4.10.б, при использовании критерия FEL: а) обучение сети без аугментации обучающей выборки, б) применение аугментации ISO-шумом*

На рисунках 4.17 и 4.18 представлены графики обучения сети для случаев с аугментацией и без аугментации обучающих данных. На рисунке 4.17 приведены

графики для дифракционно-рефракционной ОС, а на рисунке 4.18 для ДОС на основе трех линз. Видно, что использование аугментации обеспечивает более высокий уровень PSNR на валидационной выборке для случая ДОС на основе трех линз (рисунок 4.18.а), но для дифракционно-рефракционной ОС наблюдается обратная ситуация (рисунок 4.17.а). Как уже отмечалось ранее этот негативный эффект компенсируется значительным улучшением визуального качества реконструкции изображения реальной сцены с точки зрения количества артефактов, которое измеряется с помощью FEL. С точки зрения FEL по графикам на рисунках 4.17.б и 4.18.б видно, что при обучении на исходных данных без аугментации появляется проблема переобучения, при которой значение FEL достигает некоторого локального минимума (около 80 эпохи), после чего начинает расти. При использовании аугментации данная проблема исчезает.

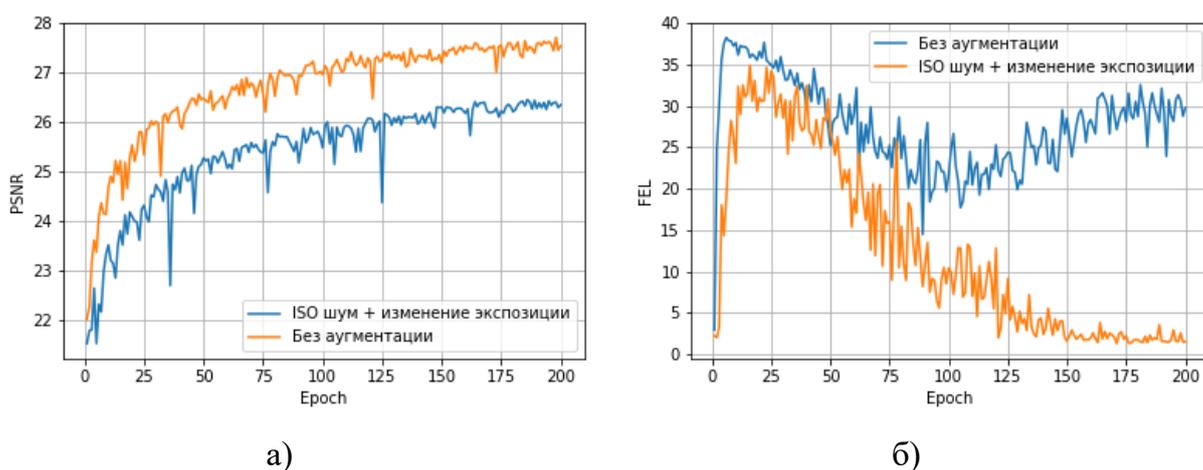
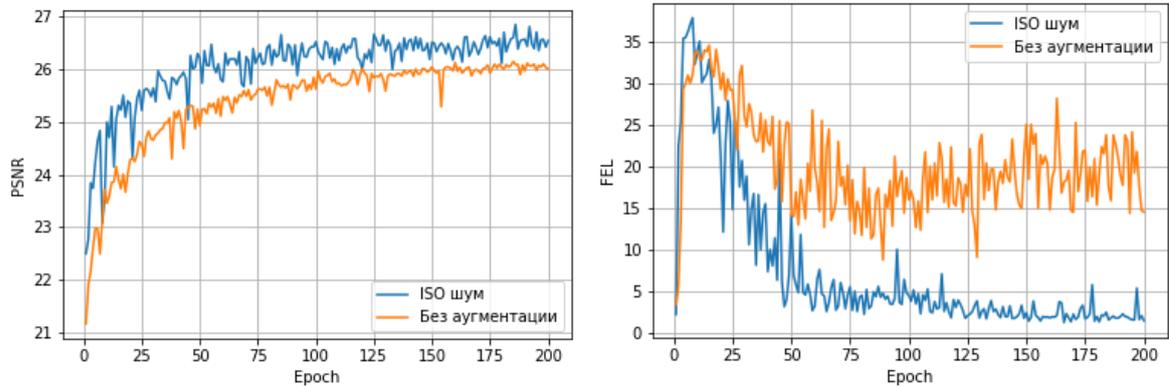


Рисунок 4.17 – Сравнение графиков обучения при отсутствии аугментации обучающей выборки и случая аугментации ISO-шумом для дифракционно-рефракционной ОС: а) PSNR на валидационной выборке, б) FEL, рассчитанный по фрагменту изображения реальной сцены (рисунок 4.10.а)

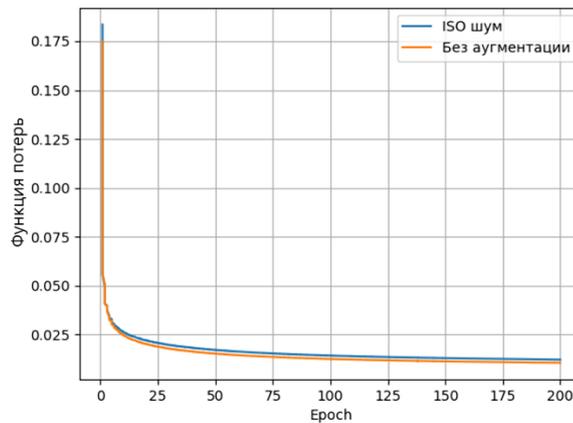
После выбора финальной точки в пространстве параметров полнокадровой сети для ДОС на основе трех линз и гибридной ОС, проводилась локальная реконструкция на основе сети RCAN с механизмом внимания. В случае гибридной ОС прирост среднего значения PSNR составил 0,23 дБ, а при использовании ДОС на основе трех линз – 0,15 дБ. Таким образом, использование более современной архитектуры локальной сети с механизмом внимания обеспечило сопоставимый прирост среднего значения PSNR по сравнению с VDSR и DRRN. В приложении Б

представлено сравнение результатов реконструкции для локальных архитектур RCAN и VDSR в случае использования критерия ранней остановки обучения PSNR.



а)

б)



в)

*Рисунок 4.18 – Сравнение графиков обучения при отсутствии аугментации обучающей выборки и случая аугментации ISO-шумом для ДОС на основе трех линз: а) PSNR на валидационной выборке, б) FEL, рассчитанный по фрагменту изображения реальной сцены (рисунок 4.10.б), в) функция потерь на валидационной выборке*

Подтверждение эффективности критерия FEL проводилось путем расчета средней экспертной оценки MOS по данным 26 участников опроса [71]. Для определения статистической значимости различий средних величин использовался t-критерий Стьюдента. Р-значение составило меньше 0,05. В таблице 4.5 приведены значения FEL и MOS для случаев использования критерия ранней остановки обучения на основе FEL и аугментации данных по сравнению с обучением без

аугментации данных и с использованием критерия ранней остановки обучения на основе PSNR.

Таблица 4.5 – Результаты экспертной оценки MOS для фрагментов восстановленных изображений

Критерий ранней остановки обучения, использование аугментации данных	FEL, %	MOS, балл	Рисунок, используемый при оценке MOS
<b>Дифракционно-рефракционная ОС</b>			
PSNR, без аугментации	30,69	1 ( $p < 0,05$ )	4.11.в
FEL, без аугментации	14,48	1,75 ( $p < 0,05$ )	4.11.г
FEL, аугментация	1,29	3 ( $p < 0,05$ )	4.13.в
<b>ДОС на основе трех линз</b>			
PSNR, без аугментации	17,51	1,5 ( $p < 0,05$ )	4.12.в
FEL, аугментация	0,51	3 ( $p < 0,05$ )	4.15.б

### **4.3 Исследование алгоритма адаптации нейросетевой модели к новым условиям формирования изображений на основе метода крупношагового метаобучения**

В данном разделе представлены экспериментальные исследования алгоритма адаптации полнокадровой сети U-Net (раздел 2.2) под различные условия формирования изображений на основе метода крупношагового метаобучения, описанного в разделе 3.4. Для проведения исследований использовался набор данных, сформированный на основе МДЛ ( $m=7$ ), формирование которого описано в разделе 4.1.3.

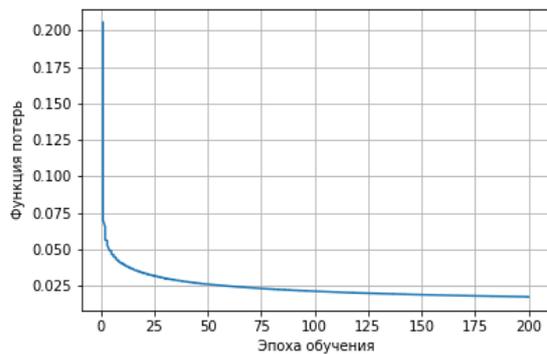
#### **4.3.1 Обучение на исходном наборе полуреальных лабораторных данных**

Используя результаты исследования аугментации, представленные в разделе 4.2.3, на первом этапе алгоритма обучение сети проводилось на аугментированном

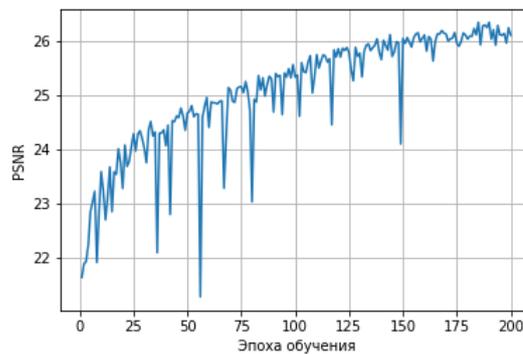
наборе данных. Проведено зашумление обучающих данных с вероятностью 0,5. Параметр интенсивности яркостного шума случайным образом выбирался из множества {0,1; 0,2; 0,3}. Изменение экспозиции (увеличение и уменьшение) применялось с вероятностью 0,1 после аугментации ISO-шумом. Обучение проводилось в рамках 200 эпох, после которых наблюдается стагнация среднего значения PSNR на валидационной выборке. На рисунке 4.19 представлены графики обучения полнокадровой сети. График FEL для фрагмента реального изображения показан, начиная с 30 эпохи обучения, поскольку из-за проблем с масштабированием может быть немного затруднительно увидеть общую тенденцию: на первых эпохах обучения уровень ложных контуров значительно выше, чем на последующих, что не позволяет наглядно представить график на интересующем временном периоде, где FEL достигает минимальных значений. На рисунке 4.20 представлено реальное изображение с выделенным фрагментом, используемым для оценки FEL в процессе обучения.

По критерию FEL значения параметров модели были зафиксированы на 140 эпохе, где FEL достиг значения 0,61%. Среднее значение PSNR на тестовой выборке составило 25,14 дБ. На рисунке 4.20 представлен результат обработки реального изображения, представленного на рисунке 4.19, моделью, обученной в рамках первого этапа алгоритма, а также указан фрагмент изображения, по которому рассчитывалось значение FEL.

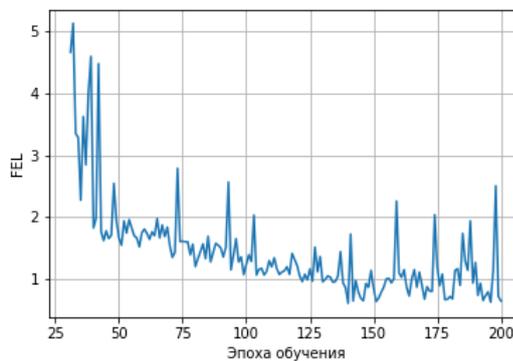
На рисунках 4.22-4.23 представлены примеры обработки реальных изображений, не участвовавших в процессе обучения сети. На восстановленных изображениях (рисунок 4.22.б и 4.22.г) наблюдаются артефакты реконструкции, которые особенно заметны на однотонных фрагментах (стена, фон текста).



а)



б)

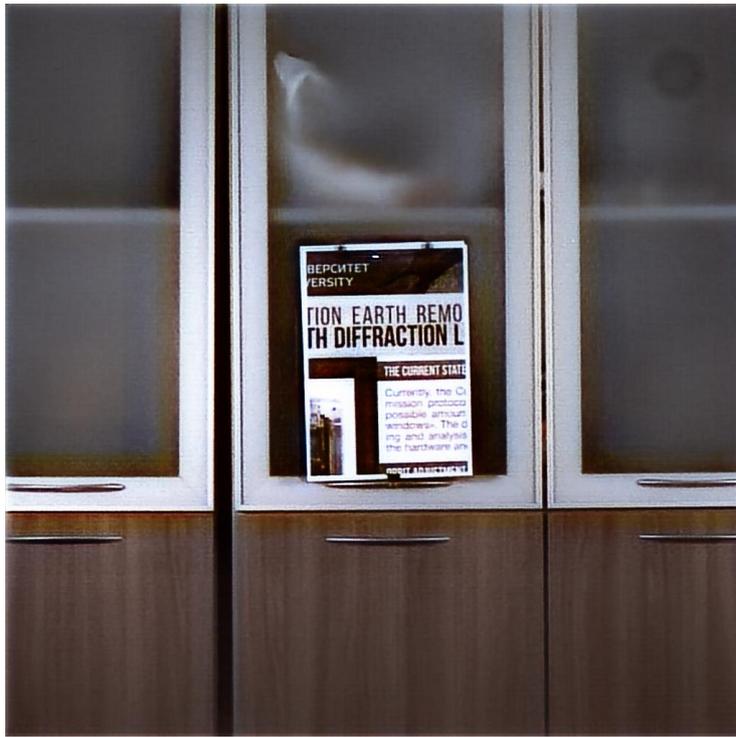


в)

*Рисунок 4.19 – Графики функции потерь (а), среднего значения PSNR на валидационной выборке (б) и FEL на фрагменте реального изображения в процессе I этапа обучения*



*Рисунок 4.20 – Изображение реальной сцены, используемое в процессе обучения сети, где фрагмент размера 200×240 пикселей, выделенный черной ограничивающей рамкой, используется при оценке FEL*



а)



б)

*Рисунок 4.21 – Результат обработки изображения реальной сцены, представленного на рисунке 4.20 на первом этапе алгоритма (а) и фрагмент изображения, используемый для расчета FEL (б).*

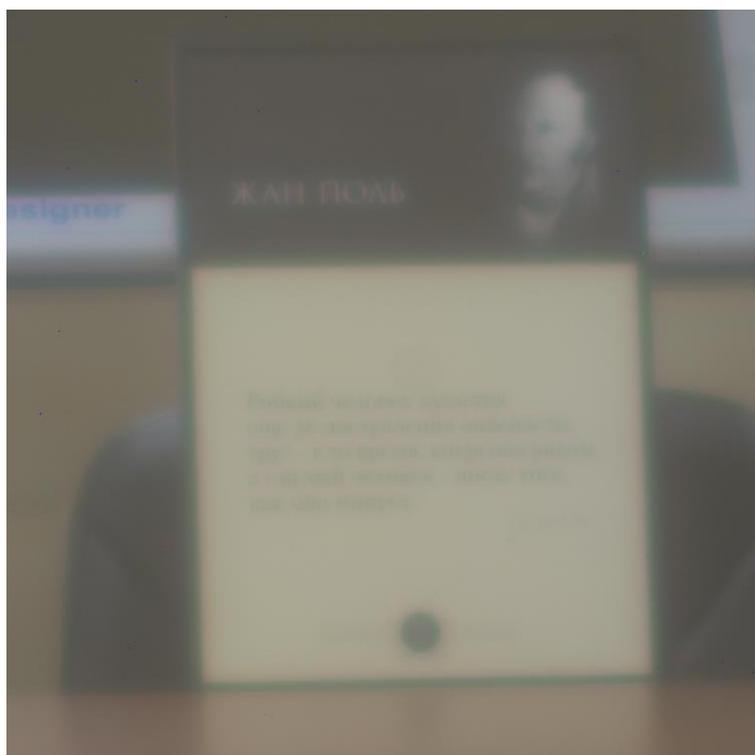


а)

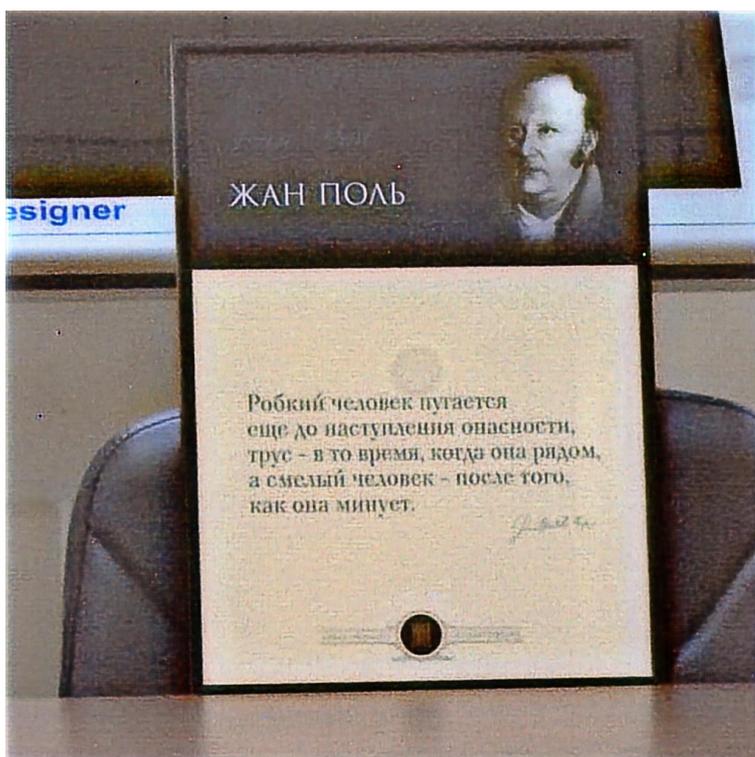


б)

*Рисунок 4.22 – Пример обработки изображения реальной сцены, не используемого в процессе обучения сети на 1 этапе алгоритма: а) изображение, сформированное МДЛ, б) результат обработки моделью на 1 этапе алгоритма*



а)



б)

*Рисунок 4.23 – Пример обработки изображения реальной сцены, не используемого в процессе обучения сети на 1 этапе алгоритма: а) изображение, сформированное МДЛ, б) результат обработки моделью на 1 этапе алгоритма*

### 4.3.2 Обучение на N доменах с разными моделями искажений

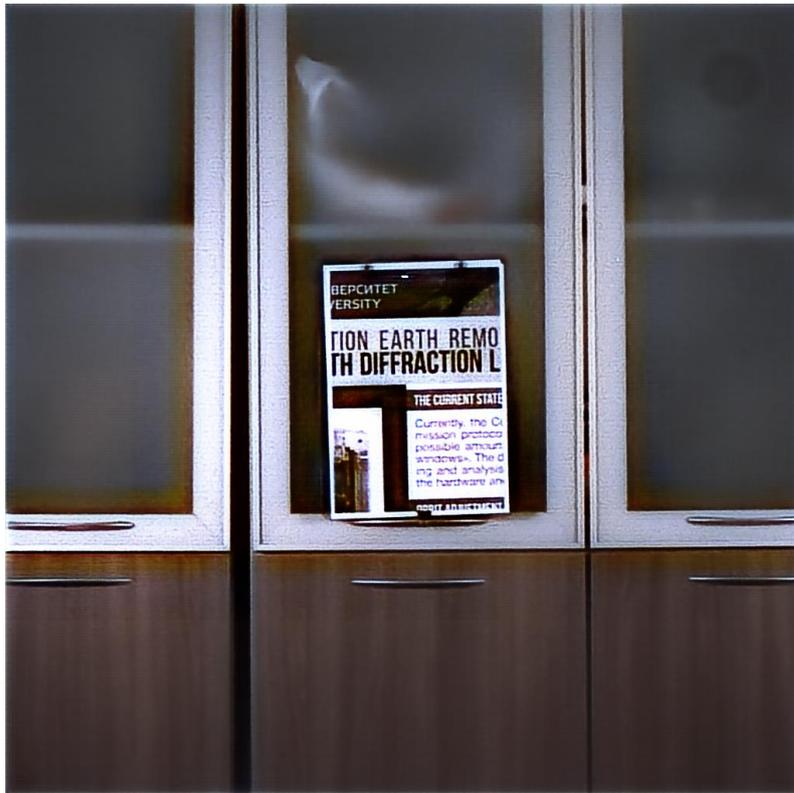
На втором шаге алгоритма проводилось дообучение модели  $\theta_T$ , выбранной в разделе 4.3.1 на пяти наборах данных ( $N=5$ ), сформированных применением следующих типов искажений ко всем изображениям исходного обучающего набора данных:

1. Увеличение экспозиции.
2. Уменьшение экспозиции.
3. Моделирование ISO-шума. Параметр интенсивности случайным образом выбран из множества  $\{0,1; 0,2; 0,3\}$ .
4. Моделирование ISO-шума. Параметр интенсивности случайным образом выбран из множества  $\{0,4; 0,5; 0,6\}$ .
5. Уменьшение экспозиции и моделирование ISO-шума. Параметр интенсивности случайным образом выбран из множества  $\{0,1; 0,2; 0,3\}$ .

Использовались те же гиперпараметры обучения, что и на первом шаге. Средняя модель  $\theta_M$  для пяти точек в пространстве параметров сети рассчитана по формуле 3.2. Результат обработки реального изображения, используемого в процессе обучения (рисунок 4.20), усредненной моделью  $\theta_M$  представлено на рисунке 4.23.

На рисунке 4.24 представлены графики обучения сети на наборах данных с пятью описанными выше типами искажений. Поскольку критерий ранней остановки обучения FEL не информативен при обучении на наборе с одним типом искажения, то на данном этапе используется критерий PSNR и на рисунке 4.24.б представлены графики среднего значения PSNR на валидационной выборке. Нумерация эпохи обучения начинается с номера 140, поскольку на 1 этапе алгоритма обучение было остановлено на данной эпохе.

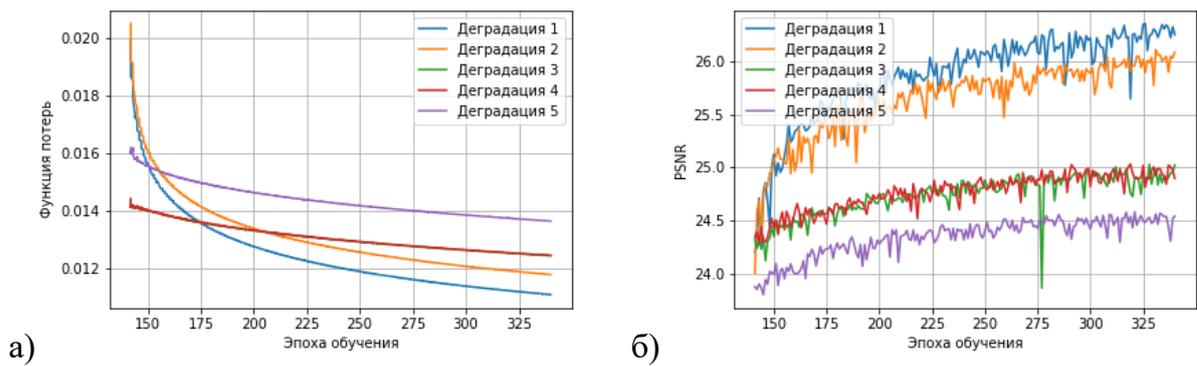
На рисунке 4.25 представлен пример обработки реального изображения, не используемого в процессе обучения. При обработке средней моделью  $\theta_M$  артефакты реконструкции сохраняются.



а)

б)

Рисунок 4.23 – Результат обработки изображения реальной сцены, представленном на рисунке 4.20 усредненной моделью, полученной на втором этапе алгоритма (а) и фрагмент обработанного изображения, используемого для расчета FEL



а)

б)

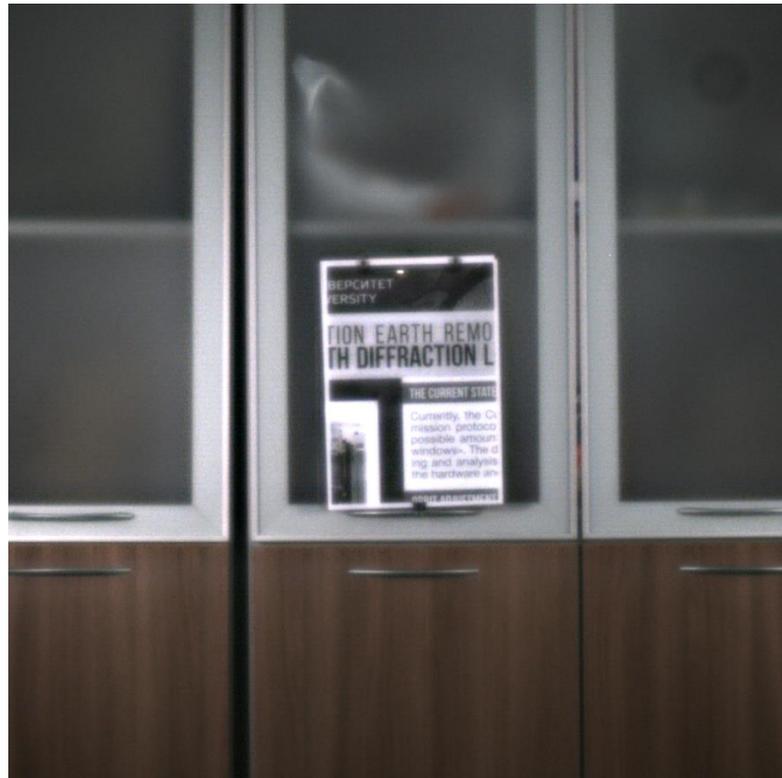
Рисунок 4.24 – Графики обучения сети на наборах данных, соответствующие описанным выше типам искажений: а) функция потерь, б) среднее значение PSNR на валидационной выборке



*Рисунок 4.25 – Пример обработки изображения реальной сцены (рисунок 4.22.а), не участвовавшего в процессе обучения сети на 2 этапе алгоритма*

### **4.3.3 Дообучение на одном изображении реальной сцены**

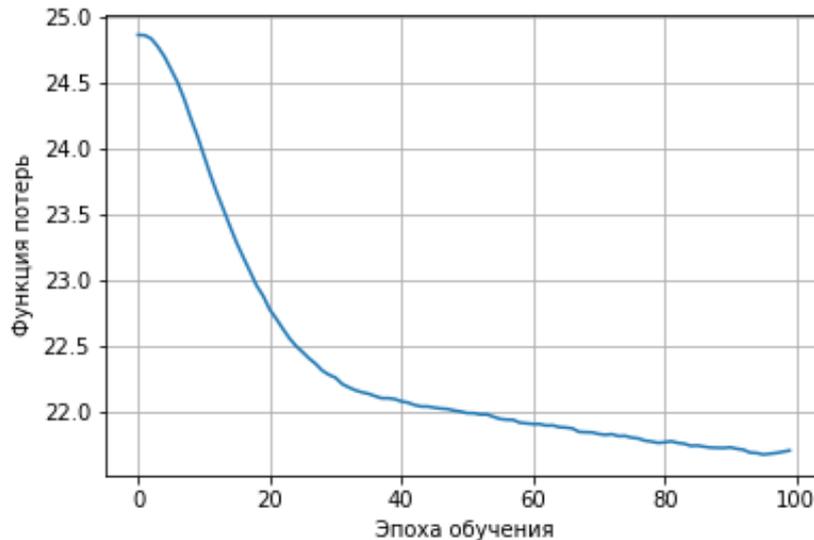
На третьем этапе алгоритма проводится дообучение «средней» модели  $\theta_M$  на одном изображении реальной сцены с неизвестной моделью искажений, представленном на рисунке 4.20. Изображение для расчета функции потерь (рисунок 4.26) получено экспертом средствами Photoshop. Дообучение проводилось в течение 100 эпох. На рисунке 4.27.а представлен результат обработки реального изображения, используемого в обучении, а на рисунке 4.27.б – изображения, не используемого в процессе обучения. Видно, что после обработки финальной моделью артефакты реконструкции исчезли. На рисунке 4.28 представлен график функции потерь в процессе обучения сети.



*Рисунок 4.26 – Результат обработки изображения реальной сцены, представленного на рисунке 4.20, средствами Photoshop*



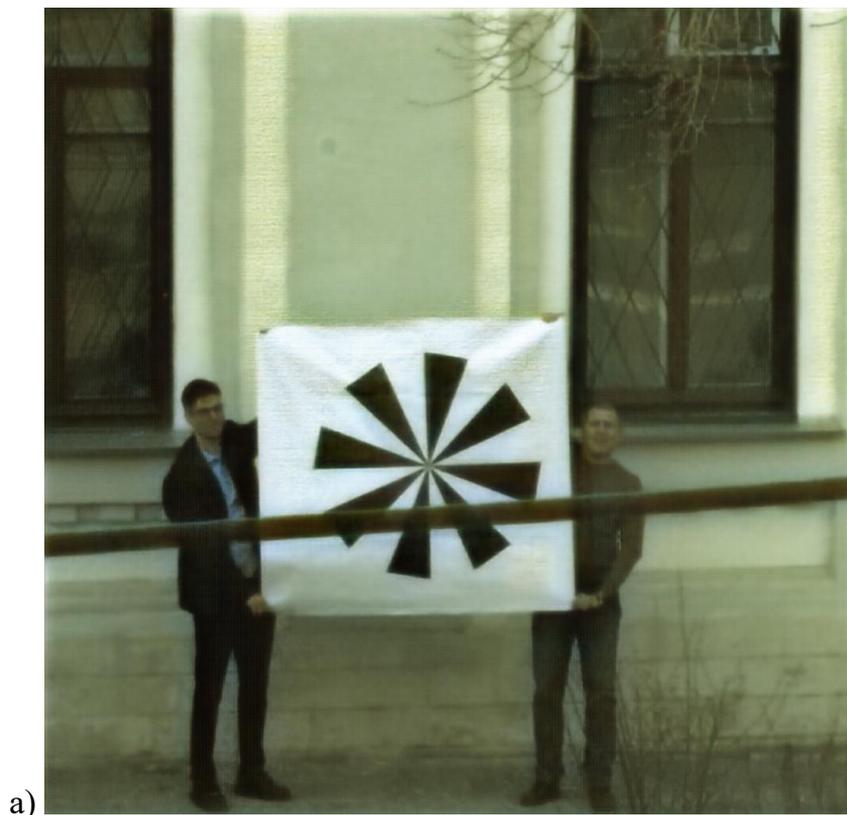
*Рисунок 4.27 – Пример обработки изображения реальной сцены финальной моделью: а) изображение, используемое в процессе обучения (рисунок 4.20), б) изображение, не используемое в процессе обучения (рисунок 4.22.а)*



*Рисунок 4.28 – График функции потерь в процессе обучения сети на 3 этапе*

Также проведено сравнение модели, полученной в результате трех этапов обучения, со случаем пропуска 2 этапа алгоритма (этапа крупношагового метаобучения). В таком случае проводилось дообучение на нулевой выборке (3 этап алгоритма) модели, полученной на первом этапе, то есть использовалась комбинация трансферного обучения и обучения по нулевой выборке.

Обучение проводилось на такой же паре реального (рисунок 4.20) и улучшенного в Photoshop (рисунок 4.26) изображений в течение 100 эпох. На рисунке 4.29 представлен результат обработки реального изображения, не используемого в процессе обучения. Видно, что артефакты реконструкции сохраняются по сравнению с обработкой финальной моделью, полученной на основе крупношагового метаобучения (рисунок 4.27.б). Но по сравнению с обработкой моделями  $\theta_M$  и  $\theta_T$  количество артефактов значительно сокращается за счет дообучения на реальном изображении.



а)



б)

*Рисунок 4.29 – Пример обработки изображения реальной сцены, не используемого в процессе обучения сети, моделью, полученной при пропуске 2 этапа алгоритма (этапа крупношагового метаобучения)*

На рисунках 4.30-4.31 приведено сравнение результатов реконструкции реального изображения, не используемого в процессе обучения для трех случаев:

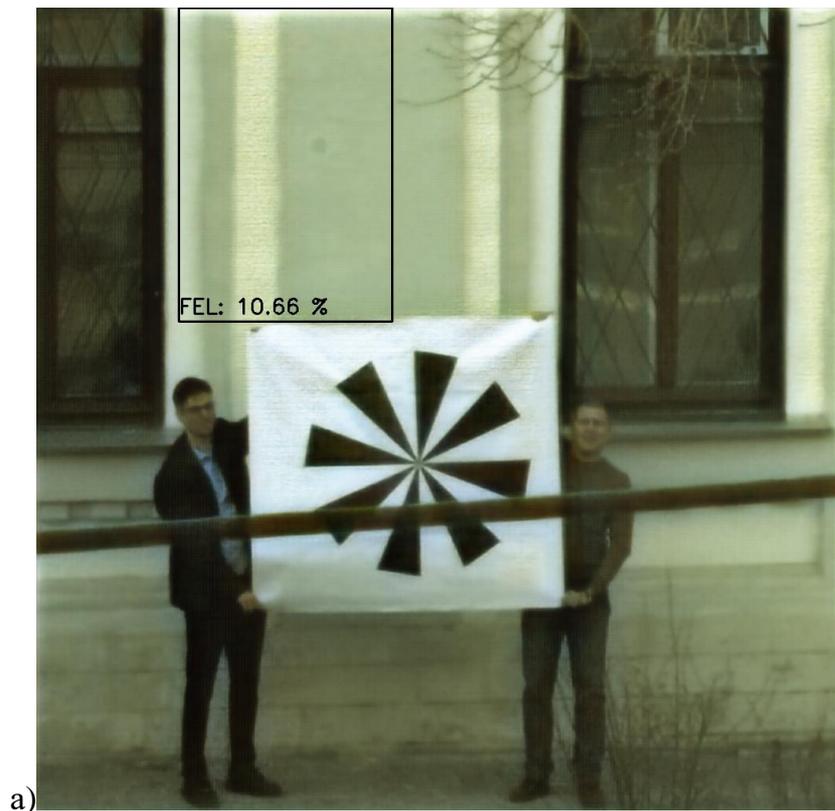
1. Обучение на исходном наборе полуреальных лабораторных данных (этап 1, рисунок 4.30).

2. Дообучение модели, полученной на первом этапе, на одном реальном изображении (этапы 1 и 3, рисунок 4.31.а) – комбинация трансферного обучения и обучения по нулевой выборке.
3. Трехэтапный подход на основе крупношагового метаобучения, предложенный в данной работе (этапы 1, 2, 3, рисунок 4.31.б).

На изображениях выделен фрагмент, по которому оценивался уровень артефактов реконструкции на основе FEL. В данном случае выделенный фрагмент используется только для сравнения результатов реконструкции, поэтому он может содержать «полезные» контуры изображения. Видно, что предложенный подход на основе метаобучения обеспечивает минимальный уровень артефактов – 1,56%.



*Рисунок 4.30 – Результат реконструкции после обучения в рамках этапа 1 алгоритма*



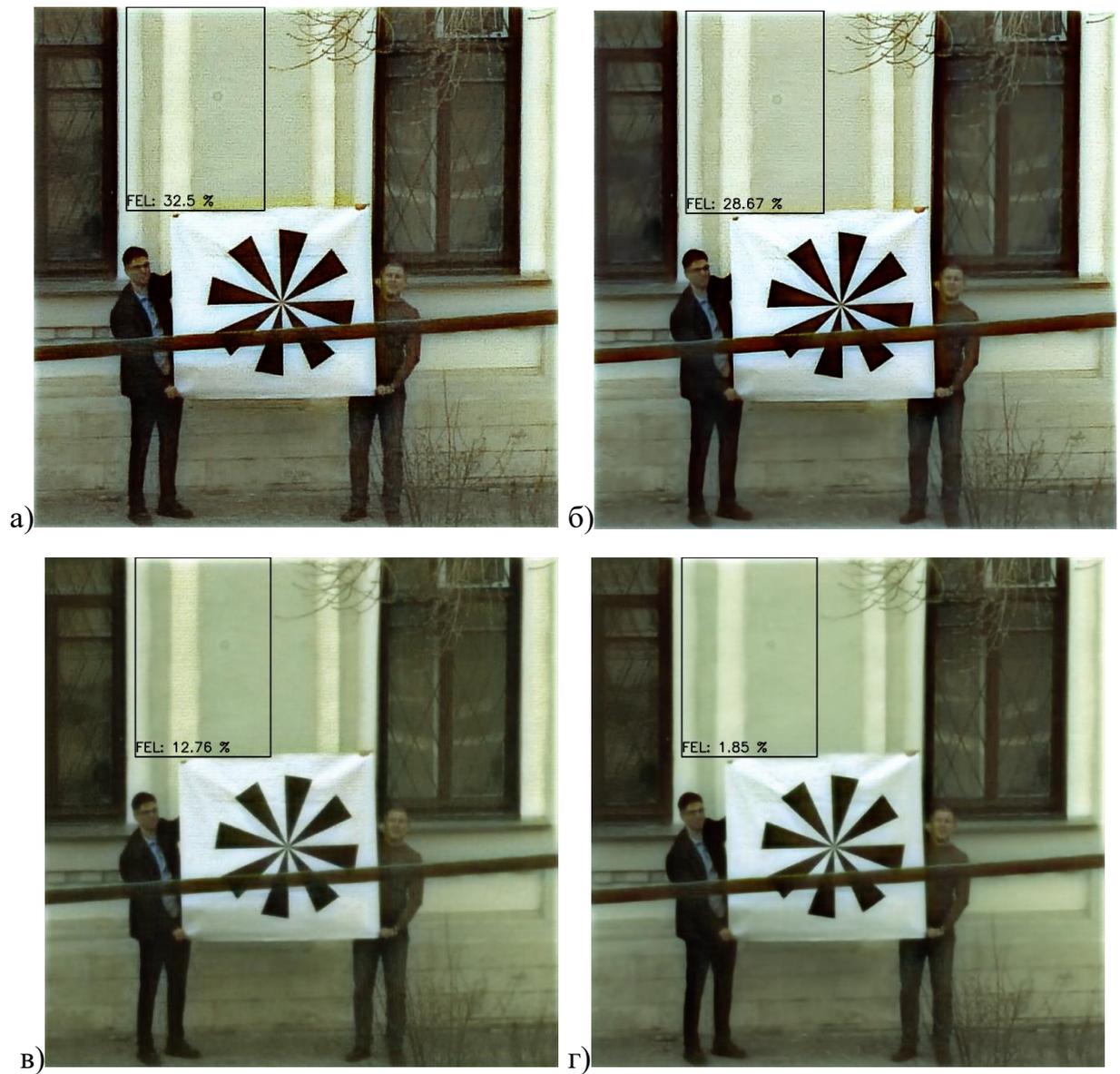
*Рисунок 4.31 – Сравнение результатов реконструкции реального изображения, не используемого в процессе обучения: а) обучение в рамках этапов 1 и 3 без крупношагового метаобучения, б) обучение на основе трех этапов.*

#### 4.3.4 Добавление нового домена

В рамках данного эксперимента на втором этапе алгоритма в стек из 5 моделей добавляются две новых модели, обученные на наборах изображений, сжатых алгоритмом JPEG. Параметр качества в первом случае был выставлен от 20 до 50%, а во втором от 20 до 90%. К реальным изображениям добавляется искажение сжатия с параметром качества 90%. На рисунке 4.32 представлен пример сжатого изображения реальной сцены, не используемого в процессе обучения. На рисунках 4.33-4.34 представлены результаты реконструкции для модели, обученной в рамках первого этапа, усредненной модели, модели, полученной с пропуском этапа метаобучения (1 и 3 этапы) и финальной модели (1, 2 и 3 этапы). Использование крупношагового метаобучения (рисунок 4.31.г и 4.32.в) обеспечивает минимальный уровень артефактов реконструкции как визуально, так и с точки зрения FEL.



*Рисунок 4.32 – Результат JPEG сжатия реального изображения (рисунок 4.22.а) с параметром качества 90%.*



*Рисунок 4.33 – Сравнение результатов реконструкции реального изображения, сжатого алгоритмом JPEG (рисунок 4.31): а) обучение в рамках этапа 1, б) реконструкция усредненной моделью, в) обучение в рамках этапов 1 и 3, г) обучение на основе трех этапов алгоритма*

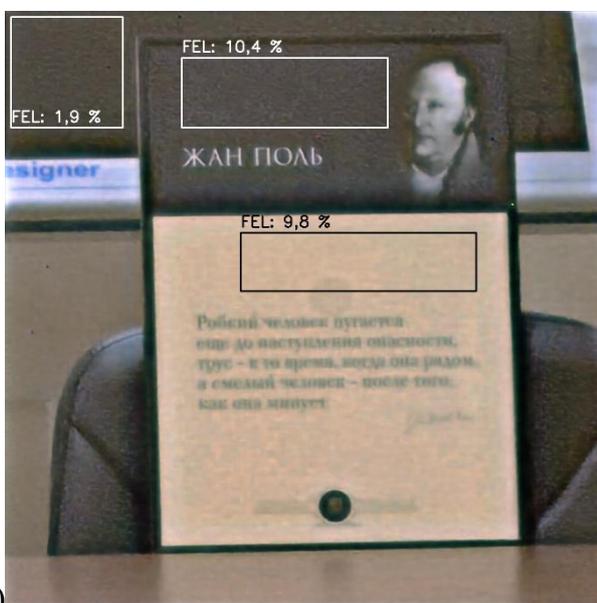
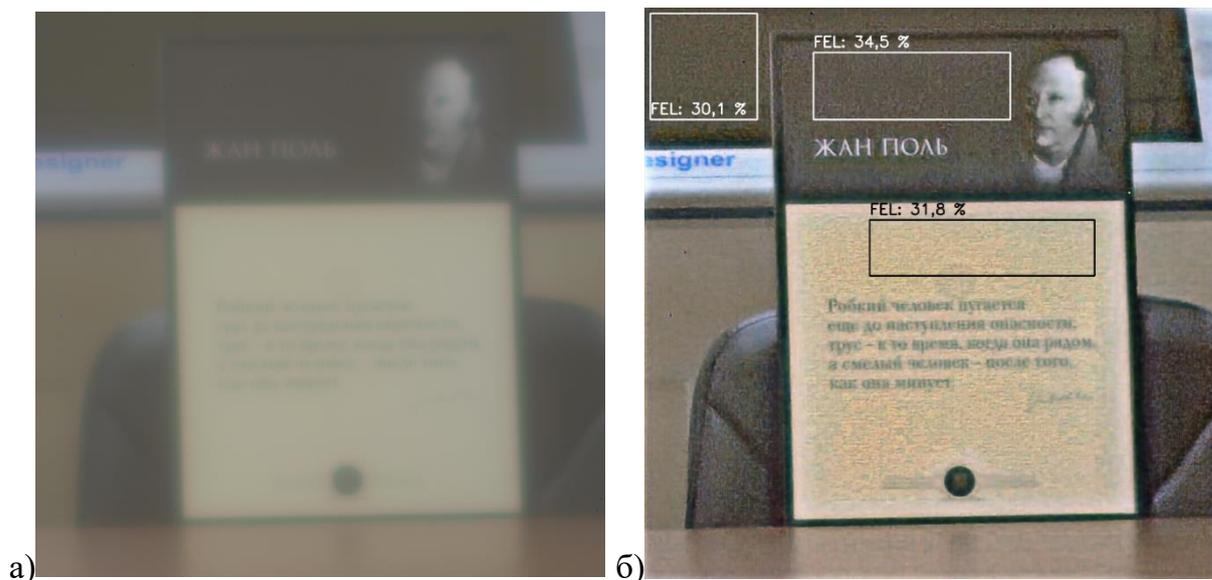


Рисунок 4.34 – Сравнение результатов реконструкции реального изображения, не используемого в процессе обучения: а) реальное изображение, дополнительно сжатое алгоритмом JPEG с параметром качества 90%, б) обучение в рамках этапов 1 и 3 без крупношагового метаобучения, в) обучение на основе трех этапов с крупношаговым метаобучением

В приложении Г приведены другие примеры реконструкции реальных изображений без сжатия, не участвовавших в процессе обучения.

#### 4.4 Сравнение качества реконструкции изображений для разных ИДОС

В главе 1 (таблица 1.1) представлена таблица, объединяющая различные ИДОС с указанием среднего значения PSNR формируемых изображений. В данном разделе приведено расширение таблицы 1.1 с указанием среднего значения PSNR после полнокадровой и локальной реконструкции, а также общий прирост относительно входного значения. Для полнокадровой реконструкции дополнительно указаны архитектура полнокадровой сети и критерий ранней остановки обучения. Для локальной реконструкции дополнительно указаны архитектура локальной сети и прирост среднего значения PSNR относительно полнокадровой реконструкции. Таблица 4.6. обобщает результаты исследований, приведенные в главе 4.

Как правило, нейросетевая реконструкция обеспечивает хорошее качество реконструкции как с точки зрения PSNR, так и при визуальной оценке изображений для систем с входным PSNR больше 17 дБ. Но как видно из таблицы 1.1, предлагаемая в данной работе двухэтапная схема сквозной нейросетевой реконструкции обеспечила наибольший прирост (13,03 дБ) среднего значения PSNR для трехапертурной системы с входным PSNR, равным 12,77 дБ. И как показано в разделе 4.2.3, текст на изображении реальной сцены после реконструкции реконструкции становится читаемым. Однако для повышения естественности цветов остаются полезными методы процедурной коррекции.

При сравнении результатов реконструкции предлагаемой двухэтапной схемы нейросетевой реконструкции (таблица 4.6) с ранними работами (таблица 4.7) видно, что предлагаемая двухэтапная схема нейросетевой реконструкции обеспечивает лучшее качество реконструкции с точки зрения среднего значения PSNR: при использовании металинз и нейросетевой реконструкции [73] прирост среднего значения PSNR составил 6,879 дБ (21 дБ после реконструкции); при использовании МДЛ ( $m=13$ ) и реконструкции на основе обратной свертки [2, 7] прирост среднего значения PSNR составил 1,2 дБ (17,21 дБ после реконструкции); при использовании МДЛ и нейросетевой реконструкции командой Стэнфордского университета [31] прирост среднего значения PSNR составил 4,61 дБ (25,89 дБ после реконструкции).

Таблица 4.6 – Оценка качества реконструкции для различных типов ИДОС

№	Тип ИДОС	Объем выборки (обучение/ тест)	Входной PSNR, дБ	Полнокадровая рек-я (PSNR, дБ, архитектура, критерий)	Локальная рек-я (архитектура, прирост PSNR)	Общий прирост PSNR, дБ
1	Одно-апертурная система на основе МДЛ (m=7) [14]	1226/605	<b>18,13</b>	<b>26,65</b> (U-Net, PSNR)	<b>26,82</b> (VDSR; 0,17дБ)	<b>8,69</b>
2	Одно-апертурная система на основе МДЛ (m=13) [2, 7]	321/29	<b>17,21</b>	<b>25,51</b> (Pix2Pix, PSNR)	<b>25,96</b> (DRRN; 0,45 дБ)	<b>8,75</b>
		1226/605	<b>19,68</b>	<b>27,02</b> (U-Net, PSNR)	<b>27,26</b> (VDSR, 0,24 дБ)	<b>7,58</b>
3	Двух-апертурная система на основе МДЛ (m=13) [57]	119/29	<b>17,21</b>	<b>25,41</b> (Pix2Pix, PSNR)	<b>26,07</b> (VDSR; 0,66 дБ)	<b>8,86</b>
4	Трех-апертурная система на основе МДЛ (m=1) [8]	1244/605	<b>12,77</b>	<b>25,67</b> (U-Net, PSNR)	<b>25,8</b> (RCAN, 0,13 дБ)	<b>13,03</b>
				25,32 (U-Net, FEL)	25,47 (RCAN, 0,15 дБ)	
5	Гибридная система (m=1) [33]	1244/605	<b>17,85</b>	<b>28,09</b> (U-Net, PSNR)	<b>28,32</b> (RCAN, 0,23 дБ)	<b>10,47</b>
				26,91 (U-Net, FEL)	27,11 (RCAN; 0,2 дБ)	

Таблица 4.7 – Качество реконструкции для различных типов ИДОС в работах [2, 7, 31, 73]

Тип ИДОС	Входной PSNR, дБ	PSNR после реконструкции, дБ (метод реконструкции)	Общий прирост PSNR, дБ
Одно-апертурная система на основе МДЛ (m=13) [2, 7]	<b>17,21</b>	<b>18,41</b> (алгоритм на основе обратной свертки)	<b>1,2</b>
Металинзы [73]	14,568	21,447	<b>6,879</b>
МДЛ (m неизвестно) [31]	21,28	25,89	<b>4,61</b>

## Выводы к главе 4

1. Двухэтапная схема сквозной нейросетевой реконструкции обеспечивает больший прирост среднего PSNR на тестовой выборке по сравнению с классическим подходом на основе обратной свертки, а также показывает визуально сопоставимое с эталоном качество восстановленных изображений.
2. Значимость небольшого прироста PSNR (0,45 дБ и 0,66 дБ) при локальной коррекции по сравнению с полнокадровой коррекцией подтверждается исследованием публикаций по задаче сверхразрешения. В данной работе на этапе локальной коррекции значимым считается прирост PSNR, превышающий 0,1 дБ.
3. Двухэтапная схема сквозной нейросетевой реконструкции обеспечивает высокое качество реконструкции независимо от технологии изготовления ИДОС.
4. Использование критерия FEL для ранней остановки обучения обеспечивает меньший уровень артефактов реконструкции по сравнению с результатами для критерия ранней остановки обучения PSNR.
5. Аугментация обучающей выборки позволяет сократить уровень артефактов реконструкции по сравнению с обучением сети без аугментации данных. Аугментация обучающих данных в комбинации с критерием FEL для ранней остановки обучения обеспечивают наименьший уровень артефактов на изображениях реальных сцен.
6. Использование алгоритма адаптации полнокадровой сети U-Net на основе крупношагового метаобучения позволяет адаптировать уже существующую модель к новому домену при наличии одного реального изображения из нового домена, а также обеспечивает сокращение уровня артефактов реконструкции на изображениях сцен реального мира. По сравнению с комбинацией трансферного обучения и обучения по нулевой выборке (без крупношагового метаобучения) предложенный подход позволяет уменьшить уровень артефактов с точки зрения FEL от 3 до 7 раз в зависимости от выбранного фрагмента реального изображения.

## 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.

## СПИСОК СОКРАЩЕНИЙ И УСЛОВНЫХ ОБОЗНАЧЕНИЙ

ДОС – дифракционно-оптическая система

ИДОС – изображающая дифракционно-оптическая система

ДОЭ – дифракционно-оптический элемент

МДЛ – многоуровневая дифракционная линза

PSNR – peak signal-to-noise ratio, пиковое отношение сигнала к шуму

SSIM – индекс структурного сходства

FEL – false edge level, уровень ложных контуров

ФРТ – функция рассеяния точки

СКО – среднеквадратичное отклонение, root-mean-square error

СНС – сверточная нейронная сеть

ГСС – генеративно-состязательная нейронная сеть

VDSR – название архитектуры сети, предложенной в работе [34]

U-Net – название архитектуры сети, предложенной в работе

## СПИСОК ЛИТЕРАТУРЫ

1. Banerji, S. Imaging with flat optics: metalenses or diffractive lenses? / S. Banerji, M. Meem, A. Majumder, F.G. Vasquez, B. Sensale-Rodriguez, R. Menon // *Optica*. – 2019. – Vol. 6, no 6. – P. 805-810.
2. 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.
3. Peng, Y. Computational imaging using lightweight diffractive-refractive optics / Y. Peng, Q. Fu, H. Amata, Sh. Su, F. Heide, W. Heidrich // *Optics Express*. – 2015. – Vol. 23, no. 24. – P. 31393-31407.
4. Sahin, F.E. Light L16 Computational Camera / F.E. Sahin, R. Laroia // *Imaging and Applied Optics*. – 2017.
5. Genevet, P. Recent advances in planar optics: from plasmonic to dielectric metasurfaces / P. Genevet, F. Capasso, F. Aieta, M. Khorasaninejad, R. Devlin // *Optica*. – 2017. – V. 4, no. 1. – P. 139-152/
6. Rostami, S.R.M. Extended DoF and achromatic inverse imaging for lens and lensless MPM camera based on Wiener filtering of defocused OTFs / S.R.M. Rostami, V.Y. Katkovnik, K.O. Eguiazarian // *Optical Engineering*. – 2021. – Vol. 60, no. 5. – p. 051204.
7. 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.
8. 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.
9. 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.
10. 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.
11. 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.
12. Nikonorov, A. Fresnel Lens Imaging with Post-Capture Image Processing / A. Nikonorov, R. Skidanov, V. Fursov, M. Petrov, S. Bibikov, Y. Yuzifovich // Proceedings of IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW 2015). – 2015. – P. 33-41.
13. Никоноров, А.В. Реконструкция изображений в дифракционно-оптических системах на основе сверточных нейронных сетей и обратной свертки / А.В. Никоноров, М.В. Петров, С.А. Бибииков, В.В. Кутикова, А.А. Морозов, Н.Л. Казанский // Компьютерная оптика. – 2017. – Т. 41, № 6. – С. 875-887.
14. 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.
15. 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.
16. 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.

17. 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.
18. Nikonorov, A.V. Deconvolution and Correction Based Approach to Restore Images Captured Using Simple Diffractive Lenses/ A.V. Nikonorov, S.A. Bibikov, M.V. Petrov // *9-th Open German-Russian Workshop on Pattern Recognition and Image Understanding*. – 2014. – P. 1-4.
19. Nikonorov, A. Computational correction for imaging through single fresnel lenses / A.V. Nikonorov, S.A. Bibikov, M.V. Petrov, Y. Yuzifovich, V. Fursov // *12th International Joint Conference on e-Business and Telecommunications (ICETE 2015)*. – 2015. – P. 68-75.
20. Heide, M. High-quality computational imaging through simple lenses / M. Heide, M. Rouf, B. Hullin, B. Labitzke, W. Heidrich, A. Kolb // *ACM Trans. Graph.* – 2013. – vol. 32, no. 5.
21. Pan, J. Blind image deblurring using dark channel prior / J. Pan, D. Sun, H. Pfister, and M.-H. Yang // *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR 2016)*. – 2016. – p. 1628–1636.
22. 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.
23. Фирсов, Н. Разработка и исследование алгоритмов нейросетевой реконструкции изображений в дифракционных оптических системах / Ню Фирсов, В. Евдокимова, А. Никонов // *XIX Международная научно-техническая конференция «Оптические технологии в телекоммуникациях ОТТ-2021»*. – 2021. – С. 101-102.
24. Chen, C. Learning to See in the Dark / C. Chen, Q. Chen, J. Xu, V. Koltun // *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. – 2018.

25. Zhang, X. Zoom to Learn, Learn to Zoom / X. Zhang, Q. Chen, R. Ng, V. Koltun // IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). – 2019.
26. Yan, Q. Attention-guided Network for Ghost-free High Dynamic Range Imaging / Q. Yan, D. Gong, Q. Shi, A. Hengel, C. Shen, I. Reid, Y. Zhang // IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). – 2019.
27. Chen, S. Extreme-Quality Computational Imaging via Degradation Framework / S. Chen, H. Feng, K. Gao, Z. Xu, Y. Chen // ICCV. – 2021.
28. Nayar, S.K. Towards Self-Powered Cameras / S.K. Nayar, D.C. Sims, M. Fridberg // Proceedings of IEEE International Conference on Computational Photography (ICCP). – 2015. – P. 1-10.
29. Sun, T. Revisiting cross-channel information transfer for chromatic aberration correction / T. Sun, Y. Peng, and W. Heidrich // CVPR. – 2017. – P. 3248-3256.
30. Харитонов, С.И., Компьютерное моделирование дифракционных изображающих линз с использованием гиперспектральных изображений / С.И. Харитонов, В.А. Фурсов // Компьютерная оптика. – 2023. – Т. 47, №5. – С. 725-733.
31. Peng, Y. Learned large field-of-view imaging with thin-plate optics / Y. Peng, Q. Sun, X. Dun, G. Wetzstein, W. Heidrich, F. Heide // ACM Transactions on Graphics. – 2019. – Vol. 38. – p. 1-14.
32. Dun, X. Supplementary document for Learned Rotationally Symmetric Diffractive Achromat for Full-Spectrum Computational Imaging / X. Dun, H. Ikoma, G. Wetzstein, Z. Wang, X. Cheng, Y. Peng // Optica. – 2020.
33. 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.
34. 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.
35. Quinonero-Candela, J. Dataset shift in machine learning / J. Quinonero-Candela, M. Sugiyama, N. D. Lawrence, A. Schwaighofer // Mit Press, 2009. – 246 с.

36. Wang, Z. Deep Learning for Image Super-resolution: A Survey / Z. Wang, J. Chen, S. C. H. Hoi // IEEE Transactions on Pattern Analysis and Machine Intelligence. – 2021. – Vol. 43, no. 10. – p. 3365-3387.
37. Wang, J. Generalizing to Unseen Domains: A Survey on Domain Generalization / J. Wang, C. Lan, C. Liu, Y. Ouyang, T. Qin, W. Lu, Y. Chen, W. Zeng, P. S. Yu // IEEE Transactions on Knowledge and Data Engineering. – 2023. – Vol. 35, p. 8052-8072.
38. Wang, L. Unsupervised degradation representation learning for blind superresolution / L. Wang, Y. Wang; X. Dong; Q. Xu; J. Yang; W. An, Y. Guo // CVPR. – 2021.
39. Soh, J. W. Meta-transfer learning for zero-shot super-resolution / J. W. Soh, S. Cho, N. I. Cho // Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. –2020. – p. 3516-3525.
40. Wang, X. Real-ESRGAN: Training Real-World Blind Super-Resolution with Pure Synthetic Data / X. Wang; L. Xie; C. Dong, Y. Shan. // IEEE/CVF International Conference on Computer Vision Workshops (ICCVW). – 2021. – p. 1904-1914.
41. Wang, M. Deep visual domain adaptation: A survey / M. Wang, W. Deng // Neurocomputing. – 2018. – Vol. 312. – p. 135–153.
42. Visual domain adaptation: A survey of recent advances / V. M. Patel, R. Gopalan, R. Li, and R. Chellappa // IEEE signal processing magazine. – 2015. – Vol. 32, no. 3. – p. 53–69.
43. Finn, C. Model-agnostic meta-learning for fast adaptation of deep networks / C. Finn, P. Abbeel, S. Levine // ICML. – 2017.
44. Antoniou, A. How to train your MAML / A. Antoniou H. Edwards, A.J. Storkey // ICLR. – 2018.
45. Pan, S. A survey on transfer learning / S. Pan, Q. Yang, // IEEE TKDE. – 2010. – Vol. 22. – p. 1345–1359.
46. Weiss, K. A survey of transfer learning / K. Weiss, T. M. Khoshgoftaar, D. Wang, // Journal of Big data. – 2016. – Vol. 3, no. 1. – p. 1–40.
47. Zhuang, F. A comprehensive survey on transfer learning / F. Zhuang, Z. Qi, K. Duan, D. Xi, Y. Zhu, H. Zhu, H. Xiong, Q. He // Proceedings of the IEEE. – 2020. – Vol. 109, no. 1. – p. 43–76.

48. Wang, W. A survey of zeroshot learning: Settings, methods, and applications / W. Wang, V. W. Zheng, H. Yu, C. Miao // ACM Trans.Intelligent Systems and Technology. – 2019. – Vol. 10, no. 2. – p. 1–37.
49. Wang, Z. A decadal survey of zero-shot image classification / Z. Wang, Y. Yu, Y. Pang, // Scientia Sinica Informationis. – 2019. – Vol. 49, no. 10. – p. 1299–1320.
50. Wang, P. Chromatic-aberration-corrected diffractive lenses for ultra-broadband focusing / P. Wang, N. Mohammad, R. Menon // Scientific Reports. – 2016. – Vol.6.
51. Lenkova, G.A. Diffractive-Refractive Intraocular Lenses with Binary Structures // Optoelectronics, Instrumentation and Data Processing. – 2018. – Vol. 54, no.5. – p. 469-476.
52. Schuler, C. J. Non-stationary correction of optical aberrations / C. J. Schuler, M. Hirsch, S. Harmeling, B. Scholkopf // Proc. Int. Conf. Comput. Vis. – 2011. – p. 659–666.
53. Peng, Y. The diffractive achromat full spectrum computational imaging with diffractive optics / Y. Peng, Q. Fu, F. Heide, W. Heidrich // Proc. SIGGRAPH ASIA. – 2016.– Vol. 4.
54. Poleshchuk, A.G. Diffractive optical elements: fabrication and application / A.G. Poleshchuk, V.P. Korolkov, R.K. Nasyrov // Proceedings of SPIE. – 2014. – Vol. 9283 – p. 928302.
55. Stone, T. Hybrid diffractive-refractive lenses and achromats / T. Stone, N. George // Applied Optics. –1988. – Vol. 27, no. 14. – p. 2960-2971.
56. Mao, S. Design and analysis of a hybrid optical system containing a multilayer diffractive optical element with improved diffraction efficiency / S. Mao, J. Zhao // Applied Optics. – 2020. – Vol. 59, no. 20. – p. 5888-5895.
57. Piao, M. Keyan Design of achromatic annular folded lens with multilayer diffractive optics for the visible and near-IR wavebands / M. Piao, B. Zhang, // Optics Express. – 2020. – Vol. 28, no. 20. – p. 29076-29085.
58. Dubik, B. Hybrid lens with corrected sphero-chromatic aberration / B. Dubik, S. Koth, J. Nowak, M. Zajac // Optics and Laser Technology. – 1995. – Vol. 27, no. 5. – p. 315-319.

59. Choi, H. Design of Hybrid Lens for Compact Camera Module Considering Diffraction Effect / H. Choi, Y.J. Yoon, B. Kim, S.H. Lee, W.C. Kim, N.C.h. Park, Y.P. Park, S. Kang // *Jpn. J. Appl. Phys.* – 2008. – Vol. 47. – p. 6678.
60. Van De Weijer, J. Edge-based color constancy / J. Van De Weijer, T. Gevers, A. Gijsenij // *IEEE Transactions on Image Processing.* – 2007. – V. 16(9). – P. 2207-2214.
61. Nikonorov, A. Comparative Evaluation of Deblurring Techniques for Fresnel Lens Computational Imaging / A. Nikonorov, M. Petrov, S. Bibikov, Y. Yuzifovich, P. Yakimov, N. Kazanskiy, R. Skidanov, V. Fursov // *ICPR.* – 2016. – P. 775-780.
62. Heide, F. FlexISP: A Flexible Camera Image Processing Framework / F. Heide, M. Steinberger, Y-T. Tsai, M. Rouf, D. Pająk, D. Reddy, O. Gallo, J. Liu, W. Heidrich, K. Egiazarian, J. Kautz, K. Pulli // *ACM Trans. Graph.* – 2014. – Vol. 33, no. 6. – P. 231.
63. Chambolle, A. A first-order primal-dual algorithm for convex problems with applications to imaging / A. Chambolle, T. Pock // *J Math Imaging Vision.* – 2011. – Vol. 40. – p. 120-145.
64. Schuler, C.J. Learning to deblur / C.J. Schuler, M. Hirsch, S. Harmeling, B. Schölkopf // *IEEE Trans Pattern Anal Mach Intel.* – 2016. – Vol. 38, no. 7. – p.1439-1451.
65. Dong, C. Image superresolution using deep convolutional networks / C. Dong, C.C. Loy, K. He, X. Tang // *IEEE Trans. Pattern Anal. Mach. Intel.* – 2015. – Vol. 38, no. 2. – p. 295-307.
66. Kim, J. Accurate image super-resolution using very deep convolutional networks / J. Kim, J. Lee, and K. Lee // *IEEE Conference on Computer Vision and Pattern Recognition (CVPR-2016).* – 2016. – P. 1646-1654.
67. Lai, W.-S. Deep laplacian pyramid networks for fast and accurate super-resolution / W.-S. Lai, J.-B. Huang, N. Ahuja, M.-H. Yang // *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR).* – 2017.
68. Deqing, S. Secrets of optical flow estimation and their principles / S. Deqing, S. Roth, M.J. Black // *IEEE Conference on Computer Vision and Pattern Recognition (CVPR).* – 2010. – P. 2432-2439.

69. 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.
70. Chen, S. Computational Optics for Mobile Terminals in Mass Production / S. Chen, T. Lin, H. Feng, Z. Xu, Q. Li, Y. Chen // IEEE Transactions on Pattern Analysis and Machine Intelligence. – 2022. – Vol. 45, № 4. – p. 4245-4259.
71. Ledig, C. Photorealistic single image super-resolution using a generative adversarial network / C. Ledig, L. Theis, F. Huszar, J. Caballero, A. Cunningham, A. Acosta, A. P. Aitken, A. Tejani, J. Totz, Z. Wang // CVPR. – 2017.
72. Mean opinion score interpretation and reporting. ITU-T Recommendation P.800.2 [Электронный ресурс]. – Режим доступа: <https://handle.itu.int/11.1002/1000/12973-en>.
73. Dong, Y., et al. High-fidelity achromatic metalens imaging via deep neural network / Y. Dong, B. Zheng, H. Li, H. Tang, Y. Huang, S. An, H. Zhang // arXiv:2308.00211v1. – 2023.
74. Huang, L. Design and analysis of extended depth of focus metalenses for achromatic computational imaging / L. Huang, J. Whitehead, S. Colburn, A. Majumdar // Photonics Research. – 2020. – Vol. 8. – p. 1613-1623.
75. Zhang, Y. Image super-resolution using very deep residual channel attention networks / Y. Zhang, K. Li, K. Li, L. Wang, B. Zhong, Y. Fu // Proceedings of the European Conference on Computer Vision (ECCV). – 2018. – pp. 286–301.
76. Zhang, Y. Residual dense network for image super-resolution / Y. Zhang, Y. Tian, Y. Kong, B. Zhong, Y. Fu // CVPR. – 2018.
77. Dong, C. Learning a deep convolutional network for image super-resolution / C. Dong, C. C. Loy, K. He, X. Tang // ECCV. – 2014.
78. Lai, W.-S. Deep laplacian pyramid networks for fast and accurate superresolution/ W.-S. Lai, J.-B. Huang, N. Ahuja, M.-H. Yang // CVPR. – 2017.
79. Ahn, N. Fast, accurate, and lightweight super-resolution with cascading residual network / N. Ahn, B. Kang, K.-A. Sohn // ECCV. – 2018.
80. Johnson, J. Perceptual Losses for Real-Time Style Transfer and Super-Resolution / J. Johnson, A. Alahi, L. Fei-Fei // ECCV. – 2016.

81. Sajjadi, M. S. Enhancenet: Single image super-resolution through automated texture synthesis / M. S. Sajjadi, B. Scholkopf, M. Hirsch // ICCV. – 2017.
82. Lim, B. Enhanced deep residual networks for single image super-resolution / B. Lim, S. Son, H. Kim, S. Nah, K. M. Lee // CVPRW. – 2017.
83. Wang, Y. A fully progressive approach to single-image super-resolution / Y. Wang, F. Perazzi, B. McWilliams, A. Sorkine-Hornung, O. Sorkine-Hornung, C. Schroers // CVPRW. – 2018.
84. Tai, Y. Image Super-Resolution via Deep Recursive Residual Network / Y. Tai, J. Yang, X. Liu // CVPR. – 2017.
85. Zhang, Y. Image Super-Resolution Using Very Deep Residual Channel Attention Networks / Y. Zhang, K. Li, K. Li, L. Wang, B. Zhong, Y. Fu // European Conference on Computer Vision (ECCV). – 2018. – Vol. 11211. – p. 294-310.
86. Isola, P. Image-to-image translation with conditional adversarial networks / P. Isola, J.-Y. Zhu, T. Zhou, A. Efros // IEEE Conference on Computer Vision and Pattern Recognition (CVPR). – 2017.
87. Ronneberger, O. U-net: Convolutional networks for biomedical image segmentation / O. Ronneberger, P. Fischer, T. Brox // International Conference on Medical image computing and computer-assisted intervention. – 2015. – P. 234-241.
88. Araujo, A. Computing Receptive Fields of Convolutional Neural Networks / A. Araujo, W. D. Norris, J. Sim // Distill. – 2019.
89. Zhao, H. Loss Functions for Image Restoration with Neural Networks / H. Zhao, O. Gallo, I. Frosio, J. Kautz // IEEE Transactions on Computational Imaging. – 2016. – Vol. 3. – P. 47-57.
90. Huber, P.J. Robust Estimation of a Location Parameter / P.J. Huber // The Annals of Mathematical Statistics. – 1992. –Vol. 35, no. 1. – P. 73-101.
91. Wang, Z. Multiscale structural similarity for image quality assessment. / Z. Wang, E.P. Simoncelli, A.C. Bovik // Conference Record of the Thirty-Seventh Asilomar Conference on Signals, Systems and Computers. – 2003. – Vol. 2. – P. 1398–1402.
92. Ying, Z. A New Low-Light Image Enhancement Algorithm Using Camera Response Model / Z. Ying; G. Li; Y. Ren; R. Wang, W. Wang // IEEE International Conference on Computer Vision Workshops (ICCVW). – 2017. – p. 3015-3022.

93. .Weijer, V.D. Edge-based color constancy / V.D. Weijer, T. Gevers, and A. Gijsenij // IEEE Transactions on Image Processing. – 2007. – Vol. 16, no. 9. – p. 2207–2214.
94. Gijsenij, A. Improving color constancy by photometric edge weighting / A. Gijsenij, T. Gevers, and J. van de Weijer // IEEE Transactions on Pattern Analysis and Machine Intelligence. – 2012. – Vol. 34, no. 5. – p. 918– 929.
95. Afifi, M. Deep White-Balance Editing / M. Afifi, M. S. Brown // IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). – 2020. – p. 1394–1403.
96. Buades, A. A non-local algorithm for image denoising / A. Buades, B. Coll, J.-M. Morel // IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR). – 2005. – Vol. 2. – p. 60–65.
97. Kwok, N. Simultaneous image color correction and enhancement using particle swarm optimization / N. Kwok, H. Shi, Q. Ha, G. Fang, S. Chen, X. Jia // Engineering Applications of Artificial Intelligence. – 2013. – Vol. 26, no. 10. – p. 2356–2371.
98. Foi; A. Practical Poissonian-Gaussian Noise Modeling and Fitting for Single-Image Raw-Data / A. Foi; M. Trimeche; V. Katkovnik, K. Egiazarian // IEEE Transactions on Image Processing. – 2008. – Vol. 17, no.10. – p. 1737-1754.
99. Guo, S. Toward Convolutional Blind Denoising of Real Photographs / S. Guo, Z. Yan, K. Zhang, W. Zuo, L. Zhang // CVPR. – 2019.
100. Albumination library [Электронный ресурс]. – Режим доступа: <https://albuminations.ai/>.
101. Полное кроссплатформенное решение для записи, преобразования и потоковой передачи аудио и видео [Электронный ресурс]. – Режим доступа: <https://ffmpeg.org>.
102. Goodfellow, I. Explaining and harnessing adversarial examples / I. Goodfellow, J. Shlens, C. Szegedy // International Conference on Learning Representations (ICLR). – 2015.
103. Kingma, D.P. Adam: A method for stochastic optimization / D. P. Kingma, J. Ba // International Conference on Learning Representations. – 2014.
104. Carles, G. Compact multi-aperture imaging with high angular resolution / G. Carles, G. Muyo, N. Bustin, A. Wood, A.R. Harvey // Journal of the Optical Society of America A. – 2015. – Vol. 32, no. 3. – P. 411-419.

105. Tanida, J. Multi-aperture optics as a universal platform for computational imaging / J. Tanida // *Optical Review*. – 2016. – Vol. 23, no. 5. – P. 859-864.
106. Sweeney, D.W. Harmonic diffractive lenses / D.W. Sweeney, G.E. Sommargren // *Applied Optics*. – 1995. – V. 34, no. 14. – P. 2469-2475.
107. Soifer, V.A. *Computer Design of Diffractive Optics* / V.A. Soifer // Cambridge International Science Publishing Limited in association with Woodhead Publishing Limited. – 2012. – 896 p.
108. Wang, Z. Image quality assessment: from error visibility to structural similarity / Z. Wang, A.C. Bovik, H.R. Sheikh, E.P. Simoncelli // *IEEE Transactions on Image Processing*. – 2004. – Vol. 13, no. 4. – P. 600–612.
109. Li, C. Precomputed real-time texture synthesis with markovian generative adversarial networks / C. Li, M. Wand // *European Conference on Computer Vision*. – 2016.
110. Yu, X. Underwater-GAN: Underwater Image Restoration via Conditional Generative Adversarial Network / X. Yu, Y. Qu, M. Hong // *ICPR*. – 2019. – Vol. 11188. – P. 66-75.

## ПРИЛОЖЕНИЕ А – АКТ О ВНЕДРЕНИИ



### ООО ЛОКУС

443051, Самарская область, г. Самара, Кировский район,  
ул. Республиканская, дом 65, кв. 38  
ИНН 7731315489 КПП 631201001 ОГРН 1167746437990  
ОКПО 2337363 ОКВЭД 72.19.2, 72.19.3  
Тел./факс: +7 (846) 2702385  
[artniko@gmail.com](mailto:artniko@gmail.com)

### **АКТ о внедрении результатов диссертационной работы по теме «Метод крупношагового метаобучения в сквозной нейросетевой реконструкции одного класса изображений»**

Настоящий Акт свидетельствует, что результаты диссертационной работы В.В. Евдокимовой по теме «Метод крупношагового метаобучения в сквозной нейросетевой реконструкции одного класса изображений» внедрены в ООО «Локус» и используются 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» (договор № №557ГРНТИС5/49477 от 01 октября 2019 г.).

Разработки В.В. Евдокимовой по нейросетевой реконструкции изображений, сформированных в изображающих дифракционно-оптических системах, а также по адаптации нейросетевых моделей реконструкции к новым условиям формирования изображений используются при реконструкции изображений, сформированных на борту 3U CubeSat, который запущен на низкую околоземную орбиту Земли 22 марта 2021 года. Разработки В.В. Евдокимовой позволяют использовать комбинацию изображающих дифракционно-оптических систем и нейросетевой реконструкции изображений в прикладных задачах, где изображения сформированы на больших расстояниях, в небольшом количестве и отсутствует возможность формирования репрезентативной выборки большого объема для обучения нейросетевых моделей.

Директор по науке  
ООО «Локус», д.ф.-м.н.



Р.В. Скиданов

## ПРИЛОЖЕНИЕ Б – РЕЗУЛЬТАТЫ РЕКОНСТРУКЦИИ ИЗОБРАЖЕНИЙ ДЛЯ РАЗНЫХ ИДОС

В таблице А.1 представлено сравнение результатов реконструкции для разных ИДОС и комбинаций нейросетевых архитектур в случае использования критерия ранней остановки обучения PSNR.

Таблица А.1 – Результаты реконструкции для разных комбинаций архитектур

Архитектуры	Входное изображение (PSNR, дБ)	Объем выборки (обучение/тест)	Полнокадровая реконструкция (PSNR, дБ)	Локальная реконструкция (PSNR, дБ, прирост PSNR)	Общий прирост PSNR, дБ
<b>Однопертурная система на основе МДЛ (m=13)</b>					
Pix2pix и трехканальная VDSR	17,21	321/29	25,51	25,59 (0,08)	8,38
Pix2pix и трехканальная DRRN	17,21	321/29	25,51	25,96 (0,45)	8,75
U-Net и трехканальная VDSR	19,68	1226/605	27,02	27,26 (0,24)	7,58
U-Net и трехканальная DRRN	19,68	1226/605	27,02	27,11 (0,09)	7,43
<b>Двухпертурная система на основе МДЛ (m=13)</b>					
Pix2Pix и шестиканальная VDSR	17,21	119/29	25,41	26,07 (0,66)	8,86
<b>Однопертурная система на основе МДЛ (m=7)</b>					
U-Net и трехканальная VDSR	18,13	1226/605	26,65	26,82 (0,17 дБ)	8,69
U-Net и трехканальная DRRN	18,13	1226/605	26,65	26,7 (0,05)	8,57
<b>Трехпертурная система на основе МДЛ (m=1)</b>					
U-Net и трехканальная VDSR	12,77	1244/605	25,67	25,78 (0,11)	13,01
U-Net и RCAN	12,77	1244/605	25,67	25,8 (0,13)	13,03
<b>Гибридная система на основе МДЛ (m=1)</b>					
U-Net и трехканальная VDSR	17,85	1244/605	28,09	28,28 (0,19)	10,43
U-Net и RCAN	17,85	1244/605	28,09	28,32 (0,23)	10,47

## ПРИЛОЖЕНИЕ В – ПРИМЕРЫ РЕКОНСТРУКЦИИ РЕАЛЬНЫХ ИЗОБРАЖЕНИЙ ДЛЯ РЕФРАКЦИОННО-ДИФРАКЦИОННОЙ ОС

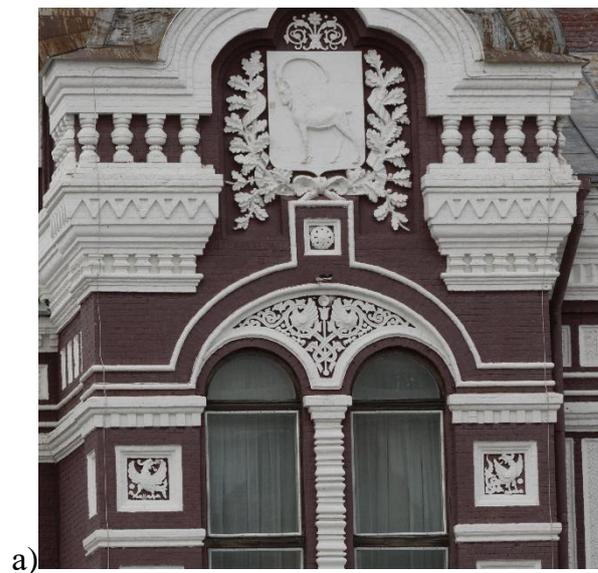
На рисунках Б.1-Б.3 представлены дополнительные примеры реконструкции реальных изображений для рефракционно-дифракционной ОС.



*Рисунок Б.1 – Результат реконструкции реального изображения, не участвовавшего в обучении: а) изображение, сформированное рефракционно-дифракционной ОС, б) случай без аугментации обучающей выборки, в) случай с аугментацией обучающей выборки (таблица 4.4 строка 4), г) фрагмент б), д) фрагмент в)*



*Рисунок Б.2 – Результат реконструкции реального изображения, не участвовавшего в обучении: а) изображение, сформированное рефракционно-дифракционной ОС, б) случай без аугментации обучающей выборки, в) случай с аугментацией обучающей выборки (таблица 4.4 строка 4), г) фрагмент б), д) фрагмент в)*



*Рисунок Б.3 – Изображения, сформированные классическим рефракционным объективом*

## ПРИЛОЖЕНИЕ Г – ДОПОЛНИТЕЛЬНЫЕ ПРИМЕРЫ РЕКОНСТРУКЦИИ РЕАЛЬНЫХ ИЗОБРАЖЕНИЙ ДЛЯ МЕТОДА КРУПНОШАГОВОГО МЕТАОБУЧЕНИЯ

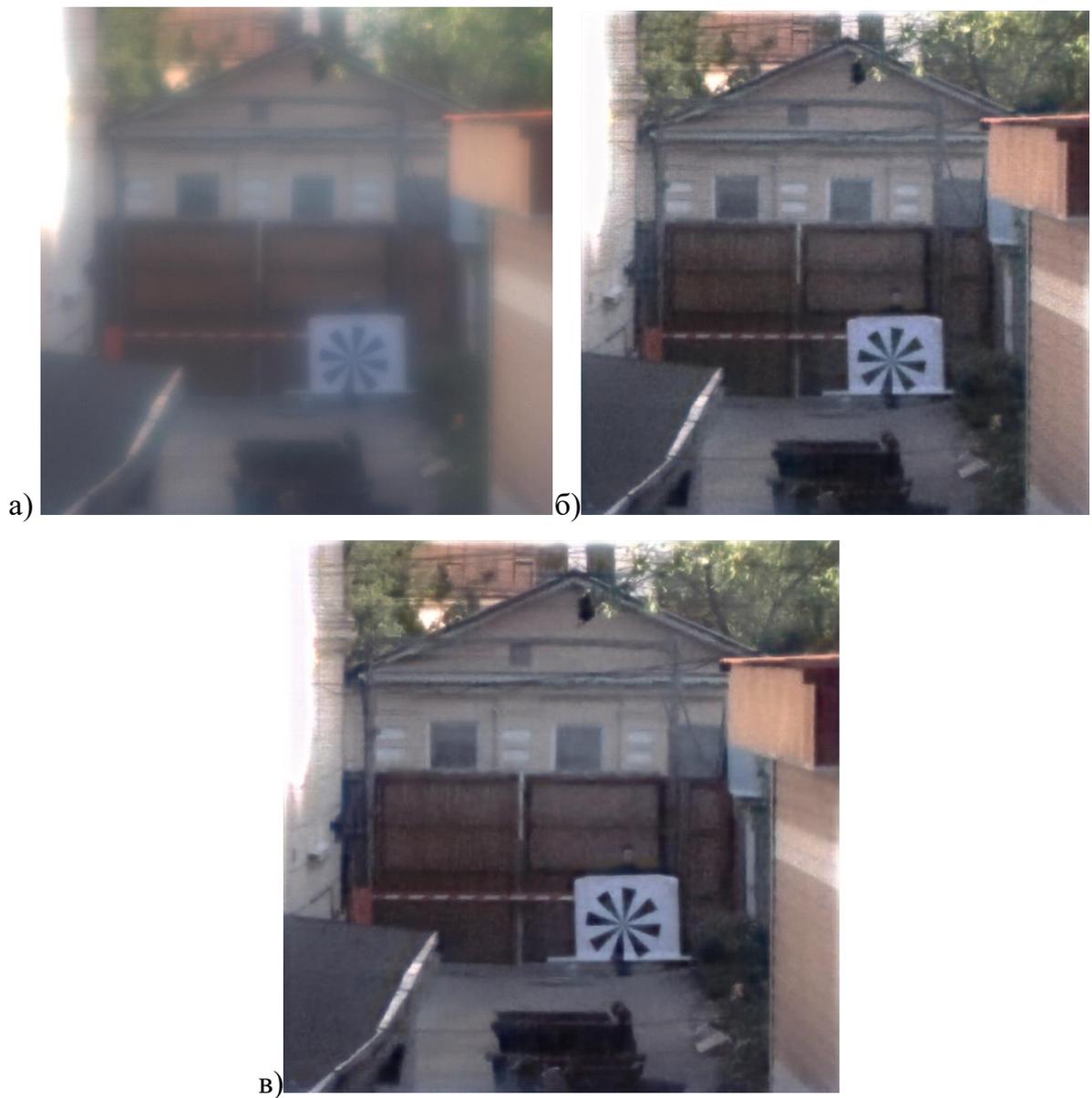
На рисунках В.1-В.4 представлены дополнительные примеры реконструкции реальных изображений для метода крупношагового метаобучения.



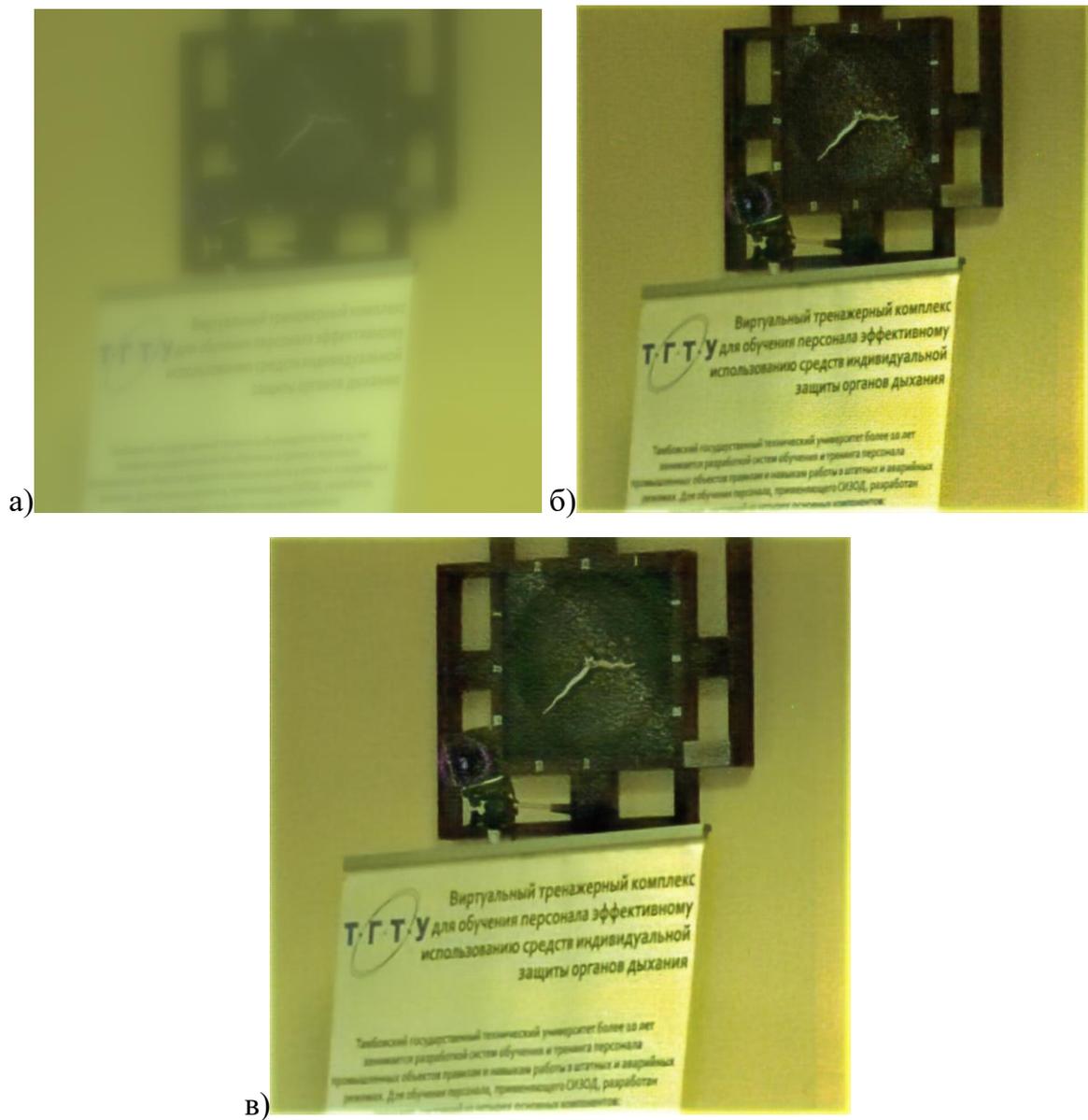
*Рисунок В.1 – Сравнение результатов реконструкции реального изображения, не участвовавшего в процессе обучения: а) реальное изображение, б) обучение в рамках этапов 1 и 3 без крупношагового метаобучения, в) обучение на основе трех этапов с крупношаговым метаобучением*



*Рисунок В.2 – Сравнение результатов реконструкции реального изображения, не участвовавшего в процессе обучения: а) реальное изображение, б) обучение в рамках этапов 1 и 3 без крупношагового метаобучения, в) обучение на основе трех этапов с крупношаговым метаобучением*



*Рисунок В.3 – Сравнение результатов реконструкции реального изображения, не участвовавшего в процессе обучения: а) реальное изображение, б) обучение в рамках этапов 1 и 3 без крупношагового метаобучения, в) обучение на основе трех этапов с крупношаговым метаобучением*



*Рисунок В.4 – Сравнение результатов реконструкции реального изображения, не участвовавшего в процессе обучения: а) реальное изображение, б) обучение в рамках этапов 1 и 3 без крупношагового метаобучения, в) обучение на основе трех этапов с крупношаговым метаобучением*

а

## ПРИЛОЖЕНИЕ Д – СВИДЕТЕЛЬСТВА О РЕГИСТРАЦИИ ПРОГРАММЫ ДЛЯ ЭВМ

На рисунках Г.1-Г.2 представлены свидетельства о регистрации программы для ЭВМ.

РОССИЙСКАЯ ФЕДЕРАЦИЯ



**RU**

**2021612591**

ФЕДЕРАЛЬНАЯ СЛУЖБА  
ПО ИНТЕЛЛЕКТУАЛЬНОЙ СОБСТВЕННОСТИ  
(12) ГОСУДАРСТВЕННАЯ РЕГИСТРАЦИЯ ПРОГРАММЫ ДЛЯ ЭВМ

Номер регистрации (свидетельства):

**2021612591**

Дата регистрации: **19.02.2021**

Номер и дата поступления заявки:

**2021611563 11.02.2021**

Дата публикации: **19.02.2021**

Контактные реквизиты:

**443110, обл. Самарская, г.  
Самара, ул. Мичурина, д. 23,  
офис 433 +79272008091  
artniko@gmail.com**

Авторы:

**Никоноров Артем Владимирович (RU),  
Евдокимова Виктория Витальевна (RU),  
Петров Максим Витальевич (RU),  
Бибииков Сергей Алексеевич (RU)**

Правообладатель:

**Общество с ограниченной ответственностью  
"ЛОКУС" (RU)**

Название программы для ЭВМ:

**Программное обеспечение нейросетевой реконструкции изображений в дифракционно-оптических системах**

**Реферат:**

Программное обеспечение предназначено для решения задачи реконструкции изображений в дифракционно-оптических системах. Программа позволяет устранять хроматические искажения, характерные для данных систем. Функциональные возможности: обучение, тестирование нейронных сетей реконструкции, вычисление метрик качества реконструкции. Работа создана при финансовой поддержке ФГБУ «Фонд содействия развитию малых форм предприятий в научно-технической сфере», соглашение 557ГРНТИС5/49477 от 01.10.2019. Тип ЭВМ: IBM PC-совмест. ПК; ОС: Linux Ubuntu/CentOS.

**Язык программирования: Python**

**Объем программы для ЭВМ: 192 КБ**

*Рисунок Г.1 – Свидетельство о регистрации программы для ЭВМ 2021612591*



ФЕДЕРАЛЬНАЯ СЛУЖБА  
ПО ИНТЕЛЛЕКТУАЛЬНОЙ СОБСТВЕННОСТИ  
(12) ГОСУДАРСТВЕННАЯ РЕГИСТРАЦИЯ ПРОГРАММЫ ДЛЯ ЭВМ

Номер регистрации (свидетельства): <u>2021615738</u>	Авторы: Евдокимова Виктория Витальевна (RU), Петров Максим Витальевич (RU), Бибиков Сергей Алексеевич (RU), Никоноров Артем Владимирович (RU)
Дата регистрации: 12.04.2021	
Номер и дата поступления заявки: 2021614769 07.04.2021	Правообладатель: Общество с ограниченной ответственностью "ЛОКУС" (RU)
Дата публикации: <u>12.04.2021</u>	
Контактные реквизиты: 443110, Самарская обл., г. Самара, ул. Мичурина, 23, кв. 433 +79272008091 artniko@gmail.com	

Название программы для ЭВМ:  
**Облачная платформа нейросетевой реконструкции изображений**

**Реферат:**

Программа представляет собой клиент-серверную платформу для решения задачи реконструкции изображений в дифракционно-оптических системах. Программа имеет следующий функционал: отправка изображений клиентом на обработку, прием изображений и управление очередью задач на сервере, реконструкция изображений на сервере. Работа выполнена при финансовой поддержке ФГБУ «Фонд содействия развитию малых форм предприятий в научно-технической сфере», соглашение 557ГРНТИС5/49477 от 01.10.2019. Тип ЭВМ: IBM PC-совмест. ПК; ОС: Ubuntu, CentOS.

**Язык программирования:** Python

**Объем программы для ЭВМ:** 205 КБ

*Рисунок Г.2 – Свидетельство о регистрации программы для ЭВМ 2021615738*