

Санкт-Петербургский национальный исследовательский университет  
информационных технологий механики и оптики

На правах рукописи

**Щербаков Олег Викторович**

**Реконструкция изображений в отсутствующих или повреждённых  
областях на основе критерия алгоритмической вероятности**

05.13.01 – Системный анализ, управление и обработка информации (в  
технических системах)

Диссертация на соискание ученой степени

кандидата технических наук

Научный руководитель

д.т.н, доцент

Потапов Алексей Сергеевич

**Санкт-Петербург – 2015**

## Оглавление

Введение.....	5
Глава 1. Обзор современного состояния проблем реконструкции изображений в невидимой области .....	13
1.1. Введение.....	13
1.2. Задача реконструкции изображений в утерянных областях .....	14
1.3. Восстановление структуры объектов на изображении в процессе реконструкции .....	16
1.4. Субъективизм в оценке качества реконструкции больших участков изображений .....	18
1.5. Анализ предметной области .....	21
1.5.1. Условная классификация существующих методов.....	22
1.5.2. Текстурные методы реконструкции.....	22
1.5.3. Методы реконструкции, привлекающие аппарат дифференциальных уравнений в частных производных.....	24
1.5.4. Методы реконструкции, использующие поиск по экземпляру .....	28
1.5.5. Гибридные методы реконструкции.....	31
1.5.6. Полуавтоматические методы и методы быстрой реконструкции ...	32
1.5.7. Визуальное восприятие и оценка качества результатов реконструкции.....	34
Выводы по главе 1.....	38
Глава 2. Критерий алгоритмической вероятности и разработка методов на его основе для решения задачи реконструкции изображений в невидимой области.....	40
2.1. Введение.....	40
2.2. Предсказание на основе алгоритмической вероятности.....	44
2.3. Практическая алгоритмическая вероятность .....	47
2.4. Восстановление изображения в утерянной области как частный случай задачи предсказания .....	48

2.5. Метод реконструкции изображений на базе алгоритмической вероятности с использованием спектрального представления .....	52
2.6. Учет предыдущего опыта наблюдателя для решения задачи реконструкции изображений в невидимой области .....	54
2.7. Метод реконструкции изображений на базе алгоритмической вероятности с использованием систем, способных к обучению представлениям .....	58
Выводы по главе 2.....	61
Глава 3. Алгоритмы реконструкции изображений в невидимой области на основе критерия алгоритмической вероятности.....	63
3.1. Введение.....	63
3.2. Алгоритм реконструкции изображения с использованием спектрального представления.....	63
3.3. Об алгоритмах обучения систем представлениям.....	66
3.3.1. Недостаточность критерия максимизации взаимной информации для выделения информативных признаков.....	70
3.3.2. Регуляризация как способ повышения информативности выученных признаков .....	71
3.3.3. Разреженное кодирование.....	73
3.3.4. Обучение шумоподавлению .....	75
3.3.5. Геометрическая интерпретация критерия шумоподавления .....	78
3.3.6. Регуляризация и компактность кода.....	83
3.4. Сверточный автоэнкодер.....	84
3.4.1. Разреженное кодирование для сверточного автоэнкодера.....	86
3.5. Алгоритмы реконструкции изображений в невидимой области, использующие обучаемые представления системы .....	87
Выводы по главе 3.....	89
Глава 4. Практическое применение алгоритмов реконструкции изображений в невидимой области на основе критерия алгоритмической вероятности. ....	90

4.1. Введение.....	90
4.2. Практическое применение алгоритма реконструкции изображений на основе критерия информативности их спектров.....	90
4.3. Практическое применение алгоритмов реконструкции изображений на основе использования систем, способных к обучению представлениям ..	104
Выводы по главе 4.....	114
Заключение .....	115
Список литературы .....	117

## **Введение**

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

В результате развития компьютерной техники и цифровых фото- и видеокамер стали массово востребованными многие задачи автоматической обработки и анализа изображений, которые хотя и начали изучаться еще пол века назад, в настоящее время приобретают конкретное наполнение, диктуемое практикой. Примером одной из таких задач является задача восстановления изображений в утеранных или поврежденных областях.

Решение данной задачи оказывается нужным в самых разных приложениях. Примерами может служить:

- редактирование пользовательских фотографий в целях удаления из них нежелательных объектов (с заполнением их областей тем фоном, который мог бы быть на фотографии в их отсутствие), портящих композицию снимка;
- удаление изображений людей на публично доступных снимках в системах типа Google Street View;
- реставрация старых фильмов.

Задача восстановления изображений в утеранных или поврежденных областях актуальна не только в прикладном, но и научном плане. Действительно, данная задача не имеет в настоящее время удовлетворительного решения и вызывает ряд научных вопросов. Существует большое число методов ее частичного решения, разработанных в рамках разных подходов и использующих различные типы представлений изображений. Примерами таких методов могут служить алгоритмы реконструкции на основе дифференциальных уравнений в частных производных, предложенные Bertalmio M., Sapiro G., Caselles V. и Ballester C., алгоритмы реконструкции на основе поиска по экземпляру, предложенные Criminisi P. и Toyama K., а также Hays J. и Efros A., алгоритмы полуавтоматической реконструкции, предложенные Sun J., Yuan L., Jia J. и

Shum H. При этом не просто отсутствует обобщающая теория для всех этих методов, но даже нет единых критериев их оценивания в связи с тем, что задача «угадывания» отсутствующего содержания выглядит как математически некорректная.

Тем не менее, данную задачу можно попытаться рассмотреть как специфический случай проблемы предсказания битовых строк, для которой существует общетеоретическое решение. Исследование возможности применения этого решения к задаче восстановления изображений в утерянных или поврежденных областях может позволить создать единую теоретическую основу для существующих методов и разработать новые эффективные методы, что и обуславливает актуальность настоящей работы.

## **Цель работы**

Разработка методов восстановления информации в утерянных или поврежденных областях изображений, подверженных влиянию аддитивного и структурного шума, с использованием критерия алгоритмической вероятности применительно к различным типам представления изображений, в том числе обучаемым.

## **Основные задачи:**

- постановка задачи восстановления изображений в утерянных областях как задачи предсказания в терминах теории универсального предсказания;
- вывод критерия оптимальности предсказания для решения задачи восстановления изображения на основе алгоритмической вероятности;
- вывод критерия оптимальности предсказания для решения задачи восстановления изображения на основе моделей, способных к обучению представлениям;

- разработка методов заполнения отсутствующих областей изображения на основе критерия алгоритмической вероятности;
- разработка методов восстановления утерянных участков изображения с использованием методов, способных к обучению представлениям.

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

1. Формализация задачи реконструкции изображений в утерянной или поврежденной области, содержащая критерий качества решения этой задачи на основе алгоритмической вероятности, в виде частного случая проблемы универсального предсказания.

2. Аналитическое обоснование использования и непосредственное применение генеративных моделей специального типа для решения задачи реконструкции изображений в утерянной или поврежденной области.

3. Метод реконструкции изображений в утерянной или поврежденной области, основанный на критерии информативности его спектра, способный к восстановлению структурной и текстурной информации.

4. Метод реконструкции изображений в утерянной или поврежденной области, основанный на применении методов, способных к обучению представлениям.

5. Алгоритмы итеративной реконструкции и реконструкции на базе эволюционных стратегий поиска для методов восстановления изображений, использующих специальные архитектуры сетей глубокого обучения.

### **Научная новизна работы**

- Впервые были применены сверточные нейронные сети, способные к обучению представлениям, при решении задачи восстановления утерянных областей изображения.

- Впервые применен подход на основе критерия алгоритмической вероятности в задаче заполнения отсутствующих участков изображения.
- Разработан алгоритм восстановления изображения с применением генетических алгоритмов и шумоподавляющих автоэнкодеров.
- Разработан алгоритм восстановления изображения на основе критерия информативности спектра этого изображения.

### **Практическая значимость**

- автоматическое редактирование изображений с целью удаления или восстановления объектов интереса;
- применимость метода к широкому классу задач восстановления утерянной информации на изображении, включая такие примеры, как база лиц, база рукописных цифр, фотографии различных объектов реального мира и др;
- алгоритм также может быть применен к восстановлению изображений, передаваемых по зашумленным каналам связи в системах передачи данных, к примеру, при потоковой передаче видео.

### **Личный вклад автора**

Все результаты получены автором диссертации лично или при его непосредственном участии.

### **Реализация результатов работы**

Результаты диссертационной работы были использованы при выполнении работ по НИР «Разработка теории анализа изображений на основе принципа репрезентационной минимальной длины описания» (проект РНП 2.1.2/9645 по аналитической ведомственной целевой программе "Развитие научного потенциала высшей школы по заказу Федерального



агентства по образованию), грантам Президента РФ (проект МД-2040.2010.9 и МД-1072.2013.9) «Разработка теории обучаемых систем анализа изображений и распознавания образов на основе принципа репрезентационной минимальной длины описания» и «Разработка теории вычислимых аппроксимаций алгоритмической вероятности в моделях машинного обучения и восприятия», НИР «Исследование проблем распознавания изображений в информационных системах и построение теории синтеза алгоритмов распознавания на основе интеллектуальных технологий» (Государственное задание образовательным организациям высшего образования, базовая часть) и НИР «Исследование методов формирования и интерпретации изображений объектов в обучаемых интеллектуальных системах» (Государственное задание образовательным организациям высшего образования, проектная часть).

Акты внедрения приложены к диссертации.

### **Апробация работы**

Основные результаты работы докладывались на следующих научных форумах:

1. XLII научная и учебно-методическая конференция НИУ ИТМО. НИУ ИТМО Санкт-Петербург. 29.01.2013 – 01.02.2013;
2. 11<sup>th</sup> International Conference on Quality Control by Artificial Vision (QCAV 2013). Fukuoka, Japan. 30.05.2013 – 01.06.2013;
3. Sixth International Conference on Machine Vision (ICMV 2013). London, UK. 16.11.2013-17.11.2013;
4. XLIII научная и учебно-методическая конференция НИУ ИТМО. НИУ ИТМО, Санкт-Петербург. 28.01.2014 - 31.01.2014;
5. III Всероссийский конгресс молодых учёных. НИУ ИТМО, Санкт-Петербург. 08.04.2014 – 11.04.2014;
6. 2014 International Conference on Future Communication Technology and Engineering (FCTE2014);

7. XLIV научная и учебно-методическая конференция Университета ИТМО. Университет ИТМО Санкт-Петербург.

**Публикации по теме диссертации**

По материалам диссертации опубликовано 8 научных работ, из которых 5 опубликованы в изданиях, рекомендованных ВАК, в том числе 3 статьи в российских журналах, входящих в перечень ведущих рецензируемых научных журналов и изданий, формируемый ВАК РФ, и 2 статьи, опубликованные в журналах, индексируемых базой цитирования Scopus, а также 2 свидетельства о государственной регистрации программ для ЭВМ.

1. Щербаков О.В. Сверточный автоэнкодер как генеративная модель изображений для задач выделения признаков и восстановления изображений в утерянных областях [Текст] / О.В. Щербаков, И.Н. Жданов, Я.А. Лушин // Оптический журнал. – 2015. – Т. 82. – № 8. – С. 48-53.
2. Потапов А.С. Метод предсказания на основе алгоритмической вероятности в задаче восстановления изображений в утерянных областях [Текст] / А.С. Потапов, О.В. Щербаков, И.Н. Жданов // Оптический журнал. – 2013. – Т. 80. – №11. – С. 51-56.
3. Potapov A. Practical algorithmic probability: an image inpainting example [Text] / A. Potapov, O. Scherbakov, I. Zhdanov // Proc. SPIE, Sixth International Conference on Machine Vision (ICMV 2013). – 2013. – Vol. 9067. – P. 906719-906719-5.
4. Shcherbakov O. Image inpainting based on stacked autoencoders [Text] / O. Shcherbakov, V. Batishcheva // Journal of Physics: Conference Series. – 2014. – V. 536. – P. 012020.
5. Жданов И. Н. Метод сжатия трёхмерных биомедицинских изображений на основе представления информации в виде октодерева [Текст] / И. Н. Жданов, А. С. Потапов, О. В. Щербаков // Научно-технический вестник

информационных технологий, механики и оптики. – 2012. – Т. 79. – № 3. – С. 100-104.

6. Потапов А.С. Реконструкция изображения в невидимой области по критерию информативности его пространственного спектра [Текст] / А.С. Потапов, О.В. Щербаков, И.Н. Жданов // Сборник трудов II конференции «Будущее оптики» для молодых специалистов, кандидатов наук, аспирантов и студентов оптической отрасли и смежных дисциплин. 2–3 апреля 2013. – С. 48–50.
7. Щербаков О.В. Программный модуль «Система восстановления утраченных областей на изображении на базе стека автоэнкодеров». Жданов И.Н., Потапов А.С. Свидетельство о государственной регистрации программы для ЭВМ № 2014660857. Роспатент. 17.10.2014.
8. Щербаков О.В. Программный модуль «Сжатие трёхмерных биомедицинских изображений с использованием их представления в виде октодерева». Потапов А.С., Жданов И.Н. Свидетельство о государственной регистрации программы для ЭВМ № 2012618609. Роспатент. 21.09.2012.

### **Структура и объём диссертации**

Диссертация состоит из введения, четырёх глав и списка цитируемой литературы. Она содержит 128 страниц машинописного текста, 49 рисунков и 3 таблицы. Список цитируемой литературы содержит 106 наименований. Нумерация формул и рисунков сквозная по всей диссертации.

**Во введении** обосновывается актуальность темы и сформулированы цель, научная новизна и практическая значимость работы. Представлены основные результаты и положения, выносимые на защиту, а также кратко изложено содержание разделов диссертации.

**В первой главе** проведён анализ существующих методов, показаны их преимущества и недостатки, указаны трудности, связанные с решением

задачи реконструкции изображений в отсутствующих или повреждённых областях.

**Во второй главе** проводится постановка и аналитическое обоснование задачи реконструкции изображений в отсутствующих или повреждённых областях как частного случая более общей задачи универсального предсказания. Вводится критерий качества ее решения на основе алгоритмической вероятности. Показывается необходимость замены универсальной опорной машины неуниверсальными, которым могут быть поставлены в соответствие представления изображений, задаваемые в форме генеративных моделей. Приводится аналитическое обоснование использования спектральных и обучаемых представлений изображений для решения задачи реконструкции изображений в невидимой области, что дает возможности для разработки методов с указанными типами представлений на основе объективного критерия.

**В третьей главе** показывается процесс воплощения метода реконструкции изображений, основанного на критерии информативности спектра изображения, а также методов, способных к обучению представлениям на основе стековых сверточных и несверточных автоэнкодеров в форму алгоритмов итеративной реконструкции и реконструкции на базе эволюционных стратегий поиска.

**В четвертой главе** обсуждаются различные аспекты практического применения алгоритмов, предложенных в третьей главе. Проводится сравнение предлагаемых алгоритмов с программными реализациями уже известных методов.

**В заключении** проводится обобщение материала и результатов, изложенных в настоящей работе.

# **Глава 1. Обзор современного состояния проблем реконструкции изображений в невидимой области**

## **1.1. Введение**

Процесс реконструкции изображений в утеранных областях подразумевает заполнение этих областей содержанием с использованием информации из остальных участков изображения. На практике утеранные области описываются пользователем в виде специальных масок или могут задаваться автоматическим или полуавтоматическим способом. Процесс реконструкции изображений в утеранных областях или, как его еще называют, врисовывания отсутствующих областей (от английского названия данной задачи – inpainting) является распространенной задачей, которая возникает, например, во время реставрации поврежденных старых картин и фотографий, удаления нежелательных объектов и надписей на фотографиях, восстановления изображений и видеопоследовательностей, прошедших по зашумленным каналам связи, мультимедийной обработки изображений, а также различных процедур по защите визуальной информации. Цель такой реконструкции – замена поврежденных областей на изображении таким образом, чтобы замененная область была бы незаметной для стороннего наблюдателя.

Немного иначе можно определить данную задачу как процесс воссоздания отсутствующих или поврежденных областей таким способом, чтобы сделать общую картину более отчетливой, понятной и согласованной. В зависимости от контекста, может требоваться как получение результата восстановления поврежденного изображения наиболее похожего на некий оригинал, так и получение альтернативного от оригинального заполнения незаметного для человека-наблюдателя.

С разработкой алгоритмов реконструкции изображений в утеранных областях связан ряд сложностей. Во-первых, такие алгоритмы должны производить качественное текстурное и структурное заполнение отсутствующих областей. Во-вторых, желательно, чтобы время работы таких

алгоритмов было относительно небольшим. В-третьих, такие алгоритмы должны быть устойчивыми к изменению различных функциональных условий. И, в-четвертых, нетривиальной задачей представляется количественная оценка качества результатов восстановления с точки зрения визуального восприятия.

Далее мы несколько подробнее рассмотрим суть задачи реконструкции изображений в утеранных областях, а также связанные с ней задачи и трудности, возникающие при их решении.

## **1.2. Задача реконструкции изображений в утеранных областях**

Исторически задача реконструкции изображений в утеранных областях решалась художниками и реставраторами при восстановлении поврежденных произведений изобразительного искусства и фотографий. Как правило, это подразумевало устранение незначительных дефектов вроде царапин, трещин, капель, а также нежелательных эффектов, возникающих в процессе фотосъемки. Но с развитием цифровых технологий появилась насущная необходимость в автоматизации некоторых из этих процедур.

Может показаться, что часть описанных выше дефектов можно рассматривать как некий шум на изображении и, соответственно, применить к его устранению техники шумоподавления, разработанные, в частности, в рамках дисциплины цифровой обработки изображений. Но, несмотря на это, стоит особо отметить, что на текущий момент задача реконструкции изображений в утеранных областях и классическая задача шумоподавления имеют фундаментальные отличия между собой. В случае классического шумоподавления исходный сигнал искажен шумом, и задача алгоритма шумоподавления состоит в восстановлении исходного сигнала посредством построения моделей шума и различных статистических оценок.

Типичными примерами моделей шумов могут быть аддитивный гауссовский шум, спекл-шум, который можно наблюдать на интерференционных изображениях и томограммах, шум типа «соль и перец» и другие. В качестве модели сигнала используют представление изображения

в виде кусочно-непрерывной двумерной функции. Существует множество алгоритмов фильтрации как в пространственной, так и в других областях (например, частотной), которые призваны устранить шум и восстановить исходное, истинное изображение.

С другой стороны, в задаче реконструкции изображений в утеранных областях регионы, в которых отсутствует информация об исходном сигнале, имеют достаточно большой размер, и алгоритм врисовки пытается воссоздать эти регионы изображения, используя всю имеющуюся вокруг информацию. Таким образом, алгоритмы шумоподавления напрямую, как правило, неприменимы при решении задачи реконструкции изображений в утеранных областях. Примеры изображений для алгоритмов шумоподавления и реконструкции приведены на рисунке 1.



Рисунок 1 – Исходные данные для задач шумоподавления и реконструкции. Слева – изображение искаженное шумом, которое может подаваться на вход алгоритмам шумоподавления. Справа – изображение с отсутствующим фрагментом, который представлен квадратной «дыркой» и который требуется реконструировать

Чтобы продемонстрировать применение алгоритмов реконструкции для целей реставрации изображений, приведем пример изображения, содержащего дефекты в виде царапин и трещин (рисунок 2 слева), а также пример соответствующей данным дефектам пользовательской маски (рисунок 2 в центре). Используя эти изображения в качестве входных данных, алгоритм реконструкции пытается заполнить отсутствующие

области, обозначенные маской, на основе статистической информации, содержащейся в оставшихся фрагментах исходного изображения. Результат применения такого алгоритма реконструкции, описанного в работе [1], представлен на рисунке 2 справа.

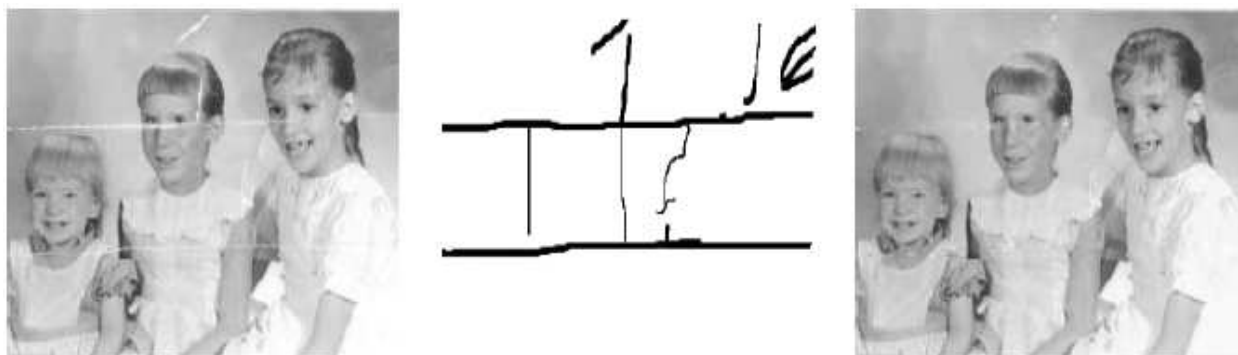


Рисунок 2 – Пример применения алгоритма реконструкции для реставрации старых фотоснимков: слева – исходное изображение, в центре – задаваемая пользователем маска поврежденных областей, справа – результат реконструкции, полученный после работы алгоритма описанного в [1]

Несмотря на существенные успехи, сделанные в области исследования алгоритмов реконструкции, в ряде приложений этих алгоритмов к обработке реальных изображений возникают некоторые сложности. Далее мы рассмотрим основные проблемы, которые возникают в процессе работы существующих алгоритмов реконструкции изображений в невидимой области.

### **1.3. Восстановление структуры объектов на изображении в процессе реконструкции**

Существующие методы реконструкции изображений в невидимой области показывают неплохие результаты в случаях, когда области, подлежащие реконструкции, имеют сравнительно небольшой размер и практически полностью не имеют текстуры. Примером такого случая можно считать изображения на рисунке 2. Когда область реконструкции имеет сравнительно большой размер, а также должна содержать внутри себя



предполагаемые границы объектов, то в таких случаях существующие алгоритмы и методы реконструкции не дают удовлетворительных результатов. Зачастую это связано с тем, что многие из существующих методов реконструкции пытаются заполнить невидимые области за счет процессов сглаживания или фильтрации, которые естественным образом размывают границы объектов и не позволяют обрабатывать эти границы в явном виде.

Чтобы проиллюстрировать данный эффект, можно привести пример работы алгоритма, описанного в работе [2]. На рисунке 3 слева приведен пример исходного изображения, в центре находится изображение с круговой областью, которая искусственно была удалена из исходного изображения пользователем, справа же показан результат применения алгоритма описанного в [2] к изображению в центре.



Рисунок 3 – Иллюстрация проблем связанных с реконструкцией больших областей на изображении. Слева – исходное изображение. В центре – изображение с отсутствующей областью. Справа – результат работы алгоритма описанного в работе [2]

Полученный результат вряд ли можно назвать хорошим с точки зрения визуального восприятия, особенно в сравнении с оригиналом. Стоит заметить, что данный метод, столкнувшись с необходимостью реконструкции относительно большой области, не справился с восстановлением структурной границы между двумя сегментами на изображении. Для примера на рисунке 3 такой структурной границей являлся бы кусочек кривой второго порядка.

Одним из интуитивных способов решения данной проблемы были бы проведение сегментации изображения, замыкание соответствующих границ сегментов и заливка каждого из полученных сегментов подходящей текстурной информацией. Несмотря на то, что данный процесс «мысленно» легко осуществляет человек, для систем и алгоритмов компьютерного зрения такая задача до сих пор представляет собой существенную трудность.

Как можно увидеть из приведенного выше примера, процесс замыкания контуров играет важную роль для получения визуально удовлетворительной реконструкции изображений. Данная задача относится к классу задач замыкания структур и решается в рамках структурных представлений изображений среднего уровня.

Также в рамках данного подхода рассматриваются и другие важные структурные компоненты изображений, которые помогают более точно оценивать и замыкать границы сегментов. Среди таких компонент бывают информация о глубине, повторяющихся элементах или паттернах на изображении, учет различных симметрий и контурных представлений. Зачастую это помогает точнее оценивать параметры границ, а значит, и повышать качество реконструкции [3].

#### **1.4. Субъективизм в оценке качества реконструкции больших участков изображений**

Оценка качества результатов реконструкции изображений в невидимой области представляет собой сложную научную задачу в силу своих принципиальных отличий от базового понятия качества изображения вообще. Иными словами изображение, не содержащее никаких визуальных артефактов, тем не менее, может быть расценено как изображение неудовлетворительного качества с точки зрения реконструкции.

К примеру, для синтетических изображений, содержащих геометрические фигуры, результат дорисовки границ этих самых фигур может быть воспринят более удовлетворительным, нежели простая

подстановка в отсутствующую область окружающей текстуры (рисунок 4). И, наоборот, для естественных изображений заливка текстурой внутрь естественных границ объектов может давать более хороший результат (рисунок 5).

Таким образом, текстура, структура, размер отсутствующей области и прочие характеристики содержимого изображения играют ключевую роль в определении эффективности работы алгоритма реконструкции. Требование учета такого разнообразия характеристик делает процесс оценки качества реконструкции крайне сложным.

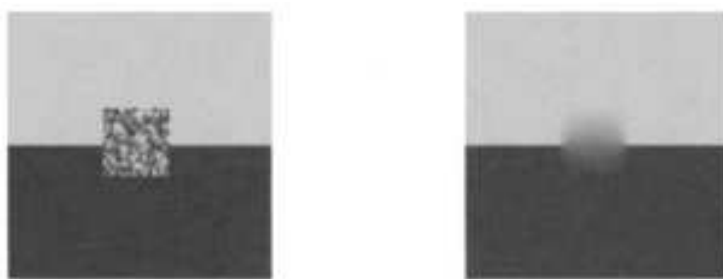


Рисунок 4 – Иллюстрация визуально неудовлетворительных результатов реконструкции для синтетических изображений при текстурной заливке области реконструкции



Рисунок 5 – Иллюстрация визуально хороших результатов реконструкции для естественных изображений при текстурной заливке области реконструкции

Существует несколько работ, посвященных оценке качества результатов реконструкции изображений в невидимой области. Эти работы мы обсудим далее, здесь только лишь отметим, что в них делаются попытки сравнительной оценки качества работы различных алгоритмов без систематического обоснования выбираемых критериев. Существует ряд исследований зрительной системы человека, в которых показано, что интересные или необычные для человеческого восприятия области привлекают зрительное внимание людей, заставляя их более тщательным образом рассматривать такой участок изображения [4, 5]. Соответственно, можно предположить, что области, которые будут некорректно или не соответствующим образом реконструированы, привлекут внимание наблюдателей, вызвав отклонение в поведении зрительной системы человека. Можно попытаться дать количественную оценку такого рода отклонениям,

измеряя изменения в паттернах движения глаз по изображениям средствами окулографии и пытаюсь выявить взаимосвязь со зрительным восприятием качества реконструкции. Именно таким методикам посвящена, в частности, работа [6]. Аргументируется такой подход тем, что, несмотря на активное участие в процессе формирования внимания высших когнитивных функций, средне- и низкоуровневые зрительные механизмы играют важную роль в формировании зрительного восприятия, например, от результатов реконструкции. И, действительно, в работе [6] показано наличие сильных корреляций между паттернами движения взора наблюдателя и качеством реконструкции изображений в невидимой области. Существуют основания полагать, что данные методики при условии активного привлечения численного и математического моделирования могут стать наиболее перспективными способами оценки качества восстанавливаемых фрагментов изображений. Однако, по причине необходимости привлечения к процедурам оценки зрительной системы человека, пусть даже и на основе обширных репрезентативных выборок, такие техники будут обладать изрядной долей субъективизма. Именно этот фактор является одной из основных проблем в обосновании критериев качества результатов автоматической компьютерной врисовки недостающих областей на изображениях.

### **1.5. Анализ предметной области**

Данный раздел посвящен подробному анализу разнообразных существующих на текущий момент подходов, алгоритмов и техник, предназначенных для решения задачи реконструкции изображений в невидимой области. Также будут обсуждаться их преимущества и недостатки. Помимо этого несколько слов будет сказано и о работах, в которых затрагивается вопрос оценки качества при визуальном восприятии результатов реконструкции изображений в невидимой области.

### **1.5.1. Условная классификация существующих методов**

На текущий момент существует несколько подходов к решению задачи реконструкции изображений в невидимой области. Несмотря на то, что всякая классификация условна, можно выделить следующие основные направления развития в понимании процесса решения, обозначенной выше задачи.

1. Текстурные методы реконструкции.
2. Методы реконструкции, привлекающие аппарат дифференциальных уравнений в частных производных.
3. Методы реконструкции, использующие поиск по экземпляру.
4. Гибридные методы.
5. Полуавтоматические методы и методы быстрой реконструкции.

### **1.5.2. Текстурные методы реконструкции**

Одной из самых ранних тенденций в разработке алгоритмов реконструкции было использование алгоритмов синтеза текстуры для заполнения недостающих фрагментов. Такие алгоритмы пытались подобрать новые пиксели исходя из начального предположения, при этом пытаясь сохранить локальные структуры на изображении. Наиболее ранние из таких методик просто-напросто заполняли недостающие фрагменты при помощи сэмплирования или копирования пикселей из соседних областей [7-12]. Например, в работе [7] марковская сеть использовалась для моделирования локального распределения пикселей. Новая текстура синтезировалась из имеющейся, а после производилось уточнение значений соседних пикселей. Далее решалась задача обеспечения условия непрерывности между границами реконструируемой области и восстановленным содержимым. Такие методы показывают неплохие результаты лишь для ограниченного множества изображений, для которых заполнение отсутствующей области

приведет к более или менее естественным для данных изображений результатам.

Позднее данный подход был расширен до, так называемого, алгоритма быстрого синтеза текстур, который заключался в сшивке между собой маленьких фрагментов имеющейся области изображения [8]. Такой процесс получил название «стёжки» изображения. Также известны алгоритмы параметрического синтеза текстуры, которые способны порождать соответствующую задаче реконструкции текстуру по фрагментам данной опорной текстуры [10]. Синтез текстур с мультиразрешением тоже применялся для решения задачи реконструкции изображений в невидимой области [13]. При этом в работе [13] отмечается возможность генерации такой текстуры с различными параметрами яркости. Также стоит отметить метод, который для синтеза текстур использует фрагментарный анализ изображения при помощи алгоритма анализа главных компонент (АГК) [14]. Различные методы синтеза текстур, представленные в данном обзоре, и которые применяются для решения задачи реконструкции изображений, рознятся между собой своими способностями создавать текстуры при помощи различных статистических характеристик и варьировать полученные текстуры по различным параметрам, вроде яркости, цвета и другим. Существует бесчисленное количество методов синтеза текстуры, но здесь мы ограничимся рассмотрением тех из них, которые имеют отношение к решению задачи реконструкции изображений в невидимой области.

Несмотря на способность текстурных методов реконструкции аппроксимировать текстуры, у таких методов все же возникают сложности с восстановлением естественных изображений, так как такие изображения представляют собой помимо всего прочего еще и набор структурных элементов, как правило, описываемых границами областей занимаемых объектами. Такой тип границ и границы текстур обычно представляются двумя разными множествами, пусть и частично пересекающимися между

собой. Соответственно, методы реконструкции, базирующиеся на синтезе текстур, могут описать лишь малую часть случаев, в которых возникает задача восстановления утерянной информации на изображении, и существует большое число случаев, в которых такие методы просто неприменимы.

### **1.5.3. Методы реконструкции, привлекающие аппарат дифференциальных уравнений в частных производных**

Одним из первых алгоритмов, использовавших техники дифференциальных уравнений в частных производных для решения задачи реконструкции изображений в невидимой области, был алгоритм, описанный в работе [1]. Данный подход позволил по-новому взглянуть на задачу восстановления утерянной информации на изображении. Результат работы данного алгоритма был уже нами продемонстрирован на рисунке 2. Идеи для реализации данного метода были заимствованы из практики ручной реставрационной работы художников, а также из теоретической теплофизики. Интуитивно понятным первым шагом в данном подходе, используя уравнение теплопроводности (1), могло бы стать рассмотрение модели распространения интенсивности (2) по аналогии с моделью тепловой диффузии.

$$\frac{\partial u(x,t)}{\partial t} = a^2 \frac{\partial^2 u(x,t)}{\partial x^2}, \quad (1)$$

где  $u$  – температура стержня (так как здесь всего лишь одна пространственная координата  $x$ ),  $t$  – время,  $a$  – коэффициент температуропроводности.

$$I_t(t, x, y) = a^2 (I_{xx} + I_{yy}) = a^2 \Delta I, \quad (2)$$

где  $I$  – интенсивность пикселя изображения, которая зависит от его местоположения, задаваемого координатами  $x$  и  $y$ , а также времени  $t$ . Коэффициент  $a$  в данном случае будет играть роль темпа восстановления.



Но такой первый интуитивный способ описания будет давать некорректные результаты для больших по размеру областей, подлежащих восстановлению. Пример такого рода результатов можно наблюдать на рисунке 6.

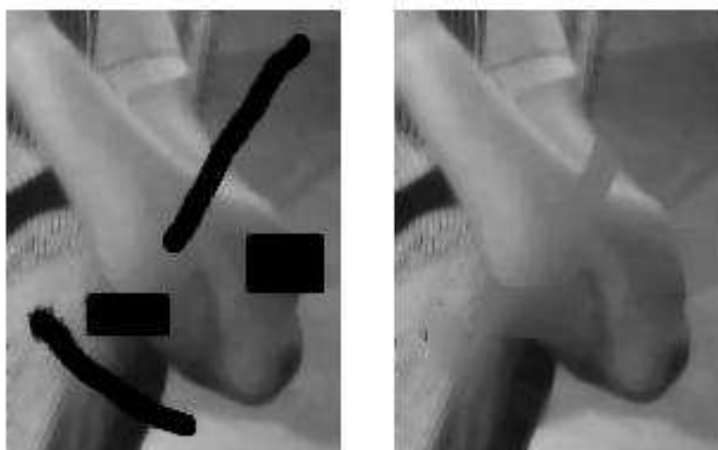


Рисунок 6 – Пример некорректных результатов восстановления изображения методом диффузии интенсивности при больших размерах областей восстановления

Можно увидеть, что основной трудностью здесь является процесс восстановления границ. Соответственно, следующим шагом в модернизации данного алгоритма был учет опыта работы реставраторов. Этот опыт подсказывал необходимость заполнения восстанавливаемой области в направлении границ объектов на изображении. В результате получается итеративный процесс продления структурных границ областей объектов на изображении (такие границы получили название – изофоты) внутрь отсутствующей области. Формально он записывается так:

$$I_t = \nabla(\Delta I) \cdot \vec{N}, \quad (3)$$

где  $N$  – вектор, указывающий направление изофоты и определяющийся как  $\vec{N} = [-I_y, I_x]$ .

Наглядная иллюстрация схемы работы метода показана на рисунке 7.

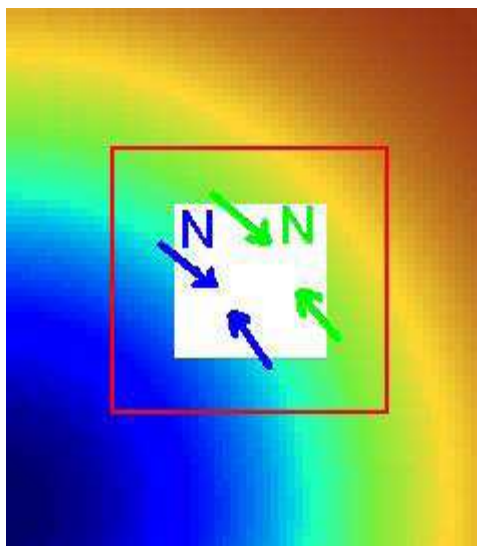


Рисунок 7 – Наглядная иллюстрация схемы работы метода реконструкции изображений в невидимой области на основе уравнений анизотропной диффузии

Такая методика восстановления утерянных областей изображений получила название анизотропной диффузии, так как учитывает специфическую информацию о направлении изофот. Интересные результаты работы таких алгоритмов продемонстрированы на рисунках 8 и 9 для синтетических и естественных изображений соответственно.

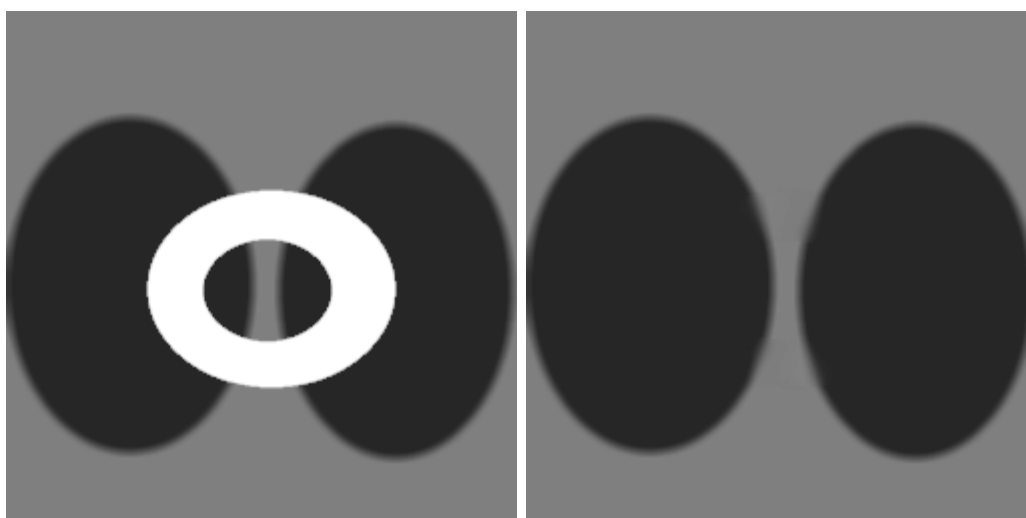


Рисунок 8 – Результаты работы метода реконструкции изображений в невидимой области на основе уравнений анизотропной диффузии

применительно к синтетическим изображениям. Слева – изображение с отсутствующей кольцевидной областью. Справа – результат реконструкции

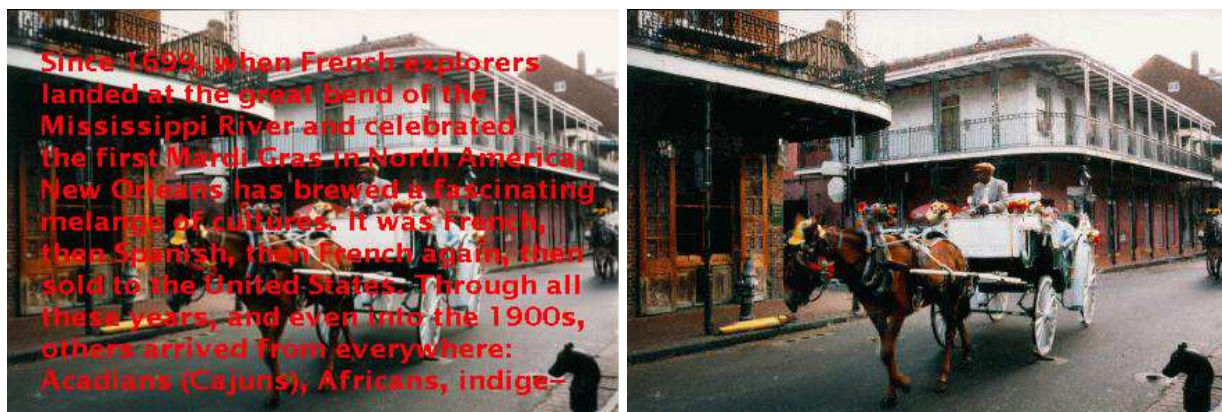


Рисунок 9 – Результаты работы метода реконструкции изображений в невидимой области на основе уравнений анизотропной диффузии применительно к естественным изображениям. Слева – исходное изображение, искаженное наложенным текстом. Справа – результат реконструкции

Одним из очевидных недостатков данного подхода является невозможность восстановления с его помощью фрагментов изображения, требующих репликации больших текстурных областей, поскольку процесс диффузии неминуемо приведет к размытию, а также неправильным образом будут трактованы интенсивности пикселей, находящихся на границе с восстанавливаемым фрагментом.

Дальнейшим развитием представленных выше идей послужила работа [15]. В этой работе предлагается модель реконструкции на основе вариационного исчисления. В ней рассматривается некоторая расширенная область реконструкции, включающая части оригинального изображения. На основании информации имеющейся в данной области модель пытается найти такую функцию  $u$ , для которой минимизировался бы функционал  $R(u)$ . При этом для областей смежных с областью подлежащей восстановлению вводится специальный критерий шумоподавления. В общем виде подобный функционал можно записать так:

$$R(u) = \int_{E \cup D} r(|\nabla u|) dx dy, \quad (4)$$

где  $E$  – части оригинального изображения, соприкасающиеся с областью реконструкции,  $D$  – область реконструкции,  $r$  – специальная вещественная функция, которая будет давать неотрицательные значения при неотрицательных аргументах.

Как и предыдущий метод, данный подход хорошо восстанавливает небольшие фрагменты изображений, а также крайне неплохо решает частный случай задачи реконструкции изображений – задачу интерполяции изображений, например, при многократном увеличении. Также стоит отметить, что данный метод справляется в некоторых случаях и с решением задачи шумоподавления. К недостаткам метода также можно отнести его неспособность эффективно восстанавливать текстурные фрагменты изображения.

Путем наложения различных регуляризационных ограничений описанная выше техника реконструкции была расширена для более качественного восстановления фрагментов, содержащих границы структурных элементов повышенной кривизны. Примеры подобных модификаций описаны в работе [16]. Ожидаемым недостатком таких модификаций является все та же невозможность реконструкции текстур.

#### **1.5.4. Методы реконструкции, использующие поиск по экземпляру**

Важным классом алгоритмов реконструкции являются алгоритмы поиска по экземпляру. При этом сам класс таких алгоритмов может быть разделен на два подкласса на основании различий в методологии, заложенной в каждом из этих подклассов.

Тем не менее, оба подхода оказались довольно эффективными в решении задачи реконструкции изображений в невидимой области. Первый способ восстановления больших фрагментов изображений в какой-то степени совмещает в себе техники синтеза текстур и диффузии и был предложен в работе [2]. В данном алгоритме порядок реконструкции области, подлежащей

восстановлению, определяется механизмом задания приоритетности и может быть схематически проиллюстрирован на рисунке 10.

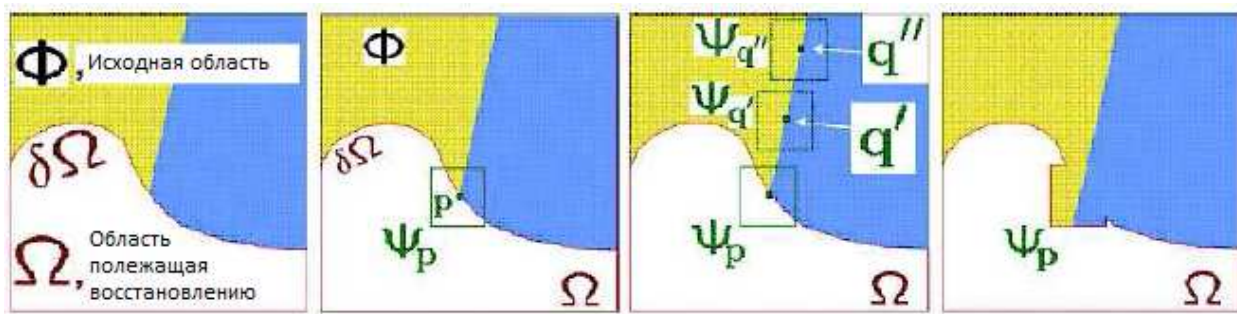


Рисунок 10 – Экземплярная реконструкция изображений (по аналогии с рисунком из [2]). Слева направо: исходное изображение с отсутствующей областью; заполняемый участок, содержащий пиксель с наивысшим приоритетом; наиболее подходящие кандидаты для заполнения; пиксель с наивысшим приоритетом заполнен наиболее подходящим экземпляром

Точки, которые лежат на границе областей, будут обладать большим приоритетом и, следовательно, окрестности вокруг них будут заполнены раньше, чем вокруг других пикселей. На рисунке 10 показана такая точка с высоким приоритетом, которая лежит на границе области реконструкции. Окрестность вокруг точки с наивысшим приоритетом заполняется наиболее подходящим фрагментом, выбираемым из исходной области, как показано на рисунке 10. Размер фрагмента для заполнения может варьироваться на основе характеристик изображения.

Такой способ реконструкции показывает неплохие результаты для сравнительно большого количества разнообразных изображений, восстанавливая текстуру и структуру изображений в отсутствующих областях. Правда, он испытывает некоторые сложности при восстановлении искривленных структурных границ.

Но главный недостаток данного метода проявляется в случаях, когда в процессе работы будет неправильно или недостаточно точно оценен наиболее подходящий фрагмент, тогда это может привести к

лавинообразному накоплению систематической ошибки от итерации к итерации и, в конце концов, к абсолютно некорректным результатам работы.

Несколько более усложненный подход был предложен в работе [17]. В ней делалась попытка итеративной аппроксимации отсутствующих фрагментов на изображении посредством адаптации уже имеющихся. Соответственно, для заполнения выбирался не только наиболее подходящий фрагмент, но также определялась его наилучшая ориентация и разрешение.

Стоит отметить, что данный метод требует достаточно больших вычислительных затрат именно в силу необходимости уточнения дополнительных параметров для фрагментов заливки. Но существуют близкие к только что описанному методы, которые частично решают проблему быстродействия. К таким методам, например, можно отнести алгоритм, описанный в [18]. Этот алгоритм также использует методы синтеза текстур, что улучшает качество результатов и уменьшает время работы.

Ко второму подклассу алгоритмов поиска по экземпляру можно отнести алгоритмы, которые осуществляют поиск наиболее подходящих текстур и структур не на самом исходном изображении, а среди совершенно других, возможно никак не относящихся к данному, изображений.

До текущего момента при анализе задачи реконструкции изображений в невидимой области мы опирались на предположение, что вся необходимая для восстановления отсутствующих фрагментов информация может быть извлечена из доступной наблюдателю части этого же исходного изображения.

Несмотря на то, что такой подход имеет свои преимущества, в работе [19] утверждается, что, используя миллионы изображений в качестве базы данных, существует возможность получить намного более естественный и семантически богатый результат реконструкции, нежели при использовании одного единственного изображения.

Такой взгляд на задачу реконструкции изображений в невидимой области дал толчок развитию новых идей и алгоритмов, которые, в свою

очередь, показывают очень неплохие результаты для различных классов изображений. В одном из таких методов для каждого изображения из базы строится дескриптор, использующий информацию о пространственном спектре данного изображения и инвариантный к ряду преобразований, в частности, поворота [20]. Наилучшее семантическое соответствие результата достигается путем поиска по всей базе данных, состоящей из миллионов изображений.

После того как соответствующий фрагмент будет найден, происходит бесшовное вписывание найденного фрагмента в исходное изображение при помощи преобразования Пуассона. Результат применения алгоритма к пейзажному изображению с большой отсутствующей областью представлен на рисунке 11.



Рисунок 11 – Результаты метода реконструкции изображений в невидимой области, взятые из работы [19]. Слева направо: исходное изображение; изображение вырезанной пользователем областью; ближайшие семантические совпадения, полученные при помощи дескриптора из [20]; результат наложения с использованием преобразования Пуассона

### **1.5.5. Гибридные методы реконструкции**

Гибридные методы воплощают в себе подходы на основе синтеза текстур и диффузных методов реконструкции. Основной идеей таких методов является разделение изображения на непосредственно текстурную и структурную части. Соответствующие части заполняются последовательно при помощи алгоритмов основанных на использовании изофот и алгоритмов синтеза текстур [21-23]. Как правило, такие алгоритмы обладают серьезной



вычислительной сложностью за исключением необходимости реконструкции областей небольшого размера.

Стоит отметить одно из важных направлений структурной реконструкции изображений, которое использует алгоритмы сегментации. Пример такого подхода описан в работе [24]. Алгоритм состоит из двух этапов: достраивание всех необходимых структурных элементов и покрытие полученных областей синтетической текстурой.

На этапе дорисовки структурных элементов выполняется сегментация изображения на основе алгоритма [25], затем все незамкнутые границы экстраполируются для достижения полноценной сегментации исходного изображения при помощи тензорного голосования [26].

На втором этапе происходит генерация текстурной и цветовой информации для каждого полученного сегмента на основе той же процедуры голосования, описанной в [26].

#### **1.5.6. Полуавтоматические методы и методы быстрой реконструкции**

Полуавтоматические методы реконструкции являются одними из наиболее многообещающих с точки зрения показываемых результатов, несмотря на требование активного участия пользователя в процессе их работы. В основном, участие пользователя сводится к помощи в структурной реконструкции границ: пользователь должен наметить их приблизительное расположение.

Метод, предложенный в работе [27], требует от пользователя вручную указать важную отсутствующую информацию внутри области реконструкции, начертив границы объектов, соединяющие наблюдаемые участки изображения. После этого алгоритм, сходный экзemplарному, используется для синтеза окружающей текстуры. Отсутствующие фрагменты изображения синтезируются вдоль указанных пользователем границ на основе решения задачи глобальной оптимизации с различными граничными условиями однородности, согласованности и др.



Также для нахождения оптимума может использоваться и простая техника динамического программирования, но только лишь в случае, когда кривая, описывающая границу, единственна. Для случая многих границ задача оптимизации существенно усложняется, и решение ищется при помощи алгоритма распространения доверия. Процесс реконструкции наглядно показан на рисунке 12.

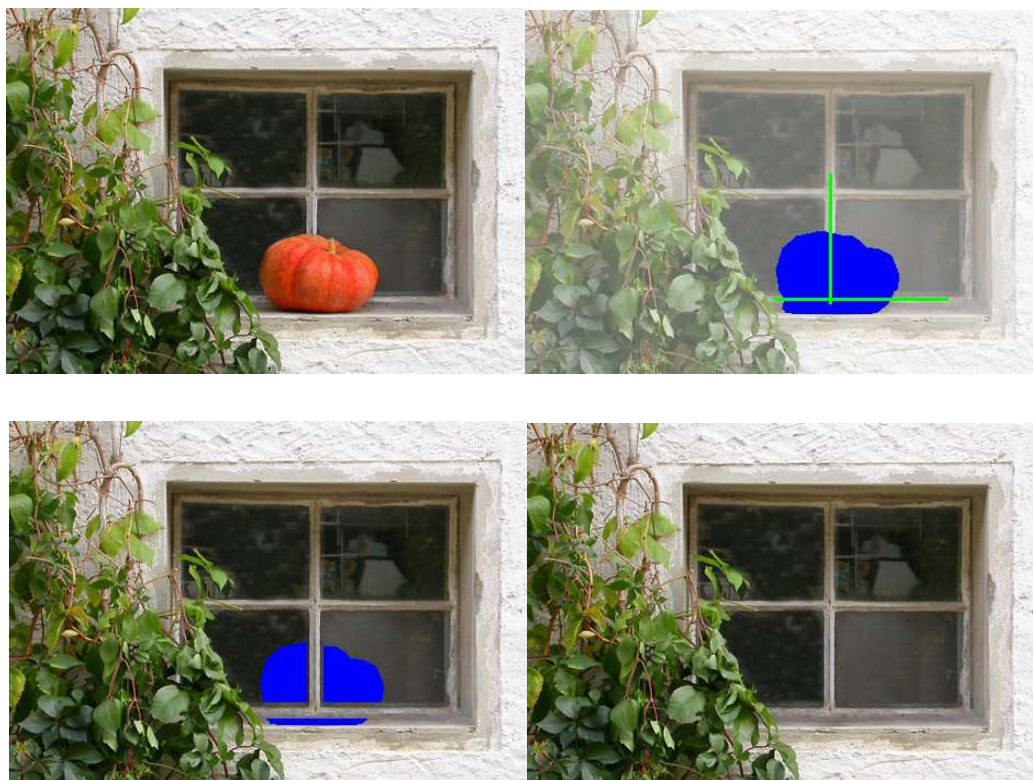


Рисунок 12 – Этапы и результаты работы алгоритма, представленного в работе [27]. В левом верхнем углу – исходное изображение. В правом верхнем – результат указания пользователем области реконструкции, а также предполагаемых границ объектов внутри этой области. В левом нижнем углу – результат реконструкции границ. В правом нижнем – конечный результат реконструкции с учетом синтеза текстур в отсутствующих областях

В зависимости от размеров области реконструкции время работы всех перечисленных нами методов может варьироваться от нескольких минут до нескольких часов, что зачастую делает их непригодными для использования в интерактивных приложениях. Для ускорения работы методов реконструкции был разработан ряд техник, получивших название быстрых

алгоритмов реконструкции. В частности, для ускорения алгоритмов, использующих уравнения диффузии, было предложено многократно производить операцию свертки между областью, подлежащей реконструкции, и ядром диффузии [28].

Также существуют техники, которые используют быстрый маршевый метод (от англ. Fast Marching Method, сокращенно – FMM) для распространения информации внутри области реконструкции [29]. Отметим, что такие ускоренные методы не подходят для реконструкции больших областей на изображениях, так как уступают в точности определения границы классическим методам, при этом обладая еще большей тенденцией к порождению артефактов с эффектом размытия.

#### **1.5.7. Визуальное восприятие и оценка качества результатов реконструкции**

Поскольку большинство работ в области реконструкции изображений в невидимой области сосредоточены на разработке новых алгоритмов для каких-либо специфических приложений, достаточно редко можно встретить работы, посвященные проблемам оценки того, что можно назвать визуальным качеством результатов реконструкции. Если совсем упростить формулировку, то мало кто из исследователей задается вопросом: что же с формальной точки зрения можно считать визуально хорошей или приемлемой реконструкцией? Большинство методов реконструкции, которые мы обсудили выше, пытались описать свою эффективность как способность восстанавливать большие участки изображений, границы большой кривизны, сложные текстурные фрагменты; также особое внимание уделялось времени работы того или иного алгоритма [1]. Многие методы, вообще, пытались в качестве результата получить изображение как можно более похожее на опорный оригинал. Естественно, что такие методы, равно как и многие другие, используют в процессе своей работы лишь информацию, доступную из видимой части изображения. Хотя для некоторых приложений такая оценка качества работы алгоритмов реконструкции, пусть и в ограниченном

количестве случаев, но все же может являться приемлемой. Невзирая на изначально субъективную природу процесса реконструкции, во многих работах качество реконструкции измерялось при помощи простых объективных соотношений типа сигнал-шум между реконструированным изображением и оригинальным или, просто, делалась визуальная оценка на основе сравнения этих двух изображений. В некоторых работах было показано, что такие объективные метрики как среднеквадратичное отклонение и отношение сигнал-шум не могут адекватно характеризовать визуальное восприятие результатов реконструкции [30].

Работа [31] является, наверно, одной из первых работ, в которых указывается на взаимосвязь ошибки, возникающей в результате реконструкции, и формой области на изображении, подлежащей восстановлению. В данной работе было показано, что качество реконструкции в большей степени зависит от формы невидимой области, чем от ее размера или количества таких областей. Иными словами форма, а не площадь области реконструкции является более важным фактором при визуальной оценке результатов восстановления информации на изображении. По сути, в работе было показано, что алгоритмы, использующие подход на базе дифференциальных уравнений в частных производных, эффективны при восстановлении малых или узких областей на изображении, так как в основе их лежит техника сглаживания. Тем не менее, в данной работе не делалось попыток ответить на главный вопрос: что же отличает от других алгоритм по-настоящему качественной реконструкции? В качестве примера можно привести изображения, представленные на рисунке 13. Человек-наблюдатель с легкостью заметит, что результат реконструкции в левом нижнем углу рисунка 13 визуально выглядит гораздо лучше, чем результат реконструкции в правом нижнем углу того же рисунка. Скорее всего, зрительная система человека сделала вывод о качестве реконструкции на основе иерархической последовательности низко-, средне- и высокоуровневых когнитивных процессов. Поэтому большим научным достижением будет выявление тех

факторов, которые влияют на визуальное восприятие результатов реконструкции, и которые помогут в конечном итоге разработать объективную методику оценки качества с точки зрения визуального восприятия.



Рисунок 13 – Зависимость качества реконструкции от формы области реконструкции. На верхней строке показаны два одинаковых исходных изображения с разными областями, подлежащими реконструкции. На нижней строке – соответствующие результаты реконструкции. Можно заметить, что изображение слева имеет более высокое качество, чем изображение справа. Изображения взяты из [31]

Как уже отмечалось ранее, область интересов специалистов, решающих задачу реконструкции, также включает в себя автоматическое редактирование различных мультимедийных объектов, что порой подразумевает реконструкцию достаточно больших областей на

изображениях. Для таких целей были разработаны методы, которые используют информацию о многих и многих изображениях из базы для подбора оптимального заполнения в противовес методам, ограничивающимся рассмотрением лишь неполной информацией, доступной на одном единственном изображении. Соответственно, для таких методов критерии качества, основанные на сравнении с опорным изображением, уже никак не смогут работать. Результаты работы таких методов должны быть просто незаметны для человеческого глаза, и в этом их главная цель. Такой подход неминуемо влечет за собой необходимость учета свойств и субъективной природы зрительной системы человека. В работе [19] был проведен эксперимент, в котором люди должны были оценивать качество результатов реконструкции. На основе рейтинга, полученного в результате эксперимента, исследователи делали выводы о сравнительной эффективности своего метода по отношению к другим. Для количественной оценки качества реконструкции в эксперименте участвовали 20 сторонних наблюдателей, которых просили найти поддельные изображения, разбросанные среди множества неизмененных изображений, при этом среди «поддельных» изображений были изображения, реконструированные как методом [19] так и методом [2]. В результате десятисекундного ознакомления с каждым изображением 34% изображений модифицированных алгоритмом [19], 69% изображений модифицированных алгоритмом [2] и 3% реальных фотографий были забракованы в качестве «подделок» участниками эксперимента. На основе этих результатов авторы в [19] утверждают, что их алгоритм лучше по сравнению с алгоритмом, описанным в [2]. Несмотря на довольно интересные результаты, данная работа все же не отвечает основной вопрос, возникающий при оценке качества результатов реконструкции.

В заключение, опишем результаты работы [32], в которой делается попытка соотнесения карт зрительного внимания, построенных при помощи численной модели зрительной системы, и визуального восприятия качества

реконструкции изображений. В процессе анализа небольших областей реконструкции, исследователи обнаружили некоторое соответствие между субъективными предпочтениями наблюдателей во время оценки качества реконструкции и картами зрительного внимания, построенными на базе математической модели. Правда, для подтверждения наличия такого соответствия требуется проведение дополнительных исследований, так как в работе [32] эксперимент проводился на ограниченной выборке изображений (4 штуки) и с ограниченным количеством людей (5 человек). Помимо этого исследователям еще только предстоит понять реальную взаимосвязь между подобными моделями и количественными параметрами для оценки качества реконструкции изображений в невидимой области.

## **Выводы по главе 1**

Анализ предметной области показывает, что можно выделить методы восстановления изображений в утерянных и поврежденных областях по типу привлекаемого представления изображений, в частности низкоуровневых (в виде дифференцируемых функций), текстурных и структурных представлений. Соответствующие методы оказываются неспособными реконструировать те аспекты изображения, которые не описываются в рамках выбранного представления (например, текстурные методы не воспроизводят структуру изображения и наоборот).

Методы, использующие поиск по экземпляру, решают эту проблему за счет переноса всей утерянной информации из изображения, сходного с реконструируемым. Однако применение этого подхода ограничено необходимостью априорного наличия изображений той же сцены, которая присутствует на реконструируемом изображении.

В гибридных методах осуществляется попытка учесть как текстурные, так и структурные особенности изображений, однако существенной трудностью для этих (как и для других) методов является отсутствие адекватных объективных критериев качества реконструкции, которые бы могли заменить субъективную оценку человеком.

Таким образом, актуальной задачей является выработка объективного критерия оценки качества реконструкции в утерянных или поврежденных областях изображений и разработка на его основе методов реконструкции на основе различных представлений изображений, в том числе обучаемым.

## **Глава 2. Критерий алгоритмической вероятности и разработка методов на его основе для решения задачи реконструкции изображений в невидимой области**

### **2.1. Введение**

Как уже было показано в предыдущей главе, одной из основных трудностей сопровождающих процесс решения задачи реконструкции изображений в утерянных областях является установление критериев качества полученных результатов. Существующие методы заполнения отсутствующих на изображении фрагментов в процессе своей работы используют достаточно мощные эвристики, применение которых, как правило, мотивируется произвольно выбираемым представлением изображений, подвергающихся обработке. Например, методы, применяющие аппарат дифференциальных уравнений в частных производных, опираются на функциональное представление изображений, методы структурного и текстурного заполнения, соответственно, используют структурное и текстурно-статистическое представление изображений.

В данной работе предлагается найти такой объективный критерий оценки качества результатов врисовывания отсутствующих фрагментов на изображении, который, будучи формализованным, мог бы позволить корректно предсказывать содержимое этих фрагментов и, следовательно, послужить общим методологическим принципом для разработки методов реконструкции изображений в утерянных областях. Слово «предсказывать» здесь было использовано не случайно, поскольку далее мы попробуем описать задачу реконструкции изображений в утерянных областях как частный случай более сложной научной проблемы универсального предсказания.

Мощными инструментами для формального описания задач предсказания являются алгоритмическая теория информации и близко прилегающая к ней теория универсальной индукции. Ключевым понятием,



связывающим обе этих теории, является алгоритмическая или колмогоровская сложность объекта, подлежащего описанию. В алгоритмической теории информации с данным понятием ассоциируется количество информации, которая содержится в описывающем объект сообщении, к примеру, передаваемом по каналу связи между источником и приемником и, как правило, состоящем из нулей и единиц. Таким образом, основными объектами, с которыми работают в рамках данной теории, являются битовые строки, индивидуальную сложность которых и приходится оценивать. Соответственно, задача предсказания при такой постановке вопроса сводится к экстраполяции этих самых битовых строк. Но, тем не менее, такое представление никак не ограничивает тот круг индуктивных задач, который может быть формализован при помощи данного подхода, помимо очевидного применения к продлению временных последовательностей данных. Иными словами любая практически значимая задача обобщения, предсказания или, в общем случае, индукции может быть описана в терминах поиска наилучшего с некоторой точки зрения продолжения уже имеющейся битовой строки на произвольную длину. Это можно наглядно продемонстрировать на примере формализации подобным образом задач классификации и регрессии.

В дисциплине машинного обучения под задачей классификации понимается процесс отнесения некоторого объекта  $x$  к соответствующему ему классу на основании признакового описания данного объекта, а также имеющихся обучающих примеров. Как правило, такие примеры представляются в виде пар объект-класс  $(x_i, c_i)$ . Целью же является правильная классификация некоторого нового образа  $x_{n+1}$  за счет нахождения соответствующей метки класса  $c_{n+1}$ . Очевидно, такая задача может быть представлена в виде последовательности с признаковым описанием объекта  $x_{n+1}$  в конце. Таким образом задача классификации превращается в процесс ответа на вопрос о том, какое же число должно быть следующим в последовательности  $x_1 c_1 x_2 c_2 \dots x_n c_n x_{n+1}$ ?

В свою очередь задача регрессии подразумевает поиск такой функции, которая ответственна за порождение результатов наблюдения некоторой зависимой величины. Как правило, в данной задаче также учитывается влияние шумов и погрешностей наблюдения на полученные данные. Обучающие примеры обычно представляются в виде пар независимая переменная - зависимая переменная:  $\{(x_1, f(x_1)) (x_2, f(x_2)) \dots (x_n, f(x_n))\}$ . В дисциплине машинного обучения решение данной задачи зачастую сводится к построению оценок, которые подбирают функцию, наилучшим образом аппроксимирующую имеющиеся данные. Альтернативным подходом может стать попытка описать данную задачу в терминах предсказания последовательностей (в данном случае совершенно не принципиально битовые это строки или числа как таковые, поскольку это повысит наглядность, но не уменьшит общность). Это можно сделать, записав входные данные ряд друг за другом и добавив новое значение независимой переменной, для которой мы хотим предсказать значение функции, в конец полученной строки. При этом наша задача аналогично предыдущей может быть переформулирована как вопрос: каков же следующий элемент в последовательности  $x_1, f(x_1), x_2, f(x_2), \dots x_n, f(x_n), x_{n+1}$ ? Отметим, что данный подход не дает искомой функции в явном виде, но по существу является таковой, так как значение искомой функции  $f(x)$  для любого произвольного  $x$  может быть получено из предположения  $x_{n+1} = x$ .

Несмотря на свою теоретическую мощь, непосредственная практическая применимость алгоритмической теории информации и базирующейся на ней алгоритмической теории вероятности кажется весьма проблематичной в силу принципиальной невычислимости основных величин, формирующих данные теории. Однако данную проблему частично снимают разрабатываемые вычислимые аппроксимации для этих величин, что, как правило, ведет к сужению области применимости решения.

Учитывая все выше сказанное, кажется интуитивно приемлемым попытаться описать в том числе и задачу реконструкции изображений в

утерянных областях как процесс предсказания содержимого внутри отсутствующего на изображении фрагмента с привлечением алгоритмической теории вероятности, тем более, что попытки разработки методик практического использования идей универсального предсказания для анализа изображений если и есть, то крайне редки. Такой дефицит работ в области разработки вычислимых аппроксимаций алгоритмической вероятности, скорее всего, связан с неспособностью этих аппроксимаций к решению практически значимых задач, включая задачи анализа изображений, являющихся, по сути, очень длинными строками данных. Хотя, стоит отметить, что теория универсальной индукции, базирующаяся на понятии колмогоровской сложности, дала обоснование принципу минимальной длины описания (МДО) [33-35], который нашел свое применение во многих задачах анализа изображений и компьютерного зрения.

Принцип МДО утверждает, что процесс индуктивного вывода может быть сведен к поиску наилучшего сжатия исходных данных, таким образом результатом предсказания в соответствии с этим принципом будут данные, которые позволят максимизировать степень компрессии (сжатия) или, говоря иначе, предельно уменьшить суммарный размер уже имеющихся и, собственно, предсказываемых данных [34]. Для принципа МДО вычислимость достигается за счет сужения алгоритмически полного пространства моделей источников информации до некоторого пространства моделей, представляющих алгоритмы сжатия, которые, например, базируются на вероятностных схемах кодирования по Шеннону-Фано [35]. Стоит отметить, что в основе данного принципа лежит исторически ранее разработанный подход, который именовался как принцип минимальной длины сообщения (МДС) [36]. Суть этого подхода состоит в том, что размер сообщения, подлежащего передаче, рассматривается как сумма размеров модели, по которой осуществляется сжатие, и данных, сжатых при помощи данной модели [37].

Несмотря на то, что принцип МДО уже нашел свое применение в области анализа изображений, в частности, при решении задач сегментации изображений [38-40], построения структурных элементов [41] и описания формы границ областей изображений [42-44], распознавания объектов на изображении [45] и распознавания рукописных символов [46, 47], оценивания параметров пространственного преобразования между парой изображений [48, 49], оценивания поля движения по видеоизображениям [50-52], а также многих других задач [53-57], в данной работе будет сделана попытка показать, что теория универсального предсказания сама по себе имеет свою методологическую значимость при решении практических задач не универсальными методами.

Исходя из уже написанного, можно предположить, что такая аппроксимация как принцип МДО напрямую вряд ли применима для решения задачи реконструкции изображений в невидимой области. Возможно, но в то же время мы беремся показать, что понятие алгоритмической вероятности является для этой задачи весьма естественным. Начнем с формальной постановки задачи предсказания на основе алгоритмической вероятности.

## **2.2. Предсказание на основе алгоритмической вероятности**

В алгоритмической теории информации А.Н. Колмогорова строго формализуются понятия простоты и сложности [58]. Кажется интуитивно очевидным, что строка (символов, бит или чисел) является простой, если она может быть описана при помощи всего лишь нескольких слов, например «строка из тысячи нулей», и, наоборот, является сложной, если такое емкое описание отсутствует, и приходится поэлементно перечислять содержимое такой строки. Соответственно этому индивидуальная сложность  $K(\alpha)$  некоторой бинарной строки  $\alpha$ , или количество содержащейся в ней информации, определяется как длина  $l(\mu)$  наикратчайшей программы-описания  $\mu$ , порождающей данную строку при выполнении этой программы

$\mu$  на некоторой опорной машине  $U$ :  $U(\mu)=\alpha$ . Формально это можно записать следующим образом:

$$K(\alpha) = \min_{\mu} \{l(\mu) : U(\mu) = \alpha\} . \quad (5)$$

В рамках проблематики индуктивного вывода данная программа трактуется как наилучшая модель источника строки. Если вспомнить уравнение, связывающее количество информации  $I$  и вероятность  $P$ :  $P=2^{-I}$ , то кажется естественным приписать строке  $\alpha$  вероятность  $P(\mu)=2^{-K(\alpha)}$ , однако более обоснованным является рассмотрение  $2^{-l(\mu)}$  в качестве априорной вероятности программы  $\mu$ , а с учетом того, что строка  $\alpha$  может быть порождена разными программами, ее (алгоритмическую) вероятность  $P_U$  следует рассчитывать в соответствии с выражением:

$$P_U(\alpha) = \sum_{\mu: U(\mu)=\alpha} 2^{-l(\mu)} , \quad (6)$$

где индекс  $U$  указывает на зависимость этого распределения от опорной машины. Заметим, что существует бесчисленное количество строк, которые будучи поданными на вход машине  $U$  дадут на выходе строку  $\alpha$  (возможно сгенерировав после нее еще что-нибудь). При этом большинство таких входных «программ» будут обладать свойством избыточности в том смысле, что, удалив из конца такой программы один или более бит, это никак не повлияет на генерацию строки  $\alpha$ . Соответственно, в уравнении для алгоритмической вероятности  $P_U$  под суммированием по  $\mu$ :  $U(\mu)=\alpha$  мы будем понимать только те входные строки-программы, которые не будут обладать свойством избыточности, то есть те программы  $\mu$ , которые при удалении из их конца хотя бы одного бита, фактически, перестают быть программами для  $\alpha$ .

Данное определение вероятности появления строки обладает несколькими интересными свойствами:

1. В соответствии с таким определением наибольшую вероятность будут иметь строки с наименьшим программным описанием.

2. Данная величина носит объективный характер и в некоторой степени не зависит от выбора конкретной универсальной машины, используемой для декодирования, поскольку код для одной универсальной машины может быть выполнен на другой универсальной машине с добавлением конечной последовательности бит (или символов) необходимых для эмуляции исходной машины.

3. Алгоритмическая вероятность обладает свойством полноты. Это означает принципиальную возможность найти любую описываемую регулярность в исходных данных [59].

Данное определение может быть использовано в качестве основы теории универсального предсказания [60], для чего необходимо ввести следующее определение условной вероятности того, что в качестве продолжения строки  $\alpha$  будет наблюдаться строка  $\beta$ :

$$P_U(\beta | \alpha) = P_U(\alpha\beta) / P_U(\alpha), \quad (7)$$

где  $\alpha\beta$  – конкатенация двух строк.

Стоит отметить, что при использовании алгоритмической вероятности в контексте предсказания равенство  $U(\mu) = \alpha$  следует заменить проверкой того, что программа  $\mu$  порождает некоторую строку, для которой  $\alpha$  является префиксом (то есть программа печатает  $\alpha$  с произвольным продолжением). Такая интерпретация, очевидно, обеспечит выполнение неравенства  $P_U(\alpha\beta) \leq P_U(\alpha)$ , и вероятность конкатенации  $P_U(\alpha\beta)$  будет обладать свойствами классической совместной вероятности, а вероятность продолжения  $P_U(\beta | \alpha)$  – свойствами условной вероятности [61, 62].

Важным следствием, проистекающим из свойства полноты, для алгоритмической вероятности является ее принципиальная невычислимость, а значит и неприменимость на практике. Тем не менее, и колмогоровская сложность является невычислимой (в силу алгоритмической неразрешимости

проблемы останова), но методологические следствия из нее как критерия качества модели имеют не только важное теоретическое значение, но и, как отмечалось выше, большую практическую пользу. Такую же пользу можно извлечь и из аналогичного использования алгоритмической вероятности.

### 2.3. Практическая алгоритмическая вероятность

Как уже отмечалось ранее, применение алгоритмической сложности на практике достигается за счет сужения пространства моделей и использования эвристических схем кодирования, что более строго можно рассматривать как замену универсальной опорной машины  $U$  на неуниверсальную  $S$  (не полную по Тьюрингу). Более того, можно ввести дополнительный уровень индуктивного вывода и рассматривать эмуляцию неуниверсальной машины  $S$ , играющей роль представления информации [67], на универсальной машине  $U$ , тогда можно разделить две задачи:

1. задачу поиска оптимальной модели  $\mu$  данных  $\alpha$  в рамках представления  $S$ :

$$\mu^* = \arg \min_{\mu: U(S\mu)=S(\mu)=\alpha} l(\mu)$$

2. и задачу поиска оптимального представления  $S$  для некоторой выборки строк  $\alpha_1 \dots \alpha_n$ :

$$S^* = \arg \min_S \sum_{i=1}^n \min_{S(\mu)=\alpha_i} l(\mu).$$

Большой интерес представляют методы автоматической оптимизации представлений (что пока выполняется только для узких классов представлений, зачастую не связанных с обработкой изображений), а также автоматические методы построения эффективных алгоритмов поиска моделей в рамках заданных представлений (так называемые методы метавычислений). Однако на данный момент такие автоматические методы метаиндукции все еще далеки от возможности практического применения, поэтому, как правило, этот процесс метаиндукции неявно выполняется разработчиками соответствующих прикладных методов. В данной работе мы

не будем касаться вопроса автоматического синтеза оптимальных специализированных методов, а будем выполнять подобную специализацию вручную для нескольких наперед заданных представлений при решении задачи реконструкции изображений в невидимой области.

## 2.4. Восстановление изображения в утерянной области как частный случай задачи предсказания

Попытаемся сформулировать критерий алгоритмической вероятности для рассматриваемой нами задачи реконструкции изображений в невидимой области.

Пусть дано изображение  $f(x, y)$ , которое должно быть реконструировано по некоторой области  $R$  на основе яркостей в остальной части изображения  $R'$ . Будем обозначать изображения в соответствующих областях через  $f|_R$  и  $f|_{R'}$ . Через  $f|_{R'} g|_R$  будет обозначать «конкатенацию» двух изображений:

$$(f|_{R'} g|_R)(x, y) = \begin{cases} f(x, y), (x, y) \notin R, \\ g(x, y), (x, y) \in R. \end{cases} \quad (8)$$

Тогда задача определения наилучшего заполнения изображения в области  $R$  может быть сведена к задаче оптимального предсказания:

$$g^*|_R = \arg \max_g P_U(g|_R | f|_{R'}) = \arg \max_g P_U(f|_{R'} g|_R) \quad (9)$$

В силу предсказательной универсальности алгоритмической вероятности данную формулу (9) можно рассматривать как универсальный критерий для оценки качества результатов реконструкции изображения в невидимой области. Но вследствие своей универсальности он оказывается невычислимым, что, по сути, не так уж плохо. Кто-то может задаться вопросом: «Какой толк от критерия качества, который невозможно численно оценить?» Ответ кроется в сущности самой задачи: во введении было показано, что имеющиеся на текущий момент методы в существенной мере опираются на специфику своих подходов к постановке исходной задачи как таковой. Иными словами, мы можем попытаться использовать лишь



информацию вокруг отсутствующей области, а можем привлечь весь свой жизненный опыт для реконструкции изображения. Таким образом, все будет зависеть от тех упрощений или аппроксимаций, а также априорного знания (с которым можно эффективно манипулировать в рамках теорий универсального предсказания и индуктивного вывода), которое привлекается для решения рассматриваемой задачи.

Говоря об аппроксимациях, то, как и отмечалось ранее, принцип минимальной длины описания здесь напрямую не применим, но причина этого заключается не в том, что в индукции строится одна лучшая модель, а в том, что в ней модель строится только по имеющимся данным, и эта модель может не осуществлять никакого предсказания. Предсказание также можно пробовать выполнять на основе одной лучшей модели, которая порождала бы  $f|_{R'} g|_R$  с некоторым случайным  $g$ , но при этом нужно будет найти такое заполнение области  $R$ , при котором длина описания всего изображения в совокупности с реконструированной областью окажется минимальной. Однако учет различных моделей мог бы привести к более точной оценке неопределенности в гипотетическом содержимом области  $R$  и, как следствие, к более точным результатам реконструкции.

Может показаться вообще сомнительным, что заполнение изображения в утерянной области какой-то «содержательной» информацией может уменьшить длину описания всего изображения по сравнению, скажем, с его заполнением просто однородным фоном. Однако для детерминистских моделей изображений это действительно будет таковым, если мы не будем ограничиваться какими-то простыми способами сжатия изображений. Поясним это на простом примере. На рисунке 14 представлено исходное изображение, составленное из наложенных друг на друга прямоугольников разной формы, и это же изображение с удаленной областью.



Рисунок 14 – Исходное изображение (слева) и изображение с вырезанной областью (справа)

Очевидно, самое простое (то есть самое вероятное) описание данного изображения должно состоять просто из указания координат углов и яркостей накладывающихся прямоугольников, например, отсортированных по «дальности». Рассмотрим изображение с вырезанной областью. Действительное содержание этой области не вполне тривиально и включает пять подобластей с разной яркостью. Однако если оставить вырезанную область заполненной пикселями постоянной (например, нулевой) яркости, то полученное изображение будет иметь более сложное описание. Если же заполнить ее пикселями разной яркости так, чтобы полное изображение представлялось совокупностью наложенных прямоугольников, длина описания полного изображения сократится. В то же время, если мы попытаемся распространить подобный подход на более сложные модели, то есть шанс столкнуться с определенными трудностями.

И действительно, представим некоторое зашумленное изображение с вырезанным на нем фрагментом (рисунок 15 слева). Очевидно, что исходное изображение (рисунок 15 в центре) будет иметь намного меньшую алгоритмическую вероятность, чем изображение на рисунке 15 справа.

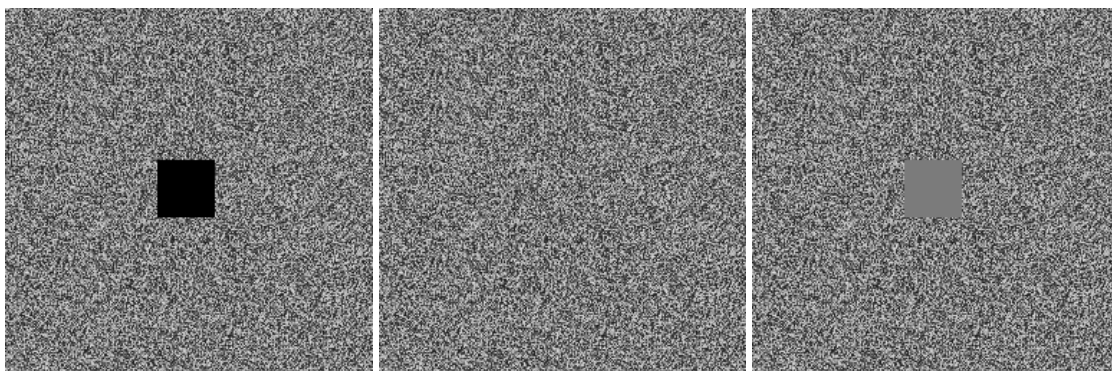


Рисунок 15 – Реконструкция зашумленного изображения. Слева на право: зашумленное изображение с вырезанным на нем фрагментом, исходное изображение, реконструкция изображения в отсутствующей области при помощи ее заполнения однородным фоном

С одной стороны такой результат может быть и не так уж плох с точки зрения реконструкции, особенно если вспомнить о том, что истинно случайный процесс не может быть предсказан, но с другой стороны на это можно посмотреть как ошибку реконструкции в силу того, что бесшумная реконструкция не будет корректно описывать изображение как набор случайных яркостей пикселей, не говоря уже о произвольности в структурном и семантическом содержании изображения. Можно даже подумать, что задача реконструкции не совсем соответствует задаче предсказания (особенно, если считать что изображение на рисунке 15 в центре лучше изображения на рисунке 15 справа). Тем не менее, если перенести подход на основе универсального предсказания с детерминистских моделей на модели стохастические, то окажется возможным индуктивно «вывести» стохастический процесс, порождающий содержимое изображения представленного на рисунке 15 в центре, как наиболее компактную модель. Правда, в таком случае во время расчета  $P_U(f|_R, g|_R)$  пришлось бы игнорировать гипотетическое случайное содержимое  $g$  в области  $R$ . На практике же стохастические модели зачастую используются вместе с принципом МДО [68-70]. В таком случае оказывается удобным в явном виде разделять регулярную и случайную части, чтобы далее работать с ними по отдельности, применяя соответствующую модель.

Для практического применения критерия (9), как отмечалось выше, мы должны сузить пространство моделей, заменив универсальную машину некоторым ограниченным представлением изображений [67, 71]. Естественно, чем богаче представление изображений, чем больше типов регулярностей (текстурных, структурных, семантических) оно использует, тем более качественным может быть восстановление. В частности, для реализации реконструкции из примера, приведенного на рисунке 14, необходимо использование структурных представлений. И, как можно увидеть из предыдущей главы, в практике реконструкции изображений большое внимание уделяют статистическим (текстурным) и структурным представлениям [63, 64], а также возможности их совместного использования [65, 66]. Однако сначала для демонстрации перспективности подхода на основе алгоритмической вероятности мы воспользуемся простейшим функциональным представлением. Требование, которое следует предъявить к такому представлению, заключается в том, чтобы оно было распределенным, а не локальным, то есть каждый элемент описания изображения зависел бы от яркостей различных пикселей. Этому требованию вполне удовлетворяет представление изображения в форме пространственного спектра.

## 2.5. Метод реконструкции изображений на базе алгоритмической вероятности с использованием спектрального представления

Рассмотрим дискретное преобразование Фурье от изображения  $f(x, y)$ :

$$F(u, v) = \sum_{y=0}^{N-1} \sum_{x=0}^{N-1} f(x, y) e^{-\frac{2\pi i}{N}(ux+vy)},$$

где  $N$  – линейный размер изображения (для упрощения записи будем рассматривать квадратные изображения). При неизвестном (произвольном) содержании изображения в области  $S$  это преобразование примет вид:

$$F(u, v, g|_R) = \sum_{y=0}^{N-1} \sum_{x=0}^{N-1} (f|_R, g|_R)(x, y) e^{-\frac{2\pi i}{N}(ux+vy)}.$$

Интересно, что попытка поиска заполнения  $g|_R$  с использованием критерия минимизации энергии спектра  $F(u, v, g|_R)$  не приводит к положительному результату, поскольку «наилучшим» по такому критерию оказывается заполнение  $g|_R(x, y)=0$ .

Рассмотрим сужение критерия (9) на данное представление. Генеративным представлением  $S$ , которое позволяет по описанию изображения  $F$  породить само изображение  $S(F)=f$ , очевидно, является обратное преобразование Фурье. То есть обратное преобразование Фурье является «машиной», для которой спектр изображения является «программой», и «исполнение» этой программы на выходе дает исходное изображение. Очевидно, такая машина является неполной по Тьюрингу, поэтому термин «программа» для входных данных такой машины весьма условен. Удобной особенностью этого представления является то, что каждому изображению в нем соответствует лишь одно описание, которое можно получить, используя прямое преобразование Фурье, которое мы обозначим через  $S^{-1}$ :  $S^{-1}(f)=F$ , поскольку оно является обратным к  $S$ . Перепишем общий критерий (3) для такого представления  $S$ :

$$\begin{aligned} g^*|_R &= \arg \max_g P_S(f|_{R'}, g|_R) = \arg \max_g \sum_{F: S(F)=f|_{R'}} 2^{-l(F)} = \\ &= \arg \max_g 2^{-l(S^{-1}(f|_{R'} g|_R))} = \arg \min_g l(S^{-1}(f|_{R'} g|_R)). \end{aligned} \quad (10)$$

Отметим, что данный результат верен для любой обратимой программы  $S$  (а не только программы, которая восстанавливает изображение по его спектру). При этом если обращение программы  $S$  является задачей класса  $P$ , то возможен синтез эффективного практического метода.

Таким образом, нам необходимо найти такое заполнение  $g|_R(x, y)$  области  $R$ , что сложность спектра минимальна. Критерий сложности спектра требует некоторого уточнения. Поскольку сам спектр является «моделью» источника изображения, к нему не должны применяться дополнительные

методы поиска закономерностей. Здесь мы ограничиваемся энтропийной оценкой для величины  $I(F)=H(F)$ :

$$H(F) = - \sum_y p(y | F(u, v) = y) \log_2 p(y | F(u, v) = y) \quad (11)$$

где суммирование ведется по всем возможным значениям (амплитудам и фазам) гармоник. Как будет показано в следующей главе уже для такого, сравнительно простого, представления можно получить достаточно хорошие результаты реконструкции изображений в невидимой области [72]. При этом такой метод предсказания окажется способным восстанавливать как структурную, так и текстурную информацию на отсутствующих фрагментах. Далее рассмотрим более продвинутый класс представлений, для которого, в том числе, возможен учет предыдущего опыта наблюдателя.

## **2.6. Учет предыдущего опыта наблюдателя для решения задачи реконструкции изображений в невидимой области**

Представим себе изображение человеческого лица и предположим, что некоторая часть этого изображения отсутствует, требуя реконструкции. Предположим, что эта часть закрывает от нас один глаз изображенного лица, в таком случае критерий на основе алгоритмической вероятности в чистом виде скорее всего даст адекватный результат реконструкции в силу симметрии лица (для того чтобы изображение обладало подобной регулярностью необходимо, чтобы оно включало лицо целиком). Но, допустим, что теперь на изображении отсутствует нос, тогда критерий на основе алгоритмической вероятности без дополнительных допущений вряд ли позволит получить желаемый результат. В то же самое время мы знаем, что люди способны делать самые общие предположения об отсутствующих частях на изображениях произвольных лиц (или в некоторых экстремальных случаях не обязательно на изображениях) и особенно точно могут восстанавливать знакомые им лица. Несомненно, что люди используют для этого свой предыдущий опыт. Для корректной формализации идеи учета предыдущего опыта наблюдателя нам необходимо поставить задачу

реконструкции изображений в невидимой области как процесс максимизации условной алгоритмической вероятности при некоторой заданной истории. Можно записать это следующим образом:

$$P_U(g|_R | f_1 \dots f_n f|_{\bar{R}}),$$

где  $f_1 \dots f_n$  представляют собой длинную последовательность ранее виденных наблюдателем изображений.

Однако, как и в случае с использованием лишь информации, содержащейся на видимых фрагментах изображения, такая постановка исходной задачи не допускает прямого решения, тем более что данный процесс усложняется из-за необходимости хранить и обрабатывать большие массивы изображений [73]. Для того чтобы появилась возможность решить исходную задачу в такой ее постановке нам, как и в случае спектральных представлений, следует воспользоваться некоторой опорной машиной:

$$P_S(g|_R | f|_{R'}) = P_U(g|_R | f_1 \dots f_n f|_{R'}) \text{ для любых } f|_{R'} \text{ и } g|_R.$$

Опять же такая опорная машина будет соответствовать некоторому представлению изображений, и мы можем обозначить те желаемые свойства, которыми должно обладать такое представление:

1. Поскольку в рамках данного представления мы хотели бы учитывать некий предыдущий опыт системы, то весьма естественным было бы реализовать это через механизмы обучения. Иными словами наша опорная машина должна быть способной к обучению представлениям.
2. Результат обучения такой опорной машины должен быть представлен в более компактном виде, нежели исходная последовательность изображений.
3. Такое представление должно быть обратимым, то есть между каждой парой «программа»-изображение должно быть взаимно однозначное соответствие:

$$\forall S : S(code) = image, \exists S^{-1} : S^{-1}(image) = code$$

Тогда мы сможем воспользоваться частным критерием (11) для нашего предполагаемого представления.

4. Такое представление должно обладать достаточно большой выразительной силой, то есть способностью учитывать различные регулярности и высокоуровневые взаимосвязи между элементами на изображении.

Всем вышеописанным критериям удовлетворяет обширный класс моделей, решающих задачу автоматического выбора признаков. Типичным примером таких моделей является метод анализа главных компонент (АГК). Известно, что при выборе количества главных компонент меньшим, чем количество независимых направлений в исходном пространстве, в результате работы метода будет получено более компактное, т. е. сжатое представление исходного образа. А учитывая тот факт, что сама модель представляется всего лишь набором коэффициентов, определяющих степень влияния какого-либо из входных признаков на выходную компоненту, то можно заметить, что данное свойство удовлетворяет второму критерию эффективности представления, описанному выше. Другими словами, применив данный метод, мы получим более компактное представление исходных данных при минимальных потерях информации, что обусловлено моделью ошибок данного метода, которая является лишь частным случаем критерия максимизации взаимной информации (так называемый принцип инфомакса используемый, в частности, для АГК). Существенным недостатком данного подхода является то, что все полученные в результате работы метода главные компоненты будут находиться в линейной зависимости от компонент векторов исходного пространства. Таким образом, можно предположить, что при сложных взаимосвязях между исходным признаковым описанием и искомым представлением, каковыми могут быть соотношения между яркостями пикселей на изображении и характеристиками, например, симметрии лица, данная модель АГК окажется неспособной к нахождению



столь нетривиальных взаимозависимостей [74]. Еще одним примером моделей, применяемых для решения сходных задач, могут служить модели, пришедшие из области нейросетевых вычислений, которые в некоторой литературе принято называть моделями автоассоциативной памяти, но на сегодняшний день они известны под более характерным названием – автоэнкодеры. Данный класс моделей, несмотря на свою схожесть с АГК, более интересен тем, что способен обучаться нелинейным взаимосвязям между входным и выходным представлением, которое иногда именуют кодом входного сигнала (отсюда и название: автоэнкодер – автоматический кодировщик). Данной особенностью автоэнкодер обязан нелинейной активационной функции, которая применяется к взвешенной сумме сигналов идущих с входа. Также автоэнкодер еще интересен и тем, что при наложении определенных ограничений на тип представлений, которые могут быть получены на его выходе, данная модель способна обучаться в том числе и параметрам функции распределения, в соответствии с которой распределены образы в исходном пространстве [75]. А эта особенность автоэнкодера указывает на пригодность применения данной модели для эффективного выделения признаков в соответствии с четвертым требованием, о котором шла речь выше. Помимо этого способность восстанавливать параметры функции распределения позволяет рассматривать автоэнкодер как особую генеративную модель, реализующую возможность сэмплирования (генерирования наблюдений случайной величины) синтетических примеров представителей классов, параметрам которых автоэнкодер сумел обучиться. Стоит отметить, что данный тип нейронных сетей является полносвязным. Это обстоятельство не позволяет автоэнкодеру, не вводя дополнительных ограничений, учитывать наличие взаимосвязи внутри локальных областей изображений. Рассмотрим подробнее архитектуру данной модели для автоматического выделения признаков.

## 2.7. Метод реконструкции изображений на базе алгоритмической вероятности с использованием систем, способных к обучению представлениям

Автоэнкодер (AE) представляет собой алгоритм обучения без учителя, представляемый в виде нейронной сети, который подбирает значения на своих выходах таким образом, чтобы они были как можно более близкими (в идеальном случае тождественными) к значениям на входах. Другими словами автоэнкодер пытается обучиться тождественному отображению между входным и выходным векторами:  $AE(x)=x$ . Такое тождественное отображение может оказаться крайне тривиальным, и процесс обучения как таковой может отсутствовать, если, конечно, не наложить некоторые ограничения на данную нейронную сеть. Такими ограничениями могут служить как лимитирование количества нейронов на скрытом слое, так и наложение критериев разреженности на активации нейронов на скрытом слое. Архитектура автоэнкодера представлена на рисунке 16а.

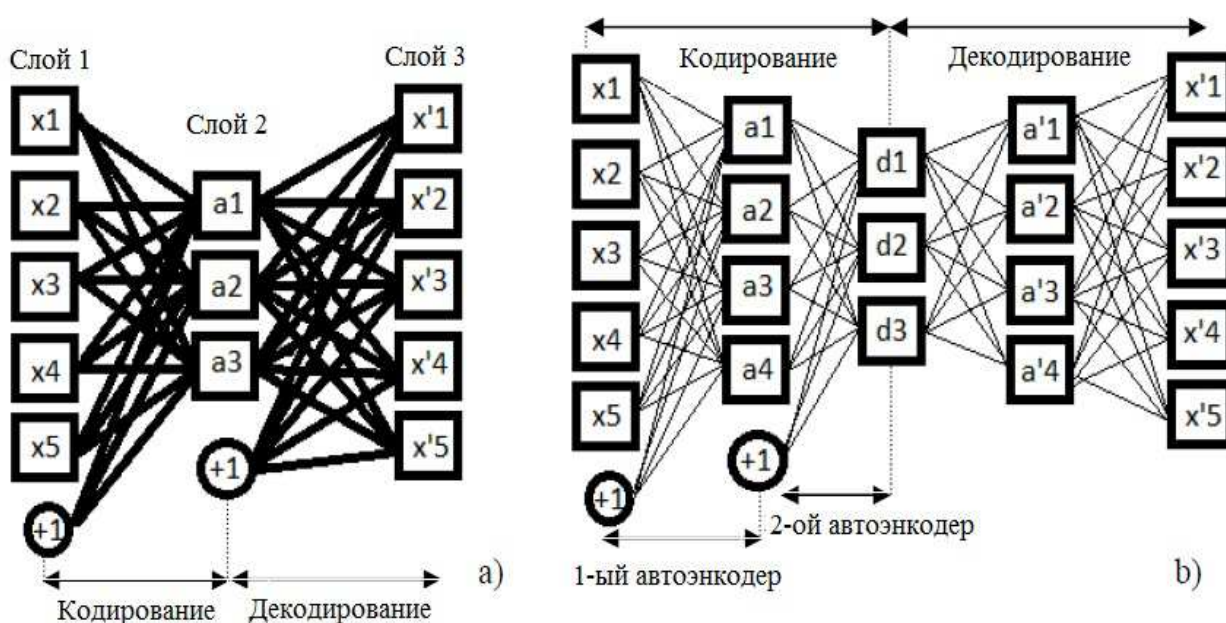


Рисунок 16 – Схемы архитектур автоэнкодера (а) и стекового автоэнкодера (b)

Отметим, что после обучения автоэнкодер может работать в двух режимах: кодирования – когда на вход подается вектор-образ некоторого объекта, а в результате получается его «код»; декодирования - когда на

скрытый слой подается «код» объекта, а в результате получается «сам» исходный объект.

Если рассматривать автоэнкодер в совокупности с накладываемыми ограничениями, то такая система будет удовлетворять всем критериям, предъявленным нами для опорных машин способных к обучению представлениям. Стоит также отметить, что в случаях, когда количество нейронов на скрытом слое меньше количества нейронов на входе и выходе, автоэнкодер, по существу, выполняет сжатие поступающей на вход информации. Если ко всему прочему мы учтем свойство обратимости данной модели, становится очевидной непосредственная корреляция между такой моделью и понятием алгоритмической сложности, хотя произвол в выборе количества нейронов на скрытом слое все же оставляет место для некоторой неоднозначности.

Можно предложить несколько более усложненный вариант архитектуры автоэнкодеров. Если каждый скрытый слой автоэнкодера рассматривать как гипотетический вход для другого автоэнкодера, то можно конструировать многослойные архитектуры обучающихся послойно стековых автоэнкодеров (*SAE*) [76]. Такая архитектура должна позволить стековому автоэнкодеру обучаться все более и более абстрактным наборам признаков по мере удаления от входного слоя, представляющего самый низкий – пиксельный уровень описания изображения [77]. На рисунке 16b представлена схема стекового автоэнкодера, составленного из двух простых, сжимающих автоэнкодеров. Рассмотрим применение подобной модели к задаче реконструкции изображения в невидимой области.

Предположим, что у нас имеется предобученный стековый автоэнкодер (*SAE*). Воспользуемся данной моделью автоматического выбора признаков и перепишем общий критерий (9) для этой опорной машины:

$$\begin{aligned} g^*|_R &= \arg \max_g P_{SAE^d} (f|_{R'}, g|_R) = \arg \max_g \sum_{code: SAE^d (code)=f|_{R'}, g|_R} 2^{-l(code)} = \\ &= \arg \max_g 2^{-l(SAE^c (f|_{R'}, g|_R))} = \arg \min_g l(SAE^c (f|_{R'}, g|_R)), \end{aligned} \quad (12)$$

где  $SAE^d$  - стековый автоэнкодер, работающий в режиме декодирования;  $SAE^c$  - стековый автоэнкодер, работающий в режиме кодирования;  $code$  – вектор выделенных стековым автоэнкодером признаков (активаций нейронов на скрытом слое). При этом  $f|_R, g|_R$  следует трактовать как развернутую в вектор функцию-изображение, заданную на дискретной двумерной сетке, и значения которой в узлах этой сетки могут быть найдены по формуле:  $i*w+j$ , где  $i$  и  $j$  – координаты узла сетки,  $w$  – ширина изображения.

Следуя выражению (12), нам необходимо найти такое заполнение  $g|_R(x, y)$  области  $R$ , чтобы сложность полученного кода была минимальной. Но тут сразу встает вопрос: если количество нейронов на скрытом слое заранее задано, то как измерить компактность такого «кода»? По аналогии со спектральным представлением можно попытаться оценить энтропию полученного кода, но такой способ, как мы увидим в дальнейшем, не будет обладать оптимальностью с точки зрения затрат вычислительных ресурсов, необходимых для поиска решения. В данном же случае было бы удобнее воспользоваться еще одним замечательным свойством рассматриваемых нами генеративных моделей. Это свойство заключается в том, что с помощью автоэнкодеров можно непосредственным образом осуществлять процедуру байесовского вывода. Таким образом, мы сможем построить конкретные алгоритмы поиска в виде аппроксимаций процедуры вывода апостериорных оценок содержимого области подлежащей реконструкции. Компактность кода при этом не будет требовать никаких дополнительных оценок в силу регуляризационных ограничений, накладываемых на процесс обучения автоэнкодера. Соответственно, после прохождения изображения через автоэнкодер, работающий в режиме кодирования, полученный код уже будет являться «оптимальным», останется лишь сгенерировать по данному коду изображение, пропустив этот код через автоэнкодер в режиме декодирования. При этом в процессе реконструкции на начальном этапе (перед первым проходом реконструируемого изображения через автоэнкодер в режиме кодирования) необходимо ввести некоторое априорное

предположение о содержимом восстанавливаемой области, после этого апостериорная оценка результата декодирования может стать новой, уточненной, оценкой априорного распределения вероятности интенсивности пикселей в области, подлежащей реконструкции.

Отметим, что в качестве опорных машин для критерия (12) могут также использоваться и более сложные модели, способные к обучению представлениям. Примером таких моделей могут служить так называемые сверточные автоэнкодеры, но принципиального влияния на сам критерий алгоритмической вероятности архитектурная сложность этих моделей оказывать не будет, оставив его без изменений. Модели сверточных автоэнкодеров и их применение для решения задачи реконструкции изображений в невидимой области, а также конкретные алгоритмы решения этой задачи для рассмотренных нами представлений более подробно будут обсуждаться в следующей главе.

## **Выводы по главе 2**

В данной главе было показано, что задача реконструкции изображений в невидимой области может быть поставлена как частный случай проблемы универсального предсказания. Это позволило ввести критерий качества ее решения на основе алгоритмической вероятности.

Для достижения практической применимости методов предсказания на основе алгоритмической вероятности была показана необходимость замены универсальной опорной машины неуниверсальными, которые могут быть поставлены в соответствие представлениям изображений, заданным в форме генеративных моделей. Приведено аналитическое обоснование использования и непосредственное применение представлений изображений для решения задачи реконструкции изображений в невидимой области.

Выведена проекция критерия алгоритмической вероятности на спектральное представление изображений, а также на обучаемые представления изображений на основе стека автоэнкодеров, что открывает

перспективы по разработке методов с указанными типами представлений на основе объективного критерия.

## **Глава 3. Алгоритмы реконструкции изображений в невидимой области на основе критерия алгоритмической вероятности.**

### **3.1. Введение**

В предыдущей главе был рассмотрен критерий на основе алгоритмической вероятности для решения задачи реконструкции изображений в невидимой области, а также были предложены методы на базе исходного критерия, которые при помощи некоторых аппроксимаций делали задачу предсказания содержимого в отсутствующих на изображении областях вычислимой. В данной главе мы рассмотрим конкретные алгоритмы, воплощающие предложенные ранее методики, также чуть более подробно коснемся свойств моделей специального вида, способных к обучению представлениям.

### **3.2. Алгоритм реконструкции изображения с использованием спектрального представления**

Ранее отмечалось, что наиболее простым представлением изображений, для которого можно сформулировать аппроксимацию критерия алгоритмической вероятности, является функциональное представление. Также мы установили, что такое представление должно быть распределенным, дабы значение каждого элемента подобного представления зависело бы от значений яркостей различных пикселей исходного изображения. В результате для апробации предлагаемого нами критерия на практике было решено воспользоваться спектральным представлением изображений. Если с помощью данного, сравнительно бедного по своей выразительной силе, представления удастся синтезировать алгоритм, который позволит получить какие-либо интересные результаты, то в дальнейшем мы рассмотрим чуть менее тривиальные типы представлений с целью получения более эффективных и качественных решений.

Как уже было показано в предыдущей главе, базовой оценкой длины генеративной по отношению к изображению программы в случае спектрального представления является энтропия спектра. Как следствие, энтропийная оценка сложности спектра будет являться ключевым параметром, влияющим на величину алгоритмической вероятности. Но, даже имея возможности эффективной оценки сложности генеративного «кода», задача поиска минимума критерия (10) не может быть решена в явном виде. Так или иначе, но для эффективного синтеза алгоритмов реконструкции изображений в невидимой области оказывается необходимым введение дополнительных эвристик поиска. Таким образом, даже на этапе реализации конкретной методики на практике из одного метода возможно получение нескольких алгоритмов решения, разнящихся между собой процедурами поиска минимума для критерия (10).

В качестве одной из самых простых эвристик можно предложить следующий итерационный алгоритм [78] (который, однако, не гарантирует сходимости к лучшему решению).

1. Восстанавливаемая область заполняется начальными значениями путем, например линейной интерполяции.

2. Итерационно выполняется: для каждого пикселя в восстанавливаемой области подбирается такое значение яркости, что энтропия спектра всего изображения (вычисленная при фиксированных текущих значениях яркостей всех остальных пикселей в области) минимальна.

Уже в таком виде данный алгоритм позволяет получить довольно неплохие результаты, в чем можно будет убедиться в следующей главе. Тем не менее, существуют возможности для его модификации. Во-первых, в качестве начального предположения о содержимом восстанавливаемой области может выступать результат более продвинутой предобработки и оценки яркостей пикселей. Во-вторых, важным моментом в процессе поиска решения в соответствии с итерационной процедурой из пункта 2 алгоритма



является порядок обхода пикселей внутри области реконструкции. Можно заполнять область, двигаясь внутри нее построчно, либо вдоль столбцов, но, возможно, наиболее эффективным вариантом обхода будет являться движение по спирали – сначала вдоль границ, постепенно уменьшая радиус и сходясь к центру восстанавливаемой области. Пример варианта подобного обхода представлен на рисунке 17.

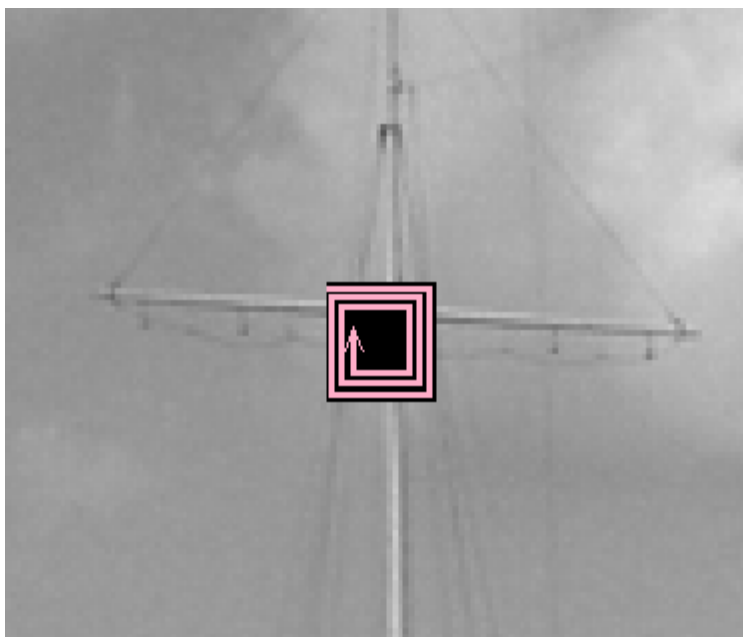


Рисунок 17 – Пример спиралевидного обхода пикселей области, подлежащей реконструкции

Отметим, что предложенные выше алгоритмы становятся крайне требовательными с точки зрения вычислительных мощностей при увеличении размера области, подлежащей реконструкции.

Также возможно допустить глобальную оптимизацию критерия (10) за счет использования более мощных метаэвристик поиска, привлекающих методы Монте-Карло. Примерами таких метаэвристик могут быть алгоритм имитации отжига, а также генетические алгоритмы. Подобная глобальная оптимизация, скорее всего, позволит добиться сходимости решения к наилучшему возможному, но ни коем образом не снимет проблемы бысродействия, свойственной итерационным алгоритмам перебора.

### 3.3. Об алгоритмах обучения систем представлений

Как уже обсуждалось ранее, важным классом моделей, позволяющим системе реконструкции изображений в невидимой области имитировать обработку, аккумуляцию и использование предыдущего опыта, являются модели, способные к обучению представлений. Важным семейством или подклассом подобных моделей являются системы автоассоциативной памяти или автоэнкодеры. Интересными для нас их делает свойственная им способность выделять на изображениях признаки, находящиеся в нелинейной взаимосвязи с интенсивностями пикселей исходного изображения или признаками более низкого уровня.

Рассмотрим простейшую модель автоэнкодера, который можно описать как нечто состоящее из кодировщика и декодировщика. Кодировщик представляет собой функцию, однозначно отображающую входной вектор  $\mathbf{x}$  в скрытое описание или представление  $\mathbf{y}$ . Как правило, это выражается во взятии нелинейной функции от результата аффинного преобразования:

$$\mathbf{y} = f_{\theta=(\mathbf{W}, \mathbf{b})}(\mathbf{x}) = s(\mathbf{W}\mathbf{x} + \mathbf{b}),$$

где  $\theta$  – объединенный вектор параметров для функции  $f(\mathbf{x})$ ,  $\mathbf{W}$  – матрица параметров размера  $d' \times d$ , которые обозначают размеры кодируемого и исходного векторов признаков соответственно;  $\mathbf{b}$  – вектор смещений размерностью  $d'$ ,  $s$  – некоторая нелинейная функция. В случае, когда  $d' < d$ , то говорят о сжимающем представлении, если же  $d' > d$ , то говорят об избыточном представлении.

Под декодером (или декодировщиком) обычно понимают функцию, которая отображает скрытые признаки  $\mathbf{y}$  обратно в вектор  $\mathbf{z}$  исходного пространства размерностью  $d$ . По аналогии с кодировщиком данная функция – это, опять же, результат взятия нелинейной функции от аффинного преобразования:

$$\mathbf{z} = g_{\theta'=(\mathbf{W}', \mathbf{b}')}(\mathbf{y}) = s(\mathbf{W}'\mathbf{y} + \mathbf{b}'),$$

где  $\mathbf{W}'$  и  $\mathbf{b}'$  – матрица и вектор параметров соответствующих размеров.

В общем случае вектор  $\mathbf{z}$  не обязательно должен интерпретироваться как точная реконструкция вектора  $\mathbf{x}$ , например, в терминах вероятностного подхода к данной модели вводится условная плотность вероятности  $p(\mathbf{x}|\mathbf{z})$ , в которой  $\mathbf{z}$  интерпретируется как значение параметра, при котором  $\mathbf{x}$  генерируется из данного распределения с наибольшей вероятностью. В рамках данного подхода к интерпретации моделей автоэнкодеров для получения скрытого вектора  $\mathbf{y}$ , содержащего наиболее интересные признаки, зачастую прибегают к применению принципа *инфомакса* [79, 80]. Суть данного принципа состоит в том, чтобы полученное представление  $\mathbf{y}$  оставляло внутри себя как можно больше информации об исходном векторе признаков  $\mathbf{x}$ . В терминах теории информации это будет эквивалентно максимизации взаимной информации  $I(\mathbf{x}, \mathbf{y})$  для двух распределений случайного входного вектора  $\mathbf{X}$  и высокоуровневого представления  $\mathbf{Y}$ , соответственно.

Можно показать, что критерий максимизации взаимной информации может быть сведен к максимизации условной апостериорной вероятности:

$$\max(I(\mathbf{x}, \mathbf{z})) \propto \max(p(\mathbf{x} | \mathbf{z})).$$

Но, как правило, для удобства решения задачи оптимизации вводят понятие функции ошибки реконструкции:

$$L(\mathbf{x}, \mathbf{z}) \propto -\log(p(\mathbf{x} | \mathbf{z})).$$

Таким образом, максимизация взаимной информации становится эквивалентной минимизации функции ошибки реконструкции:

$$\max(I(\mathbf{x}, \mathbf{z})) \propto \min(-\log(p(\mathbf{x} | \mathbf{z}))).$$

Далее в зависимости от предположения о принадлежности плотности распределения вероятностей  $p(\mathbf{x}|\mathbf{z})$  к тому или иному семейству параметрических функций могут строиться гипотезы относительно конкретного вида функции  $L(\mathbf{x}, \mathbf{z})$ , описывающей поверхность ошибок реконструкции.

Наиболее распространены следующие варианты предположений о виде плотности распределения апостериорной вероятности.

1. Для входных векторов, каждый элемент которых описывается множеством действительных чисел, то есть для  $\mathbf{x} \in R^d$ :  $X|\mathbf{z} \sim N(\mathbf{z}, \sigma, \mathbf{I})$ , где  $N$  соответствует нормальному закону распределения вероятностей,  $\sigma$  – среднеквадратичному отклонению (или в общем случае ковариационной матрице),  $\mathbf{I}$  – вектору средних. Данный случай приводит к следующему виду функции ошибок:  $L(\mathbf{x}, \mathbf{z}) = L_2(\mathbf{x}, \mathbf{z}) = C(\sigma^2) \|\mathbf{x} - \mathbf{z}\|^2$ , где  $C(\sigma)$  – константа, зависящая исключительно от  $\sigma$ , и которую можно игнорировать в процессе оптимизации. Очевидно, что таким способом получается уже ставший классическим метод наименьших квадратов (МНК), традиционно использующийся при обучении различных регрессионных моделей, а также нейронных сетей.

2. Для входных векторов, каждый элемент которых описывается множеством бинарных значений, то есть для  $\mathbf{x} \in \{0,1\}^d$ :  $X|\mathbf{z} \sim B(\mathbf{z})$ , где  $B(\cdot)$  соответствует распределению Бернулли. Для данного случая необходимо, чтобы значения компонент вектора  $\mathbf{z}$ , получаемые в результате декодирования, лежали в диапазоне  $[0,1]$ , то есть  $\mathbf{z} \in [0,1]^d$ . Как правило, для этих целей используется сигмоидальная активационная функция  $s(\mathbf{x}) = 1/(1 + \exp(-\mathbf{x}))$  на выходах декодера. В результате функция ошибки принимает следующий вид:  $L(\mathbf{x}, \mathbf{z}) = L_H(\mathbf{x}, \mathbf{z}) = -\sum_j [x_j \log z_j + (1 - x_j) \log(1 - z_j)] = H(B(\mathbf{x}) \| B(\mathbf{z}))$ , где  $H(\cdot \| \cdot)$  обозначает кросс-энтропию. Отсюда и название для этой функции ошибки – кросс-энтропийная, где величина кросс-энтропии измеряется для двух независимых многомерных распределений Бернулли. Для первого из этих распределений вектором математический ожиданий будет являться  $\mathbf{x}$ , а для второго,

соответственно,  $\mathbf{z}$ . Отметим, что требование  $\mathbf{x} \in \{0,1\}^d$  не является строгим, то есть данное допущение будет работать и в случае, когда  $\mathbf{x} \in [0,1]^d$ .

Здесь мы описали лишь самые распространенные гипотезы о виде плотности распределения апостериорной вероятности для автоэнкодера. Существует множество других параметрических функций, как для кодировщика, так и для декодировщика, не говоря уже о непараметрических методах оценки плотности вероятности [81].

В рамках данной работы нами были реализованы архитектуры автоэнкодеров, которые воплощали как МНК функцию ошибки, так и кросс-энтропийный вариант этой функции.

Соответственно, в обоих вариантах кодировщик состоял из последовательного применения по отношению к входному вектору  $\mathbf{x}$  аффинного преобразования и сигмоидальной функции. А вот архитектура декодера разнилась в зависимости от типа функции ошибки: для МНК ошибки декодер состоял из применения к полученному вектору признаков  $\mathbf{y}$  лишь аффинного преобразования, в то время как кросс-энтропийная оценка ошибки, как и в случае с кодировщиком, требовала последовательного применения по отношению к полученному вектору признаков  $\mathbf{y}$  аффинного преобразования и сигмоидальной функции.

Еще одним решением, существенно упрощавшим архитектуру реализованного нами автоэнкодера, было введение общих для кодировщика и декодера весов связи. Иными словами в процессе работы автоэнкодера входной вектор  $\mathbf{x}$  и скрытый вектор признаков  $\mathbf{y}$  проходили через одну и ту же матрицу параметров так, что  $\mathbf{W}' = \mathbf{W}^T$ . Такой подход позволял существенно сэкономить вычислительные ресурсы, требуемые для обучения сети.

Таким образом, обучение автоэнкодера сводилось к минимизации функции ошибки реконструкции  $L(\mathbf{x}, \mathbf{z})$ , задаваемой в том или ином виде. Для

этих целей использовался уже ставший типовым алгоритм стохастического градиентного спуска. Данный алгоритм подразумевает корректировку параметров модели непосредственно каждый раз после предъявления обучающего примера, вместо расчета среднего значения изменения параметров по всей выборке.

### **3.3.1. Недостаточность критерия максимизации взаимной информации для выделения информативных признаков**

Стоит еще раз отметить важное следствие, проистекающее из обучения автоэнкодера в соответствии с критерием инфомакса. Заключается оно в том, что, если полученное представление  $y$  дает в результате декодирования качественную реконструкцию входного вектора  $x$ , то это значит, что такое представление сохранило большинство той необходимой информации, которая в достаточной мере характеризует входной вектор.

Однако критерий, говорящий о том, что представление  $y$  должно сохранять информацию о входе  $x$  явно не является достаточным для того, чтобы такое представление  $y$  можно было интерпретировать как программу или «код» с точки зрения колмогоровской сложности. И действительно, в самом тривиальном случае максимума взаимной информации можно достигнуть, просто приравняв  $x = y$ .

Без каких-либо дополнительных ограничений данный критерий сам по себе вряд ли позволит построить эффективные признаковые описания, а также настроить параметры модели так, чтобы она могла служить опорной машиной для решения задачи реконструкции изображений в утеранных областях. Таким образом, из общей информации, содержащейся в исходном векторе, нам необходимо выделить компактное описание наиболее существенной информации об объекте или классе, к которому данный объект может принадлежать. Правда, это неминуемо приведет к тому, что значение ошибки реконструкции будет ненулевым.

Традиционным подходом к решению данной задачи является построение автоэнкодеров, которые в процессе кодирования давали бы

сжатое представление исходного вектора, то есть такие, в которых  $d' < d$  [82]. Альтернативным подходом может считаться использование избыточных представлений (когда  $d' > d$ ), на которые наложены ограничения связанные с разреженным кодированием. Соответственно, такие представления принято называть разреженными. Большой интерес, проявляемый к разреженным представлениям, связан во многом с тем, что некоторое время назад из области нейрофизиологии были получены свидетельства того, что нейронная активность мозга носит в некотором роде разреженный характер [83, 86, 93]. Так же важным свойством разреженных представлений является относительная легкость при их интерпретации, в отличие от сжатых представлений, где при изменении какого либо значения входного вектора существенным образом изменяются все компоненты представления.

Основной областью применения подобных разреженных представлений являются сети глубокого обучения, в которых они используются, как правило, для повышения результативности работы при классификации [84, 85, 90].

Стоит отметить, что разреженное избыточное представление может рассматриваться, в том числе, и как альтернативный подход к получению сжатых представлений в том смысле, что сжатие информации там происходит неявно вследствие наличия в таком представлении большого количества нулевых компонент. Зачастую в таких представлениях количество значащих компонент бывает существенно меньшим, нежели в сжимающих представлениях.

### **3.3.2. Регуляризация как способ повышения информативности выученных признаков**

Часто даже в случае сжимающих представлений для получения сглаженных, то есть плавно изменяющихся отсчетов внутри компонент скрытого вектора прибегают к наложению ограничений на сами параметры модели. Одним из наиболее популярных методов для решения этой задачи является процедура «тушения» или «разложения» весов связи, которую в

математической статистике принято называть  $L_2$  регуляризацией [87, 92]. Суть этой методики состоит в следующем: предполагается, что наиболее вероятные значения параметров модели сосредоточены вокруг некоторого значения, являющегося математическим ожиданием плотности распределения априорных вероятностей для этих самых параметров. При этом чаще всего данное математическое ожидание помещают в точку ноль. В соответствие с этим процесс обучения модели переформатируется в процедуру поиска условного экстремума:

$$L_2(\mathbf{x}, \mathbf{z}) = \|\mathbf{x} - \mathbf{z}\|^2 + \|\mathbf{W}\|^2. \quad (13)$$

Данное выражение записано с точностью до нормирующих множителей перед слагаемыми, что никак не сказывается на процессе оптимизации. Соответствующая такой функции ошибки априорная плотность вероятности значений параметров автоэнкодера выглядит следующим образом:

$$p(\mathbf{W}) = N(\mathbf{W} | 0, \alpha) \propto \exp\left(\frac{\alpha}{2} \mathbf{W}^T \mathbf{W}\right),$$

где параметр  $\alpha$ , отвечающий за «точность» данного распределения, принято называть гиперпараметром. Геометрически такой тип регуляризации можно интерпретировать в соответствии с тем, как это показано на рисунке 18.



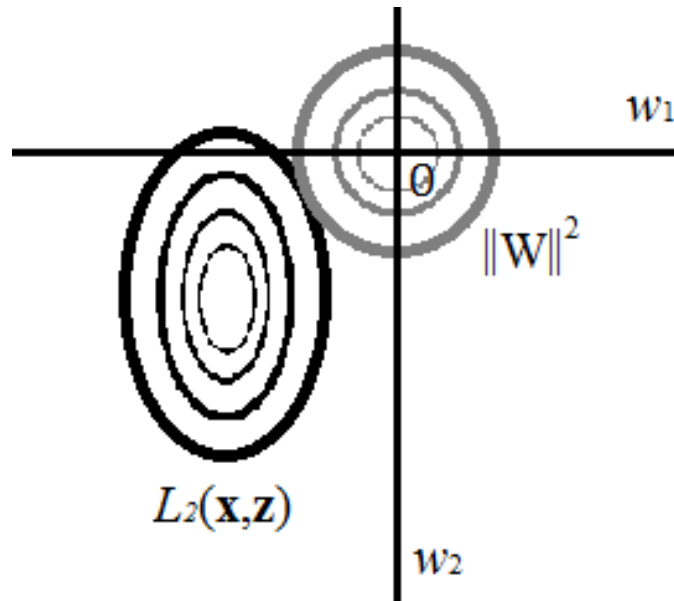


Рисунок 18 – Геометрическая интерпретация «разложения» весов связи для двумерного пространства параметров

### 3.3.3. Разреженное кодирование

Одним из способов реализации механизма разреженного кодирования является наложение ограничений на среднюю по всей выборке активацию каждого из признаков на скрытом, кодирующем, слое автоэнкодера [89]. Обозначим  $y_j(\mathbf{x})$  значение активации  $j$ -го признака на скрытом слое при условии, что на вход автоэнкодеру подается вектор  $\mathbf{x}$ . Тогда через

$$\bar{\rho}_j = \frac{1}{m} \sum_{i=1}^m (y_j(\mathbf{x}_i))$$

будем обозначать среднюю по обучающей выборке активацию  $j$ -го признака на кодирующем слое, где  $m$  – количество образов в обучающей выборке. Зачастую для реализации необходимых ограничений на значения  $\bar{\rho}_j$  в процессе обучения автоэнкодера в целевую функцию ошибки вводят дополнительные «штрафующие» слагаемые. При этом можно руководствоваться следующей теоретико-информационной логикой: представим, что требуемое достаточно малое значение средне выборочной активации для каждого признака принадлежит некоторому неизвестному

распределению  $\rho(v)$ , которое мы пытаемся моделировать, используя аппроксимацию в виде  $\bar{\rho}_j(v)$ .

Если  $\bar{\rho}_j(v)$  применяется для построения схемы кодирования с целью передачи значений  $v$  приемнику сообщений, то среднее количество информации, которое требуется добавить к сообщению, чтобы однозначно определить значение  $v$  (при условии кодирования без потерь) в силу использования  $\bar{\rho}_j$  вместо истинного распределения  $\rho(v)$ , можно выразить так:

$$\begin{aligned} KL(\rho \parallel \bar{\rho}_j) &= -\sum_{v \in V} \rho(v) \log(\bar{\rho}_j(v)) - (-\sum_{v \in V} \rho(v) \log(\rho(v))) = \\ &= -\sum_{v \in V} \rho(v) \log\left(\frac{\bar{\rho}_j(v)}{\rho(v)}\right). \end{aligned}$$

Данное выражение обычно называют относительной энтропией или расстоянием Кульбака-Лейблера [88] (также распространен термин «информационная дивергенция») между распределениями  $\rho(v)$  и  $\bar{\rho}_j(v)$ . Стоит отметить, что данная величина не является симметричной, то есть  $KL(\rho \parallel \bar{\rho}_j) \neq KL(\bar{\rho}_j \parallel \rho)$ .

Данная мера обладает замечательным свойством, а именно она удовлетворяет условию  $KL(\rho \parallel \bar{\rho}_j) \geq 0$ , при этом данное неравенство обращается в тождество тогда и только тогда, когда  $\rho(v) = \bar{\rho}_j(v)$ . Подобное свойство, очевидно, является полезным в случае введения дополнительных слагаемых на базе относительной энтропии в оптимизируемые функции.

Учитывая тот факт, что в архитектуре автоэнкодера при проектировании кодировщика нами использовалось сигмоидальное «сплющивание» результатов аффинного преобразования, то получаемые значения признаков на кодирующем слое (которые в силу «сплющивания» всегда будут лежать в диапазоне  $[0,1]$ ) могут с легкостью, как это уже было показано ранее, моделироваться при помощи распределения Бернулли. Тогда с учетом того, что математическими ожиданиями, а значит и параметрами,

для распределений  $\rho(v)$  и  $\bar{\rho}_j(v)$  будут являться соответствующие вероятности  $\rho$  и  $\bar{\rho}_j$ , можно записать формулу относительной энтропии для  $j$ -го признака на кодирующем слое:

$$KL(\rho \parallel \bar{\rho}_j) = \rho \log\left(\frac{\rho}{\bar{\rho}_j}\right) + (1 - \rho) \log\left(\frac{1 - \rho}{1 - \bar{\rho}_j}\right).$$

Далее можно предложить следующую «штрафующую» добавку к функции ошибки реконструкции:

$$\sum_{j=1}^{d'} KL(\rho \parallel \bar{\rho}_j) = \sum_{j=1}^{d'} \rho \log\left(\frac{\rho}{\bar{\rho}_j}\right) + (1 - \rho) \log\left(\frac{1 - \rho}{1 - \bar{\rho}_j}\right),$$

при этом в качестве  $\rho$  обычно выбирают некоторое, достаточно близкое к нулю, число, которое должно определять степень «разреженности» кода в виде требования равенства этому числу средней активации каждого признака в получаемом коде.

В соответствии с требованиями разреженного кодирования итоговая целевая функция ошибки примет следующий вид:

$$L_{sparse}(\mathbf{x}, \mathbf{z}) = L_2(\mathbf{x}, \mathbf{z}) + \sum_{j=1}^{d'} KL(\rho \parallel \bar{\rho}_j). \quad (14)$$

Получаемый в результате оптимизации данной целевой функции автоэнкодер позволяет выделять на изображениях интересные наборы признаков даже в случаях, когда  $d' > d$ , то есть при избыточном кодировании, позволяя тем самым сохранить компактность кода в смысле колмогоровской сложности.

### 3.3.4. Обучение шумоподавлению

Существует еще один подход, который позволяет автоэнкодеру в результате кодирования получать признаки, формирующие интересные с точки зрения реконструкции изображений представления.

В отличие от методик, накладывающих ограничений на получаемые представления  $\mathbf{u}$  или на набор параметров  $\mathbf{W}$ , в данном подходе меняется сам критерий реконструкции, который превращается в процесс удаления шумов с

частично зашумленного изображения. Таким образом, изменяются сами требования, предъявляемые к «интересному» или полезному представлению. Для данного подхода эти требования можно сформулировать так: хорошим представлением может называться такой результат кодирования, который при поступлении на вход зашумленного образа неизменно давал бы один и тот же набор признаков, способный восстановить исходный образ, но уже без шума. В основе данного подхода лежат две главные идеи:

1. Достаточно естественное желание того, чтобы представления более высокого уровня были бы устойчивыми к воздействию извне, оказываемому на входные образы.
2. Ожидание того, что в процессе успешного решения задачи шумоподавления система выделит лишь нужные и полезные признаки, в достаточной мере характеризующие исходное распределение.

При этом, как правило, задача шумоподавления не ставится как цель работы системы, а используется лишь как средство, для получения устойчивых (робастных) признаков более высокого уровня [97]. Обычно такие признаки далее используются в качестве входных описаний объектов интереса для различного рода классификаторов [90, 91]. А, собственно, в системах для решения задачи классификации данные признаковые описания уже успели зарекомендовать себя с самой положительной стороны [94, 96].

Таким образом, шумоподавляющий автоэнкодер обучается реконструкции «чистого» исходного образа на основе зашумленной версии этого образа, подаваемой на вход. Для этого необходимо подвергнуть первоначальный образ  $\mathbf{x}$  воздействию шума и получить его искаженную версию  $\tilde{\mathbf{x}}$ .

Далее искаженный вектор  $\tilde{\mathbf{x}}$  подвергается тому же набору преобразований, который применяется для входного вектора в случае обычного автоэнкодера. Соответственно, скрытый вектор признаков теперь будет определяться так:

$$\mathbf{y} = f_{\theta}(\tilde{\mathbf{x}}) = s(\mathbf{W}\tilde{\mathbf{x}} + \mathbf{b}).$$

При помощи вектора  $y$  можно получить результат реконструкции  $z = g_{\theta}(y)$ . Параметры, обозначенные нами через  $\theta$ , в процессе обучения настраиваются таким образом, чтобы минимизировать среднюю ошибку реконструкции по всей обучающей выборке, при этом восстановленный в процессе декодирования вектор  $z$  должен быть как можно более похожим на исходный, неискаженный образ  $x$ . Схематическое описание данной процедуры представлено на рисунке 19.

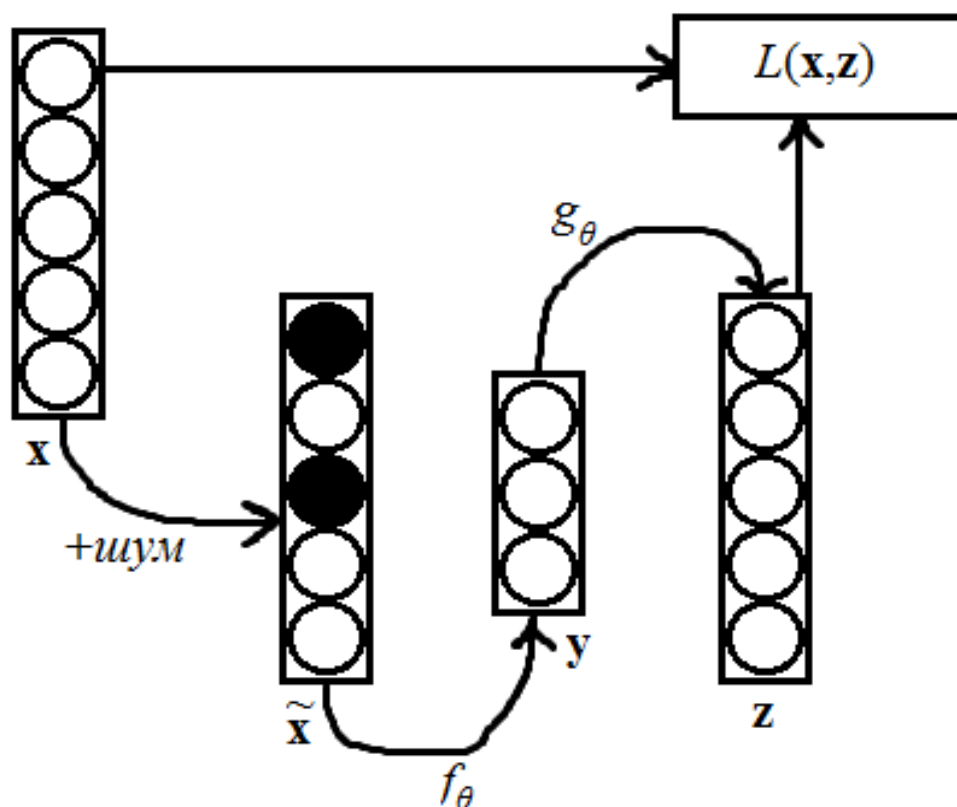


Рисунок 19 – Архитектура автоэнкодера, обучаемого в стиле шумоподавления. Образ  $x$  подвергнут воздействию шума, искажающего его до  $\tilde{x}$ . Автоэнкодер на основании образа  $\tilde{x}$  строит признаковое описание  $y$  при помощи кодировщика  $f_{\theta}$ , далее пытается восстановить исходный  $x$  при помощи декодера  $g_{\theta}$ . Ошибка реконструкции оценивается при помощи функции  $L(x, z)$

Как и в случае обычного автоэнкодера настройка параметров модели, изначально задаваемых случайными числами, может осуществляться при помощи процедуры стохастического градиентного спуска.

Стоит особо отметить, что каждый раз, когда в процессе обучения мы на входе получаем образ  $x$ , под воздействием шума генерируются все новые и новые искаженные версии  $\tilde{x}$ , которые не совпадают с  $\tilde{x}$ , полученными на предыдущих эпохах обучения.

Примечателен тот факт, что автоэнкодеры, обучаемые в стиле шумоподавления, также стремятся максимизировать взаимную информацию, то есть используют принцип инфомакса в своей основе, в силу того, что оптимизация функции ошибки нацелена на минимизацию потерь реконструкции между выборочным случайным вектором  $X$  и результатом его реконструкции из случайного вектора  $Y$ . Основная разница между классическим автоэнкодером и данной модификацией состоит в том, что представление  $y$  теперь получается за счет применения детерминированной функции  $f_\theta$  к стохастически искаженному вектору  $x$ .

Таким образом, выучиваемая шумоподавляющим автоэнкодером функция, а значит и кодируемое представление, представляют собой нечто намного менее тривиальное нежели простое тождественное отображение.

### **3.3.5. Геометрическая интерпретация критерия шумоподавления**

Процесс реконструкции по критерию шумоподавления может иметь интуитивную геометрическую интерпретацию. Как уже отмечалось ранее, между входным признаковым описанием образов и высокоуровневыми характеристиками классов, кластеров и групп, к которым принадлежат данные образы, существует нелинейная взаимосвязь.

С геометрической точки зрения это означает, что естественные данные, задаваемые в пространствах высокой размерности, как правило, концентрируются возле нелинейных многообразий (подпространств локально сходных с евклидовым) более низких размерностей. Такая

геометрическая интерпретация получила название гипотезы об обучении многообразиям [95].

Соответственно, во время обучения автоэнкодера в режиме шумоподавления мы пытаемся найти такой стохастический оператор  $p(X | \tilde{X})$ , который отобразил бы множество искаженных образов  $\tilde{X}$  во множество неискаженных образов  $X$ . Скорее всего, искаженные примеры, подаваемые системе в процессе обучения, будут находиться гораздо дальше от общего для множества  $X$  целевого многообразия, нежели сами исходные образы. Поэтому оператор  $p(X | \tilde{X})$  обучается отображению, которое переводит точки с низкой вероятностью  $\tilde{x}$  в точки с более высокой вероятностью  $x$ , находящиеся на или достаточно близко к искомому многообразию.

Успешное решение данной задачи должно приводить к тому, что выученный оператор окажется способным к переносу даже далеко расположенных точек в области, достаточно близкие к целевому многообразию.

Графическая иллюстрация данной геометрической интерпретации представлена на рисунке 20.

Очевидно, что такая интерпретация явным образом применима к представлениям  $Y$ , чья размерность меньше размерности  $X$ . В таком случае это представление  $Y = f_{\theta}(X)$  можно трактовать как координатную сетку для точек на многообразии. Иными словами  $Y$  является таким представлением для  $X$ , которое достаточно хорошо улавливает изменения в исходных данных, лежащие и описываемые на определенном многообразии.

С точки же зрения обработки изображений этот процесс можно проиллюстрировать с помощью серии рисунков 21-23.

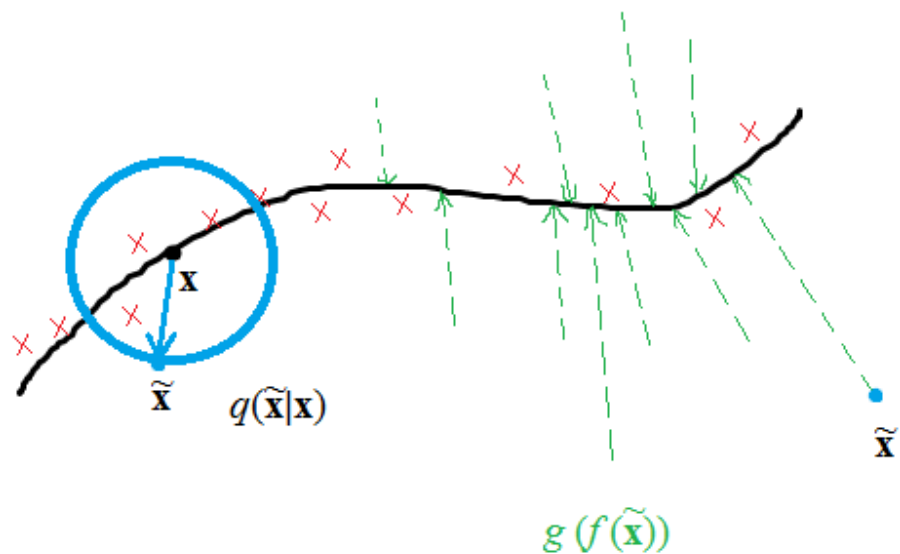


Рисунок 20 – Иллюстрация геометрической интерпретации критерия шумоподавления для автоэнкодера, выраженной в виде гипотезы об обучении многообразиям. Данные, представленные в виде обучающего набора примеров и обозначаемые красным значком  $\times$ , концентрируются вблизи поверхности (многообразия) более низкой по сравнению с исходным пространством размерности. Искаженные образы, обозначаемые голубыми точками ( $\cdot$ ), могут быть получены посредством применения к исходному вектору стохастического оператора  $q(\tilde{\mathbf{x}} | \mathbf{x})$ . В результате искаженные образы в общем случае будут находиться дальше от границ поверхности. Модель автоэнкодера с шумоподавлением  $g_{\theta}(f_{\theta}(\tilde{\mathbf{x}}))$  будет обучаться тому, чтобы спроецировать эти искаженные образы обратно на многообразие



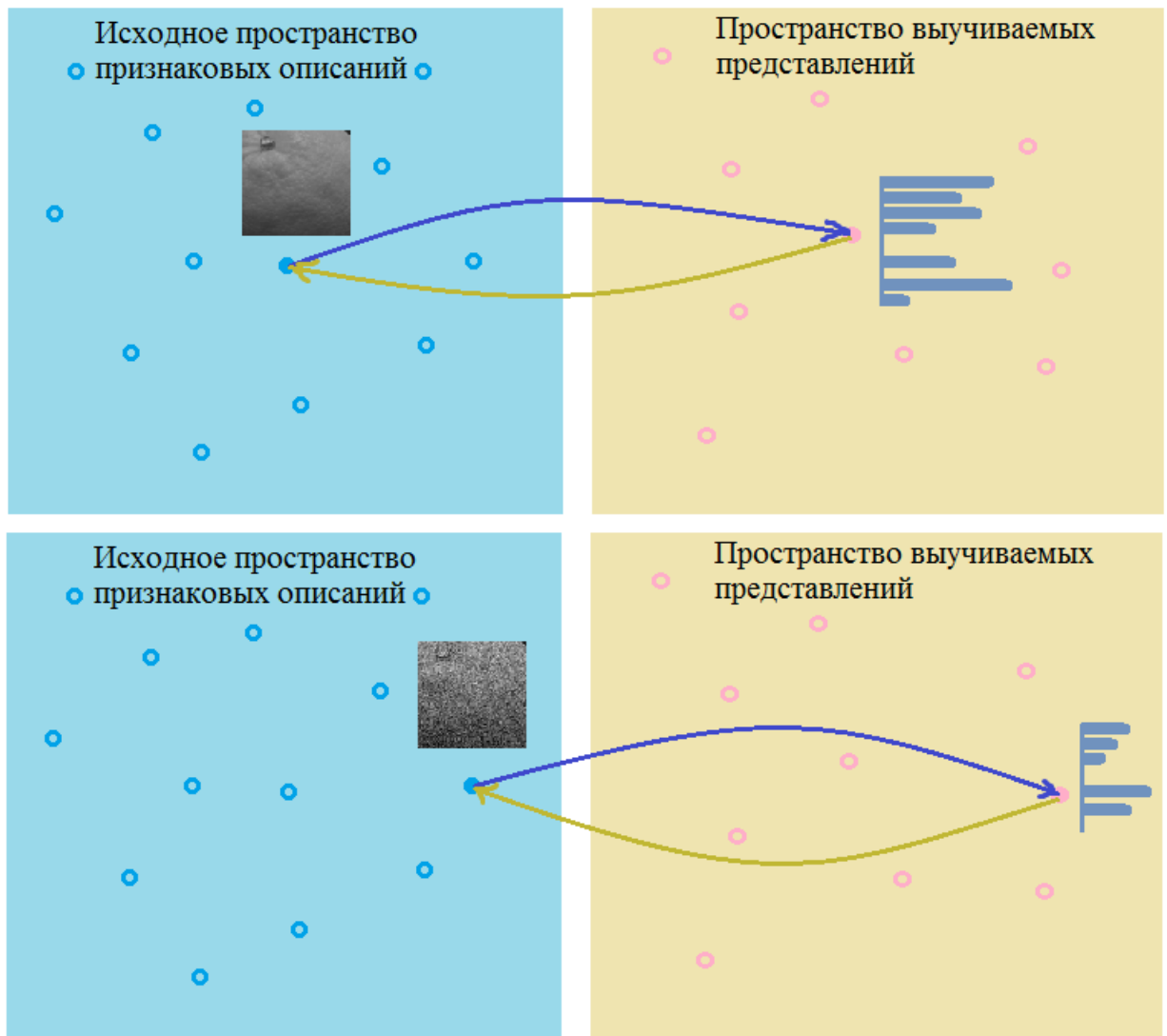


Рисунок 21 – Иллюстрация работы автоэнкодера, обученного на основе классической минимизации ошибки реконструкции по обучающей выборке. Различным изображениям соответствуют различные представления, несмотря на то, что изображения различаются только лишь наличием шума на одном из них. Таким образом, система оказывается неспособной выявить регулярностей, содержащихся в исходных данных. Она просто осуществляет копирование данных в новом пространстве признаков

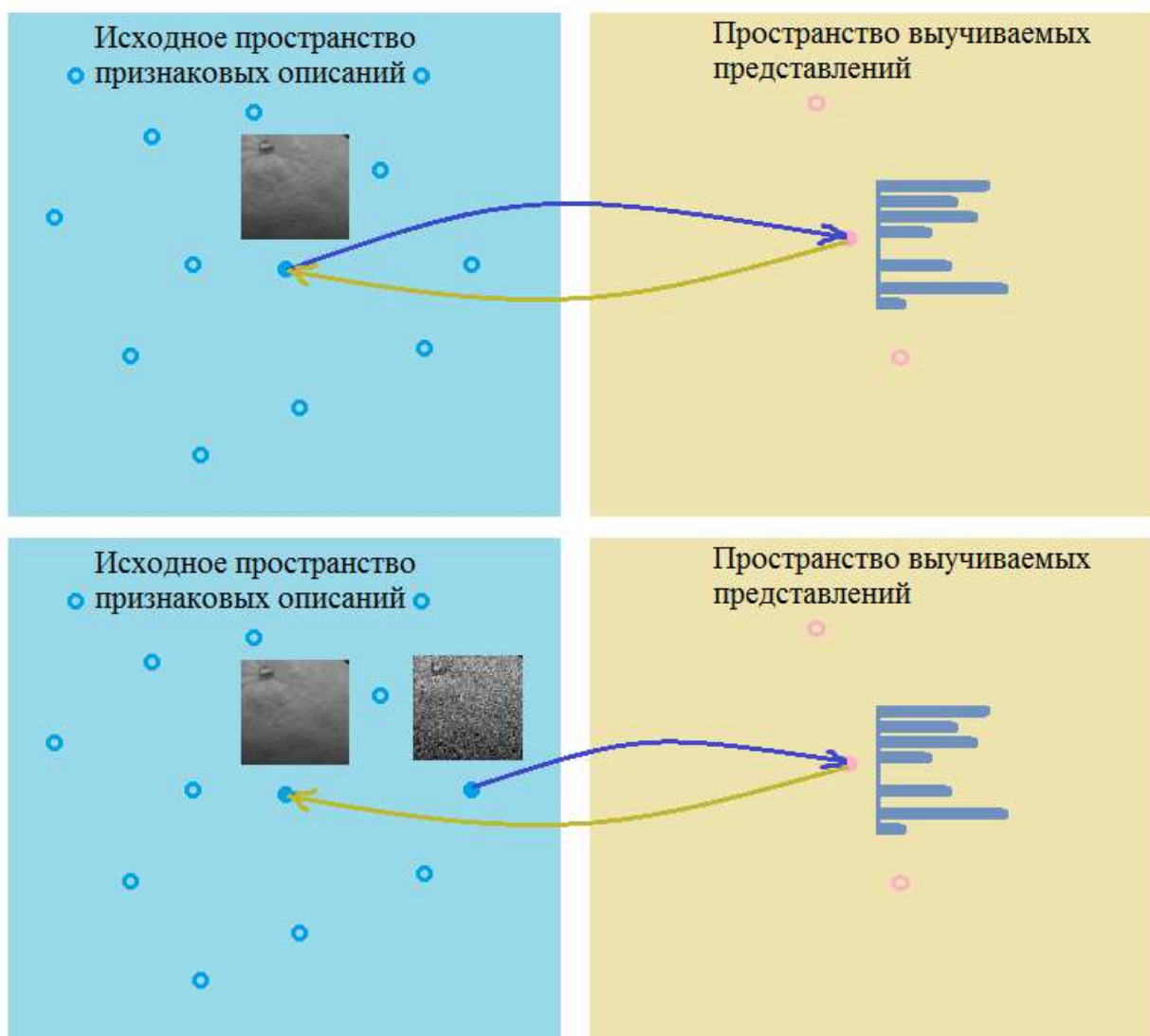


Рисунок 22 – Иллюстрация работы автоэнкодера, обученного в условиях ограниченного количества доступных кодирующих признаков. Такой стиль обучения свойственен как разреженному кодированию, так и при обучении шумоподавлению. Можно увидеть, что в данном случае двум различным входным описаниям одного и того же объекта соответствует единственный вектор в пространстве представлений более высокого уровня. Таким образом, уменьшение количества доступных «кодовых» описаний объекта автоматически влечет компактность получаемого кода, в некоторой степени устраняя необходимость дополнительной оценки сложности

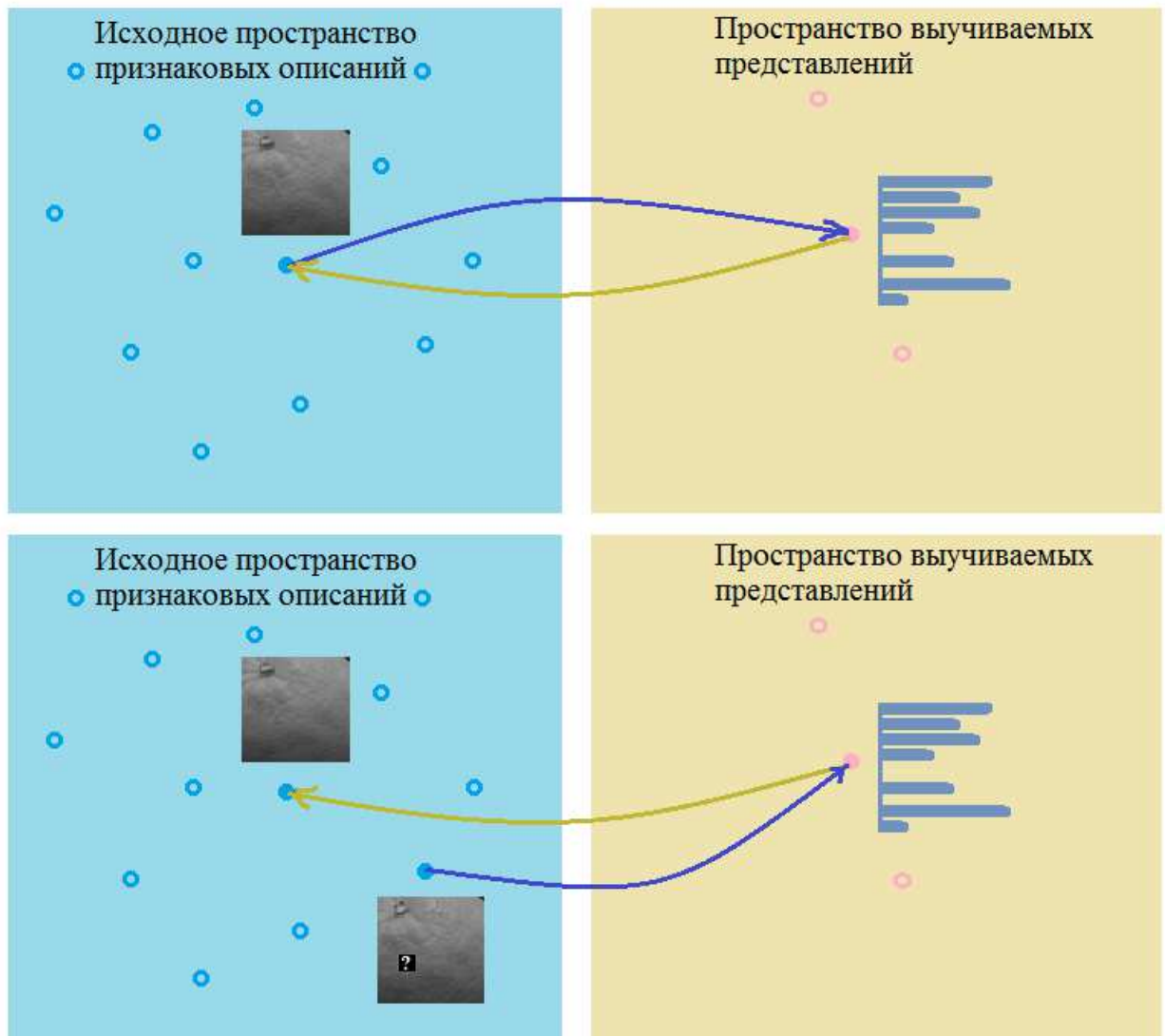


Рисунок 23 – Иллюстрация работы автоэнкодера, обученного в условиях ограниченного количества доступных кодирующих признаков, с целью реконструкции изображений в невидимой области. Подобные системы являются мощными опорными машинами, которые могут быть эффективно использованы для задач реконструкции содержимого внутри отсутствующих на изображении областей. Нами предложено несколько алгоритмов решающих данную задачу (подробнее см. в разделе 3.5)

### 3.3.6. Регуляризация и компактность кода

Можно увидеть, что подобные модели представляют собой класс достаточно мощных и в тоже время удобных в использовании опорных машин, аккумулирующих свой предшествующий опыт. Важной особенностью таких машин является возможность достаточно легкого

внесения в их структуру любой дополнительной априорной информации, позволяющей существенно ограничивать пространство представлений для объектов интереса, а следовательно упрощать поиск в этом пространстве. Данная возможность достигается за счет введения ограничений для получаемых признаков, введения критерия шумоподавления в целевую функцию ошибки реконструкции, а также при помощи «штрафов» на «сложность» параметров модели, что выражается в виде классических механизмов регуляризации.

Но, как мы увидим далее, получаемое в результате кодирования представление не является компактным в универсальном смысле этого слова. Такой код будет лишь «локально»-компактным, и, соответственно, для наиболее эффективного решения задачи реконструкции изображений в невидимой области требуются дополнительные механизмы поиска.

Отметим, что такая необходимость в привлечении дополнительных инструментов поиска в том или ином виде будет возникать практически для любой опорной машины, которая может быть использована при решении задачи реконструкции содержимого внутри отсутствующих на изображении областей.

### **3.4. Сверточный автоэнкодер**

Еще рассмотрим один интересный вариант систем способных к обучению представлениям, который может быть полезным при решении задачи реконструкции изображений в невидимой области, и который для этих целей мы реализовали в рамках данной работы. Эта система носит название сверточного автоэнкодера [98]. Суть работы сверточного автоэнкодера может быть продемонстрирована на рисунке 24.

Каждый выходной, т. е. кодирующий, нейрон активируется в результате нелинейного преобразования над взвешенной суммой элементов внутри фрагмента входного изображения, покрытого маской весов связи, которая движется по всему изображению со сдвигом на единицу в том или ином направлении, тем самым делая процесс активации нейронов выхода

идентичным операции свертки [99]. При этом количество масок с весами связи и их размер задаются произвольно. В результате получатся несколько (по количеству масок) двумерных наборов активаций выходных признаков. Такие результаты свертки входного изображения с матрицами весов связи принято называть картами признаков. Так работает сверточный автоэнкодер в режиме кодирования.

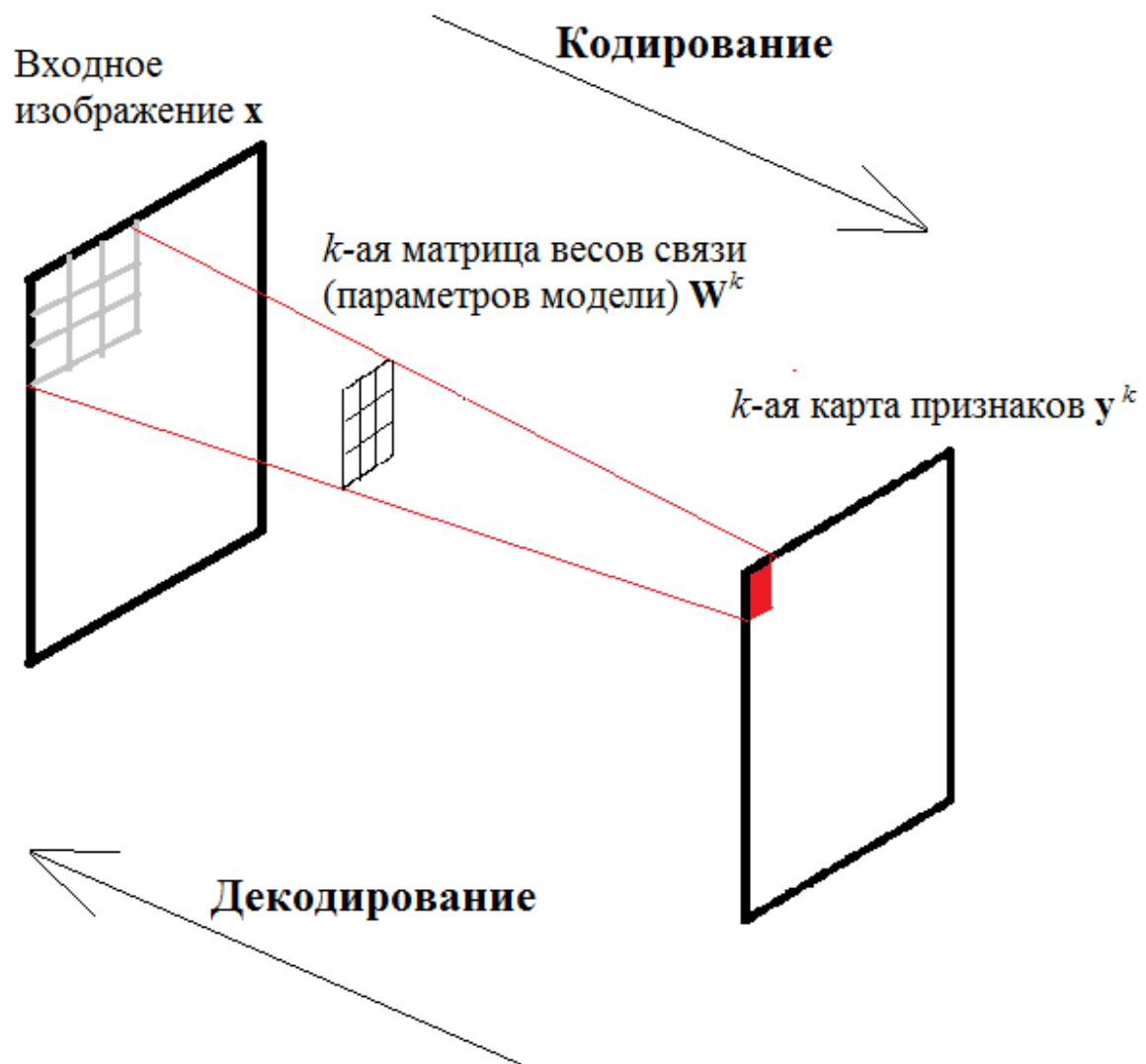


Рисунок 24 – Схема работы сверточного автоэнкодера

Декодирование же осуществляется за счет нелинейной активации поэлементной суммы результатов противосвертки карт признаков с матрицами весов связи. Формально это можно записать так:

$$y^k = \sigma(x * W^k + b^k)$$

$$\mathbf{z} = \sigma\left(\sum_{k \in H} \mathbf{y}^k * \tilde{\mathbf{W}}^k + \mathbf{c}\right),$$

где  $\mathbf{y}^k$  –  $k$ -ая карта признаков,  $\sigma$  – функция-сигмоида,  $\mathbf{x}$  – входное изображение,  $*$  – оператор свертки,  $\mathbf{W}^k$  –  $k$ -ая матрица весов (ядро свертки),  $\mathbf{b}^k$  –  $k$ -ый вектор смещения кодирующего слоя,  $\mathbf{z}$  – результат декодирования (массив размерности исходного изображения),  $\tilde{\mathbf{W}}^k$  –  $k$ -ое ядро свертки, отраженное по вертикали и горизонтали,  $\mathbf{c}$  – смещение декодирующего слоя.

Процесс обучения сверточных автоэнкодеров сводится к нахождению таких ядер свертки (матриц весов связи), которые позволят минимизировать соответствующую функцию ошибки.

Как и для большинства архитектур автоассоциативной памяти при обучении сверточных автоэнкодеров также используется принцип инфомакса.

### 3.4.1. Разреженное кодирование для сверточного автоэнкодера

Поскольку в случае, когда количество карт признаков больше единицы, и признаковое описание изображения будет иметь существенно большую размерность, нежели исходное представление, тогда для получения полезных признаков приходится накладывать дополнительные ограничения на модель сверточного автоэнкодера. Как уже обсуждалось ранее, полезность признаков в случае их переизбытка в результате кодирования по сравнению с количеством признаков в исходном представлении напрямую ассоциируется с их разреженностью в новом пространстве.

Помимо уже упомянутых техник разреженного кодирования для сверточного автоэнкодера существует еще одна полезная и в тоже время простая методика получения разреженного кода. Эту методику можно назвать максимизационным прореживанием (*max-pooling*) [100].

Суть данной техники состоит в том, что на каждой карте признаков происходит сжатие исходного пространства в некоторое количество раз, при этом в получившейся карте меньшего разрешения в ячейках окажутся значения максимальные для области, по которой происходило сжатие.

Иными словами при уменьшении размеров карты признаков в  $n$  раз в уменьшенной карте останутся значения, максимальные внутри смежных областей размером  $n \times n$  исходной карты.

### **3.5. Алгоритмы реконструкции изображений в невидимой области, использующие обучаемые представления системы**

Для решения задачи реконструкции изображений в невидимой области нами предложены несколько вариантов алгоритмов, использующих обучаемые представления системы автоэнкодеров [101].

Первый алгоритм представляет собой итеративную процедуру восстановления, которая может быть описана следующей последовательностью шагов.

1. Обучить модель автоэнкодера, используя выборку изображений (если выборка достаточно мала или сам автоэнкодер имеет сравнительно не большой размер, то лучше использовать изображения принадлежащие одному классу объектов, изображение представителя которого необходимо будет восстанавливать). Этот шаг эквивалентен получению «жизненного» опыта системой, на основании виденных ею до этого изображений.
2. Подать на вход обученному автоэнкодеру изображение, которое необходимо реконструировать, например, изображение с отсутствующим фрагментом.
3. Получить результат реконструкции, то есть результат последовательного применения операций кодирования и декодирования входного изображения.
4. Поврежденные области входного изображения заменить на результаты реконструкции, то есть на те фрагменты выхода автоэнкодера, положение которых соответствует положению фрагментов, подлежащих восстановлению.
5. Изображение с замененными областями идет на вход автоэнкодеру

6. Повторять шаги 3-5 пока не наступит критерий останова, например, на количество итераций.

Заметим, что подобный итеративный алгоритм может быть реализован для большинства модификаций концепции автоэнкодера. Также стоит отметить, что данный алгоритм не гарантирует сходимости к лучшему решению. По всей видимости, такое свойство характерно для большинства итеративных алгоритмов реконструкции, которые можно предложить для различных опорных машин.

Можно модифицировать данный алгоритм, предложив на входе в качестве начальной гипотезы о содержимом восстанавливаемой области не нулевые значения яркости пикселей (0-гипотеза), а, к примеру, некоторую модель шума. Такой подход, в особенности для моделей, обученных шумоподавлению, будет в среднем давать более высокую скорость сходимости по сравнению с 0-гипотезой. Однако всегда существует шанс того, что некоторое случайное заполнение может привести алгоритм в неустойчивые режимы работы и, как следствие, увести его в сторону от лучшего или хотя бы локально-лучшего решения.

Использование случайного заполнения также может иметь трактовку с позиций байесовского вывода. В таком случае начальное шумовое заполнение есть процесс введения априорной плотности распределения яркостей. А сам алгоритм есть процедура байесовского вывода, каждую итерацию дающая новую, уточненную оценку этой априорной-апостериорной вероятности. Особенно показательна такая интерпретация для моделей, обученных шумоподавлению, где в явном виде можно задавать закон распределения яркости пикселей в реконструируемой области, на основе модели шума, использовавшейся при обучении системы.

Учитывая тот факт, что итеративный алгоритм не гарантирует сходимости к лучшему решению, можно предложить алгоритм глобальной оптимизации для реконструкции изображений в невидимой области с использованием обучаемых представлений моделей.



Предлагаемый нами второй алгоритм реконструкции базируется на предположении о том, что правильно восстановленная область на изображении может быть представлена в качестве неподвижной точки  $\mathbf{x}_r$  для некоторой очень сложной многомерной функции  $f_r$ :

$$\mathbf{x}_r : f_r(\mathbf{x}_r) = \mathbf{x}_r.$$

Однако такая функция может обладать не единственной неподвижной точкой. Вследствие этого основная идея алгоритма заключается в применении классической поисковой метаэвристики под названием генетические алгоритмы для поиска наиболее оптимальной неподвижной точки  $\mathbf{x}_r$ .

Соответственно, особями будут являться различные по содержанию фрагменты изображения  $\{\mathbf{x}_r^i\}_{i \in H}$ , подлежащие восстановлению. Фитнесс-функцией в свою очередь может служить, к примеру, оценка среднеквадратичного отклонения (СКО) между входным изображением, содержащим особь, и результатом реконструкции данного изображения автоэнкодером.

### **Выводы по главе 3**

Предложенный в предыдущей главе подход к синтезу методов реконструкции изображений в утерянной или поврежденной области на основе проекции алгоритмической вероятности на представления изображений позволил разработать метод, основанный на критерии информативности спектра изображения, а также методы, способные к обучению представлениям на основе стековых сверточных и несверточных автоэнкодеров.

Методы были воплощены в форме алгоритмов итеративной реконструкции и реконструкции на базе эволюционных стратегий поиска для методов восстановления изображений.

## **Глава 4. Практическое применение алгоритмов реконструкции изображений в невидимой области на основе критерия алгоритмической вероятности.**

### **4.1. Введение**

В предыдущих главах были предложены методы реконструкции изображений в невидимой области на базе критерия алгоритмической вероятности, который предоставляет возможности для объективной оценки качества результатов реконструкции, а также были рассмотрены алгоритмы, предлагаемые в качестве воплощений описанных методов. В данной главе будут представлены результаты применения предлагаемых алгоритмов для различных изображений.

### **4.2. Практическое применение алгоритма реконструкции изображений на основе критерия информативности их спектров**

Рассмотрим применение алгоритма реконструкции изображений на базе спектрального представления. Как уже отмечалось ранее в главе 1, одними из ключевых требований, предъявляемых к алгоритмам реконструкции изображений в невидимой области, являются требования восстановления структурной и текстурной информации внутри отсутствующих фрагментов. Тем не менее, зачастую разработчики такого рода алгоритмов демонстрируют результаты работы своих методов на изображениях, подвергнутых структурным искажениям на подобие тех, что приведены на рисунке 25.

Стоит заметить, что подобного рода структурные искажения не позволяют в полной мере оценить способность того или иного алгоритма к восстановлению утерянной текстурной или структурной информации. Несмотря на то, что общая площадь таких искажений может быть достаточно большой, специфическая форма каждого из локальных участков, содержащих искажения, как правило, имеет достаточно малую площадь.



Рисунок 25 – Пример структурных искажений на изображении в виде текста нанесенного поверх. Информацию внутри участков, заслоненных текстом, требуется восстановить

При этом еще раз стоит оговориться, что обычно под оценкой качества понимают субъективное восприятие результатов реконструкции конкретным наблюдателем или экспертом. На рисунках 26 и 27 показаны результаты реконструкции изображения, приведенного на рисунке 25, при помощи одного из наиболее распространенных на текущий момент алгоритмов реконструкции, содержащихся в программных библиотеках и графических редакторах, а также предлагаемого в данной работе алгоритма реконструкции изображений на основе спектрального представления соответственно.

При визуальной оценке представленных результатов будет достаточно сложно найти какие-либо различия, поэтому есть подозрение, что для корректного сравнения восстанавливающих способностей алгоритмов требуются более подходящие примеры.



Рисунок 26 – Пример результатов работы одного из алгоритмов реконструкции изображений в невидимой области на базе дифференциальных уравнений в частных производных. Визуально результаты воспринимаются наблюдателями как удовлетворительные. Но в силу малой площади локальных областей повреждений сложно что-либо сказать о характере текстурной или структурной информации внутри восстановленных участков



Рисунок 27 – Пример результатов работы предлагаемого в данной диссертации алгоритма реконструкции изображений на основе спектрального представления (то есть использующего в качестве опорной машины обратное преобразование Фурье). Визуально результаты воспринимаются наблюдателями как удовлетворительные. Но в силу малой площади локальных областей повреждений также как и для рисунка 26 сложно что-либо сказать о характере текстурной или структурной информации внутри восстановленных участков

Теперь рассмотрим несколько менее тривиальный, а также более характерный и показательный пример изображения для реконструкции, который позволит оценить качество восстановления структурной информации внутри отсутствующего участка. Такой пример представлен на рисунке 28.

Можно увидеть, что на данном изображении отсутствует область, которая должна соответствовать месту пересечения мачты и реи, изображенного на фотографии корабля.



Рисунок 28 – Пример изображения содержащего область повреждения, которая требует заполнения структурной информацией о пресечении реи и мачты

Такой отсутствующий фрагмент при правильной реконструкции должен быть заполнен информацией о структуре пересечения между двумя объектами, представляющими собой рею и мачту соответственно.

Рассмотрим результат применения к изображению, представленному на рисунке 28, алгоритма реконструкции изображений на основе спектрального представления, предлагаемого в данной работе. Этот результат отражен на рисунке 29.

Можно заметить, что в результате работы алгоритма структура области пресечения мачты и реи восстановлена успешно с точностью до тени, падающей на мачту от реи, что вызвано условиями освещения при съемке, и, конечно же, вряд ли может быть предсказано при помощи данной опорной машины, а также, возможно, доступной на изображении информации. Хотя последний довод может показаться сомнительным в силу того, что на другой области пересечения мачты и реи, которая расположена на изображении ниже, мы все же можем увидеть тень и каким-то образом пытаться строить

предположения на счет наличия или отсутствия тени внутри восстанавливаемой области.

Но здесь сразу можно привести контраргумент, который будет касаться интерпретации наблюдаемой на фрагменте ниже тени (является ли эта тень в действительности тенью, или может это полоса покраски). Правда подобное обсуждение носит сугубо умозрительный характер и выходит за пределы данной работы.



Рисунок 29 – Слева: исходное изображение. Справа: результата реконструкции исходного изображения с отсутствующей областью (рисунок 28) при помощи алгоритма, основанного на информативности спектра. Структурная информация восстановлена успешно с точки зрения визуального восприятия

Для большей наглядности ниже приведен увеличенный вариант реконструированной области (рисунок 30). При помощи этого рисунка проще и удобнее оценить результат реконструкции отсутствовавшей области с точки зрения визуального восприятия.

Визуально кажется, что помимо восстановления структуры перекрестия между мачтой и реей алгоритму также удалось «угадать» яркости данной структуры и области вокруг нее.

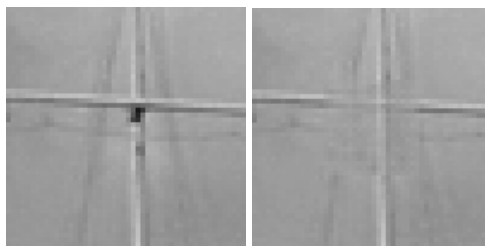


Рисунок 30 – Слева: увеличенный участок исходного изображения соответствующий структуре пересечения мачты и реи корабля. Справа: соответствующий увеличенный участок изображения, полученного в результате реконструкции

Но, как уже неоднократно отмечалось ранее, визуальное восприятие носит крайне субъективный характер и может существенным образом различаться от одного наблюдателя к другому. Поэтому предлагается сопоставить результатам некоторую объективную меру качества реконструкции на основе алгоритмической вероятности этих результатов.

Здесь для удобства отображения численных характеристик результатов мы будем пользоваться не самой оценкой алгоритмической вероятности, а отрицательно логарифмированной величиной от данной оценки:

$$Compl = -\log(P_S(g|_R | f|_{R'})) \cong l(f|_{R'} g|_R) \propto H(F) \cdot N,$$

что для опорной машины, выполняющей обратное преобразование Фурье, выразится в виде произведения энтропии спектра  $H(F)$  и линейного размера  $N$  области пространственных частот, который в случае дискретного преобразования Фурье совпадает с размерами исходного изображения.

Для приведенного на рисунке 30 примера оценка значения  $Compl$  будет равняться 471799,36. При этом разница между начальным значением оценки  $Compl_{init}$  для изображения, требующего реконструкции (рисунок 28), и значением оценки  $Compl_{result}$ , полученной в результате работы алгоритма будет составлять:

$$Compl_{init} - Compl_{result} = 472023 - 471799,4 = 223,6$$

При этом в процессе работы алгоритма могут наблюдаться следующие промежуточные результаты (рисунок 31), которые возникают на различных



итерациях процедуры реконструкции описанной в разделе 3.2 главы 3 настоящей работы.

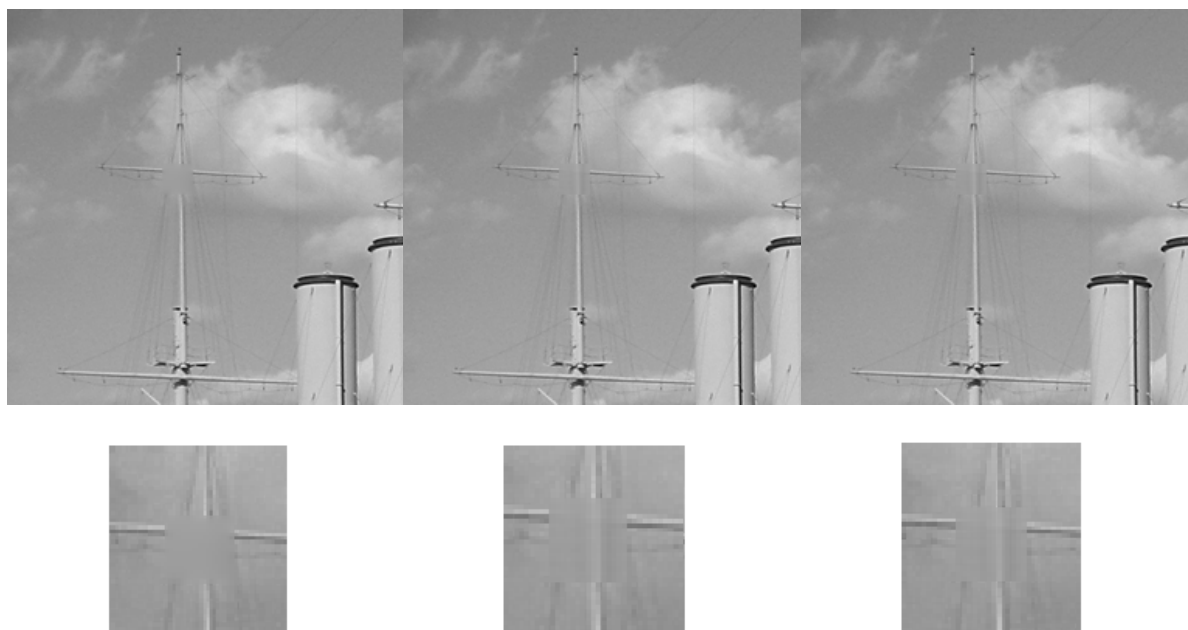


Рисунок 31 – Промежуточные результаты работы алгоритма реконструкции изображений в невидимой области на основе критерия информативности спектра. Верхний ряд, слева направо: результаты, полученные на первой, третьей и пятой итерациях, соответственно. Нижний ряд, слева направо: увеличенные фрагменты изображений сверху, содержащие области реконструкции. Оценка алгоритмической сложности *Compl* для этих изображений будет (слева направо): 471997; 471931; 471806

Из рисунка 31 видно, что ростом качества реконструкции при визуальном восприятии значения оценки *Compl* неизменно падают, то есть оценка алгоритмической вероятности от итерации к итерации растет.

Для сравнения рассмотрим результаты работы на данном примере двух наиболее популярных алгоритмов реконструкции, которые можно найти в программных библиотеках для обработки изображений общего и специального назначения, например, в библиотеке OpenCV. Эти два алгоритма – быстрый маршевый метод (Fast Marching Method – сокр. FMM) [29] и алгоритм, использующий уравнения Навье-Стокса [102]. Результаты применения этих алгоритмов к рассматриваемому на рисунке 28 примеру изображены на рисунке 32.

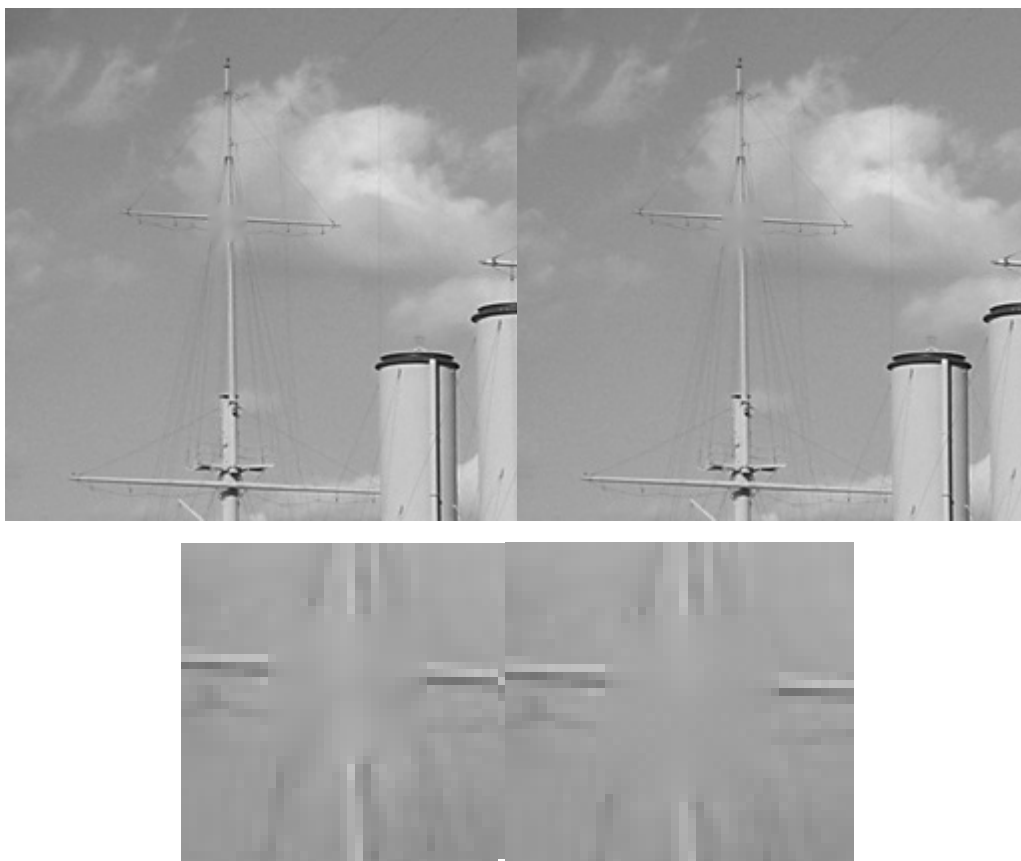
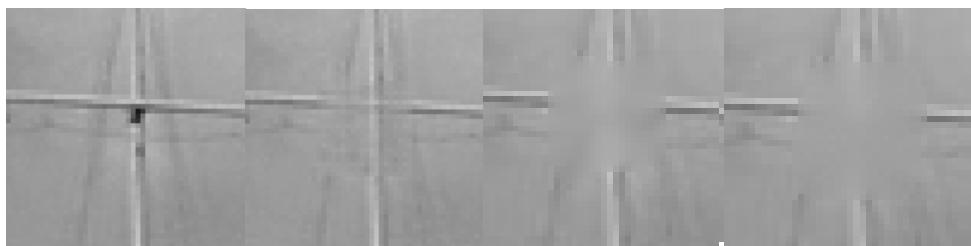


Рисунок 32 – Результаты реконструкции изображения на рисунке 28 по алгоритму, использующему уравнений Навье-Стокса, (слева) и по быстрому маршевому методу (справа). Снизу приведены увеличенные фрагменты изображений сверху, содержащие области реконструкции

Можно уже заметить, что подобные результаты даже при визуальной оценке качества реконструкции вряд ли получится назвать удовлетворительными. Естественно будет попытаться дать этим результатам оценку при помощи меры алгоритмической сложности *Compl*. Окажется, что оценка алгоритмической сложности для спектрального представления окажется меньшей для результатов, полученных с помощью предлагаемого в данной работе алгоритма, нежели для обоих библиотечных методов. Наглядно этот факт продемонстрирован на рисунке 33. Подобный результат совпадает не только с теоретическими ожиданиями, но и с нашим субъективным пониманием качества реконструкции.



471793,7      471799,4      471995      471996

Рисунок 33 – Слева направо: исходное изображение, результат реконструкции алгоритмом на базе спектрального представление, результат реконструкции алгоритмом на базе уравнений Навье-Стокса, результат реконструкции алгоритмом на базе быстрого маршевого метода. Ниже для каждого из результатов дана соответствующая оценка алгоритмической сложности с учетом опорной машины в виде обратного преобразования Фурье

Правда, стоит отметить, что главным преимуществом алгоритмов на базе уравнений Навье-Стокса и быстрого маршевого метода является их быстроедействие. Целью же данной работы была попытка выявить объективный критерий для количественной оценки результатов реконструкции и систематизировать при помощи данного критерия подходы к реализации алгоритмов реконструкции, дав для этого единую теоретическую базу. Поэтому, в частности, предлагаемый в данной работе алгоритм, базирующийся на критерии информативности спектра, существенно уступает библиотечным алгоритмам в быстроедействии, но позволяет получить качественно более высокие результаты.

Для иллюстрации целенаправленности процесса поиска, используемого в предлагаемом алгоритме на базе информативности спектра, можно привести график сходимости, который изображен на рисунке 34. Из графика видно, что алгоритм за достаточно малое количество итераций сходится к некоторому устойчивому решению, которое, еще раз заметим, не обязательно будет являться глобально оптимальным.

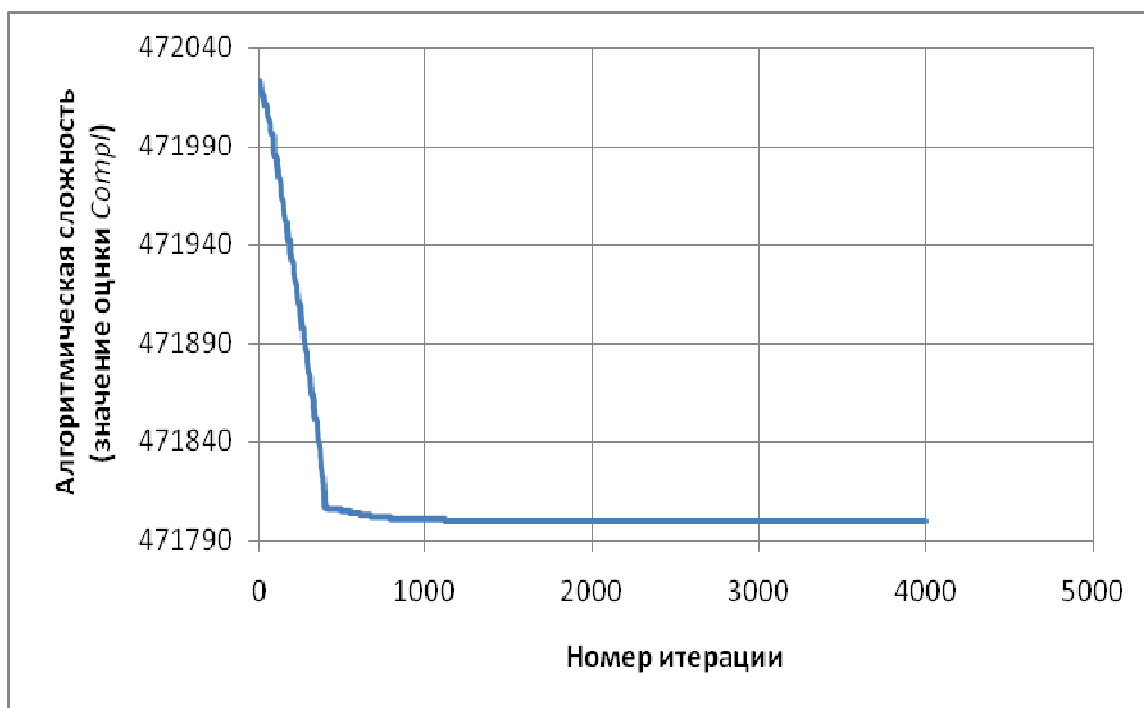


Рисунок 34 – График сходимости для предлагаемого алгоритма на основе критерия информативности спектра изображения

Как можно было убедиться, алгоритм, использующий достаточно несложное спектральное представление изображений, оказывается способным к восстановлению структурной информации внутри отсутствующих участков. Теперь посмотрим на способности данного алгоритма к восстановлению текстурной информации. Для этого возьмем следующий характерный пример (рисунок 35).

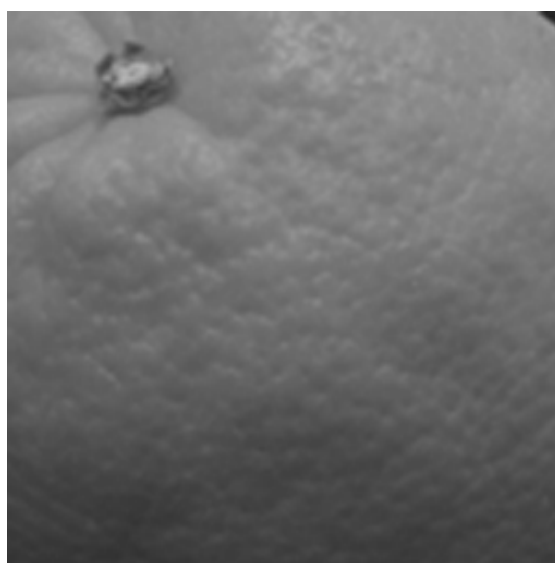


Рисунок 35 – Пример изображения для оценки качества текстурной реконструкции

На рисунке 36 справа показан результат реконструкции изображения представленного на рисунке 35, но только с удаленным фрагментом (рисунок 36 слева).

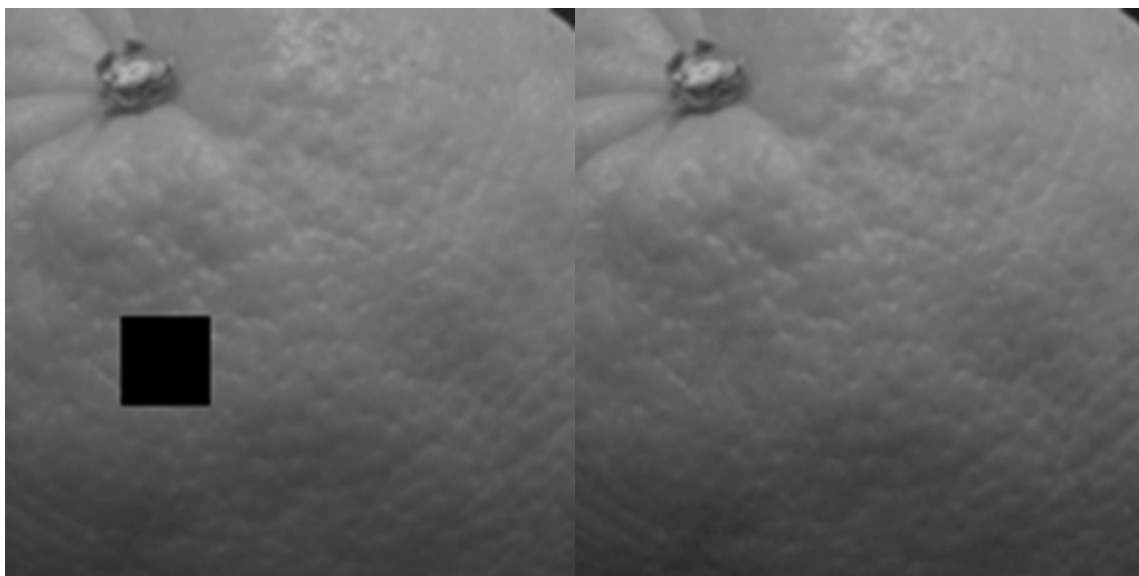


Рисунок 36 – Слева: изображение с отсутствующей текстурной областью, которое требуется реконструировать. Справа: результат реконструкции изображения слева при помощи предлагаемого алгоритма на основе информативности спектра изображения

Очевидно, что даже при оценке подобного результата с точки зрения визуального восприятия, такая реконструкция кажется вполне удовлетворительной. Опять же для сравнения приведем результаты работы двух библиотечных алгоритмов, о которых уже говорилось ранее. Примеры соответствующих результатов работы алгоритмов на основе уравнений Навье-Стокса, а также быстрого маршевого метода по отношению к изображению представленному на рисунке 36 слева можно увидеть на рисунке 37.

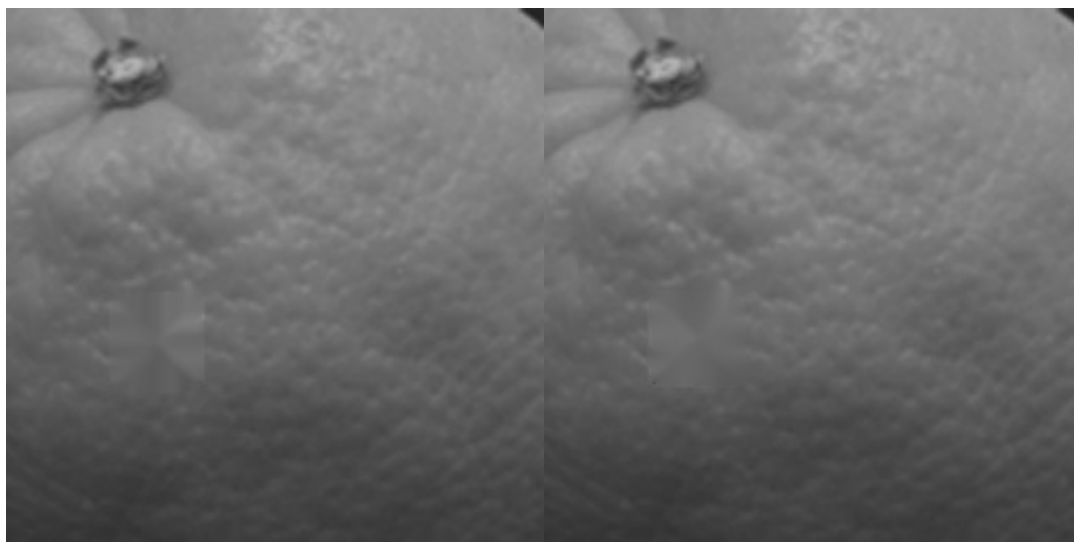


Рисунок 37 – Результаты реконструкции изображения на рисунке 36 слева при помощи алгоритма на основе уравнений Навье-Стокса (слева) и алгоритма на основе быстрого маршевого метода (справа)

Из рисунка 37 видно, что оба стандартных алгоритма не справились с задачей восстановления текстуры. Об этом же говорят цифры характеризующие оценки алгоритмической сложности *Compl* для всех трех результатов. Соответствующие оценки значения *Compl* для текстурной реконструкции изображения, представленного на рисунке 36 слева, алгоритмами на основе уравнений Навье-Стокса, быстрого маршевого метода, а также критерия информативности спектра сведены в таблицу 1.

Таблица 1. Сравнительная оценка алгоритмической сложности для результатов реконструкции текстурной информации, выполненной различными алгоритмами.

Алгоритм на основе уравнений Навье-Стокса	Алгоритм на основе быстрого маршевого метода	Алгоритм на основе информативности спектра изображения
451507,8	451069,6	370124

Нет сомнений, что подобная численная оценка имеет тесную корреляцию с визуальным восприятием результатов реконструкции

наблюдателями, но для выявления конкретной зависимости требуются дальнейшие исследования с привлечением фокус-групп.

В качестве альтернативы оценке алгоритмической вероятности (например, через спектральную алгоритмическую сложность) можно попробовать рассмотреть критерий СКО для оценки качества результатов реконструкции. Стоит сразу оговориться, что данный критерий будет возможно применить только лишь в случаях, когда известно исходное содержимое внутри отсутствующих на изображении областей.

Соответствующие оценки СКО для результатов реконструкции изображения на рисунке 28 приведены в таблице 2. Такая же оценка, но уже для результатов реконструкции изображения на рисунке 36 слева приведена в таблице 3.

Таблица 2. Сравнительная оценка СКО для результатов реконструкции структурной информации для изображения на рисунке 28, выполненной различными алгоритмами.

Алгоритм на основе уравнений Навье-Стокса	Алгоритм на основе быстрого маршевого метода	Алгоритм на основе информативности спектра изображения
5,4509	6,2356	7,2180

Таблица 3. Сравнительная оценка СКО для результатов реконструкции текстурной информации для изображения на рисунке 28, выполненной различными алгоритмами.

Алгоритм на основе уравнений Навье-Стокса	Алгоритм на основе быстрого маршевого метода	Алгоритм на основе информативности спектра изображения
3,3414	2,0304	7,3328

Как можно увидеть из таблиц 2 и 3, критерий СКО имеет мало общего с интуитивным понятием визуально «правильной», корректной или

качественной реконструкции изображения. Наоборот, данный критерий дает наилучшие оценки для визуально наименее приемлемых результатов, как при восстановлении текстурной, так и при восстановлении структурной информации на изображении.

#### **4.3. Практическое применение алгоритмов реконструкции изображений на основе использования систем, способных к обучению представлениям**

Теперь рассмотрим применение предлагаемого в данной диссертации алгоритма реконструкции, использующего системы способные к обучению представлениям.

Возьмем следующий пример. Пускай у нас имеется изображение лица (рисунок 38 слева), отсутствующий фрагмент которого требуется восстановить. Допустим, что этим фрагментом оказался глаз (рисунок 38 в центре). В таком случае алгоритм, использующий критерий информативности спектра, даст результат подобный тому, что изображен на рисунке 38 справа.



Рисунок 38 – Слева: исходное изображение. В центре: изображение с отсутствующей областью. Справа: результат реконструкции алгоритмом, использующим критерий информативности спектра. Изображения лиц взяты из базы изображений бразильского университета FEI, которая находится в открытом доступе [103]

Несмотря на использование алгоритмом критерия алгоритмической вероятности, недостаточная выразительная сила используемого в основе алгоритма представления не позволяет осуществить детальную



реконструкцию отсутствующего фрагмента на изображении. Хотя стоит отметить, что даже в этом случае в результате реконструкции угадывается нечто похожее на сильно размытые очертания глаза.

Куда более тяжелым случаем для такого алгоритма будет являться попытка реконструкции изображения представленного на рисунке 39. В данном случае информации, содержащейся на изображении, явным образом не будет хватать для осуществления хотя бы какой-нибудь реконструкции. Также ситуация осложняется тем, что данное изображения является бинарным.



Рисунок 39 – Изображение рукописной цифры (слева), отсутствующий фрагмент которой требуется восстановить (справа)

Соответственно, для решения подобных задач требуется привлечение более мощных с точки зрения выразительной силы представлений изображений. В роли таких представлений, как уже упоминалось в предыдущих главах, могут выступить сети глубокого обучения.

Рассмотрим применение итеративного алгоритма, описанного в разделе 3.5 главы 3 настоящей диссертации, для реконструкции изображений рукописных цифр из базы MNIST (примеры на рисунках 39 и 40). Для этого был реализован стэковый автоэнкодер с четырьмя скрытыми слоями, которым соответствовало 150, 100, 50 и 20 нейронов, а также стэковый автоэнкодер с тремя скрытыми слоями, которым соответствовало 150 и 50 нейронов. Входной слой состоял из вектора признаков равного линейному размеру изображения рукописной цифры, что составляло  $28 \times 28 = 784$  нейрона. Результаты приведены на рисунке 41. Видно, что с увеличением объема обучающей выборки качество получаемых результатов растет.

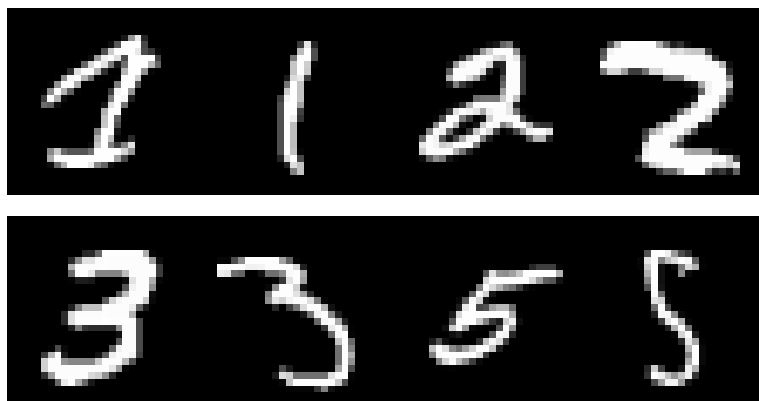


Рисунок 40 – Примеры изображений из базы рукописных цифр MNIST

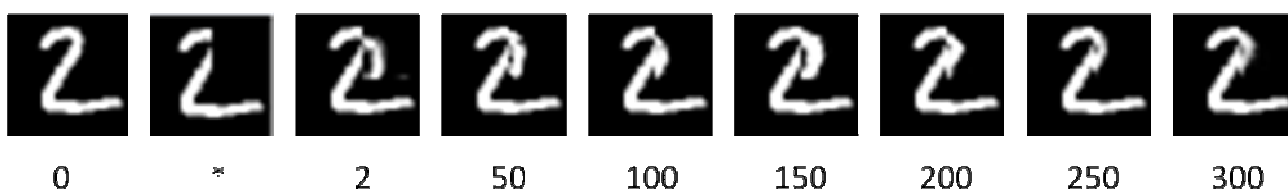


Рисунок 41 – Влияние количества образов в обучающей выборке на качество результата реконструкции. «0» обозначает исходное изображение; «\*» - изображение с удаленным фрагментом, требующее реконструкции; остальные числа соответствуют количеству образов в обучающей выборке, которая использовалась для настройки параметров автоэнкодера, чтобы получить результат реконструкции в виде изображения, расположенного над числом. Результаты получены при помощи итеративного алгоритма реконструкции для автоэнкодера с двумя скрытыми слоями

Теперь сравним результаты, полученные с помощью итерационного алгоритма, с результатами, полученными при помощи алгоритма на базе эволюционных стратегий поиска. Соответствующее сравнение представлено на рисунке 42 для одного исходного изображения, но для двух различных по размеру областей реконструкции.

Можно заметить, что алгоритм на базе эволюционных стратегий поиска на выходе дает более удовлетворительные и «осмысленные» с точки зрения визуального восприятия результаты.

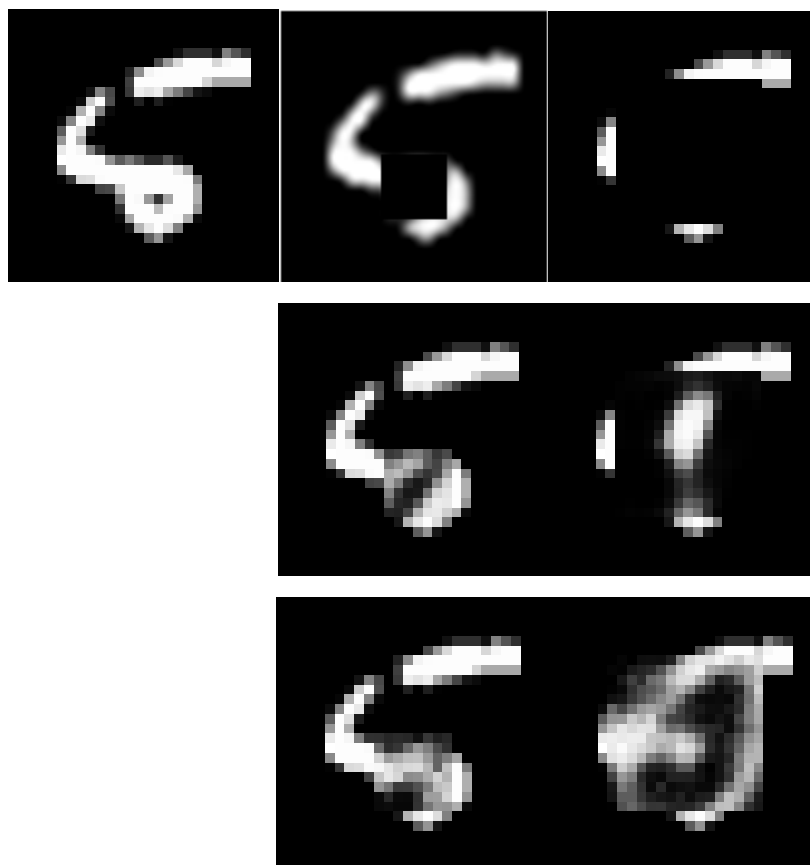


Рисунок 42 – Верхний ряд слева направо: исходное изображение, изображение с вырезанным малым фрагментом, изображение с вырезанным большим фрагментом. Средний ряд слева направо: результаты реконструкции малого и большого фрагментов при помощи предлагаемого итеративного алгоритма. Нижний ряд слева направо: результат реконструкции малого и большого фрагментов при помощи предлагаемого алгоритма на базе эволюционных стратегий поиска. Все результаты получены для одинакового размера обучающей выборки равного 200 элементам на автоэнкодере с двумя скрытыми слоями

Далее попытаемся оценить влияние степени абстрактности выучиваемых системой признаков через увеличение количества скрытых слоев в стэковом автоэнкодере на результаты реконструкции изображений лиц (примеры изображений приведены на рисунке 43). Результаты реконструкции изображения лица, полученные разными алгоритмами с различными по количеству скрытых слоев опорными машинами (автоэнкодерами) представлены на рисунке 44.



Рисунок 43 – Примеры изображений из базы фотографий лиц FEI, использованной при обучении [103]



Рисунок 44 – Верхний ряд слева направо: исходное изображение, изображение с отсутствующей областью, результат реконструкции итерационным алгоритмом с трехслойным автоэнкодером. Нижний ряд слева направо: результат реконструкции итерационным алгоритмом на базе автоэнкодера с пятью слоями, результат реконструкции генетическим

алгоритмом с трехслойным автоэнкодером, результат реконструкции генетическим алгоритмом на базе автоэнкодера с пятью слоями

Визуально, кажется, что наиболее приемлемый результат достигается в случае использования наиболее абстрактных признаков наряду с эволюционной стратегией поиска. Хотя, стоит оговориться, что полученные результаты весьма проблематично сравнивать между собой из-за сложностей связанных с оценкой «компактности» нейронных сетей. Но можно установить совершенно достоверно то, что увеличение степени абстрактности извлекаемых из изображения признаков ведет к иным результатам реконструкции.

Теперь рассмотрим применение к решению задачи реконструкции изображений специальной архитектуры нейронной сети под названием сверточный автоэнкодер [104-106]. Особенностью процесса реконструкции на базе сверточного автоэнкодера является то, что восстановление происходит на основе локальной информации об изображении вокруг обрасти реконструкции, но при помощи ядер свертки, или иначе масок (рисунок 45), которые содержат глобальную информацию о структуре изображений выученных сверточным автоэнкодером. Тем самым процесс реконструкции становится схожим с работой реставратора, который применяет весь свой опыт к восстановлению изображений. Примеры результатов реконструкции изображении лиц при помощи итерационного алгоритма на базе сверточного автоэнкодера представлены на рисунках 46-49.

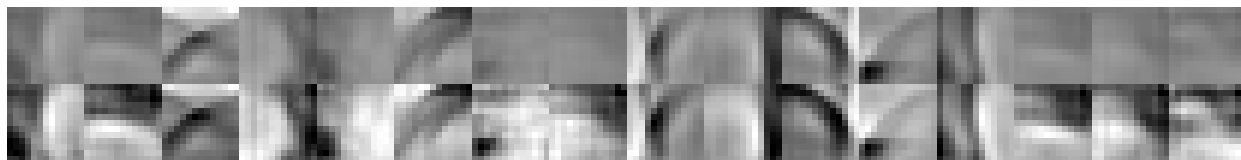


Рисунок 45 – Пример результата визуализации выученных сверточным автоэнкодером масок при одновременном тушении весов, обучении в стиле подавления шума и использовании техники максимизационного

прореживная. Верхний ряд – изображение масок, нормализованных по общему для всех масок максимуму, нижний ряд – независимая нормировка

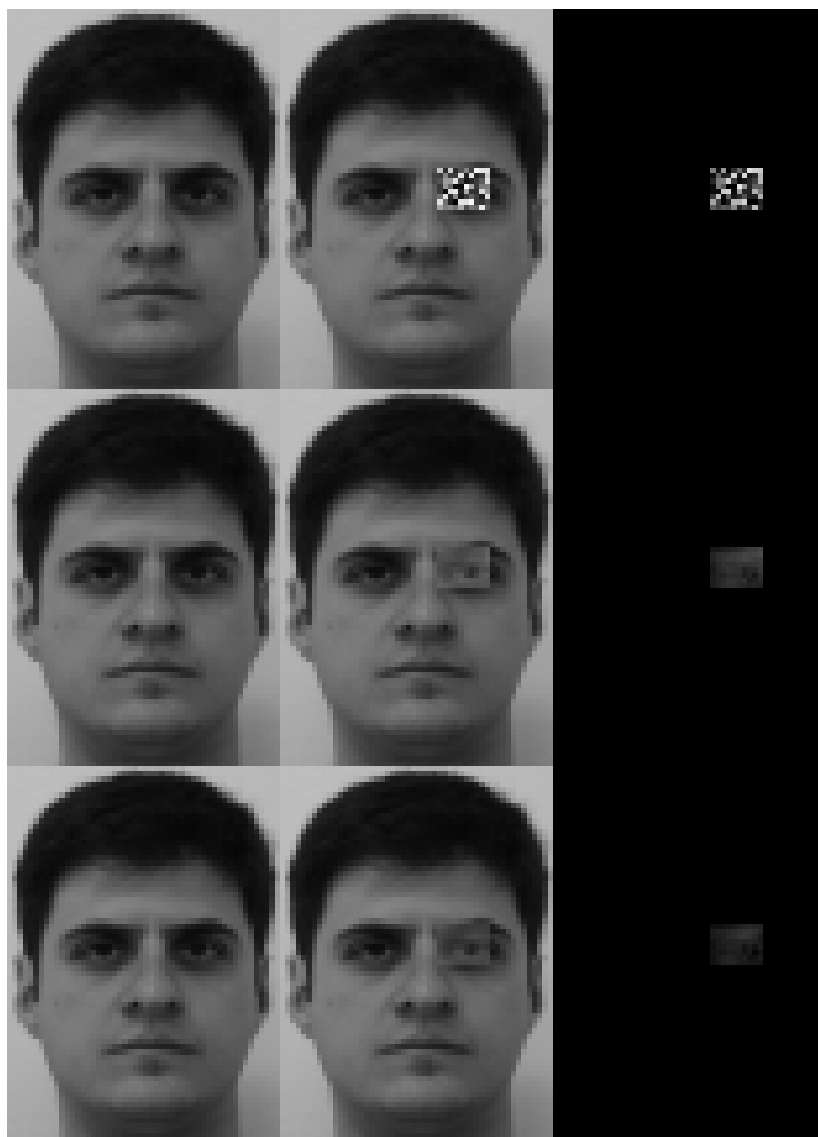


Рисунок 46 – Результат работы итерационного алгоритма реконструкции вырезанной области изображения с использованием сверточного автоэнкодера, обученного шумоподавлению с уровнем шума 0.5. Здесь и далее, левый столбец – неповрежденное изображение, средний столбец – результат работы метода на разных итерациях алгоритма (1, 10, 100), правый столбец – разница первых двух

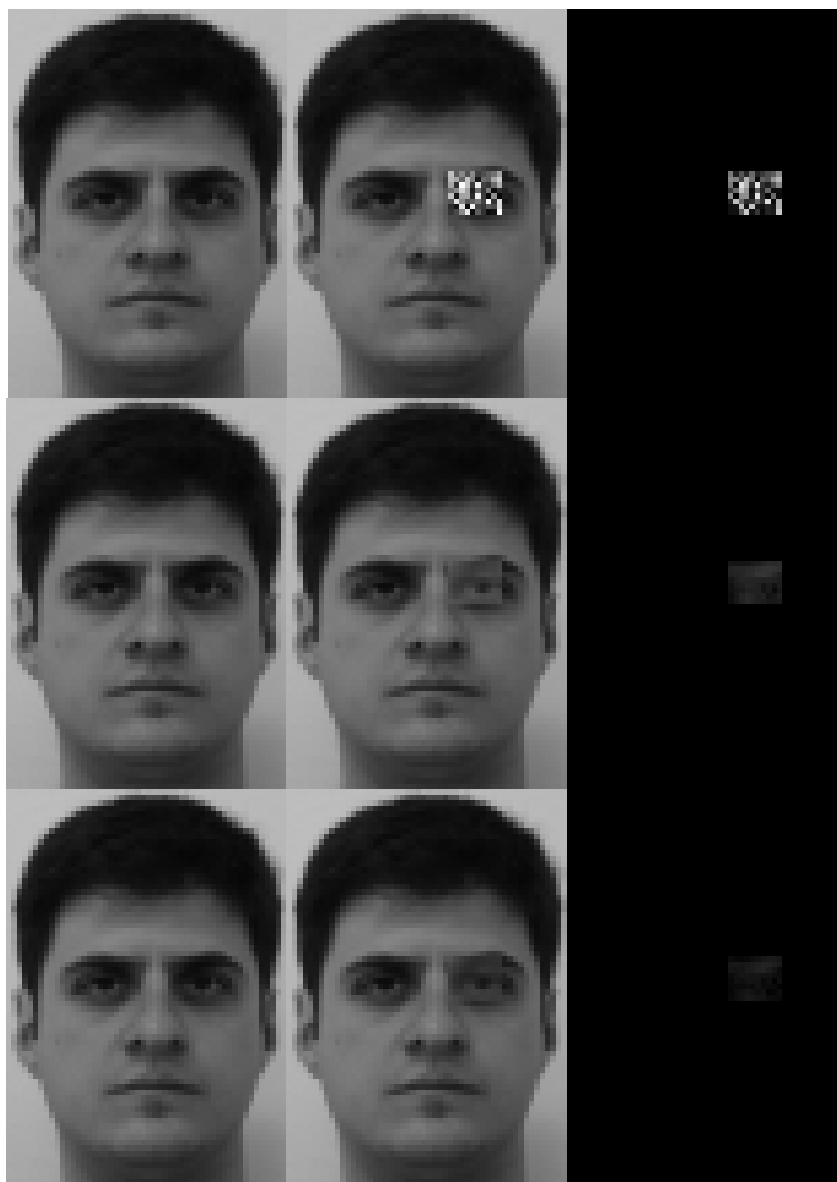


Рисунок 47 – Результат работы итерационного алгоритма реконструкции вырезанной области изображения с использованием сверточного автоэнкодера, обученного шумоподавлению с уровнем шума 0.7

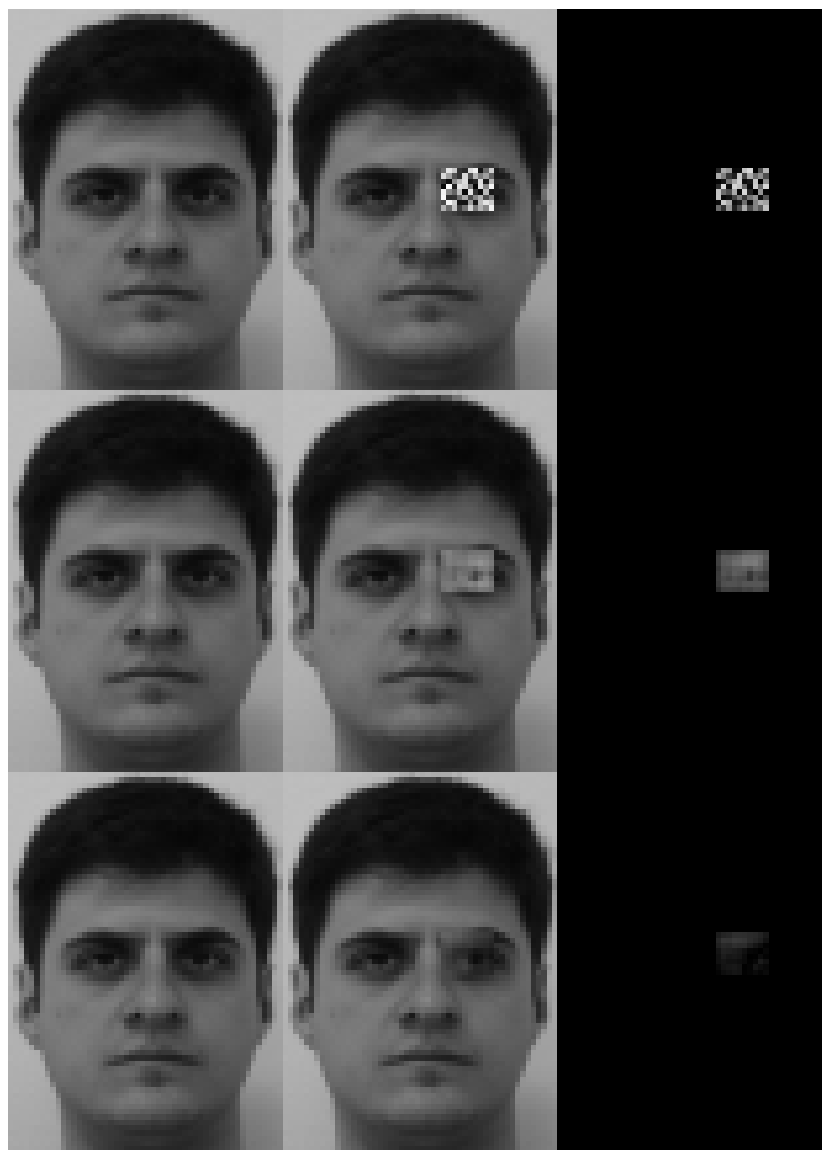


Рисунок 48 – Результат работы итерационного алгоритма реконструкции вырезанной области изображения с использованием следующих модификаций сверточного автоэнкодера: обучение шумоподавлению, тушение весов связей и максимизационное прореживание в окне размером  $2 \times 2$



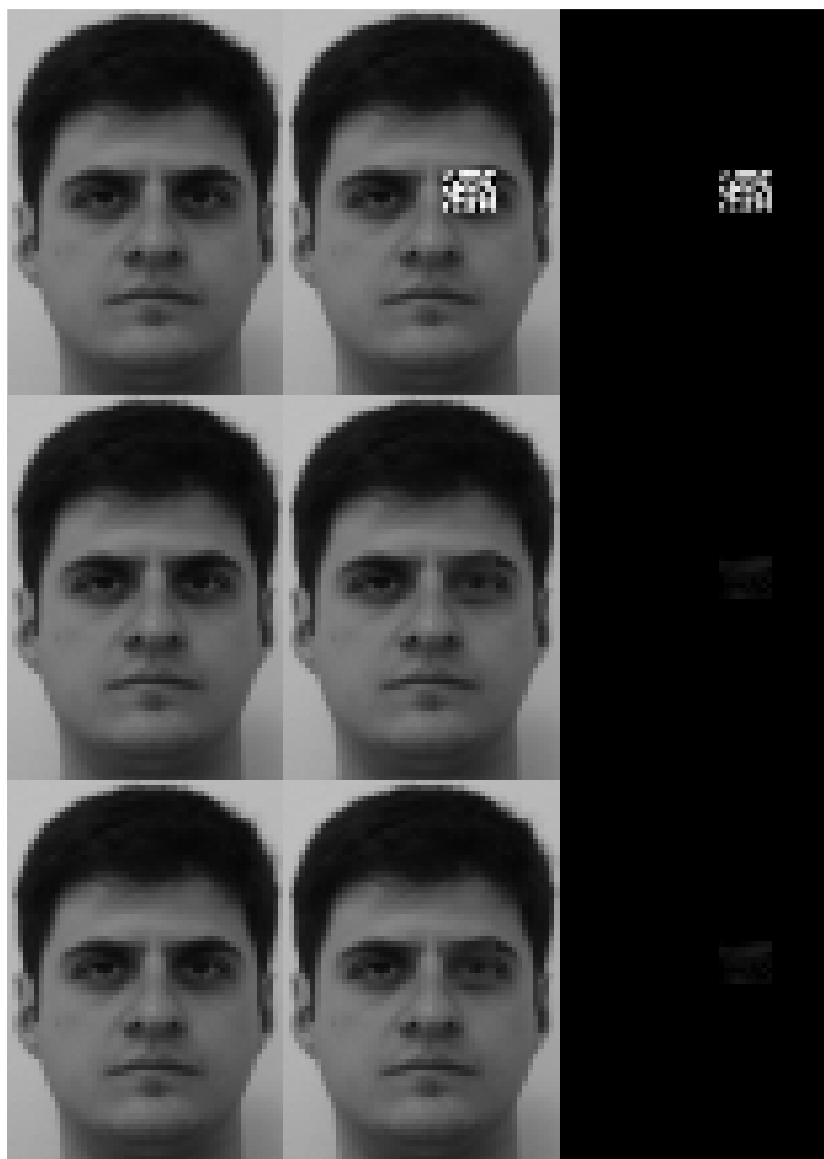


Рисунок 49 – Результат работы итерационного алгоритма реконструкции вырезанной области изображения для сверточного автоэнкодера с линейным выходом и среднеквадратичной ошибкой обучения, с шумоподавлением, тушением весов связей

Вырезанная область инициализировалась той же моделью шума, которая использовалась при обучении системы. Все результаты, представленные на изображениях 46-49, были получены при одинаково малом размере обучающей выборки, которая состояла из 10 изображений лиц. Тем не менее, в результате работы алгоритму удалось получить достаточно неплохие качественные оценки содержимого области реконструкции.

## **Выводы по главе 4**

В результате экспериментальных исследований была показана несостоятельность критерия среднеквадратичного отклонения для оценки качества результатов реконструкции. В то же время было показано наличие согласованности визуальной оценки качества результата реконструкции и значения критерия алгоритмической вероятности (точнее, ее проекции на тот или иной тип представления изображений) данного результата.

Экспериментально показано, что метод реконструкции изображений в утерянной или поврежденной области, основанный на критерии информативности его спектра, способен к восстановлению структурной и текстурной информации.

Показано, что алгоритмы реконструкции, основанные на обучаемых представлениях, способны заполнять утерянные области на изображении релевантной информацией (например, «дорисовать» глаз человека в соответствующей области на изображении его лица), которую общие методы реконструкции (такие, как методы на основе дифференциальных уравнений) восстановить по самому изображению не могут. В этом данные методы похожи на методы на основе поиска экземпляра, но при этом они не требуют наличия в обучающей выборки изображений реконструируемой сцены, а выполняют перенос из выборки информации через более высокие уровни абстракции ее представления.

## Заключение

В ходе исследований были достигнуты следующие результаты.

1. Выполнен аналитический обзор предметной области восстановления изображений в утерянных или поврежденных областях. При этом выделены методы на основе типа используемых в них представлений изображений (низкоуровневые, текстурные, структурные), а также методы на основе поиска экземпляра и гибридные методы. Показано, что существенным ограничением в развитии этих методов является отсутствие объективного критерия оценки качества реконструкции, что мешает корректному одновременному учету различных аспектов содержания изображений (в частности, текстурного и структурного).
2. Показано, что задача реконструкции изображений в невидимой области может быть поставлена как частный случай проблемы универсального предсказания, и выведен критерий качества решения этой задачи на основе алгоритмической вероятности. Разработаны частные критерии алгоритмической вероятности для случаев спектральных представлений изображений, а также обучаемых представлений изображений на основе стека автоэнкодеров.
3. Разработаны методы реконструкции, основанные на критерии информативности спектра изображения, а также на основе обучаемых представлений на базе стековых сверточных и несверточных автоэнкодеров. Методы были воплощены в форме алгоритмом итеративной реконструкции и реконструкции на базе эволюционных стратегий поиска для методов восстановления изображений.
4. Экспериментально показаны:
  - несостоятельность критерия среднеквадратичного отклонения для оценки качества результатов реконструкции и согласованность визуальной оценки качества результата

реконструкции и значения критерия алгоритмической вероятности;

- способность метода реконструкции изображений в утерянной или поврежденной области, основанного на критерии информативности его спектра, к восстановлению структурной и текстурной информации;
- способность алгоритмов реконструкции, основанных на обучаемых представлениях, заполнять утерянные области на изображении релевантной информацией.

## Список литературы

1. Bertalmio M. Image inpainting [Text] / M. Bertalmio, G. Sapiro, V. Caselles and C. Ballester // In Proceedings of ACM Conf. Comp. Graphics (SIGGRAPH), –2000. – P. 417-424.
2. Criminisi A. Region filling and object removal by exemplar-based inpainting [Text] / A. Criminisi, P. Perez and K. Toyama // IEEE Transactions on Image Processing. – 2004. – Vol.13. – No.9. – P. 1200-1212.
3. Vijay Venkatesh M. Cheung. Symmetric shape completion under severe occlusions [Text] / M. Vijay Venkatesh and S. S. Cheung // International conference on image processing (ICIP), – 2006. – P. 7009-7012.
4. Elazary L. Interesting objects are visually salient [Text] / L. Elazary and L. Itti // Journal of Vision. – 2008. – Vol.8. – No.3. – P. 1-15.
5. Itti L. Bayesian surprise attracts human attention [Text] / L. Itti and P. F. Baldi // In Advances in Neural Information Processing Systems. – 2006. – Vol. 19. – P. 547-554.
6. Vijay Venkatesh M. Eye tracking based perceptual image inpainting quality analysis [Text] / M. Vijay Venkatesh and Sen ching S. Cheung // International conference on image processing (ICIP), – 2010.
7. Efros A. Texture synthesis by non-parametric sampling [Text] / A. Efros and T. K. Leung // In IEEE International conference on Computer Vision (ICCV), Corfu, Greece, – 1999. – P. 1033-1038.
8. Efros A. Image quilting for texture synthesis and transfer [Text] / A. Efros and W.T. Freeman // In Proceedings of ACM Conf. Comp. Graphics (SIGGRAPH), – 2001. – P. 341-346.
9. Harrison P. A non-hierarchical procedure for re-synthesis of complex texture [Text] // In WSCG Winter School of Computer Graphics Conf. Proc. (WSCG), Univ. of West Bohemia, – 2001. – P. 190-197.

10. Heeger D.J. and Bergen J.R. Pyramid-based texture analysis/synthesis [Text]  
// In Proceedings of ACM Conf. Comp. Graphics (SIGGRAPH), –1995. –  
Vol.29. – P. 229-233.
11. Ashikhmin M. Synthesizing natural textures [Text] // In Proceedings of  
ACM Symposium on Interactive 3D Graphics, – 2001. – P. 217-226.
12. Igehy H. and Pereira L. Image replacement through texture synthesis [Text]  
// In Proceedings of International Conference on Image Processing (ICIP), –  
1997. – Vol.3. – P. 186-190.
13. Yamauchi H. Image restoration using multiresolution texture synthesis and  
image inpainting [Text] / H. Yamauchi, J. Haber, and H.-P. Seidel // In  
Computer Graphics International (CGI 2003), Tokyo, Japan, IEEE. – 2003.  
– P. 120-125.
14. Fang C.W. and Lien J.J.J.. Fast image replacement using multi-resolution  
approach [Text] // Lecture Notes in Computer Science. – 2006. – Vol. 3852.  
– P. 509-520.
15. Chan T. and Shen J. Mathematical Models for Local Nontexture Inpaintings  
[Text] // SIAM Journal on Applied Mathematics, – (Dec., 2001 - Feb.,  
2002). – Vol. 62. – No. 3. – P. 1019-1043.
16. Tschumperl D. and Deriche R. Vector-valued image regularization with  
pde's : A common framework for different applications [Text] // IEEE  
Transactions on Pattern Analysis and Machine Intelligence, – 2005. – Vol.  
27. – No.4. – P. 506-517.
17. Drori I. Fragment-based image completion [Text] / I. Drori, D. Cohen-Or D.  
and H. Yeshurun // In Proceedings of ACM Conf. Comp. Graphics  
(SIGGRAPH), – 2003. – Vol. 22. – P. 303-312.
18. Chih-Wei Fang and Lien J.J. Rapid image completion system using multi-  
resolution patch-based directional and non-directional approaches [Text] //  
IEEE Transactions on Image Processing, – 2009. – Vol. 18. – No. 11.

- 19.Hays J. and Efros A.A. Scene completion using millions of photographs [Text] // ACM Transactions on Graphics (SIGGRAPH 2007), – 2007. – Vol. 26. – No. 3.
- 20.Oliva A. and Torralba A. Building the gist of a scene: The role of global image features in recognition [Text] // Progress in Brain Research: Visual perception, – 2006. – P. 155: 23-36.
- 21.Bertalmio M. Simultaneous structure and texture image inpainting [Text] / M. Bertalmio, L. Vese, G. Sapiro and S. Osher // IEEE Transactions on Image Processing, – 2003. – Vol. 12. – No.8. – P. 882-889.
- 22.Starck J.-L. Image decomposition via the combination of sparse representation and a variational approach [Text] / J.-L. Starck, M. Elad and D.L. Donoho // IEEE Transaction on Image Processing, – 2005.
- 23.Elad M. Simultaneous cartoon and texture image in-painting using morphological component analysis (mca) [Text] / M. Elad, J.-L. Starck, D. Donoho and P. Querre // Applied and Computational Harmonic Analysis, – 2005.
- 24.Jiaya Jia and Chi keung Tang. Inference of segmented color and texture description by tensor voting [Text] // IEEE Transactions Pattern Analysis and Machine Intelligence (PAMI), – 2004. – Vol. 26. – No. 6. – P. 771-786.
- 25.Yining Deng and B. S. Manjunath. Unsupervised segmentation of color-texture regions in images and video [Text] // IEEE Transaction on Pattern Analysis and Machine Intelligence (PAMI), – 2001. Vol. 23. – No. 8. – P. 800-810.
- 26.Medioni G. A Computational Framework For Segmentation And Grouping [Text] / G. Medioni, Lee Mi-Suen, and Tang Chi-Keung // Elsevier, – 2000.
- 27.Sun J. Image completion with structure propagation [Text] / J. Sun, L. Yuan, J. Jia, and H.-Y. Shum // In Proceedings of ACM Conf. Comp. Graphics, – 2005. – Vol. 24. – P. 861-868.

- 28.Oliviera M. Fast digital image inpainting [Text] / M. Oliveira, B. Bowen, R. McKenna, and Y.-S. Chang // In Proc. of Intl. Conf. on Visualization, Imaging and Image Processing, – 2001. – P. 261266.
- 29.Telea. An image in-painting technique based on the fast marching method [Text] // Journal of Graphics Tools, – 2004. – Vol. 9.
- 30.Girod. What's wrong with mean-squared error // In A.B. Watson, editor, Digital Images and Human Vision, chapter 15. MIT Press, – 1993. – P. 207-220.
- 31.Chan T.F. and Kang S.H. Error analysis for image inpainting [Text] // Journal of Mathematical imaging and Vision, – 2006. – Vol. 26. – No.1-2. – P. 85-103.
- 32.Ardis P. and Singhal A. Visual salience metrics for image inpainting [Text] // Proceedings of the SPIE, – 2009. – Vol. 7257.
- 33.Rissanen J.J. Modeling by the shortest data description [Text] // Automatica-J.IFAC. – 1978. – Vol. 14. – P. 465-471.
- 34.Rissanen J.J. Stochastic Complexity and Statistical Inquiry [Text] // World Scientific Publishers. – 1989.
- 35.Vitanyi P.M.B., Li M. Minimum description length induction, Bayesianism, and Kolmogorov complexity [Text] // IEEE Trans. on Information Theory. – 2000. – Vol. 46. – No. 2. – P. 446-464.
- 36.Wallace C.S., Boulton D.M. An information measure for classification [Text] // Comput. J. – 1968. – Vol. 11. – P. 185-195.
- 37.Wallace C.S., Freeman P.R. Estimation and inference by compact coding [Text] // J. Royal Stat. Soc. – 1987. – Series B. – Vol. 49. – No. 3. – P. 240-251. Discussion: ibid. P. 252-265.
- 38.Luo Q., Khoshgoftaar T.M. Unsupervised Multiscale Color Image Segmentation Based on MDL Principle [Text] // IEEE Trans. on Image Processing. – 2006. – V. 15. – № 9. – P. 2755–2761.
- 39.Galland F. Multi-component image segmentation in homogeneous regions based on description length minimization: Application to speckle, Poisson



- and Bernoulli noise [Text] / F. Galland, N. Bertaux, Ph. Réfrégier // Pattern Recognition. – 2005. – V. 38. – Iss. 11. – P. 1926–1936.
40. Lee T.C.M. A Minimum Description Length Based Image Segmentation Procedure, and Its Comparison with a Cross-Validation Based Segmentation Procedure [Text] // Journal of the American Statistical Association. – 2000. – V. 95. – P. 259–270.
  41. Lindeberg T., Li M.-X. Segmentation and classification of edges using minimum description length approximation and complementary junction cues [Text] // Computer Vision and Image Understanding. – 1997. – V. 67. – № 1. – P. 88–98.
  42. Cazorla M.A. et al. Bayesian models for finding and grouping junctions [Text] // Energy Minimization Methods in Computer Vision and Pattern Recognition. – 1999. – P. 70–82.
  43. Ward A., Hamarneh Gh. Statistical Shape Modeling using MDL Incorporating Shape, Appearance, and Expert Knowledge [Text] // In Lecture Notes in Computer Science, Medical Image Computing and Computer-Assisted Intervention (MICCAI). – 2007. – P. 278–285.
  44. Davies R.H. A minimum description length approach to statistical shape modeling [Text] / R.H. Davies, C.J. Twining, T.F. Cootes, J.C. Waterton, C.J. Taylor // IEEE Trans. Medical Imaging. – 2002. – V. 21. – № 5. – P. 525–537.
  45. Lanterman A. Minimum Description Length understanding of infrared scenes [Text] // Proc. SPIE. – 1998. – V. 3371. – P. 375–386.
  46. Li M., Gao Q. and Vitanyi P.M.B. Recognizing on-line handwritten characters using MDL [Text] // Proc. IEEE Information Theory Workshop. – 1993. – P. 24–25.
  47. Gao Q., Li M., Vitanyi P.M.B. Applying MDL to learning best model granularity [Text] // Artificial Intelligence. – 2000. – V. 121. – P. 1–29.
  48. Yvan G. et al. Self-Consistency and MDL: A Paradigm for evaluating point-correspondence algorithms, and its application to detecting changes in

- surface elevation [Text] // Int. J. of Computer Vision. – 2003. – V. 51. – № 1. – P. 63–83.
49. Maybank S.J. and Sturm P.F. MDL, Collineations and the Fundamental Matrix [Text] // Proc. 10th British Machine Vision Conference. – 1999. – P. 53–62.
  50. Ayer S., Sawhney H. Layered representation of motion video using robust maximum-likelihood estimation of mixture models and MDL encoding [Text] // ICCV. – 1995. – P. 777–784.
  51. Mansouri A.-R. and Konrad J. Minimum description length region tracking with level sets [Text] // Proc. SPIE Image and Video Communications and Process. – 2000. – V. 3974. – P. 515–525.
  52. Mansouri A.-R. and Konrad J. Motion segmentation with level sets [Text] // Proc. IEEE Int. Conf. Image Processing. – 1999. – V. II. – P. 126–130.
  53. Radke R.J. Image change detection algorithms: a systematic survey [Text] / R.J. Radke, S. Andra, O. Al-Kofahi, B. Roysam // IEEE Transactions on Image Processing. – 2005. – V. 14. – Iss. 3. – P. 294–307.
  54. Maybank S.J., Sturm P.F. Minimum description length and the inference of scene structure from images [Text] // IEE Colloquium on Applied Statistical Pattern Recognition. – 1999. – P. 9–16.
  55. Feldman J. Perceptual grouping by selection of a logically minimal model [Text] // Int. J. of Computer Vision. – 2003. – V. 55. – № 1. – P. 5–25.
  56. Pilu M., Fisher R.B. Part segmentation from 2D edge images by the MDL criterion [Text] // Image and Vision Computing. 1997. – V. 15. – № 8. – P. 563–573.
  57. Lillholm M. Feature-Base Image Analysis [Text] / M. Lillholm, M. Nielsen, L.D. Griffin // Int. J. of Computer Vision. – 2003. – V. 52. – № 2/3. – P. 73–95.
  58. Колмогоров А.Н. Три подхода к определению понятия «количество информации» [Текст] // Проблемы передачи информации. – 1965. – Т. 1. – № 1. – С. 3-11.

- 59.Solomonoff R. Does Algorithmic Probability Solve the Problem of Induction? [Text] // Oxbridge Research, P.O.B. 391887, Cambridge, Mass. 02139. – 1997.
- 60.Solomonoff R. Algorithmic Probability, Heuristic Programming and AGI [Text] // In: Baum, E., Hutter, M., Kitzelmann, E. (eds). Advances in Intelligent Systems Research. – 2010. – V. 10 (proc. 3<sup>rd</sup> Conf. on Artificial General Intelligence). – P. 151–157.
- 61.Solomonoff R. A formal theory of inductive inference, part 1 and part 2 [Text] // Information and Control. – 1964. – Vol. 7. – P. 1-22, 224-254.
- 62.Solomonoff R.J. The Discovery of Algorithmic Probability [Text] // J. of Computer and System Sciences. – 1997. – Vol. 55. – No. 1. – P. 73-88.
- 63.Arias P. A Variational Framework for Non-local Image Inpainting [Text] / P. Arias, V. Caselles, G. Sapiro // Proc. EMMCVPR'09. – 2009. – P. 345–358.
- 64.Shibata T. Fast and Structure-preserving Inpainting Based on Probabilistic Structure Estimation [Text] / T. Shibata, A. Iketani, Sh. Senda // MVA2011 IAPR Conference on Machine Vision Applications, Nara, JAPAN. – 2011. – P. 22–25.
- 65.Berthomme J.-M. How to use Information Theory for Image Inpainting and Blind Spot Filling-in? [Text] / J.-M. Berthomme, Th. Chateau1, M. Dhome1 // Proc. VISIGRAPP (2013). – 2013.
- 66.Cislariu M. A Fuzzy Set Generalization Of The Exemplar-Based Image Inpainting [Text] / M. Cislariu, M. Gordan, A. Vlaicu // Acta Technica Napocensis Electronics And Telecommunications. – 2011. – Vol. 52. – No. 2. – P. 54–59.
- 67.Март Д. Зрение. Информационный подход к изучению представления и обработки зрительных образов [Текст] // М., Радио и связь. – 1987. – 400 с.

- 68.Потапов А.С. Выбор представлений изображений на основе минимизации репрезентационной длины их описания [Текст] // Изв. вузов. Приборостроение. – 2008. – Т. 51. – № 7. – С. 3–7.
- 69.Potapov A.S. New paradigm of learnable computer vision algorithms based on the representational MDL principle [Text] / A.S. Potapov, I.A. Malyshev, A.E. Puysya, A.N. Averkin // Proc. SPIE. – 2010. – V. 7696. Automatic Target Recognition XX; Acquisition, Tracking, Pointing, and Laser Systems Technologies XXIV; and Optical Pattern Recognition XXI, Firooz A. Sadjadi; Abhijit Mahalanobis; Steven L. Chodos; William E. Thompson; David P. Casasent; Tien-Hsin Chao, Editors. – P. 769606.
- 70.Potapov A.S. Principle of Representational Minimum Description Length in Image Analysis and Pattern Recognition [Text] // Pattern Recognition and Image Analysis. – 2012. – V. 22. – No. 1. – P. 82–91.
- 71.Жданов И. Н. Метод сжатия трёхмерных биомедицинских изображений на основе представления информации в виде октодерева [Текст] / И. Н. Жданов, А. С. Потапов, О. В. Щербаков // Научно-технический вестник информационных технологий, механики и оптики. – 2012. – Т. 79. – № 3. – С. 100-104.
- 72.Потапов А.С. Метод предсказания на основе алгоритмической вероятности в задаче восстановления изображений в утерянных областях [Текст] / А.С. Потапов, О.В. Щербаков, И.Н. Жданов // Оптический журнал. – 2013. – Т. 80. – №11. – С. 48-53.
- 73.Potapov A. Practical algorithmic probability: an image inpainting example [Text] / A. Potapov, O. Scherbakov, I. Zhdanov // Proc. SPIE, Sixth International Conference on Machine Vision (ICMV 2013). – 2013. – Vol. 9067. – P. 906719-906719-5.
- 74.Baldi P. and Hornik K. Neural networks and principal component analysis: Learning from examples without local minima [Text] // Neural Networks, – 1989. – No. 2. – P. 53–58.

- 75.Japkowicz N. Nonlinear autoassociation is not equivalent to PCA [Text] / N. Japkowicz, S.J. Hanson and M.A. Gluck // Neural Computation, – 2000. –Vol. 12. – No. 3. – P. 531–545.
- 76.Larochelle H. Exploring strategies for training deep neural networks [Text] / H. Larochelle, Y. Bengio, J. Louradour, and P. Lamblin // Journal of Machine Learning Research, – January 2009a. – Vol. 10. – P. 1–40.
- 77.Larochelle H. An empirical evaluation of deep architectures on problems with many factors of variation [Text] / H. Larochelle, D. Erhan, A. Courville, J. Bergstra, and Y. Bengio // In Z. Ghahramani, editor, Proceedings of the Twenty-fourth International Conference on Machine Learning (ICML'07, 2007). – 2007. – P. 473–480.
- 78.Потапов А.С. Реконструкция изображения в невидимой области по критерию информативности его пространственного спектра [Текст] / А.С. Потапов, О.В. Щербаков, И.Н. Жданов // Сборник трудов II конференции «Будущее оптики» для молодых специалистов, кандидатов наук, аспирантов и студентов оптической отрасли и смежных дисциплин. 2–3 апреля 2013. – 2013. – С. 48–50.
- 79.Linsker R. An application of the principle of maximum information preservation to linear systems [Text] // In D.S. Touretzky, editor, Advances in Neural Information Processing Systems 1 (NIPS'88). Morgan Kaufmann, –1989.
- 80.Bell A.J. and Sejnowski T.J. An information maximisation approach to blind separation and blind deconvolution [Text] // Neural Computation, – 1995. – Vol. 7. – No. 6. – P. 1129–1159.
- 81.Larochelle H. Deep learning using robust interdependent codes [Text] / H. Larochelle, D. Erhan, and P. Vincent // In Proceedings of the Twelfth International Conference on Artificial Intelligence and Statistics (AISTATS2009), – 2009. – P. 312–319.

82. Bourlard H. and Kamp Y. Auto-association by multilayer perceptrons and singular value decomposition [Text] // Biological Cybernetics, – 1988. – Vol. 59. – P. 291–294.
83. Olshausen B.A. and Field D.J. Emergence of simple-cell receptive field properties by learning a sparse code for natural images [Text] // Nature, 1996. – Vol. 381. – P. 607–609.
84. Ranzato M. Efficient learning of sparse representations with an energy-based model [Text] / M. Ranzato, C.S. Poultney, S. Chopra, and Y. LeCun // In B. Scholkopf, J. Platt, and T. Hoffman, editors, Advances in Neural Information Processing Systems 19 (NIPS'06). MIT Press, – 2007. – P. 1137–1144.
85. Ranzato M. Sparse feature learning for deep belief networks [Text] / M. Ranzato, Y. Boureau, and Y. LeCun // In J.C. Platt, D. Koller, Y. Singer, and S. Roweis, editors, Advances in Neural Information Processing Systems 20 (NIPS'07), Cambridge, MA, MIT Press, – 2008. – P. 1185–1192.
86. Olshausen B.A. and Field D.J. Sparse coding with an overcomplete basis set: a strategy employed by V1? [Text] // Vision Research, – 1997. – Vol. 37. – P. 3311–3325.
87. Bishop C.M. Training with noise is equivalent to Tikhonov regularization [Text] // Neural Computation, – 1995. – Vol. 1. – P. 108–116.
88. Kullback S., Leibler R.A. On information and sufficiency [Text] // Annals of Mathematical Statistics 22, – 1951. – Vol. 1. – P. 79–86.
89. Perez-Cruz F. Kullback-Leibler divergence estimation of continuous distributions [Text] // Information Theory, 2008. ISIT 2008. IEEE International Symposium:, – 2008. – P. 1666–1670.
90. Erhan D. Why Does Unsupervised Pre-training Help Deep Learning? [Text] / D. Erhan, Y. Bengio, A. Courville, P.-A. Manzagol, P. Vincent, S. Bengio // Journal of Machine Learning Research. – 2010. – V. 11. – P. 625–660.
91. Vincent P. Stacked Denoising Autoencoders: Learning Useful Representations in a Deep Network with a Local Denoising Criterion [Text]

- / P. Vincent, H. Larochelle, I. Lajole, Y. Bengio, P.-A. Manzagol // Journal of Machine Learning Research. – 2010. – V. 11. – P. 3371–3408.
92. Ng Y. Feature selection, L1 vs. L2 regularization, and rotational invariance [Text] // In ICML, – 2004.
93. Olshausen B.A. and Field D.J. Sparse coding of sensory inputs [Text] // Cur. Op. Neurobiology, – 2004. – Vol. 14. – No. 4.
94. Krizhevsky. ImageNet Classification with deep convolutional neural networks [Text] / Krizhevsky, Sutskever, Hinton // NIPS 2012.
95. Chapelle O., Scholkopf B., and Zien A., editors. Semi-Supervised Learning [Text] // MIT Press, Cambridge, MA, – 2006.
96. LeCun Y. Backpropagation applied to handwritten zip code recognition [Text] / Y. LeCun, B. Boser, J.S. Denker, D. Henderson, R.E. Howard, W. Hubbard, and L.D. Jackel // Neural Computation, – 1989. – Vol. 1. – No. 4. – P. 541–551.
97. Gallinari P. Memoires associatives distribuees [Text] / P. Gallinari, Y. LeCun, S. Thiria, and F. Fogelman-Soulie // In Proceedings of COGNITIVA 87, Paris, La Villette, – 1987.
98. Masci J. Stacked Convolutional Auto-Encoders for Hierarchical Feature Extraction [Text] / J. Masci, U. Meier, D. Ciresan, J. Schmidhuber // Lecture Notes in Computer Science. – 2011. – V. 6791. – P. 52–59.
99. Ciresan D.C. Flexible, high performance convolutional neural networks for image classification [Text] / D.C. Ciresan, U. Meier, J. Masci, J. Schmidhuber // In International Joint Conference on Artificial Intelligence, IJCAI. – 2011
100. Scherer D. Evaluation of pooling operations in convolutional architectures for object recognition [Text] / D. Scherer, A. Muller, S. Behnke // In International Conference on Artificial Neural Networks. – 2010.

101. Shcherbakov O. Image inpainting based on stacked autoencoders [Text] / O. Shcherbakov, V. Batishcheva // Journal of Physics: Conference Series. – 2014. – V. 536. – P. 012020.
102. Bertalmio M. Navier-Stokes, fluid dynamics, and image and video inpainting [Text] / M. Bertalmio, A. Bertozzi, G. Sapiro // Proc. IEEE Computer Vision and Pattern Recognition (CVPR), – 2011.
103. FEI Face Database, [Электронный ресурс]. Режим доступа: <http://fei.edu.br/~cet/facedatabase.html>, свободный. Яз. англ. (дата обращения 04.09.2015).
104. Щербаков О.В. Сверточный автоэнкодер как генеративная модель изображений для задач выделения признаков и восстановления изображений в утерянных областях [Текст] / О.В. Щербаков, И.Н. Жданов, Я.А. Лушин // Оптический журнал. – 2015. – Т. 82. – № 8. – С. 48-53.
105. Щербаков О.В. Программный модуль «Система восстановления утерянных областей на изображении на базе стека автоэнкодеров». Жданов И.Н., Потапов А.С. Свидетельство о государственной регистрации программы для ЭВМ № 2014660857. Роспатент. 17.10.2014.
106. Щербаков О.В. Программный модуль «Сжатие трёхмерных биомедицинских изображений с использованием их представления в виде октодерева». Потапов А.С., Жданов И.Н. Свидетельство о государственной регистрации программы для ЭВМ № 2012618609. Роспатент. 21.09.2012.