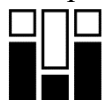


Министерство науки и высшего образования Российской Федерации
федеральное государственное автономное образовательное учреждение
высшего образования



**«НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ
ТОМСКИЙ ПОЛИТЕХНИЧЕСКИЙ УНИВЕРСИТЕТ»**

На правах рукописи

ДАНИЛОВ ВЯЧЕСЛАВ ВЛАДИМИРОВИЧ

**МЕТОДЫ И АЛГОРИТМЫ СЕГМЕНТАЦИИ МЕДИЦИНСКИХ
ИЗОБРАЖЕНИЙ НА ОСНОВЕ МАШИННОГО ОБУЧЕНИЯ**

Специальность 05.13.01

Системный анализ, управление и обработка информации
(научные исследования)

ДИССЕРТАЦИЯ

на соискание учёной степени
кандидата технических наук

Научный руководитель:
доктор технических наук, доцент
Гергет Ольга Михайловна

Томск – 2020

ОГЛАВЛЕНИЕ

ВВЕДЕНИЕ.....	4
1 Обзор существующих средств обработки медицинских данных	14
1.1 Представление изображений в медицине.....	14
1.2 Сегментация и локализация медицинского инструмента.....	17
1.3 Сегментация и локализация анатомических структур.....	31
1.4 Программные продукты для обработки медицинских данных.....	46
1.5 Математическая формулировка задачи	50
1.6 Выводы по первой главе.....	50
2 Методы и алгоритмы сегментации области интереса.....	53
2.1 Исходные данные и оборудование.....	53
2.2 Алгоритм синтеза данных	58
2.3 Алгоритм аугментации данных	64
2.4 Архитектура нейронной сети.....	66
2.4.1 Стандартная U-net архитектура.....	67
2.4.2 Разреженная U-net архитектура.....	68
2.4.3 Слой разреженной свёртки	70
2.4.4 Слой активации	72
2.4.5 Слой нормализации	73
2.4.6 Dropout слои и регуляризация	74
2.4.7 Пробросы признаков.....	76
2.4.8 V-net архитектура с пробросами признаков.....	79
2.5 Метод обучения модели	82
2.5.1 Функция потерь.....	82
2.5.2 Алгоритм изменения скорости обучения	84
2.5.3 Алгоритм изменения размера батча.....	89
2.6 Алгоритм выбора оптимальных гиперпараметров.....	90
2.7 Выводы по второй главе.....	99

3	Оценка полученных результатов.....	102
3.1	Синтез медицинского инструмента	102
3.1.1	Результаты сегментации синтезированных данных.....	102
3.1.2	Оценка результатов сегментации стандартной U-net модели....	106
3.1.3	Оценка результатов сегментации разреженной U-net модели...	112
3.2	Выбор оптимальных гиперпараметров.....	118
3.3	Сегментация медицинского инструмента	132
3.3.1	Результаты сегментации медицинского инструмента	132
3.3.2	Оценка результатов сегментации V-net модели	138
3.4	Сегментация анатомических структур	144
3.4.1	Сегментация левого предсердия	144
3.4.2	Сегментация поджелудочной железы.....	146
3.4.3	Сегментация селезёнки	147
3.4.4	Сегментация гиппокампа	149
3.4.5	Сегментация печени	150
3.5	Сравнительный анализ	152
3.6	Выводы по третьей главе	156
	ЗАКЛЮЧЕНИЕ	159
	СПИСОК СОКРАЩЕНИЙ И ОПРЕДЕЛЕНИЙ	162
	СПИСОК ЛИТЕРАТУРЫ.....	164
	ПРИЛОЖЕНИЕ А	182
	ПРИЛОЖЕНИЕ Б.....	185

ВВЕДЕНИЕ

Актуальность работы. На сегодняшний день существует множество работ, посвящённых сегментации изображений в различных сферах деятельности человека [1–6]. В свою очередь, работы, направленные на сегментацию медицинских инструментов или анатомических структур, представлены значительно меньшим числом, что объясняется низким качеством выходных данных, большой вариабельностью признаков, малым объёмом выборок, высокой зашумлённостью и т.д. Существующие в настоящее время классические подходы для выполнения сегментации медицинских изображений имеют множество ограничений, одним из которых является каскадность [7–10]. Алгоритмы на основе каскадности имеют блочную последовательную структуру и в большинстве случаев обладают низкой обобщающей способностью. Способность модели к обобщению заключается в распространении обнаруженных на обучающем множестве зависимостей и закономерностей на новые данные. В дополнении стоит отметить, что каскадные алгоритмы обладают негативным эффектом накопления ошибки.

В ряде исследований алгоритмы сегментации используют предварительно обработанные данные. Подобная реализация хоть и показывает в ряде случаев свою эффективность, тем не менее обладает множеством недостатков, а именно: высоким временем выполнения, низкой вариативностью параметров настройки и слабой адаптацией к предсказанию новых данных.

Другим недостатком классических алгоритмов сегментации, в том числе использующих подход машинного обучения, является использование признаков, синтезированных вручную. Несмотря на то, что существуют эффективные методы по выбору релевантных признаков, сам процесс их создания не всегда является приемлемым для решения медицинских задач. Важным будет отметить, что трёхмерные статические и трёхмерные динамические алгоритмы сегментации представлены небольшим числом исследований. Из-за высокой сложности методов сегментации трёхмерных объектов требуются большие временные затраты, что не

позволяет внедрять подобные методы для выполнения обработки данных в реальном масштабе времени.

Большинство современных медицинских модальностей, выполняющих сбор и визуализацию данных в режиме реального времени, не позволяют эффективно распознавать и анализировать сложные динамические структуры. В свою очередь, недостатки программно-аппаратной составляющей минимально инвазивной технологии значительно сокращают спектр возможных операций на сердце. В случае выполнения катетерных операций, актуальность выполнения сегментации обусловлена необходимостью локализации хирургического инструмента относительно оперируемой области. Поскольку направление работы связано с проведением исследований на работающем сердце, существует необходимость в технологии извлечения структурированной информации об объекте интереса из последовательности изображений.

Говоря о сфере разработки классических алгоритмов сегментации, значительный вклад в эту область внесли следующие учёные и исследователи: S. Horowitz и T. Pavlidis (Split-and-Merge) [11], R. M. Haralick и L.G. Shapiro (Region Growing) [12], S. Lloyd (K-means) [13], A. P. Dempster (Expectation-Maximization Clustering) [14], D. Comaniciu и P. Meer (Mean shift clustering) [15,16], F. Meyer (Watershed) [17] и другие. На развитие области разработки алгоритмов машинного, а также методов сегментации и локализации на основании глубокого обучения оказали большое влияние такие исследователи, как Y. LeCun (LeNet), D.E. Rumelhart, G.E. Hinton, A. Krizhevsky (AlexNet) [18], I.J. Goodfellow, A. Ng, J. Long (Fully Convolutional Network) [19], W. Liu (ParseNet) [20], H. Noh [21], O. Ronneberger (U-net) [22], T.-Y. Lin (Feature Pyramid Network) [23], R. Girshick (R-CNN, Fast R-CNN и Faster R-CNN) [24–26], H. Zhao (Pyramid Scene Parsing Network) [27], K. He (Mask R-CNN) [28], L.-C. Chen (DeepLab) [29, 30], S. Liu (Path Aggregation Network) [31], H. Zhang (Context Encoding Network) [32]. Отдельно стоит отметить деятельность отечественных учёных, таких как В.Н. Вапник, В.А. Князь, В.Н. Колмогоров (GrabCut) [33], Ю.И. Журавлёв, В.Г. Спицын, А.С. Конушин и другие, внесших свой вклад в исследования по обработке изображений.

Несмотря на наличие разработанных методов и программных средств, задача сегментации малых объектов, размер которых не превышает 1% от объёма изображения, на зашумлённых трёхмерных данных остаётся не решённой. В связи с этим становится актуальной разработка методов сегментации изображений и визуализации объектов на трёхмерной сцене, в том числе хирургического инструмента и анатомических структур на входных данных низкого качества.

Целью данной работы является создание методов и алгоритмов сегментации на основе машинного обучения, позволяющих повысить точность распознавания и визуализации медицинского инструмента и анатомических структур.

Для достижения поставленной цели в диссертации необходимо решить следующие исследовательские **задачи**:

- Провести анализ существующих методов и алгоритмов локализации и сегментации медицинских инструментов и анатомических структур.
- Разработать алгоритм синтеза трёхмерных эхокардиографических изображений с медицинским инструментом для случаев малого объёма данных, гомогенности и слабой репрезентативности выборки.
- Разработать архитектуру нейронной сети, которая сохраняет градиенты и решает проблему чрезмерной подгонки параметров к зависимостям, содержащимся в обучающем множестве данных.
- Реализовать алгоритм выбора оптимальных гиперпараметров для глубоких нейронных сетей.
- Реализовать метод обучения нейронной сети с использованием циклической скорости обучения и динамического изменения размера батча, позволяющего ускорить сходимость процесса обучения.
- Разработать программное обеспечение и провести вычислительный эксперимент с целью оценки результатов сегментации, полученных на основе предлагаемых методов и алгоритмов.

Объектом исследования является процесс сегментации медицинских изображений.

Предметом исследования являются методы машинного обучения, с помощью которых осуществляется локализация и сегментация изображений, содержащих медицинский инструмент и анатомические структуры.

Методы исследования. Выполненные в диссертационной работе исследования основаны на использовании методов цифровой обработки изображений, математического анализа и моделирования, алгоритмов компьютерного зрения, математической статистики, методов преобразования и анализа изображений, машинного обучения, распознавания образов и прикладного программирования.

Научная новизна диссертационной работы представлена следующими составляющими:

- Предложена модификация U-net архитектуры в виде V-net архитектуры свёрточной нейронной сети с дополнительными пробросами признаков, наличие которых позволяет решить проблему затухающих градиентов глубоких нейронных сетей. V-net архитектура в среднем на 10% превосходит стандартную U-net архитектуру по точности.
- Разработан алгоритм синтеза медицинского инструмента на трёхмерных ультразвуковых изображениях, который позволяет увеличить гетерогенность используемых данных и автоматизировать процесс разметки. Для точной реконструкции данный алгоритм использует кинематику гибких роботов при совмещении изображений анатомических структур и медицинского инструмента.
- Предложен метод обучения нейронных сетей, основанный на интеграции динамического изменения размера батча и циклической скорости обучения с алгоритмом test-time аугментации на этапе предсказания новых данных. Такая реализация позволяет повысить скорость сходимости по сравнению с классическими методами обучения, использующими статический размер батча и/или постоянное значение скорости обучения.
- Разработан алгоритм выбора гиперпараметров, основанный на последовательном подборе значений параметров посредством использования

t-критерия Стьюдента для связанных выборок. Предлагаемый алгоритм позволяет учитывать динамику на прошлых итерациях и не выполнять полный перебор по подмножеству пространства гиперпараметров, что значительно сокращает время подбора оптимальных значений.

Достоверность результатов диссертационного исследования обеспечивается успешным тестированием обученных моделей, а достигнутые результаты коррелируют и согласуются с эталонными, полученными квалифицированными специалистами в следующих организациях: Boston Children's Hospital (Бостон, США), King's College London (Лондон, Великобритания), Memorial Sloan Kettering Cancer Center (Нью-Йорк, США), Vanderbilt University Medical Center (Нэшвилл, США), IRCAD Hôpitaux Universitaires (Страсбург, Франция). Принятые в работе авторские решения основаны на известных и проверенных методах разработки глубоких нейронных сетей для обработки медицинских графических данных. Положения диссертации базируются на анализе и практических результатах использования машинного обучения, обобщении опыта в области математического моделирования и распознавания образов. Достоверность результатов, работоспособность разработанных методов и алгоритмов подтверждается соответствием экспериментальных данных, полученных в ходе исследований и опытного внедрения.

Практическая значимость работы заключается в следующем:

- Разработанное программное обеспечение позволяет сопровождать хирургические вмешательства, в том числе минимально инвазивные операции, а также снизить вероятность врачебной ошибки.
- Научные исследования по визуализации медицинских инструментов в анатомических структурах могут быть внедрены в медицинские образовательные программы и использоваться в научно-исследовательских целях. Полученные трёхмерные модели использованы студентами и медицинскими работниками в целях моделирования поведения хирургического инструмента при проведении

малоинвазивных операций, а также для симуляционного и имитационного обучения новым методам диагностики и вмешательств.

- Модели и методы, использующиеся для обработки медицинских изображений, могут применяться при выполнении минимально инвазивных вмешательств в хирургической онкологии, нейрохирургии и сосудистой хирургии.

- В связи с отсутствием в настоящее время предварительно обученных трёхмерных нейронных сетей для выполнения сегментации, все полученные модели могут использоваться в качестве первичной инициализации весов. В таком случае время обучения новых моделей значительно снизится.

Положения, выносимые на защиту:

- Метод сегментации на основе V-net архитектуры с плотными пробросами признаков позволяет решить задачу сегментации не только медицинского инструмента, но и анатомических структур с высокой точностью. При тестировании используемого метода точность сегментации катетера составила $93.6 \pm 2.4\%$, а левого предсердия – $92.2 \pm 2.7\%$.

- Алгоритм синтеза ультразвуковых данных позволяет повысить точность сегментации за счёт увеличения репрезентативности обучающей выборки и автоматизировать процесс разметки сгенерированных данных. Тестирование стандартной и разреженной U-net архитектур выявило положительную тенденцию изменения точности за счёт увеличения количества синтетических примеров в обучаемых выборках. Точность сегментации для разреженной U-net архитектуры составила $86.5 \pm 3.6\%$, $88.3 \pm 3.1\%$, $89.8 \pm 3.5\%$ и $92.6 \pm 2.2\%$ при RDR равном 1.00, 0.50, 0.33 и 0.25 соответственно.

- Метод обучения нейронных сетей на основе использования динамического размера батча и циклической скорости обучения, а также предсказания на основе алгоритма test-time аугментации позволяет в среднем на 6.0% точнее выполнять предсказания новых данных, а также повысить обобщающую способность и надёжность модели. Использование данного метода увеличило точность сегментации медицинского инструмента и анатомических структур в среднем на 3.0% и 6.3% соответственно.

– Алгоритм выбора гиперпараметров позволяет снизить время поиска оптимальных значений параметров модели по сравнению с такими алгоритмами, как Grid Search и Randomized Search. Предлагаемый алгоритм учитывает динамику точности модели на прошлых итерациях и не проходит по всему подмножеству пространства гиперпараметров. При использовании предлагаемого подхода с 11 гиперпараметрами количество моделей для оптимизации составило 690 вместо 204 374 016 для Grid Search и Randomized Search.

Внедрение. Разработанные методы и алгоритмы сегментации используются в федеральном государственном бюджетном научном учреждении «Научно-исследовательский институт комплексных проблем сердечно-сосудистых заболеваний» (Кемерово, Россия) и адаптированы для выполнения сегментации стеноза при проведении диагностики коронарных артерий на основе данных ангиографии (акт о внедрении от 07.05.2019). Материалы диссертационной работы используются при проведении практических занятий по дисциплинам «Машинное обучение», «Технология обработки информации», «Теория распознавания образов», «Нейронные сети», «Методы интеллектуальной обработки и анализа изображений» для магистрантов направлений 230100 "Информатика и вычислительная техника" и 230400 "Информационные системы и технологии" в Томском политехническом университете (акт о внедрении от 05.09.2019).

Реализация результатов. Результаты диссертационного исследования использованы при выполнении следующих научно-исследовательских проектов:

- Государственное задание "Наука" №4003/БЧ «Разработка новых моделей и методов для диагностики и лечения заболеваний сердца и нарушений мозгового кровообращения».
- Государственное задание "Наука" №17.8205.2017/БЧ «Моделирование и алгоритмизация процессов обработки, анализа и репрезентации графических данных в режиме реального времени для проведения эндоваскулярных операций на сердце».
- Грант РНФ №18-75-10061 "Исследование и реализация концепции роботизированного малоинвазивного протезирования клапана аорты".

– Государственное задание "Наука" №FFSWW-2020-0014 «Разработка научных основ технологии роботизированной мультипараметрической томографии на основе методов обработки больших данных и машинного обучения для исследования перспективных композиционных материалов».

Апробация. Работа прошла апробацию на следующих конференциях:

– III International Scientific Conference on Information Technologies in Science, Management, Social Sphere and Medicine (Томск, Россия, 2016).

– XV Международная научно-практическая конференция студентов, аспирантов и молодых учёных (Томск, Россия, 2017).

– 27-я Международная конференция по компьютерной графике и машинному зрению «ГрафиКон» (Пермь, Россия, 2017).

– Computer Assisted Radiology and Surgery International Conference and Exhibition (Берлин, Германия, 2018).

– V International workshop “Critical infrastructures: Contingency management, Intelligent, Agent-based, Cloud computing and Cyber security” (Иркутск, Россия, 2018).

– 28-я Международная конференция по компьютерной графике и машинному зрению «ГрафиКон» (Томск, Россия, 2018).

– 18th International Conference on Mechatronics (Брно, Чехия, 2018).

– 14th International Conference on Electromechanics and Robotics "Zavalishin's Readings" (Курск, Россия, 2019).

– International Workshop on “Photogrammetric and computer vision techniques for video surveillance, biometrics and biomedicine” (Москва, Россия, 2019).

– 29-я Международная конференция по компьютерной графике и машинному зрению «ГрафиКон» (Брянск, Россия, 2019).

– 14th International Forum on Strategic Technology (IFOST 2019) (Томск, Россия, 2019).

– I Всероссийский Конгресс с международным участием «Физиология и тканевая инженерия сердца и сосудов: от клеточной биологии до протезирования» (Кемерово, Россия, 2019).

Публикации. Основные результаты диссертационного исследования нашли своё отражение в 22 печатных работах, из которых:

- 4 публикации в рецензируемых журналах, рекомендованных ВАК РФ;
- 13 публикаций в изданиях, индексируемых в базах научного цитирования Scopus и Web of Science (в том числе, 4 работы типа Article, 9 работ типа Conference paper);
- 5 публикаций в других изданиях;

Дополнительно получены:

- 5 свидетельств о государственной регистрации программ для ЭВМ;
- 1 решение о выдаче патента на изобретение.

Личный вклад соискателя. Основные научные результаты, выносимые на защиту, получены автором лично. Данная работа выполнена в сотрудничестве с учёными и исследователями, оказавшими содействие при проведении настоящего исследования, а именно:

- Постановка задач исследования по теме диссертации выполнена совместно с научным руководителем Гергет О.М. (Томский политехнический университет) и кардиохирургом Васильевым Н.В. (Boston Children's Hospital, Harvard Medical School).
- Обзор существующих алгоритмов анализа, обработки и визуализации медицинских данных выполнен совместно со Скирневским И.П. (Центр Финансовых Технологий) и научным руководителем Гергет О.М. (Томский политехнический университет).
- Сбор, разметка и валидация ультразвуковых данных при проведении минимально инвазивных вмешательств выполнены совместно с кардиохирургом Васильевым Н.В. (Boston Children's Hospital, Harvard Medical School).
- Разработка алгоритма синтеза ультразвуковых данных сердца и медицинского инструмента выполнена совместно с Манаковым Р.А. (Научно-исследовательская лаборатория «Гамма технологии») и Колпащиковым Д.Ю. (Лаборатория дизайна медицинских изделий, Томский политехнический университет).

– Разработка архитектуры нейронной сети выполнена совместно с María J. Ledesma-Carbayo, David Bermejo Peláez и Luis Alfonso Hernández Gómez (Biomedical Imaging Technologies Laboratory, Universidad Politécnica de Madrid).

Благодарности. Автор выражает искреннюю благодарность:

– María J. Ledesma-Carbayo (Biomedical Imaging Technologies Laboratory, Universidad Politécnica de Madrid) за предоставленный сервер с видеокартами Titan X и Titan Xp для проведения исследования;

– Farid Melgani (Signal Processing and Recognition Laboratory, Università degli Studi di Trento) за курирование исследования, посвящённого извлечению и выявлению наиболее релевантных и информативных признаков

– Alejandro F. Frangi (Center for Computational Imaging and Simulation Technologies in Biomedicine, University of Leeds) за оборудование для выполнения обучения/валидации моделей и доступ к платформам MULTI-X, Amazon Web Services и UK Biobank.

Структура и объём диссертации. Диссертация состоит из введения, 3 глав, заключения, списка использованной литературы из 179 наименований и 2 приложений. Работа изложена на 191 странице и включает 85 рисунков и 36 таблиц.

1 Обзор существующих средств обработки медицинских данных

1.1 Представление изображений в медицине

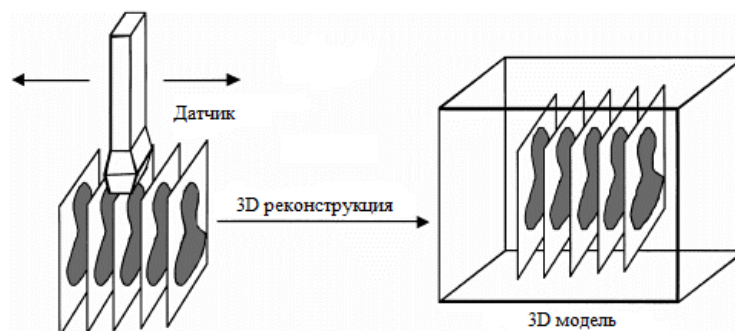
В настоящее время существует несколько универсальных стандартов для хранения и передачи медицинских данных, одними из которых являются Analyze [34, 35], Minc [36], DICOM [37] и NIfTI [38]. Использование этих форматов оказывает непосредственное влияние на этапы и эффективность обработки информации (в частности, медицинских изображений). Некоторые из этих форматов носят узконаправленный характер и служат исключительно для хранения изображений, которые были получены при специфических способах исследования. Другие являются общепринятыми и обеспечивают хранение практически любых типов медицинских изображений. Наиболее распространённым считается формат DICOM. Большая часть производимых в настоящий момент аппаратных комплексов совместимы с данным форматом. Его широкое распространение в России во многом связано с тем, что значительная часть диагностической аппаратуры, покупаемой и используемой в России, производится за рубежом в тех странах, где DICOM является de-facto стандартом. При этом переход на собственный формат хранения данных не является рациональным. Таким образом, широкое распространение стандарта DICOM послужило толчком к созданию целого ряда программных продуктов, обеспечивающих обработку и просмотр медицинской графической информации. Зачастую, DICOM-визуализаторы идут в комплекте с аппаратными комплексами, но могут использоваться и в качестве самостоятельных приложений.

Современные томографы (МРТ, КТ, ПЭТ) и эхокардиографы (ЭхоКГ), в отличие от видеокамер и фотоаппаратов, не производят готовых изображений. Взамен этого создаётся DICOM файл, включающий в себя метаданные. Каждый такой файл представляет собой срез выбранной части тела в определённой плоскости, чаще всего в горизонтальной. В свою очередь, дата исследования, данные о пациенте, данные об изображении, информация для рендеринга и визуализации, а также сведения об оборудовании непосредственно содержатся в

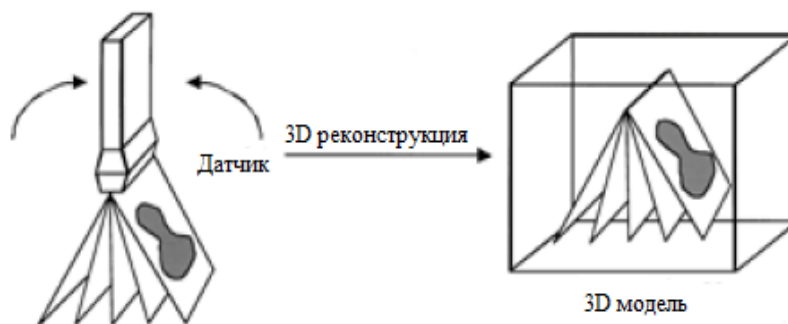
метаданных. В каждом DICOM файле представлена информация о плотности и интенсивности тканей в определённом срезе, на основе которой происходит построение итогового изображения. Необходимо учитывать, что плотность и интенсивность – разные понятия. К примеру, рентгеновская плотность, сохраняемая в файлах компьютерной томографии, зависит от физической плотности тканей. У костей физическая плотность больше, а у крови и мягких тканей меньше. В свою очередь, эхокардиографы и магнитно-резонансные томографы сохраняют интенсивность обратного сигнала. Информация о плотности в DICOM файле представлена как обычное изображение со своим разрешением, цветовой моделью, размером пикселя/вокселя, форматом, методом представления и прочими данными. Однако вместо информации о цвете пиксель хранит в себе информацию о плотности тканей. Как уже было сказано ранее, основу для рендеринга DICOM файла составляет плотность тканей, в связи с чем визуализация изображения невозможна без сопоставления значения плотности со значением цветов. Обеспечение данного процесса происходит благодаря передаточной функции визуализатора, которая может иметь как открытый, так и закрытый код. Помимо этого, существует множество различных готовых шаблонов для рендеринга разных по плотности тканей разными цветами.

На сегодняшний день подавляющее большинство подходов по реконструкции 3D моделей используют наборы двумерных срезов (слайсов), которые могут быть получены с использованием различных типов модальностей: МРТ, ПЭТ, КТ и т.д. Однако стоит иметь в виду то, что способы записи изображений у различных медицинских модальностей могут быть разными. Так, например, сканер компьютерной томографии всегда производит параллельные слайсы с высоким контрастом, а изображения от ультразвукового сканера или эхокардиографа могут собираться при низком уровне контраста как параллельно, так и расходясь в стороны под определённым углом (см. рисунок 1.1). В связи с этим, методы обработки и 3D реконструкции медицинских данных зачастую подстраиваются под конкретный тип модальности.

Основными модальностями, широко применяемыми сегодня при выполнении эндоваскулярных и минимально инвазивных процедур на сердце, являются флюороскопия и эхокардиография. Зачастую, выходными данными этих модальностей являются двухмерные изображения в оттенках серого. Важно также отметить возможность передачи графических данных этих модальностей в режиме реального времени. Однако, существенными недостатками этих модальностей являются: низкое качество выходных данных, ограниченная область видимости, наличие значительного числа артефактов и теней, высокая зашумлённость и низкий контраст. Очевидно, что использование классических систем двухмерной динамической визуализации, таких как эхокардиография, ангиография, флюороскопия, не даёт возможности расширения спектра реконструктивных операций с использованием эндоваскулярной технологии. Низкое качество выходных данных при эхокардиографии не позволяет выполнять сложные и точные манипуляции внутри сердечных отделов. Доза облучения, полученная пациентами при ангиографии и флюороскопии, ограничивает время визуализации и вынуждает хирурга работать со статичным изображением, а использование двухмерных данных не позволяет выполнять сложные пространственные



(а) Параллельный тип слайсов



(б) Расходящийся тип слайсов

Рисунок 1.1 – Типы сбора данных ультразвуковым датчиком

манипуляции. Стоит также добавить, что флюороскопия не позволяет оценить глобальный контекст, поскольку на изображениях данной модальности мягкие ткани представляют собой серый зернистый фон без наличия границ объектов. В свою очередь, двумерный сигнал, использующийся в подавляющем большинстве кардиохирургических интервенций, не позволяет адекватно отобразить пространственное положение хирургического инструмента и визуализировать область интереса, обладая при этом значительным уровнем цифрового шума, тенями и артефактами.

Таким образом, широкий спектр кардиохирургических операций осуществляется только на открытом «сухом» сердце с применением искусственного кровообращения. Подобные методы коррекции патологий имеют ряд недостатков, связанных с тяжёлыми послеоперационными осложнениями и длительной реабилитацией, что представляет собой важную проблему кардиологии.

1.2 Сегментация и локализация медицинского инструмента

Одна из главных задач сегментации заключается в разделении изображения на регионы, для которых должен соблюдаться определённый критерий гомогенности. Регионом изображения является его область, элементы которой имеют общее свойство или признак. На сегодняшний день можно выделить пять основных классов методов сегментации [39, 40]. Классификация этих методов представлена ниже на рисунке 1.2. Далее в этой главе рассматриваются одни из наиболее популярных, современных и эффективных методов сегментации как медицинских инструментов, так и различных анатомических структур.

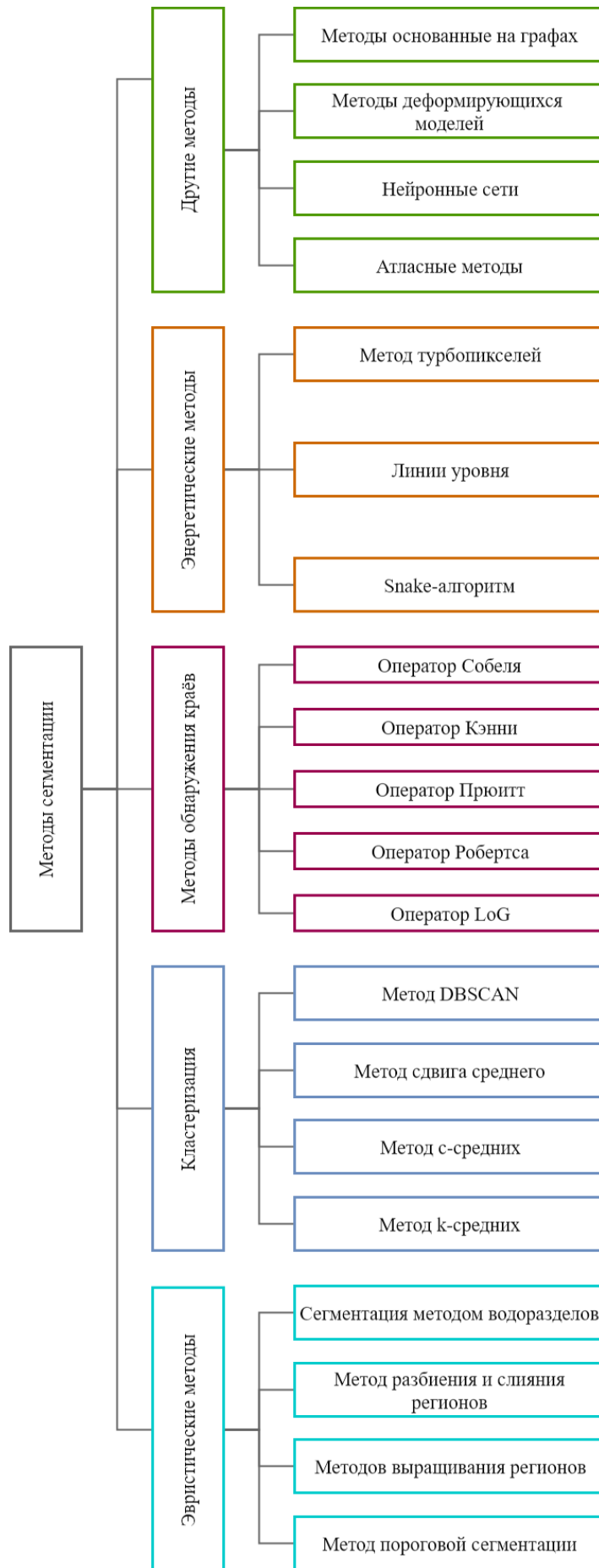


Рисунок 1.2 – Классификация методов сегментации

В настоящее время одной из актуальных задач информационных технологий в медицине является разработка методов и алгоритмов, позволяющих выполнить точную, пациентоориентированную трёхмерную реконструкцию отделов сердца вместе с хирургическим инструментом. В целом, задача трёхмерной реконструкции анатомических структур сердца в режиме реального времени актуальна уже несколько десятков лет и на данный момент активно используется в электрофизиологии. Например, в работах [41, 42] рассматриваются методы трёхмерной реконструкции для эффективного выполнения абляции в камерах сердца и поиска координат катетера. Однако все известные на сегодня методы и системы, включая современные системы электроанатомического картирования, не решают задачу трёхмерной динамической визуализации, а также точной сегментации области интереса. Несмотря на то, что в области электрофизиологии на текущем этапе не требуется отображение динамической сцены и хирургу достаточно статической трёхмерной поверхности эндокарда во время систолы, существуют исследования и операции, которые учёные и врачи вынуждены проводить на открытом сердце из-за отсутствия механизмов точного позиционирования и визуализации. К таким операциям стоит отнести шунтирование коронарной артерии; установку медицинских приспособлений и устройств для последующего восстановления правильного ритма работы сердца; восстановление повреждённых и/или аномальных участков сердца; восстановление и/или замена сосудов.

Кардиохирургическая катетеризация как метод интервенционной терапии получила свою популярность за последнее десятилетие благодаря снижению рисков и уменьшению времени восстановления пациентов. В связи с этим было предложено множество методов и алгоритмов по детектированию медицинского инструмента внутри сердца, таких как роботизированная локализация или трекинг на основе электромагнитных сенсоров. Однако, подобные техники требуют наличия дополнительного оборудования, что влечёт за собой большие затраты при их использовании в клинической практике. Менее затратной альтернативой является обнаружение медицинского инструмента на основе изображений. Один из

предложенных методов локализации катетера на основе изображений описан в работе [43]. Авторы предлагают проводить локализацию инструмента при помощи сопоставления с шаблоном. Однако, данный метод требует предварительных знаний о направлении катетера и не обладает дискриминативной способностью. В связи с этим, этот метод не может применяться в сложных анатомических условиях и различных положениях катетера. Другой техникой, показавшей приемлемые результаты, является метод, обучающийся на основе интенсивности изображения и отклика фильтра Frangi [44]. Однако более детальное исследование по классификации вокселей инструмента показало, что использование не только интенсивности, но и признака формы инструмента даёт лучший результат [45]. Подобные методы были верифицированы на фантомном *in-vitro* наборе данных, в то время как эффективность на *ex-vivo* или *in-vivo* тканях проведена не была.

Как отмечалось выше, в отличие от магнитно-резонансной и компьютерной томографии, трёхмерная эхокардиография обеспечивает более гибкий и недорогой подход для контроля изображений в реальном масштабе времени. Однако присущий ультразвуковым изображениям спекл-шум и низкое разрешение усложняют задачу локализации объекта исследования. Кроме того, медицинские инструменты, в том числе катетеры, обычно следуют по изогнутой траектории и на изображениях могут иметь нерегулярный вид, особенно в тех случаях, когда их дистальный конец находится перпендикулярно к ультразвуковому датчику.

Неоднократно предпринимались попытки усовершенствования контроля за такими инструментами, как иглы, электроды, катетеры посредством их сегментации. Так, например М. Varva предложил метод, который позволяет автоматически определять положение тонкого прямолинейного объекта [46]. Применив данный метод для локализации металлического электрода в биологической ткани, авторы показали, что ось электрода можно найти посредством максимизации преобразования параллельной интегральной проекции. Однако стоит отметить, что при выполнении хирургического вмешательства медицинский инструмент большую часть времени имеет нерегулярную форму, что значительно ограничивает область применения данного метода. Другой метод,

описанный в работе [47], определяет позицию криволинейного электрода на ультразвуковых изображениях. Авторы предложили сегментировать возможные местоположения электрода при помощи метода пороговой бинаризации и метода оценки параметров модели на основе случайных выборок (сокр. RANSAC от англ. Random Sample Consensus). Именно по рандомизированному методу RANSAC происходило моделирование траектории объекта с использованием трёхмерной кубической кривой. Одним из ограничений алгоритма являлась гладкость кривой объекта. В аналогичном исследовании Н. Neshat продемонстрировал технику обнаружения изогнутых игл на ультразвуковых изображениях в режиме реального времени [48]. Данная техника основана на новом подходе параметризации для обобщённого преобразования Радона, а сам алгоритм реализован на GPU с использованием программно-аппаратной архитектуры параллельных вычислений CUDA. Однако стоит заметить, что во многих исследованиях использовались лишь фантомные изображения, а тестирование надёжности на данных медицинского инструмента отсутствует.

Ряд научных и медицинских групп, используя результаты сегментации, выполняют исследования по реконструкции и детектированию катетера на основе бипланарной флюороскопии, в том числе с использованием С-дуги. Важно отметить, что данная модальность использует ионизирующее излучение. Несмотря на то, что радиация у флюороскопии минимальна, все же существует вероятность повреждения из-за радиационного облучения, а присутствие связанных с ним рисков невозможно избежать. К другим недостаткам стоит отнести низкое разрешение, плохое отображение мягких тканей, перекрывающуюся анатомию, а также ограничение подвижности пациента.

Сложность алгоритмов обработки и реконструкции 3D моделей за счёт совмещения проекций не позволяет выполнять обработку и визуализацию графических данных в режиме реального времени. Однако, на сегодняшний день бипланарная флюороскопия с использованием С-дуги является одним из наиболее часто (до 60%) применяемых методов визуализации [49]. Главной тому причиной является отсутствие современных программно-аппаратных решений,

выполняющих точную автоматическую обработку и визуализацию. Так, например, в работах Matthias Hoffman выполнена полуавтоматическая реконструкция катетера по двум проекциям [50, 51]. В своём методе авторы свели взаимодействие с пользователем к минимуму, однако для локализации и реконструкции всё же требуется указание хотя бы одной точки, принадлежащей к области катетера. Несмотря на относительно высокую точность распознавания, составляющую примерно 94%, суммарное время выполнения алгоритмов локализации и реконструкции составило 8.85 секунд, что говорит о том, что данная методика может использоваться лишь в условиях предоперационной подготовки. Стоит также учесть факт неполной автоматизации данного метода, что, в свою очередь, требует постоянного отвлечения врача от своих основных обязанностей при выполнении медицинских операций. Дополнительно стоит отметить, что метод реконструкции с помощью скелетонизации, применяемый в работе, не даёт возможности полноценно оценить размеры катетера внутри анатомических структур, а определяет лишь положение центральной линии инструмента. Исходя из этого, данный метод имеет ограниченную область применения, в частности, для оффлайн визуализации катетера внутри кровеносных сосудов, где знание положения катетера относительно стенок вен или артерий не так критично.

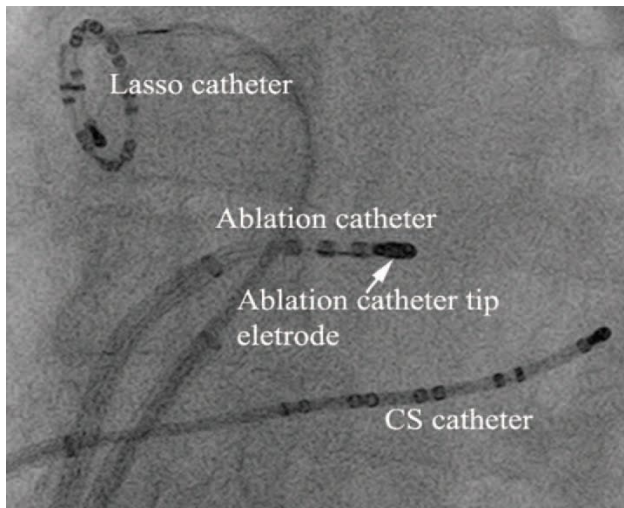
Из-за высокого времени сбора информации на сегодняшний день сложно использовать модальности с высоким разрешением, такие как МРТ или КТ для оценки положения катетера. Однако М. Vock совместно со своими коллегами попытались реализовать метод получения данных катетера внутри сердца в режиме реального времени [52] с помощью модифицированного алгоритма GRAPPA (от англ. Generalized Autocalibrating Partially Parallel Acquisitions). С помощью представленного метода время сбора данных было сокращено на 34%. Для ускорения реконструкции, чувствительность катушек обновлялась только после переориентации слайсов. Следует подчеркнуть, что двухмерная trueFISP реконструкция в режиме реального времени была получена с помощью 6 катушек. При заданном количестве активных катушек время сбора данных составило 209 мс.

Аналогичные исследования по реконструкции и визуализации катетера в трёхмерном пространстве сосудистой сети проводились В. Shirley [53]. В представленной работе исследователи протестировали собственный метод на внутричерепном антропоморфном сосудистом фантоме. Средняя ошибка для реконструированного инструмента составила 0.42 мм, а ошибка реконструкции дистального конца медицинского инструмента составила 0.65 мм. Необходимо выделить, что предложенный подход является каскадным, поэтому для достижения положительных результатов требуется выполнение ряда дополнительных алгоритмов, таких как геометрическая калибровка, корректировка изображений, удаление искажений и шума. Последнее ведёт к тому, что предложенный метод носит узконаправленный характер из-за своей гомогенности исследуемых данных.

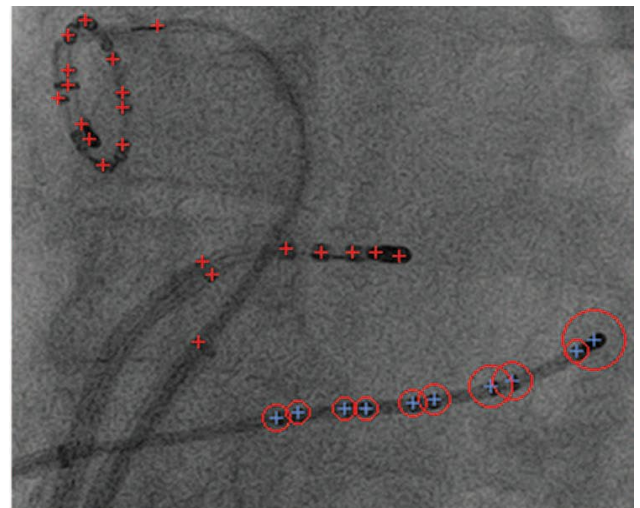
В настоящее время ведутся работы по детектированию и трекингу катетера без осуществления реконструкции. В работе [54] исследователи Y. L. Ma, N. Gogin, P. Cathier и др. использовали алгоритм трекинга на трёх различных типах катетеров, не выполняя реконструкцию медицинского инструмента. Стоит подчеркнуть, что точность детектирования для всех трёх катетеров остаётся высокой, ошибка детектирования на основе двухмерных бипланарных изображений составляет $0.50 \text{ мм} \pm 0.29 \text{ мм.}$, $0.92 \text{ мм} \pm 0.61 \text{ мм}$ и $0.63 \pm 0.45 \text{ мм}$, а вероятность точной локализации составляет 99.4%, 97.2% и 88.9% соответственно для катетера коронарного синуса, абляционного катетера и катетера-лассо. Отметим высокую скорость работы данного алгоритма, что позволяет обрабатывать данные в режиме реального времени. Однако на сегодняшний день одного лишь детектирования недостаточно, так как даже не опытный глаз увидит катетер на изображениях флюороскопии. Необходима ещё реконструкция и визуализация для оценки положения медицинского инструмента внутри анатомических структур сердца для проведения более безопасных и эффективных интервенций. Так, например, на рисунке 1.3 отражена визуализация и локализация медицинского инструмента при проведении кардиохирургического вмешательства. Исходя из этих данных, видно, что на изображениях видны лишь медицинские инструменты, а зернистый фон не несёт какой-либо важной информации.

Становится очевидным, что данная модальность не полноценно отражает ситуацию, и врач, в свою очередь, будет ориентироваться только на свой опыт и ощущения при проведении операции. Тестирование своей методологии на модальностях другого типа, в том числе ультразвуковых данных, исследователи не приводят.

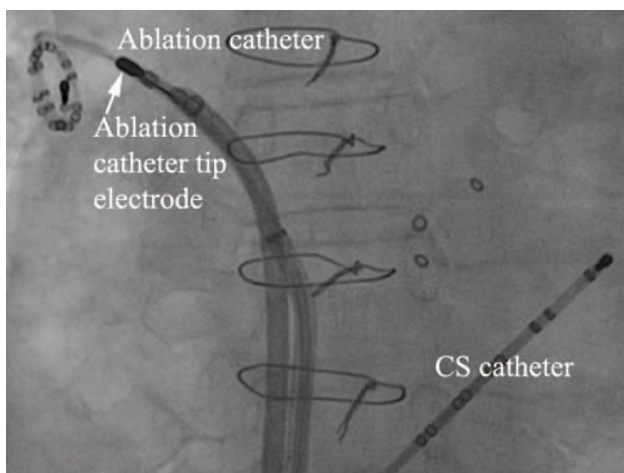
Обнаружение катетера на данных флюороскопии является довольно изученной проблемой. На сегодняшний день существует ряд методов для извлечения графических данных различных инструментов в интерактивном режиме [54, 55]. Так, в работе [56] описывается система трекинга катетера на данных флюороскопии для выполнения дальнейшей регистрации. В некоторых случаях при использовании ультразвуковых данных или данных ангиографии,



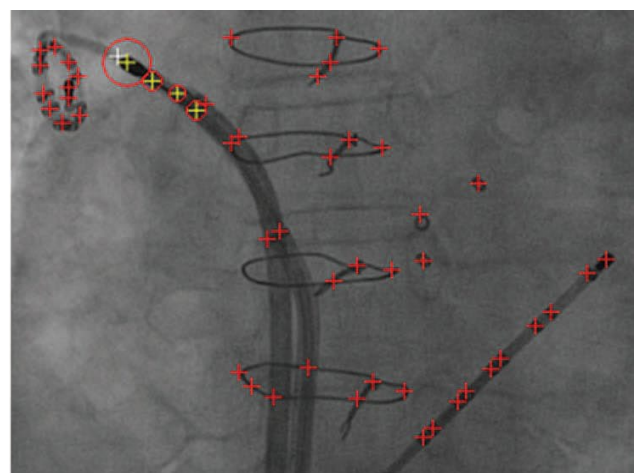
(а) Исходное изображение с низкой дозой облучения



(б) Детектирование катетера коронарного синуса



(в) Исходное изображение с нормальной дозой облучения



(г) Детектирование абляционного катетера

Рисунок 1.3 – Пример детектирования катетеров на основе рентгеновских изображений

также требуется трёхмерное позиционирование устройства. Например, система CartoSound использует электромагнитное отслеживание, встроенное в С-образную дугу, для локализации катетера при эпикардиальной эхокардиографии [57], а в работе [58] представлен подход роботизированного самодетектирования для проведения кардиохирургических вмешательств. Отметим, что внедрение дополнительного оборудования не всегда представляется возможным, а системы регистрации, основанные на обработке изображений без дополнительного оборудования, набирают популярность в связи с их простотой и точностью [59].

При проведении хирургических интервенций на сердце, медицинские устройства хорошо различимы на данных флюороскопии. Однако С-дуга, используемая данной модальностью, может иметь произвольные углы наклона, инъекции могут значительно изменять контрастность, а сама модальность не может предоставить глобальный контекст для решения задачи детектирования. Альтернативой флюороскопии может стать эхокардиография. Отметим, что в отличие от чреспищеводной эхокардиографии, датчик эпикардиальной эхокардиографии может быть доставлен из различных направлений в целевую область, не ограничиваясь при этом 6 степенями свободы. Учитывая данные преимущества в работе [60] К. Ralovich предложил метод детектирования медицинского инструмента на основе гибридной модальности: флюороскопия и эхокардиография. Для этой системы разработан метод оценки положения катетера, оснащённого специальным рентгеноконтрастными маркерами. С целью локализации катетера исследователи использовали Байесовский подход и метод уточнения местоположения маркеров посредством шаблонного сопоставления. Эксперименты были проведены как на искусственных данных, так и на *in-vivo* данных свиных сердец. Ошибка регистрации на основании 703 случаев составила 8.06 ± 7.20 мм, а результат работы данного метода отражён на рисунке 1.4.

Возможность получения достоверных результатов на основе только данных трёхмерной эхокардиографии представлена исследователями из General Electric [43]. Они разработали автоматизированный метод локализации катетера, основным этапом которого является использование карты вероятности катетера. Для более

быстрого детектирования авторы предлагают использовать метод жадной оптимизации, где вероятностная карта катетера проецируется на изображение, используя подход проецирования максимальной интенсивности. В качестве дополнительных этапов используются алгоритмы пост-обработки, направленные на сокращение ложных и улучшение истинных срабатываний. Результат детектирования катетера на основе подобного подхода отражён на рисунке 1.5. По заявлению авторов, метрика *sensitivity* для данного метода составляет 0.77 ± 0.11 , а метрика *precision* – 0.97 ± 0.11 . Стоит отметить, что несмотря на такое высокое значение точности, выборка данного метода являлась весьма ограниченной, а именно 26 изображений от 7 пациентов. Последнее говорит о недостаточном тестировании метода и о возможной гомогенности данных. Представленный метод обладает рядом ограничений, а именно катетер должен обязательно находиться параллельно к плоскости ультразвукового датчика, а также появляться на изображениях в виде яркого трубчатого объекта. Такие ограничения вкупе с валидацией на малой выборке могут негативно сказаться на работе представленного метода на новых данных.

Н. Fazlali представил работу, посвящённую методу сегментации катетера и коронарных сосудов для извлечения центральных линий артерий [10]. В качестве метода сегментации исследователи использовали подход суперпикселей. В свою очередь, для решения задачи детектирования катетера предлагается использовать подход, описанный в работе А. Frangi [61], который основан на преобразовании Хафа (от англ. Hough Transform) и скелетонизации. Предлагаемый Н. Fazlali метод применяется на 10 самых длинных сегментах скелета. Одним из ограничений локализации катетера является то, что требуется детектировать катетер до введения

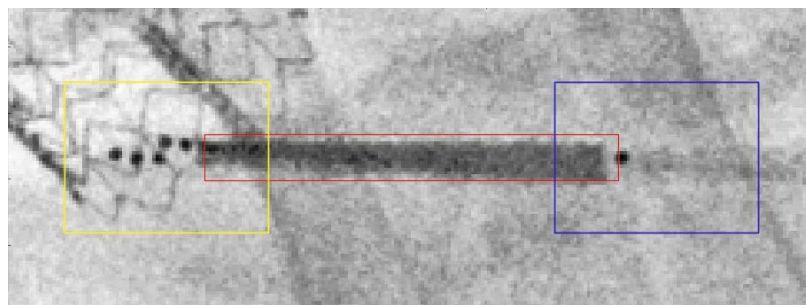
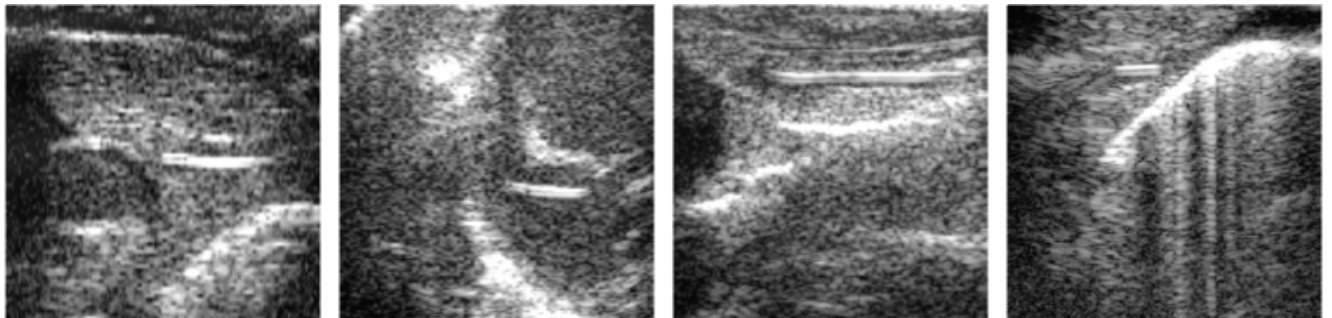


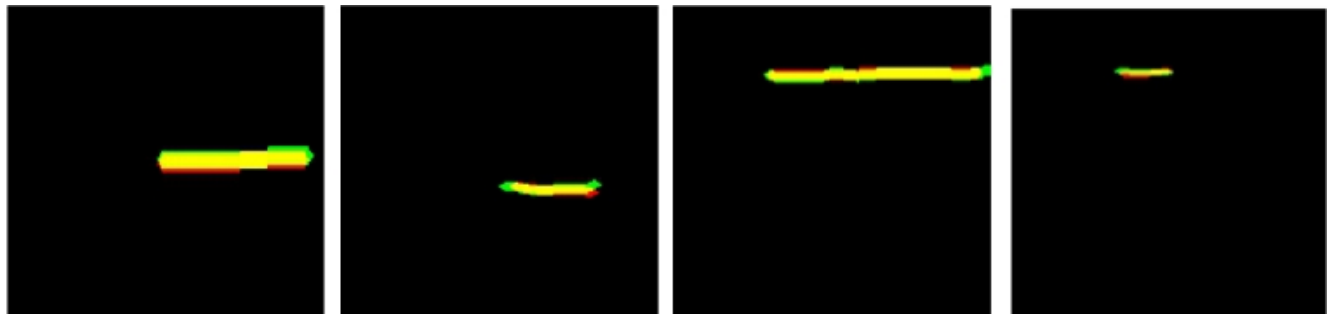
Рисунок 1.4 – Локализация катетера с использованием рентгеноконтрастных маркеров

рентгеноконтрастного препарата, и только после этого осуществляется его отслеживание. Результат детектирования скелета катетера отражён на рисунке 1.6.

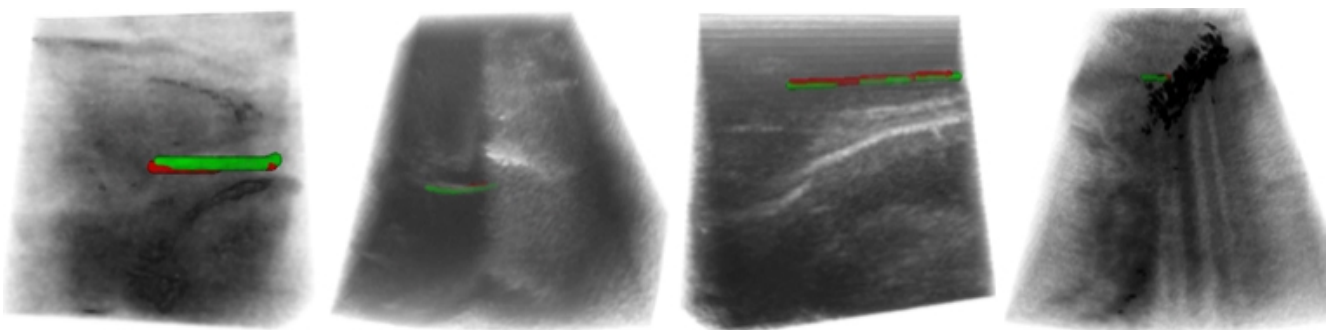
Н. Fazlali и его коллеги отмечают относительно высокую точность сегментации, а именно 16% случаев были сегментированы с отметкой «Limited», т.е. основные сосуды идентифицированы, но результаты не включают существенные ветви или неправильно включают обширные площади без сосудов; 56% сегментированы с отметкой «Good», т.е. крупные сосуды и существенные ветви включены, но при этом отсутствуют только клинически незначительные ветви; 28% сегментированы с отметкой «Excellent», т.е. включены и основные и небольшие ветви. Стоит отметить, что предлагаемый подход содержит большое количество вспомогательных блоков. Так, например, модуль предварительной



(а) Исходные данные



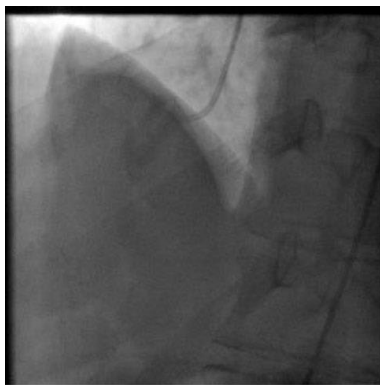
(б) Размеченные данные (отмечены зелёным цветом), детектирования (отмечены красным цветом) и идеальные совпадения (отмечены жёлтым цветом)



(в) Трёхмерный рендеринг

Рисунок 1.5 – Пример детектирования катетера на основе проецирования вероятностных карт

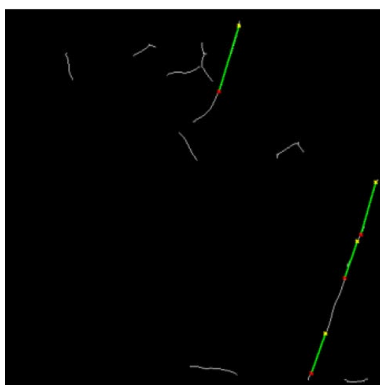
обработки содержит следующие блоки: блок улучшения контраста, блок измерения сосудов, блок скелетизации, блок сглаживания и т.д. Основной этап сегментации артерий и локализации катетера проходит также в несколько этапов. Главный недостаток такого подхода – накопление ошибки, которая может значительно увеличиваться при прохождении по отдельным блокам алгоритма. Дополнительно стоит отметить, что процесс сегментации проходит также итеративно при различных масштабах суперпикселей. Только такой подход комбинирования разномасштабных суперпикселей позволил решить проблему сегментации сосудов различных размеров. Другим недостатком такого метода является проблема сегментации и локализации катетера при пересечении его с другими анатомическими структурами. В связи с этим исследователи предлагают начинать фазу локализации инструмента до введения в организм радиоактивных инъекций, так как на данной стадии медицинский инструмент хорошо различим и не имеет наложений и пересечений с другими объектами или артефактами. В подобных случаях возникает ошибка классификации и катетер классифицируется как артерия



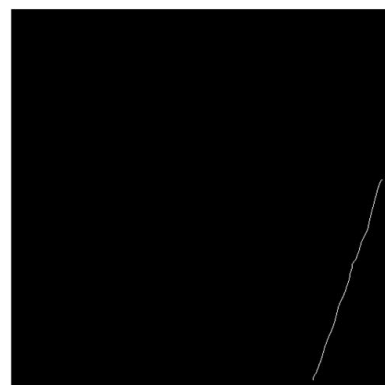
(а) Исходное изображение



(б) Скелетонизация



(в) Применение преобразования Хафа

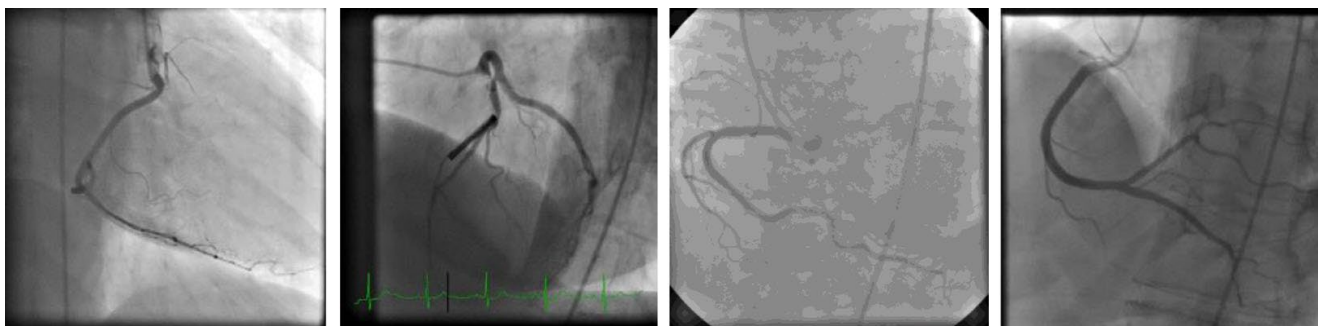


(г) Скелет катетера

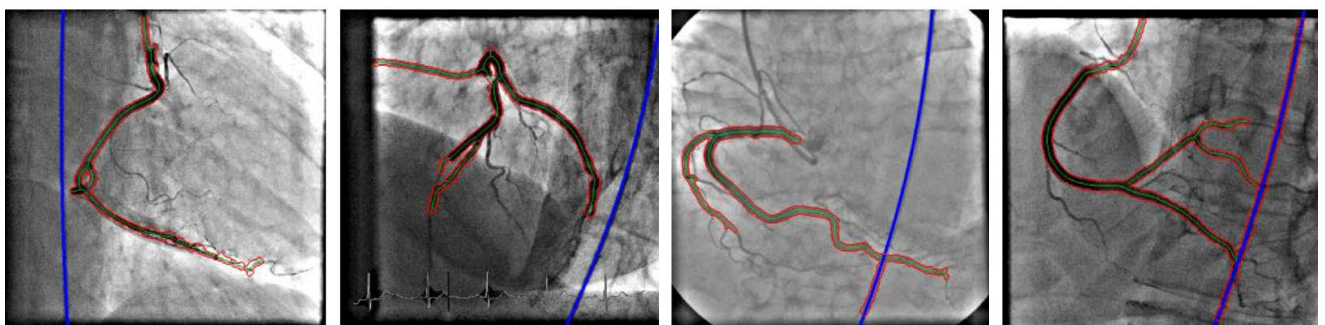
Рисунок 1.6 – Результат детектирования скелета катетера

(см. крайний правый рисунок 1.7 (б)). Результат работы данного метода отражён на рисунке 1.7.

Другой подход по локализации медицинских устройств относится к применению нейронных сетей. В работе [9] исследователи использовали комбинацию метода RANSAC и свёрточных нейронных сетей для выполнения локализации катетера на данных трёхмерного ультразвука. Для классификации вокселей исследователи использовали две нейронных сети, с помощью которых предсказывали принадлежность данного вокселя к катетеру. Схема предложенного алгоритма, состоящая из двух этапов, отражена на рисунке 1.8. Стоит отметить, что в данном алгоритме используются 2.5-мерные данные, т.е. нейронная сеть на входе принимала не трёхмерный куб, а три двухмерные ортогональные проекции с центром в текущем пикселе. По заявлению авторов метода, использование такого типа данных основано на компромиссе между сложностью архитектуры и точностью предсказания. При получении ложных срабатываний нейронных сетей дополнительно использовался метод RANSAC, который производил оценку параметров модели. Данный метод минимизирует количество случайных включений и ложных срабатываний с предыдущего этапа, а именно с этапа



(а) Исходные данные



(б) Результат сегментации

Рисунок 1.7 – Результат сегментации коронарных артерий сердца и катетера

классификации. Результат локализации инструмента с помощью такого подхода отражён на рисунке 1.9, где оба случая представляют ортогональные проекции. В качестве оценки точности данного метода использовалась метрика F2, значение которой составило 0.71 ± 0.18 . Такая точность является относительно низкой, особенно для медицины, где средний порог точности находится на уровне 85%. Важно также отметить, что используемая метрика F2 имеет смещённый характер, так как параметр recall имеет больший вес, чем параметр precision. Авторы сделали упор на максимизацию обобщающей способности сети для предсказания катетера. При одинаковых значениях precision и recall метрика F2 в среднем на 5% выше метрики F1, что говорит о более низкой обобщающей способности модели к предсказанию пикселей заднего фона.

В аналогичном исследовании [9] авторы использовали разреженные данные для локализации медицинского инструмента, принимая во внимание только три ортогональные проекции. Такой подход может быть усилен заменой классификации отдельных вокселей на классификацию полноценных трёхмерных объектов. В таком случае будет эффективнее учитываться пространственная информация. В качестве примера можно использовать архитектуру на основе энкодера и декодера, которая учитывает пространственную информацию, а также такие атрибуты, как форма, размер, цвет, компактность, гомогенность и т.д. Исследование на основе подобного подхода было реализовано для сегментации

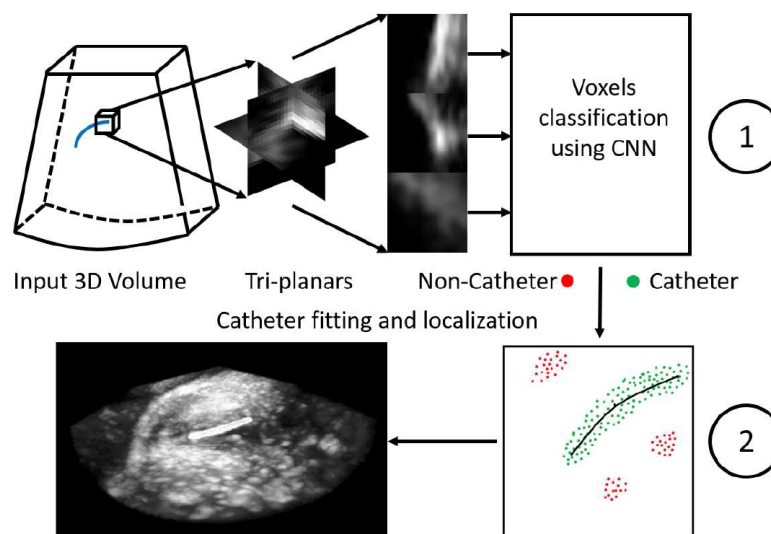
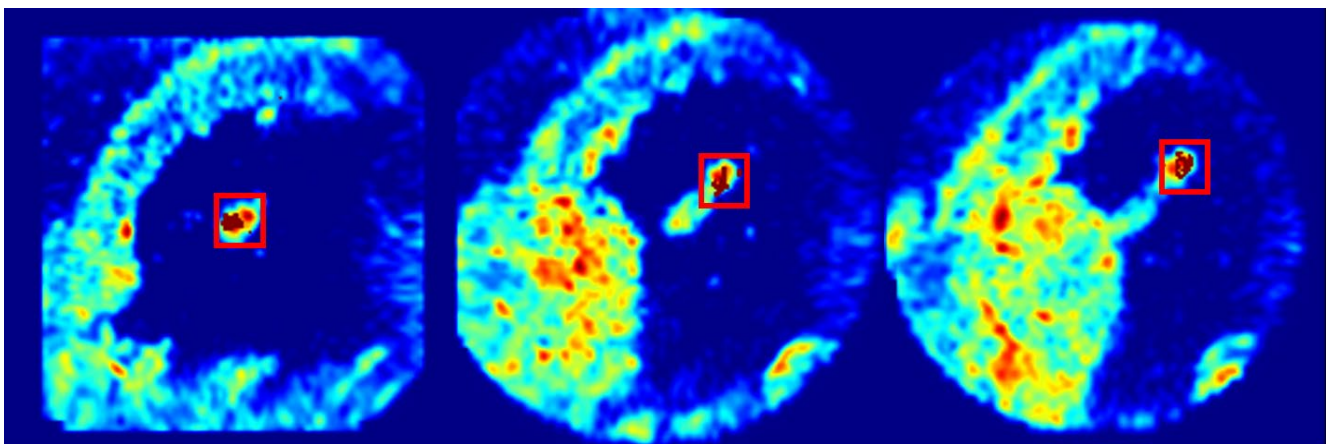


Рисунок 1.8 – Схема детектирования катетера на основе свёрточной нейронной сети и алгоритма RANSAC

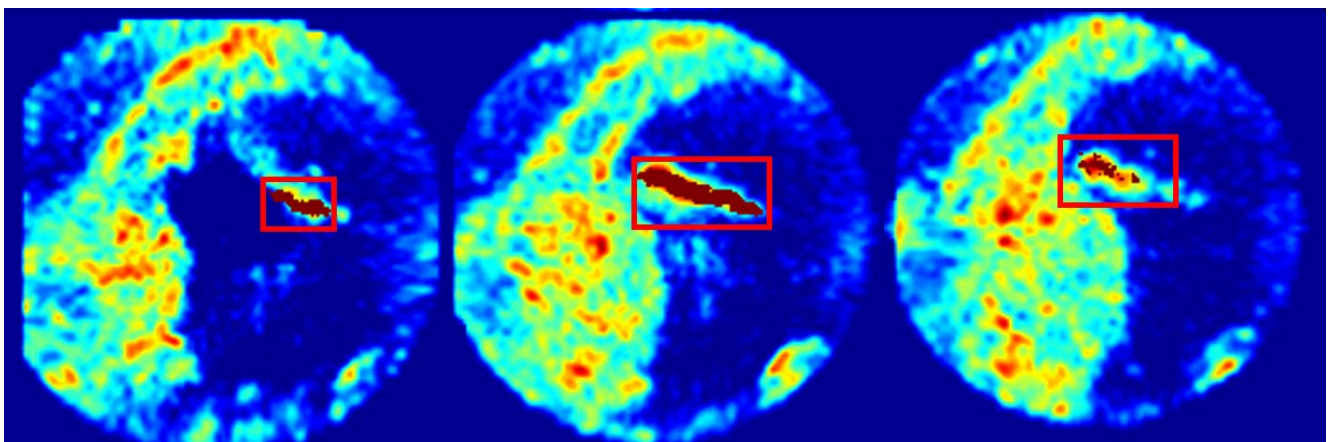
катетера с помощью нейронной сети FCN и предобученной модели VGG16 [62]. В результате тестирования обученной модели на *ex-vivo* наборе данных исследователи получили значение индекса Дайса равное 57.7%.

1.3 Сегментация и локализация анатомических структур

Анализ медицинских изображений на основе сегментации области интереса имеет большое клиническое значение, так как его основной целью является локализация органов и патологий с последующим повышением качества постановки диагноза. Однако нельзя сказать, что существует единый метод сегментации превосходящий все остальные. Связано это с тем, что существует ряд обстоятельств, по-разному влияющих на выходное изображение, а именно качество входных данных, наличие шума и его тип, качество предобработки и т.п. Поэтому на сегодняшний день для минимизации потерь выходных данных и изображений,



(a) Recall = 0.12, Precision = 0.36



(б) Recall = 0.92, Precision = 0.68

Рисунок 1.9 – Результат работы алгоритма детектирования катетера на основе свёрточной нейронной сети и алгоритма RANSAC

а также для достижения наилучшего результата при выделении области интереса существует множество методов по сегментации и контурированию анатомических структур, которые справляются с определённым спектром задач. Несмотря на то, что обработка изображений и компьютерное зрение являются молодыми направлениями науки, они достаточно востребованы. За последние два десятилетия методы обработки и анализа изображений стали активно применяться в медицинских исследованиях. Будет важным отметить, что данная предметная область до сих пор не получила достаточного освещения в научном сообществе. Если речь идёт о визуализации кардиохирургических данных, то наиболее востребованными являются работы, посвящённые внедрению информационных технологий в область анализа левого желудочка сердца, а также сердечных клапанов. Так, в работе К.Н. Lim [63] рассматриваются методы ассиметричного моделирования митрального клапана на основе трёхмерной эхокардиографии. Схожие исследования также встречаются в работах R. Levine и M. Handschumacher [64]. Большое количество исследований, направленных на изучение клапанов сердца, а также левого предсердия и левого желудочка обусловлено широким спектром заболеваний, вызванных нарушениями в работе данных анатомических структур [65]. Тем не менее, вопрос о выполнении сегментации правого желудочка при проведении минимально инвазивных операций также является актуальным. Например, в работе M. Nillesen [66] выполняется автоматическая оценка разнородных характеристик правого желудочка на основе трансторакальной трёхмерной эхокардиографии. Отличительной особенностью работы является то, что исследователи не использовали априорные выборки с выполненной ручной сегментацией.

Небольшое количество исследований, относящихся к реконструкции и сегментации правого желудочка, связано со сложной геометрической формой данной анатомической структуры. В силу специфического расположения правый желудочек сложнее сегментировать и в последствии выполнять трёхмерную реконструкцию. Тем не менее, в работе A. Tandon, N. Byrne и др. описывается подход полуавтоматической сегментации правого желудочка и предсердия [67].

Важно заметить, что разработчики данного подхода использовали изображения магнитно-резонансной томографии, специфика которых даёт значительный прирост качества входных данных, а также снижает уровень цифрового шума. Таким образом, при выполнении диагностики сердца с использованием ультразвука перед исследователями встаёт вопрос о предварительной обработке входных изображений путём снижения уровня цифрового шума и других побочных эффектов с целью улучшения их качества.

Предварительной обработке ультразвуковых данных посвящено множество работ, в том числе работы Y. Chen, S. Broschat и др. [68] и J. R. Sanchez и M. Oelze [69]. В представленных работах исследователями проводятся серии экспериментов с применением таких фильтров, как фильтр Гаусса, медианный фильтр, идеальный фильтр и другие. Описанные фильтры широко распространены и являются de-facto стандартами в области подавления цифрового шума. Результаты их работ отражены во множестве статей, но, как показала практика, данные фильтры не всегда дают приемлемые результаты с точки зрения такой характеристики, как отношение контраст-шум, являющейся одной из важнейших для выполнения точной сегментации объектов. Другими словами, при значительном снижении уровня шума классические фильтры размывают границы областей объектов, что в будущем делает невозможным использование автоматических методов точной сегментации.

В области предварительной обработки цифровых изображений существует ряд работ, в которых описаны отличные от классических подходы к фильтрации. Так, Y. Yue, M. Croitoru и M. Bidani в своём исследовании применяют вейвлет-преобразования для выполнения процедуры шумоподавления [70]. Важной характеристикой вычислений на основе преобразований Фурье или вейвлет-преобразований является высокая эффективность алгоритмов, а к недостаткам можно отнести их высокую сложность и трудоёмкость.

Среди российских и зарубежных источников не было найдено работ, посвящённых внедрению аналитической системы оценки качества предварительной обработки для автоматизированного выбора алгоритма или

входных характеристик. Однако, среди русскоязычного сообщества интерес к изучению представленных выше тематик нашёл своё отражение в работах следующих отечественных учёных-исследователей: В.А. Абрамов, Ю.С. Андреев, Ю.М. Баяковский, А.А. Богуславский, В.П. Вежнев, И.М. Журавель, Ю.И. Журавлев, М. Казанов, Н.Н. Красильников, В.Ф. Нестерук, В.В. Сергеев, В.А. Сойфе, С.В. Яблонский, Л.П. Ярославский. Фундаментальные исследования в области обработки цифровых изображений были проведены следующими авторами: Brown Russel, Yu-Jin Zhang, Richard E. Woods, Paul Viola, Michael Jones, Larry S. Davis, Rafael C. Gonzalez, B. Lippel, M. Kurland, Thomas Shi-Tao Huang, Judith M. S. Prewitt, Azriel Rosenfeld, John F. Canny, Theo Pavlidis, Irwin Sobel, William K. Pratt и др. Большинство представленных выше исследователей не специализируются на исследованиях в области медицины, но тем не менее описанные ими подходы и принципы могут быть применены для решения задач данного исследования.

Область сегментации цифровых изображений получила широкое распространение среди медицинского сообщества, а фундаментальные принципы сегментации отражены в широком спектре статей и книг. Однако, подавляющее большинство работ, описывающих алгоритмы сегментации медицинских изображений, базируются на данных, полученных вследствие проведения магнитно-резонансной или компьютерной томографии. Значительно меньше исследований посвящено обработке и сегментации эхокардиографических данных. Одной из наиболее интересных и актуальных работ, где в качестве входных данных используется ультразвук, является исследование учёного Rob Schneider из компании Philips [71]. Работы R. Schneider направлены на реконструкцию митрального клапана, все показания сняты в динамике, а в исследовании используется множество нестандартных решений как в области сегментации, так и при решении задачи реконструкции. Совместная работа учёных J. Hung, R. Lang и др., связанная с обзором текущего положения 3D эхокардиографии и перспектив её развития, описывает преимущества данного подхода и выделяет сложности, с которыми могут столкнуться исследователи [72]. Значимый вклад в область

реконструкции и сегментации медицинских изображений вносят следующие исследования:

- Работа по созданию алгоритма сегментации левого желудочка сердца на основе активных моделей внешнего вида, разработанного такими исследователями, как Nikos Paragios, Marie-Pierre Jolly и др. [73].
- Исследование по сегментации правого и левого желудочков в режиме реального времени на основе трёхмерной эхокардиографии [74]. Данная работа выполнена совместно учёными из США и Франции, а именно: Elsa D. Angelini, Shunichi Homma, Gregory Pearson, Jeffrey W. Holmes и Andrew F. Laine.
- Работа по созданию метода автоматической сегментации правого желудочка на основе 3D ЭхоКГ, созданного такими учёными, как Jørn Bersvendsen, Fredrik Orderud и др. [75].
- Монография Европейского Общества Кардиологов, посвящённая основным рекомендациям по получению и репрезентации трёхмерной ЭхоКГ [76].
- Сравнительное исследование по сегментации правого желудочка на основе данных МРТ, выполненное такими исследователями, как Caroline Petitjean, Maria A. Zuluaga, Wenjia Bai, Jean-Nicolas Dacher и т.д. [77].

Важно отметить, что многие исследователи в своих работах оперируют либо статическим потоком данных, либо двухмерными данными, что не актуально для представленного исследования. Такое ограничение связано с тем, что трёхмерные данные позволяют выполнять пространственную оценку и измерение анатомических структур сердца, а динамические данные обеспечивают непрерывную обратную связь между врачом/исследователем и пациентом.

Задачами трёхмерной реконструкции, как и сегментацией медицинских данных занимается широкий спектр учёных. К заслуживающим внимания работам стоит отнести исследования, проводившиеся такими учёными, как Navaneetha Krishnan Rajan, Zeying Song и др. [78], Shyh-Jye Chen, Ming-Tai Lin и др. [79], а также работы [80, 81]. Важно подчеркнуть, что в представленных выше исследованиях реконструкция выполнялась либо на основе данных магнитно-резонансной томографии, либо с использованием статических данных

эхокардиографии сердца. Задача трёхмерной реконструкции правого желудочка сердца является наименее освещённой в силу описанных выше причин, а именно геометрически сложной формы данной анатомической структуры и низким числом прикладных исследований, связанных с изучением и проведением операций на правом желудочке до недавнего времени.

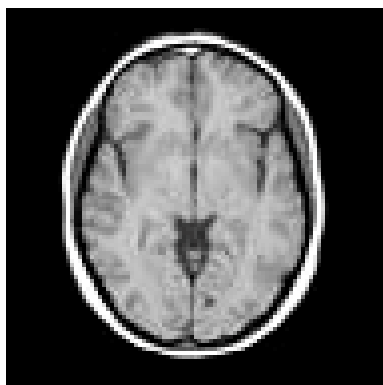
Одной из популярных и используемых техник сегментации является метод выращивания регионов. В одной из самых простых реализаций метода выращивания регионов центр региона вручную выбирается оператором, что делает использование взаимодействия с пользователем одним из главных недостатков данного подхода. Другими словами, для того чтобы сегментировать регионы исходного изображения пользователь вручную должен определить центр каждой области. Поэтому для минимизации взаимодействия с пользователем существуют иные стратегии выбора зёрен регионов и их выращивания [82]. Стоит отметить, что техника выращивания регионов имеет высокую чувствительность к шуму, что приводит к тому, что сегментированные регионы имеют артефакты типа «дырки» или вообще отсоединяются от региона интереса. При обработке изображения метод выращивания регионов редко применяется в качестве самостоятельного. Чаще всего он используется совместно с другими подходами, особенно если требуется определить небольшую и сравнительно простую структуру. Тем не менее, данный подход применяется в медицине и с целью оценки очагов поражений или опухолей. Так, например, при сравнении нескольких методов сегментации D. Junfeng и Y. Yunyang получили достаточно результативный алгоритм сегментации высококонтрастных МРТ изображений головного мозга [83]. Результат работы данного алгоритма отображён на рисунке 1.10.

Алгоритмически более сложным подходом является сегментация методом дробления-слияния. Сегментация на основе данного метода проходит в две стадии: разделение и последующее объединение. Однако в сравнении с методом выращивания регионов, этот алгоритм не требует определения центров регионов [84]. Процедура дробления начинается с разбиения изображения на регионы, причём не обязательно однородные. Первый этап длится до момента получения

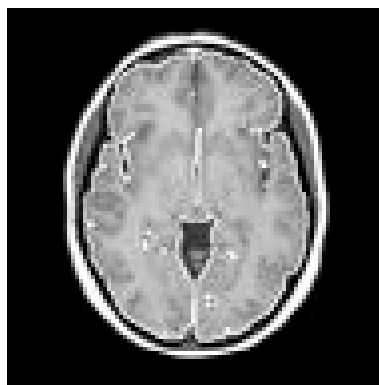
полной пересегментации. После чего начинается процесс объединения схожих соседних сегментов до тех пор, пока изображение не будет разделено на гомогенные области максимального размера.

Множество методов автоматической сегментации имеют сложности с выделением регионов из-за наличия различных факторов: схожие значения интенсивности для разных анатомических структур, наличие артефактов, неравномерный контраст и т.п. Поэтому для решения подобных задач необходима исходная информация. Иногда она может быть представлена экспертами в виде ряда правил, однако более эффективным подходом является сравнение с атласом [85]. В том случае, когда априорная информация представлена в виде атласа, процесс сегментации сводится к сопоставлению входного изображения и данных атласа. Данный процесс обычно проходит в два этапа. Первый этап заканчивается глобальным преобразованием, быстро компенсирующим основные различия атласа и входного изображения. Во время второго этапа происходят локальные преобразования для точной регистрации структур. После совмещения обоих изображений разметку из атласа переносят на входное изображение для получения сегментируемых элементов и контуров.

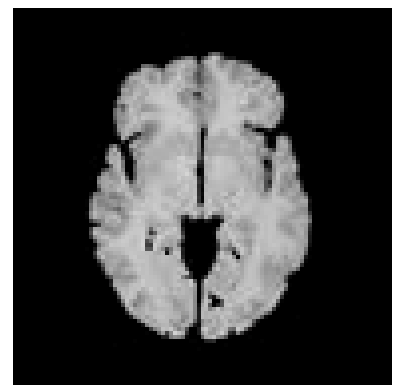
В целях повышения эффективности и точности сегментации применяют несколько атласов [86, 87]. Мультиатласная сегментация начинается с сопоставления входного изображения со всеми атласами, после чего происходит слияние разметок атласов в целях получения контура изображения. Также, данные о количестве исходных изображений могут быть объединены в один средний атлас



(а) Исходное изображение



(б) Детектирование границ



(в) Сегментация

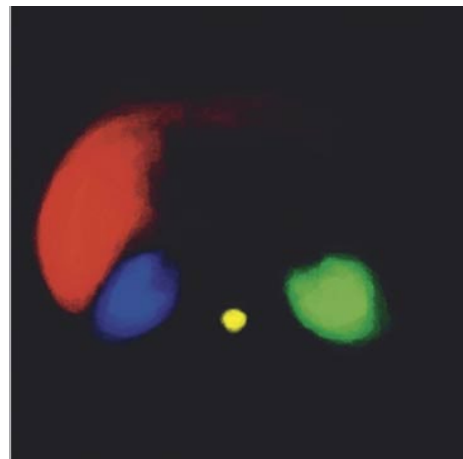
Рисунок 1.10 – Результат работы метода выращивания регионов

или, так называемый, вероятностный атлас, который содержит значения вероятностей для каждого региона. Так, например, Н. Park в своих работах [88, 89] проводит сегментацию на основе вероятностного атласа, а результат такой сегментации представлен ниже на рисунке 1.11.

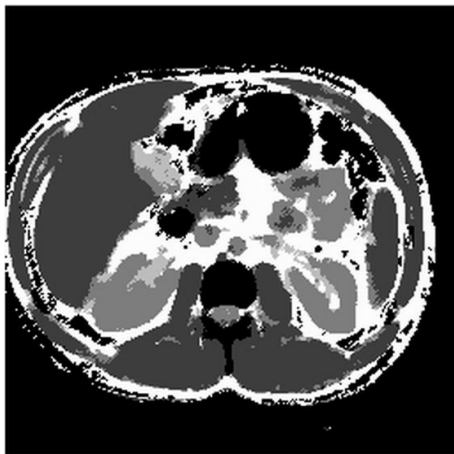
Подобными исследованиями в сфере сегментации медицинских данных на основе атласов занимались А.Ю. Зубов и О.В. Сеньюкова. В своей работе они предлагают модификацию метода слияния разметок нескольких атласов для увеличения точности сегментируемых регионов [90]. Важно отметить, что выборка являлась небольшой, в ней использовались 43 изображения магнитно-резонансной томографии мозга. Исходные данные были взяты из открытой базы Internet Brain Segmentation Repository. Полученные результаты сегментации и точность предлагаемого метода указаны в таблице 1.1 и на рисунке 1.12. Стоит отметить высокую точность сегментации практически всех анатомических структур мозга.



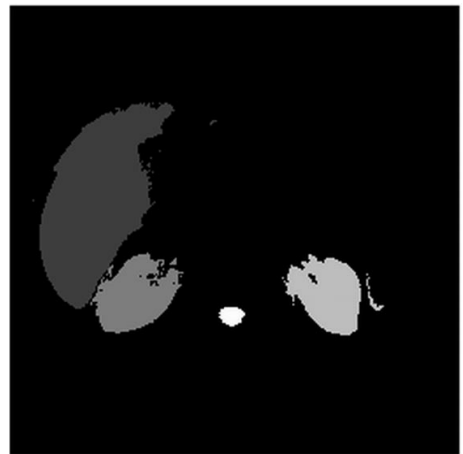
(а) Исходное изображение



(б) Атлас



(в) Сегментация без использования атласа



(г) Сегментация с использованием атласа

Рисунок 1.11 – Сегментация на основе вероятностного атласа

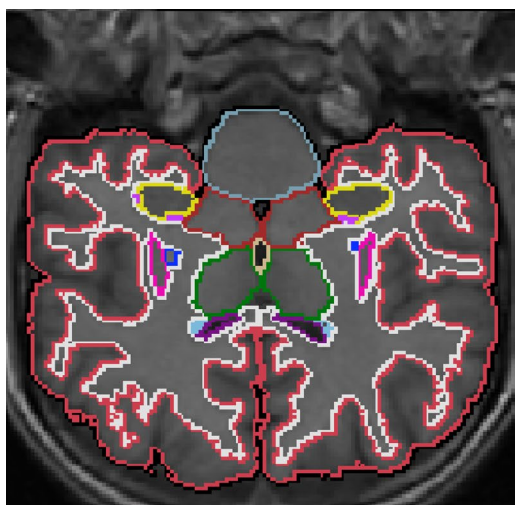
Таблица 1.1 – Точность выделения анатомических структур мозга

Анатомическая структура	Индекс Дайса
Мозжечок	0.89 ± 0.05
Желудочек	0.90 ± 0.02
Средний мозг	0.85 ± 0.007
Паллидиум	0.74 ± 0.009
Кровеносные сосуды	0.20 ± 0.001

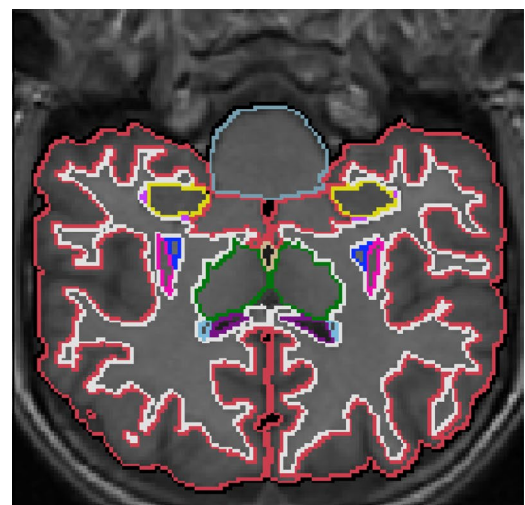
Однако главным недостатком такого подхода является время работы. В методе Зубова А.Ю. и Сенюковой О.В. оно составляет около 190 секунд для одного изображения, что приемлемо только для обработки изображений, используемых при диагностике заболеваний.

Другим примером сегментации анатомических структур является подход, описанный в работе [91]. В данной работе проводилось контурирование лёгких на основе не жёсткой регистрации с атласами. После осуществления сегментации проведена оценка контурированного изображения по нескольким параметрам. Исходные данные для алгоритма взяты из нескольких баз данных, а именно JSRT, Montgomery и India. Результат сегментации области интереса и точность данного подхода отражены в таблице 1.2 и на рисунке 1.13 соответственно.

Сравнивая метод сегментации на основе атласов с другими подходами, можно сказать, что такая техника обладает рядом достоинств, одним из которых



(а) Экспертная разметка



(б) Мульти-атласная сегментация

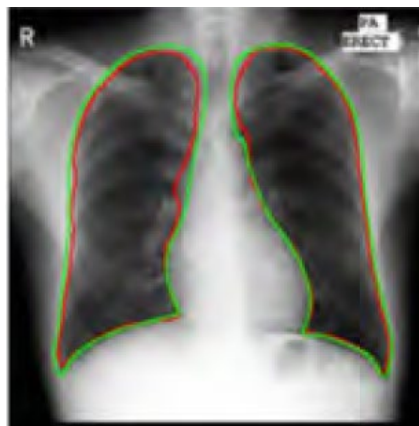
Рисунок 1.12 – Сегментация на основе атласов

Таблица 1.2 – Оценка качества сегментации на основе не жёстко регистрированных атласов

База данных	Индекс Дайса	Дистанция между контурами
JSRT	$0.97 + 0.008$	$1.32 + 0.32$
Montgomery	$0.96 + 0.018$	$1.60 + 0.74$
India	$0.95 + 0.028$	$2.57 + 1.46$



Индекс Дайса: 0.975



Индекс Дайса: 0.948



Индекс Дайса: 0.946



Индекс Дайса: 0.974



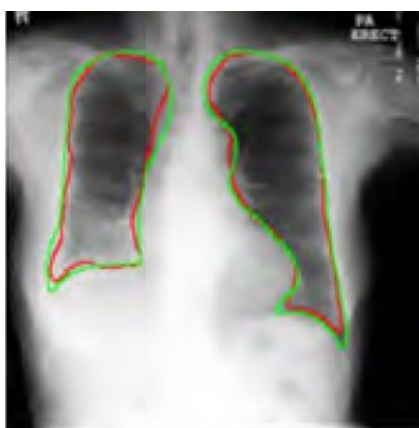
Индекс Дайса: 0.938



Индекс Дайса: 0.947



Индекс Дайса: 0.959



Индекс Дайса: 0.942



Индекс Дайса: 0.935

(а) База данных JSRT

(б) База данных Montgomery

(в) База данных India

Рисунок 1.13 – Сегментация на основе не жёстко регистрированных атласов

является достаточно высокая точность сегментации. Это достигается за счёт использования априорных контурированных данных. Однако при использовании методов на основе атласов возникают проблемы с детектированием структур малого размера, что приводит к большой ошибке сегментации. Чаще всего размер медицинского инструмента или небольшой опухоли представляет 2-5% от размера всего изображения. Ещё одним минусом атласной техники является время сегментации. При использовании многоатласной техники время сегментации увеличивается кратно, что делает маловероятным использование данного подхода в качестве визуализации медицинских операций в реальном масштабе времени.

На сегодняшний день методы нейронных сетей применяются практически повсеместно, в том числе и при сегментации изображений. Одним из примеров использования нейронных сетей является работа M. Lai [92]. Исходными данными в этой работе являются магнитно-резонансные изображения для гиппокампа мозга. В ходе выполнения данного исследования сравнивалась эффективность сегментации разных свёрточных архитектур при варьировании таких гиперпараметров, как количество слоёв, размеры свёрток для двухмерных, трипланарных и трёхмерных данных. Результат работы нейронной сети отражён ниже на рисунке 1.14.

Аналогичное исследование было проведено для сегментации лёгких на изображениях компьютерной томографии. После 12 часового обучения

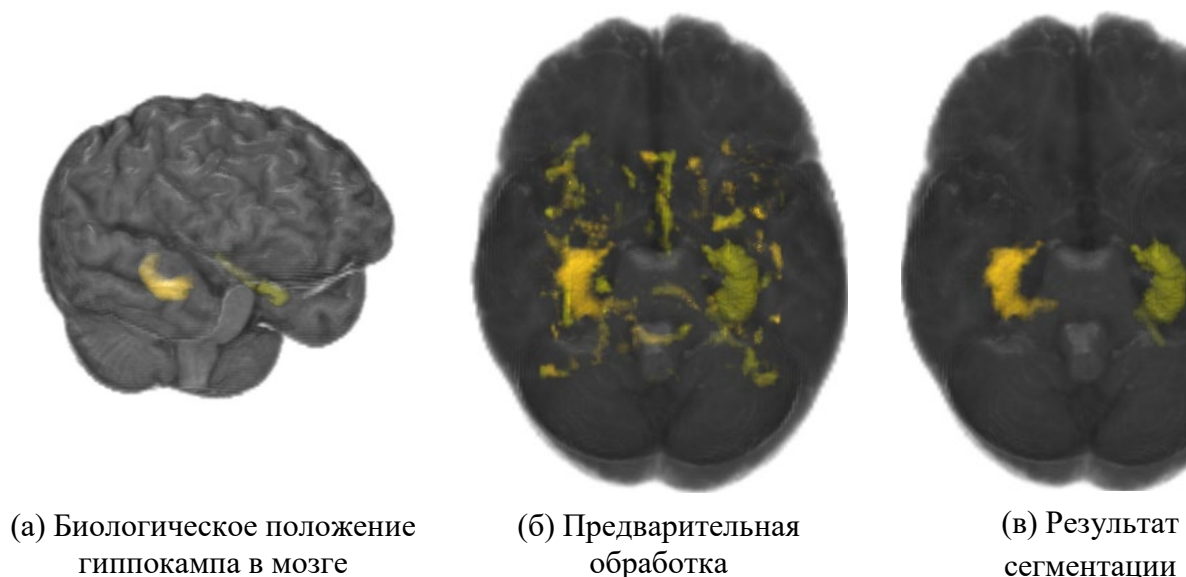
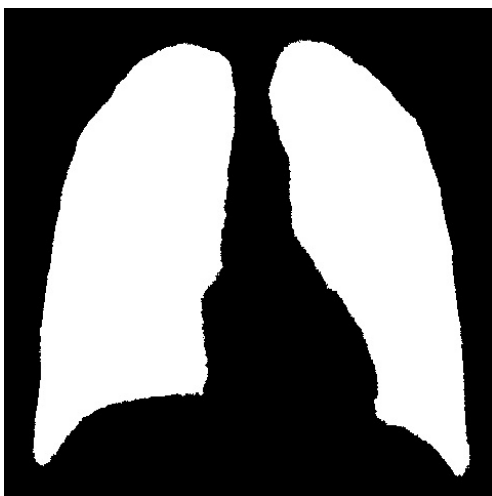


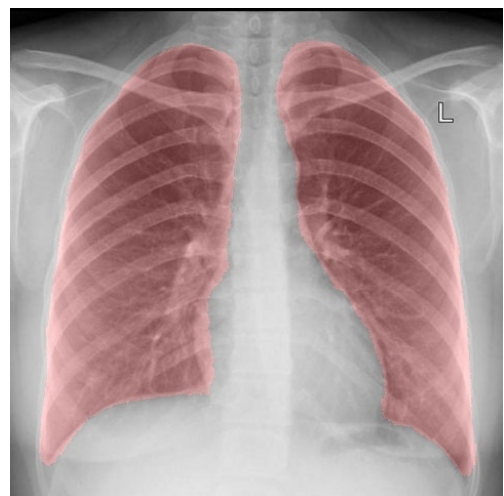
Рисунок 1.14 – Сегментация гиппокампа на основе нейронной сети

нейронной сети на 480 исходных изображениях точность сегментации оказалась на достаточно высоком уровне (см. рисунок 1.15). В целом, подход с использованием нейронных сетей показывает высокую работоспособность и гибкость для широкого спектра задач. Одним из достоинств данного подхода является возможность проведения параллельных вычислений. Так, например, при реализации данной методики на видеокарте NVIDIA Tesla K80, обработка одного слайса происходит в среднем за 12 мс, а на встраиваемой платформе NVIDIA Jetson TK1 – за 1.4 секунды.

Другим классом методов сегментации являются методы сегментации, основанные на активных контурах. Идея таких контуров заключается в их передвижении по изображению и минимизации их энергии. Контур может двигаться в сторону объектов интереса на изображении, имея при этом низкое значение внешней энергии на самих объектах и высокое значение в других областях [93]. Один из примеров сегментации на основе такого подхода описан в совместной работе датских и шведских исследователей [94]. Авторы детектировали и затем диагностировали степень поражения почек на основе изображений, полученных с помощью сцинтиграфии. Процесс сегментации и классификации почек состоял из нескольких шагов: сегментация, удаление фона, анализ объектов переднего фона и классификация. В качестве основных методов для решения поставленной задачи были использованы искусственные нейронные



(а) Бинарная маска
сегментации

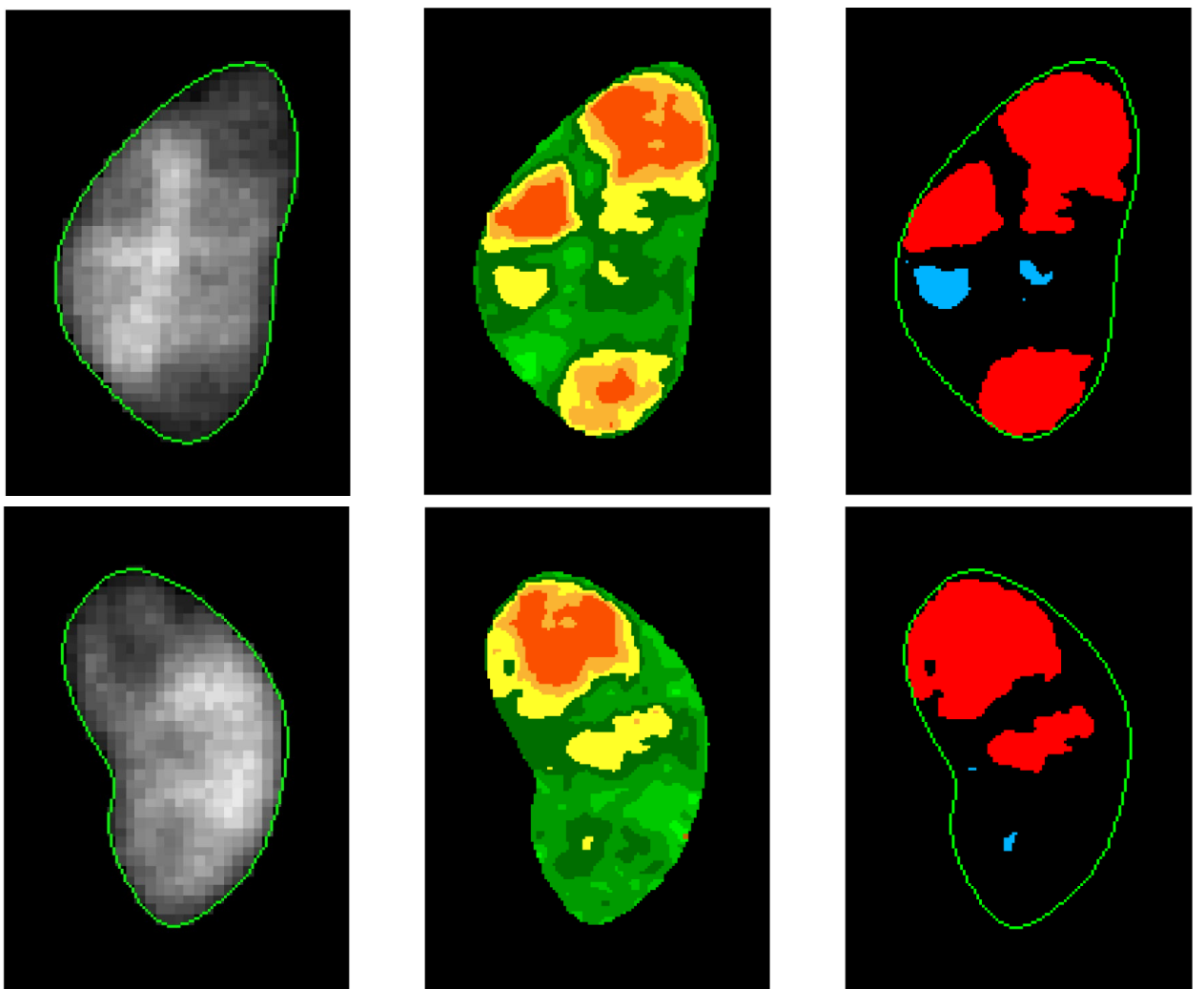


(б) Маска, наложенная на исходное
изображение

Рисунок 1.15 – Сегментация лёгких на основе сверточной нейронной сети

сети, линейный и квадратичный анализ для классификации и активные модели формы для контурирования, которые могут многократно динамически деформироваться для подгонки к форме объекта. В результате исследования получена автоматическая система для проведения сегментации и постановки диагноза. Пример работы данной системы отражён на рисунке 1.16. Стоит отметить, что из-за сильной каскадности алгоритма и большого количества обрабатываемых блоков время обработки одного изображения варьируется в диапазоне от 5 до 9 сек.

К аналогичному исследованию можно отнести работу [95], в которой учёные проводили сегментацию левого желудочка на основе изображений, полученных с помощью эхокардиографии (см. рисунок 1.17). В данном случае, основным методом сегментации являлся snake-алгоритм. Несмотря на



(а) Детектирование почки

(б) Статистическая карта

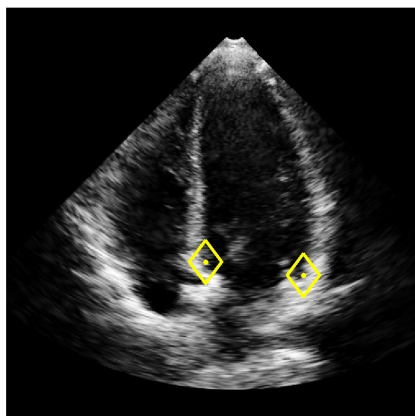
(в) Результат классификации

Рисунок 1.16 – Сегментация и классификация скintiграфических изображений

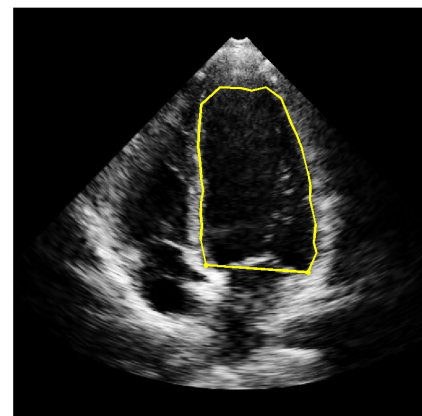
относительно высокую эффективность и точность данного подхода, он является полуавтоматическим, т.к. для инициализации процесса сегментации требуется участие пользователя, которому необходимо выставлять якорные точки на изображениях.

Другим классом методов по распознаванию и сегментации объектов являются методы на основе графов [96]. Главным минусом таких методов является трудоёмкость процедуры сортировки всех рёбер, которая выполняется за $O(n * \log n)$, где n – количество рёбер в графе. Таким образом, для изображения $N * M$ пикселей количество рёбер составит $|n| = 4 * N * M$. Данный недостаток сильно ограничивает применение данного метода сегментации на практике.

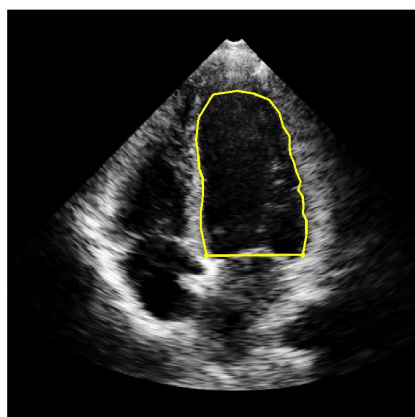
Зачастую для задач сегментации, и в частности кластеризации, часто используется итеративный метод k -средних, описанный в работах [13] и [97]. Алгоритм k -средних может использоваться как самостоятельно, так и при совместной работе с другими алгоритмами обработки изображений. Например, в



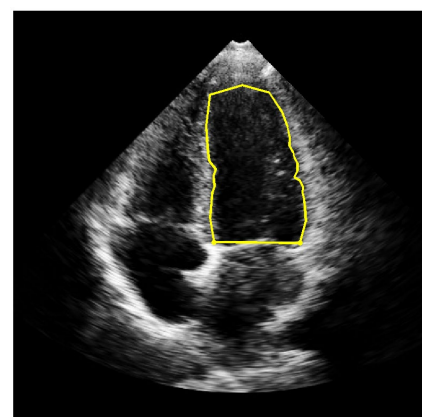
(а) Якорные точки



(б) Сегментация



(в) Сегментация

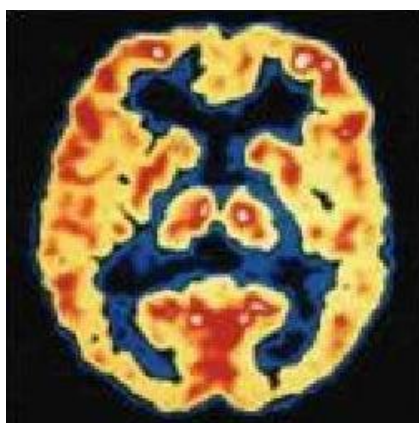


(г) Сегментация

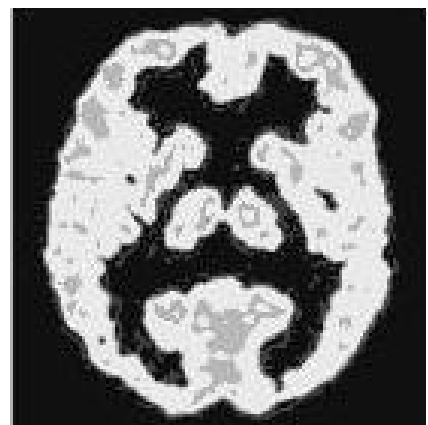
Рисунок 1.17 – Сегментация на основе snake-алгоритма

работе [98] исследователи провели сегментацию и кластеризацию данных для определения и постановки диагноза болезни Альцгеймера на основе изображений позитронно-эмиссионной томографии. Результат сегментации данного подхода отражён на рисунке 1.18.

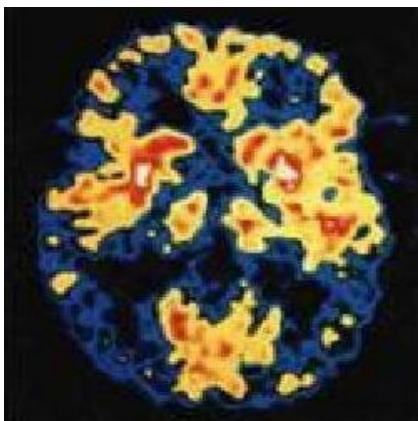
Другим примером внедрения алгоритма k -средних для сегментации является работа [99]. В своей работе авторы использовали алгоритм k -средних вместе с методом сегментации на основе водоразделов. Результат совместной работы алгоритмов отражён на рисунке 1.19. Исследователи провели сравнение совместной работы алгоритма k -средних с улучшенным и стандартным методом управляемого водораздела. В результате, при использовании стандартного метода водоразделов на финальном изображении получалось 259 регионов, а при использовании улучшенного метода водоразделов – 172 региона. Таким образом, алгоритм позволяет в какой-то степени избежать пересегментации изображения. Одной из особенностей метода k -средних является то, что алгоритм всегда



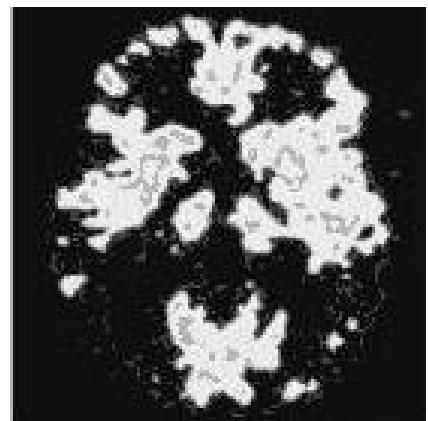
(а) Исходное изображение



(б) Сегментация



(в) Исходное изображение



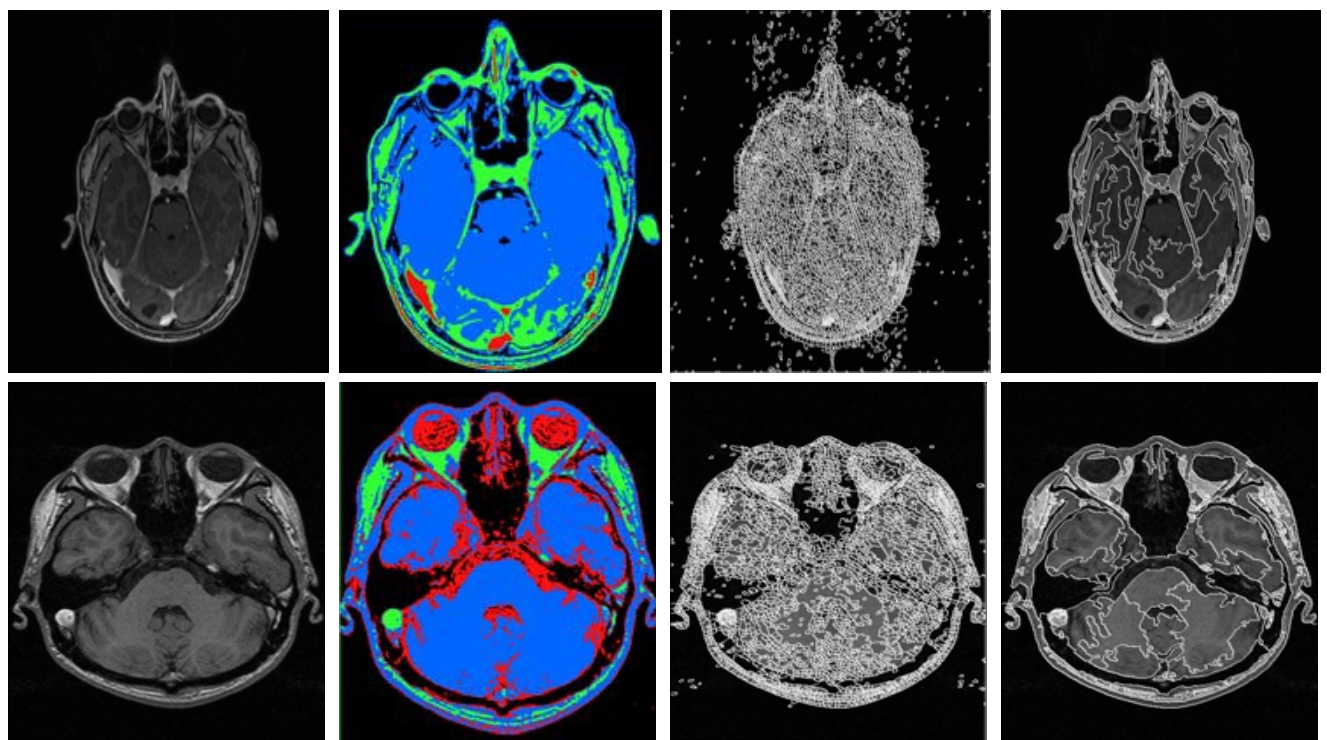
(г) Сегментация

Рисунок 1.18 – Сегментация алгоритмом k -средних

сходится, однако не всегда даёт оптимальный результат, а качество решения всегда зависит от начального количества кластеров и значения k . Значение k может быть получено как случайно или на основе определённого алгоритма, так и вручную.

1.4 Программные продукты для обработки медицинских данных

Решения для обработки медицинских данных сегодня представлены различными программными продуктами, которые отличаются по функциям, направленности и сложности. Можно выделить два ключевых способа реализации подобных программных комплексов. Первый способ предполагает, что комплексы обработки медицинских изображений встраиваются в аппаратную часть, а затем поставляются совместно с ней. Второй способ заключается в реализации и установке подобных систем в качестве самостоятельного программного обеспечения. Более актуальными и перспективными системами являются те, что поставляются и реализуются независимо от аппаратной части, так как они позволяют выполнять обработку изображений независимо от оборудования, на основе которого они были получены. Необходимо заметить, что зачастую



(а) Исходное изображение

(б) Кластеризация методом k -средних

(в) Сегментация методом водоразделов

(г) Окончательная сегментация

Рисунок 1.19 – Результат сегментации и кластеризации МРТ изображений

подобные программные продукты разрабатываются под разные операционные системы и устройства. На сегодняшний день акцент ставится на кроссплатформенность и мобильность программного обеспечения. На данный момент существует множество программных продуктов, предназначенных для операционных систем семейства Windows, Linux и OS X. В целях обеспечения переносимости такие приложения часто проектируются под платформы Java или .NET. Поэтому одним из наиболее важных вопросов является обеспечение совместимости систем. В свою очередь диверсификация и непрерывное развитие аппаратных средств, предназначенных для медицинской диагностики, вынуждают производителей аппаратного и программного обеспечения применять универсальные стандарты и протоколы передачи информации. К популярным программным решениям по обработке и визуализации медицинских данных можно отнести:

- QLab от Philips (специализированный программный пакет для количественного анализа сердца и сосудов [100]).
- Syngo от Siemens (современный PACS и платформа для анализа медицинских изображений [101]).
- Innova EPVision 2.0 от General Electric (система трёхмерной визуализации [102]).
- TRENION 3D HD от Zeiss (система трёхмерной визуализации [103]).

Решения, описанные выше и подобные им, изначально встраиваются в системы этих компаний и идут в комплекте с оборудованием компании-производителя. Данный факт значительно ограничивает спектр применения подобного программного обеспечения. Тем не менее, существуют компании, занимающиеся разработкой программных продуктов, которые позволяют расширить список доступных функций оборудования и улучшить качество визуализации для производителей оборудования. Однако коммерческая составляющая деятельности подобных групп не позволяет активно внедрять новые разработки. Ряд подобных групп и их решений отражён ниже:

- Open Inventor от FEI VSG (набор инструментов для анализа и обработки медицинских изображений).
- LEADTOOLS (набор инструментов для обработки медицинских изображений).
- RSIP Vision (набор инструментов для обработки и анализа медицинских изображений).

Помимо коммерческих компаний, множество некоммерческих исследовательских групп реализуют проекты по обработке, сегментации и реконструкции анатомических структур и медицинских инструментов, а также повышению качества визуализации. Ниже отражён список подобных научных групп, которые ведут свою деятельность на базе университетов:

- Harvard Biorobotics Lab (сегментация митрального клапана на основе четырёхмерных ультразвуковых исследований [104, 105], обработка трёхмерных ультразвуковых изображений и трекинг медицинских приборов [106]).
- Hamlyn Centre Imperial College London (адаптивная визуализация сердечно-сосудистой системы [107], визуализация потока в миокарде [108]).
- University of North Carolina (инструментарий по сегментации и регистрации [109]).
- Technical University of Wien (проблемы трёхмерной медицинской визуализации [110], фильтрация объёмных потоков данных [111]).
- State University of Campinas (инструментарий трёхмерной визуализации медицинских данных [112]).
- Otto-von-Guericke University (оценка восприятия трёхмерной медицинской визуализации [113]).

Важным будет упомянуть программные решения с открытым исходным кодом, которые часто используются в исследовательских и диагностических целях. Наиболее популярными решениями с открытым исходным кодом являются:

- MITK от German Cancer Research Center (система программного обеспечения, использующаяся для обработки медицинских изображений [114])

- 3D Slicer от Harvard Medical School (программная платформа, предназначенная для обработки медицинских изображений и трёхмерной визуализации [115]).
- ITK-SNAP от University of Pennsylvania и University of Utah (интерактивное программное приложение для предварительной обработки и визуализации медицинских данных [116]).

Некоторые из задач динамической обработки и визуализации медицинских данных частично решаются современными установками, например, такими как Medison Accuvix A30, Philips HD9 Ultrasound System и др. Ряд крупных коммерческих компаний, таких как Philips, Siemens, General Electric и Hitachi поставляют на рынок программно-аппаратные комплексы, способные выполнять не только сбор трёхмерных данных в режиме реального времени, но и их частичную предварительную обработку. Ярким примером являются современные трёхмерные ультразвуковые датчики. При использовании подобной технологии шумоподавление и реконструкция происходят за счёт встроенного программного обеспечения. Однако, врачи и клиницисты работают с трёхмерным эхо-сигналом, интерпретация которого для неподготовленного пользователя является затруднительной, так как трёхмерная эхокардиография не отображает классических парастернальной, апикальной и других проекций. Компании Philips, в частности, разрабатывает программный комплекс, который извлекает из трёхмерного эхо-сигнала спектр двумерных проекций в режиме реального времени, чтобы упростить работу оператора. Последнее подчёркивает сложность работы с трёхмерной эхокардиографией в чистом виде. Более того, выходные трёхмерные данные не содержат координат инструментов и расстояний до области интереса. Вычисления других характеристик, таких как текущий объем камеры, фракции выброса и толщина стенки также производится отдельно на уже скомпонованной выборке данных.

В связи с вышеописанным становится очевидным, что на сегодняшний день существует множество методов, позволяющих выполнять сегментацию медицинского инструмента или анатомической структуры. Однако большинство

подобных методов чаще всего носят исследовательский характер и на практике не применяются. Стоит также отметить множество ограничений у разрабатываемых методов, в том числе их комплексность, тонкую настройку и слабую валидацию на реальных данных. Несмотря на существующие разработки и программные решения, остаётся необходимость в развитии сферы обработки и визуализации медицинских данных с использованием аппаратно-независимых и универсальных методов и подходов.

1.5 Математическая формулировка задачи

Сегментация изображения представляет собой разбиение совокупности пикселей/вокселей на связанные определённым образом области. Как правило, операция сегментации проводится на основе интенсивности для одноцветного изображения и цветовых компонент для цветного. Тем не менее, конечной целью сегментации является разбиение всего поля зрения N на области объектов $N_1, N_2, \dots, N_S$ и области фона N_B . Математически задача сегментации выражается следующим образом:

$$\begin{cases} f[x, y, p(x, y)] \rightarrow f(n, p'(n)), \\ f(N^k) \rightarrow N^l, \\ l < k, \end{cases} \quad (1.1)$$

где x, y – координаты пикселя исходного изображения, $p(x, y)$ – интенсивность пикселя, n – регион интереса, $p'(n)$ – признак или группа признаков сегментируемой области, k – количество уровней интенсивности изображения, l – количество сегментов после сегментации.

1.6 Выводы по первой главе

1. Анализ современного состояния науки в сфере эффективной и быстрой сегментации области интереса позволяет утверждать, что данная тематика так и не получила качественного освещения в отечественных и зарубежных работах, несмотря на большое количество публикаций. Причина недостаточного внимания исследователей к обработке и анализу медицинских изображений обусловлена различными факторами, а именно: сложным типом данных, аппаратной

зависимостью структуры файлов хранения для медицинской техники, большим объёмом информации, высокими требованиями к точности сегментированных изображений, временем обработки. Важным будет отметить, что модальности, подобные эхокардиографии, накладывают более жёсткие ограничения на методы по распознаванию и анализу данных. Кроме того, медицинские исследования часто осложняются различными факторами, начиная от цифрового шума и произвольных движений самого пациента (к примеру, во время дыхания) и заканчивая нестандартными методами проведения исследований врачом.

2. Литературный обзор методов сегментации позволил выделить подход с использованием методов глубокого обучения как наиболее эффективный. Одним из преимуществ нейронных сетей является их способность решения той или иной задачи в условиях, когда зависимость между входом и выходом неизвестна. Нейронные сети также устойчивы к шумам и самостоятельно выбирают наиболее релевантные признаки, отсеивая при этом малоинформативные. При неблагоприятных условиях выходная точность математической модели, основанной на нейронной сети, падает незначительно, что говорит об их отказоустойчивости. Так, например, получение информации затрудняется, если отсутствует связь с нейроном. Однако, принимая во внимание распределённый характер хранения информации и большое количество нейронов, только потеря связи с большим количеством нейронов может существенно повлиять на работоспособность модели. В дополнение стоит отметить то, что за счёт параллелизма обработки данных быстрое действие нейронных сетей остаётся на высоком уровне.

3. Выявлена потребность в развитии методов, алгоритмов и моделей, позволяющих выделять медицинский инструмент на трёхмерных данных. Последнее объясняется тем, что низкая надёжность и точность средств трёхмерной визуализации области вмешательства, а именно отсутствие обратной связи между средствами визуализации и средствами позиционирования инструмента является одной из причин повреждения мягких тканей. С учётом особенностей адаптации

может быть решена проблема визуализации и контроля медицинского инструмента при выполнении эндоваскулярных вмешательств.

4. Оценка программно-аппаратных средств показала, что большой спектр реконструктивных вмешательств сегодня осуществляется только на открытом «сухом» сердце с применением искусственного кровообращения. Последнее вызывает множество рисков при проведении интервенций. Несмотря на тот факт, что сегодня существуют решения по навигации медицинских устройств, например, система CARTO3, подобные решения весьма сложны технологически, обладают высокой стоимостью и носят узкоспециализированный характер. Более того, такая система выполняет сегментацию лишь некоторой части анатомической структуры, которая находится в статическом положении. Ни одна из современных систем не решает задачу визуализации инструмента, а также не предоставляет дополнительной информации о нём.

2 Методы и алгоритмы сегментации области интереса

В данной главе представлены разработанные методы и алгоритмы, используемые для решения задачи сегментации медицинского инструмента и анатомических структур при проведении диагностики и минимально инвазивных вмешательств. В связи с малой репрезентативностью и недостаточностью данных, в том числе размеченных, дополнительно был разработан алгоритм синтеза трёхмерных данных, описанный в пункте 2.2. Ключевой частью работы является глубокая свёрточная нейронная сеть, которая представлена в пункте 2.4. Архитектура предлагаемой нейронной сети основана на V-net архитектуре глубокого свёрточного энкодер-декодера и плотно-соединённых слоях. Для выбора оптимальных гиперпараметров в пункте 2.6 предложен алгоритм на основе статистической значимости сравниваемых моделей нейронных сетей и значений их t-статистик.

2.1 Исходные данные и оборудование

Исходные данные были получены средствами эхокардиальной трёхмерной эхокардиографии при выполнении кардиохирургических операций у Йоркширских свиней весом 65-75 кг. Медицинский инструмент (абляционный катетер) был введён в полость левого желудочка (ЛЖ). Трансторакальный датчик X7-2t был размещён на эпикарде верхушки ЛЖ таким образом, чтобы получить плоскость сечения максимально похожую на стандартную, а именно на парастернальную плоскость по короткой оси. В стандартных исследованиях с помощью трансторакальной эхокардиографии (ТТЭ) данная плоскость сечения используется для визуализации желудочков, их верхушек и папиллярных мышц. Датчик помещался непосредственно на верхушку ЛЖ и слегка наклонялся к правому желудочку (ПЖ). Подобная ориентация датчика позволила фиксировать непрерывную форму анатомических структур, а именно ЛЖ, ПЖ и межжелудочковой перегородки.

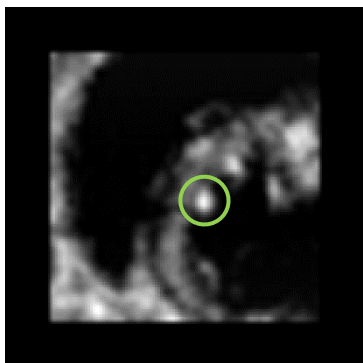
Помимо трансторакального датчика X7-2t, для получения данных использовались эхокардиограф Philips iE33 и программное обеспечение PMS5.1

Ultrasound Software. В процессе сбора информации было получено 75 трёхмерных изображений с катетером. Каждое изображение представляло из себя 208 слайсов в оттенках серого с размерами 176*176 пикселей. Примеры исходных трёхмерных изображений, взятых с трёх разных экспериментов, указаны на рисунке 2.1. Стоит заметить, что катетер слабо различим на двухмерных данных. Сама область катетера находится внутри зелёного круга/эллипса на примерах рисунка 2.1. Дополнительно были получены данные пустых полостей человеческого сердца, где медицинский инструмент отсутствовал. Общее количество данных с пустыми полостями сердца составило 600 трёхмерных изображений. Примеры таких изображений отражены на рисунке 2.2.

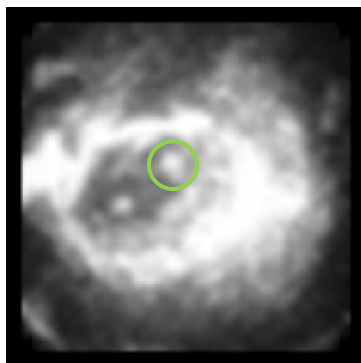
Данные были получены в Бостонском Детском Госпитале, где экспериментальные протоколы для тестирования животных были одобрены IACUC (сокр. от Institutional Animal Care and Use Committee). Все животные получали гуманный уход в соответствии с Руководством 1996 года по уходу и использованию лабораторных животных, рекомендованным Национальными институтами здравоохранения США [117].

Предварительная обработка и разметка данных осуществлялась на компьютере со следующими характеристиками: Intel Core i7-4820K CPU 3.7GHz, Nvidia GeForce 960 GT 2, 16 GB RAM и Windows 10. Обучение нейронных сетей осуществлялось на компьютере со следующими характеристиками: Intel Core i7-6850K CPU 3.6GHz, Nvidia TITAN X и Nvidia TITAN Xp, 32 GB RAM и Linux 14. Дополнительно для обучения моделей и сокращения времени обучения использовались инстансы P2 (на базе Nvidia Tesla K80) и P3 (на базе Nvidia Tesla V100) от Amazon Web Services.

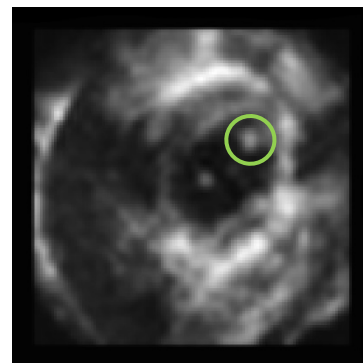
Эксперимент 1



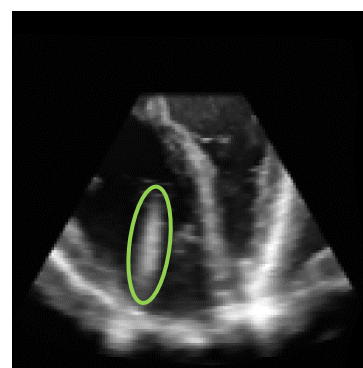
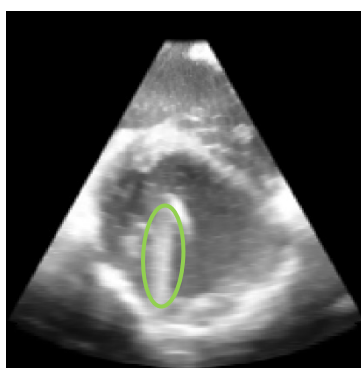
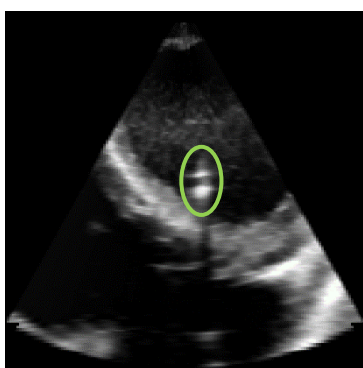
Эксперимент 2



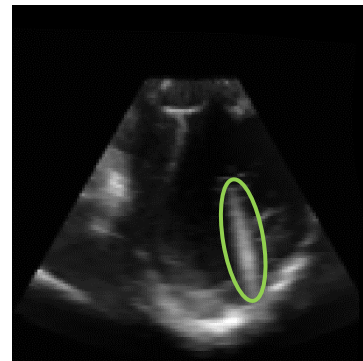
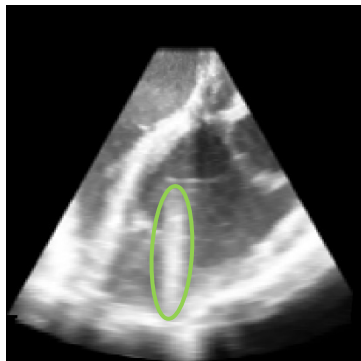
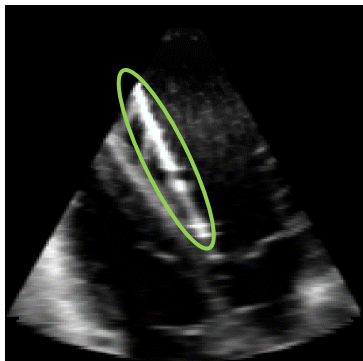
Эксперимент 3



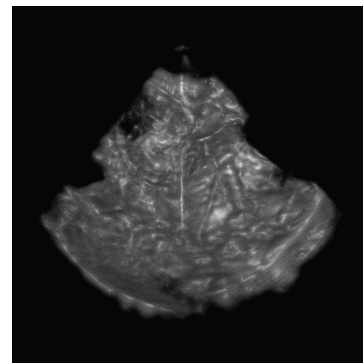
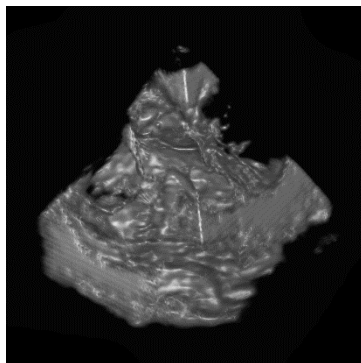
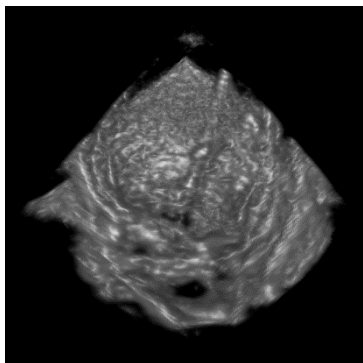
(а) Аксиальная плоскость



(б) Сагиттальная плоскость



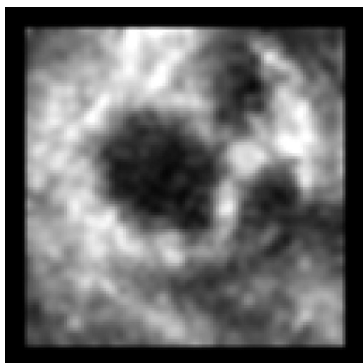
(в) Корональная плоскость



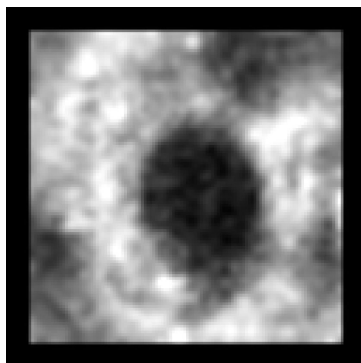
(г) 3D визуализация

Рисунок 2.1 – Пример трёхмерных изображений с введённым в полость ЛЖ катетером

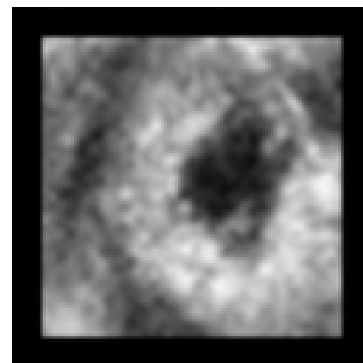
Эксперимент 1



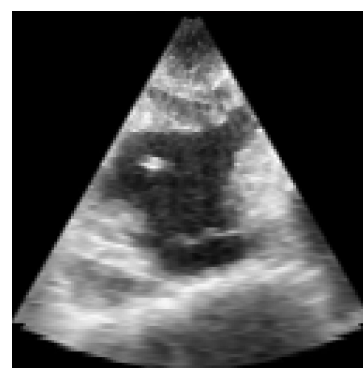
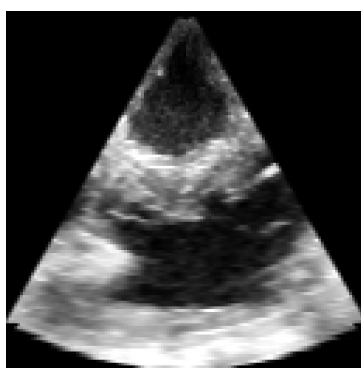
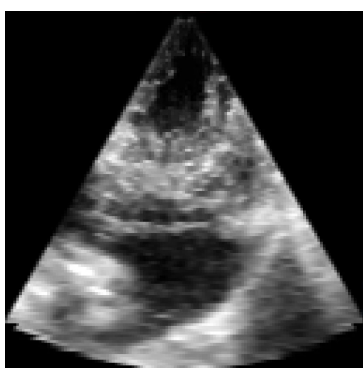
Эксперимент 2



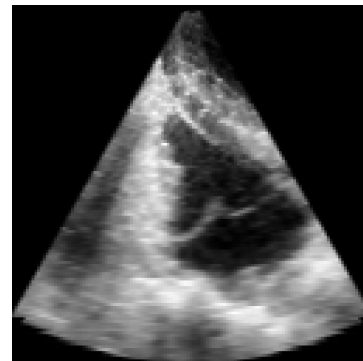
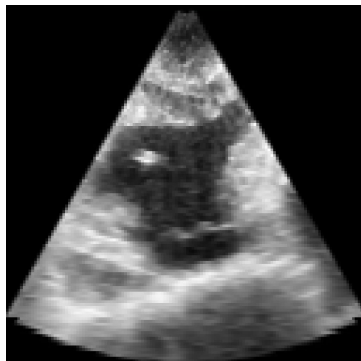
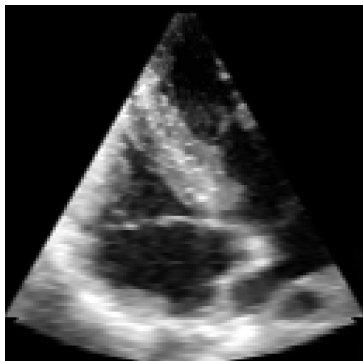
Эксперимент 3



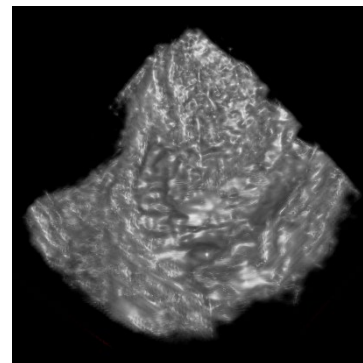
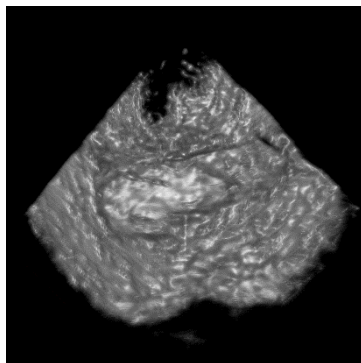
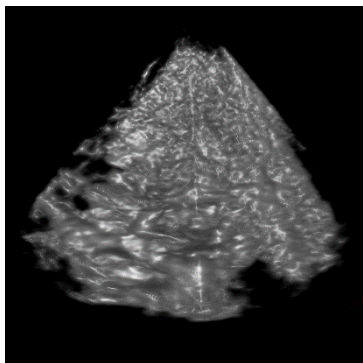
(а) Аксиальная плоскость



(б) Сагиттальная плоскость



(в) Корональная плоскость



(г) 3D визуализация

Рисунок 2.2 – Пример трёхмерных изображений с пустыми полостями сердца

Для создания и обучения модели было размечено 75 трёхмерных изображений, каждый из которых состоит из 208 слайсов (15600 двухмерных изображений). Дополнительно к этому набору данных было синтезировано ещё 175 трёхмерных изображений (36400 двухмерных изображений). С целью создания дополнительного набора данных был разработан алгоритм синтеза искусственных изображений эхокардиографии, описанный ниже в пункте 2.2. Для синтеза использовались анонимизированные данные, полученные при диагностике сердечных отклонений человека. Дополнительно для расширения обучающей, валидационной и тестовой выборок к данным применялась аугментация с пятью различными трансформациями (повороты на 90° и зеркальные отражения). Распределение данных по типам отражено на рисунке 2.3. Полный набор, включающий исходные и аугментированные данные, состоит из 1500 изображений. Для обучающей выборки использовались 1200 (80%) изображений, для валидационной выборки – 150 (10%) изображений, для тестовой выборки – 150 (10%) изображений.

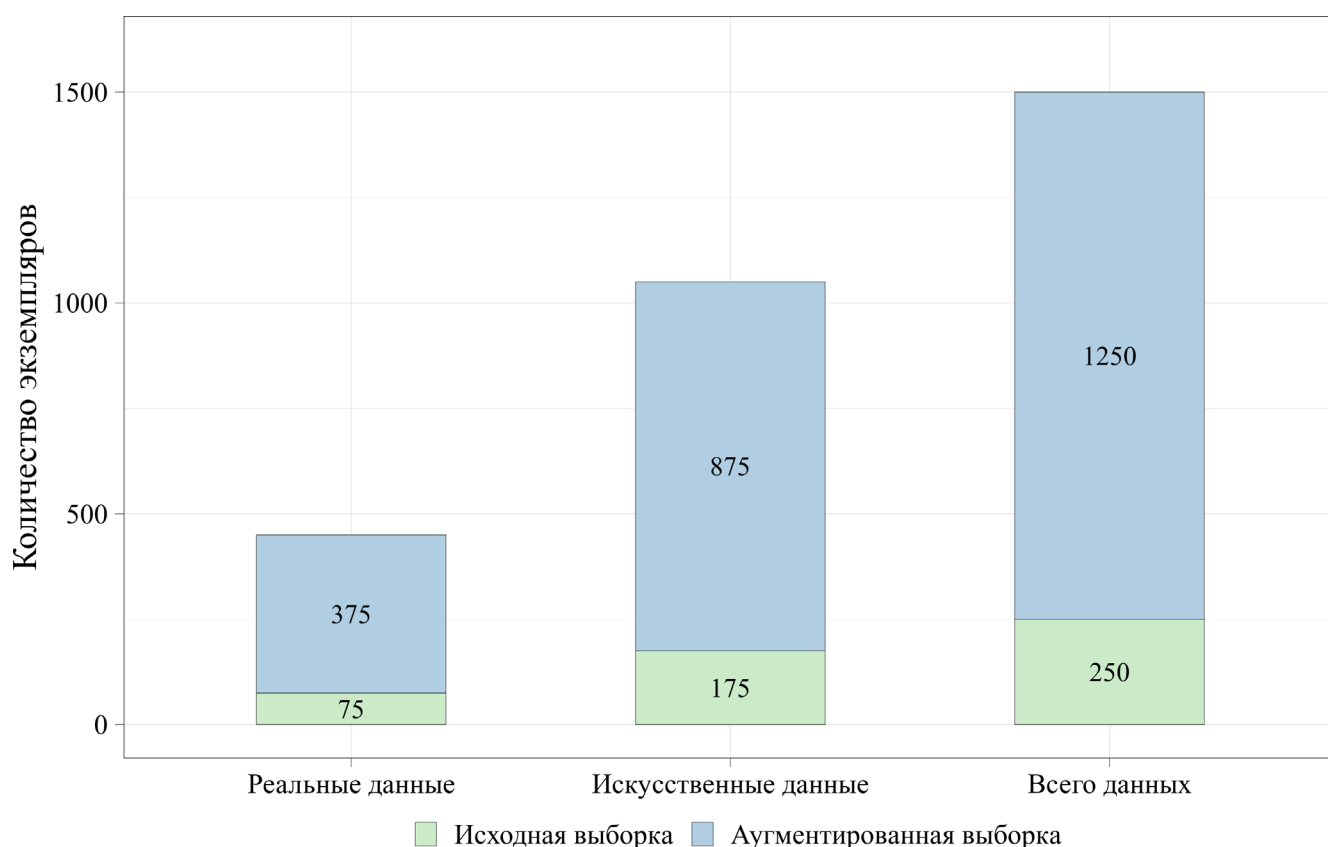


Рисунок 2.3 – Распределение исходных данных

2.2 Алгоритм синтеза данных

Представленный алгоритм используется для синтеза, наложения и аугментации данных. Его важной особенностью является возможность генерации и аугментации трёхмерных коррелирующих данных, что является важной составляющей для медицинской отрасли. Последнее обусловлено тем, что множество графических данных в медицине представлены в статическом и динамическом трёхмерном форматах. Важным будет отметить, что современные методы аугментации данных не учитывают корреляции соседних 2D и/или 3D изображений, а просто накладывают ту или иную трансформацию. В свою очередь, предлагаемый алгоритм призван трансформировать объект исследования с учётом взаимосвязи данных. Последнее позволит минимизировать рутинность процесса разметки данных за счёт автоматизации. Важным будет отметить, что обучение нейронных сетей для локализации и сегментации медицинского инструмента на трёхмерных данных эхокардиографии требует наличия большой обучающей выборки. Отсутствие изображений в достаточном количестве является одной из ключевых проблем. Одним из решений данной проблемы является генерация новых данных из ранее полученных. В основу такой генерации заложена концепция трансформации изображений медицинского инструмента на исходных данных и его перенос на реальные данные сердца, где медицинский инструмент отсутствует. В качестве трансформации использованы изгиб, скручивание, масштабирование и смещение.

Таким образом, для генерации искусственной трёхмерной эхокардиографии сердца, содержащей дистальный конец катетера, необходимы следующие входные данные: трёхмерное изображение катетера в оттенках серого, трёхмерное изображение сердца в оттенках серого, начальная точка и ориентация катетера. Результатом работы генератора являются трёхмерные изображения, содержащие катетер внутри анатомических структур сердца. Таким образом, модель генератора имеет 4 входа и может быть концептуально представлена так, как это указано на рисунке 2.4. Стоит добавить, что пара векторов $(\bar{P}, \bar{V})$ задаёт положение и ориентацию катетера в новом пустом изображении.

Для осуществления переноса и трансформации катетера была разработана система координат, в основу которой положен кубический сплайн. Такая система позволяет гибко работать с точками катетера и осуществлять все необходимые трансформации. Сплайн проходит через продольную ось катетера и задаёт его конфигурацию, а все точки катетера выражаются относительно сплайна через координаты (ρ, φ, z) по принципу цилиндрической системы координат. Отличием от классической цилиндрической системы координат является то, что роль оси Z в данной системе выполняет сплайн, а не прямая. Таким образом, z координата является длиной сплайна от его начала до точки O на сплайне, в которой строится ось X и определяется угол и удаление некоторой точки катетера M (см. рисунок

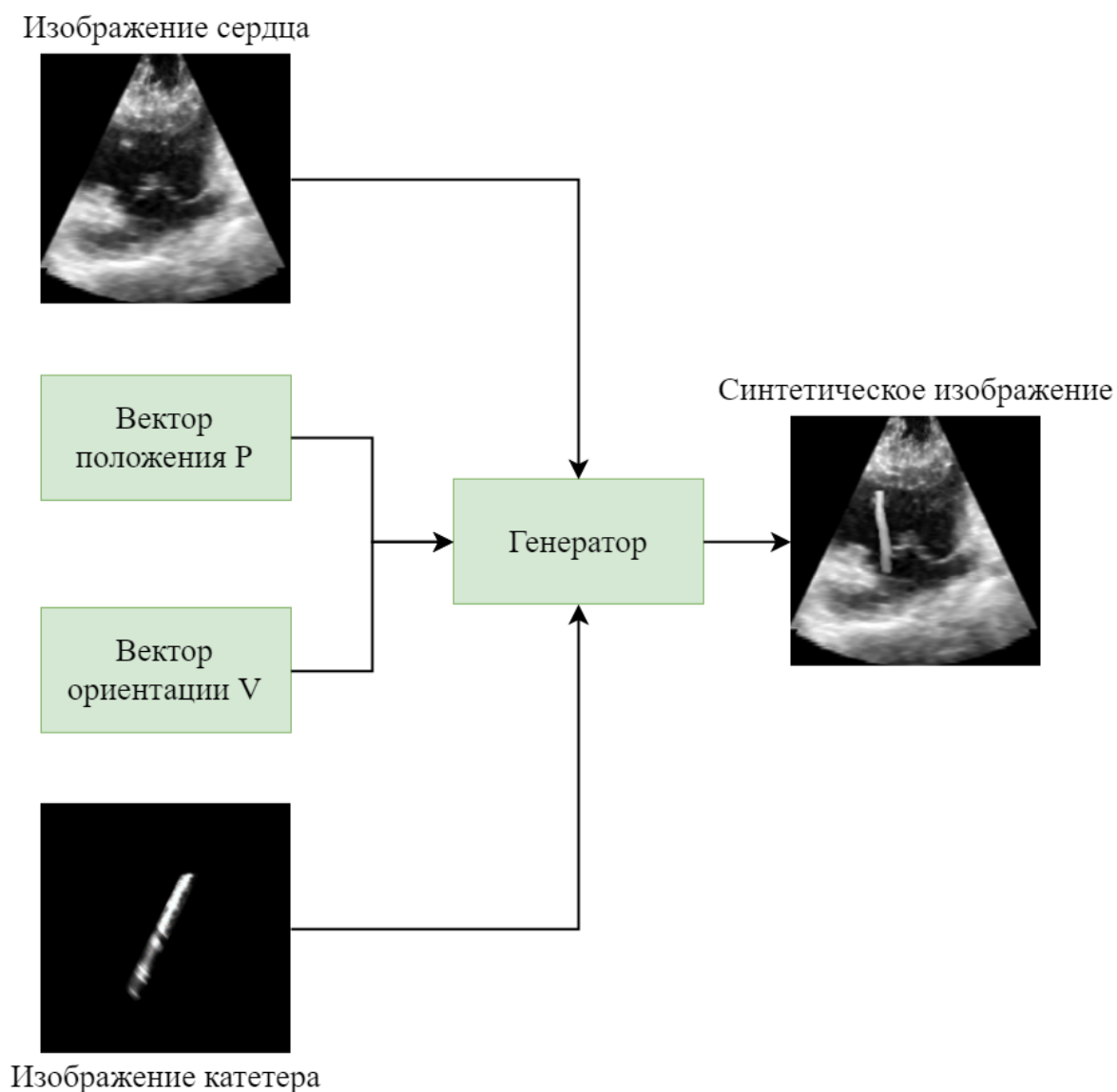


Рисунок 2.4 – Общая схема генератора искусственных данных

2.5). Точка O является ближайшей точкой на сплайне относительно точки M , а отрезок OM перпендикулярен касательной к сплайну в точке O .

Подобный подход позволяет позиционировать точки катетера относительно его оси, независимо от формы сплайна. Перенос точки с одного сплайна происходит по следующей формуле:

$$M' = F^{-1}(F(M, S_1), S_2) \quad (2.1)$$

где F – описанное выше преобразование, F^{-1} – обратное преобразование. Данное выражение позволяет производить отображение точек со сплайна S_1 на сплайн S_2 , что решает задачу трансформации общей формы катетера. Растяжение катетера может быть достигнуто путём нормализации z координаты на длину сплайна, используя следующую трансформацию:

$$z' = \frac{z}{l_1} l_2 \quad (2.2)$$

где l_1 и l_2 – длины соответствующих сплайнов S_1 и S_2 соответственно.

Однозначно произвести построение оси X в точке O без дополнительной информации о её направлении невозможно, так как условию перпендикулярности к касательной в данной точке удовлетворяет бесконечное множество векторов, лежащих в плоскости, перпендикулярной касательной. Для точного построения этой оси предлагается использовать некоторую табличную вектор-функцию $\bar{D}(z)$,

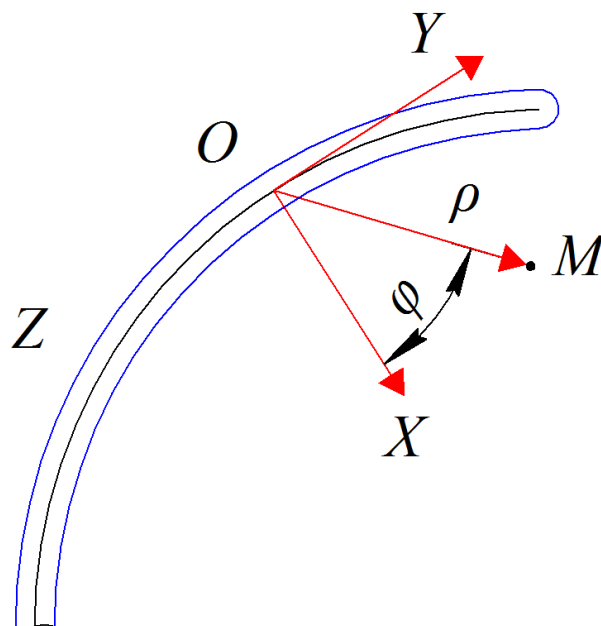


Рисунок 2.5 – Определение координат точки

которая будет определять направление оси в узлах основного сплайна, а в промежутках интерполироваться сплайном. Однако, интерполяция этой функции не гарантирует перпендикулярность вектора касательной в точках, отличных от узловых. Для этого происходит поправка этого вектора так, чтобы с минимальными его отклонениями выполнялось условие перпендикулярности к касательной. Таким образом, ось X может быть найдена по следующим формулам:

$$\alpha = \frac{-(\overline{K(z)}, \overline{D(z)})}{(\overline{K(z)}, \overline{K(z)})} \quad (2.3)$$

$$\bar{X} = \overline{D}(z) + \alpha \overline{K}(z) \quad (2.4)$$

где $\overline{K}(z)$ – вектор касательной, $\overline{D}(z)$ – функция, задающая скручивание катетера вокруг своей оси, что является одним из способов трансформации данных.

Генерация конфигурации катетера происходит с использованием алгоритмов прямой кинематики, которые строят осевой сплайн, а также вектор-функцию $\overline{D}(z)$ по заданным углам изгиба и длинам соответствующих плеч катетера. Более подробно алгоритм кинематики катетера описан в работе [118]. Генерация новой конфигурации осуществляется случайным образом до тех пор, пока не будет получена такая конфигурация, которая полностью помещается в требуемой анатомической структуре сердца. Контроль осуществляется путём проверки облака внешних точек катетера по соответствующей маске сегментации для пустого трёхмерного изображения, в которое происходит вставка. Также случайным образом генерируется изменение положения и ориентации катетера внутри камер сердца.

Процесс отображения происходит для каждого вокселя отдельно. Стоит заметить, что результатом отображения в общем случае являются вещественные декартовы координаты, в то время как значения вокселей трёхмерного изображения определены только для целых значений координат. В этом случае, взамен метода округления до ближайшего целого предлагается использовать трилинейную интерполяцию, которая позволит улучшить качество изображения. На основе трилинейной интерполяции функция $f(x, y, z)$ в трёхмерном евклидовом пространстве аппроксимируется следующим образом:

$$\begin{aligned}
f(x, y, z) \approx & \frac{f(x_1, y_1, z_1)}{(x_2 - x_1)(y_2 - y_1)(z_2 - z_1)} (x_2 - x)(y_2 - y)(z_2 - z) \\
& + \frac{f(x_1, y_1, z_2)}{(x_2 - x_1)(y_2 - y_1)(z_2 - z_1)} (x_2 - x)(y_2 - y)(z - z_1) \\
& + \frac{f(x_1, y_2, z_1)}{(x_2 - x_1)(y_2 - y_1)(z_2 - z_1)} (x_2 - x)(y - y_1)(z_2 - z) \\
& + \frac{f(x_1, y_2, z_2)}{(x_2 - x_1)(y_2 - y_1)(z_2 - z_1)} (x_2 - x)(y - y_1)(z - z_1) \\
& + \frac{f(x_2, y_1, z_1)}{(x_2 - x_1)(y_2 - y_1)(z_2 - z_1)} (x - x_1)(y_2 - y)(z_2 - z) \\
& + \frac{f(x_2, y_1, z_2)}{(x_2 - x_1)(y_2 - y_1)(z_2 - z_1)} (x - x_1)(y_2 - y)(z - z_1) \\
& + \frac{f(x_2, y_2, z_1)}{(x_2 - x_1)(y_2 - y_1)(z_2 - z_1)} (x - x_1)(y - y_1)(z_2 - z) \\
& + \frac{f(x_2, y_2, z_2)}{(x_2 - x_1)(y_2 - y_1)(z_2 - z_1)} (x - x_1)(y - y_1)(z - z_1)
\end{aligned} \tag{2.5}$$

где $f(x_i, y_j, z_k)$ – значения функции в точках (x_i, y_j, z_k) , $i = 1, 2, j = 1, 2, k = 1, 2$, причём $x_1 < x < x_2, y_1 < y < y_2, z_1 < z < z_2$.

Схематически алгоритм синтеза данных медицинского инструмента внутри анатомической структуры представлена на рисунке 2.6. Реализация данного алгоритма произведена в среде MATLAB R2017a. Выбор математического пакета обусловлен возможностями работы со сплайнами, матричными операциями, а также наличием обширного списка библиотек для работы с изображениями. Ещё одним преимуществом данной среды является простота работы с медицинским форматом данных DICOM и форматом сырых данных NRRD. Однако из недостатков стоит выделить относительно невысокую скорость работы данного пакета и слабую оптимизацию.

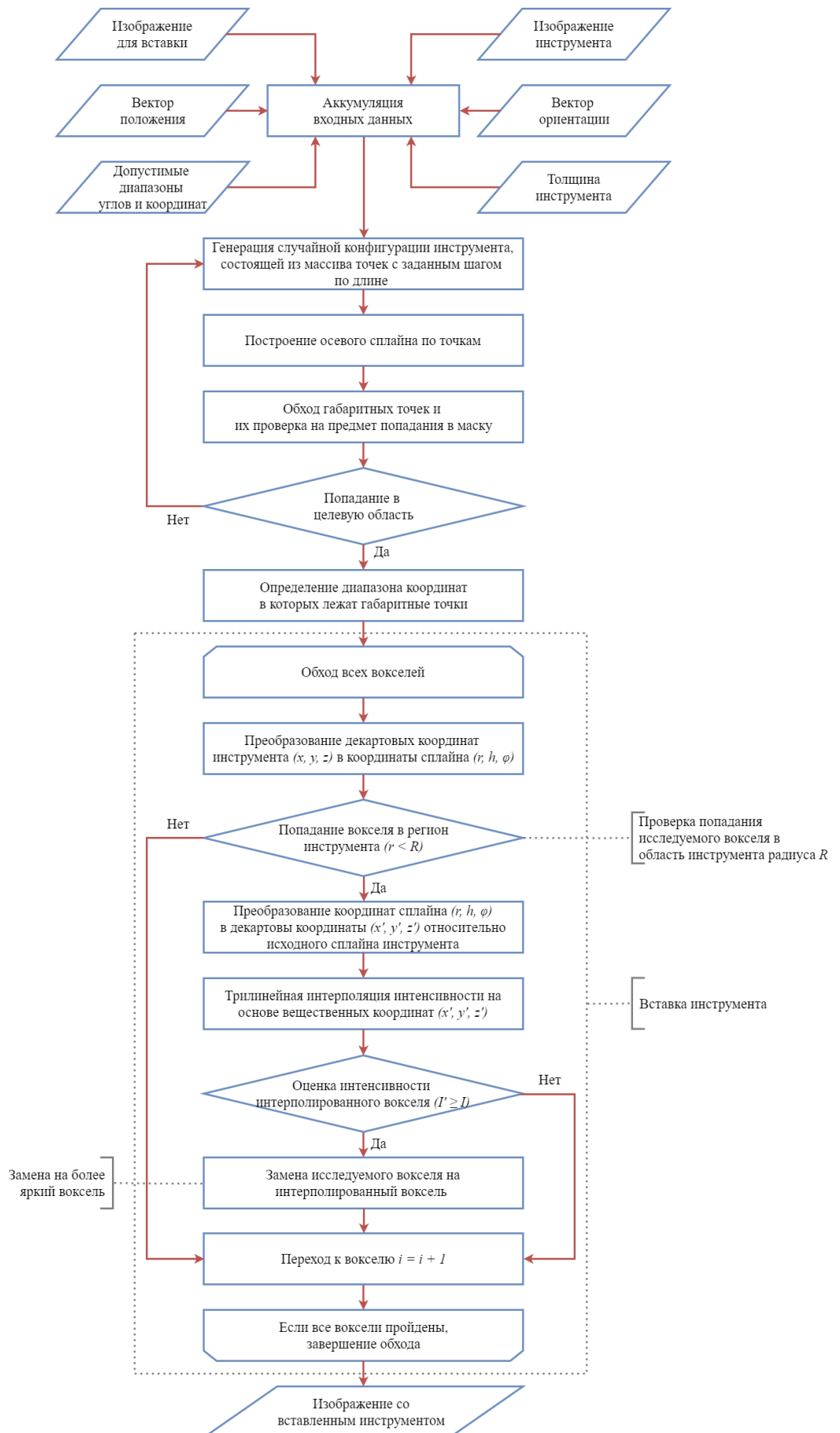


Рисунок 2.6 – Схема алгоритма генерации искусственных данных

2.3 Алгоритм аугментации данных

Одной из популярных техник при недостатке данных является аугментация. В процессе аугментации происходит искусственная генерация новых данных на основе уже существующих. При наложении различных искажений и трансформаций на исходные реальные изображения, первоначальная выборка значительно расширяется, что приводит к существенному улучшению работы модели. Наиболее популярными типами искажений являются яркостные и цветовые изменения, замена фона, геометрические (проективные и аффинные) трансформации, наложение бликов, шумов, размытия и т.д. Стоит отметить, что данный подход отличается сравнительной дешевизной и отсутствием необходимости большого количества человеческих ресурсов, поскольку не нуждается в разметке и сборе данных. Это обусловлено тем, что весь процесс создания базы изображений зависит от выбора алгоритма и его параметров. Однако, могут возникнуть сложности при аккуратном подборе параметров аугментации обучающей выборки из первоначальных экземпляров. С одной стороны, необходимо такое количество экземпляров, которое позволит научить нейронную сеть распознавать зашумлённые примеры, а с другой стороны, необходимо при этом поддерживать качество на других типах сложных изображений.

В целях оптимального расширения выборки алгоритм аугментации осуществляется итеративно, а именно, различные виды искажений, соответствующие частым типам ошибок, постепенно добавляются в обучающую выборку. К популярным типам ошибок относятся ошибки на изображениях с шумом и размытием, ошибки на повёрнутых изображениях, ошибки на нецентрированных данных, ошибки на комбинированных трансформациях и т.д. Количество добавленных трансформированных изображений варьируется и подбирается на основании обратного отклика валидационной выборки. Алгоритм по подбору оптимальной расширенной выборки изображён на рисунке 2.7. Стоит отметить, что в целях достижения необходимого качества подобная итеративная

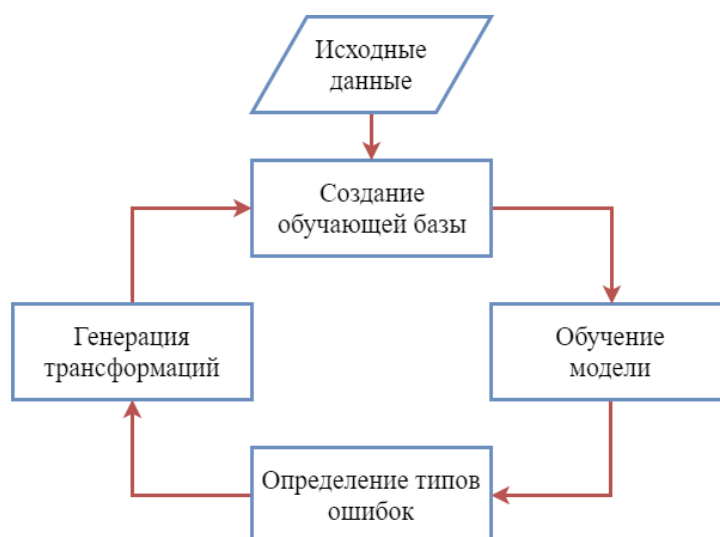


Рисунок 2.7 – Алгоритм создания расширенной обучающей выборки

процедура проводится многократно или до тех пор, пока функция потерь не перестанет снижаться.

Помимо использования аугментации на этапе обучения, предлагается использовать расширение выборки на этапе предсказания новых данных. Такая процедура называется test-time аугментация (ТТА). Основное её предназначение заключается в нивелировании и уменьшении ошибки предсказания. Процедура предсказания на основе ТТА выполняется с использованием N аугментированных примеров, в отличие от стандартного предсказания новых данных, использующего один пример. В качестве примера на рисунке 2.8 показана схема использования алгоритма ТТА на 5 трансформированных изображениях. Обычно результатом

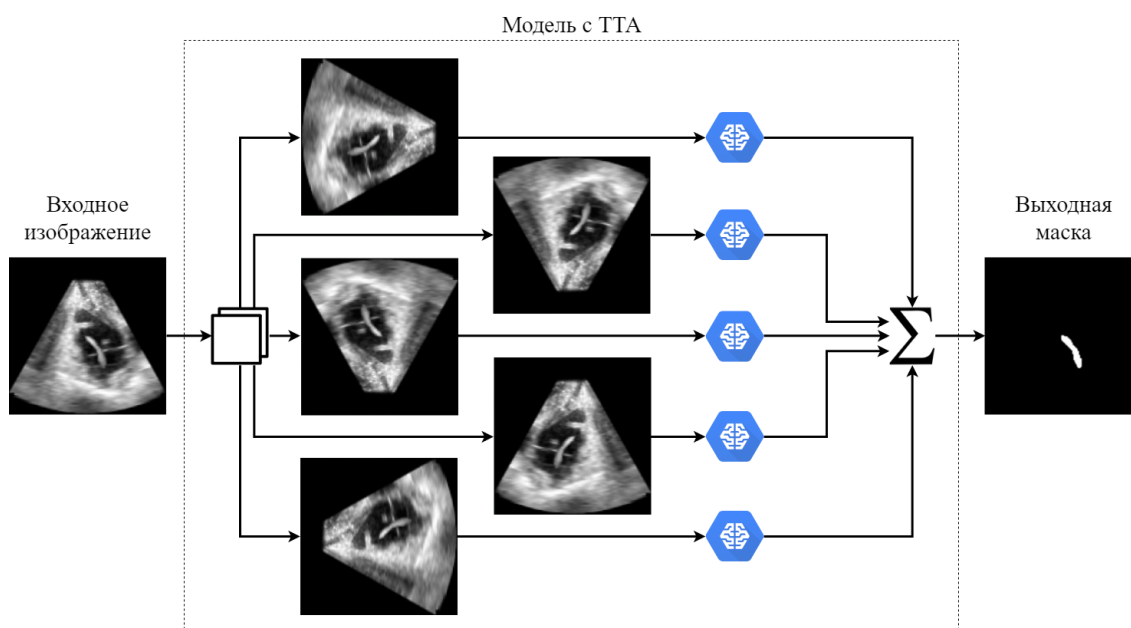


Рисунок 2.8 – Процесс предсказания изображения на основе обученной модели с ТТА

ТТА является среднее значение N карт вероятностей. Однако альтернативной методикой может быть использование взвешенного среднего, где вес каждой трансформации определяет её влияние. В свою очередь, значение веса может быть оценено на валидационной выборке. В случае если модель выполняет плохие предсказания на новых данных, использование ТТА зачастую повышает шансы модели на более корректное прогнозирование, тем самым увеличивая точность сегментации.

2.4 Архитектура нейронной сети

На сегодняшний день существуют архитектуры нейронных сетей, использующие две части для эффективной репрезентации данных, а именно энкодер и декодер. Энкодер используется для преобразования входных данных в некоторый набор признаков меньшей размерности. В свою очередь, декодер использует эти признаки в качестве входных данных, преобразует их и выдаёт окончательное решение. На сегодняшний день существует множество различных архитектур нейронных сетей, направленных на решение задач сегментации. Некоторыми из них являются FCN-8s [119], FCN-16s [119], FCN-32s [119], SegNet [120], U-net [22], Mask-RCNN [28], DeepLab [121], DeepMedic [122], DeconvNet [21], RedNet [123], TernausNet [124] и LinkNet [125]. Однако по ряду причин в качестве базовой архитектуры нейронной сети была выбрана архитектура U-net. Во-первых, U-net архитектура использует множество слоёв развёртки. Подобная имплементация позволяет итеративно восстановить изображение до исходного размера с минимальными потерями в точности и качестве. В свою очередь, такие архитектуры, как FCN-8s, FCN-16s и FCN-32s используют однократный блок развёртки, что зачастую ведёт к потере некоторых регионов интереса. Во-вторых, U-net архитектура использует конкатенации, в отличие от стандартного суммирования слоёв. В-третьих, при недостатке данных обучение нейронной сети с такой архитектурой возможно на основе стратегии перекрытия, т.е. изображение разбивается на несколько частей, которые потом используются как входные данные для обучения. При обучении оригинальной имплементации U-net

архитектуры использовался именно такой подобный подход, так как обучающая выборка состояла всего из 30 изображений.

2.4.1 Стандартная U-net архитектура

Первоначальная попытка по оценке данной архитектуры заключалась в использовании оригинальной имплементации U-net архитектуры и оценки качества сегментации. При стандартной имплементации U-net архитектура использует следующий набор слоёв: стандартный свёрточный слой, активационный слой ReLU, max-pooling слой (слой субдискретизации) и слой развёртки на основе билинейного масштабирования. Архитектура оригинальной имплементации U-net сети изображена ниже на рисунке 2.9.

Однако при реализации стандартной архитектуры алгоритм градиентного спуска сходил плохо и, как следствие, точность сегментации не превышала 5% порога. Поэтому в первой модификации было принято решение использовать слои

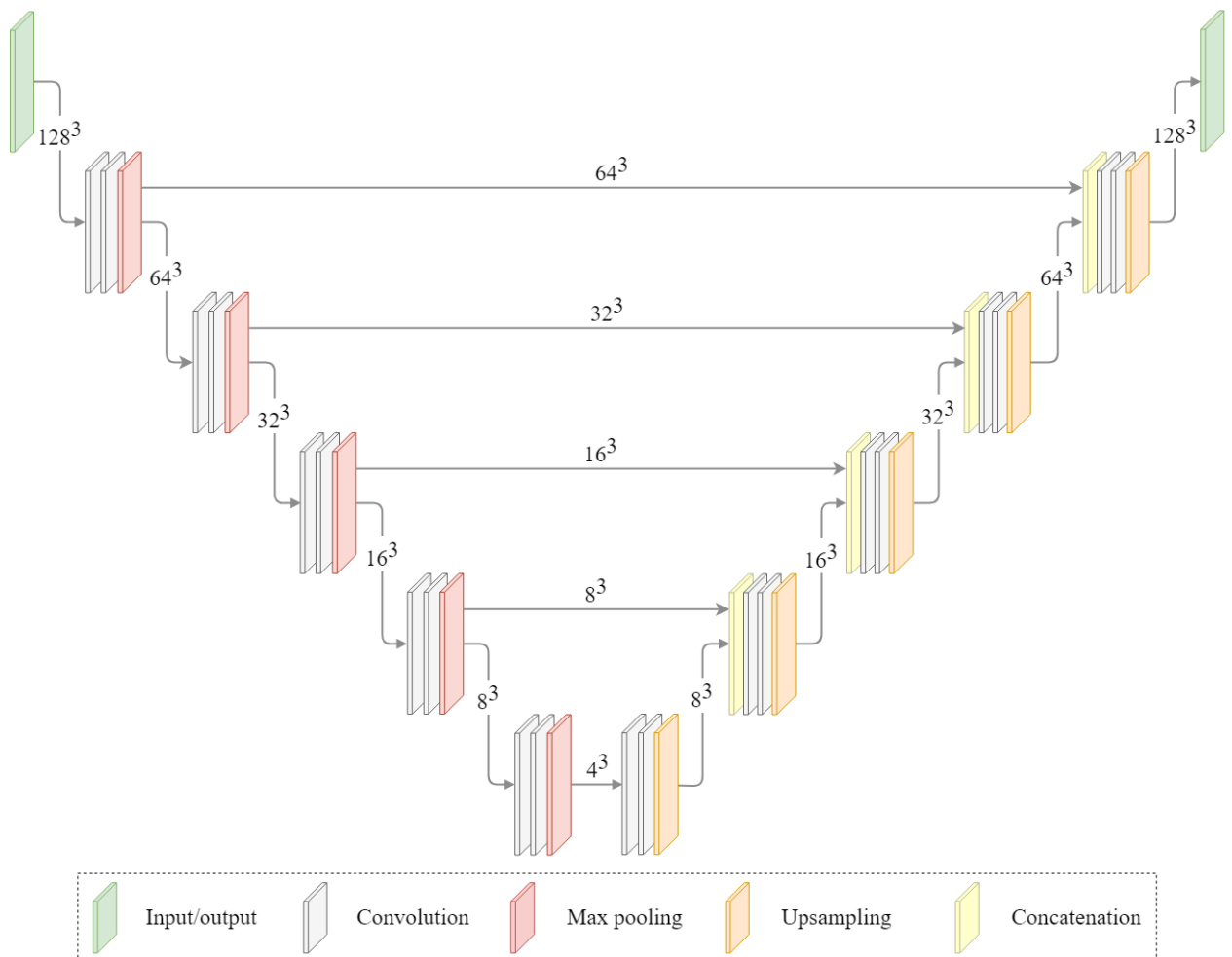


Рисунок 2.9 – Стандартная U-net архитектура

batch нормализации [126]. Слой batch нормализации нормализует значения активации предыдущего слоя таким образом, что среднее значение выходных данных с предыдущего слоя смещается к 0, а их стандартное отклонение – к 1. Архитектура оригинальной имплементации U-net сети с batch нормализацией отражена ниже на рисунке 2.10. В дальнейшем в качестве сравнения будет использоваться именно эта архитектура. Далее в работе используется название «Стандартная U-net архитектура» для данного типа нейронной сети.

2.4.2 Разреженная U-net архитектура

Добавление нормализации на прошлом шаге помогло методу градиентного спуска находить минимум функции потерь эффективнее по сравнению с классической реализацией U-net архитектуры. Однако точность всё ещё оставалась

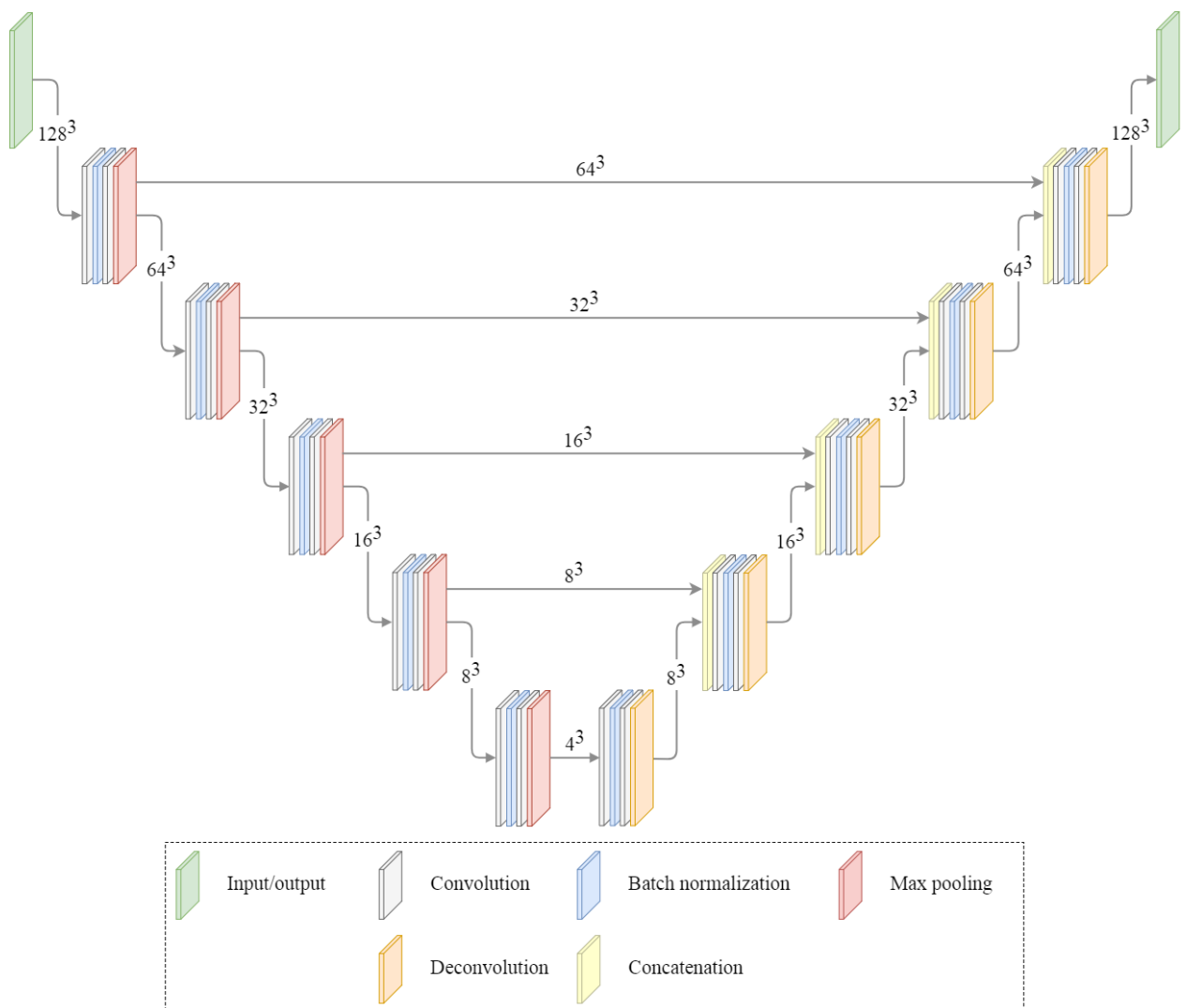


Рисунок 2.10 – Стандартная U-net архитектура с batch нормализацией

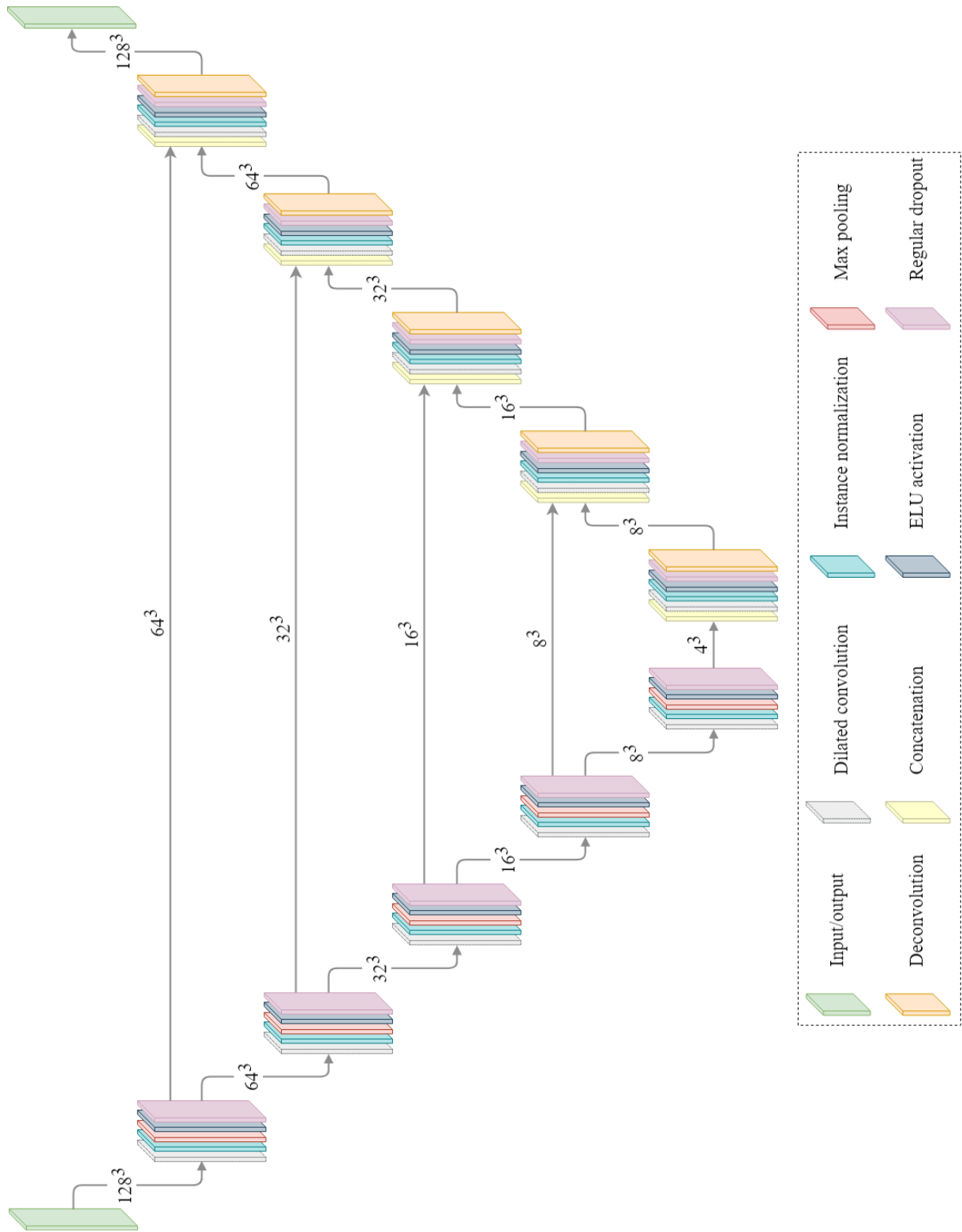


Рисунок 2.11 – Разреженная U-net архитектура

на низком уровне. Для потенциального увеличения точности и улучшения обобщающей способности был внесён ряд изменений в стандартную U-net архитектуру. Архитектурно были переработаны блоки энкодера и декодера. Во второй модификации большинство данных слоёв было заменено, и предлагаемая U-net архитектура содержала разреженные свёрточные слои, instance нормализацию, активационный слой ELU, max-pooling слой и транспонированный свёрточный слой (см. рисунок 2.11). Использование именно таких слоёв объясняется в пунктах 2.4.3, 2.4.4 и 2.4.5. В дальнейшем для данной модификации U-net архитектуры используется название «Разреженная U-net архитектура» или «Dilated U-net».

2.4.3 Слой разреженной свёртки

Введение слоя разреженной свёртки [127] в данную работу связано с минимизацией количества обучаемых весов текущей модели. Также, внедрение данного типа слоёв связано с уравниванием локального свойства (попиксельная/повоксельная точность и точное определение краёв объекта) и глобального свойства (интеграция знаний в более широком, глобальном контексте) обучаемой модели. Для разреженного свёрточного слоя с фильтром размера $k \times k$, выходное рецептивное окно будет иметь размер $k + (k - 1)(l - 1)$, где l – фактор разрежения/дилатации. Стоит отметить, что обычная свёртка является частным случаем её разреженной версии при $l = 1$. Стандартная и разреженная свёртки определяются следующим образом:

$$(F * k)(p) = \sum_{s+t=p} F(s) \cdot k(t) \quad (2.6)$$

$$(F *_l k)(p) = \sum_{s+lt=p} F(s) \cdot k(t) \quad (2.7)$$

где F – дискретная функция (сигнал), k – фильтр размером $(2r + 1) \times (2r + 1)$, l – фактор разрежения, $*$ и $*_l$ – обычный и разреженный свёрточный оператор соответственно. Видно, что при суммировании $s + lt = p$, функция разреженной свёртки пропускает $l - 1$ точек между рядами/столбцами исходного окна свёртки. Важным свойством разреженной свёртки является то, что при экспоненциальном

увеличении фактора разрежения на каждом слое сети, количество параметров растёт линейно, в свою очередь, рецептивное окно растёт так же, как и фактор разрежения, экспоненциально [127]. Подобное свойство динамики поля зрения разреженной свёртки отражено ниже на рисунке 2.12. Важным будет отметить, что количество фильтров для каждого слоя не выбиралось фиксированным. С увеличением глубины, количество фильтров/карт признаков увеличивалось. Расчёт количества фильтров на определённой глубине нейронной сети осуществлялся по следующей формуле:

$$F = 2^{level} * initial\ filter\ rate \quad (2.8)$$

где *level* – уровень/глубина сети, *initial filter rate* – исходное значение количества свёрточных фильтров. Значение *initial filter rate* сильно влияет на сложность модели и, в свою очередь, на время её обучения. Подробное описание выбора оптимальных значений, в том числе и количества свёрточных фильтров, отражено в пункте 2.6.

Легко заметить, что при использовании фактора разрежения $l > 1$, количество весов по сравнению со стандартной свёрткой остаётся на прежнем уровне, в то время как размер рецептивного окна значительно увеличивается. Так, например, стандартная свёртка с фильтром, размером 5×5 , имеет 25 весов. В свою очередь, разреженная свёртка с таким же фильтром имеет 9 не нулевых весов. Данная модификация позволяет использовать фильтры с гораздо большим размером, что увеличивает поле зрения ядра свёртки. Важным будет отметить, что функция разреженной свёртки обобщает свёрточные фильтры Кронекера, для

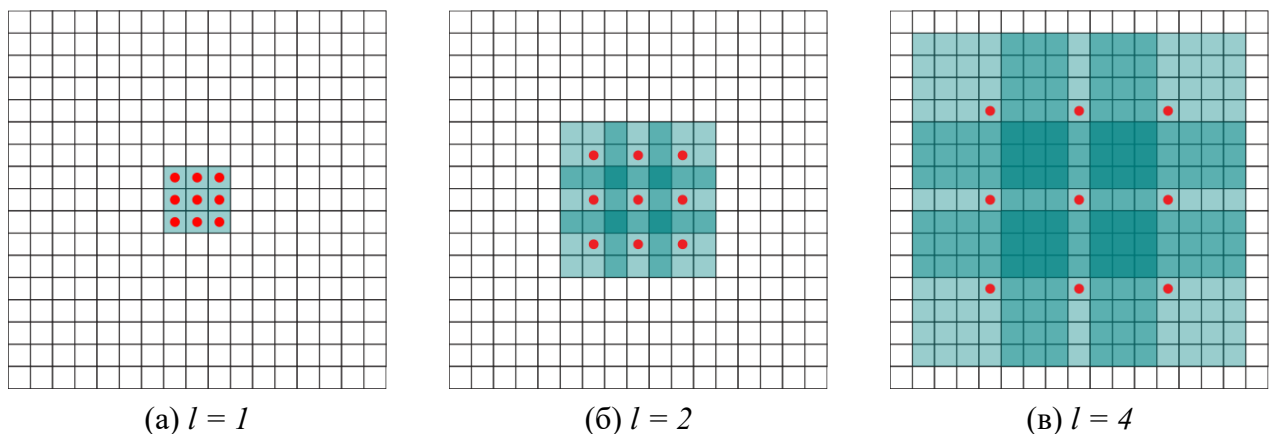


Рисунок 2.12 – Динамика изменения рецептивного окна в зависимости от фактор разрежения

которых количество обучаемых параметров растёт логарифмически в зависимости от размера рецептивного окна. Так, например, исследование [128] показало, что при использовании свёрточных слоёв с фильтрами Кронекера, скорость обучения возрастает в 3.3 раза, а количество весов уменьшается в 3.6 раз при уменьшении точности на 1%.

2.4.4 Слой активации

Использование активационных слоёв ELU ведёт к ускорению процесса обучения, положительно влияет на решение проблемы исчезающих градиентов, а также увеличивает точность классификации нейронных сетей [129]. В отличие от функции активации ReLU, функция активации ELU имеет не нулевую отрицательную составляющую при $x \leq 0$ (см. (2.9)). Использование отрицательных градиентов позволяет смещать к нулю среднее значение активации, что, в свою очередь, помогает избегать лишних сдвигов и смещений. Подобную процедуру осуществляет слой batch нормализации, однако активационный слой ELU выполняет это с меньшей вычислительной сложностью. В свою очередь, активационные функции Leaky ReLU и PReLU также имеют отрицательные составляющие, однако они не устойчивы к шумам и выбросам. Это объясняется тем, что у этих функций отсутствует насыщение, в то время как на отрицательной половине функция активации ELU насыщается при небольших значениях входов, что в дальнейшем уменьшает вариативность информации, передаваемой вперёд. Дополнительно стоит отметить, что область значений функции ELU находится в диапазоне значений $(-\alpha; +\infty)$, тогда как области значений функций Leaky ReLU и PReLU находятся в диапазоне $(-\infty; +\infty)$. Показано, что если множество значений функции активации ограничено, то методы обучения на основе градиентного спуска более стабильны и в ряде случаев обеспечивают нейронным сетям более высокую степень обобщающей способности. Функция активации ELU и её градиент определяется следующим образом:

$$f(x) = \begin{cases} x, & \text{если } x > 0 \\ \alpha(e^x - 1), & \text{если } x \leq 0 \end{cases} \quad (2.9)$$

$$f'(x) = \begin{cases} 1, & \text{если } x > 0 \\ f(x) + \alpha, & \text{если } x \leq 0 \end{cases} \quad (2.10)$$

где α – масштабирующий коэффициент для отрицательной составляющей. Принято брать значения коэффициента α в диапазоне значений $[0; 1]$.

2.4.5 Слой нормализации

В настоящее время batch нормализация имеет большую популярность, однако её альтернативой является instance нормализация. В целом, оба метода нормализации выполняют практически одну и ту же процедуру, однако сам подход к выполнению этой процедуры отличается. Так, batch нормализация рассчитывает одно значение среднего и одно значение стандартного отклонения, преобразуя входные значения в гауссово распределение [130–132]. В свою очередь, instance нормализация производит расчёт T значений среднего и стандартного отклонения, преобразуя каждое входное значение таким образом, чтобы оно являлось независимым гауссовым распределением. Формулы для расчёта значений на основе batch нормализации (см. (2.11), (2.12), (2.13)) и instance нормализации (см. (2.14), (2.15), (2.16)) указаны ниже:

$$y_{nchw} = BN(x_{nchw}) = \gamma_c \left(\frac{x_{nchw} - \mu_c}{\sigma_c} \right) + \beta_c \quad (2.11)$$

$$\mu_c = \frac{1}{HWT} \sum_{t=1}^T \sum_{w=1}^W \sum_{h=1}^H x_{nchw} \quad (2.12)$$

$$\sigma_c = \sqrt{\frac{1}{HWT} \sum_{t=1}^T \sum_{w=1}^W \sum_{h=1}^H (x_{nchw} - \mu_c)^2 + \epsilon} \quad (2.13)$$

$$y_{nchw} = IN(x_{nchw}) = \gamma_{nc} \left(\frac{x_{nchw} - \mu_{nc}}{\sigma_{nc}} \right) + \beta_{nc} \quad (2.14)$$

$$\mu_{nc} = \frac{1}{HW} \sum_{w=1}^W \sum_{h=1}^H x_{nchw} \quad (2.15)$$

$$\sigma_{nc} = \sqrt{\frac{1}{HW} \sum_{w=1}^W \sum_{h=1}^H (x_{nchw} - \mu_{nc})^2 + \epsilon} \quad (2.16)$$

где n – размер батча, c – количество каналов, h – высота, w – ширина, γ и β – обучаемые аффинные параметры, которые смещают моменты первого и второго порядка в процессе обучения, ϵ – малая по значению константа, μ и σ^2 – среднее и дисперсия карты признаков.

В данном случае batch нормализация отчасти не применима по ряду причин, а именно в связи с невозможностью использования большого значения батча. Последнее связано с большой размерностью входных данных ($128 \times 128 \times 128$); большим количеством обучаемых весов (более 10 миллионов); ограничением видеопамати. Будет важным отметить, что при размере батча равному 1, дисперсия равна 0 и тогда такой тип нормализации для обучения не применим. Увеличение размера батча отчасти нивелирует подобную проблему. Однако небольшие размеры батча вносят шум из-за того, что результат нормализации зависит от соседних экземпляров, что негативно влияет на процесс обучения.

2.4.6 Dropout слои и регуляризация

В целях минимизации переобучения для предлагаемой архитектуры сети использовалась техника dropout [133]. Главная идея данной техники заключается в том, чтобы вместо одной нейронной сети обучить ансамбль нескольких сетей, а затем усреднить полученные результаты. Новые сети для обучения получаются с помощью исключения нейронов из базовой сети с вероятностью p . Исключение нейрона означает, что при любых входных данных или параметрах он возвращает 0. Исключённые нейроны не вносят свой вклад в процесс обучения ни на одном из этапов метода обратного распространения ошибки, поэтому исключение хотя бы одного из нейронов равносильно обучению новой нейронной сети.

При тестировании использовались две версии dropout слоёв, а именно стандартная и пространственная. В отличие от стандартной версии dropout слоя, пространственная версия обнуляет не отдельные элементы, а полностью карту

признаков. В предлагаемой работе вероятность обнуления весов была не фиксированной, а увеличивалась по мере углубления архитектуры. Для расчёта вероятности обнуления весов на текущем уровне сети использовалась следующая формула:

$$\text{Dropout rate} = 5 * \text{level} * \text{initial dropout rate} \quad (2.17)$$

где *level* – уровень/глубина сети, *initial dropout rate* – начальное значение вероятности обнуления нейронов. Значение *initial drop rate* выбрано на уровне 0.005, что говорит о том, что на первом уровне будет обнуляться 2.5% весов, на втором – 5% весов и так далее. Несмотря на то, что формула (2.17) явно не отражает зависимость от количества весов, для разрабатываемой архитектуры V-net значение вероятности обнуления нейронов на уровнях сети было квазипропорционально количеству свёрточных фильтров.

Другой используемой техникой в данной работе является процедура регуляризации. Регуляризация является распространённым методом уменьшения сложности модели, а также позволяет предотвратить переобучение или исправить некорректно поставленную задачу. Обычно регуляризация представляет собой некоторую добавку $R(w)$ к функции потерь $L(f(x, w), y)$, где w – параметры модели. Тем самым задача минимизации функции потерь сводилась к следующему виду:

$$\min_w \sum_{i=1}^N L(f(x_i, w), y_i) + R(w) \quad (2.18)$$

В качестве типов регуляризации были рассмотрены LASSO регуляризация (L_1 -регуляризация) [134, 135], Ridge регуляризация (L_2 -регуляризация или регуляризация Тихонова) [135–137] и Elastic net регуляризация (комбинация L_1 и L_2 регуляризаций) [138]. Формально данные типы регуляризации имеют следующий вид:

$$\text{LASSO} \quad R(w) = \lambda \|w\|_1 = \lambda \sum_{i=1}^d |w_i| \quad (2.19)$$

$$\text{Ridge} \quad R(w) = \lambda \|w\|_2^2 = \lambda \sum_{i=1}^d w_i^2 \quad (2.20)$$

$$\text{Elastic net} \quad R(w) = \lambda_1 \|w\|_1 + \lambda_2 \|w\|_2^2 = \lambda_1 \sum_{i=1}^d |w_i| + \lambda_2 \sum_{i=1}^d w_i^2 \quad (2.21)$$

2.4.7 Пробросы признаков

Одним из проблемных моментов глубоких нейронных сетей является потеря точности после насыщения. Заключается она в том, что при увеличении глубины сети значение точности начинает насыщаться, а затем, доходя до какого-то определённого значения, резко деградирует. Подобная деградация точности или функции потерь не связана с переобучением, а добавление большего количества слоёв ведёт к увеличению ошибки при обучении [139]. Данная проблема относится к исчезающим градиентам. Несмотря на то, что D.E. Rumelhart, G.E. Hinton и R.J. Williams показали сходимость метода обратного распространения ошибки [140], метод сходится при скорости обучения близкой к нулю. Последнее ведёт к затуханию градиентов, то есть к бесконечно малому изменению весов. Не глубокие сети, имеющие малое количество слоёв, как правило, не обладают этим недостатком, так как подобный эффект не успевает аккумулироваться при обратном проходе метода. При обучении глубоких архитектур эффект обучения на основе метода обратного распространения ошибки не доходит до начальных слоёв. Это объясняется тем, что градиенты весов в начальных слоях становятся предельно малыми за счёт постепенно уменьшающихся градиентов на более поздних слоях. При прямом прохождении, функции активации позволяют скомпенсировать данный недостаток. Однако обратный проход метода обратного распространения ошибки линеен, поэтому для того, чтобы скомпенсировать эффект затухания градиентов предлагается внедрить пробросы. Пробросы позволяют скомпенсировать затухание градиентов за счёт увеличения значений весов и производной в точке.

На этапе обучения наиболее общие признаки данных извлекаются на начальных уровнях сети. При использовании архитектур, основывающихся на

комбинации энкодера и декодера, подобные признаки впоследствии участвуют в сверхдискретизации (повышении размерности) данных. При использовании обычной архитектуры сети без использования плотно-соединённых слоёв информация может теряться. Поэтому информацию, используемую на предыдущих слоях, особенно на начальных, эффективнее передавать на последующие слои напрямую, используя при этом пробросы. Подобная топология позволяет информации быстрее проходить по сети, а также передавать общие признаки на более низкие уровни.

Использование пробросов признаков может быть реализовано двумя способами: суммированием и конкатенацией слоёв. Основная разница заключается в том, что при суммировании слоёв их тензоры должны иметь одинаковую размерность, а результирующий тензор будет иметь ту же размерность, что и входной. В случае с конкатенацией, результирующий тензор будет шире и будет состоять из двух и более карт признаков.

Для улучшения градиентного потока через архитектуру сети, внутри нейронной сети были реализованы плотно-соединённые слои или, так называемые, пробросы. При прохождении изображения x через свёрточную нейронную сеть, состоящую из L слоёв, каждый слой накладывает определённую нелинейную трансформацию $H_l()$, где l – индекс слоя. Для нелинейной трансформации могут использоваться различные операции, например, такие как instance и/или batch нормализация, активация, субдискретизация, стандартная и/или разреженная свёртка и т.д. Традиционные свёрточные сети прямого распространения соединяют выход слоя l с входом слоя $(l+1)$. В результате такой архитектуры происходит следующая трансформация $x_l = H_l(x_{l-1})$. В некоторых архитектурах, например, ResNet, внедрён подход «skip-connection», который обходит нелинейные преобразования с помощью тождественной функции, в результате чего выходная трансформация имеет следующий вид:

$$x_l = H_l(x_{l-1}) + x_{l-1} \quad (2.22)$$

Преимущество такого подхода заключается в том, что градиенты могут дополнительно проходить непосредственно через ветвь тождественной функции,

что положительно влияет на решение проблемы затухающих градиентов. Однако тождественная функция и выход $H_l()$ объединяются суммированием, что не всегда способствует эффективному потоку информации в сети. Поэтому для улучшения информационного потока между слоями и в целях решения проблемы исчезающих градиентов предлагается использовать подход, в котором каждый слой соединяется со всеми последующими слоями с помощью конкатенации. Тем самым, слой l получает карты признаков всех предшествующих слоёв $x_0, \dots, x_{l-1}$:

$$x_l^1 = H_l^1([x_l^1, \dots, x_{l-1}^1]) \quad (2.23)$$

$$x_l^2 = H_l^2([x_l^2, \dots, x_{l-1}^2]) \quad (2.24)$$

где $[x_l^1, \dots, x_{l-1}^1]$ и $[x_l^2, \dots, x_{l-1}^2]$ – конкатенации карт признаков для энкодера и декодера соответственно. Подобное разделение вводится ввиду того, что в двух половинах используются различные нелинейные трансформации и разный порядок следования слоёв.

Аналогичная реализация выполнена в работе [139], где авторы предложили внедрить дополнительный поток признаков, идущий через блок. Однако слой l на входе имел карты признаков со слоёв $l-1$, а данная архитектура использует признаки со всех предыдущих слоёв. Стоит отметить, что использование подобной топологии нейронных сетей позволяет сгладить рельеф функции потерь и решить проблему исчезающих градиентов, что в последующем увеличит обобщающую способность. В зависимости от архитектуры нейронной сети ландшафт функции потерь может значительно отличаться. Так, неглубокие нейронные сети зачастую имеют выпуклые широкие регионы. В свою очередь, более глубокие архитектуры сетей становятся хаотичными, спонтанными и сильно невыпуклыми, что негативно сказывается на процессе обучения [141]. На рисунке 2.13 отражены ландшафты функции потерь для разных топологий нейронных сетей: VGG-56 (плоская топология сети), VGG-110 (плоская топология сети), ResNet-56 (топология с «остаточным» блоком), DenseNet-121 (топология с пробросом признаков). В таком случае становится очевидным, что задача минимизации функции потерь и поиска глобального минимума значительно упрощается за счёт использования пробросов признаков.

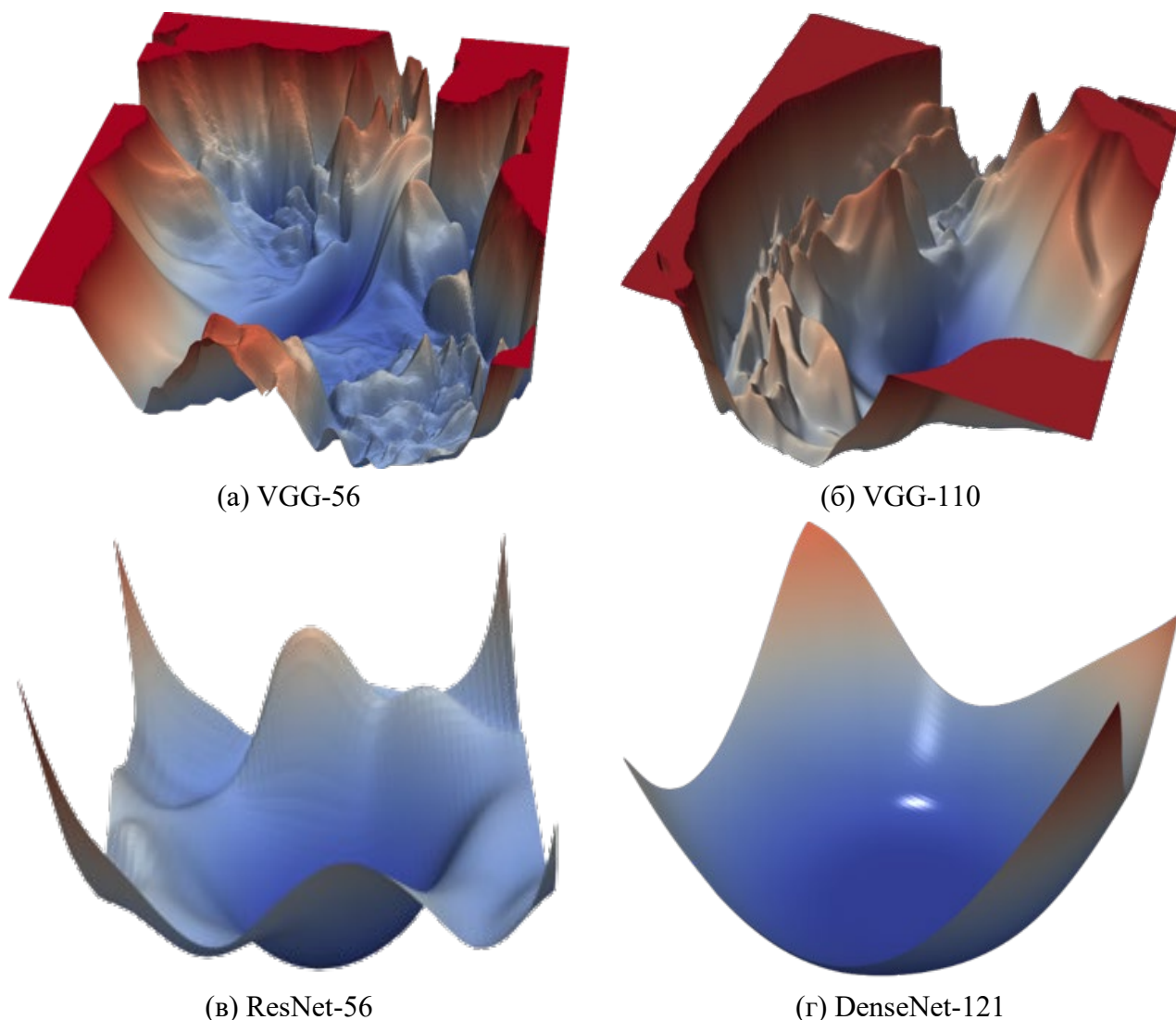


Рисунок 2.13 – Ландшафты функции потерь на основе набора данных CIFAR-10 для разных архитектур нейронных сетей

2.4.8 V-net архитектура с пробросами признаков

При формировании окончательной V-net архитектуры учитывались изменения, выполненные для разреженной U-net архитектуры, а также описанная выше методика проброса признаков. Как отмечалось в пункте 2.4.7, в предлагаемой архитектуре V-net использовалась конкатенация карт признаков взамен их поэлементного суммирования. Зачастую архитектуры нейронных сетей, использующих суммирование признаков [139, 142], сохраняют количество признаков в рамках одного блока, а также имеют более компактную структуру и меньшее количество слоёв, что сказывается на более быстром времени обучения и предсказания. Однако архитектуры, использующие конкатенацию признаков,

обеспечивают архитектуру сети большим потоком информации, что сказывается на более высокой точности и лучшей обобщающей способности [143].

Важным будет отметить, что архитектура, представленная в работе [144] и схожая по названию, значительно отличается от предлагаемой. Одним из ключевых отличий в предлагаемой архитектуре V-net является использование плотных пробросов признаков, в том числе и на более глубокие уровни. В свою очередь, авторы вышеупомянутой работы использовали так называемые глубокие «остаточные» блоки для передачи признаков через нелинейные трансформации. Другим отличием является построение блоков энкодера и декодера. В предлагаемой архитектуре, данные блоки строятся по концепции двойных свёрточных блоков VGG архитектуры со слоями instance нормализации, ELU активации, слоёв субдискретизации и dropout слоёв. С другой стороны, авторы сравниваемой работы ставили акцент на стандартные свёрточные слои, слои субдискретизации и параметрическую функцию активации PReLU. Важно отметить, что последняя значительно увеличивает количество весов по сравнению с используемой функцией активации ELU. Ещё одним отличием предлагаемой архитектуры является использование конкатенации признаков, а не их суммирования. Сходство данных архитектур объясняется лишь тем, что они обе похожи на букву «V» благодаря наличию сжимающего и расширяющего путей, а также свёрточного блока на самом нижнем уровне архитектуры. Окончательная версия предлагаемой V-net архитектуры с плотно-соединёнными слоями, использующимися для проброса большего количества признаков, отражена на рисунке 2.14. Большинство гиперпараметров предлагаемой архитектуры, в том числе типы слоёв, определены на основе алгоритма выбора оптимальных гиперпараметров, описанного в пункте 2.6.

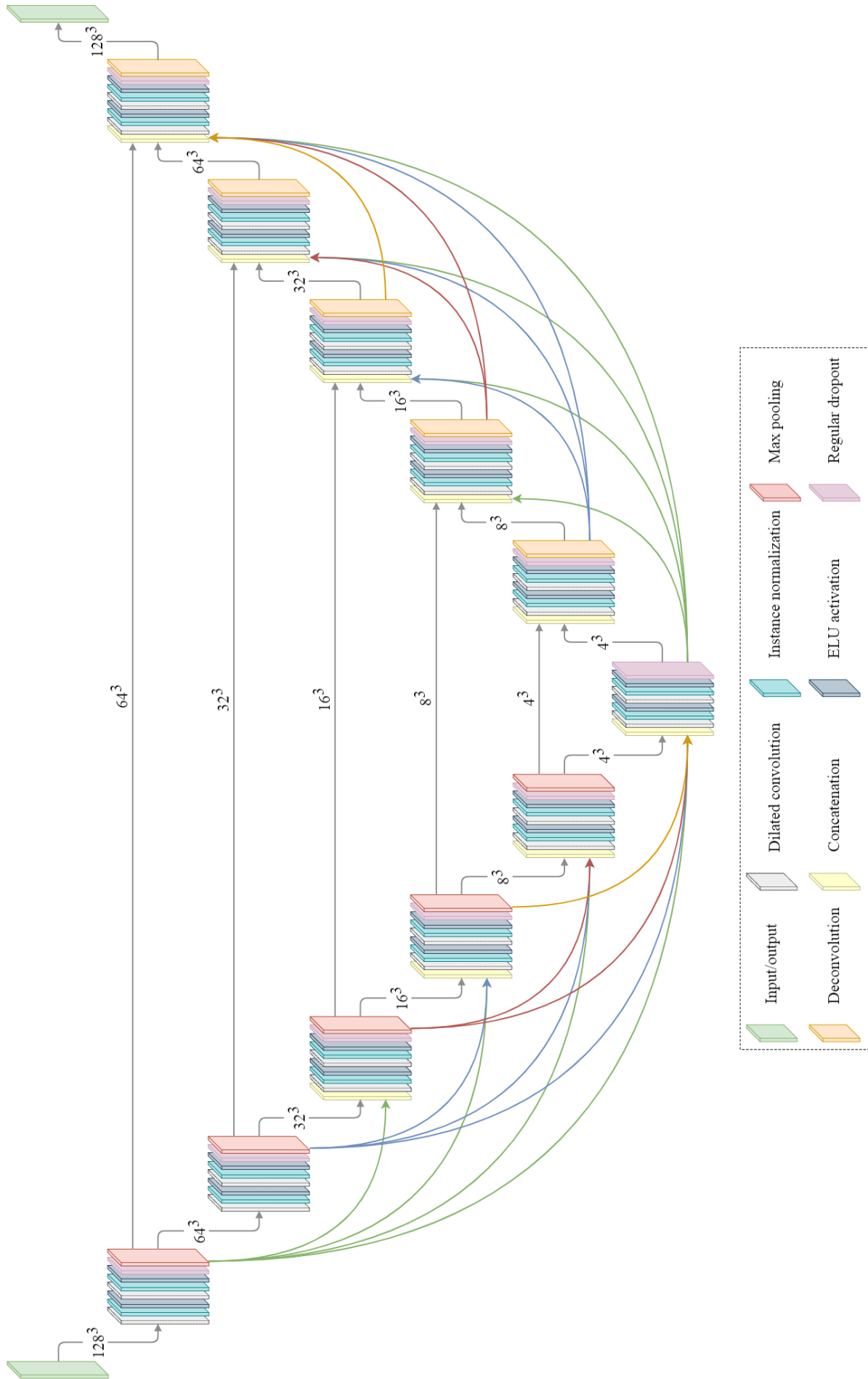


Рисунок 2.14 – V-net архитектура с пропусками признаков

2.5 Метод обучения модели

Одними из главных характеристик модели являются её обобщающая способность и точность предсказания. Для достижения требуемых значений данных параметров необходимо выполнить ряд операций. Одними из таких операций являются выбор оптимальной функции потерь, аугментация данных и оптимизации гиперпараметров. Поэтому в данном пункте описываются наиболее значимые процедуры по достижению требуемых значений точности и обобщающей способности модели.

2.5.1 Функция потерь

В подавляющем большинстве случаев при решении задач семантической сегментации в качестве основной функции потерь используется функция потерь на основании бинарной меры сходства Дайса (DSC) [145, 146]. Рассчитывается данная функция потерь следующим образом:

$$L(DSC) = 1 - DSC \quad (2.25)$$

$$DSC = \frac{2 \cdot |X \cap Y|}{|X| + |Y|} = \frac{2 * TP}{2 * TP + FP + FN} \quad (2.26)$$

где X – множество выходных значений модели, Y – множество исходных размеченных значений, $|X|$ и $|Y|$ – мощности множеств, TP – количество верноположительных случаев, TN – количество верноотрицательных случаев, FP – количество ложноположительных случаев, FN – количество ложноотрицательных случаев.

Однако, в связи с тем, что исходный набор данных не сбалансирован, предлагается ввести и использовать функцию потерь, основанную на взвешенном индексе Дайса. Вклад каждого класса корректируется обратной величиной его объёма, тем самым уменьшая корреляцию с метрикой сходства. Расчёт взвешенной функции потерь производится следующим образом:

$$L(DSC_w) = 1 - DSC_w \quad (2.27)$$

$$DSC_w = 2 \cdot \frac{\sum w_i \cdot |X_i \cap Y_i| + \epsilon}{\sum w_i \cdot (|X_i| + |Y_i|) + \epsilon} = 2 \cdot \frac{w_0 \cdot |X_0 \cap Y_0| + w_1 \cdot |X_1 \cap Y_1| + \epsilon}{w_0 \cdot (|X_0| + |Y_0|) + w_1 \cdot (|X_1| + |Y_1|) + \epsilon} \quad (2.28)$$

$$rate = \frac{positive\ cases}{all\ cases} \quad (2.29)$$

$$w_0 = \frac{1}{rate^2} \quad (2.30)$$

$$w_1 = \frac{1}{(1 - rate)^2} \quad (2.31)$$

где ϵ – предельно малая величина, необходимая для обеспечения стабильности функции потерь и избегания деления на 0 при пустых множествах, w_i – весовой коэффициент для i -го класса, $rate$ – относительный показатель экземпляров положительного класса в выборке.

При проведении экспериментов по обучению сети экспериментально была получена сигмоидоподобная функция потерь, более чувствительная на начальном и конечном отрезках функции. Подобная функция потерь задаётся либо кусочно на основании (2.32), либо непрерывно на основании (2.33):

$$L(DSC_w, z) = \begin{cases} 0.5 * (2 * DSC_w)^z, & 0 \leq DSC_w \leq 0.5 \\ 1 - 0.5 * (2 - 2 * DSC_w)^z, & 0.5 < DSC_w \leq 1.0 \end{cases} \quad (2.32)$$

$$L(DSC_w, z) = 1 - (0.5 + 0.5 * (DSC_w - 0.5) * |2 * (DSC_w - 0.5)|^z) \quad (2.33)$$

где z варьируется от 0 до ∞ . Для более точной подстройки кривизны функции потерь в (2.33), степень z предпочтительнее заменить на тангенциальную функцию $\tan(\frac{\pi}{2} * z)$, где $z \in [0, 1]$. В результате такой замены, взвешенная функция потерь на основании коэффициента z примет вид:

$$L(DSC, z) = 1 - (0.5 + 0.5 * (DSC - 0.5) * |2 * (DSC - 0.5)|^{\tan(\frac{\pi}{2} * z)}) \quad (2.34)$$

Стоит отметить, что стандартная функция потерь, основанная на индексе Дайса (см. (2.25)) и используемая в подавляющем большинстве для решения задачи семантической сегментации, является частным случаем предлагаемой функции (2.34) при $z = 0.5$. При граничных значениях $z = 0.0$ и $z = 1.0$, функция потерь принимает вид ступенчатой функции. Зависимость функции потерь от коэффициента z отражена на рисунке 2.15.

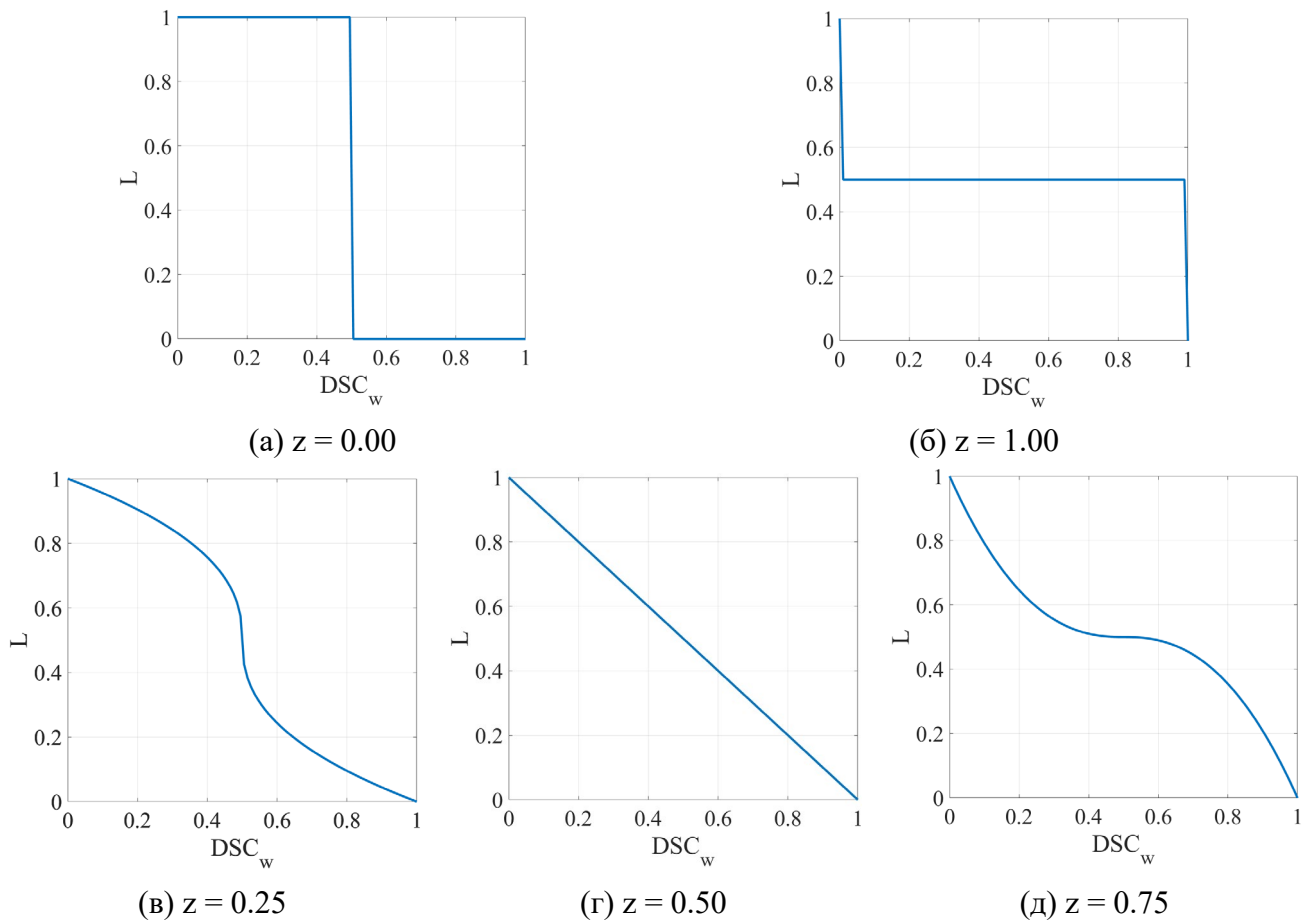


Рисунок 2.15 – График изменения взвешенной функции потерь при различных значениях z

2.5.2 Алгоритм изменения скорости обучения

Одним из ключевых параметров при обучении нейронных сетей является скорость обучения (LR), а также стратегия её изменения. Данный параметр определяет то, насколько сильно будут обновляться весовые коэффициенты модели на каждой итерации метода обратного распространения ошибки. В случае небольшого значения скорости обучения, время обучения модели будет достаточно большим, что вызвано малым обновлением весов модели. При больших значениях скорости обучения имеется шанс того, что алгоритм не будет сходиться и не попадёт в область глобального минимума, что негативно скажется на точности и обобщающей способности модели. Основной проблемой настройки данного параметра, как и многих других, является то, что заведомо сложно определить наилучшее значение и стратегию его изменения для получения оптимальной модели.

Немаловажным фактором для оптимизации весов является архитектура сети и выбранная функция потерь, на основе которых формируется ландшафт функции потерь. Становится очевидным, что при ландшафте, представленном на рисунке 2.16 (б), метод градиентного спуска будет менее эффективен, чем при ландшафте, отражённом на рисунке 2.16 (а). В случае остроконечных и невыпуклых ландшафтов метод обратного распространения ошибки потенциально имеет более высокую вероятность попасть в зону локального минимума, что негативно скажется на обобщающей способности модели. Ряд подходов, в том числе использование алгоритмов изменения скорости обучения, позволяют нивелировать проблему, связанную с подобными случаями.

Одним из часто используемых подходов при обучении нейронных моделей является отжиг скорости обучения. Наиболее популярными методами отжига является стратегия ступенчатого и экспоненциального уменьшения скорости обучения. Данные стратегии отражены на рисунке 2.17 (а) и (б), где по оси ординат отражено значение скорости обучения, а по оси абсцисс отражены итерации обучения. Основная идея такого подхода заключается в том, чтобы начать процесс обучения с наибольшего значения скорости обучения, а затем постепенно его уменьшать на основе определённого алгоритма. Большое начальное значение скорости обучения используется для того, чтобы максимально быстро спуститься от исходных параметров в область оптимальных параметров. В свою очередь, в

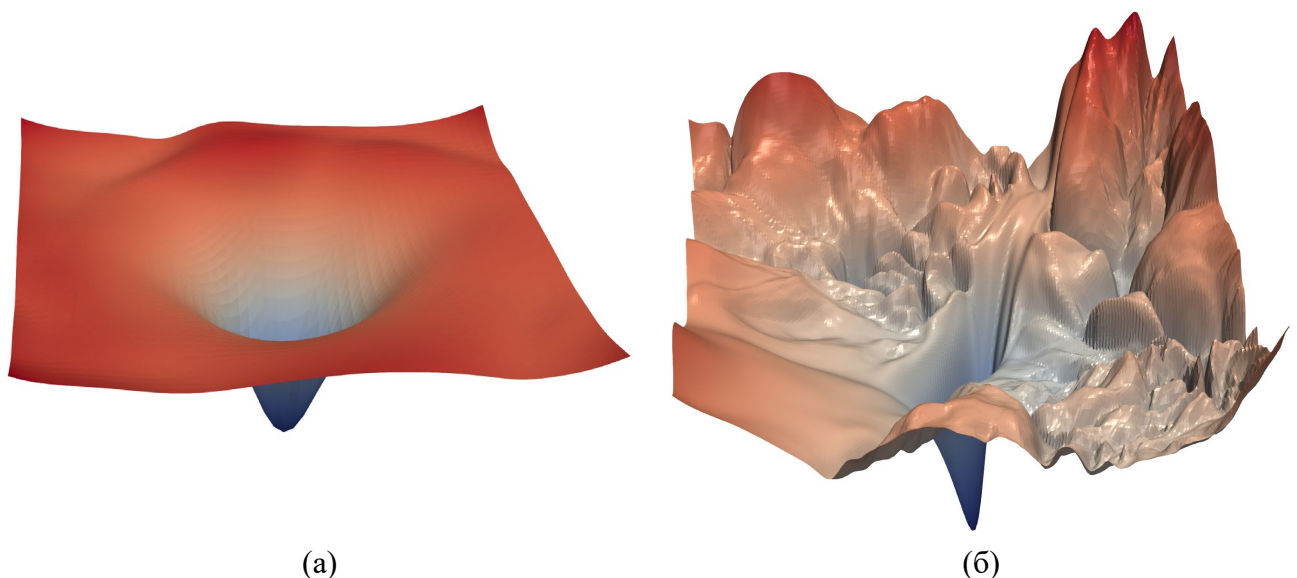


Рисунок 2.16 – Ландшафты функции потерь

области глобального минимума алгоритм градиентного спуска стремится к поиску оптимальных параметров модели.

В связи с проблемой техник отжига, малоэффективно проходящих седловые точки и локальные минимумы, предлагается использовать собственную функцию циклического изменения скорости обучения. Основная концепция данного подхода предполагает то, что скорость обучения не просто уменьшается, а варьируется по определённому алгоритму в рамках двух границ [147]. При проведении экспериментов были использованы 7 алгоритмов скорости обучения из которых:

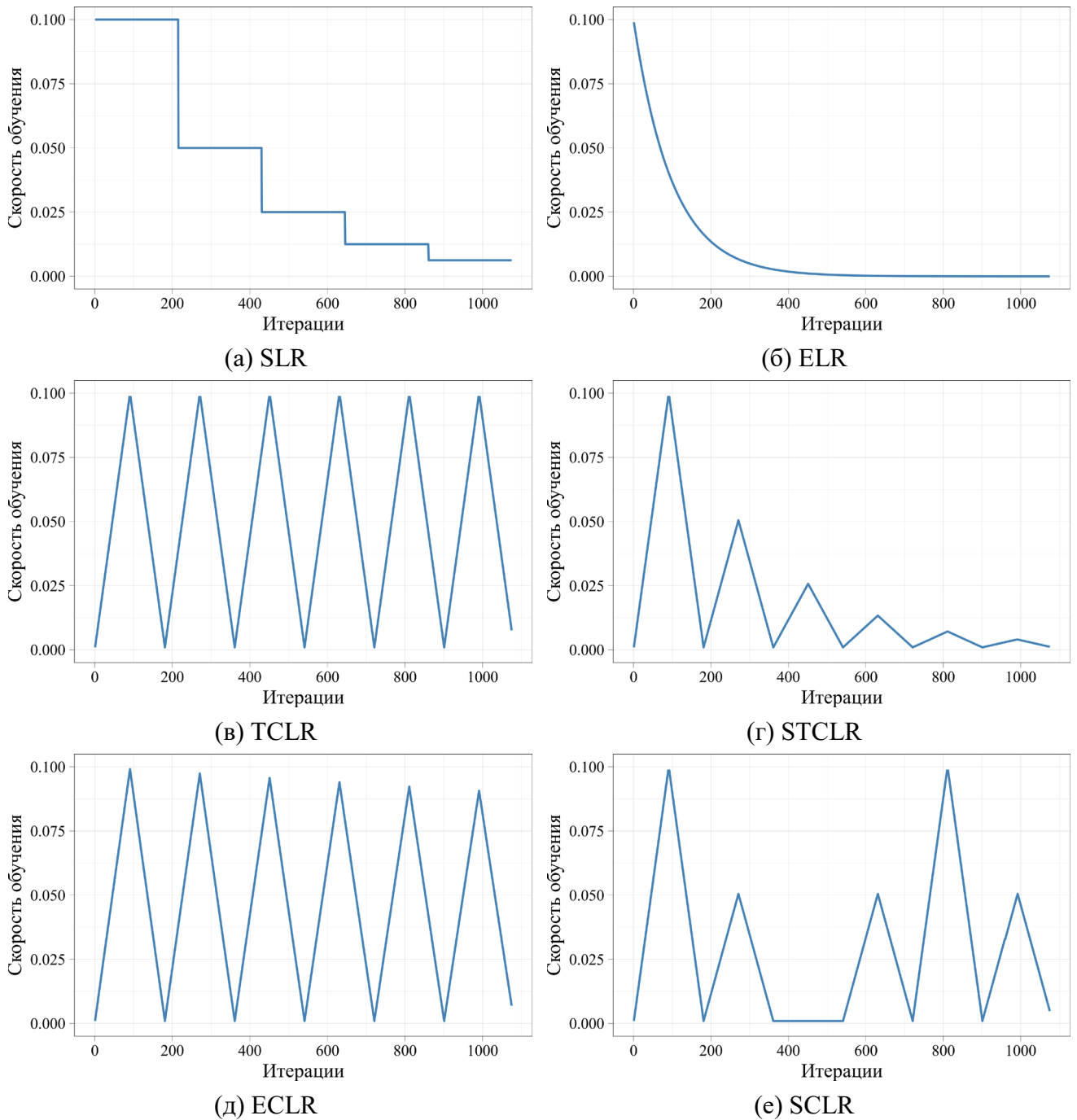


Рисунок 2.17 – Алгоритмы изменения скорости обучения

- Один алгоритм с постоянной скоростью обучения (ConLR).
- Два алгоритма базирующиеся на техники отжига: ступенчатая скорость обучения (SLR) и экспоненциальная скорость обучения (ELR). Графики данных законов отражены на рисунке 2.17 (а) и (б).
- Четыре циклических алгоритма изменения скорости обучения: триангулярно-циклическая скорость обучения (TCLR), ступенчатая триангулярно-циклическая скорость обучения (STCLR), экспоненциально-циклическая скорость обучения (ECLR) и синусоидально-циклическая скорость обучения (SCLR). Графики данных законов отражены на рисунке 2.17 (в), (г), (д) и (е).

Несмотря на то, что увеличение скорости обучения в краткосрочной перспективе может носить негативный характер, в целом, при длительном процессе обучения негативный эффект значительно компенсируется. Так, в работе [147] показано несколько примеров эволюции функции потерь, которая временно значительно отклоняется, но в конечном итоге сходится к более низким значениям функции потерь по сравнению с эталонной фиксированной скоростью обучения. Важным будет отметить, что начальная цель заключается в том, чтобы обученная модель имела высокую обобщающую способность, что позволит точно предсказывать новые данные. В свою очередь, модель с хорошими обобщающими свойствами должна быть робастной и малочувствительной. Последнее говорит о том, что небольшие изменения параметров сети не должны вызывать резких изменений в точности. Стоит иметь в виду, что резкие и глубокие минимумы функции приводят к плохому обобщению, поскольку небольшие изменения значений параметров могут привести к значительному увеличению функции потерь [148]. Поэтому использовании циклической скорости обучения позволяет алгоритму градиентного спуска выпрыгнуть из локального минимума и в конечном итоге сойтись и достичь глобального минимума (см. рисунок 2.18). Дополнительно при обучении глубоких сетей рекомендуется нивелировать резкие и глубокие минимумы, а также стремиться к поиску плоских минимумов для лучшего обобщения [149]. Однако, обычные алгоритмы отжига не могут миновать глубокие

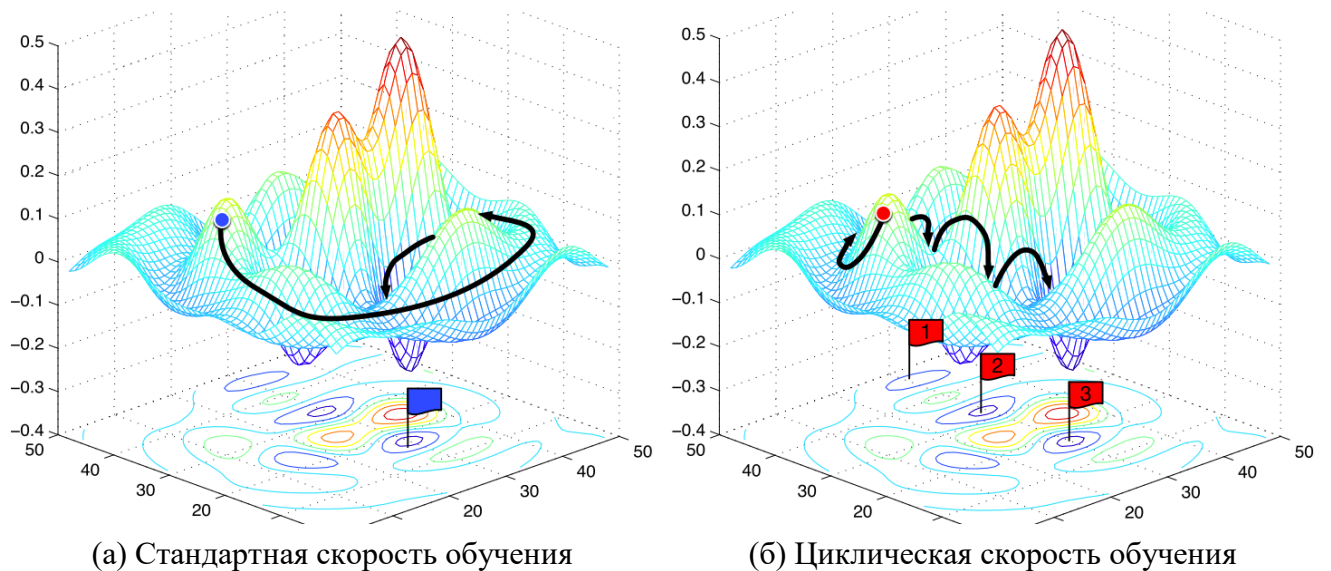


Рисунок 2.18 – Сравнение работы алгоритма градиентного спуска на основе стандартной и циклической скоростей обучения

минимумы, что приводит к нестабильности модели и её использованию в достаточно узком диапазоне данных.

Ещё одним достоинством циклической скорости обучения является то, что она позволяет быстрее миновать плато в области седловых точек. Так, например, значения градиентов в седловой точке (см. рисунок 2.19) имеют небольшие по модулю значения. В связи с тем, что функция обновления параметров зависима от

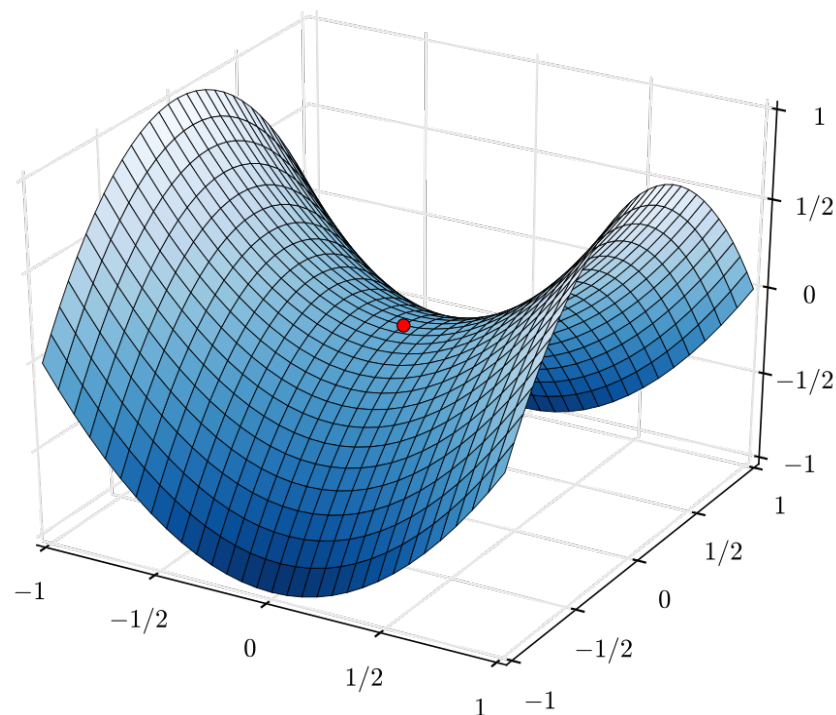


Рисунок 2.19 – График гиперболического параболоида $z = x^2 - y^2$ с отмеченной седловой точкой

градиентов, это приведёт к медленной оптимизации в области плато. Поэтому увеличение скорости обучения могло бы ускорить прохождение алгоритма градиентного спуска по этому участку функции потерь. Однако алгоритмы отжига ограничиваются только уменьшением шага вдоль ландшафта функции потерь, что ещё больше замедляет процесс обучения.

2.5.3 Алгоритм изменения размера батча

На сегодняшний день при обучении различных моделей нейронных сетей, обучающий набор данных разбивается на фиксированное значение батчей. Каждый батч обрабатывается последовательно в рамках одной эпохи. Однако, конкретные экземпляры в рамках одного батча могут быть обработаны параллельно. Важно отметить, что при использовании статического батча с фиксированным значением возникает дилемма. С одной стороны, небольшие значения батча ведут к лучшей сходимости модели. С другой стороны, большие значения батча обеспечивают более производительное распараллеливание, что в свою очередь улучшает вычислительную эффективность и масштабируемость, а также сокращает время обучения. В связи с этой проблемой предлагается использовать стратегию динамического изменения размера батча. На основе такой стратегии процесс обучения начинается с большого значения батча B , что обеспечивает быстрое схождение алгоритма градиентного спуска в область минимума. Большие батчи, используемые на ранних эпохах, демонстрируют достаточно высокий показатель параллелизма. При отсутствии улучшения функции потерь на валидационной выборке происходит уменьшение размера батча. После чего процесс итеративно повторяется. Финальное значение батча может варьироваться в зависимости от типа данных, их количества и времени обучения. В данной работе окончательный процесс поиска глобального минимума функции потерь происходил при значении батча равному 1. Для задания размера батча модели, выполняющей семантическую сегментацию, использовалась последовательность Фибоначчи (см. рисунок 2.20). Стоит отметить, что при использовании полноценных трёхмерных данных размер возможного батча был ограничен физической памятью видеокарты.

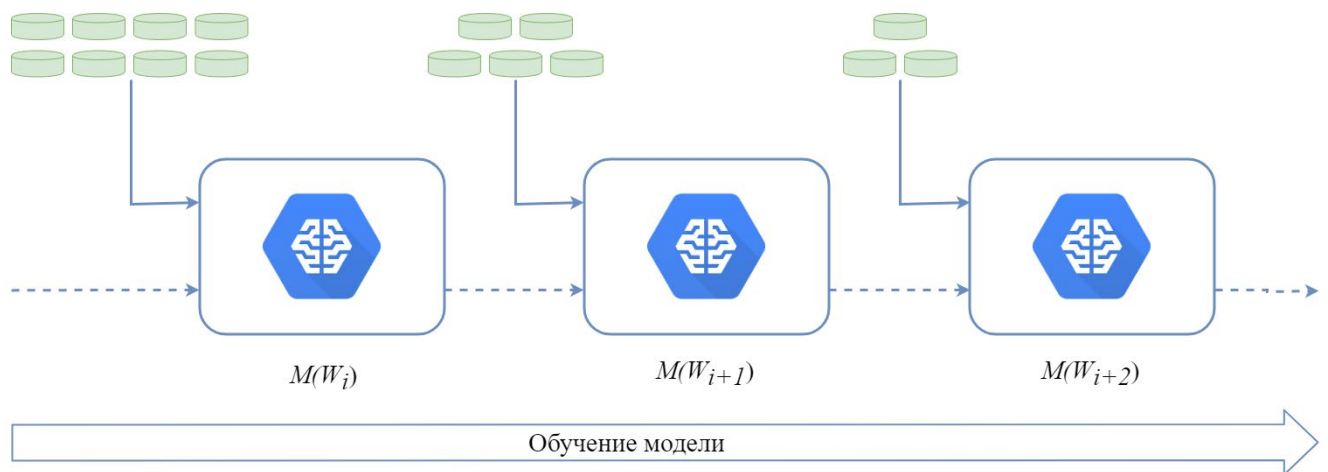


Рисунок 2.20 – Метод обучения на основе переменного размера батча

Дополнительно, для ускорения процесса обучения без потери точности возможно выдерживать постоянным соотношение скорости обучения и размера батча. Необходимо отметить, что направление изменения размера батча может также варьироваться. В данной имплементации размер батча может как расти, так и уменьшаться. Аналогично алгоритму изменения скорости обучения, возможно задать изменение размера батча на основе определённого закона. Несмотря на то, что в научном сообществе преобладает идея увеличения размера батча в процессе обучения [150], необходимо критически отнестись к направлению изменения батча. Использование кросс-валидации может использоваться как один из подходов для оценки оптимального направления изменения батча.

2.6 Алгоритм выбора оптимальных гиперпараметров

При современном развитии подходов машинного обучения всё чаще приходится сталкиваться с оптимизацией гиперпараметров математических моделей. Гиперпараметры представляют собой параметры моделей (чаще всего нейронных сетей и алгоритмов бустинга), значения которых заданы до начала обучения и не изменяются в процессе обучения. Стоит отметить, что гиперпараметры критически влияют на динамику и функцию потерь модели. В связи с этим, выбор значений гиперпараметров является ключевым моментом, стоящим наравне с выбором архитектуры нейронной сети при использовании подхода глубокого обучения. В подавляющем большинстве работ при оценке оптимальности гиперпараметров используется кросс-валидация [151–153]. При

проведении кросс-валидации оценка модели чаще происходит на основании метрики точности, реже на основании функции потерь. Процедура проведения кросс-валидации оценивает большое количество моделей, из которых определяется оптимальная. Схематически подход выбора оптимальной модели на основании кросс-валидации отражён на рисунке 2.21. Однако при попадании выбросов в разбиения при тестировании, модель, вероятнее, будет иметь большой разброс по значениям исследуемой метрики. В таком случае, стандартное отклонение по разбиениям может быть настолько велико, что сравнение средних теряет смысл. Более негативное влияние могут оказывать сами метрики, чувствительные к выбросам, одной из которых, например, является коэффициент детерминации R^2 .

Другими подходами для выбора и оптимизации гиперпараметров являются поиск по решётке (от англ. Grid Search) [154, 155] и случайный поиск (от англ. Randomized Search) [156]. Метод Grid Search (GS) предполагает выполнение процедуры полного перебора по заданному вручную подмножеству гиперпараметров (см. рисунок 2.22 (а)). Этот подход к оптимизации многомерных задач предполагает обучение и оценку модели для каждой комбинации гиперпараметров, указанных в сетке. В подавляющем большинстве случаев оптимизация осуществляется при использовании 5-кратной или 10-кратной кросс-валидации. В качестве результата выбираются значения тех гиперпараметров, с помощью которых было достигнуто наилучшее значение выбранной метрики.

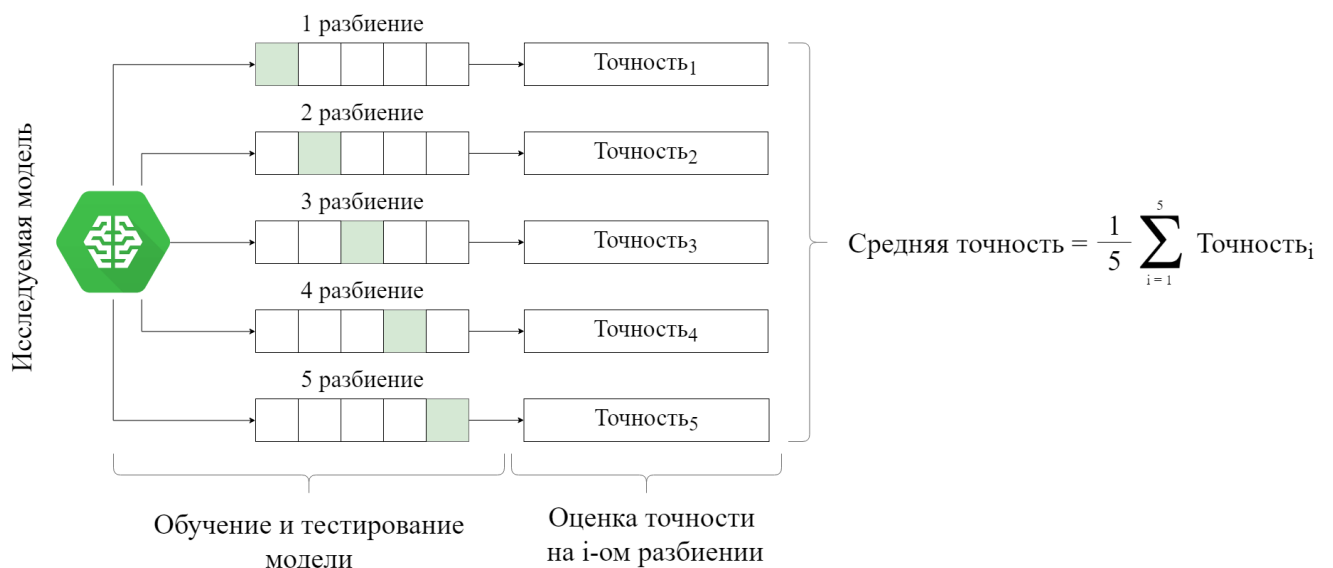


Рисунок 2.21 – Схема тестирования модели на основе кросс-валидации

Метод Randomized Search заменяет перебор всех комбинаций гиперпараметров на их выборку случайным образом (см. рисунок 2.22 (б)). Такой подход выбирает гиперпараметры из случайного распределения для фиксированного числа итераций. Модель обучается и оценивается для каждой выбранной комбинации гиперпараметров.

В современной науке о данных методы Grid Search и Randomized Search нашли своё применение для классических алгоритмов машинного обучения, а именно для метода опорных векторов, регрессии, метода К-ближайших соседей, деревьев решений, случайного леса и других методов. Оценка гиперпараметров на основе методов Grid Search и Randomized Search стала популярной благодаря тому, что вышеописанные алгоритмы машинного обучения имеют низкую сложность и применяются в основном на небольших наборах данных, что не ведёт к большим временным затратам. В связи с чем, процедура выбора гиперпараметров даже на основе полноценного прохода по всему подмножеству пространства гиперпараметров считается приемлемой. Однако в эпоху доминирования нейронных сетей, имеющих миллионы весов, использование классических подходов по оптимизации гиперпараметров приводит к экспоненциальному увеличению временных затрат, что делает невозможным использование данной

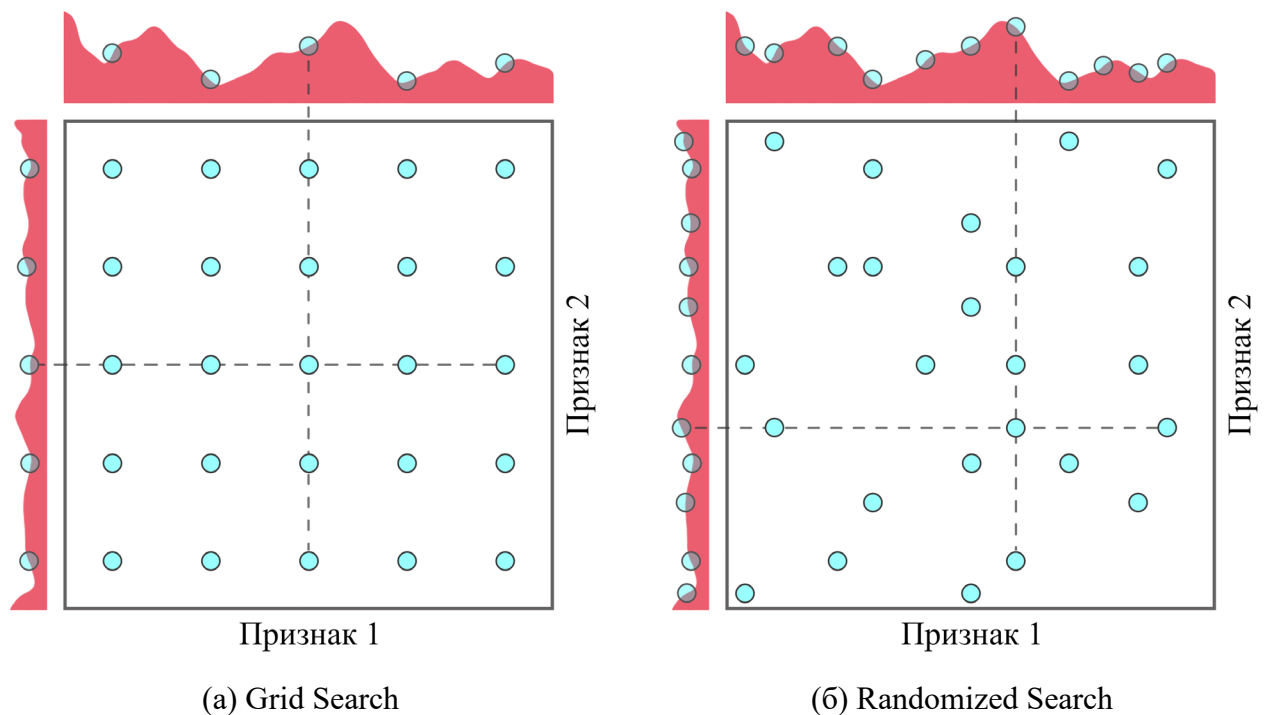


Рисунок 2.22 – Принцип построения подмножества пространства гиперпараметров

стратегии оптимизации. Так, например, при обучении модели в течении суток, оптимизация даже пары гиперпараметров, принимающих 3-4 значения, становится невероятно затратной по времени. В связи с чем, данный недостаток накладывает значительные ограничения на количество оптимизируемых гиперпараметров. Другим недостатком методов Grid Search и Randomized Search является то, что в многомерном пространстве оптимизация значительно затрудняется по причине возникновения большого количества локальных минимумов и седловых точек. В случае, если требуется оптимизировать большое количество гиперпараметров, размер ячейки (расстояние между значениями гиперпараметра) приходится увеличивать, что приводит к снижению вероятности поиска оптимума параметров. Стоит также отметить, что стратегии Grid Search и Randomized Search заведомо перебирают множество неудачных положений и не учитывают предшествующую динамику. В связи с вышеописанными недостатками и ограничениями некоторых современных стратегий оптимизации предлагается алгоритм по выбору оптимальных значений гиперпараметров, учитывающий динамику точности модели на прошлых итерациях и менее подверженный выбросам на разбиениях.

С учётом вышеописанных проблем и ограничений для нахождения оптимальных гиперпараметров и выбора наилучшей модели, предлагается алгоритм, отражённый на рисунке 2.23. В рамках данного алгоритма выполняется следующее:

- Вместо однократного деления данных на n разбиений, выполняется m делений на n разбиений. В результате такого деления общее количество разбиений станет равным $n \times m$, причём размер разбиения (количество экземпляров в разбиении) кросс-валидации остаётся прежним. Массив тестовых данных также будет состоять из $n \times m$ экземпляров.

- Сравниваться будет не средняя точность моделей, подверженная выбросам, а то, насколько одна из моделей превосходит другую на соответствующем разбиении. Математически такая процедура выполняется с помощью сравнения с нулём среднего попарных разностей результатов двух моделей на соответствующих разбиениях.

– Процедура выбора оптимальных значений гиперпараметров выполняется строго последовательно.

При использовании данного подхода значение метрики на определённом разбиении не усредняется с остальными, а сравнивается исключительно со значением метрики на том же разбиении второй модели. Соответственно, выбросы не оказывают влияние на сравниваемые модели из-за попарного сравнения моделей на одинаковых разбиениях. При использовании подобного подхода необходимо оценить статистическую значимость и ответить на вопрос: значимо ли отличаются оценки метрик сравниваемых моделей на $n \times m$ тестовых разбиениях?

Метрики сравниваемых моделей представляют собой связанные выборки. Для оценки таких выборок в работе использован t-критерий Стьюдента для связанных выборок. Однако перед его использованием стоит убедиться в выполнении гипотезы нормальности с достаточной точностью и применимости данного критерия. В случае ненормальности распределения метрик предлагается использовать непараметрический аналог парного критерия Стьюдента – W-критерий Вилкоксона. Для оценки статистики t-критерия Стьюдента [157, 158] используется следующая формула:

$$T(X_1^n, X_2^n) = \frac{M_d}{\sigma_d / \sqrt{n}} = \frac{E(X_1) - E(X_2)}{\sigma_d / \sqrt{n}} \quad (2.35)$$

где X_1^n, X_2^n – массивы значений метрики по тестовым разбиениям для сравниваемых моделей, M_d – среднее арифметическое разностей показателей, σ_d – среднее квадратическое отклонение попарных разностей, n – число исследуемых экземпляров.

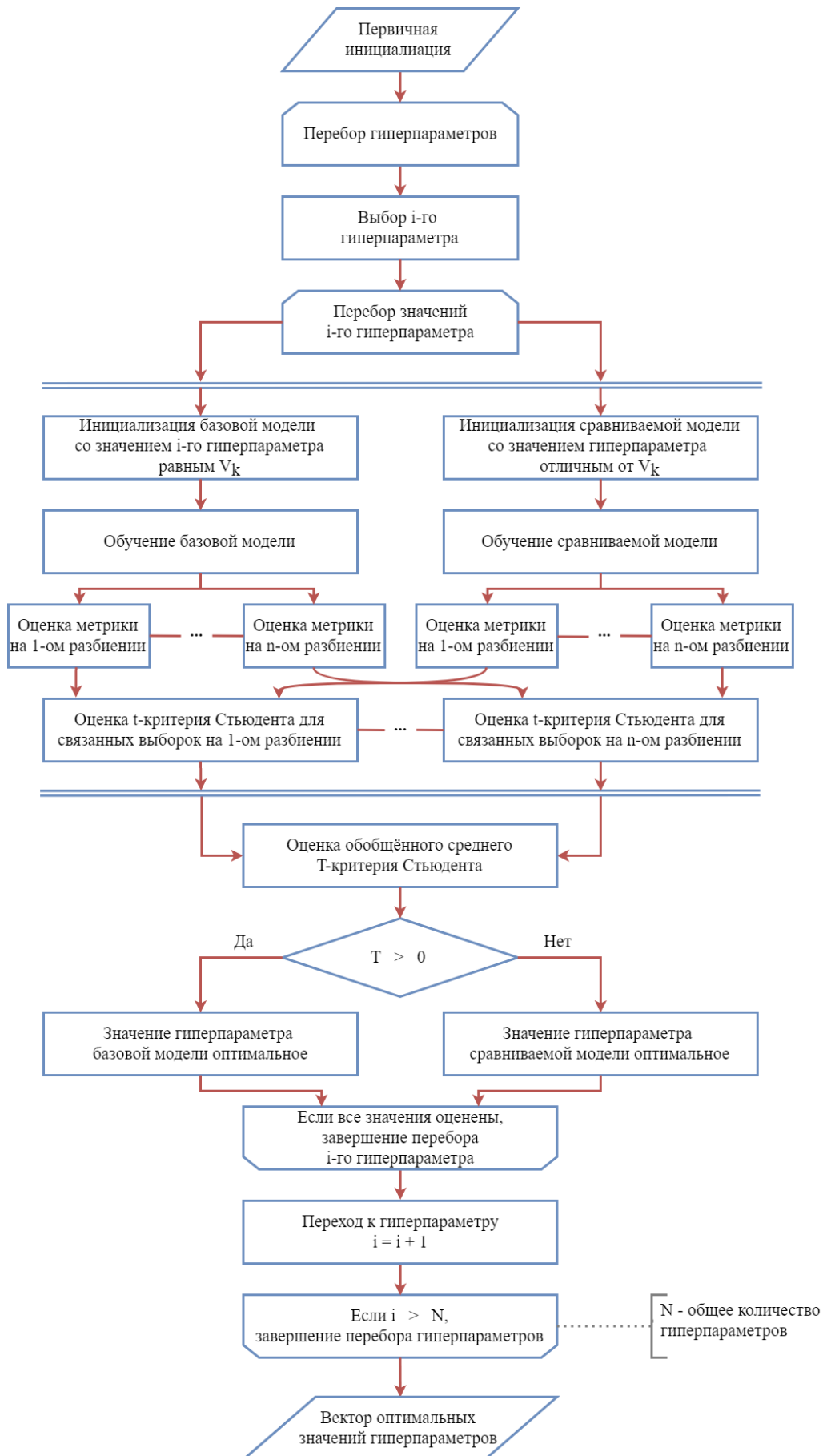


Рисунок 2.23 – Алгоритм выбора оптимальных значений гиперпараметров

Нулевая гипотеза H_0 предлагаемого подхода состоит в том, что сравниваемые модели дают одинаковые результаты. В таком случае t-статистика распределена по Стьюденту с математическим ожиданием в нуле. Соответственно, чем больше она отклоняется от нуля, тем меньше вероятность выполнения нулевой гипотезы и случайного совпадения средних в числителе дробей. Стоит отметить, что знаменатель (2.35) содержит стандартное отклонение попарных разностей σ_d . Таким образом, стандартное отклонение определённой метрики на разбиениях кросс-валидации для каждой модели в отдельности не сравнивается. Предлагаемый подход оценивает то, насколько одна модель эффективнее выполняет предсказание по сравнению с другой.

Для выбора оптимальных значений гиперпараметров модели вычисляется t-статистика для каждого разбиения, после чего вычисляется обобщённая T-статистика путём усреднения последних. Если рассчитанный результат положителен, то базовая модель выполняет предсказание эффективнее, чем сравниваемая модель. Последнее означает, что гиперпараметры базовой модели позволяют получить более высокую точность. Поэтому модель с наибольшей отрицательной обобщённой T-статистикой эффективнее влияет на точность. Тем самым можно заключить, что чем больше T-статистика по модулю и чем меньше значение p , тем существеннее отличия. Схематично сравнение двух моделей и оценка обобщённой T-статистики отражены на рисунке 2.24.

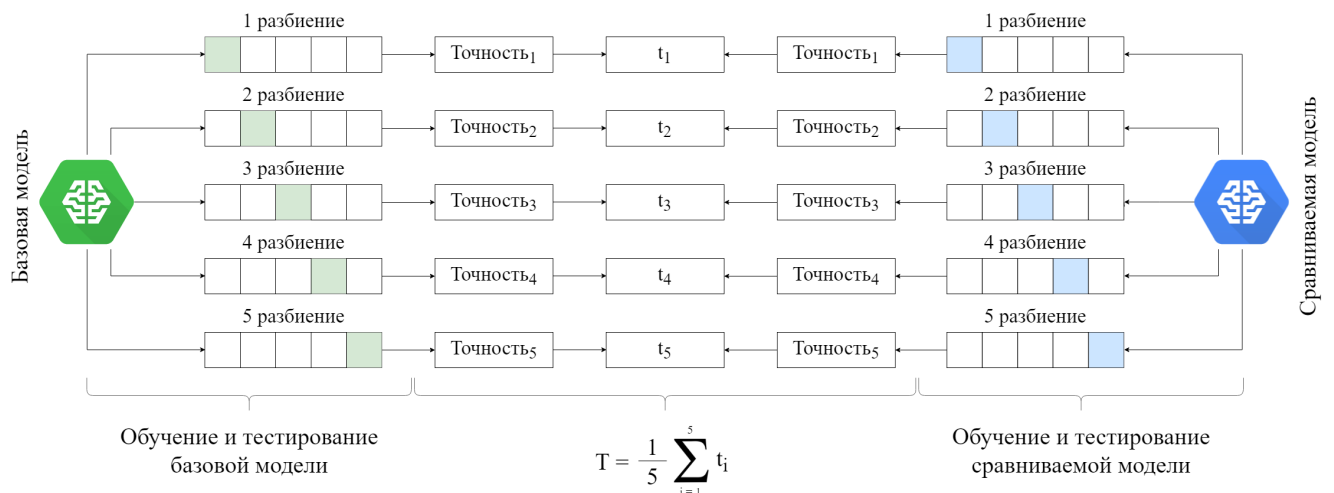


Рисунок 2.24 – Оценка базовой и сравниваемой моделей на основе t-критерия Стьюдента

Работа предлагаемого алгоритма выбора оптимальных значений гиперпараметров схематично представлена на рисунке 2.25. В представленном примере предполагается использование 5 гиперпараметров, количество принимаемых значений которых варьируется от 3 до 6. При использовании классического алгоритма Grid Search общее количество тестируемых моделей составит 1080 ($3 \times 3 \times 5 \times 4 \times 6$) значений. В свою очередь, для поиска оптимальных значений гиперпараметров предлагаемый подход предполагает оценку 21 ($3 + 3 + 5 + 4 + 6$) модели. Подобное сокращение числа тестируемых моделей обусловлено учётом динамики при оптимизации, в том числе отрицательной, а также последовательной оценкой гиперпараметров. В представленном ниже примере выбор оптимальных значений начинается с гиперпараметра 1. Стоит отметить, что первичная инициализация гиперпараметров может выполняться как вручную, так и алгоритмически. При выполнении процедуры оптимизации возможно совпадение

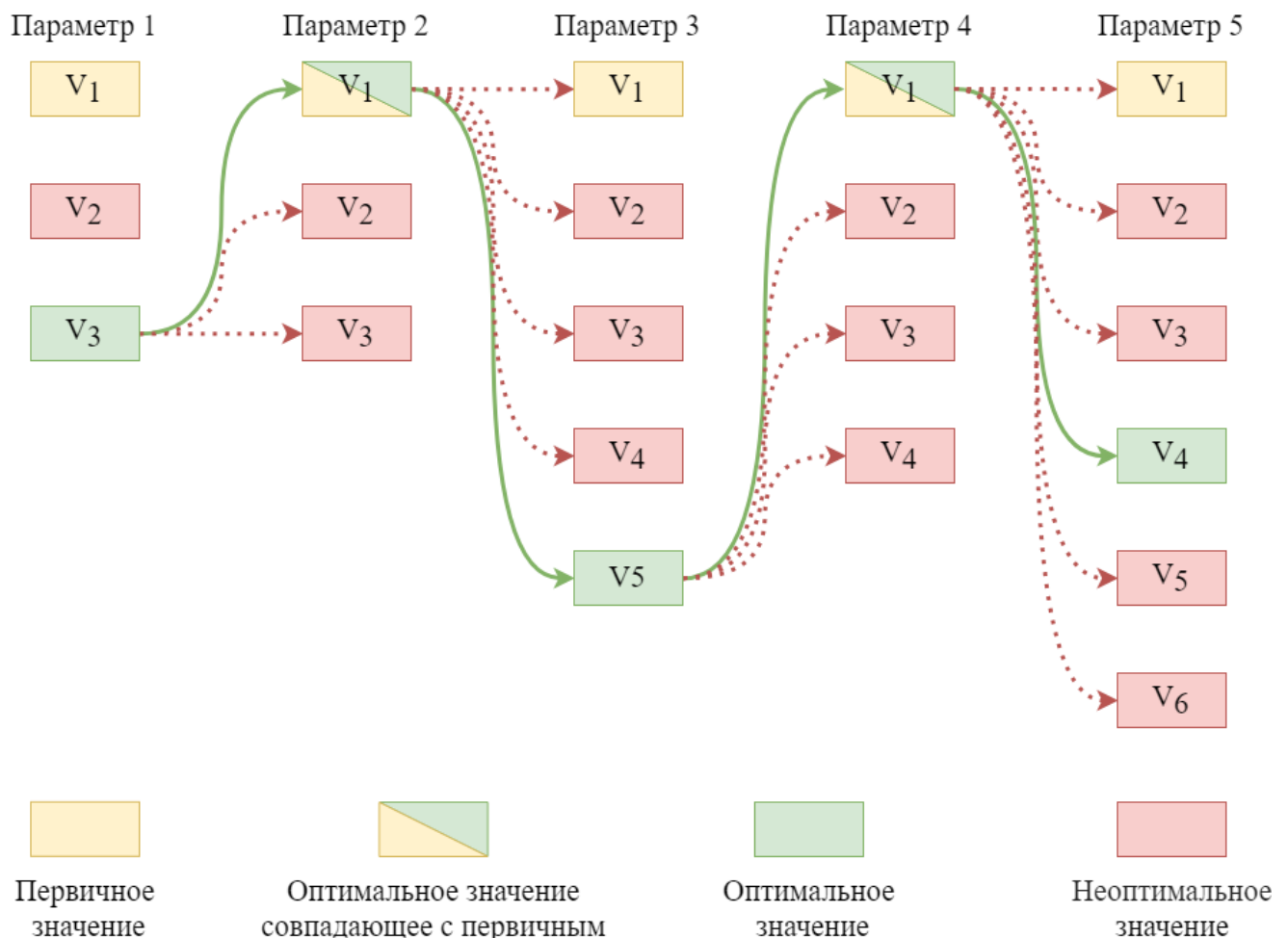


Рисунок 2.25 – Процедура выбора оптимальных значений гиперпараметров

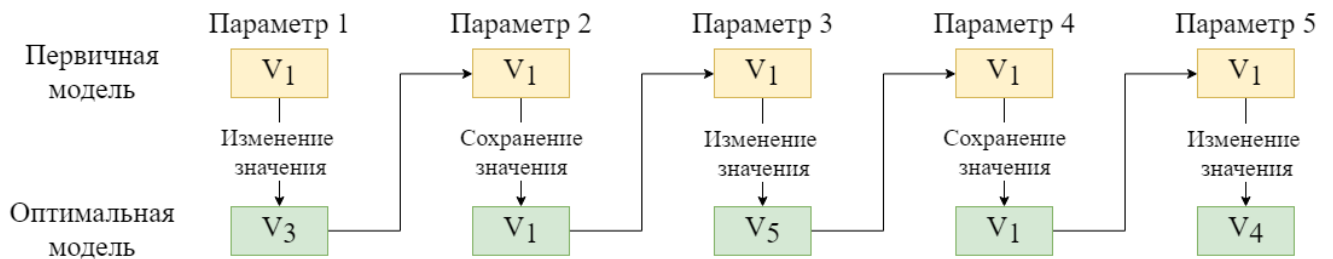


Рисунок 2.26 – Сравнение значений гиперпараметров первичной и оптимальной моделей

первичного инициализированного и оптимального значений (см. гиперпараметр 2 и гиперпараметр 4 на рисунке 2.25).

После того, как была полностью выполнена процедура выбора, значения гиперпараметров первичной модели, вероятнее всего, будут отличаться от оптимальной модели. В рассматриваемом гипотетическом примере с 5 гиперпараметрами значения 3 из них изменились. Изменёнными являются гиперпараметры 1, 3 и 5 (см. рисунок 2.26).

Важным будет отметить, что под термином «оптимальность» подразумевается квазиоптимальность, то есть решения или значения, близкие к оптимальным, которые в последующем могут использоваться в качестве оптимальных в практических целях. Связано это с тем, что проверка абсолютно всех принимаемых значений гиперпараметров невозможна, поэтому исследуемая область значений ограничивается.

При проведении выбора оптимальных значений модели использовались следующие гиперпараметры:

- Тип dropout слоя. Исследуемые значения (3 значения): без dropout слоя, стандартный dropout, пространственный dropout.
- Метод нормализации. Исследуемые значения (3 значения): нет нормализации, batch нормализация, instance нормализация.
- Метод оптимизации. Исследуемые значения (6 значений): SGD, RMSprop, Adagrad, Adadelata, Adam, Adamax.
- Метод регуляризации. Исследуемые значения (4 значения): нет регуляризации, LASSO регуляризация (L1-норма), Ridge регуляризация (L2-норма), Elastic net регуляризация (L1, L2-норма).
- Размер батча. Исследуемые значения (8 значений): 1, 2, 3, 4, 5, 6, 7, 8.

- Тип активации. Исследуемые значения (4 значения): ReLU, ELU, Leaky ReLU, PReLU.
- Скорость обучения. Исследуемые значения (8 значений): 0.001, 0.005, 0.01, 0.05, 0.10, 0.15, 0.20, 0.25.
- Метод изменения скорости обучения. Исследуемые значения (7 значений): постоянная скорость обучения, ступенчатая скорость обучения, экспоненциальная скорость обучения, треугольно-циклическая скорость обучения, ступенчатая треугольно-циклическая скорость обучения, экспоненциально-циклическая скорость обучения, синусоидально-циклическая скорость обучения.
- Метод инициализации. Исследуемые значения (10 значений): ones, random normal, random uniform, Glorot normal, Glorot uniform, He normal, He uniform, Lecun normal, Lecun uniform, orthogonal.
- Дилатация. Исследуемые значения (8 значений): 1, 2, 3, 4, 5, 6, 7, 8.
- Коэффициент свёрточных фильтров. Исследуемые значения (8 значений): 1, 2, 3, 4, 5, 6, 7, 8. Коэффициент фильтров используется для вычисления количества фильтров свёртки/деконволюции на определенном уровне сети.

Исходя из такого большого количества значений гиперпараметров ни GS, ни RS стратегии не способны справиться с поставленной задачей выбора оптимальных значений. При использовании 11 гиперпараметров, каждый из которых принимает от 3 до 10 значений, общее количество моделей составит 204 374 016. Также стоит принять во внимание, что GS и RS стратегии используют кросс-валидацию, что кратно увеличивает количество моделей в зависимости от числа разбиений. В свою очередь, при использовании предлагаемого подхода количество моделей для оптимизации гиперпараметров составит 690.

2.7 Выводы по второй главе

1. В связи с гомогенностью выборки и небольшим количеством примеров в исходном наборе данных был предложен алгоритм синтеза трёхмерных ультразвуковых данных. Для создания случайных трансформаций объекта

исследования предлагается использовать систему координат, которая, вне зависимости от изгиба и формы объекта, позволит однозначно идентифицировать каждую точку. Это позволит трансформировать изображение под любую заданную конфигурацию. За основу предлагается взять цилиндрическую систему координат, модифицировав её заменой оси Z на сплайн, вдоль которого отсчитывается h координата. Предлагаемый алгоритм генерации данных позволяет уменьшить субъективность при разметке, которая может достигать до 20%, а также минимизировать рутинность за счёт автоматизации.

2. Разработана V-net архитектура свёрточной нейронной сети на основе плотных пробросов признаков. Для понижения размерности входных данных используется энкодер, в свою очередь декодер преобразует закодированную информацию, получая тем самым сегментацию. Между энкодером и декодером формируется закодированный набор признаков, который обладает меньшей размерностью по сравнению с входными данными. Это минимальный набор признаков, достаточный для выполнения эффективной сегментации. Дополнительно на каждом уровне сети используются слои обратной свёртки, которые итеративно разворачивают карты признаков до размеров исходного изображения и получают финальную сегментацию. Слои обратной свёртки обучаемы, что позволяет модели учитывать специфику данных точнее и эффективнее слоёв повышающей дискретизации.

3. Для улучшения прохода информации внутри архитектуры сети были реализованы плотные пробросы признаков. Использование топологии нейронных сетей с подобными пробросами позволяет сгладить рельеф функции потерь, что в свою очередь ведёт к более эффективной минимизации при использовании метода обратного распространения ошибки. Дополнительно проброс признаков позволяет решить проблему исчезающих градиентов для глубоких архитектур, которые имеют хаотичный и невыпуклый ландшафт функции потерь.

4. Для более эффективного обучения модели было предложено использовать алгоритм циклического изменения скорости обучения, а также алгоритм динамического изменения размера батча. Подобные алгоритмы

позволяют миновать локальные минимумы или быстрее из них выходить, успешно проходить седловые точки, а также ускорять процесс обучения. Изменение скорости обучения и размера батча возможно задавать функционально. Альтернативным вариантом обучения без потери точности является обучение при поддержании постоянного соотношения скорости обучения и размера батча. В процессе обучения, в качестве функции потерь использовалась функция, основанная на индексе Дайса. Модели на основе такой функции потерь эффективнее справляются с сегментацией небольших объектов, чем, например, функция потерь на основе кросс-энтропии. В исследуемом наборе данных размер объекта интереса (медицинского инструмента) составляет 1415 ± 529 вокселей, что представляет 0.07% изображения, имеющего размер $128 \times 128 \times 128$ вокселей.

5. Предлагаемый алгоритм оптимизации на основе t-критерия Стьюдента позволяет с меньшими затратами выбрать оптимальные значения гиперпараметров для исследуемой модели, чем такие подходы, как Grid Search или Randomized Search. Данный алгоритм отличается от существующих тем, что он учитывает динамику на прошлых итерациях и не выполняет полный перебор по подмножеству пространства гиперпараметров. Нулевая гипотеза H_0 при использовании такого подхода состоит в том, что базовая и сравниваемая модели дают одинаковые результаты. Соответственно, чем больше t-статистика отклоняется от нуля, тем меньше вероятность выполнения нулевой гипотезы и случайного совпадения средних значений точности. Как результат, предлагаемый алгоритм оценивает то, насколько одна модель точнее и эффективнее выполняет сегментацию, чем другая.

6. На этапе предсказания была использована стратегия предсказания на основе алгоритма test-time аугментации. Точность предсказания, полученная с использованием данного алгоритма, в худшем случае будет равна точности на основе стандартного предсказания одного примера. В случае, если модель слабо выполняет предсказания на новых данных, использование алгоритма test-time аугментации повышает шансы модели на более корректное прогнозирование, тем самым увеличивая точность сегментации.

3 Оценка полученных результатов

В данной главе описываются результаты сегментации, полученные на основе предлагаемых методов и алгоритмов. Используя данные медицинского инструмента для обучения, предлагаемая V-net модель с плотными пробросами признаков сравнивалась с рядом популярных моделей, предназначенных для сегментации, а именно FCN, SegNet, Deep Medic, Pseudo 3D U-net, U-net и разреженная U-net. Для оценки работоспособности и надёжности предложенный подход был протестирован не только для сегментации медицинского инструмента, но и для сегментации различных анатомических структур, таких как левое предсердие, поджелудочная железа, селезёнка, гиппокамп и печень.

3.1 Синтез медицинского инструмента

3.1.1 Результаты сегментации синтезированных данных

Для оценки влияния синтетических данных на точность и обобщающую способность нейронных сетей были обучены 8 моделей на основе стандартной и разреженной U-net архитектуры, специфика которых описана в параграфах 2.4.1 и 2.4.2. Обучение проводилось с учётом вариации значений параметра RDR (сокр. от Real Data Ratio). Параметр RDR отражает соотношение синтетических и реальных примеров в обучающей выборке и рассчитывается следующим образом:

$$RDR = \frac{RS}{RS + SS} \quad (3.1)$$

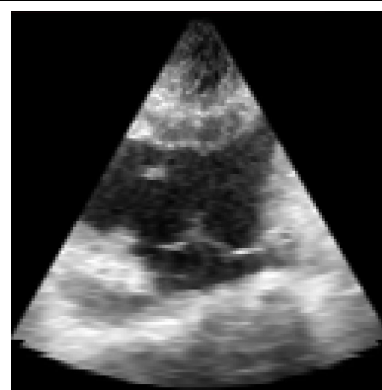
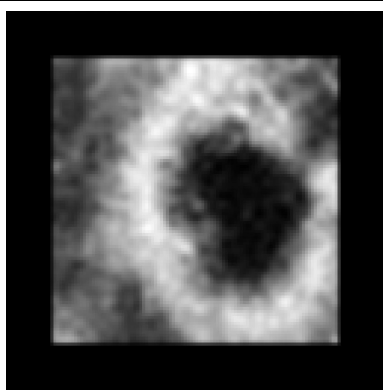
где RS и SS – количество реальных и синтетических изображений соответственно.

Основная идея выполнения экспериментов заключалась в том, чтобы оценить влияют ли дополнительные синтетические данные на обобщающую способность модели, которая в последующем тестируется и используется для сегментации реальных данных. Для выявления положительной, нейтральной или отрицательной тенденций влияния синтетических данных каждая из моделей обучалась на 4 наборах данных при различном соотношении реальных и синтетических изображений. Описание распределения данных для каждого эксперимента отражено в таблице 3.1.

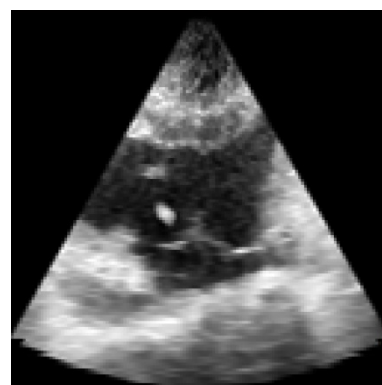
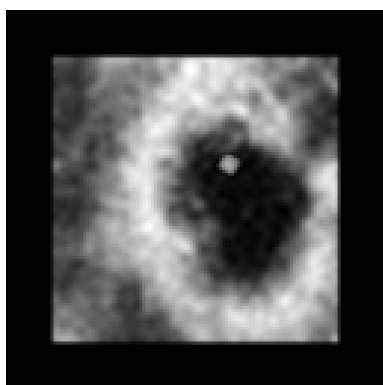
Используя данные с пустыми полостями, было синтезировано 225 трёхмерных изображений эхокардиографии с хирургическим инструментом внутри полости сердца. Один из примеров сгенерированного катетера внутри левого предсердия отражён на рисунке 3.1.

Таблица 3.1 – Распределение данных для проводимых экспериментов

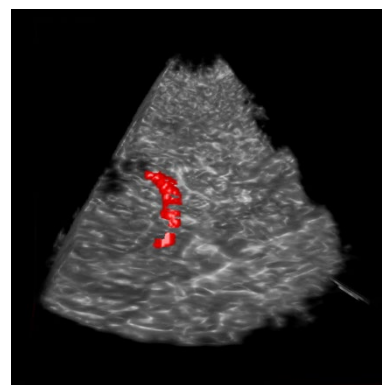
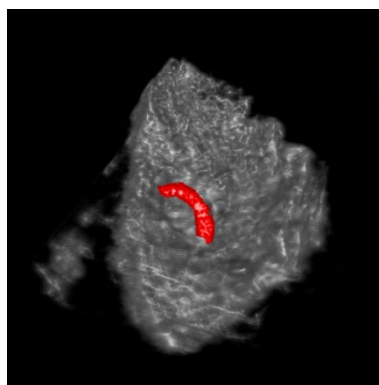
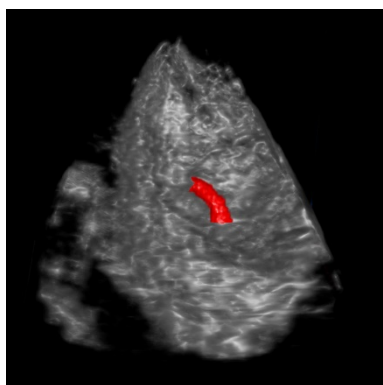
Эксперимент	RS	SS	Общее число изображений	RDR
1	75	0	75	1.00
2	75	75	150	0.50
3	75	150	225	0.33
4	75	225	300	0.25



(а) Исходные данные



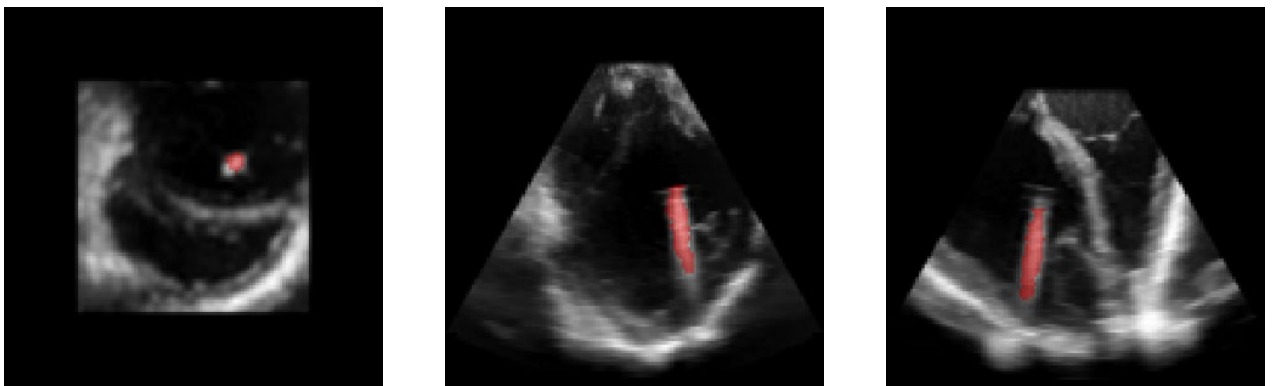
(б) Синтезированные данные



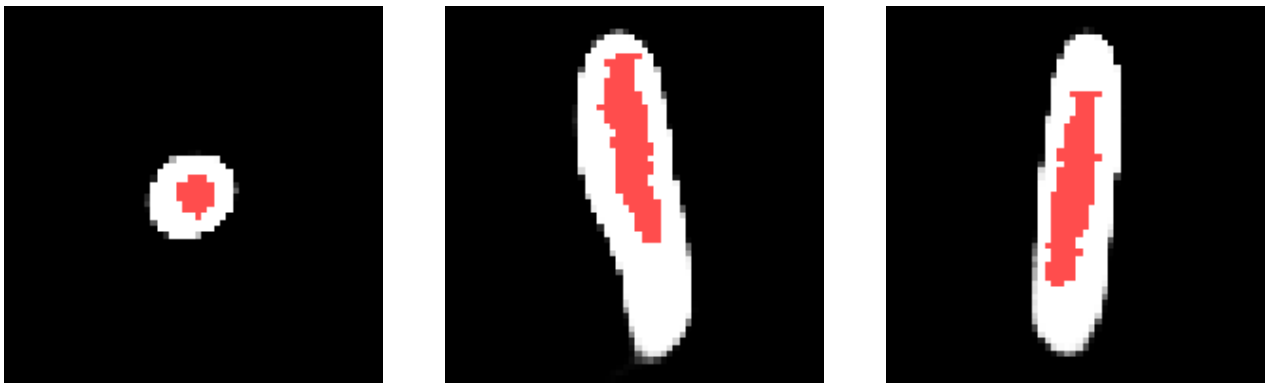
(в) Трёхмерная визуализация катетера внутри сердца

Рисунок 3.1 – Результат синтеза искусственных данных предложенным алгоритмом

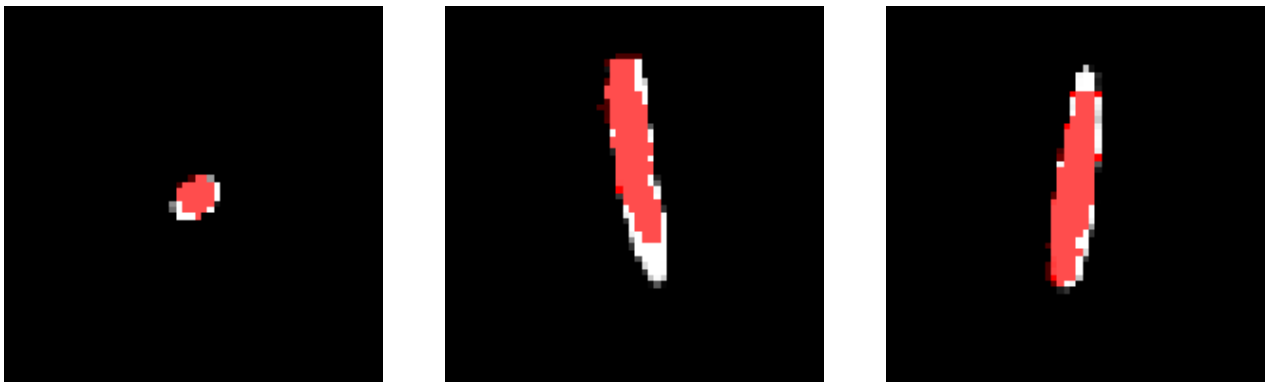
После получения синтетических данных, обе версии U-net архитектуры были обучены согласно экспериментам, описанным в таблице 3.1. В качестве визуального сравнения исследуемых моделей используется рисунок 3.2. Как видно, модель на основе стандартной U-net архитектуры сегментирует катетер недостаточно точно. Это объясняется меньшей сложностью сети по сравнению с разреженной версией. Для более подробной оценки точности сегментации были получены диаграммы размаха, отражённые на рисунках 3.3 и 3.4. Данные диаграммы показывают динамику изменения точности в зависимости от параметра RDR. Видно, что разреженная версия U-net архитектуры выполняет сегментацию



(а) Исходные данные с эталонной разметкой (выделена красным цветом)



(б) Сегментация стандартной U-net (белый цвет) и эталонная сегментация (красный цвет)



(в) Сегментация разреженной U-net (белый цвет) и эталонная сегментация (красный цвет)

Рисунок 3.2 – Сегментация медицинского инструмента моделями обученных при $RDR = 0.25$

качественнее. Стоит также отметить, что наблюдается обратно пропорциональная зависимость между индексом Дайса и параметров RDR, т.е. чем меньше RDR, тем больше индекс Дайса. Основные результаты точности по архитектурам в зависимости от параметра RDR отражены в таблице 3.2.

Несмотря на то, что данный алгоритм направлен на усиление гетерогенности и репрезентативности данных, он имеет ряд ограничений. Одно из ограничений относится к форме объекта. Данная версия алгоритма протестирована на объектах цилиндрической формы. Другое ограничение относится к тому, что алгоритм недостаточно точно оценивает эффекты исследуемой модальности, а именно шум, текстуру и структуру. Однако подобный эффект частично решается использованием трилинейной интерполяции, встроенной в алгоритм. Для более полного решения возможно использование дополнительных алгоритмов обработки

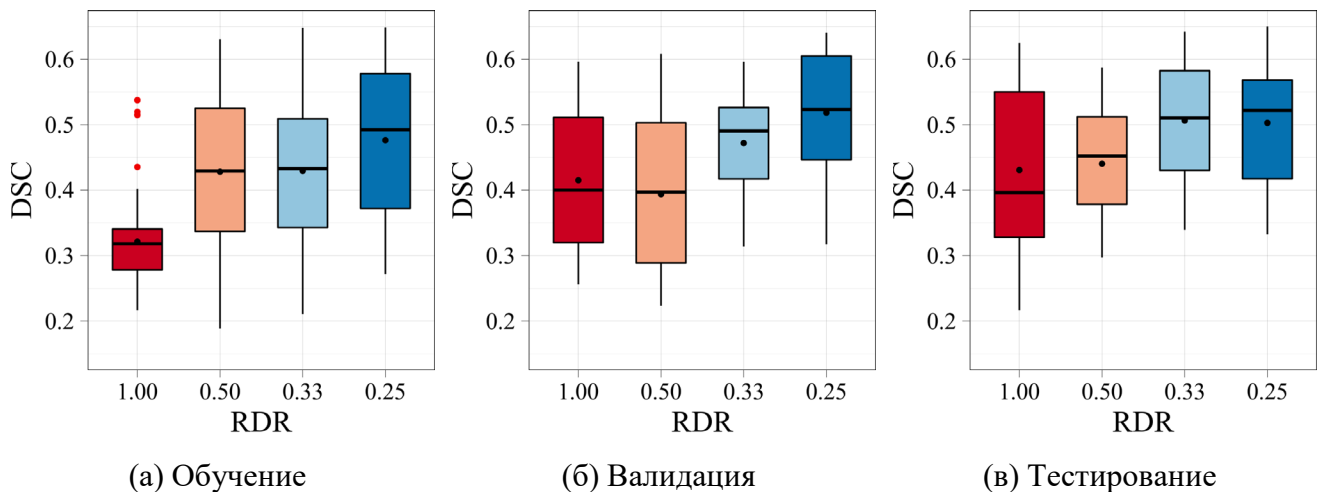


Рисунок 3.3 – Индекс Дайса для стандартной U-net архитектуры при разных значениях RDR

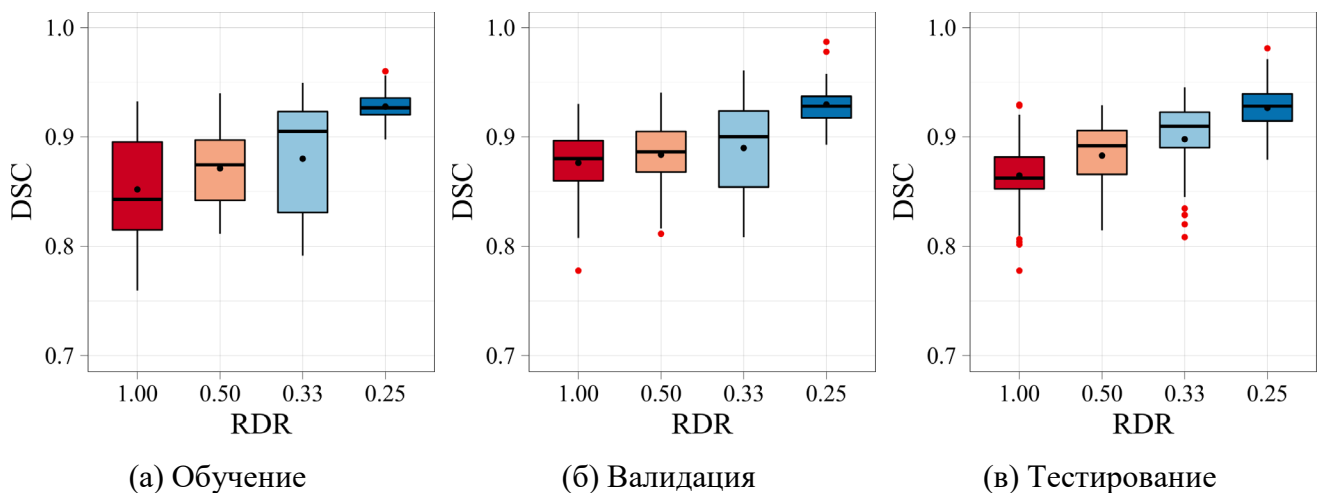


Рисунок 3.4 – Индекс Дайса для разреженной U-net архитектуры при разных значениях RDR

Таблица 3.2 – Индекс Дайса при различных значениях параметра RDR

Этап	Архитектура	RDR			
		1.00	0.50	0.33	0.25
Обучение	Стандартная U-net	0.32±0.06	0.43±0.11	0.44±0.12	0.49±0.12
	Разреженная U-net	0.85±0.05	0.87±0.03	0.88±0.05	0.93±0.01
Валидация	Стандартная U-net	0.44±0.13	0.39±0.12	0.47±0.09	0.52±0.09
	Разреженная U-net	0.88±0.03	0.88±0.03	0.89±0.04	0.93±0.02
Тестирование	Стандартная U-net	0.43±0.12	0.44±0.09	0.51±0.09	0.53±0.10
	Разреженная U-net	0.86±0.04	0.88±0.03	0.90±0.03	0.93±0.02

изображений, а также имплементации генеративно-сопоставительных сетей и вариационных автоэнкодеров. Если рассматривать скорость работы алгоритма, то для синтеза одного трёхмерного изображения размером $128 \times 128 \times 128$ требуется в среднем 54 секунды. Это связано с достаточно длительной операцией интегрирования сплайн-функции по длине, которая производится для каждого вокселя. Однако данный фактор не является критически важным, потому как в подавляющем большинстве случаев подготовка данных производится оффлайн. В свою очередь, использование реальных входных данных позволяет получать изображения с высокой степенью правдоподобия, а прямая кинематика позволяет моделировать форму реального медицинского инструмента.

3.1.2 Оценка результатов сегментации стандартной U-net модели

В целях первичной оценки значимости различий моделей при разных значениях метрики RDR использовался дисперсионный анализ (метод ANOVA). Данный метод не определяет в каких именно группах наблюдаются значимые отличия, но при этом показывает их наличие или отсутствие. Нулевая гипотеза H_0 дисперсионного анализа указывает на отсутствие отличий между группами. В таком случае подразумевается, что все группы по индексу Дайса относятся к одной генеральной совокупности и соответственно их средние равны друг другу, то есть $\mu_1 = \mu_2 = \mu_3 = \mu_4$. В свою очередь, альтернативная гипотеза H_1 выдвигает предположение о том, что средние в исследуемых группах не равны.

После проведения дисперсионного анализа были получены результаты, которые отражены в таблице 3.3. Как видно, значение F больше критического значения $F_{кр}$ при выбранном уровне значимости равном 0.05 для каждой подвыборки. В связи с этим есть все основания, чтобы отвергнуть нулевую гипотезу и констатировать, что, по крайней мере, одна из четырёх выборок имеет существенно различающееся среднее значение и, таким образом, принадлежит к совершенно иной генеральной совокупности. Стоит отметить, что вычисленные значения p также меньше уровня значимости, что говорит о том, что нулевая гипотеза может быть отклонена.

В качестве более детальной оценки был проведён дополнительный статистический анализ значений индекса Дайса при разных значениях метрики RDR. При визуальном инспектировании видно, что распределения четырёх рассматриваемых выборок отличаются. Причём распределения индекса Дайса для трёх выборок распределены бимодально, т.е. выборки при $RDR = 1.00$, $RDR = 0.50$ и $RDR = 0.33$. В свою очередь, распределение индекса Дайса при $RDR = 0.25$ характеризуется отрицательной (правосторонней) асимметрией. Распределения индекса Дайса стандартной U-net архитектуры при разных значениях метрики RDR отражены на рисунке 3.5. Дополнительно на этих графиках отмечены средние значения распределений (штриховые линии), а также стандартные отклонения от среднего в положительную и отрицательную сторону (пунктирные линии).

В связи с тем, что проводится оценка зависимости индекса Дайса от метрики RDR, то есть от степени смешивания данных, ниже приводятся сравнения распределений индекса Дайса. В качестве базового распределения для сравнения используется распределение при $RDR = 1.00$, то есть случай обучения и

Таблица 3.3 – Результаты ANOVA тестирования стандартной U-net архитектуры

Параметр	Обучение	Валидация	Тестирование
F	83.2	10.9	10.4
$F_{кр}$	2.6	2.7	2.7
p	3.6E-46	1.3E-06	2.4E-06
Значимы ли различия?	Да	Да	Да

тестирования стандартной U-net модели исключительно на реальных данных. Остальные распределения, полученные при $RDR = 0.50$, $RDR = 0.33$ и $RDR = 0.25$, сравниваются с базовым. Данное сравнение отражено на рисунке 3.6. При визуальной оценке видно, что среднее значение точности увеличивается при уменьшении метрики RDR. Однако при значениях RDR близких к 0.25 наблюдается асимптотическое насыщение точности. Это проявляется тем, что среднее значение индекса Дайса растёт медленнее при уменьшении RDR. Так среднее значение индекса Дайса при $RDR = 0.50$ отличается незначительно от среднего значения базового распределения. Затем при $RDR = 0.33$ наблюдается значительное увеличение точности. А при сравнении средних значений при $RDR = 0.33$ и $RDR = 0.25$, уже наблюдается замедление при изменении точности.

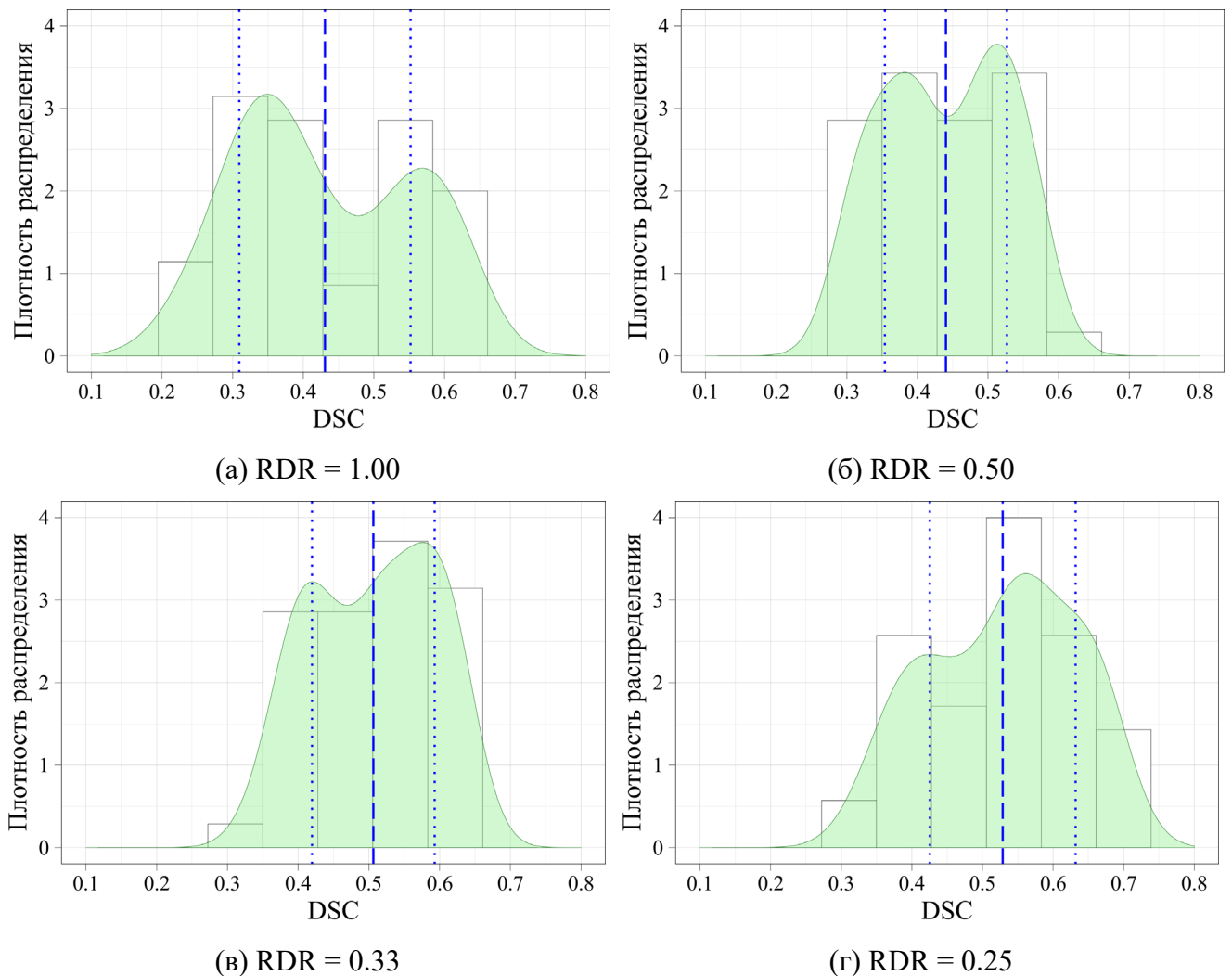
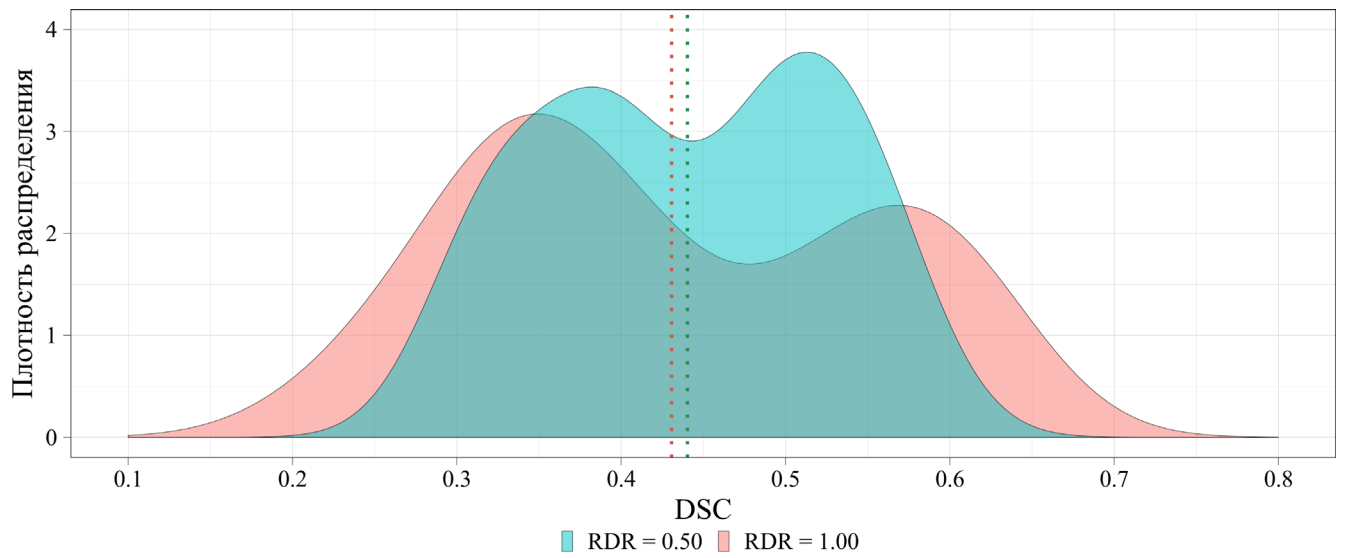
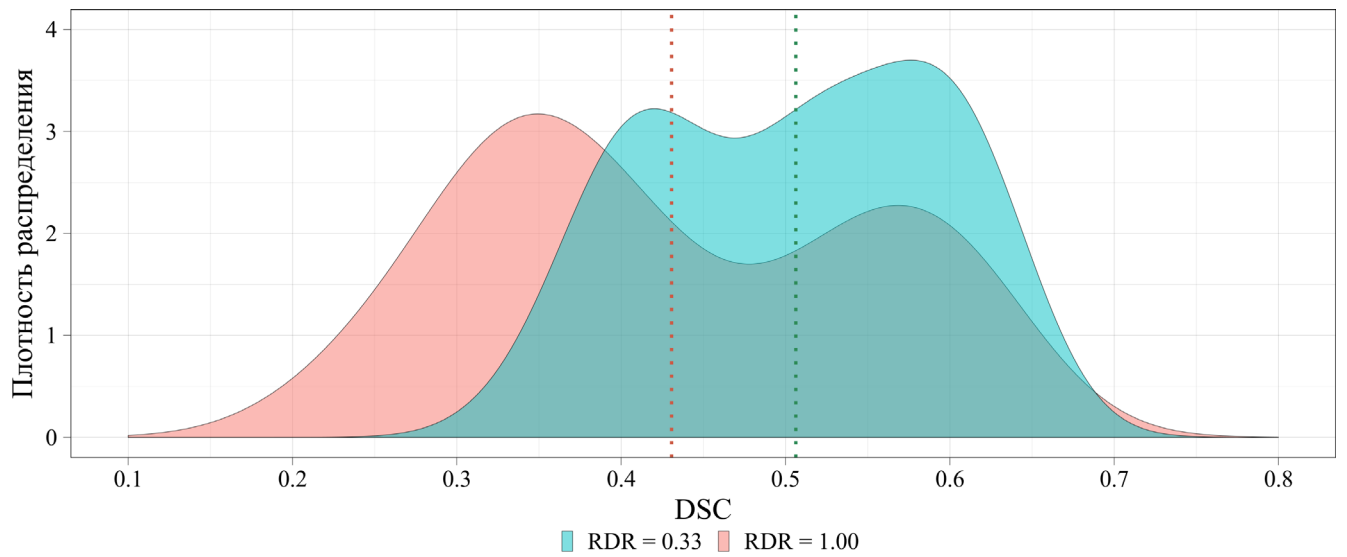


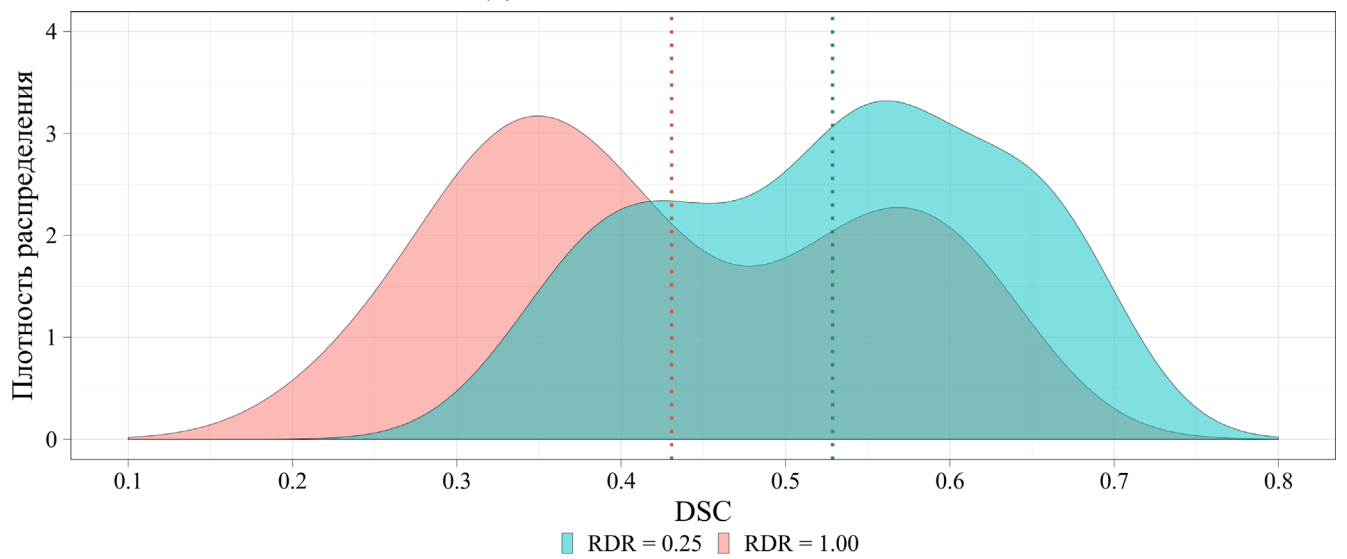
Рисунок 3.5 – Распределение индекса Дайса стандартной U-net архитектуры при разных значениях метрики RDR



(a) RDR = 0.50 и RDR = 1.00



(б) RDR = 0.33 и RDR = 1.00



(в) RDR = 0.25 и RDR = 1.00

Рисунок 3.6 – Сравнение распределений индекса Дайса стандартной U-net архитектуры при разных значениях метрики RDR

Перед проверкой формальных статистических гипотез была проведена оценка типа распределения. Визуальная проверка на нормальность каждого из распределений стандартной U-net архитектуры выполнялась с помощью квантильных графиков, отражённых на рисунке 3.7. В связи с тем, что для большинства из рассматриваемых распределений характерно бимодальное поведение, исследуемые квантильные распределения S-образно изогнуты, а точка их перегиба наблюдается в области нулевого квантиля нормального распределения. В дополнении стоит отметить значительные отклонения исследуемых квантилей от нормальных при низких и высоких значениях индекса Дайса. Также видно, что некоторые квантили выходят за пределы границ доверительной зоны. Тем самым можно предварительно говорить об отличии данных распределений от нормального.

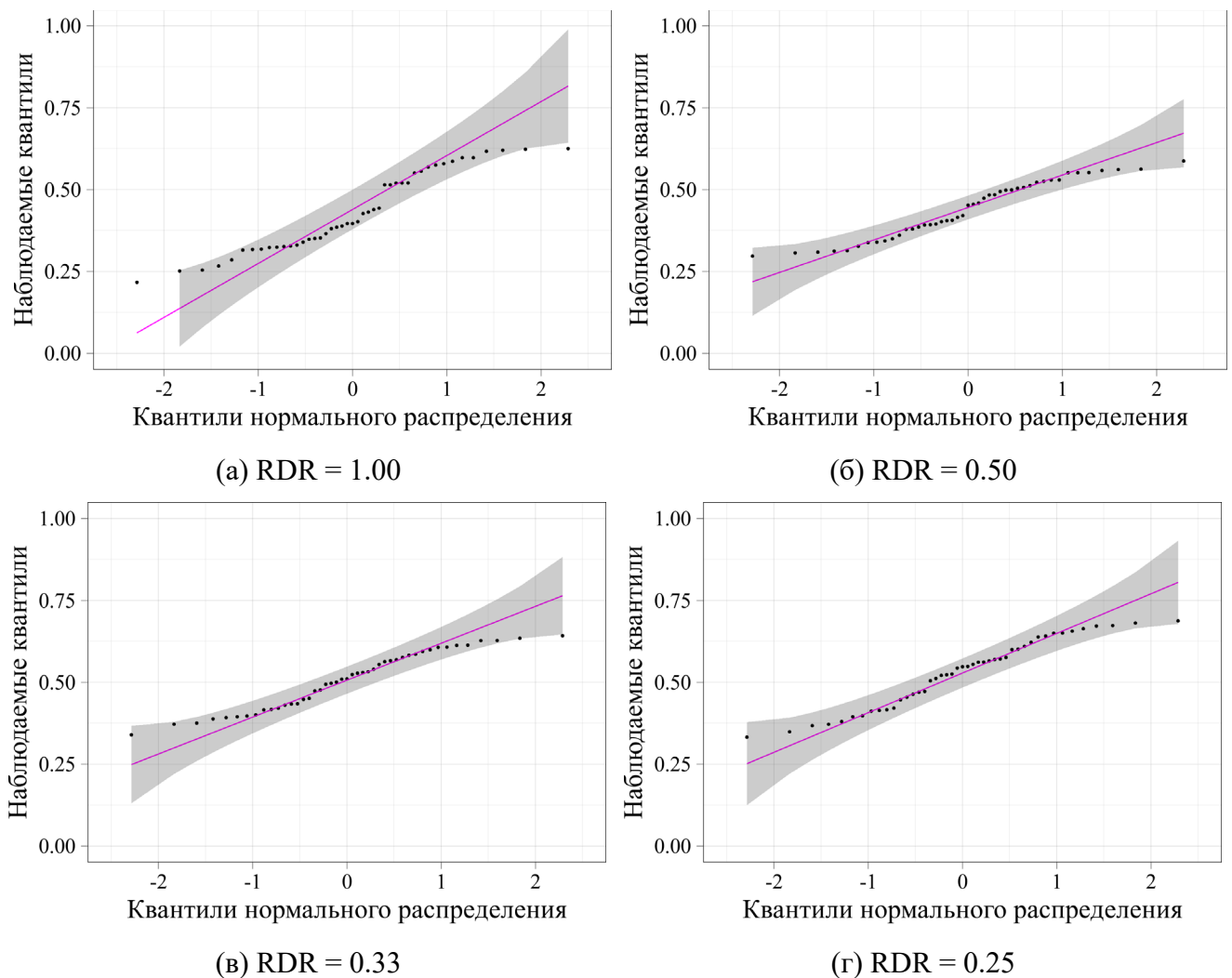


Рисунок 3.7 – Квантильный график стандартной U-net архитектуры при разных значениях метрики RDR

С целью формальной оценки типа распределения и значимости различий были выполнены два статистических теста, а именно тест Шапиро-Уилка и тест Вилкоксона. Важным будет заметить, что формальные тесты на нормальность чувствительны к размеру выборки. Чаще всего небольшие по размеру выборки проходят тесты на нормальность. Несмотря на то, что квантильные графики дали примерную визуальную картину о типе исследуемых распределений, тест Шапиро-Уилка проводился с целью формального подтверждения или отрицания факта нормальности распределения данных. Нулевая гипотеза H_0 данного теста предполагает, что исследуемая выборка распределена нормально. Согласно полученным результатам, отражённым в таблице 3.4, все выборки, за исключением выборки индекса Дайса при $RDR = 0.25$, оказались распределёнными ненормально при выбранном уровне значимости равном 0.05.

После определения типа распределения был выполнен тест Вилкоксона. Сравнение выборок при помощи критерия Вилкоксона основывается на нахождении разницы между всеми имеющимися парными выборочными

Таблица 3.4 – Результаты статистической оценки стандартной U-net архитектуры

Статистический параметр	RDR			
	0.25	0.33	0.50	1.00
Критерий Шапиро-Уилка	0.951	0.947	0.940	0.928
р-значение	5.7E-02	3.8E-02	2.2E-02	7.9E-03
Нормальное распределение?	Да	Нет	Нет	Нет
W-критерий Вилкоксона	230	230	483	—
р-значение	8.6E-04	8.6E-04	7.0E-01	—
Значимы ли различия?	Да	Да	Нет	—
Доверительный интервал (мин.)	-0.155	-0.121	-0.066	—
Доверительный интервал (макс.)	-0.042	-0.034	0.039	—
Псевдомедиана	-0.102	-0.078	-0.010	—

наблюдениями с целью проверки нулевой гипотезы H_0 , которая заключается в том, что медиана полученных разностей равна нулю. Дополнительно в таблице 3.4 отражены минимальная и максимальная границы доверительного интервала, а также медиана разницы между парными выборочными наблюдениями (псевдомедиана). Как отмечалось ранее, базовой выборкой для сравнения является выборка при $RDR = 1.00$. На основании полученных результатов видно, что выборка при $RDR = 0.50$ не отличается значимо от базовой выборки. Последнее означает, что обучение модели на основе стандартной U-net архитектуры на данных, смешанных в соотношении 1:1 ($RDR = 0.50$), не приведёт к статистически значимому повышению точности. С другой стороны, использование обучающей выборки, смешанной в соотношении 1:2 ($RDR = 0.33$) или 1:3 ($RDR = 0.25$), приводит к статистически значимому повышению точности сегментации.

3.1.3 Оценка результатов сегментации разреженной U-net модели

Аналогичный подход по оценке статистических различий использовался для выборок разреженной U-net архитектуры. В целях первичной оценки значимости различий моделей при разных значениях метрики RDR использовался дисперсионный анализ (метод ANOVA). Как отмечалось выше, данный метод определяет наблюдаются ли значимые отличия в исследуемых группах или нет. Нулевая гипотеза H_0 дисперсионного анализа указывает на отсутствие отличий между группами. В таком случае подразумевается, что все группы по индексу Дайса относятся к одной генеральной совокупности и соответственно их средние равны друг другу, то есть $\mu_1 = \mu_2 = \mu_3 = \mu_4$. В свою очередь, альтернативная гипотеза H_1 выдвигает предположение о том, что средние в исследуемых группах не равны.

После проведения дисперсионного анализа были получены результаты, которые отражены в таблице 3.5. Как видно, значение F больше критического значения $F_{кр}$ при выбранном уровне значимости равном 0.05 для каждой подвыборки. В связи с этим есть все основания, чтобы отвергнуть нулевую гипотезу и констатировать, что, по крайней мере, одна из четырёх выборок имеет

существенно различающееся среднее значение и, таким образом, принадлежит к совершенно иной генеральной совокупности. Стоит отметить, что вычисленные значения p также меньше уровня значимости, что говорит о том, что нулевая гипотеза может быть отклонена.

При визуальном инспектировании видно, что распределения четырёх рассматриваемых выборок сильно отличаются друг от друга. В данном случае две из четырёх выборок имеют выраженную бимодальную природу, а именно распределения индекса Дайса при $RDR = 0.33$ и $RDR = 1.00$. С другой стороны выборки при $RDR = 0.25$ и $RDR = 0.50$ по характеру распределения близки к нормальному. Однако распределение индекса Дайса при $RDR = 0.50$ характеризуется отрицательной (правосторонней) асимметрией. Распределения индекса Дайса разреженной U-net архитектуры при разных значениях метрики RDR отражены на рисунке 3.8. Дополнительно на этих рисунках отмечены средние значения распределений (штриховые линии), а также стандартные отклонения от среднего в положительную и отрицательную сторону (пунктирные линии).

В связи с тем, что проводится оценка зависимости индекса Дайса от метрики RDR, то есть от степени смешивания данных, ниже приводятся сравнения распределений индекса Дайса. В качестве базового распределения для сравнения используется распределение при $RDR = 1.00$, то есть случай обучения и тестирования разреженной U-net модели исключительно на реальных данных. Остальные распределения, полученные при $RDR = 0.50$, $RDR = 0.33$ и $RDR = 0.25$, сравниваются с базовым. Данное сравнение отражено на рисунке 3.9. Аналогично случаю со стандартной U-net моделью, среднее значение точности сегментации на основе разреженной U-net модели увеличивается при уменьшении метрики RDR.

Таблица 3.5 – Результаты ANOVA тестирования разреженной U-net архитектуры

Параметр	Обучение	Валидация	Тестирование
F	127.6	22.6	31.3
F _{кр}	2.6	2.7	2.7
p	3.4E-66	1.9E-12	3.0E-16
Значимы ли различия?	Да	Да	Да

Однако для случая стандартной U-net модели наблюдалась большая прибавка в точности сегментации при смешивании данных. Тем не менее среднее значение точности сегментации разреженной U-net модели значительно выше и находится на уровне 90%. Также стоит отметить, что в данном случае не наблюдается ярко выраженное асимптотическое насыщение, как это было в случае стандартной U-net модели. Поэтому смешивание данных, например в соотношении 1:4, может предположительно привести либо к увеличению точности, либо к достижению её асимптоты.

Перед проверкой формальных статистических гипотез была проведена оценка типа распределения. Визуальная проверка на нормальность каждого из распределений разреженной U-net архитектуры выполнялась с помощью квантильных графиков, которые отражены на рисунке 3.10. В связи с тем, что двум

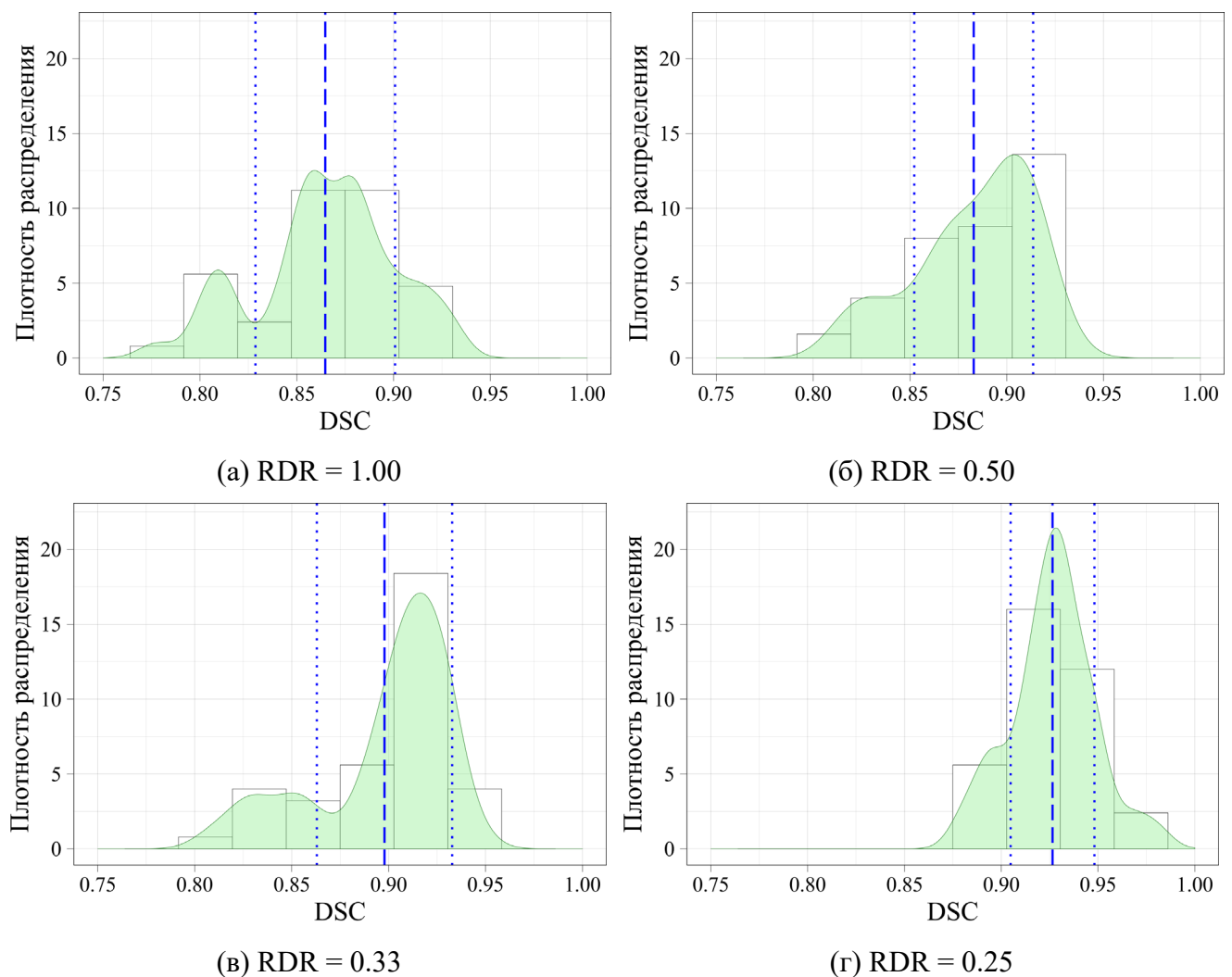
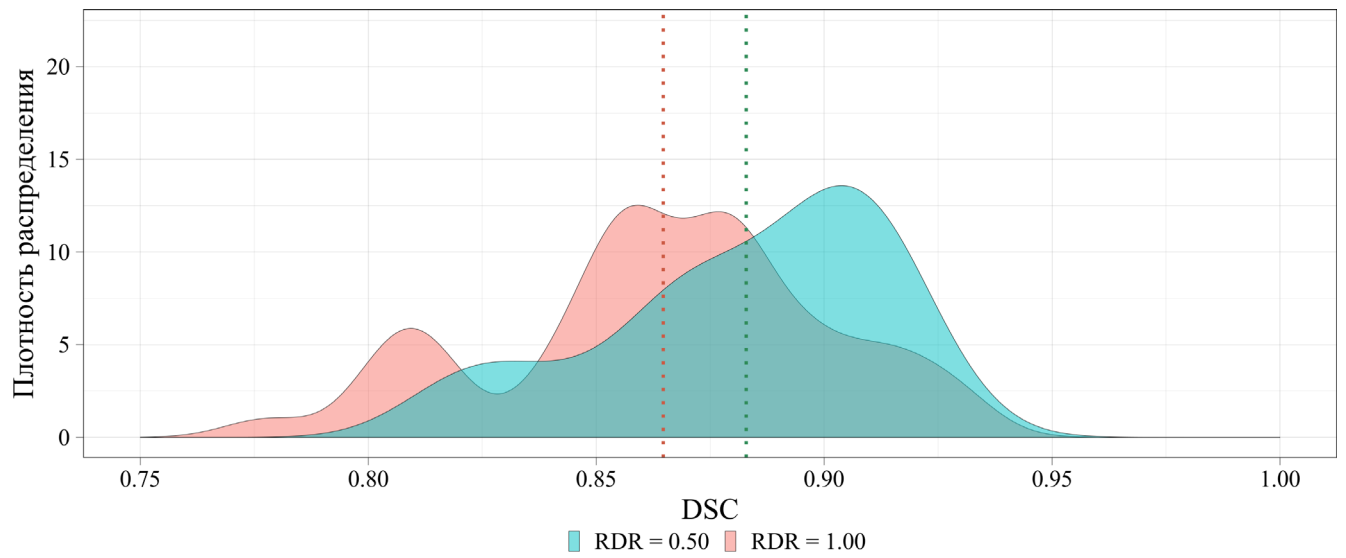
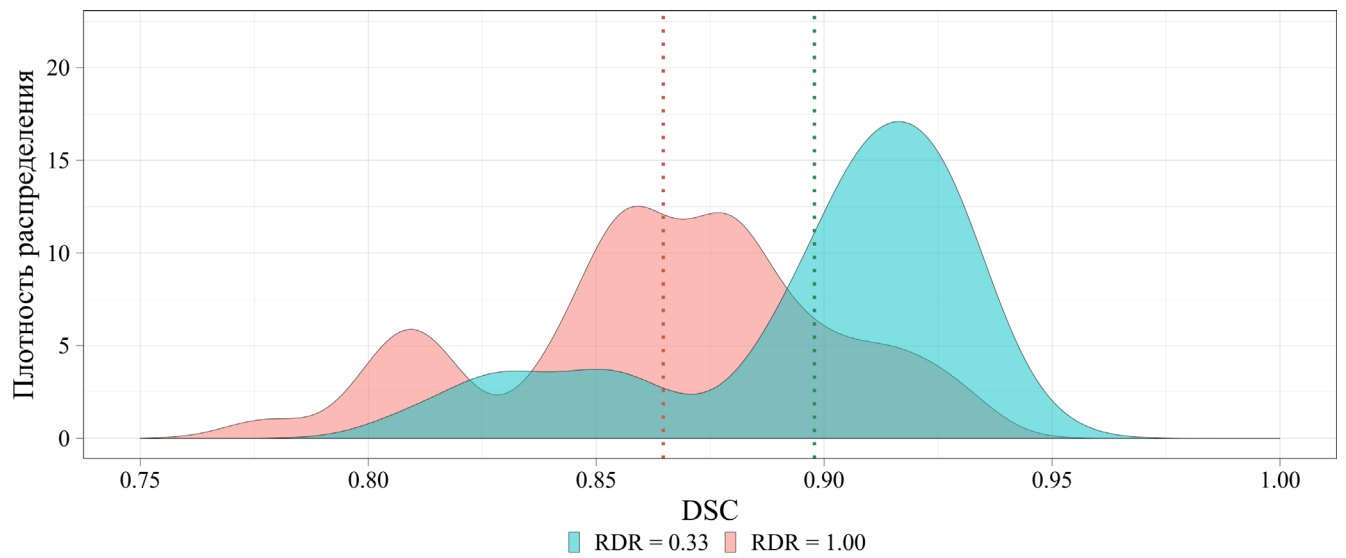


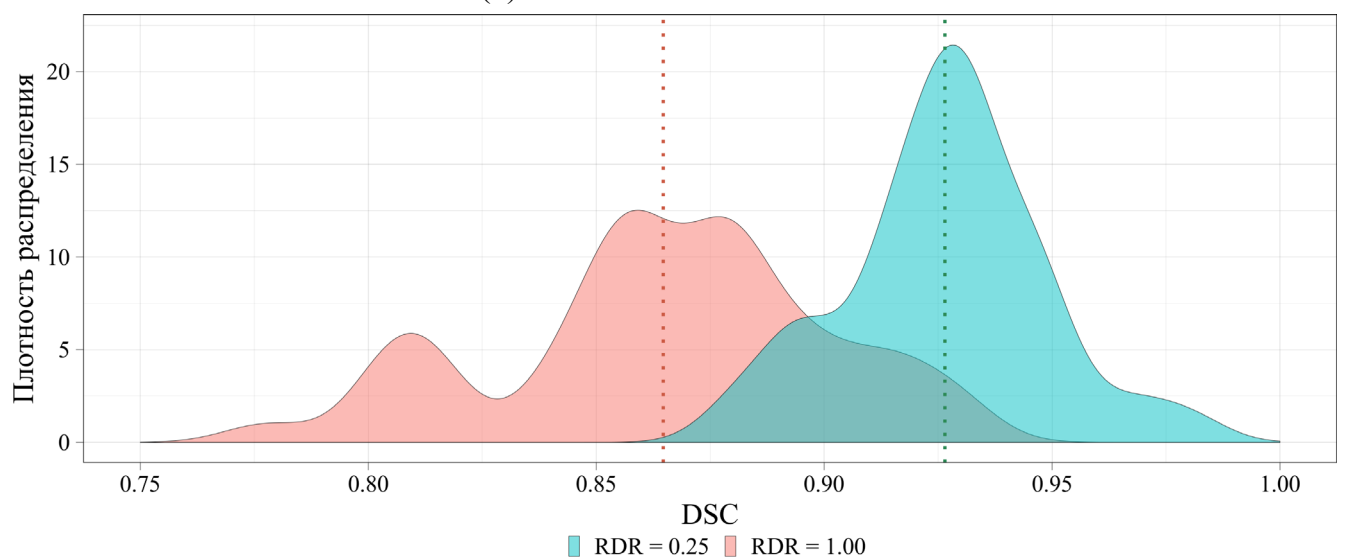
Рисунок 3.8 – Распределение индекса Дайса разреженной U-net архитектуры при разных значениях метрики RDR



(a) RDR = 0.50 и RDR = 1.00



(б) RDR = 0.33 и RDR = 1.00



(в) RDR = 0.25 и RDR = 1.00

Рисунок 3.9 – Сравнение распределений индекса Дайса разреженной U-net архитектуры при разных значениях метрики RDR

выборкам из четырёх ($RDR = 0.33$ и $RDR = 1.00$) присущ характер бимодального поведения, исследуемые квантильные распределения S-образно изогнуты, а точка их перегиба наблюдается в области близкой к нулевому квантилю нормального распределения. В дополнение стоит отметить значительные отклонения исследуемых квантилей от нормальных при низких и высоких значениях индекса Дайса. Ряд рассматриваемых квантилей значительно выходит за пределы границ доверительной зоны. Особенно это заметно для распределения индекса Дайса при $RDR = 0.33$ и $RDR = 1.00$. По визуальной оценке, часть выборок индекса Дайса для разреженной U-net архитектуры предварительно может иметь природу распределения данных близкую к нормальной.

Аналогично случаю стандартной U-net архитектуры, для выборок индекса Дайса разреженной U-net архитектуры были выполнены два формальных

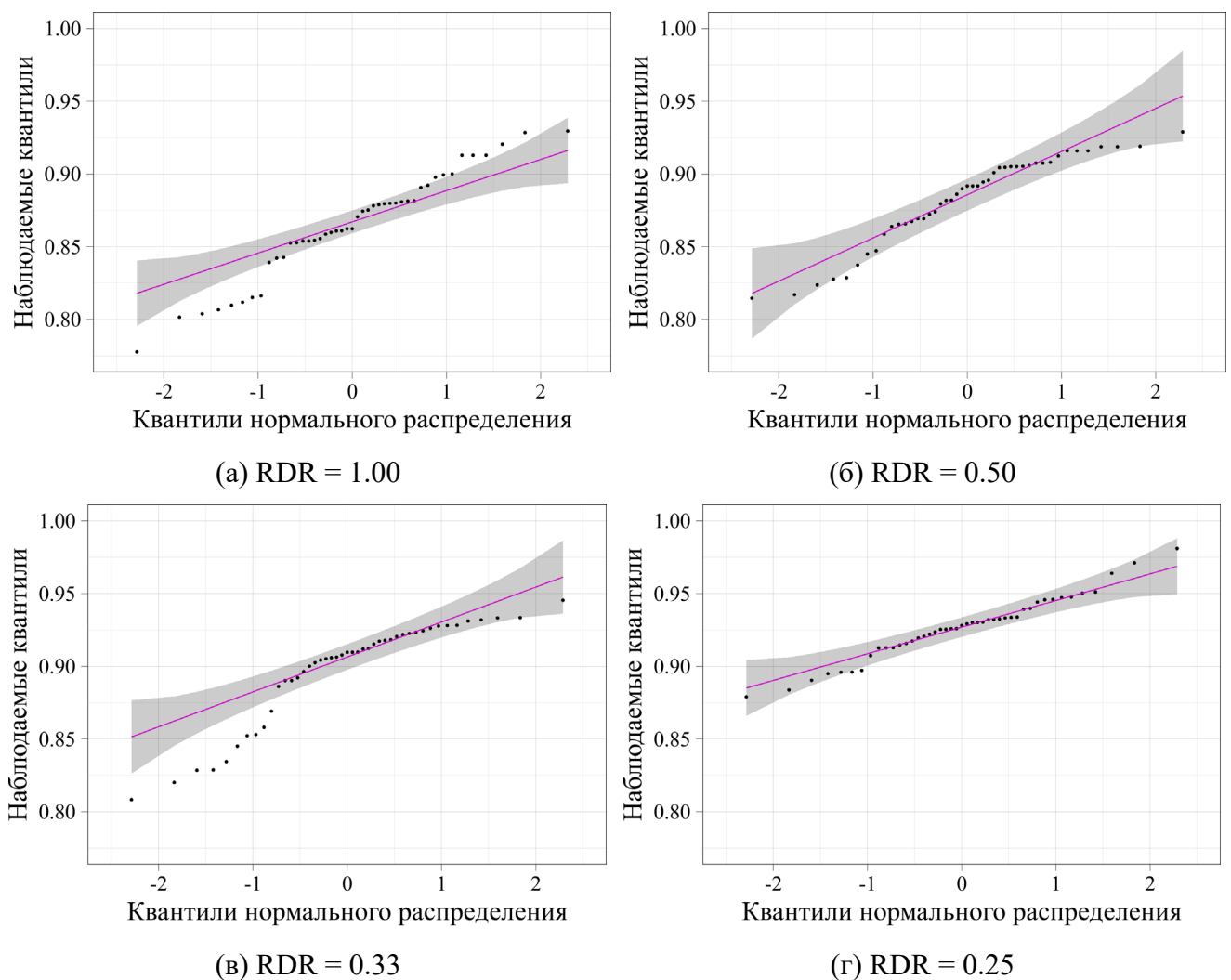


Рисунок 3.10 – Квантильный график разреженной U-net архитектуры при разных значениях метрики RDR

статистических теста: тест Шапиро-Уилка и тест Вилкоксона. Несмотря на то, что квантильные графики дали примерную визуальную картину о типе исследуемых распределений, тест Шапиро-Уилка проводился с целью формального подтверждения или отрицания факта нормальности распределения данных. Нулевая гипотеза H_0 данного теста предполагает, что исследуемая выборка распределена нормально. Согласно полученным результатам, отражённым в таблице 3.6, данные выборок индекса Дайса при $RDR = 0.33$ и $RDR = 0.50$ оказались распределёнными ненормально при выбранном уровне значимости равном 0.05.

После определения типа распределения был выполнен тест Вилкоксона. Сравнение выборок при помощи критерия Вилкоксона основано на нахождении разницы между всеми имеющимися парными выборочными наблюдениями с целью проверки нулевой гипотезы H_0 , которая заключается в том, что медиана полученных разностей равна нулю. Дополнительно в таблице 3.6 отражены минимальная и максимальная границы доверительного интервала, а также медиана

Таблица 3.6 – Результаты статистической оценки разреженной U-net архитектуры

Статистический параметр	RDR			
	0.25	0.33	0.50	1.00
Критерий Шапиро-Уилка	0.979	0.861	0.925	0.968
р-значение	6.0E-01	7.2E-05	6.0E-03	2.4E-01
Нормальное распределение?	Да	Нет	Нет	Да
W-критерий Вилкоксона	4	124	220	—
р-значение	4.0E-13	1.8E-06	5.4E-04	—
Значимы ли различия?	Да	Да	Да	—
Доверительный интервал (мин.)	-0.074	-0.048	-0.030	—
Доверительный интервал (макс.)	-0.049	-0.026	-0.011	—
Псевдомедиана	-0.062	-0.038	-0.021	—

разницы между парными выборочными наблюдениями (псевдомедиана). На основании полученных результатов видно, что все выборки значимо отличаются от базовой выборки. Последнее означает, что обучение модели на основе разреженной U-net архитектуры при смешивании реальных и искусственных данных в соотношении 1:1, 1:2 и 1:3 приведёт к статистически значимому повышению обобщающей способности модели и, как следствие, точности сегментации.

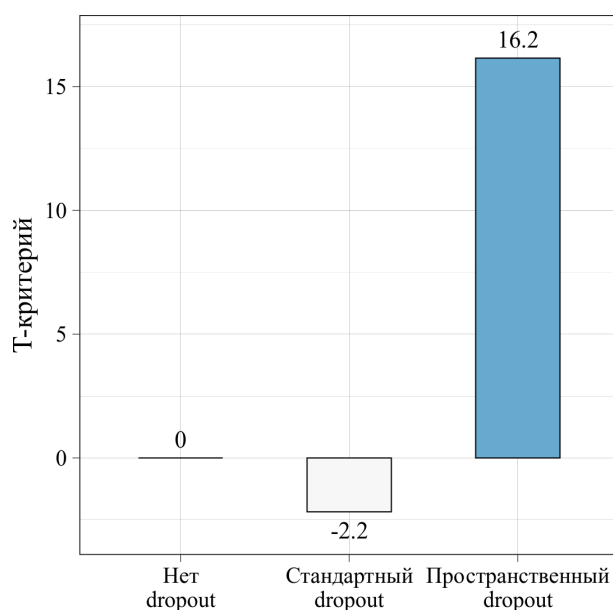
3.2 Выбор оптимальных гиперпараметров

В данном пункте освещена оценка гиперпараметров модели на основе подхода, описанного в пункте 2.6. Для оптимизации модели использовались 11 гиперпараметров, каждый из которых варьируется в определённом диапазоне. Всего было 69 значений для выбранных гиперпараметров. Для выбора оптимального гиперпараметра на основе Т-критерия использовалась кросс-валидация, в результате чего было обучено 690 моделей. Также стоит отметить, что базовая модель, на основе которой происходило сравнение, имела следующие характеристики: нет dropout слоя, нет нормализации, тип оптимизации – SGD, нет регуляризации, размер батча – 1, тип активации – ReLU, скорость обучения – 0.001, тип скорости обучения – ConLR, тип инициализации – Ones, дилатация – 1, коэффициент свёрточных фильтров – 1. Для оценки оптимальности параметра используется значение Т-критерия. Нулевая гипотеза H_0 представленного подхода заключается в том, что и базовая и сравниваемая модели имеют одинаковую точность. Соответственно, чем больше значение Т-критерия отклоняется от нуля, тем меньше вероятность выполнения нулевой гипотезы и случайного совпадения средних в числителе дробей (2.35). Далее в данном пункте в качестве описательных статистик индекса Дайса будут использоваться следующие сокращения: μ (среднее значение), σ (стандартное отклонение), Me (медиана) и IQR (интерквартильный размах).

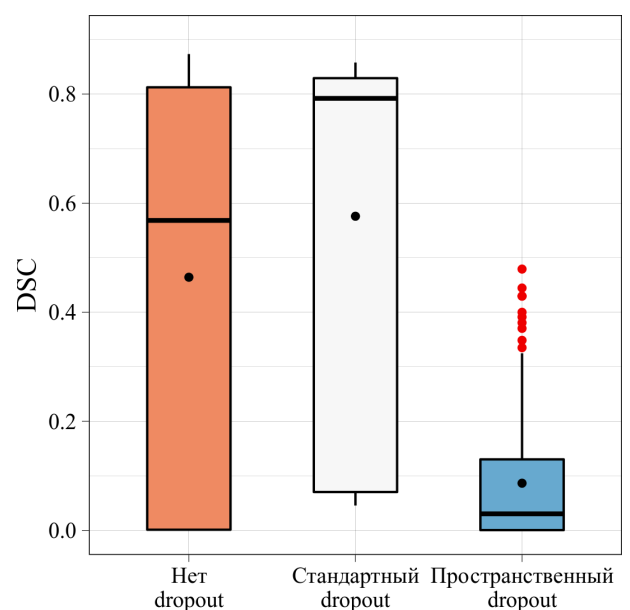
Процедура выбора оптимальных значений гиперпараметров начиналась с оценки типа dropout слоя, результаты которого отражены в таблице 3.7 и на рисунке 3.11. Согласно результатам, оптимальным является стандартный dropout слой, так как значение его Т-критерия является минимальным. Тем самым точность модели выше при использовании стандартной версии dropout слоёв. Последнее объясняется тем, что при обучении пространственный dropout слой обнуляет не нейрон в отдельности, а карту признаков в целом. Такое обнуление сильнее снижает мощность модели и, соответственно, ослабляет обобщающую способность модели. С другой стороны, отсутствие dropout слоя приводит к нестабильной работе модели. Несмотря на то, что для такого случая среднее значение и медиана индекса Дайса остаются на среднем уровне, разброс по точности остаётся высоким, что приводит к слабой обобщающей способности. Как результат, для оптимальной модели был выбран стандартный dropout слой.

Таблица 3.7 – Оценка типов dropout слоя

Значение гиперпараметра	$\mu \pm \sigma$	Индекс Дайса	
		Me	IQR
Нет dropout	0.46 ± 0.35	0.57	0.81
Стандартный dropout	0.58 ± 0.34	0.79	0.76
Пространственный dropout	0.09 ± 0.12	0.03	0.13



(а) Среднее значение Т-критерия



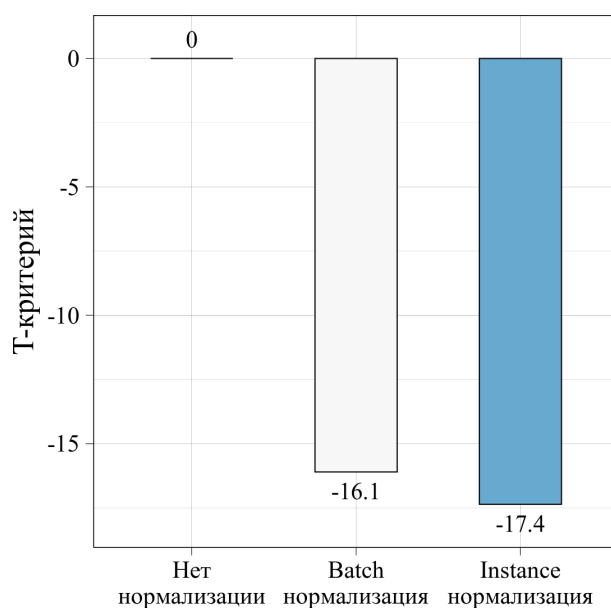
(б) Распределение индекса Дайса

Рисунок 3.11 – Оценка типов dropout слоя

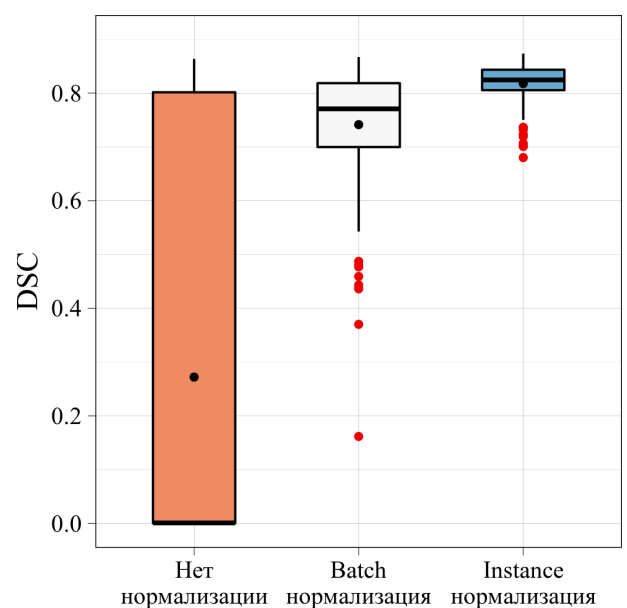
Следующим оптимизируемым гиперпараметром после типа dropout слоя являлся метод нормализации. Результаты оценки метода нормализации отражены в таблице 3.8 и на рисунке 3.12. При сравнении значений Т-критерия различных типов нормализации, оптимальной является instance нормализация. Несмотря на близкие значения индекса Дайса для instance и batch нормализаций, их значения Т-критерия отличаются практически вдвое, что приводит к выводу о лучшей обобщающей способности модели на основе instance нормализации. Стоит также отметить, что при отсутствии какой-либо нормализации, модель слабо и не точно предсказывает новые данные, что отражено большими значениями интерквартильного размаха и стандартного отклонения. Другим недостатком при обучении модели без нормализации являлось значительно большее время обучения. В качестве оптимального метода нормализации выбрана instance нормализация.

Таблица 3.8 – Оценка методов нормализации

Значение гиперпараметра	$\mu \pm \sigma$	Индекс Дайса	
		Me	IQR
Нет нормализации	0.25 ± 0.39	0.00	0.80
Batch нормализация	0.74 ± 0.11	0.77	0.12
Instance нормализация	0.82 ± 0.04	0.82	0.04



(а) Среднее значение Т-критерия



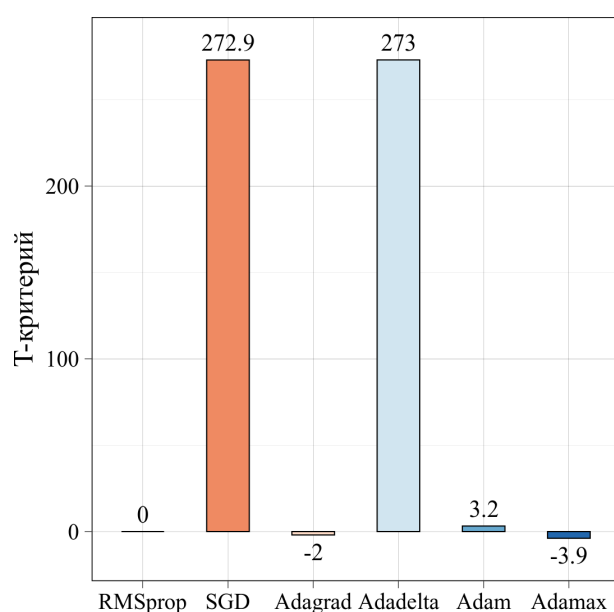
(б) Распределение индекса Дайса

Рисунок 3.12 – Оценка методов нормализации

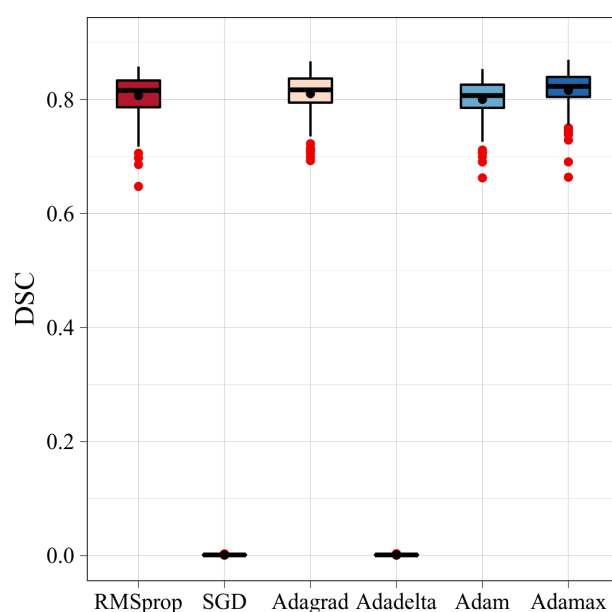
Одним из важных гиперпараметров для нейронных сетей является метод оптимизации. Результаты оценки методов оптимизации отражены в таблице 3.9 и на рисунке 3.13. При сравнении значений Т-критерия методы RMSProp, Adagrad, Adam и Adamax находятся примерно на одном уровне по эффективности. В свою очередь, модели на основе SGD и Adadelata не смогли сойтись при обучении, что привело к нулевой точности при сегментации. Предположительно, это произошло из-за сложного ландшафта функции потерь, застревания в локальных минимумах или седловых точках, или инициализации весов на плато, где градиенты приблизительно равны нулю. В результате оптимальным методом для оптимизации весов был выбран метод Adamax.

Таблица 3.9 – Оценка методов оптимизации

Значение гиперпараметра	$\mu \pm \sigma$	Индекс Дайса	
		Me	IQR
RMSprop	0.81 ± 0.04	0.82	0.05
SGD	0.00 ± 0.00	0.00	0.00
Adagrad	0.81 ± 0.04	0.82	0.04
Adadelata	0.00 ± 0.00	0.00	0.00
Adam	0.80 ± 0.04	0.81	0.04
Adamax	0.82 ± 0.03	0.82	0.04



(а) Среднее значение Т-критерия



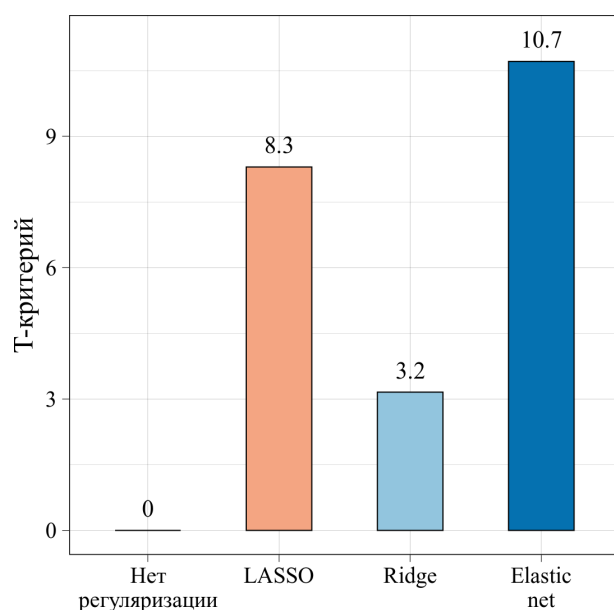
(б) Распределение индекса Дайса

Рисунок 3.13 – Оценка методов оптимизации

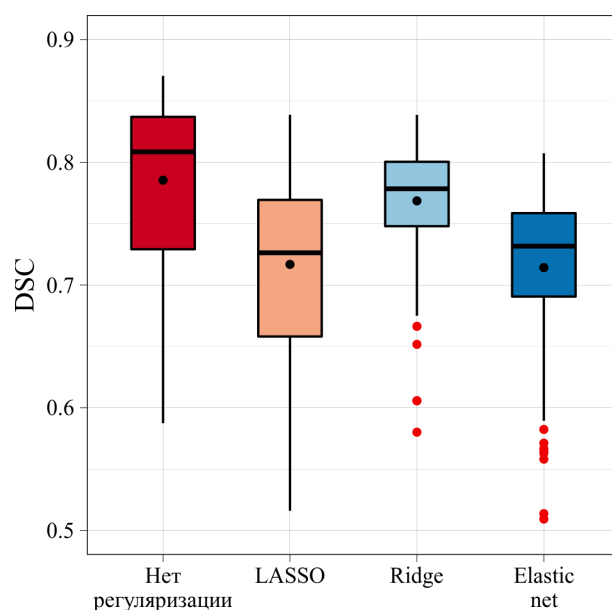
В целях потенциального уменьшения сложности модели и предотвращения переобучения исследовались методы регуляризации. Результаты оценки методов регуляризации отражены в таблице 3.10 и на рисунке 3.14. Оптимальной показала себя модель без использования какой-либо регуляризации. Несмотря на то, что штраф за большие коэффициенты был небольшой ($\lambda = 0.005$), модели с регуляризацией показали себя хуже. Среди моделей с регуляризацией оптимальной является Ridge регуляризация (L_2 -регуляризация). В свою очередь, комбинация L_1 -регуляризации и L_2 -регуляризации (Elastic net регуляризация) оказалась худшей для данной задачи. Важно отметить, что использование dropout и стандартной регуляризации (LASSO, Ridge и Elastic net) может значительно ослабить мощность и снизить точность модели. В результате, оптимальная модель для решаемой задачи обладает исключительно dropout регуляризацией.

Таблица 3.10 – Оценка методов регуляризации

Значение гиперпараметра	Индекс Дайса		
	$\mu \pm \sigma$	Me	IQR
Нет регуляризации	0.79 ± 0.06	0.81	0.11
LASSO регуляризация	0.72 ± 0.07	0.73	0.11
Ridge регуляризация	0.77 ± 0.05	0.78	0.05
Elastic net регуляризация	0.71 ± 0.07	0.73	0.07



(а) Среднее значение Т-критерия



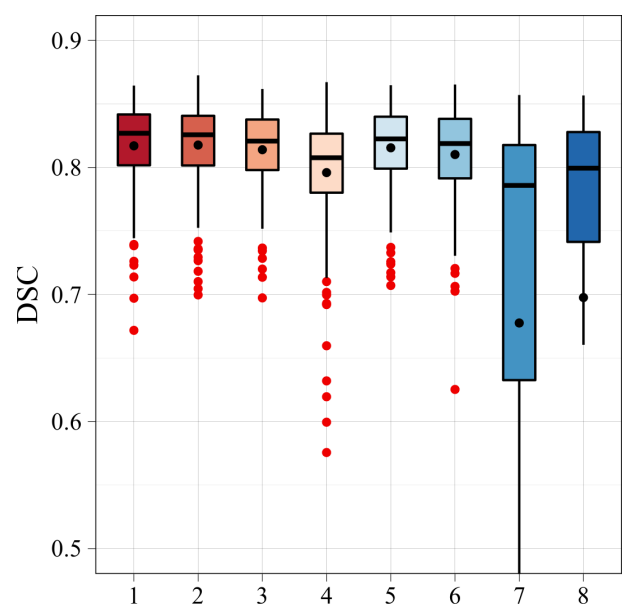
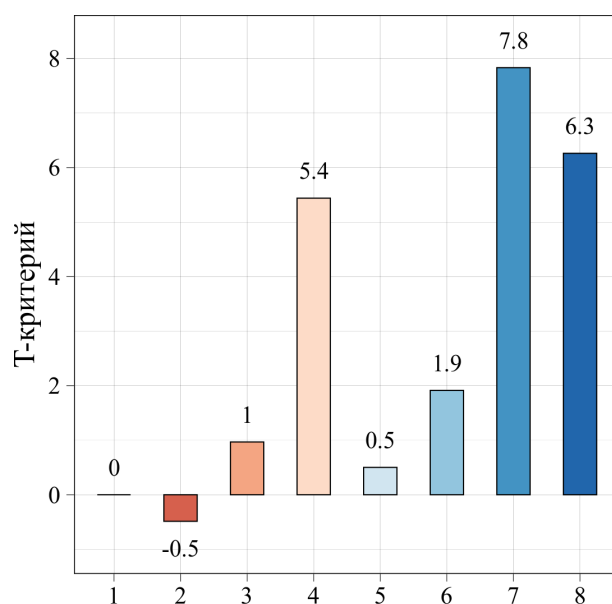
(б) Распределение индекса Дайса

Рисунок 3.14 – Оценка методов регуляризации

Для расчёта компромисса между эффективностью и скоростью обучения нейронной сети исследовались разные размеры батча. Результаты оценки размера батча отражены в таблице 3.11 и на рисунке 3.15. На основании результатов видно, что значения Т-критерия для моделей, обученных на батче размером 1, 2 и 3, незначительно отличаются друг от друга. В связи с этим, при обучении с использованием статического размера батча эффективнее выбирать его значение в диапазоне от 1 до 3. Оптимальным выбран батч размером 2, который обеспечивает компромисс между скоростью обучения и точностью сегментации.

Таблица 3.11 – Оценка размера батча

Значение гиперпараметра	Индекс Дайса		
	$\mu \pm \sigma$	Me	IQR
1	0.82 ± 0.03	0.83	0.04
2	0.82 ± 0.04	0.83	0.04
3	0.81 ± 0.03	0.82	0.04
4	0.80 ± 0.05	0.81	0.05
5	0.82 ± 0.03	0.82	0.04
6	0.81 ± 0.04	0.82	0.05
7	0.68 ± 0.21	0.79	0.19
8	0.70 ± 0.23	0.80	0.09



(а) Среднее значение Т-критерия

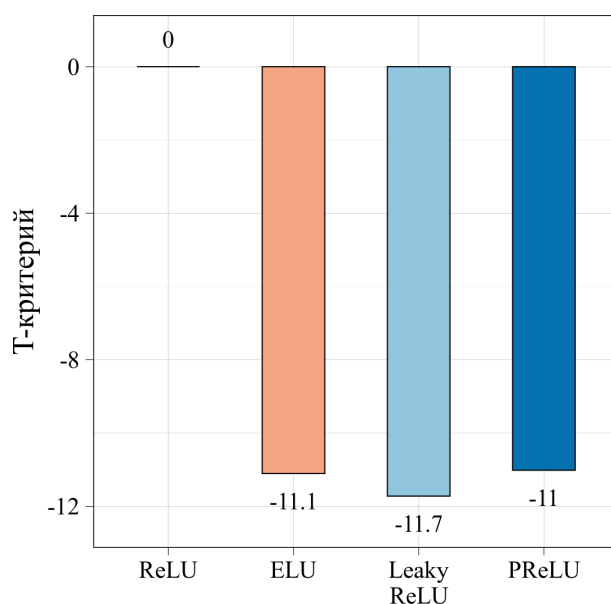
(б) Распределение индекса Дайса

Рисунок 3.15 – Оценка размера батча

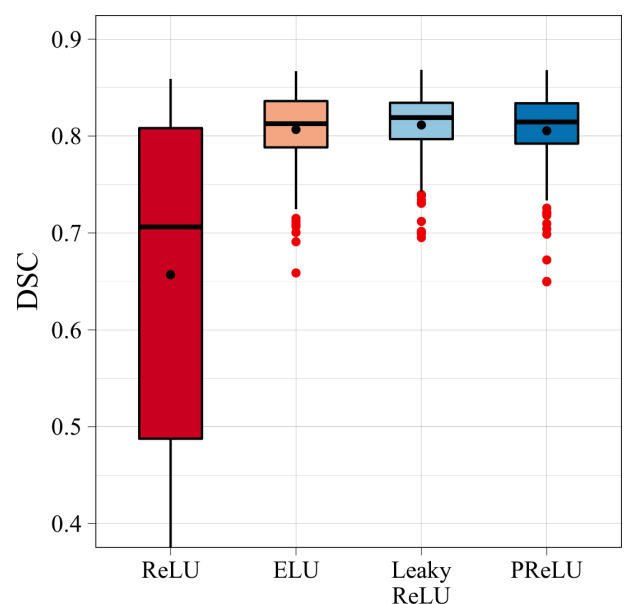
Для нелинейной аппроксимации искомой функции, имеющей сложную природу, использовались несколько популярных типов активации. Результаты оценки типов активации отражены в таблице 3.12 и на рисунке 3.16. Исходя из полученных результатов Т-критерия видно, что исследуемые функции активации находятся примерно на одном уровне, за исключением функции ReLU. Низкая эффективность этой функции активации, связана с тем, что градиент для отрицательных значений x равен нулю. Из-за равенства градиента нулю, веса модели не корректируются, а пребывающие в таком состоянии нейроны не реагируют на изменения в ошибке. С другой стороны, использование функции активации PReLU вызывает практически двукратное увеличение количества весов, причём точность данного типа активации слабо отличается от ELU или Leaky ReLU. В связи с вышесказанным, оптимальной выбрана функция активации ELU.

Таблица 3.12 – Оценка типов активации

Значение гиперпараметра	$\mu \pm \sigma$	Индекс Дайса	
		Me	IQR
ReLU	0.66 ± 0.16	0.71	0.32
ELU	0.81 ± 0.04	0.81	0.05
Leaky ReLU	0.81 ± 0.04	0.82	0.04
PReLU	0.81 ± 0.04	0.81	0.04



(а) Среднее значение Т-критерия



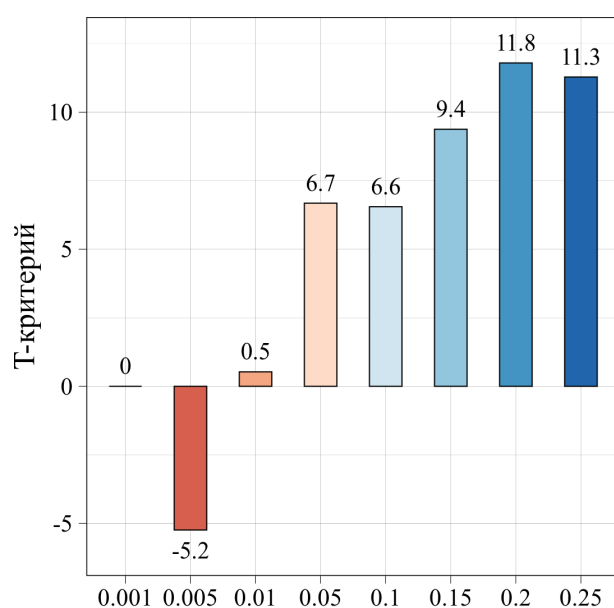
(б) Распределение индекса Дайса

Рисунок 3.16 – Оценка типов активации

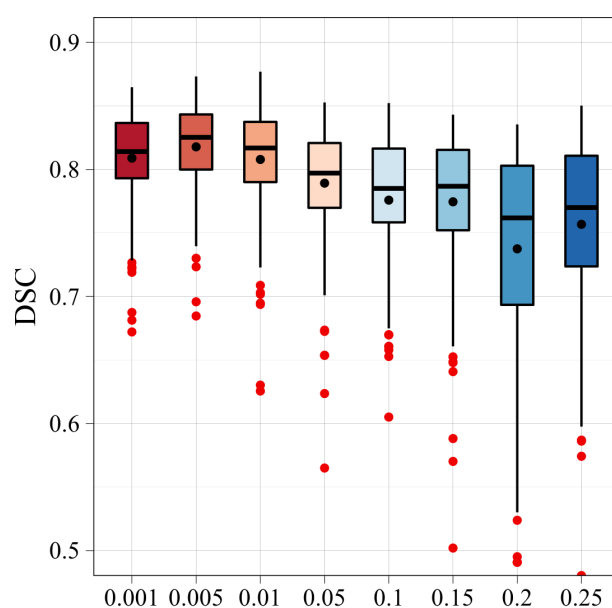
Одним из главных гиперпараметров при обучении нейронной сети является скорость обучения. Результаты оценки скорости обучения отражены в таблице 3.13 и на рисунке 3.17. На основании полученных значений Т-критерия, оптимальное значение скорости обучения для поиска глобального минимума составляет 0.005. Видно, что увеличение скорости обучения негативно влияет на минимизацию функции потерь, что приводит к снижению точности сегментации. При скорости обучения выше 0.20, обобщающая способность модели сильно снижается, что отражено большим стандартным отклонением и интерквартильным размахом.

Таблица 3.13 – Оценка значений скорости обучения

Значение гиперпараметра	Индекс Дайса		
	$\mu \pm \sigma$	Me	IQR
0.001	0.82±0.04	0.81	0.04
0.005	0.82±0.03	0.83	0.04
0.01	0.81±0.04	0.82	0.05
0.05	0.79±0.05	0.80	0.05
0.10	0.78±0.06	0.79	0.06
0.15	0.77±0.06	0.79	0.06
0.20	0.74±0.08	0.76	0.11
0.25	0.76±0.07	0.77	0.09



(а) Среднее значение Т-критерия



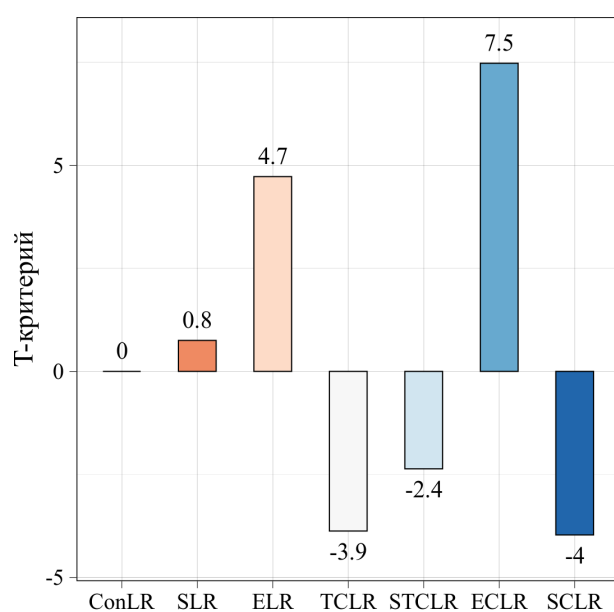
(б) Распределение индекса Дайса

Рисунок 3.17 – Оценка значений скорости обучения

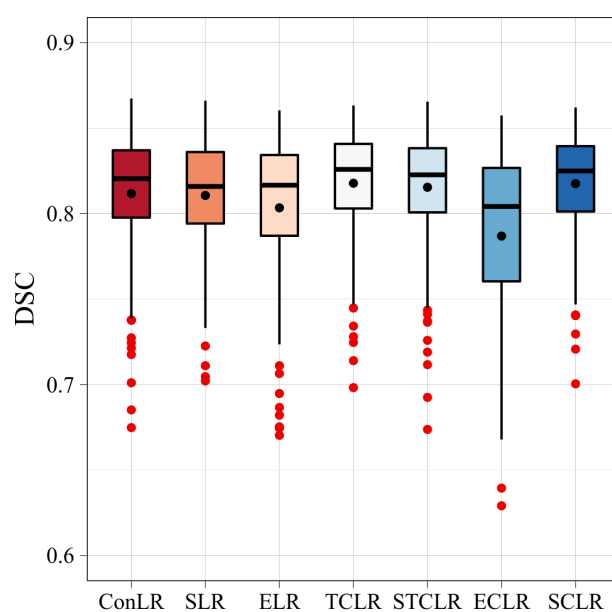
Следующим гиперпараметром для оптимизации является метод скорости обучения. Результаты оценки методов скорости обучения отражены в таблице 3.14 и на рисунке 3.18. Согласно полученным результатам, оцениваемые методы выполняют сегментацию примерно одинаково. Тем не менее, метод постоянной скорости обучения ConLR теоретически имеет высокую вероятность заикливания в локальных минимумах или на седловых точках. Частично, подобная проблема решается методами отжига SLR и ELR. Однако данные методы не всегда эффективно проходят область плато. В связи с вышесказанными доводами и полученными результатами, оптимальным выбран метод SCLR.

Таблица 3.14 – Оценка методов изменения скорости обучения

Значение гиперпараметра	Индекс Дайса		
	$\mu \pm \sigma$	Me	IQR
ConLR	0.81 ± 0.04	0.82	0.04
SLR	0.81 ± 0.03	0.82	0.04
ELR	0.80 ± 0.04	0.82	0.05
TCLR	0.82 ± 0.03	0.83	0.04
STCLR	0.82 ± 0.04	0.82	0.04
ECLR	0.79 ± 0.06	0.80	0.07
SCLR	0.82 ± 0.03	0.82	0.04



(а) Среднее значение Т-критерия



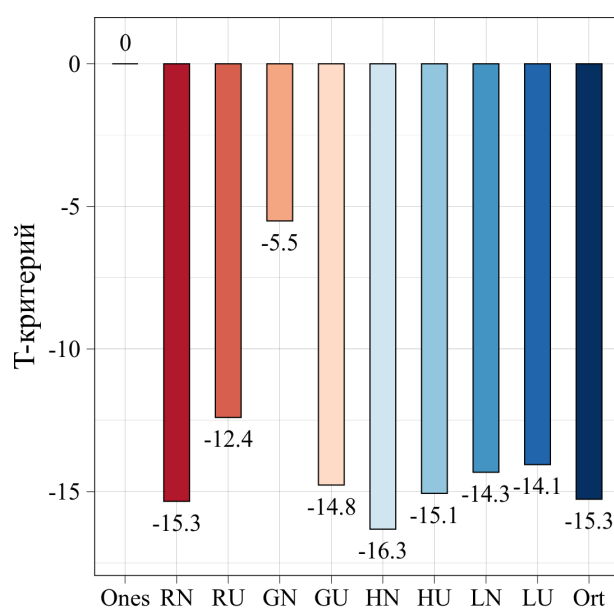
(б) Распределение индекса Дайса

Рисунок 3.18 – Оценка методов изменения скорости обучения

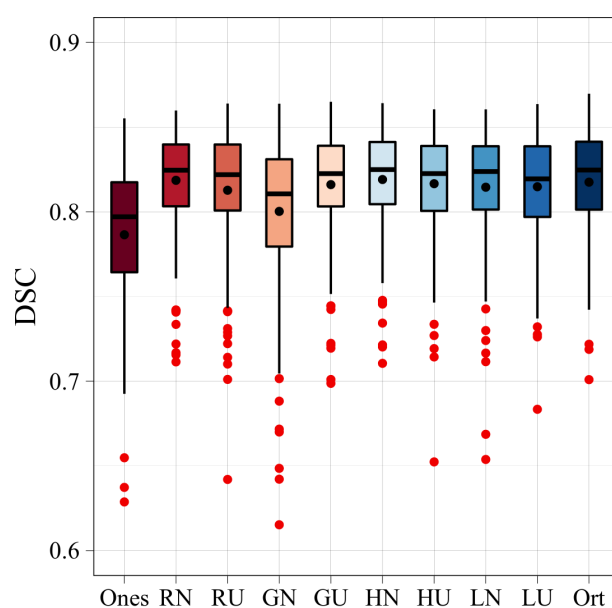
Одним из рассматриваемых гиперпараметров, слабо влияющих на обобщающую способность модели, является метод инициализации весов. Результаты оценки типа инициализации отражены в таблице 3.15 и на рисунке 3.19. Исходя из полученных результатов, можно сделать вывод о слабом влиянии типа инициализации на финальный результат модели. Практически все методы дают одинаковую точность, за исключением метода Ones. Как результат, в качестве оптимального метода инициализации выбран метод He normal.

Таблица 3.15 – Оценка методов инициализации

Значение гиперпараметра	Индекс Дайса		
	$\mu \pm \sigma$	Me	IQR
Ones	0.79±0.04	0.80	0.05
Random normal	0.82±0.03	0.82	0.04
Random uniform	0.81±0.04	0.82	0.04
Glorot normal	0.80±0.05	0.81	0.05
Glorot uniform	0.82±0.03	0.82	0.04
He normal	0.82±0.03	0.82	0.04
He uniform	0.82±0.03	0.82	0.04
Lecun normal	0.81±0.04	0.82	0.04
Lecun uniform	0.81±0.03	0.82	0.04
Orthogonal	0.82±0.03	0.82	0.04



(а) Среднее значение Т-критерия



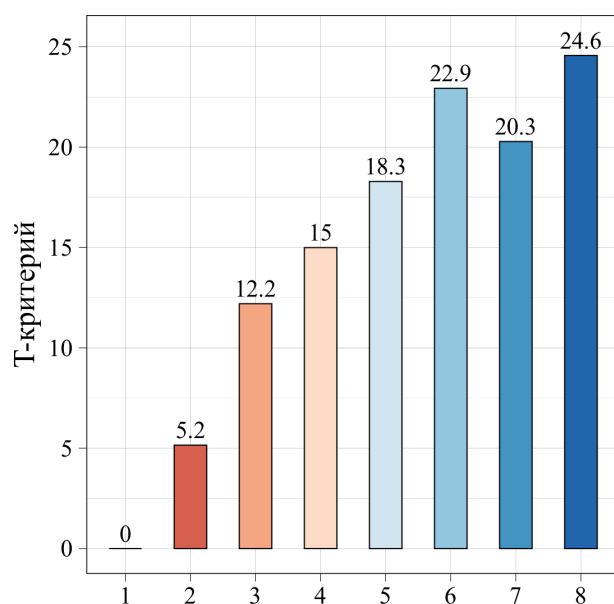
(б) Распределение индекса Дайса

Рисунок 3.19 – Оценка методов инициализации

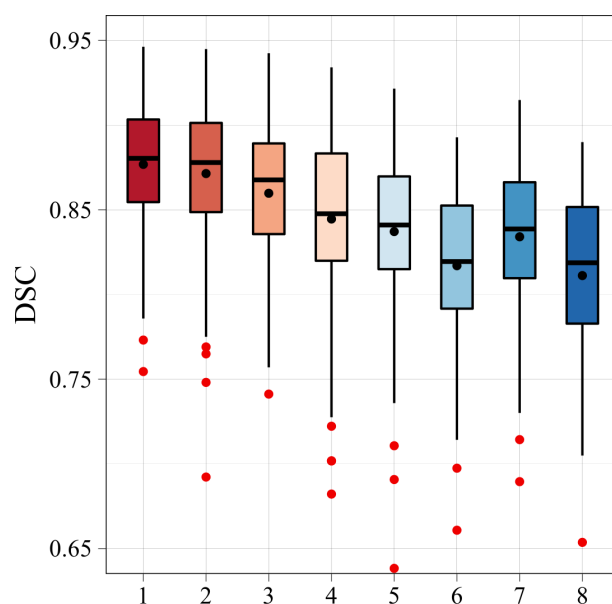
На предпоследнем этапе выбора гиперпараметров оценивалась дилатация. Результаты оценки значений дилатации отражены в таблице 3.16 и на рисунке 3.20. Исходя из полученных результатов наблюдается нисходящий тренд точности сегментации, что говорит об обратной зависимости значения дилатации и обобщающей способности модели. Наибольшая точность была достигнута при значении дилатации равном 1. При использовании значений дилатации равных 2 или 3 сохраняется компромисс между временем обучения и выходной точностью. Тем не менее, в качестве оптимального выбрано значение дилатации равное 1.

Таблица 3.16 – Оценка значений дилатации

Значение гиперпараметра	Индекс Дайса		
	$\mu \pm \sigma$	Me	IQR
1	0.88 $\pm$ 0.04	0.88	0.05
2	0.87 $\pm$ 0.04	0.88	0.05
3	0.86 $\pm$ 0.04	0.87	0.05
4	0.84 $\pm$ 0.05	0.85	0.06
5	0.84 $\pm$ 0.05	0.84	0.05
6	0.82 $\pm$ 0.05	0.82	0.06
7	0.83 $\pm$ 0.05	0.84	0.06
8	0.81 $\pm$ 0.05	0.82	0.07



(а) Среднее значение Т-критерия



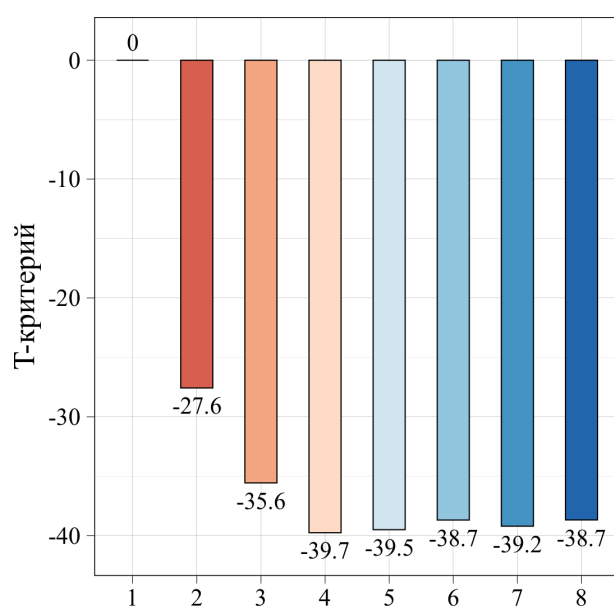
(б) Распределение индекса Дайса

Рисунок 3.20 – Оценка значений дилатации

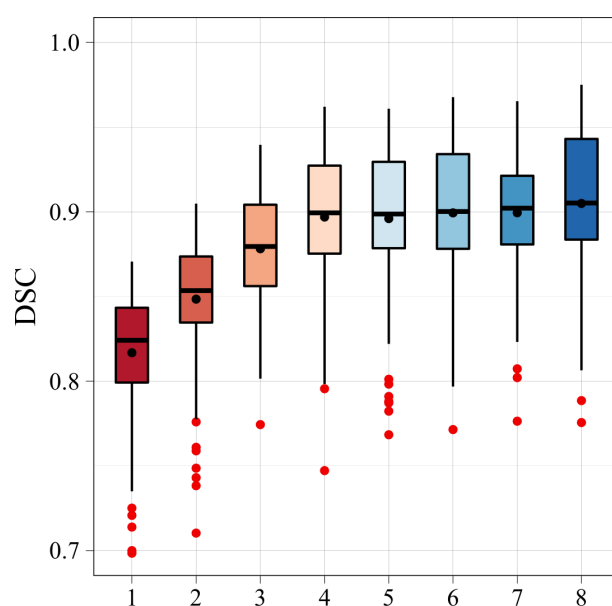
На заключительном этапе выбора оптимальных значений гиперпараметров оценивался коэффициент количества свёрточных фильтров. Результаты оценки коэффициента свёрточных фильтров отражены в таблице 3.17 и на рисунке 3.21. На основе полученных результатов видно, что точность модели асимптотически насыщается при значении коэффициента свёрточных фильтров большего 4. Дополнительно стоит отметить экспоненциальный рост количества параметров сети, что ведёт к значительному увеличению времени обучения модели. В связи с этим, коэффициент свёрточных фильтров равный 4 выбран как оптимальный.

Таблица 3.17 – Оценка значений коэффициента свёрточных фильтров

Значение гиперпараметра	Количество весов	Индекс Дайса		
		$\mu \pm \sigma$	Me	IQR
1	1 268 066	0.82±0.03	0.82	0.04
2	5 070 963	0.85±0.04	0.85	0.04
3	11 408 692	0.88±0.03	0.88	0.05
4	20 281 253	0.90±0.04	0.90	0.05
5	31 688 646	0.90±0.04	0.90	0.05
6	45 630 871	0.90±0.04	0.90	0.06
7	62 107 928	0.90±0.04	0.90	0.04
8	81 119 817	0.90±0.04	0.91	0.06



(а) Среднее значение Т-критерия



(б) Распределение индекса Дайса

Рисунок 3.21 – Оценка значений коэффициента свёрточных фильтров

В результате проведения процедуры оптимизации гиперпараметров были подобраны те значения гиперпараметров, которые эффективно выполняют поиск минимума функции потерь. Оптимальные значения для архитектуры сети и метода её обучения в дальнейшем использовались для выполнения сегментации не только медицинского инструмента, но и анатомических структур. Ниже в таблице 3.18 приведено сравнение базовой и оптимальной моделей, а пошаговая процедура выбора оптимальных значений гиперпараметров представлена на рисунке 3.22.

Таблица 3.18 – Сравнительная оценка базовой и оптимальной моделей

Гиперпараметр	Базовая модель	Оптимальная модель
Dropout	–	Стандартный
Метод нормализации	–	Instance
Метод оптимизации	SGD	Adam
Метод регуляризации	–	–
Размер батча	1	2
Тип активации	ReLU	ELU
Скорость обучения	0.001	0.005
Метод скорости обучения	ConLR	SCLR
Метод инициализации	Ones	He normal
Дилатация	1	1
Коэффициент свёрточных фильтров	1	4
Сложность	1 268 066	20 281 253

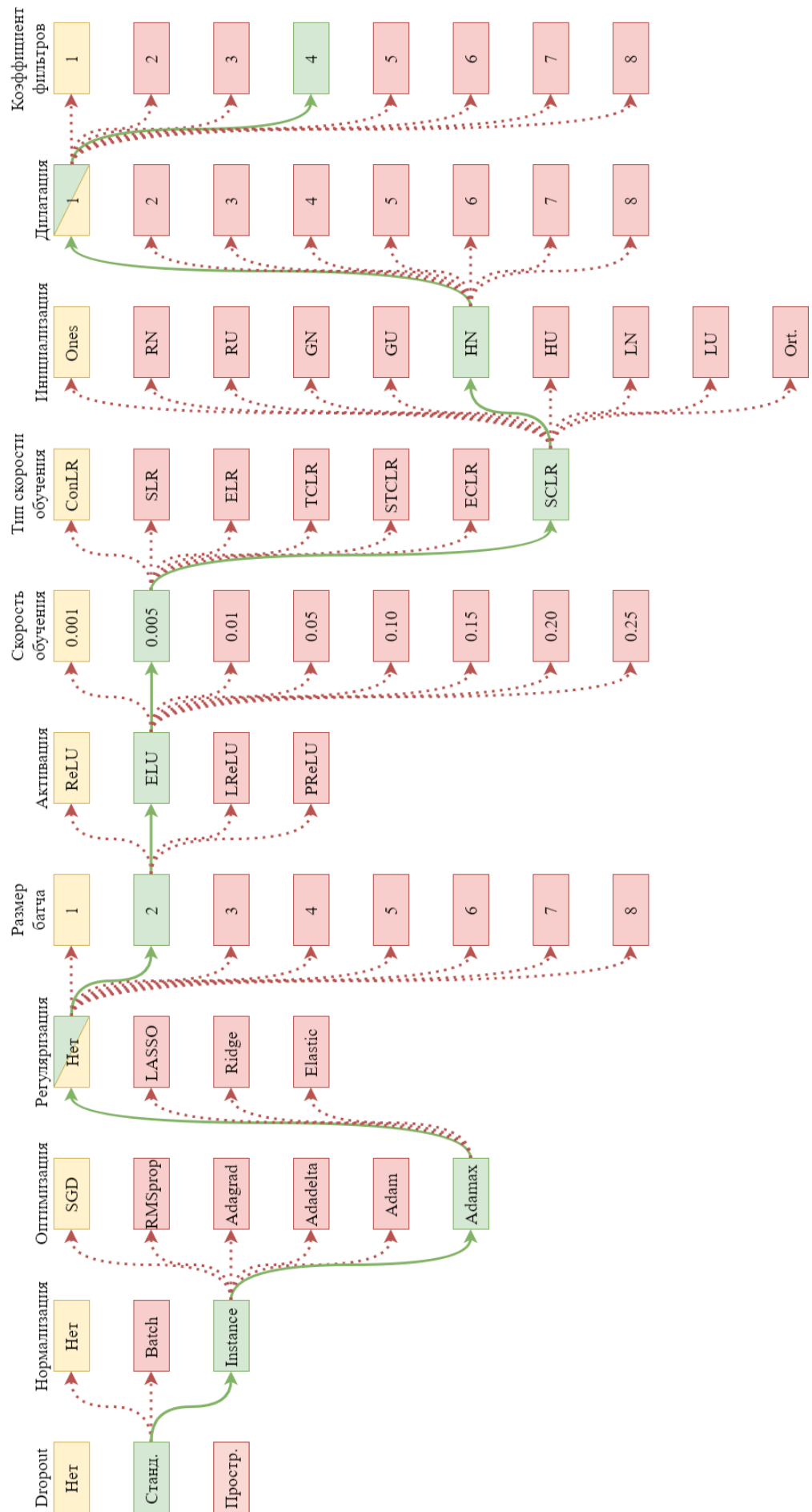


Рисунок 3.22 – Процедура выбора оптимальных значений гиперпараметров для предлагаемой модели

3.3 Сегментация медицинского инструмента

Для выполнения сегментации различных объектов использовалась V-net модель, оптимальные гиперпараметры которой получены в пункте 3.2. Данная модель была обучена на основе методов и алгоритмов, описанных в главе 2, включая аугментацию, использование динамического размера батча, циклической скорости обучения, а также алгоритма test-time аугментации на этапе предсказания. В качестве сегментируемых объектов рассматривались данные не только медицинского инструмента, но и анатомических структур.

3.3.1 Результаты сегментации медицинского инструмента

В качестве данных для сегментации медицинского инструмента использовался набор данных, краткое описание которого представлено в таблице 3.19. После обучения V-net модель была дополнительно протестирована с алгоритмом test-time аугментации. На основе полученных результатов, представленных на рисунке 3.23, видно, что алгоритм ТТА положительно влияет на предсказание новых данных. В особенности, видно уменьшение разброса на исследуемых выборках. Для стандартной модели V-net точность на обучении, валидации и тестировании составила $93.8 \pm 2.2\%$, $90.5 \pm 3.5\%$ и $90.6 \pm 3.3\%$ соответственно. В свою очередь, модель V-net с использованием алгоритма ТТА на 3% увеличила среднюю точность на этапе предсказания, а также уменьшила разброс предсказываемых значений. В результате, точность на этапе обучения, валидации и тестирования составила $94.1 \pm 2.0\%$, $93.5 \pm 2.4\%$ и $93.6 \pm 2.4\%$ соответственно.

Таблица 3.19 – Характеристика набора данных с медицинским инструментом

Исследуемый объект	Медицинский инструмент
Модальность	Трёхмерная эпикардальная эхокардиография
Размер набора данных	250 трёхмерных экземпляров, из которых:
	200 экземпляров для обучения
	25 экземпляров для валидации
	25 экземпляров для тестирования
Источник	Boston Children's Hospital

Предлагаемые методы обучения, модель V-net и модель V-net (TTA), сравнивались с рядом популярных моделей сегментации, а именно с FCN, SegNet, Deep Medic, Pseudo 3D U-net, U-net и разреженной U-net. Данные модели базируются на подходе глубокого обучения. Архитектуры этих моделей варьируются по количеству слоёв, однако концептуально они строятся на использовании комбинации энкодера и декодера. Результаты оценки точности сегментации отражены в таблице 3.20 и на рисунке 3.24.

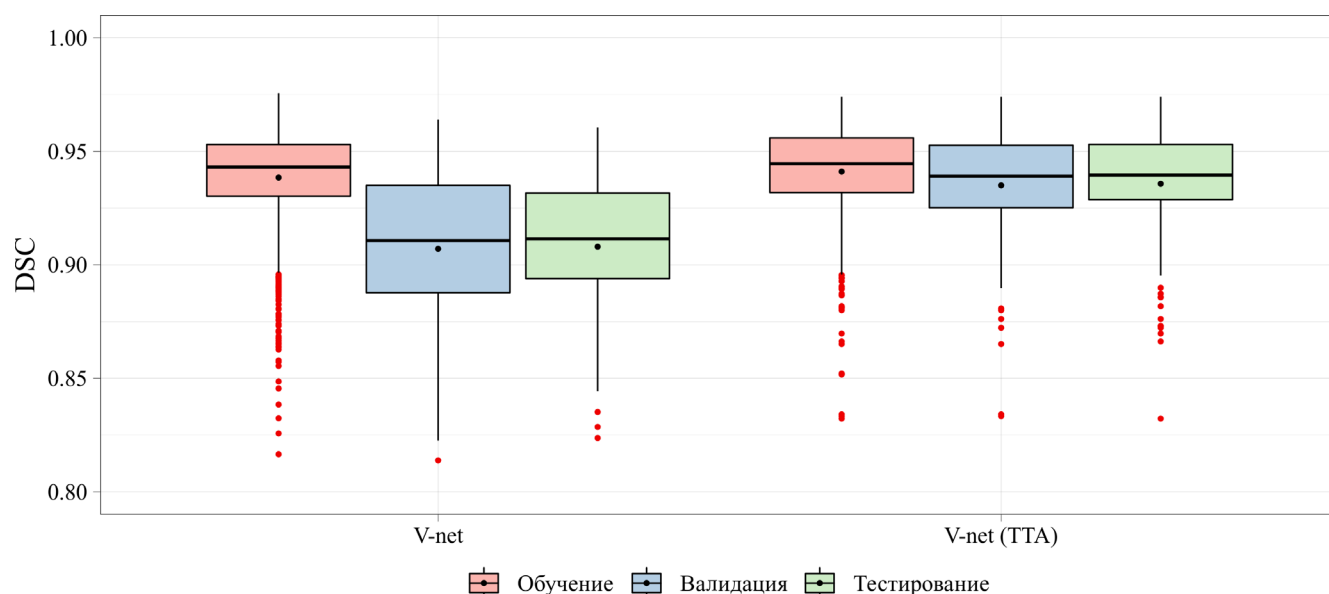
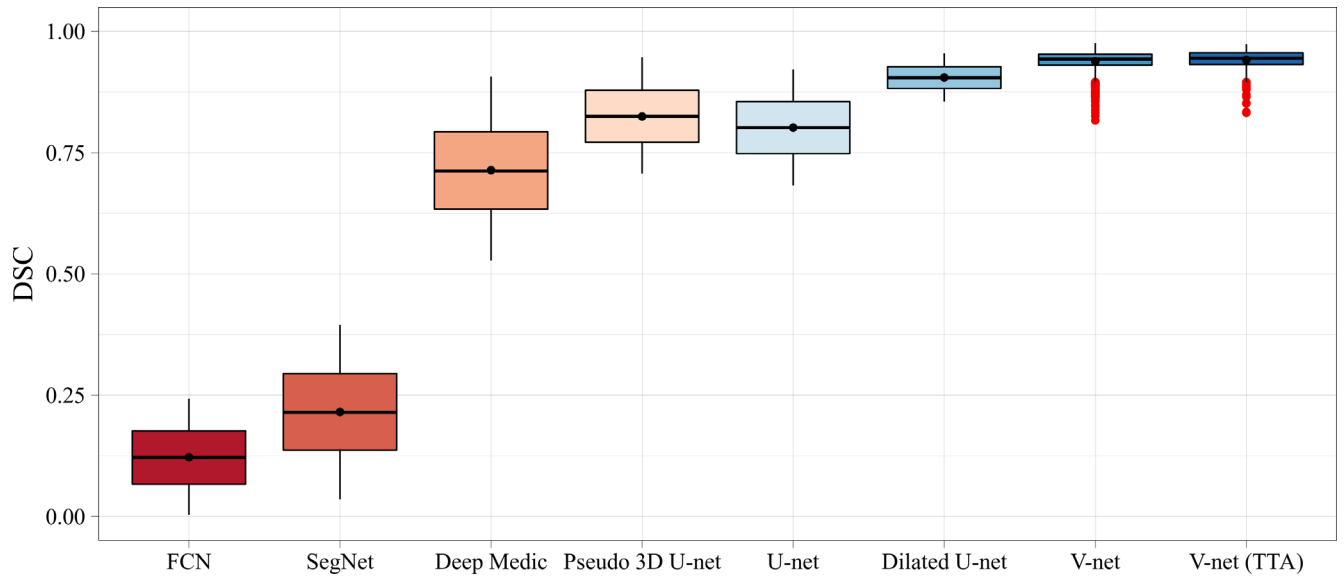


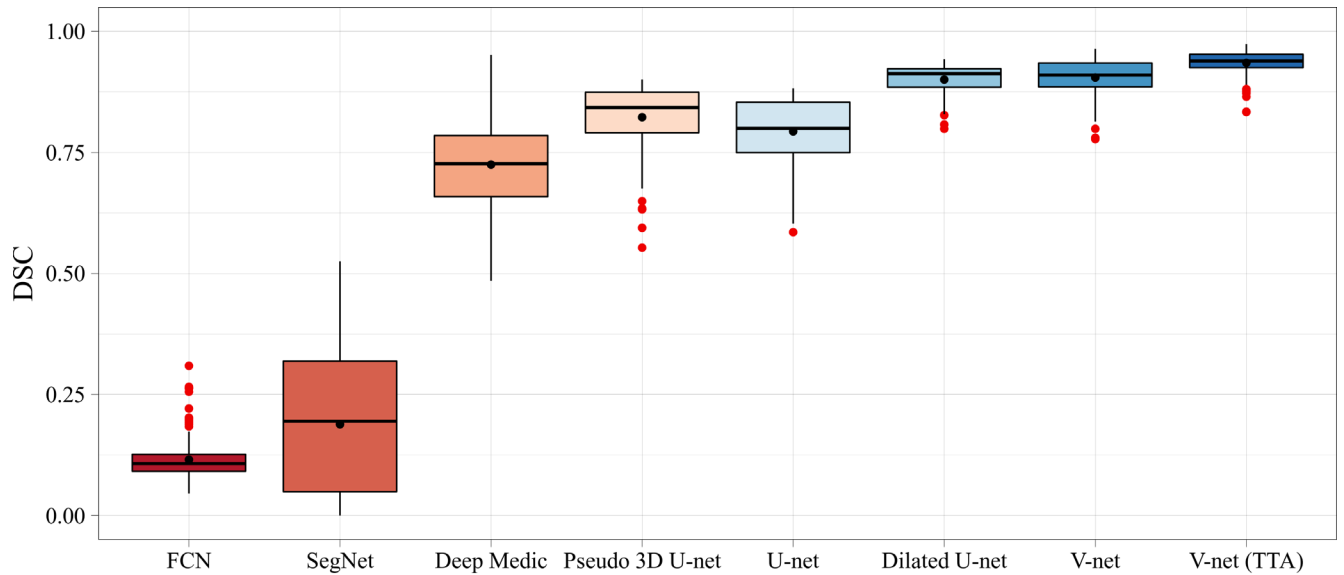
Рисунок 3.23 – Диаграммы размаха точности сегментации катетера

Таблица 3.20 – Точность сегментации медицинского инструмента

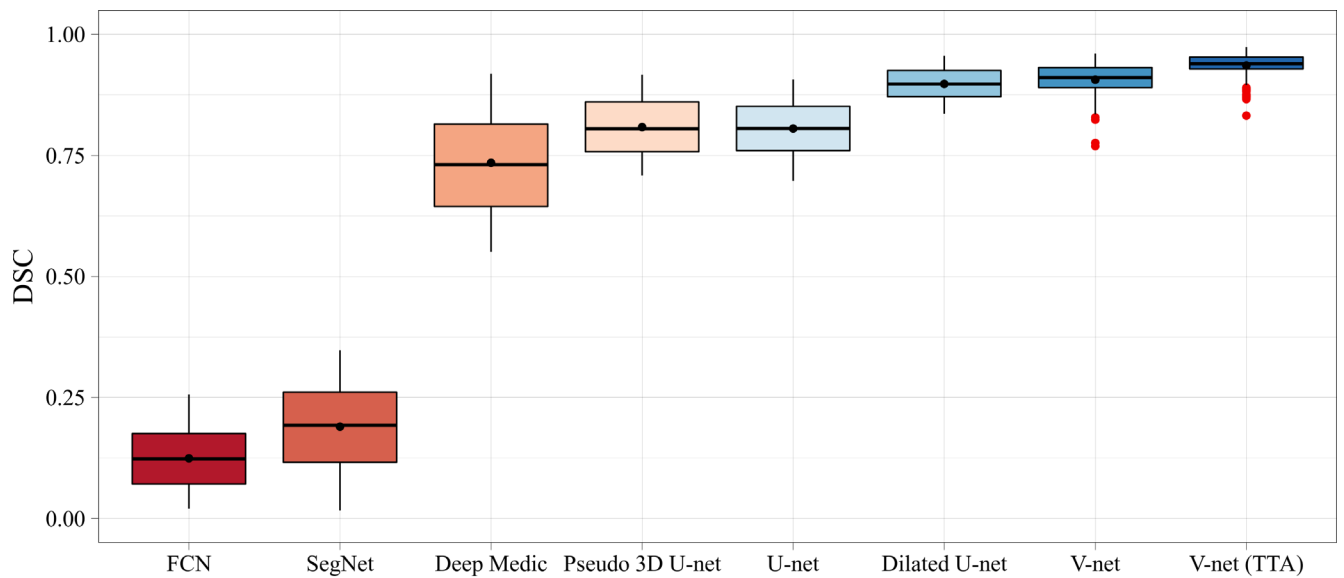
Модель	Индекс Дайса, %		
	Обучение	Валидация	Тестирование
FCN	12.2±6.6	11.5±4.3	12.4±6.4
SegNet	21.5±9.6	18.9±15.3	19.0±8.5
Deep Medic	71.4±10.0	72.5±10.5	73.5±12.5
Pseudo 3D U-net	82.5±6.5	82.3±6.8	80.8±6.0
U-net	80.2±6.5	79.3±6.4	80.5±5.8
Dilated U-net	90.5±2.7	90.0±3.0	89.8±3.4
V-net	93.8±2.2	90.5±3.5	90.6±3.3
V-net (TTA)	94.1±2.0	93.5±2.4	93.6±2.4



(a) Обучение



(б) Валидация



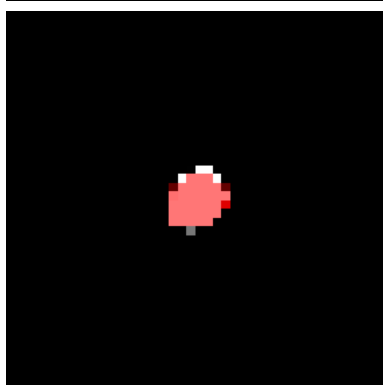
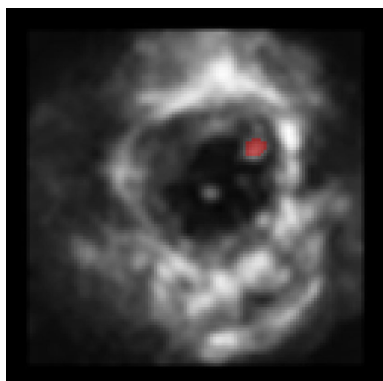
(б) Тестирование

Рисунок 3.24 – Диаграммы размаха точности сегментации катетера для разных моделей

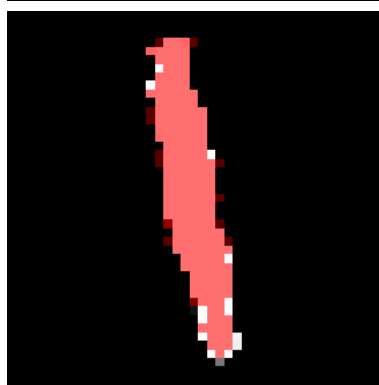
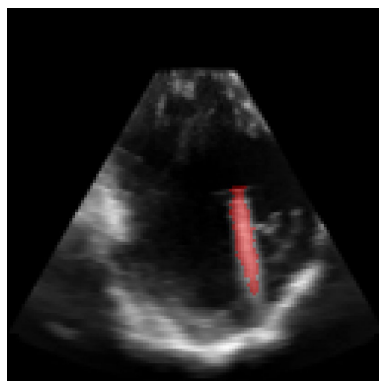
Исходя из полученных результатов, можно сделать вывод, что такие модели, как FCN и SegNet не выполняют сегментацию должным образом. Это объясняется рядом причин, а именно недостаточной сложностью моделей, затуханием градиентов при использовании глубоких архитектур, отсутствием дополнительных пробросов карт признаков и т.д. С другой стороны, архитектуры, имеющие структуру близкую к U-net, выполняют сегментацию с более высокой точностью. Стоит также отметить то, что на данных с медицинским инструментом модель Pseudo 3D U-net выполняет сегментацию эффективнее полноценной трёхмерной модели на основе стандартной U-net архитектуры. В связи с тем, что архитектура Pseudo 3D U-net использует двухмерные тензоры для предсказания новых данных, такая модель может эффективнее использоваться на устройствах с малой вычислительной мощностью. В таком случае будет требоваться меньше видеопамати для того, чтобы выполнять предсказания новых данных, чем, например, при имплементации стандартной или разреженной U-net архитектур. Что касается предлагаемой архитектуры V-net, то она превосходит U-net архитектуру на 10.1% по ряду причин, а именно: использование внедрённого дополнительного потока градиентов через пробросы признаков позволяет точнее определить глобальный минимум функции потерь; применение циклической скорости обучения позволяет эффективнее проходить локальные минимумы; обучение на основе переменного размера батча позволяет быстрее спуститься в область глобального минимум; внедрение алгоритма ТТА гарантирует более высокое значение точности сегментации.

В качестве визуальной оценки результатов сегментации на рисунках 3.25 и 3.26 представлены результаты работы предлагаемой модели на основе реальных и синтезированных данных медицинского инструмента. Дополнительно на этих рисунках (второй ряд каждого случая) отражено сравнение результатов сегментации V-net модели (белые маски) и исходных масок (красные маски). При сравнении полученных результатов видно, что разница между масками незначительная, что говорит об эффективной сегментации и, соответственно, о хорошей обобщающей способности модели.

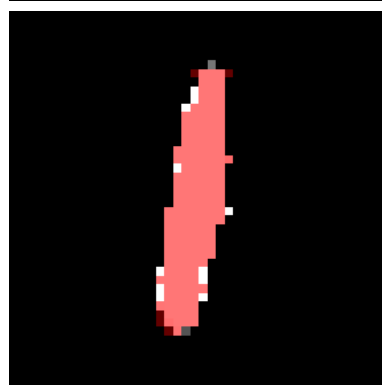
Аксиальная плоскость



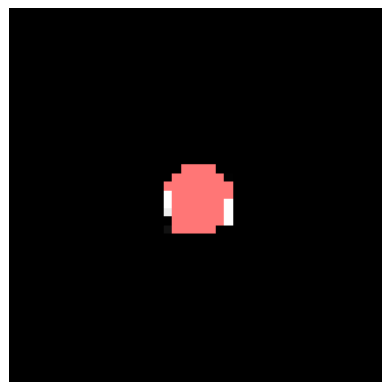
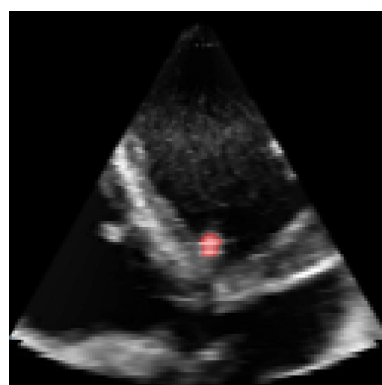
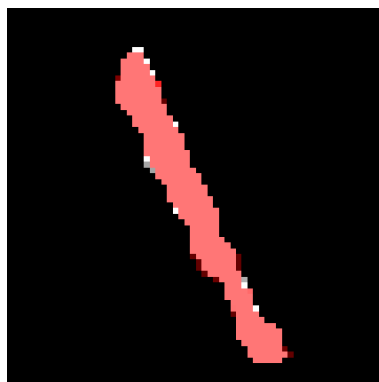
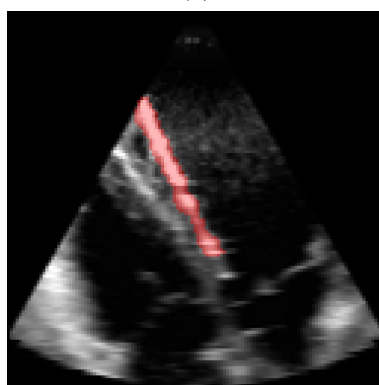
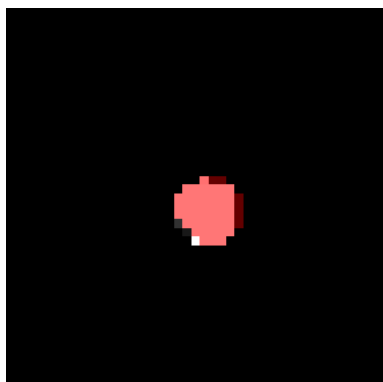
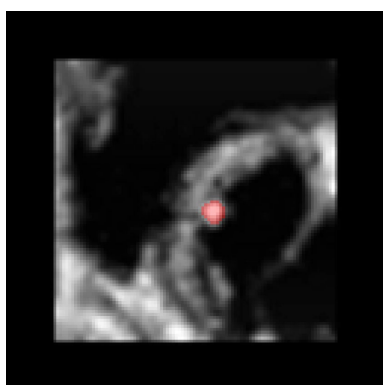
Корональная плоскость



Сагиттальная плоскость

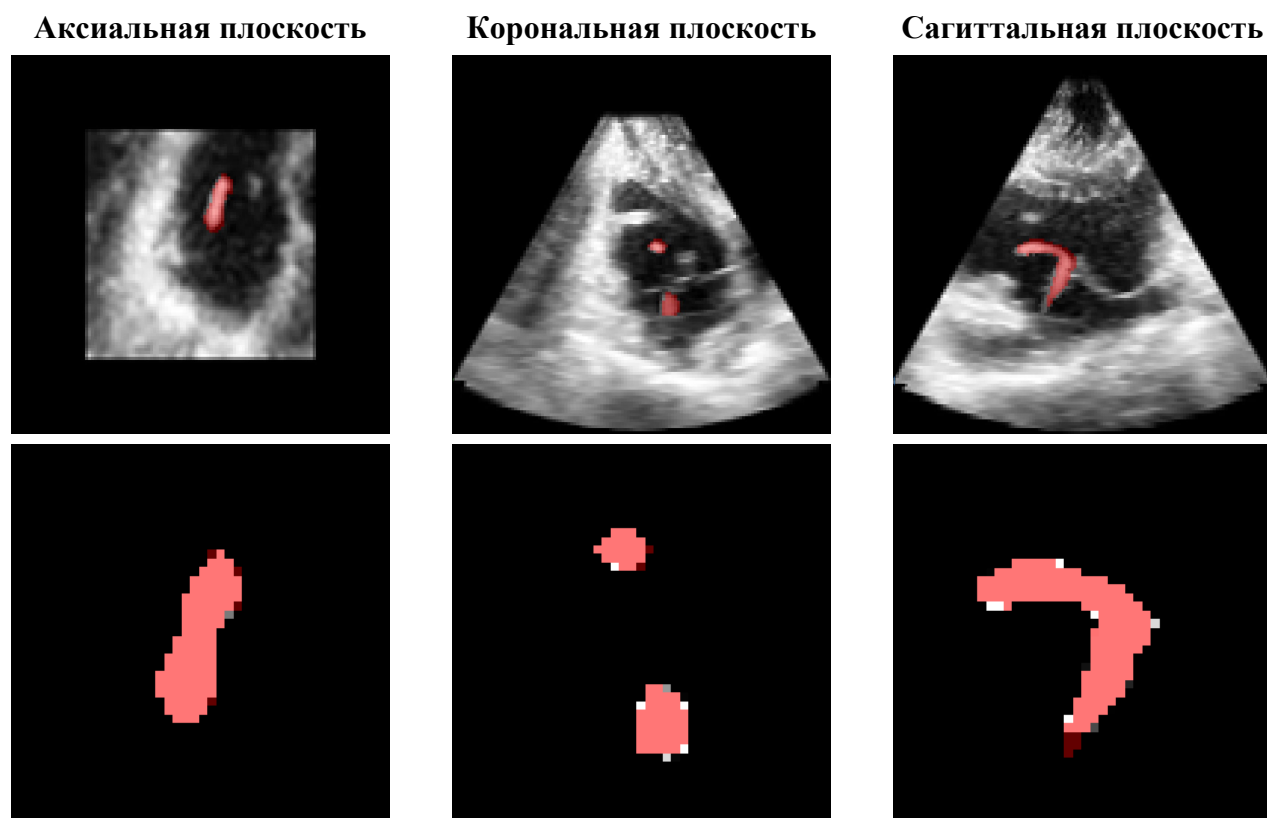


(a)

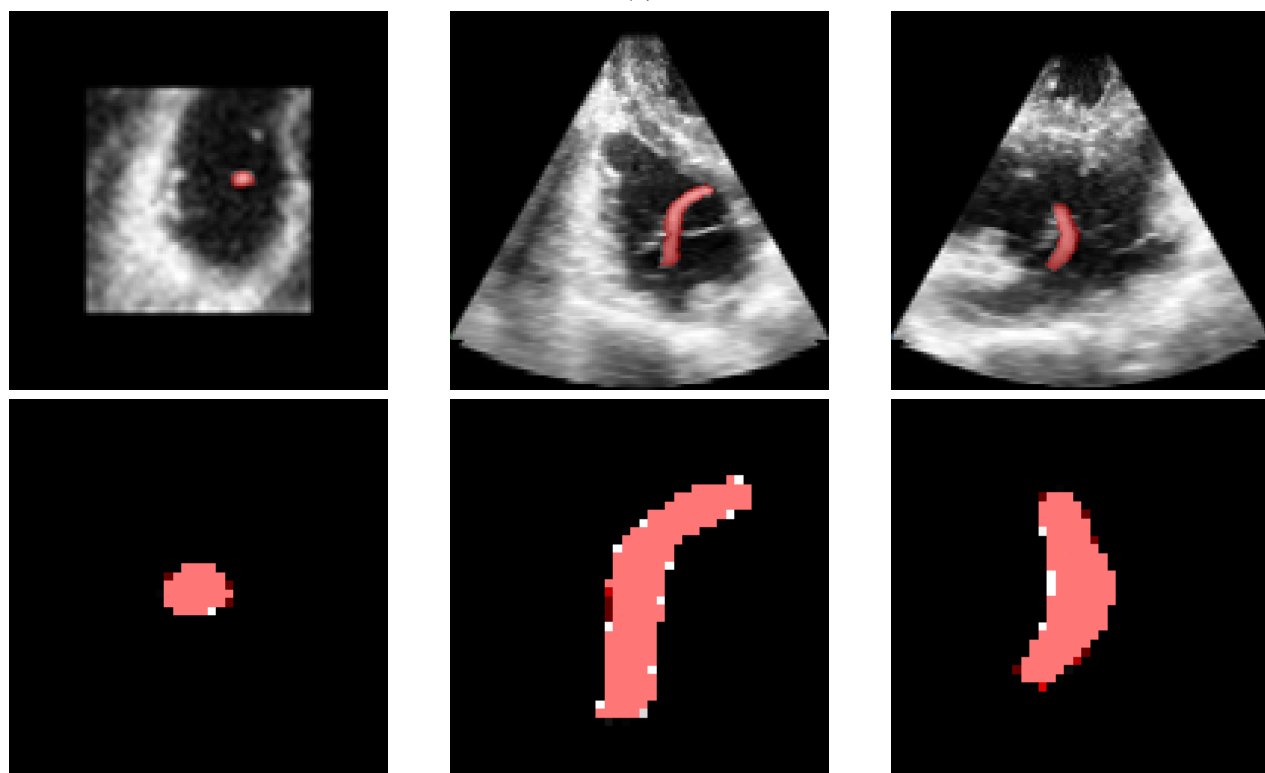


(б)

Рисунок 3.25 – Результат сегментации реальных данных катетера на основании предложенной модели V-net (TTA)



(a)



(б)

Рисунок 3.26 – Результат сегментации искусственных данных катетера на основании предложенной модели V-net (TTA)

3.3.2 Оценка результатов сегментации V-net модели

В данном пункте освещена статистическая оценка результатов сегментации медицинского инструмента, которые получены восемью моделями, базирующимися на разных архитектурах свёрточных нейронных сетей. В целях первичной оценки значимости различий между средними значениями индекса Дайса исследуемых групп использовался дисперсионный анализ (метод ANOVA), нулевая гипотеза H_0 которого указывает на отсутствие отличий между группами. Альтернативная гипотеза H_1 выдвигает предположение о том, что средние в исследуемых группах не равны. На основании результатов, отражённых в таблице 3.21, видно, что значения F на всех этапах больше критических значений $F_{кр}$ при выбранном уровне значимости равном 0.05. Вычисленные значения p также сильно меньше уровня значимости, что говорит о том, что нулевая гипотеза может быть отклонена. В связи с этим можно утверждать, что как минимум одна из выборок принадлежит к совершенно иной генеральной совокупности.

В качестве первоначальной оценки распределения индекса Дайса разных архитектур были визуально проинспектированы графики плотности распределения. Данные графики представлены на рисунке 3.27. На основе плотности распределения видно, что все выборки имеют разную природу плотности распределения. Так, распределения индекса Дайса для архитектур FCN и SegNet предположительно имеют нормальный характер распределения, обладая при этом ярко выраженным отрицательным эксцессом. С другой стороны, плотность распределения данных для архитектуры Deep Medic визуально близка к равномерному распределению. В свою очередь, распределение предлагаемой архитектуры V-net близко по характеру поведения к бета распределению.

Таблица 3.21 – Результаты ANOVA тестирования рассматриваемых архитектур

Параметр	Обучение	Валидация	Тестирование
F	30470	2736	4123
$F_{кр}$	2.0	2.0	2.0
p	8.4E-105	6.4E-83	4.3E-87
Значимы ли различия?	Да	Да	Да

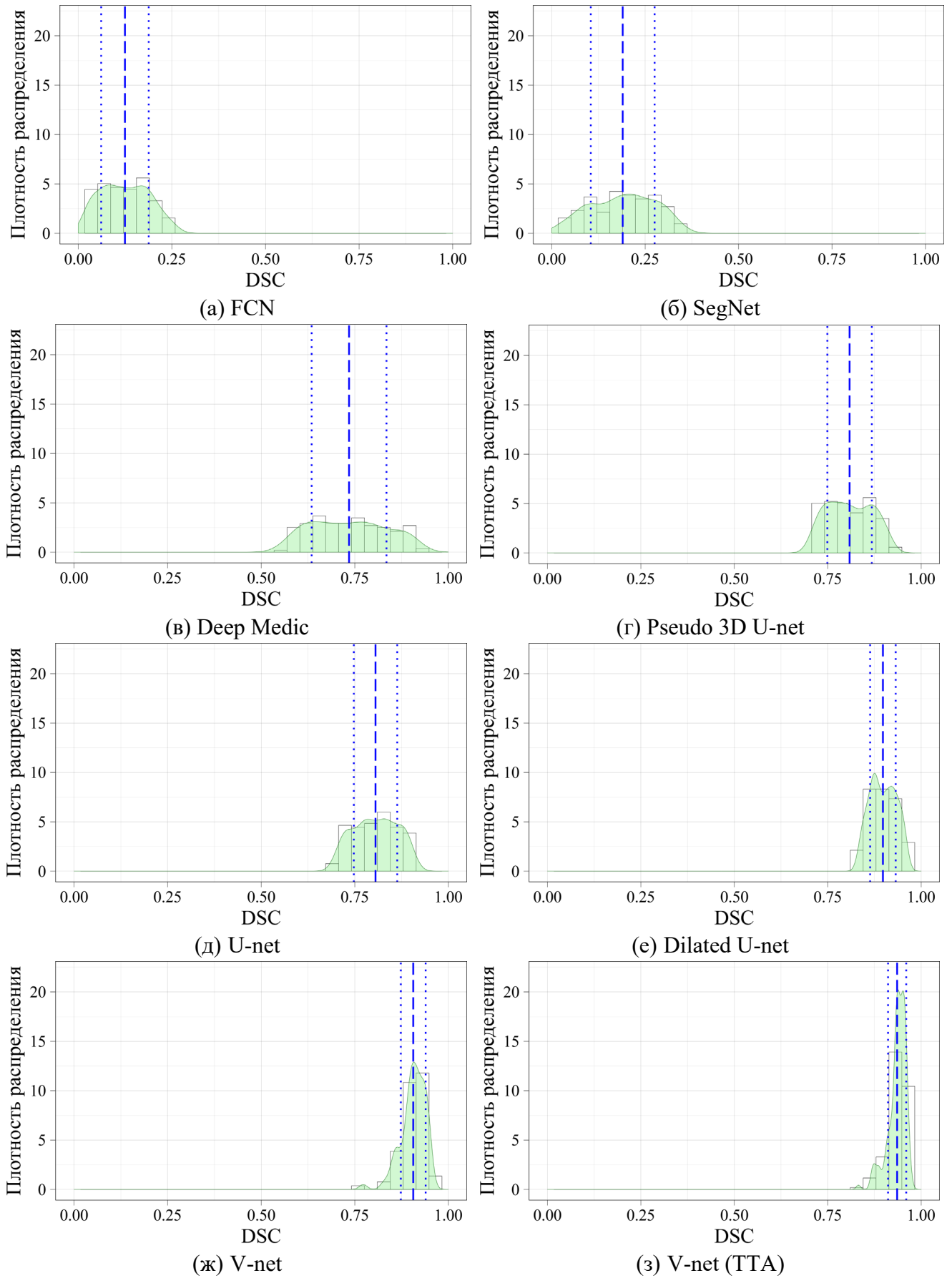


Рисунок 3.27 – Распределение индекса Дайса для рассматриваемых архитектур

В качестве дополнительной оценки полученных выборок были построены графики сравнения плотности распределения индекса Дайса разных моделей. Важным будет отметить, что базовым распределением для сравнения являлась выборка для модели V-net (TTA), а распределения остальных моделей с ней сравнивались. Данные графики представлены на рисунке 3.28. Исходя из графиков видно, что средние значения точности сегментации модели V-net (TTA) и остальных моделей достаточно сильно отличаются. Исключениями являются модель V-net без использования алгоритма TTA и разреженная (dilated) U-net. По типу распределению и среднему значению индекса Дайса разреженная U-net и V-net приближаются к модели V-net (TTA). Важным будет отметить тот факт, что отсутствие пробросов признаков архитектуры приводит к уплощению и увеличению по ширине плотности распределения. В свою очередь, это приводит к худшей обобщающей способности модели, большим значениям стандартного распределения и интерквартильного размаха точности сегментации.

Перед проверкой формальных статистических гипотез была проведена оценка типа распределения. Визуальная проверка на нормальность каждого из распределений выполнялась с помощью квантильных графиков, отражённых на рисунке 3.29. На основе данных графиков видно, что большинство распределений S-образно изогнуты в точке перегиба нулевого квантиля нормального распределения. Исключениями являются модели V-net и V-net (TTA), которые в точке перегиба нулевого квантиля нормального распределения изогнуты C-образно. В ряде случаев наблюдаются значительные отклонения исследуемых квантилей от нормальных при низких и высоких значениях индекса Дайса. Особенно это заметно для таких архитектур, как FCN, SegNet, Deep Medic и Pseudo 3D U-net. Тем не менее, для всех распределений наблюдается выход квантилей за пределы границ доверительной зоны. На основе подобных выводов можно предварительно утверждать о том, что исследуемые выборки имеют тип распределения данных, отличающийся от нормального закона распределения.

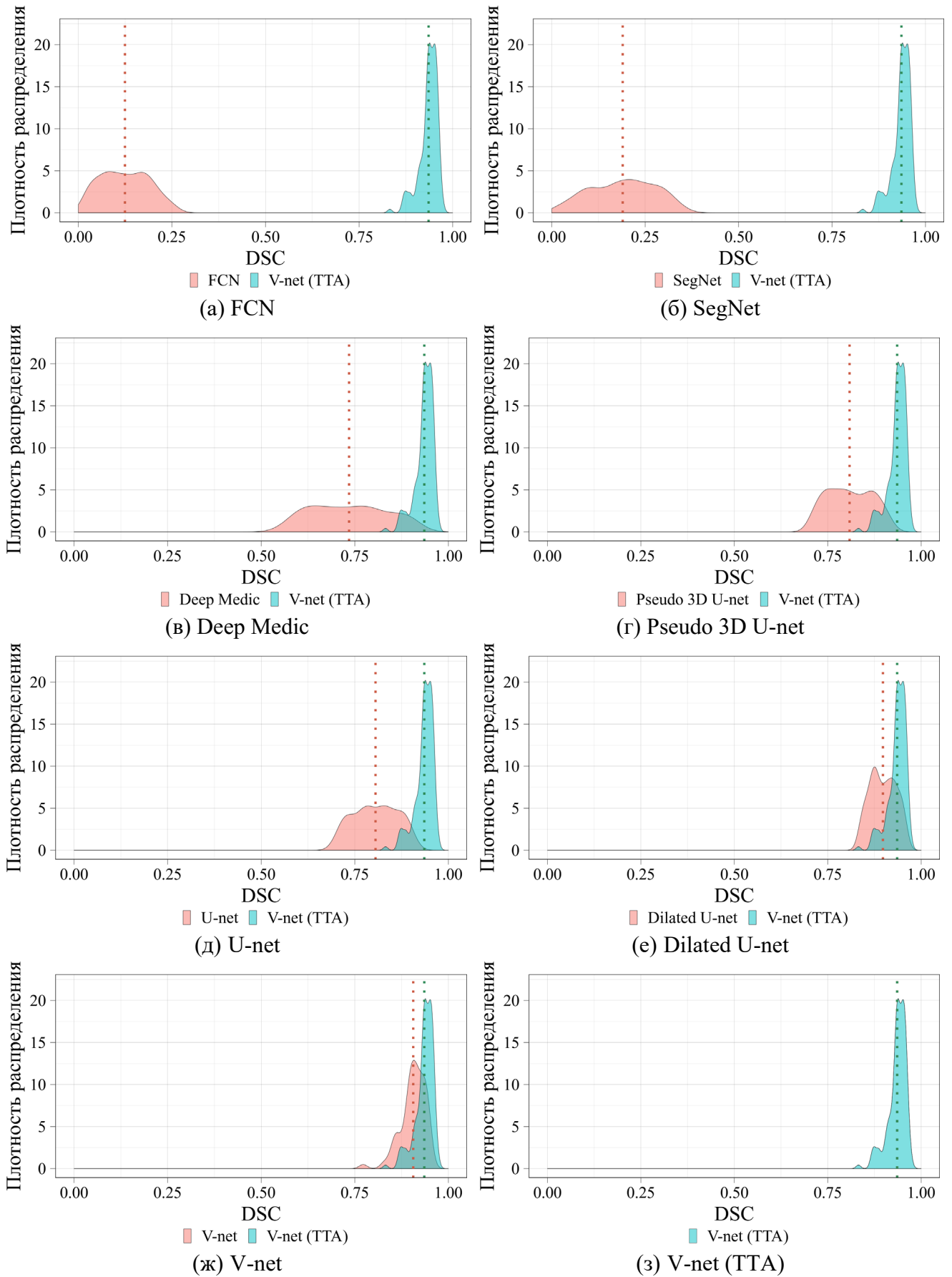


Рисунок 3.28 – Сравнение распределений индекса Дайса предлагаемой архитектуры V-net (TTA) с другими архитектурами

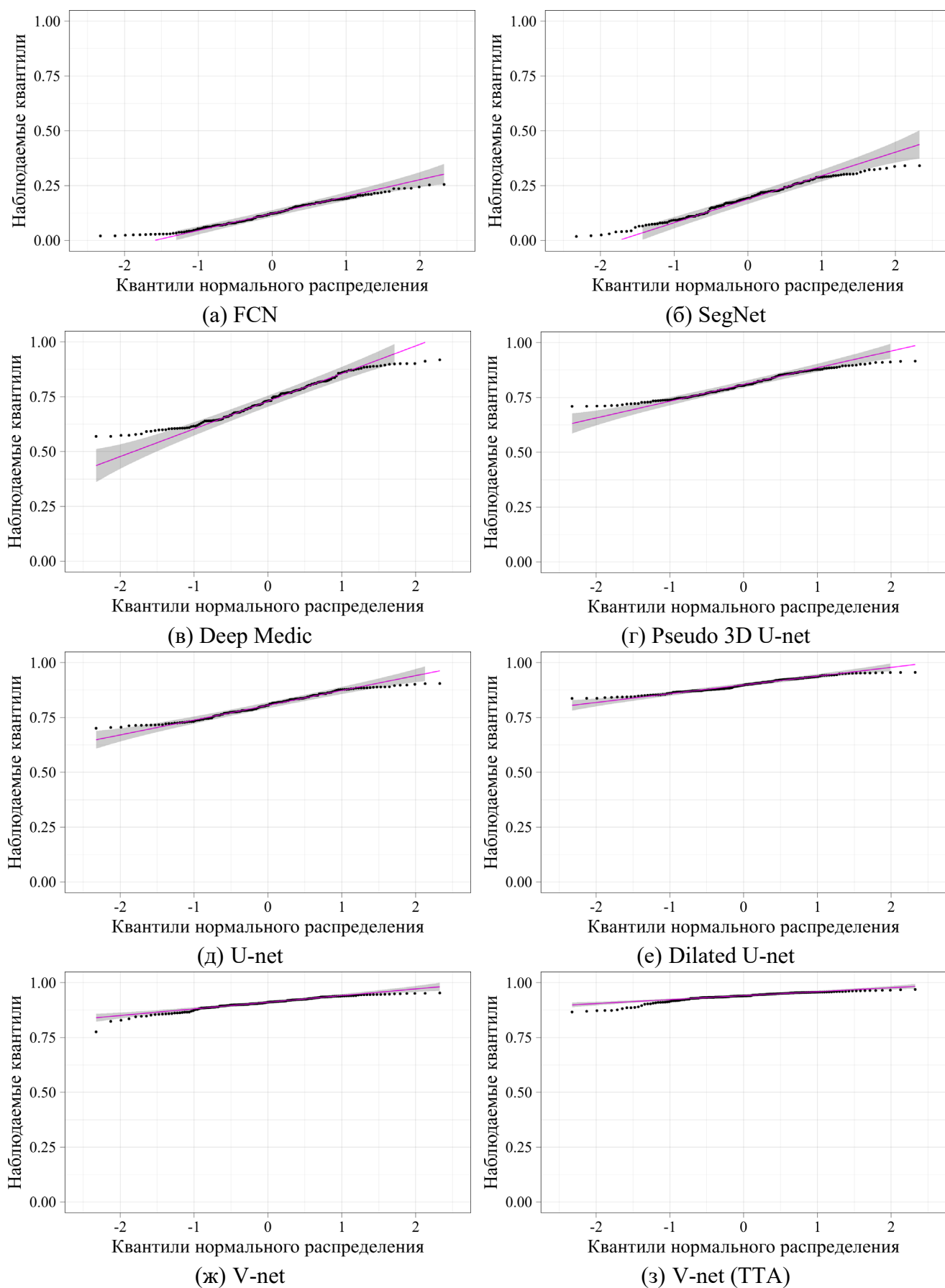


Рисунок 3.29 – Квантильный график рассматриваемых архитектур

С целью оценки значимости различий были выполнены два формальных статистических теста, а именно тест Шапиро-Уилка и тест Вилкоксона. Несмотря на то, что квантильные графики дали примерную визуальную картину о типе исследуемых распределений, тест Шапиро-Уилка проводился с целью формального дополнительного подтверждения или отрицания факта нормальности распределения данных. Нулевая гипотеза H_0 данного теста предполагает, что исследуемая выборка распределена нормально. Согласно полученным результатам, отражённым в таблице 3.22, выборки индекса Дайса всех архитектур оказались распределёнными ненормально при выбранном уровне значимости равном 0.05.

После определения типа распределения был выполнен тест Вилкоксона. Сравнение выборок при помощи критерия Вилкоксона основано на нахождении разницы между всеми имеющимися парными выборочными наблюдениями с

Таблица 3.22 – Результаты статистической оценки рассматриваемых архитектур

Статистический параметр	Архитектура							
	FCN	SegNet	Deep Medic	Pseudo 3D U-net	U-net	Dilated U-net	V-net	V-net (TTA)
Критерий Шапиро-Уилка	0.966	0.972	0.959	0.957	0.959	0.959	0.925	0.885
р-значение	9.3E-04	3.7E-03	2.1E-04	1.2E-04	2.1E-04	2.1E-04	4.3E-07	2.1E-09
Нормальное распределение?	Нет	Нет	Нет	Нет	Нет	Нет	Нет	Нет
W-критерий Вилкоксона	11325	11325	11300	11297	11320	9902	11304	–
р-значение	2.3E-26	2.3E-26	3.8E-26	4.1E-26	2.6E-26	1.8E-15	3.5E-26	–
Значимы ли различия?	Да	Да	Да	Да	Да	Да	Да	–
Доверительный интервал (мин.)	0.801	0.730	0.184	0.117	0.119	0.032	0.024	–
Доверительный интервал (макс.)	0.824	0.760	0.221	0.140	0.141	0.048	0.030	–
Псевдомедиана	0.812	0.746	0.202	0.128	0.130	0.040	0.027	–

целью проверки нулевой гипотезы H_0 , которая заключается в том, что медиана полученных разностей равна нулю. Дополнительно в таблице 3.22 отражены минимальная и максимальная границы доверительного интервала, а также медиана разницы между парными выборочными наблюдениями (псевдомедиана). Как отмечалось выше, все выборки сравниваются с базовой выборкой индекса Дайса, полученной на основе модели с архитектурой V-net (ТТА). На основании полученных результатов можно заключить, что все выборки значимо отличаются от базовой выборки. Последнее означает, что предлагаемая архитектура V-net, работающая совместно с алгоритмом ТТА, статистически значимо отличается от других архитектур по точности и обобщающей способности.

3.4 Сегментация анатомических структур

В данном пункте предлагаемые методы, алгоритмы и предложенная модель V-net с плотными пробросами признаков тестировались для выполнения сегментации различных анатомических структур. Все изображения были собраны в рамках проведения медицинских процедур, после чего переформатированы в формат NIfTI и деперсонализированы. Результаты тестирования предлагаемой модели представлены ниже в пунктах 3.4.1 – 3.4.5.

3.4.1 Сегментация левого предсердия

В качестве данных для сегментации левого предсердия использовался набор данных, характеристики которого отражены в таблице 3.23. Оценка точности сегментации показана в таблице 3.24 и на рисунке 3.30. Исходные данные, результат сегментации и сравнение масок представлены на рисунке 3.31.

Таблица 3.23 – Характеристика набора данных левого предсердия

Исследуемый объект	Левое предсердие
Модальность	Мономодальная магнитно-резонансная томография
Размер набора данных	30 трёхмерных изображений, из которых:
	20 изображений для обучения
	10 изображений для тестирования
Размер изображения	100 x 320 x 320
Источник	King's College London

Таблица 3.24 – Точность сегментации левого предсердия

Модель	Индекс Дайса, %		
	Обучение	Валидация	Тестирование
V-net	93.0±2.0	90.2±2.2	87.9±2.6
V-net (TTA)	93.7±1.9	93.4±2.8	92.2±2.7

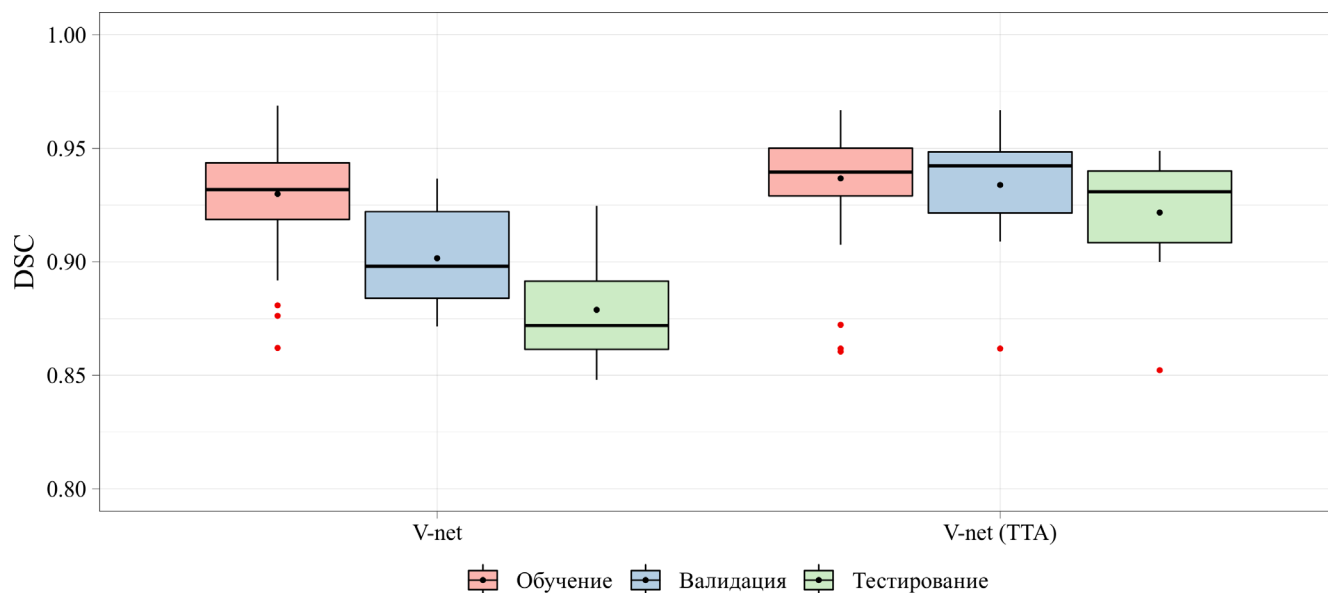


Рисунок 3.30 – Диаграммы размаха точности сегментации левого предсердия

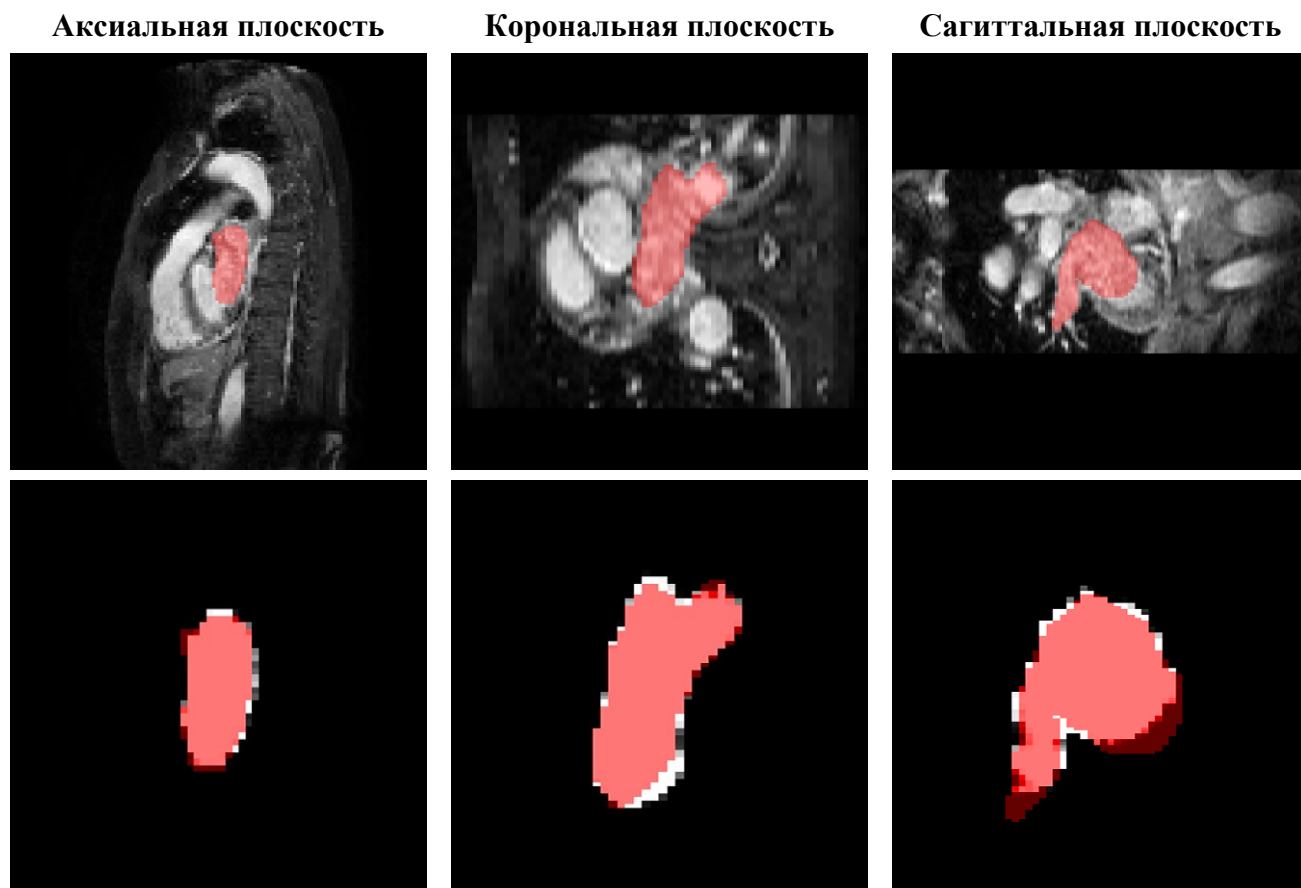


Рисунок 3.31– Результат сегментации левого предсердия на основе модели V-net (TTA)

3.4.2 Сегментация поджелудочной железы

В качестве данных для сегментации поджелудочной железы использовался набор данных, характеристики которого отражены в таблице 3.25. Оценка точности сегментации показана в таблице 3.26 и на рисунке 3.32. Исходные данные, результат сегментации и сравнение масок представлены на рисунке 3.33.

Таблица 3.25 – Характеристика набора данных поджелудочной железы

Исследуемый объект	Поджелудочная железа
Модальность	Контрастная компьютерная томография
Размер набора данных	420 трёхмерных изображений, из которых:
	281 изображение для обучения
	139 изображений для тестирования
Размер изображения	113 x 512 x 512
Источник	Memorial Sloan Kettering Cancer Center

Таблица 3.26 – Точность сегментации поджелудочной железы

Модель	Индекс Дайса, %		
	Обучение	Валидация	Тестирование
V-net	84.9±4.1	68.3±11.3	68.6±11.6
V-net (TTA)	85.7±4.4	82.9±5.7	82.9±5.8

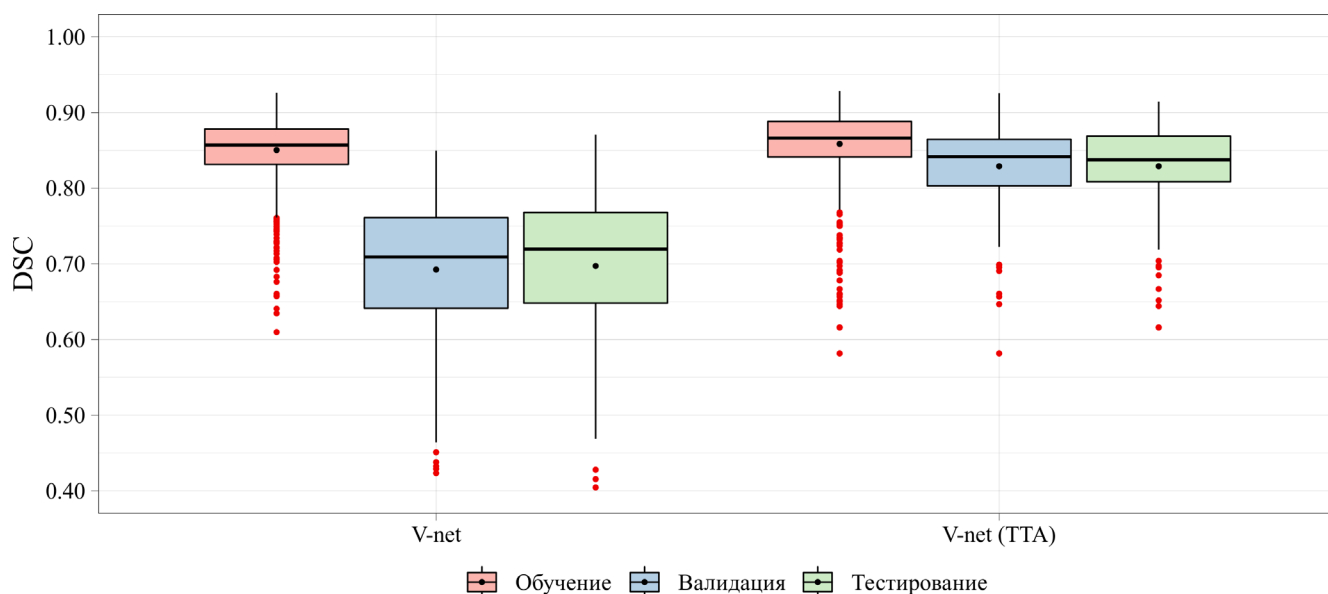


Рисунок 3.32 – Диаграммы размаха точности сегментации поджелудочной железы

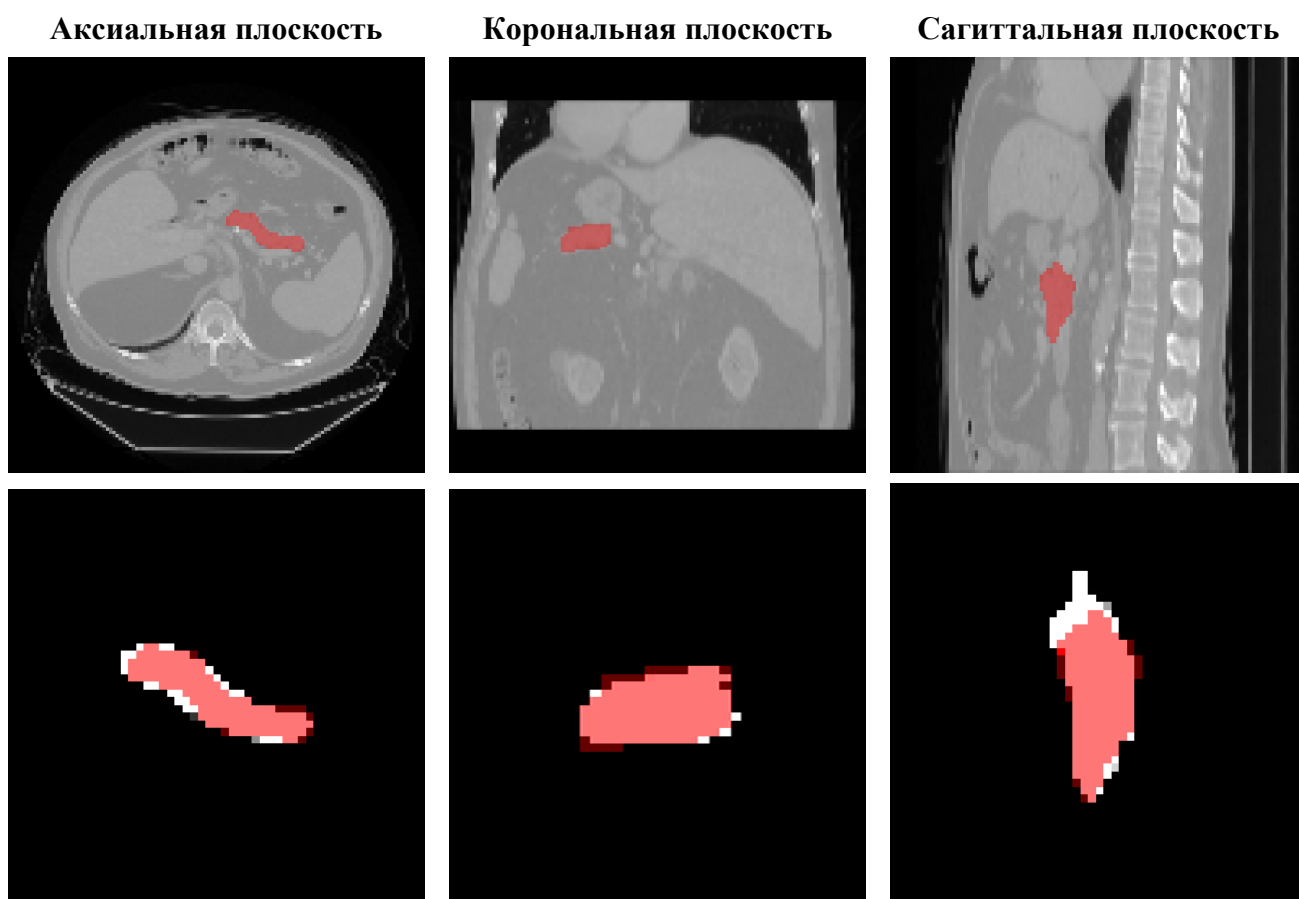


Рисунок 3.33 – Результат сегментации поджелудочной железы на основе модели V-net (TTA)

3.4.3 Сегментация селезёнки

В качестве данных для сегментации селезёнки использовался набор данных, характеристики которого отражены в таблице 3.27. Оценка точности сегментации показана в таблице 3.28 и на рисунке 3.34. Исходные данные, результат сегментации и сравнение масок представлены на рисунке 3.35.

Таблица 3.27 – Характеристика набора данных селезёнки

Исследуемый объект	Селезёнка
Модальность	Компьютерная томография
Размер набора данных	190 трёхмерных изображений, из которых: 126 изображений для обучения 64 изображения для тестирования
Размер изображения	92 x 512x 512
Источник	Memorial Sloan Kettering Cancer Center

Таблица 3.28 – Точность сегментации селезёнки

Модель	Индекс Дайса, %		
	Обучение	Валидация	Тестирование
V-net	93.5±2.0	87.2±6.0	85.4±8.0
V-net (TTA)	93.6±2.4	92.6±4.4	93.6±1.6

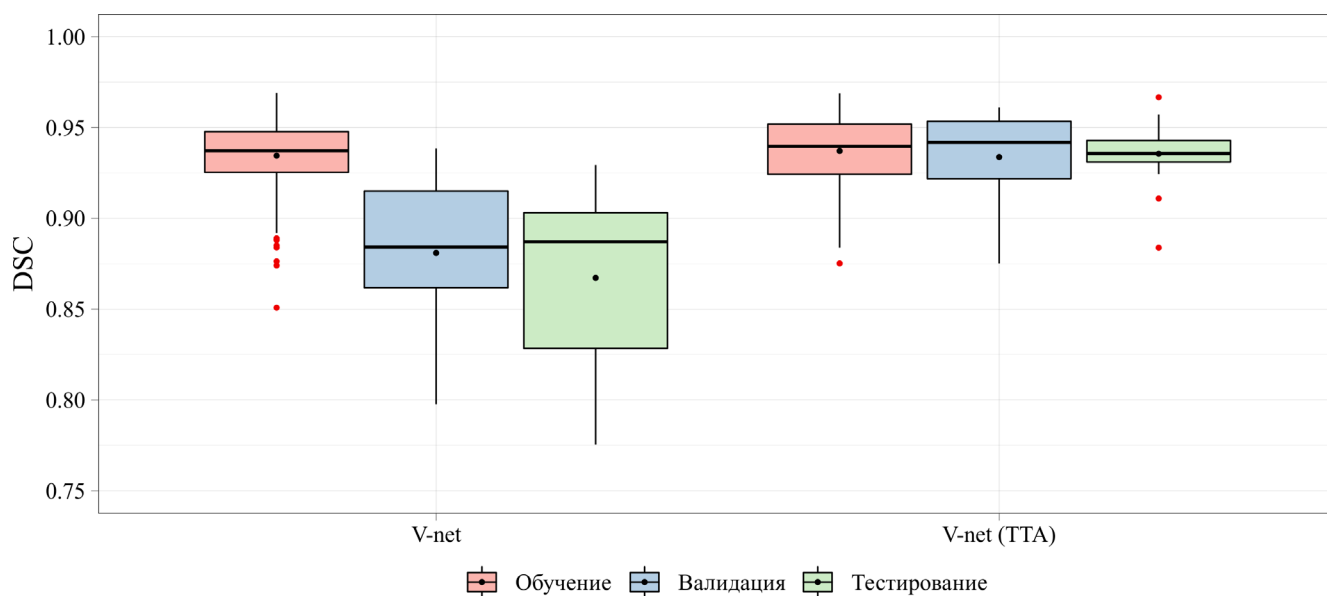


Рисунок 3.34 – Диаграммы размаха точности сегментации селезёнки

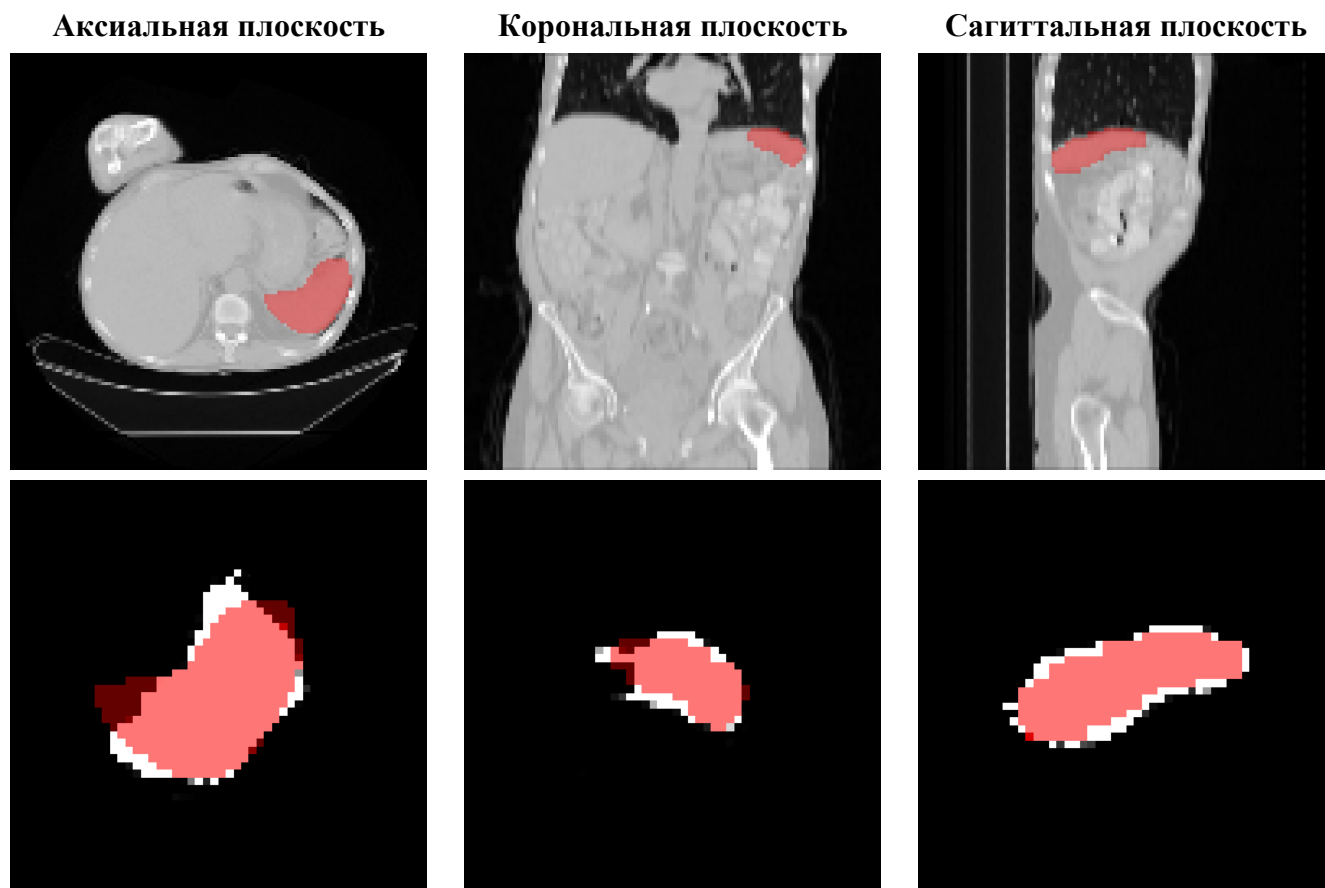


Рисунок 3.35 – Результат сегментации селезёнки на основе модели V-net (TTA)

3.4.4 Сегментация гиппокампа

В качестве данных для сегментации гиппокампа использовался набор данных, характеристики которого отражены в таблице 3.29. Оценка точности сегментации (индекса Дайса) показана в таблице 3.30 и на рисунке 3.36. Исходные данные, результат сегментации и сравнение масок представлены на рисунке 3.37.

Таблица 3.29 – Характеристика набора данных гиппокампа

Исследуемый объект	Гиппокамп
Модальность	Мономодальная магнитно-резонансная томография
Размер набора данных	394 трёхмерных изображения, из которых: 263 изображения для обучения 131 изображение для тестирования
Размер изображения	38 x 49 x 38
Источник	Vanderbilt University Medical Center

Таблица 3.30 – Точность сегментации гиппокампа

Модель	Индекс Дайса, %		
	Обучение	Валидация	Тестирование
V-net	95.9±0.7	91.5±1.5	91.7±1.7
V-net (TTA)	95.7±1.0	94.9±1.2	94.9±1.5

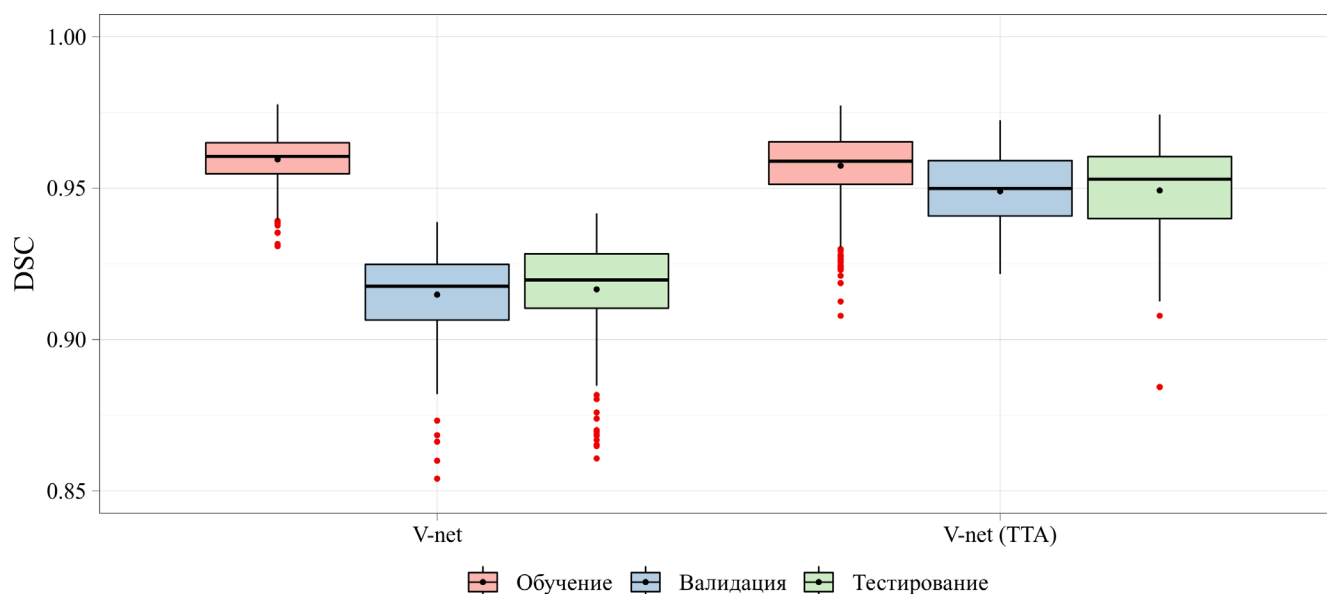


Рисунок 3.36 – Диаграммы размаха точности сегментации гиппокампа

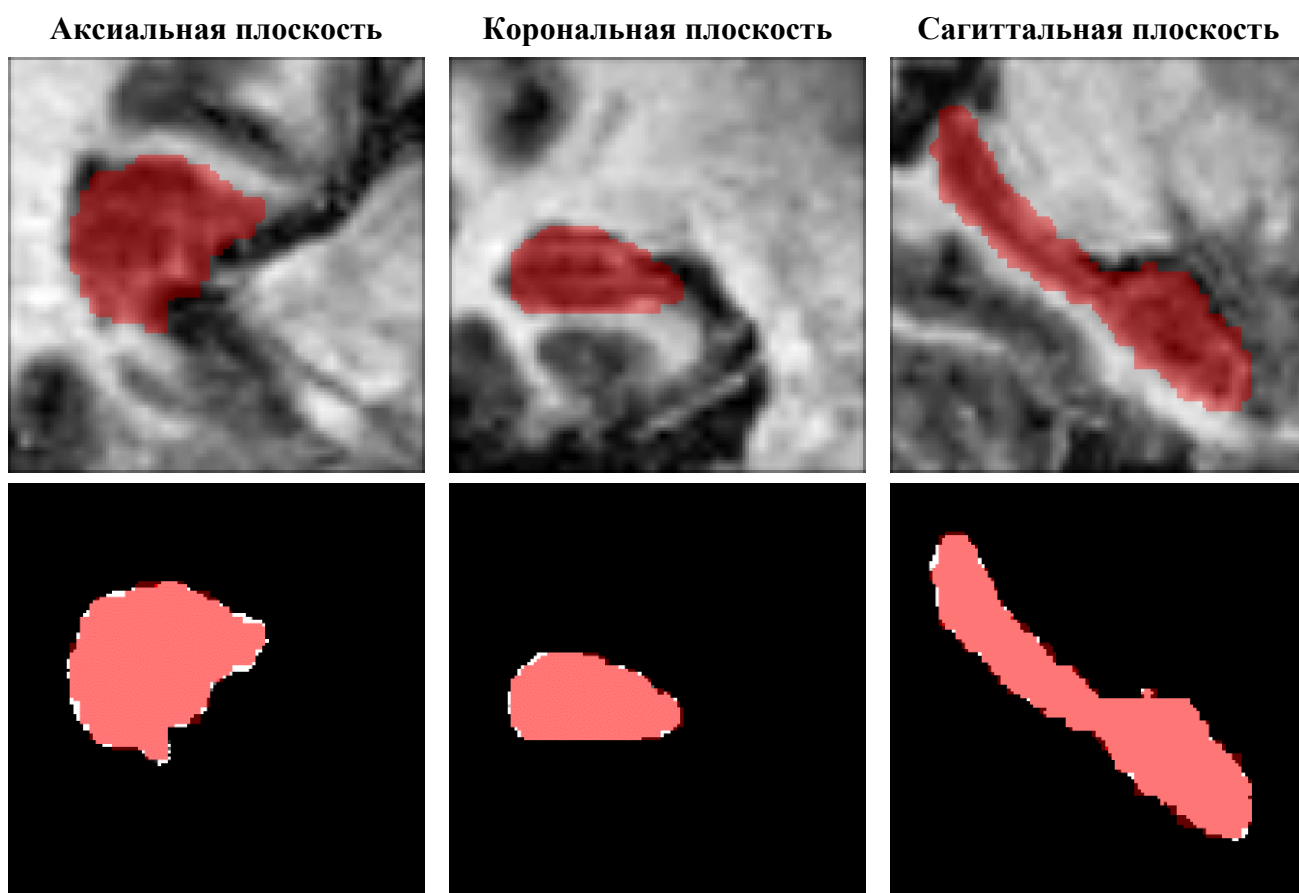


Рисунок 3.37 – Результат сегментации гиппокампа на основе модели V-net (TTA)

3.4.5 Сегментация печени

В качестве данных для печени использовался набор данных, характеристики которого отражены в таблице 3.31. Оценка точности сегментации (индекса Дайса) показана в таблице 3.32 и на рисунке 3.38. Исходные данные, результат сегментации и сравнение масок представлены на рисунке 3.39.

Таблица 3.31 – Характеристика набора данных печени

Исследуемый объект	Печень
Модальность	Контрастная компьютерная томография
Размер набора данных	201 трёхмерный изображений, из которых: 131 изображение для обучения 70 изображений для тестирования
Размер изображения	917 x 512 x 512
Источник	IRCAD Hôpitaux Universitaires

Таблица 3.32 – Точность сегментации печени

Модель	Индекс Дайса, %		
	Обучение	Валидация	Тестирование
V-net	96.2±0.8	94.7±1.0	94.5±1.6
V-net (TTA)	96.3±0.8	96.1±0.8	95.9±1.0

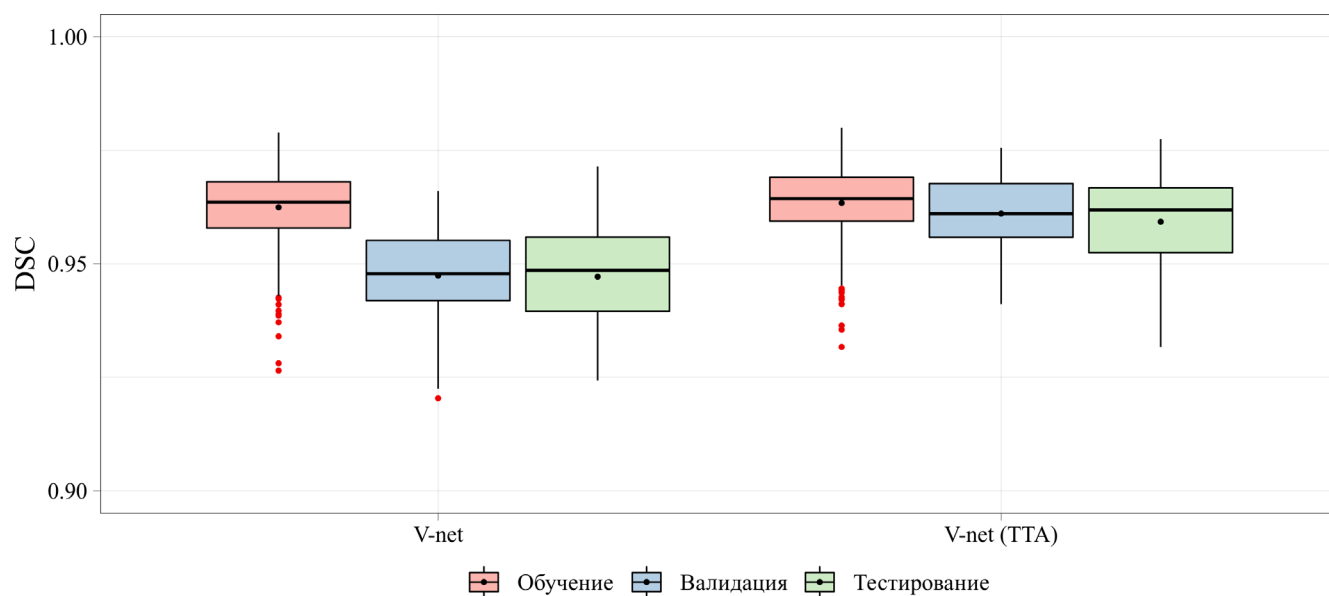


Рисунок 3.38 – Диаграммы размаха точности сегментации печени

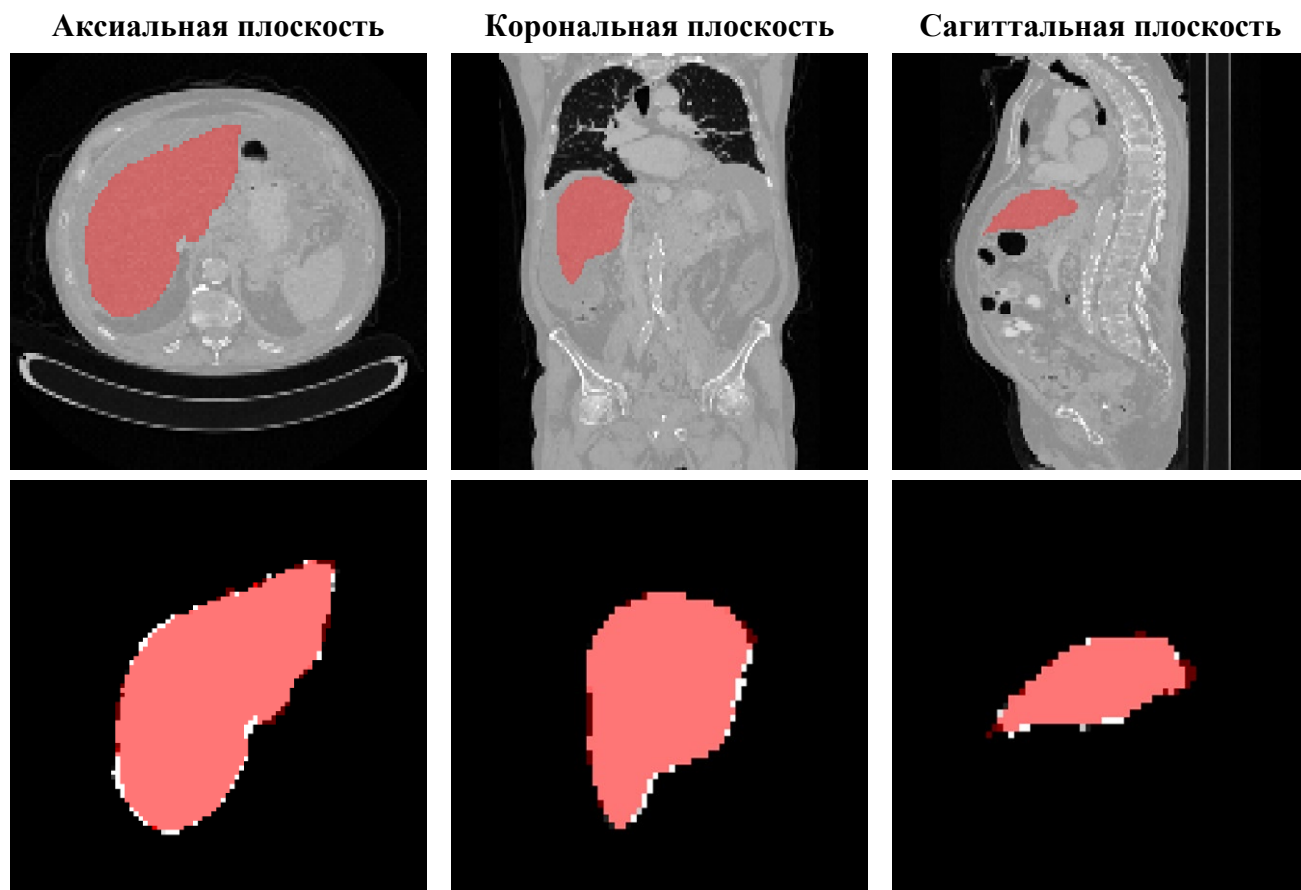


Рисунок 3.39 – Результат сегментации печени на основе модели V-net (TTA)

3.5 Сравнительный анализ

В качестве дополнительной валидации представляет интерес сравнение разработанного подхода сегментации с другими методами. Для этого результаты, полученные на основе разработанной модели V-net, сравниваются с результатами, полученными 11 лучшими группами соревнования «Medical Segmentation Decathlon» [159, 160]. Данное соревнование проходило в рамках конференции «Medical Image Computing and Computer Assisted Intervention (MICCAI)», где оценивалась универсальность алгоритмов машинного обучения применительно к различным задачам семантической сегментации анатомических структур. В качестве исходных данных использовались трёхмерные изображения анатомических структур, описанные ранее в пунктах 3.4.1 – 3.4.5. Краткая характеристика исходных наборов данных для исследуемых анатомических структуры отражена в таблице 3.33. В таблице 3.34 отражена краткая информация о группах, с которыми проводилось сравнение предлагаемого подхода.

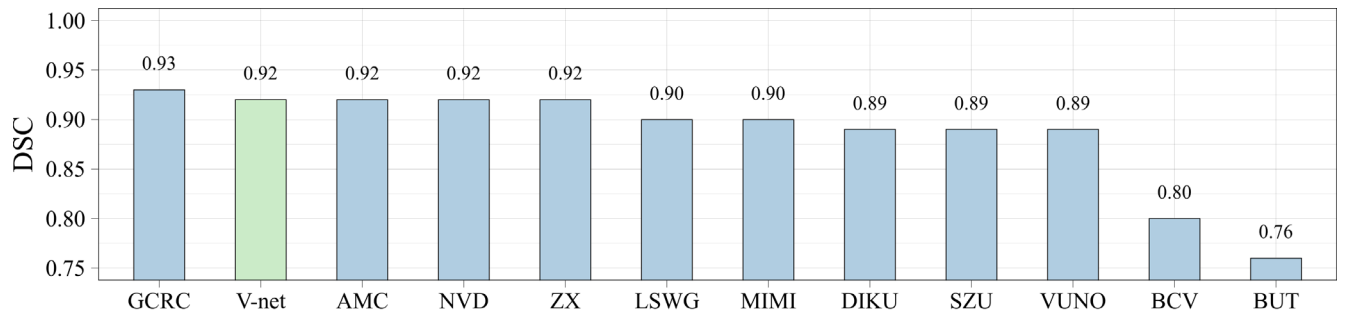
Таблица 3.33 – Краткое описание данных анатомических структур

Анатомическая структура	Модальность	Количество данных	Источник данных	Ссылка
Левое предсердие	MPT	30	King's College London	[161]
Поджелудочная железа	КТ	420	Memorial Sloan Kettering Cancer Center	[162–164]
Селезёнка	КТ	61	Memorial Sloan Kettering Cancer Center	[165]
Гиппокамп	MPT	394	Vanderbilt University Medical Center	—
Печень	КТ	201	IRCAD Hôpitaux Universitaires	[166]

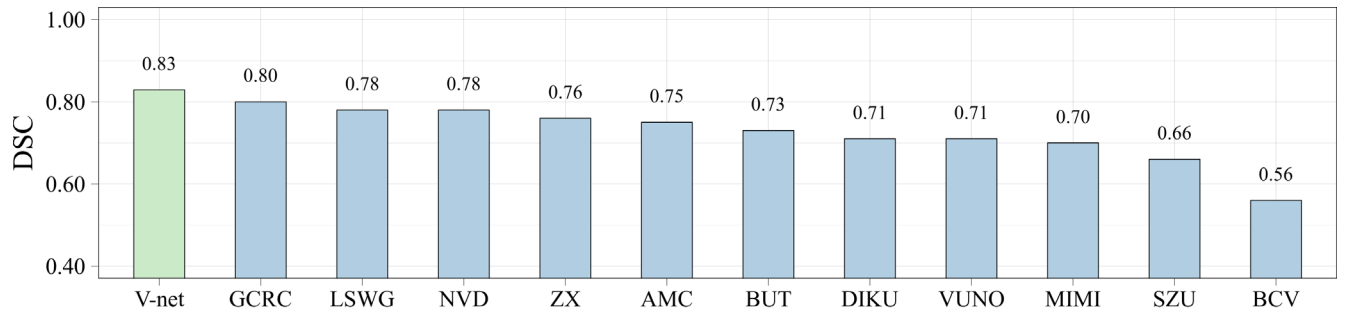
Таблица 3.34 – Описание 11 лучших подходов сегментации на основании результатов соревнования «Medical Segmentation Decathlon»

Ранг	Название модели	Организация	Город, страна	Ссылка
1	GCRC	German Cancer Research Center	Гейдельберг, Германия	[167–169]
2	AMC	Medical Imaging and Intelligent Reality Lab, Asan Medical Center	Сеул, Южная Корея	[170]
3	NVD	Nvidia	Санта-Клара, США	[171, 172]
4	ZX	Tsinghua University	Пекин, Китай	—
5	DIKU	Department of Computer Science, University of Copenhagen	Копенгаген, Дания	[173, 174]
6	LSWG	Lai-Sheng Wang's Group, Xiamen University	Сямынь, Китай	[175]
7	MIMI	Medical Imaging and Intelligent Reality Lab, Asan Medical Center	Сеул, Южная Корея	[170]
8	SZU	Biomedical Engineering Laboratory, Shenzhen University	Шэньчжэнь, Китай	-
9	VUNO	VUNO Inc.	Сеул, Южная Корея	[176]
10	BCV	Department of Biomedical Engineering, Universidad de los Andes	Богота, Колумбия	—
11	BUT	Department of Computer Graphics and Multimedia, Brno University of Technology	Брно, Чехия	—

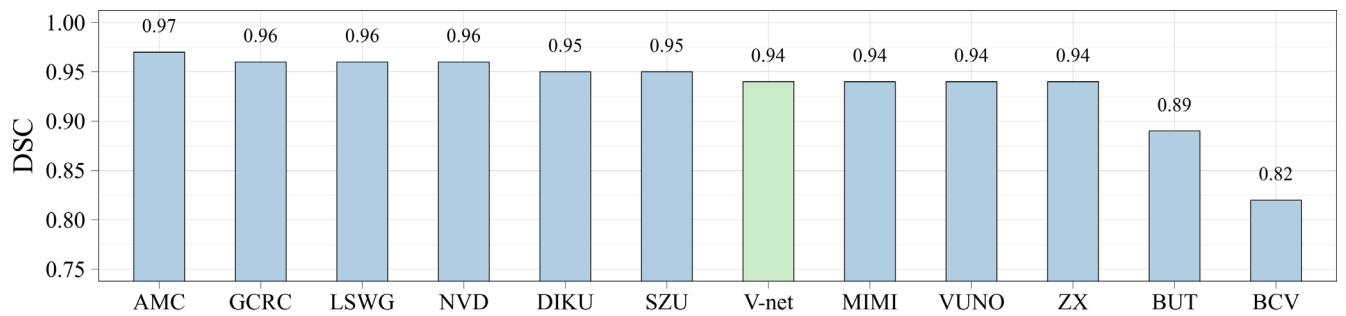
В качестве сравнения результаты сегментации 11 моделей с соревнования «Medical Segmentation Decathlon» сравнивались с результатами, полученными предлагаемой моделью. При сравнении использовалось среднее значение индекса Дайса, которое варьируется в масштабе от 0 до 1. Основные результаты сравнительного исследования продемонстрированы на рисунке 3.40. Модели на данном рисунке отсортированы и представлены в порядке убывания точности. Предлагаемая модель на основе V-net архитектуры отражена зелёным цветом, остальные модели – голубым. Полученные результаты показывают, что предлагаемая модель сегментации на основе V-net архитектуры в большинстве случаев обладает более высокой точностью, чем другие подходы. Немаловажным фактом является то, что сравниваемые подходы в основном базируются на комбинации энкодера и декодера и не используют только свёрточные слои без уменьшения размерности или полносвязные слои. Например, модель GCRC не использует сложные архитектурные модификации и вместо этого базируется на популярной архитектуре U-net. Для универсальной сегментации изображений команда Medical Imaging and Intelligent Reality Lab разработала более сложную модель AMC которая базируется на каскадной архитектуре U-net. Данная архитектура предполагает одновременное обучение модели E-net [177], которая определяет орган или заболевание на медицинских изображениях и модели P-net [178], которая сегментирует орган или заболевание в определённом месте. Модель NVD также базируется на архитектуре U-net с некоторыми правками. Энкодер и декодер модифицированной архитектуры включают ResNet-подобные свёрточные блоки с group нормализацией [179]. Однако, ключевой модификацией являлась архитектура вариационного автоэнкодера, которая использовалась при обучении в целях регуляризации энкодера основной U-net архитектуры. Для поиска оптимальной архитектуры в модели VUNO использовалась методология Neural Architecture Search [176]. Однако базовая модель, которая впоследствии была модифицирована с помощью такой методологии, также базировалась на U-net архитектуре.



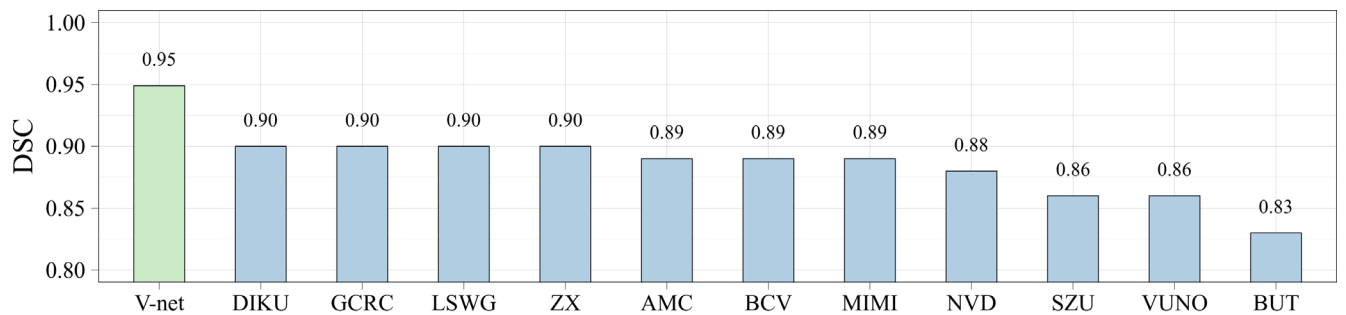
(а) Левое предсердие



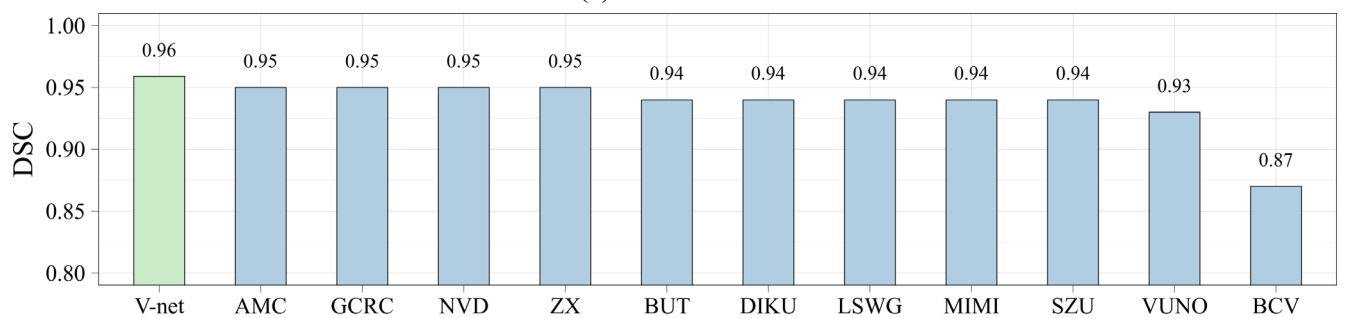
(б) Поджелудочная железа



(в) Селезёнка



(г) Гиппокамп



(д) Печень

Рисунок 3.40 – Сравнение средней точности сегментации разных моделей

3.6 Выводы по третьей главе

1. Разработанный алгоритм синтеза медицинского инструмента внутри анатомических структур позволил автоматизировать процесс разметки данных, а также увеличить точность сегментации за счёт увеличения репрезентативности и гетерогенности обучающей выборки. На основе тестирования стандартной и разреженной U-net архитектур была выявлена тенденция к увеличению точности сегментации при увеличении количества синтетических примеров в обучаемых выборках. Точность сегментации для разреженной U-net архитектуры составила $86.5 \pm 3.6\%$, $88.3 \pm 3.1\%$, $89.8 \pm 3.5\%$ и $92.6 \pm 2.2\%$ при RDR равном 1.00, 0.50, 0.33 и 0.25 соответственно.

2. Предлагаемый алгоритм по выбору оптимальных гиперпараметров на основе t-критерия позволил оценить, насколько базовая модель отличается от сравниваемой. Процедура оценки модели начиналась с оценки типа dropout слоёв, а затем итеративно проходила через остальные параметры, заканчиваясь на оценке коэффициента свёрточных фильтров. В результате была получена модель с оптимальными гиперпараметрами, которые существенно отличались от базовой модели. Важно отметить, что базовая не оптимизированная модель имела точность равную $46.4 \pm 34.6\%$. После проведения оптимизации точность оптимизированной модели составила $90.6 \pm 3.3\%$.

3. После выполнения сравнительного анализа результатов сегментации медицинского инструмента было выявлено, что модели на основе архитектуры FCN и SegNet выполняют сегментацию некачественно, а их выходная точность составила $12.4 \pm 6.4\%$ и $19.0 \pm 8.5\%$ соответственно. Это объясняется рядом причин, а именно недостаточной сложностью моделей, затуханием градиентов при использовании глубоких архитектур и отсутствием дополнительных пробросов признаков. С другой стороны, архитектуры, использующие проброс признаков, позволяют выполнить сегментацию с большей точностью. Так, модель на основе стандартной архитектуры U-net выполнила сегментацию, точность которой составила $80.5 \pm 5.8\%$. Важно отметить, что практически не наблюдается разницы между полноценной трёхмерной и псевдотрёхмерной архитектурами U-net.

Последнее позволяет применять псевдотрёхмерные архитектуры для сегментации при слабой аппаратной составляющей, не теряя в точности. Предложенная разреженная U-net архитектура показала относительно высокие показатели точности при невысокой сложности и легковесности модели. Наибольшую точность показала предлагаемая V-net архитектура, точность которой на обучающей выборке составила $93.8 \pm 2.2\%$, а на тестовой – $90.6 \pm 3.3\%$.

4. В целях оценки универсальности предлагаемых методов и алгоритмов, они были использованы не только для задачи сегментации медицинского инструмента, но и для сегментации таких анатомических структур, как левое предсердие, поджелудочная железа, селезёнка, гиппокамп и печень. Точность сегментации практически всех анатомических структур была выше минимальной пороговой точности, составляющей 85%. Так, точность сегментации левого предсердия составила $92.2 \pm 2.7\%$, поджелудочной железы – $82.9 \pm 5.8\%$, селезёнки – $93.6 \pm 1.6\%$, гиппокампа – $94.9 \pm 1.5\%$, печени – $95.9 \pm 1.0\%$.

5. При внедрении алгоритма test-time аугментации точность сегментации медицинского инструмента увеличилась в среднем на 3%. При использовании стандартной V-net архитектуры точность сегментации инструмента на тестовой выборке составила $90.6 \pm 3.3\%$, в то время как V-net архитектура с использованием ТТА выполнила процедуру сегментации с точностью равной $93.6 \pm 2.4\%$. На основании результатов, алгоритм ТТА уменьшает разброса индекса Дайса, а также повышает его среднее значение. Наиболее успешным примером использования алгоритма ТТА является случай сегментации поджелудочной железы. Так, точность сегментации стандартной V-net модели на тестовой выборке составила $68.6 \pm 11.6\%$, в то время как использование ТТА дало значительный прирост к средней точности и уменьшило стандартное отклонение. Результирующая точность сегментации такой модели составила $82.9 \pm 5.8\%$. Аналогичная ситуация наблюдается с сегментацией селезёнки, где использование ТТА увеличило точность в среднем на 8%. Индекс Дайса на тестовой выборке данной анатомической структуры составил $85.4 \pm 8.0\%$ и $93.6 \pm 1.6\%$ соответственно для V-net и V-net (ТТА) моделей.

6. Сравнительное исследование между предлагаемой V-net моделью и 11 лучшими моделями из соревнований «Medical Segmentation Decathlon» показало, что в большинстве случаев предлагаемая модель обладает большей точностью, чем сравниваемые подходы. Исключением стал случай с сегментацией селезёнки. В рейтинге предлагаемый подход находится на 7 месте, при этом точность сегментации находится на высоком уровне и составляет $93.6 \pm 1.6\%$. Другим важным фактом, который необходимо отметить, является значительное доминирование U-net архитектуры в сфере семантической сегментации. Большинство моделей, которые используются сегодня, в том числе и предлагаемая, базируются на концепции U-net.

ЗАКЛЮЧЕНИЕ

Оценка современного состояния сферы информационных технологий в медицине показала, что из-за слабой развитости средств трёхмерной визуализации большой спектр реконструктивных вмешательств осуществляется только на открытом «сухом» сердце с применением искусственного кровообращения. Такой способ проведения вмешательств обладает множеством послеоперационных осложнений, что остаётся проблемой. Несмотря на то, что существуют программно-аппаратные решения аналогичные CARTO3, их применение не всегда становится возможным из-за громоздкости, сложности, дороговизны, а также узкой сферы применения. В подавляющем большинстве операционные комнаты оборудованы флюороскопией и/или эхокардиографией. В связи с этим, использование современных средств обработки и визуализации данных на основании этих модальностей становится достаточным и эффективным. Исходя из этого, существует потребность в развитии методов, моделей и алгоритмов, позволяющих выделить медицинский инструмент и анатомические структуры на трёхмерных данных. Данная потребность объясняется низкой надёжностью и точностью существующих средств трёхмерной визуализации области вмешательства, а также отсутствием обратной связи между средствами визуализации и средствами позиционирования медицинского инструмента.

Для решения задачи сегментации области интереса предлагается использовать метод глубокого обучения на основе свёрточных нейронных сетей. Последние зарекомендовали себя в качестве эффективного инструмента при решении задач, связанных с обработкой данных. При обучении нейронные сети самостоятельно проводят процедуру выбора оптимальных признаков, что эффективно решает поставленную задачу. В качестве основной архитектуры нейронной сети предлагается использовать свёрточную нейронную сеть V-net с плотными пробросами признаков, где основные закодированные признаки находятся на нижнем уровне сети. Для более эффективного декодирования признаков и получения финальной сегментации используются обучаемые слои обратной свёртки, а не стандартные слои повышающей дискретизации.

Дополнительно стоит отметить внедрённые в архитектуру сети пробросы признаков. Данные пробросы позволяют не затухать градиентам при обратном прохождении метода обратного распространения ошибки и эффективнее минимизировать функцию потерь.

С целью эффективного обучения нейронных сетей были внедрены и использованы алгоритм циклической скорости обучения и алгоритм изменения размера батча. В случае остроконечных и невыпуклых ландшафтов функции потерь метод обратного распространения ошибки потенциально имеет более высокую вероятность попадания в зону локального минимума и седловых точек, что негативно скажется на точности и обобщающей способности модели. Подобные алгоритмы позволили нивелировать проблему, связанную с подобными случаями. Стоит также отметить, что в работе использовалась функция потерь, основанная на бинарной мере сходства Дайса. В отличие от кросс-энтропии, используемая функция потерь эффективнее справляется с сегментацией небольших объектов. Так, например, размер медицинского инструмента (катетера) в исследуемом наборе данных составляет 1415 ± 529 вокселей из 2 097 152, что представляет примерно 0.07% всего изображения.

В качестве оптимизации гиперпараметров модели использовался алгоритм, основанный на оценке t-критерия Стьюдента для связанных выборок. В отличие от таких подходов, как Grid Search или Randomized Search, предлагаемый алгоритм учитывает динамику на прошлых итерациях и не проходит по всему подмножеству пространства гиперпараметров. Согласно данному алгоритму, чем больше t-статистика отклоняется от нуля, тем меньше вероятность того, что базовая и сравниваемая модели имеют одинаковые результаты. Ярким примером является случай оптимизации коэффициента свёрточных фильтров. Несмотря на то, что средняя точность при коэффициенте равном 8 выше (0.91 ± 0.04), чем точность при коэффициенте равном 4 (0.90 ± 0.04), значение t-критерия данных моделей показывает, что они отличаются не значительно. Аналогичная ситуация наблюдается в случае с выбором оптимального типа активации, где функция

PReLU была отклонена из-за того, что она увеличивает сложность модели практически вдвое, при этом, не увеличивая её точность.

Несмотря на сравнительно высокие показатели точности на тестовой выборке, дополнительно был внедрён алгоритм test-time аугментации. Данный алгоритм производит предсказание не по одному примеру, а по N аугментированных примеров. В случае если модель слабо выполняет предсказания на новых данных, использование test-time аугментации повышает шансы модели на более корректное прогнозирование, вследствие чего увеличивая точность сегментации. Внедрённый алгоритм test-time аугментации увеличил среднюю точность предсказания на тестовой выборке данных медицинского инструмента на 3%. Схожая ситуация наблюдалась при сегментации селезёнки, где точность увеличилась в среднем на 8%. Наиболее эффективно test-time аугментация показала себя при сегментации поджелудочной железы, где точность сегментации выросла с $68.6 \pm 11.6\%$ до $82.9 \pm 5.8\%$.

Предлагаемая V-net архитектура с плотными пробросами признаков сравнивалась с некоторыми классическими архитектурами сегментации, такими как FCN, SegNet, Deep Medic и U-net. Важно отметить, что использование таких архитектур как FCN и SegNet не дало хороших результатов. В свою очередь, архитектуры на основе U-net, в том числе и псевдотрёхмерная U-net архитектура, способны выполнить сегментацию качественно. Стандартная U-net архитектура выполнила сегментацию катетера с точностью равной $80.5 \pm 5.8\%$. Однако наибольшую точность показала предлагаемая V-net архитектура, точность которой на обучающей выборке составила $93.8 \pm 2.2\%$, а на тестовой – $90.6 \pm 3.3\%$. Дополнительное внедрение алгоритма test-time аугментации увеличило точность V-net архитектуры на тестовой выборке с $90.6 \pm 3.3\%$ до $93.6 \pm 2.4\%$. Представленная архитектура V-net с плотным пробросом признаков показала высокие результаты сегментации анатомических структур по сравнению с 11 лучшими моделями соревнования «Medical Segmentation Decathlon». Таким образом, точность сегментации левого предсердия составила $92.2 \pm 2.7\%$, поджелудочной железы – $82.9 \pm 5.8\%$, селезёнки – $93.6 \pm 1.6\%$, гиппокампа – $94.9 \pm 1.5\%$, печени – $95.9 \pm 1.0\%$.

СПИСОК СОКРАЩЕНИЙ И ОПРЕДЕЛЕНИЙ

Минимально инвазивная хирургия	– направление хирургии, предусматривающее минимизацию повреждений органов и тканей, которые обусловлены операционным вмешательством
Сегментация	– процедура присвоения меток каждому пикселю изображения таким образом, что пиксели с одинаковыми метками имеют общие визуальные характеристики и образуют сегменты
Каскадные алгоритмы	– класс алгоритмов, которые базируются на использовании множества последовательно идущих блоков обработки данных
Обобщающая способность	– способность модели к распространению обнаруженных на обучающем множестве зависимостей и закономерностей на новые данные
U-net	– архитектура свёрточной нейронной сети, которая содержит сжимающий путь (энкодер), расширяющий путь (декодер) и пропуски признаков с энкодера в декодер
Pseudo 3D U-net	– архитектура свёрточной нейронной сети, аналогичная архитектуре U-net, выполняющая процесс сегментации послайсово, в отличие от полноценной трёхмерной сегментации изображений, выполняемой U-net моделью
Dilated U-net	– архитектура свёрточной нейронной сети, аналогичная архитектуре U-net, использующая разреженное рецептивное окно
V-net	– предлагаемая архитектура свёрточной нейронной сети, которая содержит сжимающий путь (энкодер),

расширяющий путь (декодер), дополнительный двойной свёрточный блок между энкодером и декодером, пропуски признаков с энкодера в декодер и пропуски признаков внутри энкодера и декодера

- | | |
|---|--|
| Батч (от англ. batch) | – группа определённого числа объектов (изображений), которая используется при обучении нейронной сети |
| Скорость обучения
(от англ. learning rate, LR) | – параметр градиентного алгоритма обучения нейронных сетей, который позволяет управлять величиной коррекции весов на каждой итерации |
| TTA (сокр. от Test-Time Аугментация) | – алгоритм аугментации данных, использующий ряд трансформированных экземпляров для предсказания на отложенной выборке |
| Гиперпараметры | – параметры, значения которых задаются до начала обучения модели и не оптимизируются/изменяются в процессе обучения |
| RDR (сокр. от Real Data Ratio) | – параметр, отражающий соотношение реальных и искусственных экземпляров в выборке |

СПИСОК ЛИТЕРАТУРЫ

1. Audebert N., Le Saux B., Lefèvre S. Segment-before-Detect: Vehicle Detection and Classification through Semantic Segmentation of Aerial Images // Remote Sens. 2017. Vol. 9, № 4. P. 368.
2. Alvarez J.M. et al. Road Scene Segmentation from a Single Image // Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics). 2012. Vol. 7578 LNCS, № PART 7. P. 376–389.
3. Alvarez J.M. et al. Semantic Road Segmentation via Multi-scale Ensembles of Learned Features // Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics). 2012. Vol. 7584 LNCS, № PART 2. P. 586–595.
4. Zhang W., Mahale T. End to End Video Segmentation for Driving : Lane Detection For Autonomous Car. 2018.
5. Lyu Y., Huang X. Road Segmentation Using CNN with GRU. 2018.
6. Xia F. et al. Joint Multi-person Pose Estimation and Semantic Part Segmentation // 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). IEEE, 2017. Vol. 2017-Janua. P. 6080–6089.
7. Aslanzadeh R., Qazanfari K., Rahmati M. An Efficient Evolutionary Based Method For Image Segmentation. 2017. P. 1–17.
8. Weiss D., Taskar B. SCALPEL: Segmentation Cascades with Localized Priors and Efficient Learning // 2013 IEEE Conference on Computer Vision and Pattern Recognition. IEEE, 2013. № Figure 1. P. 2035–2042.
9. Yang H. et al. Catheter Detection in 3D Ultrasound Using Triplanar-Based Convolutional Neural Networks // 2018 25th IEEE International Conference on Image Processing (ICIP). IEEE, 2018. P. 371–375.

10. Fazlali H.R. et al. Vessel segmentation and catheter detection in X-ray angiograms using superpixels // *Med. Biol. Eng. Comput. Medical & Biological Engineering & Computing*, 2018. Vol. 56, № 9. P. 1515–1530.
11. Horowitz S.L., Pavlidis T. Picture Segmentation by a Tree Traversal Algorithm // *J. ACM*. 1976. Vol. 23, № 2. P. 368–388.
12. Haralick R.M., Shapiro L.G. Image segmentation techniques // *Comput. Vision, Graph. Image Process.* 1985. Vol. 29, № 1. P. 100–132.
13. Lloyd S. Least squares quantization in PCM // *IEEE Trans. Inf. Theory*. 1982. Vol. 28, № 2. P. 129–137.
14. Dempster A.P., Laird N.M., Rubin D.B. Maximum Likelihood from Incomplete Data Via the EM Algorithm // *J. R. Stat. Soc. Ser. B*. 1977.
15. Comaniciu D., Meer P. Mean shift analysis and applications // *Proceedings of the IEEE International Conference on Computer Vision*. 1999.
16. Comaniciu D., Meer P. Mean shift: A robust approach toward feature space analysis // *IEEE Trans. Pattern Anal. Mach. Intell.* 2002.
17. Meyer F. Topographic distance and watershed lines // *Signal Processing*. 1994. Vol. 38, № 1. P. 113–125.
18. Krizhevsky A., Sutskever I., Hinton G.E. ImageNet classification with deep convolutional neural networks // *Advances in Neural Information Processing Systems*. 2012.
19. Long J., Shelhamer E., Darrell T. Fully Convolutional Networks for Semantic Segmentation [Electronic resource]. 2016. URL: <https://github.com/shelhamer/fcn.berkeleyvision.org>.
20. Liu W., Rabinovich A., Berg A.C. ParseNet: Looking Wider to See Better. 2015. P. 1–11.

21. Noh H., Hong S., Han B. Learning deconvolution network for semantic segmentation // Proc. IEEE Int. Conf. Comput. Vis. 2015. Vol. 2015 Inter. P. 1520–1528.
22. Ronneberger O., Fischer P., Brox T. U-Net: Convolutional Networks for Biomedical Image Segmentation. 2015. P. 234–241.
23. Lin T.-Y. et al. Feature Pyramid Networks for Object Detection // Proc. - 30th IEEE Conf. Comput. Vis. Pattern Recognition, CVPR 2017. 2016. Vol. 2017-Janua. P. 936–944.
24. Ren S. et al. Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks // IEEE Trans. Pattern Anal. Mach. Intell. 2017.
25. Girshick R. Fast R-CNN // Proc. IEEE Int. Conf. Comput. Vis. 2015. Vol. 2015 Inter. P. 1440–1448.
26. Girshick R. et al. Region-Based Convolutional Networks for Accurate Object Detection and Segmentation // IEEE Trans. Pattern Anal. Mach. Intell. 2016. Vol. 38, № 1. P. 142–158.
27. Zhao H. et al. Pyramid scene parsing network // Proc. - 30th IEEE Conf. Comput. Vis. Pattern Recognition, CVPR 2017. 2017. Vol. 2017-Janua. P. 6230–6239.
28. He K. et al. Mask R-CNN // Proc. IEEE Int. Conf. Comput. Vis. 2017. Vol. 2017-Octob. P. 2980–2988.
29. Chen L.-C. et al. Rethinking Atrous Convolution for Semantic Image Segmentation. 2017.
30. Chen L.C. et al. Encoder-decoder with atrous separable convolution for semantic image segmentation // Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics). 2018. Vol. 11211 LNCS. P. 833–851.
31. Liu S. et al. Path Aggregation Network for Instance Segmentation // Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. 2018. P. 8759–8768.

32. Zhang H. et al. Context Encoding for Semantic Segmentation // Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. 2018. P. 7151–7160.
33. Rother C., Kolmogorov V., Blake A. GrabCut - Interactive foreground extraction using iterated graph cuts // ACM SIGGRAPH 2004 Pap. SIGGRAPH 2004. 2004. P. 309–314.
34. The Analyze data format [Electronic resource]. URL: <http://imaging.mrc-cbu.cam.ac.uk/imaging/FormatAnalyze> (accessed: 31.08.2016).
35. ANALYZE 7.5 File Format [Electronic resource]. 2015. URL: <https://portal.mayo.edu/bir/> (accessed: 27.08.2016).
36. Larobina M., Murino L. Medical image file formats // J. Digit. Imaging. 2014. Vol. 27, № 2. P. 200–206.
37. DICOM PS3.5 2016c - Data Structures and Encoding [Electronic resource].
38. NIfTI-1 Data Format — Neuroimaging Informatics Technology Initiative [Electronic resource]. URL: <http://nifti.nimh.nih.gov/nifti-1> (accessed: 31.08.2016).
39. Абдулракеб А.Р.А., Сушкова Л.Т., Лозовская Н.А. Обзор методов сегментации опухолей на МРТ-изображениях головного мозга // Casp. J. Manag. High Technol. 2015. Т. 1, № 29. С. 122–138.
40. Selvaraj Assley P.S.B., Chellakkon H.S. A Comparative Study on Medical Image Segmentation Methods // Appl. Med. Informatics. 2014. Vol. 34, № 1. P. 31–45.
41. Zipes D.P. et al. ACC/AHA/ESC 2006 Guidelines for Management of Patients With Ventricular Arrhythmias and the Prevention of Sudden Cardiac Death // J. Am. Coll. Cardiol. 2006. Vol. 48, № 5. P. 746–837.
42. Iaizzo P.A. Handbook of Cardiac Anatomy, Physiology, and Devices / ed. Iaizzo P.A. Totowa, NJ: Humana Press, 2009. 659 p.

43. Cao K., Mills D., Patwardhan K.A. Automated catheter detection in volumetric ultrasound // 2013 IEEE 10th International Symposium on Biomedical Imaging. IEEE, 2013. P. 37–40.
44. Uherčík M. et al. Line filtering for surgical tool localization in 3D ultrasound images // Comput. Biol. Med. Pergamon, 2013. Vol. 43, № 12. P. 2036–2045.
45. Pourtaherian A. et al. Medical Instrument Detection in 3-Dimensional Ultrasound Data Volumes // IEEE Trans. Med. Imaging. 2017. Vol. 36, № 8. P. 1664–1675.
46. Barva M. et al. Parallel integral projection transform for straight electrode localization in 3-D ultrasound images // IEEE Trans. Ultrason. Ferroelectr. Freq. Control. 2008. Vol. 55, № 7. P. 1559–1569.
47. Uherčík M. et al. Model Fitting Using RANSAC for Surgical Tool Localization in 3-D Ultrasound Images // IEEE Trans. Biomed. Eng. 2010. Vol. 57, № 8. P. 1907–1916.
48. Neshat H.R.S., Patel R. V. Real-time parametric curved needle segmentation in 3D ultrasound images // 2008 2nd IEEE RAS & EMBS International Conference on Biomedical Robotics and Biomechatronics. IEEE, 2008. P. 670–675.
49. Tepeler A. et al. Factors affecting the fluoroscopic screening time during percutaneous nephrolithotomy // J Endourol. 2009. Vol. 23, № 11. P. 1825–1829.
50. Hoffmann M. et al. Semi-automatic catheter reconstruction from two views. // Med. image Comput. Comput. Interv. 2012. Vol. 15, № Pt 2. P. 584–591.
51. Hoffmann M. et al. Electrophysiology catheter detection and reconstruction from two views in fluoroscopic images // IEEE Trans. Med. Imaging. 2016. Vol. 35, № 2. P. 567–579.
52. Bock M. et al. Active catheter tracking using parallel MRI and real-time image reconstruction // Magn. Reson. Med. 2006. Vol. 55, № 6. P. 1454–1459.
53. Baert S.A.M. et al. Three-Dimensional Guide-Wire Reconstruction from Biplane

- Image Sequences for Integrated Display in 3-D Vasculature // IEEE Trans. Med. Imaging. 2003. Vol. 22, № 10. P. 1252–1258.
54. Ma Y. et al. Real-time x-ray fluoroscopy-based catheter detection and tracking for cardiac electrophysiology interventions // Med. Phys. 2013. Vol. 40, № 7. P. 071902.
 55. Lu X., Chen T., Comaniciu D. Robust discriminative wire structure modeling with application to stent enhancement in fluoroscopy // CVPR 2011. IEEE, 2011. P. 1121–1127.
 56. Wang P. et al. Image-Based Device Tracking for the Co-registration of Angiography and Intravascular Ultrasound Images // Medical image computing and computer-assisted intervention : MICCAI ... International Conference on Medical Image Computing and Computer-Assisted Intervention. 2011. Vol. 14, № 1. P. 161–168.
 57. Biosense Webster. The CARTOSOUND module with SOUNDSTAR catheter [Electronic resource]. URL: <https://www.biosensewebster.com/products/cartosound-3/cartosound-module.aspx>.
 58. Ma Y. et al. Hybrid echo and x-ray image guidance for cardiac catheterization procedures by using a robotic arm: a feasibility study // Phys. Med. Biol. 2010. Vol. 55, № 13. P. N371–N382.
 59. Lang P. et al. US–Fluoroscopy Registration for Transcatheter Aortic Valve Implantation // IEEE Trans. Biomed. Eng. 2012. Vol. 59, № 5. P. 1444–1453.
 60. Ralovich K. et al. 6DoF Catheter Detection, Application to Intracardiac Echocardiography // Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics). 2014. Vol. 8674 LNCS, № PART 2. P. 635–642.
 61. Frangi A.F. et al. Multiscale vessel enhancement filtering // Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and

- Lecture Notes in Bioinformatics). Springer Verlag, 1998. Vol. 1496. P. 130–137.
62. Yang H. et al. Improving Catheter Segmentation & Localization in 3D Cardiac Ultrasound Using Direction-Fused FCN // Proc. - Int. Symp. Biomed. Imaging. 2019. P. 1122–1126.
 63. Lim K.H., Yeo J.H., Duran C.M.G. Three-Dimensional Asymmetrical Modeling of the Mitral Valve: A Finite Element Study with Dynamic Boundaries // J. Heart Valve Dis. 2005. Vol. 14, № 3. P. 386–392.
 64. Levine R.A. et al. Three-dimensional echocardiographic reconstruction of the mitral valve, with implications for the diagnosis of mitral valve prolapse. // Circulation. 1989. Vol. 80, № 3. P. 589–598.
 65. Avendi M.R., Kheradvar A., Jafarkhani H. A Combined Deep-Learning and Deformable-Model Approach to Fully Automatic Segmentation of the Left Ventricle in Cardiac MRI // Med. Image Anal. Elsevier B.V., 2015. Vol. 30. P. 1–34.
 66. Nillesen M.M. et al. Automated Assessment of Right Ventricular Volumes and Function Using Three-dimensional Transesophageal Echocardiography. // Ultrasound Med. Biol. 2015. Vol. 42, № 2. P. 1–11.
 67. Tandon A. et al. Use of a semi-automated cardiac segmentation tool improves reproducibility and speed of segmentation of contaminated right heart magnetic resonance angiography // Int. J. Cardiovasc. Imaging. 2016. Vol. 32, № 8. P. 1273–1279.
 68. Chen Y., Broschat S.L., Flynn P.J. Phase insensitive homomorphic image processing for speckle reduction. // Ultrason. Imaging. 1996. Vol. 18, № 2. P. 122–139.
 69. Sanchez J.R., Oelze M.L. An ultrasonic imaging speckle-suppression and contrast-enhancement technique by means of frequency compounding and coded excitation // IEEE Trans. Ultrason. Ferroelectr. Freq. Control. 2009. Vol. 56, № 7.

P. 1327–1339.

70. Yue Y. et al. Nonlinear multiscale wavelet diffusion for speckle suppression and edge enhancement in ultrasound images // IEEE Trans. Med. Imaging. 2006. Vol. 25, № 3. P. 297–311.
71. Schneider R. HeartModel Removing the complexity of Live 3D Quantification. 2015. 2–5 p.
72. Hung J. et al. 3D Echocardiography: A Review of the Current Status and Future Directions // J. Am. Soc. Echocardiogr. 2007. Vol. 20, № 3. P. 213–233.
73. Paragios N. et al. Active shape models and segmentation of the left ventricle in echocardiography // Scale Space and PDE Methods in Computer Vision. 2005. № November 1999. P. 131–142.
74. Angelini E.D. et al. Segmentation of real-time three-dimensional ultrasound for quantification of ventricular function: a clinical study on right and left ventricles. // Ultrasound Med. Biol. 2005. Vol. 31, № 9. P. 1143–1158.
75. Bersvendsen J. et al. Automated Segmentation of the Right Ventricle in 3D Echocardiography: A Kalman Filter State Estimation Approach // IEEE Trans. Med. Imaging. 2015. Vol. 0062, № c. P. 1–1.
76. Lang R.M. et al. EAE/ASE recommendations for image acquisition and display using three-dimensional echocardiography // Eur. Heart J. Cardiovasc. Imaging. 2012. Vol. 13, № 1. P. 1–46.
77. Petitjean C. et al. Right ventricle segmentation from cardiac MRI: A collation study // Med. Image Anal. Elsevier B.V., 2015. Vol. 19, № 1. P. 187–202.
78. Rajan N.K. et al. Automated Three-Dimensional Reconstruction of the Left Ventricle From Multiple-Axis Echocardiography. // J. Biomech. Eng. 2016. Vol. 138, № 1.
79. Chen S.J. et al. Usefulness of 3D reconstructed computed tomography imaging for

- double outlet right ventricle // J. Formos. Med. Assoc. 2008. Vol. 107, № 5. P. 371–380.
80. Ryan J. Pericardial/Myocardial Disease/Pulmonary Hypertension // J. Am. Coll. Cardiol. American College of Cardiology Foundation, 2013. Vol. 61, № 10. P. E1295.
 81. Valocik G., Kamp O., Visser C.A. Three-dimensional echocardiography in mitral valve disease. // Eur. J. Echocardiogr. 2005. Vol. 6, № 6. P. 443–454.
 82. Tremeau A., Borel N. A region growing and merging algorithm to color segmentation // Pattern Recognit. 1997. Vol. 30, № 7. P. 1191–1203.
 83. Junfeng D., Yunyang Y. The Fast Medical Image Segmentation of Target Region Based on Improved FM Algorithm // Procedia Eng. Int. Work. Inf. Electron. Eng. 2012. Vol. 29. P. 48–52.
 84. Manousakas I.N. et al. Split-and-merge segmentation of magnetic resonance medical images: performance evaluation and extension to three dimensions. // Comput. Biomed. Res. 1998. Vol. 31, № 6. P. 393–412.
 85. Cabezas M. et al. A review of atlas-based segmentation for magnetic resonance brain images // Comput. Methods Programs Biomed. 2011. Vol. 104, № 3. P. 158–177.
 86. Aljabar P. et al. Multi-atlas based segmentation of brain images: Atlas selection and its effect on accuracy // Neuroimage. 2009. Vol. 46, № 3. P. 726–738.
 87. Klein A. et al. Mindboggle: automated brain labeling with multiple atlases // BMC Med. Imaging. 2005. Vol. 5, № 1.
 88. Park H., Bland P.H., Meyer C.R. Construction of an Abdominal Probabilistic Atlas and its Application in Segmentation // IEEE Trans. Med. Imaging. 2003. Vol. 22, № 4.
 89. Park H. et al. Construction of Abdominal Probabilistic Atlases and Their Value in

Segmentation of Normal Organs in Abdominal CT Scans. 2010. № 8.

90. Зубов А.Ю., Сенюкова О.В. Сегментация изображений магнитно-резонансной томографии головного мозга с помощью сопоставления с несколькими атласами // Юбилейная 25-а Международная конференция GraphiCon 2015. 2015. С. 56–61.
91. Candemir S. et al. Lung segmentation in chest radiographs using anatomical atlases with nonrigid registration // IEEE Trans. Med. Imaging. 2014. Vol. 33, № 2. P. 577–590.
92. Lai M., Rueckert D. Deep Learning for Medical Image Segmentation. 2015.
93. Kass M., Witkin A., Terzopoulos D. Snakes: Active contour models // International Journal of Computer Vision. 1988. Vol. 1, № 4. P. 321–331.
94. Landgren M. et al. An Automated System for the Detection and Diagnosis of Kidney Lesions in Children from Scintigraphy Images // Image Analysis. 2011. P. 489–500.
95. Landgren M., Overgaard N.C., Heyden A. Segmentation of the Left Heart Ventricle in Ultrasound Images Using a Region Based Snake // Proceedings of SPIE / ed. Ourselin S., Haynor D.R. 2013.
96. Felzenszwalb P.F., Huttenlocher D.P. Efficient graph-based image segmentation // Int. J. Comput. Vis. 2004. Vol. 59, № 2. P. 167–181.
97. Steinhaus H. Sur la division del corps matériels en parties // Bulletin de l'Académie Polonaise del Sciences - Classe III. 1956. Vol. IV, № 12. 801–804 p.
98. Meena A., Raja K. Segmentation of Alzheimer's disease in PET Scan Datasets using Matlab // Int. J. Comput. Appl. 2012. Vol. 57, № 10. P. 15–19.
99. Ng H.P. et al. Medical Image Segmentation Using K-Means Clustering and Improved Watershed Algorithm // 2006 IEEE Southwest Symposium on Image Analysis and Interpretation. IEEE, 2006. P. 61–65.

100. Philips. QLAB Ultrasound Cardiac Analysis | Philips Ultrasound [Electronic resource]. URL: <https://www.usa.philips.com/healthcare/product/HCNOCTN145/qlab-cardiac-analysis-cardiovascular-ultrasound-quantification-software.html> (accessed: 20.08.2017).
101. Siemens. Syngo fastView [Electronic resource]. URL: <https://www.healthcare.siemens.com/medical-imaging-it/syngo-special-topics/syngo-fastview> (accessed: 20.08.2017).
102. General Electric. Innova EPVision 2.0 [Electronic resource]. URL: http://www3.gehealthcare.com/en/education/product_education_clinical/interventional_systems/invasive/innova_epvision_2-0_training (accessed: 20.08.2017).
103. Zeiss. TRENION 3D HD – Visualization Systems - Spine Surgery - Medical Technology | ZEISS International [Electronic resource]. URL: <https://www.zeiss.com/meditec/int/products/spine-surgery/visualization-systems/trenion-3d-hd.html> (accessed: 20.08.2017).
104. Schneider R.J. et al. Patient-specific mitral leaflet segmentation from 4D ultrasound. // Med. Image Comput. Comput. Assist. Interv. NIH Public Access, 2011. Vol. 14, № Pt 3. P. 520–527.
105. Schneider R.J. Mitral Valve Segmentation from Clinical 4D Ultrasound [Electronic resource]. URL: <http://biorobotics.harvard.edu/schneider.html> (accessed: 25.04.2019).
106. Novotny P. 3D Ultrasound Image Processing and Instrument Tracking [Electronic resource]. URL: <http://biorobotics.harvard.edu/imageprocessing.html> (accessed: 25.04.2019).
107. Dowsey A.W., Merrifield R.D., Guang-Zhong Yang. Adaptive cardiovascular imaging: challenges and opportunities for real-time processing // IEEE Signal Process. Mag. 2006. Vol. 23, № 5. P. 112–116.

108. Yang G.-Z. et al. Flow and myocardial interaction: an imaging perspective. // Philos. Trans. R. Soc. Lond. B. Biol. Sci. The Royal Society, 2007. Vol. 362, № 1484. P. 1329–1341.
109. Insight Software Consortium. ITK - Segmentation & Registration Toolkit [Electronic resource]. URL: <https://itk.org/itkindex.html>.
110. König A.H. Usability Issues in 3D Medical Visualization. Technische Universität Wien, 2011. 122 p.
111. Solteszova V. et al. Output-Sensitive Filtering of Streaming Volume Data. // Comput. Graph. Forum. 2017. Vol. 36, № 1. P. 249–262.
112. Ting W.S., Watanabe L. de S., Vidalón J.E.Y. 3D Medical Data Visualization Toolkit [Electronic resource]. URL: http://www.dca.fee.unicamp.br/courses/IA369E/2s2010/projects/vidalon_watanabe/index.htm.
113. Baer A. Perception Guided Evaluation of 3D Medical Visualizations. Otto-von-Guericke University Magdeburg, 2015.
114. German Cancer Research Center. The Medical Imaging Interaction Toolkit (MITK) [Electronic resource]. URL: [http://mitk.org/wiki/The_Medical_Imaging_Interaction_Toolkit_\(MITK\)](http://mitk.org/wiki/The_Medical_Imaging_Interaction_Toolkit_(MITK)) (accessed: 16.04.2017).
115. Harvard Medical School. 3D Slicer [Electronic resource]. URL: <https://www.slicer.org/> (accessed: 16.04.2017).
116. Yushkevich P.A., Gerig G. ITK-SNAP.
117. Clark J.D. et al. The 1996 Guide for the Care and Use of Laboratory Animals // ILAR J. 1997. Vol. 38, № 1. P. 41–48.
118. Kolpashchikov D. et al. FABRIK-Based Inverse Kinematics For Multi-Section Continuum Robots // Proceedings of the 2018 18th International Conference on

Mechatronics. Brno, Czech Republic: IEEE, 2018. P. 1–8.

119. Long J., Shelhamer E., Darrell T. Fully convolutional networks for semantic segmentation // 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). IEEE, 2015. P. 3431–3440.
120. Badrinarayanan V., Kendall A., Cipolla R. SegNet: A Deep Convolutional Encoder-Decoder Architecture for Image Segmentation // IEEE Trans. Pattern Anal. Mach. Intell. 2017. Vol. 39, № 12. P. 2481–2495.
121. Chen L.C. et al. DeepLab: Semantic Image Segmentation with Deep Convolutional Nets, Atrous Convolution, and Fully Connected CRFs // IEEE Trans. Pattern Anal. Mach. Intell. 2018. Vol. 40, № 4. P. 834–848.
122. Kamnitsas K. et al. DeepMedic for Brain Tumor Segmentation. 2016. P. 138–149.
123. Jiang J. et al. RedNet : Residual Encoder-Decoder Network for. P. 1–14.
124. Iglovikov V., Shvets A. TerausNet: U-Net with VGG11 Encoder Pre-Trained on ImageNet for Image Segmentation. 2018.
125. Chaurasia A., Culurciello E. LinkNet: Exploiting encoder representations for efficient semantic segmentation // 2017 IEEE Visual Communications and Image Processing (VCIP). IEEE, 2017. P. 1–4.
126. Ioffe S., Szegedy C. Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift. 2015.
127. Yu F., Koltun V. Multi-Scale Context Aggregation by Dilated Convolutions // Proceedings at ICLR 2016. 2015.
128. Zhou S. et al. Exploiting Local Structures with the Kronecker Layer in Convolutional Networks. 2015. P. 1–17.
129. Clevert D.-A., Unterthiner T., Hochreiter S. Fast and Accurate Deep Network Learning by Exponential Linear Units (ELUs) // Conference proceedings at ICLR

2016. 2016. P. 1–14.

130. Ba J.L., Kiros J.R., Hinton G.E. Layer Normalization // CoRR. 2016. Vol. abs/1607.0.
131. Xu Z. et al. The Effectiveness of Instance Normalization: a Strong Baseline for Single Image Dehazing. 2018. P. 1–13.
132. Ulyanov D., Vedaldi A., Lempitsky V. Instance Normalization: The Missing Ingredient for Fast Stylization. 2016.
133. Sutskever I. et al. Dropout: A Simple Way to Prevent Neural Networks from Overfitting // J. Mach. Learn. Res. 2014.
134. Tibshirani R. Regression Shrinkage and Selection Via the Lasso // J. R. Stat. Soc. Ser. B. 1996. Vol. 58, № 1. P. 267–288.
135. Ng A.Y. Feature selection, L1 vs. L2 regularization, and rotational invariance // Proceedings, Twenty-First Int. Conf. Mach. Learn. ICML 2004. 2004. P. 615–622.
136. Тихонов А.Н. О некорректных задачах линейной алгебры и устойчивом методе их решения // Доклады Академии наук СССР. 1965. Т. 163, № 3. С. 591–594.
137. Hoerl A.E., Kennard R.W. Ridge Regression: Applications to Nonorthogonal Problems // Technometrics. 1970. Vol. 12, № 1. P. 69–82.
138. Zou H., Hastie T. Regularization and variable selection via the elastic net // J. R. Stat. Soc. Ser. B (Statistical Methodol. 2005. Vol. 67, № 5. P. 301–320.
139. He K. et al. Deep Residual Learning for Image Recognition // Front. Psychol. 2015. Vol. 4, № 9. P. 124.
140. Rumelhart D.E., Hinton G.E., Williams R.J. Learning representations by back-propagating errors // Nature. 1986. Vol. 323, № 6088. P. 533–536.
141. Li H. et al. Visualizing the Loss Landscape of Neural Nets // Proc. 32nd Conf.

Neural Inf. Process. Syst. (NIPS 2018). 2017.

142. Greff K., Srivastava R.K., Schmidhuber J. Highway and Residual Networks learn Unrolled Iterative Estimation. 2016.
143. Huang G. et al. Densely connected convolutional networks // Proc. - 30th IEEE Conf. Comput. Vis. Pattern Recognition, CVPR 2017. 2017. Vol. 2017-Janua. P. 2261–2269.
144. Milletari F., Navab N., Ahmadi S.-A. V-Net: Fully Convolutional Neural Networks for Volumetric Medical Image Segmentation // 2016 Fourth International Conference on 3D Vision (3DV). IEEE, 2016. P. 565–571.
145. Sørensen T. A method of establishing groups of equal amplitude in plant sociology based on similarity of species and its application to analyses of the vegetation on Danish commons // K. Danske Vidensk. Selsk. Biol. Skr. 1948. Vol. 5, № 4. P. 1–34.
146. Dice L.R. Measures of the Amount of Ecologic Association Between Species // Ecology. Ecological Society of America, 1945. Vol. 26, № 3. P. 297–302.
147. Smith L.N. Cyclical learning rates for training neural networks // Proc. - 2017 IEEE Winter Conf. Appl. Comput. Vision, WACV 2017. 2017. № April. P. 464–472.
148. Keskar N.S. et al. On Large-Batch Training for Deep Learning: Generalization Gap and Sharp Minima. 2016. P. 1–16.
149. Dinh L. et al. Sharp Minima Can Generalize For Deep Nets. 2017.
150. Smith S.L., Kindermans P.-J., Le Q. V. Don't Decay the Learning Rate, Increase the Batch Size // International Conference on Learning Representations. 2018.
151. Arlot S., Celisse A., others. A survey of cross-validation procedures for model selection // Stat. Surv. The author, under a Creative Commons Attribution License, 2010. Vol. 4. P. 40–79.

152. Payam R., Lei T., Huan L. Cross-validation // *Encycl. database Syst.* Springer, 2009. Vol. 25. P. 532–538.
153. Kirschen R.H., O’Higgins E.A., Lee R.T. The Royal London space planning: An integration of space analysis and treatment planning // *Am. J. Orthod. Dentofac. Orthop.* 2000. Vol. 118, № 4. P. 456–461.
154. Pedregosa F. et al. Scikit-learn: Machine Learning in Python // *J. Mach. Learn. Res.* 2012. Vol. 19, № 1. P. 29–33.
155. Bergstra J. et al. Algorithms for hyper-parameter optimization // *Adv. Neural Inf. Process. Syst.* 24 25th Annu. Conf. Neural Inf. Process. Syst. 2011, NIPS 2011. 2011. P. 1–9.
156. Bergstra J., Bengio Y. Random search for hyper-parameter optimization // *J. Mach. Learn. Res.* 2012. Vol. 13. P. 281–305.
157. Fisher R.A. Student // *Ann. Eugen.* Wiley Online Library, 1939. Vol. 9, № 1. P. 1–9.
158. Kalpić D., Hlupić N., Lovrić M. Student’s t-Tests // *International Encyclopedia of Statistical Science* / ed. Lovric M. Berlin, Heidelberg: Springer Berlin Heidelberg, 2011. P. 1559–1563.
159. Medical Segmentation Decathlon. Generalisable 3D Semantic Segmentation [Electronic resource]. URL: <http://medicaldecathlon.com/index.html>.
160. Simpson A.L. et al. A large annotated medical image dataset for the development and evaluation of segmentation algorithms. 2019.
161. Ecabert O. et al. Segmentation of the heart and great vessels in CT images using a model-based adaptation framework. // *Med. Image Anal.* 2011. Vol. 15, № 6. P. 863–876.
162. Chakraborty J. et al. CT radiomics to predict high-risk intraductal papillary mucinous neoplasms of the pancreas // *Med. Phys.* 2018. Vol. 45, № 11.

P. 5019–5029.

163. Attiyeh M.A. et al. Preoperative risk prediction for intraductal papillary mucinous neoplasms by quantitative CT image analysis // HPB. International Hepato-Pancreato-Biliary Association Inc., 2019. Vol. 21, № 2. P. 212–218.
164. Attiyeh M.A. et al. Survival Prediction in Pancreatic Ductal Adenocarcinoma by Quantitative Computed Tomography Image Analysis // Ann. Surg. Oncol. Springer New York LLC, 2018. Vol. 25, № 4. P. 1034–1042.
165. Simpson A.L. et al. Chemotherapy-Induced Splenic Volume Increase Is Independently Associated with Major Complications after Hepatic Resection for Metastatic Colorectal Cancer // J. Am. Coll. Surg. 2015. Vol. 220, № 3. P. 271–280.
166. Bilic P. et al. The Liver Tumor Segmentation Benchmark (LiTS). 2019.
167. Isensee F. et al. Abstract: nnU-Net: Self-adapting Framework for U-Net-Based Medical Image Segmentation // Informatik aktuell. 2019. P. 22–22.
168. Isensee F. et al. nnU-Net: Breaking the Spell on Successful Medical Image Segmentation. 2019. P. 1–8.
169. nnUNet [Electronic resource]. URL: <https://github.com/MIC-DKFZ/nnunet>.
170. Medical Imaging and Intelligent Reality Lab [Electronic resource]. URL: <https://www.mi2rl.co/>.
171. Myronenko A. 3D MRI brain tumor segmentation using autoencoder regularization // Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics). 2018. Vol. 11384 LNCS. P. 311–320.
172. Myronenko A. 3D MRI Brain Tumor Segmentation Using Autoencoder Regularization [Electronic resource]. URL: <https://github.com/IAmSuyogJadhav/3d-mri-brain-tumor-segmentation-using-autoencoder-regularization#3d-mri-brain-tumor-segmentation-using-autoencoder-regularization>.

173. Perslev M. et al. A Multi-View Fully Convolutional Neural Network for Segmentation of Medical Image Volumes.
174. Kristensen T.V. Top position in Medical Segmentation Decathlon challenge [Electronic resource]. URL: <https://ai.ku.dk/news/top-5-position-in-medical-segmentation-decathlon-challenge/>.
175. Lai-Sheng Wang's Group [Electronic resource]. URL: <http://casey.brown.edu/chemistry/research/LSWang/publications/publications.html>.
176. Bae W. et al. Resource Optimized Neural Architecture Search for 3D Medical Image Segmentation. 2019. P. 1–9.
177. Paszke A. et al. ENet: A Deep Neural Network Architecture for Real-Time Semantic Segmentation. 2016. P. 1–10.
178. Wang G. et al. DeepIGeoS: A Deep Interactive Geodesic Framework for Medical Image Segmentation // IEEE Trans. Pattern Anal. Mach. Intell. 2019. Vol. 41, № 7. P. 1559–1572.
179. Wu Y., He K. Group Normalization // Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics). 2018. Vol. 11217 LNCS. P. 3–19.

ПРИЛОЖЕНИЕ А

Акты о внедрении

Ф 2.2/01-12

Министерство науки и высшего образования
Российской Федерации



Федеральное государственное
бюджетное научное учреждение
«Научно-исследовательский институт
комплексных проблем
сердечно-сосудистых заболеваний»
(НИИ КПССЗ)

Сосновый бульвар, д. 6, г. Кемерово, 650002
тел. 8 (3842) 643-308, факс 8 (3842) 643-410
e-mail: reception@kemcardio.ru
www.kemcardio.ru
ОКПО 55608705; ОГРН 1034205024479;
ИНН/КПП 4205012290/420501001

«Утверждаю»

Директора НИИ КПССЗ

Член-корр., д.м.н. Барбараш О.Л.

« 07 » июл 2019 г

АКТ

об использовании результатов

кандидатской диссертационной работы Данилова Вячеслав Владимировича

Мы, нижеподписавшиеся, комиссия в составе председателя Кутихина А.Г. и членов: Овчаренко Е.А. и Клышникова К.Ю. удостоверяем, что результаты диссертационного исследования Данилова Вячеслава Владимировича используются в работе НИИ КПССЗ в форме инструментального средства обработки информации с января 2019 года. Данный инструмент адаптирован для выполнения сегментации стеноза при проведении диагностики коронарных артерий с помощью ангиографии. Средняя точность сегментации стеноза сосудов, полученная данным инструментом, составила 93.2%. Реализованное программное обеспечение на языке Python позволяет выделить на графических данных стеноз и указать точность его прогнозирования. Внедрение в НИИ КПССЗ данной системы предоставило аналитическую информацию для принятия более эффективных решений, а также повысило скорость принимаемых решений в среднем на 14% по сравнению с предустановленным ранее программным обеспечением.

Председатель комиссии

(подпись)

Кутихин А.Г.

Члены комиссии:

(подпись)

Овчаренко Е.А.

(подпись)

Клышников К.Ю.



Министерство науки и высшего образования Российской Федерации
федеральное государственное автономное
образовательное учреждение высшего образования
«Национальный исследовательский Томский политехнический университет» (ТПУ)



УТВЕРЖДАЮ

Директор инженерной школы

информационных технологий и робототехники

Сонькин Д.М.

« 05 » сентября 2019 г.

АКТ

**о внедрении в учебный процесс результатов
кандидатской диссертационной работы Данилова Вячеслав Владимировича**

Комиссия в составе:

Председателя:

Дёмин А.Ю.

Заместитель директора по развитию, к.т.н.

Членов комиссии:

Шерстнёв В.С.

Заведующий кафедрой – руководитель

отделения на правах кафедры ОИТ, к.т.н.

Леонов С.В.

Заведующий кафедрой – руководитель

отделения на правах кафедры ОАР, к.т.н.

Кочегуров А.И.

Доцент ОИТ ИШИТР, к.т.н.

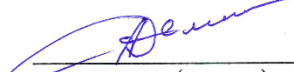
Составили настоящий акт о нижеследующем:

Материалы диссертационной работы Данилова В.В. используются:

1. При проведении практических занятий по дисциплинам «Машинное обучение», «Технология обработки информации», «Теория распознавания образов», «Нейронные сети», «Методы интеллектуальной обработки и анализа изображений» для магистрантов направлений 230100 "Информатика и вычислительная техника" и 230400 "Информационные системы и технологии" в Томском Политехническом Университете.

2. Для курирования выпускных работ студентов бакалавриата и магистратуры.

Председатель комиссии


(подпись)

Дёмин А.Ю.

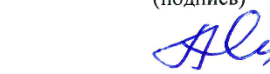
Члены комиссии


(подпись)

Шерстнёв В.С.


(подпись)

Леонов С.В.


(подпись)

Кочегуров А.И.

ПРИЛОЖЕНИЕ Б

Свидетельства о регистрации программ для ЭВМ и патент на изобретение

РОССИЙСКАЯ ФЕДЕРАЦИЯ



СВИДЕТЕЛЬСТВО

о государственной регистрации программы для ЭВМ

№ 2020613197

«Модели локализации стеноза на основе данных ангиографии»

Правообладатель: **Федеральное государственное бюджетное научное учреждение «Научно-исследовательский институт комплексных проблем сердечно-сосудистых заболеваний» (НИИ КПССЗ) (RU)**

Авторы: **Клышников Кирилл Юрьевич (RU), Овчаренко Евгений Андреевич (RU), Скирневский Игорь Петрович (RU), Данилов Вячеслав Владимирович (RU)**

Заявка № **2020611881**Дата поступления **25 февраля 2020 г.**

Дата государственной регистрации

в Реестре программ для ЭВМ **11 марта 2020 г.**

Руководитель Федеральной службы
по интеллектуальной собственности

Г.П. Ивлиев

РОССИЙСКАЯ ФЕДЕРАЦИЯ



СВИДЕТЕЛЬСТВО

о государственной регистрации программы для ЭВМ

№ 2020612121

**Наложение медицинского инструмента на графические
данные ультразвукового исследования**

Правообладатель: *федеральное государственное автономное
образовательное учреждение высшего образования
«Национальный исследовательский Томский политехнический
университет» (RU)*

Авторы: *Данилов Вячеслав Владимирович (RU), Манаков Роман
Аркадьевич (KZ), Скирневский Игорь Петрович (RU)*



Заявка № 2020610938

Дата поступления 03 февраля 2020 г.

Дата государственной регистрации
в Реестре программ для ЭВМ 17 февраля 2020 г.

Руководитель Федеральной службы
по интеллектуальной собственности

Г.П. Ивлиев Г.П. Ивлиев

РОССИЙСКАЯ ФЕДЕРАЦИЯ



СВИДЕТЕЛЬСТВО

о государственной регистрации программы для ЭВМ

№ 2020611930

Сегментация изображений на основе распространения
динамически изменяемых суперпикселей

Правообладатель: *федеральное государственное автономное
образовательное учреждение высшего образования
«Национальный исследовательский Томский политехнический
университет» (RU)*

Авторы: *Данилов Вячеслав Владимирович (RU), Скирневский Игорь
Петрович (RU), Манаков Роман Аркадьевич (KZ)*



Заявка № 2020610927

Дата поступления 03 февраля 2020 г.

Дата государственной регистрации
в Реестре программ для ЭВМ 12 февраля 2020 г.

Руководитель Федеральной службы
по интеллектуальной собственности

Г.П. Ивлиев Г.П. Ивлиев

РОССИЙСКАЯ ФЕДЕРАЦИЯ



СВИДЕТЕЛЬСТВО

о государственной регистрации программы для ЭВМ

№ 2019664820

Система автоматического тестирования гибких роботов

Правообладатель: *федеральное государственное автономное образовательное учреждение высшего образования «Национальный исследовательский Томский политехнический университет» (RU)*

Авторы: *Данилов Вячеслав Владимирович (RU), Кравченко Андрей Александрович (RU), Колпащиков Дмитрий Юрьевич (RU), Лаптев Никита Витальевич (RU), Манаков Роман Аркадьевич (KZ), Гергет Ольга Михайловна (RU)*



Заявка № 2019663014

Дата поступления 22 октября 2019 г.

Дата государственной регистрации

в Реестре программ для ЭВМ 13 ноября 2019 г.

Руководитель Федеральной службы
по интеллектуальной собственности

Г.П. Ивлиев

РОССИЙСКАЯ ФЕДЕРАЦИЯ



СВИДЕТЕЛЬСТВО

о государственной регистрации программы для ЭВМ

№ 2019664675

Система локализации и сегментации хирургического
инструмента и анатомических структур на основе
машинного обучения

Правообладатель: *федеральное государственное автономное
образовательное учреждение высшего образования
«Национальный исследовательский Томский политехнический
университет» (RU)*

Авторы: *Данилов Вячеслав Владимирович (RU),
Гергет Ольга Михайловна (RU)*



Заявка № 2019663095

Дата поступления 16 октября 2019 г.

Дата государственной регистрации

в Реестре программ для ЭВМ 12 ноября 2019 г.

Руководитель Федеральной службы
по интеллектуальной собственности

Г.П. Ивлиев

РОССИЙСКАЯ ФЕДЕРАЦИЯ

(19)

RU⁽¹¹⁾2019 127 060⁽¹³⁾

A



ФЕДЕРАЛЬНАЯ СЛУЖБА
ПО ИНТЕЛЛЕКТУАЛЬНОЙ СОБСТВЕННОСТИ

(12) **ЗАЯВКА НА ИЗОБРЕТЕНИЕ**

Состояние делопроизводства: Экспертиза по существу завершена (последнее изменение статуса: 20.03.2020)

(21)(22) Заявка: [2019127060](#), 28.08.2019

(30) Конвенционный приоритет: RU

Делопроизводство

Исходящая корреспонденция		Входящая корреспонденция	
Решение о выдаче патента	18.03.2020	Дополнительные материалы	10.03.2020
		Письмо для ответа	05.03.2020
Отчет об информационном поиске	13.02.2020		
Запрос экспертизы	13.02.2020		
Уведомление об удовлетворении ходатайства	11.09.2019	Ходатайство о проведении экспертизы заявки по существу	09.09.2019
Уведомление об удовлетворении ходатайства	10.09.2019	Ходатайство об освобождении от уплаты пошлин или уменьшении размера	09.09.2019
Уведомление о положительном результате формальной экспертизы	04.09.2019		
Уведомление об удовлетворении ходатайства	04.09.2019	Ходатайство об освобождении от уплаты пошлин или уменьшении размера	28.08.2019
Уведомление о зачете пошлины	04.09.2019	Платежный документ	28.08.2019
Уведомление о поступлении документов заявки	28.08.2019		