

Санкт–Петербургский государственный университет

КВАНЧИАНИ Карина Шавкановна

Выпускная квалификационная работа

Нейросетевые методы повышения качества медицинских изображений для обнаружения и сегментации опухолей

Уровень образования: бакалавриат

Направление 01.03.02 «Прикладная математика и информатика»

Основная образовательная программа СВ.5005.2017 «Прикладная математика, фундаментальная информатика и программирование»

Профиль «Математическое и программное обеспечение
вычислительных машин»

Научный руководитель:

доцент, кафедра технологии программирования,
к.т.н. Блеканов Иван Станиславович

Рецензент:

доцент, кафедра механики управляемого движения,
к.ф. - м.н. Шиманчук Дмитрий Викторович

Санкт-Петербург

2021 г.

Содержание

Введение	3
Актуальность работы	3
Цель и задачи работы	5
Практическая значимость	6
Глава 1. Обзор литературы и технологических решений	9
1.1. Обзор существующих технологических решений по работе с медицинскими изображениями	9
1.2. Задачи детекции и сегментации изображений	12
1.3. Обзор методов повышения качества изображений и их оцен- ка в задачах детекции и сегментации	17
1.3.1 Методы обучения с учителем	18
1.3.2 Методы обучения без учителя	27
1.3.3 Применение алгоритма повышения разрешения к зада- чам детекции и сегментации опухолей	28
1.3.4 Генеративно-состязательные сети	29
1.4. Обзор методов оценки качества изображений	33
Глава 2. Разработка программного комплекса для детекции и сег- ментации опухолей	36
2.1. Проектирование архитектуры программного комплекса	36
2.2. Разработка и реализация методов на основе описанных мо- делей	40
2.3. Описание гиперпараметров обучения	43
2.4. Тестирование и апробация разработанного решения	45
2.4.1 Постановка эксперимента	45
2.4.2 Результаты эксперимента	45
2.4.3 Выводы	48
Заключение	49
Результаты работы	49
Перспективы развития	50
Список литературы	51

Введение

Актуальность работы

Компьютерное зрение является одним из наиболее прогрессивных направлений машинного обучения. Распознавание лиц, сегментация и классификация изображений, их обработка и генерация - все это задачи компьютерного зрения. За последние 10 лет в этой области произошло немало достижений, достигнуты почти человеческие способности решения некоторых задач, а иногда алгоритмы машинного обучения даже превосходят человека - нейронная сеть способна классифицировать изображения, ошибаясь в 1.2% случаев [1], тогда как ошибка человека составляет 5.1% [2] (результаты соревнования ImageNet). Также, в последние несколько лет, исследователи активно разрабатывают state-of-the-art алгоритмы на основе машинного обучения в медицинской сфере, в особенности уделяя внимание выявлению злокачественных новообразований. В настоящее время опухоль головного мозга является одной из главных причин роста смертности среди взрослых и детей, а наличие специализированных методов обнаружения аномалий предоставляют возможность врачам помочь сохранить жизни за счет более раннего выявления болезни.

Опухоль головного мозга - это масса клеток, которые бесконтрольно растут и размножаются в различных частях мозга [3]. В нормальных условиях здоровые клетки тела периодически умирают и на их месте появляются новые. Опухолевые клетки ведут себя иначе - хотя организм в них не нуждается, они размножаются и не умирают, как здоровые. По мере продолжения этого процесса злокачественное новообразование продолжает расти и, в некоторых случаях, очень стремительно. Злокачественные опухоли головного мозга не имеют четких границ - помимо того, что они быстро растут, подобные новообразования способны перемещаться. Существуют различные методы обработки изображений, которые помогают обнаружить такие особенности опухолей, как форма, размер, текстура и кальцификация. Нахождение этих особенностей делает процесс обнаружения опухоли наиболее точным.

Качественная визуализация в диагностике рака является важным ин-

струментом для его выявления с использованием алгоритмов детекции и сегментации. Основной проблемой медицинской визуализации (и компьютерного зрения в целом) является сложность сбора необходимого количества качественных данных для наиболее эффективного анализа изображений методами искусственного интеллекта. В частности, такого рода проблемы порождаются наличием технических ограничений устройств получения изображений (камера, компьютерная томография, спутник) и устройств их обработки (оптические искажения, наличие шума, размытие линз).

Медицинские снимки можно разделить на несколько групп в зависимости от того, с помощью какого аппарата они сделаны. К таким способам получения медицинских изображений относятся ультразвуковое сканирование, магнитно резонансная томография, компьютерная и ядерная томографии. МРТ использует радиоволны и магнитные поля для создания детализированных изображений поперечного сечения внутренних органов в отличие от других способов, что делает ее наиболее применяемой. К тому же, МРТ считается наиболее безопасным способом медицинской визуализации, но применим такой способ не во всех случаях (только в тех частях тела, где много жидкости). Компьютерная томография справляется с большей площадью человеческого тела, однако существует высокая вероятность облучения пациента вредным излучением. В связи с этим в медицинской визуализации собрать множество снимков хорошего качества значит не только преодолеть трудности получения доступа к персональным данным, но и негативно воздействовать на здоровье часто ослабленных пациентов. При снижении дозы агрессивного излучения (для более щадящего воздействия), полученные снимки могут отличаться низкой контрастностью, наличием шумов и размытием. Другая не менее важная проблема - не всегда медицинское учреждение имеет ресурсы на приобретение аппаратов, способных создавать качественные снимки. Это порождает собой создание изображений низкого разрешения, которые к тому же страдают от низкой контрастности.

Стоит помнить и о человеческих факторах при получении снимка - пациент может случайно привести в движение ту или иную часть тела (на результат может повлиять даже сердцебиение), что негативно влияет на получение четкого снимка. Ошибка при настройке аппарата, отсутствие фокуси-

ровки, наличие у пациента металлических имплантов - также может повлиять на качество изображения. Не всегда существует возможность сделать вторую попытку (пациент может уже уйти, повторное воздействие излучения может оказать неблагоприятное воздействие и т.д.).

Тогда как для решения проблемы нехватки данных существуют предварительно обученные модели и возможность аугментации, то для борьбы с изображениями низкого качества существуют алгоритмы его повышения.

Повышение качества изображений - важная задача компьютерного зрения. Область применения технологий решения данной проблемы достаточно обширна - помимо использования алгоритмов повышения качества изображений в машинном обучении, они также популярны и в реальном мире - для качественной интерпретации человеком. В медицине хорошее качество изображения необходимо для правильной постановки диагноза [4], в правоохранительных органах - для распознавания лиц с камер наблюдения [5], в медиа - для повышения качества давно снятых фильмов [6].

Что представляет собой повышение качества изображений? Изображение в плохом качестве отличается от изображения в хорошем качестве наличием шумов и артефактов, размытостью, низким разрешением и другими, в основном зависящими от устройства получения изображений, недостатками. Исходя из этого можно сказать, что улучшить изображение - значит увеличить его пространственное разрешение, избавиться от шумов и артефактов, сделать изображение более четким ¹.

Цель и задачи работы

Целью данной работы является разработка методов улучшения качества изображений, основанных на использовании современных нейронных сетей, для повышения эффективности обнаружения и сегментации опухолей.

Для достижения цели в работе поставлены следующие задачи:

1. обзор автоматизированных техник повышения качества изображений и

¹Задача повышения качества изображений в данной работе описана в терминах задачи повышения разрешения (Super Resolution, SR). В литературе она часто подразумевает под собой не только увеличение изображения, но и само улучшение его качества (устранение размытия, удаление артефактов и т.д.). В связи с этим далее словосочетание "повысить разрешение" будет пониматься как "улучшить качество".

их модификаций

2. обзор существующих технологических решений в области анализа изображений
3. изучение генеративно-состязательных сетей с целью их применения в качестве основного алгоритма
4. обзор существующих методов оценки качества изображений, выбор наиболее подходящего для данной работы
5. проектирование архитектуры программного комплекса на основе методов искусственного интеллекта
6. разработка и реализация нейронных сетей для задач повышения разрешения, детекции и сегментации
7. тестирование и апробация разработанного решения на медицинских изображениях, с использованием полученных знаний об их специфичности
8. оценка качества детекции и сегментации опухолей с использованием алгоритма повышения качества изображений и без него

Практическая значимость

Использование технологий повышения разрешения в медицине не ограничивается диагностикой рака с применением методов машинного обучения. Помимо того, что устранение шумов и других артефактов с изображения может благоприятно повлиять на эффективность алгоритмов детекции и сегментации опухолей, снимок лучшего качества и бóльшего размера дает возможность специалисту увидеть мелкие детали новообразований. Все это необходимо для максимальной корректности поставленных диагнозов и снижения затрат на их получение.

Построенный в результате данной работы алгоритм может быть использован и в других доменах для решения различных задач. Однако его комплексность заключается в самом методе, тогда как в зависимости от поставленной

проблемы могут меняться данные, их предобработка, методы оценки и другие нюансы. Далее рассмотрены некоторые из тех приложений, в которых применяются технологии повышения разрешения.

1. Глубинные карты. Для оценки положения некоторых объектов в пространстве и их семантической сегментации существуют глубинные карты. Проблема таких карт - экономические и производственные ограничения при их создании, из-за которых глубинная карта может оказаться в низком разрешении с присутствием шумов и отсутствием некоторых значений. Такие сложности существенно влияют на оценку положения объекта, в частности на его границы, что делает задачу улучшения качества глубинных карт актуальной.
2. Видео. Решение задачи повышения качества отдельных изображений несложно распространить на совокупность изображений - видео. В данном случае несколько кадров представляют собой гораздо больше информации о сцене и существуют не только внутрикадровые пространственные зависимости, но и межкадровые временные зависимости (движения, изменения яркости и цвета). Актуальность повышения качества видео достаточно велика. Примером подобного приложения может служить технология Яндекса DeepHD [6] - проект, с помощью которого пользователям предоставляется возможность смотреть давно снятые фильмы в хорошем качестве.
3. Камеры наблюдения. Многим известно, что камеры, установленные в периметрах важных объектов не всегда снабжены дорогостоящими компонентами, способными создавать качественные видео. Часто, целью таких камер наблюдения является фиксация подозрительных движений, для чего качество играет малую роль. Однако при совершении серьезных преступлений, таких как кража, убийство и терроризм, для обнаружения и распознавания лица подозрительного / участвующего в действии человека необходима четкая картинка с установленных поблизости камер. В таких ситуациях - когда на полученной картинке все-таки невозможно различить лицо преступника - применяются ме-

тоды повышения качества изображений. Такой алгоритм может быть полезен и в пропускной системе сотрудников компании - при нечетком снимке система может ошибиться и впустить на территорию человека, не являющегося сотрудником компании, или не впустить сотрудника.

4. Спутниковые снимки. Часто спутниковые снимки не страдают проблемой низкого разрешения, наоборот, они достаточно огромны и могут содержать в себе 10.000×10.000 пикселей. Несмотря на это, изображения подобной природы обычно очень шумные, что негативно влияет на эффективность работы с ними. В качестве примера использования алгоритмов улучшения качества спутниковых снимков можно привести цифровую систему управления арктической логистикой Газпром Нефти "Капитан" [7], в задачи которой входит сегментация типов льда в реальном времени как раз на основе снимков со спутника, предварительно очищенных от специфичных шумов.

Стоит отметить, что в приложениях 1, 3 и 4 алгоритм повышения качества изображений является не целью, а вспомогательным действием для решения других задач машинного обучения - детекции (обнаружение объектов на изображении), распознавании лиц и сегментации (поиск границ объектов) [8]. Аналогично, идея данной работы заключается в построении алгоритма повышения разрешения для наиболее эффективного решения задач детекции и сегментации опухолей на медицинских снимках.

Глава 1. Обзор литературы и технологических решений

1.1 Обзор существующих технологических решений по работе с медицинскими изображениями

Сегодня медицинская визуализация является одной из самых развивающихся областей здравоохранения. За последние несколько десятилетий она расширилась и стала включать в себя несколько методов визуализации - компьютерную и ядерную томографии, МРТ, ультразвук и другие. Наряду с усовершенствованием аппаратного обеспечения и устройств, используемых для создания медицинских изображений, огромный прогресс был достигнут с различными типами программного обеспечения, которое обрабатывает эти изображения. Объем рынка ПО для анализа и обработки медицинских изображений в 2019 году оценивался в 2.41 миллиарда долларов США, в 2020 - 2.6 и, ожидается, достигнет 4.5 в 2027 году [9]. Годовой темп роста составляет около 8%.

Актуальность. Сейчас обработка изображений в медицине - гораздо больше, чем просто возможность просмотра изображений. Благодаря автоматизированным системам для обработки медицинских снимков, медицинские работники могут отслеживать и архивировать изображения, манипулировать ими, обмениваться связанной информацией не тратя на это много времени [10]. Способы медицинской визуализации способны до составления плана лечения снизить потребность в хирургии, госпитализации и длительном пребывании в больнице. Тем самым, внедрение ПО приносит большую пользу как прямым пользователям (медицинским работникам), так и конечным (пациентам). В настоящее время обработка медицинских изображений является обычной практикой в диагностических, амбулаторно-хирургических, академических и исследовательских центрах.

Использование алгоритмов искусственного интеллекта. Искусственный интеллект имеет огромный потенциал в секторе здравоохранения с точки зрения будущих перспектив. Области внедрения приложений искусственного интеллекта в здравоохранении охватывают медицинские испытания, открытие лекарств, клинические исследования, хирургическую робототехнику, биоме-

дицинскую инженерию и т.д. Более того, его глубокий потенциал может быть направлен на создание персонализированной медицины и достижение нулевого уровня ошибок при определении стратегий лечения. Ожидается, что к 2025 году мировой рынок искусственного интеллекта в здравоохранении вырастет до 28 миллиардов долларов, а внедрение приложений искусственного интеллекта в клинической медицине может сэкономить до 150 миллиардов долларов в год экономике здравоохранения США к 2026 году [11]. Почти четверть (24%) приложений на основе искусственного интеллекта на мировом рынке здравоохранения относятся к развитию личного здоровья и помощникам медсестер, а 20% приложений обеспечивают мониторинг решения для клиентов в режиме реального времени, 11% доступных приложений предназначены для облегчения административного рабочего процесса и оптимизации управления задачами в секторе здравоохранения. Искусственный интеллект может повысить качество и скорость медицинского обслуживания практически во всех областях медицины за счет автоматизации административных задач, упрощения процедур и улучшения коммуникации между врачами и пациентами.

В настоящее время искусственный интеллект в здравоохранении также используется для анализа и обработки медицинских изображений и сбора данных в медицинских картах пациентов. В отличие от ПО для обработки медицинских изображений, ПО для их анализа все еще находится в зачаточном состоянии.

Сложности внедрения. Несмотря на то, что решения, основанные на искусственном интеллекте положительно влияют на качество предоставляемых медицинскими работниками услуг, сложность инноваций в области здравоохранения может создать нормативные и технологические барьеры для внедрения приложений ИИ в медицине. Одна из проблем заключается в том, что результаты работы построенных алгоритмов недостаточно гибки и не могут на данном этапе полностью заменить человека. Алгоритм может ошибаться там, где медицинский работник никогда не ошибется и наоборот.

Коммерческие решения:

1. IBM (США) [12]. В поисках способов постановки точных и быстрых

диагнозов IBM была одной из первых компаний, которые внедрили ИИ для медицинских услуг, предложив клиницистам систему Watson Health, чтобы помочь онкологам выбирать оптимальные планы лечения на основе индивидуальных данных пациента.

2. Aidoc (Израиль) [13]. Главной целью компании является поддержка и увеличение эффективности диагностических возможностей радиологов, помогая им ускорить лечение пациентов и улучшить качество медицинской помощи. Решения компании помогают радиологам сократить время обработки изображений, повысить качество и эффективность за счет выявления визуальных нарушений на снимках в режиме реального времени. Особенность: основу обучения составляет инкрементное онлайн-обучение, что дает возможность алгоритму непрерывно получать новые знания из вновь добавленных данных.
3. Arterys (США) [14]. Изначально разработанное ПО предназначалось только для МРТ сердца, но теперь Arterys разработала аналогичные приложения для МРТ печени, МРТ легких и маммограммы и помогает выявлять патологические поражения в этих регионах. Особенность: сочетание возможностей глубокого обучения и облачных вычислений для автоматизированного анализа изображений.

Открытые решения:

1. MeVisLab (Германия) [15]. Приложение включает в себя расширенные модули для сегментации, регистрации, волюметрии, а также количественного морфологического и функционального анализа. Его часто используют для нейровизуализации, динамического анализа изображений, планирования операций, а также для анализа сердечно-сосудистой системы.

1.2 Задачи детекции и сегментации изображений

Выше упоминалось, что задача повышения разрешения в данной работе считается низкоуровневой (вспомогательной) задачей, цель которой улучшить качество работы алгоритмов двух других проблем машинного обучения - детекции и сегментации. Так как сами эти задачи не являются предметом исследования (они скорее выступают в роли оценщика алгоритма повышения разрешения), решено рассмотреть их кратко.

Задача детекции [26]. Детекция (или обнаружение) объектов на изображениях считается одной из самых важных и сложных направлений компьютерного зрения. Детекция нацелена на поиск объектов на изображениях с их последующей идентификацией. Описываемая задача применяется для обнаружения лиц по снимкам с камер наблюдения, идентификации тех или иных препятствий на пути у беспилотного автомобиля, детекции опухолей на медицинских снимках и для других.

Задача сегментации [27]. Сегментация (или семантическая сегментация) - это процесс присвоения каждому пикселю изображения определенной метки, чем она и отличается от классификации, где одному изображению соответствует единственный класс. Область применения рассматриваемой задачи почти идентична области применения задачи детекции. Сегментация препятствий на пути беспилотника в данном случае говорит не только об их наличии, но и о том, где они находятся, насколько они велики и, используя полученную информацию, имеется возможность решить, например, в какую сторону необходимо повернуть автомобиль, чтобы выехать на свободную от нежелательных объектов дорогу. То же касается и медицинских снимков - сегментация новообразований дает понять хирургу о размере таковых, об их положении и распространенности.

Ранее исследования показали, что характеристики недавно выявленных опухолей головного мозга с помощью методов медицинской визуализации можно использовать для определения вероятной стратегии диагностики и лечения. Важность приведения алгоритма детекции опухоли к полной автоматизации очевидна - не всегда медицинские работники имеют достаточно опыта для определения наличия опухоли и ее типа (доброкачественная или

злокачественная). В случае ошибки второго рода при гипотезе "опухоль на снимке есть" (то есть когда специалист неверно решил, что опухоль присутствует) проблема не слишком велика, однако когда допущенная ошибка - ошибка первого рода (специалист не заметил новообразование или счел его доброкачественным, каковым оно не является), результат часто оказывается неисправимым и печальным. Сегментация изображений является критически важным шагом в процессе использования медицинских снимков в исследованиях опухолей головного мозга: (1) сегментированная протяженность опухоли головного мозга может устранить мешающие структуры из других тканей мозга и, следовательно, обеспечить более точную классификацию подтипов опухолей головного мозга и сообщить последующий диагноз; (2) точное разграничение имеет решающее значение при лучевой терапии или хирургическом планировании, из которого не только выделяется опухоль головного мозга, но и тщательно исключаются окружающие опухоль здоровые ткани, чтобы избежать повреждения участков языковой, двигательной и сенсорной функции во время терапии; (3) сегментация продольных снимков может эффективно контролировать рецидив, рост или уменьшение опухоли головного мозга.

В современной клинической практике сегментация по-прежнему основана на ручном разграничении, выполняемом специалистами. Сегментация вручную - очень трудоемкая задача, которая обычно включает в себя пошаговые процедуры, а результаты во многом зависят от опыта специалистов и их субъективных решений. Для мультимодального и продолжительного клинического исследования очень востребован полностью автоматический, объективный и воспроизводимый метод сегментации [28]. Раннее и точное обнаружение опухолей головного мозга обеспечивает эффективную и действенную терапию, что, в свою очередь, может привести к увеличению времени жизни человека. Однако стоит помнить, что автоматизация решения таких важных задач, как детекция и сегментация опухолей, может привести к серьезным ошибкам. При возможности (когда не требуется срочное вскрытие пациента и т.д.) результат алгоритма следует проанализировать специалисту.

Основными трудностями, возникающими во время решения задач детекции опухолей и сегментации изображений, являются: шум, размытие с

низкой контрастностью, низкое качество снимка, артефакты и другие. Это, вероятно, оказывает влияние на качество работы рассматриваемых алгоритмов, поэтому далее (в разделе 1.3.3) решено подробнее описать влияние улучшения качества медицинских снимков на задачи детекции и сегментации. Ниже описаны state-of-the-art модели для решения описываемых задач.

UNet [33]. Типичное использование сверточных нейронных сетей - задачи классификации, где выходом сети является метка класса. Однако в задаче сегментации желаемый выход должен включать локализацию, то есть метка класса должна быть присвоена каждому пикселю. В связи с этим была разработана нейронная сеть, которая предсказывала метку класса для всех пикселей путем учета локальной области (патча) вокруг каждого пикселя. Очевидно, такой подход имел некоторые недостатки. Сеть работала медленно ввиду того, что она должна была “пробежаться” по всем маленьким патчам. Помимо этого возникала избыточность из-за наложения патчей друг на друга.

Архитектура. Для решения описанных проблем исследователи предложили UNet-образную архитектуру нейронной сети. Такой тип архитектуры в основном состоит из двух частей - слоев понижающей дискретизации и слоев повышающей дискретизации. Первая представляет собой типичную архитектуру сверточной нейронной сети, то есть содержит в себе повторяющееся наложение двух сверток с размером ядра 3×3 , за каждой из которых следует слой линейной ректификации (Rectified Linear Unit, ReLU) и операцию пулинга с размером ядра 2×2 и с шагом 2 для понижения дискретизации. Таким образом на каждом этапе первой части количество каналов изображения удваивается. Каждый шаг второй части состоит из слоя повышающей дискретизации путем свертки 2×2 (вместо увеличения количества каналов здесь свертка их уменьшает), конкатенации с соответствующей картой признаков из первой части (таким образом сеть может научиться использовать больше информации для более точного вывода) и двух сверток 3×3 , после каждой из которых следует слой линейной ректификации.

Данные. Для обучения нейронной сети используются изображения и соответствующие им сегментационные маски, то есть пары (объект; метка) представлены парами (изображение; маска). Аугментация данных необходима для обучения сети желаемым свойствам инвариантности и устойчивости,

когда доступна только маленькая обучающая выборка. В случае медицинских снимков нам в первую очередь важна инвариантность к сдвигу и вращению, а также устойчивость к деформациям и вариациям значений серого. Аугментация обучающих выборок считается ключевой концепцией для обучения нейронной сети сегментации с очень небольшим количеством аннотированных изображений, что часто касается медицинской визуализации.

Оценка результата в задаче сегментации. Dice score - одна из наиболее часто используемых метрик для оценки качества решения задачи сегментации в медицинской визуализации. Было отмечено [34], что dice score отражает соответствие как размера, так и локализации, в большей степени ориентируясь на качество восприятия, чем на точность соответствия пикселей. Мера схожести двух масок сегментации (y - ground truth метка, \bar{y} - предсказанная метка) выглядит следующим образом:

$$D(y, \bar{y}) = \frac{2 |y \cap \bar{y}|}{|y| + |\bar{y}|}$$

Для того, чтобы сравнивать полностью пустые маски, к числителю и знаменателю добавляют сколь угодно малое значение

VGG. VGG - архитектура сверточной нейронной сети, предложенная в статье [35]. Модель достигает 92,7% точности на наборе данных ImageNet [1] и считается одной из самых популярных нейронных сетей для классификации изображений.

Архитектура. Архитектура VGG может меняться в зависимости от сложности поставленной задачи и наличия вычислительных ресурсов (чем больше в нейронной сети слоев, тем более сложную задачу она, потенциально, может решить и больше ресурсов потребовать). В частности, VGG-16 состоит из 5 блоков, в первых двух из которых по два сверточных слоя с применением слоя линейной ректификации и последующим применением пулинга, в следующих трех блоках вместо двух сверточных слоев - три. Завершается архитектура приведением полученной матрицы к вектору и последующими полносвязными слоями.

Данные. В задаче детекции обучение происходит с использованием пар (изображение; класс изображения). Аналогично задаче сегментации в зада-

че детекции часто можно улучшить эффективность решения используя в обучении аугментацию данных.

Оценка результата в задаче детекции. Как и любая задача классификации, детекцию принято оценивать метрикой f1 score.

1.3 Обзор методов повышения качества изображений и их оценка в задачах детекции и сегментации

Цель задачи повышения разрешения (Super Resolution, SR) - получить изображение в высоком разрешении (High Resolution, HR) из того же изображения в низком разрешении (Low Resolution, LR), не уменьшая его качества. Для того, чтобы обучить будущий алгоритм, необходимы пары изображений (изображение в низком разрешении; изображение в высоком разрешении). В данной задаче изображения LR являются обучающими данными, а изображения HR - метками. Однако не всегда существует возможность собрать требуемое количество подобных пар. Поэтому принято создавать LR изображения (далее LR) с помощью деградации HR изображений (далее HR):

$$LR = D(HR, \delta),$$

где D - функция отображения деградации (например, бикубическая интерполяция)², а δ - ее параметр (коэффициент масштабирования, шум или др.) [16]. Тогда, после получения необходимой выборки, становится возможным обучить модель генерировать HR из LR :

$$HR = F(LR, \gamma),$$

где F - модель повышения разрешения, а γ - ее параметр.

Иначе говоря стоит задача создать такую функцию, которая будет отображать пространство полученных изображений LR в пространство HR , то есть функцию, обратную функции деградации (далее сгенерированные изображения HR будут обозначаться через HRg - High Resolution generated):

$$HRg = F(LR, \gamma) = D^{-1}(LR) + r,$$

где r - остаток реконструкции. Задачу повышения разрешения часто называют некорректной, так как таковой является задача оценки инверсии функции

²Деградация подразумевает под собой не только уменьшение изображения, но и размытие и добавления шума. Такие неточности в определении терминов задачи SR существуют по причине неточной постановки самой проблемы.

деградации.

Существует еще одна формулировка задачи повышения разрешения, цель которой получить изображение в высоком разрешении $X \in R^m$ из соответствующего ему изображения в низком разрешении $Y \in R^n$, где $m > n$. Другими словами - получить изображение в пространстве большей размерности чем изначальное, повторив полностью сцену изображения.

Выше была описана одна из нескольких структур наборов данных для решения данной задачи. Помимо выборок, состоящих только из изображений HR (что обязывает нас создавать самостоятельно LR изображения), используются также и выборки с заранее подготовленными парами $(LR; HR)$. Другие отличия между наборами данных заключаются в количестве изображений, их разрешении и формате.

1.3.1 Методы обучения с учителем

Часто для решения задачи SR используются методы обучения с учителем, то есть обучение происходит на парах $(LR; HR)$. В тех случаях, когда изображение LR отсутствует, нам необходимо его создать с помощью деградации изображения HR , что дает повод описать в данном разделе методы повышения разрешения на основе интерполяции (не являются методами обучения с учителем), которые чаще используются для его понижения.

Повышение разрешения на основе интерполяции. Самыми доступными в смысле вычислительной сложности методами увеличения изображений являются методы, основанные на интерполяции. Интерполяция - поиск неизвестных промежуточных значений некоторой функции по уже вычисленным ее значениям. Если интерпретировать данное определение к задаче повышения разрешения изображений, то можно дать следующее пояснение.

Дана матрицей пикселей (изображением) и цель состоит в его увеличении, для чего необходимо получить бóльшую по размеру матрицу. Сделать это можно “раздвинув” значения матрицы и заполнив полученные пропуски значениями определенного рода (см. рис. 1.1). В терминах определения интерполяции изображение является функцией, пропущенные значения которой необходимо восстановить по известным значениям. В частности, пустая

ячейка матрицы заполняется усредненным значением ее соседей.

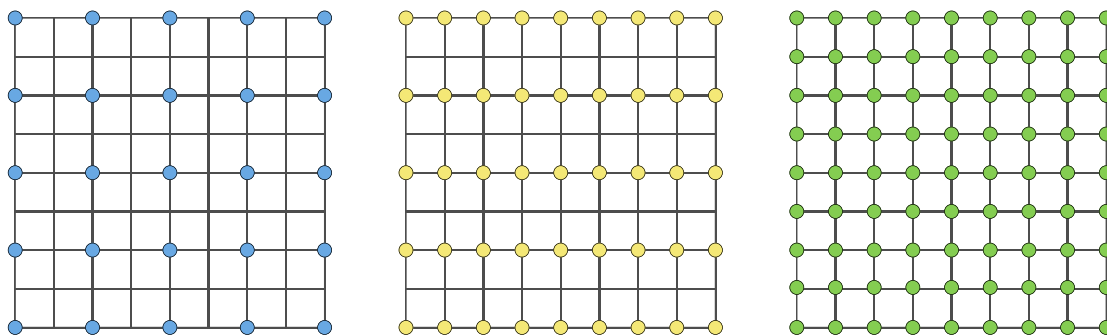


Рис. 1.1: Повышение размерности на основе интерполяции. Сетка обозначает координаты пикселей, голубые, желтые и зеленые точки представляют собой начальное, промежуточное и конечное положение пикселей, соответственно

Наиболее используемые методы интерполяции - интерполяция методом ближайшего соседа, билинейная и бикубическая интерполяции. Все эти методы хоть и справляются с задачей повышения разрешения за относительно короткое время, изображение в результате их применения по качеству часто хуже исходного.

Несмотря на свою непригодность в тех задачах, в которых требуется получить четкое изображение высокого разрешения, методы интерполяции часто используются для сравнения более способных алгоритмов. Также, как описано выше, методы повышения разрешения на основе интерполяции (а обычно именно бикубическая интерполяция) применяются в обратной задаче - в построении изображения низкого разрешения для создания пар $(LR; HR)$.

Различные стратегии построения и обучения нейронной сети. Архитектуры существующих моделей можно разделить на 4 модельных фреймворка на основе используемых операций повышения дискретизации и их местоположения в модели. Рассмотрим кратко каждую из них:

1. Предварительное увеличение разрешения (pre-upsampling). Сперва изображение подвергается повышающей дискретизации до некачественных HR изображений с желаемым размером, например, с помощью бикубической интерполяции (рис. 1.2). После, к полученному изображению применяются глубокие сверточные нейронные сети (Convolution Neural Networks, CNN) с целью обогащения отсутствующих текстур. Такая по-

следовательность действий была предложена в статье [17] и считается одной из самых популярных, однако предварительная дискретизация часто приводит к таким побочным эффектам, как усиление шума и размытие.

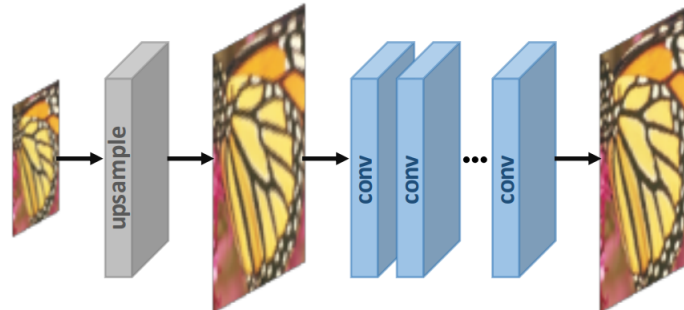


Рис. 1.2: Предварительное увеличение разрешения. Источник: [16]

2. Пост-увеличение разрешения (post-upsampling). Для повышения эффективности вычислений в статье [18] авторы предложили выполнять большую их часть в низкоразмерном пространстве. Сперва LR изображение подается на вход глубокой CNN, после чего применяется повышение дискретизации (рис. 1.3). Также считается одним из самых популярных фреймворков.

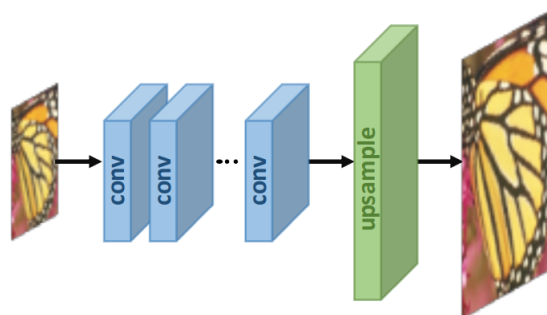


Рис. 1.3: Пост-увеличение разрешения. Источник: [16]

3. Прогрессивное увеличение разрешения (progressive upsampling). При увеличении изображения крайним методом во много раз (например, 4, 8), результат может оказаться очень некачественным. Для устранения

этого недостатка авторы статьи [19] использовали структуру прогрессивного повышения дискретизации (рис. 1.4). Модели в рамках этой структуры основаны на каскаде CNN и постепенно восстанавливают изображение с более высоким разрешением (на каждом этапе изображение увеличивается до более высокого разрешения и уточняется CNN). Такие модели отличаются сложностью проектирования и нестабильностью обучения.

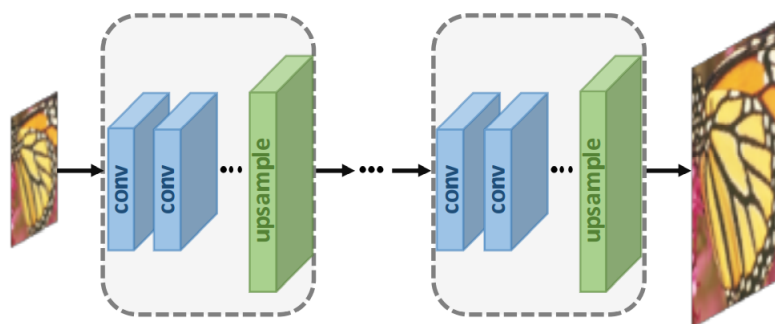


Рис. 1.4: Прогрессивное увеличение разрешения. Источник: [16]

4. Итеративное увеличение и уменьшение разрешения (iterative up-and-down sampling). Суть данного метода состоит в том, чтобы итеративно применять повышение и снижение дискретизации для уточнения обратной проекции, после чего результаты объединяются для настройки интенсивности изображения HR (рис. 1.5). Модели, использующие такую структуру, способны изучить глубокие взаимосвязи внутри пар (LR ; HR) и, таким образом, обеспечить наиболее качественные результаты реконструкции.

Помимо выбора места применения апсемплинга необходимо решить какую архитектуру блоков нейронной сети использовать. Ниже описаны некоторые из подходящих архитектур:

1. Остаточное обучение (residual learning). Поскольку задача улучшения качества изображений представляет собой задачу преобразования картинку в картинку, в которой входное изображение сильно коррелирует

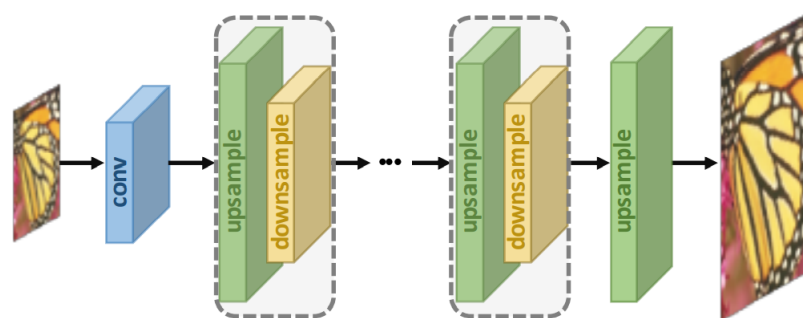


Рис. 1.5: Итеративное увеличение и уменьшение разрешения. Источник: [16]

с целевым изображением, авторы статьи [20] предложили изучать только разницу между ними. В этом случае появляется возможность избежать обучения сложному преобразованию одного изображения в другое. Вместо этого требуется обучить алгоритм вычислять остаточную карту признаков для восстановления недостающих высокочастотных деталей.

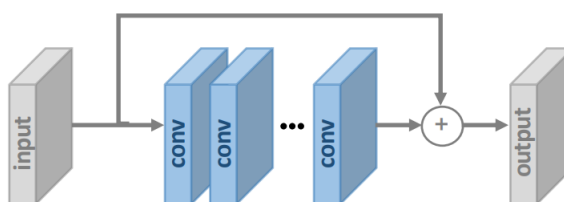


Рис. 1.6: Остаточное обучение. Источник: [16]

2. Полносвязное обучение (dense connections). Каждый слой полносвязного блока принимает в качестве входных данных карты признаков всех предыдущих слоев. Цель такого подхода состоит в объединении признаков низкого и высокого уровней для получения более общей информации о сцене изображения [21].
3. Вейвлет преобразование (wavelet transformation). Преобразование с использованием вейвлет является высокоэффективным представлением изображения путем разложения его сигнала на высокочастотные поддиапазоны (детали текстуры) и низкочастотные (глобальная топологическая информация). В статье [22] модель SR принимает поддиапазоны

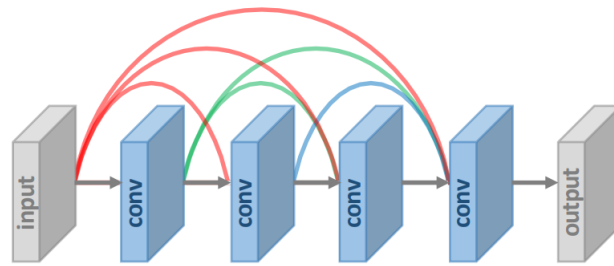


Рис. 1.7: Полносвязное обучение. Источник: [16]

интерполированного вейвлета LR в качестве входных данных и прогнозирует остатки соответствующих поддиапазонов HR. Прямое и обратное вейвлет-преобразования применяются для разложения входа LR и восстановления выхода HR соответственно. Благодаря эффективному представлению с помощью вейвлет-преобразования часто является возможным снизить вычислительные затраты, сохраняя при этом конкурентоспособную производительность.

4. Десубпиксель (desubpixel). Для увеличения скорости вычислений, было предложено выполнять трудоемкое извлечение признаков в пространстве меньшей размерности. Операция десубпикселя разделяет изображения пространственно, складывает их в виде дополнительных каналов и, таким образом, позволяет избежать потери информации. Таким образом, авторы статьи [23] снижают дискретизацию входных изображений с помощью десубпикселей в начале модели, изучают представления в пространстве с меньшей размерностью и повышают дискретизацию до целевого размера в конце. Предлагаемая модель способна в реальном времени получить целевое изображение в достаточно хорошем качестве.

Хорошая архитектура нейронной сети не только определяет пространство гипотез с хорошей верхней границей производительности, но и помогает эффективно изучать представления без излишней пространственной и вычислительной избыточности.

Далее будут рассмотрены различные стратегии обучения нейронной сети. Начать обзор решено с обсуждения используемых в задаче SR функ-

ций потерь. В области повышения разрешения изображений функции потерь используются для измерения ошибки их восстановления. Для лучшего измерения ошибки восстановления и получения более реалистичных и высококачественных результатов используются следующие функции потерь:

1. Пиксельная функция потерь (pixel loss). Описываемая функция потерь измеряет пиксельную разницу между двумя изображениями и, в основном, включает потерю L1 (среднюю абсолютную ошибку) и потерю L2 (среднеквадратичную ошибку):

$$Loss_{L1}(\hat{I}, I) = \frac{1}{hwc} \sum_{i,j,k} |\hat{I}_{i,j,k} - I_{i,j,k}|,$$

$$Loss_{L2}(\hat{I}, I) = \frac{1}{hwc} \sum_{i,j,k} (\hat{I}_{i,j,k} - I_{i,j,k})^2,$$

где \hat{I} и I - сравниваемые изображения (в данном случае увеличенное изображение LR и изображение HR), h , w и c - их высота, ширина и количество каналов. Пиксельная функция потерь ограничивает сгенерированное изображение HR так, чтобы оно было достаточно близко к LR по значениям пикселей. Поскольку определение PSNR (peak signal-to-noise ratio - метрика качества в задаче SR, будет описана ниже) сильно коррелирует с пиксельной разницей, а минимизация пиксельной функции потерь напрямую увеличивает PSNR, pixel loss является наиболее широко используемой функцией подобного рода. Однако в ее результате применения целевым изображениям часто не хватает высокочастотных деталей.

2. Потеря контента (content loss). Эта функция измеряет семантические различия между изображениями с помощью предварительно обученной нейронной сети классификации изображений (часто это VGG или ResNet). Обозначая эту сеть как ϕ , а извлеченные высокоуровневые представления на n -м уровне как $\phi^{(n)}(I)$, потеря контента описывается как евклидово расстояние между высокоуровневыми представлениями

двух изображений, как показано ниже:

$$Loss_{content}(\hat{I}, I, \phi, n) = \frac{1}{h_n w_n c_n} \sqrt{\sum_{i,j,k} (\phi_{i,j,k}^{(n)}(\hat{I}) - \phi_{i,j,k}^{(n)}(I))^2},$$

где h_n , w_n и c_n - высота, ширина и количество каналов представлений на слое n . В отличие от pixel loss, потеря контента способствует тому, чтобы выходное изображение было перцептивно похожим на целевое изображение, вместо того, чтобы заставлять их точно соответствовать пикселям. Таким образом, он дает визуально более заметные результаты и также широко используется в описываемой области.

3. Потеря реконструкции стиля (texture loss). Тот факт, что реконструированное изображение должно иметь тот же стиль (например, цвета, текстуры, контраст), что и у целевого изображения, стал причиной разработки потери текстуры. Текстура изображения рассматривается как корреляции между различными каналами признаков и определяется как матрица Грама:

$$G_{i,j}^n(I) = \text{vec}(\phi_i^{(n)}(I)) \times \text{vec}(\phi_j^{(n)}(I)),$$

где $\text{vec}()$ обозначает операцию векторизации, а $\phi_i^{(n)}(I)$ обозначает i -й канал карты признаков на слое n изображения I . Сама потеря реконструкции стиля определяется выражением:

$$Loss_{texture}(\hat{I}, I, \phi, n) = \frac{1}{c_n^2} \sqrt{\sum_{i,j} (G_{i,j}^n(\hat{I}) - G_{i,j}^n(I))^2},$$

Отдельного внимания заслуживает состязательная потеря (adversarial loss), подробно она будет описана в разделе 1.3.4 как часть генеративно состязательных моделей.

На практике описанные функции потерь часто взвешенно объединяются, так как все они по-разному влияют на реконструкцию изображения. Поэтому, одним из перспективных направлений в данной теме является изу-

чение потенциальной корреляции между изображениями и поиск наиболее точных функций потерь.

Следующая стратегия обучения нейронных сетей - пакетная нормализация (batch normalization) - вообще как таковым алгоритмом не является. Это метод адаптивной перепараметризации; считается одной из наиболее интересных новаций в области оптимизации глубоких нейронных сетей, появившейся из-за трудностей их обучения. Задачами нормализации считаются ускорение и стабилизация обучения глубоких CNN.

Другие возможные улучшения алгоритма решения задачи.

1. Аугментация данных. Одним из самых распространенных способов повышения эффективности нейронной сети в задачах компьютерного зрения считается увеличение количества ее обучающих данных. Аугментация способна решить описываемую проблему путем добавления в набор данных уже существующих объектов, но с некоторыми модификациями. Под модификациями в данном случае понимаются такие преобразования как отражение изображений, поворот по различным осям, масштабирование, выделение фрагмента, колебание цветов и другие.
2. Контекстное слияние сетей (context-wise network fusion). Аналогично существованию случайного леса в классическом машинном обучении в глубоком обучении существуют свои методы объединения предсказаний нескольких нейронных сетей. Конкретнее - каждая из моделей SR отдельно обучается со своей уникальной архитектурой, после чего все предсказания передаются отдельным сверточным слоям, выходы которых суммируются для получения финального предсказания.

Фактически, большинство state-of-the-art SR моделей в основном можно отнести к комбинации нескольких стратегий, описанных выше. Например, самая большая производительность сетей RCAN (Residual Channel Attention Networks) достигается с помощью комбинации таких методов как механизм внимания, остаточное обучение, повышение дискретизации с использованием субпикселя (subpixel upsampling) и попиксельная L1 функция потерь. Выбор

подходящих стратегий и их верная комбинация помогают не только увеличить точность алгоритма, но и повысить его эффективность.

1.3.2 Методы обучения без учителя

Существующие SR исследования фокусируются на обучении с учителем, то есть на обучении с использованием тренировочного набора, состоящего из пар (LR ; HR). Как было описано ранее, проблема отсутствия пар изображений решается деградацией изображений в высоком разрешении. Тем самым появляется новая сложность - создать объект в низком разрешении, полностью отражающий все проблемы LR объектов в реальном мире: полученное изображение должно содержать достаточное, но не чрезмерное количество шума, если требуется обучить алгоритм ставить изображениям “заплатки”, то необходимо создать картинку с намеренными пропусками. Для того, чтобы избежать вышеописанных проблем, исследователи все больше внимания уделяют обучению SR моделей без учителя, где для обучения предоставляются непарные изображения в различных разрешениях. Далее описаны несколько из проведенных исследований неконтролируемого подхода повышения разрешения изображений:

1. Так как предварительная деградация неоптимальна, в статье [24] исследователи сперва обучают алгоритм понижению дискретизации с помощью небольшого количества парных (HR ; LR) изображений, что дает возможность создавать релевантные по отношению к реальному миру изображения в низком разрешении. Далее, для построения алгоритма, способного создавать изображения HR из полученных ранее изображений LR , была обучена генеративно-сопоставительная сеть.
2. Следующий метод заключается в рассмотрении LR и HR пространств, как двух разных доменов, с последующим изучением сопоставлений между ними. В статье [25] исследователи предложили CinCGAN (cycle-in-cycle GAN), состоящий из 4-х генераторов и 2-х дискриминаторов. Для первого CycleGAN зашумленное LR изображение подается на вход генератору и от выхода требуется, чтобы он согласовывался с реальным

распределением “чистых” (не зашумленных) LR изображений. Затем полученное изображение подается на вход другому генератору и требуется восстановить вход (зашумленное LR изображение). Для обучения используются несколько функция потерь - adversarial loss, cycle consistency loss и identity loss. Необходимо это для обеспечения согласованности цикла, согласованности распределения и достоверности отображения. Второй CycleGAN построен аналогично, за исключением того, что области отображения различны. CinCGAN достигает сравнимых результатов с методами обучения с учителем.

Несмотря на успехи описанных подходов неконтролируемого обучения, чаще используются методы обучения с учителем по причине нехватки данных для построения алгоритма HR-to-LR. Однако эта область активно развивается и считается многообещающей для будущего развития. Данная работа основана на методах обучения с учителем.

1.3.3 Применение алгоритма повышения разрешения к задачам детекции и сегментации опухолей

Одной из многочисленных проблем в анализе медицинских снимков является низкое анизотропное (т.е. различное) разрешение данных [29]. Причина состоит в том, что низкое разрешение ограничивает способность врачей хорошо рассмотреть снимок и верно поставить диагноз - очень маленькие опухоли могут быть незаметны. Поэтому, в большинстве случаев, данные сначала проходят через повышающую дискретизацию. Однако увеличение изображения плохо сказывается на его качестве, что может только ухудшить ситуацию. В процессе решения этой проблемы используются методы предварительной обработки для улучшения качества изображения и удаления мелких артефактов и шума для точного обнаружения аномалий на медицинских снимках.

После проведения реконструкции снимка специалисту намного легче рассмотреть изображение в деталях и, тем самым, составить наиболее точный план лечения или операции. Помимо того, что снимок большего размера и лучшего качества благоприятно влияет на верность решения врачей, такая

реконструкция может оказаться решающим действием для анализа структур изображения в задачах детекции и сегментации.

1.3.4 Генеративно-сопязательные сети

Сперва необходимо описать генеративные модели в целом. Генеративные модели делятся на две категории - вариационные автокодировщики (Variational AutoEncoders, VAE) и генеративно-сопязательные сети (Generative Adversarial Networks, GAN). VAE состоит из кодировщика и декодера, где задача первого - сжать входное пространство признаков для его представления в скрытом пространстве, а цель второго - воссоздать исходное пространство с некоторыми модификациями. GAN состоит из генератора (аналог декодера) и дискриминатора. Генератор получает на вход вектор случайного шума (как получает некоторый вектор декодер после кодировщика) и пытается сгенерировать реальное пространство признаков с помощью дискриминатора. Используются генеративные модели в таких сервисах, как Adobe (для создания фильтров фотошопа), Google (для генерации текстов), IBM (для создания большего количества новых данных), Snapchat и TikTok (для фильтрации изображений), Disney (для повышения разрешения). Далее подробно описана идея и особенности генеративно-сопязательных сетей, которые активно используются в реализации к данной работе.

GAN - структура для оценки генеративных моделей через сопязательный процесс, в котором одновременно обучаются две нейронные сети - генератор G (генеративная модель, пытается сгенерировать объекты из реального распределения данных или, что тоже самое, воссоздать признаки по классу: $P(X|Y)$) и дискриминатор D (классифицирующая нейронная сеть, оценивает вероятность того, что сгенерированный объект принадлежит обучающим данным, а не генератору: $P(Y|X)$) [31] [32]. Обучение генеративной модели нацелено на максимизацию вероятности того, что классификационная модель ошибется. Для этого поочередно выполняются два шага: (а) исправить генератор и обучить дискриминатор, чтобы лучше различать изображения, (б) исправить дискриминатор и обучить генератор, чтобы лучше генерировать изображения (то есть с большей вероятностью обманывать дискрими-

натор). Благодаря описанному итеративному обучению противников, результирующий генератор способен выдавать выходные данные, наиболее близкие к распределению реальных данных, в то время как дискриминатор к концу обучения становится хорошим классификатором реальных и сгенерированных данных. Конкуренция двух моделей дает возможность обеим улучшать свою производительность до тех пор, пока сгенерированные объекты станут неотличимы от реальных образцов. Если рассмотреть описанную структуру обучения в терминах задачи повышения разрешения, то генератором будет считаться сама модель SR, а дискриминатор необходимо определить.

Подробнее о связной работе генератора и дискриминатора. Для того, чтобы узнать распределение p_g генератора по данным x , следует определить априорный входной шум $p_z(z)$, затем представить отображение в пространство данных как $G(z; \theta_g)$, где G - дифференцируемая функция, представленная сверточной нейронной сетью с параметрами θ_g . Также следует определить вторую сверточную нейронную сеть $D(x; \theta_d)$, которая выводит одно скалярное значение - вероятность того, что x получено из реальных данных, а не из p_g . Дискриминатор обучается путем максимизации вероятности присвоения верной метки как реальным данным, так и выборке из G . Одновременно с этим происходит обучение генератора, цель которого минимизировать $\log(1 - D(G(z)))$. В литературе такое противостояние двух субъектов называется минимаксной игрой - участники имеют противостоящие интересы (в данном случае дискриминатор нацелен на увеличение вероятности верных ответов, в то время как генератору необходимо эту вероятность уменьшить). Целевая функция минимаксной задачи в вышеописанных обозначениях выглядит следующим образом:

$$\min_G \max_D V(D, G) = E_{x \sim p_{data}(x)} \log D(x) + E_{z \sim p_z(z)} \log(1 - D(G(z))),$$

где $V(D, G)$ - функция значений. Минимаксная игра имеет глобальный минимум тогда и только тогда, когда $p_g = p_{data}$ (распределение генератора не отличить от распределения реальных данных).

Решение поставленной задачи обычно реализуется с использованием численного итеративного подхода. На практике, численное вычисление урав-

нения целевой функции может не привести к градиенту, достаточному для хорошего обучения G . В начале обучения, когда G не обучен (генерирует близкие к шуму изображения), D может отвергать его образцы с высокой уверенностью, потому что они четко отличимы от реальных данных. В этом случае $\log(1 - D(G(z)))$ насыщается. Вместо обучения G минимизацией $\log(1 - D(G(z)))$, в таком случае лучше обучать его максимизацией $\log(D(G(z)))$. Это обеспечивает более сильные градиенты на ранних этапах обучения.

Более подробное описание алгоритма обучения (рис. 1.8). Генеративно-состязательные сети обучаются путем обновления распределения дискриминатора D (голубая линия) так, чтобы отличать образцы из реальных данных p_x (черная) от образцов распределения генератора p_g (зеленая).

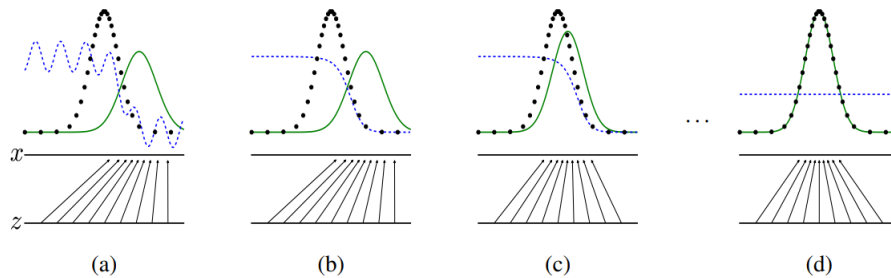


Рис. 1.8: Синяя пунктирная - дискриминатор, черная пунктирная - распределение реальных данных, зеленая - распределение сгенерированных данных. Нижняя горизонтальная линия - это область, из которой выбирается z (шум). Верхняя горизонтальная линия - часть области реальных данных x . Направленные вверх стрелки показывают отображение $x = G(z)$. G сжимается в областях с высокой плотностью и разжимается в областях с низкой плотностью p_g . (a) Рассматривается состязательная пара, близкая к сходимости: p_g похож на p_{data} и D - частично точный классификатор. (b) D обучается различать реальные данные сходясь к $D^*(x) = \frac{p_{data}(x)}{p_{data}(x) + p_g(x)}$. (c) После обновления G , градиент D направляет $G(z)$ в регионы, где с большой вероятностью данные будут классифицированы как реальные. (d) После нескольких шагов обучения, если G и D достаточно обучились, они достигнут точки, когда оба не смогут стать еще лучше, потому что $p_g = p_{data}$. D не способен различить два распределения, то есть $D(x) = 0.5$.

Функция потерь. Функция потерь, используемая для обучения генеративно-состязательной сети, состоит из двух частей - потери реконструкции и состязательной потери. Первая рассчитывает расстояние между реальным и сгенерированным изображениями, тогда как состязательная потеря выражает описанную выше минимаксную игру между генератором и дискриминатором,

оценивая то, насколько реалистичное изображение создал генератор.

Архитектуры генератора и дискриминатора. Генератор основан на сверточных слоях и содержит остаточные блоки, то есть представляет собой подобие ResNet. На входе он принимает интерполированное изображение LR . Дискриминатор представляет собой полностью сверточную нейронную сеть.

Сравнение с другими подходами. Обширные тесты MOS показывают [16], что даже несмотря на то, что модели SR, обученные с состязательной потерей и потерей реконструкции, достигают более низкого PSNR по сравнению с моделями, обученными с потерей пикселей - они приносят значительный выигрыш в качестве восприятия. Фактически, дискриминатор извлекает некоторые трудные для изучения скрытые паттерны реальных изображений HR и подталкивает сгенерированные изображения HR к более реалистичным признакам. Преимущество использования GAN состоит в том, что его можно оптимизировать без заранее разработанной функции потерь для конкретной задачи. Также, важным достоинством генеративно состязательных сетей считается возможность не использовать в обучении цепи Маркова (вместо этого для обновления градиентов используется метод обратного распространения ошибки), что вычислительно менее дорого.

1.4 Обзор методов оценки качества изображений

Качество изображений относится к визуальным характеристикам изображений и фокусируется на оценках восприятия зрителей. Как правило, методы оценки качества изображения (Image Quality Assessment, IQA) включают субъективные методы (основанные на восприятии людей) и объективные вычислительные методы. Первое больше соответствует нашим потребностям, но на это требуется немалое количество временных и денежных ресурсов, поэтому второе в настоящее время является основным направлением. Однако эти методы не обязательно согласованы между собой, потому что объективные методы часто не могут достаточно точно уловить визуальное восприятие человека, что может привести к большой разнице в результатах оценки. Кроме того, вычислительные методы IQA делятся на три типа: методы с полным эталоном, выполняющие оценку с использованием реальных изображений, методы с уменьшенным эталоном, основанные на сравнении извлеченных функций, и методы без эталона (слепой IQA). Далее рассмотрены субъективный метод оценки качества изображений и наиболее используемые объективные методы с полным эталоном.

1. Максимальное отношение сигнала к шуму (Peak Signal-to-Noise Ratio, PSNR). PSNR считается одним из самых популярных методов измерения качества реконструкции. Описываемая метрика определяется через максимальное значение пикселя L и среднеквадратичную ошибку между двумя изображениями. Сравнивая исходное изображение I с N пикселями и реконструкцию \hat{I} с тем же количеством пикселей, метрику между I и \hat{I} можно записать следующим образом:

$$PSNR = 10 \times \log_{10} \frac{L^2}{\frac{1}{N} \sum_{i=1}^N (I(i) - \hat{I}(i))^2},$$

где $L = 255$ (в случаях с использованием 8-битных представлений, т.е. для изображений 256 цветов). Поскольку PSNR берет в расчет только различия между соответствующими пикселями и не учитывает визуальное восприятие, это часто приводит к нерепрезентативным результатам на изображениях с реальными сценами, то есть в тех случаях, где нам

особенно важно восприятие людей.

2. Индекс структурного сходства (Structural Similarity Index, SSIM). Так как зрительная система человека (Human Visual System, HVS) хорошо приспособлена к извлечению структур изображений, логично учитывать при оценке качества яркость и контрастность изображения. С этой целью в статье [30] был предложен и успешно применен индекс структурного сходства, который сравнивает изображения не только с точки зрения яркости и контрастности, но и учитывает их структуру. Для изображения I с N пикселями яркость μ_I и контраст σ_I оцениваются как среднее и стандартное отклонение интенсивности изображения, соответственно, то есть $\mu_I = \frac{1}{N} \sum_{i=1}^N (I(i))$ и $\sigma_I = (\frac{1}{N-1} \sum_{i=1}^N (I(i) - \mu_I)^2)^{\frac{1}{2}}$, где $I(i)$ - интенсивность i -го пикселя изображения I . Сравнения яркости, контрастности и структуры, обозначенные как $C_l(I, \hat{I})$, $C_c(I, \hat{I})$ и $C_s(I, \hat{I})$ соответственно, определяются следующим образом:

$$C_l(I, \hat{I}) = \frac{2\mu_I\mu_{\hat{I}} + C_1}{\mu_I^2 + \mu_{\hat{I}}^2 + C_1}, \quad C_c(I, \hat{I}) = \frac{2\sigma_I\sigma_{\hat{I}} + C_2}{\sigma_I^2 + \sigma_{\hat{I}}^2 + C_2},$$

$$C_s(I, \hat{I}) = \frac{\sigma_{I\hat{I}} + C_3}{\sigma_I\sigma_{\hat{I}} + C_3},$$

где C_1, C_2, C_3 - константы устойчивости, $\sigma_{I\hat{I}}$ - ковариация между I и \hat{I} . Наконец, SSIM определяется следующим образом:

$$SSIM(I, \hat{I}) = C_l(I, \hat{I})^\alpha C_c(I, \hat{I})^\beta C_s(I, \hat{I})^\gamma,$$

где α, β, γ - управляющие параметры для регулировки относительной важности. Поскольку SSIM оценивает качество реконструкции с точки зрения HVS, он лучше PSNR отвечает требованиям визуальной оценки, и также широко используется.

3. Средняя экспертная оценка (Mean Opinion Score, MOS). MOS - широко используемый субъективный метод IQA, для реализации которого люди-оценщики присваивают тестируемым изображениям оценки качества восприятия. Чаще всего оценки принимают значения от 1

(плохое качество) до 5 (хорошее качество). Средняя экспертная оценка равна среднему арифметическому полученных значений. Несмотря на то, что MOS интуитивно кажется верной метрикой, ему присущи такие недостатки, как нелинейное восприятие шкал, смещение и дисперсия критериев оценки.

4. Оценка на основе задач (task-based evaluation). В тех случаях, когда алгоритм решения задачи повышения разрешения применяется для более качественного решения других задач компьютерного зрения, оценку эффективности построенной модели необязательно вычислять при помощи метрики. Нам важно не то, насколько повысилось качество изображения после работы алгоритма, а увеличилась ли точность решения целевых задач. Рассмотрим детекцию в роли задачи, производительность которой хотим улучшить с помощью алгоритма повышения разрешения. Сперва необходимо вычислить долю правильно классифицированных LR изображений. После чего следует сделать то же самое для полученных после алгоритма SR изображений HR . Если качество повысилось - построенный алгоритм справился со своей задачей.

Несмотря на то, что PSNR склоняется к чрезвычайно размытому результату, а SSIM не способен точно оценить качество восприятия, они считаются наиболее широко распространенными метриками в задачах SR. Субъективный метод MOS хоть и ближе всех остальных к человеческим требованиям, он требует больших затрат.

Из вышеописанного обзора можно сделать вывод, что единой и всеми признанной метрики оценки качества в задачах SR не существует. Если улучшение качества изображения / видео является целью использования алгоритма SR, то целесообразно использовать, если позволяют ресурсы, среднюю экспертную оценку. Если таковых недостаточно - хорошим выбором станет комбинация метрик PSNR и SSIM. В данном случае - при применении алгоритма повышения разрешения для улучшения производительности проблем детекции и сегментации - решено использовать оценку на основе задач.

Глава 2. Разработка программного комплекса для детекции и сегментации опухолей

2.1 Проектирование архитектуры программного комплекса

Ниже описан пошаговый план проектирования архитектуры для построения решения поставленной задачи (то есть задачи повышения качества детекции и сегментации опухолей с помощью методов увеличения разрешения используемых в этих задачах изображений). Реализация будет состоять из трех основных частей: построение алгоритма повышения качества изображений, тестирование алгоритма на задаче детекции, тестирование алгоритма на задаче сегментации. Сперва необходимо описать данные, с которыми предстоит работать.

Набор данных. Задача SR сама по себе не особенно требовательна к данным по сравнению с другими задачами компьютерного зрения - необходимы лишь изображения одного размера (хотя и это не обязательно) в относительно большом разрешении. Однако детекция и сегментация добавляют некоторые нюансы - для детекции опухолей требуются изображения с новообразованиями и без них, для сегментации же необходимо иметь изображения с опухолями и их масками. Учитывая вышеописанные ограничения, подходящий для поставленной задачи набор данных состоит из изображений с опухолями и их масками, и изображений без опухолей, все хорошего качества.

Для данной работы было решено использовать набор данных “Brain MRI Segmentation” с сайта kaggle [36]. Он состоит из 110 папок, каждая из которых содержит в себе данные одного пациента - снимки с опухолью (если пациент болен) или снимки без опухоли (если здоров) и их маски (если изображение не содержит в себе опухоль, то соответствующая маска - изображение черного цвета без каких-либо пятен). Однако “пустые” маски нам не понадобятся - сегментация будет производиться только на изображениях с опухолями, поэтому их решено не использовать. Всего в наборе данных содержится 1373 изображения с опухолями, их маски и чуть больше 2500 изображений без опухолей, все изображения одного размера - 256×256 пикселей.

Так как данная задача никак не использует знания о пациентах соответ-

ствующим изображениям, было решено отсортировать имеющиеся снимки по 3-м папкам - снимки без опухолей (далее *clean*), снимки с опухолями (далее *tumor*) и с их масками (далее *masks*). Необходимо учесть, что номера изображений в папках *tumor* и *masks* должны соответствовать (маска с номером 1 в папке *masks* должна принадлежать снимку 1 в папке *tumor*). Для наиболее интерпретируемой оценки качества алгоритма детекции решено уравнивать количество изображений в папках с опухолями и без них. В результате, каждая из них будет содержать 1373 изображения (очевидно, в папке *masks* также будет 1373 изображения).

Для упрощения следующих этапов разработки к каждой из трех папок *tumor*, *masks* и *clean* необходимо создать еще две - *tumor_train* и *tumor_test* для первой и так далее. Нужно это для того, чтобы после обучения алгоритма сгенерированные изображения (далее *HR*), изображения в низком разрешении (далее *LR*) и изображения изначальные (далее *GT*, где *GT* - Ground Truth) можно было сохранить в 3 соответствующие их происхождению папки с одинаковыми индексами (сгенерированное изображение с номером 1 в папке *HR* должно соответствовать изображению 1 в папке *LR*, которое в свою очередь соответствует изображению 1 в папке *GT*). Сохранить их с соответствием, в свою очередь, следует для того, чтобы в задачах детекции и сегментации протестировать алгоритмы на полученных данных (*HR*), на изначальных (*LR*) и на эталонных (*GT*). То есть стоит цель ответить на вопрос: "каким образом сгенерированные изображения влияют на эффективность решения задач детекции и сегментации по сравнению с изображениями в низком разрешении и эталонами?".

Повышение качества изображений. Данный блок реализации состоит из нескольких этапов, которые в подробности будут описаны в следующем разделе. Сперва изображения для обучения из папок *tumor_train* и *clean_train* следует подвергнуть понижению дискретизации (получить снимки в низком разрешении). После чего алгоритм на полученных парах обучается. В результате тестирования алгоритма, для каждой из папок *tumor_test* и *clean_test* необходимо сохранить сгенерированные изображения в высоком разрешении и изображения в низком разрешении (их номера будут соответствовать номерам изображений в папках *tumor_test* и *clean_test*, то есть эталонам).

После применения алгоритма SR имеется 10 папок: 1) *tumor* (все изображения с опухолями); 2) *clean* (все изображения без опухолей); 3) *masks* (все маски для изображений с опухолями); 4) *tumor_test*; 5) *tumor_LR* (подверженные понижению дискретизации изображения *tumor_test*); 6) *tumor_HR* (сгенерированные алгоритмом изображения из *tumor_test*); 7) *masks_test* (маски изображений *tumor_test*); 8) *clean_test*; 9) *clean_LR* (подверженные понижению дискретизации изображения *clean_test*); 10) *clean_HR* (сгенерированные алгоритмом изображения из *clean_test*) (рис. 2.1).

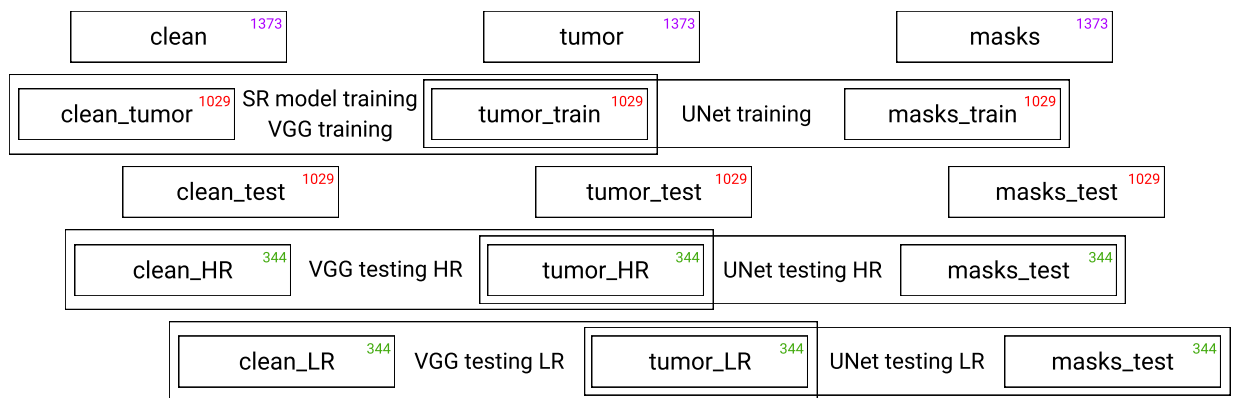


Рис. 2.1: Структура хранения папок с данным. Цветом обозначено количество изображений в соответствующей папке. Ссылка на Google Диск: [37]

Детекция. Для того, чтобы избежать погрешностей в обучении, для детекции опухолей решено с нуля обучить нейронную сеть на тренировочных данных *tumor_train* и *clean_train*. Для сравнения качества работы алгоритма на различных наборах данных (на изображениях в низком разрешении, на сгенерированных изображениях в высоком разрешении и, для сравнения с эталоном, на изначальных изображениях в высоком разрешении) изображения будут выбираться из соответствующих папок. Например, для вычисления качества работы алгоритма детекции на сгенерированных данных, изображения следует "достать" из папок *clean_HR* и *tumor_HR*. Сравнение полученных значений точности классификации дает возможность судить о производительности построенного алгоритма SR.

Сегментация. Аналогично детекции, для решения задачи сегментации будет использована с нуля обученная нейронная сеть на данных из *tumor_train* и *masks_train*. Сравнение качества сегментации так же будет

производиться на трех наборах данных - *tumor_test*, *tumor_HR* и *tumor_LR* (маски для каждого из них одни и те же - *masks_test*).

На рисунке 2.2 представлена визуализация описанной архитектуры программного комплекса.

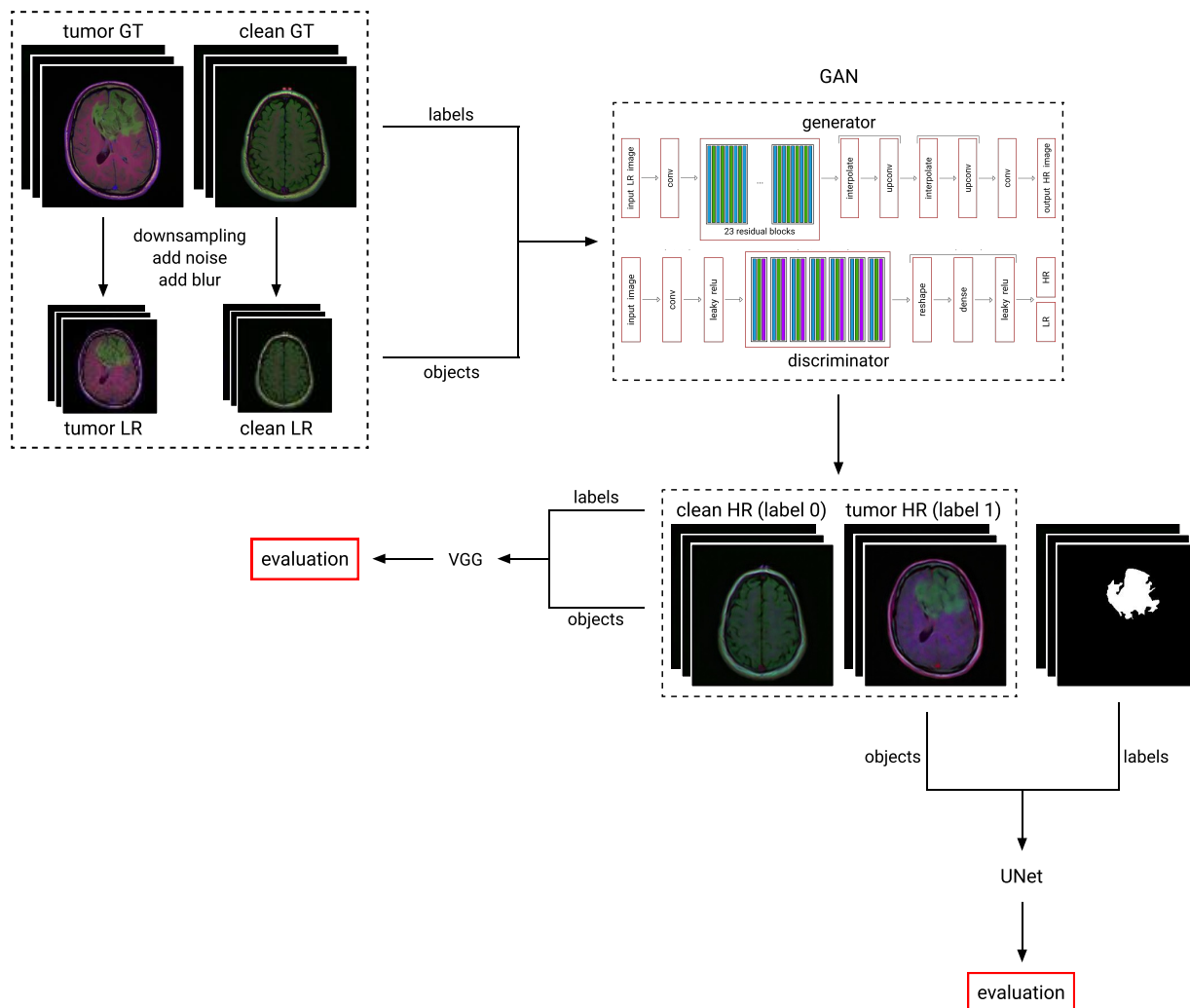


Рис. 2.2: Визуализация архитектуры программного комплекса

2.2 Разработка и реализация методов на основе описанных моделей

Как было указано в разделе 1.3.4, классические генеративно-состязательные сети состоят из ResNet-подобного генератора и сверточной нейронной сети (дискриминатора). Данная работа как раз основана на классическом варианте с некоторыми модификациями.

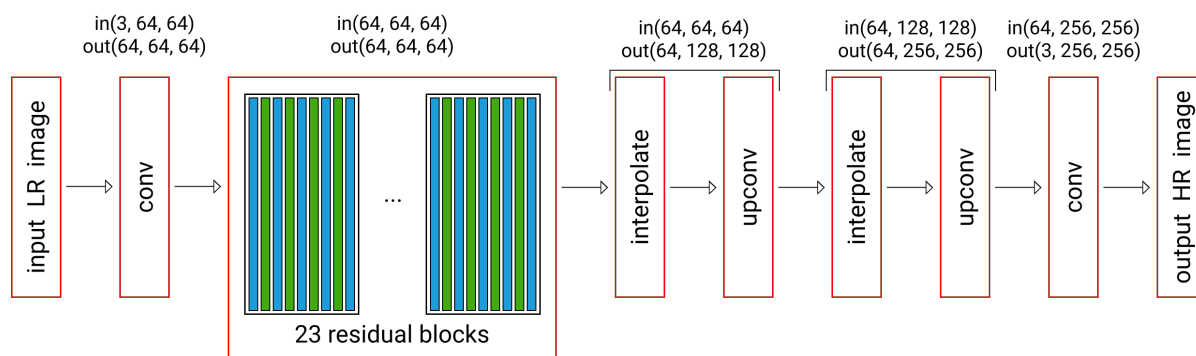


Рис. 2.3: Архитектура генеративной нейронной сети. Синим обозначены сверточные слои, зеленым - слои линейной ректификации. Над слоями указаны размерности входного тензора и выходного после их применения

Генератор (см. рис. 2.3). Сперва входное изображение LR размерности (3, 64, 64) "проходит" через слой свертки с параметрами: (input channels = 3, output channels = 64, kernel size = 3, stride = 1, padding = 1), далее параметры сверточных слоев будут использоваться без указаний (т.е. в данном случае - (3, 64, 3, 1, 1)). После полученный тензор поступает на вход последовательности из 23-х остаточных блоков. Каждый из них состоит из 5 сверточных слоев с параметрами $(64 + i * 32, 32, 3, 1, 1)$, где $i \in [0, 1, 2, 4]$ - номер слоя, при $i = 3$ параметры $(64 + i * 32, 64, 3, 1, 1)$. В отличие от первого слоя, остальные слои применяются к конкатенации всех предыдущих выходов слоев (т.е. третий слой свертки принимает на вход конкатенацию выходов первого и второго слоев), проиллюстрировано на рисунке 2.4. К каждому из четырех первых слоев свертки впоследствии применяется слой линейной ректификации. Полученный результат умножается на эмпирическим путем выбранную константу и прибавляется к входному тензору, что дает итоговый результат с размерностью (64, 64, 64). Полученная конструкция применяется ко входу 3 раза и, тем самым, представляет собой готовый остаточный блок. После того,

как тензор "прошел" через все 23 остаточных блока, его размерность осталась прежней - (64, 64, 64). Далее тензор 2 раза попеременно интерполируется и поступает на вход транспонированной свертке. После применения последней свертки (уменьшающей количество каналов с 64-х до 3-х) размерность итогового изображения составляет (3, 256, 256), что в 4 раза больше входа. В отличие от классической архитектуры генератора данная концепция оснащена более объемным остаточным блоком. Количество блоков подбиралось эмпирическим путем. Также, последние слои интерполяции и транспонированной свертки добавлены для повышения разрешения изображения.

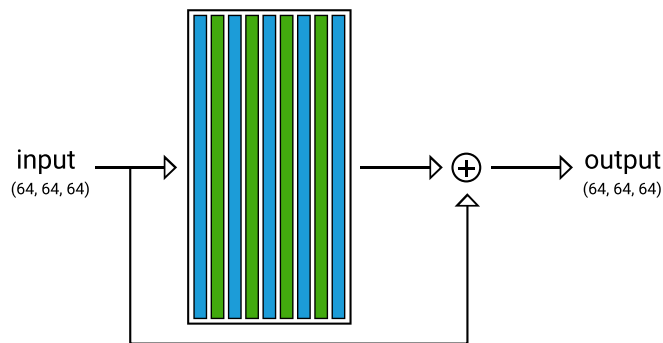


Рис. 2.4: Архитектура остаточного блока. Синим обозначены сверточные слои, зеленым - слои линейной ректификации, знаком сложения обозначена операция конкатенации

Дискриминатор (см. рис. 2.5). В отличие от генератора, дискриминатор принимает на вход изображения разной природы размера (3, 256, 256) - изначальные и сгенерированные. После первого слоя свертки (3, 64, 3, 1, 1) и

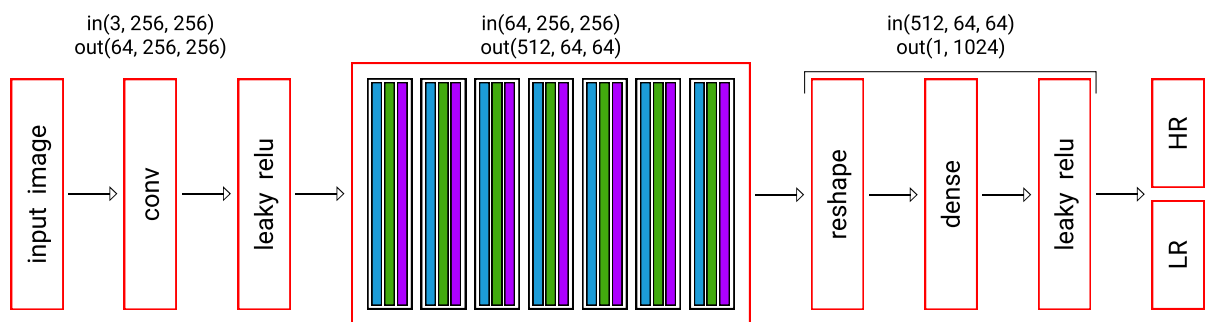


Рис. 2.5: Архитектура дискриминатора. Синим обозначены сверточные слои, зеленым - слои линейной ректификации, фиолетовым - слои нормализации. Над слоями указаны размерности входного тензора и выходного после их применения

следующего за ним слоя линейной ректификации, тензор размером уже (64,

256, 256) поступает на вход семи одинаковым блокам, каждый из которых состоит из слоя свертки $(n, n, 3, 2, 1)$ для нечетных блоков и слоя свертки $(n, n \times 2, 3, 1, 1)$ для четных, где n - количество каналов после применения предыдущих слоев, слоев линейной ректификации и нормализации. В результате получается тензор размера $(512, 64, 64)$. Далее тензор приводится к размеру $(1, 512 \times 64 \times 64)$, который в свою очередь является входом для полносвязного слоя и слоя линейной ректификации. Полученный вектор считается выходом дискриминатора и, с помощью функции потерь, вычисляется разница между выходом, где вход - изначальное изображение и выходом, где вход - сгенерированное.

2.3 Описание гиперпараметров обучения

С математической точки зрения алгоритм обучения генеративно-сопоставительной сети был описан в разделе 1.3.4, а так как сегментация и детекция не являются темой данной работы, обучение построенных алгоритмов описываться более не будет. Тем самым в данном разделе будут описаны лишь параметры обучения 4-х нейронных сетей - генератора, дискриминатора, сети UNet для сегментации и сети VGG для детекции.

Обучение генератора и дискриминатора основано на методе обратного распространения ошибки (backpropagation) с использованием оптимизатора Adam. Значение параметра скорости обучения является константным и равняется $1e-4$. Функция потерь генератора состоит из взвешенной суммы трех функций потерь: средняя квадратичная ошибка для оценки расстояния сгенерированного изображения от Ground Truth изображения, средняя квадратичная ошибка для вычисления расстояния между картами признаков сети VGG и комбинация сигмоидальной функции и бинарной кросс-энтропии - BCEWithLogitsLoss в PyTorch [38] - в качестве сопоставительной потери:

$$\begin{aligned} generator_loss = & \alpha \times \log(1 - D(G(LR))) + \beta \times \frac{1}{N} \sum_{i=1}^N (HR(i) - GT(i))^2 + \\ & + \gamma \times \frac{1}{N} \sum_{i=1}^N (HR_feature_map(i) - GT_feature_map(i))^2, \end{aligned}$$

где $HR_feature_map$ и $GT_feature_map$ - полученные карты признаков для сгенерированного и изначального изображений. Функция потерь дискриминатора включает в себя только сопоставительную потерю:

$$discriminator_loss = \log(D(HR)) + \log(1 - D(G(LR))).$$

Размер батча подбирался экспериментальным путем, наиболее оптимальным оказалось значение 32. Обучение производилось на протяжении 10 эпох. Время, затраченное на прохождение одной эпохи, в среднем, стремилось к часу.

Обучение UNet также основано на методе backpropagation с оптимизатором Adam. Используемая функция потерь - DiceLoss, для ее вычисления

следует сперва рассчитать метрику DiceScore [34] (немного отличающаяся метрика качества в задаче сегментации от также используемой в данной работе - IoU (Intersection over Union) [39]) и отнять полученное значение от единицы.

$$DiceScore = \frac{2 \times (A \cap B)}{A + B}, \quad IoU = \frac{A \cap B}{A \cup B},$$

где A и B - маски изображений. Обучающая выборка предварительно была разделена на тренировочную и валидационную выборки 80% и 20% соответственно. Оптимальный размер батча - 16, количество эпох - 200, обучение производилось на протяжении 5 часов.

Обучение VGG, аналогично предыдущим моделям, построено на методе обратного распространения с оптимизатором Adam. В качестве функции потерь была использована кросс-энтропия - CrossEntropyLoss в PyTorch [40]. Данные были разделены на две выборки: 80% от общего количества для обучения и 20% для валидации. Обучение производилось на протяжении 10 эпох, время обучение - 30 минут, размер батча - 16.

2.4 Тестирование и апробация разработанного решения

2.4.1 Постановка эксперимента

Для оценки эффективности использования техник повышения качества медицинских изображений в задачах детекции и сегментации опухолей был поставлен эксперимент, в рамках которого на тестовом наборе данных требовалось получить значения метрик качества IoU для сегментации и f1 score для детекции.

Подготовка данных. Для получения изображений в низком разрешении были предприняты следующие характерные для медицинских снимков трансформации: подвержение уменьшающей дискретизации в четыре раза, размытие снимка и добавление спекл-шума. Также, учитывая тот факт, что в скомпанованном наборе данных около 2000 (в папках *clean* и *tumor* примерно по 1000 изображений) для его увеличения будет применена аугментация с специфичными в области медицины модификациями, описанным в разделе 1.2. В результате применения процесса аугментации набор данных увеличился в два раза: чуть более 4000 снимков в высоком разрешении и столько же в низком.

2.4.2 Результаты эксперимента

Сегментация. В таблице 2.1 приведены значения метрики IoU для трех наборов данных - LR, HR, GT. Из таблицы видно, что использование методов повышения разрешения изображений улучшило качество сегментации опухолей на 2.7% по сравнению с подходом, не учитывающим предобработку данных.

Таблица 2.1: Значение метрики IoU на тестовых наборах данных

	LR	HR	GT
IoU	0.787	0.814	0.899

На рис. 2.6 представлен пример визуализации численных результатов (изображение GT, соответствующая ему маска, маски, сгенерированные из изображений LR, HR и GT соответственно). Видно, что о качестве сегментации опухолей в целом на отдельных изображениях судить нельзя. Сгенерированное алгоритмом повышения разрешения изображение может как превосходить соответствующее изображение в низком разрешении по эффективности применения алгоритма сегментации, так и быть хуже. Примеры на рисунках выбирались с целью показать о различных ситуациях, однако, так как общая оценка на изображениях HR выше, чем на изображениях LR, то примеров с положительным результатом (когда качество сегментации на изображении HR выше, чем на изображении LR) больше.

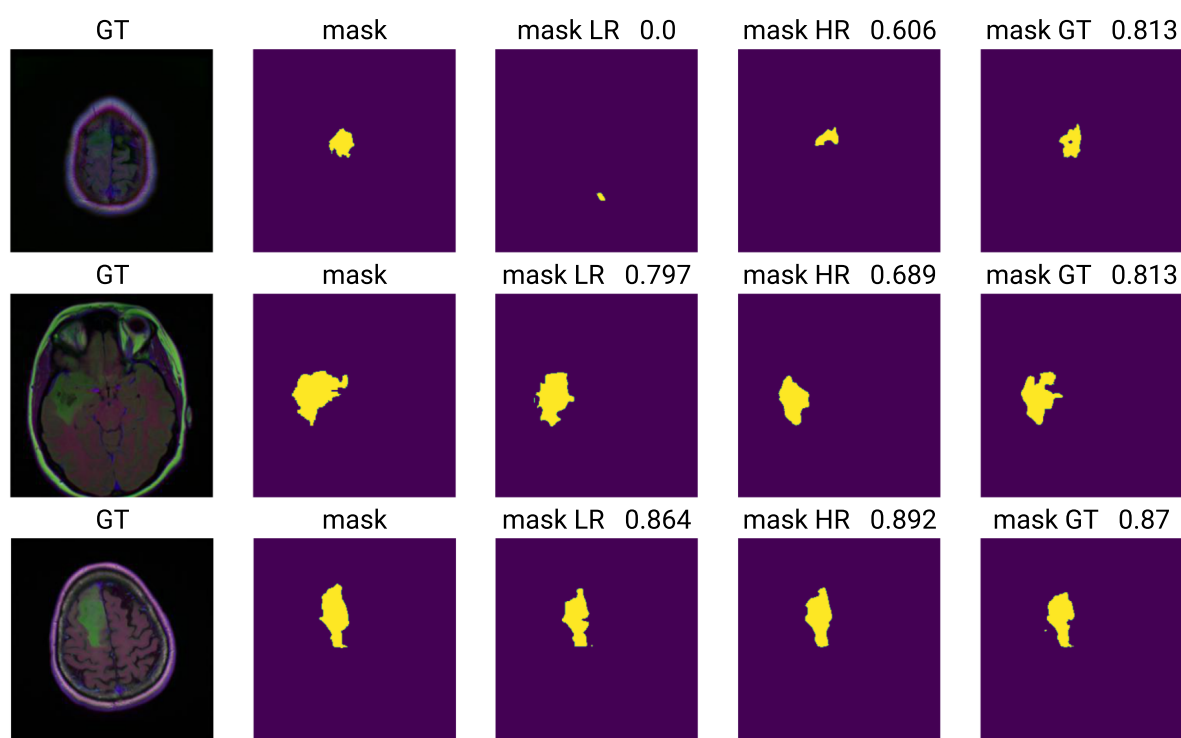


Рис. 2.6: Примеры визуализации численных результатов. Числа – значения метрики на представленных данных

Детекция. Исходя из значений метрики $f1$ score в таблице 2.2 можно сделать вывод, что увеличение разрешения изображений негативно повлияло на качество детекции опухолей. Однако в задаче обнаружения нас больше

беспокоит не общее количество правильно классифицированных изображений, а количество ошибок первого рода при гипотезе "опухоль на снимке есть", то есть те случаи, когда алгоритм классифицировал снимок с опухолью как снимок без опухоли. Такие ошибки наиболее опасны в медицинской диагностике и их важно сводить к минимуму.

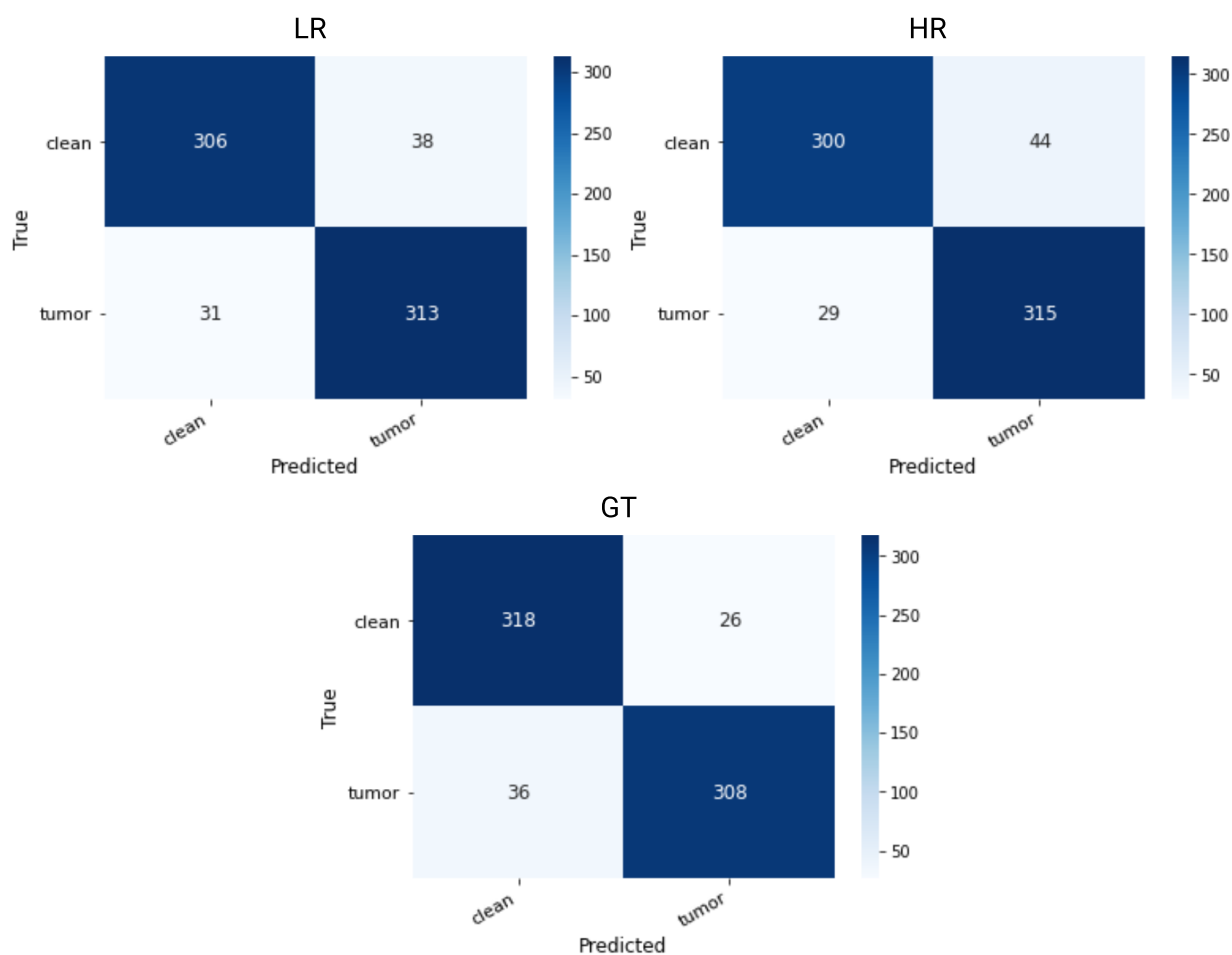


Рис. 2.7: Матрицы ошибок для изображений LR, HR и GT. Количество ошибок первого рода указано в левом нижнем углу в каждой из матриц (31, если тестировать на изображениях LR, 29 - HR, 36 - GT)

Таблица 2.2: Значение метрики f1 score на тестовых наборах данных

	LR	HR	GT
f1 score	0.901	0.896	0.909

На рис. 2.7 можно проанализировать три матрицы ошибок, полученные в результате тестирования алгоритма детекции. Несмотря на то, что общее количество верных ответов, полученных на изображениях HR меньше, чем на изображениях LR, количество ошибок первого рода обратно пропорционально общим результатам.

2.4.3 Выводы

По завершении эксперимента можно сделать вывод, что качество изображений в задачах детекции и сегментации опухолей играет важную роль, так как даже маленький прирост эффективности способен повлиять на жизни людей.

Заключение

Результаты работы

В процессе выполнения данной работы были получены следующие результаты:

1. проведен обзор существующих техник решения задачи повышения разрешения и их вариаций для различных целей, задач и доменов
2. проведен обзор существующих технологических решений по работе с медицинскими изображениями, что говорит об актуальности разработки методов искусственного интеллекта в сфере здравоохранения
3. была подробно изучена концепция генеративно-состязательных сетей, описаны особенности их обучения
4. описаны методы оценки качества изображений, сложности их применения и возможные способы комбинации
5. проектирование архитектуры программного комплекса на основе методов искусственного интеллекта
6. разработан комплекс методов в результате использования которого возможно оценить влияние техники повышения разрешения медицинских изображений на задачи детекции и сегментации опухолей (реализацию всего программного комплекса можно найти по ссылке [41])
7. поставлен эксперимент, в результате которого были получены необходимые для оценки эффективности метрики качества

Несмотря на то, что прирост в эффективности решений двух высокоуровневых задач оставляет желать лучшего, описанная предварительная обработка изображений в виде улучшения их качества является перспективным направлением не только в сфере медицинской визуализации, но и в других.

Перспективы развития

Результаты, полученные в данной работе, дают мотивацию продолжить изучение описанного влияния техник улучшения качества изображений на другие задачи машинного обучения. Следующие этапы исследования:

1. Переход к реальным данным: получение доступа к данным, работа с изначально шумными снимками;
2. Разделение больших снимков на патчи: в реальности разрешение МРТ снимков может достигать нескольких тысяч пикселей;
3. Использование предварительно обученных моделей: адаптация к такому домену, как медицинские снимки является достаточно сложной задачей;
4. Наиболее качественная симуляция шума: очевидно, что для того, чтобы избавиться от шума, необходимо воссоздать шум близкий к реальному и использовать его при генерации снимков в низком разрешении. Помимо этого, важно научиться обобщать способность алгоритма удалять один вид шума на способность удалять несколько различных;
5. Предварительное контрастирование снимков: важно понять, влияет ли данная модификация на результат сегментации;
6. Одна из частых проблем задач с генерацией изображений - интерпретируемость метрик качества: одни из самых популярных метрик данной области - PSNR и SSIM - ориентируются на восприятие картинки человеком. Однако при удалении шума и увеличении разрешения мелкие детали могут смазаться (человек такую мелочь может не заметить, а для решения алгоритма сегментации это может оказаться важным).

Список литературы

- [1] Image classification on ImageNet [Электронный ресурс]: URL: <https://paperswithcode.com/sota/image-classification-on-imagenet> (дата обращения: 10.05.2021).
- [2] Russakovsky O., Deng J., Su H., et al. ImageNet large scale visual recognition challenge // International Journal of Computer Vision, V. 115, 2015. P. 211-252.
- [3] Shantta K., Basir O. Brain tumor detection and segmentation: a survey // IRA-International Journal of Technology Engineering, V. 10, 2018. P. 55-61.
- [4] Chahal P. K., Pandey S., Goel S. A review of super resolution and tumor detection techniques in medical imaging // International Journal of Trend in Scientific Research and Development (IJTSRD), V. 3, 2019. P. 1785-1787.
- [5] Fookes C., Lin F., Chandran V. Evaluation of image resolution and super-resolution on face recognition performance // Journal of Visual Communication and Image Representation, V. 23, 2012. P. 75-93.
- [6] DeepHD technology: improving video quality with computer vision [Электронный ресурс]: URL: <https://toloka.ai/blog/deephd> (дата обращения: 18.05.2021).
- [7] Капитан арктической нефти [Электронный ресурс]: URL: <https://www.gazprom-neft.ru/press-center/sibneft-online/archive/2019-march/2628800/> (дата обращения 18.05.2021).
- [8] Dai D., Wang Y., Chen Y., et al. Is image super resolution helpful for other vision tasks? // IEEE Winter Conference on Applications of Computer Vision (WACV), V. 1, 2016. P. 1-9.
- [9] Medical image analysis software market report, 2020-2027 [Электронный ресурс]: URL: <https://www.grandviewresearch.com/industry-analysis/medical-image-analysis-software-market> (дата обращения: 10.05.2021)

- [10] Roy S., Nag S., Maitra I., et al. A review on automated brain tumor detection and segmentation from MRI of brain // International Journal of Advanced Research in Computer Science and Software Engineering, 2013.
- [11] 131 Myth-busting statistics on artificial intelligence in 2021 [Электронный ресурс]: URL: <https://research.aimultiple.com/ai-stats/> (дата обращения: 18.05.2021).
- [12] Watson [Электронный ресурс]: URL: <https://www.ibm.com/ru-ru/watson> (дата обращения: 18.05.2021).
- [13] Aidoc [Электронный ресурс]: URL: <https://www.aidoc.com/> (дата обращения: 18.05.2021).
- [14] Arterys [Электронный ресурс]: URL: <https://arterys.com/> (дата обращения: 18.05.2021).
- [15] Mevislab [Электронный ресурс]: URL: <https://www.mevislab.de/> (дата обращения: 18.05.2021).
- [16] Wang Z., Chen J., Hoi S. C. H. Deep learning for image super-resolution: A survey // CoRR, 2019.
- [17] Dong C., Loy C. C., He K., et al. Learning a deep convolutional network for image super-resolution // European Conference on Computer Vision (ECCV), 2014. P. 184-199.
- [18] Dong C., Loy C. C., Tang X. Accelerating the superresolution convolutional neural network // European Conference on Computer Vision (ECCV), 2016. P. 391-407.
- [19] Lai W.-S., Huang J.-B., Ahuja N., et al. Deep laplacian pyramid networks for fast and accurate superresolution // IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2017. P. 5835-5843.
- [20] He K., Zhang X., Ren S., et al. Deep residual learning for image recognition // IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016. P. 770-778.

- [21] Tong T., Li G., Liu X., et al. Image super-resolution using dense skip connections // IEEE International Conference on Computer Vision (ICCV), 2017. P. 4799-4807.
- [22] Bae W., Yoo J. J., Ye J. C. Beyond deep residual learning for image restoration: Persistent homology-guided manifold simplification // IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2017. P. 145-153.
- [23] Vu T., Van Nguyen C., Pham T. X., et al. Fast and efficient image quality enhancement via desubpixel convolutional neural networks // European Conference on Computer Vision (ECCV), 2018.
- [24] Bulat A., Yang J., Tzimiropoulos G. To learn image superresolution, use a gan to learn how to do image degradation first // European Conference on Computer Vision (ECCV), 2018. P. 187-202.
- [25] Zhu J.-Y., Park T., Isola P., et al. Unpaired image-to-image translation using cycle-consistent adversarial networks // IEEE International Conference on Computer Vision (ICCV), 2017. P. 2242-2251.
- [26] Jiao L., Zhang F., Liu F., et al. A survey of deep learning-based object detection // IEEE Access, V. 7, 2019. P. 128837-128868.
- [27] Minaee S., Boykov Y., Porikli F., et al. Image segmentation using deep learning: A survey. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2021.
- [28] Dong H., Yang G., Liu F., et al. Automatic brain tumor detection and segmentation using U-Net based fully convolutional networks // Annual Conference on Medical Image Understanding and Analysis, 2017. P. 506-517.
- [29] Chen Y., Shi F., Anthony G., et al. Efficient and accurate MRI super-resolution using a generative adversarial network and 3D multi-level densely connected network // Medical Image Computing and Computer Assisted Intervention (MICCAI), 2018. P. 91-99.

- [30] Wang Z., Bovik A. C., Sheikh H. R., et al. Image quality assessment: From error visibility to structural similarity // IEEE Transactions on Image Processing, V. 13, 2004. P. 600-612.
- [31] Goodfellow I. J., Pouget-Abadie J., Mirza M., et al. Generative Adversarial Nets // International Conference on Neural Information Processing Systems, V. 2, 2014. P. 2672–2680.
- [32] Zhu Y., Zhou Z., Liao G., et al. CsrGAN: medical image super-resolution using a generative adversarial network // IEEE 17th International Symposium on Biomedical Imaging (ISBI), 2020. P. 1-4.
- [33] Ronneberger O., Fischer P., Brox T. U-Net: convolutional networks for biomedical image segmentation // International Conference on Medical Image Computing and Computer-Assisted Intervention, 2015. P. 234-241.
- [34] Bertels J., Eelbode T., Berman M. Optimizing the dice score and jaccard index for medical image segmentation: theory and practice // International Conference on Medical Image Computing and Computer-Assisted Intervention, 2019. P. 92-100.
- [35] Simonyan K., Zisserman A. Very deep convolutional networks for large-scale image recognition // CoRR, 2015.
- [36] Buda M. Brain MRI segmentation [Электронный ресурс]: URL: <https://www.kaggle.com/mateuszbeda/lgg-mri-segmentation> (дата обращения: 10.09.2020)
- [37] https://drive.google.com/drive/folders/1DjUYccsprL1v_3HqakYaBD9C-g5a2ev2?usp=sharing
- [38] BCEWithLogitsLoss [Электронный ресурс]: URL: <https://pytorch.org/docs/stable/generated/torch.nn.BCEWithLogitsLoss.html#torch.nn.BCEWithLogitsLoss> (дата обращения: 16.05.2021)

- [39] Rezatofighi H., Tsoi N., Reid I. Generalized Intersection Over Union: A Metric and a Loss for Bounding Box Regression // IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2019.
- [40] CrossEntropyLoss [Электронный ресурс]: URL: <https://pytorch.org/docs/stable/generated/torch.nn.CrossEntropyLoss.html#torch.nn.CrossEntropyLoss> (дата обращения: 19.05.2021)
- [41] diploma [Электронный ресурс]: URL: <https://github.com/karinakvanchiani/diploma> (дата обращения: 23.05.2021)