

**Санкт–Петербургский государственный университет**

***Сермягин Никита Владимирович***

**Выпускная квалификационная работа**

***Мониторинг айсбергов и кораблей по спутниковым  
снямкам***

Уровень образования: магистратура

Направление 02.04.02 «Фундаментальная информатика и информационные  
технологии

Основная образовательная программа ВМ.5503.2021 «Технологии баз  
данных»

Научный руководитель:

кандидат технических наук, доцент  
Гришкин Валерий Михайлович

Рецензент:

кандидат технических наук,  
профессор каф. ИТП ГУАП  
Космачев Валентин Михайлович

Санкт-Петербург

2023 г.

# Содержание

<b>Введение</b> . . . . .	3
<b>Постановка задачи</b> . . . . .	4
<b>Обзор литературы</b> . . . . .	5
<b>Глава 1. Теоретические основы SAR снимков</b> . . . . .	6
1.1. Введение в SAR . . . . .	6
1.2. Основные этапы обработки SAR снимков . . . . .	9
1.3. Инструменты для работы с SAR снимками . . . . .	12
<b>Глава 2. Сбор и подготовка данных</b> . . . . .	15
2.1. HRSID . . . . .	15
2.2. Сбор датасета для задачи сегментирования айсбергов . . . . .	16
<b>Глава 3. Модели для обнаружения объектов</b> . . . . .	24
3.1. UNet3+ . . . . .	25
3.2. TransUNet . . . . .	27
3.3. Механизмы внимания в CNN . . . . .	27
<b>Глава 4. Реализация и результаты</b> . . . . .	30
<b>Заключение</b> . . . . .	34
<b>Список литературы</b> . . . . .	34
<b>Приложение 1. Примеры сегментации кораблей</b> . . . . .	37
<b>Приложение 2. Примеры сегментации айсбергов</b> . . . . .	38

## Введение

Одной из самых больших проблем для судоходства являются айсберги. Столкновение с ними может стать причиной кораблекрушения и гибели экипажа. Их обнаружение может позволить создавать более безопасные и эффективные маршруты.

Среди айсбергов и кораблей, первые представляют большую проблему в задаче обнаружения. Для обнаружения различных типов айсбергов могут быть использованы радиолокационные станции (РЛС), установленные непосредственно на судне. РЛС позволяет обнаружить айсберги на пути следования судна, однако данный метод не всегда надежен. Интенсивность отражения сигнала от многих айсбергов вдвое меньше, чем от участка суши или судна тех же размеров. Дальность обнаружения айсбергов судовыми РЛС также существенно зависит от формы объектов. Обнаружение айсбергов может стать затруднительным также по причине того, что на расстоянии как айсберг, так и торос (нагромождение обломков льда) могут давать одинаковый сигнал на экране радара.

Для обнаружения айсбергов также можно использовать аэрофотоснимки или спутниковые снимки. В данной работе будут использоваться снимки со спутников, с установленными на них радиолокаторами с синтезированной апертурой (англ. Synthetic Aperture Radar, SAR). Такие спутники позволяют получать изображения поверхности земли, на которые не влияют погодные условия или время суток, однако данные, полученные с них, требуют предварительной обработки, чтобы их можно было использовать в дальнейшем в различных прикладных задачах.

Возможность обнаружения айсбергов с суднами на спутниковых снимках может решать не только задачу предотвращения их столкновения друг с другом, но и другие проблемы. Например, это может использоваться для обнаружения судов, занимающихся нелегальной рыбной ловлей, что негативно сказывается на обеспечении продовольствием или наносит вред морской экосистеме. Также это позволит отслеживать состояние отдельных вод, а именно количество айсбергов в них, изменение их объема в процессе таяния и т. д.

## **Постановка задачи**

Целью данной работы является разработка методов обнаружения айсбергов и кораблей на спутниковых снимках, и в процессе ее выполнения было принято решение сосредоточиться на нейросетевых методах обнаружения объектов на изображениях. Для достижения вышеупомянутой цели были поставлены следующие задачи:

- Рассмотрение существующих моделей нейронных сетей, которые могут быть применены в рамках данной работы;
- Сбор и подготовка данных;
- Обучение отобранных моделей и сравнение полученных результатов.

## Обзор литературы

В процессе работы над ВКР были рассмотрены различные методы детектирования объектов на изображениях. В работе [13] описывается алгоритм, традиционно применяемый для детектирования айсбергов и кораблей на спутниковых SAR снимках, однако данный метод имеет ряд недостатков, и поэтому были изучены нейросетевые методы способные решать данную задачу. В работах [16], [14] представлены методы сегментирования объектов на основе сверточных нейронных сетей из семейства U-Net. В работе [15] представлен метод основанный на новом подходе, использующем механизм внимания [17] и методику из [18], позволяющую применить данный механизм к изображениям. Также были рассмотрены механизмы внимания [20] и [19], которые можно применить к сверточным нейронным сетям, что в данной работе использовалось для модификации одной из моделей. Похожую идею использовали в работах [21], [22], [23].

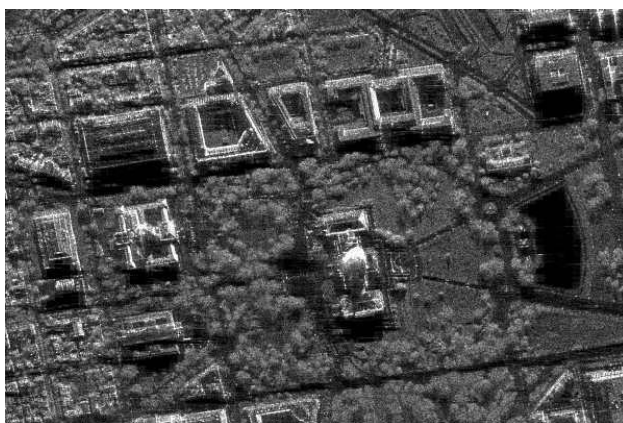
Также возникла необходимость сбора датасета с айсбергами, для этого были изучены особенности SAR снимков и инструменты для работы с ними. Ресурсы [1], [2], [3], [4], [5], [11] можно использовать для получения и загрузки SAR снимков, а ПО [6] и [7] для их обработки. В сборе датасета с айсбергами для задачи сегментации помогли ресурс [10], содержащий базу данных местоположений айсбергов в Атлантике и инструмент для разметки изображений [12]. Для обучения модели, сегментирующей корабли, использовался готовый датасет [8]. Собранный и использованный датасеты представлены в формате [9].

При реализации моделей были использованы Python библиотеки: [24] – для реализации моделей глубокого обучения, и [25] – для подготовки данных к обучению.

# Глава 1. Теоретические основы SAR снимков

## 1.1 Введение в SAR

Радиолокатор с синтезированной апертурой (SAR) - один из типов радарного датчика, которые обычно устанавливают на спутники или самолеты. SAR посылает электромагнитные волны на земную поверхность и принимает отраженный от нее сигнал, при этом волны посылаются не вертикально вниз, а вбок под определенным углом. Электромагнитные волны, принимаемые датчиком называются измеренным обратным рассеянием, используя SAR снимок можно получить двумерную визуализацию измеренного обратного рассеяния (рис. 1). Сила амплитуды измеренного обратного рассеяния используется для различения объектов на земле, а временная задержка между переданной и полученной электромагнитной волной позволяет определить местоположение объекта.

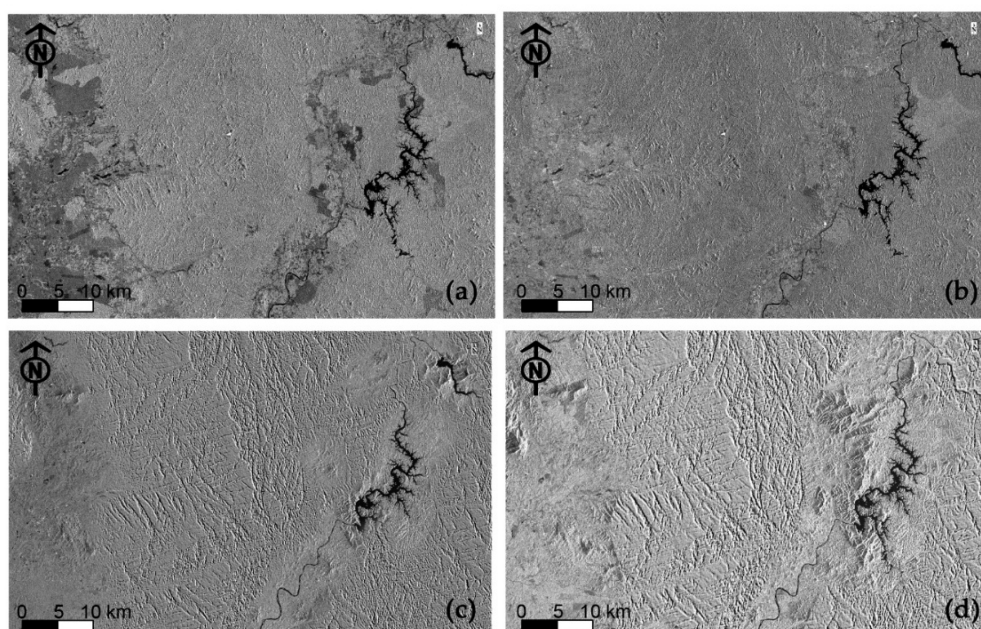


**Рис. 1:** Пример визуализации данных SAR снимка

По сравнению с обычными оптическими снимками SAR снимки обладают рядом преимуществ. Так как в SAR используется активный датчик, который не только принимает, но и отправляет сигнал, в отличие от оптических систем можно получать снимки как в дневное, так и в ночное время. Также, благодаря активному зондированию датчики в SAR могут получать изображения с использованием более длинных волн (датчик использует микроволновые длины волн от К-канала -  $7,5 \times 10^{-3}$  м, до Р-канала - 1 м), это позволяет радиолокаторам SAR "пробивать" облака, туман, пыль, смог, дым

и другие погодные условия. Еще одной значимой характеристикой, особенно в задаче обнаружения айсбергов, является способность этих волн проникать в такие субстанции, как почва, снег и лед. При этом содержание влаги или воздуха в материале будет влиять на глубину проникновения и на полученный сигнал соответственно, что может быть использовано для дифференцирования объектов.

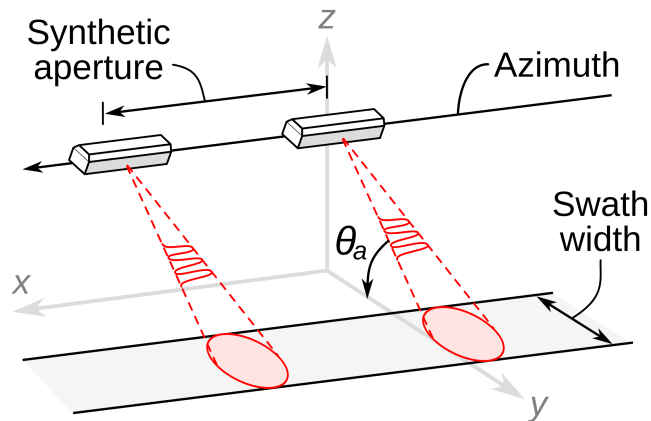
Кроме излучения более длинных волн активное зондирование также предоставляет возможность управлять поляризацией этих электромагнитных волн. Это возможно благодаря тому, что датчик SAR определяет как излучаемую, так и принимаемую поляризацию, вследствие чего на итоговом изображении SAR на основе обратного рассеяния можно выделить различные объекты на поверхности земли, и, как показала практика, в зависимости от окружения (открытые воды или льды) заметность айсберга на снимке изменится при использовании разных поляризаций. Поляризация данных SAR обозначается двумя буквами, первая из которых означает излучаемую поляризацию, а вторая — принимаемую. Обычно в SAR используется двойная поляризация и снимки содержат либо данные с поляризацией VV и VH, либо данные с поляризацией HH и HV.



**Рис. 2:** Пример визуализации данных SAR снимка при использовании разных поляризаций HV (a), HH (b), VH (c), VV (d)

По мере того, как спутник движется вокруг Земли, датчик многократно захватывает сигнал (одни и те же объекты при этом также захватываются несколько раз). Сочетание этого с методами обработки сигнала позволяет датчику SAR искусственно удлинить свою антенну для получения большей апертуры, и как следствие увеличить пространственное разрешение снимка.

## Synthetic aperture radar



**Рис. 3:** Визуализация принципа синтезирования апертуры

Снимки SAR обычно предоставляются в виде данных двух типов: ground range detected (GRD) или single look complex (SLC). Данные в типе GRD усредняют для получения изображения, спроецированного на земную поверхность с использованием модели эллипсоида Земли. Снимки формата GRD хранятся в виде массива действительных чисел, в котором значения каждого пикселя является значением амплитуды измеренного сигнала обратного рассеяния. Данные в типе SLC представляет собой изображение в плоскости снимка в процессе съемки, называемой плоскостью наклонной дальности. Снимок SLC хранится в виде массива комплексных значений (амплитуды и фазы измеренного сигнала обратного рассеяния). SLC и GRD относятся к данным первого уровня обработки, кроме него существуют еще нулевой и второй. Данные первого уровня обработки представляют собой сжатые и необработанные наборы значений, полученных напрямую с радара, и являются основой для получения данных следующих уровней обработки. Данные второго уровня обработки могут содержать информацию о направлениях ветра, волн и течений.



Несмотря на то, что характеристики датчика SAR обеспечивают уникальную визуализацию наземных объектов, такой метод получения снимков земной поверхности также приводит к сложностям при обработке данных. Среди проблем, которые необходимо решить, можно отметить устранение тепловых шумов, калибровку величины обратного рассеяния, устранение спекл-шума и устранение радиометрических и геометрических искажений.

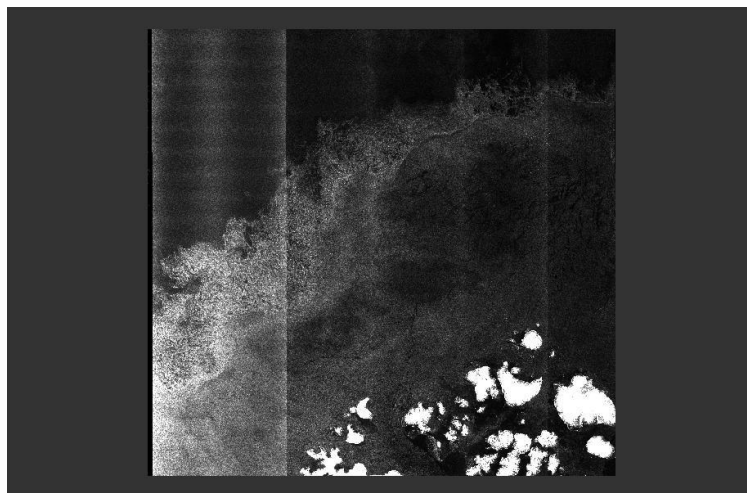
## **1.2 Основные этапы обработки SAR снимков**

Перед тем, как снимок с радиолокатора с синтезированной апертурой первого уровня обработки можно будет использовать для анализа или визуализации, он должен быть обработан. Обработка, которую необходимо выполнить, зависит исходного уровня обработки снимка и его типа. В данной работе использовались снимки типа GRD первого уровня обработки и следующие этапы предварительной обработки:

1. Коррекция орбиты;
2. Удаление тепловых шумов;
3. Удаление шумов на границах;
4. Радиометрическая калибровка;
5. Фильтрация спекл-шума;
6. Геометрическая коррекция;
7. Преобразование единиц измерения.

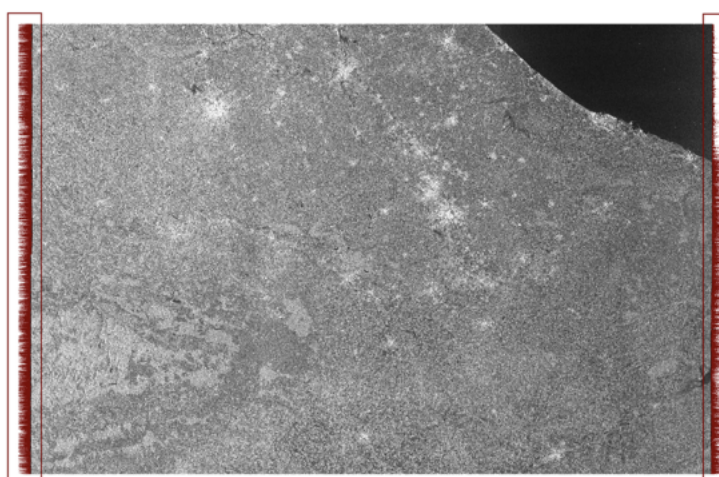
Коррекция орбиты используется для корректировки геоданных снимка. Стандартные метаданные, предоставляемые вместе со снимком не всегда точны, поэтому на данном этапе происходит обновление векторов состояния орбиты (англ. *Orbital State Vectors*, OSV). Для этого запрашивается и скачивается более актуальная информация об OSV, которая обычно доступна в течение от нескольких часов до нескольких недель с момента получения снимка.

Удаление тепловых шумов используется для удаления аддитивного шума, вызванного нагревом SAR датчика (рис. 4). Этот этап необходим, чтобы нормализовать сигнал обратного рассеяния в пределах целого SAR снимка. Тепловой шум наиболее заметен в режимах кросс-поляризации (HV или VH) или на снимках с низким обратным рассеянием, например на снимках местности океана.



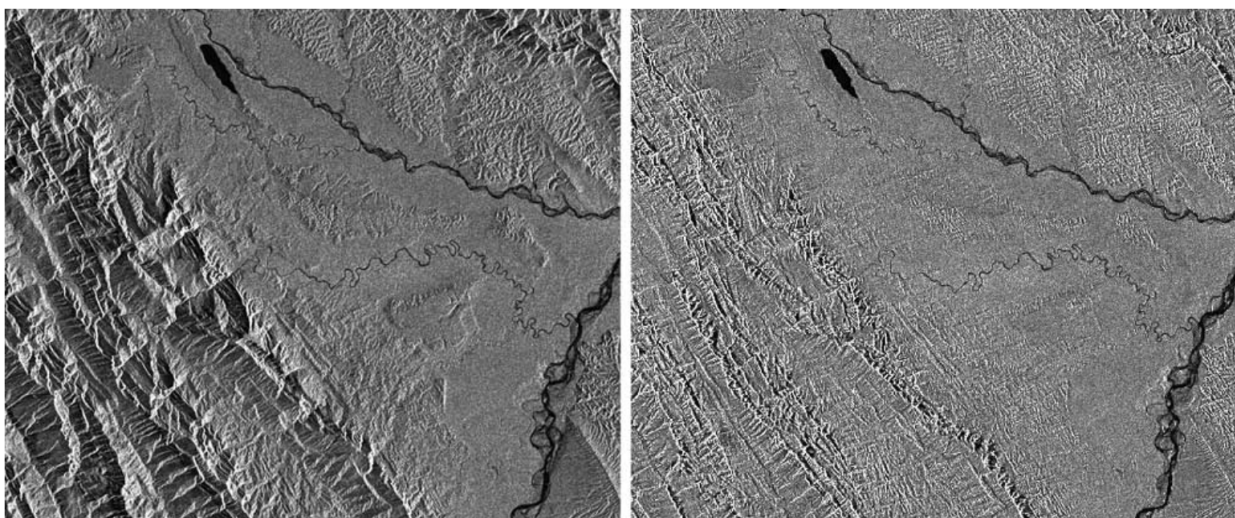
**Рис. 4:** Пример теплового шума на SAR снимке

Удаление шумов на границах необходимо, чтобы получить бесшовное изображение снимка. Один SAR снимок составляется из множественных за-меров разных участков поверхности, при этом на границе этих участков обычно появляется тонкая полоса шума (рис. 5).



**Рис. 5:** Пример граничного шума на SAR снимке (пиксели шума выделены красным цветом)

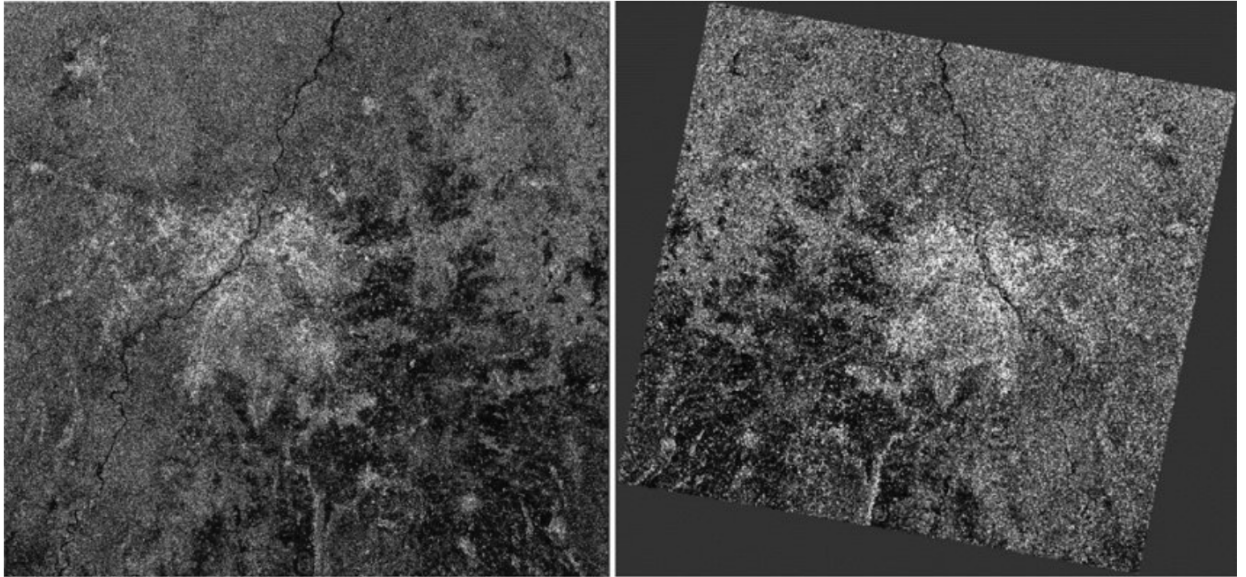
Радиометрическая калибровка и радиометрическое сглаживание необходимы, чтобы преобразовать данные с датчиков в значения интенсивности отраженного сигнала обратного рассеяния на единицу площади и скорректировать интенсивность полученного сигнала на основе угла наклона датчика, а также исправить радиометрические искажения, вызванные сложной топографией местности или геометрией объектов. Искажения появляются из-за того, что датчики SAR смотрят на объекты сбоку, из-за чего объекты, обращенные к датчику, кажутся ярче, чем объекты, обращенные в сторону от датчика.



**Рис. 6:** SAR снимок до и после радиометрической обработки

Фильтрация спекл-шума необходима из-за присутствия на снимках высокочастотного шума, характерного изображениям с радаров. Наличие данного шума может затруднить дальнейший анализ снимков и для его подавления и улучшения качества изображения могут быть использованы различные алгоритмы фильтрации, которые используют локальную статистику пикселей изображения.

Геометрическая коррекция необходима из-за того, что в зависимости от расположения объектов по отношению к датчику они могут выглядеть более сжатыми или растянутыми по сравнению с действительностью. Эта особенность возникает из-за того, что датчики SAR направлены вбок. Геометрическая коррекция устраняет эти искажения и также смещает пиксели в их правильное географическое положение.



**Рис. 7:** SAR снимок до и после геометрической коррекции

Последним шагом подготовки данных является преобразование единиц измерения. На данном шаге исходное значение интенсивности обратного рассеяния сигнала преобразуется в децибелы при помощи простого логарифмического преобразования (1). Логарифмическое преобразование уменьшает диапазон значений интенсивности обратного рассеяния и улучшает процессы визуализации и интерпретации изображения.

$$dB = 10 * \log_{10}(\sigma^0)$$

$\sigma^0$  – исходное значение интенсивности сигнала обратного рассеяния (1)

$dB$  – интенсивность сигнала обратного рассеяния в децибелах

### **1.3 Инструменты для работы с SAR снимками**

Различные типы SAR снимков, могут быть получены бесплатно с некоторых платформ. Однако каждая из них имеет свои особенности, которые необходимо учитывать при выборе под конкретные задачи. Наиболее известные и используемые из этих платформ это:

- Copernicus Open Access Hub [1] (ранее известный как Sentinel Scientific Data Hub), предоставляет снимки полученные со спутников миссий

Sentinel, в частности необходимые SAR снимки со спутников Sentinel-1A и Sentinel-1B всех уровней обработки и различных поляризаций. Все данные на сервисе предоставляются бесплатно и становятся доступными в тот же момент когда они будут получены со спутников. Орбитальный период спутников составляет 12 дней по отдельности, а при использовании их вместе повторный цикл снимков достигает 6 дней. По истечении срока в 6 месяцев снимки отправляются в архив и для получения к ним доступа необходимо отправить запрос на временный возврат снимка в “online” режим.

- The Copernicus Dedicated Access Center [2] предоставляет доступ по запросу к старым снимкам, которые больше не доступны на главном сервере. Однако данные услуги данной платформы становятся доступны только после регистрации в программе Copernicus Service Projects, что может быть затруднительным.
- Alaska Satellite Facility [3] – часть программы NASA’s Earth Science Data Systems, платформа предоставляет доступ к огромному количеству SAR снимков, при этом сделанными не только спутниками Sentinel-1. Данные предоставляются бесплатно. При этом важно отметить, что все снимки доступны в режиме “online”, т.е. не требуется запрашивать доступ на получение старых снимков, потому что они не отправляются в архив.
- EO BROWSER [4] предоставляет доступ к снимкам со всех упомянутых выше платформ.
- ICEYE SAR Data [5] – платформа компании ICEYE, которая предоставляет доступ к SAR снимкам, полученным с их собственных спутников. Особенность данной платформы в том, что благодаря большому числу спутников можно получать снимки одного и того же места с периодом 24 часа, или даже меньше. Однако для бесплатного доступа существует лимит на количество снимков, которые можно получить с одного аккаунта. Кроме того предоставляются снимки только первого уровня обработки с поляризацией VV.

После более детального изучения указанных выше вариантов было решено использовать платформу Alaska Satellite Facility. Решающими факторами стали отсутствие необходимости запрашивать и ожидать доступ к старым снимкам и наличие удобного поиска по снимкам (фильтрация снимков по дате, географическому местоположению). Для поиска и скачивания снимков у платформы имеется удобное и хорошо документированное API, которое также доступно в виде пакета для языка Python, что будет удобно использовать для автоматизации процесса сбора SAR снимков.

Инструментов для обработки SAR снимков намного меньше. Среди них можно выделить Xarray-sentinel [6] и ESA SNAP (The Sentinel Application Platform) [7].

Xarray-sentinel это библиотека для языка Python с открытым исходным кодом, которая позволяет просматривать и обрабатывать SAR снимки. Среди ее плюсов можно отметить простую установку, настройку и эффективное использование вычислительных ресурсов благодаря интеграции с модулем Xarray по работе с многомерными массивами. Среди минусов важно выделить отсутствие какой-либо документации.

ESA SNAP это программное обеспечение которое было разработано для работы со снимками со спутников Sentinel, включая SAR снимки. SNAP предоставляет как графический интерфейс, так и более удобный вариант работы с большим количеством снимков через интерфейс командной строки или через модули Python. SNAP использует большое количество сторонних библиотек, основные из которых Geospatial Data Abstraction Library (GDAL), NetBeans Platform, GeoTolls и Java Advanced Imaging (JAI). Из-за этого процесс установки и настройки может вызвать затруднения, которые чаще всего связаны с проблемами в совместимости. Можно отметить, что данное ПО активно развивается и предоставляет большое количество инструментов для обработки снимков с гибкой настройкой параметров их обработки.

## Глава 2. Сбор и подготовка данных

Решить проблему обнаружения объектов на снимке можно несколькими способами. Например, самым простым решением было бы решать данную проблему как задачу классификации, а именно определять отсутствие или присутствие объекта на изображении. Другой способ – детектирование объектов, в данном случае дополнительно необходимо на исходном изображении определить область с границами в котором расположен искомый объект (например найти координаты, ширину и высоту ограничивающей объект прямоугольной рамки). И последний способ использовать сегментацию, которая в свою очередь делится на два вида: семантическую сегментацию и сегментацию экземпляров. В первом случае для каждого пикселя входного изображения определяется класс к которому он принадлежит, во втором дополнительно отделяются друг от друга отдельные экземпляры объектов, даже если они относятся к одному и тому же классу.

Самым сложным и предоставляющим наибольшее количество информации является способ сегментирования объектов на изображении. В данной работе было решено использовать именно его. Между семантической сегментацией и сегментацией экземпляров был выбран первый вариант, так как на снимках со спутников объекты не перекрывают друг друга, хотя и могут возникнуть определенные сложности с отделением малых объектов.

Для обучения модели, сегментирующей корабли, был выбран датасет HRSID [8], а для модели, сегментирующей айсберги, собран собственный.

### 2.1 HRSID

HRSID (High-resolution SAR Images Dataset) предоставляет данные как для решения задачи детектирования, так и сегментации кораблей на SAR снимках. Он состоит из снимков, полученных со спутников TanDEM-X SAR, Sentinel-1B и TerraSAR-X. Среди других схожих датасетов его отличает наличие достаточно большого количества снимков со сложным фоном (не только снимки кораблей в открытом море, но и около участков суши). Снимки в данном датасете представляют собой обработанные SAR снимки размера  $800 \times 800$  пикселей в оттенках серого в формате JPEG, с пространственным

разрешением от 0.5 м до 3 м, и в трех вариантах поляризации (HH, VV и HV). Однако каждый снимок доступен только в одном из трех режимов поляризации. Аннотации к снимкам предоставляются в формате MS COCO [9], маски для сегментации в котором представлены в виде координат полигонов. Всего в датасете 5604 снимков с 16951 кораблями.

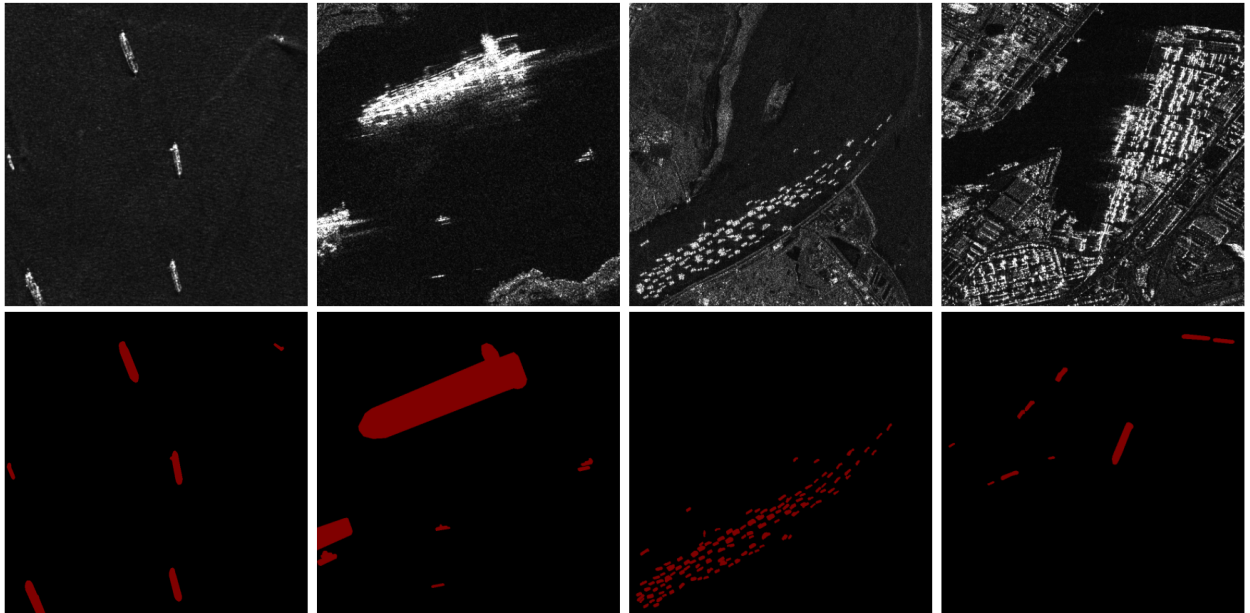


Рис. 8: Пример участков снимков с масками из датасета HRSID

## 2.2 Сбор датасета для задачи сегментирования айсбергов

Не удалось найти датасет с SAR снимками айсбергов, подходящий для задачи сегментации, поэтому было принято решение собрать такой датасет собственноручно. Стоит отметить, что также отсутствует какой-либо датасет для сегментирования айсбергов на оптических спутниковых снимках.

На первом этапе было необходимо разобраться с процессом обработки SAR снимков. Инструментом для их обработки был выбран ESA SNAP. Однако с его установкой и настройкой возникли некоторые сложности, связанные с конфликтами версий ПО необходимых для SNAP и уже установленных на оборудовании. На устранение конфликтов было потрачено достаточно большое количество времени, поэтому, чтобы избежать подобных проблем в будущем, был настроен и создан Docker образ, на основе которого можно при необходимости создать контейнер с окружением готовым для обработки SAR



снимков с использованием ESA SNAP версии 9.0 и Python версии 3.6 с настроенным модулем snappy. Получившийся образ можно собрать самостоятельно из репозитория (<https://github.com/nyaaneet/esa-snap-docker>) или загрузить уже собранный (<https://hub.docker.com/r/nyaaneet/esa-snap>). Дальнейшая разработка и обработка SAR снимков производилась внутри контейнера Docker.

Для обработки SAR снимков был написан Python скрипт, принимающий на вход следующие входные аргументы:

- `product_path` – путь к SAR снимку на диске;
- `wkt_aoi` – область снимка, которую необходимо обработать, описанная в формате Well-known text (WKT);
- `polarizations` – поляризации, которые необходимо обработать, в случае если данный аргумент пропущен будут обработаны все доступные поляризации текущего снимка;
- `out_path` – путь, куда необходимо сохранить обработанный снимок;
- `out_name` – шаблон имени для обработанного снимка.

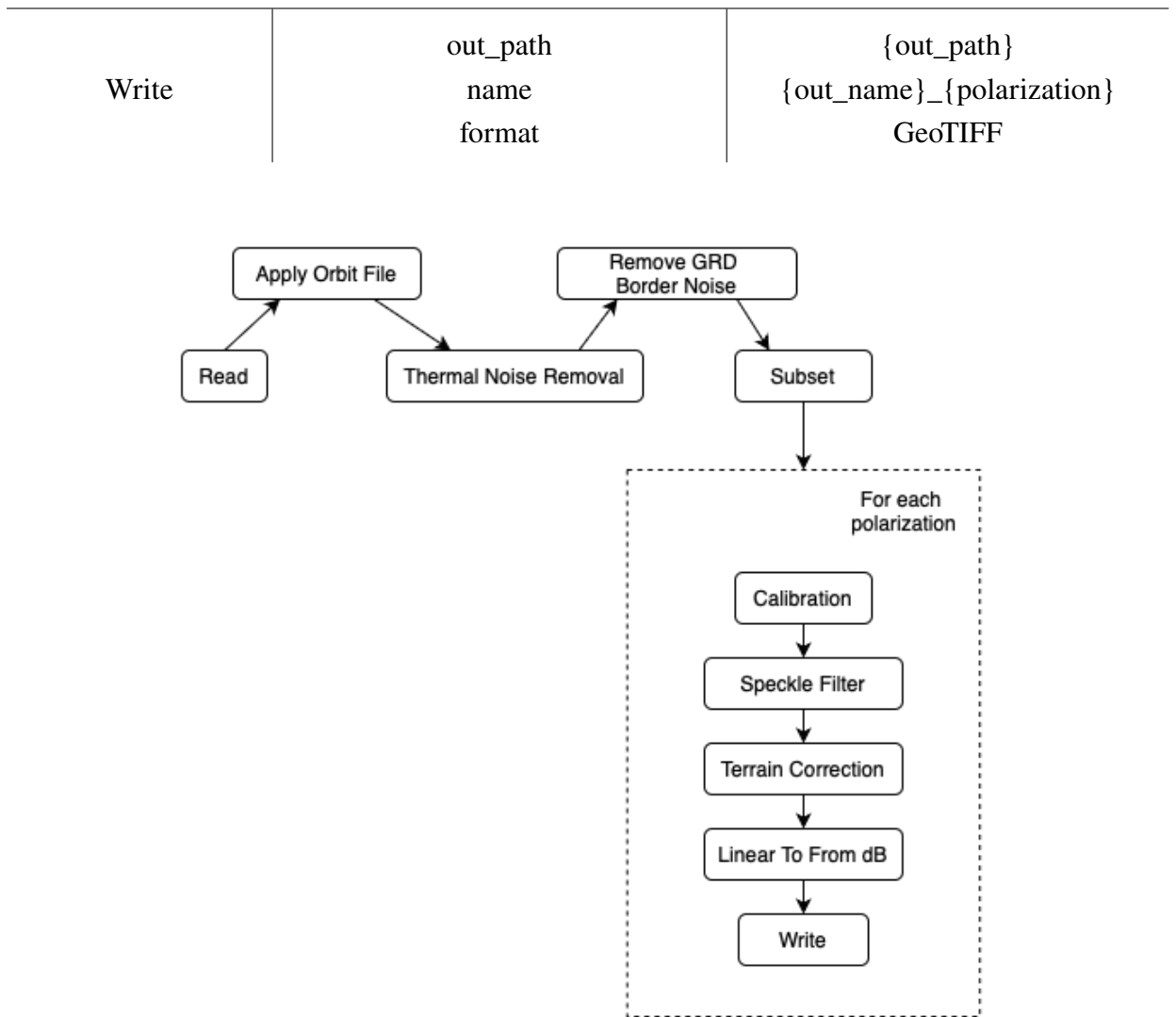
Скрипт обрабатывает SAR снимок типа GRD по шагам, описанным в разделе 1.2, однако добавляется несколько дополнительных этапов:

- Загрузка данных предварительно загруженного SAR снимка, используя оператор "Read" из SNAP;
- Использование оператора "Subset" для ограничения области обработки до интересующей зоны. Обработка снимка целиком потребовала бы огромных вычислительных мощностей, поэтому снимки обрабатывались по частям;
- Использование оператора "Write" для записи обработанного снимка на диск.

Более подробно схема обработки изображена на рис. 9 и табл. 1. В процессе написания скрипта корректность обработки снимка на разных этапах проверялась с помощью юнит-тестов, а на финальной стадии результат обработки каждого этапа и конечный результат обработки более тщательно были проверены вручную. Код скриптов доступен по ссылке в репозитории <https://github.com/nyaneet/sar-grd-processing>.

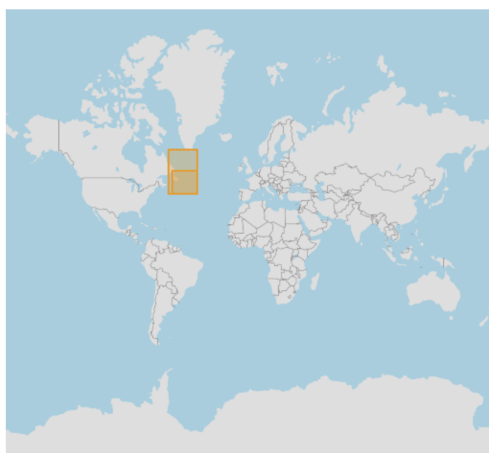
**Таблица 1:** Этапы обработки снимка с использованием ESA SNAP

Оператор	Параметры	Значения
Read	path	{product_path}
Apply Orbit File	orbitType polyDegree continueOnFail	Sentinel Precise (Auto Download) 3 False
Thermal Noise Removal	removeThermalNoise	True
Remove GRD Border Noise	borderLimit trimThreshold	500 0.5
Subset	copyMetadata geoRegion	True {wkt_aoi}
Calibration	outputSigmaBand sourceBands selectedPolarisations	True Intensity_{polarization} {polarization}
Speckle Filter	sourceBands filter	Sigma0_{polarization} Refined Lee
Terrain Correction	demName imgResamplingMethod sourceBands saveProjectedLocalIncidenceAngle saveSelectedSourceBand	GETASSE30 BILINEAR_INTERPOLATION Sigma0_{polarization} True True
Linear To From dB	–	–



**Рис. 9:** Процесс обработки SAR снимка с использованием ESA SNAP

Далее для получения итогового датасета использовался другой датасет The International Ice Patrol (ИП) Iceberg Sightings Database [10], что позволило упростить задачу поиска айсбергов на снимках. ИП собирала информацию об активности айсбергов в определенной части Атлантического океана. Данные, представленные в датасете собирались из множества различных источников и содержат записи о наблюдениях с 1960 по 2021 год.



**Рис. 10:** Зона покрытия датасета IP Iceberg Sightings Database

В зависимости от года формат и количество доступной об айсбергах информации могут различаться. В данной работе использовались наблюдения, сделанные с 2019 по 2021 год, и следующие характеристики айсбергов:

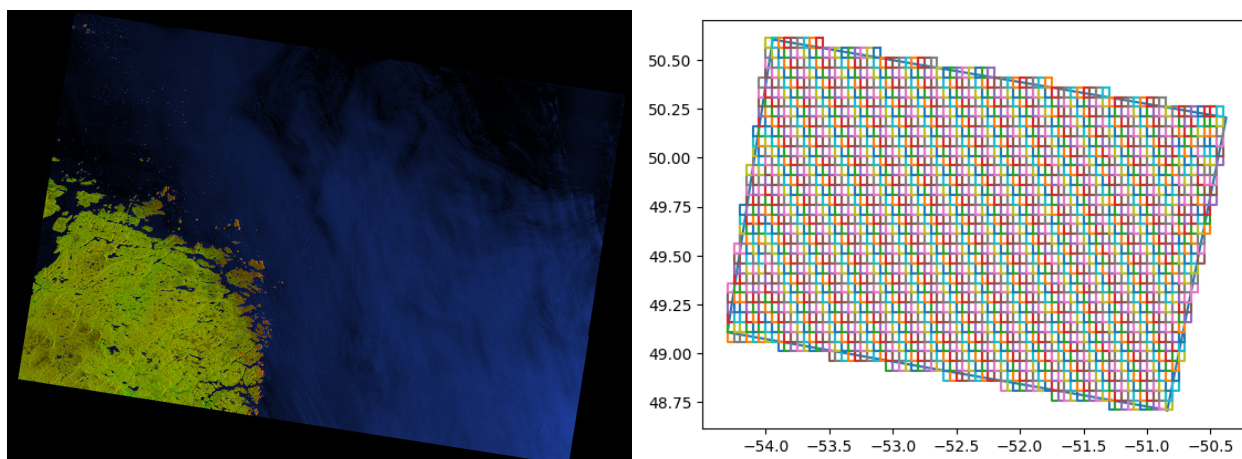
- SIGHTING\_DATE – дата наблюдения в формате MM/DD/YYYY;
- SIGHTING\_TIME – UTC время наблюдения в формате HHmm;
- ICEBERG\_NUMBER – уникальный идентификатор, выданный каждому айсбергу на год;
- SIGHTING\_LATITUDE и SIGHTING\_LONGITUDE – широта и долгота айсберга на момент когда было сделано наблюдение;
- SIZE – размер айсберга (GR: growler, BB: bergy bit, SM: small, MED: medium, LG: large, VLG: very large, RAD: radar berg/undetermined size);
- SHAPE – форма айсберга (TAB: tabular, NTB: nontabular, BLK: blocky, DD: dry dock, DOM: domed, GEN: general, PIN: pinnacled, RAD: radar berg/undermined shape, WDG: wedge).

Для поиска снимков, содержащих указанные в датасете айсберги, был написан Python скрипт. Скрипт использует модуль asf-search [11] и для каждого айсберга находит данные обо всех доступных для скачивания с Alaska Satellite Facility снимках SAR и также суммирует информацию об айсбергах для каждого снимка. Чтобы избежать блокировки аккаунта на сервисе поиск

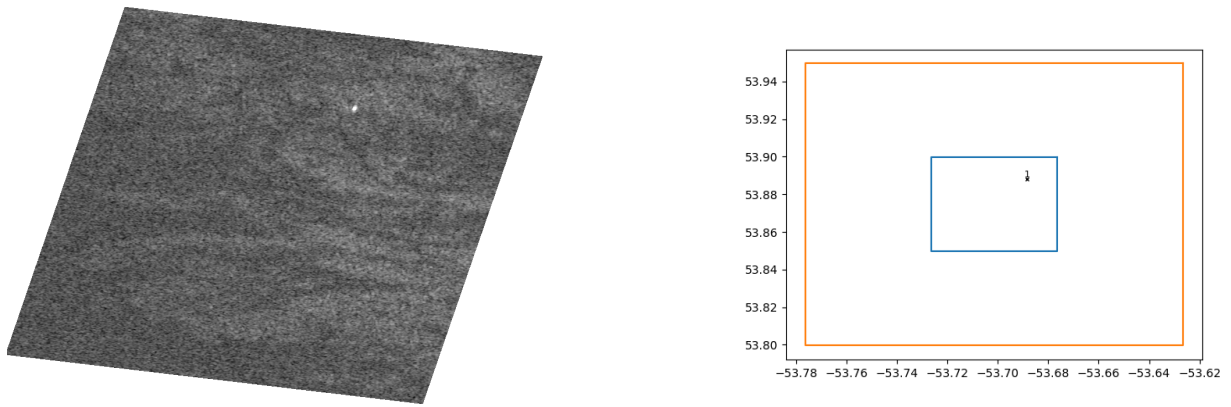
по записям производился последовательно и обработка каждого года в среднем занимала день. Собранные данные доступны для скачивания по ссылке [https://drive.google.com/drive/folders/1HGcyehgumJ9D761U8mz\\_sj1R4GqzMYPt?usp=share\\_link](https://drive.google.com/drive/folders/1HGcyehgumJ9D761U8mz_sj1R4GqzMYPt?usp=share_link). Далее вручную была отобрана группа снимков для обработки. Критериями для отбора являлись: количество айсбергов на снимке, время года, разрешение, уровень и тип обработки SAR снимка.

Также был написан еще один Python скрипт, который используя тот же модуль `asf-search` скачивает выбранные SAR снимки, отбирает интересные участки этих снимков (содержащие хотя бы один айсберг), запускает их обработку и генерирует вспомогательные графики для более удобного поиска айсбергов при разметке. На загрузку и обработку 40 снимков в сумме ушло около четырех дней.

Для обработки снимков использовалось следующее оборудование: ПК с ОС Ubuntu 22.04 LTS 64-bits, 32 GB RAM и процессором AMD Ryzen 7 2700X (8 × 3700MHz).



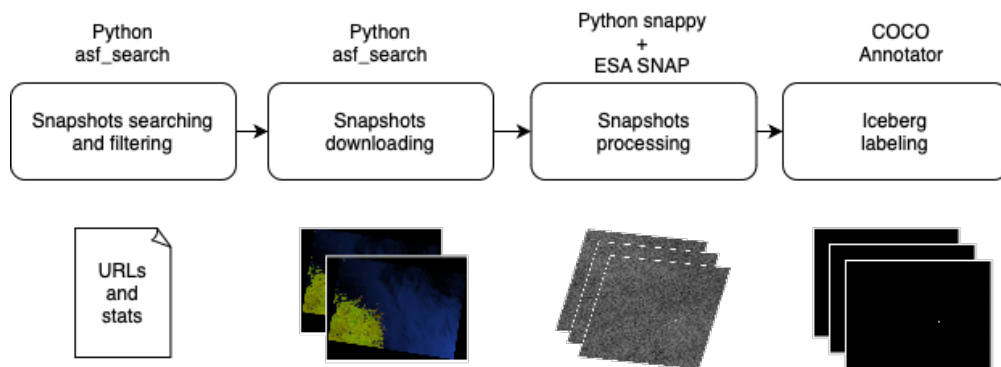
**Рис. 11:** Пример SAR снимка и его разбиение на участки



**Рис. 12:** Пример изображения обработанного участка и его вспомогательного графика (синей рамкой выделена область обработанного участка)

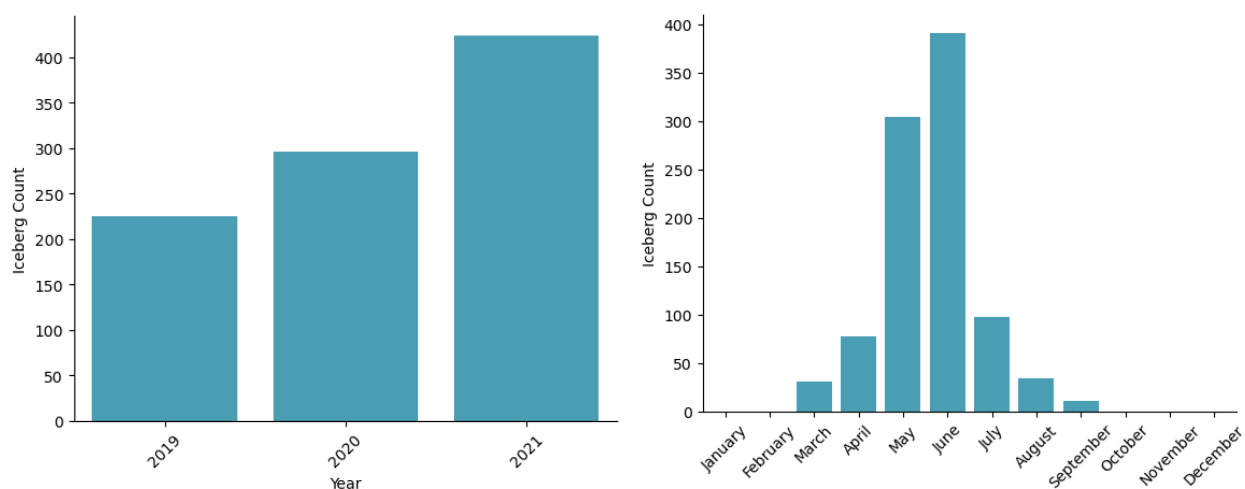
Далее для разметки датасета использовался инструмент COCO Annotator [12]. Для айсбергов на снимках были созданы маски и присвоены уникальные идентификаторы, аналогичные номерам из датасета IP Iceberg Sightings Database.

Всего в размеченном датасете 957 изображений (1914 если учитывать, что для каждого снимка доступны два варианта поляризации) с 1106 айсбергами. Изображения имеют пространственное разрешение  $20 \times 20$  м и средний размер  $900 \times 600$  пикселей. На размеченных снимках присутствуют айсберги разного размера (от SM/small до VLG/very large), в различном окружении (в открытых водах, среди льдов, около суши), с разного времени года в режимах поляризации HH и HV. Датасет доступен по ссылке [https://drive.google.com/file/d/14K0z1Rip2bqUoufeCuyKY9dbVb1Nhphe/view?usp=share\\_link](https://drive.google.com/file/d/14K0z1Rip2bqUoufeCuyKY9dbVb1Nhphe/view?usp=share_link), а аннотации к нему предоставляются в формате MS COCO. Получившийся процесс сбора датасета изображен на рис. 13.

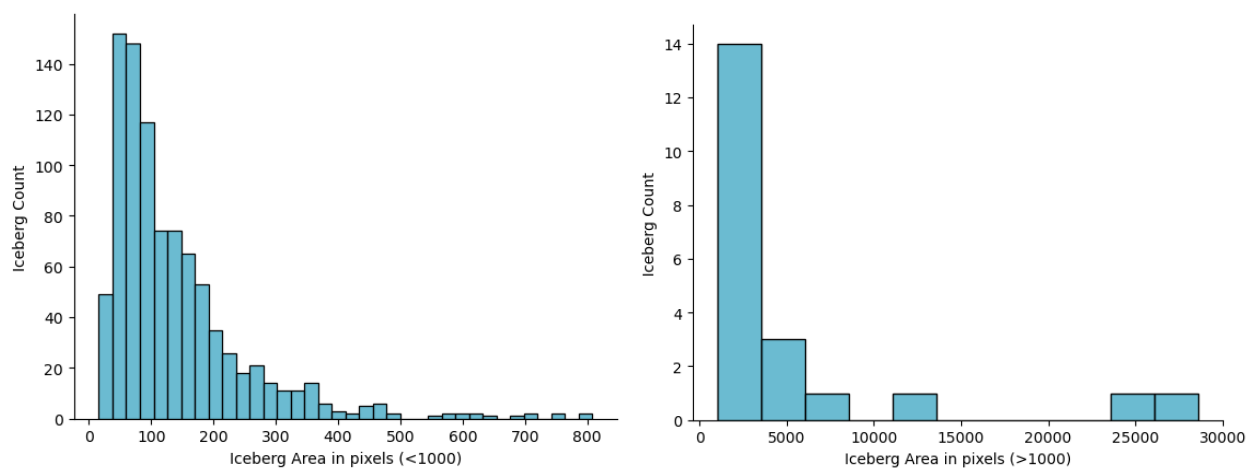


**Рис. 13:** Процесс сбора данных с айсбергами

На рис. 14 и 15 изображена статистика по собранному датасету.



**Рис. 14:** Распределение айсбергов в датасете по годам и месяцам

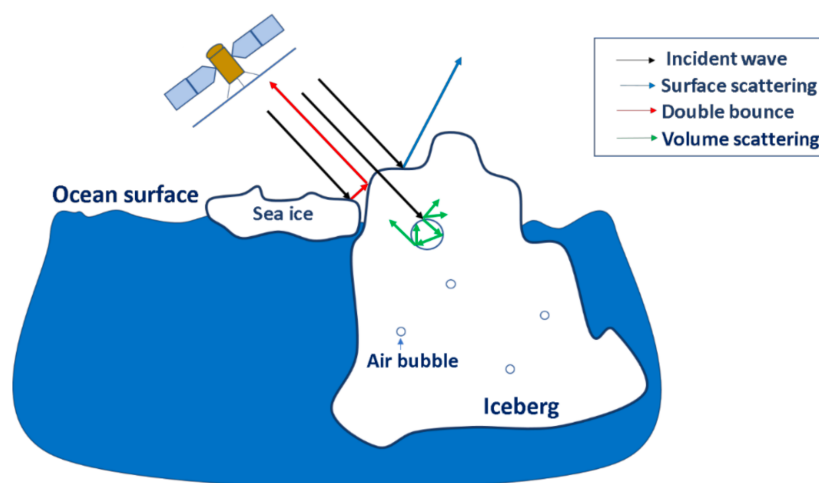


**Рис. 15:** Распределение айсбергов в датасете по их площади (каждый пиксель соответствует области  $10 \times 10$  м)

## Глава 3. Модели для обнаружения объектов

Задача обнаружения объектов на радиолокационных снимках чаще всего решается применением алгоритмов с адаптивным порогом, одним из традиционных и наиболее часто применяемых методов при обнаружении объектов на SAR снимках является CFAR (constant false alarm rates) детектор [13] и другие улучшенные методы, основанные на нем. CFAR – алгоритм с адаптивным порогом, используемый в радиолокационных системах для обнаружения цели в условиях шума, помех и интерференции сигналов.

Все эти методы анализируют интенсивность обратного рассеяния от исследуемых объектов. Сигнал обратного рассеяния в основном приходит от поверхностного или объемного рассеяния, при чем на эти два механизма влияют различные параметры. Часть из этих параметров зависят от целевых объектов, а именно от их формы, геометрии, температуры, шероховатости или структуры поверхности. Другие зависят от датчика, такие как угол падения сигнала, его частота и используемая поляризация.



**Рис. 16:** Визуализация механизма обратного рассеяния на примере айсберга и морских льдов

В CFAR алгоритм сравнивает интенсивность каждого исследуемого пикселя на изображении с его локальными фоновыми помехами, и если значение пикселя превосходит некоторый посчитанный порог он помечается как выброс. Предполагается, что кластеры этих выбросов являются интересующими объектами. Порог определяется на основе функции плотности вероятности локальных помех.



У данного алгоритма есть ряд недостатков. Для его работы необходимо достаточно точно подобрать используемые параметры, а также сильная зависимость от статистического распределения локального шума снижает надежность данного метода. Из-за этого настроенный алгоритм может относительно хорошо работать в одних условиях, но выдавать плохие результаты при их изменении (например изменения в погоде, изменение времени или местоположения получения снимков). Также данный алгоритм плохо подходит для детектирования объектов в условиях сложного фона (например на участках близких к суше или с большим количеством морских льдов).

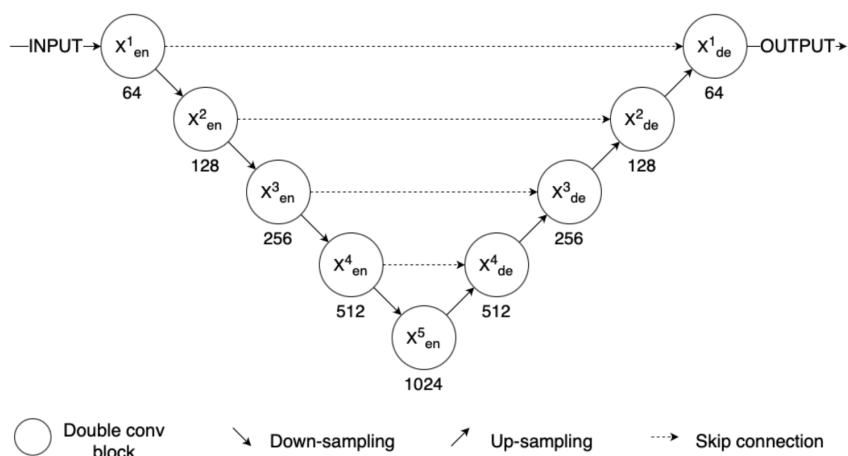
В данной работе для решения задачи обнаружения объектов на спутниковых SAR снимках было решено использовать нейросетевые методы. Так как эти методы обладают хорошей обобщающей способностью, будет легче избежать многих проблем характерных для методов на основе адаптивного порога. Была выбрана модель UNet3+[14] и TransUNet[15]. Также дополнительно рассмотрена возможность применения различных механизмов внимания к UNet3+.

### 3.1 UNet3+

Сверточная нейронная сеть U-Net[16] изначально была разработана для сегментации объектов на медицинских изображениях, однако хорошо показала себя и в других сферах.

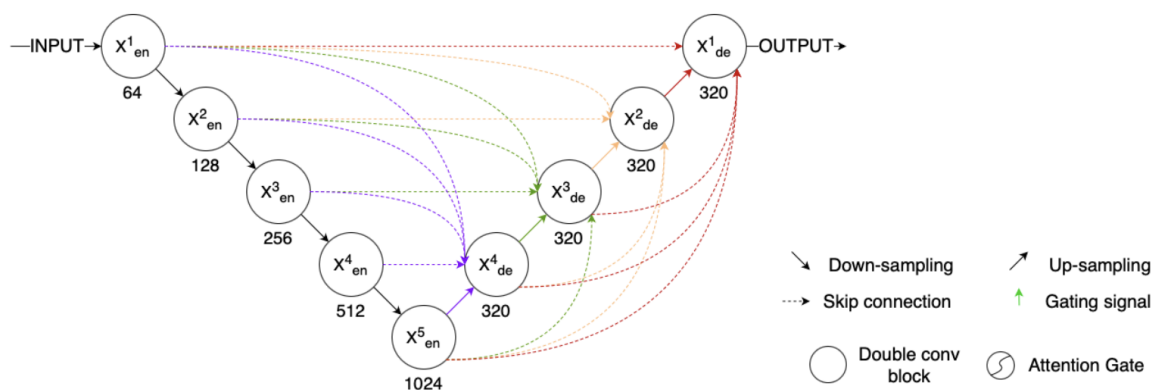
U-Net соответствует архитектуре кодер-декодер. Кодер уменьшает пространственное разрешение изображения, а после декодер увеличивает его. Последний слой дополнительно классифицирует пиксели получившегося изображения по заранее заданным классам. Также существуют сквозные соединения от кодера к декодеру, чтобы помочь декодеру лучше восстановить детали объектов. Более подробно архитектура U-Net изображена на рис. 17.

Можно отметить, что так как в сети помимо слоев подвыборки (pooling слоев) и активации применяются только сверточные слои, поэтому U-Net может принимать на вход и обрабатывать изображения различных размеров, не зависимо от того на каких обучалась.



**Рис. 17:** Архитектура сети U-Net

Основной особенностью UNet 3+ является наличие дополнительных сквозных соединений, а именно помимо соединений между кодером и декодером одного уровня добавляются соединения между декодерами и кодерами более высокого уровня и между декодерами более низкого уровня. Авторы назвали такие соединения "Full-Scale Skip Connections". Они позволяют блокам декодера получить больше низкоуровневых деталей и семантической информации на каждом уровне. Более подробно архитектура сети изображена на рис. 18.



**Рис. 18:** Архитектура сети UNet3+

Помимо добавления новых сквозных соединений авторы предложили механизм, названный "Full-scale Deep Supervision". В нем для получения итоговых масок используется не только выход последнего блока декодера, но и выходы блоков декодера перед ним. Однако в данной работе этот механизм не использовался.

## 3.2 TransUNet

TransUNet, так же как и остальные модели включает в себя кодер и декодер, однако в отличие от традиционных сетей U-Net она использует гибридную архитектуру CNN-трансформер в качестве кодировщика. Трансформер[17], используемый в ее архитектуре является ViT трансформером[18], который был разработан, чтобы применить механизм внимания, изначально разработанный для работы с текстовой информацией в задачах обработки естественного языка, к изображениям. Одна из основных идей ViT это разделение исходного изображения на малые участки (патчи), которые после преобразования в эмбединги могут быть использованы в составных блоках трансформера. Такой подход позволяет объединить преимущества U-Net с возможностью трансформеров извлекать и учитывать информации о глобальном контексте. Архитектура такой нейронной сети изображена на рис. 19.

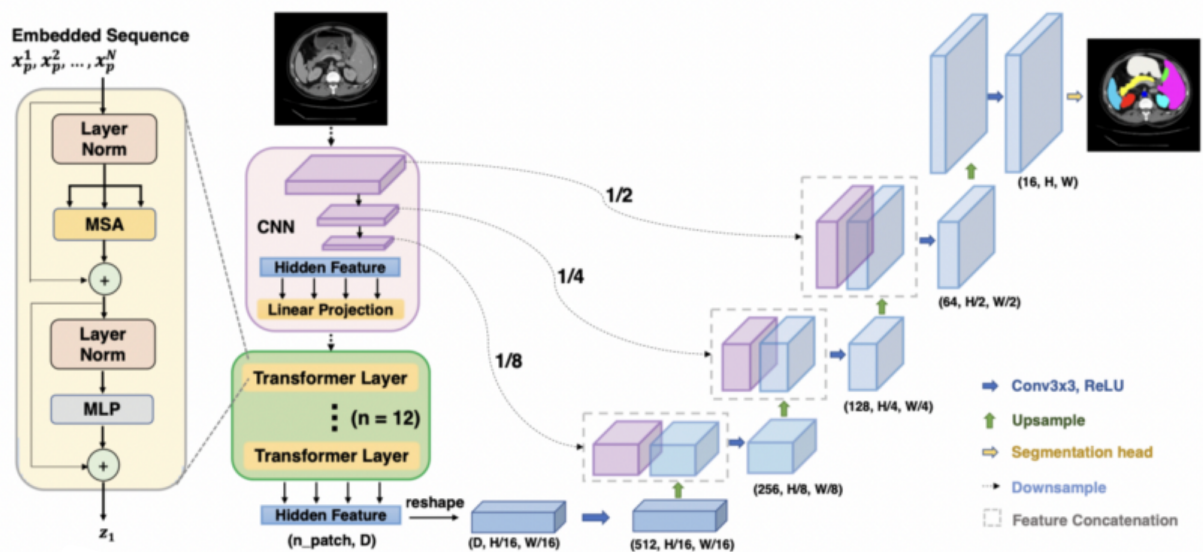


Рис. 19: Архитектура сети TransUNet

## 3.3 Механизмы внимания в CNN

Механизм внимания, позволяет модели абстрагироваться от ненужной ей фоновой или шумовой информации и сосредоточиться на целевых объектах. Данный механизм применяется не только в трансформерах и может быть

использован в сверточных нейронных сетях. В данной работе было рассмотрено два таких способа, один из которых применяется исключительно в сетях класса U-Net, а другой может быть применен к любой сверточной нейронной сети.

Первый механизм был предложен авторами архитектуры Attention Unet [19]. В этой сети используются особые Attention-блоки, добавляющие в оригинальную архитектуру U-Net механизм внимания, что позволяет сосредотачиваться на наиболее релевантных для сегментации областях исходного изображения. Схема и пример работы Attention-блока представлены на рис. 20.

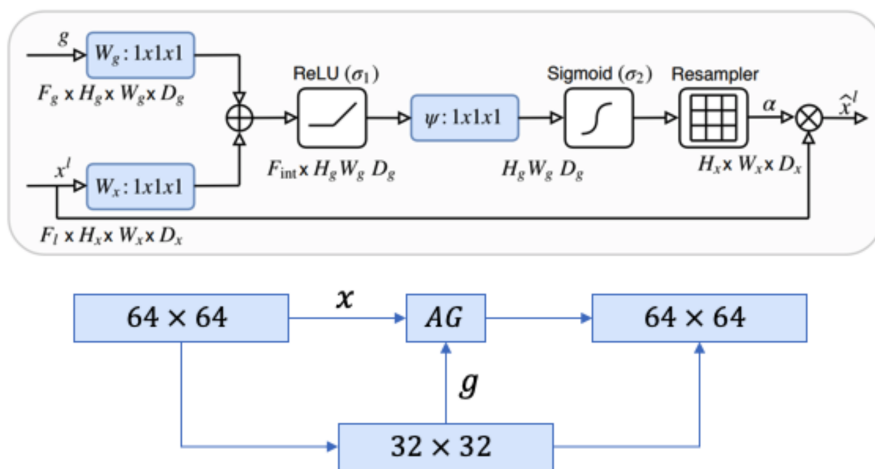


Рис. 20: Схема работы Attention-Gate блока

На вход данный блок принимает тензоры  $x$  (из сквозного соединения) и  $g$  (из нижележащего блока декодера), которые при помощи операций сверток и апсемплинга приводятся к одинаковой размерности и поэлементно складываются, затем применяется функция активации ReLU и еще одна свертка с сигмоидной функцией активации для получения матрицы весов, которая после преобразования к нужной размерности поэлементно перемножается с входным тензором  $x$ .

В качестве второго подхода применения механизма внимания был выбран механизм СВММ (Convolutional Block Attention Module)[20]. В СВММ состоит из двух последовательных модулей: Channel Attention Module (СММ) и Spatial Attention Module (SAM). Авторы предложили применять данный блок к каждому сверточному блоку CNN. СММ позволяет сосредоточиться

на наиболее значимых в текущей задачи фильтрах, а SAM на определенных участках входного изображения. Во многом операции и этапы, применяемые в СВМ и Attention-Gate, схожи. В СВМ добавляются операции пулинга, чтобы получить карты признаков с информацией об фильтрах или пространственной информации соответственно, и также добавляются полносвязные слои в САМ.

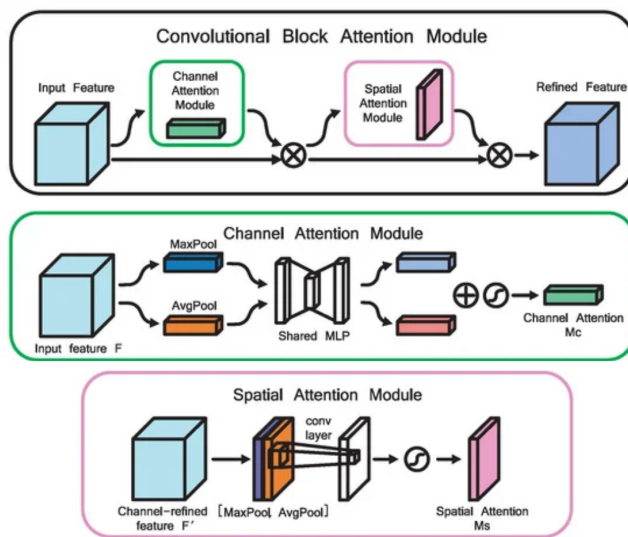


Рис. 21: Схема работы СВМ блока

В данной работе эти механизмы были добавлены в UNet3+, получившие архитектуры изображены на рис. 22 и 23. Идея добавлять механизмы внимания в различные модели семейства U-Net не нова и была предложена, например, в работах [21] [22] [23].

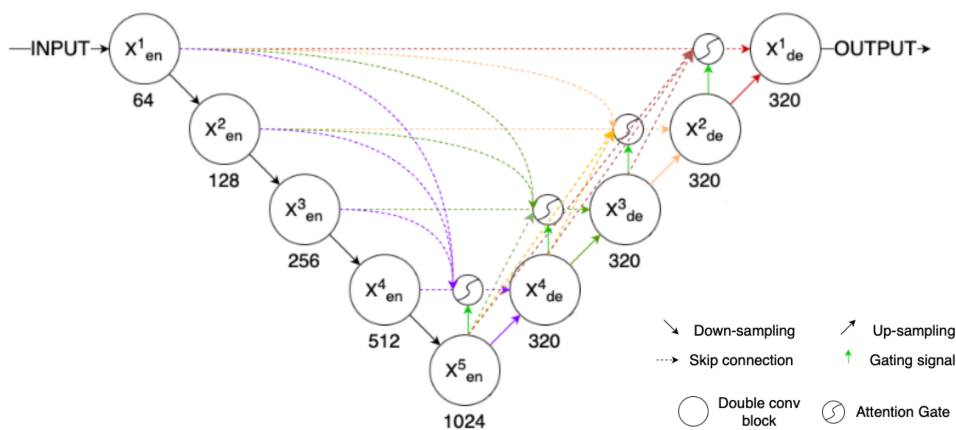


Рис. 22: Архитектура UNet3+ с Attention-Gate

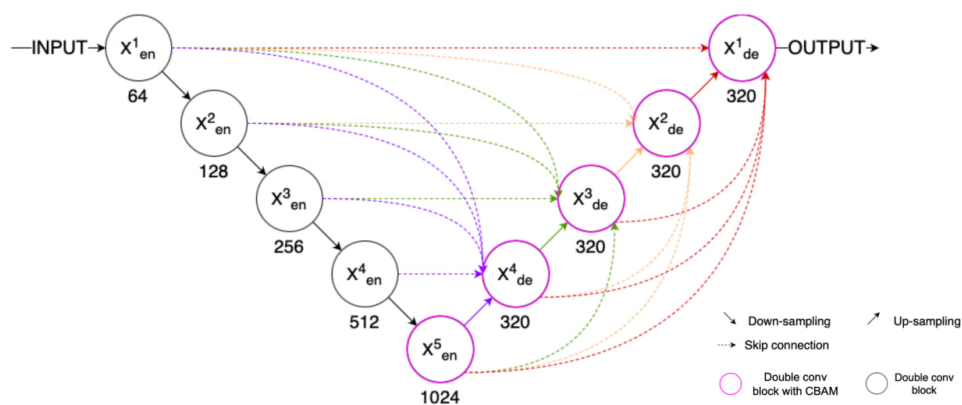


Рис. 23: Архитектура UNet3+ с CBAM

## Глава 4. Реализация и результаты

Все модели были реализованы и обучены с использованием языка программирования Python и открытой программной библиотеки для машинного обучения PyTorch[24].

В процессе обучения моделей появилась проблема несбалансированности данных в используемых датасетах, а именно преобладание фоновых пикселей, не относящихся к целевым классам. Для ее устранения сначала были предприняты попытки использовать весовые коэффициенты в функциях потерь или использовать особые функции потерь, которые могли бы решить данную проблему (например Focal Loss). Однако при таких подходах не вышло получить удовлетворительные результаты и модели не обучались, либо обучались очень медленно.

В конечном итоге для решения этой проблемы было решено использовать специальный модуль Python TorchIO[25], для того, чтобы подавать в процессе обучения не все данные со снимков, а лишь малую часть. В результате на каждой эпохе обучения с каждого снимка в датасете подается определенный участок, с наибольшей вероятностью этот участок будет содержать целевые объекты, если же все-таки подается участок изображения с фоном, то с большей вероятностью он будет содержать пиксели с высоким значением интенсивности сигнала обратного рассеяния. Такой подход позволяет сосредоточиться на наиболее значимой для обучения модели части данных и устранить проблему несбалансированности классов в данных.

Используемые датасеты были разделены на тренировочную, валидационную и тестовую выборки в соотношении 8:1:1. К данным, подаваемым моделям в процессе обучения применялись аугментации, чтобы повысить обобщающую способность моделей и избежать переобучения. Однако из-за особенностей SAR снимков применять к ним большую часть привычных в задачах обработки изображений аугментаций очень сложно, так как они могут слишком сильно изменить исходные данные и отрицательно повлиять на процесс обучения. Поэтому использовались простые пространственные аугментации: горизонтальное отзеркаливание изображений, поворот на угол в пределах от -5 до 5 градусов и масштабирование с коэффициентом от 0.9 до 1.1.

Во всех моделях использовался оптимизатор RMSprop и комбинированная функция ошибки (2).

$$Loss = BCE_{Loss} + Dice_{Loss},$$

where

$$BCE_{Loss}(y, \hat{y}) = -(y * \log(\hat{y}) + (1 - y) * \log(1 - \hat{y})) \quad (2)$$

$$Dice_{Loss}(y, \hat{y}) = 1 - \frac{2 * \sum y_i \hat{y}_j + \varepsilon}{\sum (y_i^2 + \hat{y}_j^2) + \varepsilon}, \quad \varepsilon - \text{some small value}$$

Обучение каждой модели проводилось 100 эпох, после чего модель с лучшими значениями на валидации тестировалась на тестовой выборке. При этом модели для обнаружения айсбергов или кораблей обучались отдельно. Так было сделано из-за того, что в датасете HRSID каждый снимок представлен только в одном режиме поляризации. Это может быть не критично при обнаружении кораблей, однако при детектировании айсбергов, в зависимости от их окружения может возникнуть ситуация когда они заметны только на одном из режимов поляризации, при этом каждый раз этот режим может быть разным. Таким образом, в модель для сегментирования кораблей подавалось одноканальное изображение, а в модель для айсбергов – двухканальное. Все подаваемые в модели изображения имеют размер  $128 \times 128$ .

Для оценки использовались метрики  $IoU$  (Intersection over Union) и  $F_1$ , вычисляемые по формулам (3)-(4).  $F_1$  также известна как коэффициент Сёренсена (Dice similarity coefficient). Результаты обучения моделей представлены в таблицах 2-3. Примеры сегментации моделей с лучшей точностью можно посмотреть в приложениях 1-2.

$$IoU = \frac{|A \cap B|}{|A \cup B|} = \frac{|A \cap B|}{|A| + |B| - |A \cap B|} \quad (3)$$

$$F_1 = DSC = \frac{2|A \cap B|}{|A| + |B|} \quad (4)$$

**Таблица 2:** Результаты обучения моделей на датасете HRSID

Модель	$IoU$	$F_1$
UNet3+	0.7495	0.8409
UNet3+ (AG)	<b>0.7745</b>	<b>0.8610</b>
UNet3+ (CBAM)	0.7681	0.8547
TransUNet	0.7168	0.8161
UNet3+ (AG, $224 \times 224$ )	<b>0.7760</b>	<b>0.8623</b>

**Таблица 3:** Результаты обучения моделей на собранном датасете с айсбергами

Модель	$IoU$	$F_1$
UNet3+	0.4840	0.6131
UNet3+ (AG)	<b>0.5271</b>	<b>0.6587</b>
UNet3+ (CBAM)	0.5111	0.6382
TransUNet	0.4765	0.6046
UNet3+ (AG, $224 \times 224$ )	<b>0.5529</b>	<b>0.6793</b>



Результаты обучения показывают, что лучше всего с сегментированием справляется модель UNet3+ с добавленным в нее механизмом внимания Attention-Gate. После эта модель была дополнительно обучена с входным размером изображения  $224 \times 224$ , что позволило значительно улучшить метрику на датасете с айсбергами и немного на HRSID. В целом добавление механизмов внимания в UNet3+ показывает улучшение метрик и хуже всего справилась модель TransUNet. Здесь стоит уточнить, что первоначально она обучалась с загрузкой заранее натренированных весов для ViT на датасете ImageNet21k[26], однако получающиеся метрики не сильно отличались от конечного результата, представленного в таблицах. В итоге она обучалась с нуля, что при использовании трансформеров представляет очень сложный процесс, так как они требуют огромного набора данных для получения метрик, превосходящих обычные модели. Возможно обычные изображения слишком сильно отличаются от SAR снимков по своей природе, что затрудняет использование заранее обученных на них весов для обучения моделей в исследуемой области.

Репозиторий проекта доступен по ссылке: <https://github.com/nyanet/iceberg-ship-segmentation.git>.

## Заключение

В результате работы были изучены SAR снимки, а именно их особенности и преимущества перед оптическими снимками в рассматриваемой области. Разработан инструмент, позволяющий обрабатывать большое количество снимков удобным способом, что использовалось при сборе датасета с айсбергами. Также, были изучены стандартные методы, которые могли бы быть использованы, но имеют ряд ограничений, из-за чего было решено использовать в работе нейросетевые методы сегментирования объектов на изображениях для решения задачи обнаружения айсбергов и кораблей на спутниковых снимках. Были рассмотрены и реализованы архитектуры существующих моделей и предложена их вариация с использованием дополнительных блоков с механизмом внимания. В конечном итоге были получены модели, способные сегментировать айсберги и корабли на спутниковых SAR снимках.

## Список литературы

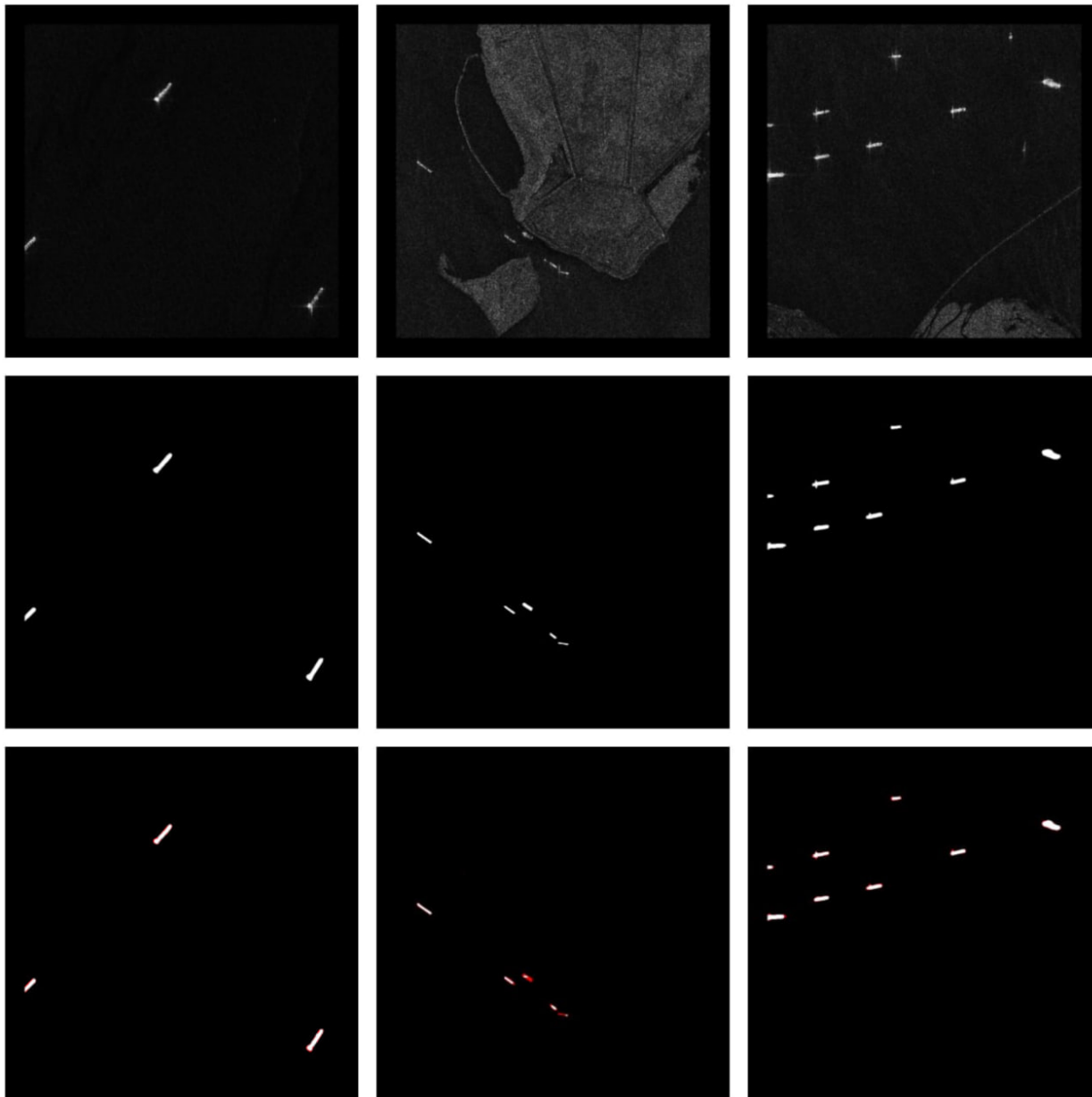
- [1] Copernicus Open Access Hub [Электронный ресурс] URL : <https://scihub.copernicus.eu> (дата обращения: 22.05.2023).
- [2] Copernicus Services Data Hub [Электронный ресурс] <https://cophub.copernicus.eu> (дата обращения: 22.05.2023).
- [3] ASF Data Search [Электронный ресурс] <https://search.asf.alaska.edu> (дата обращения: 22.05.2023).
- [4] EO Browser [Электронный ресурс] <https://apps.sentinel-hub.com/eo-browser> (дата обращения: 22.05.2023).
- [5] ICEYE SAR Data [Электронный ресурс] <https://www.iceye.com/sar-data> (дата обращения: 22.05.2023).
- [6] Xarray backend for Copernicus Sentinel-1 satellite data products [Электронный ресурс] <https://github.com/bopen/xarray-sentinel> (дата обращения: 22.05.2023).

- [7] The Sentinel Application Platform and the Sentinel Toolboxes [Электронный ресурс] <https://step.esa.int/main/download/snap-download/> (дата обращения: 22.05.2023).
- [8] Shunjun Wei, Xiangfeng Zeng, et al. HRSID: A High-Resolution SAR Images Dataset for Ship Detection and Instance Segmentation // *IEEE Access*, vol. 8, P. 120234-120254. 2020.
- [9] Tsung-Yi Lin, Michael Maire, et al. Microsoft COCO: Common Objects in Context // *ArXiv*, abs/1405.0312. 2015.
- [10] International Ice Patrol Iceberg Sightings Database [Электронный ресурс] <https://nsidc.org/data/g00807/versions/1> (дата обращения: 22.05.2023).
- [11] Python wrapper for the ASF SearchAPI [Электронный ресурс] [https://github.com/asfadmin/Discovery-asf\\_search](https://github.com/asfadmin/Discovery-asf_search) (дата обращения: 22.05.2023).
- [12] COCO Annotator, a web-based image annotation tool [Электронный ресурс] <https://github.com/jsbroks/coco-annotator> (дата обращения: 22.05.2023).
- [13] F. C. Robey, D. R. Fuhrmann, E. J. Kelly and R. Nitzberg. A CFAR adaptive matched filter detector. *IEEE Trans. Aerosp. Electron. Syst.* 1992, vol. 28, no. 1, P. 208-216.
- [14] Huimin Huang, Lanfen Lin, et al. UNet 3+: A Full-Scale Connected UNet for Medical Image Segmentation // *ArXiv*, abs/2004.08790. 2020.
- [15] Jieneng Chen, Yongyi Lu, et al. TransUNet: Transformers Make Strong Encoders for Medical Image Segmentation // *ArXiv*, abs/2102.04306. 2021.
- [16] Ronneberger O., Fischer P., Brox T. U-Net: Convolutional Networks for Biomedical Image Segmentation. // *Medical Image Computing and Computer-Assisted Intervention. MICCAI*. 2015. P. 234—241

- [17] Ashish Vaswani, Noam Shazeer, et al. Attention Is All You Need // ArXiv, abs/1706.03762. 2021.
- [18] Alexey D., Lucas B., Alexander K., et al. An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale // ArXiv, abs/2010.11929. 2021.
- [19] Ozan Oktay, Jo Schlemper, et al. Attention U-Net: Learning Where to Look for the Pancreas // ArXiv, abs/1804.03999. 2018.
- [20] Sanghyun Woo, Jongchan Park, et al. CBAM: Convolutional Block Attention Module // ArXiv, abs/1807.06521. 2018.
- [21] Chaowei Jiang, Hong Zhang, et al. Water Surface Mapping from Sentinel-1 Imagery Based on Attention-UNet3+: A Case Study of Poyang Lake Region // Remote Sens. 2022.
- [22] Pooya Mohammadi Kazaj, Alireza Vafaei Sadr, et al. U-Net-based Models for Skin Lesion Segmentation: More Attention and Augmentation // 2022.
- [23] Kevin Trebing, Tomasz Stanczyk, et al. SmaAt-UNet: Precipitation Nowcasting using a Small Attention-UNet Architecture // ArXiv, abs/2007.04417 2020.
- [24] PyTorch [Электронный ресурс] URL:<https://pytorch.org> (дата обращения: 22.05.2023).
- [25] TorchIO, an open-source Python library for efficient loading, preprocessing, augmentation and patch-based sampling of images in deep learning [Электронный ресурс] URL:<https://torchio.readthedocs.io> (дата обращения: 22.05.2023).
- [26] Tal Ridnik, Emanuel Ben-Baruch, et al. ImageNet-21K Pretraining for the Masses // ArXiv, abs/2104.10972. 2021.

## Приложение 1. Примеры сегментации кораблей

Сверху вниз: снимок, маска, предсказание. Красным цветом выделены различия между предсказанием и маской.



## Приложение 2. Примеры сегментации айсбергов

Сверху вниз: снимок (в поляризации HV), маска, предсказание. Красным цветом выделены различия между предсказанием и маской.

