|  |  |
| --- | --- |
| Gerb-BMSTU_01 | **Министерство науки и высшего образования Российской Федерации**  **Федеральное государственное бюджетное образовательное учреждение**  **высшего образования**  **«Московский государственный технический университет**  **имени Н.Э. Баумана**  **(национальный исследовательский университет)»**  **(МГТУ им. Н.Э. Баумана)** |

ФАКУЛЬТЕТ ИНФОРМАТИКА И СИСТЕМЫ УПРАВЛЕНИЯ

КАФЕДРА СИСТЕМЫ ОБРАБОТКИ ИНФОРМАЦИИ И УПРАВЛЕНИЯ

**РАСЧЕТНО-ПОЯСНИТЕЛЬНАЯ ЗАПИСКА**

***К НАУЧНО-ИССЛЕДОВАТЕЛЬСКОЙ РАБОТЕ***

***НА ТЕМУ:***

**Исследование методов сегментации спутниковых снимков**

Студент группы ИУ5-71б **\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_** П.К. Шимолина

(код группы) (подпись, дата) (инициалы и фамилия)

Научный руководитель **\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_** А.Н. Нардид

(подпись, дата) (инициалы и фамилия)

Оценка \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Руководитель от кафедры **\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_** В.И. Терехов

(подпись, дата) (инициалы и фамилия)

Оценка \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Москва, 2023 г.

**Министерство науки и высшего образования Российской Федерации**

**Федеральное государственное бюджетное образовательное учреждение**

**высшего образования**

**«Московский государственный технический университет имени Н.Э. Баумана**

**(национальный исследовательский университет)»**

**(МГТУ им. Н.Э. Баумана)**

УТВЕРЖДАЮ

Заведующий кафедрой ИУ5

\_\_\_\_\_\_\_\_\_\_\_\_  В.И. Терехов

« \_\_\_\_ » \_\_\_\_\_\_\_\_\_\_ 2023 г.

**ЗАДАНИЕ**

**на выполнение научно-исследовательской работы**

по теме: Исследование методов сегментации спутниковых снимков

**Студент группы** ИУ5-71б Шимолина Полина Кирилловна   
 (Фамилия имя отчество)

**Направленность НИР** (учебная, исследовательская, практическая, производственная, др.) исследовательская

**Источник тематики** (кафедра, предприятие, НИР) учебная тематика

**График выполнения НИР**: 25% к 5 нед., 50% к 9 нед., 75% к 13 нед., 100% к 16 нед.

***Техническое задание*:**Исследование методов бинаризации изображений, исследование и сравнение сегментации спутниковых снимках с использованием нейронных сетей

***Оформление научно-исследовательской работы:***

Расчетно-пояснительная записка, минимальный объем 12 листов формата А4.

Приложения: графический (иллюстративный) материал (чертежи, схемы, диаграмма и т.п.)

Дата выдачи задания «15» октября 2023 г.

Научный руководитель **\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_** А.Н. Нардид

(подпись, дата) (инициалы и фамилия)

Студент группы ИУ5-71б **\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_** П.К. Шимолина

(код группы) (подпись, дата) (инициалы и фамилия)

Примечание: Задание оформляется в двух экземплярах: один выдается студенту, второй хранится на кафедре.

**Содержание**

[ВВЕДЕНИЕ 3](#_Toc154015589)

[1. Типы задач сегментации изображений 4](#_Toc154015590)

[2. Бинаризация изображений 6](#_Toc154015591)

[2.1. Метод Оцу 6](#_Toc154015592)

[2.2. Метод Бернсена 9](#_Toc154015593)

[2.3. Метод Эйквеля 9](#_Toc154015594)

[3. Архитектуры нейронный сетей, применяемые для сегментации 11](#_Toc154015595)

[3.1. U-Net 11](#_Toc154015596)

[3.2. Mask R-CNN 12](#_Toc154015597)

[3.3. SegNet 15](#_Toc154015598)

[3.4. Сравнение архитектур 16](#_Toc154015599)

[ЗАКЛЮЧЕНИЕ 18](#_Toc154015600)

# **ВВЕДЕНИЕ**

Сегментация изображений на спутниковых снимках является важной задачей в области обработки и анализа геопространственных данных. Автоматическое распознавание спутниковых снимков представляет собой наиболее перспективный подход для получения информации о расположении различных объектов на местности. Особенно актуальным становится отказ от ручной сегментации снимков, когда необходимо обработать большие участки земной поверхности в ограниченные сроки. Автоматизированный процесс распознавания позволяет существенно сократить время и усилия, а также минимизировать ошибки, связанные с человеческим фактором.

Распознавание объектов — это процесс нахождения конкретного объекта на изображении или видеопоследовательности. Каждому объекту на изображении соответствуют уникальные характеристики, которые описывают его особенности и свойства. Эти характеристики извлекаются из обучающего набора данных и используются для идентификации местоположения объекта, а затем сравниваются с множеством других объектов из тестового набора данных. Таким образом, главная цель обучения сверточных нейронных сетей (СНС) заключается в определении и извлечении характеристик объекта на изображении.

Цели работы: исследовать существующие методы и арзитектуры, применяемые для сегментации изображений на спутниковых снимках.

# **Типы задач сегментации изображений**

Задачи сегментации изображения можно разделить на типы на основе количества и типа информации, которую они передают. (см. рис.1)

1. Семантическая сегментация.
2. Сегментация экземпляра
3. Паноптическая сегментация

Семантическая сегментация относится к классификации пикселей изображения по семантическим классам. Пиксели, принадлежащие к определенному классу, просто относятся к этому классу без учета другой информации или контекста. [10] Сегментация работает со множеством объектов одного класса как с единым целым.

Модели сегментации экземпляров классифицируют пиксели по категориям на основе «экземпляров», а не классов. Алгоритм сегментации экземпляров не знает к какому классу принадлежит классифицируемая область, но может разделить перекрывающиеся или очень похожие области объекта на основе их границ.

Паноптическая сегментация, самая последняя разработанная задача сегментации, может быть выражена как комбинация семантической сегментации и сегментации экземпляров, при которой каждый экземпляр объекта на изображении выделяется и предсказывается идентичность объекта.

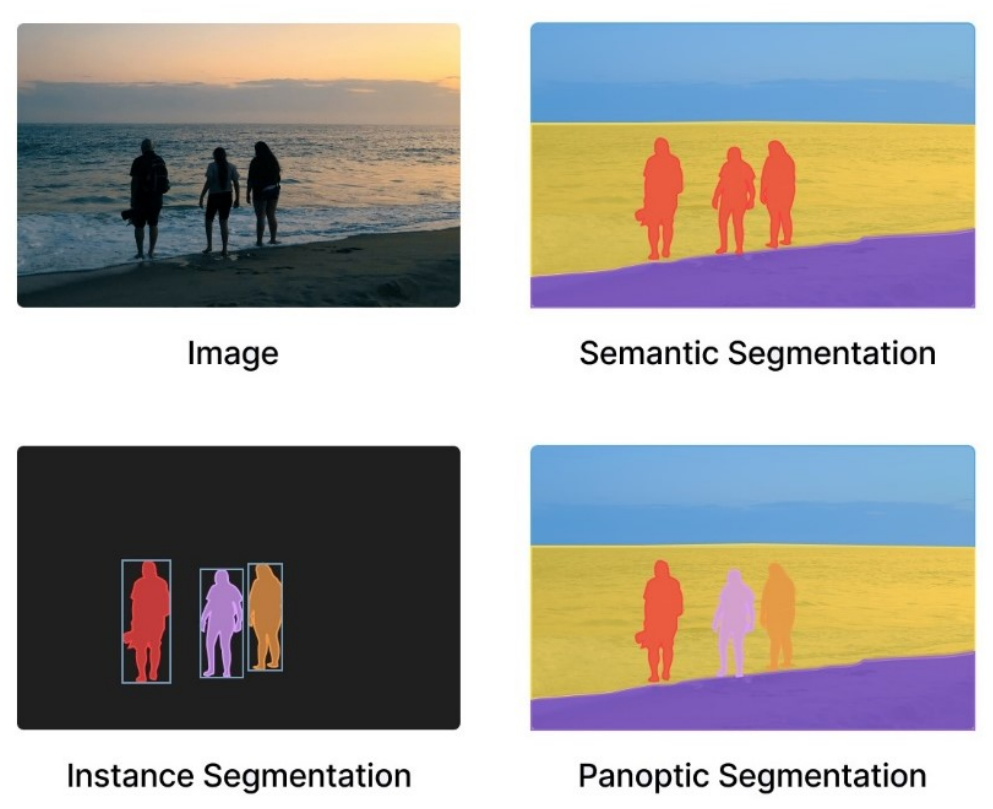


Рисунок 1 – Типы сегментации

# **Бинаризация изображений**

Бинаризация — это процесс преобразования цветного изображения или изображения в градациях серого в двухцветное черно-белое. Яркость каждого пикселя сравнивается с пороговым значением, и ему присваивается либо 0, если пиксель является границей объекта, либо 1 если не является. Бинаризация используется для уменьшения количества информации, с которой будет производиться работа. Существуют пороговые (глобальные) и адаптивные (локальные) методы бинаризации. В пороговых происходит работа со всем изображением сразу. Величина порога постоянна для всего процесса бинаризации.

## Метод Оцу

Метод Оцу (Отсу) – метод глобальной пороговой обработки, минимизирующий среднюю ошибку от принятия решения о принадлежности пикселей изображения объекту или фону.

Простейший метод пороговой обработки – использование гистограммы изображения. На основании общего для всего изображения порога каждый пиксель обозначается как объект или фон. Такой способ подходит для случаев, когда гистограмма носит четко выраженный бимодальный характер (см. рис.2).

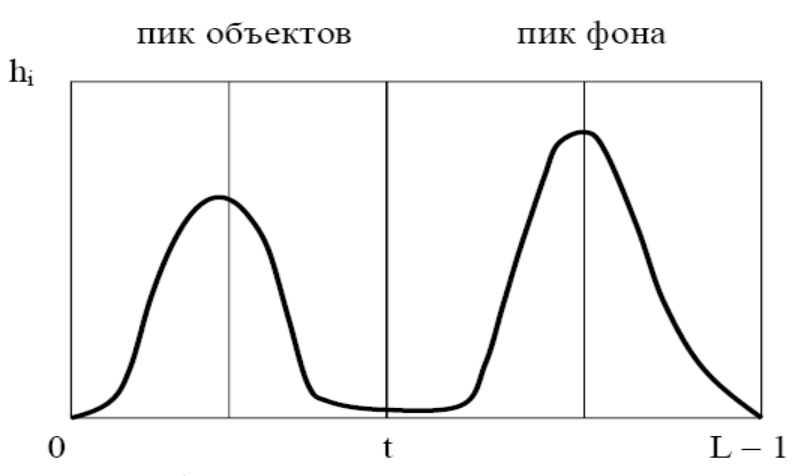


Рисунок 2 – Гистограмма с двумя пиками (t – глобальный порог)

Для определения оптимального порога бинаризации T предложено огромное количество различных алгоритмов. Как правило, такие методы опираются на использование локальной или глобальной гистограммы изображения. Одним из наиболее широко используемых подходов является подход, предложенный Оцу; преимуществами данного алгоритма являются понятный статистический смысл и низкая вычислительная сложность. [2] В идеальном случае гистограмма имеет глубокую и четкую впадину между двумя пиками, представляющими объекты и фон соответственно, так что порог может быть выбран в нижней части этой впадины. Однако на большинстве реальных снимков часто бывает трудно точно определить дно долины, особенно в таких случаях, когда долина плоская и широкая, наполненная шумом, или когда две вершины чрезвычайно неравны по высоте, часто не образуя прослеживаемой долины. Предложен новый метод с точки зрения дискриминантного анализа; он непосредственно приближается к возможности оценки "добротности" порога и автоматического выбора оптимального порога. [4]

Метод бинаризации Оцу используется для определения порога T, который минимизирует среднюю ошибку сегментации изображения, возникающую при принятии решения о принадлежности пикселей изображения объекту или фону. В данном подходе значения яркостей пикселей рассматриваются как случайные величины, а их гистограмма - как оценка плотности распределения вероятностей. Если мы знаем плотности распределения вероятностей, то мы можем определить оптимальный порог, минимизирующий ошибку для разделения изображения на два класса: объекты и фон.

Рассмотрим относительные частоты для этих классов: [1]

где w1 и w2 – области объектов и фона соответственно; i – номер элемента матрицы изображения; i p – значения гистограммы; k – количество элементов в первой или второй области (объект или фон); L – максимальное значение, которое может принять T; µ1, µ2 – средняя яркость всего изображения.

Тогда дисперсия внутри классов следующая:

Дисперсия внутри классов может быть рассчитана как взвешенная сумма дисперсий каждого отдельного класса:

Дисперсия между двумя классами:

Дисперсия всего изображения:

Метод Оцу выбирает оптимальный порог T, c помощью вычисления максимизации между классом дисперсии:

Метод ОЦУ прост в реализации, быстр в выполнении и адаптивен к различным изображениям, но чувствителен к неравномерной яркости изображения.

## Метод Бернсена

Метод Бернсена реализован и протестирован для различных значений соседства и пределов контрастности.

Изображение делится на квадраты r × r (r – нечетное) с центром в точке (x, y). [3] Для каждого пикселя изображения в пределах квадрата считается порог согласно равенству:

Где Zlow и Zhigh – самые низкие и самые высокие значения пикселей уровня серого в квадрате r x r с центром в точке (x, y) в соответствии с уравением:

Если показатель контрастности C(x, y) < l, то окрестности состоят только из одного класса, переднего плана или фона. Кроме того, значения r и l варьируются в зависимости от используемых изображений и областей.

Применение метода включает следующе этапы: [5]

1. Считать изображение в оттенках серого
2. Определить значение близости (r) и предел контрастности (l)
3. Найти наименьшее значение пикселя уровня серого (Zlow) в квадрате r x r с центром в точке (x, y)
4. Найти наибольшее значение пикселя уровня серого (Zhigh) в квадрате r x r с центром в точке (x, y)
5. Посчитать пороговое значение T(x, y)
6. Посчитать показатель контрастности C(x, y)
7. Обозначить пиксель (x, y) как передний план или фон в соответствии с рассчитанным пороговым значением и мерой контрастности
8. Повторить пункты 3 – 7 для каждого пикселя

## Метод Эйквеля

В этом методе используются два окна r и R, из которых большее по размерам (окно L) служит для вычисления значения порога, в то время как меньшее (окно S) определяет область изображения, в которой будет использоваться полученный порог. Оба окна перемещаются параллельно по изображению с шагом, равным размеру меньшего окна S, и каждый раз для всех элементов окна L вычисляется оптимальный порог по методу Отса (см. рис.3). [3]

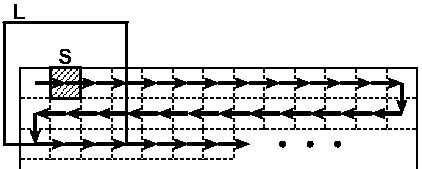


Рисунок 3 – Перемещение окон в методе Эйквеля

Если вычисленные средние значения µ0 и µ1 различаются сильно и для них выполняется условие |µ0 - µ1| ≥ ε, где ε – заданный параметр, то пиксели внутри окна S подвергаются бинаризации в соответствии с вычисленным порогом t. Если же |µ0 - µ1| < ε, то все пиксели внутри окна S относятся к классу с ближайшим средним значением/

# **Архитектуры нейронный сетей, применяемые для сегментации**

Сверточные нейронные сети являются мощным инструментом в области компьютерного зрения и имеют широкое применение в задачах сегментации изображений. В последние годы было предложено множество различных моделей СНС для сегментации, каждая из которых имеет свои особенности и преимущества.

## U-Net

U-Net является одной из основных архитектур сверточных нейронных сетей (СНС) для задач сегментации изображений. Применяется не только для определения класса изображения, но и для разделения на области и создания маски для каждого класса. U-Net – очень быстрая сеть, способная сегментировать изображение 512x512 менее чем за секунду.

Архитектура состоит из стягивающего пути для захвата контекста и симметричного расширяющегося пути, который позволяет осуществить точную локализацию (см. рис.4). [6] То есть стягивающий путь содержит слои, фиксирующие информацию с изображений, уменьшающие пространственное разрешение входных данных, а расширяющийся путь содержит слои, декодирующие закодированные данные, определяющие расположение объектов и создающие маски сегментации, используя информацию, полученную на стягивающем пути.

Рисунок 3 демонстрирует преобразование изображение размером 572 на 572 в оттенках градации серого в двоичную сегментированную выходную карту размером 388×388×2. Размер выходных данных меньше размера входных данных, потому что не используется заполнение. Однако, если использовать “padding”, можно сохранить размер входных данных.

В процессе сжатия пути входное изображение постепенно уменьшается по высоте и ширине, но увеличивается количество каналов. Это увеличение каналов позволяет сети захватывать объекты высокого уровня по мере продвижения по пути. В узком месте выполняется заключительная операция свертки для создания карты объектов размером 30 × 30 × 1024. Затем расширенный путь извлекает карту объектов из узкого места и преобразует ее обратно в изображение того же размера, что и исходные входные данные. Это делается с использованием слоев с повышающей дискретизацией, которые увеличивают пространственное разрешение карты объектов при одновременном сокращении количества каналов. Наконец, каждый пиксель на выходном изображении представляет собой метку, соответствующую определенному объекту или классу на входном изображении. [7]

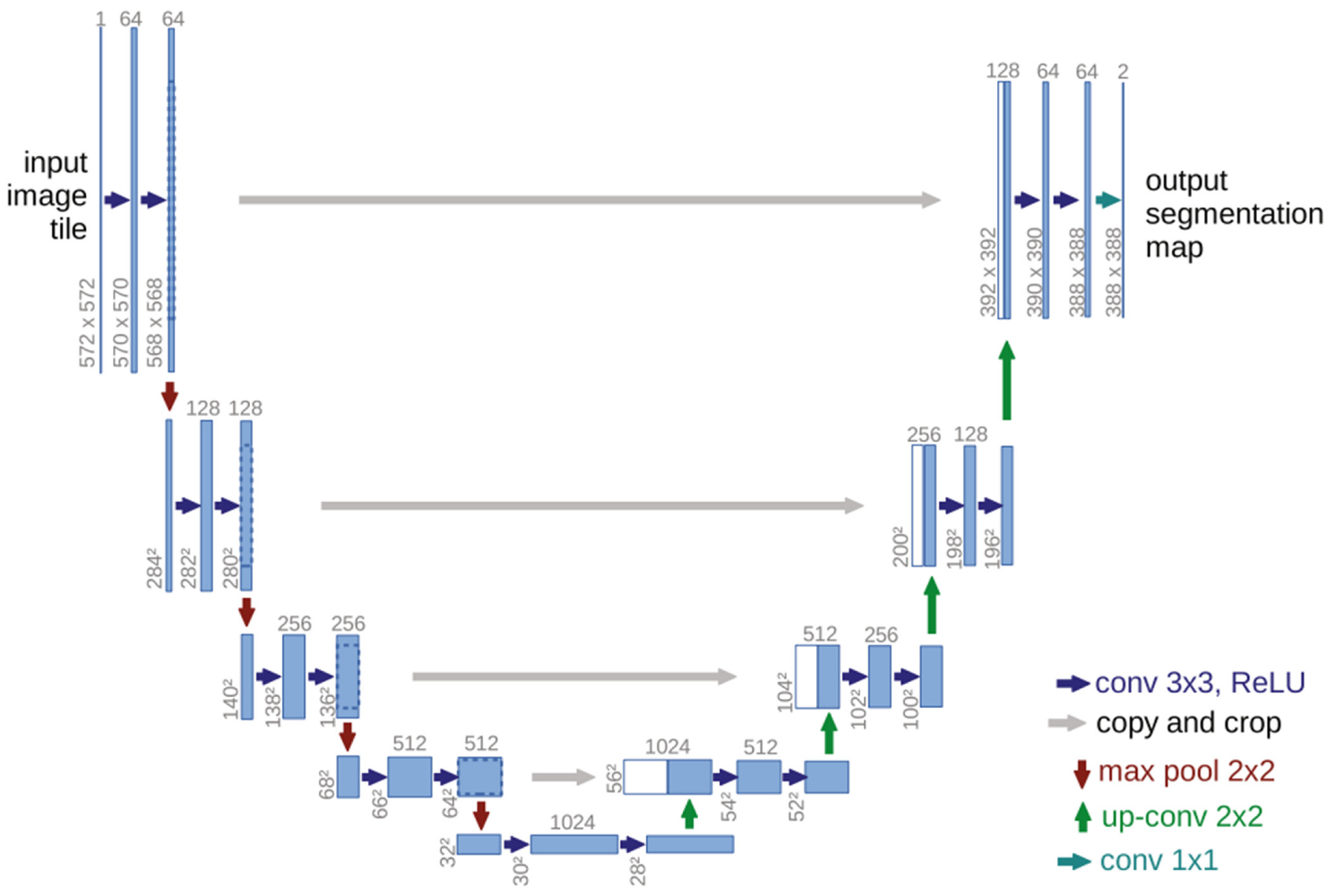


Рисунок 4 – Архитектура U-Net

## Mask R-CNN

R-CNN или RCNN, расшифровывается как Region-Based Convolutional Neural Network – сверточная нейронная сеть на основе регионов, это тип модели машинного обучения, которая используется для задач компьютерного зрения, в частности для обнаружения объектов.

Рассмотрим алгоритм работы. Сначала изображение разделяется на множество прямоугольных регионов, которые могут содержать объекты. Эти регионы выбираются с помощью алгоритма селекции регионов. Затем каждый выбранный регион подвергается процессу извлечения признаков с помощью сверточной нейронной сети (CNN). В качестве CNN-сети используется готовая CaffeNet, классифицирующая на 1000 классов. CNN анализирует содержимое каждого региона и извлекает характеристики, которые могут помочь в определении наличия объекта. Полученные признаки передаются в классификатор, который определяет, принадлежит ли регион к одному из заданных классов объектов или относится к фону (не содержит объекта). Наконец, если регион классифицирован как содержащий объект, применяется дополнительная обработка, чтобы точнее определить границы объекта и создать окончательную маску или охватывающую рамку объекта. (см. рис.5)

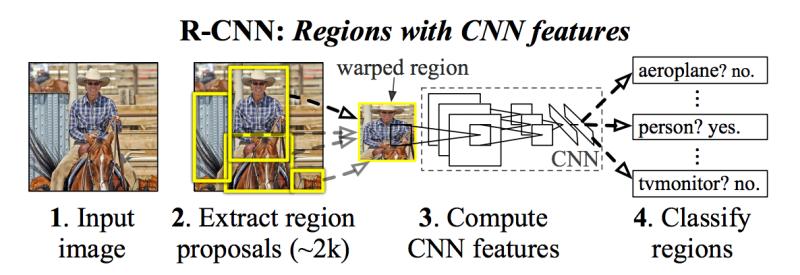


Рисунок 5 – Схема R-CNN

Далее была разработана Fast R-CNN, которая отличалась тем, что через CNN пропускалось все изображение целиком. (см. рис.6)

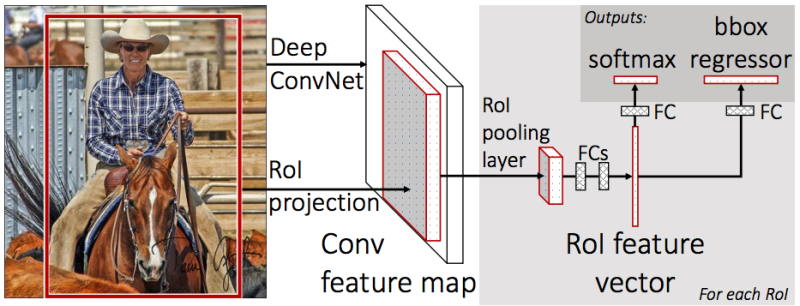


Рисунок 6 – Схема Fast R-CNN

Следующей была создана Faster R-CNN. Которая вычислять регионы не по изначальному изображению, а опять же по карте признаков, полученных из CNN. Для этого был добавлен модуль под названием Region Proposal Network (RPN). (см. рис.7)

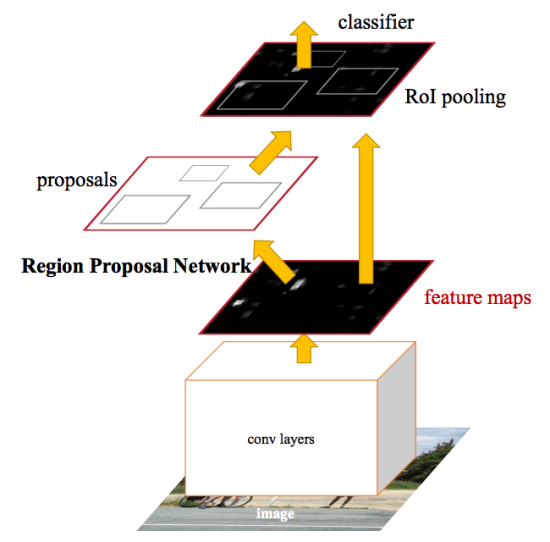


Рисунок 7 - Схема Faster R-CNN

Mask R-CNN развивает архитектуру Faster R-CNN. Существует 2 основных типа сегментации изображений, которые подпадают под Mask R-CNN:  
1. Семантическая сегментация  
2. Сегментация экземпляров

Семантическая сегментация классифицирует каждый пиксель по фиксированному набору категорий без дифференциации экземпляров объектов. То есть семантическая сегментация имеет дело с идентификацией/классификацией похожих объектов как единого класса на уровне пикселей.

Сегментация экземпляров имеет дело с правильным обнаружением всех объектов на изображении, а также с точной сегментацией каждого экземпляра. Таким образом, это комбинация обнаружения объектов, локализации объектов и классификации объектов. Другими словами, этот тип сегментации идет дальше, чтобы провести четкое различие между каждым объектом, классифицированным как похожие экземпляры.

Маска R-CNN является расширением более быстрого R-CNN и работает путем добавления ветви для прогнозирования маски объекта (интересующей области) параллельно с существующей ветвью для распознавания ограничивающих рамок. [8]

Маска представляет собой матрицу, где каждый пиксель имеет значение 1 или 0. (см. рис.8)

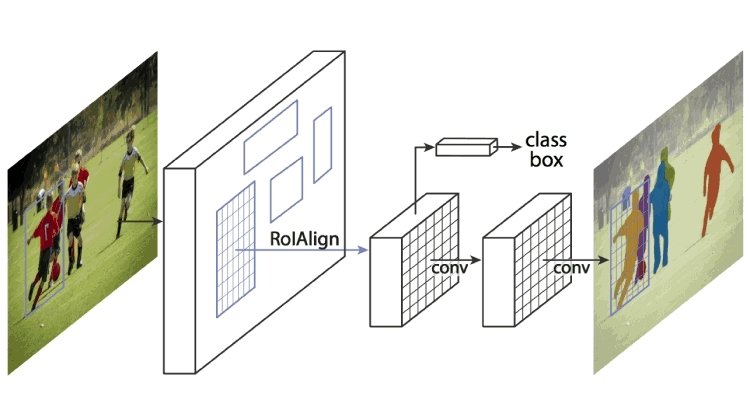


Рисунок 8 - Схема Mask R-CNN

## SegNet

SegNet - это эффективная архитектура для пиксельной семантической сегментации.

SegNet имеет сеть кодировщиков и соответствующую сеть декодеров, за которыми следует последний слой классификации по пикселям. (см. рис.9)

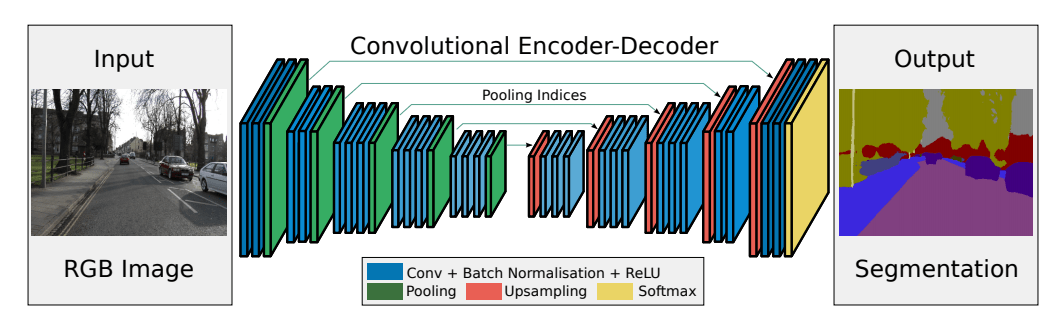


Рисунок 9 – Архитектура SegNet

Сеть кодировщика состоит из 13 сверточных слоев которые соответствуют первым 13 сверточным слоям в сети VG16, предназначенной для классификации объектов. Каждый уровень кодера имеет соответствующий уровень декодера, и, следовательно, сеть декодера имеет 13 уровней. Окончательный вывод декодера подается на мульти-классификатор soft-max для независимого получения вероятностей классов для каждого пикселя.

SegNet является эффективным и гибким инструментом для сегментации изображений, который обладает рядом преимуществ и особенностей, делающих его популярным в области компьютерного зрения.

## Сравнение архитектур

В таблице 1 представлены сравнительные характеристики методов для сегментации изображений, описанных выше.

Таблица 1 – Сравнение методов для сегментации изображений

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **№** | **Метод** | **Архитектура** | **Типы изображений** | **Производительность** |
| 1 | U-Net­ | Энкодер-декодер | Разработана для медиинских изображений с возможностью адаптации для других типов | Хорошая производительность для разных типов данных |
| 2 | Mask R-CNN | Расширенная Faster R-CNN и полносвязная нейросеть | Для любых типов, включая спутниковые снимки | Хорошо работает для сложных форм и нескольких экземпляров |
| 3 | SegNet | Энкодер-декодер | Для любых типов, включая спутниковые снимки | Хорошо справляется с сегментацией небольших классов |

# **ЗАКЛЮЧЕНИЕ**

В данной работе был проведен анализ методов сегментации спутниковых снимков. Сегментация спутниковых снимков имеет большое значение в различных областях, таких как геология, геодезия, сельское хозяйство и многое другое. Она позволяет выделять и классифицировать различные объекты и структуры на спутниковых изображениях, что помогает в понимании и анализе окружающей среды.

В работе были рассмотрены различные типы задач сегментации спутниковых снимков. Каждая из этих задач требует специфических методов и подходов для достижения точной и качественной сегментации

Также были изучены различные архитектуры нейронных сетей, применяемых для сегментации спутниковых снимков. В работе были рассмотрены такие архитектуры, как U-Net, Mask R-CNN и SegNet. Каждая из этих архитектур имеет свои особенности и преимущества в контексте сегментации спутниковых снимков.

В заключение, сегментация спутниковых снимков является важным инструментом для анализа и интерпретации данных, полученных с помощью спутниковых систем. Выбор подходящего метода и архитектуры для сегментации зависит от конкретной задачи, типа спутниковых снимков и требований к производительности. Понимание различных методов и архитектур помогает в выборе оптимального подхода и достижении высокой точности и качества сегментации спутниковых снимков.

**СПИСОК ИСПОЛЬЗУЕМЫХ ИСТОЧНИКОВ**

1. Годунов А. И., Баланян С. Т., Егоров П. С. Сегментация изображений и распознавание объектов на основе технологии сверточных нейронных сетей // Надежность и качество сложных систем. 2021. № 3. С. 62–73. doi:10.21685/2307-4205-2021-3-8
2. Хаустов П.А. Алгоритмы распознавания рукописных символов на основе построения структурных моделей // КО., 2017. № 1.
3. Янковский А.А., Бугрий А.Н. Критерии выбора метода бинаризации при обработке изображений лабораторных анализов // АСУ и приборы автоматики., 2010. № 153.
4. *N. Otsu.* A threshold selection method from gray-level histograms (англ.) // IEEE Trans. Sys., Man., Cyber. : journal. — 1979. — Vol. 9. — P. 62—66
5. Can Eyupoglu Implementation of Bernsen’s locally adaptive binarization method for gray scale images [Текст] / Can Eyupoglu // The Online Journal of Science and Technology. — 2017. — № Volume 7, Issue 2. — С. 68-72.
6. Павел Глек U-Net: нейросеть для сегментации изображений / Павел Глек [Электронный ресурс] // Neurohive : [сайт]. — URL: https://neurohive.io/ru/vidy-nejrosetej/u-net-image-segmentation/ (дата обращения: 20.12.2023).
7. U-Net Architecture Explained / [Электронный ресурс] // GeeksForGeeks : [сайт]. — URL: https://www.geeksforgeeks.org/u-net-architecture-explained/ (дата обращения: 20.12.2023).
8. Elisha Odemakinde Everything about Mask R-CNN: A Beginner’s Guide Read more at: https://viso.ai/deep-learning/mask-r-cnn/ / Elisha Odemakinde [Электронный ресурс] // viso.ai : [сайт]. — URL: https://viso.ai/deep-learning/mask-r-cnn/ (дата обращения: 20.12.2023).
9. Vijay Badrinarayanan, Alex Kendall, Roberto Cipolla SegNet: A Deep Convolutional Encoder-Decoder Architecture for Image Segmentation [Текст] / Vijay Badrinarayanan, Alex Kendall, Roberto Cipolla — 2016.
10. Д.В. Лукашик Анализ современных методов сегментации изображений [Текст] / Д.В. Лукашик // ЭКОНОМИКА И КАЧЕСТВО СИСТЕМ СВЯЗИ. — 2022. — № 2. — С. 57-65.