

Федеральное государственное автономное образовательное учреждение
высшего образования ”Московский физико-технический институт
(национальный исследовательский университет)”

На правах рукописи

Емельянов Антон Владимирович

**Дешифрирование и векторизация аэро- и космofотоснимков
методами машинного обучения для обновления
геопространственной информации**

Специальность 1.2.1.

Искусственный интеллект и машинное обучение

Диссертация на соискание учёной степени
кандидата технических наук

Научный руководитель:
кандидат технических наук
Князь Владимир Александрович

Долгопрудный — 2025

Оглавление

	Стр.
Введение	5
 Глава 1. Анализ существующих современных методов сегментации и векторизации аэро- и космofотоснимков	
1.1 Семантическая сегментация спутниковых изображений	11
1.1.1 Свёрточные нейронные сети для семантической сегментации	13
1.1.2 Графовая сверточная нейронная сеть	16
1.2 Методы извлечения контуров зданий со спутниковых изображений	17
1.2.1 Нейросетевые подходы	18
1.2.2 Наборы данных со спутниковыми снимками	20
1.3 Обучение с подкреплением	21
1.4 Выводы	25
 Глава 2. Усовершенствование извлечения контуров зданий с помощью методов регуляризации	
2.1 Математическая постановка задач	27
2.1.1 Задача семантической сегментации	27
2.1.2 Задача векторизации	29
2.2 Извлечение контуров зданий с использованием свойства линейной связности	30
2.2.1 Сегментация экземпляров с помощью Mask R-CNN	32
2.2.2 Применение регуляризации к предсказаниям	33
2.3 Анализ взаимосвязи пикселей для извлечения контуров зданий . . .	35
2.3.1 Бинарная сегментация с U-netFormer	36
2.3.2 Применение регуляризации к предсказаниям	37

2.4	Извлечения контуров зданий с помощью методов многоэтапной регуляризации	40
2.4.1	Семантическая сегментация с U-NetFormer	41
2.4.2	Применение разработанного метода регуляризации	43
2.5	Результаты работы методов	46
2.5.1	Метрики оценки	46
2.5.2	Эксперимент	46
2.6	Выводы	49

Глава 3. Нейросетевой алгоритм на основе обучения с подкреплением для векторизации изображений дистанционного зондирования Земли

	дистанционного зондирования Земли	51
3.1	Архитектура системы	51
3.2	Оптимизация обучения с подкреплением	53
3.2.1	Усовершенствованный алгоритм DQN	53
3.2.2	Реализация оптимизации проксимальной политики	54
3.3	Система вознаграждений	54
3.4	Результаты проведённых экспериментов	55
3.4.1	Метрики оценки	55
3.4.2	Эксперимент	57
3.5	Выводы	59

Глава 4. Семантическая сегментация изображений дистанционного зондирования на основе визуального трансформера и

	графовой нейронной сети	61
4.1	Выборка SVAI	63
4.2	Архитектура нейросетевой модели	66
4.2.1	Извлечение глубоких признаков	67
4.2.2	Представление графа изображения	68

	Стр.
4.2.3 Кластеризация графовых нейронных сетей	71
4.3 Результаты	72
4.4 Выводы	73
Заключение	75
Словарь терминов	77
Список литературы	79
Список рисунков	95
Список таблиц	98

Введение

Число задач, в которых требуется актуальное и достоверное геоинформационное обеспечение расширяется, увеличивается с каждым годом. Основным источником геоинформационного обеспечения в настоящее время является аэро- и космофотосъёмка. Однако сложность первичного дешифрирования аэро- и космических снимков вызывает значительное количество ошибок и создаёт временную задержку в доставке актуальных векторных карт потребителям геоинформационных данных.

С другой стороны, в последнее десятилетие произошёл технологический и научный прорыв в методах обработки изображений, который связан с появлением высокопроизводительных массово-параллельных вычислителей и средствами эффективного обучения глубоких свёрточных нейронных сетей. Несмотря на значительный прогресс нейросетевых методов, существующие на текущий момент методы не обеспечивают качества обработки спутниковых и аэрофотоснимков, сравнимого с качеством построения векторных карт человеком-оператором.

Данный факт обуславливает актуальность и создаёт предпосылки для разработки методов и алгоритмов интеллектуального анализа аэро- и космоснимков для обновления и векторизации геопространственной информации в обеспечение создания отечественного программного комплекса обновления картографического обеспечения.

Развитие методов семантической сегментации и векторизации аэро- и космофотоснимков, берущее начало на заре компьютерного зрения, прошло последовательную эволюцию от аналитических методов к глубоким свёрточным сетям, где ключевую роль сыграло наращивание вычислительных возможностей. Аналитические методы сегментации аэрофотоснимков включают пороговые алгоритмы (Оцу, адаптивное пороговое разделение), операторы выделения краев (Собеля, Кэнни) и морфологические операции (эрозия, дилатация). Методы векторизации основаны на алгоритмах аппроксимации контуров (Рамера-Дугласа-Пекера), ске-

летизации (метод медиальной оси, алгоритм Чжана-Суэня) и регуляризации через минимизацию энергетических функционалов. Для анализа текстур применяются статистики Харалика и матрицы смежности, позволяющие выделять однородные зоны по спектральным и пространственным признакам.

Основу современных исследований составляют работы в области семантической сегментации, где фундаментальный вклад внесли Дж. Лонг с разработкой полностью сверточных сетей (FCN), О. Роннебергер с архитектурой U-Net, специально оптимизированной для биомедицинских и спутниковых изображений, а также Л. Чэнь с серией моделей DeepLab. Параллельно развивается направление детектирования объектов на снимках, где значимыми являются работы Р. Гиршика по R-CNN, Дж. Редмона по YOLO и В. Лю по SSD-архитектурам. В области прямого прогнозирования векторных данных перспективные разработки связаны с применением графовых нейронных сетей и методов обучения с подкреплением, где можно отметить исследование Я. Лина по распознаванию инженерных чертежей. Для решения специализированных задач геоинформатики сохраняют актуальность классические алгоритмы машинного обучения, такие как случайный лес и метод k -ближайших соседей, теоретические основы которых заложены в работах А.С. Потапова по искусственному интеллекту и Д. Черниховского по применению k -NN для оценки лесных массивов. Данные алгоритмы позволили достигнуть высокого качества решения задач.

Целью данной работы является разработка методов по повышению качества дешифрирования и векторизации аэрокосмических изображений земной поверхности на основе машинного обучения в области разработки интеллектуальных методов анализа и обновления геопространственной информации.

Для достижения поставленной цели необходимо было решить следующие **задачи**:

1. Провести анализ существующих современных методов сегментации и векторизации аэро- и космофотоснимков.

2. Разработать метод точного воссоздания полигональных границ зданий для создания векторизованной маски для использования в различных приложениях.
3. Разработать методику обучения нейросетевых алгоритмов векторизации аэро- и космоснимков на основе технологии обучения с подкреплением.
4. Разработать метод сегментации аэрофотоснимков с применением теории графов для использования существующих связей между объектами и областями на снимках.
5. Собрать выборку, включающую в себя аэро- и космоснимки из открытых источников, а также полученных в ходе натурных экспериментов, с целью проведения вычислительных испытаний для разработанных методов.

Научная новизна:

1. Впервые разработан метод сегментации и векторизации спутникового снимка для построения иерархического графа семантической разметки изображения, учитывающего не только семантическую роль объектов на изображении, но и связи и взаимоотношения между ними.
2. Впервые разработан алгоритм сегментации и векторизации снимка, имитирующий логику и действия оператора-разметчика и развивающий методы обучения с подкреплением.
3. Создана обширная обучающая выборка, содержащая не только пары аэро- и космофотоснимков и разметки к ним, но и обширную информацию о хронологии и этапах разметки, выполняемых оператором-дешифровщиком, включая масштабы изображений, выбранные для разметки заданного объекта, порядок обхода объектов и другие виды метаданных. На разработанной выборке проведены вычислительные эксперименты для реализованных методов.

Научная и практическая значимость решения обозначенной проблемы состоит в создании методов анализа данных космической и аэрофотосъёмки, позволяющих автоматически и с высоким качеством решать задачи семантической

сегментации снимка, выявления изменений в наблюдаемой сцене и получения производной (векторной) информации, необходимой для оперативного обновления геоданных. Разработанные методы позволят, с одной стороны, более эффективно решать задачи обновления геопространственной информации, такие как выявление изменений в ландшафте и инфраструктуре, корректировка и пополнение базы геоданных, а с другой стороны, станут научно-методической основой для решения более сложных задач анализа геопространственной информации, таких как: создание иерархических моделей геопространственной инфраструктуры, интеллектуальное автоматическое оперативное создание карт, автоматическое интеллектуальное планирование сельскохозяйственных работ, автоматическое интеллектуальное городское планирование, автоматический прогноз потенциальных природных и техногенных происшествий.

Методология и методы исследования. В работе использовались методы цифровой обработки изображений, анализа данных и машинного обучения. Для определения работоспособности реализованных методов и сравнения их с ведущими современными подходами использовались открытые выборки, являющиеся стандартом для тестирования алгоритмов сегментации и векторизации, содержащие реальные данные. Разработка программного кода производилась на языке программирования Python, с использованием фреймворка для глубокого обучения PyTorch, а так же библиотеки OpenCV, scikit-learn и Pillow.

Основные положения, выносимые на защиту:

1. Разработанный метод многоэтапной регуляризации обеспечивает повышение точности извлечения контуров зданий на изображениях дистанционного зондирования Земли по сравнению с современными ведущими подходами на 5-10% по метрикам средней точности и полноты и около 1% по коэффициенту Жаккара.
2. Разработанный метод векторизации аэрофотоснимков на основе обучения с подкреплением обрабатывает изображения быстрее (на 60%) и с меньшими использованиями памяти вычислительных средств (на 21%),

чем многоуровневые архитектуры, при незначительном снижении качества, что допустимо для задач векторизации в реальном времени.

3. Разработанная двухэтапная архитектура на основе нейронных сетей и теории графов демонстрирует высокий уровень эффективности при обработке аэрофотоснимков сложной структуры и обеспечивает увеличение точности семантической сегментации за счет комбинации методов глубокого обучения и графовых подходов. По результатам эксперимента прирост коэффициента Жаккара 5-10% по сравнению с современными автоматическими методами.

Достоверность полученных результатов подтверждается проведенными экспериментами для используемых методов, однородностью разметки аэро- и космofотоснимков, составляемой для одного снимка различными независимыми профессиональными дешифровщиками, а также формализованностью структуры векторной карты местности и процесса разметки, позволяющей создать единообразное математическое представление в виде графа сцены.

Апробация работы. Основные результаты по теме диссертации изложены в 6 печатных изданиях, индексируемых Web of Science и Scopus, 1 из которых изданы в журналах списка K1 и входят в собственный перечень МФТИ. Основные результаты работы докладывались на следующих научных конференциях:

1. ISPRS TC I Mid-term Symposium on Intelligent Sensing and Remote Sensing Application, Changsha, 2024;
2. ISPRS TC III Mid-term Symposium “Beyond the canopy: technologies and applications of remote sensing”, Belém, 2024;
3. International Workshop on “Photogrammetric Data Analysis” – PDA24, Moscow, 2024;
4. International workshop on Photogrammetric techniques for environmental and infraStructure monitoring, Biometry and Biomedicine – PSBB25, Moscow, 2025.

Также результаты диссертационного исследования использованы при реализации проекта российского научного фонда №24-21-00269 ”Дешифрование и

векторизация аэро- и космофотоснимков методами машинного обучения для обновления геопространственной информации”, научно-исследовательских работ, таких как проект ”САПФИР” для разработки программного обеспечения, использующего высокопроизводительные вычислительные средства, выполняемый в рамках субсидии Минпромторга России, и проект ”СИНТЕЗ” для разработки технологии повышения реалистичности тепловизионных изображений, формируемых на основе высокодетализированных трехмерных моделей местности. В первом случае для семантической сегментации использовался разработанный метод на основе визуального трансформера и графовой нейронной сети, во втором случае - разработанные методы векторизации для улучшения качества построения трехмерных моделей местности.

Личный вклад. Все основные результаты получены автором лично. Также автор участвовал в постановке исследовательских задач, выборе методов их решения, анализе экспериментальных данных, интерпретации и обсуждении результатов, и формулировке выводов работы. Во всех выносимых на защиту результатах личный вклад диссертанта является основным и определяющим.

Объем и структура работы. Диссертация состоит из введения, 4 глав, заключения и библиографии. Полный объем диссертации составляет 98 страниц, включая 26 рисунков и 3 таблицы. Список литературы содержит 130 наименований.

Глава 1. Анализ существующих современных методов сегментации и векторизации аэро- и космофотоснимков

1.1 Семантическая сегментация спутниковых изображений

Методы семантической сегментации изображений разрабатываются научным сообществом со времени появления первых цифровых изображений, и первоначально развивались в трех направлениях решения проблемы сегментации: 1) анализ локального вида семантических классов [1—3], 2) анализ локальной согласованности признаков между областями изображения [4—6] и 3) использование априорной информации в алгоритмах обработки изображения для улучшения качества сегментации [7—9].

Локальное представление моделируется на различных уровнях, таких как пиксели [1], патчи (части изображения) [2], положение точки интереса [10] или суперпиксели [11], что позволяет ускорить обработку и повысить точность. Для этого используются локальные признаки, которые кластеризуются в визуальные слова или применяются для построения высокоуровневых представлений. Тематические модели, например, латентное распределение Дирихле [3], дополняют эти методы, учитывая пространственную согласованность и комбинируя их с вероятностными подходами.

Для обеспечения локальной и глобальной согласованности широко применяются модели случайных полей (марковского (MRF) [4] и условного (CRF) [5]), где зависимости между метками определяются через минимизацию энергии, а априорная информация задается с помощью потенциалов, таких как модель Поттса [12]. В CRF также используются потенциалы высшего порядка для согласованности на уровне областей, а методы вроде плотного CRF [6] или иерархических моделей повышают точность и эффективность. В качестве аль-

тернативы, изображения разбиваются на суперпиксели, что упрощает обработку, но требует дополнительных решений для работы со сложными регионами.

Априорные знания, такие как глобальная классификация изображений [7], данные о форме объектов [8], ограничивающие рамки [9] и пользовательские теги [13], помогают улучшить локальную оценку и направлять процесс сегментации. Дополнительно используются априорные распределения классов и моделирование пространственных отношений между объектами для повышения точности и качества результатов.

Задача сегментации аэро- и космофотоснимков появилась почти одновременно с зарождением компьютерного зрения. Алгоритмы сегментации снимков преодолели долгий путь от аналитических методов до методов на основе марковских полей, прежде чем развитие вычислительных ресурсов позволило решать задачу сегментации с помощью глубоких свёрточных нейронных сетей. Одной из первых успешных работ в области сегментации изображений стала нейросеть SegNet [14], использующая архитектуру типа энкодер-декодер. Отличительной особенностью данной сети являлось то, что она состояла целиком из свёрточных слоёв (fully convolutional neural network). Это позволило повысить разрешение выходной сегментации и, как следствие, повысить качество формируемой сегментации.

На начальном этапе развития нейросетевых методов выделения специализированных методов сегментации для работы непосредственно с аэро- и космофотоснимками не проводилось. К числу универсальных нейросетей для сегментации, показывавших высокие результаты, как в части высокого качества формируемой разметки, так и в части производительности можно отнести нейросети UPerNet [15] и HRNetV2 [16]. Отличительной особенностью сети UPerNet [15] стала методика совместного обучения на несколько задач (сегментацию объектов, текстур, категорий). Подобное совместное обучение позволило повысить качество сегментации и повысить устойчивость работы сети на новых данных.

Авторы нейросети HRNetV2 [16] предложили архитектуру со сложными перекрёстными связями, которые анализирует изображение сразу в нескольких

масштабах, что позволяет повысить эффективность обнаружения как крупных, так и мелких структур в кадре.

Значительной проблемой при обучении нейросетей на задачу сегментации аэро- и космофотоснимков является дисбаланс площадей классов для мелких объектов (здания, дороги) и крупных объектов (леса, поля). Для решения данной задачи предложена модель представления семантических объектов в виде точек [17] (PointFlow). В сочетании с пирамидальным декодером изображения данный подход позволяет существенно повысить качество распознавания на выборках с большим дисбалансом классов. Альтернативный подход к устранению дисбаланса классов на основе фокусировки внимания связях между объектами и фоном предложен в работе [18]. Также проблемы контекстных связей между объектами и фоном рассмотрены в работе [19], где предложена сеть, использующая многомасштабные перекрёстные связи между слоями. Для расширения обучающих выборок предложен подход, использующий веб-сервисы и веб-карты для обучения нейросетей на задачу сегментации аэрофотоснимков [20].

Многомасштабные конволюционные слои (atrous convolution) позволяют повысить качество сегментации изображений, в том числе для задачи семантической сегментации аэро- и космофотоснимков. В последние годы активно развиваются методы построения разметки аэро- и космофотоснимков в векторном виде [21; 22], которые в основном направлены на работу с двумя основными классами объектов: здания и дороги. Значительное внимание в последнее время уделяется созданию обширных обучающих выборок по тематике семантической сегментации аэро- и космофотоснимков [23; 24].

1.1.1 Свёрточные нейронные сети для семантической сегментации

В последние годы достижения в области глубокого обучения [25—27] существенно повлияли на развитие сетей глубокой семантической сегментации

(DSSN) для обработки изображений дистанционного зондирования (RS) [28]. Эти методы продемонстрировали превосходство над традиционными подходами, такими как случайные леса (RF), деревья решений (DT) и машины опорных векторов (SVM) [29]. Важным этапом в развитии DSSN стало внедрение полностью сверточных сетей (FCN) [30], которые позволили реализовать сквозную семантическую сегментацию за счет интеграции слоев деконволюции в сверточные нейронные сети (CNN).

Одной из ключевых архитектур, получивших широкое распространение, является U-Net [31]. Ее уникальная структура, основанная на схеме кодера-декодера, использует пропускные соединения для эффективного захвата многомасштабной информации. Это позволяет объединять низкоуровневые детали с высокоуровневыми семантическими признаками, что значительно улучшает качество карт признаков. Аналогично, SegNet [14] применяет индекс максимального пула в кодере для выполнения нелинейной передискретизации в декодере, что также способствует повышению точности сегментации.

В последние годы наблюдается активное развитие методов семантической сегментации изображений RS на основе DSSN. Многочисленные исследования демонстрируют применение FCN для решения задач сегментации в дистанционном зондировании [32; 33]. Например, в работе [34] предложена новая DSSN, ориентированная на картографирование городского земельного покрова. В [35] представлен многомасштабный ансамблевый метод остаточного глубокого обучения, вдохновленный архитектурой U-Net, для эффективного извлечения зданий. Кроме того, в [36] предложен модифицированный вариант SegNet, который включает многоядерные сверточные слои для улучшения захвата признаков на различных масштабах.

В работе [37] предложена сеть ранжирования двойного многомасштабного многообразия (DMSMR), направленная на повышение производительности сегментации. Авторы [38] использовали детализированную сеть сегментации для аннотирования объектов на аэрофотоснимках высокого разрешения. Для интеграции мультисенсорных данных, таких как цифровые модели поверхности (DSM),

в DSSN, изображения RGB комбинируются с мультимодальной информацией для улучшения контекстного анализа. В [39] предложен подход на основе сиамской сети для совместной обработки изображений и данных DSM, что позволяет интегрировать обнаружение границ и семантическую сегментацию в единую модель. В [40] представлен метод, включающий информацию о границах в DSSN для повышения точности сегментации. Кроме того, в [41] предложен подход, основанный на U-Net, для быстрой и точной классификации пикселей с последующей постобработкой, использующей экспертные знания.

Важным аспектом улучшения качества сегментации является повышение выразительности сети. Механизмы внимания, аналогичные зрительной системе человека, позволяют выделять наиболее значимые признаки, минимизируя влияние менее важных. В [42] предложен блок сжатия и возбуждения (squeeze-and-excitation), который уточняет ответы признаков по каналам за счет моделирования их взаимозависимостей. Пространственное внимание, в свою очередь, назначает различные веса пикселям в зависимости от их местоположения, что обеспечивает дополнительный пространственный контекст. В [43] представлена модель U-Net с модулем шлюза внимания, который модулирует важность признаков в различных пространственных областях. В [44] предложена семантическая сегментационная сеть SCAttNet, которая использует сверточный блочный модуль внимания (CBAM) [45], объединяющий пространственное и каналное внимание для улучшения сегментации изображений дистанционного зондирования.

Таким образом, современные исследования в области семантической сегментации изображений дистанционного зондирования демонстрируют значительный прогресс, связанный с развитием архитектур глубокого обучения, интеграцией мультимодальных данных и применением механизмов внимания для улучшения качества сегментации.

1.1.2 Графовая сверточная нейронная сеть

Особенности распределения наземных объектов и их взаимосвязей, характеризующихся сложными пространственными отношениями, приводят к формированию графовых структур. В таких графах узлы представляют отдельные объекты, а ребра отражают пространственные связи между ними, такие как соседство, пересечение или разделение. Эти структуры позволяют эффективно моделировать взаимозависимости между объектами, что особенно важно для задач анализа изображений дистанционного зондирования.

Несмотря на значительные успехи глубоких пространственно - структурированных сетей в обработке евклидовых данных, их применение к графовым данным в неевклидовых пространствах остается ограниченным. Графовые сверточные сети (GCN) представляют собой специализированный подход глубокого обучения, адаптированный для работы с графовыми структурами. Они демонстрируют высокую эффективность в извлечении признаков из нерегулярных графов за счет процесса, известного как свертка графа. Основная идея свертки графа заключается в агрегировании информации от соседних узлов с использованием связей между ребрами, что позволяет создавать новые представления узлов. Это делает GCN мощным инструментом для моделирования сложных взаимозависимостей в графах.

Развитие методов анализа графов [46] привело к появлению ряда инновационных подходов. В работе [47] предложена модель многоуровневого распространения графа, использующая спектральную свертку для эффективной обработки графовых данных. В то же время, авторы [48] применили пространственную свертку графа, собирая информацию непосредственно от соседних узлов. В [49] представлена Deer-GCN, которая интегрирует концепции из CNN, такие как остаточные связи и расширенные свертки, для преодоления проблемы исчезновения градиентов, характерной для неглубоких архитектур GCN.

Особое внимание в исследованиях уделяется методам агрегации информации о соседних узлах. Графовая сеть внимания (GAT) [50] позволяет назначать веса каждому соседнему узлу на основе его значимости относительно центрального узла. В отличие от GAT, которая фокусируется на весах сходства, GCN акцентирует внимание на пространственных связях. В [51] предложена пиксельная модель GCN для семантической сегментации, которая использует точное местоположение каждого пикселя и инициализируется с помощью FCN. Однако, несмотря на сохранение локального положения пикселей, этот подход не всегда обеспечивает точное представление пространственных отношений между объектами.

Для задач классификации сцен на многомаркированных аэрофотоснимках в [52] предложен фреймворк CNN-GCN, который сочетает извлечение признаков объектов с учетом их топологических отношений. Хотя абстрактные признаки, генерируемые CNN, полезны для классификации сцен, семантическая сегментация на уровне пикселей требует более детализированного подхода для точной категоризации каждого пикселя. Таким образом, интеграция графовых нейронных сетей с традиционными методами глубокого обучения открывает новые возможности для повышения точности и эффективности анализа изображений дистанционного зондирования.

1.2 Методы извлечения контуров зданий со спутниковых изображений

В течение многих лет исследователи разрабатывают автоматизированный метод, способный заменить человека при создании векторных контуров отдельных зданий, которые играют важную роль в производстве ГИС, мониторинге окружающей среды, городском планировании, оценке плотности населения и энергоснабжении. Безусловно, это чрезвычайно сложная задача, не только из-за трудоемкости разработки такого высокоинтеллектуального алгоритма, но и из-за

проблем, связанных с неидеальными условиями съемки, разнообразной архитектурой зданий и сложностью фона.

Автоматическое обнаружение зданий по аэрофотоснимкам на протяжении десятилетий считается важным средством повышения эффективности генерации векторных карт [53—55]. В последние годы, при поддержке обширных данных для обучения и достаточной вычислительной мощности, методы глубокого обучения, такие как сверточные нейронные сети (CNN) [56] и полностью сверточные сети (FCN) [30], значительно повысили точность обнаружения зданий на изображениях дистанционного зондирования [57—59]. Однако автоматическое создание высококачественных векторных карт зданий по аэрофотоснимкам пока не стало реальностью для большинства сообществ. Отчасти это связано с тем, что подходы к обнаружению зданий на основе глубокого обучения по-прежнему сталкиваются с такими проблемами, как низкая точность распознавания крыш, закрытых деревьями или тенями [60], и относительно слабая способность к обобщению для определенных географических регионов [61]. Одна из современных задач при извлечении контуров зданий заключается в точном воссоздании полигональной границы при одновременном извлечении векторизованной маски строения в качестве выходных данных для непосредственного использования в различных приложениях.

1.2.1 Нейросетевые подходы

Современные методы извлечения зданий в основном используют два доминирующих подхода: семантическую сегментацию [58; 59; 62] и сегментацию экземпляров [63; 64]. Хотя эти подходы работают на уровне пикселей, они сталкиваются с присущими ограничениями: недостаток глобального контекста может приводить к разрывам в предсказаниях контуров зданий, а недостаточное разрешение локальных деталей может вызывать пропуск мелких структур.

Для решения этих проблем исследователи разработали несколько инновационных решений. В [62] представили многомасштабную FCN-архитектуру, которая интегрирует признаки зданий на разных масштабах для повышения точности предсказаний. Авторы [59] предложили комплексный рабочий процесс, сочетающий бинарную семантическую сегментацию, регуляризацию и векторизацию, продемонстрировав эффективность на новом наборе данных зданий с собственной реализацией векторизации. В [65] разработали технику маскирования для сегментации повторяющихся структур на изображениях, что позволило улучшить производительность сегментации на 11%.

В области сегментации экземпляров [63] предложили многоэтапную модель, исследующую влияние качества детекции на сохранность масок. Их подход сочетает детекцию объектов с уточнением сегментации для улучшения точности границ и геометрической правильности контуров зданий. Авторы [64] корректировали маски сегментации, упрощая обнаруженные границы с помощью алгоритма Дугласа-Пьюкера с последующим уточнением полигонов методом минимальной длины описания (Minimum Description Length).

Альтернативное направление исследований рассматривает инстанс-сегментацию как задачу регрессии контуров [66; 67], предсказывая координаты вершин полигонов. Традиционные методы используют активные контурные модели [68; 69], оптимизирующие эмпирически разработанные энергетические функции. Современные подходы повышают robustness за счёт интеграции CNN с активными контурами [70; 71].

Среди современных контурно-ориентированных решений доминируют CNN-архитектуры, такие как PolarMask [72], PolarMask++ [73] и LSNet [74] — эффективные одноэтапные системы, использующие глубинные признаки из центров инстансов, но производящие лишь приблизительные контуры объектов. Некоторые исследования [57; 75] применяли рекуррентные нейронные сети (RNN) [76] для последовательного предсказания углов зданий по часовой стрелке, однако эти методы часто страдают от исчезновения вершин и нерегулярного распределения точек.

Существенным ограничением существующих подходов является необходимость обширной постобработки: семантическая сегментация не различает отдельные здания, а инстанс-сегментация может производить ограничивающие рамки, включающие части соседних структур, что осложняет генерацию масок. Для преодоления этих ограничений в [77] предложили PolyWorld — нейронную сеть, непосредственно предсказывающую вершины зданий из изображений и строящую точные полигоны путём установления связей между ними. Фреймворк использует графовую нейронную сеть для оценки вероятностей связей между парами вершин, с окончательным назначением через дифференцируемую задачу оптимального транспорта, а также уточняет позиции вершин совместной оптимизацией точности сегментации и согласованности углов полигонов.

Отдельного внимания заслуживает PolygonCNN от [58], который сначала выполняет сегментацию для извлечения начальных контуров зданий, а затем использует модифицированный PointNet для изучения априорной формы и предсказания вершин полигонов, кодируя вершины полигонов зданий и объединяя признаки изображения, извлеченные на этапе сегментации.

1.2.2 Наборы данных со спутниковыми снимками

С момента публикации набора данных Zeebrugge [78] в рамках Конкурса по слиянию данных IEEE GRSS 2015 года было выпущено множество наборов данных для обнаружения и сегментации зданий. Следует отметить, что наборы данных, используемые для оценки традиционных методов, обычно имеют небольшой размер, а обучающие и тестовые наборы собираются из одного локального региона (или изображения), что приводит к слабой способности к обобщению. В эпоху глубокого обучения более продвинутые наборы данных позволяют достичь пространственной независимости обучающих и тестовых наборов, более

широкого пространственного охвата и большего объема данных, что соответствует реальным условиям.

Учитывая размер зданий ($> 10^2$), отмечаются некоторые эталонные наборы спутниковых данных/аэрофотоснимков [79—83], большинство из которых имеют пространственное разрешение от сантиметрового уровня до 2 м, за исключением относительно грубого разрешения SpaceNet 7 (4 м). Помимо общеиспользуемых каналов RGB, некоторые наборы данных также предоставляют дополнительную полезную информацию для дальнейшего анализа зданий. С точки зрения спектральной информации, наборы данных Potsdam и WHU-Satellite имеют каналы RGB/ближнего инфракрасного диапазона (NIR), а наборы данных SpaceNet и SpaceNet 4 состоят из восьми спектральных каналов с датчиков WorldView 2/3. Что касается вертикальной информации, наборы данных Potsdam, Vaihingen, Zeebrugge и DFC19-JAX предоставляют несимметричные цифровые модели рельефа, полученные с помощью бортового лазерного сканирования (LIDAR), в то время как набор данных SpaceNet 4 состоит из 27 уникальных изображений, для которых углы обзора варьируются от -32.5^0 до 54.0^0 [84]. Некоторые наборы данных (например, DFC19-JAX) также пытались улучшить сети глубокого обучения путем объединения планарных и стерео наблюдений дистанционного зондирования. С точки зрения временных свойств, наборы данных WHU Building Change Detection, SECOND, Hi-UCD и ZKXT_2021 содержат многовременные наблюдения дистанционного зондирования, контуры зданий для каждой даты и записи об изменениях зданий.

1.3 Обучение с подкреплением

Обучение с подкреплением (Reinforcement Learning, RL) стало мощным инструментом для решения сложных задач принятия решений и оптимизации, демонстрируя впечатляющие достижения в различных областях — от робото-

техники до стратегических игр [85]. Фундаментальная работа Саттона и Барто заложила теоретическую основу методов RL, введя марковские процессы принятия решений (Markov Decision Processes, MDPs) и алгоритмы обучения по временным разностям (temporal difference, TD) [86]. Прорывные разработки DeepMind, включая DQN для игр Atari [87] и AlphaGo для игры Го [88], продемонстрировали потенциал комбинации глубоких нейронных сетей с методами обучения с подкреплением. Особого внимания заслуживают современные вариации этих подходов, такие как алгоритмы актор-критик, включая PPO [89] и SAC [90], которые демонстрируют улучшенную стабильность обучения и эффективность в непрерывных пространствах действий. В приложениях компьютерного зрения RL нашел успешное применение в задачах сегментации и детектирования объектов, где его способность к последовательному принятию решений позволяет эффективно учитывать пространственные зависимости и глобальный контекст сцены [91].

Недавние достижения в RL для визуальных задач представили инновационные архитектуры, которые преодолевают разрыв между восприятием и принятием решений. RL-агенты на основе трансформеров [92] и политики на основе графовых нейронных сетей [93] теперь обеспечивают лучшее управление дальнодействующими зависимостями в данных изображений, в то время как мета-RL подходы [94] решают проблему адаптации к новым окружениям с ограниченным количеством образцов. Эти разработки особенно актуальны для геопространственных приложений, где нестационарный характер аэрофотоснимков требует надежных возможностей адаптации. Появление самоконтролируемых RL-парадигм [95] дополнительно снижает зависимость от дорогостоящих ручных аннотаций, открывая новые пути для приложений дистанционного зондирования в крупном масштабе.

Методы обучения с подкреплением предназначены для обучения агентов (алгоритмов принятия решений) в тех случаях, когда среда, в которой оперирует агент не может быть представлена в виде дифференцируемой математической

модели. Данные методы позволяют обучать алгоритмы принятия решений без явного вычисления производной на основе функции потерь.

Одной из успешных работ в области обучения с подкреплением является модель на основе связки вариационного автоэнкодера, рекуррентной нейросети с долгой краткосрочной памятью (World Models) [96]. Авторы предложили создать нейросетевую дифференцируемую модель среды, в которой оперирует агент, на основе вариационного автоэнкодера и рекуррентной нейронной сети. Авторы показали, что обучение агента в связке с нейросетевой моделью позволяет достичь того же качества управления в значительно более краткий срок.

Интересной задачей является обратное обучение с подкреплением, в котором функция потерь агента задана неявно и подбирается по ходу обучения алгоритма. В нейросетевом алгоритме GAIL [97] был предложен подход, имитирующий работу генеративно-состязательной сети. При обучении сеть дискриминатора оценивает правдоподобность действий агента как сходство с действиями реальных экспертов, записанными в обучающей выборке.

Распространенной проблемой в обучении с подкреплением является определение «функции потерь» сети. Для решения этой проблемы в алгоритме V-MAIL [98] авторы предлагают сети обучаться не только с помощью взаимодействия с окружающей средой, но и на базе ограниченного количества визуальных примеров желаемого результата. Этот подход позволил добиться большей стабильности и сходимости вариационного обучения, а также обеспечил более высокую производительность.

Однако применение глубокого обучения с подкреплением для обработки изображений остаётся ограниченным.

Интересный подход PixelRL [99] предполагает наличие у каждого пикселя своего агента, который при обучении изменяет его значения, выполняя определённые действия. Благодаря этому способу («попиксельное вознаграждение») становится возможным учёт не только будущих состояний данного пикселя, но и соседних с ним. Также данный подход позволяет визуализировать операцию,

используемую для каждого пикселя на каждой итерации, что даёт возможность повысить понимание работы глубоких сетей.

Часто некоторые части изображения несут более подробную и важную информацию, чем остальные (например, при обнаружении объектов область, на которой располагается сам объект, является важнее фона). Разработанный метод [100] позволяет выполнять иерархическое обнаружение объектов на изображении. Идея заключается в обучении агента, который получает на вход изображение и решает, на какой области сосредоточить внимание и увеличить её масштаб. Данная процедура повторяется к уже успешно обработанной области, обеспечивая иерархичный анализ изображения и повышая точность работы алгоритма.

Нейросетевые алгоритмы для семантической сегментации изображений имеют два существенных недостатка. Во-первых, получение точной попиксельной разметки - трудоёмкий процесс. Во-вторых, выборки изображений сильно несбалансированны: некоторые категории объектов более многочисленны, чем другие, из-за чего при обучении алгоритмов распознавания происходит переобучение нейронной сети. Метод обучения с подкреплением для сегментации изображений [101] использует идею иерархического анализа изображения, определяя наиболее важную область изображения из пяти представленных областей. Данный подход позволяет улучшить точность обнаружения объектов, а его вычислительные затраты являются незначительными благодаря увеличению просматриваемых частей изображения.

В области векторизации аэрофотоснимков современные методы можно разделить на три эволюционных этапа. Традиционные подходы компьютерного зрения (например, детектирование краев Канни [102] или водораздельная сегментация) страдают от низкой точности и требуют обширной ручной постобработки. Появление глубокого обучения принесло качественный скачок вперед — такие архитектуры, как U-Net и DeepLab, позволили автоматизировать извлечение объектов с точным сохранением границ [31; 103]. Однако эти методы требуют больших размеченных наборов данных и часто показывают недостаточную устойчивость к изменяющимся условиям съемки. Современные исследования

сосредоточены на гибридных подходах, сочетающих преимущества CNN и трансформеров, вместе с методами, сохраняющими топологическую согласованность [30; 104]. Особенно ценными являются исследования по слабо контролируемому обучению и активному уточнению модели, которые значительно сокращают потребности в аннотациях [105]. Перспективное направление предполагает интеграцию семантической информации с геометрическими ограничениями, что особенно важно для картографических приложений, где важны как точность границ, так и структурные взаимосвязи между объектами [106].

Интеграция RL с современными методами векторизации открывает новые возможности для разработки адаптивных систем, способных обрабатывать сложные пространственные зависимости и динамически оптимизировать рабочие процессы обработки изображений. В отличие от традиционных методов статической обработки, подходы на основе RL формулируют векторизацию как процесс последовательного принятия решений, где каждое действие учитывает как локальные особенности изображения, так и глобальный контекст. Это оказывается особенно ценным для аэрофотоснимков, где объекты часто обладают сложными структурами и пространственными отношениями. Недавние исследования демонстрируют, что комбинация глубокого обучения с подкреплением и компьютерного зрения [107] может преодолеть многие ограничения существующих методов, предлагая более гибкие и адаптивные решения для задач автоматической векторизации.

1.4 Выводы

В данной главе произведен обзор литературы, посвященной методам сегментации изображений земной поверхности, извлечения границ зданий, векторизации аэро- и космofотоснимков, а также использование обучения с подкреплением и теории графов для решения этих задач. Выделены основ-

ные подходы к решению данных проблем, произведен анализ их достоинств и недостатков.

Анализ современных методов обработки аэрофотоснимков показывает эволюцию от традиционных подходов компьютерного зрения к глубоким нейросетевым архитектурам. Ключевой тенденцией является переход от семантической сегментации к комплексным решениям, сочетающим CNN, трансформеры и графовые нейронные сети. Особое внимание уделяется решению проблемы дисбаланса классов и улучшению качества векторизации через регуляризацию границ.

Одной из перспективных направлений признана интеграция обучения с подкреплением (RL) с методами компьютерного зрения, что позволяет формулировать задачу векторизации как процесс последовательного принятия решений. Это обеспечивает адаптивность к сложным пространственным зависимостям и улучшает обработку объектов со сложной структурой. Современные исследования демонстрируют эффективность гибридных подходов, сочетающих преимущества различных архитектур для достижения высокой точности при работе с реальными данными дистанционного зондирования.

Глава 2. Усовершенствование извлечения контуров зданий с помощью методов регуляризации

В данной главе представлены 2 последовательных алгоритма, каждый из которых автоматически извлекает контуры зданий посредством комбинации методов бинарной сегментации, регуляризации и векторизации. Однако, начнём с математической постановки задач сегментации и векторизации аэро- и космостримков.

2.1 Математическая постановка задач

2.1.1 Задача семантической сегментации

Аэрофотоснимок представляется как математическая функция, отображающая координаты пикселей в цветовое пространство. Это позволяет применять аппарат функционального анализа для обработки изображений.

1. Входное изображение: $I : \Omega \rightarrow \mathbb{R}^3$

- $\Omega \subset \mathbb{Z}^2$ - дискретная прямоугольная сетка пикселей размером $H \times W$, где каждый элемент (i, j) представляет координаты пикселя
- \mathbb{R}^3 - трёхмерное цветовое пространство (RGB), где каждому пикселю сопоставлен вектор интенсивностей красного, зеленого и синего каналов

2. Множество семантических классов: $\mathcal{C} = \{c_1, c_2, \dots, c_K\}$

- c_1 = "здания", c_2 = "дороги", c_3 = "растительность", c_4 = "водоёмы", c_5 = "прочее"

- Каждый класс представляет категорию объектов, которые необходимо выделить на изображении

Задача сводится к нахождению оптимального отображения, которое каждому пикселю ставит в соответствие наиболее вероятный семантический класс.

1. Требуется найти функцию отображения:

$$F_{seg} : I \rightarrow S \quad (2.1)$$

где $S : \Omega \rightarrow \mathcal{C}$ - семантическая разметка, функция, присваивающая каждому пикселю $x \in \Omega$ метку класса из множества \mathcal{C} , F_{seg} - искомая функция, обычно реализуемая с помощью глубокой нейронной сети.

2. Минимизируемый функционал:

$$\mathcal{L}_{seg} = - \sum_{x \in \Omega} \sum_{k=1}^K \mathbb{I}[S_{gt}(x) = c_k] \log P(S(x) = c_k | I) + \lambda R(S) \quad (2.2)$$

где:

- Первый компонент $-\sum_{x \in \Omega} \sum_{k=1}^K \mathbb{I}[S_{gt}(x) = c_k] \log P(S(x) = c_k | I)$: $\mathbb{I}[S_{gt}(x) = c_k]$ - индикаторная функция, равная 1 если истинная метка пикселя x равна c_k , и 0 иначе, $P(S(x) = c_k | I)$ - вероятность того, что пиксель x принадлежит классу c_k при заданном изображении I . Логарифмическая функция усиливает штраф за уверенные, но неправильные предсказания, а суммирование по всем пикселям и классам обеспечивает глобальную оптимизацию.
- Второй компонент $\lambda R(S)$: $R(S)$ - функция регуляризации, штрафующая "нереалистичные" конфигурации разметки, а λ - гиперпараметр, балансирующий влияние регуляризации. Наиболее типичные формы регуляризации: сглаживание границ (штраф за резкие изменения меток в однородных областях), согласованность размеров (предотвращение появления слишком мелких или слишком крупных областей), геометрические ограничения (учет ожидаемой формы объектов).

2.1.2 Задача векторизации

Векторизация преобразует растровое представление объектов (бинарные маски) в векторное представление (полигоны), которое более удобно для ГИС-приложений и картографирования.

1. Исходные данные:

- M_i - бинарная маска, полученная на этапе сегментации, где $M_i(x) = 1$ если пиксель x принадлежит зданию i , и 0 иначе
- ∂M_i - граница маски M_i , множество пикселей, где происходит переход между объектом и фоном

2. Требуется найти отображение:

$$V : M_i \rightarrow P_i = \{v_1, v_2, \dots, v_m\} \quad (2.3)$$

где P_i - полигон, представляющий контур i -го здания, $v_j \in \mathbb{R}^2$ - вершины полигона в координатах изображения, а также вершины упорядочены (обычно против часовой стрелки) и образуют замкнутую ломаную линию.

Для определения качества векторизации необходимо минимизировать функционал:

$$\mathcal{L}_{vec} = \mathcal{L}_{boundary}(P_i, \partial M_i) + \mathcal{L}_{regularity}(P_i) \quad (2.4)$$

где:

1. Первый компонент (соответствие границам) $\mathcal{L}_{boundary}(P_i, \partial M_i)$ измеряет, насколько хорошо полигон P_i аппроксимирует исходную границу ∂M_i . Обычно определяется как среднее расстояние между точками границы маски и ближайшими точками полигона:

$$\mathcal{L}_{boundary} = \frac{1}{|\partial M_i|} \sum_{x \in \partial M_i} \min_{y \in P_i} \|x - y\| + \frac{1}{|P_i|} \sum_{y \in P_i} \min_{x \in \partial M_i} \|y - x\| \quad (2.5)$$

где первое слагаемое гарантирует, что все точки границы маски хорошо аппроксимированы полигоном, а второе слагаемое предотвращает появление ”лишних” вершин в полигоне.

2. Второй компонент (геометрическая регулярность) $\mathcal{L}_{regularity}(P_i)$ учитывает архитектурные особенности зданий (прямые углы, параллельность сторон). Наиболее типичные составляющие:

– Регуляризация углов: штраф за отклонение углов от 90°

$$\sum_{j=1}^m (1 - \cos(4(\theta_j - \frac{\pi}{2}))) \quad (2.6)$$

где θ_j - внутренний угол при вершине v_j

- Регуляризация длин: поощрение ребер примерно равной длины
- Регуляризация кривизны: штраф за излишне ”зубчатые” контуры
- Регуляризация площади: сохранение площади объекта при векторизации

Минимизация этого функционала обеспечивает нахождение полигона, который точно следует исходному контуру здания, но при этом обладает ”реалистичной” геометрией с прямыми углами и гладкими границами, что соответствует реальным архитектурным объектам.

2.2 Извлечение контуров зданий с использованием свойства линейной связности

Первый метод предоставляет комплексный рабочий процесс для извлечения зданий и улучшает предсказанную область зданий посредством регуляризации границ. Сначала сверточная нейронная сеть используется для обучения модели сегментации экземпляров. Затем свойство линейной связности множества применяется для организации предсказанных контуров зданий и улучшения их геометрии. Заключительный шаг — процесс векторизации, преобразующий регуляризованные маски зданий в полигоны для использования в любых приложениях. Схема алгоритма представлена на Рисунке 2.1.

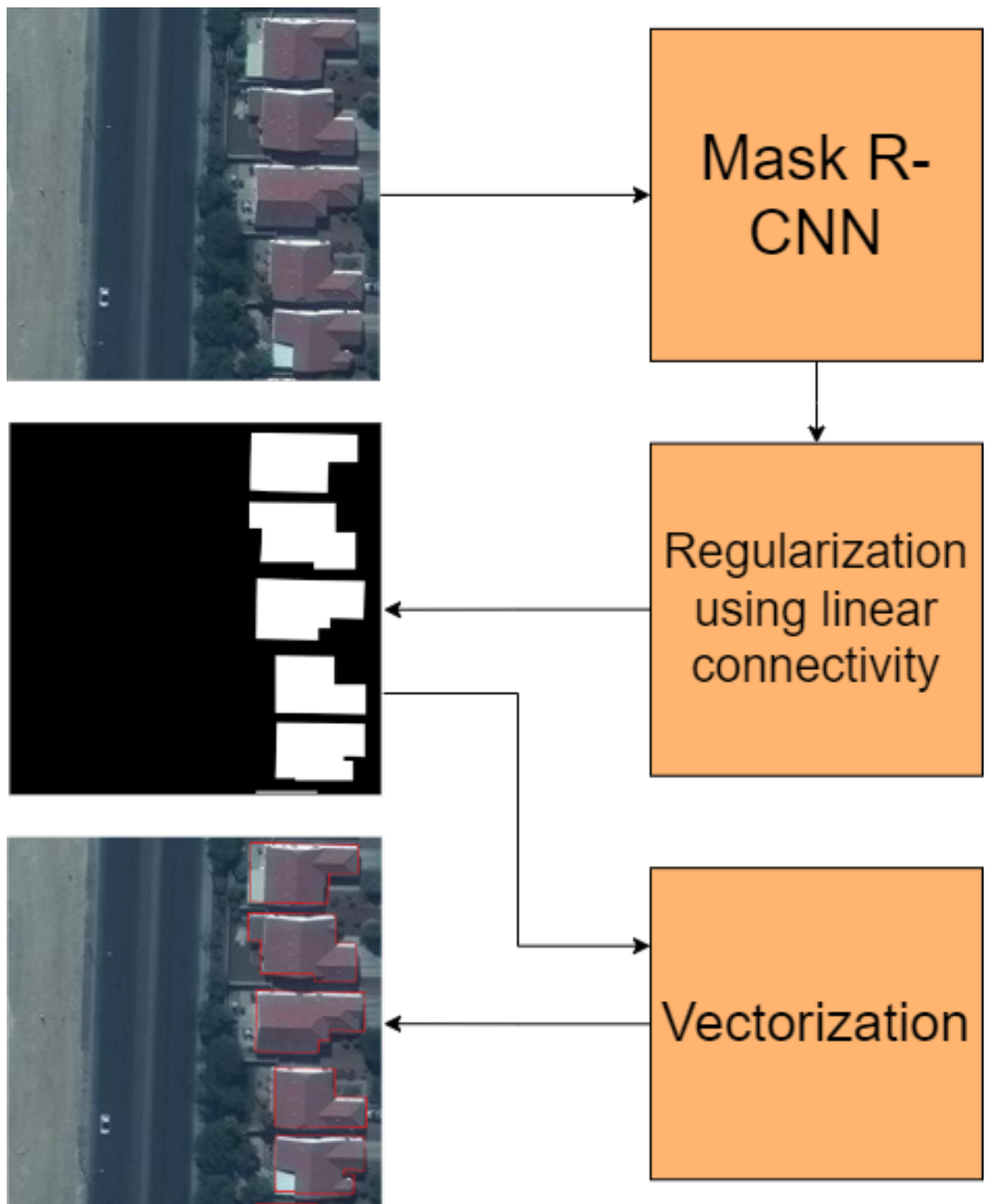


Рисунок 2.1 — Схема предложенного алгоритма.

2.2.1 Сегментация экземпляров с помощью Mask R-CNN

Начальный этап методологии включает выявление и определение границ зданий, представленных на аэрофотоснимках. Для выполнения этой задачи использовалась нейронная сеть Mask R-CNN.

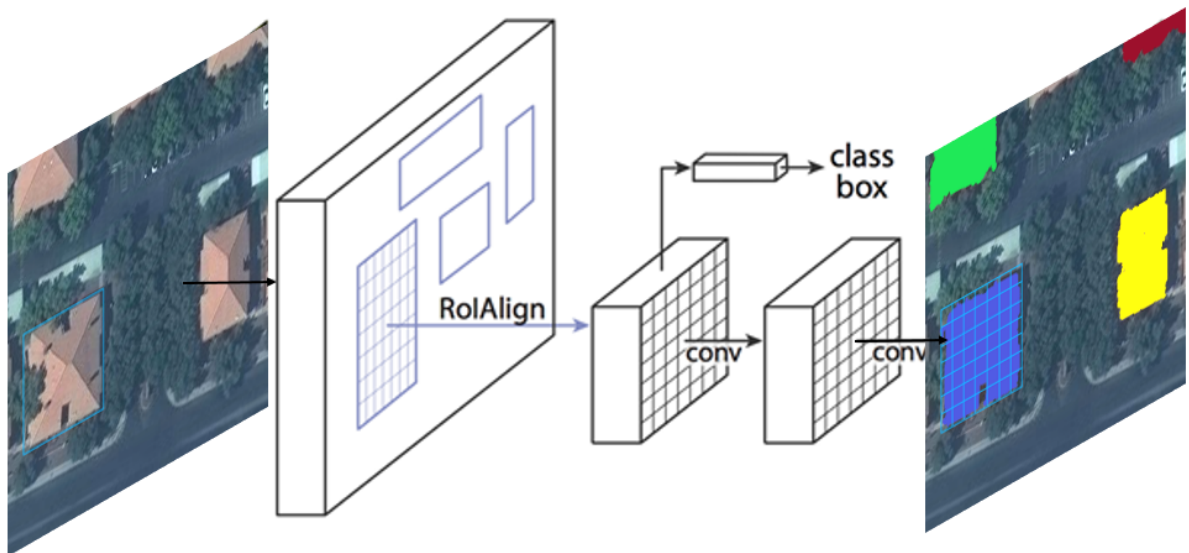


Рисунок 2.2 — Структура Mask R-CNN.

Mask R-CNN [108] расширяет Faster R-CNN [109] путем добавления ветви для предсказания масок сегментации для каждого региона интереса (RoI), параллельно с существующей ветвью для классификации и регрессии ограничивающей рамки. Ветвь маски представляет собой небольшую полностью сверточную сеть (FCN), применяемую к каждому RoI, и предсказывает маску сегментации в пиксель-к-пиксельной манере. Mask R-CNN проста в реализации и обучении благодаря фреймворку Faster R-CNN, что способствует широкому спектру гибких архитектурных решений. Кроме того, ветвь маски добавляет лишь небольшие вычислительные затраты, обеспечивая быструю работу системы и возможность оперативного экспериментирования.

Во время обучения использовался оптимизатор Adam с бинарной кросс-энтропией между логитами в качестве функции потерь для измерения разницы между предсказанным результатом и эталонными данными. Функция потерь

определяется как:

$$L = -\frac{1}{N} \sum_{i=1}^N [x_i \log(\sigma(y_i)) + (1 - x_i) \log(1 - \sigma(y_i))] \quad (2.7)$$

где N — размер батча, x_i — эталонные данные изображения для образца i , y_i — логит-выход модели для образца i , и σ — сигмоидная функция. Сигмоидная функция — это любая математическая функция, график которой имеет характерную S-образную кривую (сигмоида). В качестве сигмоидной функции используется логистическая функция, которая определяется как:

$$\sigma(x) = \frac{1}{1 + e^{-x}} \quad (2.8)$$

2.2.2 Применение регуляризации к предсказаниям

После генерации предсказаний с помощью обученной модели, шаг постобработки применяет регуляризацию для дальнейшего улучшения геометрии и точности предсказанных масок зданий. Поскольку пиксельная классификация приводит к закругленным углам и замкнутым границам предсказаний, регуляризация является важным шагом для дальнейшего улучшения предсказаний. Кроме того, после процесса сегментации предсказанная ограничивающая рамка может содержать дополнительные экземпляры, что затрудняет обучение масковой головы сети.

Учитывая, что идентификация границ зданий проводится на изображениях дистанционного зондирования, мы будем предполагать, что здания на изображениях не пересекаются и не перекрывают друг друга. В этом случае каждое здание представляет собой замкнутое ограниченное множество пикселей. Таким образом, используя свойство линейной связности, из ограничивающих прямоугольников удаляются лишние точки, не принадлежащие основному объекту.

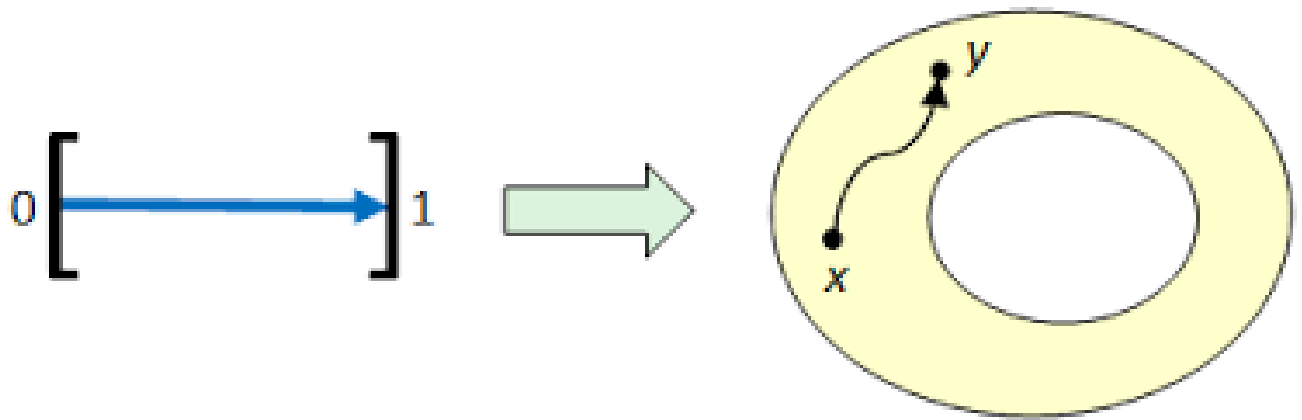


Рисунок 2.3 — Схематическое определение линейно связного множества.

В топологии и смежных разделах математики, связное пространство — это топологическое пространство, которое не может быть представлено как объединение двух или более непересекающихся непустых открытых подмножеств. Линейно связное пространство — это более сильное понятие связности, требующее структуры пути. Путь от точки x к точке y в топологическом пространстве X — это непрерывная функция f из единичного интервала $[0, 1]$ в X с $f(0) = x$ и $f(1) = y$. Компонента линейной связности пространства X — это класс эквивалентности X по отношению эквивалентности, которое считает x эквивалентным y , если существует путь из x в y . Пространство X называется линейно связным, если существует ровно одна компонента линейной связности. Для непустых пространств это эквивалентно утверждению, что существует путь, соединяющий любые две точки в X . Определение линейно связного множества аналогично определению для пространства.

Таким образом, разработанный алгоритм имеет следующую структуру:

1. Детекция объектов класса "здание" на изображении
2. Сегментация объектов класса "здание" в каждой обнаруженной ограничивающей рамке
3. Поиск и удаление точек, не принадлежащих основному зданию в ограничивающей рамке, но сегментированных как "здание"
4. Векторизация полученных изображений



Рисунок 2.4 — Использование свойства линейной связности для удаления ненужных точек с изображений. На первом изображении обнаружен объект класса «здание». На втором — объекты класса «здание» сегментируются в ограничивающем прямоугольнике. На третьем — пиксели, не принадлежащие основному объекту, но ошибочно сегментированные, выделяются красным цветом и удаляются. На последнем изображении представлен конечный результат сегментации и регуляризации.

2.3 Анализ взаимосвязи пикселей для извлечения контуров зданий

Данный метод представляет тщательный процесс извлечения зданий и повышает точность предсказаний площадей зданий за счёт применения техник регуляризации границ. В начале сверточная нейронная сеть используется для обучения модели бинарной сегментации. Матрица окрестности также используется для организации предсказанных контуров зданий и улучшения их геометрии. Заключительный этап включает преобразование регуляризованных масок зданий в полигоны через процесс векторизации, что позволяет использовать их в различных приложениях. Рисунок 2 отображает схему алгоритма.

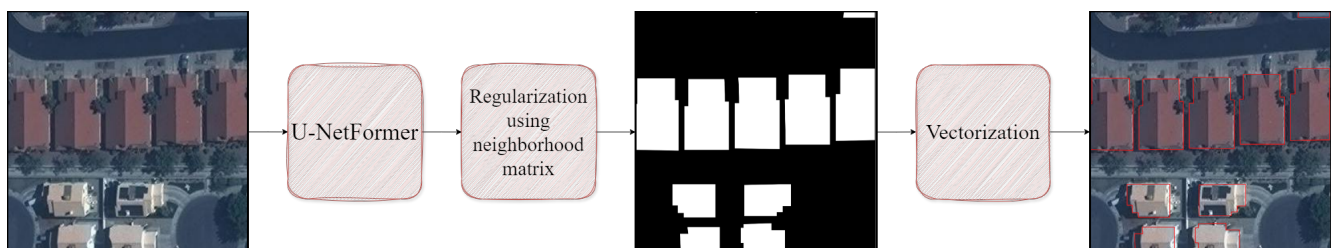


Рисунок 2.5 — Схема предложенного алгоритма.

2.3.1 Бинарная сегментация с U-netFormer

Начальный этап методики включает бинарную сегментацию изображений дистанционного зондирования для идентификации зданий. Для выполнения этой задачи использовалась нейронная сеть UNetFormer.

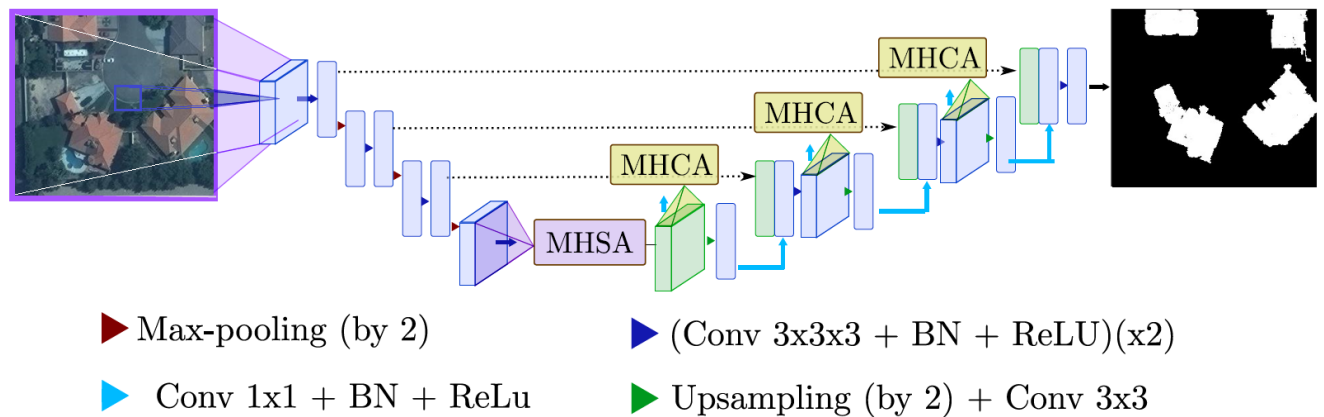


Рисунок 2.6 — Структура сверточной нейронной сети, используемой для бинарной сегментации.

Сеть U-NetFormer (U-Transformer) [110] дополняет U-Nets [31] модулями внимания, построенными на основе многоголовых трансформеров. U-образный трансформер моделирует контекстуальные взаимодействия на больших расстояниях и пространственные зависимости с помощью двух типов модулей внимания (см. Рисунок 3): многоголового самовнимания (MHSA) и многоголового перекрёстного внимания (MHCA). Модули разработаны для выражения нового представления входных данных на основе собственного внимания в первом случае или внимания, уделяемого функциям более высокого уровня, во втором.

Модуль MHSA предназначен для извлечения структурной информации из изображений на больших расстояниях. Основная цель MHSA — соединить каждый элемент карты признаков высшего уровня с каждым другим, обеспечивая таким образом доступ к рецептивному полю, включающему всё входное изображение. Таким образом, решение для одного конкретного пикселя может быть подвержено влиянию любого входного пикселя. Идея модуля MHCA заключает-

ся в отсечении ненужных или зашумленных областей из пропускаемых функций и выделении областей, представляющих значительный интерес для приложения.

Во время обучения использовался оптимизатор Adam с бинарной кросс-энтропией между логитами в качестве функции потерь для измерения разницы между предсказанным результатом и ground truth. Функция потерь определяется как:

$$L = -\frac{1}{N} \sum_{i=1}^N [x_i \log(\sigma(y_i)) + (1 - x_i) \log(1 - \sigma(y_i))] \quad (2.9)$$

где N — размер батча, x_i — эталонные данные изображения для образца i , y_i — логит-выход модели для образца i , и σ — сигмоидная функция. Сигмоидная функция — это любая математическая функция, график которой имеет характерную S-образную кривую (сигмоида). В качестве сигмоидной функции используется логистическая функция, которая определяется как:

$$\sigma(x) = \frac{1}{1 + e^{-x}} \quad (2.10)$$

2.3.2 Применение регуляризации к предсказаниям

После генерации предсказаний с помощью обученных моделей реализуется шаг постобработки для применения регуляризации с целью улучшения геометрии и точности предсказанных масок зданий. Пиксельная классификация часто приводит к закруглённым углам и замкнутым границам предсказаний, поэтому применение регуляризации крайне важно для повышения точности этих предсказаний. Кроме того, после процесса бинарной сегментации класс "здание" может включать дополнительные пиксели, в то время как некоторые ключевые элементы могут быть потенциально опущены.

Поскольку идентификация границ зданий основывается на изображениях дистанционного зондирования, предполагается, что здания на изображениях не пересекаются и не перекрываются. В этой ситуации каждое здание представлено

замкнутым ограниченным множеством пикселей. Используя матрицу окрестности, можно устранить точки, не принадлежащие классу "здание" и также включить любые пиксели, которые классификатор мог пропустить. Также отметим, что из-за специфики изображений дистанционного зондирования большинство зданий на изображениях имеют прямые грани, а углы при вершинах составляют 90^0 , т.е. здания являются композицией прямоугольников.

Все точки на изображении присваиваются значения 1 и 0 в зависимости от их принадлежности к классу "здание" (если точка сегментирована, то значение 1, если нет — 0). Матрица окрестности точки представляет собой матрицу 3×3 , заполненную соответствующими значениями самой точки в центре и её соседей. Аналогично, матрица окрестности 2-го порядка точки представляет собой матрицу 5×5 , заполненную соответствующими значениями точки в центре и её соседей, и т.д.

Обозначим через K сумму всех элементов в матрице окрестности определённого пикселя. K показывает, сколько элементов в окрестности пикселя принадлежат классу "здание" (включая принадлежность рассматриваемой точки). На основе результатов изучения различных вариаций матриц окрестности можно заключить, что наименьшее значение K для точки, принадлежащей зданию, равно 4 (это вершины выпуклых углов здания, у них наименьшее количество соседей из нужного класса — 3), а значения для точек, являющихся границей здания, варьируются от 5 до 8. Также отметим, что для внутренних точек зданий $K = 9$. Примеры матриц окрестности для граничных точек показаны на Рисунке 2.7. Следовательно, все точки, для которых $K \leq 3$, считаются некорректно сегментированными и удаляются из класса "здание".

Также отметим, что на первом этапе алгоритма точки со значением $K = 8$ добавляются в класс "здание" но нейронная сеть не распознала их как представителей нужного класса. Вышеописанные шаги регуляризации представлены в виде алгоритма на Рисунке 2.9.

$$\begin{array}{ccc}
 \begin{pmatrix} 0 & 0 & 0 \\ 0 & 1 & 1 \\ 0 & 1 & 1 \end{pmatrix} & \begin{pmatrix} 1 & 1 & 0 \\ 1 & 1 & 0 \\ 0 & 0 & 0 \end{pmatrix} & \begin{pmatrix} 0 & 0 & 0 \\ 0 & 1 & 0 \\ 1 & 1 & 1 \end{pmatrix} \\
 \text{(a)} & \text{(b)} & \text{(c)} \\
 \begin{pmatrix} 1 & 1 & 1 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \end{pmatrix} & \begin{pmatrix} 1 & 1 & 0 \\ 1 & 1 & 0 \\ 1 & 1 & 0 \end{pmatrix} & \begin{pmatrix} 0 & 0 & 0 \\ 1 & 1 & 0 \\ 1 & 1 & 1 \end{pmatrix} \\
 \text{(d)} & \text{(e)} & \text{(f)}
 \end{array}$$

Рисунок 2.7 — Матрицы окрестности для граничных точек. Матрицы (а-с, f) показывают основные типы вершин выпуклых углов зданий, а (d, e) — типы границ.



Рисунок 2.8 — Результат процесса регуляризации. Слева: входное изображение. В центре: сегментированное изображение. Справа: сегментированное изображение после процесса регуляризации.

Algorithm 1: Regularization using neighborhood matrices

Input:

An image with set of points belonging to the class
"building" $\mathbf{B} = \{x_i\}$,

Output:

An image with set of points belonging to the class
"building" $\mathbf{B} = \{x'_i\}$,

```

1  Search and adding missing points to the set of points  $\mathbf{B}$  ;
2  Procedure Add( $\mathbf{B}$ ):
3      for each point  $x_j$  not in class "building"  $\mathbf{B}$  do
4          Construct a neighborhood matrix  $A_j$  for point  $x_j$ 
5          if  $K = 8$  then
6              add  $x_j$  to  $\mathbf{B}$  ;
7          else
8              skip;
9      return  $\mathbf{B}$ ;
10 Search and removal unnecessary points from the set of points  $\mathbf{B}$  ;
11 Procedure Remove( $\mathbf{B}$ ):
12     while  $|\mathbf{B}_i| < |\mathbf{B}_{i-1}|$  do
13         for each point  $x_i$  of class "building"  $\mathbf{B}$  do
14             Construct neighborhood matrix  $A_i$  for point  $x_i$ 
15             if  $K \leq 3$  then
16                 delete  $x_i$  from  $\mathbf{B}$  ;
17             else
18                 skip;
19         return  $\mathbf{B}_i$ ;
20     return  $\mathbf{B}$ ;
21 Search particularly important points to the set of points  $\mathbf{B}$  ;
22 Procedure Search( $\mathbf{B}$ ):
23     for each point  $x_i$  of class "building"  $\mathbf{B}$  do
24         Construct a neighborhood matrix  $A_i$  for point  $x_i$ 
25         if  $K = 4$  and  $\det(A_i) = 0$  and  $(a_{ij} = a_{ji}$  or
26              $a_{ij} = a_{4-j,4-i})$  then
27             add  $x_i$  to vertices of  $\mathbf{B}$  ;
28         if  $K = 6$  and  $\det(A_i) = 0$  and  $A_i^2 = 2A_i$  then
29             add  $x_i$  to boundary of  $\mathbf{B}$  ;
30         else
31             skip;
32     return  $\mathbf{B}$ ;
  
```

Рисунок 2.9 — Алгоритм регуляризации и выбора особо значимых точек.

2.4 Извлечения контуров зданий с помощью методов многоэтапной регуляризации

Данный метод представляет оптимизированный конвейер для автоматиче-

с помощью инновационных стратегий регуляризации границ. Наша методология реализует многоэтапную цепочку обработки, которая последовательно выполняет: (1) бинарную семантическую сегментацию с использованием глубокой сверточной нейронной сети, (2) комплексное уточнение границ с помощью интегрированных методов регуляризации и (3) геометрическую векторизацию улучшенных выходных данных. Схематичное представление алгоритма, интегрирующая эти методы регуляризации, проиллюстрирована на Рисунке 2.10.

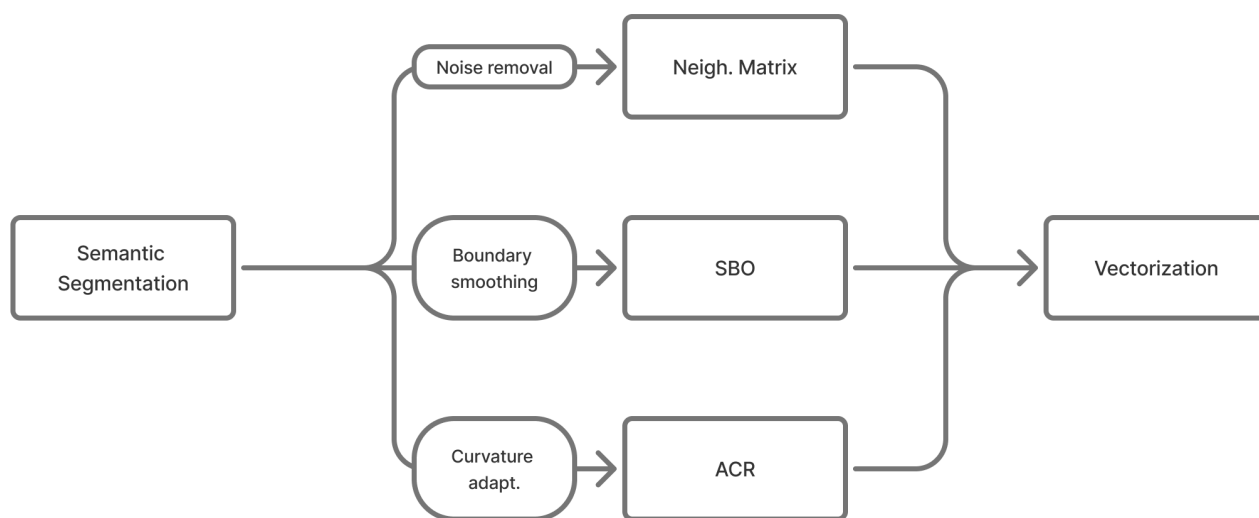


Рисунок 2.10 — Схема предлагаемого алгоритма, объединяющая регуляризацию матрицы соседства, оптимизацию спектральной границы и адаптивное уточнение контура.

2.4.1 Семантическая сегментация с U-NetFormer

Разработанная методология использует сложный подход глубокого обучения для обнаружения контуров зданий из изображений дистанционного зондирования, начиная с бинарной сегментации с использованием улучшенной архитектуры UNetFormer. Эта продвинутая нейронная сеть расширяет возможности традиционных фреймворков U-Net за счёт интеграции мощных механизмов

внимания на основе трансформеров, обеспечивая более комплексный анализ пространственных отношений на множестве масштабов.

Архитектура UNetFormer (U-Transformer) [110] фундаментально улучшает оригинальный дизайн U-Net [31] через свои инновационные модули внимания, которые работают комплементарным образом. Механизм многоголового самовнимания (MHSA) обеспечивает глобальное контекстуальное понимание, устанавливая дальнотянущие зависимости по всей карте признаков, эффективно соединяя каждый пиксель со всеми остальными в изображении. Это создает полностью комплексное рецептивное поле, которое позволяет решениям по сегментации в любом месте включать релевантную информацию из удаленных областей входа.

Работая совместно с MHSA, модуль многоголового перекрестного внимания (MNSA) выполняет интеллектуальный отбор признаков, динамически фильтруя нерелевантную или зашумленную информацию в skip-соединениях, одновременно точно выделяя наиболее значимые области для обнаружения зданий. Эта система двойного внимания генерирует оптимизированные представления признаков через два различных, но комплементарных пути: MHSA обрабатывает внутренние отношения признаков в данных, в то время как MNSA стратегически включает в себя контекстуальную информацию более высокого уровня, чтобы сфокусировать обработку на наиболее диагностически ценных областях изображения.

Модель была оптимизирована с использованием алгоритма Adam с бинарной кросс-энтропией между логитами (BCEWithLogitsLoss) в качестве функции потерь. Эта метрика потерь количественно определяет расхождение между предсказаниями сети и эталонными аннотациями через следующую формулу:

$$L = -\frac{1}{N} \sum_{i=1}^N [x_i \log(\sigma(y_i)) + (1 - x_i) \log(1 - \sigma(y_i))] \quad (2.11)$$

где N представляет размер батча, x_i обозначает бинарную маску эталонных данных для i -го образца, y_i соответствует сырым выходным логитам модели для образца i , σ указывает на сигмоидную функцию активации.

Сигмоидная функция, характеризующаяся своей характерной S-образной кривой, преобразует значения логита в вероятностные выходы. В частности, используется стандартная логистическая функция, определенная как:

$$\sigma(x) = \frac{1}{1 + e^{-x}} \quad (2.12)$$

Эта функция активации отображает неограниченные выходы логитов в диапазон $[0, 1]$, позволяя вероятностную интерпретацию предсказаний модели, сохраняя дифференцируемость для обратного распространения.

2.4.2 Применение разработанного метода регуляризации

1. Начальная регуляризация на основе матрицы окрестности

Первый этап нашего конвейера регуляризации использует подход локальной пространственной согласованности для исправления ошибок классификации на уровне пикселей. Для каждого пикселя $p_{i,j}$, классифицированного как "здание" (метка класса 1), мы строим матрицу окрестности 3×3 $N(p_{i,j})$, охватывающую центральный пиксель и его восемь непосредственных соседей. Эта матрица служит окном локального контекста, позволяя нам:

– Обнаруживать неправильно классифицированные пиксели:

Пиксель помечается как потенциально неправильно классифицированный, если его метка противоречит большинству его соседей (например, изолированный пиксель "здания" окруженный пикселями "не-здания").

- **Применять вероятностную коррекцию:** Конечная метка $p_{i,j}$ переопределяется на основе взвешенного голоса его соседей, где веса обратно пропорциональны их евклидовому расстоянию от $p_{i,j}$. Этот шаг эффективно удаляет salt-and-pepper шум, сохраняя законные структуры малого масштаба.

2. Спектральная оптимизация границ (SBO) для глобальной согласованности

Для обеспечения топологической согласованности между сегментами зданий мы моделируем выход сегментации как ненаправленный граф $G = (V, E)$, где Узлы V соответствуют пикселям, помеченным как "здание". Рёбра E соединяют пространственно смежные пиксели (8-соседство), с весами w_{ij} , определенными гауссовым ядром сходства:

$$w_{ij} = \exp \left(-\frac{|p_i - p_j|_2^2}{2\sigma_d^2} - \frac{|I_i - I_j|_2^2}{2\sigma_I^2} \right), \quad (2.13)$$

где σ_d и σ_I контролируют чувствительность к пространственному расстоянию и вариации интенсивности соответственно.

Лапласиан графа $L = D - W$ (где D — степень матрицы, а W — взвешенная матрица смежности) кодирует глобальную структуру областей зданий. Мы решаем задачу спектральной оптимизации:

$$\min_f (f^T L f + \lambda |f - y|^2), \quad (2.14)$$

где f — регуляризованное поле меток, y — начальная сегментация, и λ — параметр компромисса. Этот шаг устраняет фрагментированные области и сглаживает нерегулярные границы, уважая края изображения.

3. Адаптивное уточнение контура (ACR) для геометрической точности

Финальный этап уточняет границы зданий через явную адаптацию кривизны. Для каждого извлеченного контура границы C :

- **Вычисляем локальную кривизну** κ_i в каждой точке i с помощью производно-ориентированного оценщика:

$$\kappa_i = \frac{|x'_i y''_i - y'_i x''_i|}{(x'^2_i + y'^2_i)^{3/2}}, \quad (2.15)$$

где производные аппроксимируются через центральные конечные разности (например, $x' = \frac{x_{i+1} - x_{i-1}}{2}$).

- **Регулируем силу сглаживания** с помощью кривизно-адаптивного ядра:

$$h_i = \exp\left(-\frac{\kappa_i^2}{2\sigma_c^2}\right), \quad \sigma_c = \text{толерантность кривизны.} \quad (2.16)$$

Области с высокой кривизной (например, углы с $\kappa_i > \sigma_c$) подвергаются слабому сглаживанию, сохраняя резкие черты, в то время как сегменты с низкой кривизной (прямые края) агрессивно регуляризуются.

- **Передискретизируем контур** используя В-сплайн интерполяцию для обеспечения равномерного расстояния между точками, критически важного для высококачественной векторизации. Конечный контур реконструируется как:

$$C_{\text{refined}} = \sum i h_i \cdot K(|C_i - C|) * C, \quad (2.17)$$

где K — гауссово ядро сглаживания.



Рисунок 2.11 — Результат процесса регуляризации. Слева: исходное изображение. В центре: сегментированное изображение. Справа: сегментированное изображение после процесса регуляризации.

2.5 Результаты работы методов

2.5.1 Метрики оценки

В работах используются несколько стандартных метрик для оценки производительности модели.

Intersection-over-Union (IoU), также известный как индекс Жаккара, количественно оценивает точность сегментации, измеряя перекрытие между предсказанными и истинными масками. Он вычисляется как:

$$IoU = \frac{Intersection}{Union} = \frac{TP}{TP + FP + FN} \quad (2.18)$$

Дополнительно мы вычисляем метрики точности и полноты для получения оценок Average Precision (AP) и Average Recall (AR):

$$Precision = \frac{TP}{TP + FP} \quad (2.19)$$

$$Recall = \frac{TP}{TP + FN} \quad (2.20)$$

где TP (True Positives) представляет правильно идентифицированные пиксели зданий, FP (False Positives) указывает на пиксели не-зданий, ошибочно классифицированные как здания, FN (False Negatives) обозначает пропущенные пиксели зданий в предсказании.

2.5.2 Эксперимент

Модели обучались с использованием общедоступного набора данных CrowdAI Mapping Challenge [83], который содержит более 280 000 спутниковых изображений для обучения и дополнительные 60 000 для тестирования. Мы



Рисунок 2.12 — Некоторые изображения из набора данных CrowdAI Mapping Challenge.

использовали разделение 80-20 для обучающих данных, выделяя 80% для обучения модели и оставляя 20% для целей валидации. Все процедуры обучения были реализованы с использованием CUDA 11.7 на GPU NVIDIA GeForce RTX 3070 с 8GB VRAM.

Для оценки производительности в Таблице 1 представлены количественные результаты предложенных методов вместе с сравнительными эталонными показателями современных подходов. Этот сравнительный анализ позволяет оценить относительную производительность алгоритмов по сравнению с существующими решениями на сопоставимых данных.

Метод	AP	AR	IoU
Mask R-CNN	41.9	47.6	-
PolyMapper	55.7	62.1	-
PolyWorld	63.3	75.4	91.3
Линейная связность	65.2	74.9	91.4
Матрица окрестности	64.1	75.1	91.2
Многоэтапная регуляризация	69.8	80.3	92.0

Таблица 1 — Результаты на тестовом наборе данных CrowdAI для всех экспериментов по извлечению и полигонизации зданий.



Рисунок 2.13 — Результаты эксперимента для метода с регуляризацией на основе свойства линейной связности.



Рисунок 2.14 — Результаты эксперимента для метода с регуляризацией на основе матрицы окрестности.

2.6 Выводы

В данной главе экспериментальные результаты демонстрируют, что предложенный сначала метод по извлечению контуров зданий достигает конкурентоспособной производительности с современными подходами, такими как PolyWorld [77], предлагая при этом несколько уникальных преимуществ, а последующий его превосходит. Регуляризация с использованием матрицы окрестности показывает высокую точность сегментации: $AP = 64.1$ и $AR = 75.1$. Кроме того, разработанный метод регуляризации не только улучшает прогнозы, но и геометрическую форму контуров зданий.

Ключевые преимущества второго метода включают в себя: многомасштабный подход к регуляризации, поддержание топологической целостности



Рисунок 2.15 — Результаты эксперимента для метода с многоэтапной регуляризацией.

посредством обработки на основе графов, которая предотвращает чрезмерное сглаживание сложных строительных конструкций, и адаптивную геометрическую обработку, которая интеллектуально сохраняет критические особенности, такие как острые углы, при этом устраняя шум. Эта структура доказала свою эффективность при обработке спутниковых снимков высокого разрешения, успешно устраняя распространённые неточности сегментации (зазубренные края, ложные пиксели), которые обычно снижают качество последующих приложений, включая 3D-моделирование городов и ГИС-анализ. Надежность метода обусловлена сбалансированной интеграцией локальной и глобальной обработки, сочетающей точность операций на уровне пикселей со структурной осведомлённостью графовых и адаптивных к кривизне методов.

Глава 3. Нейросетевой алгоритм на основе обучения с подкреплением для векторизации изображений дистанционного зондирования Земли

В данной главе предлагается новый подход к векторизации аэрофотоснимков, основанный на обучении с подкреплением (RL). Основная цель — разработать адаптивную модель, способную принимать последовательные решения при выделении объектов, тем самым повышая точность векторизации по сравнению с традиционными методами. Ключевое отличие предлагаемого подхода заключается в применении RL к задачам векторизации, что позволяет модели учитывать пространственный контекст и динамически адаптироваться к различным типам объектов. В отличие от методов, основанных на сверточных нейронных сетях, наш подход использует механизм вознаграждения для оптимизации выделения границ, что способствует более точному распознаванию сложных структур.

В разработанном подходе используется комбинация сверточных нейронных сетей (CNN) для извлечения признаков и рекуррентных нейронных сетей (RNN) для учета временных зависимостей (последовательный выбор объектов). В статье также используются следующие алгоритмы обучения с подкреплением (RL): Deep Q-Network (DQN) (для обучения агента на основе Q-функции) и Proximal Policy Optimization (PPO) (для оптимизации политики агента). Функция вознаграждения рассчитывается на основе метрик качества IoU (Intersection over Union), а именно, агент получает положительное вознаграждение за правильный выбор объекта и штраф за ошибки. На рисунке 3.1 представлена схема алгоритма.

3.1 Архитектура системы

Разработанная система представляет собой интегрированную архитектуру нейронной сети, сочетающую современные достижения в области глубокого

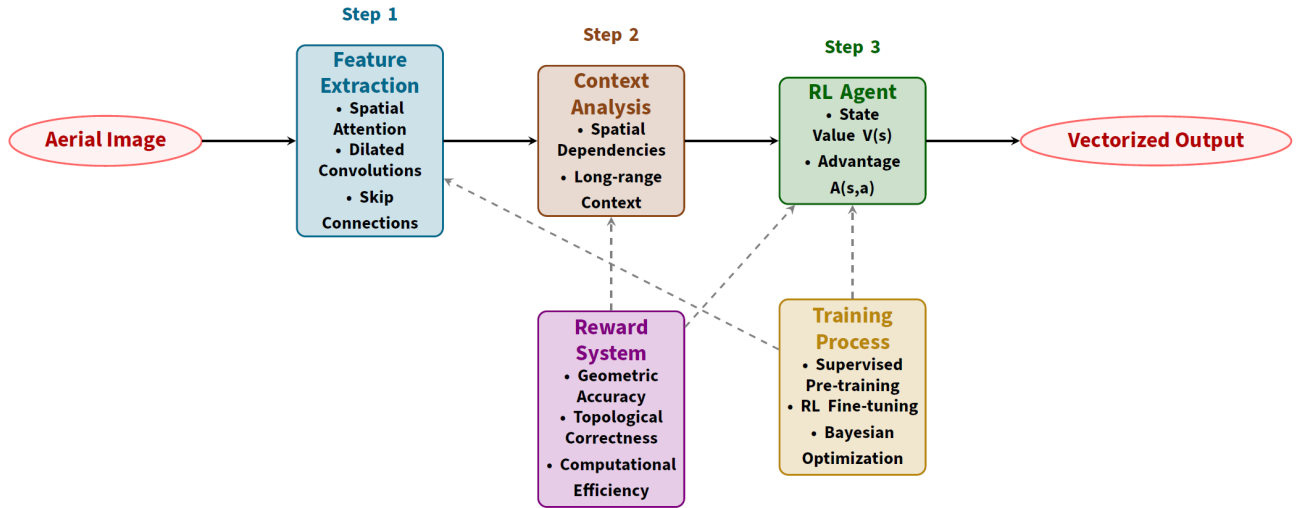


Рисунок 3.1 — Структура предложенного алгоритма.

обучения и обучения с подкреплением. Предлагаемая структура использует гибридную нейронную архитектуру, интегрирующую три ключевых компонента:

- **Модуль извлечения признаков:** Реализован с использованием улучшенного базового блока ResNet-50 [111], модифицированного для обработки аэрофотоснимков высокого разрешения с:
 - Механизмами пространственного внимания между сверточными блоками.
 - Многоуровневым пулингом с пирамидальными свертками (ASPP) для многомасштабного контекста.
 - Skip-connection с механизмами внимания между энкодером и декодером.
- **Модуль последовательного контекста:** Двухнаправленная сеть GRU [112] с 256 скрытыми юнитами с нормализацией слоёв, с дропаутом ($p = 0.2$) для регуляризации, сканирующим обходом с 8 шаблонами связности и эффективной реализацией с использованием групповых сверток.
- **Агент обучения с подкреплением:** Архитектура Dueling Double DQN [113]:

$$Q(s,a) = V(s) + \left(A(s,a) - \frac{1}{|A|} \sum A(s,a') \right) \quad (3.1)$$

где $V(s)$ — поток оценки состояния, $A(s,a)$ — поток функции преимуществ, а целевая сеть обновляется через усреднение Поляка ($\tau = 0.01$).

3.2 Оптимизация обучения с подкреплением

Фреймворк оптимизации обучения с подкреплением использует два взаимодополняющих подхода для обеспечения стабильного и эффективного обучения:

3.2.1 Усовершенствованный алгоритм DQN

Разработанная модифицированная архитектура Deep Q-Network включает три ключевых нововведения:

- **Механизм Double Q-learning:** Расчет целевого значения разделен для предотвращения смещения переоценки:

$$\text{target} = r + \gamma Q_{\text{target}}(s', \argmax_a Q(s', a; \theta); \theta') \quad (3.2)$$

где θ представляет параметры онлайн-сети, θ' обозначает параметры целевой сети, а γ — коэффициент дисконтирования (0.99), определяющий важность будущих вознаграждений.

- **Приоритизированное воспроизведение опыта:** Буфер воспроизведения выбирает переходы в соответствии с:

$$P(i) = |\delta_i| + \varepsilon, \quad \delta_i = r + \gamma \hat{Q}(s', a') - Q(s, a) \quad (3.3)$$

где δ_i — ошибка временной разницы (TD), ε — малая константа ($1e-6$), обеспечивающая ненулевую вероятность всех переходов.

- **Параметризованное исследование:** Noisy Nets вводят шум в веса сети, а именно, полностью связанные слои заменяются на их зашумленные аналоги, параметры шума изучаются вместе с обычными весами, и масштаб шума автоматически адаптируется в процессе обучения.

3.2.2 Реализация оптимизации проксимальной политики

Оптимизация политики использует функцию цели с отсечением:

$$L(\theta) = \mathbb{E}[\min(r_t(\theta)\hat{A}t, \text{clip}(r_t(\theta), 1 - \varepsilon, 1 + \varepsilon)\hat{A}t)] \quad (3.4)$$

где $r_t(\theta)$ — отношение вероятностей, $\hat{A}t$ представляет обобщенную оценку преимущества, а ε — параметр отсечения (0.2):

$$r_t(\theta) = \frac{\pi\theta(a_t|s_t)}{\pi\theta_{\text{old}}(a_t|s_t)}, \quad \hat{A}t = \sum_k k = 0^{T-t+1}(\gamma\lambda)^k \delta_{t+k} \quad (3.5)$$

3.3 Система вознаграждений

Многокомпонентная функция вознаграждения включает:

- Геометрическую точность:

$$R_{\text{geo}} = 0.7 \cdot \text{IoU} + 0.2 \cdot \text{Precedge} + 0.1 \cdot \text{Shapereg} \quad (3.6)$$

с дифференцированными штрафами для различных типов ошибок.

- Топологическую корректность:

$$R_{\text{topo}} = \sum_{v \in V} \text{conn}(v) - 0.5 \sum_{e \in E} \text{cross}(e) \quad (3.7)$$

включая +0.05 за правильные соединения вершин, +0.1 за сохранение топологии и -0.2 за самопересечения.

– Вычислительную эффективность:

$$R_{\text{eff}} = -0.01 \cdot t_{\text{step}} + 0.1 \cdot \mathbb{I}(\text{terminate_early}) \quad (3.8)$$

Финальная формулировка вознаграждения:

$$R(s, a) = \alpha R_{\text{geo}} + \beta R_{\text{topo}} + \gamma R_{\text{eff}} \quad (3.9)$$

где $\alpha = 0.6$, $\beta = 0.3$, $\gamma = 0.1$ являются адаптивными коэффициентами.

На этапе обучения используется составная функция потерь:

$$L = 0.7 \cdot \text{Dice} + 0.3 \cdot \text{Focal}(\gamma = 2, \alpha = 0.8) \quad (3.10)$$

где Dice loss улучшает региональную сегментацию, а Focal loss решает проблему дисбаланса классов:

$$\text{Dice} = 1 - \frac{2|X \cap Y|}{|X| + |Y|}, \quad \text{Focal} = -\alpha(1 - p)^\gamma \log(p) \quad (3.11)$$

Для оптимизации используется оптимизатор AdamW со скоростью обучения $\text{lr} = 3 \times 10^{-4}$ и весовым затуханием 1×10^{-5} .

.

3.4 Результаты проведённых экспериментов

3.4.1 Метрики оценки

Предлагаемая методика проходит тщательную количественную и качественную оценку с помощью комплексной системы оценки, разработанной для охвата всех критических аспектов производительности векторизации аэрофотоснимков. Оценка геометрической точности использует среднее значение Intersection over Union (mIoU) в качестве основной метрики.

Intersection-over-Union (IoU) или индекс Жаккара — это отношение площади пересечения предсказанной и истинной маски к их объединению:

$$IoU = \frac{TP}{TP + FP + FN} \quad (3.12)$$

В данной работе mIoU вычисляется для всех классов объектов как:

$$mIoU = \frac{1}{C} \sum_{c=1}^C \frac{TP_c}{TP_c + FP_c + FN_c} \quad (3.13)$$

где C представляет количество классов (здания, дороги, водоемы и т.д.), TP_c обозначает истинно положительные срабатывания для класса c , FP_c указывает ложноположительные срабатывания, а FN_c учитывает ложноотрицательные срабатывания.

Качеству границ уделяется особое внимание через оценку Boundary F1-score (BD-F1), которая вычисляется с использованием значений точности и полноты:

$$Precision = \frac{TP}{TP + FP}, \quad Recall = \frac{TP}{TP + FN} \quad (3.14)$$

$$BD - F1 = \frac{2 \cdot Precision \cdot Recall}{Precision + Recall} \quad (3.15)$$

Оценка топологической корректности включает две специализированные метрики: Junction Assurance измеряет процент правильно идентифицированных точек пересечения между линейными объектами:

$$J_{acc} = \frac{1}{N} \sum_{i=1}^N \mathbb{I}(\|J_{pred}^i - J_{gt}^i\| < \varepsilon_{junction}) \quad (3.16)$$

где $\varepsilon_{junction}$ представляет радиус принятия в 5 пикселей, а Cycle Consistency оценивает целостность замыкания полигонов с помощью анализа связности вершин на основе теории графов.

3.4.2 Эксперимент

Алгоритм обучался на открытой базе данных CrowdAI Mapping Challenge [83], состоящей из более чем 280 тыс. спутниковых изображений для обучения и 60 тыс. изображений для тестирования. Обучающие изображения были разделены на две части: 80% изображений использовались для обучения алгоритма, а оставшиеся 20% были выделены для валидации. Обучение проводилось с использованием CUDA 11.8 на GPU NVIDIA GeForce RTX 3090 с 16 ГБ памяти.



Рисунок 3.2 — Некоторые изображения из набора данных CrowdAI Mapping Challenge.

Значения вычисленных метрик приведены в таблице 2. Для понимания уровня эффективности алгоритма в таблицу также включены результаты ведущих методов на аналогичных данных.

Комплексная оценка предлагаемой методики демонстрирует значительные улучшения по всем ключевым метрикам производительности. Качественная оценка показывает надежную работу системы для различных категорий объектов, причем контуры зданий демонстрируют особенно точное определение границ

Метод	mIoU	BD-F1	Время (мс)	Память (ГБ)
U-Net	0.75	0.66	120	6.2
Предложенный RL	0.82	0.79	85	9.8
HRNet	0.84	0.81	210	12.4

Таблица 2 — Результаты на тестовом наборе данных CrowdAI для всех экспериментов по векторизации.

(средняя ошибка позиционирования вершин 1.2 пикселя), а дорожные сети сохраняют превосходную топологическую связность (Junction Accuracy 93.4% и Cycle Consistency 94.1%). Выходные данные векторизации сохраняют картографическое качество с четко определенными вершинами и минимальным количеством избыточных точек, уменьшая среднюю сложность полигонов на 28% по сравнению с традиционными подходами на основе сегментации.

Количественный анализ подтверждает значительное повышение производительности: разработанный метод демонстрирует улучшение на 9.3% по среднему значению Intersection over Union (mIoU) по сравнению с подходом на основе U-Net (0.82 против 0.75). Метрики качества границ показывают еще более значительные улучшения: оценка F1 для границ увеличивается на 19.6% (0.79 против 0.66) по сравнению с базовым показателем U-Net. Измерения вычислительной эффективности показывают сокращение времени обработки одного изображения на 29% (85 миллисекунд против 120 миллисекунд) при сохранении более высокой точности, что достигнуто благодаря оптимизированному процессу принятия решений с подкреплением, сокращающему избыточные вычисления. Однако стоит отметить, что предложенный метод немного уступает по эффективности многоуровневым архитектурам, таким как HRNet, по mIoU (0.82 против 0.84) и по оценке F1 для границ (0.79 против 0.81), но обрабатывает изображение на 60% быстрее (85 миллисекунд против 210 миллисекунд), что может быть важно для решения задач векторизации в реальном времени. Анализ использования памяти показывает особенно впечатляющие результаты для обработки больших

областей: наш метод требует всего 9.8 ГБ памяти GPU по сравнению с 12.4 ГБ для вариантов HRNet, что позволяет проводить практическое развертывание на стандартном оборудовании.



Рисунок 3.3 — Результаты эксперимента.

3.5 Выводы

В данной главе предложена методология на основе обучения с подкреплением, которая демонстрирует ряд убедительных преимуществ, которые решают ключевые проблемы векторизации аэрофотоснимков, но также имеет определённые ограничения, которые следует учитывать. Наиболее существенное преимущество системы заключается в её исключительной адаптивности к разнообразным типам объектов, достигаемой благодаря синергетическому сочетанию многомасштабного извлечения признаков в опорной сети сверточной нейронной

сети (CNN) и контекстно-зависимого принятия решений в компоненте обучения с подкреплением (RL). Эта адаптивность особенно хорошо проявляется при работе с неоднородной природой городских ландшафтов, где метод поддерживает стабильную производительность для всех зданий и дорожных сетей, превосходя специализированные узкоспециализированные алгоритмы в тестах на кросс-категориальное обобщение. Способность архитектуры к обучению на ограниченных наборах данных обусловлена тремя ключевыми элементами дизайна: механизмами внимания, которые фокусируют обучение на семантически важных областях, гибридным протоколом обучения, который использует синтетические данные, и формированием вознаграждения, которое обеспечивает плотные обучающие сигналы, что в совокупности обеспечивает эффективное обучение на небольшом количестве аннотированных образцов, сохраняя при этом большую часть производительности, достигнутой с полными наборами данных.

Глава 4. Семантическая сегментация изображений дистанционного зондирования на основе визуального трансформера и графовой нейронной сети

Актуальность картографической информации является критически важным фактором для эффективного функционирования многих организаций и служб. Благодаря последним достижениям в способах получения данных дистанционного зондирования и значительному прогрессу в методах их обработки достигнут значительный прогресс в эффективности алгоритмов автоматического обновления карт. Ключевые усовершенствования в автоматическом обновлении карт опираются на большие объемы собранных данных дистанционного зондирования и применении методов машинного обучения для анализа данных.

Высококачественная сегментация аэрофотоснимков имеет решающее значение для создания точных карт. Применение методов машинного обучения для семантической сегментации изображений обусловило значительное повышение качества сегментации аэрофотоснимков, и в настоящее время модели нейронных сетей глубокого обучения демонстрируют наилучшее качество решения данной задачи. Следует отметить, высокое качество сегментации методами глубокого машинного обучения имеет с своей основе большой объём представительных аннотированных данных.

С целью сокращения времени и ресурсов для аннотации данных были предложены методы, не требующие большого объёма размеченных данных, такие как обучение с частичным привлечением учителя или полуавтоматическое обучение (semi-supervised learning) [114], а использование механизма внимания [115] (AttentionMap) позволило создать методы обучения без учителя или автоматические (self-supervised) [116] методы сегментации изображений.

Применение метода автоматического обучения визуальных трансформеров [116] показало, что автоматические визуальные трансформеры агрегируют явную информацию о семантике изображения в форме глубоких признаков и превосхо-

дят в данной задаче как визуальные трансформеры, обучаемые с учителем, так и сверточные нейронные сети. Полученные глубокие признаки содержат важную семантическую информацию, применимую для задач семантической сегментации, в том числе, и для автоматического обновления карт.

Представление на карте конкретного участка земной поверхности в векторной форме отражает не только визуальные детали выбранного района, но и существующие связи между объектами на карте. Такая форма представления информации делает карту более понятной для пользователя, чем исходный аэрофотоснимок (растровое изображение). Кроме того, карта позволяет не только составить представление о взаимосвязи между объектами, но также отражает иерархию объектов.

Нейросетевые модели способны извлекать и обобщать неявную информацию из изображений. При этом, чем точнее модель отражает существенные характеристики изучаемого процесса (объекта), тем более эффективно она будет решать поставленную задачу. Недавно предложенные графовые нейронные сети [117] демонстрируют наилучшие результаты в задачах, где рассматриваются данные, представимые графами. В частности, в таких областях, как обработка естественного языка (классификация документов, генерация текста, ответы на вопросы, анализ эмоций), биоинформатика (прогнозирование взаимодействия белок-белок, анализ геномной последовательности, исследования лекарственных средств), а также анализ и прогнозирование движения транспорта.

В данной главе представлена разработанная нейросетевая модель для извлечения семантической и векторной информации о местности на основе анализа аэрофотоснимка. Разработанная нейросетевая модель на первом этапе использует визуальный трансформер для извлечения глубоких признаков из входного изображения. Затем графовая нейронная сеть выполняет кластеризацию извлеченных глубоких признаков для получения семантической сегментации изображения.

Для обучения и оценки разработанной нейросетевой модели была собрана и аннотирована специальная выборка, содержащая более 10 тысяч аэрофото-

снимков, сделанных в разные годы и сезоны и представляющих различные типы объектов.

Основой методов машинного обучения, как и других методов основанных на данных (data-driven methods), является представительная выборка, содержащая примеры для обучения. Поэтому в рамках разработки нейросетевых методов семантической сегментации была подготовлена специальная выборка для разработки и исследований методов сегментации и векторизации аэрофотоснимков и обнаружения изменений SVAI (Segmentation and Vectorization of Aerial imagery dataset). Данная выборка затем использовалась для обучения и исследований разработанной нейросетевой модели.

4.1 Выборка SVAI

При создании выборки учитывались следующие основные требования:

- большой объём данных: 10 000 изображений или более;
- разнообразие данных: аэрофотоснимки должны содержать различные участки земной поверхности (городская застройка, транспортная сеть, сельскохозяйственные угодья и т.п.);
- разнообразие условий: выборка должна содержать данные, полученные в разные сезоны, время суток;
- разнообразие применяемых платформ: снимки должны быть получены с различных носителей и различными сенсорами.

Подготовленная выборка содержит более 12 000 изображений, приведённых к одному размеру (512×512 пикселей) и одному масштабу съёмки (0,34 м/пиксель). Для подготовки выборки использованы материалы спутниковой съёмки (Bing и Google Earth) и съёмки с беспилотных летательных аппаратов (БПЛА). Собранные данные охватывают период 12 лет и включают как появление новых объектов, так и исчезновение существовавших ранее.

В выборке SVAI представлены различные типы зданий и сооружений, автомобильные и железные дороги, различные типы ландшафтов (лесные участки, городская застройка различных типов, водные участки), полученные в разное время.

Для аннотации данных использован формат бинарных масок, традиционно применяемый в задачах обнаружения изменений [118]. На рис. 4.1 показаны примеры изображений и аннотаций из выборки SVAI. Временной период изменений: середина 2018 года – середина 2024 года.

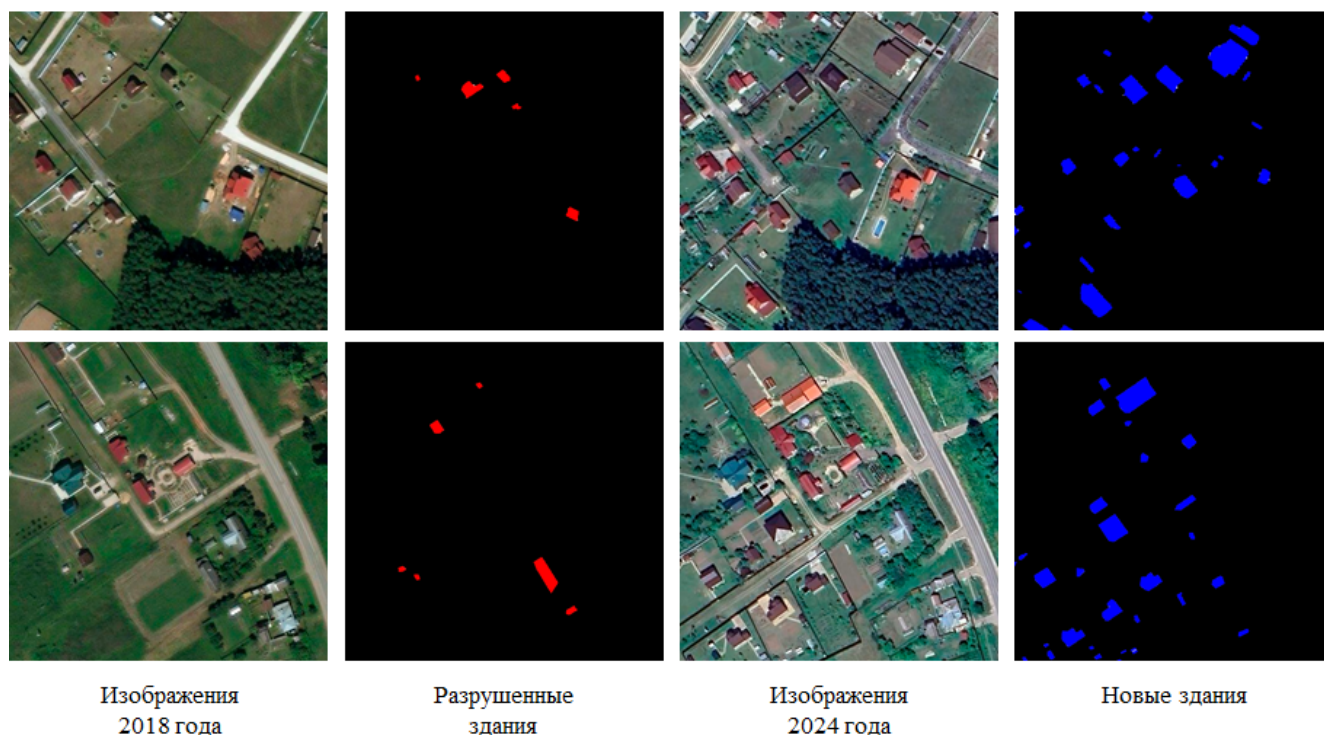


Рисунок 4.1 — Примеры изображений из набора данных Segmentation and Visualization Aerial Images. Изображения в столбцах 1 и 3 были получены со спутников Bing и Google в 2018 и 2024 годах соответственно; столбцы 2 и 4 показывают изменения (разрушенные объекты показаны красным, а новые появившиеся объекты показаны синим)

Для исследований в области векторизации аэрофотоснимков в выборку SVAI включены векторные данные об объектах. Векторные аннотации формировались двумя способами. Первый – ручная разметка оператором-дешифровщиком. Второй – создание векторной разметки с привлечением ресурса «Открытая карта улиц» (OpenStreetMap, OSM) [119]. OpenStreetMap – проект

веб-картографирования на некоммерческой основе, направленный на создание подробной, бесплатной и открытой географической карты мира силами добровольцев. Для создания карт используются данные с персональных GPS-трекеров, аэрофотосъемка, видеозаписи, спутниковые снимки и панорамы улиц, предоставляемые участниками проекта.

Данные OpenStreetMap содержат информацию об объектах в векторном формате и использует топологическую структуру данных, состоящую из объектов:

- node (точка) — точка с указанными координатами;
- way (линия) — упорядоченный список точек, составляющих линию или полигон;
- relation (отношение) — группы точек, линий и других отношений, которым назначаются некоторые свойства;
- tag (тег) — пары «ключ — значение», могут назначаться точкам, линиям и отношениям.

Фрагмент файла OSM для объекта типа отношение (relation) приведён на рис. 4.2.

```
<relation id="7069781" version="1" timestamp="2017-03-14T17:13:22Z" changeset="0">
  <member type="way" ref="6052497" role="from"/>
  <member type="node" ref="50168872" role="via"/>
  <member type="way" ref="475420031" role="to"/>
  <tag k="restriction" v="no_u_turn"/>
  <tag k="type" v="restriction"/>
</relation>
```

Рисунок 4.2 — Фрагмент файла OSM

Пример данных из выборки, предназначенной для задач векторизации аэрофотоснимков, показан на рис. 4.3. Для первичной разметки данных использовались данные OpenStreetMap, которые впоследствии проверялись (и корректировались в случае необходимости) оператором-дешифровщиком с формированием журнала действий оператора, необходимым для исследований методов обучения с подкреплением в задачах обнаружения изменений и векторизации аэрофотоснимков.

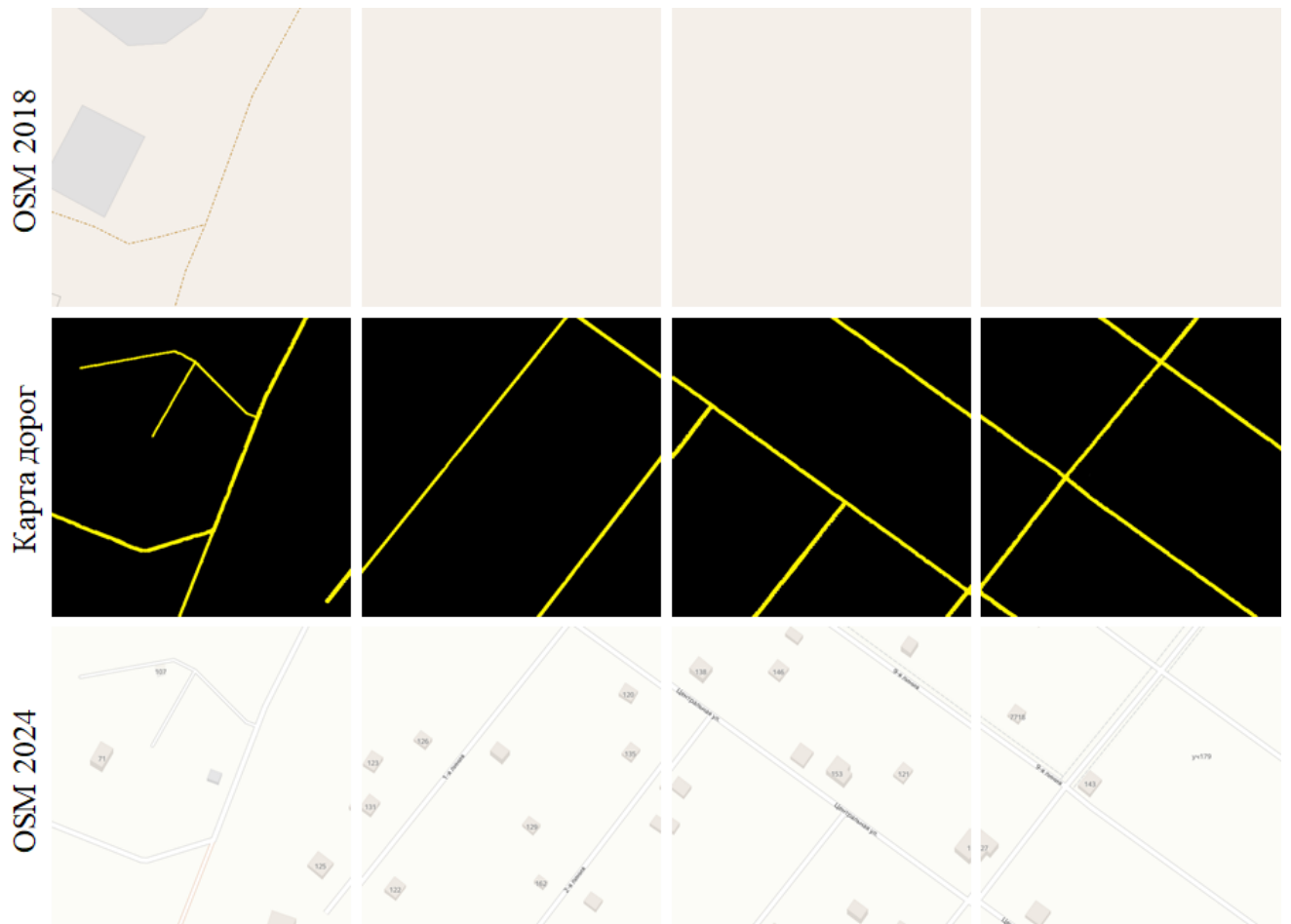


Рисунок 4.3 — Примеры изображений из выборки SVAI. Верхний и нижний ряды — это изображения OSM за 2018 и 2024 годы соответственно, а средний ряд — это изменения в дорожной карте за этот период времени

4.2 Архитектура нейросетевой модели

При выборе подхода к решению задачи семантической сегментации аэро-фотоснимков исследовалась гипотеза о том, что качество работы нейросетевой модели сегментации может быть улучшено при использовании схемы анализа изображения, сходной с работой оператора- дешифровщика, решающего аналогичную задачу. А именно, на первом этапе оператор анализирует весь снимок в целом и выделяет области, которые требуется более внимательно изучить. На втором этапе проводится детальный анализ данных областей и из них формируются семантические объекты сцены.

В соответствии с выбранным подходом принята двухэтапная схема сегментации. Сначала нейросеть - визуальный трансформер используется для извлечения глубоких признаков объектов из входного аэрофотоснимка на основе механизма внимания. Затем графовая нейронная сеть применяется для кластеризации найденных глубоких признаков в однородные сегментированные области.

4.2.1 Извлечение глубоких признаков

Для извлечения глубоких признаков объектов из аэрофотоснимка используется визуальный трансформер (ViT) [120], обученный с помощью модели DINO (self-DIstillation with NO labels) [116].

Нейросетевая модель DINO является представителем современных самообучающихся моделей [121—123], не требующих размеченных данных. Модель напрямую прогнозирует выход нейросети-учителя с использованием стандартной функции потерь перекрёстной энтропии (cross-entropy loss).

DINO использует преимущества дистилляции знаний [124], что обеспечивает повышение качества решения задачи в таких областях как сегментация изображений, обнаружение объектов, сегментация видеоэкземпляров и т. д. Данная функция позволяет DINO выделять различные области на изображении, которые затем могут быть использованы для точной семантической сегментации. Предложенная архитектура показана на рис. 4.4.

Визуальный трансформер делит входное изображение I на k^2 фрагментов (токенов) размером

$$s = \frac{h}{k} \times \frac{w}{k}, \quad (4.1)$$

где h и w высота и ширина изображения I соответственно.

Для векторного представления (эмбеддинга) аэрофотоснимка размерностью t , размерность выходного вектора признаков Z составит $s \times t$.

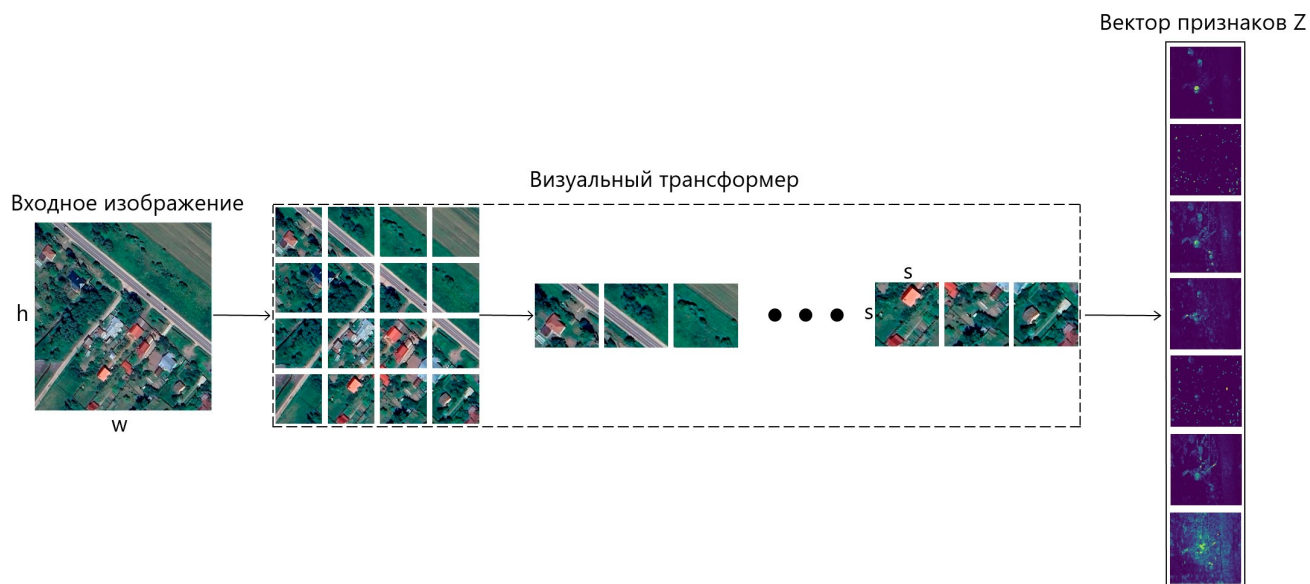


Рисунок 4.4 — После извлечения глубоких признаков из предварительно обученной модели ViT создается матрица подобия, рассматривая сходства между признаками на уровне патча. При использовании этой матрицы и глубоких признаков в качестве признаков узлов, строится граф

Образцы карт внимания, формируемые DINO для нескольких выходов визуального трансформера, показаны на рис. 4.5.

Формируемые DINO карты внимания содержат значительную информацию об изображении, которая может быть использована для семантической сегментации. Если рассматривать сходство между фрагментами изображения как граф, то задачу сегментации изображения можно сформулировать как задачу разреза графа.

4.2.2 Представление графа изображения

При решении задачи сегментации на основе теории графов изображение представляется в виде неориентированного графа $G = \{\mathcal{V}, \mathcal{E}\}$ с множеством узлов \mathcal{V} и множеством ребер \mathcal{E} , где каждый узел соответствует области изображения. Соответствие между областями изображения характеризуется с помощью матри-

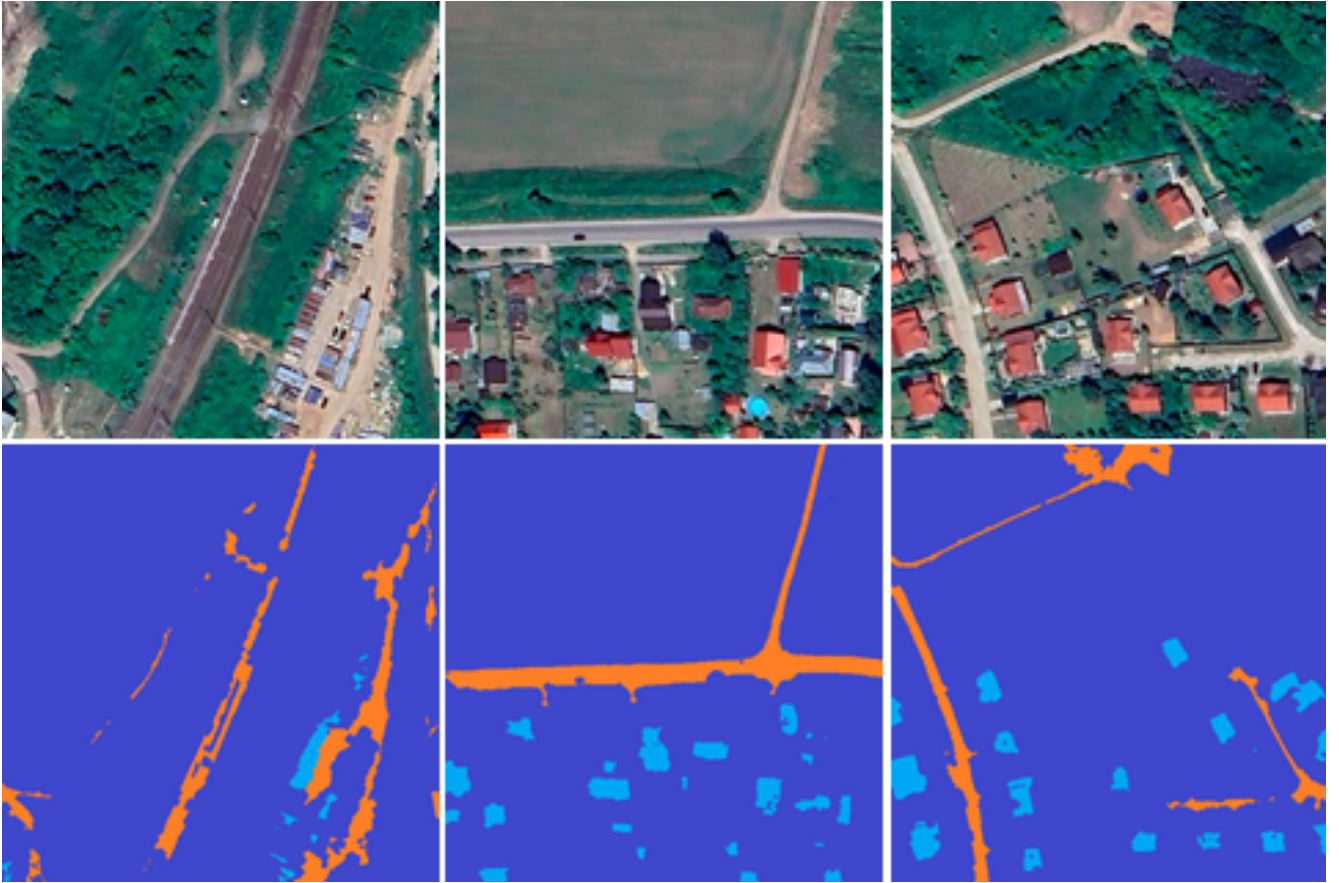


Рисунок 4.5 — Карты внимания из нескольких головок ViT-S/8 для запроса токена [CLS]

цы W размера $n \times n$, где $n = |\mathcal{V}|$, в которой элементы w_{ij} представляют сходство между областями изображения i и j , где $i, j = 1, \dots, n$. Тогда задача сегментации изображения ставится как задача группировки узлов по признаку их схожести, или, другими словами, разбиения графа на множество отдельных однородных частей.

Данную задачу можно сформулировать следующим образом. Требуется разбить граф на k непересекающихся множеств A_1, A_2, \dots, A_k таких, что $\cup_i A_i = \mathcal{V}$ и $\forall_{j \neq i} A_i \cap A_j = \emptyset$. Такое разбиение можно представить в виде бинарной матрицы $S \in \{0, 1\}^{n \times k}$, где $S_{ic} = 1$, если $i \in A_c$.

В качестве критерия корректности разбиения в данном исследовании принимается метрика количества связей между разбиениями, которая должна быть минимизирована для оптимального разреза (минимальный разрез).

Количество связей между частью A и частью B заданного графа задается разрезом $cut(A, B)$ графа G :

$$cut(A, B) = \sum_{u \in A, v \in B} w(u, v). \quad (4.2)$$

Лучший способ разбить граф на два — достичь минимально возможного значения разреза. Однако, как показано в работе [125], критерий минимального разреза способствует выделению внутри графов изолированных узлов. Рис. 4.6 иллюстрирует это явление. Фактически, любое деление, которое изолирует отдельные узлы на правой стороне, даст меньшее значение разреза по сравнению с разбиениями, которые разделяют узлы на левую и правую части.

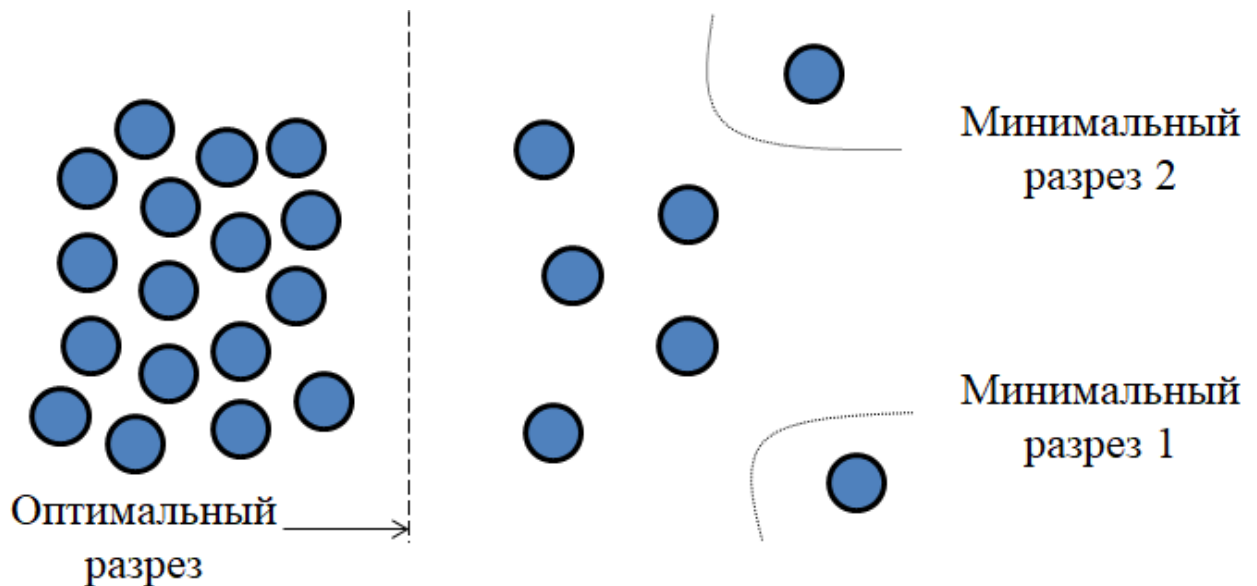


Рисунок 4.6 — Случай, когда минимальный разрез дает плохое разбиение

С целью предотвращения такого искусственного смещения при разделении небольших наборов точек, в данной работе используется подход *нормализованного разреза*. Вместо того, чтобы учитывать общий вес ребер, соединяющих два раздела, мера оценивает стоимость разреза как долю от общего количества реберных соединений со всеми узлами в графе.

Из этого следует определение *нормализованного разреза* $Ncut(A, B)$ графа G :

$$Ncut(A, B) = \frac{\sum_{u \in A, v \in B} w(u, v)}{\sum_{u \in A, v \in V} w(u, v)} + \frac{\sum_{u \in A, v \in B} w(u, v)}{\sum_{u \in B, v \in V} w(u, v)}, \quad (4.3)$$

Таким образом, для выполнения семантической сегментации изображения с использованием графового подхода необходимо определить, как установить сходство между областями изображения. Для подобных целей строится матрица сходства W с использованием выходного глубокого вектора признаков Z из визуального трансформера.

$$W = ZZ^T \in \mathbb{R}^{k \times k}. \quad (4.4)$$

4.2.3 Кластеризация графовых нейронных сетей

Пусть \hat{X} будет матрицей представлений узлов, полученной одним или несколькими слоями свертки графовой нейронной сети на графе G с матрицей сходства W . В данном случае используется однослойная GCN и строится граф, используя матрицу корреляции из извлеченных признаков визуального трансформера. Затем с помощью многослойного персептрона (MLP) с функцией активации на выходном слое вычисляется матрица кластерной принадлежности C , которая содержит векторы, представляющие вероятность принадлежности узла к определенному кластеру:

$$\hat{X} = GNN(n, W; \Theta_{GNN}), C = MLP(\hat{X}; \Theta_{MLP}), \quad (4.5)$$

где Θ_{MLP} и Θ_{GNN} — обучаемые параметры. Функция активации MLP гарантирует, что $c_{ij} \in [0, 1]$.

GNN оптимизируется с использованием доработанной модели нормализованного разреза, предложенной в [126]. Таким образом, функция потерь:

$$\mathcal{L}_{Ncut} = -\frac{Tr(C^T W C)}{Tr(C^T D C)} + \left\| \frac{C^T C}{\|C^T C\|_F} - \frac{\mathbb{I}_3}{\sqrt{3}} \right\|_F, \quad (4.6)$$

где $\|\cdot\|_F$ обозначает норму Фробениуса, а D — матрицу степеней W . Число 3 обозначает количество непересекающихся множеств, на которые разбивается граф, а \mathbb{I}_3 — единичная матрица.

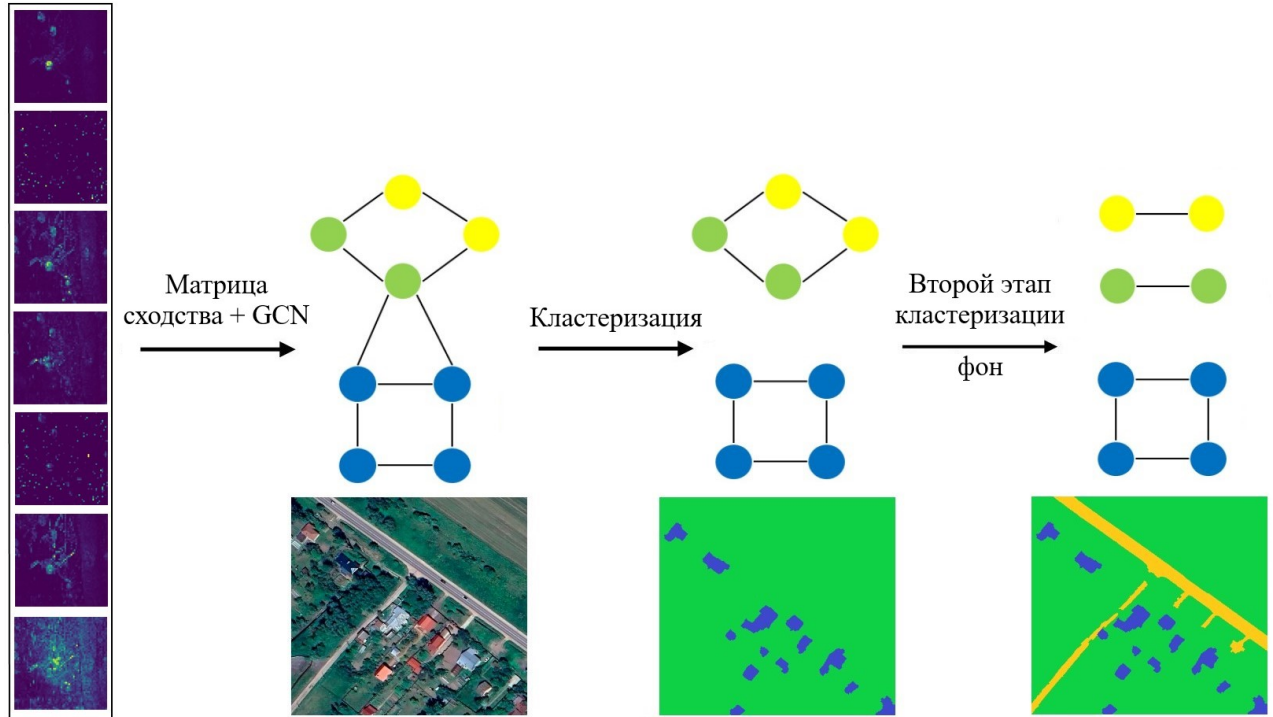


Рисунок 4.7 — Предлагаемая двухэтапная кластеризация. В начале изображение разделяется на два отдельных класса, а затем применяется кластеризация конкретно к фону

Первый член целевой функции способствует кластеризации сильно связанных компонентов, тогда как второй член направлен на то, чтобы полученные кластеры были ортогональны и имели схожие размеры.

4.3 Результаты

Оценка автоматической семантической сегментации была выполнена на разработанном наборе данных SVAI. Предложенный алгоритм сравнивается

с современными автоматическими методами. Среднее пересечение-над-объединением (mIoU) использовалось в качестве меры точности семантической сегментации. Сравнение количественных значений mIoU для каждого из методов показано в табл. 1.

Метод	Набор данных SVAI
OneGAN [127]	57,48
Voynov [128]	69,12
Spectral Methods [129]	75,81
TokenCut [130]	72,36
Разработанный метод	79,91

Таблица 3 — Значения метрики mIoU на разработанном наборе данных для различных алгоритмов.

Качественные результаты семантической сегментации по предлагаемой нейросетевой модели показаны на рис. 4.8.

Результаты оценки разработанного алгоритма на основе набора данных SVAI показывают, что предлагаемая методика демонстрирует современную производительность в задаче семантической сегментации аэрофотоснимков.

4.4 Выводы

В данной главе используется графовая нейронная сеть для извлечения семантической и векторной информации о выбранной области из ее аэрофотоснимка. Разработанная нейросетевая модель на первом этапе использует визуальный трансформер для извлечения глубоких признаков из входного аэрофотоснимка. Затем графовая нейронная сеть выполняет кластеризацию извлеченных глубоких признаков для получения семантической сегментации изображения. Для

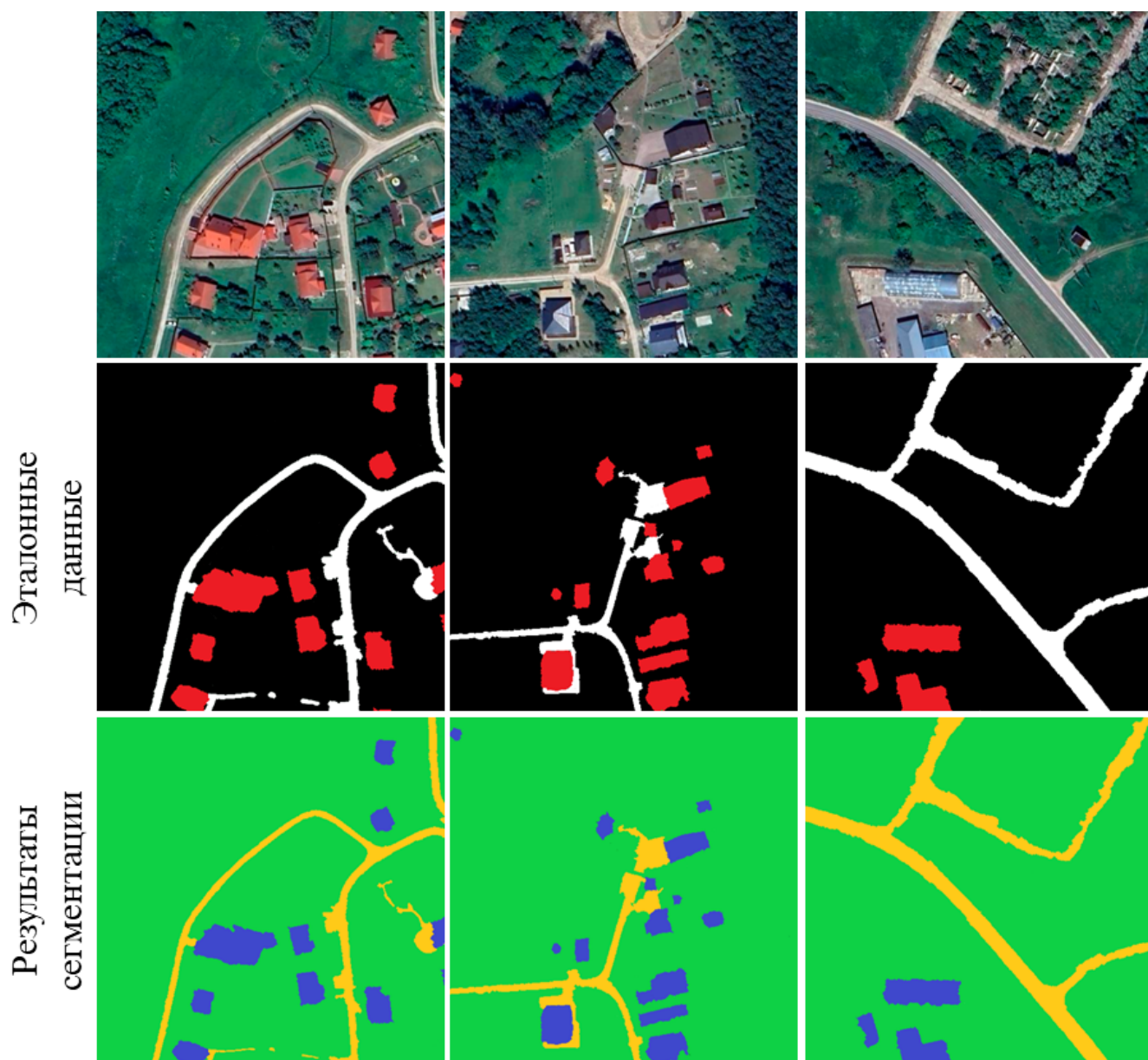


Рисунок 4.8 — Семантическая сегментация. Верхний ряд: исходное изображение; средний ряд: истинные данные; нижний ряд: результат работы предлагаемого алгоритма

выполнения поставленных задач был собран и аннотирован специальный набор данных, который содержит более 10 тыс. аэрофотоснимков, представляющих различные типы объектов, сделанных в разные годы и сезоны. Результаты оценки предлагаемой нейросетевой модели на созданном наборе данных подтвердили современную производительность, а именно прирост точности порядка 10% по сравнению с ведущими методами.

Заключение

Разработанные в исследовании методы сегментации и векторизации аэрофотоснимков вносят комплексный вклад в модернизацию процессов обновления геопространственной информации. Предложенные алгоритмы на основе бинарной сегментации с матрицей окрестности и многоэтапной регуляризации позволяют автоматизировать выявление пространственных изменений, включая детектирование новых объектов капитального строительства и трансформацию контуров существующих сооружений. Метод построения иерархического графа семантической разметки обеспечивает точное векторное представление данных с сохранением топологических отношений между объектами и учетом контекстных связей.

Существенное ускорение процессов обновления достигается за счет алгоритма на основе обучения с подкреплением, который имитирует логику оператора-дешифровщика и позволяет сократить временные затраты на обработку данных. Нейросетевая модель на основе визуального трансформера и графовой сети способствует улучшению детализации и структурирования данных через выделение объектов сложной геометрической формы и их классификацию по семантическим признакам. Применение единой методологической основы для сегментации и векторизации гарантирует согласованность пространственных данных и их соответствие стандартам ГИС.

Практическая значимость разработанных методов заключается в создании технологического фундамента для построения систем автоматизированного обновления геопространственной информации, способных оперативно отражать изменения на местности и поддерживать актуальность цифровых карт, кадастровых данных и тематических геоинформационных ресурсов.

Основные результаты работы заключаются в следующем.

1. Разработан алгоритм на основе бинарной сегментации с использованием матрицы окрестности для упорядочивания прогнозируемых контуров зданий и улучшения их геометрии.
2. Разработан алгоритм автоматического извлечения границ зданий на основе бинарной сегментации с использованием многоэтапной регуляризации, включающей метод матрицы окрестности, оптимизацию спектральных границ и адаптивное уточнение контуров.
3. Разработан алгоритм векторизации аэрофотоснимков, основанный на обучении с подкреплением.
4. Разработана нейросетевая модель для точной сегментации аэрофотоснимков на основе визуального трансформера и графовой нейронной сети.
5. Создана представительная выборка, содержащая разновременные аэрофотоснимки различных участков земной поверхности.

Таким образом, поставленная цель, а именно разработка методов по повышению качества дешифрирования и векторизации аэрокосмических изображений земной поверхности на основе машинного обучения в области разработки интеллектуальных методов анализа и обновления геопространственной информации, успешно достигнута.

В заключение автор выражает благодарность и большую признательность научному руководителю Князю В. А. за поддержку, помощь, обсуждение результатов и научное руководство. Автор также благодарит всех, кто сделал настоящую работу автора возможной.

Словарь терминов

MRF : Марковское случайное поле

CRF : Условное случайное поле

DSSN : Сеть глубокой семантической сегментации

DS : Дистанционное зондирование

RF : Случайные леса

DT : Деревья решений

SVM : Машины опорных векторов

FCN : Полностью сверточные сети

CNN : Сверточные нейронные сети

DSM : Цифровые модели поверхности

GNN : Графовые нейронные сети

GCN : Графовые свёрточные сети

GAT : Графовая сеть внимания

RNN : Рекуррентные нейронные сети

RL : Обучение с подкреплением

DQN : Глубокая сеть на основе Q-функции

PPO : Оптимизация проксимальной политики

MHSA : Модуль многоголового самовнимания

MHSA : Модуль многоголового перекрестного внимания

IoU : Пересечение над объединением, индекс Жаккара

AP : Средняя точность

AR : Средняя полнота

TP : Истинно положительные

FP : Ложно положительные

FN : Ложно отрицательные

SVAI : Выборка для сегментации и векторизации аэрофотоснимков и обнаружения изменений

OSM : Открытая карта улиц (OpenStreetMap)

ViT : Визуальный трансформер

MLP : Многослойный персептрон

Список литературы

1. *He, X.* Multiscale conditional random fields for image labeling [Текст] / X. He, R. Zemel, M. Carreira-Perpinan // Proceedings of the 2004 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2004. CVPR 2004. T. 2. — 2004. — С. II—II.
2. *Verbeek, J.* Region Classification with Markov Field Aspect Models [Текст] / J. Verbeek, B. Triggs //. — 06.2007.
3. *Blei, D.* Latent Dirichlet Allocation [Текст] / D. Blei, A. Ng, M. Jordan //. T. 3. — 01.2001. — С. 601—608.
4. Multi-Class Segmentation with Relative Location Prior [Текст] / S. Gould [и др.] // International Journal of Computer Vision. — 2008. — Т. 80. — С. 300—316. — URL: <https://api.semanticscholar.org/CorpusID:9779450> ; (Visited on: 25.08.2025).
5. TextonBoost: Joint Appearance, Shape and Context Modeling for Multi-class Object Recognition and Segmentation [Текст] / J. Shotton [и др.] // Computer Vision – ECCV 2006 / под ред. A. Leonardis, H. Bischof, A. Pinz. — Berlin, Heidelberg : Springer Berlin Heidelberg, 2006. — С. 1—15.
6. *Krähenbühl, P.* Efficient Inference in Fully Connected CRFs with Gaussian Edge Potentials [Текст] / P. Krähenbühl, V. Koltun. — 2012. — arXiv: 1210.5644 [cs.CV]. — URL: <https://arxiv.org/abs/1210.5644> ; (Visited on: 25.08.2025).
7. *Csurka, G.* An Efficient Approach to Semantic Segmentation [Текст] / G. Csurka // International Journal of Computer Vision. — 2011. — Нояб. — Т. 95. — С. 198—212.
8. *Kumar, M.* OBJ CUT [Текст] / M. Kumar, P. Ton, A. Zisserman // 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05). T. 1. — 2005. — 18—25 vol. 1.

9. Image segmentation with a bounding box prior [Текст] / V. Lempitsky [и др.] // 2009 IEEE 12th International Conference on Computer Vision. — 2009. — С. 277—284.
10. *Leibe, B.* Combined object categorization and segmentation with an implicit shape model [Текст] / B. Leibe, A. Leonardis, B. Schiele // Proc. 8th Eur. Conf. Comput. Vis. (ECCV). — 2004. — ЯНВ. — Т. 2.
11. *Borenstein, E.* Learning to Segment [Текст] / E. Borenstein, S. Ullman //. — 05.2004. — С. 315—328.
12. *Boykov, Y.* Interactive graph cuts for optimal boundary & region segmentation of objects in N-D images [Текст] / Y. Boykov, M.-P. Jolly // Proceedings Eighth IEEE International Conference on Computer Vision. ICCV 2001. Т. 1. — 2001. — 105—112 vol.1.
13. *Li, L.-J.* Towards total scene understanding: Classification, annotation and segmentation in an automatic framework [Текст] / L.-J. Li, R. Socher, L. Fei-Fei // 2009 IEEE Conference on Computer Vision and Pattern Recognition. — 2009. — С. 2036—2043.
14. *Badrinarayanan, V.* Segnet: A deep convolutional encoder-decoder architecture for image segmentation [Текст] / V. Badrinarayanan, A. Kendall, R. Cipolla // IEEE transactions on pattern analysis and machine intelligence. — 2017. — Т. 39, № 12. — С. 2481—2495.
15. Unified Perceptual Parsing for Scene Understanding [Текст] / T. Xiao [и др.]. — 2018. — arXiv: 1807.10221 [cs.CV]. — URL: <https://arxiv.org/abs/1807.10221> ; (Visited on: 25.08.2025).
16. High-Resolution Representations for Labeling Pixels and Regions [Текст] / K. Sun [и др.]. — 2019. — arXiv: 1904.04514 [cs.CV]. — URL: <https://arxiv.org/abs/1904.04514> ; (Visited on: 25.08.2025).

17. PointFlow: Flowing Semantics Through Points for Aerial Image Segmentation [Текст] / X. Li [и др.]. — 2021. — arXiv: 2103.06564 [cs.CV]. — URL: <https://arxiv.org/abs/2103.06564> ; (Visited on: 25.08.2025).
18. Foreground-Aware Relation Network for Geospatial Object Segmentation in High Spatial Resolution Remote Sensing Imagery [Текст] / Z. Zheng [и др.]. — 2020. — arXiv: 2011.09766 [cs.CV]. — URL: <https://arxiv.org/abs/2011.09766> ; (Visited on: 25.08.2025).
19. Context Contrasted Feature and Gated Multi-scale Aggregation for Scene Segmentation [Текст] / H. Ding [и др.] // 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. — 2018. — С. 2393—2402.
20. Learning Aerial Image Segmentation From Online Maps [Текст] / P. Kaiser [и др.] // IEEE Transactions on Geoscience and Remote Sensing. — 2017. — Ноябрь. — Т. 55, № 11. — С. 6054—6068. — URL: <http://dx.doi.org/10.1109/TGRS.2017.2719738>.
21. DenseASPP for Semantic Segmentation in Street Scenes [Текст] / M. Yang [и др.] // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR). — 06.2018.
22. *Sahu, M.* VECTOR MAP GENERATION FROM AERIAL IMAGERY USING DEEP LEARNING [Текст] / M. Sahu, A. Ohri // ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences. — 2019. — Т. IV—2/W5. — С. 157—162. — URL: <https://isprs-annals.copernicus.org/articles/IV-2-W5/157/2019/>.
23. DOTA: A Large-Scale Dataset for Object Detection in Aerial Images [Текст] / G.-S. Xia [и др.] // 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. — 2018. — С. 3974—3983.
24. iSAID: A Large-scale Dataset for Instance Segmentation in Aerial Images [Текст] / S. W. Zamir [и др.]. — 2019. — arXiv: 1905.12886 [cs.CV]. — URL: <https://arxiv.org/abs/1905.12886> ; (Visited on: 25.08.2025).

25. *LeCun, Y.* Deep learning [Текст] / Y. LeCun, Y. Bengio, G. Hinton // nature. — 2015. — Т. 521, № 7553. — С. 436—444.
26. *Li, Y.* Error-tolerant deep learning for remote sensing image scene classification [Текст] / Y. Li, Y. Zhang, Z. Zhu // IEEE transactions on cybernetics. — 2020. — Т. 51, № 4. — С. 1756—1768.
27. *Emelyanov, A.* Extracting building outlines based on convolutional neural networks using the property of linear connectivity [Text] / A. Emelyanov, V. A. Knyaz, V. V. Kniaz // The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences. — 2024. — Vol. XLVIII-1—2024. — P. 147—152. — URL: <https://isprs-archives.copernicus.org/articles/XLVIII-1-2024/147/2024/>.
28. *Basaeed, E.* Supervised remote sensing image segmentation using boosted convolutional neural networks [Текст] / E. Basaeed, H. Bhaskar, M. Al-Mualla // Knowledge-Based Systems. — 2016. — Т. 99. — С. 19—27.
29. Advances in hyperspectral image classification: Earth monitoring with statistical learning methods [Текст] / G. Camps-Valls [и др.] // IEEE signal processing magazine. — 2013. — Т. 31, № 1. — С. 45—54.
30. *Long, J.* Fully convolutional networks for semantic segmentation [Текст] / J. Long, E. Shelhamer, T. Darrell // Proceedings of the IEEE conference on computer vision and pattern recognition. — 2015. — С. 3431—3440.
31. *Ronneberger, O.* U-net: Convolutional networks for biomedical image segmentation [Текст] / O. Ronneberger, P. Fischer, T. Brox // Medical image computing and computer-assisted intervention—MICCAI 2015: 18th international conference, Munich, Germany, October 5-9, 2015, proceedings, part III 18. — Springer. 2015. — С. 234—241.
32. Semantic segmentation of slums in satellite images using transfer learning on fully convolutional neural networks [Текст] / M. Wurm [и др.] // ISPRS journal of photogrammetry and remote sensing. — 2019. — Т. 150. — С. 59—69.

33. *Sherrah, J.* Fully convolutional networks for dense semantic labelling of high-resolution aerial imagery [Текст] / J. Sherrah // arXiv preprint arXiv:1606.02585. — 2016.
34. *Kampffmeyer, M.* Semantic segmentation of small objects and modeling of uncertainty in urban remote sensing images using deep convolutional neural networks [Текст] / M. Kampffmeyer, A.-B. Salberg, R. Jenssen // Proceedings of the IEEE conference on computer vision and pattern recognition workshops. — 2016. — С. 1—9.
35. *Wang, C.* Multi-scale residual deep network for semantic segmentation of buildings with regularizer of shape representation [Текст] / C. Wang, L. Li // Remote Sensing. — 2020. — Т. 12, № 18. — С. 2932.
36. *Audebert, N.* Semantic segmentation of earth observation data using multimodal and multi-scale deep networks [Текст] / N. Audebert, B. Le Saux, S. Lefèvre // Asian conference on computer vision. — Springer. 2016. — С. 180—196.
37. Learning dual multi-scale manifold ranking for semantic segmentation of high-resolution images [Текст] / M. Zhang [и др.] // Remote Sensing. — 2017. — Т. 9, № 5. — С. 500.
38. Semantic labeling of high resolution aerial imagery and LiDAR data with fine segmentation network [Текст] / X. Pan [и др.] // Remote sensing. — 2018. — Т. 10, № 5. — С. 743.
39. Semantic segmentation of aerial images with an ensemble of CNSS [Текст] / D. Marmanis [и др.] // ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences, 2016. — 2016. — Т. 3. — С. 473—480.
40. Remote sensing image semantic segmentation based on edge information guidance [Текст] / С. He [и др.] // Remote Sensing. — 2020. — Т. 12, № 9. — С. 1501.

41. Semantic referee: A neural-symbolic framework for enhancing geospatial semantic segmentation [Текст] / M. Alirezaie [и др.] // Semantic Web. — 2019. — Т. 10, № 5. — С. 863—880.
42. *Hu, J.* Squeeze-and-excitation networks [Текст] / J. Hu, L. Shen, G. Sun // Proceedings of the IEEE conference on computer vision and pattern recognition. — 2018. — С. 7132—7141.
43. *Oktay, O.* Attention u-net: Learning where to look for the Pancreas [Текст] / O. Oktay // arXiv preprint arXiv:1804.03999. — 2018.
44. SCAttNet: Semantic segmentation network with spatial and channel attention mechanism for high-resolution remote sensing images [Текст] / H. Li [и др.] // IEEE Geoscience and Remote Sensing Letters. — 2020. — Т. 18, № 5. — С. 905—909.
45. Cbam: Convolutional block attention module [Текст] / S. Woo [и др.] // Proceedings of the European conference on computer vision (ECCV). — 2018. — С. 3—19.
46. *Gori, M.* A new model for learning in graph domains [Текст] / M. Gori, G. Monfardini, F. Scarselli // Proceedings. 2005 IEEE international joint conference on neural networks, 2005. Т. 2. — IEEE. 2005. — С. 729—734.
47. *Welling, M.* Semi-supervised classification with graph convolutional networks [Текст] / M. Welling, T. N. Kipf // J. International Conference on Learning Representations (ICLR 2017). — 2016.
48. *Niepert, M.* Learning convolutional neural networks for graphs [Текст] / M. Niepert, M. Ahmed, K. Kutzkov // International conference on machine learning. — PMLR. 2016. — С. 2014—2023.
49. Deepgcns: Can gcns go as deep as cnns? [Текст] / G. Li [и др.] // Proceedings of the IEEE/CVF international conference on computer vision. — 2019. — С. 9267—9276.

50. Graph attention networks [Текст] / P. Veličković [и др.] // arXiv preprint arXiv:1710.10903. — 2017.
51. Graph-FCN for image semantic segmentation [Текст] / Y. Lu [и др.] // International symposium on neural networks. — Springer. 2019. — С. 97—105.
52. A CNN-GCN framework for multi-label aerial image scene classification [Текст] / Y. Li [и др.] // IGARSS 2020-2020 IEEE International Geoscience and Remote Sensing Symposium. — IEEE. 2020. — С. 1353—1356.
53. Building Detection and Reconstruction from Mid- and High-Resolution Aerial Imagery [Текст] / N. Paparoditis [и др.] // Computer Vision and Image Understanding. — 1998. — Т. 72, № 2. — С. 122—142.
54. *Persson, M.* Automatic building detection from aerial images for mobile robot mapping [Текст] / M. Persson, M. Sandvall, T. Duckett // 2005 International Symposium on Computational Intelligence in Robotics and Automation. — 2005. — С. 273—278.
55. Building Extraction at Scale Using Convolutional Neural Network: Mapping of the United States [Текст] / H. L. Yang [и др.] // IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing. — 2018. — Т. 11, № 8. — С. 2600—2614.
56. Backpropagation Applied to Handwritten Zip Code Recognition [Текст] / Y. LeCun [и др.] // Neural Computation. — 1989. — Дек. — Т. 1, № 4. — С. 541—551. — eprint: <https://direct.mit.edu/neco/article-pdf/1/4/541/811941/neco.1989.1.4.541.pdf>. — URL: <https://doi.org/10.1162/neco.1989.1.4.541>.
57. *Li, Z.* Topological Map Extraction From Overhead Images [Текст] / Z. Li, J. D. Wegner, A. Lucchi // Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV). — 10.2019.
58. An end-to-end shape modeling framework for vectorized building outline generation from aerial images [Текст] / Q. Chen [и др.] // ISPRS Journal of Photogrammetry and Remote Sensing. — 2020. — Т. 170. — С. 114—126.

59. AN END-TO-END DEEP LEARNING WORKFLOW FOR BUILDING SEGMENTATION, BOUNDARY REGULARIZATION AND VECTORIZATION OF BUILDING FOOTPRINTS [Текст] / S. Šanca [и др.] // The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences. — 2023. — T. XLVIII-4/W7—2023. — С. 169—175. — URL: <https://isprs-archives.copernicus.org/articles/XLVIII-4-W7-2023/169/2023/>.
60. TEMPORARY REMOVAL: Aerial imagery for roof segmentation: A large-scale dataset towards automatic mapping of buildings [Текст] / Q. Chen [и др.] // ISPRS Journal of Photogrammetry and Remote Sensing. — 2019. — T. 147. — С. 42—55.
61. Can semantic labeling methods generalize to any city? the inria aerial image labeling benchmark [Текст] / E. Maggiori [и др.] // 2017 IEEE International Geoscience and Remote Sensing Symposium (IGARSS). — 2017. — С. 3226—3229.
62. *Wei, S.* Toward automatic building footprint delineation from aerial images using CNN and regularization [Текст] / S. Wei, S. Ji, M. Lu // IEEE Transactions on Geoscience and Remote Sensing. — 2019. — T. 58, № 3. — С. 2178—2189.
63. *Zhao, W.* Building Instance Segmentation and Boundary Regularization from High-Resolution Remote Sensing Images [Текст] / W. Zhao, C. Persello, A. Stein // IGARSS 2020 - 2020 IEEE International Geoscience and Remote Sensing Symposium. — 2020. — С. 3916—3919.
64. Building Extraction From Satellite Images Using Mask R-CNN With Building Boundary Regularization [Текст] / K. Zhao [и др.] // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR) Workshops. — 06.2018.
65. 3D Reconstruction of a Complex Grid Structure Combining UAS Images and Deep Learning [Текст] / V. A. Кныаз [и др.] // Remote Sensing. — 2020. — T. 12, № 19. — URL: <https://www.mdpi.com/2072-4292/12/19/3128>.

66. *Huang, W.* OEC-RNN: Object-Oriented Delineation of Rooftops With Edges and Corners Using the Recurrent Neural Network From the Aerial Images [Текст] / W. Huang, H. Tang, P. Xu // IEEE Transactions on Geoscience and Remote Sensing. — 2021. — Май. — Т. PP. — С. 1—12.
67. DANCE: A Deep Attentive Contour Model for Efficient Instance Segmentation [Текст] / Z. Liu [и др.] // Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision (WACV). — 01.2021. — С. 345—354.
68. *Kass, M.* Snakes: Active contour models [Текст] / M. Kass, A. Witkin, D. Terzopoulos // International journal of computer vision. — 1988. — Т. 1, № 4. — С. 321—331.
69. *Chan, T.* Active contours without edges [Текст] / T. Chan, L. Vese // IEEE Transactions on Image Processing. — 2001. — Т. 10, № 2. — С. 266—277.
70. Learning Deep Structured Active Contours End-to-End [Текст] / D. Marcos [и др.] // 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. — 2018. — С. 8877—8885. — URL: <https://api.semanticscholar.org/CorpusID:3939983> ; (Visited on: 25.08.2025).
71. *Hatamizadeh, A.* End-to-End Deep Convolutional Active Contours for Image Segmentation [Текст] / A. Hatamizadeh, D. Sengupta, D. Terzopoulos // ArXiv. — 2019. — Т. abs/1909.13359. — URL: <https://api.semanticscholar.org/CorpusID:203593984> ; (Visited on: 25.08.2025).
72. PolarMask: Single Shot Instance Segmentation With Polar Representation [Текст] / E. Xie [и др.] // 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). — 2020. — С. 12190—12199.
73. PolarMask++: Enhanced Polar Representation for Single-Shot Instance Segmentation and Beyond [Текст] / E. Xie [и др.] // IEEE Transactions on Pattern Analysis and Machine Intelligence. — 2021. — Т. 44. — С. 5385—5400. — URL: <https://api.semanticscholar.org/CorpusID:233739844> ; (Visited on: 25.08.2025).

74. Location-Sensitive Visual Recognition with Cross-IOU Loss [Текст] / K. Duan [и др.] // ArXiv. — 2021. — Т. abs/2104.04899. — URL: <https://api.semanticscholar.org/CorpusID:233210422>; (Visited on: 25.08.2025).
75. *Liu, Z.* Building Outline Delineation From VHR Remote Sensing Images Using the Convolutional Recurrent Neural Network Embedded With Line Segment Information [Текст] / Z. Liu, H. Tang, W. Huang // IEEE Transactions on Geoscience and Remote Sensing. — 2022. — Т. 60. — С. 1—13.
76. A Review of Recurrent Neural Networks: LSTM Cells and Network Architectures [Текст] / Y. Yu [и др.] // Neural Computation. — 2019. — ИЮЛЬ. — Т. 31, № 7. — С. 1235—1270. — eprint: https://direct.mit.edu/neco/article-pdf/31/7/1235/1053200/neco_a_01199.pdf. — URL: https://doi.org/10.1162/neco%5C_a%5C_01199.
77. PolyWorld: Polygonal Building Extraction With Graph Neural Networks in Satellite Images [Текст] / S. Zorzi [и др.] // Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). — 06.2022. — С. 1848—1857.
78. Processing of Extremely High-Resolution LiDAR and RGB Data: Outcome of the 2015 IEEE GRSS Data Fusion Contest—Part A: 2-D Contest [Текст] / M. Campos-Taberner [и др.] // IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing. — 2016. — Т. 9, № 12. — С. 5547—5559.
79. THE ISPRS BENCHMARK ON URBAN OBJECT CLASSIFICATION AND 3D BUILDING RECONSTRUCTION [Текст] / F. Rottensteiner [и др.] // ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences. — 2012. — Т. I—3. — С. 293—298. — URL: <https://isprs-annals.copernicus.org/articles/I-3/293/2012/>.
80. *Ji, S.* Fully Convolutional Networks for Multisource Building Extraction From an Open Aerial and Satellite Imagery Data Set [Текст] / S. Ji, S. Wei, M. Lu //

- IEEE Transactions on Geoscience and Remote Sensing. — 2019. — Т. 57, № 1. — С. 574—586.
81. Asymmetric Siamese Networks for Semantic Change Detection in Aerial Images [Текст] / К. Yang [и др.] // IEEE Transactions on Geoscience and Remote Sensing. — 2022. — Т. 60. — С. 1—18.
 82. Hi-UCD: A Large-scale Dataset for Urban Semantic Change Detection in Remote Sensing Imagery [Текст] / S. Tian [и др.]. — 2020. — arXiv: 2011.03247 [cs.CV].
 83. Deep Learning for Understanding Satellite Imagery: An Experimental Survey [Текст] / S. P. Mohanty [и др.] // Frontiers in Artificial Intelligence. — 2020. — Т. 3.
 84. SpaceNet MVOI: A Multi-View Overhead Imagery Dataset [Текст] / N. Weir [и др.] // Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV). — 10.2019.
 85. Human-level control through deep reinforcement learning [Текст] / V. Mnih [и др.] // nature. — 2015. — Т. 518, № 7540. — С. 529—533.
 86. Sutton, R. S. Reinforcement Learning: An Introduction [Текст] / R. S. Sutton, A. G. Barto. — Second. — The MIT Press, 2018. — URL: <http://incompleteideas.net/book/the-book-2nd.html>.
 87. Playing Atari with Deep Reinforcement Learning [Текст] / V. Mnih [и др.]. — 2013. — arXiv: 1312.5602 [cs.LG]. — URL: <https://arxiv.org/abs/1312.5602>.
 88. Mastering the game of Go with deep neural networks and tree search [Текст] / D. Silver [и др.] // nature. — 2016. — Т. 529, № 7587. — С. 484—489.
 89. Proximal Policy Optimization Algorithms [Текст] / J. Schulman [и др.]. — 2017. — arXiv: 1707.06347 [cs.LG]. — URL: <https://arxiv.org/abs/1707.06347>.

90. Soft Actor-Critic: Off-Policy Maximum Entropy Deep Reinforcement Learning with a Stochastic Actor [Текст] / T. Haarnoja [и др.]. — 2018. — arXiv: 1801.01290 [cs.LG]. — URL: <https://arxiv.org/abs/1801.01290>.
91. DeepLab: Semantic Image Segmentation with Deep Convolutional Nets, Atrous Convolution, and Fully Connected CRFs [Текст] / L.-C. Chen [и др.]. — 2017. — arXiv: 1606.00915 [cs.CV]. — URL: <https://arxiv.org/abs/1606.00915>.
92. *Tao, T.* Evaluating Vision Transformer Methods for Deep Reinforcement Learning from Pixels [Текст] / T. Tao, D. Reda, M. van de Panne. — 2022. — arXiv: 2204.04905 [cs.LG]. — URL: <https://arxiv.org/abs/2204.04905>.
93. Graph Neural Network Reinforcement Learning for Autonomous Mobility-on-Demand Systems [Текст] / D. Gammelli [и др.]. — 2021. — arXiv: 2104.11434 [eess.SY]. — URL: <https://arxiv.org/abs/2104.11434>.
94. Few-Shot Classification of Aerial Scene Images via Meta-Learning [Текст] / P. Zhang [и др.] // Remote Sensing. — 2021. — Т. 13, № 1. — URL: <https://www.mdpi.com/2072-4292/13/1/108>.
95. Self-supervised Learning in Remote Sensing: A Review [Текст] / Y. Wang [и др.]. — 2022. — arXiv: 2206.13188 [cs.CV]. — URL: <https://arxiv.org/abs/2206.13188>.
96. *Ha, D.* Recurrent world models facilitate policy evolution [Текст] / D. Ha, J. Schmidhuber // Proceedings of the 32nd International Conference on Neural Information Processing Systems. — Montréal, Canada : Curran Associates Inc., 2018. — С. 2455—2467. — (NIPS'18).
97. *f*-GAIL: Learning *f*-Divergence for Generative Adversarial Imitation Learning [Текст] / X. Zhang [и др.]. — 2020. — arXiv: 2010.01207 [cs.LG]. — URL: <https://arxiv.org/abs/2010.01207> ; (Visited on: 25.08.2025).
98. Visual Adversarial Imitation Learning using Variational Models [Текст] / R. Rafailov [и др.]. — 2022. — arXiv: 2107.08829 [cs.LG]. — URL: <https://arxiv.org/abs/2107.08829> ; (Visited on: 25.08.2025).

99. *Furuta, R.* PixelRL: Fully Convolutional Network with Reinforcement Learning for Image Processing [Текст] / R. Furuta, N. Inoue, T. Yamasaki. — 2019. — arXiv: 1912.07190 [cs.CV]. — URL: <https://arxiv.org/abs/1912.07190> ; (Visited on: 25.08.2025).
100. Hierarchical Object Detection with Deep Reinforcement Learning [Текст] / M. Bellver [и др.]. — 2016. — arXiv: 1611.03718 [cs.CV]. — URL: <https://arxiv.org/abs/1611.03718> ; (Visited on: 25.08.2025).
101. Reinforced active learning for image segmentation [Текст] / A. Casanova [и др.]. — 2020. — arXiv: 2002.06583 [cs.CV]. — URL: <https://arxiv.org/abs/2002.06583> ; (Visited on: 25.08.2025).
102. *Canny, J.* A Computational Approach To Edge Detection [Текст] / J. Canny // Pattern Analysis and Machine Intelligence, IEEE Transactions on. — 1986. — Дек. — Т. PAMI—8. — С. 679—698.
103. Encoder-Decoder with Atrous Separable Convolution for Semantic Image Segmentation [Текст] / L.-C. Chen [и др.]. — 2018. — arXiv: 1802.02611 [cs.CV]. — URL: <https://arxiv.org/abs/1802.02611>.
104. An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale [Текст] / A. Dosovitskiy [и др.]. — 2021. — arXiv: 2010.11929 [cs.CV]. — URL: <https://arxiv.org/abs/2010.11929>.
105. Weakly Supervised Semantic Segmentation in Aerial Imagery via Cross-Image Semantic Mining [Текст] / R. Zhou [и др.] // Remote Sensing. — 2023. — Т. 15, № 4. — URL: <https://www.mdpi.com/2072-4292/15/4/986>.
106. *Li, Z.* Topological Map Extraction from Overhead Images [Текст] / Z. Li, J. D. Wegner, A. Lucchi. — 2019. — arXiv: 1812.01497 [cs.CV]. — URL: <https://arxiv.org/abs/1812.01497>.
107. Deep Reinforcement Learning in Computer Vision: A Comprehensive Survey [Текст] / N. Le [и др.]. — 2021. — arXiv: 2108.11510 [cs.CV]. — URL: <https://arxiv.org/abs/2108.11510>.

108. Mask R-CNN [Текст] / К. Хе [и др.]. — 2018. — arXiv: 1703.06870 [cs.CV].
109. Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks [Текст] / S. Ren [и др.]. — 2016. — arXiv: 1506.01497 [cs.CV].
110. U-Net Transformer: Self and Cross Attention for Medical Image Segmentation [Текст] / O. Petit [и др.] // ArXiv. — 2021. — T. abs/2103.06104. — URL: <https://api.semanticscholar.org/CorpusID:232170496> ; (Visited on: 25.08.2025).
111. Deep Residual Learning for Image Recognition [Текст] / К. Хе [и др.]. — 2015. — arXiv: 1512.03385 [cs.CV]. — URL: <https://arxiv.org/abs/1512.03385>.
112. Learning Phrase Representations using RNN Encoder-Decoder for Statistical Machine Translation [Текст] / К. Cho [и др.]. — 2014. — arXiv: 1406.1078 [cs.CL]. — URL: <https://arxiv.org/abs/1406.1078>.
113. Dueling Network Architectures for Deep Reinforcement Learning [Текст] / Z. Wang [и др.]. — 2016. — arXiv: 1511.06581 [cs.LG]. — URL: <https://arxiv.org/abs/1511.06581>.
114. Semi-supervised learning of visual features by non-parametrically predicting view assignments with support samples [Текст] / M. Assran [и др.] // Proceedings of the IEEE/CVF International Conference on Computer Vision. — 2021. — С. 8443—8452.
115. *Zagoruyko, S.* Paying More Attention to Attention: Improving the Performance of Convolutional Neural Networks via Attention Transfer [Текст] / S. Zagoruyko, N. Komodakis. — 2017. — arXiv: 1612.03928 [cs.CV]. — URL: <https://arxiv.org/abs/1612.03928>.
116. Emerging properties in self-supervised vision transformers [Текст] / М. Caron [и др.] // Proceedings of the IEEE/CVF International Conference on Computer Vision. — 2021. — С. 9650—9660.

117. Graph neural networks: A review of methods and applications [Текст] / J. Zhou [и др.] // AI Open. — 2020. — Т. 1. — С. 57—81. — URL: <https://www.sciencedirect.com/science/article/pii/S2666651021000012>.
118. *LEVIR-CD*. Source: <https://justchenhao.github.io/LEVIR/> [Текст] / LEVIR-CD.
119. *OpenStreetMap*. Source: <https://www.openstreetmap.org> [Текст] / OpenStreetMap.
120. An image is worth 16x16 words: Transformers for image recognition at scale [Текст] / A. Dosovitskiy [и др.] // arXiv preprint arXiv:2010.11929. — 2020.
121. Unsupervised learning of visual features by contrasting cluster assignments [Текст] / M. Caron [и др.] // NeurIPS. — 2020.
122. Bootstrap your own latent: A new approach to self-supervised learning [Текст] / J.-B. Grill [и др.] // NeurIPS. — 2020.
123. Momentum contrast for unsupervised visual representation learning [Текст] / K. He [и др.] // CVPR. — 2020.
124. *Hinton, G.* Distilling the knowledge in a neural network [Текст] / G. Hinton, O. Vinyals, J. Dean // preprint arXiv:1503.02531. — 2015.
125. *Wu, Z.* An Optimal Graph Theoretic Approach to Data Clustering: Theory and Its Application to Image Segmentation [Текст] / Z. Wu, R. M. Leahy // IEEE Trans. Pattern Anal. Mach. Intell. — 1993. — Т. 15. — С. 1101—1113. — URL: <https://api.semanticscholar.org/CorpusID:2595046>.
126. *Bianchi, F. M.* Spectral clustering with graph neural networks for graph pooling [Текст] / F. M. Bianchi, D. Grattarola, C. Alippi // International Conference on Machine Learning. — PMLR. 2020. — С. 874—883.
127. *Benny, Y.* Onegan: Simultaneous unsupervised learning of conditional image generation, foreground segmentation, and fine-grained clustering [Текст] / Y. Benny, L. Wolf // European Conference on Computer Vision. — Springer. 2020. — С. 514—530.

128. *Voynov, A.* Object segmentation without labels with large-scale generative models [Текст] / A. Voynov, S. Morozov, A. Babenko // International Conference on Machine Learning. — PMLR. 2021. — С. 10596—10606.
129. Deep Spectral Methods: A Surprisingly Strong Baseline for Unsupervised Semantic Segmentation and Localization [Текст] / L. Melas-Kyriazi [и др.] // Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. — 2022. — С. 8364—8375.
130. Self-supervised transformers for unsupervised object discovery using normalized cut [Текст] / Y. Wang [и др.] // Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. — 2022. — С. 14543—14553.

Список рисунков

2.1	Схема предложенного алгоритма.	31
2.2	Структура Mask R-CNN.	32
2.3	Схематическое определение линейно связного множества.	34
2.4	Использование свойства линейной связности для удаления ненужных точек с изображений. На первом изображении обнаружен объект класса «здание». На втором — объекты класса «здание» сегментируются в ограничивающем прямоугольнике. На третьем — пиксели, не принадлежащие основному объекту, но ошибочно сегментированные, выделяются красным цветом и удаляются. На последнем изображении представлен конечный результат сегментации и регуляризации.	35
2.5	Схема предложенного алгоритма.	35
2.6	Структура сверточной нейронной сети, используемой для бинарной сегментации.	36
2.7	Матрицы окрестности для граничных точек. Матрицы (a-c, f) показывают основные типы вершин выпуклых углов зданий, а (d, e) — типы границ.	39
2.8	Результат процесса регуляризации. Слева: входное изображение. В центре: сегментированное изображение. Справа: сегментированное изображение после процесса регуляризации.	39
2.9	Алгоритм регуляризации и выбора особо значимых точек.	40
2.10	Схема предлагаемого алгоритма, объединяющая регуляризацию матрицы соседства, оптимизацию спектральной границы и адаптивное уточнение контура.	41
2.11	Результат процесса регуляризации. Слева: исходное изображение. В центре: сегментированное изображение. Справа: сегментированное изображение после процесса регуляризации.	45

2.12	Некоторые изображения из набора данных CrowdAI Mapping Challenge.	47
2.13	Результаты эксперимента для метода с регуляризацией на основе свойства линейной связности.	48
2.14	Результаты эксперимента для метода с регуляризацией на основе матрицы окрестности.	49
2.15	Результаты эксперимента для метода с многоэтапной регуляризацией. .	50
3.1	Структура предложенного алгоритма.	52
3.2	Некоторые изображения из набора данных CrowdAI Mapping Challenge.	57
3.3	Результаты эксперимента.	59
4.1	Примеры изображений из набора данных Segmentation and Visualization Aerial Images. Изображения в столбцах 1 и 3 были получены со спутников Bing и Google в 2018 и 2024 годах соответственно; столбцы 2 и 4 показывают изменения (разрушенные объекты показаны красным, а новые появившиеся объекты показаны синим)	64
4.2	Фрагмент файла OSM	65
4.3	Примеры изображений из выборки SVAI. Верхний и нижний ряды — это изображения OSM за 2018 и 2024 годы соответственно, а средний ряд — это изменения в дорожной карте за этот период времени	66
4.4	После извлечения глубоких признаков из предварительно обученной модели ViT создается матрица подобия, рассматривая сходства между признаками на уровне патча. При использовании этой матрицы и глубоких признаков в качестве признаков узлов, строится граф	68
4.5	Карты внимания из нескольких головок ViT-S/8 для запроса токена [CLS]	69
4.6	Случай, когда минимальный разрез дает плохое разбиение	70
4.7	Предлагаемая двухэтапная кластеризация. В начале изображение разделяется на два отдельных класса, а затем применяется кластеризация конкретно к фону	72

4.8 Семантическая сегментация. Верхний ряд: исходное изображение; средний ряд: истинные данные; нижний ряд: результат работы предлагаемого алгоритма	74
--	----

Список таблиц

1	Результаты на тестовом наборе данных CrowdAI для всех экспериментов по извлечению и полигонизации зданий.	48
2	Результаты на тестовом наборе данных CrowdAI для всех экспериментов по векторизации.	58
3	Значения метрики mIoU на разработанном наборе данных для различных алгоритмов.	73