

МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ
РОССИЙСКОЙ ФЕДЕРАЦИИ
ФГБОУ ВО «РЯЗАНСКИЙ ГОСУДАРСТВЕННЫЙ
РАДИОТЕХНИЧЕСКИЙ УНИВЕРСИТЕТ ИМ. В.Ф. УТКИНА»

На правах рукописи

Тарасова

ТАРАСОВА ВАЛЕНТИНА ЮРЬЕВНА

**МЕТОД СОВМЕЩЕНИЯ РАЗНОРОДНЫХ ИЗОБРАЖЕНИЙ
ПОДСТИЛАЮЩЕЙ ПОВЕРХНОСТИ С ИСПОЛЬЗОВАНИЕМ
НЕЙРОСЕТЕВЫХ ТЕХНОЛОГИЙ**

Специальность 2.3.8. – Информатика и информационные процессы

Диссертация
на соискание учёной степени
кандидата технических наук

Научный руководитель:
доктор технических наук,
профессор КОСТРОВ Б.В.

Рязань, 2026

ОГЛАВЛЕНИЕ

ВВЕДЕНИЕ	4
ГЛАВА 1 АНАЛИЗ ИЗВЕСТНЫХ МЕТОДОВ СОВМЕЩЕНИЯ ИЗОБРАЖЕНИЙ	12
1.1 Корреляционный метод совмещения изображения	14
1.2 Метод на основе дробно-линейного преобразования	15
1.3 Метод совмещения на основе ключевых точек	18
1.4 Нейросетевые методы совмещения изображений	28
1.4.1 Определение визуального сходства	39
1.5 Выводы по первой главе	42
ГЛАВА 2 МЕТОД СОВМЕЩЕНИЯ РАЗНОРОДНЫХ ИЗОБРАЖЕНИЙ ПОДСТИЛАЮЩЕЙ ПОВЕРХНОСТИ	44
2.1 Метод совмещения разнородных изображений подстилающей поверхности и алгоритмы его реализующие.....	44
2.1.1 Алгоритм обучения свёрточной нейронной сети	46
2.2.2 Алгоритм формирования вектора-признаков изображения	53
2.2.3 Алгоритм поиска спутникового изображения, позволяющего определить потенциальное положение летательного аппарата.....	67
2.2 Способ предварительной обработки изображений подстилающей поверхности с использованием спектральных преобразований	70
2.3 Выводы по второй главе.....	78
ГЛАВА 3 АЛГОРИТМ ОПРЕДЕЛЕНИЯ ПОЛОЖЕНИЯ ТЕКУЩЕГО ИЗОБРАЖЕНИЯ ОТНОСИТЕЛЬНО СПУТНИКОВОГО ИЗОБРАЖЕНИЯ. 80	
3.1 Задача уточнения положения ЛА по спутниковому снимку, полученному в результате работы нейросетевого алгоритма	81
3.1.1 Формирование и структура базы данных	83
3.2 Алгоритм детектирования и сопоставления ключевых точек	88
3.4 Алгоритм уточнения на основе ключевых точек	90
3.4.2 Оценка вычислительной сложности	94
3.5 Вычисление географических координат методом билинейной интерполяции.....	97
3.6 Выводы по третьей главе	98

ГЛАВА 4 АПРОБАЦИЯ И ИССЛЕДОВАНИЕ МЕТОДА СОВМЕЩЕНИЯ РАЗНОРОДНЫХ ИЗОБРАЖЕНИЙ ПОДСТИЛАЮЩЕЙ ПОВЕРХНОСТИ И АЛГОРИТМОВ ЕГО РЕАЛИЗУЮЩИХ.....	100
4.1 Описание программно-алгоритмического комплекса	100
4.2 Метрики качества совмещения изображений	102
4.3 Экспериментальные исследования	103
4.3.1 Исследования алгоритмов предварительной обработки изображений подстилающей поверхности.....	103
4.3.2 Экспериментальные исследования различных моделей нейронных сетей	110
4.3.3 Экспериментальные исследования сходимости нейронной сети в зависимости от контекста обучающей выборки.....	113
4.3.4 Экспериментальные исследования поиска изображения, соответствующего изображениям из базы данных	116
4.3.5 Экспериментальные исследования детектирования ключевых точек	120
4.4 Оценка точности и погрешности совмещения изображений	121
4.5 Выводы.....	127
ЗАКЛЮЧЕНИЕ	129
СПИСОК ЛИТЕРАТУРЫ.....	131
ПРИЛОЖЕНИЕ 1 Акт внедрения результатов кандидатской диссертации в ООО «Квантрон Групп»	146
ПРИЛОЖЕНИЕ 2 Акт внедрения результатов кандидатской диссертации в учебном процессе ФГБОУ ВО «РГРТУ».....	147
ПРИЛОЖЕНИЕ 2. Свидетельства о государственной регистрации программ для ЭВМ	148
ПРИЛОЖЕНИЕ 3. Патент на изобретение.....	153

ВВЕДЕНИЕ

Актуальность темы исследования. В настоящее время существует большое количество сервисов, которые предоставляют доступ к базам данных изображений подстилающей поверхности, снятыми с помощью различных устройств, отличающиеся временем проведения съемки, различными углами ориентации устройств, производящих съемку, различными высотами. Совмещение разнородных изображений позволяет решить ряд задач, связанных с использованием беспилотной воздушной техники в различных сферах функционирования общества, таких как: сельское хозяйство, органы охраны и надзора, мониторинг чрезвычайных ситуаций, системы автономной доставки грузов, поиск пропавших людей. Использование данных технологий приводит к существенному увеличению объемов обрабатываемой информации, которая в большинстве затронутых областей представлена цифровыми базами данных разнородных изображений районов использования беспилотной воздушной техники. Под разнородными изображениями будем понимать изображения, полученные с борта летательного аппарата (ЛА) и спутниковые изображения.

Совмещение разнородных изображений может служить основой для систем определения положения беспилотной воздушной техники, что является актуальной задачей в условиях отсутствия или вызванного помехами искажения сигнала глобальных систем навигации. Определение положения воздушной беспилотной техники возможно с помощью совмещения разнородных изображений, полученных с бортовых сенсоров и с заранее сформированной базы данных цифровых спутниковых изображений района её использования.

В настоящее время для решения данной задачи применяются ряд альтернативных подходов таких как: астронавигация, зависящая от условий видимости небесных светил, геофизические методы навигации, навигация по рельефу и по изображениям местности. Одним из наиболее перспективных направлений в данной области является визуальная навигация, основанная на сравнении данных бортовых сенсоров со спутниковыми изображениями, что позволяет использовать недорогое оборудование. Её частным случаем является

определение положения беспилотной воздушной техники на основе совмещения разнородных изображений подстилающей поверхности.

В настоящее время существуют законченные технические решения для оцифровки и хранения больших объемов визуальной информации, которые обладают существенным недостатком в виде чрезмерного объема данных, занимаемого как отдельным изображением, так и всей базой данных в целом. В области разработки алгоритмов хранения и предоставления информации из базы данных достигнуты некоторые положительные результаты, в области разработки алгоритмов поиска и анализа визуальной информации, исследователям не удалось достичь технических решений, полностью удовлетворяющих потребностям пользователей.

В данной работе рассматривается определение местоположения на основе совмещения разнородных изображений подстилающей поверхности, ключевым подходом к решению данной задачи является поиск изображений по заданному сюжету (Content-based Image Retrieval, CBIR). В настоящее время существует научная задача определения и сопоставления аэрокосмических и спутниковых снимков в условиях отсутствия или ненадёжности сигналов/данных с глобальных навигационных систем. Существуют подходы, основанные на сопоставлении ключевых точек, которые позволяют формировать координаты в локальной системе координат. При этом остаётся актуальной задача привязки получаемых локальных координат к глобальным. Основой решения данной задачи может служить метод совмещения разнородных изображений подстилающей поверхности.

Степень проработанности темы исследования

Задачей совмещения разнородных изображений в отечественной научной школе занимались такие исследователи, как Новиков А.И., Ефимов А.И., работы которых охватывают вопросы сопоставления разнородных изображений, отличающихся по ракурсу, разрешению. В работах Визильтера Ю.В., Жел-

това С.Ю., Елесиной С.И. были предложены методы корреляционно-экстремального совмещения изображений, широко применяемые для решения данной задачи.

В настоящее время для решения задачи определения местоположения беспилотной воздушной техники применяются ряд альтернативных подходов: астронавигация, зависящая от условий видимости небесных светил, геофизические методы навигации (AQNav – Atomic Quantum Navigation, атомная квантовая навигация), навигация по рельефу (TERCOM – TERrain Contour Matching, сопоставление с контуром рельефа местности) и по изображениям местности (DSMAC – Digital Scene-Matching Area Correlation, цифровая корреляция по площади со сценой, VISNAV – VISual NAVigation, визуальная навигация). Одним наиболее перспективным направлением в данной области является визуальная навигация, основанная на сравнении данных бортовых датчиков (камер, лидаров) с эталонами картами, так как позволяет использовать недорогое оборудование. Ее частным случаем является навигация по совмещению разнородных изображений подстилающей поверхности.

Поиск изображений по заданному сюжету и обнаружение целевых объектов могут быть взяты за основу визуальной навигации, осуществляемой на основе анализа данных с видеопотока, выделения на них ключевых признаков, а навигационные решения (корректировка курса, привязка к местности) принимаются в результате использования методов машинного обучения и алгоритмов сопоставления признаков. В области поиска изображений по заданному сюжету известны работы Игнатова А.К., Рогова А.А., Бабенко А.В., Горы С.Ю., Популина С.Ю. Спектрально-текстурным анализом занимались, в частности, Злобин В.К., Костров Б.В., Сюзев В.В., Трахтман А.М., Левашкина А.О., который направлен повысить надежность классификации и распознавания образов.

Нейронные сети для разноракурсной геолокации (cross-view geolocalization). Зарубежные исследовательские группы, такие как Albert Gordo, Jon Almazan, Jerome Revaud, Diane Larlus, а также Tzu-Heng Henry Lee, Noa

Garcia, George Vogiatzis, сосредоточены на создании нейросетевых архитектур (СVM-Net, модели с механизмами внимания, графовые сети), способных находить семантическое соответствие между изображениями, сделанными с разных точек – например, между наземной фотографией и спутниковым снимком. Их работы направлены на преодоление семантического и геометрического разрыва между разнородными изображениями.

Научная задача заключается в определении положения ЛА при отсутствии сигналов глобальных навигационных систем на основе совмещения разнородных изображений подстилающей поверхности для заданного района полёта.

Целью настоящей работы является разработка метода совмещения текущего изображения подстилающей поверхности со спутниковыми изображениями района полёта и алгоритмов, его реализующих.

Для достижения этой цели необходимо решить следующие задачи:

1. Выявить недостатки существующих методов совмещения изображений в контексте сравнения изображений подстилающей поверхности.
2. Разработать метод и алгоритмы совмещения разнородных изображений подстилающей поверхности с текущим изображением района полёта.
3. Разработать способ предварительной обработки изображений подстилающей поверхности с использованием спектральных преобразований.
4. Разработать алгоритм определения местоположения летательного аппарата за счёт использования ключевых точек.
5. Экспериментальное исследование разработанного метода и алгоритмов, его реализующих с применением реальной информации.

Научная новизна заключается в разработке:

1. Метода совмещения разнородных изображений подстилающей поверхности, и алгоритмов, его реализующих, отличающихся от известных подходов следующим:

1.1) алгоритм обучения свёрточной нейронной сети отличается применением последующей декомпозиции ее архитектуры;

1.2) алгоритм формирования вектора-признаков отличается способом извлечения признаков, обеспечивающим степень сжатия визуальной информации фрагмента изображений с 20 килобайт до 256 байт при сохранении семантического содержания;

1.3) алгоритм поиска спутникового изображения впервые адаптирован для сопоставления текущего кадра с сенсора ЛА с эталонным изображением из цифровой базы данных, что позволяет обеспечить первичное позиционирование ЛА с точностью до 300 пикселей.

2. Способа предварительной обработки изображений подстилающей поверхности, отличающегося применением фильтрации низкочастотных секвент в спектральном пространстве Уолша совместно с нормализацией и аугментацией, что обеспечивает рост точности сходимости сети на 7,8 %.

3. Алгоритма определения положения текущего изображения относительно спутникового изображения, отличающегося механизмом использования ключевых точек, сужающим радиус точного совмещения до 20 пикселей.

Практическая значимость работы заключается в том, что разработанный программно-алгоритмический комплекс, внедрённый в ООО «Квантрон Групп», обеспечивает определение местоположения ЛА с точностью 10–20 м в режиме реального времени (время обработки ≤ 207 мс) в условиях недоступности сигналов ГНСС на основе разнородных изображений подстилающей поверхности.

Теоретическая значимость работы заключается в развитии теории распознавания образов и цифровой обработки изображений путем применения спектральных преобразований Уолша в задачах подготовки данных для метрического обучения сиамских сверточных нейронных сетей.

Объект исследования – процесс автоматического совмещения разнородных изображений подстилающей поверхности.

Предмет исследования – методы и алгоритмы цифровой обработки и нейросетевого анализа разнородных изображений для определения местоположения ЛА. Предмет исследования определен предметной областью пунктов

4, 13 паспорта специальности 2.3.8. – Информатика и информационные процессы:

– разработка методов и технологий цифровой обработки аудиовизуальной информации с целью обнаружения закономерностей в данных, включая обработку текстовых и иных изображений, видеоконтента. Разработка методов и моделей распознавания, понимания и синтеза речи, принципов и методов извлечения требуемой информации из текстов. (п.4);

– разработка и применение методов распознавания образов, кластерного анализа, нейро-сетевых и нечетких технологий, решающих правил, мягких вычислений при анализе разнородной информации в базах данных. (п.13).

Личный вклад автора. Все выносимые на защиту положения получены автором лично.

Апробация работы. Основные результаты диссертации докладывались:

1. Международная научная конференция студентов, аспирантов и молодых ученых «Ломоносов-2025», Российская Федерация, Москва, 2025;

2. X международная конференция по когнитивной науке, Российская Федерация, Пятигорск, 2024;

3. III всероссийская научная конференция «Искусственный интеллект в автоматизированных системах управления и обработки данных (ИИАСУ'24)», Российская Федерация, Москва, 2024;

4. Международная научно-техническая конференция «Автоматизация» (RusAutoCon-2024), Российская Федерация, Сочи, 2024;

7. Конгресс «Интеллектуальные системы и информационные технологии» (IS&IT'22), Российская Федерация, Краснодарский край, с. Дивноморское, 2022;

8. III международный научный форум по компьютерным и энергетическим наукам (WFCES 2022), Казахстан, Алматы, 2022;

9. XXII международная научно-техническая конференция «Нейроинформатика-2020», Российская Федерация, Москва, 2020;

10. 13-я международная конференция ELEKTRO 2020, Италия, Таормина, 2020;

11. 8-я средиземноморская конференция «Встраиваемые вычислительные системы» (MECO-2019), Черногория, Будва, 2019;

12. Международная конференция «Proceedings of IEEE East-West Design & Test Symposium (EWDTS-2018)», Российская Федерация, Казань, 2018.

Основные результаты диссертации внедрены:

1. В компанию ООО «Квантрон Групп» в виде программного модуля, предназначенного для определения местоположения аэрофотосъемки.

2. В учебный процесс в курсе «Системы технического зрения», который изучается в рамках направления 09.04.01 Информатика и вычислительная техника (программа магистратуры) Федерального государственного бюджетного образовательного учреждения высшего образования «Рязанский государственный радиотехнический университет имени В. Ф. Уткина» и в электронный образовательный ресурс дистанционного учебного курса «Теоретические основы информатики» для направления 09.06.01 «Информатика и вычислительная техника» № 24967, дата регистрации 24.03.2022.

Публикации. По теме диссертации опубликована 21 работа. Из них 5 опубликованы в журналах перечня ВАК, 5 работ в международных журналах, рецензируемых в Scopus и Web of Sciences, 5 статей, опубликованных в других изданиях, 5 программы для ЭВМ, 1 патент на изобретение.

Методы исследования. При проведении исследований и разработке алгоритмического и программного обеспечения с целью решения поставленных в диссертационной работе задач использованы: методы вычислительной математики, методы линейной алгебры и математического анализа, цифровые методы обработки изображений и распознавания образов, методы математического моделирования, системного анализа, теории обработки изображений, методы обучения искусственных нейронных сетей.

Положения, выносимые на защиту:

1. Метод совмещения разнородных изображений подстилающей поверхности на базе сверточных нейронных сетей, позволяющий осуществлять первичное позиционирование ЛА в условиях отсутствия сигналов ГНСС, обеспечивающий существенное уменьшение объема хранимой информации об эталонных фрагментах в 80 раз (до 256 байт на фрагмент) и определяющий потенциальное положение ЛА с точностью до 300 пикселей.

2. Способ предварительной обработки изображений подстилающей поверхности на основе фильтрации низкочастотных секвент в пространстве Уолша, обеспечивающий повышение точности сходимости сверточной нейронной сети на 7,8 % за счет снижения влияния изменяющихся условий проведения съемки.

3. Алгоритм уточнения положения текущего изображения относительно спутникового изображения методом сопоставления ключевых точек, позволяющий компенсировать взаимные искажения и сократить конечную область совмещения до 20 пикселей (что соответствует 20 метрам на местности при высоте полета 300 м).

Достоверность полученных результатов и выводов подтверждается экспериментальными результатами, которые были получены при решении поставленных задач, а также их согласованностью с результатами работ, представленных в данной области.

Структура и объем работы. Диссертация состоит из введения, четырёх глав, заключения, списка используемых источников и приложения. Текст работы содержит 153 страницы, 56 рисунков и 23 таблицы. Список использованных источников включает 112 наименований (из них 54 зарубежных источника).

ГЛАВА 1 АНАЛИЗ ИЗВЕСТНЫХ МЕТОДОВ СОВМЕЩЕНИЯ ИЗОБРАЖЕНИЙ

В настоящее время для решения задачи определения местоположение летательного аппарата помимо совмещения разнородных изображений применяются ряд альтернативных подходов: астронавигация, зависящая от условий видимости небесных светил, геофизические методы навигации, навигация по рельефу и по изображениям местности.

Система AQNav, разработанная компанией SandboxAQ в 2024 году, использует алгоритмы искусственного интеллекта, квантовые сенсоры и магнитное поле земной коры для обеспечения геомагнитной (геофизической) навигации в реальном времени в отсутствие сигнала GPS [1]. Система AQNav является одним из перспективных решений автономной навигации. Согласно исследованиям авторов [2] точность данной системы от ~70 м до ~1.8 км. Однако AQNav требует дорогостоящих и пока не массовых квантовых сенсоров, в то время как предложенный метод основан на стандартном и недорогом оборудовании.

Алгоритм навигации по небесным светилам с использованием визуального анализа в условиях отсутствия систем глобальной навигации для ЛА [3] рассматривается вариант астрономической навигации. По данным исследования [3] точность навигации составляет порядка 4 км. Для разработанной системы требуется специализированная оптика для астронавигации, а также чистое звёздное небо. Астрономическая навигация не имеет возможности работать в условиях тумана, облачности, городской засветки и днем.

Программный комплекс «Навигация» [4] для автоматизации научных исследований в области создания перспективных систем навигации беспилотных летательных аппаратов по данным видеонаблюдения обеспечивает имитационное моделирование видеосъемки с борта ЛА, подготовку полетных заданий и исходных данных для проведения экспериментальных исследований. Комплекс «Навигация», решает задачу локального уточнения положения вблизи известной точки.

Системы и технологии определения местоположения без использования глобальных навигационных систем связи можно классифицировать:

1. Системы навигации по сигналам искусственных источников (приборный навигационный комплекс):

- инерциальные навигационные системы (ИНС) используют гироскопы и акселерометры для отслеживания перемещения от известной начальной точки;

- доплеровские измерители скорости и угла сноса (ДИСС) применяются чаще на вертолетах и самолетах, измеряют скорость, могут работать совместно с ИНС для коррекции накопленной ошибки;

- системы воздушных сигналов (СВС) измеряют истинную воздушную скорость и угол наклона.

2. Системы навигации по естественным полям Земли:

- радиотехнические систем:

- радионавигация использует наземные радиомаяки (например, система VOR (Very High Frequency Omnidirectional Range – высокочастотный радиомаяк) и DME (Distance Measuring Equipment – дальномерное оборудование) для гражданской авиации);

- спутниковая связь: определение местоположения через геолокацию сигналов спутников связи (например, Iridium);

- астронавигация использует звездные датчики для определения ориентации и, косвенно, положения, применяется для космических аппаратов и самолетов;

- магнитная навигация использует сравнение бортового магнитометра с магнитной картой местности;

- гравиметрическая навигация использует карты аномалий гравитационного поля Земли;

- барометрическая высотометрия осуществляет измерение высоты по атмосферному давлению.

3. Системы навигации по наблюдаемым объектам местности:
 - навигация по рельефу;
 - навигация по изображению;
4. Гибридные и комплексные системы [5-8].

Представлена классификация методов навигации в условиях отсутствия или ненадёжности глобальных навигационных систем связи. Для автономного позиционирования в условиях отсутствия сигналов GPS/ГЛОНАСС применяются альтернативные подходы: геофизические (AQNav), астрономические, инерциальные и визуальные методы. Визуальная навигация, основанная на сравнении бортовых снимков с эталонными картами, является наиболее перспективной для ЛА, так как позволяет использовать относительно недорогое оборудование (камеры) и не зависит от внешних сигналов или погодных условий (в отличие от астронавигации).

1.1 Корреляционный метод совмещения изображения

При решении задач совмещения изображений, особенно когда отсутствуют четкие контуры или однозначные ключевые точки, например, при наложении спутниковых снимков одной местности, полученных в разное время или разными датчиками, часто применяется корреляционный метод. Его основное преимущество – возможность работы непосредственно с яркостными характеристиками изображений без необходимости предварительно выделять характерные точки. Этот метод широко используется в системах технического зрения для точного позиционирования и отслеживания объектов.

Учет локальных сдвигов и поиск области максимального сходства реализуется при помощи функции взаимной корреляции. Наиболее эффективной по критерию точности является нормированная кросс-корреляция (Normalized Cross-Correlation, NCC), устойчивая к изменениям яркости и контраста [9].

Функция взаимной корреляции для изображений $f(x, y)$ (эталонное изображение) и $g(x, y)$ (текущего изображений) вычисляется следующим образом:

$$R(u, v) = \sum_x \sum_y f(x, y) \cdot g(x + u, y + v), \quad (1)$$

где (u, v) – вектор смещения, суммирование по области перекрытия.

Идея совмещения двух изображений с помощью корреляционного метода заключается в последовательном сканировании целевого изображения шаблоном на основе фрагмента опорного изображения по целевому изображению в заданной области поиска и вычислении в каждой позиции меры сходства (NCC – Normalized Cross-Correlation – нормализованная кросс-корреляция). Позиция с максимальным значением NCC считается наилучшим совпадением [10].

$$NCC(u, v) = \frac{\sum_x \sum_y [f(x, y) - \bar{f}] \cdot [g(x + u, y + v) - \bar{g}_{u,v}]}{\sqrt{\sum_x \sum_y [f(x, y) - \bar{f}]^2} \cdot \sqrt{\sum_x \sum_y [g(x + u, y + v) - \bar{g}_{u,v}]^2}}, \quad (2)$$

где $NCC(u, v)$ – значение меры сходства в позиции смещения (u, v) , $f(x, y)$ – интенсивность пикселя с координатами (x, y) в эталонном изображении (шаблоне), $g(x + u, y + v)$ – интенсивность пикселя с координатами $(x + u, y + v)$ в целевом изображении, куда сместили шаблон, \bar{f} – средняя яркость всего шаблона f . Вычисленная позиция является локальными координатами совмещения.

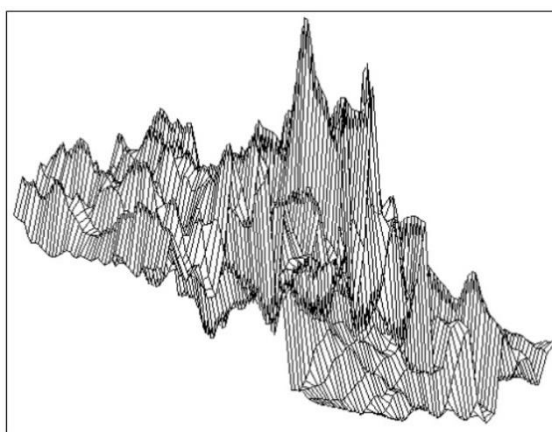


Рисунок 1 – Вид корреляционной экстремальной функции

1.2 Метод на основе дробно-линейного преобразования

При решении некоторых задач совмещения изображений, полученных при различных видах съемки (тепловизорами, телевизионными камерами), например, при проведении работ по реконструкции здания специалистам

удобно визуально определять положение скрытых в стенах объектов путем совмещения реальных изображений и чертежей (схем) о расположении электрической проводки, телефонных линий и т.п. В некоторых случаях необходимо совмещать реальные изображения одной и той же сцены, полученные с видеокамер, работающих в разных спектральных диапазонах. Это может потребоваться для определения положения теплых или нагретых тел, скрытых некоторым препятствием. Так же совмещение изображений применяется при разработке систем технического зрения [11].

Учет влияния проекционных искажений возможен с помощью матрицы гомографии. Матрица гомографии является невыраженной матрицей размерностью 3×3 и имеет следующий вид:

$$\mathbf{H} = \begin{pmatrix} h_{11} & h_{12} & h_{13} \\ h_{21} & h_{22} & h_{23} \\ h_{31} & h_{32} & 1 \end{pmatrix} \quad (3)$$

Элементы матрицы $h_{11}, h_{12}, h_{21}, h_{22}$ отвечают за вращение, элементы h_{31}, h_{32} - за сдвиг, h_{13}, h_{23} - за проективные преобразования, h_{11}, h_{22} - за масштаб.

Идея совмещения двух изображений с помощью проективного дробно-линейного преобразования одного изображения к плоскости другого заключается:

- в оценивании элементов матрицы гомографии \mathbf{H} ,
- в умножении матрицы \mathbf{H} на векторы $X_i = (x_i, y_i, 1)^T$, первыми двумя элементами которых являются координаты i -ой точки совмещаемого изображения [12].

Элемент h_{ij} матрицы гомографии находится путем решения следующего СЛАУ:

$$\mathbf{A}g_h = \mathbf{B} \quad (4)$$

где $\mathbf{A} = \begin{pmatrix} \mathbf{Z} & \mathbf{O} & \mathbf{W}_1 \\ \mathbf{O} & \mathbf{Z} & \mathbf{W}_2 \end{pmatrix}$ – основная матрица системы, $\mathbf{Z} = (x'_i \quad y'_i \quad 1)$ – матрица коэффициентов $i = \overline{1,4}$, $\mathbf{O} = (\mathbf{0})_{4 \times 3}$ – вектор свободных членов, \mathbf{W}_i – координаты то-

чек на первом изображении W_2 – координаты точек на втором изображении, основная матрица системы, размера 8×8 :

$$\mathbf{A} = \begin{pmatrix} x'_1 & y'_1 & 1 & 0 & 0 & 0 & -x_1x'_1 & -x_1y'_1 \\ x'_2 & y'_2 & 1 & 0 & 0 & 0 & -x_2x'_2 & -x_2y'_2 \\ x'_3 & y'_3 & 1 & 0 & 0 & 0 & -x_3x'_3 & -x_3y'_3 \\ x'_4 & y'_4 & 1 & 0 & 0 & 0 & -x_4x'_4 & -x_4y'_4 \\ 0 & 0 & 0 & x'_1 & y'_1 & 1 & -y_1x'_1 & -y_1y'_1 \\ 0 & 0 & 0 & x'_2 & y'_2 & 1 & -y_2x'_2 & -y_2y'_2 \\ 0 & 0 & 0 & x'_3 & y'_3 & 1 & -y_3x'_3 & -y_3y'_3 \\ 0 & 0 & 0 & x'_4 & y'_4 & 1 & -y_4x'_4 & -y_4y'_4 \end{pmatrix}, \quad (5)$$

где \mathbf{V} – вектор-столбец координат ключевых точек первого изображения $\mathbf{V} = (x_1, x_2, x_3, x_4, y_1, y_2, y_3, y_4)^T$, \mathbf{g}_h – вектор-столбец, подлежащих определению элементов матрицы гомографии $\mathbf{g}_h = (h_{11}, h_{12}, h_{13}, h_{21}, h_{22}, h_{23}, h_{31}, h_{32})^T$.

Сравнивая полученные результаты, можем наблюдать, что в первом случае, наложение не является гладким, что объясняется тем, что не все точки исходного изображения отображаются на результирующем. В результате операции округления пиксели могут накладываться друг на друга из-за чего образуются незаполненные области («пустоты») на результирующем изображении.

При использовании обратной матрицы, напротив, каждому пикселю второго изображения в обязательном порядке соответствует некоторый пиксель первого изображения, что приводит к более корректному совмещению изображений [13].



Рисунок 2 – Результат совмещения пары изображений

Выбор оптимального набора точек, который обеспечивает корректное определение элементов матрицы гомографии является одной из самых сложных задач данного метода, так как от их выбора зависит качество совмещения.

1.3 Метод совмещения на основе ключевых точек

В дополнении к ручному заданию контрольных точек, для автоматизации и повышения точности задач совмещения широко применяются методы, основанные на автоматическом обнаружении и сопоставлении ключевых точек. Изображение заменяется математической моделью и представляется набором ее ключевых точек. Ключевая точка – это точка на изображении, которая наиболее вероятно будет найдена на другом изображении этого же объекта [14]. Детектором называется метод извлечения ключевых точек на изображении. Детектор должен обнаруживать одни и те же ключевые точки инвариантно преобразований (углы поворота, смещение, растяжение/сжатие, изменение положения камеры). Детектор (detector – обнаруживающий) определяет только координаты ключевых точек, а так как на разных изображениях один и тот же объект имеет разные координаты, после применения детектора используют дескриптор. Дескриптор (descriptor – описывающий) – это идентификатор ключевой точки, позволяющий отличить ее от других ключевых точек. Дескрипторы также должны обеспечивать инвариантность к преобразованиям. Дескриптор представляет собой вектор признаков фиксированной размерности. Его формат можно классифицировать по способу кодирования информации о локальном окружении ключевой точки.

Существует два основных типа дескрипторов: вещественные и бинарные. Вещественные дескрипторы представляют собой векторы, элементами которых являются действительные числа, которые формируются на основе статистических характеристик региона, таких как гистограммы градиентов.

Наиболее распространенным подходом является разделение окрестности ключевой точки на блоки. Для каждого блока вычисляется гистограмма ори-

ентированных градиентов (Histogram of Oriented Gradients, HOG) [15]. Значения этих гистограмм преобразуются в единый вектор. Для сопоставления таких дескрипторов обычно используется евклидово расстояние (формула 11).

Бинарные дескрипторы формируются путем выполнения операции попарного сравнения интенсивности пикселя в окрестности ключевой точки. Преимущество данного формата представления заключается в высокой скорости сопоставления. Сравнение двух бинарных дескрипторов выполняется с помощью расстояния Хэмминга (формула 6).

$$d(A, B) = \sum_{i=1}^n A_i \oplus B_i, \quad (6)$$

где $d(A, B)$ – расстояние Хэмминга между векторами A и B , \oplus - операция исключающее ИЛИ.

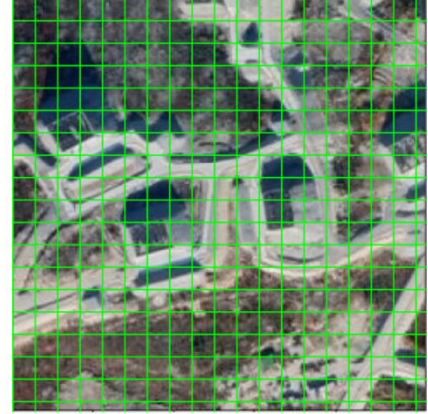
На рисунке 3 приведена визуализация метода гистограмм ориентированных градиентов. На рисунке 3,г цветные линии в ячейках показывают преобладающие направления градиентов. Цветовое кодирование направлений: красный – горизонтальные градиенты, синий – вертикальные градиенты, зеленый – диагональные направления.

Ключевые точки отвечают следующим свойствам:

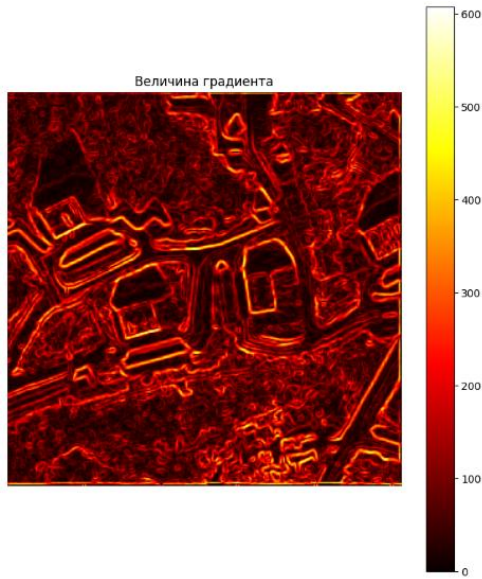
- определенность точки – точка должна выделяться среди соседних точек;
- устойчивость точки – изменение яркости, контрастности и цветовой гаммы не должны влиять на место особой точки на объекте или сцене;
- стабильность точки – зашумлённость, не превышающая определенный порог, не должна влиять на работу детектора;
- интерпретируемость точки – особые точки должны быть представлены в формате, пригодном для дальнейшей работы;
- количество точки – количество обнаруженных особых точек должно обеспечивать требуемому их количеству для обнаружения объектов.



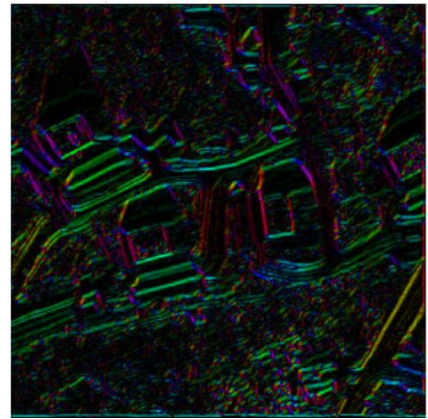
а



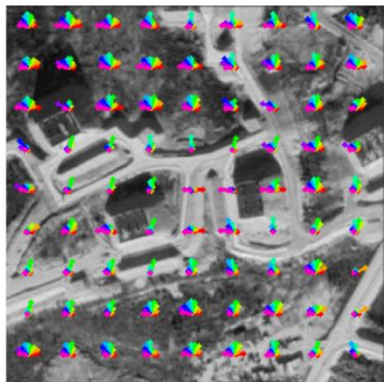
б



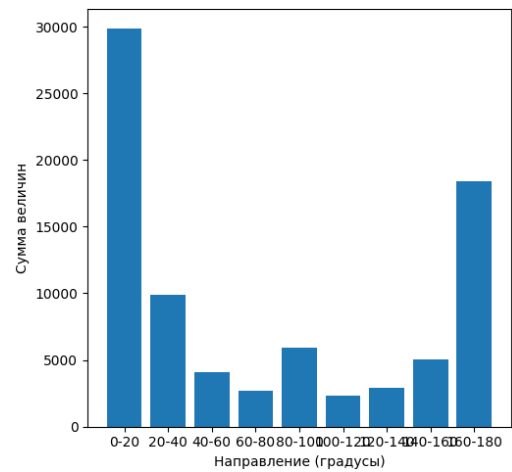
в



г



д



е

Рисунок 3– Визуализация метода гистограмм ориентированных градиентов
 а – исходное изображение в градации серого, б – разбиение на блоки,
 в – величина градиента, г – направление градиента, д – НОГ векторы направ-
 ления в блоках б, е – гистограмма градиентов для одного блока.

В качестве локальных особенностей на изображении выделяют:

– углы – особые точки, которые формируются из двух или более граней, которые обычно определяют границу между различными объектами и / или частями одного и того же объекта;

– края – множество точек на изображении, которые имеют высокую величину градиента;

– пятна – множество точек, имеющих одинаковую яркость [16, 17].

В зависимости от количества пересекаемых граней существуют различные виды углов: L, Y, T, X.

Критерии значимости точек:

– response – отклик детектора (чем выше, тем лучше точка);

– size – размер особенности (большие обычно устойчивее);

– distance – расстояние между дескрипторами (меньше = лучше совпадение);

– spatial distribution – равномерное распределение по изображению;

– octave – уровень пирамиды (более высокие уровни обычно устойчивее).

Ключевые точки – это локальные особенности изображения, которые обладают свойствами уникальности, устойчивости и повторяемости. Они соответствуют участкам с высокой информативностью, такими как углы, границы, текстуры или области с выраженными изменениями яркости или цвета. Ключевые точки описываются дескрипторами – числовыми векторами, которые кодируют их локальное окружение и позволяют сравнивать их между различными изображениями даже при изменениях масштаба, освещения, ракурса [18,19].

Классические локальные детекторы:

– SIFT (Scale-Invariant Feature Transform) – эталонный метод, отличающийся высокой точностью и инвариантностью к масштабу и повороту, но имеющий низкую скорость работы [20, 21]. На рисунках 4-11 приведены примеры сопоставления ключевых точек для различных дескрипторов.

– ORB (Oriented FAST and Rotated BRIEF) – быстрый бинарный детектор, обеспечивающий хороший компромисс между скоростью и точностью [22].

– AKAZE (Accelerated-KAZE) – современный метод, сочетающий устойчивость к нелинейным искажениям (как у KAZE) с повышенной скоростью работы [22].

– BRISK (Binary Robust Invariant Scalable Keypoints) – быстрый бинарный детектор, оптимизированный для скорости вычислений [22].

Нейросетевые методы:

– LOFTR (LoFTR: Detector-Free Local Feature Matching) – детектор-фри подход на основе трансформеров, который напрямую сопоставляет плотные признаки, демонстрируя высокую эффективность при слабой текстурности сцены, но требующий значительных вычислительных ресурсов [23].

– Deep Features – подход, использующий активации предобученной сверточной нейронной сети (СНС) (например, ResNet50) в качестве дескрипторов ключевых точек [24].

– TIMM Features – метод, основанный на использовании разнообразных предобученных архитектур из библиотеки PyTorch Image Models (timm), что позволяет гибко выбирать глубину и тип извлекаемых признаков [24].

– Multi-Detector (Multi-Detector Network) – нейросетевой метод, предназначенный для извлечения нескольких непересекающихся множеств ключевых точек на одном изображении. Данный метод делит все ключевые точки на заданное количество множеств. Сравнение дескрипторов выполняется в пределах одного множества [25].



а

б

Рисунок 4 – Пример сопоставления ключевых точек для локального дескриптора SIFT

а – снимок с ЛА, б – спутниковое изображение

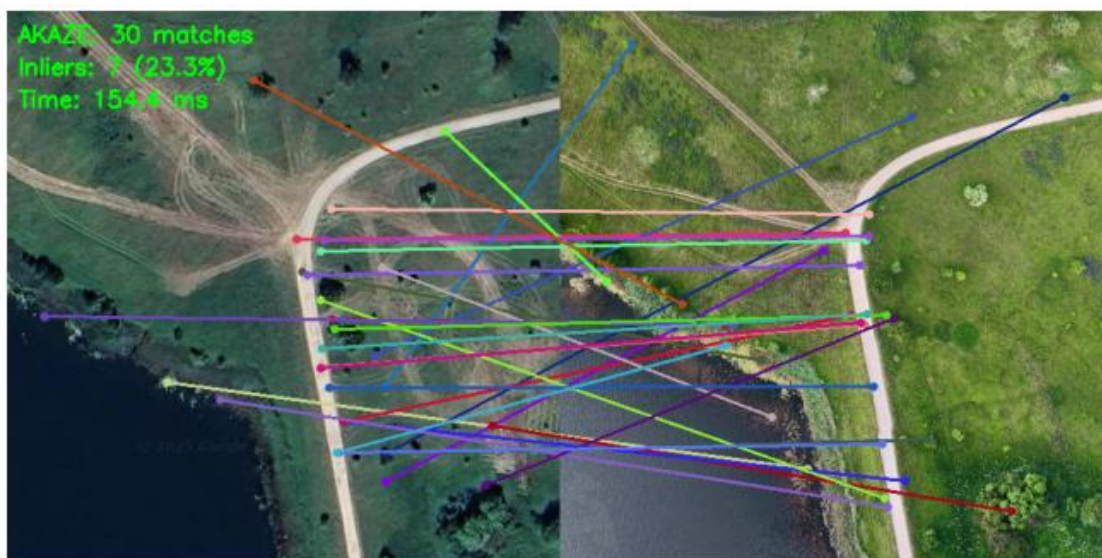


а

б

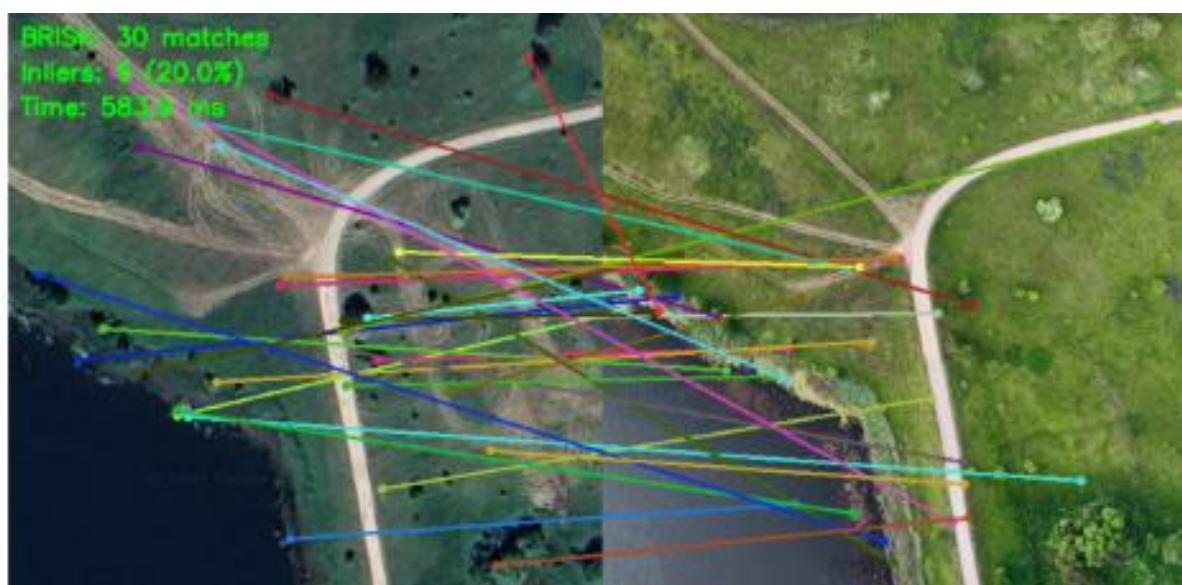
Рисунок 5 – Пример сопоставления ключевых точек для бинарного дескриптора ORB

а – снимок с ЛА, б – спутниковое изображение



а) б)

Рисунок 6 – Пример сопоставления ключевых точек для дескриптора AKAZE
а – снимок с ЛА, б – спутниковое изображение



а б

Рисунок 7 – Пример сопоставления ключевых точек для дескриптора BRISK
а – снимок с ЛА, б – спутниковое изображение

Нейросетевые методы детектирования, такие как LOFTR, Deep Features и TIMM Features, представляют собой современный подход, в котором СНС напрямую обучаются предсказывать положения ключевых точек и их дескрипторы [26].

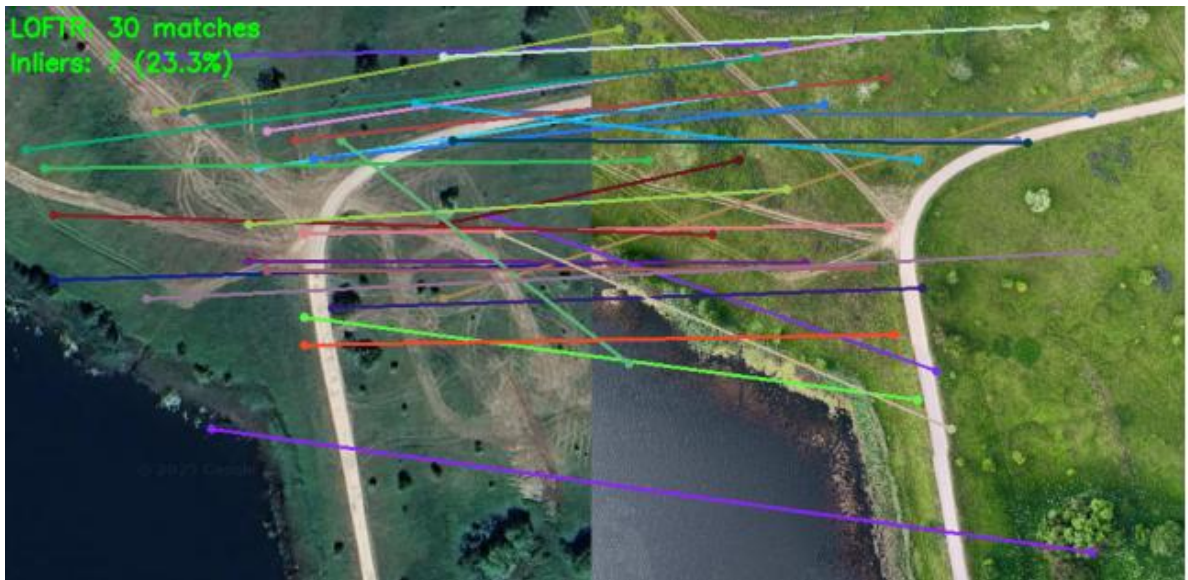
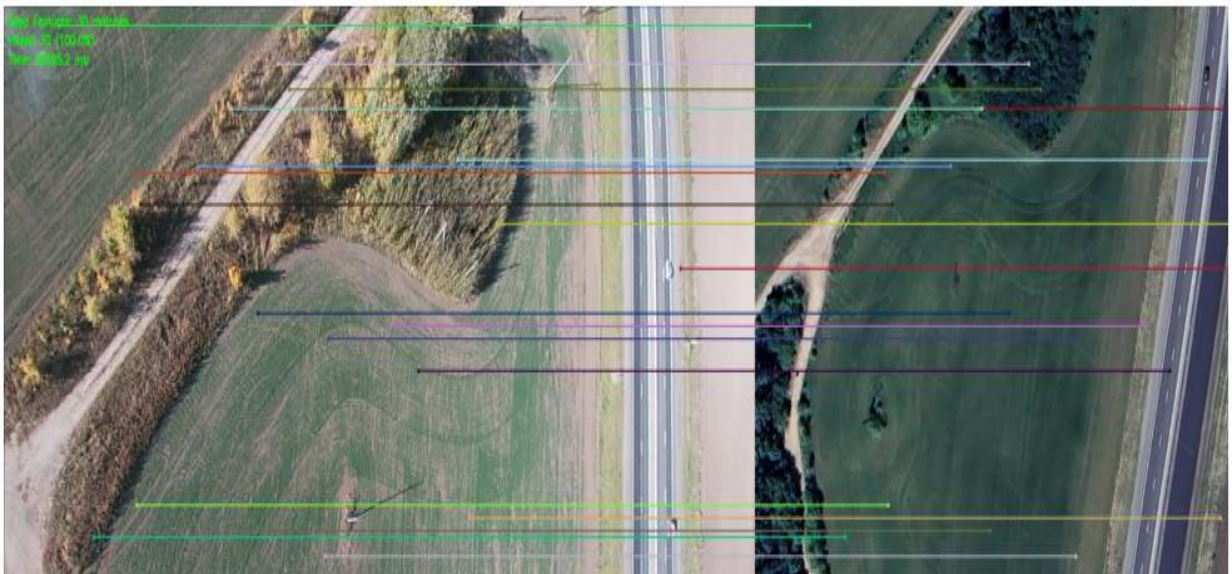


Рисунок 8 – Пример сопоставления ключевых точек для нейросетевого дескриптора LOFTR

а – снимок с ЛА, б – спутниковое изображение

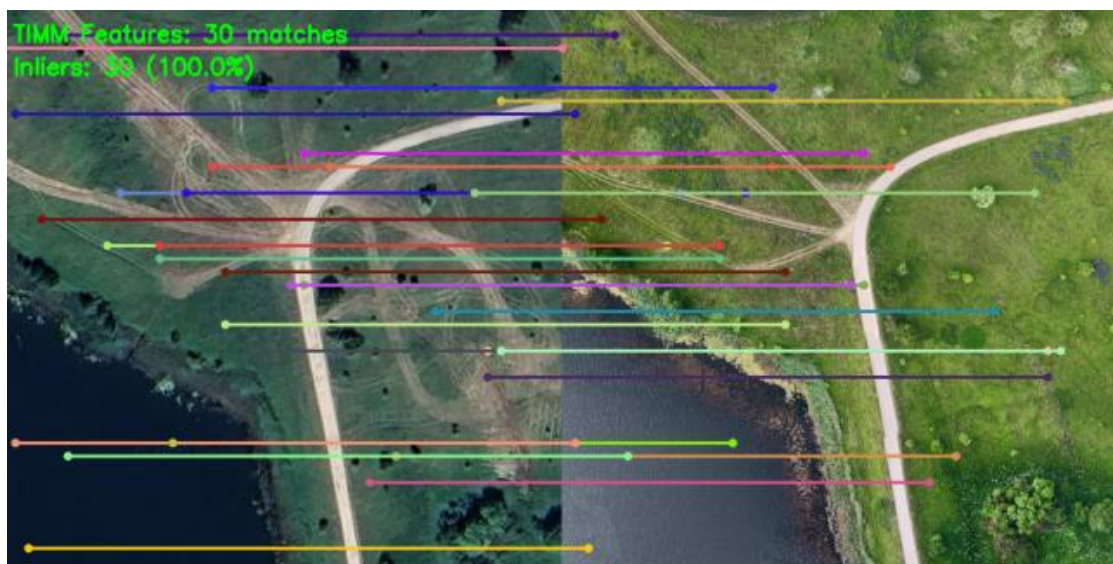


а

б

Рисунок 9 – Пример сопоставления ключевых точек для нейросетевого метода Deep Features

а – снимок с ЛА, б – спутниковое изображение



а б

Рисунок 10 – Пример сопоставления ключевых точек для нейросетевого дескриптора TIMM Features

а – снимок с ЛА, б – спутниковое изображение



а б

Рисунок 11 – Пример сопоставления ключевых точек для нейросетевого дескриптора Multi-Detector

а – снимок с ЛА, б – спутниковое изображение

Определение ключевых точек базируется на построении гауссианов и разностей гауссианов. Гауссианом или изображением размытым гауссовым фильтром является изображение:

$$L(x, y, \sigma) = G(x, y, \sigma) * I(x, y), \quad (7)$$

где L – значение гауссиана в точке с координатами (x, y) , σ — радиус размытия. G – гауссово ядро, I – значение исходного изображения, $*$ – операция свертки. Ядро G задается как:

$$G(x, y, \sigma) = \frac{1}{2\pi\sigma^2} e^{-\frac{x^2+y^2}{2\sigma^2}}. \quad (8)$$

Под разностью гауссианов понимают изображение, полученное путем по-пиксельного вычитания одного гауссиана исходного изображения из гауссиана с другим радиусом размытия.

$$\begin{aligned} D(x, y, \sigma) &= (G(x, y, k\sigma) - G(x, y, \sigma)) * I(x, y) = \\ &= L(x, y, k\sigma) - L(x, y, \sigma). \end{aligned} \quad (9)$$

Ключевые точки определяются как локальные экстремумы функции $D(x, y, \sigma)$. Каждая точка проверяется на устойчивость (контрастность и отношение главных кривизн) для отсева слабых и лежащих на контурах. Для каждой устойчивой ключевой точки в соответствующей ей гауссиане $L(x, y, \sigma)$ вычисляется градиент и его направление:

$$\begin{aligned} m(x, y) &= \sqrt{(L(x+1, y) - L(x-1, y))^2 + (L(x, y+1) - L(x, y-1))^2}, \\ \theta(x, y) &= \arctan(L(x, y+1) - L(x, y-1), L(x+1, y) - L(x-1, y)), \end{aligned} \quad (10)$$

где m – величина градиента, θ – направление градиента [27].

На основе ориентации строится ориентированная гистограмма градиентов в окрестности точки. Полученный 128-мерный вектор нормируется, что обеспечивает инвариантность к изменению освещенности и контраста.

Для дескрипторов ключевых точек эталонного f и текущего g изображений вычисляется евклидово расстояние. Пара считается соответствием, если отношение расстояний до ближайшего и следующего соседа меньше порога T (обычно $T \approx 0.7 - 0.8$) [20]:

$$\frac{d_{\text{ближайший}}}{d_{\text{второй ближайший}}} < T. \quad (11)$$

На множестве найденных пар соответствий с помощью алгоритма RAN-SAC (Random Sample Consensus) оцениваются параметры модели преобразования [27].

$$\begin{bmatrix} x' \\ y' \\ 1 \end{bmatrix} \sim \mathbf{H} \begin{bmatrix} x \\ y \\ 1 \end{bmatrix}, \quad (12)$$

где \mathbf{H} – матрица гомографии, (x, y) – координаты на эталонном изображении, (x', y') – на текущем изображении.

Найденная матрица \mathbf{H} применяется для проективного преобразования одного изображения в систему координат другого, что завершает процесс совмещения.

1.4 Нейросетевые методы совмещения изображений

С развитием моделей машинного обучения, таких как нейронные сети, и возрастанием вычислительной мощности, появляются новые подходы и решения для задач визуального поиска. Модели глубокого обучения позволяют более точно и эффективно сравнивать, и анализировать изображения [28].

Изначально [20,21] для представление визуальной информации использовались методы проектирования ручных признаков, такие как SIFT LOWE, затем [29] путем сжатия этих локальных признаков в единый глобальный дескриптор изображения, использовались различные технологии, такие как переход к глобальному описанию и путем извлечения глубоких представлений изображений из нейронных сетей [30]. Однако, как только два изображения описаны векторами признаков, визуальное сходство обычно измеряется путем вычисления стандартных метрик между ними. Несмотря на то, что стандартные метрики расстояния, такие как Евклидово расстояние или косинусное сходство быстрые и простые в реализации, они не принимают во внимание возможную взаимосвязь в наборе данных, что означает то, что даже если возникает сильная нелинейная зависимость данных в визуальной поисковой коллекции, они могли бы быть не в состоянии обнаружить это. Изучение оценки сходства непосредственно из визуальных данных может повысить производительность в задачах поиска изображений, при условии, что вероятные нелинейные зависимости внутри набора данных точно изучены функцией подобия [31].

Согласно некоторым исследованиям [32], стандартные метрические аксиомы не действительны для восприятия человеком визуального сходства и, следовательно, функции визуального сходства необязательно должны удовлетворять условиям метрик расстояния. Глубокое обучение на основе методов изучения подобия в основном сосредоточены на изучении оптимального отображения от пикселей до линейного пространства, в котором может быть применено Евклидово расстояние. Вместо этого [33] предлагаю подход, основанный на нейронных сетях, для изучения неметрической оценки сходства в пространстве признаков. Для оценки визуального сходства двух изображений на основе СНС [33] используется неметрическая оценка. На рисунке 12 представлена схема определения визуального сходства между двумя изображениями с использованием сиамских сетей [33].

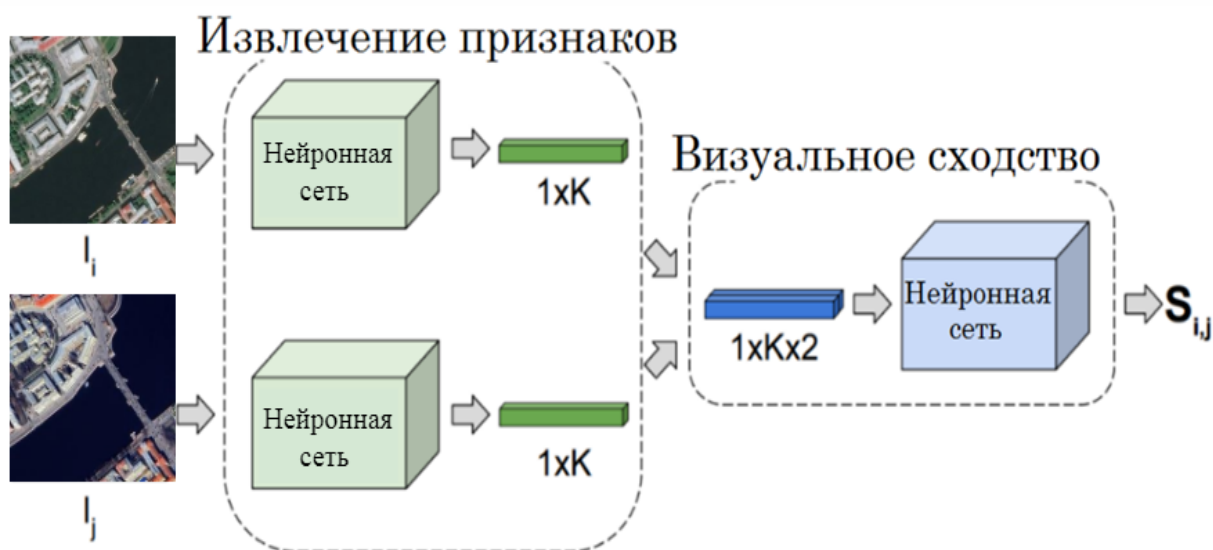


Рисунок 12 – Схема определения визуального сходства между изображениями

На вход СНС поступает пара изображений I_i, I_j , которая на последнем слое СНС преобразуется в вектор-признаков ($1 \times K$), далее эти вектора ($1 \times K \times 2$) передаются на блок оценки визуального сходства. Степень визуального сходства оценивается с использованием нейронной сети.

С помощью блока извлечения признаков нейросетевой алгоритм вычисляет визуальное представление изображения. С помощью блока вычисления

визуального сходства формируется числовая оценка показателя схожести, на основе нейронной сети.

Обучая модель глубокого обучения [33], можно оценить функцию визуального сходства, которая превосходит методы на основе стандартных метрических вычислений. Одна СНС извлекает представления изображений из входных изображений, а вторая – формирует показатель визуального сходства. Используется сямская архитектура нейронной сети, на вход которой подается два изображения. Выходная оценка сходства нейронной сети может быть непосредственно применена в качестве оценки сходства для ранжирования изображений в задаче их поиска. Экспериментальные результаты по стандартным наборам данных показывают, что нейронная сеть способна различать, когда пара изображений похожа или непохожа, и улучшить стандартную оценку метрики в дополнение к этому.

Так же известны подходы [34], которые используют методы машинного обучения для визуального поиска, например, применение сверточных автоэнкодеров, которое заключается в применении предобученных моделей сверточных нейронных сетей и библиотек. Согласно исследованиям авторов [34] недостатком применения предобученных моделей является то, что они могут быть обучены для одних целей и не подойдут при решении других. Использование сверточных автоэнкодеров требует точной настройки параметров для слоев и большое количество исходных данных, которое измеряется в миллионах фотографий. Подход [34] исправляет указанные недостатки за счет использования сямских сетей, которые на основе пары идентичных нейронных сетей формируют векторы-признаков. Далее вычисляется разница между векторами, что и является показателем сходства. Однако данный подход работоспособен только для простейших пар изображений (из набора MNIST) [35].

Поиск схожих изображений с использованием глубоких нейронных сетей применяют Яндекс и Гугл. Подход [36] описывает эффективные алгоритмы поиска по большим коллекциям изображений. В задачах поиска по изображениям авторы применяют нейросетевые дескрипторы с полносвязными слоями.

Качество нейросетевых дескрипторов повышают путем дообучения нейросети на поисковой коллекции, содержащей изображения семантически близкие изображениям из тестовой выборки. Также данные дескрипторы могут быть сжаты без потери качества, что делает их применимыми в поисковых коллекциях большого объема данных. Сопоставление схожести изображений осуществляется с использованием поиска ближайших соседей.

Классические методы подвыборки (операции уменьшения размерности) могут быть заменены [37-40] для сохранения большего объема информации. Интерполяционный пулинг [37] учитывает веса ближайших пикселей и направлен на минимизацию потерь важных признаков при уменьшении размерности. Вейвлет-пулинг [39, 40] сохраняет структурную информацию за счёт спектрального разложения. Однако он не адаптирован для сравнения изображений. Для уменьшения потери информации в СНС применяют вейвлет-преобразования [41-43]. Для восстановления изображений в нейронную сеть U-Net может быть встроено вейвлет-преобразование [41]. Так же вейвлет-преобразование используют для предобработки и ускорения СНС [42]. В [43] используется вейвлет-преобразование Хаара для шумоподавления и выделения признаков. Для анализа одномерных сигналов в нейросетях [44] применяется ряд Фурье.

Для улучшения сходимость СНС [41-43] используют спектральные преобразования. Данные работы направлены на сохранение информации при уменьшении размерности.

В [45] исследуется вейвлет-анализ для обработки изображений с ЛА, включая задачи навигации. Авторы комбинируют СНС с частотным разложением для улучшения выделения признаков. В [45] применяется бинарная классификация (обнаружение объектов).

Выбор архитектуры нейронной сети и количество характеристик в векторе обусловлен форматом входных данных и обучающей выборки. Поэтому для данной задачи были исследованы существующие модели такие как Xception (Extreme Inception), MobileNet v2, VGG16.

Архитектура Xception (Extreme Inception) представляет собой глубокую свёрточную нейронную сеть, развивающую идеи модуля Inception [46]. Исследовательский центр Google создал новый алгоритм свертки под названием Xception [47] для ускорения обработки. Его модель была адаптирована из InceptionV3, и вместо оригинального модуля Inception использовалась свертка с разделяемой глубиной для разделения обычной свертки на пространственную свертку и поточечную свертку. Её ключевая инновация – это глубоко разделяемые свёртки (depthwise separable convolutions), которые используются как основная строительная единица вместо классических свёрточных слоёв. Эта архитектура предлагает компромисс между высокой точностью и эффективным использованием параметров.

Основные характеристики Xception:

- глубоко разделяемые свёртки: каждая свёртка разделена на два этапа:
- пространственная свёртка (Depthwise Convolution): применяется по каждому входному каналу отдельно для захвата пространственных признаков.
- точечная свёртка (Pointwise Convolution, 1x1): проецирует каналы в новое пространство для комбинирования признаков. Этот подход значительно снижает вычислительную сложность и количество параметров при сохранении или даже улучшении эффективности.
- линейные остаточные связи (Linear Residual Connections): Xception построена как последовательность модулей с остаточными (residual) связями, что облегчает обучение очень глубоких сетей, решая научную задачу затухающего градиента. Связи, как правило, линейные (без активации между свёртками внутри блока).
- поток данных: входной поток данных сначала проходит через начальную серию обычных свёрточных слоёв, затем через несколько (обычно 14) основных блоков, построенных на глубоко разделяемых свёртках, и завершается полносвязными слоями классификации.

Применение в задаче поиска: вектор-признаков (эмбединг), извлечённый из предпоследних слоёв Xception, обладает высокой дискриминативной

способностью благодаря сложным комбинациям признаков, извлечённым разделяемыми свёртками. Это делает Xception мощным инструментом для создания компактных и информативных векторных представлений изображений.

На рисунке 13 приведена архитектура нейронной сети Xception.

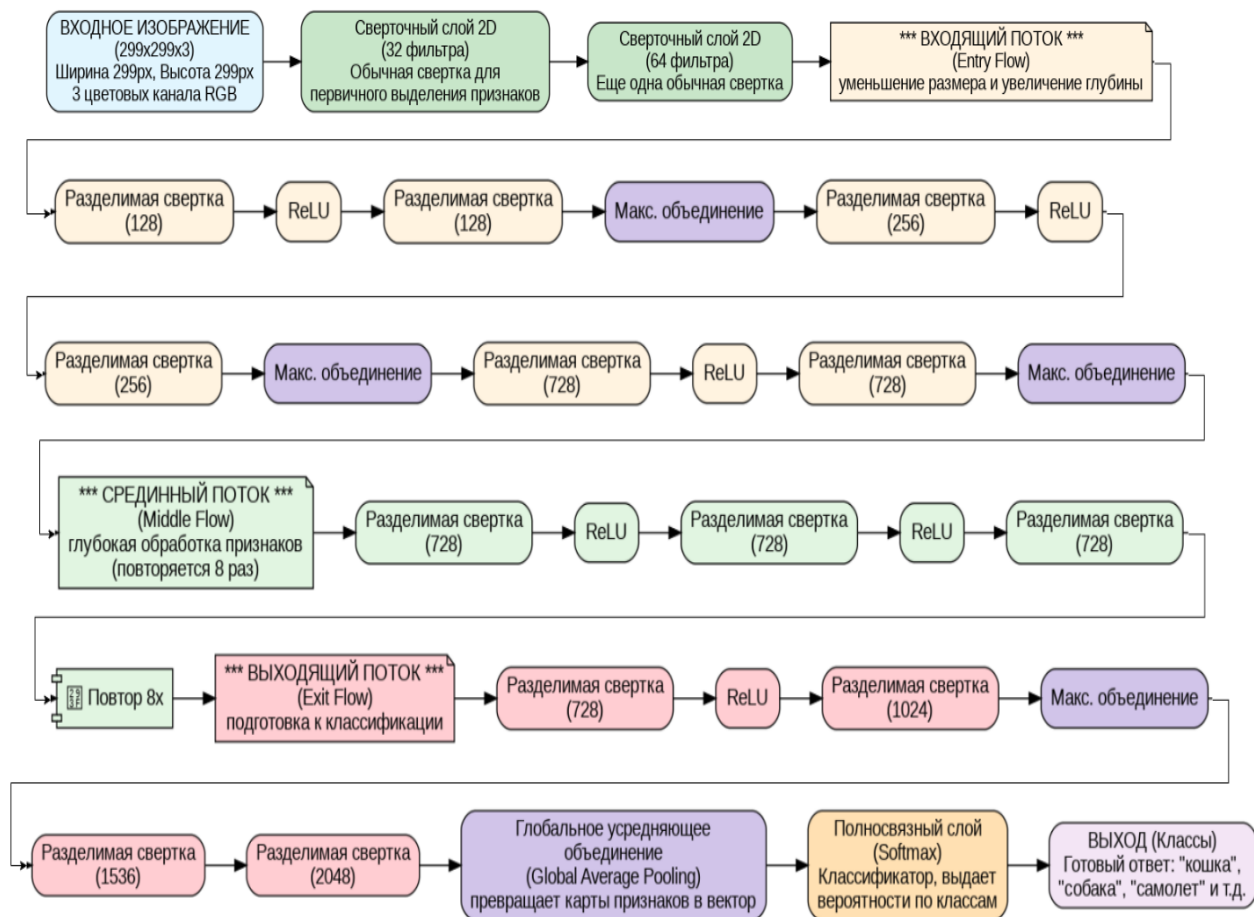


Рисунок 13 – Архитектура нейронной сети Xception

Архитектура MobileNet v2 – это архитектура, спроектированная специально для мобильных и встраиваемых устройств с ограниченными вычислительными ресурсами. Её цель – максимальное снижение количества операций и параметров при сохранении приемлемой точности, что делает её идеальной для систем реального времени [48].

Основные характеристики MobileNet v2:

- инвертированные остатки (Inverted Residuals): в отличие от классических остаточных блоков, которые сначала сжимают, а потом расширяют количество каналов, блок MobileNet v2 сначала расширяет количество каналов

(обычно в 6 раз) с помощью свёртки 1×1 , затем применяет глубоко разделяемую свёртку 3×3 для пространственной фильтрации, и в конце сжимает каналы обратно другой свёрткой 1×1 . Это позволяет сохранить больше информации в сжатом пространстве.

– линейное узкое место (Linear Bottleneck): в блоке MobileNet v2 не используется функция активации (ReLU) на выходе последней свёртки 1×1 (сжимающего слоя). Было обнаружено, что нелинейность разрушает информацию в низкоразмерном пространстве, поэтому здесь применяется линейная активация для сохранения важных признаков.

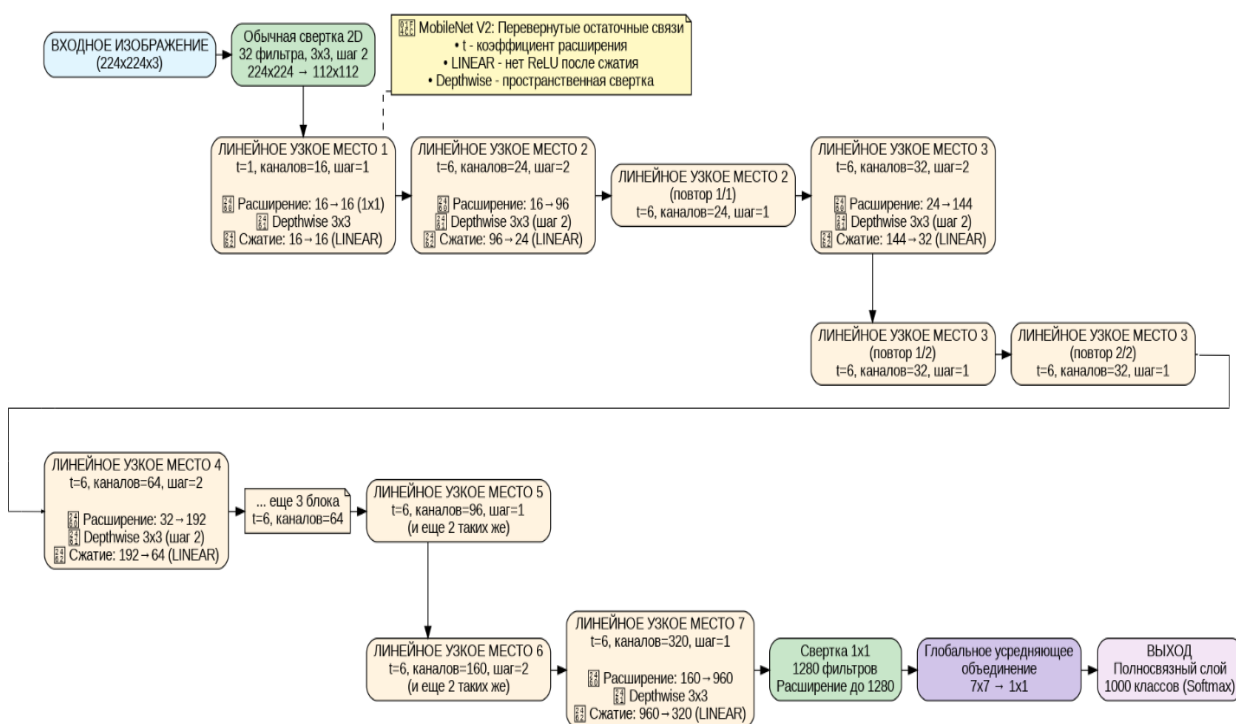


Рисунок 14 – Архитектура сверточной сети MobileNet

Архитектура СНС VGG16 (Visual Geometry Group) используется для задач классификации изображений, и может быть применен в задачах визуального поиска. На рисунке 15 приведена модель свёрточной нейронной сети, на рисунке 16 изображена ее архитектура.

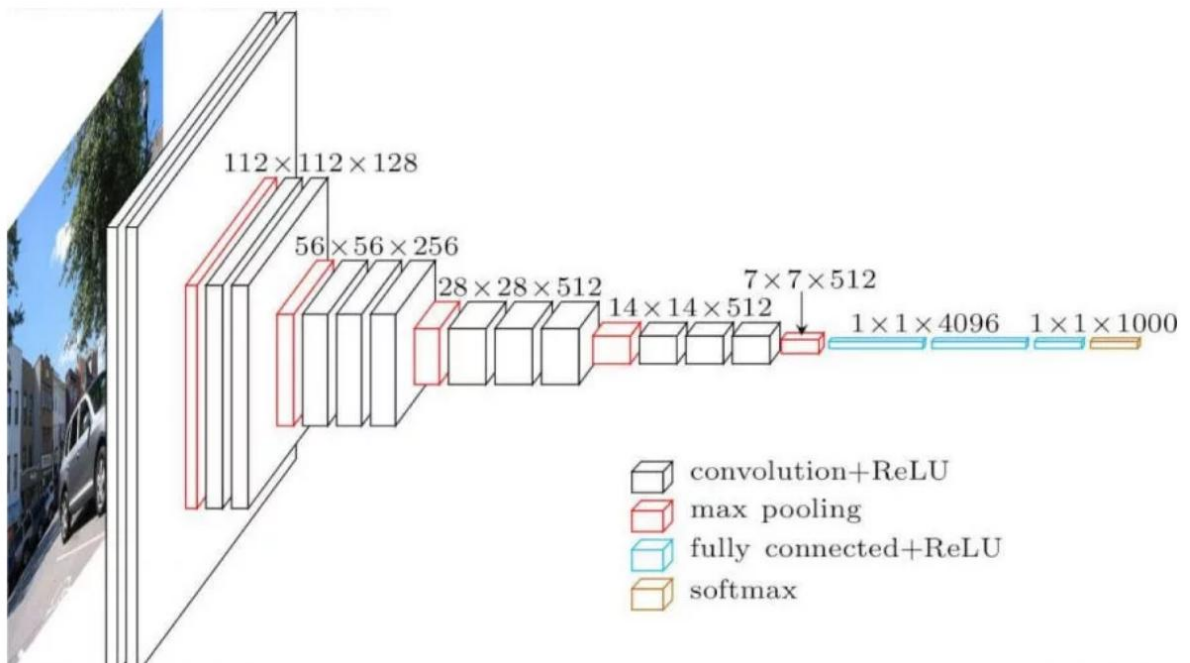


Рисунок 15 – Модель сверточной нейронной сети VGG16

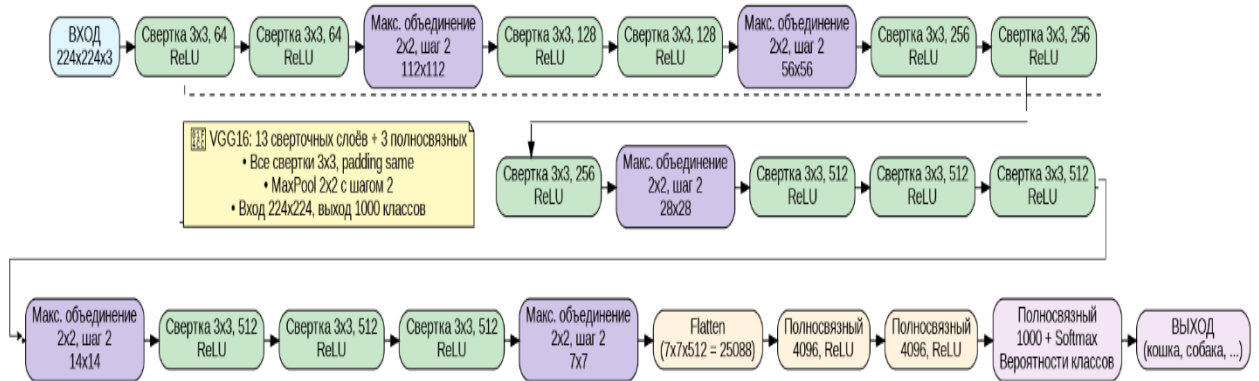


Рисунок 16 – Архитектура сверточной нейронной сети VGG-16

На рисунке 15 изображены блоками:

- convolution + ReLu – сверточный слой + активационная функция ReLu;
- max pooling – операция уменьшения размерности max pooling;
- fully connected + ReLu – полносвязный слой + активационная функция ReLu;
- softmax – финальный слой нейросети, функция, преобразующая логиты (числа) в вероятности.

$$\text{Softmax}(z_i) = \frac{e^{z_i}}{\sum_j e^{z_j}} \quad (13)$$

где z_i – логит (выходное значение для i -го класса), e^{z_i} – экспонента от логита.

Архитектура СНС VGG16 имеет следующие основные характеристики:

- глубина сети: 16 слоев, 13 из которых сверточные слои и 3 полносвязных;
- свёрточные слои: применяются слои с ядром свертки размером 3x3, позволяющие выделить детали на изображениях;
- слои уменьшения размерности: применяется Max Pooling размером 2x2 для извлечения наиболее значимых признаков;
- функция активации: применяется ReLU (Rectified Linear Unit), что помогает ускорить обучение и улучшить производительность сети.
- полносвязные слои: в данной сети используются три полносвязных слоя, где на последнем слое используется активационная функция softmax для классификации.
- параметры: около 138 миллионов параметров, из-за чего данная нейронная сеть является вычислительно сложной, но при этом она демонстрирует высокую точность в различных задачах [49, 50].

Слои организованы таким образом, что после свёрточных слоев идут слои уменьшения размерности (подвыборки).

Архитектуры нейронных сетей VGG16 и VGG19, хоть и принадлежат одному семейству, имеют принципиальные отличия в количестве слоев [51,52]. Три дополнительных свёрточных слоя в VGG19 интегрированы в конец третьего, четвертого и пятого блоков сети, позволяя модели извлекать иерархически более сложные и абстрактные признаки. VGG19 предпочтительна в задачах, где максимальное качество векторных представлений критично, а вычислительные возможности позволяют использовать более тяжёлую модель.

Косинусное сходство между векторами-признаков рассчитывается по формуле (14) и представляет собой меру, используемую для оценки схожести

между двумя векторами-признаками в пространстве, основанную на косинусе угла между ними. В задачах визуального поиска, данный вид сходство часто применяется [53]. Косинусное сходство между двумя векторами-признаков A и B можно вычислить по следующей формуле:

$$\text{cosine}_{\text{similar}}(A, B) = \frac{(A, B)}{\|A\| \cdot \|B\|}, \quad (14)$$

где (A, B) – скалярное произведение векторов-признаков A и B , $\|A\|$ – норма вектора-признаков A , вычисляемая как $\sqrt{\sum_{i=1}^m A_i^2}$, i – индекс суммирования, M – размер каждого из векторов-признаков A и B , $\|B\|$ – норма вектора-признаков B , вычисляемая аналогично.

Значение косинусного сходства изменяется от -1 до 1:

- в том случае, если векторы-признаки идентичны значение косинусного сходства 1 (максимальная схожесть);
- в том случае, если векторы-признаков ортогональны значение косинусного сходства 0 (нет схожести);
- в том случае, если векторы-признаков имеют противоположные направления значение косинусного сходства -1 (в контексте поиска похожих изображений это обычно не имеет смысла, так как векторы-признаков обычно не имеют отрицательных значений).

В задачах визуального поиска косинусное сходство применяют для сравнения векторных представлений изображений, полученных, например, с помощью свёрточных нейронных сетей [54]. Пользователем загружается изображение-запрос, на его основе формируется векторное представление, которое сравнивается с векторным представлением изображений из базы данных. Результатом данного сравнения является численная оценка сходства загруженного пользователем изображения и изображений из базы данных. Результаты сортируются по степени схожести. Это позволяет находить визуально похожие изображения на основе их признаков.

Евклидово расстояние представляет собой меру, позволяющую вычислить «дистанцию» между двумя точками в пространстве. В задачах визуального поиска и анализа изображений, евклидово расстояние может применяться при сравнении векторных представлений изображений [55,56].

Евклидово расстояние вычисляется по формуле (15) и может применяться для измерения различия между векторными представлениями изображений. Чем меньше Евклидово расстояние между векторами двух изображений, тем более похожими они считаются.

$$d_{euclid}(A, B) = \sqrt{\sum_{i=1}^m (A_i - B_i)^2} \quad (15)$$

где A_i и B_i – компоненты векторов-признаков A и B соответственно, m – размерность векторов-признаков.

В задачах визуального поиска также применяется Манхэттенское расстояние, которое вычисляет сумму абсолютных разностей между компонентами векторов. В отличие от евклидовой метрики, возводящей разности в квадрат, манхэттенское расстояние менее чувствительно к выбросам и может быть более устойчивым при работе с разреженными векторными представлениями [57]. Манхэттенское расстояние между векторами A и B рассчитывается по следующей формуле:

$$d_{\text{манхэттен}}(A, B) = \sum_{i=1}^m |A_i - B_i| \quad (16)$$

Выбор метрики оценки расстояния между векторами-признаков обуславливается тем, что косинусное сходство фокусируется на направлении векторов, что позволяет лучше оценивать, насколько пара изображений визуально схожи. Выбор косинусного сходства в задачах поиска по содержанию обусловлен его способностью более точно отражать семантическое сходство между векторами.

1.4.1 Определение визуального сходства

Визуальный поиск (поиск по заданному сюжету, контекстный поиск) относят к области информационного поиска и предназначен для обнаружения идентичных по содержанию объектов на разных изображениях. Поиск основан на алгоритмах компьютерного зрения и позволяет анализировать объекты на изображении. Методы визуального поиска позволяют без использования концепции индексирования, метода полного перебора или просмотра миниатюр, осуществлять обнаружение схожих по содержанию изображений на основании пользовательского запроса. Данные методы, как правило, учитывают только графическое отображение изображения, без использования ключевых слов, метаданных и прочего.

На рисунке 17 представлена схема работы систем визуального поиска [58].

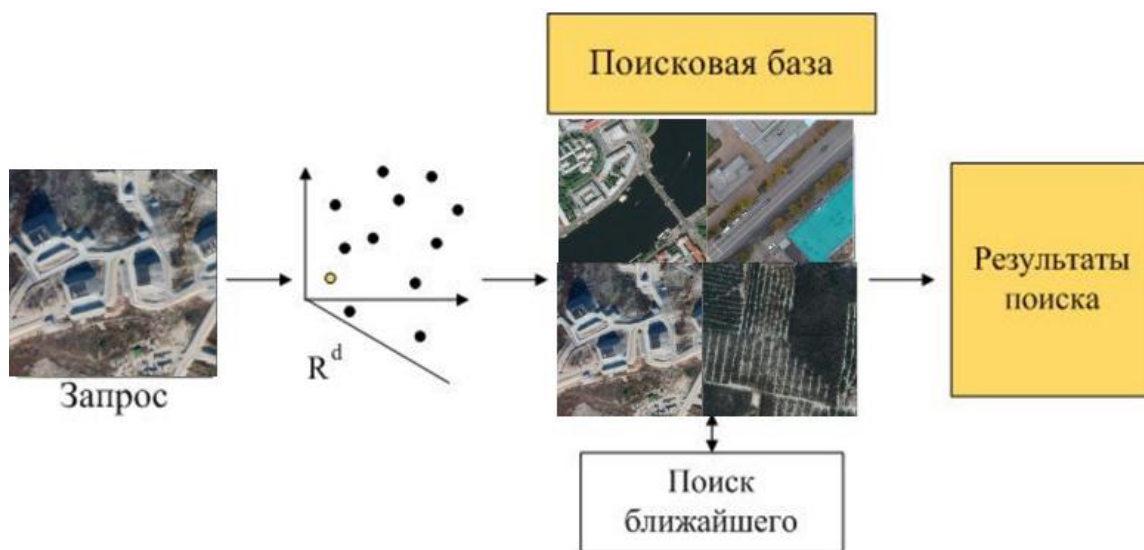


Рисунок 17 – Схема работы систем визуального поиска

Пользователем загружается изображение-запрос, то есть заданный сюжет, на основании которого будет осуществлен поиск. Исходное изображение преобразуют к набору метрик (в случае использования метрической оценки сходства) [59]. Для корректного поиска сформированные метрики R^d должны быть инвариантны к аффинным преобразованиям. Характеристики вычислен-

ные для изображений формируются в виде вектора фиксированной длины. Далее система находит изображения из поисковой базы данных, применяя следующий подход: используя Евклидово или другое расстояние, определяются такие вектора, которые являются ближайшими к вектору изображения. То есть вычисляется разность векторов. В результате поисковая система ранжирует изображения в соответствии с тем, насколько они похожи по отношению к входным данным.

Визуальное сходство – это количественная мера, отражающая степень схожести пары изображений на основе визуального содержания. В задаче совмещения разнородных изображений визуальное сходство используется для оценки географической близости текущего аэрофотоснимка с ЛА (I_k) и изображения из базы данных (I_s).

Определение визуального сходства – это задача, которая определяет, насколько идентична пара изображений, на основе визуального анализа содержания. Тогда определим $\tau_{k,s} = \tau(I_k, I_s)$ как показатель сходства, который определяет, насколько похожи пара изображений $\tau_{k,s} = \tau(I_k, I_s)$ как показатель сходства. Даны m -изображений из поисковой базы данных $G = \{I_1, I_2, \dots, I_M\}$. Каждое изображение $I_j, j = \overline{1, M}$, характеризуется вектором числовых значений признаков $\Phi(I) = \{\varphi_1(I), \varphi_2(I), \dots, \varphi_N(I)\}$, где N – мощность пространства признаков R^N , полученных с предпоследнего слоя сверточной нейронной сети. Решаемая задача состоит в определении визуального сходства τ , которое определяет значение показателя сходства, основываясь на векторах числовых значений признаков, как:

$$\tau_{k,s} \geq \tau_{min} , \quad (17)$$

где τ_{min} – пороговое значение сходства.

Решение данной задачи состоит из следующих этапов:

- формирование вектора-признаков изображения, которые определяют наиболее заметную визуальную информацию;

– численная оценка визуального сходства (мера близости) на основе сформированных векторов [60].

Изначально системы визуального поиска использовали следующие метрики: информацию о среднем цвете, гистограмме цветов, разбиение на классы, описание текстур [61]. Однако, простейшие метрики не обеспечивали необходимую инвариантность к изменениям освещенности, поворотам, изменениям масштаба, растяжения/сжатия. Глобальные дескрипторы, например, SIFT (scale-invariant feature transform), обладали некоторыми типами инвариантности, но имели высокую размерность, что делает невозможным их применение для больших по объему коллекций изображений. Требовалось применять сжатия глобальных дескрипторов [62].

Метрики расстояния необходимо сжимать до компактного вектора-признаков для быстрого вычисления расстояния между метрикой изображения-запроса и сжатой метрикой изображения из эталонной базы данных.

Поиск ближайших осуществляется на основе вычисленных метрик (в случае использования метрической оценки сходства). Самыми известными подходами для решения этой задачи являются разделяющие деревья поиска, иерархический метод k-means, локально-чувствительное хэширование, метрические свойства гауссовых смесей, множественный анализ классов с использованием расстояний хи-квадрат [63-65].

Существует несколько подходов определения визуально схожих пар изображений:

– на основе попиксельного сравнения;

$$D = \sum_{(x,y)} |(I_1(x, y) - I_2(x, y))| \quad (18)$$

где D – разница между двумя изображениями, I_1, I_2 – сопоставляемые изображения, x, y – значение координат.

– с использованием среднеквадратического отклонения (СКО):

$$D = \sqrt{\sum_{(x,y)} (I_1(x,y) - I_2(x,y))^2} \quad (19)$$

- с использованием экспертных систем [66];
- с использованием компьютерного зрения и машинного обучения [67-71].

Первые два подхода определяют идентичные изображения, то есть имеют одинаковую сцену и их визуальное различие для пользователя не заметно. Однако, если изображение повернуто на определенный угол, снято в другое время суток, то используются экспертные системы или методы на основе машинного обучения и алгоритмов компьютерного зрения. Экспертные системы могут использовать предварительно заданные правила для определения сходства между изображениями, учитывая такие факторы, как цвет, текстура, количество классов и другие характеристики объектов и фона на изображении.

Другой подход – это использование алгоритмов компьютерного зрения и машинного обучения для извлечения визуальных признаков и их сравнения. Эти методы могут включать в себя анализ контуров, цветовых гистограмм, текстурных признаков, а также глубину и перспективу объектов на изображении. Поиск по образцу может быть реализован с использованием разных технологий, таких как методы сопоставления шаблонов, дескрипторы особых точек или нейронные сети, комбинирование которых может обеспечить более точный и надежный поиск схожих по содержанию изображений [72].

Современные поисковые системы работают на основе облака тегов – текстовых меток, которые формируются для изображения на основе текстового описания изображения.

1.5 Выводы по первой главе

Рассмотренные методы совмещения разнородных изображений обладают существенными ограничениями, такими как высокая чувствительность к геометрическим искажениям, низкая устойчивость на зашумленных и динамиче-

ски изменяющихся сценах, чрезмерная вычислительная сложность для корреляционного метода совмещения, необходимость ручной или сложной предобработки для выделения опорных точек для совмещения разнородных изображений на основе дробно-линейного преобразования, для нейросетевых методов – зависимость от специфики обучающих данных и риск переобучения, затрудняющими их применение для решения задачи. Таким образом, существует потребность в разработке специализированного метода, сочетающего точность, устойчивость к разнородности данных и приемлемую вычислительную сложность.

Кроме того, рассмотренные методы совмещения изображений обладают существенными ограничениями, затрудняющими их применение в задачах определения местоположения ЛА. Корреляционный метод является вычислительно сложным для обработки изображений высокого разрешения в реальном времени и чрезмерно чувствителен к изменениям ракурса, масштаба и нелинейных фотометрических искажений между разнородными снимками. Метод на основе дробно-линейного преобразования требует ручного или высокоточного автоматического выбора четырёх пар опорных точек, что не обеспечивает необходимой автономности и устойчивости в реальных условиях. Методы на основе ключевых точек (SIFT, ORB и др.) часто оказываются неэффективными на слаботекстурированных или зашумлённых снимках подстилающей поверхности.

Нейросетевые подходы основаны на определении визуального сходства изображений и, демонстрируя высокий потенциал, склонны к переобучению на конкретных наборах данных и требуют значительных вычислительных ресурсов. Таким образом, существует потребность в разработке специализированного метода, сочетающего точность, устойчивость к разнородности данных и приемлемую вычислительную сложность.

ГЛАВА 2 МЕТОД СОВМЕЩЕНИЯ РАЗНОРОДНЫХ ИЗОБРАЖЕНИЙ ПОДСТИЛАЮЩЕЙ ПОВЕРХНОСТИ

2.1 Метод совмещения разнородных изображений подстилающей поверхности и алгоритмы его реализующие

Для решения задачи повышения эффективности поиска разнородных изображений в цифровых базах данных, соответствующих заданному изображению подстилающей поверхности, был разработан метод совмещения разнородных изображений. Данный метод позволяет совместить изображение, полученное в текущий момент времени с ЛА и изображение из базы данных спутниковых снимков. Формирование базы данных подстилающей поверхности выполняется предварительно. Спутниковое изображение имеет высокое разрешение и предварительно разбивается на фрагменты (тайлы) и записывается в БД. Тайл – это геопривязанный фрагмент карты, используемый как эталон для сравнения с текущим кадром подстилающей поверхности. Ограничивающим фактором разработанного метода является то, что совмещать надо изображения, снятые в одно и то же время года и время суток, точность совмещения пары изображений, полученных в межсезонье, не рассматривалась.

Метод совмещения разнородных изображений подстилающей поверхности состоит из комбинации алгоритмов: алгоритма обучения СНС, алгоритма формирования вектора-признаков, предназначенного для извлечения визуальной информации из изображения и алгоритма поиска изображения, который, используя сформированный на предыдущем этапе вектор-признаков изображения, с помощью косинусного сходства обнаруживает изображение из БД со схожим содержанием. Разработанный метод позволяет определить потенциальное положение ЛА на основе совмещения пары разнородных изображений [73].

Основные составляющие метода совмещения разнородных изображений представлены диаграммой потоков данных, изображенной на рисунке 18.

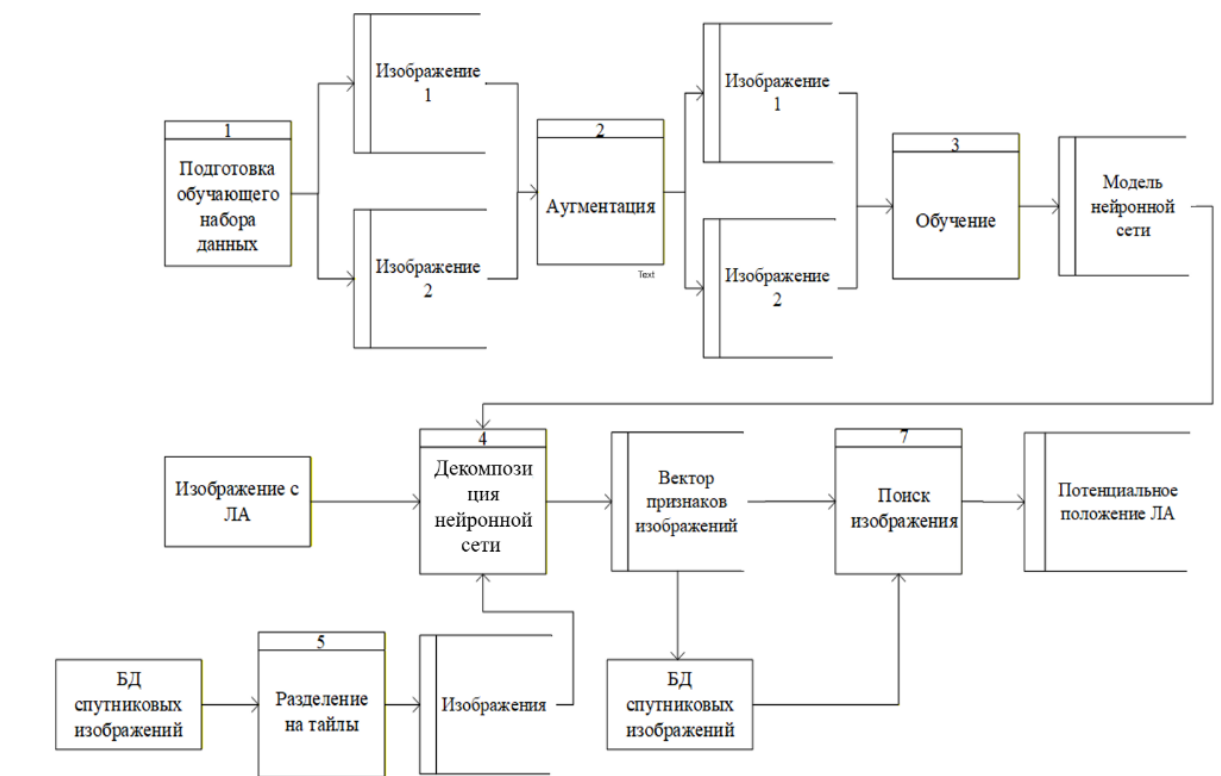


Рисунок 18 – Диаграмма потоков данных предложенного метода совмещения разнородных изображений подстилающей поверхности

Для того, чтобы определить потенциальное положение ЛА с помощью совмещения разнородных изображений, необходимо с помощью СНС преобразовать изображения в вектор-признаков фиксированной длины и осуществить поиск изображений с помощью косинусного сходства [74]. Выбирается изображение с косинусным сходством ≈ 1 .

Для преобразования изображений в компактные вектора-признаки используется сиамская архитектура, состоящая из двух идентичных ветвей с общими весами. Каждая ветвь после декомпозиции может рассматриваться как самостоятельная нейронная сеть и представляет собой свёрточную нейронную сеть, которая принимает на вход предобработанное изображение размером 256×256 и выдаёт вектор признаков размером 256 [75]. В результате на целевом устройстве нет необходимости хранить исходные изображения (средний объём которых составляет 20 КБ). Вместо них храниться лишь вектор признаков.

2.1.1 Алгоритм обучения свёрточной нейронной сети

Отдельную задачу представляет обучение данной нейронной сети. Для ее решения нейронная сеть дополнена парной, и в качестве последних слоев добавлена функция вычисления расстояния, чтобы в качестве выходной характеристики формировалась разница между парой изображений.

Сиамская архитектура – это особый тип нейронной сети, предназначенный для сравнения двух входных объектов (например, изображений) и определения степени их сходства. Данная архитектура применяется в задачах, где необходимо измерить сходство или расстояние между объектами [76, 77].

Все пары изображений, используемые в процессе обучения сиамских нейронных сетей, можно разделить:

1. По критерию классификации:

- простая пара: расстояние явно меньше или явно больше порога при сравнении объектов одного или разных классов;

- сложная пара: расстояние «на грани», вводящее сеть в заблуждение (одинаковые классы далеки, разные — близки);

2. С точки зрения цели обучения:

- простая пара: подтверждает текущие веса;

- сложная пара: полезнее для обучения, «заставляет» сеть уточнять признаки.

3. В контексте потерь:

- простая пара: даёт малый вклад в функцию потерь;

- сложная пара: даёт основной вклад в градиент, «подталкивая» веса к изменениям.

Таким образом выделяют несколько классов для обучения пар изображений: простые позитивные пары, простые негативные пары, сложные позитивные пары, сложные негативные пары [78].

Под простой парой будем понимать два изображения, которые подаются на обучения и:

- правильно классифицируются текущей моделью;
- не требуют дообучения модели, так как ошибка на таких примеров $\rightarrow 0$.

Под сложной парой будем понимать два изображения, которые подаются на обучение и в процессе обучения:

- трудно классифицируются текущей моделью;
- ошибочно принимаются моделью за похожие, хотя имеют принадлежность к разным классам;
- ошибочно принимаются моделью за различные, хотя принадлежат одному классу.

Для обучения сиамской нейронной сети использовались сложные пары с целью корректного выделения всех возможных значимых характеристик [79].

Пусть даны:

- сиамская сеть с функцией расстояния $D(x_i, x_j)$;
- порог τ , определяющий, считаются ли пара изображений визуально похожими ($D < \tau$) или разными ($D \geq \tau$).

Тогда:

- простая позитивная пара:

$$\begin{aligned} D(x_i, x_j) &\ll \tau; \\ y_i &= y_j. \end{aligned} \tag{20}$$

- простая негативная пара:

$$\begin{aligned} D(x_i, x_j) &\gg \tau; \\ y_i &\neq y_j. \end{aligned} \tag{21}$$

- сложная негативная пара:

$$\begin{aligned} D(x_i, x_j) &\approx \tau; \\ y_i &\neq y_j. \end{aligned} \tag{22}$$

- сложная позитивная пара:

$$\begin{aligned} D(x_i, x_j) &\gg 0; \\ y_i &= y_j. \end{aligned} \tag{23}$$

где y_i – метка (класс) объекта x_i , y_j – метка (класс) объекта x_j / [80].

Для обучения нейронной сети необходимо использовать: простые и сложные, негативные и позитивные пары. Простые негативные пары в обучении применять не целесообразно, так как это вносят дисбаланс в обучение (в реальных данных простых негативных пар значительно больше, чем сложных), что приведет к тому, что модель будет перегружена тривиальными примерами, из-за этого сложные пары окажутся в меньшинстве и модель не научится правильно их обрабатывать. Задачей метрического обучения является научить модель различать похожие, но разные объекты (сложная негативная пара) и объединять разные, но одинаковые по классу (сложная позитивная пара) [81,82].

На рисунке 19 приведен пример сложной позитивной пары, а – фрагменты со съёмки на малой высоте, б – спутниковое изображение.

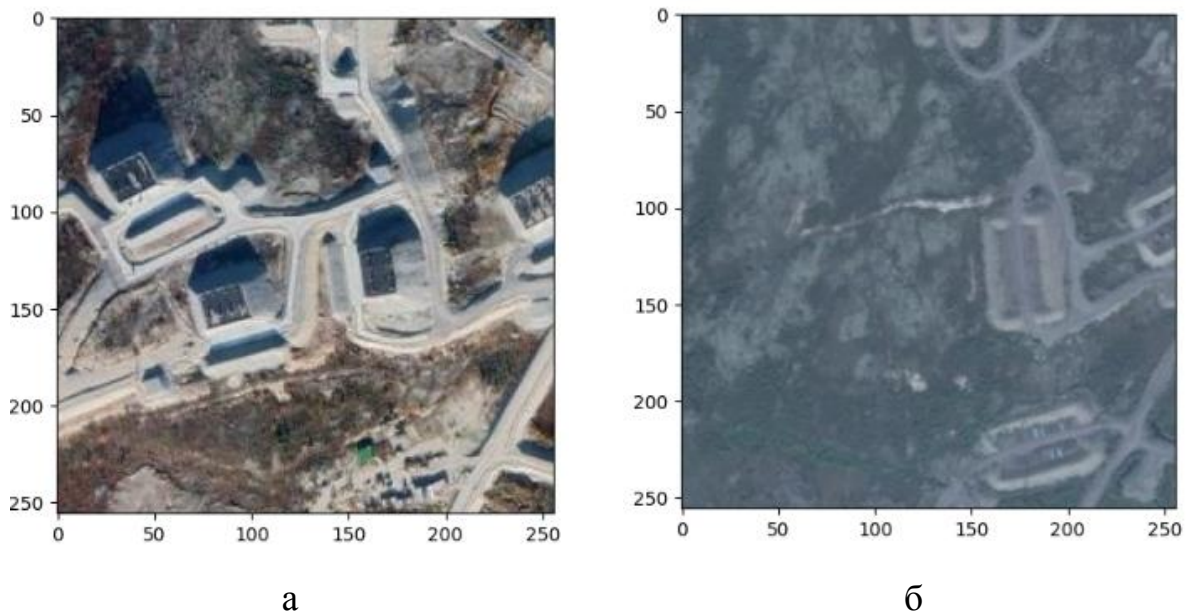


Рисунок 19 – Пример сложной позитивной пары изображений
а – фрагменты со съёмки на малой высоте, б – спутниковое изображение

Формирование данных для СНС может быть выполнено следующим образом:

1. Аугментирование данных – процесс создания новых обучающих данных путем искусственного видоизменения уже существующих изображений, применяется с целью увеличения разнообразия обучающей выборки, чтобы

нейронная сеть была устойчивой к помехам и изменениям условий съемки и не переобучалась [83]. Аугментирование состоит из следующих шагов:

Шаг 1: Нормализация в диапазоне $[0, 1]$.

Нормализация приводит значения пикселей от 0 до 255 в дробные числа от 0.0 до 1.0. Это стандарт для нейронных сетей, так как модели эффективнее работают с маленькими значениями.

Шаг 2: Применение аугментации:

- изменение контраста в обучающей выборке;
- изменение яркости в обучающей выборке;
- добавление Гауссова шума для изображений из обучающей выборке.

Шаг 3: Возврат в исходный формат. На рисунке 20 приведены примеры изображений после аугментации.

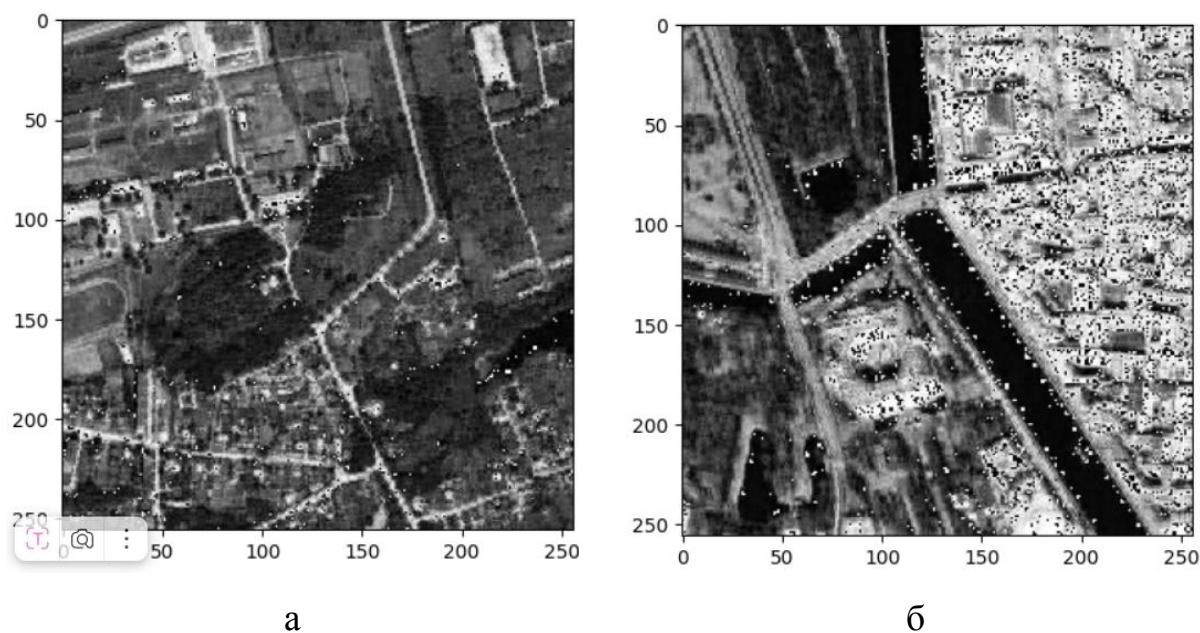


Рисунок 20 – Пара изображений после аугментирования данных

2. Поворот и масштабирование изображения – геометрическая трансформация изображений, изменяющие его пространственную ориентацию и размер, применяемая с целью привести все изображения к одному размеру и одновременно создать дополнительные вариации объектов для повышения устойчивости модели к разным ракурсам съемки. Данный этап состоит из следующих шагов:

Шаг 1: Поворот изображения на заданный угол (в градусах).

Изображение вращается вокруг своего центра на указанный (случайный) угол. При повороте углы изображения могут выходить за границы исходного кадра, поэтому пустые области обычно заполняются нулями.

Шаг 2: Масштабирование изображения с сохранением центра.

Масштабирование изображений с сохранением пропорций, то есть используется одинаковый коэффициент по ширине и высоте.

Шаг 3: Приведение изображения к фиксированному размеру 256x256.

После шагов 1 и 2 изображение приводится к единому размеру, так как входные данные в СНС должны иметь одинаковый размер.

3. Генерация выборки для обучения. Так как применяется сиамская архитектура используется генерация парных выборок, что представляет собой процесс создания обучающих примеров, состоящих из двух изображений и метки, указывающей на их принадлежности к одному или разным классам [84]. Такой формат используется в метрическом обучении при решении задач проверки подлинности или поиска визуально схожих изображений [85]. Генерация выборки для обучения состоит из следующих шагов:

Шаг 1: Выбор стратегии с равной вероятностью. Алгоритм случайным образом выбирает пары изображений:

- с вероятностью 50% — парные изображения (метка 0);
- с вероятностью 50% — непарные изображения (метка 1).

Под парными изображениями будем понимать пару, снятую в одной локации (метка 0). Непарные изображения – пары, снятые в разных локациях.

Шаг 2: Создание однолокационных пар. Если выпали парные изображения, то алгоритм берет одно и тоже исходное изображение и применяет к нему две разные аугментции (например, разные углы поворотов, изменения яркости). В результате получаются два визуально разных изображения, но принадлежащие одной локации. Устанавливается метка 0.

Шаг 3: Создание разнолокационных пар. Если выпали непарные изображения, алгоритм берет два разных изображения из набора данных, к каждому

из них применяются случайные аугментации. Устанавливается метка 1, что означает, что изображения относятся к разным классам.

Шаг 4: Формирование выходных данных. На выходе формируется структура, состоящая из тройки значений:

- первое преобразованное изображение;
- второе преобразование изображений;
- метка, указывающая на схожесть.

На вход для обучение модели подается структура, состоящая из тройки значений, чтоб нейронная сеть обучилась различать визуально похожие и непохожие объекты.

Наибольшую эффективность демонстрирует изображение подстилающей поверхности с применением фильтрации низкочастотных секвент в спектральном пространстве Уолша совместно с нормализацией и аугментации (глава 4, таблица 3). Исследована точность сходимости разработанной сверточной нейронной сети на большем объеме данных и использован следующий подход к обучению:

1. Обучение на парах без искажений на 3х наборах случайных пар из выборки.
2. Дообучение на парах с незначительными искажениями цветовой составляющей.
3. Фиксация весов верхнего уровня нейронной сети.
4. Дообучение на геометрических искажениях.
5. Анализ результатов на реальных данных с ЛА.

Использовалось следующее разбиение: обучающая выборка составляла 70-80 % от всех данных, тестовая выборка 20-30 % от всех данных.

На рисунке 21 приведена схема алгоритма обучения нейронной сети.

Входные данные поступают с двух источников: изображений с борта ЛА и базы данных спутниковых снимков. После поступления входных данных осуществляется формирование обучающих пар: данные аугментируются (нор-

мализация, изменение цветовых характеристик, добавление шума) и подвергаются геометрическим преобразованиям (поворот, масштабирование). Затем генерируются пары: 50% "похожих" (парные, метка 0) и 50% "непохожих" (непарные, метка 1).

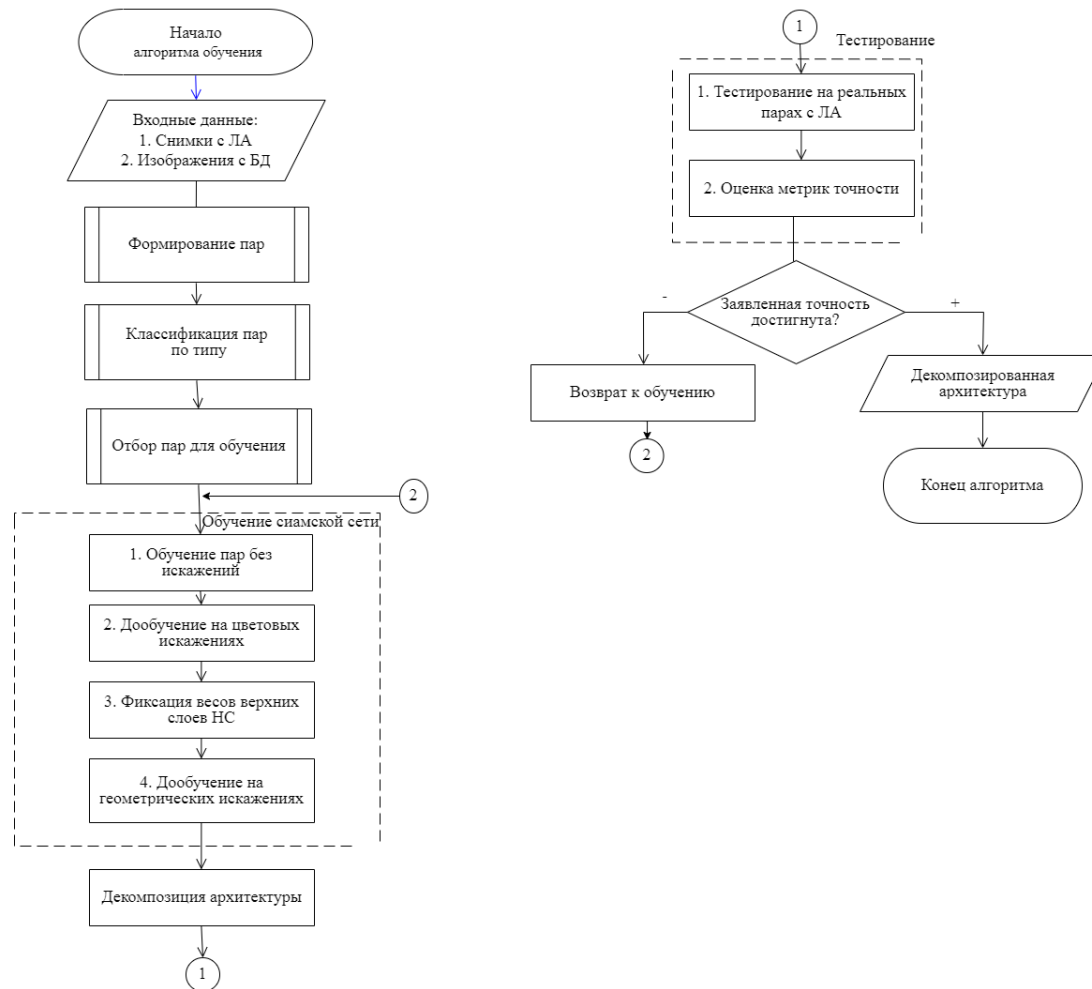


Рисунок 21 – Схема алгоритма обучения свёрточной нейронной сети

Классификация и отбор пар: сгенерированные пары классифицируются на типы на основе расстояния между их векторами и порога τ . Для обучения отбираются только простые позитивные, сложные позитивные и сложные негативные. Простые негативные пары исключаются, чтобы избежать дисбаланса и перекоса обучения в сторону тривиальных примеров.

Многоэтапное обучение сямской сети: отобранные пары используются для обучения сети в несколько этапов:

- начальное обучение на данных без искажений;

- дообучение на данных с цветовыми искажениями;
- фиксация весов верхних слоев для сохранения выученных признаков;
- финальное дообучение на геометрических искажениях для повышения устойчивости.

После обучения сямской архитектуры, от нее отделяется и сохраняется только одна ветвь (декомпозиция). Эта ветвь способна преобразовывать изображение в вектор-признаков.

Результатом работы алгоритма является полученная модель (одна ветвь) готова к использованию. На практике изображение, полученное с сенсора с ЛА преобразуется в вектор-признаков, который затем сравнивается (например, по косинусному расстоянию или L2) с базой данных изображений для поиска соответствий [90, 91]. Это эффективнее, чем постоянное использование полной парной сети.

2.2.2 Алгоритм формирования вектора-признаков изображения

На основе нейросетевого подхода изображения преобразуются в вектор-признаков [92]. На рисунке 22 приведена структурная схема модуля вычисления сходства. Так как на вход такой нейронной сети требуется два изображения I_i – изображение подстилающей поверхности, I_j – изображение с ЛА, а на выходе нужно сформировать численную оценку схожести D , архитектуру сети необходимо расширить до сямской.

Изображение преобразуется в вектор-признаков $h(I_i)$, $h(I_j)$, затем вычисляется расстояние между этими векторами.

Парные нейронные сети возможно использовать для выделения закономерностей в исходных парах изображений для формирования характеристик, на основе которых в дальнейшем можно сравнивать эти изображения. Веса в парных сетях, могут быть как общими, так и различными. Независимые веса используются для сравнения объектов разной природы, (например, текст и изображение, фото и чертеж).

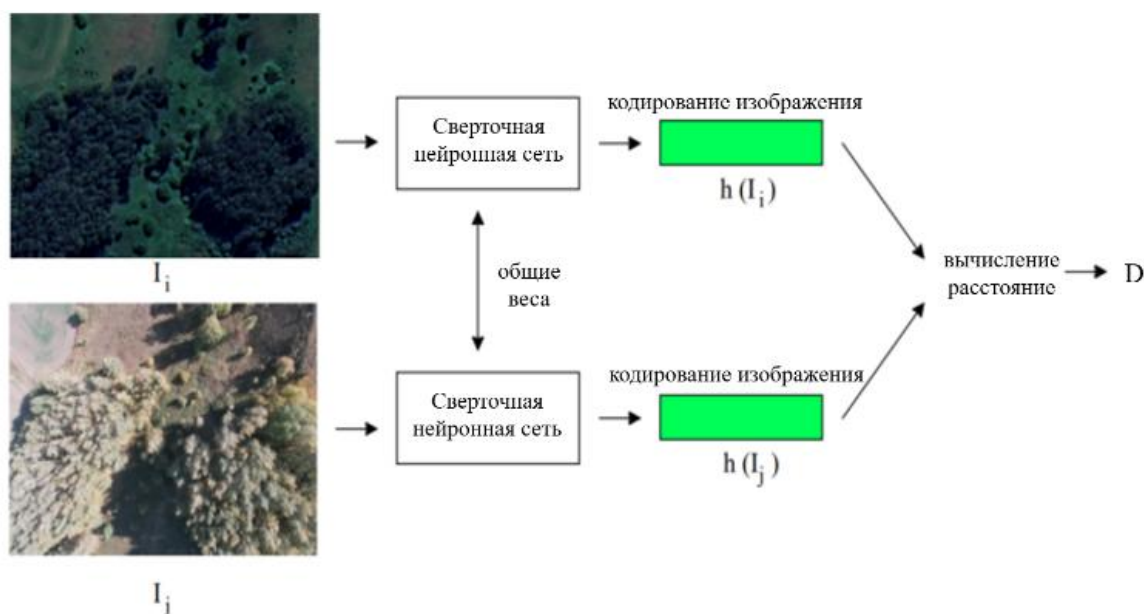


Рисунок 22 – Структурная схема модуля вычисления сходства с использованием сиамских сетей

Ключевые особенности сиамской архитектуры:

1. Две идентичные подсети (близнеца):

- сеть состоит из двух параллельных ветвей с одинаковой архитектурой и общими весами;

- каждая ветвь обрабатывает одно из входных изображений, преобразуя его в вектор признаков (эмбединг).

2. Функция расстояния (метрика сходства):

- после получения векторных представлений вычисляется расстояние между ними (например, евклидово расстояние, косинусная расстояние).

- чем меньше расстояние, тем более похожи объекты.

3. Обучение с парными примерами.

Формирование вектора-признаков основано на операциях свертки и подвыборки (операции уменьшения размерности). Основной идеей свёрточной сети является то, что обработка участка изображения должна не зависеть от расположения этого участка на изображении. Например, необходимо найти объект на изображении, и абсолютно не важно на 100 или на 200 пикселей этот объект расположен от края [84].

Важную роль играет взаимное расположением объектов на изображении. Распознавание этих объектов локальное и не зависит от взаимного расположения этих объектов внутри изображения. Операция свертка – это линейное преобразование входных данных особого вида. Входные данные для операции свертки не должны быть велики, так как это повысит вычислительную сложность. В тоже время, если размер входного изображения будет слишком мал, то ключевые признаки на данном изображении выявить не удастся. Операция подвыборки – это операция уменьшения размерности, которая позволяет выделять наиболее значимые признаки на изображении.

Операции свертки и подвыборки организованы таким образом, что за M итерации можно преобразовать изображение в некоторый вектор-признаков, который позволяет получить характеристики изображений. Некоторое ядро свертки обрабатывает предыдущее ядро по фрагментам. На рисунке 23 представлена модель алгоритма свертки.

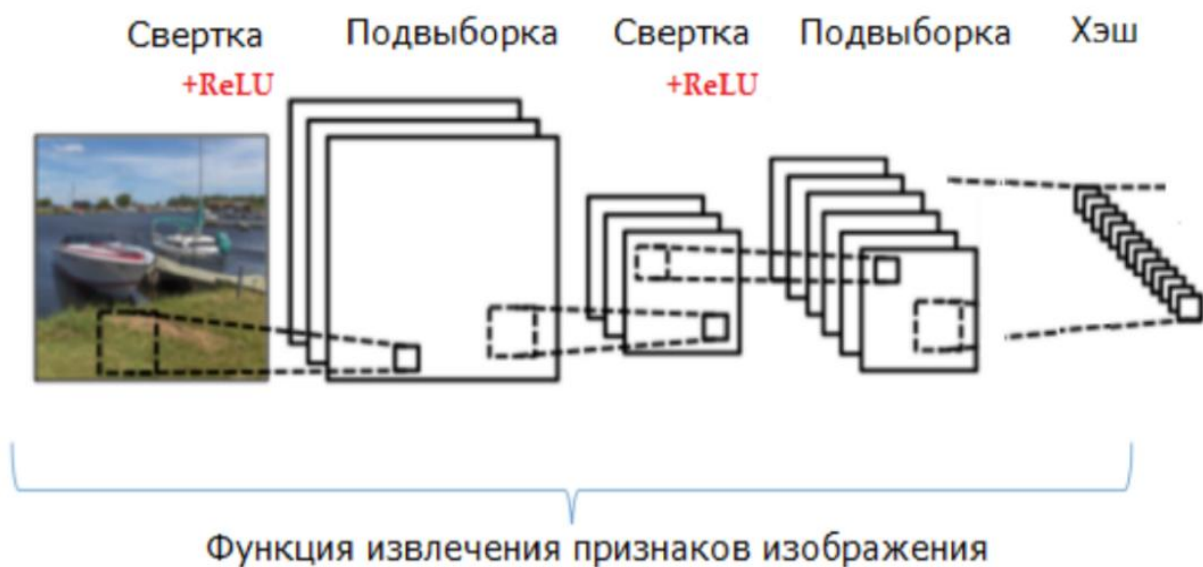


Рисунок 23 – Модель алгоритма свертки

Ядро свертки представляет собой фильтр (скользящее окно), которое находит определенные признаки объектов [31]. Ядро скользит по предыдущей итерации и производит операцию свертки. Операция ReLU – это активационная функции, которая обнуляет все отрицательные элементы. Обычно в сверточном слое применяется функция активации ReLU.

$$\text{ReLU}(z) = \max(0, z) = \begin{cases} z, & \text{если } z > 0 \\ 0 & \text{если } z \leq 0 \end{cases} \quad (24)$$

Также если модификации ReLU: LeakyReLU и ELU. Leaky ReLU вычисляется следующим образом:

$$\text{LeakyReLU}(z) = \max(\alpha z, z), \alpha \approx 0.01 \quad (25)$$

В данной работе предлагается использовать ELU (Exponential Linear Unit) в качестве функции активации, что позволяет решить задачу затухающих нейронов. Активационная функция ELU определяется следующей формулой:

$$\text{ELU}(z) = \begin{cases} z, & \text{если } z \geq 0 \\ \alpha(e^x - 1) & \text{если } z < 0 \end{cases} \quad (26)$$

где z – входной сигнал, α (обычно $\alpha \geq 0$) – гиперпараметр, определяющий насыщение функции для отрицательных значений [86].

На рисунке 24 приведено отличие активационных функций

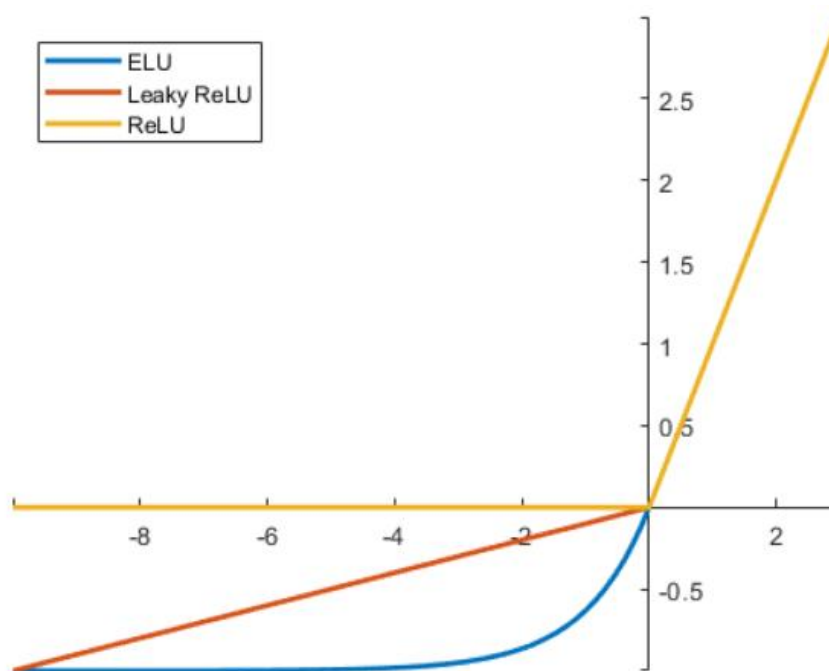


Рисунок 24 – Отличие активационных функций

Операции уменьшения размерности

Подвыборка представляет собой нелинейное уплотнение карты признаков, при котором группа пикселей (2x2) уплотняется до 1 пикселя, проходя

операцию подвыборки [31,87]. Цель операции подвыборки – уменьшение размерности результата предыдущей операции (свертки). В качестве преобразования наиболее часто используется функция поиска максимума.

1. Max Pooling – классический метод подвыборки, который осуществляет уменьшение размерности карт признаков, выбирая максимальное значение в пределах заданного окна. На рисунке 25 представлена операция подвыборки (Max Pooling).

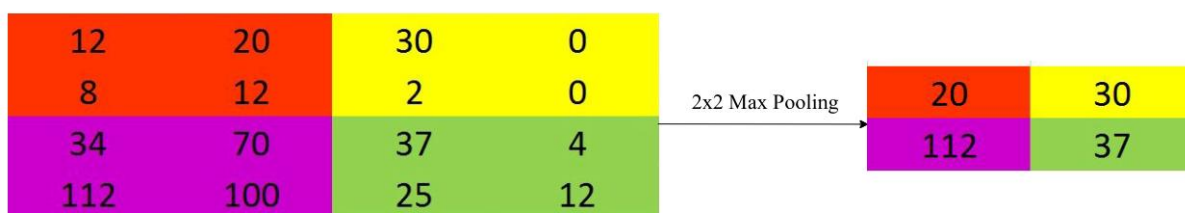


Рисунок 25 – Демонстрация операции подвыборки

Свойства:

- выбирает максимальное значение в пределах окна;
- сохраняет наиболее яркие (важные) признаки;
- уменьшает размерность карты признаков;
- устойчив к малым сдвигам и шуму.

2. AvgPooling – классический метод подвыборки, который осуществляет уменьшение размерности карт признаков, выбирая среднее значение в пределах заданного окна.

Свойства:

- вычисляет среднее значение в окне;
- сглаживает признаки, уменьшая влияние выбросов;
- часто используется в завершающих слоях (например, перед полносвязным слоем в VGG).

На рисунке 26 представлена операция подвыборки (Avg Pooling).

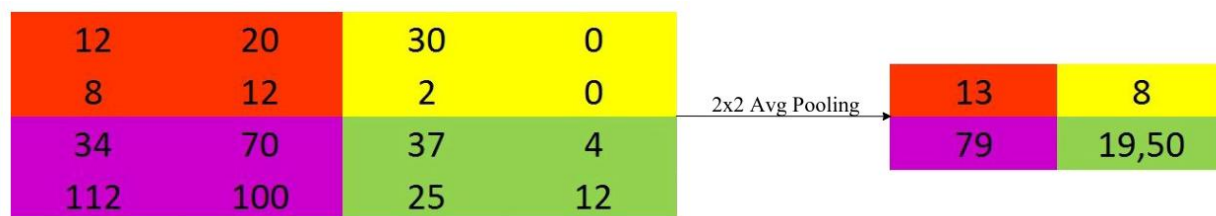


Рисунок 26 – Демонстрация операции подвыборки

3. Линейная интерполяция – альтернативный подход к подвыборке, при котором используется взвешенная линейная интерполяция значений к некоторой точке [88]. Модель на основе обучения выбирает к какой точке ближе аппроксимировать.

Алгоритм работы линейной интерполяции с обучаемым параметром основывается на тензорных вычислениях. Тензор – математический объект, который представляет собой многомерный массив чисел, характеризующийся рангом (количество осей, размерностей) и формой (количеством элементов по каждой оси). В нейросетевых технологиях, как правило используется тензор четвертого ранга.

Иерархия тензоров: тензор нулевого ранга (скаляр), например, отдельно взятое обучаемое смещение b , отдельно взятый вес w_{11} ; тензор первого ранга (вектор), например, набор весов для одного фильтра: $[w_{11}, w_{12}, w_{21}, w_{22}]$; тензор второго ранга (матрица), например, один канал изображения размером 4x4 пикселя; тензор третьего ранга (куб или последовательность), например, цветное изображение (высота, ширина, каналы); тензор четвертого ранга (стандарт для нейросетей), например, пакет тензоров третьего ранга.

Алгоритм работы линейной интерполяции с обучаемым параметром состоит из следующих шагов:

Шаг 1. Разделить входной тензор на блоки 2×2 .

Шаг 2. Вычислить взвешенные коэффициенты для каждого пикселя в блоке, чтобы приблизить значение, например, к точке x_{11} .

Формула линейной интерполяции с обучаемыми весами (для одного выходного пикселя):

$$y' = w_{11}x_{11} + w_{12}x_{12} + w_{21}x_{21} + w_{22}x_{22} + b, \quad (27)$$

где w_{11} , w_{12} , w_{21} , w_{22} – обучаемые веса, b – обучаемое смещение.

Шаг 3. Задаем смещение, как обучаемый параметр, что позволяет модели автоматически выбирать, к какой точке внутри блока аппроксимировать.

Шаг 4. Для каждой точки в блоке задаем обучаемые веса, которые определяют их вклад в распределении.

Шаг 5. Сформировать выходной тензор меньшего размера.

Шаг 6. Передать выходной тензор дальше по нейронной сети [88].

Данный подход уменьшения размерности является альтернативой AvgPooling, но применяется тогда, когда важна именно плавность изменений, например, при спектральных преобразованиях. Данный подход может быть применен к нейронным сетям с малым количеством данных, так как в них меньше риск к переобучению. К преимуществам использования можно отнести то, что данная операция меньше искажает границы объектов, чем AvgPooling. Веса адаптивны к данным.

На рисунке 27 приведено распределение значений при использовании разных операций для уменьшения размерности.

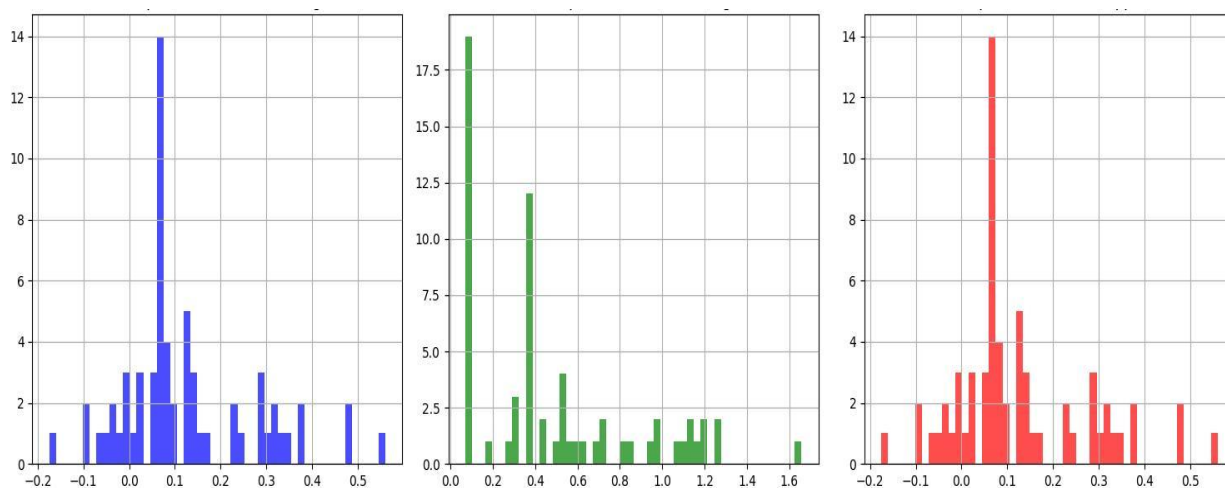


Рисунок 27 – Распределение значений при использовании разных операций для уменьшения размерности

а – распределение AvgPooling, б – распределение MaxPooling,

в – распределение LinearApprox

На рисунке 28 приведена гистограмма сравнения операций уменьшения размерности по ключевым характеристикам. В качестве критериев использовались: чувствительность, устойчивость, сохранение границ, адаптивность, вычислительная сложность, градиенты.

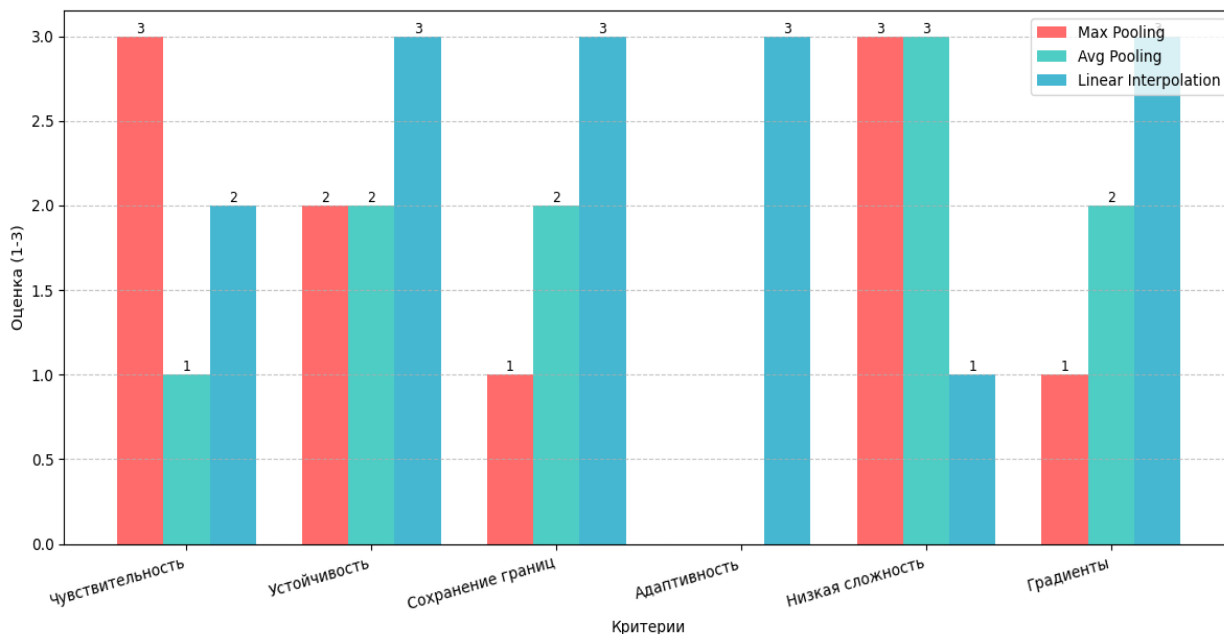


Рисунок 28 – Гистограмма сравнения операций уменьшения размерности по ключевым характеристикам

Качественные характеристики переведем в числовые баллы от 1 до 3, где 3 – наилучший показатель по данному критерию. Чувствительность (к ярким признакам): MaxPool=3 (акцентирует яркие признаки), AvgPool=1 (учитывает все значения), Linear=2 (адаптивно взвешивает значения, приближая к ключевой точке). Устойчивость: Max=2 (к шуму), Avg=2 (к плавным), Linear=3 (и к тому, и к другому). Сохранение границ (обратный показатель "Искажение границ"): Max=1, Avg=2, Linear=3. Адаптивность: Max=0, Avg=0, Linear=3. Сложность: Max=3 (низкая), Avg=3, Linear=1 (высокая, т.к. есть обучение). Градиенты (эффективность распределения): Max=1 (одному), Avg=2 (всем), Linear=3 (обучаемым).

На рисунке 29 приведена демонстрация операций подвыборки; а) – исходное изображение (256x256), б) – после операции свёртки (256x256), в) – после

применения MaxPooling к свертке б) (64x64), г) – после применения AvgPooling к свертке б) (64x64), д) – после применения линейной аппроксимации с обучаемым параметром к свертке б) (64x64).

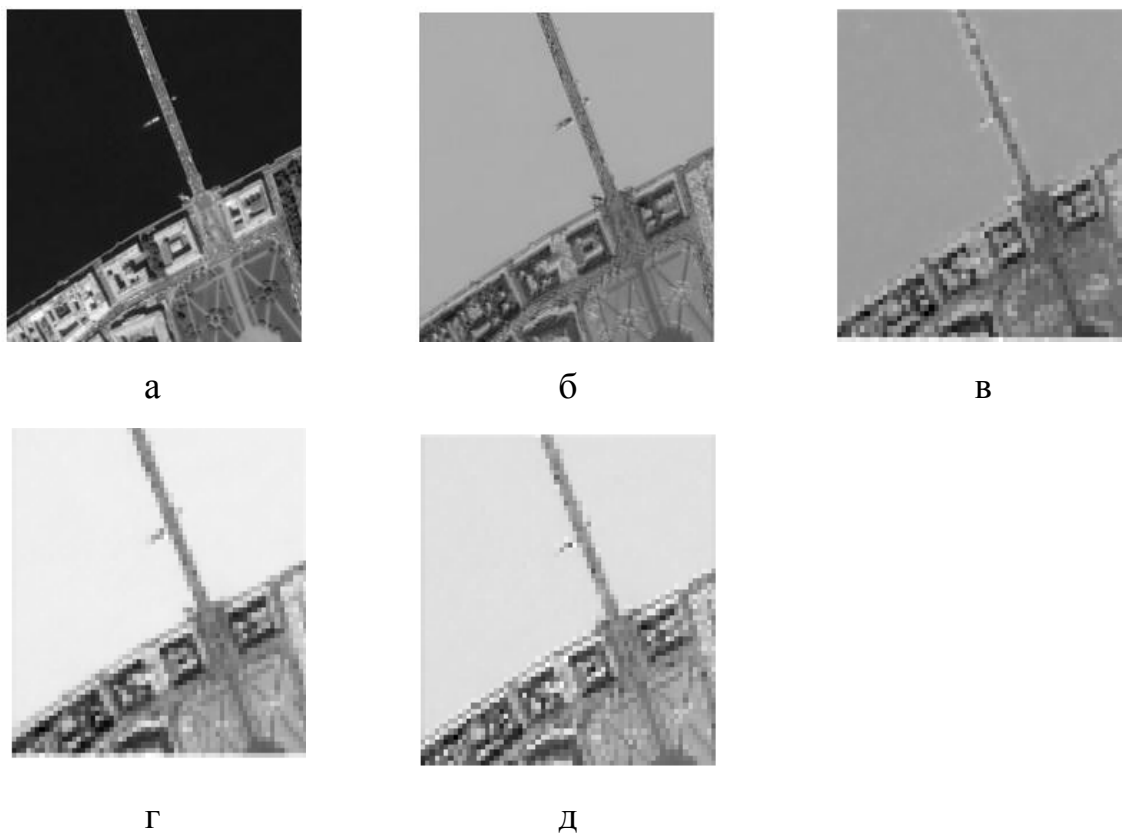


Рисунок 29 – Демонстрация операций подвыборки

На вход нейронной сети поступает изображение, размера 256x256, которого достаточно для обнаружения и классификации наиболее значимых частей изображения и при этом обработка изображения такого размера имеет низкую вычислительную сложность.

Исходное изображение 256x256 преобразуется в спектр Уолша, удаляются низкочастотные компоненты и восстанавливается изображение [89]. Таким образом осуществляется предобработка.

Архитектура СНС:

1. Применяем операцию свертки к предобработанному изображению с ядром свёртки, вычисленным на основе обучения нейронной сети [90].

2. К полученной после операции свёртки 256×256 применяем линейную интерполяцию. Тем самым получаем 64×64 значения операции свёртки после усечения.

3. Применяем операцию свёртки 64×64 для полученной на шаге 3 сверки (после усечения) с ядром свёртки, вычисленным на основе обучения нейронной сети.

4. Применяем операцию свёртки 64×64 для полученной на шаге 4 свёртке.

5. К полученной после операции свёртки 64×64 применяем операцию усечения (MaxPooling 4×4). Получим 16×16 значения операции свёртки после усечения.

6. Применяем к полученной после усечения MaxPooling операцию свёртки с ядром свёртки, вычисленным на основе обучения нейронной сети.

7. Применяем к полученной свёртке 16×16 операцию усечения (MaxPooling 4×4). Получим 4×4 значений.

8. Полученные 4×4 значений преобразуем из многомерного пространства в одномерный вектор, чтобы его можно было подать на вход полносвязному слою (Dense) в нейронной сети. Получим 16 значений.

9. Следующие два полносвязных слоя осуществляют линейное преобразование + активацию. На первом полносвязном слое на вход поступает 16 нейронов с предыдущего слоя и 256 весов.

$$h_9 = \sigma(W_9 x_{flat} + b_9), \quad (28)$$

где x_{flat} – вектор, полученный на 8 шаге, $W_9 \in R^{256 \times 16}$ – матрица весов, $b_9 \in R^{256}$, $\sigma(z)$ – сигмоидная активация. Итого на выходе с данного слоя получим $4096 + 256 = 4352$ параметров.

10. Второй полносвязный слой Dense(128). На вход поступает 256 нейронов с предыдущего слоя. Итого на выходе $32768 + 128 = 32896$ параметров.

Dense(256) принимает 16 значений и выдаёт 256 значений \rightarrow 4352 параметра. Dense(128) принимает 256 значений и выдаёт 128 значений \rightarrow 32,896 параметров. Общее число параметров до сиамской части: 37,248.

11. После шагов 9 и 10 формируются две ветви (для изображений A и B):

$$x_{11}^a, x_{11}^b \in R^{128}, \quad (29)$$

12. Общие слои для обеих ветвей. Повторим два полносвязных слоя Dense(256), Dense(128).

13. Сравнение ветвей осуществляется с использованием вычитания. Применяется квадратичная активация.

14. Финальный слой Dense(1). На вход подается вектор из 128 значений и 1 смещения. Получаем 129 параметров, которые описывают степень соответствия.

Полные формулы разработанной архитектуры

Для двух изображений A и B :

1. Слои сверётки и подвыборка:

$$\begin{aligned} x_{flat}^a &= Flatten(Pooling(Conv3(A))) \\ x_{flat}^b &= Flatten(Pooling(Conv3(B))) \end{aligned} \quad (30)$$

2. Общие полносвязные слои:

$$\begin{aligned} x_{11}^a &= \sigma(W_{10}(\sigma(W_9 x_{flat}^a + b_9) + b_{10})) \\ x_{11}^b &= \sigma(W_{10}(\sigma(W_9 x_{flat}^b + b_9) + b_{10})) \end{aligned} \quad (31)$$

3. Сиамская сеть:

$$\begin{aligned} y &= w_{out} [\sigma(W_{14} \sigma(W_{13} x_{11}^a + b_{13}) + b_{14}) - \sigma(W_{14} \sigma(W_{13} x_{11}^b \\ &\quad + b_{13}) + b_{14})]^2 + b_{out} \end{aligned} \quad (32)$$

На рисунке 30 приведена архитектура разработанной СНС.



Рисунок 30 – Архитектура сверточной нейронной сети

На рисунке 31 приведена визуализация слоев для спутникового изображения. На рисунке 31: а – результат после первой операции свёртки; б – результат после операции уменьшения размерности (линейная аппроксимация к центру квадрата); в – результат после второй операции свёртки; г – результат после третьей операции свёртки; д – результат после подвыборки MaxPooling; е – результат после операции свёртки; ж – результат после подвыборки MaxPooling.

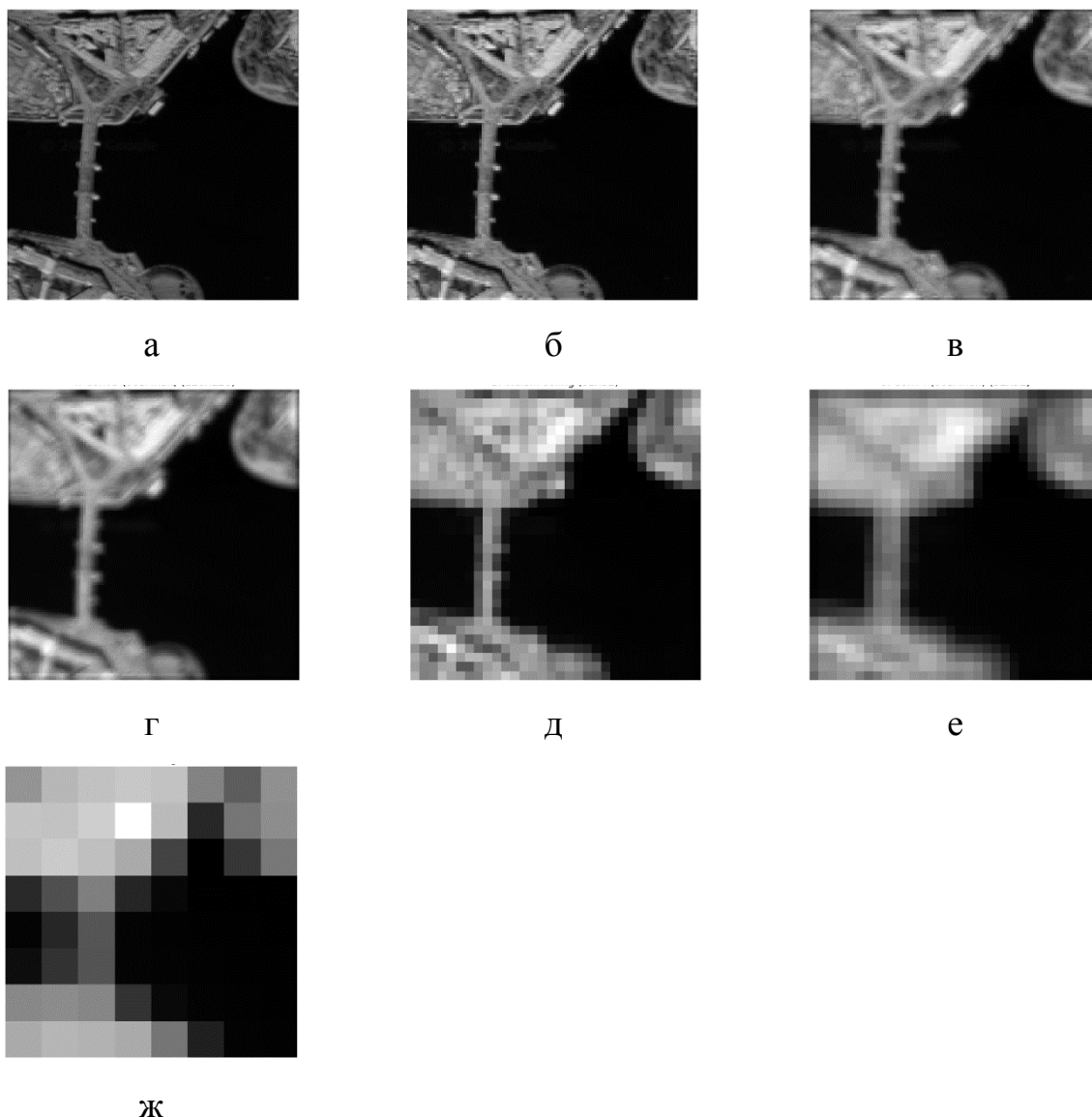


Рисунок 31 – Визуализация слоев нейронной сети

На рисунке 32 приведена схема алгоритма формирования вектора-признаков изображения.



Рисунок 32 – Схема алгоритма формирования вектора-признаков изображения

В качестве входных значений подается изображение, далее применяется предобработка, заключающаяся в нормализации, приведении изображения к размеру 256x256 и аугментации. Далее предобработанное изображение подается на блок предобработки на основе спектрального преобразования с подавлением низкочастотных секвент. Полученное изображение подается на деком-

позированную СНС для формирования вектора-признаков изображения. Результатом работы алгоритма является вычисленный вектор-признаков изображений. На рисунке 33 приведена зависимость точности работы от количества характеристик в векторе-признаке.

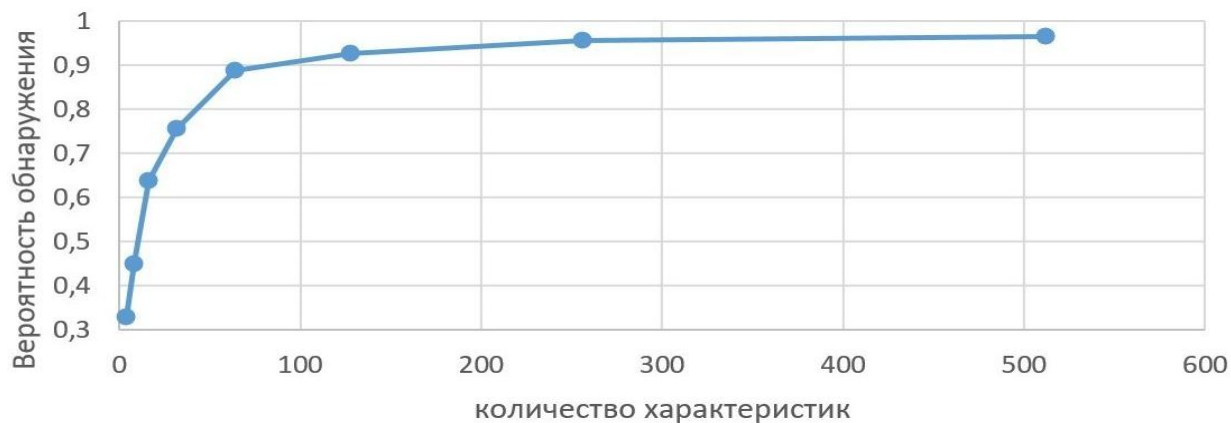


Рисунок 33 – Зависимость точности работы от количества характеристик в векторе-признаков

В результате экспериментов было выявлено, что необходимым и достаточным количеством характеристик в векторе-признаке изображений является 256 значений.

2.2.3 Алгоритм поиска спутникового изображения, позволяющего определить потенциальное положение летательного аппарата

Определять местоположение предлагается в условиях отсутствия GPS или его ненадежности, на основе визуального сходства пары изображений: полученного в текущий момент времени с ЛА и изображения подстилающей поверхности [91].

К научной задаче анализа снимков с ЛА можно отнести следующее:

- зависимость от сезона съемки;
- зависимости от времени суток съемки;
- зависимость от проекции съемки;
- искажение снимков в результате помех и шумов;
- недостаточность вычислительной мощности бортового оборудования;

– динамическое изменение анализируемой картины.

При решении задачи совмещения разнородных изображений важно, чтобы сравнение изображений было устойчивым к изменениям освещения, сезонным вариациям и другим глобальным искажениям.

На рисунке 34 приведена схема алгоритма поиска. Входным данным является изображение, полученное с сенсора ЛА. Целью данного этапа – создать БД векторов-признаков для всех тайлов карты местности.

Для формирования БД используется только одна ветвь предобученной нейронной сети. Каждый тайл пропускается через нейронную сеть для формирования вектора-признаков изображения. Векторы сохраняются вместе с географическими координатами тайлов.

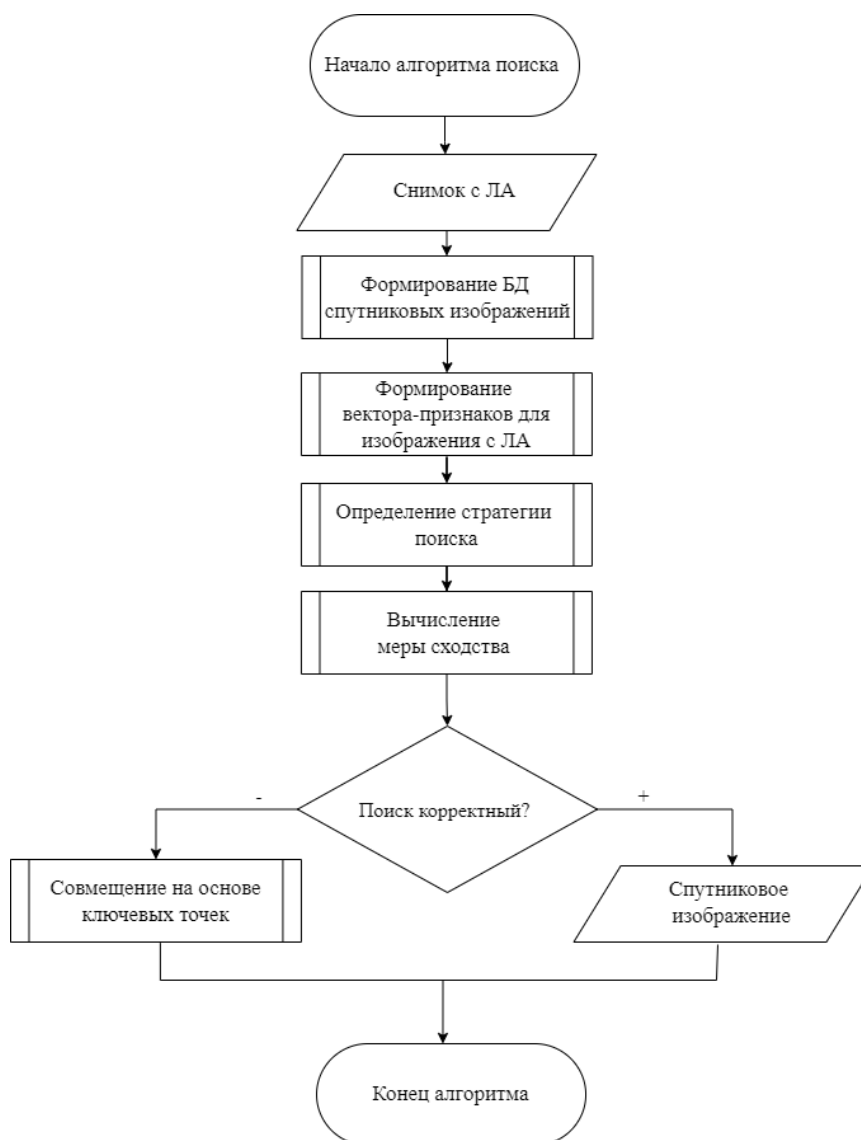
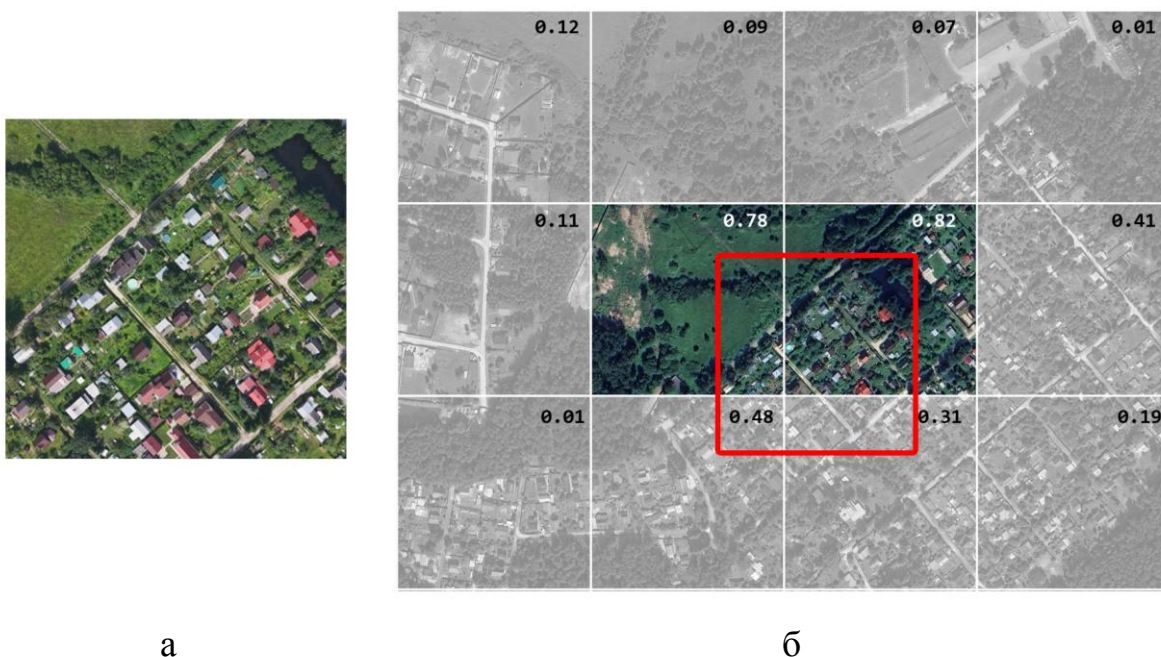


Рисунок 34 – Схема алгоритма поиска

Определение стратегии поиска: если это первый запуск осуществляется проверка в окрестности стартовой точки, если это последующие запуски, то осуществляется проверка предложенного нейросетевым алгоритмом тайла, находится в он +-3 тайла от предыдущей оценки. Выполняется доуточнение на аномальных скачках, а также периодическое доуточнение каждые 30 кадров.

На рисунке 35 приведен результат работы. На рисунке 35,а приведен снимок с ЛА, на рисунке 35,б – спутниковое изображение разбитое на тайлы.



а

б

Рисунок 35 – Результат работы алгоритма

а – снимок с ЛА, б – спутниковый снимок разбитый на тайлы

В данном случае снимок с ЛА располагается одновременно на четырех тайлах, нейросетевой алгоритм оценивает вероятность того, что изображения сняты в одном месте. В том случае когда снимок, полученный с ЛА располагается на нескольких тайлах одновременно необходимо осуществлять уточнение местоположения на основе ключевых точек.

Совмещение изображений осуществляется с учетом фильтрации низких частот, одной ветви сиадкой нейронной сети.

2.2 Способ предварительной обработки изображений подстилающей поверхности с использованием спектральных преобразований

При решении задачи локализации ЛА по аэрофотоснимкам важно, чтобы сравнение изображений было устойчивым к изменениям освещения, сезонным вариациям и другим глобальным искажениям. Низкочастотная составляющая в изображении содержит информацию о медленно меняющихся яркостных характеристиках, таких как:

- общее освещение (солнечные блики, тени, время суток);
- атмосферные эффекты (дымка, туман, облачность);
- сезонные изменения (снег, дождь, листва, высохшая трава). А инвариантность к масштабу, повороту и сдвигу.

Эти факторы не несут полезной информации о геометрии местности, однако в значительной степени влияют на метрики сходства.

Предварительную обработку применяют для:

1. Повышения инвариантности к освещению:

- низкие частоты часто связаны с изменяющимися условиями съёмки, а не с реальными особенностями местности.
- фильтрация низких частот делает векторы-признаков изображений более устойчивыми к разному освещению (утро/день/вечер).

2. Акцентирования на текстурах и контурах.

Высокочастотные компоненты (края, текстуры, детали) точнее отражают уникальные особенности местности.

3. Уменьшения влияния засветки и теней.

Если ЛА летит против солнца, одна часть кадра может быть переэкспонирована, а другая — в тени. Устранение низких частот нивелирует этот эффект.

4. Улучшения работы нейросетевых моделей:

- современные сиамские сети и методы сравнения изображений (например, VGG, ResNet) часто работают точнее, когда входные данные нормализованы по яркости.

– низкочастотные вариации могут "перегружать" признаки, делая их менее дискриминативными (не эффективно разделяют разные классы объектов).

Высокочастотная составляющая в изображении содержит информацию о резких изменениях яркости в пространстве. Эти изображения возникают на границах объектов в областях с выраженной текстурой. Высокочастотные компоненты несут информацию о:

- границах и контурах объектов (резкие перепады яркости, соответствующие краям зданий, деревьев, дорог, элементов рельефа);
- текстурах (листва деревьев, трава, кирпичная кладка, галька, асфальт);
- деталях (мелкие объекты, которые имеют контраст с окружающим фоном, например линии электропередач).

Эти факторы являются стабильными и инвариантными к изменению условий съемки, так как обусловлены объектами на местности.

С применением предобработки алгоритм будет сравнивать только геометрические паттерны (дороги, здания), игнорируя сезонные изменения. Удаление низкочастотной составляющей позволяет повысить работоспособность алгоритма сопоставления изображений, делая его менее зависимым от условий съёмки и более точным для локализации.

Геометрические паттерны – это устойчивые, повторяющиеся структуры или формы объектов на изображении, которые можно описать с помощью геометрических примитивов (линий, углов, многоугольников, кривых) и их комбинаций.

В контексте обработки спутниковых или аэрофотоснимков к ним относятся:

- линейные элементы (дороги, железнодорожные пути, реки);
- полигональные структуры (здания, парки, сельскохозяйственные поля);
- регулярные сетки (городская планировка, инфраструктурные объекты);

– симметричные или фрактальные формы (архитектурные сооружения, искусственные водоёмы).

Предварительная обработка с фильтрацией низкочастотных составляющих повышает уникальность векторов признаков изображений, так как: два участка с разной освещенностью, но одинаковыми текстурными признаками после фильтрации имеют идентичные высокочастотные компоненты.

Предварительная обработка изображений подстилающей поверхности осуществляется с использованием спектрального преобразования Уолша, а также производится аугментация набора исходных данных для обучения нейронной сети. Предварительная обработка применяется для улучшения сходимости и обобщающей способности модели нейронной сети.

По функциям Уолша можно производить разложение произвольных сигналов в ряд Уолша-Фурье, которые принимают всего два значения (+1 или -1), поэтому удобны для вычислений [9]. Спектральное преобразование Уолша использует три основные системы упорядочивания функций Уолша: систему Адамара (естественное упорядочивание), систему Пэли (диадическое упорядочивание) и систему Уолша (упорядочивание по секвенте) [92, 93].

В преобразовании Фурье базисными функциями являются синусоиды, а мерой частоты служит частота (количество полных колебаний за единицу времени), то в преобразовании Уолша базисными функциями являются прямоугольные волны (функции Уолша), а мерой их частоты служит секвента. Секвента – это половина среднего количества пересечений нуля (перемен знака) базисной функции Уолша на единичном интервале.

Функции Уолша в системе Пэли упорядочены по числу изменений знака, но в диадической (двоично-отражённой) последовательности, что делает их более удобными для анализа сигналов с резкими перепадами (например, в цифровых изображениях). Функции Уолша в системе Адамара являются наиболее вычислительно эффективными благодаря простоте алгоритма быстрого преобразования. Функции Уолша в системе Уолша строго упорядочены по возрастанию числа пересечений нуля, что обеспечивает прямую аналогию

с упорядочением тригонометрических функций в ряд Фурье по возрастанию частоты.

Преобразование Уолша в системе Пэли для дискретного сигнала $x[n]$ длины $N = 2^m$ определяются следующим образом:

– прямое преобразование Уолша-Пэли:

$$X[k] = \frac{1}{N} \sum_{n=0}^{N-1} x[n] \cdot Pal(k, n), \quad k = 0, 1, \dots, N - 1. \quad (33)$$

– обратное преобразование Уолша-Пэли:

$$x[n] = \frac{1}{N} \sum_{k=0}^{N-1} X[k] \cdot Pal(k, n), \quad n = 0, 1, \dots, N - 1. \quad (34)$$

где $Pal(k, n)$ – функция Уолша-Пэли порядка k в точке n , $N = 2^m$ – длина сигнала.

Функция Уолша-Пэли $Pal(k, n)$ определяется через произведение знаков Радемахера:

$$Pal(k, n) = \sum_{j=0}^{m-1} (Rad_j(n))^{k_j}, \quad (35)$$

где k_j – j -й бит числа k в двоичном представлении, $Rad_j(n)$ – функция Радемахера порядка j :

$$Rad_j(n) = (-1)^{\lfloor \frac{n}{2^j} \rfloor}. \quad (36)$$

Матричное представление (для $N = 4$) с использованием матрицы Уолша-Пэли \mathbf{H}_4 для $N = 4$:

$$\mathbf{H}_4 = \begin{pmatrix} 1 & 1 & 1 & 1 \\ 1 & 1 & -1 & -1 \\ 1 & -1 & 1 & -1 \\ 1 & -1 & -1 & 1 \end{pmatrix}. \quad (37)$$

Карта местности подстилающей поверхности разбивается на тайлы размером 256x256. Спектр Уолша-Пэли вычисляется как:

$$\mathbf{H} = \mathbf{H}_{256} \cdot I \cdot \mathbf{H}_{256}^T, \quad (38)$$

где \mathbf{H} – спектр Уолша-Пэли изображения I , \mathbf{H}_{256}^T – транспонированная матрица Уолша–Пэли [96].

Упорядочение по Уолшу группирует функции по числу нулей (пересечений) на интервале, что делает его более интуитивным для анализа сигналов [97, 98].

Преобразование Уолша в системе Адамара для дискретного сигнала $x[n]$ длины $N = 2^m$ определяются следующим образом:

– прямое преобразование Уолша-Адамара:

$$X[k] = \frac{1}{N} \sum_{n=0}^{N-1} x[n] \cdot Had(k, n), \quad k = 0, 1, \dots, N - 1. \quad (39)$$

– обратное преобразование Уолша-Адамара:

$$x[n] = \frac{1}{N} \sum_{k=0}^{N-1} X[k] \cdot Had(k, n), \quad n = 0, 1, \dots, N - 1. \quad (40)$$

где $Had(k, n)$ – функция Уолша-Адамара порядка k в точке n , $N = 2^m$ – длина сигнала.

Переход в систему Адамара из системы Пэли осуществляется с помощью матрицы перестановки \mathbf{P} .

$$\mathbf{H}_N^{(\text{Адамар})} = \mathbf{P} \cdot \mathbf{H}_N^{\text{Пэли}} \quad (41)$$

где \mathbf{P} строится на основе сравнения последовательностей перемен знака функций в двух упорядочениях.

Матричное представление (для $N = 4$) с использованием матрицы Уолша-Адамара \mathbf{H}_4 для $N = 4$:

$$\mathbf{H}_4 = \begin{pmatrix} 1 & 1 & 1 & 1 \\ 1 & -1 & 1 & -1 \\ 1 & 1 & -1 & -1 \\ 1 & -1 & -1 & 1 \end{pmatrix}. \quad (42)$$

Преобразование Уолша в системе Уолша вычисляется следующим образом:

$$Wal(k, n) = \sum_{j=0}^{m-1} (Rad_j(n))^{k_j(r)}, \quad (43)$$

где $Wal(k, n)$ – функция Уолша порядка k в точке n , $N = 2^m$ – длина сигнала, $k_j(r)$ – j -й бит в коде Грея числа r , который вычисляется как:

$$k_j(r) = k_j \oplus k_{j+1}, \quad (44)$$

где k_j – j -й бит двоичного представления k .

Преобразование Уолша в системе Уолша определяется следующим образом:

– прямое преобразование:

$$X[k] = \frac{1}{N} \sum_{n=0}^{N-1} x[n] \cdot Wal(k, n), \quad k = 0, 1, \dots, N-1. \quad (45)$$

– обратное преобразование:

$$X[n] = \frac{1}{N} \sum_{k=0}^{N-1} X[k] \cdot Wal(k, n), \quad n = 0, 1, \dots, N-1. \quad (46)$$

Матричное представление (для $N = 4$) с использованием матрицы Уолша \mathbf{W}_4 для $N=4$:

$$\mathbf{W}_4 = \begin{pmatrix} 1 & 1 & 1 & 1 \\ 1 & 1 & -1 & -1 \\ 1 & -1 & -1 & 1 \\ 1 & -1 & 1 & -1 \end{pmatrix}. \quad (47)$$

Спектр Уолша для изображений I размером 256×256 вычисляется аналогично (38).

На рисунке 36 распределение частотностей в спектральном преобразовании Уолша.

На рисунке 36 приведена визуализация, как именно распределены частотные компоненты в преобразовании Уолша. Рассмотрим цветовую схему и смысл каждого элемента.

Цветовая карта частотных областей:

1. Голубой - низкие секвенты.
2. Тёмно-синий - средние секвенты.

Горизонтальная полоса справа от низкочастотного блока. Отвечает за основные контуры объектов. Левый верхний квадрат 64×64 (1/4 от всего преобразования). Содержит информацию о плавных изменениях яркости. Здесь находится постоянная составляющая (средняя яркость всего изображения).

3. Зелёный и оранжевый - средние низкие секвенты.

Вертикальная полоса под низкочастотным блоком. Содержит информацию о более детализированных границах.

4. Темно красный - высокие секвенты.

Правый нижний квадрат 128×128 . Отвечает за мелкие детали и текстуры.

Предлагается проанализировать как влияет частотные компоненты на сходимость СНС в зависимости от фильтрации той или иной частотной области.

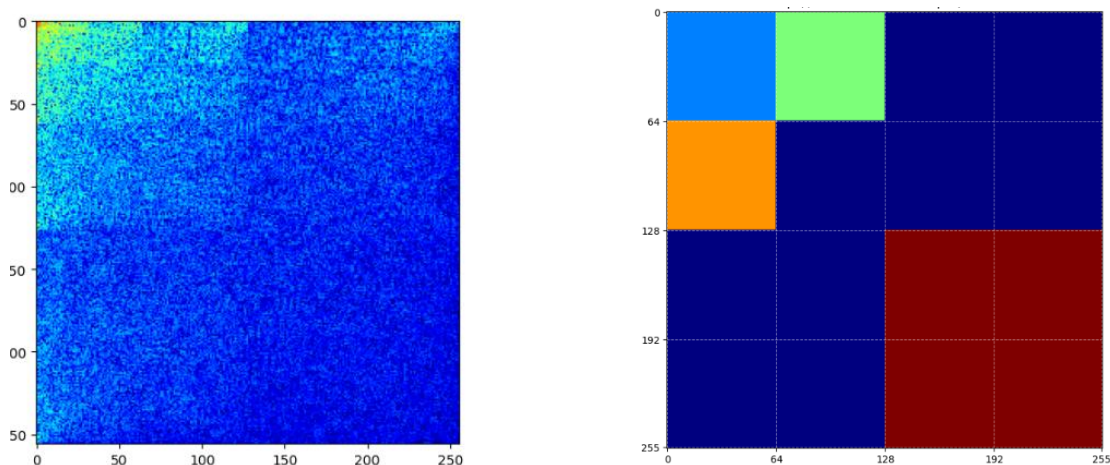


Рисунок 36 – Распределение частотностей в спектральном преобразовании

Результирующее изображение после обратного преобразования будет содержать только:

- общую структуру местности;
- основные контуры объектов;
- без мелких деталей и текстур.

Эта информация прямо отвечает на главный вопрос – какие именно детали снимка останутся после обработки, а какие будут отсеяны.

Способ предварительной обработки заключается в частотной фильтрации изображения с использованием преобразования Уолша и состоит из следующих шагов:

1. Создаем матрицу Уолша размером 256x256. Применяем преобразование Уолша для изображения размером 256x256. Результат – спектральное преобразование, вычисляемое по формуле:

$$W = W_{256} \cdot X \cdot W_{256}^T, \quad (48)$$

где W_{256} – матрица Уолша порядка 256, W — спектральная матрица после преобразования, X – входной сигнал.

2. Фильтрация низких частот. Что эквивалентно сохранению только гладких участков блока, удалению резких переходов. Для того, чтобы применить низкочастотный фильтр в W , применим маску M :

$$M_{ij} = \begin{cases} 0, & \text{если } i + j \leq K \\ 1, & \text{иначе} \end{cases} \quad (49)$$

где K – порог обнуления.

Результирующая спектральная матрица вычисляется по следующей формуле:

$$W' = W \cdot M \quad (50)$$

где \cdot – поэлементное умножение.

3. Обратное преобразование Уолша. Фильтрованное частотное представление преобразуется обратно в пространственную область. Производим нормировку:

$$X' = \frac{1}{256^2} W \cdot W' \cdot W \quad (51)$$

Способ предварительной обработки подстилающей поверхности в спектральном пространстве Уолша заключается в фильтрации низкочастотных секвент [99].

На рисунке 37,а приведено текущее изображение, 37,б – фильтр низких частот, 37,в – изображение после обратного преобразования.

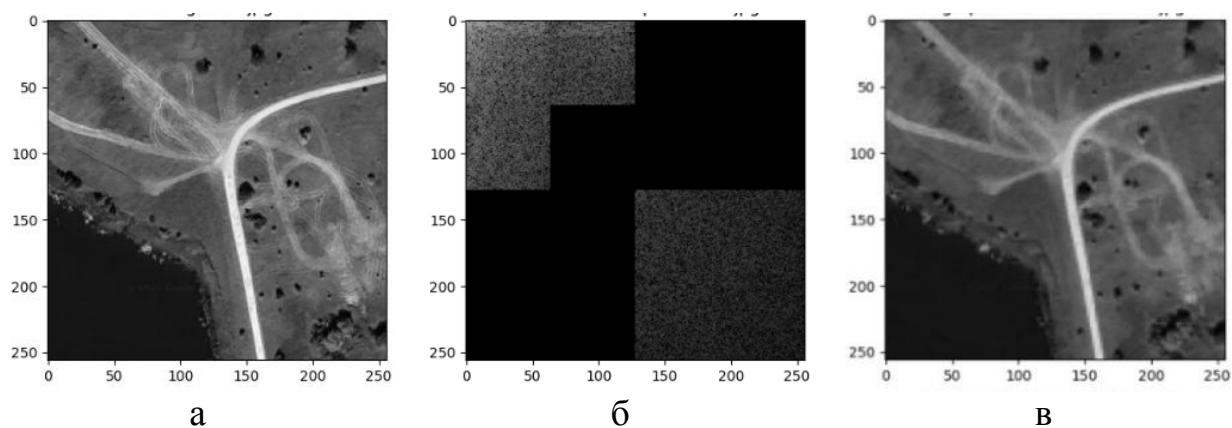


Рисунок 37 – Изображение после предварительной обработки

За счёт целенаправленной фильтрации низкочастотных секвент данных способ позволяет увеличить точность сходимости свёрточной нейронной сети в среднем на 7,8%, обеспечивая устойчивость признаков к изменению условий съёмки. Таким образом, представленный метод решает указанные задачи, сочетая высокую степень сжатия данных, адаптивность к изменениям сцены, заданную точность.

2.3 Выводы по второй главе

Предложен метод совмещения разнородных изображений и реализующие его алгоритмы, включающие: алгоритм обучения свёрточной нейронной сети, отличающийся от известных подходов введением последующей декомпозиции ее сиамской архитектуры для получения компактного кодера признаков; алгоритм формирования вектора-признаков изображения, обеспечивающий эффективное сжатие визуальной информации с 20 КБ до 256 байт на один фрагмент с сохранением ключевого семантического содержания; алгоритм поиска совпадений, позволяющий определить потенциальное положение летательного аппарата по текущему снимку с точностью до 300 пикселей на эталонной спутниковой карте. Способ предварительной обработки изображений в спектральном пространстве Уолша, который применяется в комплексе с традиционной нормализацией и аугментацией данных. За счёт целенаправленной фильтрации низкочастотных составляющих данных способ позволяет увеличить точность сходимости свёрточной нейронной сети в среднем на 7,8%,

обеспечивая устойчивость признаков к изменению условий съёмки. Теоретической основой метода является модифицированная сиамская нейронная сеть, преобразующая предобработанные изображения в инвариантные вектора-признаки, что обеспечивает эффективное сравнение разнородных бортовых и спутниковых снимков. Таким образом, представленный метод решает указанные задачи, сочетая высокую степень сжатия данных, адаптивность к изменениям сцены, заданную точность. Экспериментальная проверка эффективности и работоспособности всех предложенных решений подробно изложена в четвёртой главе.

ГЛАВА 3 АЛГОРИТМ ОПРЕДЕЛЕНИЯ ПОЛОЖЕНИЯ ТЕКУЩЕГО ИЗОБРАЖЕНИЯ ОТНОСИТЕЛЬНО СПУТНИКОВОГО ИЗОБРАЖЕНИЯ

Метод совмещения разнородных изображений подстилающей поверхности, описанный в главе 2, позволяет определить тайл из базы данных спутниковых изображений. Для получения более точного расположения предлагается найти общие признаки на двух изображениях и осуществить по ним привязку [100]. Алгоритм определения положения текущего изображения применяется в том случае, если предыдущее положение ЛА отличается больше, чем окрестность 3-5 тайлов. Данную область будем называть кандидатной областью. На рисунке 38 приведен пример кандидатной области. В качестве таких признаков предлагается использование ключевых точек.

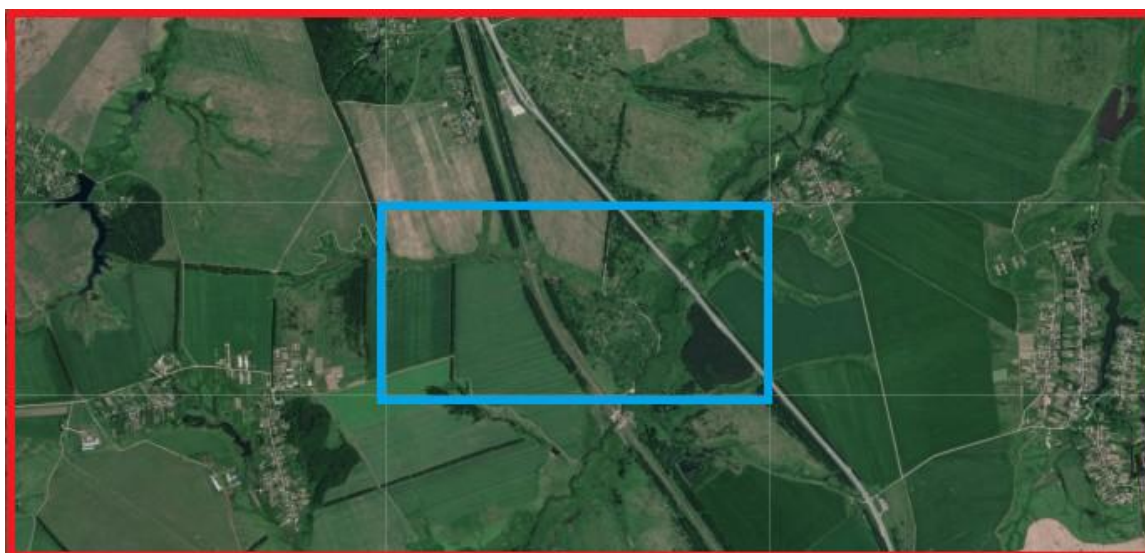


Рисунок 38 – Пример кандидатной области

Общая схема алгоритма состоит из следующих блоков:

1. Блок выделение (детектирования) особых точек.
2. Блок получения описания (дескриптора) особых точек.
3. Блок связывания (матчинга) особых точек.
4. Блок перерасчёта координат.

Для детектирования особых точек применяется алгоритм SIFT [29,105]. Он обладает высокой скоростью работы, а также наилучшими показателями

работы в условиях данной задачи (эксперимент, доказывающий это представлен в главе 4).

3.1 Задача уточнения положения ЛА по спутниковому снимку, полученному в результате работы нейросетевого алгоритма

В контексте работы с ЛА и мозаикой тайлов спутникового изображения, использование ключевых точек становится критически важным по нескольким причинам:

1. Требования к инвариантности

ЛА работает в условиях, требующих высокой инвариантности:

- изменение ракурса (углы поворота камеры);
- масштабирование (разная высота полета);
- освещение (разное время суток, погодные условия);
- частичные перекрытие объектов.

2. Пересечение тайлов

Кадр, полученный с ЛА, может одновременно располагаться на нескольких тайлах. Когда кадр с ЛА пересекает несколько тайлов, традиционные методы сопоставления, основанные на целостном изображении, становятся неэффективными. Ключевые точки позволяют:

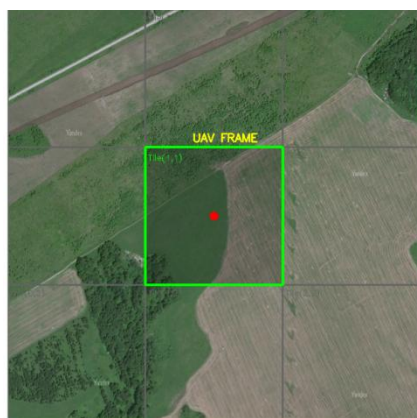
- локализовать соответствия на каждом отдельном тайле;
- объединить информацию из нескольких тайлов для точного позиционирования;
- игнорировать границы между тайлами как артефакты.

На рисунке 39 приведена визуализация пересечения кадра с ЛА и тайлами карты. Кадр с ЛА 256x256 (затемненный) накладывается поверх карты, выделенный желтой рамкой (UAV FRAME) – это границы кадра ЛА. Красная точка в центре кадра ЛА. Зеленые рамки показывают тайлы карты, которые пересекает кадр ЛА. Зеленым также отмечены номера тайлов.

Примеры позиций:

- кадр полностью в одном тайле (рис. 39,а);

- кадр пересекает два тайла (рис. 39,б, в);
- кадр пересекает три тайла (рис. 39,с, д);
- кадр находится на стыке четырех тайлов (рис. 39,е).



а



б



в



с



д



е

Рисунок 39 – Визуализация пересечения кадра с ЛА и тайлами карты

Алгоритм обработки граничных случаев состоит из нескольких шагов:

Шаг 1. Если кадр пересекает границу тайлов, формируется виртуальная мозаика из 2x2 тайлов.

Шаг 2. Детектирование ключевых точек выполняется по всей мозаике.

Шаг 3. Гомография строится для всей области [101].

Шаг 4. Географические координаты вычисляются с учетом весовых коэффициентов от каждого тайла.

Алгоритм определения положения текущего изображения, полученного с ЛА (снятого строго вертикально), относительно спутникового изображения осуществляет доуточнение после первичной (возможно, неточной) нейросетевой локализации в условиях отсутствия GPS [102,103]. В качестве исходных данных имеются:

- изображение, полученное с ЛА в текущий момент времени;
- БД изображений подстилающей поверхности разбитых на тайлы (256x256), среди которых предварительно с помощью нейросетевого алгоритма было выделено несколько кандидатов [94];
- предполагаемый маршрут: начальная точка, конечная точка.

Алгоритм уточнения положения текущего изображения относительно спутникового изображения, где с использованием сиамской нейросетевой архитектуры осуществляется отбор кандидатных областей в окрестности маршрута, а ключевые точки обеспечивают определение визуальной локализации ЛА внутри этих областей.

Алгоритм работает в условиях отсутствия систем глобальной навигации, для геопривязки и коррекции снимка используется комплексная информация: высота от радиовысотомера для масштабирования, углы ориентации (крен, тангаж) от инерциальной системы (гироскоп, акселерометры) для поворота изображения и учета перспективных искажений.

3.1.1 Формирование и структура базы данных

Спутниковое изображение подстилающей поверхности района использования ЛА имеет высокое разрешение и предварительно разбивается на тайлы. На рисунке 40 приведен алгоритм формирования базы данных спутниковых изображений. К основным свойствам тайлов в данной задаче можно отнести:

- перекрытие – соседние тайлы могут частично перекрываться для повышения точности сопоставления;
- масштабная инвариантность – тайлы могут храниться в нескольких масштабах для учета изменения высоты полета;
- ротационная инвариантность (устойчивость к изменению угла поворота / наклона камеры) – тайлы могут храниться в нескольких вариантах поворота или обрабатываться методом, устойчивым к вращению (аугментации поворота);
- структура БД – быстрый поиск по пространственному индексу (например, R-дерево) для ограничения области сравнения.

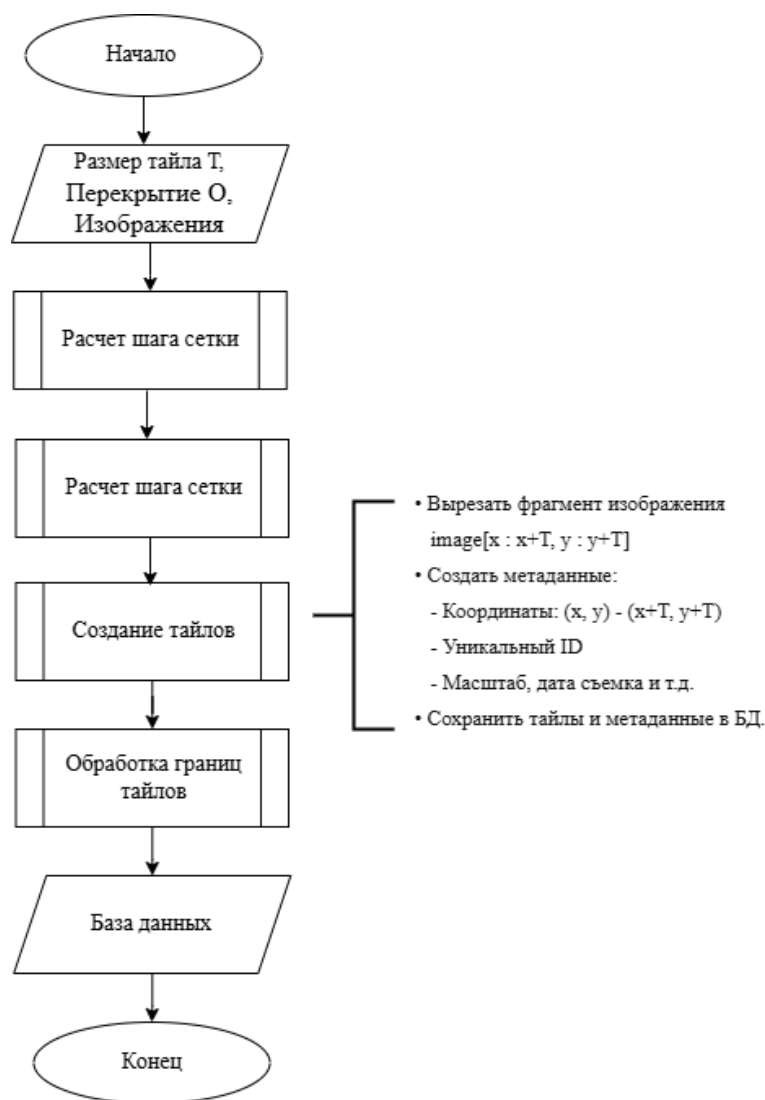


Рисунок 40 – Алгоритм формирования базы данных спутниковых изображений

На рисунке 41 приведена зависимость точности совмещения от степени перекрытия тайлов. Масштабная инвариантность обеспечивается, во-первых, с помощью обучения на разных углах поворотах, во-вторых, подстройка на основе данных с сенсоров ЛА. Ротационная инвариантность обеспечивается с помощью данных с сенсоров ЛА.

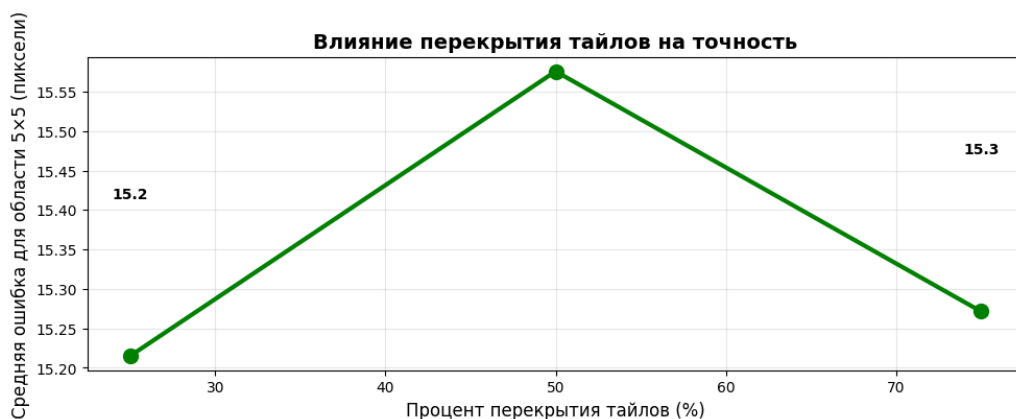


Рисунок 41 – Зависимость точности совмещения от степени перекрытия тай-
ЛОВ

Для каждого тайла сохраняется не только изображение, но и метаданные: географические координаты углов (верхнего левого и нижнего правого), уникальный идентификатор, разрешение, дата съемки.

На этапе формирования базы данных спутниковых снимков применяется предварительная обработка тайлов в пространстве Уолша [93,98].

Идея использования тайлов в задаче совмещения разнородных изображений заключается в ускорении поиска и возможности применять метод совмещения в реальном времени. Спутниковый снимок разбивается на тайлы для ускорения поиска и сравнения пары изображений нейросетевым алгоритмом [104]. Каждый тайл имеет свой уникальный идентификатор и метаданные (координаты углов, высоту и т.д.)

Использование в реальном времени заключается в следующем: ЛА делает снимок подстилающей поверхности, этот снимок подвергается предварительной обработке. Затем с использованием метода совмещения определяется

изображение из базы данных спутниковых изображений. Если находится тайл из БД с высокой степенью визуального сходства (минимальное косинусное расстояние), то местоположение ЛА считается по географическим координатам выбранного тайла. Если это не первый запуск, то оценивается предыдущее местоположение ЛА, если оно находится в окрестности 3x3 тайлов, то считаем, что совмещение осуществлено верно, если > 3 тайлов, то необходимо осуществить уточнение.

На рисунке 42 приведена зависимость точности совмещения от размера области поиска тайлов (3x3, 5x5 или 7x7).

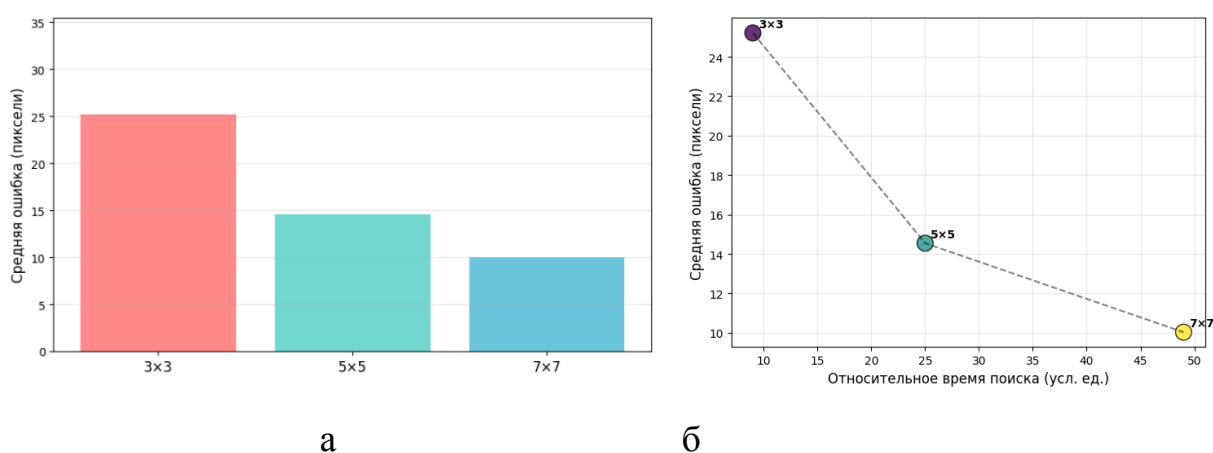


Рисунок 42 – Зависимость точности совмещения от размера кандидатной области тайлов:

а – зависимость точности от размера области поиска, б – зависимость времени поиска и средней ошибки в пикселях от размера области поиска

Для повышения эффективности работы алгоритма в различных условиях полета предлагается адаптивный выбор размера кандидатной области в зависимости от этапа полета и уверенности в предыдущей локализации. На этапах взлета, посадки или этапах с низкой уверенностью (<0.3) – используется область 7x7 тайлов для максимальной точности первичной привязки. На этапах со средней уверенностью (0.3-0.6) – используется область 5x5 тайлов, обеспечивающего баланс между скоростью и точностью совмещения. Уверенная локализация (>0.6) – считаем достаточной область 3x3 для контроля и коррекции с минимальными вычислительными затратами.

Такой подход позволяет сократить среднее время обработки на 35% по сравнению с постоянным использованием области 7x7, сохраняя требуемую точность на различных этапах полета.

Для оценки достоверности полученного уточнения используются следующие метрики: оценка репроекции – среднее расстояние между спроецированными точками изображения с ЛА и соответствующими точками на тайле спутникового снимка. Значение менее 15 пикселей соответствует высокому качеству совмещения; доля согласованных точек – отношение количества точек, подтвержденных RANSAC к общему числу найденных соответствий [105]. Значение более 0.5 говорит о геометрической согласованности найденных пар. В режиме реального полета не возможно оценить ошибку репроекции, так как нет информации про истинные координаты, поэтому данная метрика оценивалась только на тестовой выборке с известными координатами. На рисунке 43 приведена геометрическая согласованность точек в зависимости от размера области.

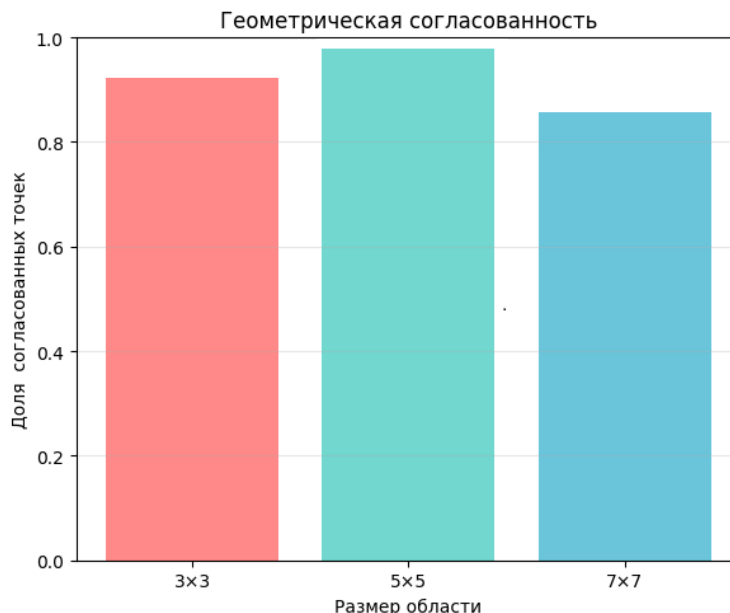


Рисунок 43 – Геометрическая согласованность точек

На представленной столбчатой диаграмме (рисунок 43) отражена зависимость доли согласованных ключевых точек от размера области поиска при уточнении совмещения методом SIFT + RANSAC [105].

3.2 Алгоритм детектирования и сопоставления ключевых точек

Сравнение методов детектирования ключевых точек для разнородных изображений подстилающей поверхности. Результаты приведены в таблице 1.

Таблица 1 – Сравнение методов детектирования ключевых точек

Метод	Время обнаружения (мс)	Найдено ключевых точек	Количество совпадающих точек	Доля совпадений %
ORB	48.76	100	10	10.0%
AKAZE	154.41	100	11	11.0%
FAST+ORB	170.9	100	8	8.0%
SIFT	254.02	11	5	45.5%
BRISK	583.93	100	7	7.0%
ORB+SIFT	1104.18	21	8	38.1%
TIMM Features	1112.4	14	9	64.3%
Deep Features	1491.8	20	6	30.0%
LOFTR	2800.15	228	22	9.6%

Для объективной оценки качества детекторов ключевых точек в задаче совмещения разнородных изображений подстилающей поверхности был проведен сравнительный анализ с использованием ROC-кривых (Receiver Operating Characteristic) [106]. Данный метод позволяет оценить соотношение между истинно положительных совпадений и долей ложных срабатываний при варьировании порога сопоставления.

ROC-анализ подтверждает, что выбор детектора ключевых точек должен определяться размером кандидатной области. Для области 3×3 , где критична точность совмещения, оптимальным является SIFT (AUC=0.94). При увеличении области до 5×5 и 7×7 целесообразно использовать ORB (AUC=0.81), что позволяет обрабатывать 49 тайлов за приемлемое время.

Схема алгоритма определения положения текущего изображения относительно спутникового изображения представлена на рисунке 44.

В качестве алгоритма получения описания и связывания особых точек используется SIFT и FLANN. А для подавления ложных пар – PROSAC [107].

После получения целевых пар точек производится анализ корректности полученных связей. Если количество согласованных точек и средняя ошибка превышает заданный порог, уточнение положения считается неудачным и оставляется положение, которое было до уточнения.

Алгоритм совмещения изображений на основе ключевых точек можно разбить на последовательные этапы:

Шаг 1. Детектирование ключевых точек.

На этом этапе на каждом из совмещаемых изображений независимо обнаруживаются точки, устойчивые к изменениям масштаба, поворота, освещенности и небольшим изменениям ракурса.

Шаг 2. Построение дескриптора.

Для каждой обнаруженной ключевой точки вычисляется ее дескриптор – числовой вектор, который характеризует локальную область изображения вокруг этой точки [108].

Шаг 3. Сопоставление ключевых точек.

Производится поиск пар соответствий между дескрипторами точек первого и второго изображения [109]. Наиболее распространенный метод – поиск ближайшего соседа (Nearest Neighbor Search). Для каждого дескриптора из первого изображения находят два наиболее похожих (ближайших по метрике, например, евклидову расстоянию) дескриптора из второго изображения. Для отсева ложных соответствий часто применяют тест отношения ближайших соседей (Lowe's ratio test): соответствие считается надежным, если расстояние до ближайшего соседа существенно меньше (например, в 0.7-0.8 раза), чем расстояние до следующего ближайшего соседа.

Шаг 4. Оценка матрицы преобразования.

Набор найденных пар соответствий как правило содержит ложные сопоставления. Для оценки параметров N применяется алгоритм RANSAC (RANDOM SAMPLE CONSENSUS - "случайная выборка и консенсус") [110].

3.4 Алгоритм уточнения на основе ключевых точек

На рисунке 43 приведена схема алгоритма уточнения положения текущего изображения относительно спутникового изображения.

Прилагается уточнить совмещение спутникового изображения и изображения с ЛА с предварительно отобранными с помощью нейросетевых технологий тайлов кандидатов спутникового изображения. Поскольку кадр может располагаться на нескольких тайлах, традиционные методы совмещения для цельных изображений не работают. Сначала с помощью метода совмещения на основе нейросетевых технологий осуществляется отбор тайлой из БД, а затем на основе ключевых точек обеспечивается точное визуальное совмещение, объединяется информация с нескольких тайлов, игнорируются границы между тайлами [111].

Алгоритм уточнения состоит из нескольких шагов:

1. Подготовка области поиска.

Для каждого тайла кандидата осуществляется склейка между соседними тайлами, формируется мозаика тайлов, что позволяет решить научную задачу пересечения границ.

2. Детектирование ключевых точек.

На текущем карде ЛА и на каждой собранной мозаике независимо детектируются ключевые точки и вычисляются их дескрипторы.

3. Сопоставление ключевых точек.

Для каждой пары кадр с ЛА и мозаика тайлов кандидатов выполняется взаимно-однозначное сопоставление дескрипторов ключевых точек.

4. Геометрическая верификация.

На основе найденных пар соответствий для каждой мозаики необходимо оценить геометрическую модель преобразования (гомографии) между кадром

с ЛА и спутниковой мозаикой, используя метод Ransac [112]. Модель с максимальным числом согласованных точек и наименьшей ошибкой считается наилучшей.

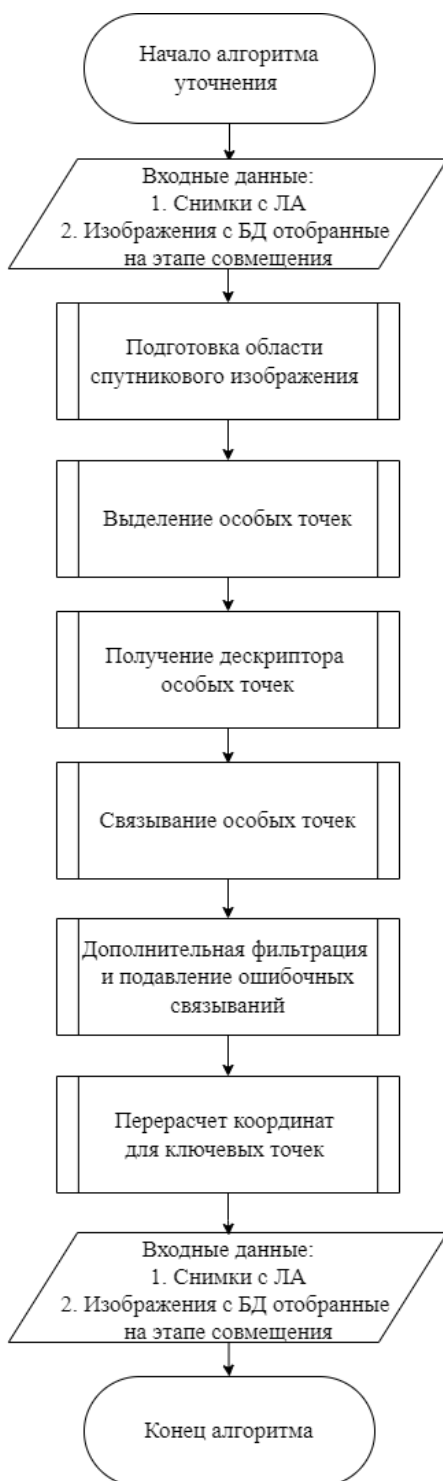


Рисунок 44 – Схема алгоритма уточнения положения текущего изображения относительно спутникового изображения

5. Вычисление локальных координат и принятие решения об уточнении.

Используя наилучшую модель гомографии, вычислить точные координаты углов кадра ЛА в системе координат спутниковой мозаики. Если количество согласованных точек и средняя ошибка превышает заданный порог, уточнение положения считается неудачным и оставляется положение, которое было до уточнения.

Для ускорения повторных совмещений в одном районе предлагается кэшировать ключевые точки тайлов. При повторном пролете над тем же районом детектирование выполняется только для кадра с ЛА, что сокращает время обработки.

Оффлайн этап включает в себя:

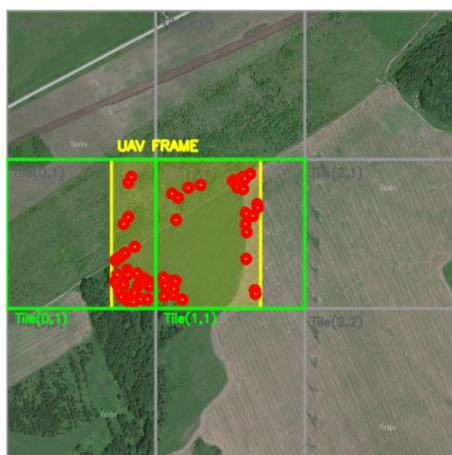
- базу данных тайлов;
- предварительную обработку, которая заключается в однократном детектировании ключевых точек и вычисление дескрипторов для каждого тайла;
- дисковый кэш, состоящий из предвычисленных данных в форматах .pru (для дескрипторов) и .json (для метаданных ключевых точек).

Онлайн этап (бортовая система ЛА) реализует:

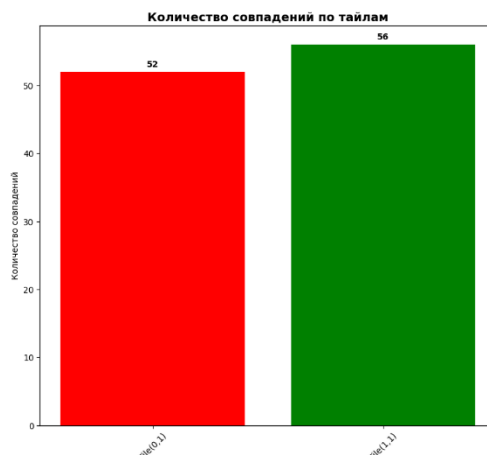
- обработку текущего кадра с ЛА;
- кэш в оперативной памяти (ОЗУ) – буфер из 50-100 наиболее часто запрашиваемых тайлов;
- загрузку из дискового кэша, при отсутствии требуемого тайла в ОЗУ производится его загрузка с диска;
- сопоставление ключевых точек;
- геометрическая верификация;
- вычисление координат совмещения на основе наилучшей модели преобразования.

На рисунках 45а,в,д приведена визуализация работы алгоритма на стыке тайлов. На рисунках 45б,г,е – столбчатая диаграмма, показывающая количе-

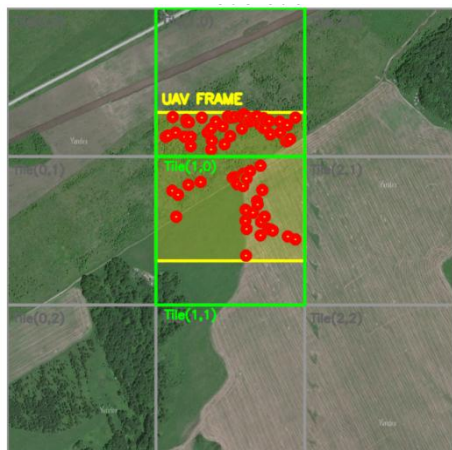
ство верно соединенных точек по тайлам. Изображения зеленым цветом разделены на тайлы и подписаны их номера. Желтым выделен прямоугольник, где в настоящий момент времени располагается ЛА.



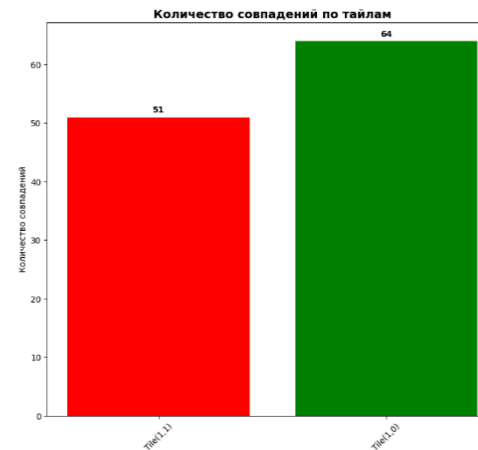
а



б



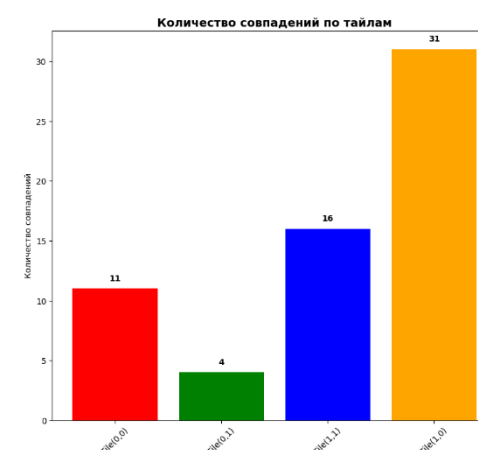
в



г



д



е

Рисунок 45 – Визуализация работы алгоритма на стыке тайлов

На рисунках 45а, в, д приведена визуализация работы алгоритма на стыке двух и четырех тайлов. Исходя из рисунка, можно сделать вывод, что ключевые точки обнаружены и соединены корректно.

Перспективным направлением развития алгоритма является использование глубоких нейронных сетей для детектирования ключевых точек, что позволит повысить устойчивость к сезонным изменениям, однако требует оптимизации для встраиваемых систем.

3.4.2 Оценка вычислительной сложности

Для обеспечения работы алгоритма на бортовом оборудовании с ограниченными вычислительными ресурсами (1 ядро ARM Cortex-A76, 2.4 GHz, RAM ≤ 200 МБ) проведен анализ вычислительной сложности различных этапов.

Вычислительная сложность преобразования Уолша с использованием быстрого алгоритма для изображения размером 256x256 вычисляется следующим способом:

Количество операций для двумерного преобразования:

$$C_{Walsh}(N) = 2 \cdot N^2 \cdot \log_2 N, \quad (52)$$

где N – размерность изображения, $\log_2 N$ – количество проходов преобразования, коэффициент 2 учитывает прямое и обратное преобразования. Для $N = 256$ получим $C_{Walsh}(N) = 1,048,576$ операций. Каждая операция представляет собой сложение и вычитание, что соответствует примерно 1.05×10^6 арифметических операций.

Декомпозированная архитектура предполагает разложение стандартных сверток на комбинацию более простых операций. Для сямской сети используются следующие виды декомпозиции:

1. Точечная свертка имеет сложность, вычисляемую следующим образом:

$$C_{pw}(N, W, C_{in}, C_{out}) = H \cdot W \cdot C_{in} \cdot C_{out}, \quad (53)$$

где H, W – пространственные размеры, C_{in} – количество входных каналов, C_{out} – количество выходных каналов.

2. Глубинные свертки:

$$C_{dw}(H, W, C) = H \cdot W \cdot C \cdot 9. \quad (54)$$

3. Полносвязные слои: для полносвязного слоя с входным слоем n_{in} и выходным слоев n_{out} вычислительная сложность вычисляется следующим образом:

$$C_{fc}(n_{in}, n_{out}) = n_{in} \cdot n_{out}. \quad (55)$$

Расчет общей сложности декомпозированной архитектуры приведен в таблице 2.

Таблица 2 – Общая сложность декомпозированной сети

Блок	Входной размер	Операция	Выходной размер	Сложность (операций)
Вход	256×256×3	-	256×256×3	-
Блок 1	256×256×3	Depthwise 3×3	256×256×3	1,769,472\$
	256×256×3	Pointwise 1×1	256×256×32	6,291,456\$
Пулинг	256×256×32	MaxPool 2×2	128×128×32	-
Блок 2	128×128×32	Depthwise 3×3	128×128×32	$128^2 \cdot 32 \cdot 9 = 4,718,592\$$
	128×128×32	Pointwise 1×1	128×128×64	$128^2 \cdot 32 \cdot 64 = 33,554,432\$$
Пулинг	128×128×64	MaxPool 2×2	64×64×64	-
Блок 3	64×64×64	Depthwise 3×3	64×64×64	2,359,296\$
	64×64×64	Pointwise 1×1	64×64×128	33,554,432\$
Пулинг	64×64×128	MaxPool 2×2	32×32×128	-
Блок 4	32×32×128	Depthwise 3×3	32×32×128	1,179,648\$
	32×32×128	Pointwise 1×1	32×32×256	33,554,432\$

Блок	Входной размер	Операция	Выходной размер	Сложность (операций)
Пулинг	32×32×256	GlobalAvgPool	256	-
FC1	256	Dense	512	131,072\$
FC2	512	Dense	256	131,072\$
FC3	256	Dense	128	32,768\$

Суммарная сложность нейросетевой части составляет 1.17×10^8 арифметических операций для обработки одного изображения.

Вычислительная сложность этапа ключевых точек. Детектирование SIFT состоит из построения масштабного пространства (DoG) $C_{DoG} \approx 26100000$ операций, детектирование ложных экстремумов $C_{extrema} \approx 10223616$ операций, уточнение положения ключевых точек = 75 000 операций, вычисление дескриптора $C_{desc} = 1\,024\,000$ операций, сопоставление дескрипторов с использованием FLANN. FLANN использует kd-деревья для ускорения поиска. Для K ключевых точек кадра и K ключевых точек тайла:

$$C_{match}^{FLANN}(K, T) = O(K \cdot \log K \cdot T), \quad (56)$$

где T – количество тайлов кандидатной области. Для кандидатной области 3×3 сложность сопоставления составляет 40 500 операций.

Для детекторов ORB для ключевых точек:

$$C_{ORB} = K \cdot n_{samples}, \quad (57)$$

где $n_{samples} = 256$, K в среднем = 500. $C_{ORB} = 128,000$ операций сравнения.

Выводы по вычислительной сложности

1. Декомпозированная нейросеть требует 117 млн операций, что в 2.73 раза меньше стандартной архитектуры.
2. SIFT-детектирование для кадра с БПЛА требует 37.8 млн операций, что составляет 24% от общего объема вычислений.

3. Благодаря кэшированию ключевых точек тайлов, сопоставление требует всего 0.22 млн операций (менее 0.2% от общего объема).

4. Преобразование Уолша добавляет всего 1.05 млн операций (менее 1% от общего объема).

5. Полная вычислительная сложность составляет 157 млн операций, что соответствует 62.8 мс на ARM Cortex-A76.

Таким образом предложенное техническое решение с декомпозированной нейросетевой архитектурой, преобразованием Уолша и уточнением SIFT (с кэшированием дескрипторов тайлов) обеспечивает работу в режиме реального времени на бортовой аппаратуре. Несмотря на то, что SIFT в 36 раз вычислительно сложнее FAST, его использование оправдано более высокой точностью и устойчивостью, а общее время обработки остается в пределах допустимого благодаря кэшированию.

3.5 Вычисление географических координат методом билинейной интерполяции

Для перевода пиксельных координат на тайле в географические используется модель билинейной интерполяции. Вычисление осуществляется на прямоугольном тайле.

Исходные данные:

- (lat_{TL}, lon_{TL}) – координаты верхнего левого угла тайла.
- (lat_{BR}, lon_{BR}) – координаты нижнего правого угла тайла.
- W, H – ширина и высота тайла в пикселях.
- (x, y) – координаты точки на тайле (в пикселях).

Выходные данные:

- (lat_{est}, lon_{est}) — расчётные географические координаты точки.

Математическая модель

Геопривязка тайла определяется координатами двух диагонально противоположных углов. Предполагается, что тайл является прямоугольником в географической проекции, а изменение координат внутри него может быть аппроксимировано линейной функцией.

Нормализация координат.

Введём безразмерные параметры u и v , описывающие положение точки относительно границ тайла:

$$u = \frac{x}{W}, v = \frac{y}{H}, \quad (58)$$

где $u \in [0,1]$ — относительная горизонтальная координата (0 — левая граница, 1 — правая граница), $v \in [0,1]$ — относительная вертикальная координата (0 — верхняя граница, 1 — нижняя граница).

Интерполяция долготы.

Долгота изменяется линейно по горизонтали от lon_{TL} до lon_{BR} :

$$lon_{est}(u) = lon_{TL} + u \cdot (lon_{BR} - lon_{TL}). \quad (59)$$

Интерполяция широты.

Широта изменяется линейно по вертикали от lat_{TL} до lat_{BR} . Учитывая, что значения широты убывают при движении сверху вниз (от lat_{TL} к lat_{BR}):

$$lat_{est}(v) = lat_{TL} - v \cdot (lat_{TL} - lat_{BR}). \quad (60)$$

Итоговая система уравнений.

$$\begin{aligned} lat_{est} &= lat_{TL} - \frac{y}{H} \cdot (lat_{TL} - lat_{BR}), \\ lon_{est} &= lon_{TL} + \frac{x}{W} \cdot (lon_{BR} - lon_{TL}). \end{aligned} \quad (61)$$

3.6 Выводы по третьей главе

Представленный во второй главе метод позволяет определить предполагаемый тайл основной карты, в котором в текущий момент времени находится ЛА. В том случае, если требуется получение более точной оценки местоположения, метод может быть дополнен алгоритмом уточнения. Поскольку кадр может располагаться на нескольких тайлах, традиционные методы совмещения для цельных изображений не работают. На основе ключевых точек обеспечивается точное визуальное совмещение, так как объединяется информация с нескольких тайлов, игнорируются границы между ними.

Предложенный в третьей главе алгоритм решает задачу пересечения кадра ЛА с границами нескольких эталонных тайлов, используя для этого детектирование, сопоставление и геометрическую верификацию локальных особенностей — ключевых точек. Для обеспечения работы в условиях строгих вычислительных ограничений бортовой аппаратуры (не более 200 МБ RAM, 1 ядро ARM Cortex-A76, 2.4 GHz,) в качестве детектора выбран алгоритм SIFT, доказавший свою эффективность в рамках данной задачи. Экспериментальные результаты приведены в четвертой главе. Основным результатом главы является алгоритм уточнения положения текущего изображения относительно спутникового изображения, который за счёт формирования мозаик из тайлов-кандидатов, точного связывания ключевых точек и оценки геометрической модели (гомографии) методом RANSAC позволяет существенно сократить зону неопределенности позиционирования — с 300 пикселей (результат метода совмещения разнородных изображений (результат метода из главы 2) до 20 пикселей, что доказывает третье положение, выносимое на защиту.

ГЛАВА 4 АПРОБАЦИЯ И ИССЛЕДОВАНИЕ МЕТОДА СОВМЕЩЕНИЯ РАЗНОРОДНЫХ ИЗОБРАЖЕНИЙ ПОДСТИЛАЮЩЕЙ ПОВЕРХНОСТИ И АЛГОРИТМОВ ЕГО РЕАЛИЗУЮЩИХ

4.1 Описание программно-алгоритмического комплекса

На рисунке 46 приведена диаграмма компонентов, которая наглядно показывает состав системы, интерфейса и потоки данных между основными модулями.

Разработанный комплекс реализован в виде модельной системы, включающей в себя следующие компоненты:

- база данных цифровой карты местности разбитой на тайлы;
- модуль предобработки изображений;
- модуль формирования вектора-признаков;
- модуль совмещения разнородных изображений;
- модуль уточнения совмещения разнородных изображений.

Поскольку метод, описанный в главе 2 требует значительных вычислительных ресурсов, для алгоритма уточнения предполагается использовать не более 200 МБ RAM и не более 1 ядра (ARM Cortex-A76, 2.4 GHz) на протяжении не более 210 мс.

Данный комплекс интегрирует современные подходы компьютерного зрения, обработки изображений. Архитектура системы и взаимодействия ее компонентов изображены на структурной диаграмме.

Таким образом предложенный комплекс представляет собой целостное программно-алгоритмическое решение, архитектура которого обеспечивает модульность, масштабируемость и возможность дальнейшего совершенствования методов и алгоритмов совмещения разнородных изображений подстилающей поверхности. Система экспериментально протестирована и оценена ее точность в реальных условиях.

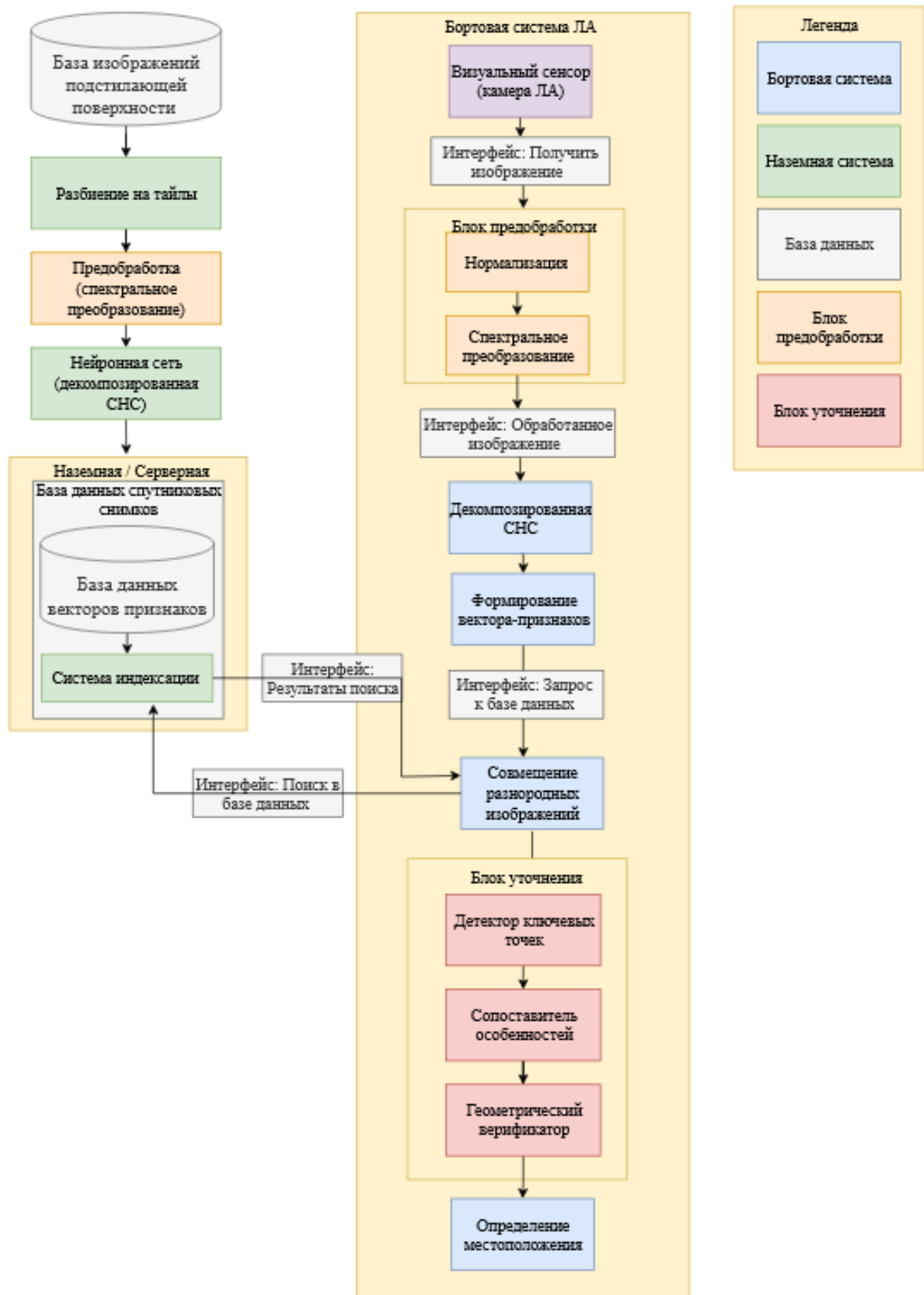


Рисунок 46 – Диаграмма компонентов программного комплекса

4.2 Метрики качества совмещения изображений

Для оценки эффективности предложенного метода совмещения использовались следующие метрики:

1. Метрики информационного поиска

– точность поиска (precision) вычисляется как отношение выданных релевантных изображений к общему количеству выданных изображений:

$$p = \frac{RF}{RF + IF}, \quad (62)$$

где RF – количество выданных релевантных изображений, IF – количество выданных нерелевантных изображений.

– полнота поиска (recall) вычисляется как отношение выданных релевантных изображений к общему числу релевантных изображений:

$$r = \frac{RF}{RF + RN}, \quad (63)$$

где RN – количество релевантных не выданных значений.

– F1-мера (сбалансированная F-мера, мера Ван Ризбергена, F1 - measure) применяется как единая мера, объединяющая полноту и точность, которая вычисляется следующим образом:

$$F_1 = \frac{2rp}{r + p}. \quad (64)$$

F1-мера (значение F1-меры не может быть ниже полноты и точности и превышать их среднеарифметическое):

$$\min(r, p) \leq F_1 \leq \frac{r+p}{2}; \quad (65)$$

2. Метрики точности совмещения

– среднеквадратичная ошибка MSE (Mean Squared Error):

$$MSE(I_k - I_s) = \frac{1}{256 \times 256} \sum_{i=1}^{256} \sum_{j=1}^{256} (I_k(i, j) - I_s(i, j)) \quad (66)$$

– максимальная ошибка (Max Error) – наибольшее отклонение оценки от истинного значения;

3. Метрики сходства вектора признаков

– косинусное расстояние:

$$\text{cosine}_{\text{similar}}(A, B) = \frac{(A, B)}{\|A\| \cdot \|B\|}, \quad (67)$$

где (A, B) – скалярное произведение векторов A и B , $\|A\|$ – норма вектора A , вычисляемая как $\sqrt{\sum_{i=1}^m A_i^2}$, i – индекс суммирования, M – размер каждого из векторов A и B , $\|B\|$ – норма вектора B , вычисляемая аналогично. Значение косинусного сходства варьируется от -1 до 1:

– значение 1 означает, что векторы идентичны (максимальная схожесть).

– значение 0 означает, что векторы ортогональны (нет схожести).

– значение -1 указывает на противоположные направления (в контексте поиска похожих изображений это обычно не имеет смысла, так как векторы признаков обычно не имеют отрицательных значений).

– евклидово расстояние (L_2 -норма):

$$d_{\text{euclid}}(A, B) = \sqrt{\sum_{i=1}^m (A_i - B_i)^2} \quad (68)$$

– манхэттенское расстояние (L_1 -норма):

$$d_{\text{manhattan}}(A, B) = \sqrt{\sum_{i=1}^m |A_i - B_i|} \quad (69)$$

4.3 Экспериментальные исследования

4.3.1 Исследования алгоритмов предварительной обработки изображений подстилающей поверхности

Алгоритм предварительной обработки включает в себя частотную фильтрацию, для того, чтоб оставить только геометрические паттерны. Рассмотрим влияние частотных компонент на сходимость СНС. Результат экспериментов по фильтрации той или иной частотной области приведен ниже. В качестве

параметров сети рассмотрено Validation Accuracy (пары спутниковых снимков в разных локациях с искажениями) и ValidationLoss (сред. квадр. ошибка).

Пусть валидационная выборка состоит из M пар данных: $\{(I_k, y_k)\}_{k=1}^M$, где I_k — входное изображение, а $y_k \in \{1, \dots, C\}$ — его истинный класс (например, индекс корректного тайла). Модель нейронной сети $f(\cdot; \theta)$ с параметрами θ для каждого входа возвращает вектор оценок $\mathbf{s}_k = f(I_k; \theta) \in \mathbb{R}^C$, который преобразуется функцией активации softmax в вектор вероятностей $\mathbf{p}_k = \sigma_{\text{softmax}}(\mathbf{s}_k)$. Прогнозируемый класс определяется как $\hat{y}_k = \arg \max_c p_{k,c}$.

1. Validation Accuracy (Валидационная точность). Точность на валидационной выборке вычисляется как доля верно классифицированных примеров:

$$\text{Validation Accuracy} = \frac{1}{M} \sum_{k=1}^M \mathbb{I}(\hat{y}_k = y_k), \quad (70)$$

где $\mathbb{I}(\cdot)$ — индикаторная функция, равная 1, если условие истинно, и 0 в противном случае.

2. Validation Loss (Валидационная функция потерь). Для задачи классификации используется функция потерь категориальная кросс-энтропия. Данная функция количественно оценивает расхождение между предсказанным распределением вероятностей и истинным распределением меток классов.

$$\text{Validation Loss} = -\frac{1}{M} \sum_{k=1}^M y_i * \ln \hat{y}_i + (1 - y_i) * \ln(1 - \hat{y}_i). \quad (71)$$

где M — количество примеров в валидационной выборке, y_i — истинная метка класса (0 или 1), \hat{y}_i — предсказанная моделью вероятность принадлежности к классу.

Общая форма передаточной функции.

Все фильтры спектральных коэффициентов являются прямоугольными масками (пропускание 1 или 0) в спектральной области Уолша. Их передаточную функцию $H(u, v)$ можно описать как функцию индикатора (прямоугольную функцию) для определённой области \mathbf{D} в спектральном пространстве:

$$H(u, v) = \begin{cases} 1, & \text{если } (u, v) \in D \\ 0, & \text{если } (u, v) \notin D \end{cases} \quad (72)$$

где $u, v \in [0; 255]$ – индексы коэффициентов в преобразовании Уолша для изображения 256×256 , D – область, которую необходимо оставить без изменения.

Операция применения фильтра к спектру W (спектральная матрица после прямого преобразования) выглядит следующим образом:

$$W_f(u, v) = H(u, v) * W(u, v) \quad (73)$$

Далее к W_f необходимо выполнить обратное преобразование Уолша для получения отфильтрованного изображения.

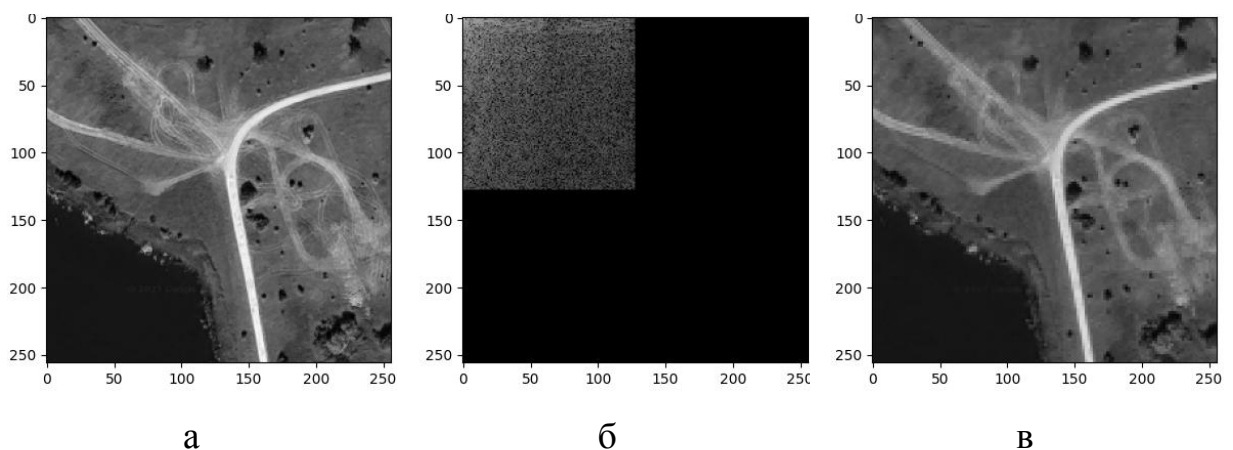
Фильтр 1. Низкочастотный (Low-Pass Filter)

Низкочастотный фильтр эквивалентен сохранению только $W_f[0:128, 0:128]$. Формула передаточной функции:

$$H_1(u, v) = \text{rect}\left(\frac{u}{128}\right) * \text{rect}\left(\frac{v}{128}\right), \quad (74)$$

где $\text{rect}(x)$ – единичный прямоугольный импульс, равный 1 при $|x| \leq 0.5$. В дискретном случае это означает, что функция равна 1 для $u = 0 \dots 127$ и $v = 0 \dots 127$.

На рисунке 47 приведено а, г – пара изображений, полученная с разных сенсоров, б, д – визуализация применения фильтра, в, е – пара изображений после обратного преобразования Уолша.



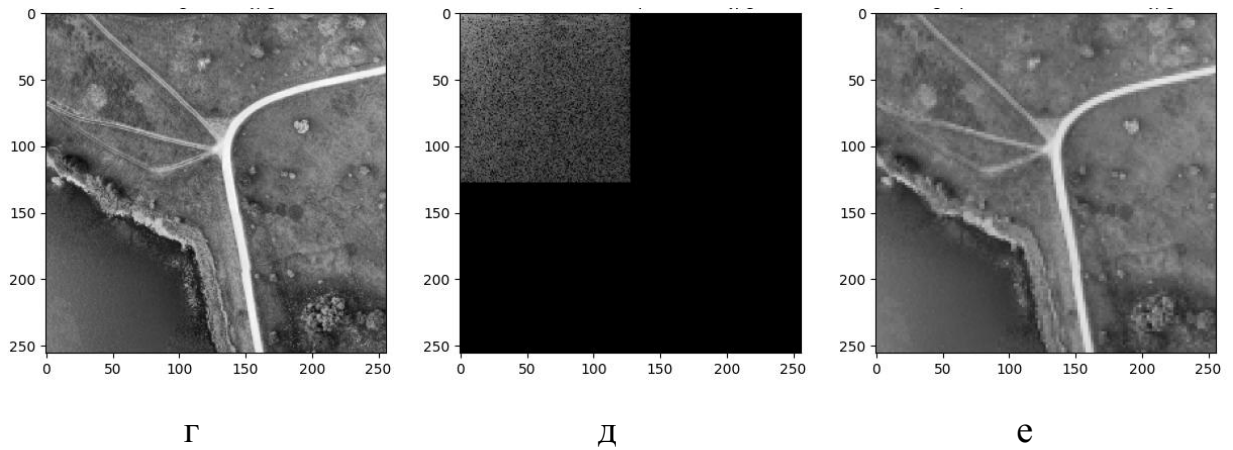


Рисунок 47 – Пара изображений после применения фильтра 1

После обратного преобразования изображение имеет размытые нечеткие контуры объектов, отфильтрованы шум, мелкие детали.

Фильтр 2. Низкочастотный + Средние частоты по осям (Low-Pass + Mid-Band Cross)

Низкочастотный фильтр эквивалентен сохранению $W_f[0:64, 0:64]$. Дополнительно к данному блоку сохранены: горизонтальная полоса низких/средних секвентностей: $W_f[0:64, 64:256]$ и вертикальная полоса низких/средних секвентностей: $W_f[64:256, 0:64]$. Формула передаточной функции:

$$H_2(u, v) = \text{rect}\left(\frac{u}{64}\right) * \text{rect}\left(\frac{v}{64}\right) + \left[\text{rect}\left(\frac{u}{64}\right) * \left(1 - \text{rect}\left(\frac{v}{64}\right)\right) \right] + \left[\left(1 - \text{rect}\left(\frac{u}{64}\right)\right) * \text{rect}\left(\frac{v}{64}\right) \right], \quad (75)$$

Данный фильтр можно описать как объединение трех областей. Данный фильтр пропускает больше областей, чем фильтр 1.

На рисунке 49 приведено а), г) – пара изображений, полученная с разных сенсоров, б), д) – визуализация применения фильтра, в), е) – пара изображений после обратного преобразования Уолша.

После обратного преобразования сохраняется общая структура, основные контуры (и горизонтальные, и вертикальные) становятся более четкими по сравнению с фильтром 1. Фильтр 2 отсеивает средние и высокие секвентности,

не лежащие на осях (диагональные паттерны и текстуры будут подавлены).

Мелкие детали по-прежнему отсутствуют.

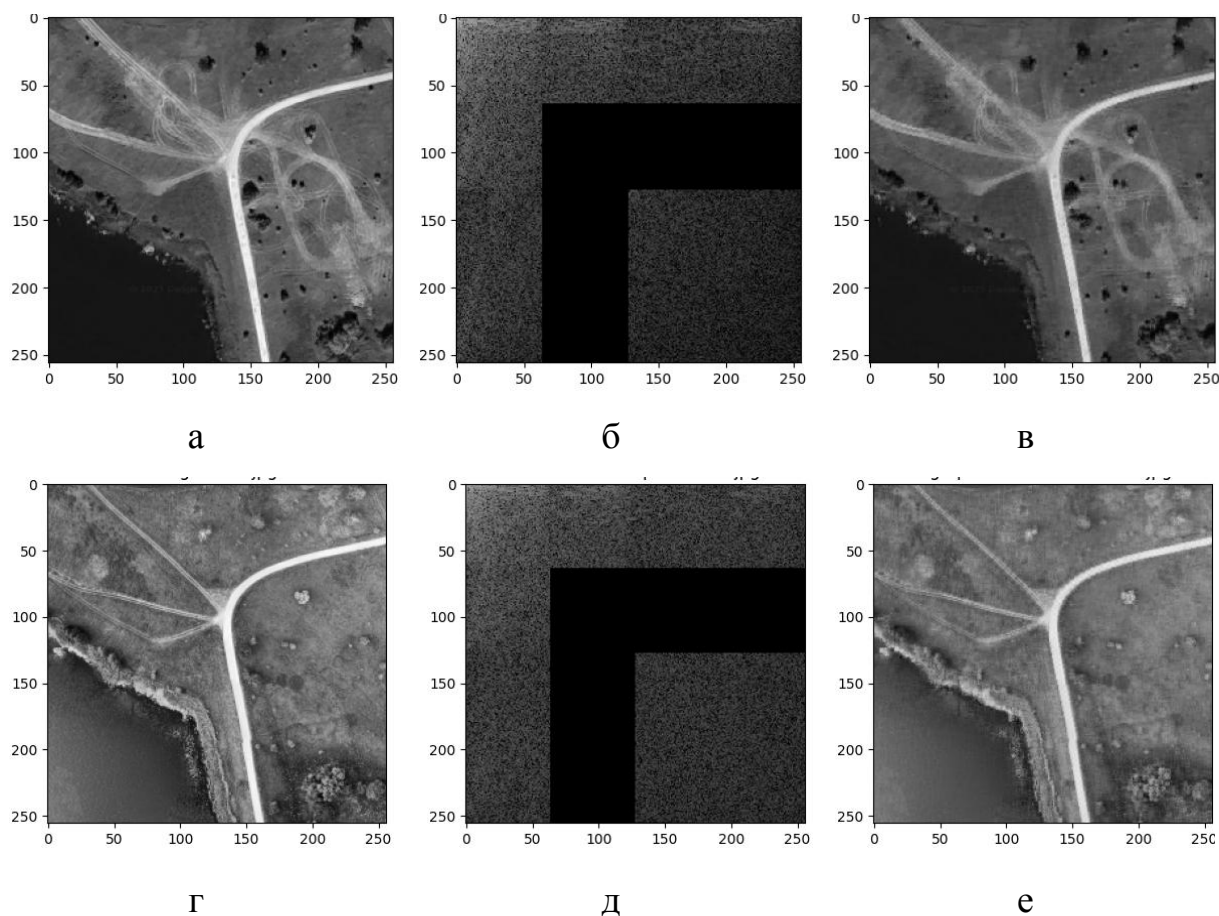


Рисунок 48 – Пара изображений после применения фильтра 2

Фильтр 3. Улучшенный низкочастотный фильтра (фильтр 2 + дополнительные ограничения)

К маске фильтра 2 (формула 75) добавлены дополнительные ограничения: $W_d[0:64, 128:256]$ и $W_d[128:256, 0:64]$. То есть, из сохранённых в Фильтре 2 горизонтальной и вертикальной полос срезаем низкие секвенты. Формула передаточной функции:

$$H_3(u, v) = \begin{cases} 1, & \text{если } (u < 64, v < 128) \text{ ИЛИ } (v < 64, u < 128) \\ 0, & \text{иначе} \end{cases} \quad (76)$$

Или тоже самое:

$$H_3(u, v) = \text{rect}\left(\frac{u}{128}\right) * \text{rect}\left(\frac{v}{64}\right) + \text{rect}\left(\frac{u}{64}\right) * \text{rect}\left(\frac{v}{128}\right) - \text{rect}\left(\frac{u}{64}\right) ** \text{rect}\left(\frac{v}{64}\right) \quad (77)$$

На рисунке 49 приведено а, г – пара изображений, полученная с разных сенсоров, б,д – визуализация применения фильтра, в,е – пара изображений после обратного преобразования Уолша.

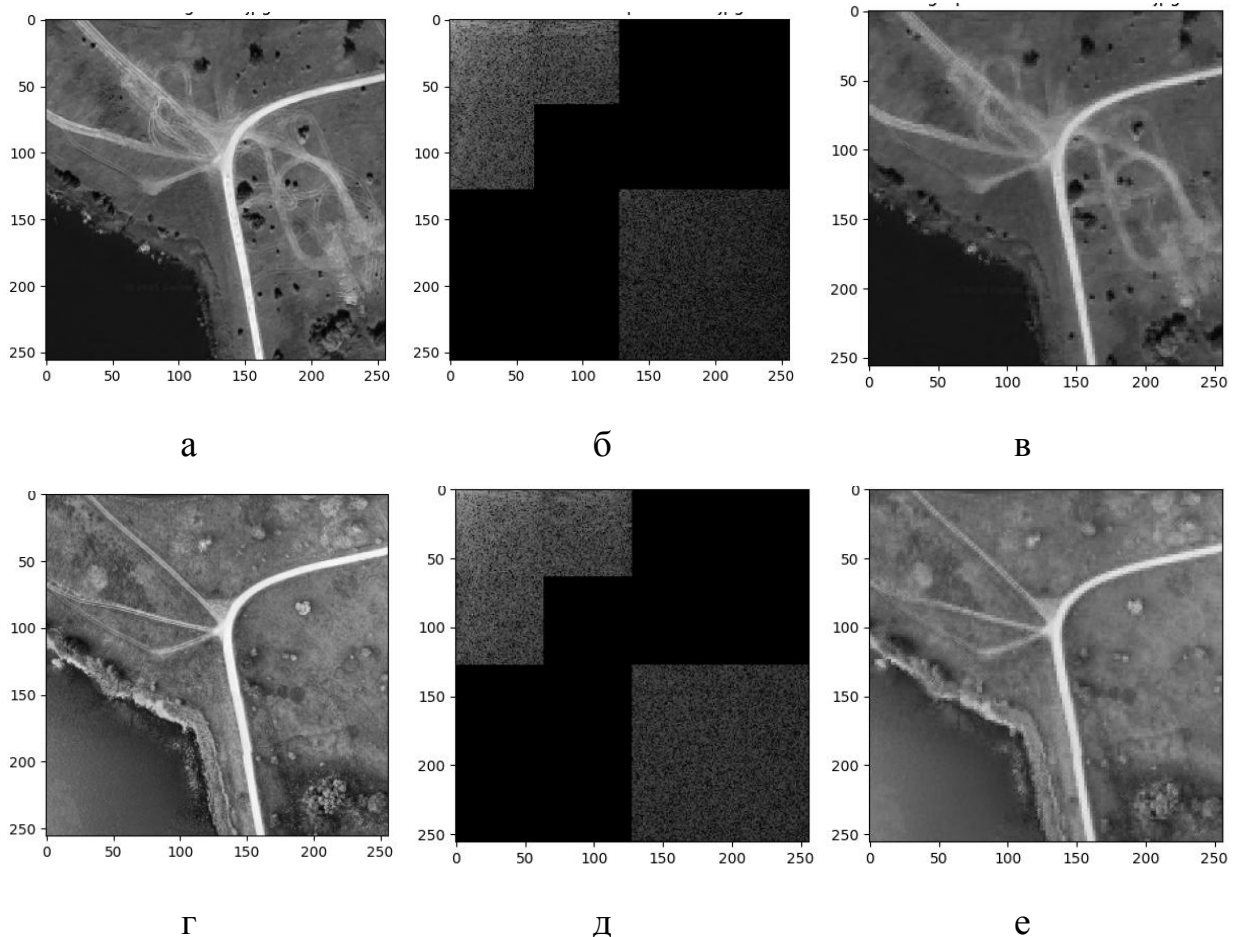


Рисунок 49 – Пара изображений после применения фильтра 3

Данный фильтр сохраняет всю информацию из блока 64x64, сохраняет также высокие частоты. Фильтр отсеивает излишнюю детализацию.

Фильтр 4 Высоких частот (Filter 3 + подавление высоких секвентностей)

К маске Фильтра 3 добавлено дополнительное ограничение – правый нижний квадрат 128x128 ($W_f[128:256, 128:256] = 0$), что соответствует подавлению высоких секвентностей. Формула передаточной функции:

$$H_4(u, v) = H_3(u, v) * (1 - \text{rect}\left(\frac{u - 128}{128}\right) * \text{rect}\left(\frac{v - 128}{128}\right)) \quad (78)$$

На рисунке 50 приведено а, г – пара изображений, полученная с разных сенсоров, б, д – визуализация применения фильтра, в, е – пара изображений после обратного преобразования Уолша.

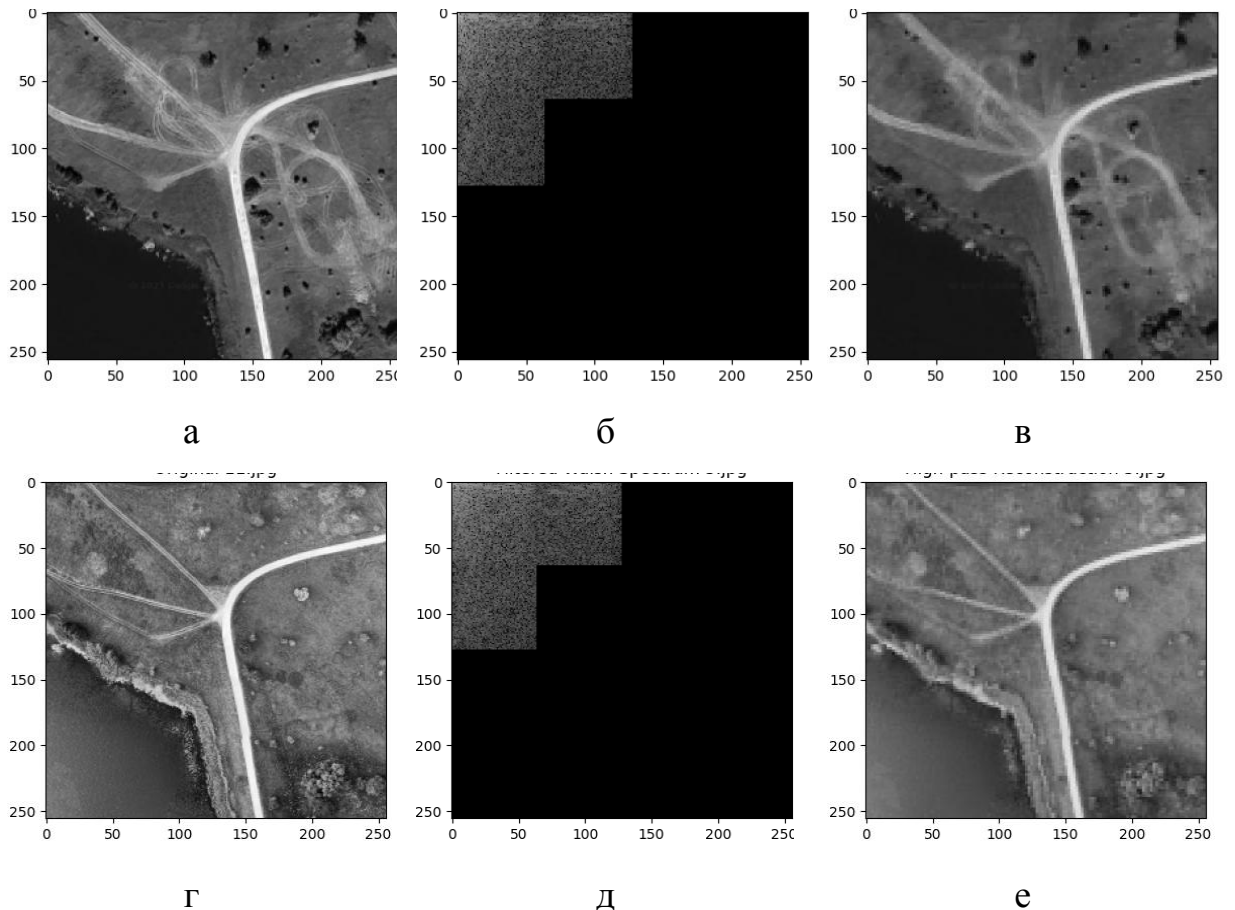


Рисунок 50 – Пара изображений после применения фильтра 4

Данный фильтр полностью удаляет область высоких секвентностей, которая отвечает за мелкие детали и текстуры. По сравнению с фильтром 3, он очищает изображение от любого высокочастотного содержимого, которое могло остаться вблизи осей. Результирующее изображение будет максимально приближено к идеалу: только общая структура и сглаженные контуры. В таблице 3 приведено влияние частотных компонент на сходимость СНС.

Таблица 3 – Влияние частотных компонент на сходимость СНС

Вид фильтра	Validation Accuracy	Validation Loss
Без фильтра	0.7300	0.6215
Фильтр 1	0.7769	0.1941
Фильтр 2	0.7925	0.8411
Фильтр 3	0.7975	0.1505
Фильтр 4	0.7068	1.1839

Фильтры 1-4 увеличивают скорость сходимости процесса обучения. Нейронная сеть быстрее учится распознавать глобальные паттерны и формы, так как набор данных не содержит шум и нерелевантные текстуры. Фильтр 1 и 4 имеют максимальную скорость, но возможна существенная потеря информации, что ограничивает итоговую точность. Фильтры 2 и 3 наиболее сбалансированный вариант. Нейронная сеть получает достаточно информации о контурах, но без зашумляющих деталей.

Предлагается использовать фильтр 3, так как он имеет наивысший показатель Validation Accuracy = 0.7975 и наименьший показатель Validation Loss = 0.1505. Тем самым производится отсечения > 50 % значений спектра.

4.3.2 Экспериментальные исследования различных моделей нейронных сетей

Выбор архитектуры СНС и количество характеристик в векторе-признаков зависит от формата входных данных и обучающей выборки. Для решения данной задачи были исследованы как существующие модели ImageNet, так и разработанная в ходе данного исследования модель. В качестве метрики оценки расстояния между векторами-признаками выбрано косинусное сходство, оно позволяет определить содержательную близость графических файлов без привязки к их абсолютным значениям.

В ходе обучения нейронных сетей установлено, что глубокие модели с большим количеством весовых коэффициентов склонны к переобучению, в то время как неглубокие архитектуры демонстрируют более высокие показатели

точности при фиксированном количестве эпох (количество полных проходов через обучающий набор данных). В таблице 4 приведены сравнительные характеристики рассмотренных моделей нейронных сетей, такие как модель, время обучения в мс, количество весов у указанных моделей, а также ошибка, вычисленная с использованием СКО.

Таблица 4 - Результаты исследования набора данных на существующих моделях ImageNet

Модель	Время, мс	Ошибка (СКО)	Кол-во весов (млн)
MobileNet v2	27	0,7557	3,5
Xception	192	0,5487	22,9
VGG16	867	0,2695	138,4
VGG19	892	0,2311	143,7
Предложенное решение	32	0,1729	0,358

Полученные в ходе эксперимента позволяют утверждать, что значительном сокращении количества весов (в 10 раз по сравнению с MobileNet v2) ошибка (СКО) ниже. В разработанной модели основное внимание сделано на замене функции активации (ELU вместо ReLU). Использование активационной функции ELU позволило устранить проблему «умирающих» нейронов, а также уменьшить количество свёрток на слоях сети. Выходным слоем является вектор-признаков изображения.

Разработанная модель представляет собой компактную и эффективную свёрточную нейронную сеть, предназначенную для задач извлечения векторов-признаков аэрокосмических изображений. Разработанная модель сбалансирована по количеству параметров и может быть использована в условиях ограниченных вычислительных ресурсов.

Было проведено исследование, результаты которого представлены на рисунке 51, как количество характеристик влияет на вероятность обнаружения пары.



Рисунок 51 - Зависимость вероятности обнаружения пары изображений от количества характеристик

Исходя из рисунка 51, было получено, что при значениях 128, 256, 512 разница в вероятности корректного обнаружения находится ниже, чем погрешность вычислений, возникающая из-за особенности алгоритма обучения. Таким образом, разработанная модель с фиксированным количеством характеристик равным 128 является наиболее оптимальным с точки зрения вероятности обнаружения пары изображений для решения рассматриваемой задачи.

Проведено сравнение предложенного метода с классическими подходами: SIFT (прямое сопоставление), предобученными CNN (VGG16) и методом CVM-Net.

Таблица 5 – Сравнительные характеристики методов

Метод	Точность в пикс	Время, мс
SIFT (прямое)	>500	180
VGG16 + L2	150	300
CVM-Net	80	450
Предложенный метод	300	106

Предложенный метод превосходит аналоги по точности и обладает приемлемым временем выполнения.

4.3.3 Экспериментальные исследования сходимости нейронной сети в зависимости от контекста обучающей выборки

В данном исследовании предложен алгоритм обучения свёрточной нейронной сети, предложен подход к формированию данных для обучения нейронной сети. В качестве искажений применяются следующие виды трансформаций, представленные в таблице 6.

Таблица 6 – Виды трансформации

Категория трансформации	Тип трансформации	Базовое значение и разброс
Яркостные искажения	Изменение яркости	$1.0 \pm 0.3\%$
	Изменение контрастности	$1.0 \pm 0.3\%$
	Добавление Гауссова шума	0.0 ± 0.2
Геометрические искажения	Поворот изображения	$0.0 \pm 5.0^\circ$
	Масштабирование	1.5 ± 0.25

Сравнение результатов работы сходимости нейронной сети производилось на:

- парах изображений без добавления шума и аугментации;
- парах изображений с добавлением шума и аугментации;
- парах изображений после применения спектральных преобразований (Уолша, Адамара, Пэли);
- парах изображений после применения спектральных преобразований (Уолша, Адамара, Пэли) усеченные на 25%, 50%, 75%;
- парах изображений после спектральных преобразований с предварительным добавлением шума, аугментации и поворота;
- парах изображений после спектральных преобразований с усеченных на 75% -1 свёрточный блок.

В таблицах 7-9 приведен результат работы сиамской нейронной сети на различных исходных данных. Обозначения:

- А) Без удаления информационной составляющей
- В) С удалением 25%
- С) С удалением 50%
- D) С удалением 75%
- 1. Без искажений.
- 2. Яркостные искажения.
- 3. Яркостные и геометрические искажения.
- x – Validation Accuracy (пары спутниковых снимков в разных локациях с искажениями).
- y – Validation Loss (сред. квадр. ошибка).
- z – Test Accuracy (пары спутник/беспилотник в идентичных локациях без искажений).

Таким образом, было выявлено, что наибольшую эффективность демонстрируют исходные изображения и спектр Уолша без удаления информационной составляющей, а также спектр Пэли с усечением на 50% значений спектра с сохранением постоянной составляющей.

Таблица 7 – Результат работы сиамской нейронной сети в качестве выходных значений, которой используется спектр

Тип входных данных	Исходное изображение			Матрица Пэли			Матрица Адамара			Матрица Уолша		
	Метрика	x	y	Z	x	y	z	x	y	z	x	y
A-1	0.8200	0.1639	4	0.6450	0.4487	1	0.6725	0.3413	4	0.6817	0.2626	6
A-2	0.7393	0.1860	5	0.6457	0.3841	2	0.6000	0.3857	1	0.6291	0.7729	2
A-3	0.6925	0.2070	4	0.6250	0.4389	1	0.5425	0.3746	2	0.6090	0.5242	7
B-1				0.6550	0.4138	1				0.7425	0.3053	7
B-2				0.6675	0.3550	0				0.5025	0.4105	3
B-3				0.6100	0.4584	1				0.5925	0.2993	3
C-1				0.6925	0.2336	3				0.6566	0.4749	3
C-2				0.6950	0.3964	5				0.6175	0.4809	1
C-3				0.5625	0.4261	1				0.6150	0.4167	2
D-1				0.6767	0.3854	3				0.6650	0.4815	2

D-2				0.5850	0.4293	3				0.6625	0.3050	3
D-3				0.6316	0.3531	2				0.5825	0.3870	1

Таблица 8 – Результат работы сиамской нейронной сети, в качестве входных значений, которой используются изображения восстановленного спектра

Тип входных данных	Исходное изображение			Матрица Пэли			Матрица Адамара			Матрица Уолша		
	х	у	z	х	у	z	х	у	z	х	у	z
A-1	0.8200	0.1639	4	0.7850	0.1727	3	0.8195	0.1442	4	0.7800	0.1918	4
A-2	0.7393	0.1860	5	0.7625	0.1858	4	0.7250	0.1822	3	0.7419	0.1817	3
A-3	0.6925	0.2070	4	0.6950	0.2034	4	0.7350	0.2150	3	0.7025	0.2142	3
B-1				0.7850	0.1622	5	0.7487	0.1882	4	0.7945	0.1616	5
B-2				0.7425	0.1854	4	0.7550	0.1869	4	0.7100	0.1952	4
B-3				0.7425	0.2032	4	0.7950	0.1620	5	0.7175	0.1980	4
C-1				0.7500	0.1862	5	0.7250	0.2162	5	0.7794	0.1968	5
C-2				0.7625	0.1783	4	0.7700	0.1818	5	0.7500	0.1712	4
C-3				0.7525	0.2005	4	0.6775	0.2099	2	0.7450	0.1856	5
D-1				0.7544	0.1756	4	0.7820	0.1999	2	0.7625	0.1725	2
D-2				0.7569	0.1679	3	0.7000	0.2093	2	0.7800	0.1664	4
D-3				0.7275	0.2094	3	0.6675	0.2698	0	0.7375	0.2018	2

Таким образом, было выявлено, что наибольшую эффективность демонстрирует исходное изображение и изображение, подверженное прямому и обратному преобразованию Уолша в системе Адамара без учета постоянной составляющей.

Таблица 9 – Результат работы сиамской нейронной сети с разным набором входных значений

Тип входных данных	Спектр Уолша			Спектр Уолша без постоянной составляющей			Восстановленное изображение			Восстановленное изображение без постоянной составляющей		
	х	у	z	х	у	z	х	у	z	х	у	z
A-1	0.4975	2.4463	1	0.5013	2.2463	1	0.7494	1.1650	7	0.7400	1.0977	5
A-2	0.5088	2.6302	1	0.4937	2.4682	1	0.7800	1.0922	6	0.7325	1.0530	5

A-3	0.5075	2.5378	1	0.4900	3.0640	1	0.6875	1.4485	4	0.6784	1.5660	5
B-1	0.5300	2.3623	2	0.4586	2.3105	1	0.8175	0.7654	5	0.7600	1.2870	5
B-2	0.5000	2.6127	1	0.4727	2.4101	1	0.7675	0.5376	7	0.7800	0.8082	6
B-3	0.5125	2.5174	1	0.4782	2.4473	0	0.7750	0.7742	6	0.7118	1.1856	7
C-1							0.7085	0.8285	6	0.7775	1.2558	4
C-2							0.7425	0.9872	7	0.7450	1.1599	6
C-3							0.7250	0.8841	5	0.7425	1.1982	7
D-1							0.7500	1.1539	8	0.7025	1.1021	4
D-2							0.7965	0.9484	6	0.7100	1.2198	6
D-3							0.7450	0.8754	4	0.7425	1.2830	6

Низкий показатель x – Validation Accuracy (пары спутниковых снимков в разных локациях с искажениями), y – ValidationLoss (сред. квадр. ошибка) на наборах данных, где в качестве входного изображения выступает спектральное представление связано с тем, что СНС в операциях свёртки оперирует не частотным представлением, поэтому необходимо переходить от пространственного представления к частотному. Наилучший результат СНС показывает в том случае, если в качестве входного слоя применяется восстановленное изображение, причем можно убирать до 75% значений, в том числе и постоянную составляющую. Лучшая конфигурация: Стандартные свёртки с восстановлением дают максимальную Validation Accuracy (0.7965) и приемлемый Loss (0.9484).

4.3.4 Экспериментальные исследования поиска изображения, соответствующего изображениям из базы данных

Для предложенной архитектуры получим следующие результаты, приведенные в таблицах 10,11.

Таблица 10 – Метрики информационного поиска для предложенной архитектуры

Метрика	Значение	Интерпретация
Precision	0.728	72.8% пар, которые модель назвала "похожими", действительно похожи

Recall	0.800	Модель находит лишь 80.0% всех реально похожих пар
F1	0.872	Значение 0.872 близко к 1, что указывает на высокое качество модели в балансе между: precision (точностью) и recall (полнотой).
AUC-ROC	0.784	Значение 0.784 означает, что в 78.4% случаев модель верно ранжирует случайно выбранный положительный пример выше, чем отрицательный.

Таблица 11 – Средние расстояния между парами

Класс	Значение	Интерпретация
Same (похожие)	0,317	Среднее расстояние для пар одной локации
Diff (разные)	1.869	Среднее расстояние для пар разных локаций

Из таблицы 11 можно сделать увидеть разделение расстояний, что подтверждает способность сети формировать дискриминативные признаки.

На рисунке 52 приведены графики сходимости сети на спектральном представлении Пэли усеченным 50% значениями (эксперимент С). Как видно из рисунка метрика точности (model accuracy) растёт, а потери (model loss) падают. Две линии – одна обучающая (train), другая – тестовая выборки (val).

Для проверки точности разработанного модели сиамской нейронной сети построим матрицу несоответствий (рис. 53): x_1 – фрагменты со съёмки на небольшой высоте после спектрального преобразования, x_2 – карта местности подстилающей поверхности после спектрального преобразования.

На пересечении строк и столбцов вычисляется евклидово расстояние между сформированными векторами. По главной диагонали вычисляется разница между спектрами, полученными с одной и той же локации, но с разных сенсоров. Данная оценка является выходом с нейронной сети и на ее основе

оценивается точность поиска. В данном эксперименте не рассматривается тот случай, когда фрагмент аэрофотосъемки с ЛА находится одновременно на нескольких тайлах (от 2 до 4).

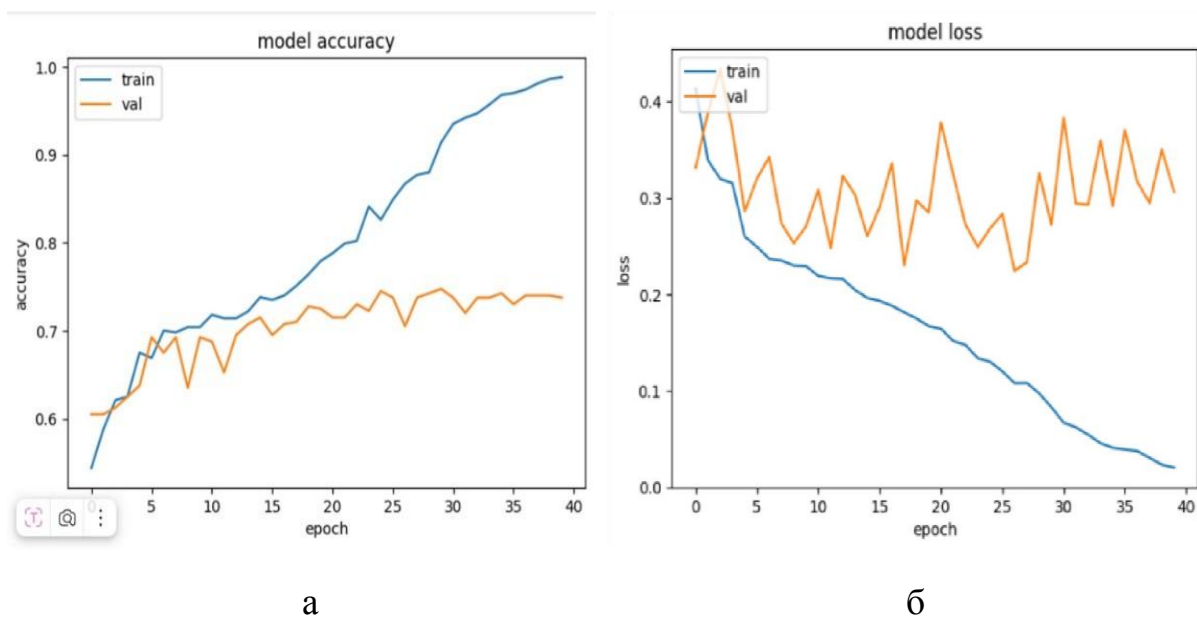


Рисунок 52 – График сходимости нейронной сети на спектральных представлениях

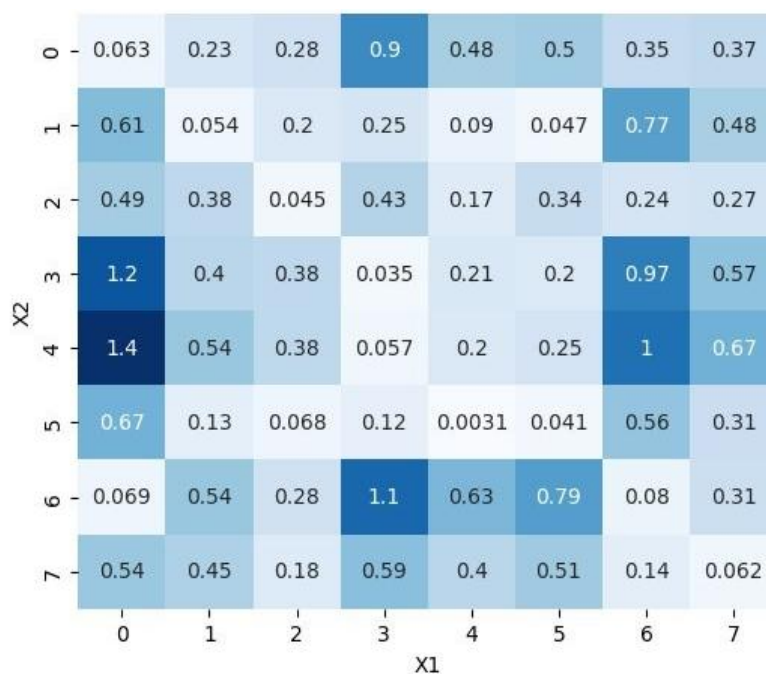


Рисунок 53 – Матрица несоответствий

Также ниже представлены результаты экспериментальных исследований пар изображений. Приведены 3 примера пар и значения показателей их схоже-

сти (рисунки 54-56).



Рисунок 54 - Пара изображений с величиной сходства - 0,63
(различный сюжет)



Рисунок 55 - Пара изображений с величиной сходства - 0,98
(идентичный сюжет)



Рисунок 56 - Пара изображений с величиной сходства - 0,93
(идентичный сюжет)

4.3.5 Экспериментальные исследования детектирования ключевых точек

В качестве алгоритма получения описания особых точек исследовались следующие: SIFT, BRISK, AKAZE, ORB, LOFTR, комбинация ORB+SIFT, Deep Features и TIMM Features [20-25].

Метрики сравнения:

- количество совпадений – сколько точек успешно сопоставлено;
- время извлечения – скорость работы детектора;
- время сопоставления – скорость связывания;
- Success Rate – процент успешных совпадений от общего числа точек;
- среднее расстояние – качество совпадений (меньше = лучше).

Дополнительные тесты:

- разные позиции UAV – тестирование в различных условиях;
- устойчивость к преобразованиям:
 - поворот;
 - размытие;
 - изменение яркости;
 - масштабирование.

В таблице 12 приведено сравнение методов детектирования ключевых точек по скорости обнаружения пар и потребляемой памяти

Таблица 12 – Сравнение методов детектирования ключевых точек

Метод	Скорость (пар/сек)	Потребление памяти (МВ)	Время инициализации (мс)
SIFT	27.6	45	0
ORB+SIFT	29.1	38	0
TIMM Features	79.5	280	860
Deep Features	14.0	450	1200

Для оценивания детекторов/дескрипторов и матчеров особых точек был подготовлен экспериментальный набор данных, состоящий из 2000 пар изображений (со спутника и ЛА). Пары изображений были выровнены по координатам, высоте и углу.

В таком наборе данных удобнее всего производить оценку работы указанных выше алгоритмов.

Используемый подход позволяет автоматизировать процесс оценки качества выбора точек и работы механизма их связывания, поскольку расстояние между парами точек в таком случае должно стремиться к нулю. Однако, наличие определённых искажений, связанных с углом съёмки, падением теней, временем года может несколько исказить расположение целевой точки относительно второго кадра. Поэтому делается допущение, что целевая точка не должна иметь расхождение с точкой на втором снимке более, чем в 10 пикселей.

Поскольку вероятность попадания точки на втором изображении очень сильно зависит от количества точек, то для чистоты эксперимента будем даже анализировать среднее количество точек на фрагменте изображения, а также проведем исследования с такими показателями порога обнаружения особой точки, чтобы проанализировать разные случаи и с малым (около 10) и большим (около 1000 количеством точек).

Для детектирования точек применялся метод SIFT.

4.4 Оценка точности и погрешности совмещения изображений

В работе было произведена оценка времени выполнения метода и алгоритмов, а также точность совмещения в пикселях (таблица 13).

Таблица 13 – Сравнительные характеристики методов

Метод	Погрешность в пикс	Время, мс
SIFT (прямое)	>500	180
VGG16 + L2	150	300
CVM-Net	80	450
Предложенный метод	300	106
Предложенный метод + уточнение	20	207

Программная реализация метода и алгоритмов обеспечивает определение местоположения менее чем за 210 мс, что удовлетворяет требованиям реального времени для ЛА.

Эксперимент проводился на 11 различных сюжетах, призванных показать работу системы в различных сценариях:

- городская застройка (небольшой город);
- мегаполис;
- типовая коттеджная застройка;
- лес с редкой дачной застройкой;
- чистый лес;
- лес с дачной застройкой;
- арктическая пустыня (на примере Гренландии);
- арктическая полупустыня (на примере Канады);
- карельский лес;
- песчаная пустыня (на примере Сахары);
- песчаная пустыня с застройкой (на примере пригорода Дубая).

Тестирование проводилось со следующими параметрами системы:

1. Размер тайла – 256 пикселей.
2. Шаг тайлов – 64 пикселей (перекрытие на 75%).
3. Коэффициент принятия тайла – 2,5.
4. Количество точек для полёта – 100.
5. Радиус поиска – 2 км.
6. В виду ряда ограничений на запуск БПЛА, для проведения эксперимента в различных сценариях, в качестве съёмки с БПЛА были выбраны спутниковые изображения с других сенсоров и за другие даты. Полученные результаты не могут свидетельствовать о качестве работы метода в целом, однако могут показать его применимость и качество работы относительно различных сценариев.

Метриками для оценки работы являлись следующие результаты:

1. Количество точек в которых алгоритм нашёл тайл с коэффициентом принятия ниже 2,5.
2. Количество истинно позитивных найденных совпадений, среди отобранных точек (True Positive).
3. Количество неверно отобранных тайлов среди отобранных точек (False Positive).
4. Количество верно найденных координат, среди непринятых точек (False Negative).
5. Количество неверно отобранных тайлов среди непринятых точек (True Negative).

«Да» — алгоритм отобрал тайл (счёл совпадение найденным / коэффициент принятия $\geq 2,5$). «Нет» — алгоритм не отобрал тайл (коэффициент принятия ниже 2,5). Строки «Факт: Да / Нет» означают, было ли на самом деле совпадение (истинное соответствие) в этой точке.

Для каждого сценария проводилось по 10 имитаций работы БПЛА. Результаты эксперимента, приведены в таблицах 14-23.

Таблица 14 – Тестирование на сценарии мегаполис

Факт	Сценарий: MEGAPOLIS	Отобрано точек: 99
	Локация: Абу-Даби, ОАЭ	
	Результат	
	Да	Нет
Алгоритм вы- брал	3	7
Нет	1	88

Разработанный метод показал низкую эффективность на сценарии мегаполис, это может быть связано с обилием похожих структур.

Таблица 15 – Тестирование на сценарии городская застройка

Факт	Сценарий: CITY	Отобрано точек: 92
	Локация: Рязань, Россия	
	Результат	
	Да	Нет
Да	8	17
Нет	3	64

Таблица 16 – Тестирование на сценарии типовая коттеджная застройка

Факт	Сценарий: VILLAGE	Отобрано точек: 100
	Локация: Детройт, США	
	Результат	
	Да	Нет
Да	15	4
Нет	30	51

Таблица 17 – Тестирование на сценарии лес с дачной застройкой

Факт	Сценарий: VILLAGE2	Отобрано точек: 85
	Локация: Рязанская обл.	
	Результат	
	Да	Нет
Да	27	9
Нет	26	23

Таблица 18 – Тестирование на сценарии чистый лес

Факт	Сценарий: FOREST	Отобрано точек: 77
	Локация: Московская обл.	
	Результат	

	Да	Нет
Да	33	0
Нет	44	0

Исходя из данных, полученный в таблице 18 можно сделать вывод, что разработанный метод обнаружил все истинные совпадения, но за счет большого количества ложных срабатываний.

Таблица 19 – Тестирование на сценарии арктическая полупустыня

Факт	Сценарий: ARCTIC	Отобрано точек: 102
	Локация: Канада, север	
	Результат	
	Да	Нет
Да	30	13
Нет	18	41

Таблица 20 – Тестирование на сценарии арктическая пустыня

Факт	Сценарий: ARCTIC2	Отобрано точек: 55
	Локация: Гренландия, Дания*	
	Результат	
	Да	Нет
Да	35	0
Нет	20	0

Для тестирования были выбраны территории около ледниковой зоны. Сами ледники на спутниковых снимках имеют практически однотонный рисунок.

Исходя из данных полученных в таблице 20, можно сделать вывод, что все истинные точки найдены, но есть ложные срабатывания

Таблица 21 – Тестирование на сценарии песчаная пустыня

Факт	Сценарий: DESERT	Отобрано точек: 89
	Локация: Пустыня Сахара	
	Результат	
	Да	Нет
Да	27	0
Нет	59	3

По результатам экспериментов можно сделать вывод, что разработанный метод совмещения разнородных изображений на основе нейросетевых технологий позволяет обнаружить все совпадения без ложных пропусков на сценарии песчаная пустыня.

Таблица 22 – Тестирование на сценарии песчаная пустыня с застройкой

Факт	Сценарий: DESERT2	Отобрано точек: 89
	Локация: Дубай, ОАЭ (пустынная часть)	
	Результат	
	Да	Нет
Да	31	4
Нет	46	8

Для тестирования были отобраны участки местности, в которых на пустынной земле присутствовала флора или активность человека.

Таблица 23 – Тестирование на сценарии поле

Факт	Сценарий: FIELDS	Отобрано точек: 84
	Локация: США, Север	
	Результат	

	Да	Нет
Да	20	18
Нет	44	2

Исходя из проведенных исследований можно сделать вывод, что разработанный метод чувствителен к типу местности, наибольшие трудности вызывает плотная городская застройка. На однородных территориях (лес, пустыня, ледник) метод совмещения разнородных изображений подстилающей поверхности находит истинные точки, но имеется тенденция к большому количеству ложных срабатываний, что может указывать на необходимость настройки пороговых параметров, что будет рассмотрено в дальнейших исследованиях. Использование пары спутниковых снимков вместо реальной съемки с ЛА не позволяет делать окончательные выводы о качестве метода при работе с различными сценариями съемки, но дает возможность сравнить относительную применимость его при различных сценариях.

В ходе эксперимента было проведено 1100 отдельных испытаний (11 сценариев по 100 испытаний на один тип сценария). Объем обучающей выборки 70 % для подбора параметров сети, тестовая выборка составляет 30 % и применяется для финальной оценки.

4.5 Выводы

Разработанный программно-алгоритмический комплекс позволяет определять местоположение ЛА с точностью до 10–20 м в условиях отсутствия сигналов ГНСС. Применение предобработки в спектральном пространстве Уолша совместно с нормализацией и аугментацией повысило точность сходимости свёрточной сети на 7,8 %. Метод совмещения разнородных изображений подстилающей поверхности, позволяет осуществлять первичное позиционирование ЛА с точностью до 300 пикселей, что при высоте 300 метров, соответствует 300 метрам. Предложенный алгоритм уточнения положения текущего изображения относительно спутникового изображения методом сопоставления ключевых точек позволяет компенсировать взаимные несоответствия на

паре разнородных изображений позволяет сократить конечную область совмещения изображений до 20 пикселей, что при высоте 300 метров, соответствует 20 метрам.

Экспериментально подтверждена эффективность разработанного метода и алгоритма определения положения текущего изображения относительно спутникового изображения в условиях отсутствия сигналов ГНСС, позволяющего существенно сократить район для точного совмещения до 10–20 пикселей, что при высоте 300 метров, соответствует 10-20 метрам (что соответствует 10-20 метрам на местности при высоте полета 300 м). Время обработки не превышает 210 мс.

ЗАКЛЮЧЕНИЕ

В данной работе выполнена разработка и исследование метода совмещения разнородных изображений подстилающей поверхности и алгоритмы его реализующие, а также способ предварительной обработки с использованием ортогональных преобразований в системе Уолша и алгоритм определения положения текущего изображения относительно спутникового изображения, позволяющего существенно сократить район для точного совмещения.

При этом получены следующие основные научные результаты:

1. Разработан метод совмещения разнородных изображений подстилающей поверхности с текущим изображением района полёта и алгоритмов его реализующие, а именно:

1.1. Разработан алгоритм обучения свёрточной нейронной сети, отличающийся от известных подходов введением последующей декомпозиции ее архитектуры;

1.2. Разработан алгоритм формирования вектора-признаков изображения, обеспечивающий степень сжатия визуальной информации для фрагмента изображений с 20 КБ до 256 байт с допустимой потерей семантического содержания;

1.3. Разработан алгоритм поиска спутникового изображения местности в цифровой базе данных спутниковых изображений района полета беспилотной воздушной техники, соответствующего текущему изображению, полученного с сенсора ЛА, позволяющего определить потенциальное положение ЛА с точностью до 300 пикселей.

2. Разработан способ предварительной обработки изображений подстилающей поверхности с использованием спектральных преобразований, который применяется в комплексе с традиционной нормализацией и аугментацией, позволяющий увеличить точность сходимости свёрточной нейронной сети на 7,8% за счет фильтрации низкочастотных секвент.

3. Разработан алгоритм определения положения текущего изображения относительно спутникового изображения в условиях отсутствия сигналов ГНСС, позволяющего существенно сократить район для точного совмещения до 10-20 пикселей за счет использования ключевых точек.

4. Проведено экспериментальное исследование разработанного метода и алгоритмов, его реализующих с применением реальной информации.

СПИСОК ЛИТЕРАТУРЫ

1. SandboxAQ Announces AQNav – World's First Commercial Real-Time Navigation System Powered by AI and Quantum to Address GPS Jamming [Электронный ресурс] <https://www.prnewswire.com/news-releases/sandboxaq-announces-aqnav--worlds-first-commercial-real-time-navigation-system-powered-by-ai-and-quantum-to-address-gps-jamming-302181462.html> (дата обращения 21.02.2026).
2. SandboxAQ and Acubed Achieve Progress in Magnetic Navigation [Электронный ресурс] <https://thequantuminsider.com/2025/07/16/sandboxaq-and-acubed-achieve-progress-in-magnetic-navigation/> (дата обращения 21.02.2026)
3. Teague, S., & Chahl, J. (2024). An Algorithm for Affordable Vision-Based GNSS-Denied Strapdown Celestial Navigation. *Drones*, 8(11), 652. [Электронный ресурс] <https://doi.org/10.3390/drones8110652> (дата обращения 21.02.2026)
4. Алпатов Б.А., Бабаян П.В., Коблов Ю.С., Муравьев В.С., Стротов В.В., Фельдман А.Б. Программный комплекс для автоматизации научных исследований в области создания перспективных систем навигации беспилотных летательных аппаратов по данным видеонаблюдения // Вестник Рязанского государственного радиотехнического университета. 2012. № 40. С. 9-15.
5. Ефимов И.П. Авиационные приборы: учебное пособие / И.П. Ефимов. – Ульяновск: УлГТУ, 2018. – 255 с.
6. Степанов О.А. Методы обработки навигационной измерительной информации. – СПб: Университет ИТМО, 2017. – 196 с.
7. Сарайский Ю.Н., Алешков И.И. Аэронавигация. Часть I. Основы навигации и применение геотехнических средств: Учебное пособие: СПб: СПбГУГА, 2010 .- 302с.
8. Шумейко В. В., Хахулина Н. Б. Спутниковые системы и технологии позиционирования: методические указания к выполнению лабораторных работ для студентов направления 21.03.03 «Геодезия и дистанционное зондирование» (профиль «Геодезия») всех форм обучения / ФГБОУ ВО «Воронежский

государственный технический университет»; сост.: В. В.Шумейко, Н. Б. Хахулина. – Воронеж: Изд-во ВГТУ, 2022. – 20 с.

9. Vizilter Yu.V., Zheltov S. Yu., Lebedev M.A. Image and shape comparison via morphological correlation // The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, Volume XLIV-2/W1-2021 4th Int. Worksh. on “Photogrammetric & computer vision techniques for video surveillance, biometrics and biomedicine”, 26–28 April 2021, Moscow, Russia <https://doi.org/10.5194/isprs-archives-XLIV-2-W1-2021-207-2021>

10. Аббасов К.Н., Елесина С.И. Сравнительный анализ метода деформируемого многогранника и нейросетевого поиска в задаче корреляционного совмещения изображений // Информационные технологии в прикладных исследованиях. Межвузовский сборник научных трудов. Рязань, 2025. С. 9-13.

11. Ефимов А.И., Новиков А.И. Алгоритм совмещения изображений на основе преобразования в комплексной плоскости // Научно-техническая конференция "Техническое зрение в системах управления-2017". Тезисы. Институт космических исследований Российской академии наук. 2017. С. 34-35.

12. Ефимов А.И., Новиков А.И. Алгоритм поэтапного уточнения проективного преобразования для совмещения изображений // Компьютерная оптика. 2016. Т. 40. № 2. С. 258-265.

13. Ефимов А.И., Логинов А.А., Никифоров М.Б., Новиков А.И. Построение матрицы гомографии на основе произвольного количества ключевых точек // Проблемы передачи и обработки информации в сетях и системах телекоммуникаций. Материалы 18-й Международной научно-технической конференции. 2015. С. 137-139.

14. Жукова К.С., Кирпичников А.П., Ляшева С.А., Шлеймович М.П. Система контекстного поиска видеоизображений // Вестник Технологического университета. 2018. Т. 21. № 9. С. 154-157.

15. Гонсалес Р, Вудс Р. Цифровая обработка изображений // 3-е издание, исправленное и дополненное М.: техносфера, 2012. – 1104с

16. Демяненко Я. М. Компьютерное зрение и обработка изображений. Лекции 13-14. Детекторы и дескрипторы [Электронный ресурс] http://edu.mmcs.sfedu.ru/pluginfile.php/26343/mod_resource/content/2/Детекторы%20и%20дескрипторы.pdf (дата обращения 19.12.18)
17. Патин М. В. Сравнительный анализ дескрипторов особых точек изображений с внедрением алгоритмов под операционной системой "Android" [Электронный ресурс] <https://dspace.spbu.ru/bitstream/11701/3991/1/st010290.pdf> (дата обращения 04.03.19)
18. Дышлюк В.О. Исследование показателей качества и быстродействия поиска опорных точек на изображениях методом SURF. Молодой ученый, 2018. № 27, с. 23-26
19. Денисов А.А., Новиков А.И. Анализ методов детектирования, описания и сопоставления ключевых точек изображений // Вестник Рязанского государственного радиотехнического университета. 2024. № 89. С. 104-116.
20. Lowe D. G. Distinctive Image Features from Scale-Invariant Keypoints // International Journal of Computer Vision, 2004.
21. Bay H., Tinne T., Luc V. G. «SURF: speeded up robust features», Computer Vision and Image Understanding, vol. 110, no. 3, pp. 346–359, 2008. DOI:10.1007/11744023_32.
22. Shuvo Kumar Paul, Pourya Hoseini, Mircea Nicolescu, and Monica Nicolescu. Performance Analysis of Keypoint Detectors and Binary Descriptors under Varying Degrees of Photometric and Geometric Transformations // Journal of Image and Graphics, Vol. 9, No. 1, March 2021 doi: 10.18178/joig.9.1.1-8.
23. J. Sun, Z. Shen, Y. Wang, H. Bao and X. Zhou. LoFTR: Detector-Free Local Feature Matching with Transformers, 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Nashville, TN, USA, 2021, pp. 8918-8927, doi: 10.1109/CVPR46437.2021.00881.
24. Daniel DeTone, Tomasz Malisiewicz, Andrew Rabinovich. SuperPoint: Self-Supervised Interest Point Detection and Description // Camera-ready version for CVPR 2018 Deep Learning for Visual SLAM Workshop

<https://doi.org/10.48550/arXiv.1712.07629>

25. E. Santellani, C. Sormann, M. Rossi, A. Kuhn and F. Fraundorfer. MD-Net: Multi-Detector for Local Feature Extraction, 2022 26th International Conference on Pattern Recognition (ICPR), Montreal, QC, Canada, 2022, pp. 3944-3951, doi: 10.1109/ICPR56361.2022.9956504.

26. Mikolajczyk K., Schmid C. A performance evaluation of local descriptors. IEEE transactions on pattern analysis and machine intelligence, pages 1615-1630, 2005.

27. Li J., Hu Q., Ai M. Point cloud registration based on one-point ransac and scale-annealing biweight estimation // IEEE Transactions on Geoscience and Remote Sensing. 2021. Т. 59. № 11. С. 9716

28. Акинин М.В., Никифоров М.Б., Таганов А.И. Нейросетевые системы искусственного интеллекта в задачах обработки изображений // Москва, 2015.

29. Perronnin, F., Liu, Y., Sánchez, J., & Poirier, H. (2010). Large-scale image retrieval with compressed Fisher vectors. In Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR), pp. 3384–3391.

30. Babenko Artem, Lempitsky Victor. Additive Quantization for Extreme Vector Compression // Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition. — Columbus, USA: 2014. — Pp. 931–938

31. Николенко С., Кадурич А., Архангельская Е. Глубокое обучение. — СПб.: Питер, 2018. — 480 с.:

32. Егорова Е. А., Киношенко Д. К., Машталир С. В., Шляхов Д. В. Метрическое сравнение результатов сегментации изображений // Радиоэлектроника и информатика. 2006. №2. URL: <https://cyberleninka.ru/article/n/metricheskoe-sravnenie-rezultatov-segmentatsii-izobrazheniy> (дата обращения: 10.03.2024).

33. Noa Garcia, George Vogiatzis. Learning Non-Metric Visual Similarity for Image Retrieval the IEEE International Conference on Computer Vision Workshops (ICCVW). 2017.
34. Рудаков И. В., Васютович И. М. Исследование перцептивных хеш-функций изображений // Машиностроение и компьютерные технологии. 2015. №8. URL: <https://cyberleninka.ru/article/n/issledovanie-pertseptivnyh-hesh-funktsiy-izobrazheniy> (дата обращения: 24.03.2026).
35. База данных MNIST с рукописными цифрами [электронный ре-сурс] <https://learn.microsoft.com/ru-ru/azure/open-datasets/dataset-mnist?tabs=azureml-opendatasets> (дата обращения 21.02.2025)
36. Бабенко А., Лемпицкий В. Эффективный алгоритм поиска ближайших соседей при больших объемах поисковой базы // Труды 54- ой науч-ной конференции МФТИ. Инновации и высокие технологии. Долгопрудный: 2011. С. 16–17.
37. Gaihua Wang, Guoliang Yuan, Meng Lv, WenZhou Liu. Application of Interpolation Pooling in Convolutional Neural Networks // 2018 Helix Vol. 8(4): 3465- 3469. DOI 10.29042/2018-3465-3469
38. Zhuo Chen, Jiyuan Zhang, Ruizhou Ding. ViP. Virtual Pooling for Accelerating CNN-based Image Classification and Object Detection // WACV 2020 P. 1180-1189 DOI:10.1109/WACV45572.2020.9093418
39. Travis Williams, Robert Li. Wavelet Pooling for Convolutional Neural Networks // Published as a conference paper at ICLR 2018 P. 1-12
40. Juan Manuel Fortuna-Cervantes, Marco Tulio Ramírez-Torres, Marcela Mejía-Carlos, José Salomé Murguía, José Martínez-Carranza, Carlos Soubervielle-Montalvo, César Arturo Guerra-García. Texture and Materials Image Classification Based on Wavelet Pooling Layer in CNN // the Special Issue Recent Advances in Deep Learning for Image Analysis <https://doi.org/10.3390/app12073592>
41. Pengju Liu, Hongzhi Zhang, Wei Lian, Wangmeng Zuo. Multi-Level Wavelet Convolutional Neural Networks // Digital Object Identifier 10.1109/ACCESS.2019.2921451

42. Н. А. Вершков, М. Г. Бабенко, А. Н. Черных, В. А. Кучуков, Н. Н. Кучеров, Н. Н. Кучукова. Оптимизация искусственных нейронных сетей с помощью вейвлет-преобразований // Программирование, 2022, № 6, стр. 22-30. DOI: 10.31857/S0132347422060073
43. V.G. Spitsyn, Yu.A. Bolotova, N.H. Phan, T.T.T. Bui. Using a Haar wavelet transform, principal component analysis and neural networks for OCR in the presence of impulse noise // Компьютерная оптика, 2016, том 40, №2 с.249-257
44. Митина О. А., Базанова Д. В. Разработка модели нейронной сети на основе ряда Фурье // НАУ. 2023. №90-2. URL: <https://cyberleninka.ru/article/n/razrabotka-modeli-neyronnoy-seti-na-osnove-ryada-furie> (дата обращения: 12.08.2025).
45. Х.М. Фортуна-Сервантес, М.Т. Рамирес-Торрес, Х. Мартинес-Карранса, Х.С. Мургуиа-Ибарра, М. Мехиа-Карлос. Обнаружение объектов в аэронавигации с использованием вейвлет-преобразования и сверточных нейронных сетей: первый подход // Труды ИСП РАН, том 33, вып. 2, 2021 г. // Trudy ISP RAN/Proc. ISP RAS, vol. 33, issue 2, 2021 DOI: 10.15514/ISPRAS–2020–33(2)–9.
46. Chollet F (2017) Xception: Deep learning with depthwise separable convolutions. In: IEEE conference on computer vision and pattern recognition. IEEE; 1251–1258
47. Мадху Г., Каутиш С., Гупта Й. XConvNet: оптимизированная сверточная нейронная сеть Xception для классификации COVID-19 по изображениям ультразвукового исследования легких, полученным непосредственно у постели больного. Multimed Tools Appl 83, 33653–33674 (2024). <https://doi.org/10.1007/s11042-023-16944-z>
48. Mark Sandler, Andrew Howard, Menglong Zhu, Andrey Zhmoginov, Liang-Chieh Chen. Inverted Residuals and Linear Bottlenecks: Mobile Networks for Classification, Detection and Segmentation // 2018, IEEE/CVF Conference on

49. Nurzarinah Zakaria, Yana Mazwin Mohmad Hassim. A Review Study of the Visual Geometry Group Approaches for Image Classification // Journal of applied sciences technology and computing e-ISSN: 3036-0250 Vol. 1 No. 1(2024) 14-28 <https://publisher.uthm.edu.my/ojs/index.php/jastec> DOI: <https://doi.org/10.30880/jastec.2024.01.01.003>

50. Karen Simonyan, Andrew Zisserman. Very Deep Convolutional Networks for Large-Scale Image Recognition // International Conference on Learning Representations (ICLR), 2015, [Электронный ресурс] <https://arxiv.org/abs/1409.1556> (дата обращения 15.03.2025)

51. Simonyan K. Very Deep Convolutional Networks for Large-Scale Image Recognition [Electronic source] / K. Simonyan, A. Zisserman // [arxiv.org](https://arxiv.org/abs/1409.1556). — 2014. — URL: <https://arxiv.org/abs/1409.1556>. (дата обращения 27.03.25)

52. Yosinski J. How transferable are features in deep neural networks? [Electronic source] / J Yosinski, J Clune, Yo Bengio et al. // [arxiv.org](https://arxiv.org/abs/1411.1792). — 2014. — URL: <https://arxiv.org/abs/1411.1792>. (дата обращения 27.03.25)

53. Liu B. A survey of recommendation systems based on deep learning. / B. Liu, Q Zeng, L Lu et al. // 3rd International Symposium on Power Electronics and Control Engineering (ISPECE 2020); — Chongqing: IEEE, 2020. — [Электронный ресурс] https://www.researchgate.net/publication/349301342_A_survey_of_recommendation_systems_based_on_deep_learning. (дата обращения 27.03.25). doi: 10.1088/1742-6596/1754/1/012148

54. Sohn K. Improved Deep Metric Learning with Multi-class N-pair Loss Objective / K. Sohn // 30th Conference on Neural Information Processing Systems (NIPS 2016) // — Barcelona: IEEE, 2016. — [Электронный ресурс] https://papers.nips.cc/paper_files/paper/2016/file/6b180037abbebea991d8b1232f8a8ca9-Paper.pdf. (дата обращения 27.03.25)

55. Distinctive Image Features from Scale-Invariant Keypoints [Electronic source] // Accepted for publication in the International Journal of Computer Vision,

2004. [Электронный ресурс] <https://link.springer.com/article/10.1023/B:VISI.0000029664.99615.94#citeas> (дата обращения 27.03.25) doi: 10.1023/B:VISI.0000029664.99615.94

56. Shahar T. Siamese-Neural-Networks-for-One-shot-Image-Recognition [Electronic source] / T Shahar, N Itzhak // [Электронный ресурс] <https://github.com/>. — 2020. — URL: <https://github.com/nevoit/Siamese-Neural-Networks-for-One-shot-Image-Recognition> (дата обращения 27.03.25)

57. Jiawei Han and Micheline Kamber. Data Mining: Concepts and Techniques / Morgan Kaufmann Publishers. P. 314, 2000

58. Babenko Artem, Lempitsky Victor. Additive Quantization for Extreme Vector Compression // Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition. — Columbus, USA: 2014. — Pp. 931–938

59. Artem Babenko and Victor Lempitsky. Aggregating local deep features for image retrieval. In Proceedings of the IEEE International Conference on Computer Vision, 2015.

60. Ustinova V.E., Lutsenko A.S., Shpak A.V., Mironenkov G.V., Ivlev V.A. A method for finding the correspondence between a railway station model and its visual representation based on graphs // Computing, Telecommunications and Control. 2024. Т. 17. № 4. С. 64-77

61. Charles E. Jacobs, Adam Finkelstein, David H. Salesin Fast Multiresolution Image Querying [Электронный ресурс] URL: <https://www.microsoft.com/en-us/research/wp-content/uploads/2016/02/mrquery.pdf> (дата обращения 10.05.17)

62. Десятников И.Е. Поиск изображений по визуальному содержанию в графических базах данных и сети интернет // Информационные технологии и вычислительные системы. 2013. № 2. С. 88-95

63. Десятников И.Е., Утробин В.А. Алгоритмы поиска изображений в базах видеоданных // Компьютерная оптика. 2011. Т. 35. № 3. С. 416-422.

64. Хромов Н.А. Разработка алгоритмов, структур данных и программных средств для поиска изображения по визуальному подобию // Информационно-телекоммуникационные технологии и математическое моделирование высокотехнологичных систем. Материалы Всероссийской конференции с международным участием. Российский университет дружбы народов. 2015. С. 179-180.
65. Левашкина А.О. Разработка методов поиска изображений на основе вычислительных моделей визуального внимания // автореферат диссертации на соискание ученой степени кандидата технических наук / Сибирский государственный университет телекоммуникаций и информатики. Новосибирск, 2009
66. Васильева Н. С. Методы поиска изображений по содержанию // Программирование. – 2009. – Т. 35. – №. 3. – С. 51-80.
67. Галахов Д.И., Гора С.Ю., Гордиенко В.В., Довгаль В.М. Алгоритм ассоциативного поиска изображений на основе хаотических последовательностей // Известия Юго-Западного государственного университета. 2011. № 3 (36). С. 105-107.
68. Левашкина А.О., Поршнева С.В. Исследование возможности использования ключевых точек в задаче поиска изображений с визуально похожими объектами // Информационные технологии. 2011. № 2. С. 62-67.
69. Васильева Н. С., Новиков Б. А. Построение соответствий между низкоуровневыми характеристиками и семантикой статических изображений // Труды седьмой всероссийской конференции RCDL. – 2005. – С. 236-240.
70. Злотников Т. Система поиска изображений по содержанию // Компоненты и технологии. – 2012. – №. 1. – С. 58-59
71. С. Е. Jacobs, A. Finkelstein, D. H. Salesin Fast Multiresolution Image Querying [Электронный ресурс] <https://grail.cs.washington.edu/projects/query/mrquery.pdf> (дата обращения 04.03.2019).
72. Elad Hoffer and Nir Ailon. Deep metric learning using triplet network. In International Workshop on Similarity-Based Pattern Recognition, 2015.

73. Тарасова В.Ю. Метод определения местоположения съемки фотографии на основе визуального поиска / Тарасова В.Ю., Никифоров М.Б. // XXI век: итоги прошлого и проблемы настоящего плюс. 2023. Т. 12. № 4 (64). С. 103-110.
74. Тарасова В.Ю. Исследование моделей сверточных нейронных сетей в задаче поиска похожих аэрокосмических изображений / Костров Б.В., Тарасова В.Ю. // сборник трудов международной научной конференции студентов, аспирантов и молодых ученых «Ломоносов-2025» секция «Вычислительная математика и кибернетика». – Москва, 11 - 25 апреля 2025. С. 264-266.
75. Tarasova V.Yu. A model for extracting significant characteristics of images / Nikiforov M.B., Tarasova V.Yu. // Десятая международная конференция по когнитивной науке: Тезисы докладов. Пятигорск, 26-30 июня 2024 г. В двух частях. Часть II / Отв. ред. Киреев М.В. - Пятигорск, 2024. - С. 328-331
76. Mathematics for Machine Learning (Marc Peter Deisenroth et al.). "Inner Products and Similarity" // [электронный ресурс] <https://mml-book.github.io/book/mml-book.pdf> дата обращения: 31.07.2025)
77. Л. Н. Конченко, О. В. Руденко. Сямская нейросеть для верификации подписей // Международный журнал гуманитарных и естественных наук. 2025. №5-1 (104). URL: <https://cyberleninka.ru/article/n/siamskaya-neyroset-dlya-verifikatsii-podpisey> (дата обращения: 31.07.2025).
78. Chopra; R. Hadsell; Y. LeCun. Learning a similarity metric discriminatively, with application to face verification // [электронный ресурс] <https://ieeexplore.ieee.org/document/1467314/authors> DOI: 10.1109/CVPR.2005.202 (дата обращения: 31.07.2025).
79. Тарасова В.Ю., Костров Б.В., Устюков Д.И. Свёрточная нейронная сеть с гибридными слоями для определения визуального сходства спутниковых снимков 20255691610, 17.11.2025. Заявка № 2025691279 от 17.11.2025.
80. Уткин Л.В., Мелдо А.А., Ипатов О.С., Рябинин М.А. Медицинские интеллектуальные системы на примере диагностики рака легкого // Известия

ЮФУ. Технические науки. 2018. №8 (202). URL: <https://cyberleninka.ru/article/n/meditsinskie-intellektualnye-sistemy-na-primere-diagnostiki-raka-legkogo> (дата обращения: 31.07.2025).

81. Тарасова В.Ю. Подходы к формированию набора данных и его предварительной обработки в задачах обучения свёрточных сиамских нейронных сетей для коррекции положения летательного аппарата по аэрофото съемке / Тарасова В.Ю., Костров Б.В., Бабаев С.И. // XXI век: итоги прошлого и проблемы настоящего плюс. 2025. Т. 14. № 3 (71). С. 70-77.

82. Тарасова В.Ю. Методика визуального поиска изображений / Тарасова В.Ю., Никифоров М.Б. // Конгресс "Интеллектуальные системы и информационные технологии" – 2022. С. 86-92.

83. К.А. Алимагадов, С.В. Умняшкин. Аугментация данных на основе вейвлет-фильтрации при обучении нейронных сетей // ГрафиКон 2023: 33-я Международная конференция по компьютерной графике и машинному зрению, 19-21 сентября 2023 г., С 437-442

84. Tarasova V. Detection Images with Similar Content / Nikiforov M., Tarasova V., Tarasov A. // III международный научный форум по компьютерным и энергетическим наукам (WFCEES 2022) С. 9.

85. Тарасова В.Ю. Разработка системы поиска похожих изображений в локальных хранилищах / Тарасова В.Ю., Тарасов А.С. // Нейроинформатика-2020. С. 286-293.

86. Л.Н. Конченко, О.В. Руденко. Сиамская нейросеть для верификации подписей // Международный журнал гуманитарных и естественных наук. 2025. №5-1 (104). URL: <https://cyberleninka.ru/article/n/siamskaya-neyroset-dlya-verifikatsii-podpisey> (дата обращения: 31.07.2025).

87. Tarasova V. Image Search Algorithm in Local Data Base / Bodrov O.A., Tarasov A.S., Tarasova V.Yu., Bodrova I.V. // 8th Mediterranean conference on embedded computing (MECO-2019). С. 453-455.

88. Tarasova V. Method for Generating Vector Image Features for Contextual Search Tasks / Tarasova V., Nikiforov M. // 2024 International Russian Automation Conference (RusAutoCon) С. 121 – 126.
89. Andrzej Materka and Michal Strzelecki. Texture Analysis Methods – A Review [Электронный ресурс] URL: http://www.eletel.p.lodz.pl/programy/cost/pdf_1.pdf (дата обращения: 10.01.2019).
90. Djork-Arné Clevert, Thomas Unterthiner, Sepp Hochreiter. Fast and Accurate Deep Network Learning by Exponential Linear Units (ELUs) // <https://arxiv.org/abs/1511.07289> (дата обращения 28.07.2025)
91. Liyanage P., Kruglova L.V., Ranasinghe N. Enhancing image processing with a wavelet-based convolutional neural network model // МНИЖ. 2024. №6 (144). URL: <https://cyberleninka.ru/article/n/enhancing-image-processing-with-a-wavelet-based-convolutional-neural-network-model> (дата обращения: 31.07.2025).
92. Тарасова В.Ю. Разработка гибридной архитектуры сверточной нейронной сети с использованием спектральных преобразований / Тарасова В.Ю., Костров Б.В., Ефимов А.И., Бабаев С.И. // Известия ЮФУ. Технические науки № 1 (249), 2026. С. 235-249
93. Костров Б.В., Гринченко Н.Н., Баранова С.Н., Трушина Е.А., Вьюгина А.А. Ортогональное кодирование бинарных изображений // Вестник Ярославского высшего военного училища противовоздушной обороны. 2023. № 2 (21). С. 82-87.
94. Тарасова В.Ю., Гринченко Н.Н., Геращенко Е.С., Тарасов А.С. Сверточная нейронная сеть для хэширования изображений 2019665048, 18.11.2019, заявка № 2019664040 от 07.11.2019.
95. Тарасова В.Ю. Алгоритмы поиска похожих изображений в задачах определения местоположения // III Всероссийскую научную конференцию «Искусственный интеллект в автоматизированных системах управления и об-

работки данных (ИИАСУ'24)»: сборник статей в трех томах. – Москва, 30 октября-1 ноября 2024. Т.1 – С. 297-302, 10.31453/kdu.ru.978-5-00247-131-7-2025-506

96. Трахтман, А.М.; Трахтман, В.А. Основы теории дискретных сигналов на конечных интер-валах; Изд-во: Советское радио, 1975 г., 208с.

97. Основы теории сигналов: учебное пособие / С. И.Кумков, А.В.Цветков; Министерство науки и высшего образования РФ.—Екатеринбург : Изд-во Урал. ун-та, 2023. — 114 с.ISBN 978-5-7996-3636-4

98. Karima Ait Bouslam; Jamal Amadid; Fatim-Zahra Bennioui; Radouane Iqdour. Zeroual Noma based Walsh Hadamard Transform and User Pairing Strategy // Published in: 2024 International Conference on Global Aeronautical Engineering and Satellite Technology (GAST) DOI: 10.1109/GAST60528.2024.10520785

99. Тарасова В.Ю., Костров Б.В., Ефимов А.И. Программа уменьшения размерности свёрточного слоя нейронной сети с использованием спектрального преобразования 2025691618, 17.11.2025. Заявка № 2025691298 от 17.11.2025.

100. Новиков А.И., Ефимов А.И., Колчаев Д.А. Алгоритмы автоматической идентификации объектов на разнородных изображениях и совмещения изображений // Информационные технологии и нанотехнологии. Сборник трудов ИТНТ-2018. Самарский национальный исследовательский университет имени академика С.П. Королева. 2018. С. 650-663.

101. Ефимов А.И., Крючкова Т.Н., Ярославцева А.И. Комбинированный алгоритм совмещения облаков точек // Труды Международной конференции по компьютерной графике и зрению "Графикон". 2023. № 33. С. 539-548.

102. Тарасова В.Ю. Программа поиска изображений по визуальной информации 2023687992, 19.12.2023, заявка № 2023687652 от 08.12.2023.

103. Тарасова В.Ю., Никифоров М.Б. Система визуального поиска в коллекции изображений 2807980, 21.11.2023, заявка № 2022116892 от 22.06.2022

104. Тарасова В.Ю. Алгоритм обнаружения визуального сходства изображений / Никифоров М.Б., Тарасова В.Ю. // Цифровая обработка сигналов. 2022. № 3. С. 53-57.
105. Li J., Hu Q., Ai M. Point cloud registration based on one-point ransac and scale-annealing biweight estimation // IEEE Transactions on Geoscience and Remote Sensing. 2021. Т. 59. № 11. С. 9716-9729.
106. Tom Fawcett. An introduction to ROC analysis // Pattern Recognition Letters, Volume 27, Issue 8, 2006, P. 861-874, ISSN 0167-8655, <https://doi.org/10.1016/j.patrec.2005.10.010>.
107. Chum O., Matas J. Matching with PROSAC — progressive sample consensus // Proceedings of the 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05). – San Diego, CA, USA, 2005. – Vol. 1. – P. 220–226.
108. Potapova V. Image search by content system development / Tarasov A.S., Potapova V., Grinchenko N.N. // Proceedings of IEEE East-West Design & Test Symposium (EWDTS'2018). С. 625-629.
109. Tarasova V. Development of a Search System for Similar Images / Tarasov A.S., Tarasova V.Yu., Grinchenko N.N., Stepanov M.A. // 13th International conference ELEKTRO 2020. С. 6
110. Sablina V.A., Efimov A.I., Novikov A.I. Combined approach to object contour superimposition in heterogeneous images // В сборнике: 2017 6th Mediterranean Conference on Embedded Computing, MECO 2017 - Including ECYPS 2017, Proceedings. 6. 2017. С. 7977175.
111. Тарасова В.Ю. Разработка и исследование свёрточных моделей нейронных сетей в задачах поиска похожих изображений /Тарасова В.Ю., Костров Б.В. // Международный научно-исследовательский журнал. – 2025. – №6 (156). – URL: <https://research-journal.org/archive/6-156-2025-june/10.60797/IRJ.2025.156.22> (дата обращения: 17.06.2025). – DOI: 10.60797/IRJ.2025.156.22

112. Ding, Y., Astermark, J., Oskarsson, M., & Larsson, V. Noisy One-point Homographies are Surprisingly Good // Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2024.

ПРИЛОЖЕНИЕ 1 Акт внедрения результатов кандидатской диссертации в ООО «Квантрон Групп»

kvan
tron

ООО «Квантрон Групп»

390000, г. Рязань, ул. Каширина, д. 1Б, офис 507,
ИНН 6234189553, КПП 622901001, ОГРН 1206200003437,
Р/С 40702810502020001647, К/С 30101810200000000593,
БИК 044525593, АО «АЛЬФА-БАНК», г. Москва,
тел. +7 (4912) 722-270
info@kvantron.com
www.kvantron.com

Акт внедрения результатов кандидатской работы Тарасовой Валентины Юрьевны в производственном процессе компании ООО «Квантрон Групп»

г. Рязань, 24.03.2025

Настоящим актом удостоверяется, что результаты экспериментальных исследований, полученные Тарасовой В. Ю. в ходе работ над кандидатской диссертацией, внедрены в научно-технической деятельности предприятия ООО «Квантрон-групп».

Предлагаемый подход сопоставления изображений нашел свое применение в программно-аппаратном комплексе определения местоположения аэрофотосъемки. Представленное решение позволило реализовать механизм определения похожих изображений. Данная модель, основанная на неглубоких нейронных сетях позволила обеспечить большую точность (36,7%) и меньшее время работы (2,7 раза быстрее) в сравнении с глубокими нейронными сетями. Вычислительная сложность разработанной модели составляет всего 0,437 GFLOPS, что делает ее доступным к применению во встраиваемых решениях за приемлемое время работы (не более 30 мс на кадр).

Генеральный директор
ООО «Квантрон групп»,
кандидат технических наук



М.В. Шадрин

ПРИЛОЖЕНИЕ 2 Акт внедрения результатов кандидатской диссертации в учебном процессе ФГБОУ ВО «РГРТУ»

УТВЕРЖДАЮ

Проректор РГРТУ по учебной
работе



к.т.н., доцент

/ Корячко А. В.

« 12 » сентября 2023

МП

АКТ

внедрения результатов кандидатской диссертационной работы
Тарасовой Валентины Юрьевны
в учебном процессе
Рязанского государственного радиотехнического университета
им. В.Ф. Уткина

Настоящим актом удостоверяется, что результаты экспериментальных исследований, полученные Тарасовой В.Ю. в ходе работ над кандидатской диссертацией, внедрены в учебном процессе кафедры «Электронные вычислительные машины» Рязанского государственного радиотехнического университета в дисциплине:

1. «Системы технического зрения», направление подготовки – 09.04.01 «Информатика и вычислительная техника» (программа магистратуры)».

Издано учебное пособие:

Никифоров М.Б., Муратов Е.Р., Тарасова В.Ю. Системы технического зрения: методическое указание к лабораторным работам и практическим занятиям // Рязанский государственный радиотехнический университет им В.Ф. Уткина. Рязань, 2023 – 40 с., которое используется при изучении указанной дисциплины.

1. «Теоретические основы информатики», направление подготовки – 09.06.01 «Информатика и вычислительная техника» (программа аспирантура)».

Опубликован: электронный образовательный ресурс дистанционного учебного курса «Теоретические основы информатики» №24967, дата регистрации: 24.03.2022

Декан факультета вычислительной техники
д.т.н., профессор

Д. А. Перепелкин

Начальник отдела аспирантуры
к.э.н., доцент

Е. Е. Нефедова

Зав. кафедрой ЭВМ,
д.т.н., профессор

Б. В. Костров

ПРИЛОЖЕНИЕ 2. Свидетельства о государственной регистрации программ для ЭВМ

РОССИЙСКАЯ ФЕДЕРАЦИЯ



СВИДЕТЕЛЬСТВО
о государственной регистрации программы для ЭВМ

№ 2025691618

**Программа уменьшения размерности свёрточного слоя
нейронной сети с использованием спектрального
преобразования**

Правообладатель: *Федеральное государственное бюджетное
образовательное учреждение высшего образования
«Рязанский государственный радиотехнический
университет имени В.Ф. Уткина» (RU)*

Авторы: *Тарасова Валентина Юрьевна (RU), Костров Борис
Васильевич (RU), Ефимов Алексей Игоревич (RU)*

Заявка № **2025691298**
Дата поступления **17 ноября 2025 г.**
Дата государственной регистрации
в Реестре программ для ЭВМ **17 ноября 2025 г.**

*Руководитель Федеральной службы
по интеллектуальной собственности*



ДОКУМЕНТ ПОДПИСАН ЭЛЕКТРОННОЙ ПОДПИСЬЮ
Сертификат 00a570a4f7ca13452164b8818e75f29506
Владелец **Зубов Юрий Сергеевич**
Действителен с 04.07.2025 по 28.11.2026

Ю.С. Зубов

РОССИЙСКАЯ ФЕДЕРАЦИЯ



СВИДЕТЕЛЬСТВО

о государственной регистрации программы для ЭВМ

№ 2025691610

**Свёрточная нейронная сеть с гибридными слоями для
определения визуального сходства спутниковых
снимков**

Правообладатель: *Федеральное государственное бюджетное
образовательное учреждение высшего образования
«Рязанский государственный радиотехнический
университет имени В.Ф. Уткина» (RU)*

Авторы: *Тарасова Валентина Юрьевна (RU), Костров Борис
Васильевич (RU), Устюков Дмитрий Игоревич (RU)*

Заявка № 2025691279

Дата поступления 17 ноября 2025 г.

Дата государственной регистрации

в Реестре программ для ЭВМ 17 ноября 2025 г.



*Руководитель Федеральной службы
по интеллектуальной собственности*

ДОКУМЕНТ ПОДПИСАН ЭЛЕКТРОННОЙ ПОДПИСЬЮ
Сертификат 00a570e4f7ca13453164b881be75f29506
Владелец **Зубов Юрий Сергеевич**
Действителен с 04.11.2025 по 28.11.2026

Ю.С. Зубов

РОССИЙСКАЯ ФЕДЕРАЦИЯ



СВИДЕТЕЛЬСТВО

о государственной регистрации программы для ЭВМ

№ 2023687992

Программа поиска изображений по визуальной информации

Правообладатель: **Федеральное государственное бюджетное образовательное учреждение высшего образования "Рязанский государственный радиотехнический университет имени В.Ф. Уткина" (RU)**

Автор(ы): **Тарасова Валентина Юрьевна (RU)**

Заявка № **2023687652**

Дата поступления **08 декабря 2023 г.**

Дата государственной регистрации

в Реестре программ для ЭВМ **19 декабря 2023 г.**



Руководитель Федеральной службы
по интеллектуальной собственности

Ю.С. Зубов

РОССИЙСКАЯ ФЕДЕРАЦИЯ



СВИДЕТЕЛЬСТВО

о государственной регистрации программы для ЭВМ

№ 2019665048

Сверточная нейронная сеть для хэширования изображений

Правообладатель: *Федеральное государственное бюджетное образовательное учреждение высшего образования "Рязанский государственный радиотехнический университет имени В.Ф. Уткина" (RU)*

Авторы: *Гринченко Наталья Николаевна (RU), Тарасов Андрей Сергеевич (RU), Тарасова Валентина Юрьевна (RU)*

Заявка № **2019664040**

Дата поступления **07 ноября 2019 г.**

Дата государственной регистрации

в Реестре программ для ЭВМ **18 ноября 2019 г.**



*Руководитель Федеральной службы
по интеллектуальной собственности*

Г.П. Ивлиев Г.П. Ивлиев

РОССИЙСКАЯ ФЕДЕРАЦИЯ



СВИДЕТЕЛЬСТВО

о государственной регистрации программы для ЭВМ

№ 2019663343

**Программа для визуального поиска изображений с
применением операции свертки**

Правообладатель: *Федеральное государственное бюджетное образовательное учреждение высшего образования "Рязанский государственный радиотехнический университет имени В.Ф. Уткина" (RU)*

Авторы: *Гринченко Наталья Николаевна (RU), Геращенко Екатерина Сергеевна (RU), Тарасов Андрей Сергеевич (RU), Тарасова Валентина Юрьевна (RU)*

Заявка № **2019662073**

Дата поступления **01 октября 2019 г.**

Дата государственной регистрации

в Реестре программ для ЭВМ **15 октября 2019 г.**

*Руководитель Федеральной службы
по интеллектуальной собственности*

 *Г.П. Ивлиев*



ПРИЛОЖЕНИЕ 3. Патент на изобретение

