

МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ
РОССИЙСКОЙ ФЕДЕРАЦИИ
ФЕДЕРАЛЬНОЕ ГОСУДАРСТВЕННОЕ АВТОНОМНОЕ
ОБРАЗОВАТЕЛЬНОЕ УЧРЕЖДЕНИЕ ВЫСШЕГО ОБРАЗОВАНИЯ
«САМАРСКИЙ НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ УНИВЕРСИТЕТ
ИМЕНИ АКАДЕМИКА С. П. КОРОЛЕВА»
(САМАРСКИЙ УНИВЕРСИТЕТ)

На правах рукописи

Борисов Алексей Николаевич

**МЕТОД ОБЩЕГО ДОМЕНА И ЕГО ПРИМЕНЕНИЕ В ЗАДАЧАХ
ОБРАБОТКИ ИЗОБРАЖЕНИЙ ДИСТАНЦИОННОГО ЗОНДИРОВАНИЯ
ЗЕМЛИ**

2.3.8. Информатика и информационные процессы

Диссертация на соискание учёной степени
кандидата технических наук

Научный руководитель:

Мясников Владислав Валерьевич,
доктор физико-математических наук,
профессор

Самара – 2026

Оглавление

	Стр.
Введение	5
Глава 1. Обработка разнородных данных ДЗЗ: анализ современного состояния и формулировка решаемой проблемы.....	11
1.1 Общие тенденции в разработке методов и алгоритмов обработки и анализа изображений ДЗЗ	11
1.1.1 Совместная обработка разнородных изображений ДЗЗ	12
1.1.2 Обработка изображений ДЗЗ на основе фундаментальных моделей	15
1.2 Методы совмещения изображений ДЗЗ.....	17
1.2.1 Методы совмещения в компьютерном зрении	17
1.2.2 Методы совмещения космических снимков ДЗЗ.....	19
1.2.3 Совмещение разнородных изображений с использованием ИНС	21
1.3 Методы обработки изображений ДЗЗ на основе доменных преобразований	22
1.3.1 Спектральное согласование оптических снимков	22
1.3.2 Доменные преобразования изображений ДЗЗ на основе генеративных нейронных сетей.....	24
1.3.3 Методы доменного преобразования изображений	24
1.4 Основные направления исследований в работе	30
1.4.1 Ограничения методов доменных преобразований на базе генеративных ИНС	30
1.4.2 Взаимная информация и гипотеза общего домена	30
1.4.3 Применение общего домена в рамках общей технологии обработки данных ДЗЗ	33

1.5 Выводы и результаты первой главы	33
Глава 2. Метод построения общего домена для цифровых изображений ДЗЗ, полученных различными сенсорами	35
2.1 Общее описание метода.....	35
2.2 Построение общего домена в задаче спектрального согласования оптических снимков.....	37
2.2.1 Описание алгоритмов согласования	37
2.2.2 Экспериментальные исследования.....	43
2.3 Построение общего домена для оптических и радарных снимков	50
2.3.1 Описание алгоритма	50
2.3.2 Экспериментальные исследования.....	53
2.4 Выводы и результаты второй главы.....	60
Глава 3. Алгоритм совмещения разнородных изображений дистанционного зондирования Земли с использованием общего домена	61
3.1 Описание алгоритма	61
3.1.1 Общая схема алгоритма.....	61
3.2 Экспериментальные исследования.....	67
3.2.1 Показатели качества совмещения.....	67
3.2.2 Совмещение разнородных оптических снимков	69
3.2.3 Совмещение оптических и радарных снимков	74
3.3 Выводы и результаты третьей главы	87
Глава 4. Технология решения прикладных задач обработки разнородных изображений ДЗЗ с использованием метода общего домена	88
4.1 Общее описание технологии.....	88
4.2 Алгоритм классификации земной поверхности по данным разнородных изображений ДЗЗ	90

4.2.1 Постановка задачи	90
4.2.2 Экспериментальные исследования.....	93
4.3 Алгоритм обнаружения изменений по данным разнородных изображений ДЗЗ	100
4.3.1 Постановка задачи	100
4.3.2 Экспериментальные исследования.....	102
4.4 Выводы и результаты четвертой главы	107
Заключение	108
Список литературы.....	111
ПРИЛОЖЕНИЕ А.....	129
ПРИЛОЖЕНИЕ Б	130

Введение

Актуальность темы исследования

В современной мировой экономике одним из ключевых факторов для ее развития является информация, ее актуальность и полнота. Для многих отраслей источником такой информации являются средства космического дистанционного зондирования Земли (ДЗЗ). В задачах сельского хозяйства сегментация снимков ДЗЗ позволяет решать задачи прогнозирования урожайности и мониторинга эффективности мер повышения урожайности. Кроме того, сегментация по типам сельскохозяйственной растительности позволяет выявлять нецелевое использование сельскохозяйственных угодий. В задачах лесного хозяйства сегментация данных ДЗЗ может использоваться для определения типов лесных массивов и мониторинга их изменений (в том числе обнаружения незаконных вырубок). В задачах экологии данные сегментации ДЗЗ позволяют выявлять ареалы обитания различных типов растительности и наблюдать за изменениями данных ареалов во времени. Кроме того, сегментация снимков ДЗЗ может использоваться для выявления участков эрозии почвы, участков, пострадавших от затопления и участков, пострадавших от лесных пожаров. В задачах гидрологии сегментация данных ДЗЗ используется для картографирования водоемов, наблюдения за изменениями береговой линии и разливами рек. Наконец, в задачах кадастрового мониторинга сегментация данных ДЗЗ может использоваться для обнаружения несоответствий в кадастровых данных, обнаружения незаконных построек и нецелевого использования земельных участков (в частности, обнаружения незаконных свалок). Приведенный список примеров, разумеется, является неполным.

Центральное место среди источников данных ДЗЗ занимают спутники с оптическими сенсорами (сюда входят панхроматические сенсоры, сенсоры видимого диапазона, инфракрасные сенсоры, мультиспектральные и гиперспектральные сенсоры) и радарные спутники, использующие в качестве сенсора радар с синтезированной апертурой. Количество спутников ДЗЗ постоянно растет по мере вовлечения в процесс ДЗЗ различных стран и крупных

коммерческих компаний. В настоящее время активно развивается направление малых и наноспутников, что дополнительно расширяет круг операторов спутников.

С ростом количества спутников ДЗЗ возрастает объем разнородной и разновременной информации, которая может использоваться для решения прикладных задач. Хотя во многих случаях для решения задачи достаточно ограничиться только данными от одного спутника (или спутников одной группировки, имеющих одинаковые сенсоры), достаточно очевидно, что качественно лучших результатов можно добиться при совместном использовании информации, полученных от сенсоров различных типов.

Главным препятствием к совместному анализу являются существенные отличия между изображениями ДЗЗ, возникающие из-за различий в характеристиках сенсоров. Отличия между снимками радарных и оптических сенсоров в первую очередь следуют из разных физических принципов функционирования (активное зондирование излучением радиодиапазона/пассивный прием отраженного излучения в оптическом диапазоне и смежных к нему). Отличия в снимках оптических сенсоров следуют из различий в пространственном разрешении и составе набора спектральных каналов. Данные отличия существенно усложняют первый этап совместного анализа изображений ДЗЗ: этап пространственного совмещения. Последующий этап совместного анализа также, как правило, требует более сложных методов и алгоритмов, учитывающих особенности изображений разных типов.

Разработка и исследование методов обработки изображений ДЗЗ ранее рассматривалась в работах как российских (Сергеев В.В., Мясников В.В., Мясников Е.В., Сойфер В.А., Фурсов В.А., Казанский Н.Л., Визильтер Ю.В., Ташлинский А.Г., Потатуркин О.И., Ханьков И.Г., Ненашев В.А., Бурнаев Е.В., Дмитриев А.В. и др.), так и зарубежных (Schmitt M., Buzzone L., Ye Y., Chen H.M., Wang L., Li J., Hedges L.H., Wegner R., Cha K. и др.) ученых. Однако, существующие методы имеют ряд недостатков. Во-первых, существующие методы совместной обработки разнородных изображений ДЗЗ чаще всего нацелены на обработку уже пространственно совмещенных снимков, оставляя задачу предварительного

совмещения изображений без внимания. Обратное также верно: методы совмещения разнородных изображений ДЗЗ решают только задачу совместного совмещения, но не дальнейшего анализа. Методы, основанные на использовании генеративных нейронных сетей, распространенные в настоящее время, хотя и могут использоваться для решения и задачи совмещения, и задачи совместной обработки, тем не менее, уступают таргетированным решениям вследствие неточностей, вносимых при преобразовании изображения из одного домена в другой. Фундаментальные модели, отличающиеся своей универсальностью, требуют больших объемов данных и вычислительных мощностей на этапе обучения, что ограничивает их применимость для решения задач, связанных с конкретными сенсорами и конкретными территориями.

Как следствие, разработка нового метода обработки разнородных изображений ДЗЗ, обеспечивающего как их совмещение, так и дальнейший совместный анализ на основе существующих решений, является актуальной.

Цель и задачи исследования

Целью диссертационной работы является разработка и исследование методов и алгоритмов обработки разнородных данных ДЗЗ, позволяющих в рамках единой методологии и с улучшенными качественными показателями решать задачу пространственного совмещения разнородных снимков и различные прикладные задачи.

Для достижения поставленной цели в диссертации решаются следующие задачи:

- 1) разработка и исследование метода и алгоритмов преобразования разнородных снимков к единому представлению (общему домену);
- 2) разработка и исследование алгоритма пространственного совмещения изображений с использованием общего домена;
- 3) разработка и исследование технологии решения прикладных задач обработки разнородных изображений ДЗЗ с использованием общего домена.

Поставленные задачи определяют структуру работы и содержание ее глав и разделов.

Практическая значимость работы

Разработанные методы и алгоритмы могут быть использованы в составе информационных систем, выполняющих обработку данных ДЗЗ от космических аппаратов с сенсорами различного типа для решения задач совмещения снимков, классификации земной поверхности, обнаружения изменений, обнаружения объектов.

Методология и методы исследования

В диссертационной работе используются методы статистического анализа, методы обработки изображений и методы машинного обучения.

Научная новизна результатов

1) Предложен метод построения общего домена для разнородных изображений ДЗЗ, минимизирующий различия между их образами в общем домене и обеспечивающий их невырожденность одновременно с сохранением пространственной информации; разработаны алгоритмы, конкретизирующие предложенный метод для различных типов данных ДЗЗ и критериев.

2) Разработан алгоритм пространственного совмещения разнородных изображений ДЗЗ, использующий их преобразование в общий домен для устранения различий между изображениями и учет данных геопривязки для увеличения точности совмещения изображений.

3) Разработана технология решения прикладных задач обработки разнородных изображений ДЗЗ, использующая представление изображений в общем домене, и алгоритмы прикладной обработки изображений ДЗЗ в рамках данной технологии.

На защиту выносятся:

1) Метод построения общего домена для разнородных изображений ДЗЗ, позволяющий на основании анализа приводить их к невырожденному представлению – общему домену – в рамках которого образы изображений обладают высокой степенью сходства; и алгоритмы, конкретизирующие метод для различных типов данных ДЗЗ и критериев.

2) Алгоритм пространственного совмещения разнородных изображений ДЗЗ, использующий преобразование изображений в общий домен для устранения различий между изображениями и учет данных геопривязки для увеличения точности совмещения изображений.

3) Технология решения прикладных задач обработки разнородных изображений ДЗЗ, использующая представление разнородных изображений в общем домене, и алгоритмы прикладной обработки изображений ДЗЗ в рамках данной технологии, обеспечивающие в рамках единой методологии решение задач классификации земной поверхности и обнаружения изменений по данным радарных и оптических изображений с качеством, превосходящим качество решения этих задач в рамках исходных доменов.

Соответствие специальности

Область исследования соответствует направлениям исследований паспорта научной специальности 2.3.8. Информатика и информационные процессы в части пунктов:

1. Разработка компьютерных методов и моделей описания, оценки и оптимизации информационных процессов и ресурсов, а также средств анализа и выявления закономерностей на основе обмена информацией пользователями и возможностей используемого программно-аппаратного обеспечения.

4. Разработка методов и технологий цифровой обработки аудиовизуальной информации с целью обнаружения закономерностей в данных, включая обработку текстовых и иных изображений, видео контента. Разработка методов и моделей распознавания, понимания и синтеза речи, принципов и методов извлечения требуемой информации из текстов.

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

Степень достоверности и апробация результатов

Достоверность полученных результатов подтверждается корректным применением математического и статистического аппарата, экспериментальными

исследованиями, обсуждением результатов диссертации на международных конференциях.

Основные результаты диссертации были представлены на 2 научных конференциях:

1) XI Международная научная конференция "Информационные технологии и нанотехнологии" (ITNT-2025, Самарканд, Узбекистан) – 2025 год;

2) LV Международная научно-практическая конференция «Актуальные проблемы науки и образования в условиях современных вызовов» (Москва, Россия) – 2026 г.

По теме диссертационной работы было опубликовано 6 работ (далее работы автора диссертации обозначаются символом «*»). Из них одна опубликована в издании, рекомендуемом ВАК, три работы опубликованы в рецензируемых изданиях, входящих в международные базы данных Scopus и/или WebOfScience. Три работы опубликованы без соавторов.

Результаты диссертационной работы использованы:

1) При выполнении работ по гранту Министерства образования и науки РФ «Фундаментальные проблемы разработки аэрокосмических транспортных систем и управления в аэрокосмической технике для обеспечения связанности территории Российской Федерации» (соглашение № 075-15-2024-558).

2) В хоздоговорных НИОКР в АО «Самара-Информспутник».

Соответствующие акты использования результатов приведены в приложениях.

Диссертация состоит из 4 глав, заключения, списка литературы из 157 наименований, содержит 18 рисунков, 24 таблицы, 2 приложения.

Глава 1. Обработка разнородных данных ДЗЗ: анализ современного состояния и формулировка решаемой проблемы

1.1 Общие тенденции в разработке методов и алгоритмов обработки и анализа изображений ДЗЗ

В настоящее время наблюдается заметный рост числа доступных источников данных дистанционного зондирования Земли (ДЗЗ). Информация, получаемая средствами ДЗЗ, позволяет решать широкий спектр задач в различных областях науки и экономики. Однако, рост объемов и разнообразия доступных данных ДЗЗ требует разработки новых подходов к обработке и анализу данных [1,2].

Широкое распространение в области обработки ДЗЗ получили методы машинного обучения, главным достоинством которых является адаптивность: решение задачи строится по мере анализа доступных данных, выявления пригодных для решения задачи закономерностей и их использования для формирования результата.

Классические алгоритмы машинного обучения до сих пор широко применяются при анализе данных ДЗЗ. К примеру, машины опорных векторов [3], деревья решений [4, 5] и случайные леса [6] широко используются для решения задач классификации земной поверхности. Методы кластеризации [7,8] активно используются при анализе мультиспектральных и гиперспектральных снимков для выполнения т.н. классификации без учителя, в рамках которой классы задаются не априорно, а формируются исходя из данных. Методы понижения размерности (PCA [9, 10], LDA [11], методы градиентного спуска [12, 13]) используются при анализе мультиспектральных и гиперспектральных снимков для формирования компактных признаковых описаний, позволяющих решать другие прикладные задачи более эффективно за счет меньшей размерности описаний.

Из методов машинного обучения в настоящее время наиболее активно развиваются методы глубокого обучения, и, в частности, аппарат искусственных

нейронных сетей (ИНС). Широчайшее применение в задачах обработки изображений ДЗЗ нашли сверточные ИНС [14], исходно предложенные для решения задач распознавания образов [15]. В области обработки данных ДЗЗ сверточные ИНС используются для классификации земной поверхности [14] и смежных к ней задачах обнаружения зон изменений [16], пожаров [17], затоплений [18] и обнаружения объектов [19, 20], повышения разрешения снимков [21].

Поскольку космические аппараты ДЗЗ осуществляют съемку территории с некоторой периодичностью, возможно выполнять анализ данных ДЗЗ, учитывая весь мультिवременной ряд данных по территории. Для анализа таких мультिवременных рядов широко применяются рекуррентные ИНС и LSTM-сети [22-24].

Среди других архитектур ИНС, используемых для решения задач ДЗЗ, следует упомянуть автоэнкодеры (автокодировщики). Типовыми задачами применения автоэнкодеров является уменьшение размерности [25] и устранение шума [26], однако, в контексте обработки ДЗЗ автоэнкодеры также применяются для обнаружения изменений на снимках [27].

Трансформеры, изначально разработанные для решения задач обработки естественного языка (англ. natural language processing, NLP) [28], также находят применение в анализе данных ДЗЗ, в первую очередь из-за возможности моделирования глобальных зависимостей в данных (сверточные ИНС, к примеру, выполняют оконную обработку изображения, и, как следствие, учитывают только локальные зависимости). ИНС данного типа широко применяются для задач классификации земной поверхности [29], обнаружения изменений [30], обнаружения объектов [31], повышения разрешения [32].

1.1.1 Совместная обработка разнородных изображений ДЗЗ

С ростом количества спутников ДЗЗ возрастает объем разнородной информации, которая может использоваться для решения прикладных задач. Хотя

во многих случаях для решения задачи достаточно ограничиться только данными от одного спутника (или спутников одной группировки, имеющих одинаковые сенсоры), достаточно очевидно, что лучших результатов можно добиться при совместном использовании информации с нескольких различных спутников.

Совместная обработка снимков, полученных сенсорами с близкими характеристиками, не представляет большой проблемы, поскольку в этом случае для обработки могут использоваться одни и те же методы (примером могут быть сенсоры спутников SPOT-6/7 и GaoFen-1). Однако, даже сенсоры одного класса могут существенным образом отличаться. К примеру, оптические сенсоры могут варьироваться по количеству спектральных каналов (один канал для панхроматических сенсоров, 3 канала для сенсоров видимого диапазона, 4 и больше каналов для мультиспектральных сенсоров, сотни каналов для гиперспектральных сенсоров), центральным длинам волн и ширине спектральной полосы индивидуальных каналов (канал ближнего инфракрасного диапазона Landsat-8 – 450-510 нм, тот же канал Sentinel-2 – 457-522 нм), чувствительности каналов, пространственному разрешению (GaoFen-7 – 2,6 м/пиксель, Landsat-8 – 30 м/пиксель на большинстве каналов, Terra MODIS – 500 м/пиксель), методам постобработки и выбираемым для представления отсчетов числовом диапазоне (диапазон значений снимков Landsat-8 – $[0; 65455]$, Sentinel-2 – $[0; 16384]$). В случае разнотипных сенсоров, помимо отличий в пространственном разрешении, возникает также существенное различие в самом характере изображения. Стоит также отметить, что время съемки территории разными спутниками чаще всего не совпадает, из чего следуют различия, вызываемые изменениями, произошедшими на территории (к примеру, сезонные изменения – паводки, выпадение снега, начало вегетации).

Указанные выше отличия являются существенным затруднением, усложняющим совместный анализ данных ДЗЗ, полученных разными сенсорами. Поскольку для проведения совместной обработки снимков требуется достаточно точное их пространственное совмещение, возникает задача совмещения снимков с

существенно отличающимися характеристиками (подробно методы решения данной задачи рассмотрены в разделе 1.2).

Задача использования данных от различных сенсоров рассматривалась и вне контекста ДЗЗ. К примеру, в задачах автономного вождения часто рассматривается использование данных обычной камеры и лидара, используемого для формирования карты глубины [33]. Процесс формирования общего признакового описания по нескольким изображениям называется слиянием изображений (англ. image fusion) [34].

С точки зрения совместного анализа изображений ДЗЗ наибольший интерес представляет обработка разнородных изображений и, в частности, радарных и оптических снимков: радарные снимки имеют высокое разрешение и нечувствительны к погодным условиям, в то время как оптические сенсоры предоставляют больше информации о территории за счет использование нескольких спектральных каналов.

Простым методом слияния изображений является их объединение (в самом простом случае – конкатенация) на ранних этапах обработки с последующим формированием нового изображения, сочетающего особенности исходных снимков – т.н. ранее (англ. early fusion) или попиксельное слияние (англ. pixel-level fusion) [35-38]. Данный тип слияния требует высокой точности совмещения изображений, кроме того, при слиянии изображений доменов с существенно различающимися характеристиками возникает проблема выбора сохраняемых особенностей [37].

Другим подходом является независимый расчет признаковых описаний для каждого изображения, с последующим слиянием признаковых описаний – т.н. слияние признаков (англ. middle fusion, feature-level fusion) [35, 37, 39]. Поскольку в рамках данного подхода нет необходимости получить новое изображение, по своим свойствам близкое к исходным доменам, слияние признаков чаще всего применяется при решении прикладных задач [24, 37, 40, 41].

Третий подход, называемый также поздним слиянием (англ. late fusion, decision-level fusion) [37] предполагает независимое решение задачи в рамках каждого домена (к примеру, сегментацию [42] или обнаружение объектов [43]), с

последующей обработкой полученных частных результатов. Важно отметить, что в рамках данного подхода совместная обработка непосредственно разнородных данных отсутствует, как следствие, нет выявления взаимных закономерностей в данных разных доменов.

1.1.2 Обработка изображений ДЗЗ на основе фундаментальных моделей

Одним из новых направлений в обработке ДЗЗ является использование т.н. фундаментальных моделей – моделей, обученных на больших объемах разнородных данных ДЗЗ и пригодных для решения разнообразных задач: классификации/сегментации земной поверхности по данным нескольких доменов или реконструкции снимков одного домена по снимкам других доменов [44]. Примерами фундаментальных моделей являются Galileo [45], OmniSat [46], AnySat [47], MAESTRO [48] и TerraMind [49].

Стоит отметить, что в случае использования фундаментальных моделей набор необходимых изображений исходных доменов на входе модели не фиксирован: к примеру, фундаментальная модель, обученная на радарных и оптических снимках, вполне может функционировать при предоставлении только оптических или только радарных снимков.

Обучение фундаментальных моделей обычно выполняется в 2 этапа. На первом этапе предварительного модель обучается без учителя с целью конструирования латентного пространства признаков, в рамках которого могут быть представлены признаки всех доменов. Распространенной техникой на данном этапе является маскирование части входных данных с последующим их восстановлением по данным других доменов фундаментальной моделью, и оценкой расхождения реальных и реконструированных данных. Обученная таким образом модель может функционировать в том числе, как генеративная ИНС, поскольку по снимкам одного домена может реконструировать снимки другого домена. На втором этапе к модели добавляется группа новых слоев – т.н. «голова» –

выполняющая решение требуемой практической задачи на основе данных из латентного пространства. Обучение «головой» выполняется методами обучения с учителем. В данном контексте стоит отметить модель TerraMind, при обучении которой данные классификации земной поверхности рассматриваются не как данные прикладной задачи, а как один из входных доменов, за счет чего TerraMind может решать задачу классификации земной поверхности (с ограничением на множество классов) сразу после первого этапа обучения.

Преимуществом фундаментальных моделей является их универсальность: обученная фундаментальная модель может использоваться для обработки снимков от нескольких разнотипных сенсоров, на которых она была обучена. Кроме того, фундаментальные модели могут легко дообучаться для обработки снимков от новых сенсоров (при условии, что эти сенсоры по характеристикам достаточно близки к исходным) и адаптируются для решения новых задач путем за счет использования специфичной для задачи «головой» без изменения основной части модели.

Недостаток фундаментальных моделей следует из процесса их обучения: требуется большой объем согласованных разнородных данных, который, в свою очередь, требует значительного количества вычислительных ресурсов для обучения фундаментальной модели за приемлемое время. К примеру, для завершения основного этапа обучения AnySat [47] потребовалось 1760 машино-часов на графическом процессоре NVIDIA H100 (номинальная стоимость единицы более 20 000\$). Как следствие, для решения прикладных задач для конкретной территории и конкретного набора сенсоров более приемлемым вариантом могут быть не-фундаментальные модели.

1.2 Методы совмещения изображений ДЗЗ

1.2.1 Методы совмещения в компьютерном зрении

Задача совмещения цифровых оптических изображений исторически возникала в приложениях фотограмметрии и смежных направлениях технического зрения [50] и рассматривалась вне контекста обработки цифровых изображений ДЗЗ. Первые методы совмещения изображений были основаны на максимизации некоторой метрики, например корреляции [51] или взаимной информации [52-54]. В этом случае сопоставление изображений приобретает характер оптимизационной задачи, что, с учетом изменения характеристик изображения на каждом шаге, требует больших вычислительных ресурсов.

Более современным подходом к решению задачи совмещения изображений является совмещение на основе сопоставления особых точек. Особые точки обычно расположены на визуальных особенностях изображения: границах, углах, центрах замкнутых контуров и т.д. В этом случае задача сопоставления изображений состоит из этапов обнаружения особых точек (выполняется т.н. детекторами) или т.н. точек интереса (англ.: RoI – Region of Interest), построения описания их локальной окрестности (выполняется т.н. дескрипторами), последующего сопоставления признаков описаний особых точек (т.н. дескрипторов точек - N-мерных векторов признаков) и формирования их пар на основе ограничений, возникающих из специфики геометрического преобразования между изображениями. Последний этап обычно выполняется с использованием алгоритмов RANSAC [55] или PROSAC [56]. При сравнении дескрипторов особых точек для формирования их пар часто используется стандартное евклидово расстояние, хотя могут быть использованы и иные метрики/меры.

Наиболее известным алгоритмом описания особых точек является запатентованный алгоритм SIFT [57], породивший ряд модификаций [58,59]. Более новые алгоритмы данного типа, например, SURF [60] и ORB [61], обещают лучшее

качество сопоставления изображений, однако SIFT до настоящего времени остается наиболее распространенным алгоритмом.

Примером алгоритма поиска особых точек является детектор углов FAST [62], который использует дерево решений, построенное на обучающей выборке, для улучшения обнаружения углов на целевом изображении.

Минусом многих детекторов и дескрипторов является отсутствие адаптивности – алгоритмы расчета, используемые для поиска и/или описания особых точек, детерминированы и фиксированы. Это порождает проблемы при использовании их на изображениях, имеющих существенные различия – особенно в случае изображений из различных доменов (например, оптических изображений и карт глубины). Проблема адаптации алгоритмов обнаружения и описания особых точек на изображениях разных доменов в настоящее время решается с помощью методов машинного обучения, включающих и нейронные сети.

Многие современные алгоритмы сопоставления изображений, в том числе с использованием особых точек, основаны на искусственных нейронных сетях. Примерами являются алгоритмы LIFT [63], R2D2 [64], Superpoint [65], DISK [66], ALIKED [67]. В зависимости от алгоритма, для поиска особых точек и расчета дескрипторов может быть использована как одна ИНС, так и несколько отдельных ИНС. Во многих случаях (особенно в случае использования сверточных ИНС) результатом работы алгоритма являются тензор дескрипторов, в котором содержатся дескриптор для каждого пикселя или группы пикселей, и маска ключевых точек, отмечающая особые точки. Отсюда следует, что может быть выполнено попиксельное сопоставление изображений, что подводит нас к следующему (бездетекторному) типу алгоритмов сопоставления изображений. Современным подтипом методов сопоставления изображений по особым точкам являются т.н. бездетекторные (англ.: detector-free) методы, которые выполняют одновременно обнаружение и сопоставление особых точек на изображении (т.е. результатом работы является набор пар точек). Обычно на одном из этапов работы метода производится расчет полей признаков для сопоставляемых изображений, за которым следует их сопоставление с последующим подавлением не-максимумов

(англ.: non-maximum suppression). Примерами алгоритмов подобного типа являются LoFTR [68], NCNet [69].

Наиболее современным подходом является т.н. попиксельное сопоставление (англ.: dense matching), при котором результатом работы алгоритма является карта соответствий между пикселями первого и второго изображений. Такой подход потенциально позволяет строить соответствия между изображениями в условиях сильных геометрических искажений. Недостатком такого подхода является высокая вычислительная сложность, однако, в отличие от классического оптимизационного подхода, алгоритмы этого типа отрабатывают за один шаг. Примером алгоритма из данного класса, не использующего машинное обучение, является SIFTFlow [70]. Примерами методов попиксельного сопоставления изображений являются DKM [71], RoMa [72]. Метод LoFTR из предыдущей категории на одном из этапов выполняет попиксельное сопоставление на большом масштабе, за которым следует определение точных координат точек.

Стоит отметить, что, если поля двух изображений связаны между собой аффинным преобразованием, то использование попиксельного преобразования является избыточным, и может приводить к обнаружению ложных соответствий. Подобная проблема может быть относительно просто решена путем отбора некоторого достаточного (для оценки параметров преобразования) количества наиболее качественных пар точек.

1.2.2 Методы совмещения космических снимков ДЗЗ

1.2.2.1. Совмещение оптических снимков

Задача совмещения оптических снимков, полученных от одного сенсоров, во многом аналогична задаче совмещения обычных изображений. В частности, совмещение одно- и трехканальных изображений оптических ДЗЗ вполне может проводиться методами совмещения обычных изображений с поправкой на

требуемый диапазон значений (требуемый диапазон – $[0, 255]$ или $[0,1]$). Стоит отметить, что при количестве каналов более 3 возникает проблема выбора оптимального с точки зрения сходства набора каналов.

Большую сложность представляет задача совмещения оптических снимков, полученных от сенсоров с существенными различиями в пространственном и/или спектральном разрешении. Среди методов сопоставления изображений, не использующих ключевые точки, можно отметить подходы на основе корреляции [51], взаимной информации [52] и нелинейного отображения тонов [73].

Также стоит отметить работы [74, 75], в которых рассматриваются методы сопоставления снимков от конкретных спутников. В данных работах для сопоставления снимков активно используются дополнительные данные (данные геопривязки, данные о положении детектора при съемке).

1.2.2.2. Совмещение радарных снимков

Задача совмещения радарных снимков от разных сенсоров ранее не рассматривалась. Однако, задача совмещения радарных снимков одного и того же космического аппарата и сенсора, вообще говоря, возникает как задача интерферометрического анализа радарных снимков – InSAR [76], однако в данном случае совмещаются снимки одного и того же сенсора, что существенно упрощает задачу.

Наиболее известным методом совмещения радарных снимков является метод SAR-SIFT [59]. Методы OS-SIFT [77], PSO-SIFT [78], RIFT [79] RIFT2 [80], появившиеся позже и используемые для сопоставления разнородных изображений, способны сопоставлять радарные снимки.

1.2.2.3. Совмещение радарных и оптических снимков

Совмещение радарных и оптических снимков представляет собой отдельную задачу в силу наличия спекл-шума на радарных изображениях. Специфика шума не позволяет автоматически перенести методы и алгоритмы, используемые для работы с оптическими изображениями, на радарные изображения.

Также как в рассмотренной выше задаче сопоставления изображений в компьютерном зрении, первые алгоритмы сопоставления пар оптических и радарных изображений ДЗЗ опирались на оптимизацию некоторой глобальной метрики, например, взаимной информации [81]. Дальнейшим развитием данной техники являются методы HOG [82], CFOG [83] и HOPC [84], в рамках которых сначала на изображении отыскиваются особые точки (например, углы с помощью детектора Харриса [85]), а затем выполняется точное совмещение фрагментов, содержащих окрестности особых точек на основе заданного метода сопоставления по шаблону. Данный подход позволяет существенно снизить вычислительную нагрузку за счет уменьшения размеров областей, по которым вычисляется оптимизируемая функция.

Для сопоставления пар оптических и радарных изображений ДЗЗ также используются и детекторы/дескрипторы особых точек. Примерами являются SIFT-подобные детекторы OS-SIFT [77], PSO-SIFT [78].

В отдельную категорию следует выделить методы RIFT [79] и RIFT2 [80], поскольку они предназначены для сопоставления разнородных изображений и потому представляет особый интерес.

1.2.3 Совмещение разнородных изображений с использованием ИНС

Интенсивное развитие в последнее десятилетие методов глубокого обучения существенным образом отразилось и на методах сопоставления снимков. Вместо детерминированных алгоритмов обнаружения и описания особых точек и

фрагментов изображений, разрабатываемых с учетом опыта людей-экспертов, существенно эффективнее оказались обучаемые алгоритмы (глубоких нейронных сетей), настраиваемые по данным конкретных доменов. Использование глубокого обучения позволило не только адаптировать детекторы и описатели к конкретным условиям наблюдения и/или доменам, но и позволило обеспечить сближение результатов конструируемых детекторов и описателей при совмещении разнородных изображений ДЗЗ.

Примеры использования методов совмещения стандартных изображений для совмещения изображений ДЗЗ можно найти в [86,87].

Примерами методов, разработанных специально для решения задачи совмещения разнородных изображений ДЗЗ, являются ACAMatch [87], CAMM-net [88], MatchosNet [89], PLISA [90].

Одним из нововведений в задаче совмещения изображений ДЗЗ, связанным с развитием методов глубокого обучения, оказалась возможность перевода изображений из одного домена в другой (англ.: domain translation). Например, при совмещении оптического и радарного изображений второе из них преобразуется к виду оптического, что позволяет для обоих изображений использовать одинаковые детекторы и дескрипторы. Подобно данный подход рассмотрен в подразделе 1.3.3.2.

1.3 Методы обработки изображений ДЗЗ на основе доменных преобразований

1.3.1 Спектральное согласование оптических снимков

Оптические сенсоры, используемые в действующих космических аппаратах, широко варьируются по пространственному разрешению, количеству и ширине спектральных каналов. Отличия в разрешении могут устраняться интерполяцией или методами повышения разрешения [91, 92].

Во многих случаях воспринимаемые сенсором спектральные каналы имеют значительное пересечение по спектральным полосам, что позволяет проводить совместный анализ снимков, ограничиваясь только этими каналами. Тем не менее, отличия в ширине и чувствительности спектральных полос, а также отличия в диапазоне значений, выбранном для представления яркости отсчетов, могут представлять собой проблему.

Отсюда возникает задача спектрального согласования снимков: приведения оптических снимков к единому числу каналов с совпадающими характеристиками.

В случае, если множество каналов выбрано заранее и фиксировано, задача состоит в оценке преобразования между яркостями отчетов соответствующих каналов. Данная задача решалась в работе [93] в ходе составления единого набора данных на основе спектрально близких каналов снимков, полученных спутниками Sentinel-2, Landsat-7 и Landsat-8. В изображения итогового набора данных были включены только 6 близких спектральных диапазонов каждого сенсора (красный, синий, зеленый и 3 инфракрасных диапазона). После совмещения снимков выполнялась корректировка значений пикселей с целью компенсации различий в ширине и центральной длине волны выбранных спектральных диапазонов. Стоит отметить, что в данном случае простое линейное преобразование оказалось достаточным для компенсации различий в спектральных полосах.

Если фиксировано только количество каналов, но не их спектральные характеристики, то вместо каналов с заданной спектральной полосой можно использовать «синтетические» каналы, яркость которых зависит не только от значения некоторого соответствующего ему канала (и, возможно, близких к нему) исходного изображения, а от всего множества каналов. Данный подход рассмотрен в работах [94*, 95*].

1.3.2 Доменные преобразования изображений ДЗЗ на основе генеративных нейронных сетей

1.3.3 Методы доменного преобразования изображений

Задача доменного преобразования изображений (в англоязычной литературе называемая image-to-image translation) состоит в поиске преобразования $G: \mathbb{X} \rightarrow \mathbb{Y}$, где $\mathbb{X}$, $\mathbb{Y}$ – исходные домены изображений, такого что $\forall X \in \mathbb{X}$ сгенерированное изображение $\hat{Y} = F(X)$ было «неотличимо» от изображений домена $\mathbb{Y}$. Способ определения сходства доменов домена сгенерированных изображений $\hat{\mathbb{Y}} = \{\hat{Y} = F(X) | X \in \mathbb{X}\}$ и домена реальных изображений $\mathbb{Y}$ зависит от используемого метода решения задачи. Стоит отметить, что доменное преобразование изображений может рассматриваться, как часть более широкой задачи генерации изображений, с тем отличием, что в рамках общей задачи генерации изображений $\hat{Y} = F(\xi)$, где ξ – некоторые входные данные, задающие начальные условия (чаще всего – некоторый псевдослучайный вектор с нормальным распределением координат), в то время как в задаче преобразования изображений в роли начальных условий выступает изображение исходного домена X .

Первые методы доменных преобразований, строившиеся на основе сверточных ИНС [96] и вариационных автоэнкодеров [97], в настоящее время практически полностью вытеснены методами на основе генеративно-состязательных сетей (англ. generative adversarial networks, GAN) [98]. В рамках генеративно-состязательных сетей одновременно обучаются 2 сети: сеть-генератор $G: \mathbb{X} \rightarrow \mathbb{Y}$ и сеть-дискриминатор $D: \mathbb{Y} \rightarrow [0; 1]$. Задача дискриминатора заключается в различении $\hat{\mathbb{Y}}$ и $\mathbb{Y}$ ($\mathbb{E}_{Y \in \mathbb{Y}}(D(Y)) + \mathbb{E}_{X \in \mathbb{X}}(1 - D(G(X))) \rightarrow \min$), задача генератора обратна задаче дискриминатора ($\mathbb{E}_{X \in \mathbb{X}}(1 - D(G(X))) \rightarrow \max$). Среди распространенных методов доменных преобразований на основе генеративно-состязательных сетей можно отметить pix2pix [99], CycleGAN [100], U-GAT-IT [101] и их модификации. Отдельно стоит упомянуть метод CycleGAN, в рамках

которого совместно обучаются сразу два генератора и два дискриминатора (рассматриваются цепочки преобразований $X \xrightarrow{G_1} \hat{Y} \xrightarrow{G_2} \hat{X}$ и $Y \xrightarrow{G_2} \hat{X} \xrightarrow{G_1} \hat{Y}$). Данный подход позволяет обучать ИНС преобразования в тех случаях, когда составление обучающей выборки в виде пар (входное изображение, целевое изображение) невозможно.

Дальнейшим развитием генеративных ИНС являются диффузионные модели [102] (которые по качеству генерируемых изображений могут превосходить классические генеративно-состязательные сети [103]), и, в последнее время, трансформеры [104].

1.3.3.1. Совмещение изображений с использованием генеративных сетей

Преобразование изображения из одного домена в другой, выполняемое генеративными сетями, позволяет использовать методы совмещения однородных изображений.

Как уже отмечалось ранее, в рамках задачи совмещения разнородных изображений наиболее часто освещается задача совмещения радарных и оптических изображений. Возможность совмещения радарных и оптических изображений на основе доменного преобразования впервые была показана в [105], в которой использовалось преобразование оптических изображений в радарный домен ИНС $pix2pix$ с последующим совмещением изображений на основе максимизации нормализованной корреляции. Эта же генеративная ИНС использовалась в работе [106] для выполнения преобразования радарных в оптический домен с последующим их совмещением с помощью метода, предназначенного для совмещения однородных изображений.

Последующие работы, использующие генеративные ИНС для приведения разнородных изображений к единому домену, используют либо модифицированные генеративные ИНС, либо модифицируют процесс обучения без изменения самой генеративной ИНС. К примеру, в работе [107] рассматривается преобразование

радарных изображений в оптические модифицированным вариантом CycleGAN, обученным с измененной функцией потерь, с последующим детектированием особых точек алгоритмом SIFT. В работе [108] при обучении CycleGAN используется дополнительная информация о кластеризации (вариант сегментации, базирующийся на выделении сходных областей на снимке, а не на априорных данных) радарного и оптического снимков; в качестве методов обнаружения и сопоставления особых точек рассматриваются SIFT, SAR-SIFT [59] и PSO-SIFT.

Интересный подход был предложен в работе [109], в рамках которого в качестве дескриптора особой точки используется выход 7-го слоя ИНС-генератора по данному соответствующего домена для окрестности особой точки 256×256 пикселей, развернутый в вектор (в случае изображений размером 256×256 пикселей данный вектор имеет размерность 2048). Сходство дескрипторов обеспечивается добавлением в функцию потерь евклидова расстояния между дескрипторами при обучении на парах совмещенных изображений разных доменов. Применение данного метода состоит в выделении особых точек на изображении некоторым внешним методом, выделении области 256×256 вокруг особой точки, вычислении дескрипторов и сопоставлении дескрипторов с помощью отдельно обученной ИНС на базе многослойного перцептрона. Использование ИНС сопоставления позволяет достичь инвариантности дескрипторов к проективным преобразованиям, которая не может быть обеспечено на этапе обучения CycleGAN (т.к. на данном этапе при обучении используются геометрически совмещенные изображения).

1.3.3.2. Применение генеративных нейронных сетей в задачах классификации и распознавания объектов

В задачах классификации, сегментации и распознавания образов приведение снимков к единому домену позволяет выполнять классификацию/сегментацию/обнаружение объектов с помощью единого метода,

работающего на данных целевого домена. Вновь отметим, что наиболее часто в данном контексте рассматривается использование радарных и оптических доменов.

Преобразование из радарного в оптический домен с целью обнаружения зон, пострадавших от пожаров, по данным Sentinel-1/2 рассматривается в работе [110]. Поскольку используемый в работе метод преобразования разработан для 3-канальных изображений, мультиспектральные снимки сводились к 3-канальным выбором каналов 11, 12 и 8А (инфракрасные каналы, пригодные для оценки урона от пожара); радарные снимки преобразовывались из 2-канальных к 3-канальным добавлением синтетического канала, содержащего показатель $NDBI = 10 \lg \frac{VV - VH}{VV + VH}$; изображения до и после пожара синтезировались по снимкам за год до пожара и год после пожара. Результирующие карты пострадавших областей, полученных по паре оптических снимков до и после пожара и по паре сгенерированные оптический снимок до пожара/реальный оптический снимок после пожара согласованы с коэффициентом каппа 0,77.

Несколько иной подход используется в работе [111], где вместо прямого преобразования из радарного в оптический домен используется преобразование из синтетического изображения, сформированного из радарных изображений до/после пожара, в домен показателя NBR (normalized burn ratio), рассчитываемого по данным оптических изображений. Дальнейшая классификация выполняется по данным сгенерированного изображения с последующей постобработкой, устраняющей малые ложно-классифицированные области, с итоговым значением показателя $F_1=0,91$.

Преобразование оптических изображений в радарные используется в задачах обнаружения кораблей и других объектов, имеющих четкий отклик на радарных изображениях. В работе [112] предлагается новый метод CDA-GAN для преобразования оптических изображений в радарные с целью детектирования кораблей. При обучении только на сгенерированных данных точность обнаружения (англ. precision) на реальных данных составила 91% при полноте (англ. recall) 48%. При обучении на сгенерированных и реальных данных в соотношении 2:1 точность

повысилась до 92%, полнота – до 87%. В работе [113] рассматривается модификация CycleGAN, при обучении которой учитывается как сходство сгенерированных и реальных изображений в радарных и оптических доменах, так и качество детектирования кораблей с целью улучшения качества генерации соответствующих изображений, с качеством детектирования кораблей на сгенерированных изображениях, превышающим качество детектирования на радарных изображениях (значения показателя *mIoU* 0,679 против 0,652 соответственно).

Генерация оптического изображения по данным радарного с помощью диффузионной ИНС используется в [41] для решения задач классификации земной поверхности и обнаружения объектов по парам снимков радарный снимок/оптический снимок. Генерация оптического изображения используется в случае, если реальное оптическое изображение недоступно либо страдает от сильной облачности или слабого освещения. Предложенный подход позволяет получить высокую точность сегментации снимков: на примере задачи сегментации снимков городских ландшафтов по 3 классам (фон, здание, растительность) показатель *mIoU* при сегментации по данным радарных снимков равен 0,82; при использовании оптических данных – 0,87; при использовании оптических данных в условиях симуляции слабого освещения (20% от исходного) – 0,83; при использовании сгенерированных оптических снимков – 0,84.

1.3.3.3. Реконструкция изображений

Одной из задач совместной обработки изображений ДЗЗ является реконструкция снимков или их частей, с целью устранения различных неинформативных особенностей (облачность, аномалии сенсоров). Данная задача может решаться как по данным целевого домена [114-117], так и по данным иных доменов [118-121].

Наиболее простым вариантом решения задачи реконструкции является выделение неинформативных областей на целевом изображении с последующим замещением пикселей этой области на пиксели опорных снимков того же домена или близких по свойствам к нему. Более сложный подход основывается на поиске для каждого пикселя внутри заполняемой области сходных пикселей вне заполняемой области на опорном изображении [114, 115], с последующей заменой спектральных откликов пикселей заполняемой области. При этом опорное и целевое могут относиться к разным доменам. К примеру, метод STARFM, предложенный в работе [119] может использоваться для реконструкции мультиспектральных изображений сенсора Landsat-7 ETM+ (30 м/пиксель) по снимкам сенсора MODIS (500 м/пиксель). В работе [120] рассмотрено применение радарных изображений для поиска сходных пикселей.

На текущий момент задача реконструкции решается с использованием генеративных ИНС. Для решения задачи используются сверточные ИНС [116], генеративно-состязательные сети [117], LSTM-сети [118] и, в последнее время, диффузионные модели [121-123] и трансформеры [124]. Популярной baseline-моделью реконструкции изображений является UnCRtainTS [125], предназначенная для реконструкции снимков Sentinel-2 по временным последовательностям снимков Sentinel-2 и Sentinel-1. При этом, как правило, в ходе обучения маска неинформативных областей не используется: предполагается, что генеративная ИНС способна самостоятельно выделять неинформативные области и выполнять их реконструкцию.

Стоит отметить, что задача реконструкции снимков решается на этапе предварительного обучения некоторых фундаментальных моделей [46, 48], как следствие, данные фундаментальные модели могут выполнять реконструкцию отсутствующих снимков.

1.4 Основные направления исследований в работе

1.4.1 Ограничения методов доменных преобразований на базе генеративных ИНС

Рассмотренные выше методы доменных преобразований на основе генеративных ИНС, выполняющие преобразования из между доменами реальных снимков, пытаются восстановить все особенности изображений целевого домена по данным исходного домена. В условиях существенного различия в характере и информационной емкости изображений данное преобразование не может быть выполнено полностью, поскольку некоторые детали просто не могут быть восстановлены. К примеру, по радарным снимкам невозможно достоверно восстановить цветность оптических изображений даже видимого диапазона, не говоря уже о полных мультиспектральных снимках [126]. Обратное также верно: спекл-шум, присутствующий на радарных изображениях, не может быть полностью воспроизведен по данным оптических изображений, поскольку является результатом интерференции, возникающей при приеме отраженных от земной поверхности импульсов. Впрочем, спекл-шум относится к тому классу деталей, которые вовсе не следует восстанавливать – однако, на этапе обучения генеративной ИНС нет опции выбора восстанавливаемых особенностей (хотя, разумеется, можно отфильтровать шум заранее, еще на этапе формирования обучающей выборки). Примером полезной особенности радарных снимков, которая не может быть в полной мере восстановлена по данным оптических снимков, является высокая видимость металлических конструкций.

1.4.2 Взаимная информация и гипотеза общего домена

В диссертационной работе предлагается новый подход на основе конструирования некоторого нового (общего) домена, в рамках которого

изображения исходно разных доменов обладают высоким сходством. При этом, общий домен конструируется таким образом, что в нем сохраняются только пространственные и спектральные особенности, присутствующие на изображениях всех исходных доменов, а особенности, частные для некоторого домена, отфильтровываются.

Рассмотрим два домена изображений ДЗЗ $\mathcal{X} \subseteq \mathbb{R}^{W \times H \times K_X}$ и $\mathcal{Y} \subseteq \mathbb{R}^{W \times H \times K_Y}$. Предположим, существует набор пар решающих правил $\mathbb{D} = \{(D_j^X, D_j^Y): D_j^X: \mathcal{X} \rightarrow \mathbb{R}^{W_j \times H_j \times K_j}, D_j^Y: \mathcal{Y} \rightarrow \mathbb{R}^{W_j \times H_j \times K_j}\}_{j=1}^T$, в рамках которого каждая пара преобразований (D_j^X, D_j^Y) решает некоторую прикладную задачу (сегментации, классификации, обнаружения объектов), причем для любой пары геометрически совмещенных снимков $X \in \mathcal{X}, Y \in \mathcal{Y}$, снятых в близкие моменты времени (т.е. для которых изменения на земной поверхности пренебрежимо малы) $D_j^X(X) = D_j^Y(Y)$.

Гипотеза, лежащая в основе предлагаемого в диссертации метода, основана на простом предположении: если одна и та же задача может решаться по любому из снимков обоих доменов, то на снимках содержится некоторая общая/взаимная информация, достаточная для решения задачи. Поскольку изображения X и Y (в идеальном случае) отличаются только методом их регистрации, в случае $\mathcal{X} \subseteq \mathcal{Y}$ данные изображения можно было бы считать изображениями одной формы по Ю.П. Пытьеву [127] и анализировать методами морфологической обработки, предложенными в работах Ю.П. Пытьева [127] и Ю.В. Визильтера [128, 129]. Отличие в количестве каналов не позволяет решать задачу подобным образом, однако стоит обратить внимание на понятие формы изображения по Пытьеву, как части информации, инвариантной к условиям регистрации изображения [127], что соответствует взаимной информации рассматриваемых снимков.

Хорошо известная формула из теории информации, определяющая объем/количество взаимной информации, имеет вид

$$I(X; Y) = H(X) + H(Y) - H(X, Y),$$

где $I(X; Y)$ – объем взаимной информации, $H(X)$ – энтропия X , $H(Y)$ – энтропия Y , $H(X, Y)$ – энтропия их совместного распределения. Но она характеризует исключительно объем/количество такой информации, не давая конструктивного способа ее выделения. Заметим также, что для решения конкретной практической задачи может быть достаточно только части этой взаимной информации.

Предположим, что возможно выделить такую взаимную/общую информацию со снимков исходных доменов, что эквивалентно конструированию синтетического домена изображений $(X; Y) \in \mathbb{R}^{W \times H \times K}$ – общему домену. Количество каналов K в общем домене должно быть достаточным для представления всей взаимной информации. Хотя точное значение априорно неизвестно, достаточно очевидно, что $K \leq \min(K_X, K_Y)$, поскольку объем взаимной информации не может превышать объем полной информации в любом из исходных доменов, как следствие, для её представления должно быть достаточно того же или меньшего количества каналов. С практической точки зрения может быть выгодно использовать большее число каналов, поэтому целесообразнее ограничить $K \leq K_X + K_Y$: наличие избыточных каналов не увеличит объем информации, хранимой в общем домене, но может расширить круг алгоритмов обработки, применимых к изображениям общего домена.

Если возможно построить общий домен, то каждая пара исходных решающих правил (D_j^X, D_j^Y) может быть заменена композициями преобразований $(F_X \circ D_j, F_Y \circ D_j)$, где F_X, F_Y – преобразования из исходных доменов в общий домен, D_j – решающее правило в общем домене. Кроме того, поскольку синтетические изображения $F_X(X)$ и $F_Y(Y)$ равны (или, как минимум, сходны), в рамках общего домена возможно выполнять совмещение разнородных изображений ДЗЗ, используя широкую номенклатуру методов совмещения однородных изображений. Важно отметить, что для большинства существующих многих методов совместной обработки разнородных изображений ДЗЗ предполагается, что разнородные изображения уже совмещены. Использование общего домена позволило бы часть

задач решать в рамках общего домена, а для остальных задач – выполнять необходимое предварительное совмещение изображений.

1.4.3 Применение общего домена в рамках общей технологии обработки данных ДЗЗ

Рассматриваемые в работе методы и алгоритмы обработки изображений ДЗЗ являются частью единого комплекса методов и алгоритмов, разработанного в рамках работ по гранту Министерства образования и науки РФ «Фундаментальные проблемы разработки аэрокосмических транспортных систем и управления в аэрокосмической технике для обеспечения связанности территории Российской Федерации» (соглашение № 075-15-2024-558) в Самарском университете. Схема обработки данных ДЗЗ в рамках комплекса приведена на рисунке 1.1.

1.5 Выводы и результаты первой главы

В рамках первой главы диссертации был проведен анализ текущего состояния исследований в области обработки разнородных снимков. Отмечено появление новых методов, сводящих обработку разнородных изображений к обработке однородных изображений с использованием генеративных ИНС, выполняющих преобразование изображений ДЗЗ одного из доменов (исходного домена) в другой (целевой) домен. Отмечен основной недостаток данного подхода, заключающийся в невозможности полного восстановления особенностей целевого домена по данным исходного домена.

Установлена целесообразность разработки нового метода, обладающего преимуществами перед существующими методами. Сформулирована гипотеза, определяющая способ построения искомого решения на основе построения домена взаимной информации данных ДЗЗ – общего домена.

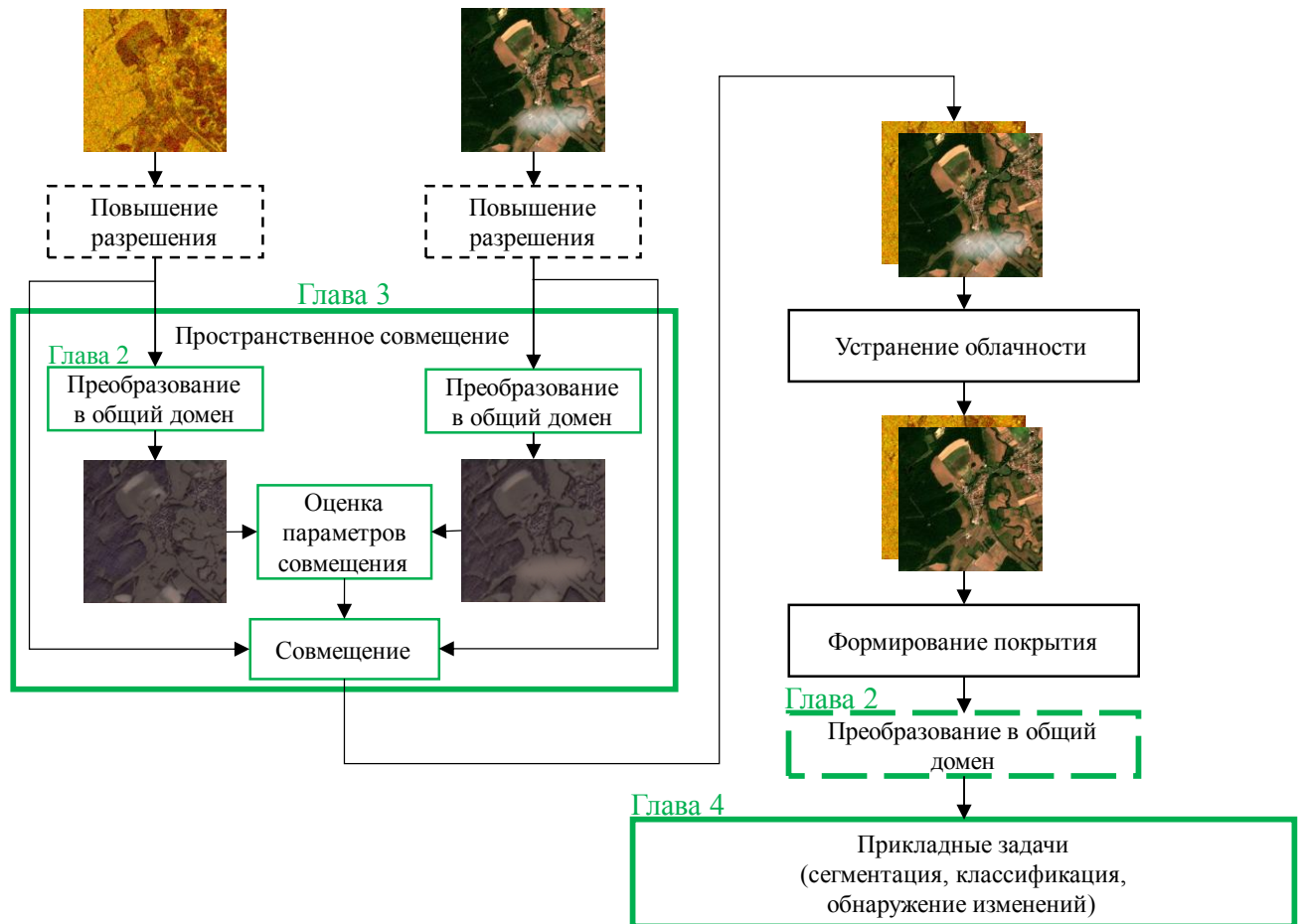


Рисунок 1.1 – Схема обработки данных ДЗЗ; этапы обработки, выполняемые методами и алгоритмами, рассмотренными в диссертации, выделены зеленым; пунктиром отмечены опциональные этапы

Глава 2. Метод построения общего домена для цифровых изображений ДЗЗ, полученных различными сенсорами

2.1 Общее описание метода

Пусть дан набор пар изображений $\mathcal{I} = \{(X^i, Y^i) \mid X^i \in \mathbb{R}^{W \times H \times K_X}, Y^i \in \mathbb{R}^{W \times H \times K_Y}\}_{i=1}^N$, содержащий пары координатно-согласованных снимков одной территории, полученных разными сенсорами с K_X и K_Y каналами соответственно и имеющих одинаковые размеры в W пикселей по горизонтали и H пикселей по вертикали. Для определенности будем считать, что $K_X \geq K_Y$.

Общим доменом является пространство изображений $\mathbb{R}^{W \times H \times K}$ с количеством каналов $K \leq K_X$. За преобразование изображений в общий домен отвечает пара преобразований $F_X: \mathbb{R}^{W \times H \times K_X} \rightarrow \mathbb{R}^{W \times H \times K}$ и $F_Y: \mathbb{R}^{W \times H \times K_Y} \rightarrow \mathbb{R}^{W \times H \times K}$. В рамках общего домена изображения должны удовлетворять следующим требованиям:

- 1) линейные размеры H и W изображений в общем домене совпадают с линейными размерами изображений в исходных доменах;
- 2) изображения общего домена для геометрически совмещенных снимков одинакового линейного размера различных доменов должны быть схожи (т.е. близки по некоторой мере);
- 3) изображения общего домена (и пара соответствующих преобразований) не являются вырожденными, то есть в них «сохраняется» значительная доля «информации» снимков исходных доменов.

Последние два требования не являются до конца определенными, более того, они в определенном смысле они противоречивы. В частности, требованию схожести изображений общего домена легко удовлетворить, построив вырожденные преобразования, которые формируют изображения в виде

константного поля. Третье требование предназначено для исключения подобных ситуаций.

Преобразования F_X и F_Y конструируются в ходе решения оптимизационной задачи

$$(F_X, F_Y) = \arg \min_{\substack{f_X \in \mathcal{F}_X, \\ f_Y \in \mathcal{F}_Y}} \frac{1}{N} \sum_{i=1}^N L_{domain}(f_X(X^i), f_Y(Y^i), X^i, Y^i), \quad (2.1)$$

где $L_{domain}: \mathbb{R}^{W \times H \times K} \times \mathbb{R}^{W \times H \times K} \times \mathbb{R}^{W \times H \times K_X} \times \mathbb{R}^{W \times H \times K_Y} \rightarrow \mathbb{R}$ – мера различия изображений, минимизируемая при переходе в общий домен, $\mathcal{F}_X$ и $\mathcal{F}_Y$ – множества функций, по которым производится поиск оптимального решения.

С точки зрения аппарата ИНС пространства $\mathcal{F}_X$ и $\mathcal{F}_Y$ задаются архитектурой нейронных сетей, а функция L_{domain} играет роль функции потерь. При этом важно отметить, что процесс обучения отличается от классического обучения с учителем, поскольку одновременно обучаются две нейросети, и эталонные значения выхода для каждой из них изначально неизвестны. Выбор $\mathcal{F}_X$, $\mathcal{F}_Y$ и L_{domain} следует производить с учетом требований к общему домену, т.е. таким образом, чтобы исключить формирование вырожденного решения (требование 3) с максимально возможным сохранением информации (требование 2).

Стоит отметить, что задача (2.1) описывает также подход на основе генеративных нейронных сетей. К примеру, обучение генеративной сети, преобразующей изображения домена Y в домен X , соответствует случаю $K = K_2$ и множеству $\mathcal{F}_Y = \{f(y) = y\}$. Однако, есть два существенных отличия предлагаемого подхода от существующих решений с преобразованием одного из изображений в альтернативный домен.

Первое, при преобразовании к существующему домену требуемый результат преобразования уже существует и может использоваться для обучения преобразования. При построении общего домена требуемый результат оказывается неизвестен. В результате и «правильный» выход генеративной сети, и итоговое преобразование снимков в общий домен определяются в результате обучения. Результат в значительной степени определяется видом критерия L_{domain} , наряду с

выбранными классами преобразований $\mathcal{F}_X$ и $\mathcal{F}_Y$ (то есть видом используемых нейронных сетей) и методами их обучения.

Вторым принципиальным отличием при использовании общего домена является отсутствие необходимости реконструировать специфические особенности изображений одного домена из изображений другого домена. Например, реконструировать спекл-шум, присущий радарным изображениям, из изображений оптического сенсора, или реконструировать цвет по данным радарного изображения. Очевидно, что исключение подобных неестественных свойств конструируемых преобразований является позитивным шагом (как покажут результаты экспериментов).

2.2 Построение общего домена в задаче спектрального согласования оптических снимков

2.2.1 Описание алгоритмов согласования

Рассмотрим применение метода общего домена в задаче спектрального согласования оптических снимков, полученных от двух оптических сенсоров со сходным пространственным разрешением, но различным числом спектральных каналов [94*, 95*].

Поскольку в данном случае свойства изображений достаточно близки, можно обойтись использованием достаточно простых ИНС (и даже обойтись вовсе без привлечения аппарата ИНС).

Для построения общего домена разнородных оптических снимков предложено 4 алгоритма: попиксельный линейный без обучения, попиксельный линейный с обучением, попиксельный нелинейный и обобщенный нелинейный.

2.2.1.1. Попиксельный линейный алгоритм без обучения

Рассмотрим базовый алгоритм построения общего домена без использования нейронных сетей.

Пусть $\mathcal{F}_X$ и $\mathcal{F}_Y$ – множества линейных функций вида

$$\begin{aligned} F_X(X) &= XA_X + b_X, \\ F_Y(X) &= YA_Y + b_Y, \end{aligned} \quad (2.2)$$

где $A_X \in \mathbb{R}^{K_X \times K}$, $A_Y \in \mathbb{R}^{K_Y \times K}$ – матрицы ранга K , $b_X, b_Y \in \mathbb{R}^K$ – вектора смещений.

Значения A_X, A_Y, b_X, b_Y можно получить без использования аппарата ИНС путем использования комбинации метода главных компонент [9] и метода наименьших квадратов.

Более подробно, пусть $P_X = \mathbb{R}^{NHW \times K_X}$ и $P_Y = \mathbb{R}^{NHW \times K_Y}$ – матрицы, составленные из значений пикселей всех изображений из наборов $\mathcal{X} = \{X_i\}_{i=1}^N$ и $\mathcal{Y} = \{Y_i\}_{i=1}^N$ соответственно. Поскольку необходимо обеспечить максимальное согласование значений пикселей в изображениях, задачу (2.1) можно переформулировать в виде

$$\|(P_X A_2 - P_Y) A_1\| \rightarrow \min_{A_1, A_2},$$

где $A_1 \in \mathbb{R}^{K_Y \times K}$, $A_2 \in \mathbb{R}^{K_X \times K_Y}$.

Матрица A_2 , минимизирующая расхождение между парами изображений, может быть найдена путем применения метода наименьших квадратов к матрицам P_X и P_Y .

Матрица A_1 понижающая размерность пикселей до целевой, может быть найдена с помощью применения метода главных компонент к матрице $P = P_X A_2 - P_Y$. Результатом работы метода главных компонент является набор $\{\lambda_i, v_i\}_{i=1}^{K_Y}$, где λ_i – собственное число, определяющее энергию компоненты, $v_i \in \mathbb{R}^{K_2}$ – собственный вектор, задающий проекцию вектора значений пикселя на компоненту.

Обозначим $C_i^X = P_X A_2 v_i$ – набор значений i -той компоненты на наборе $\mathcal{X}$; $C_i^Y = P_Y v_i$ – набор значений i -той компоненты на наборе $\mathcal{Y}$ соответственно. Поскольку в данной задаче необходимо обеспечить наименьшее расхождение

$$\sum_{j \in J} \|C_j^X - C_j^Y\| \rightarrow \min_J,$$

где $J = \{j_k \mid j_k \in \{1, \dots, K_Y\}\}_{k=1}^K$ – множество индексов отобранных компонент, логичным будет выбрать компоненты с меньшими значениями собственных чисел. При этом следует учитывать, что малое значение собственного числа может означать не только высокое сходство значений компоненты на изображениях, но и малую энергию компонент C_i^X и C_i^Y . Как следствие, при выборе компонент нужно учитывать и значение собственного числа, и энергию компоненты. Для этого будем производить отбор, исходя из значений $\tilde{\lambda}_i$:

$$\tilde{\lambda}_i = \lambda_i - \xi \sqrt{S^2[C_i^X] S^2[C_i^Y]},$$

где $S^2[C_i^X]$ и $S^2[C_i^Y]$ – выборочные дисперсии соответствующих компонент, ξ – параметр, определяющий влияние энергии компонент. В работе [94*] предлагается использовать значение $\xi = \frac{1}{K_Y} \sum_{j=1}^{K_Y} \frac{\lambda_i}{\sqrt{S^2[C_j^X] S^2[C_j^Y]}}$, что соответствует выбору

компонент, для которых соотношение собственного числа компоненты к энергии компоненты на снимках выше среднего.

Итоговая матрица $A_1 = (\tilde{v}_i)_{i=1}^K$ составляется из K собственных векторов с наименьшими значениями $\tilde{\lambda}_i$, отсортированных в порядке убывания $\tilde{\lambda}_i$.

Элементы (2.2) выражаются через A_1 и A_2 следующим образом:

$$\begin{aligned} A_X &= A_2 A_1, \\ A_Y &= A_1, \\ b_X &= 0, \\ b_Y &= \frac{1}{NHW} \sum_{i=1}^{NHW} (p_i^X A_1 - p_i^Y) A_2, \end{aligned}$$

где p_i^X и p_i^Y – строки матриц P_X и P_Y .

2.2.1.2. Алгоритмы на основе ИНС

В случае, если F_X и F_Y имеют вид (2.2), соответствующие им ИНС будут состоять из единственного линейного слоя с K выходами и K_X или K_Y входами соответственно и тривиальной активационной функцией. Значения элементов матриц A_X , A_Y и весовых коэффициентов b_X , b_Y определяются в процессе обучения полученной пары ИНС. Такой алгоритм построения F_X и F_Y называется далее попиксельным линейным алгоритмом с обучением.

Если $\mathcal{F}_X$ и $\mathcal{F}_Y$ являются множествами функций произвольного вида, выполняющих попиксельное преобразование изображения, для моделирования F_X и F_Y можно использовать многослойные персептроны с K_X и K_Y входами и K выходами. Данный алгоритм далее обозначен, как попиксельный нелинейный алгоритм.

Наконец, если $\mathcal{F}_X$ и $\mathcal{F}_Y$ являются множествами функций произвольного вида, F_X и F_Y могут быть смоделированы ИНС, которые для определения значений пикселя на выходе сети учитывают информацию из некоторой окрестности пикселя (в предельном случае – всего изображения). В данном случае в качестве архитектуры ИНС могут использоваться как сверточные ИНС, так и популярные в настоящее время трансформеры. В работах [94*, 95*] было показано, что достаточно хороший результат можно получить только с использованием простых сверточных ИНС. Данный алгоритм далее обозначен, как обобщенный нелинейный алгоритм.

Как было отмечено в подразделе 2.1, обучение пары ИНС, реализующих F_X и F_Y , осложняется тем фактом, что «истинные» выходы априорно неизвестны. Это создает особую сложность в построении критерия L_{domain} , минимизируемого в ходе обучения. Наиболее простым вариантом является использование квадрата ошибки:

$$MSE(V^X, V^Y) = \frac{1}{WHK} \sum_{i=0}^{W-1} \sum_{j=0}^{H-1} \sum_{k=0}^{K-1} (X_{ijk} - Y_{ijk})^2. \quad (2.3)$$

Однако, в случае $L_{domain}(V^X, V^Y, X, Y) = MSE(V^X, V^Y)$ оптимальным решением является вырожденное решение $F_X \equiv C$, $F_Y \equiv C$, что противоречит третьему требованию построения общего домена.

В работах [94*, 95*] предлагается применить нормализацию значений на выходе нейронных сетей:

$$Norm(V) = \frac{V - \mathbb{E}[V]}{S[V] + \varepsilon}, \quad (2.4)$$

где $V \in R^*$ – вектор нормализуемых значений (произвольной длины), $\mathbb{E}[V]$ – среднее значение V , $S[V]$ – среднеквадратичное отклонение V , $\varepsilon = 10^{-5}$ – регуляризационная константа. В случае нормализации тензоров с количеством измерений более 1 тензор разворачивается в вектор, нормализуется и затем сворачивается обратно в тензор с сохранением исходного порядка элементов.

После нормализации среднее значение данных близко к 0, а дисперсия близка к 1 (за исключением массивов, заполненных 0 или близкими к 0 значениями). Добавление нормализации на выходе модели предотвращает обучение от сведения выходных значений к 0: без нормализации уменьшение выборочной дисперсии ведет к уменьшению значения функции потерь, а при использовании нормализации дисперсия остается близкой к 1 – больше нет прямой зависимости между дисперсией результата и значением функции потерь.

В случае многоканальных изображений нормализация может применяться ко всему изображению, либо к каждому каналу по отдельности, последнее известно, как Instance Normalization [130]. Преимуществом такой нормализации является равенство средних значений и выборочной дисперсии по всем каналам (при нормализации всего изображения некоторые каналы могут зануляться). В рамках данной работы на выходе всех сетей располагается слой Instance Normalization.

Есть ещё одна проблема, наиболее актуальная для обобщённого нелинейного алгоритма. Если L_{domain} – функция, измеряющая среднее попиксельное сходство

изображений (например, среднеквадратическая ошибка или средняя абсолютная ошибка), то отфильтровывание мелких деталей из изображений увеличит сходство изображений, тем самым уменьшая значение функции потерь; поэтому оптимизация относительно L_{domain} будет приводить к излишней потере деталей. Чтобы предотвратить это, функция потерь L_{domain} должна включать член, измеряющий потерю детализации.

В работе [131], посвященной исследованию различных функций потерь в задаче реконструкции изображений с помощью ИНС, было показано, что включение индекса структурного сходства ($SSIM$, structural similarity index measure) [132] позволяет получать лучшие результаты по сравнению с использованием только MSE . Показатель $SSIM$ вычисляется по формуле:

$$SSIM(A, B) = \frac{(2\mathbb{E}[A]\mathbb{E}[B] + c_1)(2S[AB] + c_2)}{(\mathbb{E}^2[A] + \mathbb{E}^2[B] + c_1)(S^2[A] + S^2[B] + c_2)}, \quad (2.5)$$

где $A, B \in \mathbb{R}^{w \times w}$ – полутоновые изображения; $\mathbb{E}[A]$, $\mathbb{E}[B]$ – средние значения яркости изображений; $S^2[A]$, $S^2[B]$ – выборочные дисперсии изображений, $S[AB]$ – ковариация изображений; c_1 , c_2 – константы для стабилизации в области малых значений (в рамках данной работы $c_1 = 10^{-4}$, $c_2 = 9 \times 10^{-4}$). Значения индекса $SSIM$ лежат в отрезке $[-1, 0; 1, 0]$, где значение 1,0 соответствует полному совпадению изображений.

Поскольку индекс $SSIM$ определяется для полутонового изображения размером w , в качестве значения показателя для изображений произвольного размера принимается усредненный показатель по всем соответствующим фрагментам изображений размером $w \times w$. В рамках настоящей работы при обучении ИНС используется размер $w = 5$ для лучшего сохранения малых деталей на снимках (в исходной работе [132] в качестве размера окна используется $w = 11$). В случае многоканальных изображений значение показателя считается для каждого канала по отдельности, а затем усредняется.

Итоговый вид L_{domain} , используемый при решении задачи спектрального согласований:

$$g(A) = \frac{1}{C_A} \sum_{k=0}^{C_A-1} A^k, \quad (2.6)$$

$$DiSSIM(A, B) = 1 - SSIM(A, B), \quad (2.7)$$

$$\begin{aligned} DetLoss(V, X, Y) &= DiSSIM(g(V), g(Ref(X, Y))), \\ L_{domain}(V^X, V^Y, X, Y) &= MSE(V^X, V^Y) + \alpha DetLoss(V^X, X, Y) + \\ &+ \beta DetLoss(V^Y, X, Y), \end{aligned} \quad (2.8)$$

где $X \in \mathbb{R}^{W \times H \times K_X}$, $Y \in \mathbb{R}^{W \times H \times K_Y}$ – пара изображений в исходных доменах, $V^X, V^Y \in \mathbb{R}^{W \times H \times K}$ – изображения в общем домене, $g: \mathbb{R}^{W \times H \times * } \rightarrow \mathbb{R}^{W \times H \times 1}$ – функция поканального усреднения изображения, $Ref(X, Y)$ – функция конструирования опорного изображения из исходных изображений (в простом варианте, выбирающая каналы с высоким пространственным разрешением), $\alpha, \beta \in \mathbb{R}$ – параметры, определяющие влияние потери детализации на значение функции потерь.

2.2.2 Экспериментальные исследования

2.2.2.1. Набор данных

Исследование алгоритмов проводилось на парах снимков космических аппаратов Sentinel-2 и Landsat-8 из набора данных IEEE GRSS Data Fusion Contest 2021 [133]. Размеры снимков – 800×800 пикселей, снимки приведены к одному пространственному разрешению 15 м/пиксель. Количество каналов для снимков Sentinel-2 – 12 (исключен канал 10), количество каналов для снимков Landsat-8 – 11.

Следует отметить, что для пар изображений Sentinel-2 и Landsat-8 был обнаружен систематический сдвиг на 1 пиксель по вертикали и на 2 пикселя по горизонтали. Кроме того, на краях некоторых изображений наблюдались артефакты. Для их устранения, после коррекции сдвига все изображения были обрезаны до равномерного размера 792×792 .

Из оригинального набора были удалены снимки, имеющие значительную облачность, после чего для каждого снимка от одного сенсора был выбран ближайший по дате снимок от другого сенсора (для минимизации возможных отличий за счет естественных временных изменений). Итоговый набор данных был случайным образом разделен на обучающую, валидационную и тестовую выборки: обучающая выборка состоит 70 пар изображений, валидационная – из 24 пар, тестовая – из 30 пар.

2.2.2.2. Показатели качества

Следующие функции используются, как показатели качества согласования:

1) Среднеквадратическое отклонение:

$$RMSE(A, B) = \sqrt{MSE(A, B)}; \quad (2.9)$$

2) нормализованная взаимная информация [134]:

$$NMI(A, B) = \frac{H(A) + H(B)}{H(A, B)}, \quad (2.10)$$

где $H(\dots)$ – энтропия;

3) *SSIM* (2.5).

Т.к. нормализация гарантирует только нулевое среднее и дисперсию, близкую к единичной, но не ограничивает диапазон значений, необходима предобработка изображений. Для устранения выбросов, имеющих на обоих концах диапазона значений, изображение приводится к диапазону $[0; 1]$ по формуле

$$\mu(V) = clamp\left(\frac{V - p_{0,5}(V)}{p_{99,5}(V) - p_{0,5}(V)}\right), \quad (2.11)$$

где $p_{0,5}, p_{99,5}$ – функции, возвращающие 0,5- и 99,5-перцентили изображения, *clamp* – функция, ограничивающая диапазон значений аргумента до $[0; 1]$ (все значения < 0 приравниваются к 0, все значения больше 1 приравниваются к 1).

Значения показателей качества рассчитываются по преобразованным изображениям.

2.2.2.3. Обучение нейросетей

Все эксперименты проводились на компьютере с процессором AMD Ryzen 5600X, видеокартой NVIDIA GeForce RTX 3090 Ti и 32 ГБ оперативной памяти. Все модели были реализованы и обучены с использованием библиотеки PyTorch 2.3.

Для нелинейных алгоритмов в качестве функции активации использовалась LeakyReLU [135] с наклоном отрицательной части минус 0,25.

L_{domain} определена согласно (2.8) с $\beta = 0,125$. Для каждой пары входных изображений в качестве опорного изображения $Ref(X, Y)$ использовался панхроматический канал снимка Landsat-8 (исходное пространственное разрешение 10 м/пиксель).

В пределах эпохи на этапе обучения использовались 256 пар фрагментов размером 224×224 , случайным образом извлекаемых из изображений обучающей выборки. В качестве аугментаций использовались случайные повороты на 90° и горизонтальные/вертикальные отражения. В качестве оптимизатора использовался Adam [133] с темпом обучения 10^{-5} .

Этап валидации эпохи проводился на полных изображениях без аугментаций. Использовались все изображения из валидационной выборки.

Количество эпох было ограничено 128, с досрочной остановкой, если улучшение не наблюдалось более 8 эпох.

Для всех рассматриваемых алгоритмов количество каналов в общем домене равно 3.

2.2.2.4. Результаты экспериментов

Результаты попиксельного линейного алгоритма без обучения

Средние значения показателей сходства изображений тестовой выборки в общем домене, сконструированным попиксельным линейным алгоритмом без обучения приведены в таблице 2.1.

Таблица 2.1 – Результаты попиксельного линейного алгоритма без обучения

Показатель качества	$RMSE$	NMI	$SSIM$
Значение	0,095	1,093	0,69

Результаты попиксельного линейного алгоритма с обучением

Попиксельный линейный алгоритм с обучением реализован в виде пары линейных слоев с обучаемыми матрицами A_X и A_Y векторами смещений b_X , b_Y . Средние значения показателей сходства изображений тестовой выборки в общем домене, сконструированным рассматриваемым алгоритмом, приведены в таблице 2.2.

Таблица 2.2 – Результаты попиксельного линейного алгоритма с обучением

Показатель качества	$RMSE$	NMI	$SSIM$
Значение	0,075	1,112	0,78

Результаты попиксельного нелинейного алгоритма

Попиксельный нелинейный алгоритм реализован в виде многослойного персептрона. Поскольку существует множество вариантов структуры нейросети, отличающихся количеством скрытых слоев и размерностью выхода каждого слоя, необходимо рассмотреть несколько вариантов конфигураций сети. Были рассмотрены следующие варианты:

- 1) 2 слоя с выходными размерами 8,3;
- 2) 3 слоя с выходными размерами 16, 8, 3;
- 3) 4 слоя с выходными размерами 32, 16, 8, 3;
- 4) 5 слоев с выходными размерами 63, 32, 16, 8, 3.

Средние значения показателей сходства изображений тестовой выборки в общем домене, сконструированным рассматриваемым алгоритмом при различных конфигурациях ИНС, приведены в таблице 2.3.

Таблица 2.3 – Результаты попиксельного нелинейного алгоритма при различных конфигурациях сети

Конфигурация	Показатели качества		
	<i>RMSE</i>	<i>NMI</i>	<i>SSIM</i>
1	0,077	1,111	0,781
2	0,075	1,115	0,791
3	0,074	1,117	0,794
4	0,075	1,117	0,788

Как видно из результатов, наилучшей конфигурацией является конфигурация с 4 слоями и выходными размерами 32, 16, 8 и 3. Конфигурация с 5 слоями имеет более высокое значение *SSIM*, несколько более низкие значения двух других показателей (значения *NMI* до округления 1,1173 и 1,1169), но также имеет больше параметров и, следовательно, менее эффективна.

Результаты обобщенного нелинейного алгоритма

Обобщённый нелинейный алгоритм реализован в виде сверточной нейронной сети. Как и в случае с попиксельным нелинейным алгоритмом, существует несколько вариантов структуры ИНС, отличающихся по количеству слоев, размерностью выходных данных и размером обучаемых ядер фильтров. Количество слоёв и размер выходных данных слоёв взяты из наилучшего варианта попиксельного нелинейного алгоритма (4 слоя с размерами выходных данных 32, 16, 8, 3). Были рассмотрены следующие варианты размеров ядер свертки:

- 1) 4 слоя с ядрами 5,5,3,3;
- 2) 4 слоя с ядрами 5,5,5,5;
- 3) 4 слоя с ядрами 7,7,5,5;
- 4) 4 слоя с ядрами 7,7,7,7.

Средние значения показателей сходства изображений тестовой выборки в общем домене, сконструированным рассматриваемым алгоритмом при различных конфигурациях ИНС приведены в таблице 2.4.

Таблица 2.4 – Результаты обобщенного нелинейного алгоритма при различных конфигурациях сети

Конфигурация	Показатели качества		
	<i>RMSE</i>	<i>NMI</i>	<i>SSIM</i>
1	0,055	1,184	0,844
2	0,051	1,187	0,857
3	0,048	1,194	0,872
4	0,050	1,187	0,858

Как видно из результатов, наилучшей конфигурацией сверточной сети является 4 слоя с выходными размерами 32, 16, 8, 3 и размерами ядер 7,7,5,5.

Итоговые результаты

Результаты наилучших вариантов рассмотренных алгоритмов представлены в таблице 2.5.

Таблица 2.5 – Сводные результаты алгоритмов построения общего домена разнородных оптических изображений

Алгоритм построения общего домена	Показатели качества		
	<i>RMSE</i>	<i>NMI</i>	<i>SSIM</i>
-	0,202	1,068	0,38
Попиксельный линейный без обучения	0,095	1,093	0,69
Попиксельный линейный с обучением	0,075	1,112	0,78
Попиксельный нелинейный	0,074	1,117	0,79
Обобщенный нелинейный	0,048	1,194	0,87

Первая строка содержит средние значения показателей сходства изображений в исходных доменах (использованы RGB-каналы снимков Sentinel-2 и 2, 3, 4, 5, 6, 7 снимков Landsat-8 с приведением к диапазону значений $[0;1]$ согласно (2.11)). Как видно, обобщённый нелинейный алгоритм превосходит попиксельные алгоритмы.

Примеры снимков и результатов их преобразования к общему домену приведены на рисунке 2.1.

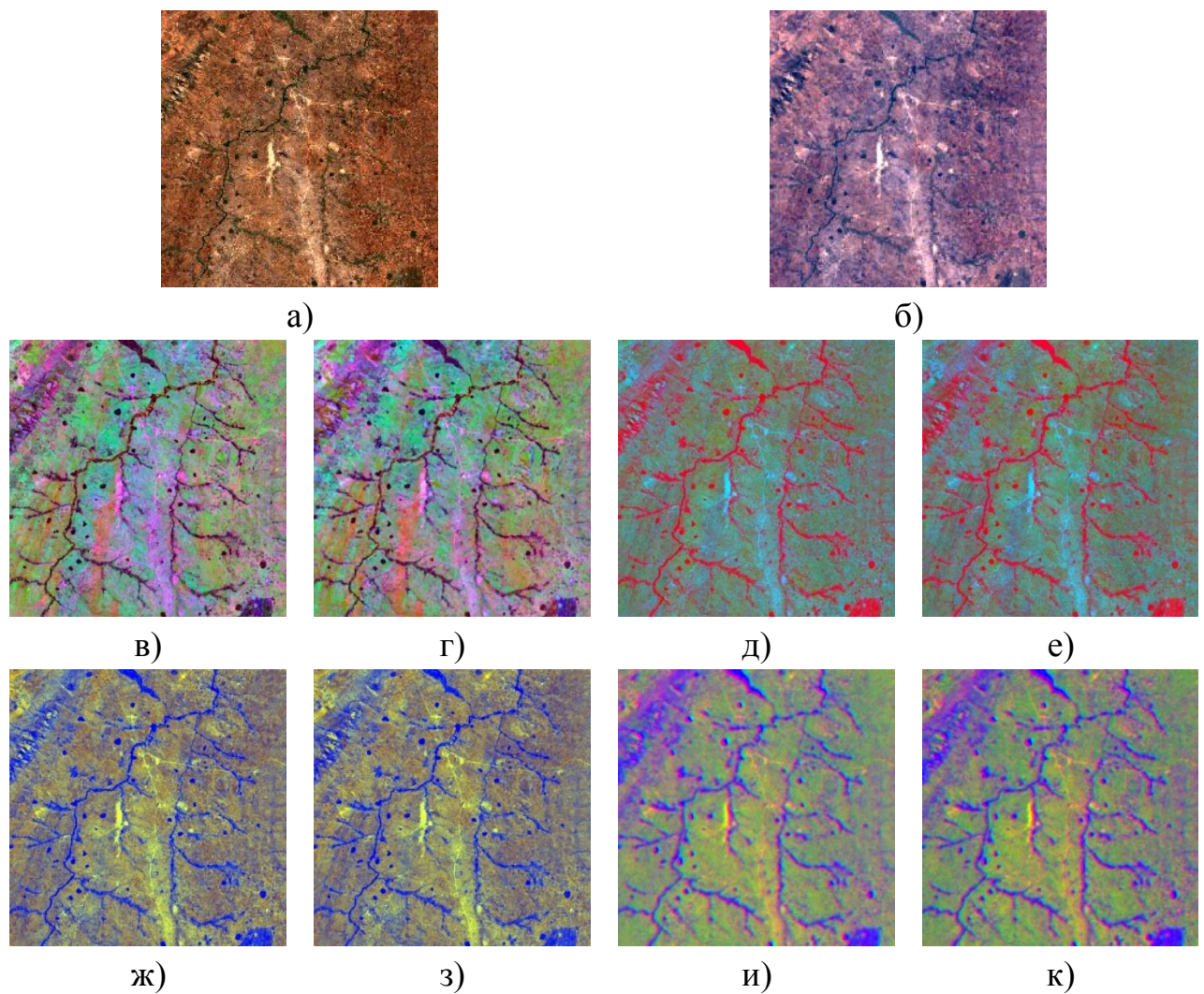


Рисунок 2.1 – Космоснимки и результаты их преобразования в общий домен; а, б – исходные снимки; в, г – результат попиксельного линейного алгоритма без обучения; д, е – результат попиксельного линейного алгоритма с обучением; ж, з – результат попиксельного нелинейного алгоритма; и, к – результат обобщенного нелинейного алгоритма

2.3 Построение общего домена для оптических и радарных снимков

2.3.1 Описание алгоритма

Как было показано в предыдущем разделе, в задаче спектрального согласования оптических снимков достаточно хорошие результаты можно получить, даже ограничившись линейными преобразованиями при конструировании общего домена за счет достаточно высокого изначального сходства между доменами.

В случае, если исходные домены существенным образом отличаются, необходимо использовать более сложные преобразования (и, что эквивалентно, более сложные архитектуры ИНС) [137*]. Возникает вопрос адаптации L_{domain} под новую, более сложную задачу.

В качестве примера задачи построения общего домена для существенно различных доменов рассмотрим задачу построения общего домена для оптических и радарных снимков.

Учитывая требования 2 и 3 к общему домену (см. раздел 2.1), представим функцию L_{domain} в следующем виде:

$$L_{domain}(V^X, V^Y, X, Y) = L_{sim}(V^X, V^Y) + \alpha L_{det}(V^X, X, V^Y, Y), \quad (2.12)$$

где $L_{sim}: \mathbb{R}^{W \times H \times K} \times \mathbb{R}^{W \times H \times K} \rightarrow \mathbb{R}$ – функция, характеризующая меру схожести двух изображений в общем домене; $L_{det}: \mathbb{R}^{W \times H \times K} \times \mathbb{R}^{W \times H \times K_X} \times \mathbb{R}^{W \times H \times K} \times \mathbb{R}^{W \times H \times K_Y} \rightarrow \mathbb{R}$ – функция, характеризующая вырожденность изображений V^X, V^Y общего домена относительно изображений X, Y исходных доменов; $\alpha \in \mathbb{R}_+$ – весовой коэффициент.

Требование 3 «невыврожденности» изображений общего домена означает, что изображения не вырождаются в (почти) постоянные области (при условии, что исходные изображения не являются таковыми). С другой стороны, требование сходства изображений в общем домене означает, что такие изображения должны иметь общий набор пространственных особенностей, а средняя яркость должна

быть одинаковой. Следовательно, при построении функции L необходимо учитывать как различия в яркости изображений, так и сохранение пространственных особенностей.

Как и при решении задачи спектрального согласования, для оценки сходства изображений в общем домене будем использовать комбинацию MSE и $SSIM$:

$$L_{sim}(V^X, V^Y) = \beta MSE(V^X, V^Y) + \gamma DiSSIM(V^X, V^Y), \quad (2.13)$$

где $DiSSIM$ – функция, определенная согласно (2.7); $\alpha, \beta \in \mathbb{R}^+$ – весовые коэффициенты. Как и при решении задачи спектрального согласования, при вычислении $SSIM$ используется размер окна $w = 5$.

При определении вида функции L_{det} , характеризующей вырожденность изображений V^X, V^Y общего домена относительно исходных изображений X, Y , рассмотрено два варианта, существенно различающихся и архитектурой, и вычислительными затратами.

Первый вариант соответствует конкретизации требования 3 к общему домена в виде возможности достаточно качественной реконструкции снимков исходных доменов по изображениям общего домена. В этом случае задача (2.1) может рассматриваться, как обучение пары автоэнкодеров [138], при этом функции f_X и f_Y будут соответствовать энкодерам. В этом случае L_{det} можно определить, как сумму двух функций потерь:

$$l(A, B) = MSE(A, B) + DiSSIM(A, B),$$

$$L_{det}(V^X, X, V^Y, Y) = l(X, f_X^{-1}(V^X)) + \eta l(Y, f_Y^{-1}(V^Y)), \quad (2.14)$$

где $f_X^{-1}: \mathbb{R}^{W \times H \times K} \rightarrow \mathbb{R}^{W \times H \times K_X}$ и $f_Y^{-1}: \mathbb{R}^{W \times H \times K} \rightarrow \mathbb{R}^{W \times H \times K_Y}$ – декодеры, $\eta \in \mathbb{R}_+$ – весовой коэффициент, определяющей вклад значений от каждого из доменов.

Учитывая, что процесс реконструкции изображений исходных доменов по изображениям общего домена архитектурно и вычислительно сложен, мы также рассматриваем и более простой (второй) вариант.

Второй вариант функции L_{det} основан на измерении сходства полутоновых изображений для обеспечения невырожденности преобразований. Для радарного

изображения, соответствующее ему полутонное изображение извлекается из его образа в общем домена. Для оптического изображения соответствующее ему полутонное изображение получается из RGB-каналов (или других каналов с наивысшим разрешением). В итоге, вторая версия показателя L_{det} выглядит следующим образом:

$$L_{det}(V^X, X, V^Y, Y) = DiSSIM\left(g(V^Y), g(Ref(X))\right), \quad (2.15)$$

где $g: \mathbb{R}^{W \times H \times \cdot} \rightarrow \mathbb{R}^{W \times H \times 1}$ – функция, усредняющая все каналы изображения, Ref – функция выделения каналов высокого разрешения из оптического изображения.

Вторая версия показателя L_2 устраняет необходимость построения обратных преобразований для обеих исходных доменов и, как правило, требует меньшего количества обученных нейронных сетей.

Стоит отметить, что изначально L_{det} рассматривалась в следующем варианте:

$$L_{det}(V^X, X, V^Y, Y) = \lambda_1 DiSSIM\left(g(V^X), g(Ref(X))\right) + \\ + \lambda_2 DiSSIM\left(g(V^Y), g(Ref(X))\right),$$

однако при значениях $\lambda_1 > 0$ показатели качества ухудшались. Данный факт можно объяснить, исходя из следующих соображений:

1) если $\lambda_1 = 0$, значение (2.12) зависит только от сходства изображений в общем домене и от структурного сходства радарного изображения в общем домене и оптического изображения. Можно сказать, что f_Y обучается восстановлению особенностей оптического изображения по радарному изображению. Поскольку полное восстановление невозможно и часть деталей все же теряется, f_X обучается упрощению оптического изображения с потерей тех деталей, которые не могут быть реконструированы f_Y ;

2) если $\lambda_1 > 0$, потеря любых деталей будет увеличивать значение (2.12), как следствие, f_X будет отбрасывать меньше деталей, что увеличит расхождение между f_X и f_Y .

В результате мы имеем следующее формальное описание преобразований в предлагаемом методе:

1) X и Y – совмещенные оптическое и радарное изображение (снимки/изображения исходных доменов);

2) $V^X = f_X(x)$ и $V^Y = f_Y(Y)$ – изображения общего домена;

3) $\hat{X} = f_X^{-1}(V^X)$ и $\hat{Y} = f_Y^{-1}(V^Y)$ – реконструированные изображения в исходных доменах.

Два рассматриваемых пайплайна обучения ИНС изображены на рисунке 2.2 (обучаемые ИНС обозначены цветом).

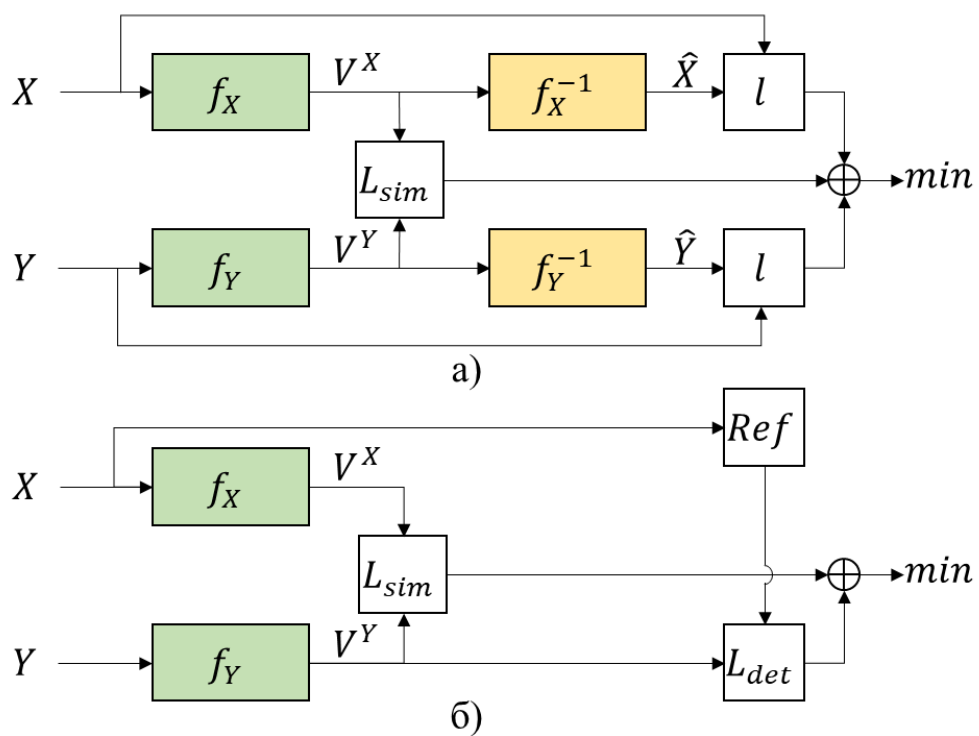


Рисунок 2.2 – Пайплайны обучения ИНС преобразования; а – пайплайн обучения на основе (2.14); б – пайплайн обучения на основе (2.15)

2.3.2 Экспериментальные исследования

2.3.2.1. Набор данных

В качестве данных – пар оптических и радарных изображений – для обучения и тестирования использовался набор данных MultiSenGE [139]. Поскольку

MultiSenGE для каждой области съемки содержит несколько радарных и оптических изображений, полученных в разное время, формирование обучающих пар выполнялось следующим образом: для каждого оптического изображения выбиралось радарное изображение для той же территории с ближайшей датой съемки. Полученный набор данных был разделен на обучающую, тестовую и валидационную выборки. Обучающая выборка состояла из 46 092 пар, валидационная – из 11 522 пар, тестовая – из 14 405 пар снимков. Значения отсчетов радарных изображений были переведены в логарифмическую шкалу.

Поскольку существующие программные решения для обучения генеративных сетей на момент проведения экспериментов поддерживали работу исключительно с RGB- или полутоновыми изображениями, был создан вспомогательный набор стандартных RGB-изображений, полученный следующим образом:

1) радарные снимки были переведены в изображения в трехканальные изображения, путем помещения VV-поляризации в красном канале и VH-поляризации в зеленом канале, с последующим приведением к логарифмической шкале;

2) оптические (мультиспектральные) изображения были преобразованы в цветные изображения путем отбора RGB-каналов.

Полученные изображения приводились к диапазону значений $[0; 1]$ согласно (2.11).

2.3.2.2. Обучение нейросетей

Для построения преобразований $\tilde{x} = f_X(x)$ и $\tilde{y} = f_Y(y)$ изображений исходных доменов в общий домен использовалась сеть U-Net [137] с 5 уровнями. Сеть Efficient-Net-B0 [141] использовалась как кодирующая часть с целью улучшения формируемых признаков описаний. В качестве активации последнего слоя сети использовалась сигмоида:

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

Использование сигмоиды позволяет гарантировать, что диапазон значений отсчетов снимков в общем домене вложен в интервал (0; 1).

Для построения реконструирующих преобразований $\hat{X} = f_X^{-1}(V^X)$ и $\hat{Y} = f_Y^{-1}(V^Y)$ изображений из общего в исходные домены использовалась аналогичная сеть U-Net с обратным числом входных и выходных каналов, но без использования сигмоиды на выходе сети.

В качестве альтернативных методов междоменного преобразования используется преобразование радарных снимков в оптический домен генеративными ИНС pix2pix [99] и LSGAN [142] (в этом случае только преобразование F_Y реализуется в виде ИНС, преобразование F_X является тривиальным).

Эксперименты проводились на машине с AMD Ryzen 5600X, 32 GB ОЗУ и Nvidia GeForce RTX 3090.

В качестве фреймворка для машинного обучения использовался PyTorch 2.7. Реализация U-Net взята из пакета segmentation-models-pytorch [143].

Для всех вариантов преобразования в общий домен итоговое число каналов $K = 3$. Другие параметры указаны ниже.

Общие параметры:

- $K_X = 10, K_Y = 2$;
- метод оптимизации: NAdam [144];
- базовый темп обучения: 2×10^{-4} ;
- метод изменения темпа обучения: OneCycle [145].
- количество эпох: 512;
- batch size: 32;
- аугментации: случайные отражения и повороты на 90° .

Параметры L_{domain} для обучения преобразований с критерием (2.14):

- $\alpha = 16$;

- $\beta = 1$;

- $\gamma = 3$;

- $\eta = 1$;

Параметры L_{domain} обучения преобразований с критерием (2.15):

- $\alpha = 16$;

- $\beta = 1$;

Для обучения pix2pix и LSGAN использовался код из репозитория [146]. Обучение pix2pix осуществлялось со стандартными параметрами без изменения кода. Обучение LSGAN проводилось с отключенной identity loss и с применением различных темпов обучения дискриминаторов и генераторов (10^{-3} и 2×10^{-4} соответственно) согласно Two Time-Scale Update Rule (TTUR) [147]. Без применения TTUR модель LSGAN сходилась к неудовлетворительным результатам.

В качестве набора данных для обучения генеративных ИНС использовался набор цветных изображений, получение которого описано в подразделе 532.3.2.1.

2.3.2.3. Показатели качества

Для оценки результатов преобразования к общему домену используются следующие показатели качества:

- $SSIM$ (2.5);

- $RMSE$ (2.9);

- NMI

- $PSNR$ (пиковое соотношение сигнал шум).

В общем случае показатель $PSNR$ определяется следующим образом:

$$PSNR(A, B) = 10 \lg \frac{MAX^2}{MSE(A, B)},$$

где MAX – максимальное по модулю значение, которое может принять отсчет изображения.

С учетом того, что и для изображений общего домена (за счет использования сигмоиды на выходе ИНС преобразования радарных/оптических изображений или процедуры нормализации для преобразования пар оптических изображений) и для RGB-изображений вспомогательного набора (за счет применения (2.11)) $MAX = 1$, выражение для вычисления $PSNR$ можно упростить до

$$PSNR(A, B) = -10 \lg MSE(A, B). \quad (2.16)$$

2.3.2.4. Результаты экспериментов

Средние значения показателей качества представлены в таблице 2.6

Обозначение «Вариант-1» соответствует обучению ИНС с критерием невырожденности (2.14), обозначение «Вариант-2» соответствует обучению ИНС с критерием (2.15).

Таблица 2.6 – Значения показателей сходства изображений радарных и оптических изображений в общем домене

Метод преобразования	$RMSE$	$PSNR$	NMI	$SSIM$
-	0,36	8,91	1,040	0,03
pix2pix	0,10	20,38	1,069	0,49
LSGAN	0,28	1,026	1,026	0,21
Вариант-1	0,13	18,66	1,078	0,53
Вариант-2	0,06	25,05	1,120	0,75

Стоит отметить, что результаты ИНС, обученных с использованием критерия (2.14), имеют низкий динамический диапазон (сигмоида гарантирует, что диапазон значений вложен в (0; 1), но не гарантирует ненулевую дисперсию значений). Как следствие, перед расчетом показателей сходства для получаемых изображений применялась нормализация по (2.11).

Результаты ИНС, обученных с использованием критерия (2.15), имеют достаточный динамический диапазон и не требуют нормализации, что является очевидным преимуществом.

Из рассматриваемой таблицы видно, что сходство между оптическими и радарными изображениями в общем домене значительно выше, чем между оптическими и псевдооптическими изображениями, сгенерированными из радарных изображений генеративными ИНС. При этом из рассмотренных вариантов лучшие результаты были достигнуты ИНС, обученными с использованием критерия (2.15).

Примеры результатов доменных преобразований изображений приведены на рисунке 2.3.

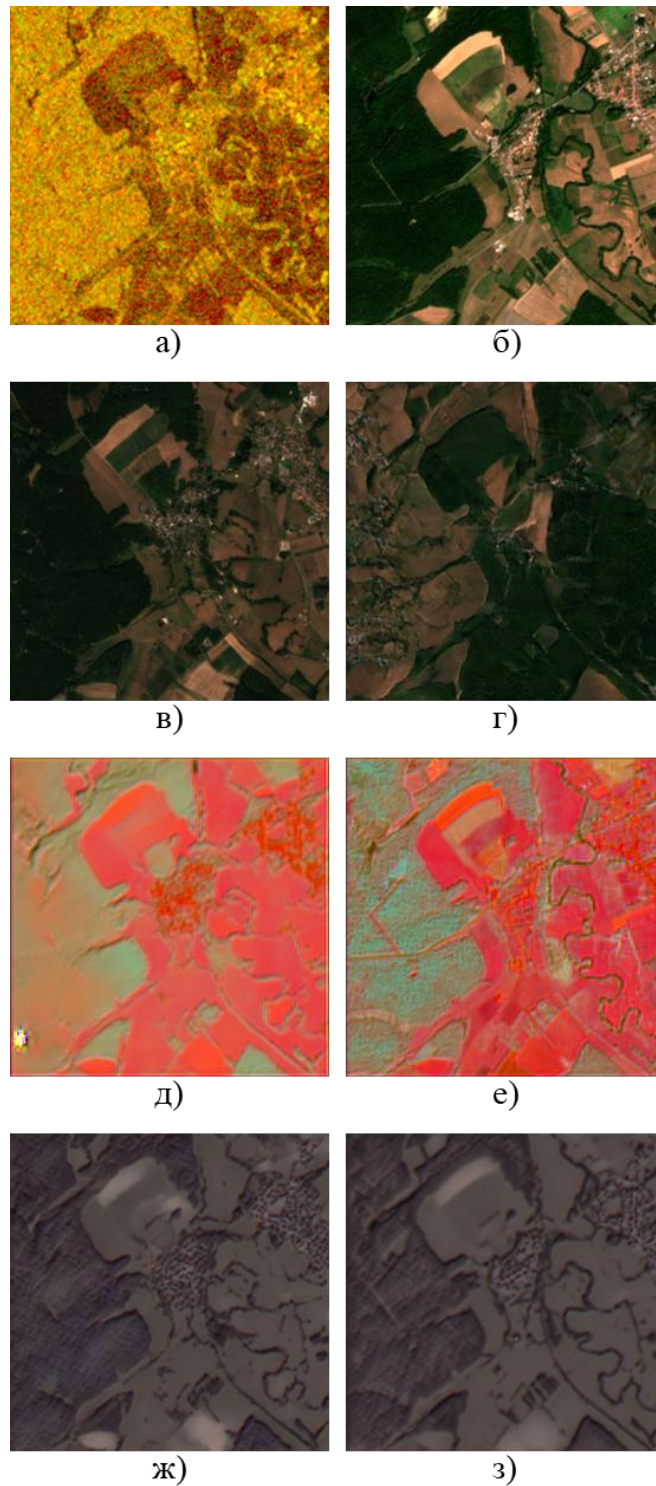


Рисунок 2.3 – Результаты преобразования оптических и радарных изображений в общий домен; а, б – исходные изображения; в – результат преобразования радарного изображения в оптический домен ИНС pix2pix; г – результат преобразования радарного изображения в оптический домен ИНС LSGAN; д, е – результаты ИНС, обученных с критерием (2.14); ж, з – результаты ИНС, обученных с критерием (2.15)

2.4 Выводы и результаты второй главы

Во второй главе диссертации была предложена концепция общего домена, конструируемого по данным исходных доменов, включающая требования к общему домену: сохранение линейных размеров изображений при их преобразовании к общему домену; высокое сходство образов разнородных изображений одной территории, получаемых при их преобразовании в общий домен; невырожденность получаемых образов изображений. Разработан метод построения общего домена для разнородных изображений ДЗЗ, а также различные алгоритмы, конкретизирующие его для различных данных ДЗЗ и критериев.

Проведены экспериментальные исследования по построению общего домена, получаемых как различными оптическими сенсорами, так и оптическим и радарным сенсорами. Для случая разнородных оптических снимков (разные оптические сенсоры) предложены четыре алгоритма построения общего домена: аналитический и три нейросетевых. Продемонстрировано улучшение показателя *SSIM* с 0,38 (при сравнении изображений исходных доменов) до 0,87. Для случая построения общего домена оптических и радарных изображений предложено два алгоритма; достигнуто улучшение показателя *SSIM* с 0,03 до 0,75. Показано преимущество предлагаемого метода перед подходом на основе преобразования радарных изображений в оптический домен: предложенный метод превосходит по показателю *SSIM* преобразование ИНС *pix2pix* на 0,25; ИНС *LSGAN* – на 0,54.

Глава 3. Алгоритм совмещения разнородных изображений дистанционного зондирования Земли с использованием общего домена

3.1 Описание алгоритма

3.1.1 Общая схема алгоритма

Предлагаемый алгоритм опирается на сходство изображений в общем домене, что позволяет использовать существующие методы совмещения 3-канальных изображений для совмещения изображений ДЗЗ [137*,148*].

На вход алгоритма поступает два изображения $I_X \in \mathbb{R}^{W_1 \times H_1 \times K_X}$, $I_Y \in \mathbb{R}^{W_2 \times H_2 \times K_Y}$. Пространственное разрешение изображений должно быть одинаковым. В случае существенных отличий в пространственном разрешении, которые не могут быть устранены простой интерполяцией (линейной или кубической), дополнительно может производиться повышение разрешения одним из существующих методов.

Задача, решаемая алгоритмом, состоит в оценке матрицы проективного преобразования $H \in R^{3 \times 3}$, связывающей координатные плоскости изображений.

Первым этапом совмещения изображений является приведение их к общему домену. При этом производится изменение количества каналов до стандартного количества – трех (в случае одноканальных изображений канал просто дублируется еще два раза для получения трехканального итогового изображения). Диапазон значений отсчетов изображения также ограничивается до $[0; 1]$.

На втором этапе по двум «стандартизированным» изображениям некоторым «оптимальным» алгоритмом находятся особые точки и делается их описание. Их множества передаются на третий этап обработки

На третьем этапе метода происходит фильтрация пар точек и формирование сокращенных подмножеств пар на основе априорных данных о геопривязке изображений.

Наконец, на последнем, четвертом этапе происходит окончательный отбор пар точек соответствия с использованием алгоритма RANSAC [55] и определение параметров проективного преобразования между координатными плоскостями изображений.

Схема предлагаемого алгоритма совмещения пар разнородных изображений ДЗЗ представлена на рисунке 3.1.

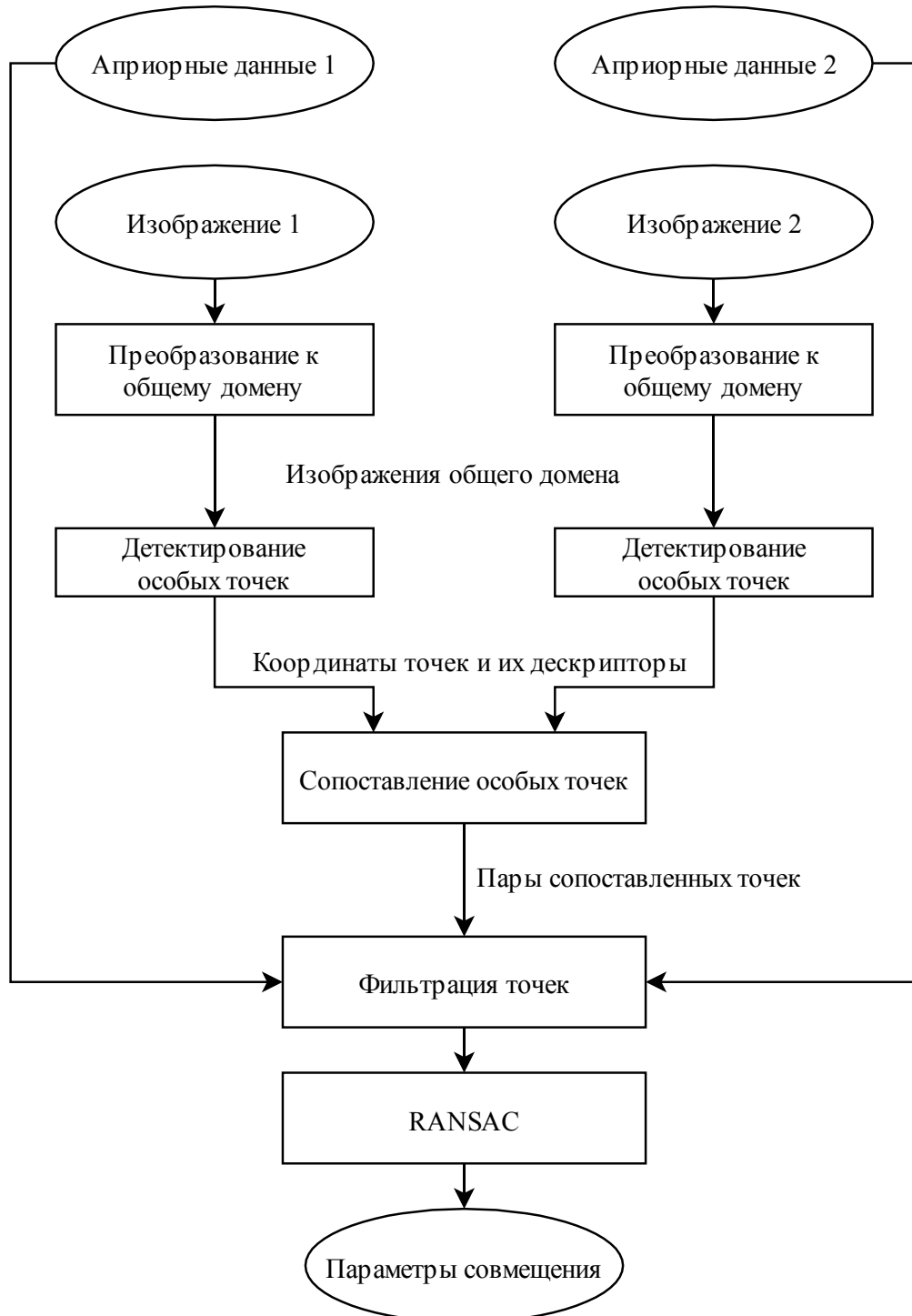


Рисунок 3.1 – Общая схема алгоритма совмещения разнородных изображений

Более формальное описание можно выполнить следующим образом. Пусть даны пара изображений $I_X \in \mathbb{R}^{W_1 \times H_1 \times K_X}$, $I_Y \in \mathbb{R}^{W_2 \times H_2 \times K_Y}$, пара преобразований к общему домену $F_X: \mathbb{R}^{W_1 \times H_1 \times K_X} \rightarrow [0; 1]^{W_1 \times H_1 \times 3}$ и $F_Y: \mathbb{R}^{W_2 \times H_2 \times K_Y} \rightarrow [0; 1]^{W_2 \times H_2 \times 3}$, метод совмещения изображений по особым точкам M и метод построения проективного преобразования R . Необходимо найти оценку матрицы проективного преобразования $H \in R^{3 \times 3}$, связывающую координатные плоскости изображений.

Результатом применения метода M к результатам предобработки I_X и I_Y является набор пар точек P_0 :

$$P_0 = M(F_X(I_X), F_Y(I_Y)),$$

где $P_0 = \{(p_X, p_Y)_i \mid p_X, p_Y \in \mathbb{R}^2\}_{i=1}^N$, p_X, p_Y – координаты особых точек 1 и 2 изображения соответственно

Методы совмещения изображений на основе особых точек могут выдавать ложные соответствия точек, при этом без дополнительной информации нельзя априорно определить, какие из соответствий являются правильными. При наличии дополнительных данных можно заранее отсеивать ложные соответствия точек.

Пусть $P \subseteq P_0$ – множество пар точек, отфильтрованных с учетом дополнительной информации (в случае, если фильтрация не применяется $P = P_0$). Тогда, оценка матрицы проективного преобразования $\hat{H} \in R^{3 \times 3}$, связывающего изображения может быть получена применением R к P .

$$\hat{H}, P_+ = R(P)$$

где $P_+ \subseteq P$ – множество пар точек, которые были признаны корректными и были учтены при построении матрицы $\hat{H}$.

Новыми элементами предложенного алгоритма являются:

- использование приведения изображений ДЗЗ к единому «формату» путем преобразования к общему домену, отличающемуся от исходных доменов;
- способ дополнительной фильтрации пар точек на основе априорной информации о снимке;
- выбор и настройка алгоритма детектирования для поиска точек соответствия и их описания;

- способ совмещения пар точек на основе предварительной их фильтрации по данным геопривязки с учетом пространственного разрешения снимков конкретного домена и априорной точности их геопозиционирования.

3.1.1.1. Преобразование к общему домену

Преобразование к общему домену выполняется алгоритмами, описанными в разделах 2.2 и 2.3.

3.1.1.2. Детектирование особых точек

Поскольку используемый общий домен изображений является 3-канальным ($K = 3$) и ограниченным по значениям отсчетов (отрезком $[0;1]$), изображения общего домена можно совместить с помощью методов совмещения обычных изображений. Преимуществом этого подхода является наличие готовых реализаций таких методов в распространенных библиотеках компьютерного зрения. Наиболее известным методом совмещения обычных изображений является SIFT [57].

В случае прямого совмещения разнородных изображений с одинаковым количеством каналов без последующего доменного преобразования для детектирования особых точек можно использовать метод RIFT2 [80]. В частности, данный метод может использоваться для совмещения радарных и оптических изображений.

Среди доступных современных (англ.: state-of-the-art) ИНС для совмещения изображений выделяются модели DeDoDe и RoMa.

ИНС DeDoDe [149] выполняет совмещение изображений на основе обнаружения особых точек. Она состоит из двух сетей: сети-детектора особых точек и сети-дескриптора, которая вычисляет описание признаков ключевых точек. Дескрипторы признаков сравниваются на основе их близости.

ИНС RoMa [72] выполняет попиксельное совмещение изображений (англ. dense matching). Результатом попиксельного совмещения является карта соответствий между всеми пикселями первого и второго изображений, а также оценка качества для каждой пары пикселей. Такой подход потенциально позволяет находить соответствия между изображениями в условиях сильных нелинейных искажений.

Стоит отметить, что координатные плоскости ортокорректированных изображений дистанционного зондирования связаны аффинным преобразованием. Использование попиксельного отображения для совмещения таких изображений излишне. Как следствие, для оценки параметров преобразования координатных плоскостей изображений достаточно ограничиться рассмотрением N пар точек с наилучшим качеством.

Для методов совмещения изображений, основанных на ИНС, возможно проведение дообучения ИНС на изображениях конкретных доменов с целью адаптации ИНС к их особенностям. В работах [134*, 148*] было рассмотрено обучение ИНС RoMa, однако сама возможность дообучения существует и для других моделей и ограничивается лишь доступностью программного кода обучения соответствующей ИНС.

3.1.1.3. Учет координатной привязки

Обычные изображения, для которых разработано большинство алгоритмов совмещения, чаще всего не имеют ассоциированной с ними дополнительной информации, пригодной для улучшения качества совмещения. Как следствие, алгоритмы совмещения обычных изображений при осуществлении совмещения опираются только на визуальные особенности изображений. В случае алгоритмов совмещения по особым точкам это может приводить к возникновению ложных сопоставлений ключевых точек, расположенных вблизи визуально сходных, но не соответствующих друг другу областей изображения.

Изображения ДЗЗ обычно имеют ассоциированную с ними информацию о геопривязке, которая позволяет определить реальное расположение пикселей в некоторой геодезической системе координат. Совмещение снимков только на основе данных геопривязки возможно, однако погрешности в геопривязке могут приводить к неточному совмещению снимков, особенно в случае совмещения снимков, полученных разными аппаратами.

Даже если геопривязка не является абсолютно точной, она может учитываться при формировании множеств пар особых точек для предварительного отсева пар, физически не допускающих такого совмещения.

Более формально, пусть P_1 – множество пар соответствующих особых точек, $T_X, T_Y \in \mathbb{R}^{3 \times 3}$ – матрицы преобразования между пиксельными и мировыми координатами соответственно, полученные по данным геопривязки изображений I_X, I_Y . Тогда процесс фильтрации пар точек на основе геопривязки может быть описан следующим образом.

Пусть $h(p)$ – оператор приведения пиксельных координат к гомогенным координатам, а $h^{-1}(\tilde{p})$ – оператор приведения гомогенных координат к пиксельным:

$$h(p) = (x, y, 1 \mid p = (x, y))^T,$$

$$h^{-1}(\tilde{p}) = \left(\frac{x}{w}, \frac{y}{w} \mid \tilde{p} = (x, y, w) \right)^T,$$

где $p \in \mathbb{R}^2$ – координаты точки на плоскости, $\tilde{p} \in \mathbb{R}^3$ – координаты точки в гомогенной системе координат.

Оператор проецирования $Pr(T, p)$ задает проективное преобразование:

$$Pr(T, p) = h^{-1}(T \cdot h(p)),$$

где $T \in \mathbb{R}^{3 \times 3}$ – матрица, задающая проективное преобразование, $p \in \mathbb{R}^2$ – координаты точки.

В данных обозначениях набор точек $P_1 \subseteq P_0$, отфильтрованных с учетом данных геопривязки, может быть определен следующим образом:

$$d(p_X, p_Y) = \|Pr(T_X, p_X) - Pr(T_Y, p_Y)\|_2,$$

$$P_1 = \{(p_X, p_Y) | (p_X, p_Y) \in P_0, d(p_X, p_Y) \leq \delta\},$$

где δ – параметр алгоритма, задающий максимальное отклонение между точками в реальных координатах. Данный параметр задается априорно, исходя из данных о точности геопривязки конкретных типов/доменов снимков: если δ_X – погрешность координат первого типа снимков, а δ_Y – второго, то итоговая погрешность $\delta = \delta_X + \delta_Y$.

Необходимо заметить, что учет данных геопривязки при решении задачи совмещения снимков ДЗЗ ранее уже рассматривался в работах [75, 150] других авторов. Но в этих работах совмещение выполнялось для конкретных пар сенсоров. В предлагаемом варианте предложен алгоритм общего вида для снимков произвольных типов/доменов.

3.2 Экспериментальные исследования

3.2.1 Показатели качества совмещения

Поскольку совмещение изображений выполняется в 2 этапа – сопоставление особых точек и вычисление параметров проективного преобразования по сопоставленным точкам – для оценки качества совмещения используются 2 группы показателей.

Первая группа показателей предназначена для оценки качества сопоставления особых точек. Напомним, что результатом алгоритма обнаружения и сопоставления пар особых точек является набор пар точек P_0 . При использовании данных геопривязки

Наиболее простым показателем является количество пар сопоставленных точек $|P|$.

Показатель Δ_P используется для оценки координатной точности сопоставления точек и равен среднему расстоянию между точками после проецирования изображения X в координатную плоскость изображения Y :

$$\Delta_P = \frac{1}{|P|} \sum_{(p_X, p_Y) \in P} \|Pr(H, p_X) - p_Y\|,$$

где H – истинная матрица проективного преобразования.

Показатель CMR_δ определяется, как доля пар точек, для которых истинная координатная ошибка не превосходит δ :

$$CMR = \frac{|P_d|}{|P|},$$

где $P_d = \{(p_X, p_Y) \mid (p_X, p_Y) \in P, \|Pr(H, p_X) - p_Y\| \leq d\}$.

Показатель LE_δ определяется, как среднее расстояние между точками в парах, для которых координатная ошибка не превосходит δ :

$$LE_d = \frac{1}{|P_d|} \sum_{(p_X, p_Y) \in P_d} \|Pr(H, p_X) - p_Y\|.$$

Вторая группа показателей используется для оценки качества совмещения изображений. Напомним, что результатом алгоритма RANSAC являются оценка матрицы проективного преобразования $\hat{H}$ и набор точек $P_+ = \{(p_X, p_Y) \mid (p_X, p_Y) \in P, \|Pr(\hat{H}, p_X) - p_Y\| \leq \delta_{RANSAC}\}$, которые согласуются с данной матрицей. Под «согласованностью» пары понимается координатная ошибка менее порога δ_{RANSAC} , который в рамках экспериментальных исследований был установлен в $\delta_{RANSAC} = 3$.

Показатель MMA (англ. mean matching accuracy), определяется, как доля пар точек, удовлетворяющих полученному проективному преобразованию:

$$MMA = \frac{|P_+|}{|P|},$$

Показатель ACE (англ. average corner error) определяется, как величина рассогласования угловых точек изображений после их согласования. Поскольку изображения набора данных имеют фиксированный размер, ACE определяется, как:

$$Corners = \{(0, 0), (0, 255), (255, 0), (255, 255)\},$$

$$ACE = \frac{1}{4} \sum_{c \in Corners} \|Pr(\hat{T}, c) - c\|,$$

где $\|\cdot\|$ – евклидова норма.

Показатель SR (success rate) определяется, как доля изображений с $ACE < 10$ (в случае провала сопоставления ACE считается равным бесконечности).

3.2.2 Совмещение разнородных оптических снимков

3.2.2.1. Используемые алгоритмы и набор данных

Для экспериментов использовались изображения тестовой выборки, ранее использовавшейся в экспериментах по построению общего домена (30 пар изображений Landsat-8 и Sentinel-2, размер изображения 792×792 , подробнее см. раздел 2.2.2.1).

В качестве алгоритма обнаружения и совмещения особых точек используется SIFT [57].

3.2.2.2. Результаты экспериментов

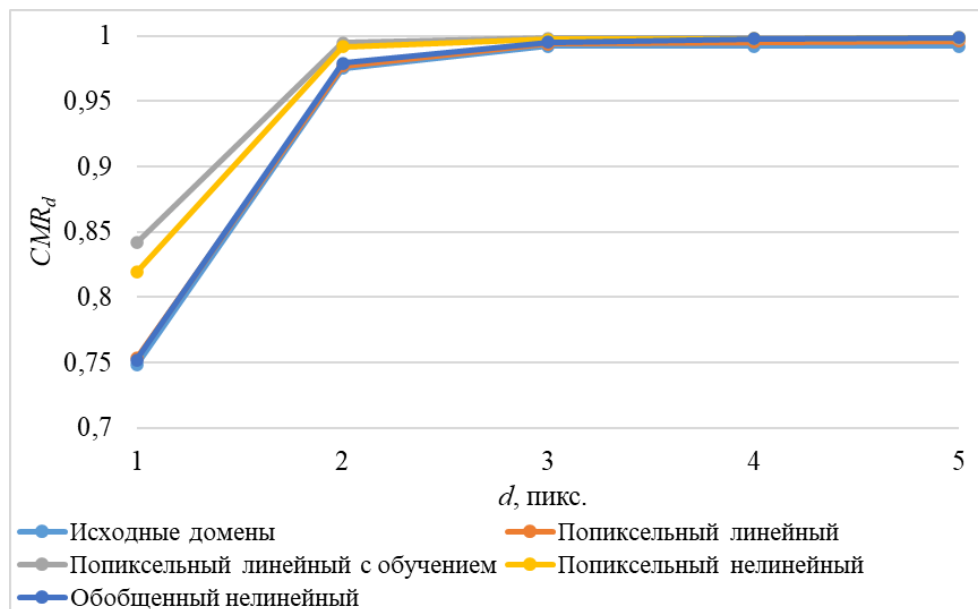
Результаты без учета геопривязки

Средние значения показателей качества совмещения приведены в таблице 3.1. Первая строка таблицы содержит результаты для совмещения изображений исходных доменов.

Графики зависимости значений показателей LE_d и CMR_d от граничного значения d приведены на рисунках 3.2 и 3.3 соответственно.

Таблица 3.1 – Значения показателей качества совмещения оптических изображений

Алгоритм построения общего домена	Показатель качества				
	SR	ACE	$ P $	Δ_P	MMA
-	1,0	0,66	32,1	3,49	0,992
Попиксельный линейный	1,0	0,63	91,4	2,05	0,993
Попиксельный линейный с обучением	1,0	0,57	81,2	1,31	0,998
Попиксельный нелинейный	1,0	0,54	83,3	0,83	0,997
Обобщенный нелинейный	1,0	0,87	57,9	1,19	0,995

Рисунок 3.2 – Значения показателя CMR_d при совмещении оптических изображений

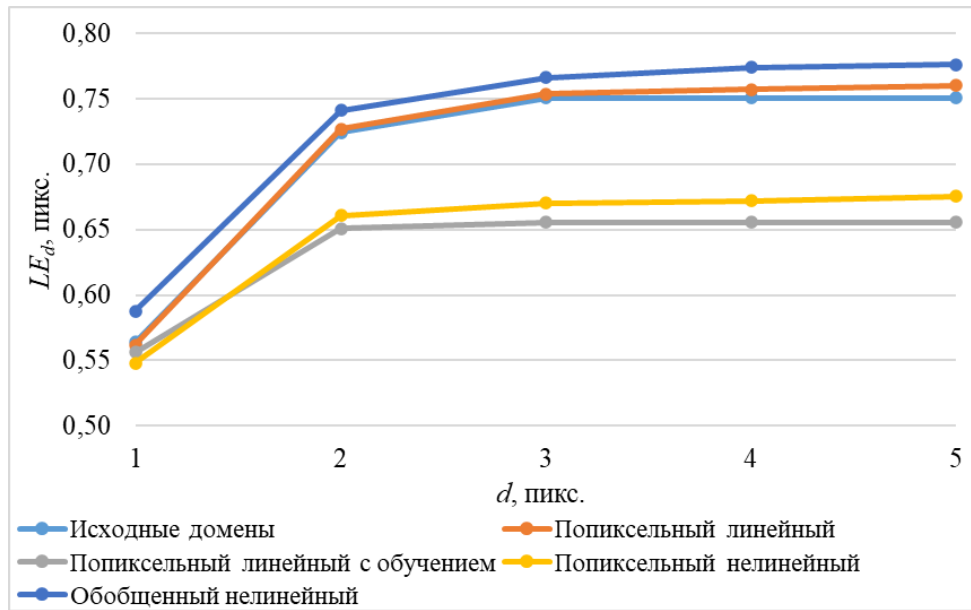


Рисунок 3.3 – Значения показателя LE_d при совмещении оптических изображений

Результаты с учетом геопривязки

Согласно [151], погрешность геопривязки снимков Sentinel-2 оценивается в 12,5 м. Для снимков Landsat-8 оценка погрешности геопривязки равна 18 м [152]. Как следствие, в рамках экспериментов по фильтрации данных на основе геопривязки используется предельное отклонение $\delta = 30,5$ м. С учетом пространственного разрешения снимков 15 м/пиксель, данное отклонение в мировых координатах соответствует отклонению в 2,03 пикселя.

Средние значения показателей качества совмещения приведены в таблице 3.2. Из рассматриваемой таблицы видно, что использование данных геопривязки позволяет повысить качество совмещения изображений. Стоит отметить, что результирующая точность совмещения изображений превышает как точность сопоставления снимков без использования данных геопривязки, так и точность самой геопривязки.

Таблица 3.2 – Значения показателей качества совмещения оптических изображений с учетом геопривязки

Алгоритм построения общего домена	Показатель качества				
	ACE	SR	$ P $	Δ_p	MMA
-	0,74	1,0	31,3	0,73	0,999
Попиксельный линейный	0,62	1,0	89,5	0,73	1,0
Попиксельный линейный с обучением	0,57	1,0	80,8	0,65	1,0
Попиксельный нелинейный	0,52	1,0	82,7	0,66	1,0
Обобщенный нелинейный	0,73	1,0	57,0	0,74	1,0

Примеры результатов сопоставления особых точек с учетом геопривязки приведены на рисунке 3.4.

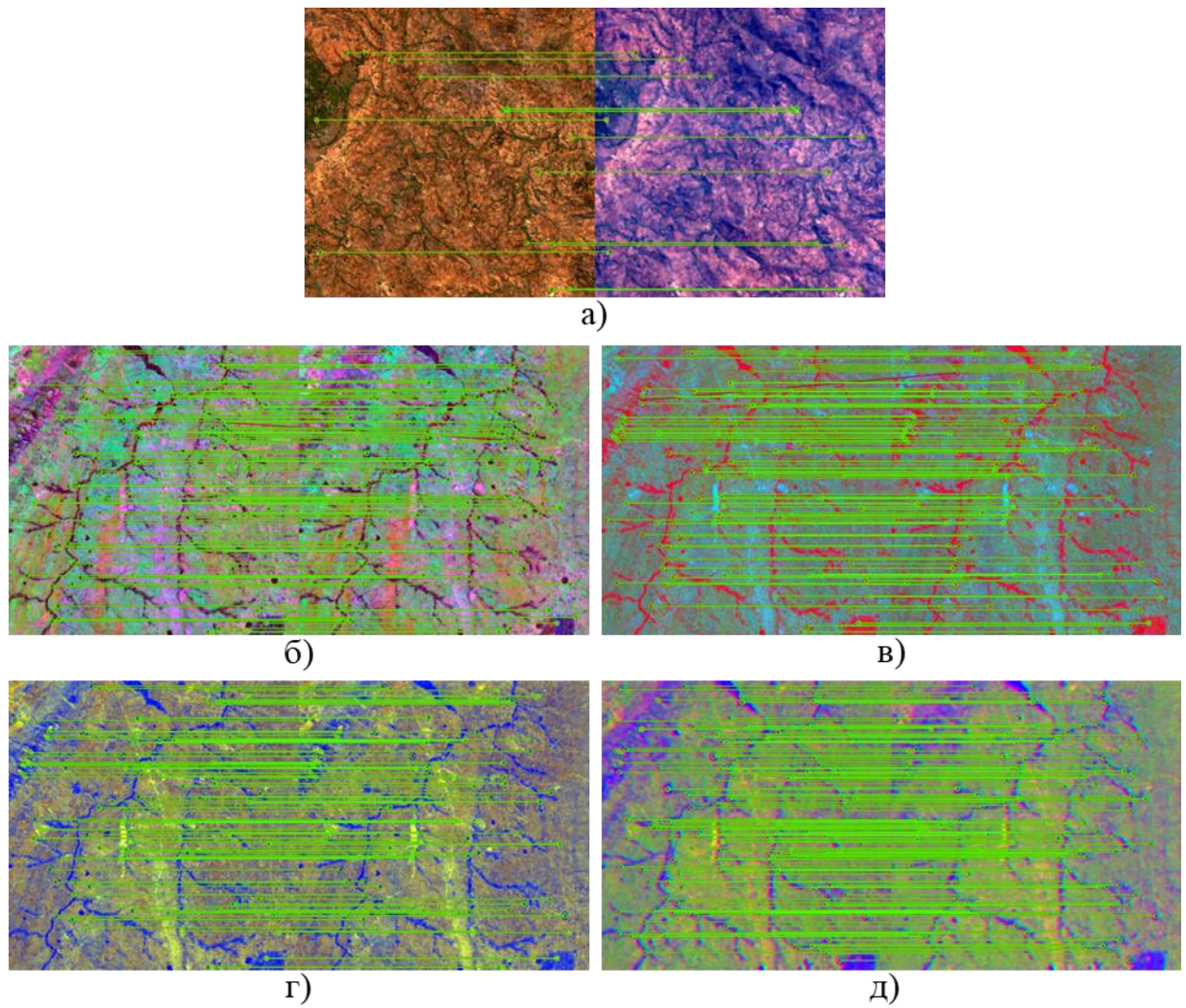


Рисунок 3.4 – Результаты детектирования и сопоставления особых точек на разнородных оптических изображениях;

а – исходные домены; б – попиксельный линейный алгоритм;

в – попиксельный линейный алгоритм с обучением; г – попиксельный нелинейный алгоритм; д – обобщенный нелинейный алгоритм

3.2.3 Совмещение оптических и радарных снимков

3.2.3.1. Используемые алгоритмы и набор данных

В качестве методов совмещения изображений используются методы SIFT, RIFT2, DeDoDe и RoMa. Количество детектируемых на изображении особых точек ограничено 4096. ИНС RoMa используется в 2 вариантах:

- без дообучения (веса сети предоставлены авторами RoMA, ИНС обучена на обычных, т.е. не-ДЗЗ, изображениях);
- в дообученном виде (ИНС дообучена на изображениях общего домена).

В качестве набора данных выступал набор MultiSenGE [136], использовавшийся ранее в экспериментах по построению общего домена в разделе 2.3. Набор состоит из мультивременных последовательностей снимков, полученных Sentinel-1 и Sentinel-2. Радарные и оптические снимки объединены в пары по близости времени съемки.

Обучающая выборка состояла из 46 092 пар, валидационная – из 11 522 пар, тестовая – из 14 405 пар снимков. Значения отсчетов радарных изображений были переведены в логарифмическую шкалу.

Для сравнения качества сопоставления снимков исходных доменов используется набор снимков, состоящих из RGB-каналов оптических снимков, и псевдо-3-канальных радарных снимков (зеленый канал – VH, красный канал – VV, синий канал занулен). Подробное описание построение вспомогательного набора приведено в подразделе 2.3.2.1.

3.2.3.2. Обучение нейросетей

Эксперименты проводились на машине с AMD Ryzen 5600X, 32 GB ОЗУ и Nvidia GeForce RTX 3090. В качестве фреймворка для машинного обучения использовался PyTorch 2.7.

В качестве доменных преобразований используются ИНС, обучение которых было рассмотрено в разделе 2.3:

- пара ИНС U-Net, обученных с использованием критерия (2.14);
- пара ИНС U-Net, обученных с использованием критерия (2.15);
- ИНС $pix2pix$, выполняющая преобразование радарных изображений в оптический домен;

Обучение ИНС RoMa проводилось со стандартными параметрами [72], предлагаемыми авторами ИНС, за исключением размера изображений при обучении, равного 224×224 (ИНС требует, чтобы размер изображений был кратен 14, кроме того, поскольку на одном из этапов работы ИНС изображение меняет разрешение до 448×448 , более выгодным является выбрать кратный размер);

Аугментации, используемые при дообучении RoMa, включали в себя

- случайный сдвиг по горизонтали/вертикали на величину до 64 пикселей;
- случайный поворот до 30° ;
- масштабирование со случайным коэффициентом из интервала $[0,9; 1,1]$.

3.2.3.3. Результаты экспериментов

Результаты без учета геопривязки

Экспериментальные исследования без учета геопривязки проводились в 2 этапа.

На первом этапе был проведен сравнительный анализ качества совмещения радарных и оптических ДЗЗ снимков при использовании различных методов совмещения изображений и методов преобразования к общему домену. Количество особых точек на изображении ограничено 4096. При расчете средних значений показателей *ACE* и *ММА* учитываются только результаты успешно сопоставленных изображений (см. определение критерия *SR*).

Средние значения показателей качества совмещения приведены в таблицах 3.3–3.7. Первая строка каждой таблицы содержит результаты для совмещения

изображений исходных доменов, вторая и третья строки – результаты совмещения оптического RGB-изображения и псевдооптического RGB-изображений, сгенерированного обученной моделями LSGAN и pix2pix соответственно. Третья и четвертая строки содержат результаты совмещения изображений в общем домене, полученными ИНС, обученными с критериями (2.14) и (2.15) соответственно (обучение ИНС преобразования было рассмотрено в разделе 2.3.2.2).

Таблица 3.3 – Доля успешно совмещенных изображений (показатель SR)

Метод преобразования	Метод совмещения			
	SIFT	RIFT2	DeDoDe	RoMa
-	0	0,01	0,05	0,003
LSGAN	0	0,01	0,05	0,001
pix2pix	0,07	0,05	0,92	0,56
Вариант-1	0,02	0,06	0,90	0,53
Вариант-2	0,35	0,24	0,98	0,96

Таблица 3.4 – Средняя ошибка совмещения изображений (показатель ACE)

Метод преобразования	Метод совмещения			
	SIFT	RIFT2	DeDoDe	RoMa
-	-	7,5	6,0	4,9
LSGAN	-	7,7	6,2	6,2
pix2pix	5,8	7,3	3,4	3,3
Вариант-1	5,0	7,3	2,9	2,4
Вариант-2	4,5	6,8	1,79	1,1

Таблица 3.5 – Среднее количество пар точек (показатель $|P|$)

Метод преобразования	Метод совмещения			
	SIFT	RIFT2	DeDoDe	RoMa
-	2,7	38,8	168,1	4096
LSGAN	3,7	49,6	148,2	4096
pix2pix	7,7	84,9	570,5	4096
Вариант-1	2,8	67,7	424,1	4096
Вариант-2	10,7	184,8	859,0	4096

Таблица 3.6 – Средняя координатная ошибка сопоставленных точек (показатель Δ_P)

Метод преобразования	Метод совмещения			
	SIFT	RIFT2	DeDoDe	RoMa
-	125,3	49,2	89,7	144,1
LSGAN	126,4	40,5	90,7	149,6
pix2pix	77,7	31,1	14,8	56,5
Вариант-1	73,8	26,9	11,4	52,1
Вариант-2	29,9	16,3	4,4	7,7

Таблица 3.7 – Среднее значение показателя MMA

Метод преобразования	Метод совмещения			
	SIFT	RIFT2	DeDoDe	RoMa
-	0	0,30	0,19	0,54
LSGAN	0	0,32	0,19	0,40
pix2pix	0,19	0,33	0,43	0,57
Вариант-1	0,06	0,37	0,55	0,66
Вариант-2	0,68	0,36	0,72	0,90

Как видно, метод построения общего домена на основе критерия (2.15) обеспечивает значительно лучшие показатели качества совмещения изображений.

Важно отметить, что на данном этапе результаты достигнуты при использовании методов сопоставления обычных изображений (не адаптированных к изображениям ДЗЗ). Из полученных результатов видно, что метод RoMa обеспечивает наилучшую точность сопоставления изображений, хотя и уступает по показателю SR методу DeDoDe.

Графики зависимости значений показателей LE_d и CMR_d от граничного значения d приведены на рисунках 3.5 и 3.6 соответственно. На данных графиках для каждого метода преобразования приведены значения показателя только наилучшего метода совмещения (наилучшим считается метод совмещения с минимальным значением $\frac{ACE}{SR}$).

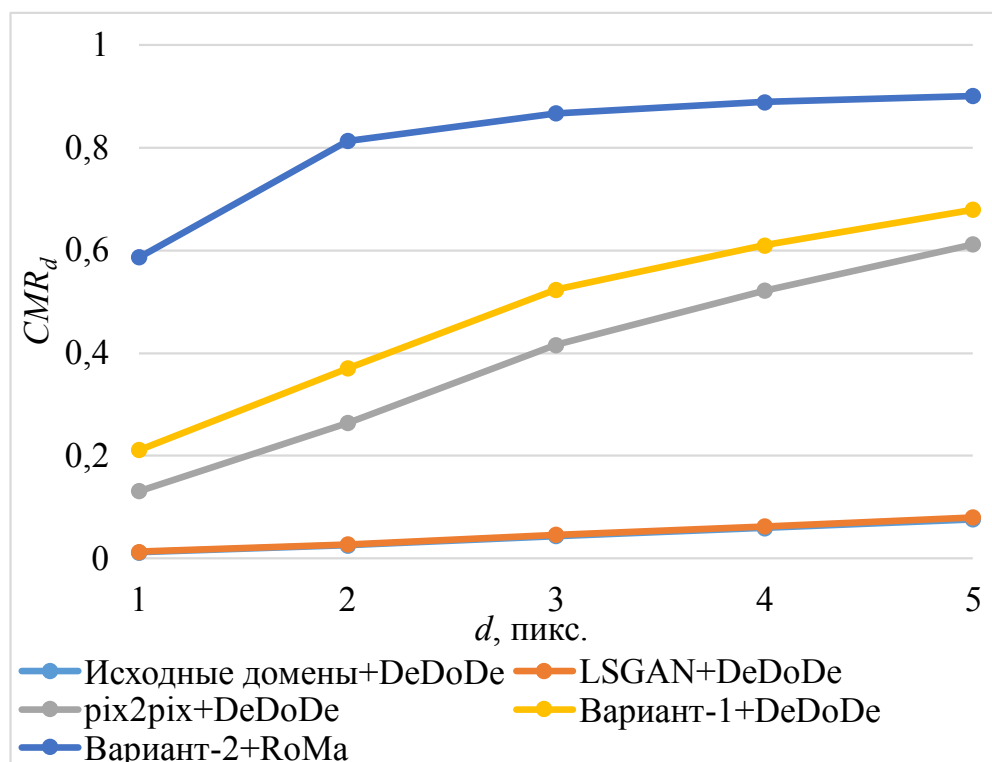
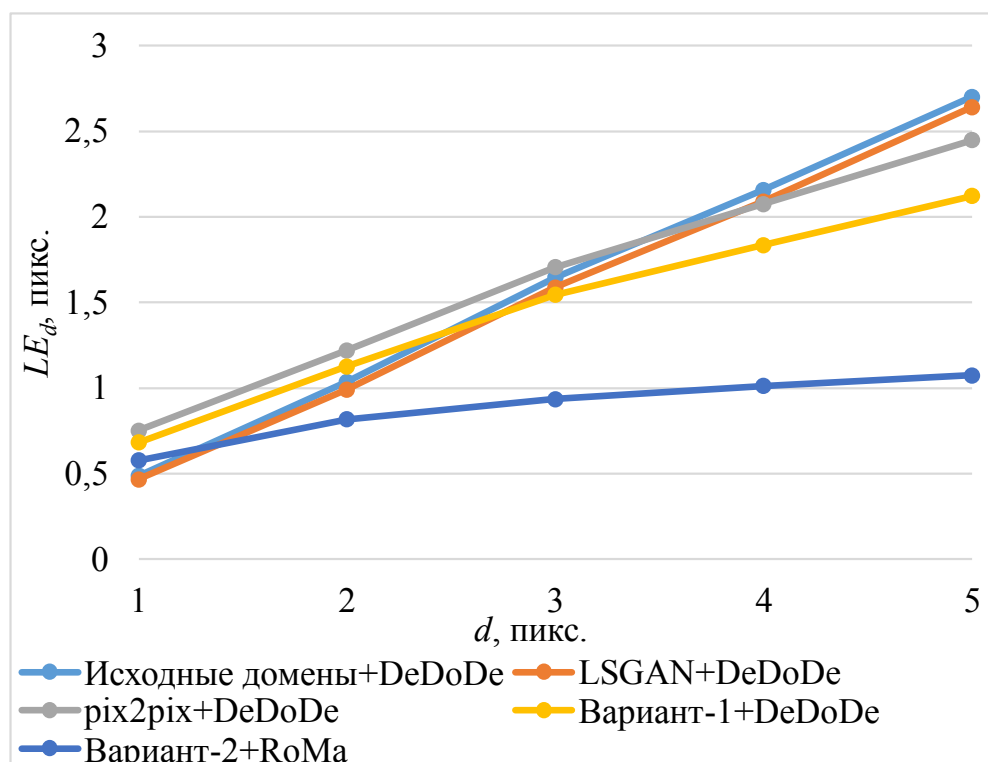


Рисунок 3.5 – Значения показателя CMR_d

Рисунок 3.6 – Значения показателя LE_d

На втором этапе ИНС RoMa была дообучена на изображениях общего домена на основе критерия (2.15). Данный метод преобразования к общему домену был выбран, как демонстрирующий наилучшее качество совмещения изображений с помощью стандартных методов.

Средние значения показателей качества совмещения изображений дообученной ИНС RoMa приведены в таблице 3.8. С целью сравнения в данных таблицах представлены также результаты других методов совмещения в том же домене.

Таблица 3.8 – Сводные значения показателей качества согласования в общем домене

Метод совмещения	Показатель качества				
	SR	ACE	$ P $	Δ_P	MMA
SIFT	0,35	4,5	10,7	29,9	0,68
RIFT2	0,24	6,8	184,8	16,3	0,36
DeDoDe	0,98	1,8	859,0	4,4	0,72
RoMa без обучения	0,96	1,1	4096	7,7	0,90
RoMa с обучением	0,999	0,83	4096	1,02	0,999

Графики зависимости значений показателей LE_d и CMR_d от граничного значения d для ИНС RoMa с дообучением и без дообучения приведены на рисунках 3.7 и 3.8.

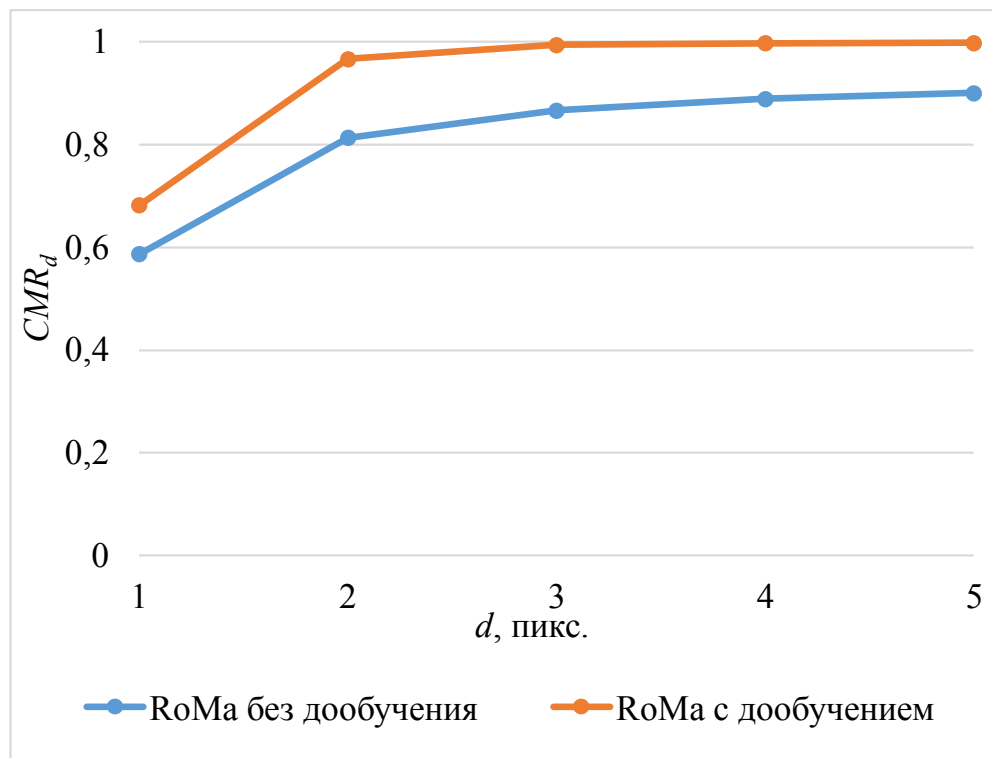


Рисунок 3.7– Значения показателя CMR_d для ИНС RoMa до и после дообучения

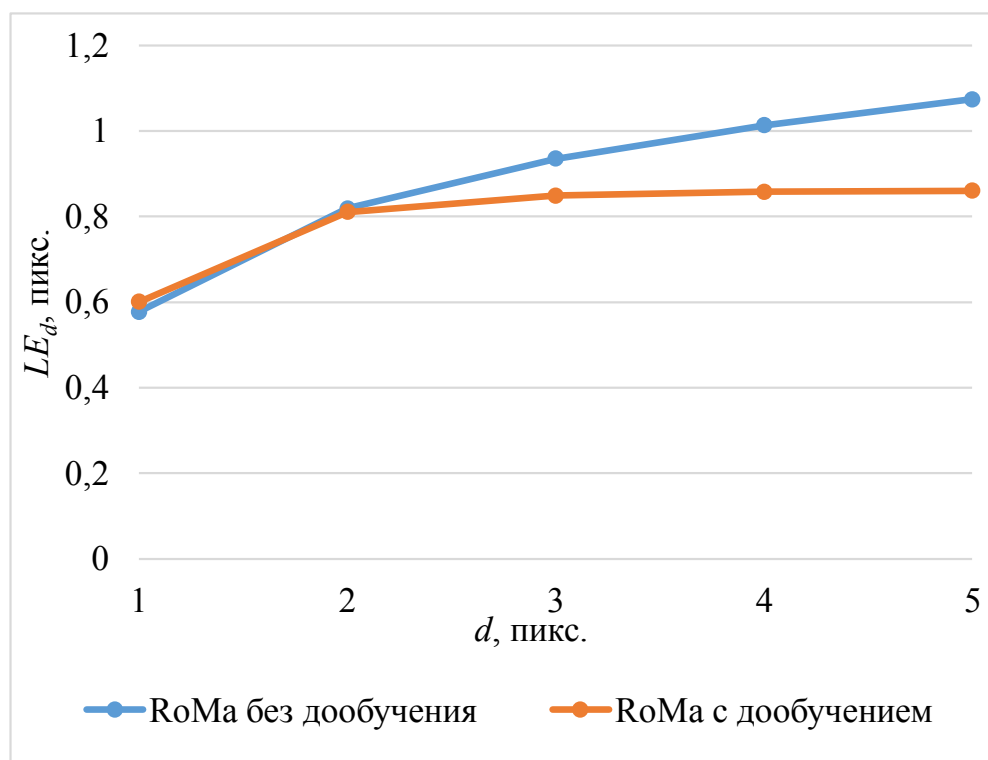


Рисунок 3.8– Значения показателя LE_d для ИНС RoMa до и после дообучения

Результаты с учетом геопривязки

Согласно [151], погрешность геопривязки снимков Sentinel-2 в большинстве случаев не превышает 12,5 м/пиксель. Номинальная точность геопривязки Sentinel-1 равна 7 м [153]. Хотя в [153] отмечается, что измеренная ошибка геопривязки не превысила 5 м, с учетом преобразования снимков Sentinel-1 к меньшему пространственному разрешению 10 м/пиксель в качестве ошибки геопривязки для снимков выбрано номинальное значение в 7 м. Итоговое предельное отклонение в рамках экспериментов по фильтрации данных на основе геопривязки равно $\delta=19,5$ м. С учетом пространственного разрешения снимков 10 м/пиксель, данное отклонение в мировых координатах соответствует отклонению в 1,95 пикселя.

Средние значения показателей качества совмещения приведены в таблицах 3.9–3.13. Первая строка каждой таблицы содержит результаты для совмещения изображений исходных доменов, вторая и третья строки – результаты совмещения оптического RGB-изображения и псевдооптического RGB-изображений, сгенерированного обученной моделью LSGAN и pix2pix соответственно. Третья и

четвертая строки содержат результаты совмещения изображений в общем домене, полученными ИНС, обученными с критериями (2.14) и (2.15) соответственно. Обозначение RoMa₊ соответствует ИНС RoMa с дообучением.

Таблица 3.9 – Доля успешно совмещенных изображений (показатель *SR*) при учете данных геопривязки

Метод преобразования	Метод совмещения				
	SIFT	RIFT2	DeDoDe	RoMa	RoMa ₊
-	0	0,05	0,13	0,01	
LSGAN	0	0,06	0,12	0,005	
pix2pix	0,06	0,22	0,98	0,73	
Вариант-1	0,02	0,21	0,97	0,67	
Вариант-2	0,33	0,60	0,99	0,98	0,999

Таблица 3.10 – Средняя ошибка совмещения изображений (показатель *ACE*) при учете данных геопривязки

Метод преобразования	Метод совмещения				
	SIFT	RIFT2	DeDoDe	RoMa	RoMa ₊
-	-	5,29	3,44	4,74	
LSGAN	-	5,66	3,70	5,28	
pix2pix	4,64	4,68	0,96	2,25	
Вариант-1	4,29	4,70	1,03	1,97	
Вариант-2	3,58	3,64	0,75	0,84	0,75

Таблица 3.11 – Среднее количество пар точек (показатель $|P|$) при учете данных геопривязки

Метод преобразования	Метод совмещения				
	SIFT	RIFT2	DeDoDe	RoMa	RoMa ₊
-	0,01	1,9	3,5	9,3	
LSGAN	0,01	2,8	3,3	3,65	
pix2pix	1,4	6,7	131,3	1091,1	
Вариант-1	0,8	6,1	142,4	1252,9	
Вариант-2	5,5	25,0	454,3	3314,2	3942,5

Таблица 3.12 – Средняя координатная ошибка сопоставленных точек (показатель Δ_P) при учете данных геопривязки

Метод преобразования	Метод совмещения				
	SIFT	RIFT2	DeDoDe	RoMa	RoMa ₊
-	1,2	1,0	1,0	1,3	
LSGAN	1,2	1,0	1,0	1,3	
pix2pix	1,2	1,0	1,0	1,1	
Вариант-1	1,0	1,0	0,9	1,0	
Вариант-2	1,0	1,0	0,9	0,8	0,8

Таблица 3.13 – Среднее значение показателя MMA при учете данных геопривязки

Метод преобразования	Метод совмещения				
	SIFT	RIFT2	DeDoDe	RoMa	RoMa ₊
-	0	0,09	0,34	0,02	
LSGAN	0	0,15	0,31	0,01	
pix2pix	0,10	0,46	0,995	0,94	
Вариант-1	0,02	0,49	0,994	0,95	
Вариант-2	0,68	0,90	0,998	0,996	1,0

Из рассмотренных таблиц видно, что использование данных геопривязки позволяет повысить качество совмещения изображений. Стоит отметить, что результирующая точность совмещения изображений превышает как точность сопоставления снимков без использования данных геопривязки, так и точность самой геопривязки.

Гистограммы распределения показателя ACE до и после фильтрации точек по данным геопривязки при совмещении изображений общего домена дообученной ИНС RoMa приведены на рисунке 3.9.

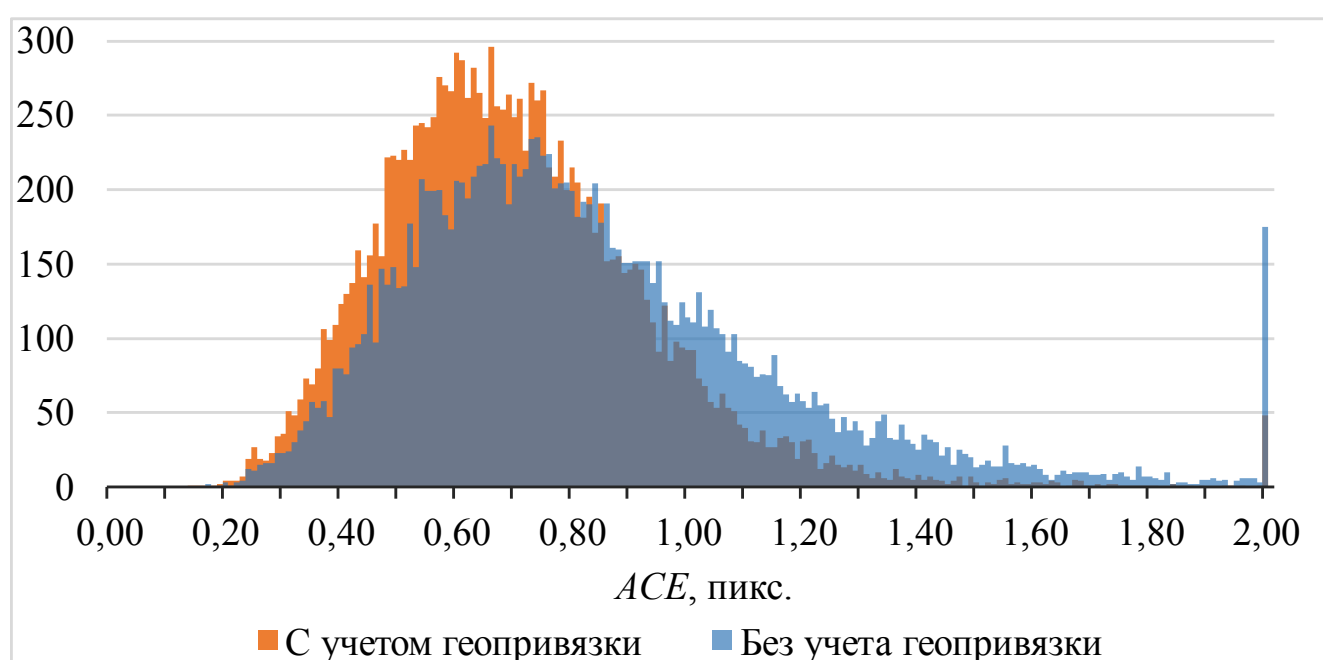


Рисунок 3.9 – Гистограммы показателя ACE до и после фильтрации точек

Итоговые результаты совмещения изображений с учетом данных геопривязки при использовании различных методов междоменного преобразования приведены в таблице 3.14. Для краткости, в данной таблице приведены только результаты методов совмещения, оптимальных для каждого из методов преобразования к общему домену.

Таблица 3.14 – Показатели качества совмещения оптических и радарных изображений с учетом данных геопривязки

Используемые методы	Показатель качества				
	ACE	SR	$ P $	Δ_P	MMA
Без преобразования + DeDoDe	3,44	0,13	3,5	1,0	0,34
LSGAN + DeDoDe	3,70	0,12	3,3	1,0	0,30
pix2pix + DeDoDe	0,96	0,98	131,3	1,0	0,995
Вариант-1+DeDoDe	1,03	0,97	142,4	0,9	0,994
Вариант-2 + RoMa	0,84	0,98	3314,2	0,8	0,996
Вариант-2 + RoMa с дообучением	0,75	0,999	3947,0	0,8	1,0

Примеры результатов сопоставления особых точек с учетом данных геопривязки приведены на рисунке 3.10.

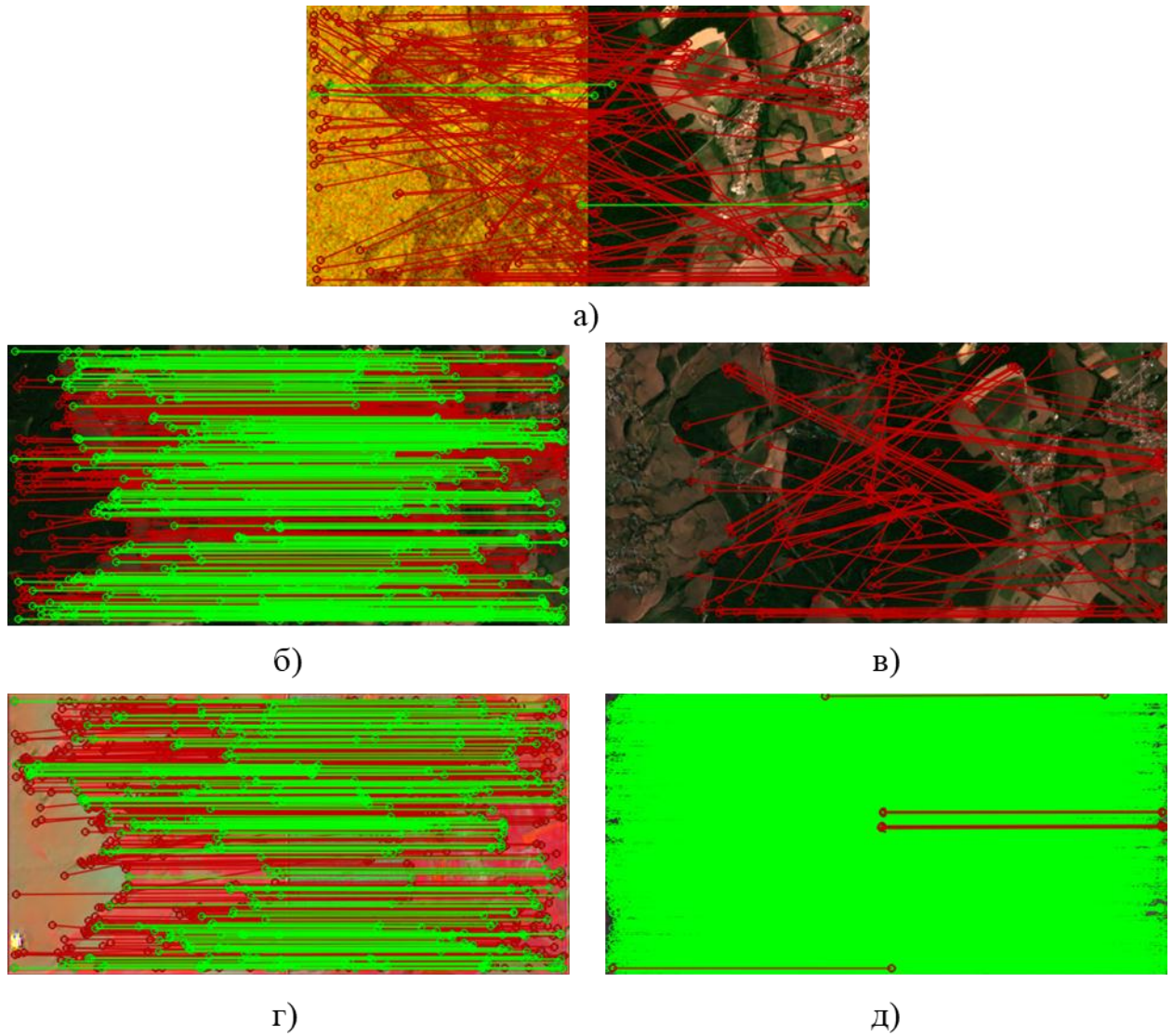


Рисунок 3.10 – Результаты детектирования и сопоставления особых точек на оптических и радарных изображениях; красным помечены пары точек, не соответствующие геопривязке;

а – исходные домены+DeDoDe; б – pix2pix + DeDoDe;

в – LSGAN+DeDoDe; г – общий домен по критерию (2.14) + DeDoDE;

д – общий домен по критерию (2.15) + RoMa с дообучением

3.3 Выводы и результаты третьей главы

В третьей главе был рассмотрен алгоритм совмещения изображений ДЗЗ с помощью их преобразования в общий домен и учета априорных данных геопривязки для устранения ложных соответствий между изображениями.

В разделе 3.1 приведено описание предложенного алгоритма, состоящего из четырех этапов: преобразование изображений к общему домену, детектирование и сопоставление особых точек, фильтрация пар точек на основе данных геопривязки и оценка параметров совмещения изображений по множеству пар точек.

В разделе 3.2 приведены описание и результаты экспериментальных исследований по сопоставлению радарных и оптических изображений с помощью предложенного метода. Было показано, что совмещение изображений в рамках общего домена может успешно проводиться существующими методами совмещения обычных изображений без их дополнительной адаптации к ДЗЗ-снимкам. Возможность адаптации методов совмещения на основе ИНС была продемонстрирована на примере дообучения ИНС RoMa на изображениях общего домена. В том же разделе приведены исследования по совмещению изображений с учетом данных геопривязки. Было показано, что фильтрация точек по данным геопривязки позволяет существенно улучшить точность совмещения радарных и оптических изображений за счет устранения заведомо ложных сопоставлений особых точек.

Глава 4. Технология решения прикладных задач обработки разнородных изображений ДЗЗ с использованием метода общего домена

4.1 Общее описание технологии

Рассмотрим обобщенную схему решения задач, опирающихся на совместную обработку пар разнородных изображений ДЗЗ [154*].

Пусть дана пара пространственно совмещенных изображений $X \in \mathbb{R}^{W \times H \times K_X}$, $Y \in \mathbb{R}^{W \times H \times K_Y}$ равного пространственного разрешения. Прикладная задача решается некоторой функцией $F_T: \mathbb{R}^{W \times H \times K_X} \times \mathbb{R}^{W \times H \times K_Y} \rightarrow \mathbb{T}$, где $\mathbb{T}$ – домен выходных данных, определяемых задачей. К примеру, в задаче классификации изображений $\mathbb{T} \subseteq \mathbb{R}^C$, где C – количество классов. В случае задачи семантической сегментации $\mathbb{T} \subseteq \mathbb{R}^{W \times H \times C}$, поскольку в данном случае классификации подвергается каждый класс по отдельности. В задаче обнаружения изменений, которую можно свести к сегментации объединенной пары снимков, $\mathbb{T} \subseteq \mathbb{R}^{W \times H}$.

Конкретный вид F_T определяется в ходе решения оптимизационной задачи

$$F_T = \arg \min_{f_T \in \mathcal{F}_T} \mathbb{E}_{(X,Y,M) \in \mathcal{Z}} L_T(f_T(X,Y), M), \quad (4.1)$$

где $\mathcal{F}_T$ – множество функций, по которому производится поиск, $\mathcal{Z}$ – обучающее множество пар изображений и ожидаемых выходных значений, $L_T: \mathbb{R}^{H_T \times W_T \times C_T} \times \mathbb{R}^{H_T \times W_T \times C_T} \rightarrow \mathbb{R}$ – функция потерь, измеряющая расхождение ожидаемым и полученным выходным значением.

Для решения (4.1) можно применить подход на основе общего домена. Для решения задачи необходимо определить 3 функции: F_X и F_Y , выполняющие преобразование изображений в общий домен, и F_T , выполняющую сегментацию в рамках общего домена. В рамках данной схемы решение итоговой задачи выполняется функцией $F_T(F_X(X), F_Y(Y))$.

Данный подход имеет несколько преимуществ. Во-первых, использование общего домена позволяет также выполнять предварительное совмещение изображений (как показано в предыдущей главе). Во вторых, независимость F_T от

исходных доменов позволяет также вычислять $F_T(F_Y(Y), F_X(X))$, $F_T(F_X(X'), F_X(X))$ и $F_T(F_Y(Y'), F_Y(Y))$, где X' и Y' – изображения исходных доменов, снятые в моменты времени, отличные от времени съемки X и Y . Особенно полезно данное свойство в задаче обнаружения изменений, поскольку в данной задаче важен временной интервал между снимками, а возможность использовать снимки от двух разных спутников существенно увеличивает количество моментов съемки. В-третьих, данный подход достаточно просто обобщается на большее число доменов: для каждого нового домена достаточно обучить преобразование к уже существующему общему домену и, опционально, дообучить ИНС, решающую практическую задачу.

На основе предложенного возможно построить технологию обработки разнородных снимков ДЗЗ, в рамках которой возможно выполнять решение различных практических задач. В рамках данной технологии обработка разнородных снимков производится в два этапа:

- 1) преобразование разнородных изображений в общий домен парой ИНС, реализующих преобразования F_X и F_Y ;
- 2) обработка изображений в рамках общего домена единой ИНС, реализующей преобразование F_T .

Для функционирования технологии необходимы преобразования F_X , F_Y и F_T . Обучение ИНС, реализующих данные преобразования, производится в 3 этапа:

- 1) построение общего домена: обучение ИНС, реализующих преобразования F_X и F_Y ;
- 2) обучение ИНС, реализующей преобразование F_T ;
- 3) совместное дообучение ИНС (опционально).

Общая схема обучения ИНС в рамках предлагаемой технологии приведена на рисунке 4.1.

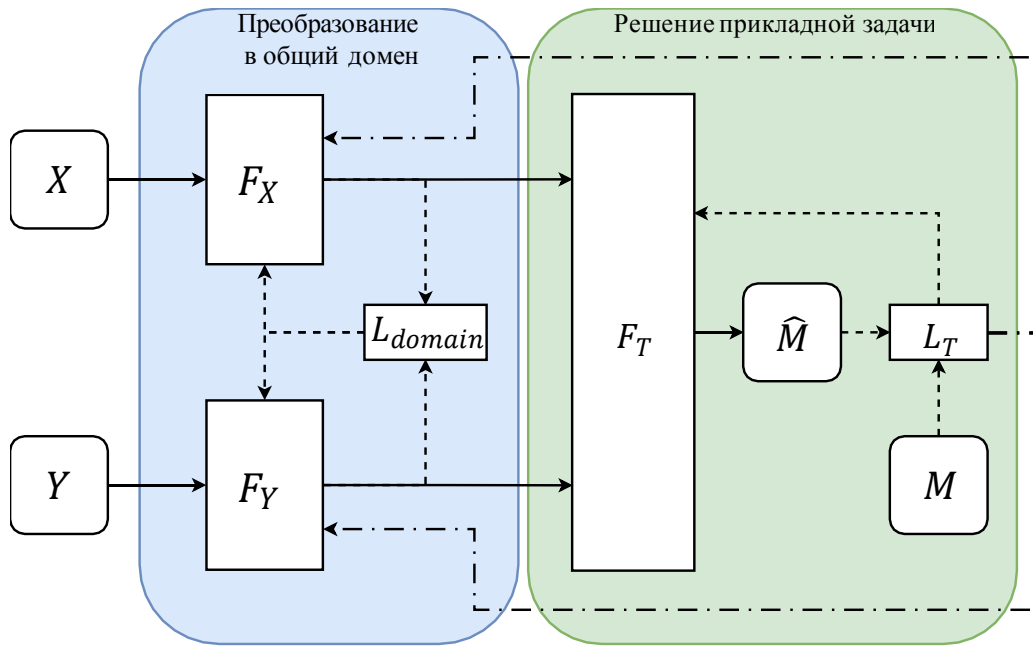


Рисунок 4.1 – Пайплайн обучения ИНС для решения прикладных задач в рамках предлагаемой технологии; пунктирные линии действуют только на этапе обучения ИНС, штрихпунктирная – при совместном дообучении

4.2 Алгоритм классификации земной поверхности по данным разнородных изображений ДЗЗ

4.2.1 Постановка задачи

Задача классификации земной поверхности является одной из самых распространенных задач прикладной обработки снимков ДЗЗ. Данная задача может быть сведена к задаче семантической сегментации изображений.

Задача семантической сегментации изображений состоит в разделении изображения на непересекающиеся части, каждая из которых отмечает пиксели, принадлежащие некоторому заданному классу. Для изображения размерами $W \times H$ пикселей сегментация может быть задана маской сегментации $M \in \mathbb{M}^{W \times H}$, где $\mathbb{M} = \{0, \dots, C - 1\}$, C – общее число классов.

Решением данной задачи является функция $F_{cls}: \mathbb{R}^{W \times H \times K} \rightarrow \mathbb{R}^{W \times H \times C}$, возвращающая оценку маски сегментации, имеющее наименьшее расхождение с

истинной маской. Данная функция отыскивается в ходе решения оптимизационной задачи

$$F_{cls} = \arg \min_{f_{cls} \in \mathcal{F}_{cls}} \mathbb{E}_{I, M \in \mathcal{Z}} L_{cls}(f_{cls}(I), M), \quad (4.2)$$

где $\mathcal{F}_{cls}$ – множество функций, по которому производится поиск, $\mathcal{Z}$ – обучающее множество пар изображений и масок сегментации, $L_{cls}: \mathbb{R}^{H \times W \times C} \times \mathbb{M}^{H \times W} \rightarrow \mathbb{R}$ – функция потерь, измеряющая расхождение между масками. В роли L_{cls} чаще всего выступает перекрестная энтропия.

В случае обучения нейросетей сегментации оценка маски $\hat{M} = F(I)$ обычно принадлежит пространству $[0, 1]^{H \times W \times C}$, или даже $\mathbb{R}^{H \times W \times C}$: оценка принадлежности пикселя к каждому из классов производится независимо. Данный подход позволяет получить более гладкий ландшафт функции потерь, поскольку в данном случае оценка идет не по бинарной шкале "правильный класс/неправильный класс". Переход $\mathbb{R}^{H \times W \times C} \rightarrow \mathbb{M}^{H \times W}$ можно выполнить путем выбора индекса максимального элемента в векторе значений пикселя ($M_{i,j} = \arg \max_k \hat{M}_{i,j,k}$).

Формулировка (4.2) хорошо применима в случае сегментации изображений из одного домена, однако в случае задачи сегментации изображений из разных доменов ее прямое применение затруднительно: в данном случае одной маске соответствуют несколько изображений, значительно отличающихся по свойствам (в первую очередь – по числу каналов). Стандартными подходами в данном случае являются решение (4.2) для каждого из доменов по отдельности (т.е. обучение отдельной нейросети сегментации для каждого домена) или выполнение т.н. слияния изображений (англ. image fusion), в рамках которого из исходных разнородных изображений формируется новое изображение, которое затем подается в сеть сегментации (в самом простом варианте изображения просто конкатенируются по каналному измерению).

Для выполнения сегментации можно использовать предлагаемую технологию обработки изображений ДЗЗ, в рамках которой на первом этапе преобразовывать разнородные изображения к общему домену, а на втором этапе –

выполнять сегментацию в рамках общего домена. Данный подход позволяет выполнять сегментацию изображений из исходно разных доменов с помощью единственной сети сегментации: в рамках данной схемы сегментация изображений из разных доменов выполняется композициями функций $F_X \circ F_{cls}$ и $F_Y \circ F_{cls}$. С точки зрения задачи (4.1) независимая сегментация разнородных изображений выполняется преобразованиями $F_T(X, Y) = \{F_{cls}(F_X(X)), F_{cls}(F_Y(Y))\}$, $L_T(\hat{M}, M) = L_{cls}(\hat{M}_X, M) + L_{cls}(\hat{M}_Y, M)$.

Более формально, пусть дан набор триплетов $\mathcal{Z} = \{(X^i, Y^i, M^i): X^i \in \mathbb{R}^{W \times H \times K_X}, Y^i \in \mathbb{R}^{W \times H \times K_Y}, M^i \in \mathbb{M}^{W \times H}\}_{i=1}^N$, каждый из которых содержит пару координатно-согласованных снимков одной территории, полученных разными сенсорами, и маску семантической сегментации.

Как описано в разделе 4.1, построение преобразований F_X , F_Y и F_{cls} выполняется в 3 этапа.

На первом этапе построения общего домена маска сегментации не используется. Подробно построение общего домена было рассмотрено ранее в главе Глава 2. Далее будем считать, что ИНС, реализующие преобразования F_X и F_Y , уже обучены. Поскольку ИНС, обученные и использованием критерия (2.14) уступают по значениям показателей сходства изображений, и показывают худшие результаты в задаче совмещения изображений (см. раздел 3.2.3.3), в рамках данного раздела рассматриваются только ИНС, полученные с использованием критерия (2.15).

На втором этапе производится обучение ИНС сегментации на изображениях общего домена. Задача (4.2), с учетом полученных на первом этапе преобразований F_X и F_Y , формулируется в виде

$$L_{cls}(\hat{M}_X, \hat{M}_Y, M) = CE(\hat{M}_X, M) + CE(\hat{M}_Y, M), \quad (4.3)$$

$$F_{cls} = \arg \min_{f_{cls} \in \mathcal{F}_{cls}} \frac{1}{N} \sum_{i=1}^N L_{cls}(f_{cls}(F_X(X^i)), f_{cls}(F_Y(Y^i)), M^i), \quad (4.4)$$

где CE – взвешенная перекрестная энтропия, определяемая в виде

$$l(\hat{m}, m, w) = - \sum_{c=0}^{C-1} w_c \ln \frac{\exp(\hat{m}_c)}{\sum_{k=0}^{C-1} \exp(\hat{m}_k)} \mathbf{1}(m = c),$$

$$CE(\hat{M}, M) = \frac{1}{WH} \sum_{i=0}^{W-1} \sum_{j=0}^{H-1} l(\hat{M}_{i,j}, M_{i,j}, w),$$

где $\mathbf{1}$ – индикаторная функция, $w \in \mathbb{R}^C$ – вектор весовых коэффициентов классов. В качестве значений вектора w берутся величины, обратные частотам встречаемости классов в наборе данных.

Наконец, на третьем этапе совместного дообучения решаемая оптимизационная задача принимает вид

$$L_{joint}(f_{cls}, f_X, f_Y, X, Y, M) = \lambda_1 L_{domain}(f_X(X), f_Y(Y), X, Y) + \lambda_2 L_{cls}(f_{cls}(f_X(X)), f_{cls}(f_Y(Y)), M), \quad (4.5)$$

$$(F_X, F_Y, F_{cls}) = \arg \min_{\substack{f_X \in \mathcal{F}_X, \\ f_Y \in \mathcal{F}_Y, \\ f_{cls} \in \mathcal{F}_{cls}}} \frac{1}{N} \sum_{i=1}^N L_{joint}(f_{cls}, f_X, f_Y, X, Y, M), \quad (4.6)$$

где $\lambda_1, \lambda_2 \in R_+$ – весовые коэффициенты.

4.2.2 Экспериментальные исследования

4.2.2.1. Набор данных

В качестве набора данных используется MultiSenGE [136], использовавшийся ранее в предыдущих главах.

При обучении ИНС преобразования использовалось ранее сформированное разбиение на обучающую, валидационную и тестовую выборки из 46 092, 11 522 и 14 405 пар соответственно.

При обучении ИНС классификации (как в общем домене, так и в исходных доменах) набор данных, как было предложено авторами набора данных в [23], ограничивается по времени съемки (июнь-ноябрь), а разделение на выборки

производится по территориальному признаку (вся территория, покрываемая набором данных, разделена на несколько непересекающихся областей, из которых формируются выборки). Итоговые обучающая, валидационная и тестовая выборки состоят из 13 476, 7 644 и 2 440 пар соответственно. Маски семантической сегментации, предоставляемые для каждой локации, содержат 14 классов. Тем не менее, в работе [23] создатели набора предлагают уменьшить количество классов путем слияния некоторых слабо представленных классов в композитные классы с большим количеством примеров. В настоящей работе используется предложенное в [23] разделение по 10 классам. Список классов и их цветовые обозначения приведены на рисунке 4.2.



Рисунок 4.2 – Классы и их цветовые обозначения

4.2.2.2. Обучение ИНС

Обучение нейронных сетей выполнялось на вычислительной машине с ЦП AMD Ryzen 5600X с 64 ГБ ОЗУ и видеокартой Nvidia GeForce RTX 3090. В качестве библиотеки машинного обучения использовалась PyTorch 2.6. Предельное количество эпох равно 128. При обучении использовалась аугментация в виде случайных поворотов на 90/180/270° и отражений по горизонтали/вертикали.

В качестве алгоритма оптимизации использовался Adam [136]. В течение первых 5 эпох темп обучения (англ. learning rate) линейно возрастал от начального до основного, в дальнейшем при отсутствии улучшения среднего значения функции

потерь на валидационной выборке в течение более чем 5 эпох темп обучения уменьшался в 10 раз. Полная остановка обучения производилась при отсутствии улучшений в течение 16 эпох.

В качестве базовой архитектуры всех ИНС (и ИНС преобразования, и ИНС сегментации) используется U-Net [140] с сигмоидой в качестве функции активации последнего слоя. В качестве кодирующей части ИНС используется сеть VGG-16 [155] (по аналогии с [23]).

При обучении ИНС использовались следующие параметры:

- $\alpha = 1$;
- $\beta = 3$;
- $\gamma = 8$;
- $\lambda_1 = 1$;
- $\lambda_2 = 4$;
- начальный темп обучения: 10^{-7} ;
- основной темп обучения: 10^{-4} .

В целях сравнения дополнительно были обучены 2 сегментирующие сети, обученные на радарных и оптических снимках по отдельности (параметры обучения те же).

4.2.2.3. Показатели качества

В качестве показателей качества семантической сегментации используются точность (англ. precision), полнота (англ. recall), F_1 и IoU (intersection over union, коэффициент Жаккара).

В случае бинарной классификации данные показатели определяются следующим образом:

$$Precision = 100 \times \frac{TP}{FP + TP},$$

$$Recall = 100 \times \frac{TP}{FP + FN},$$

$$F_1 = 100 \times \frac{2}{Precision^{-1} + Recall^{-1}},$$

$$IoU = 100 \times \frac{TP}{TP + FP + FN},$$

где TP – количество пикселей, корректно отнесенных к данному классу, FP – количество пикселей, ложно отнесенных к данному классу (ошибка I рода), FN – количество пикселей, принадлежащих к данному классу, но ложно классифицированных, как не принадлежащие к данному классу (ошибка II рода). Домножение на 100 используется для удобства отображения показателей.

В рассматриваемой задаче мультиклассовой классификации данные показатели считаются для каждого класса по отдельности. В качестве итогового значения показателя в данном случае используется среднее, взвешенное по частоте встречаемости классов в тестовой выборке.

В качестве показателей сходства изображений, используемых для оценки качества преобразования в общий домен, используются:

- СКО разности – $RMSE$ (2.9);
- нормализованная взаимная информация – NMI (2.10);
- пиковое соотношение сигнал-шум – $PSNR$ (2.16);
- индекс структурного сходства – $SSIM$ (2.5).

4.2.2.4. Результаты экспериментов

Средние значения показателей качества сходства изображений в общем домене приведены в таблице 4.1. Первая строка таблицы содержит результаты для данных без преобразования в общий домен.

Таблица 4.1 – Значения показателей сходства изображений

Тип данных	Показатель			
	<i>RMSE</i>	<i>PSNR</i>	<i>SSIM</i>	<i>NMI</i>
-	0,36	8,9	0,05	1,025
Общий домен	0,0685	23,5	0,773	1,156
Общий домен после дообучения	0,0689	23,4	0,772	1,157

Из рассматриваемой таблицы видно, что значения большинства показателей сходства незначительно снизились после дообучения. Объяснить это можно тем фактом, что при дообучении ИНС преобразования начинают сохранять большее количество деталей, важных для дальнейшей сегментации, даже если при этом несколько снижается сходство между изображениями.

Средние значения показателей качества сегментации изображений по тестовой выборке приведены в таблице 4.2.

Таблица 4.2 – Значения показателей качества сегментации

Исходный тип данных	Показатель качества			
	Точность	Полнота	F_1	IoU
Исходные домены				
Оптика	88,5	85,0	86,2	78,3
Радар	81,3	78,4	79,3	68,9
Общий домен				
Оптика	84,9	82,8	82,4	73,5
Радар	82,3	79,5	80,2	70,5
Общий домен после дообучения				
Оптика	88,3	86,2	86,8	79,1
Радар	82,0	79,8	80,5	70,7

Как видно, сегментация оптических изображений в общем домене без дообучения проигрывает по всем значениям показателей качества сегментации в рамках исходного домена. Объяснить это можно значительной потерей информации при переходе от 10 каналов к 3 каналам. С этой же точки зрения объяснимо отсутствие потери качества для радарных изображений при переходе к общему домену, т.к. количество каналов увеличивается (2 канала преобразуются в 3 канала), при этом на изображениях подавляется спекл-шум. Стоит еще раз подчеркнуть, что в рамках общего домена и для оптических, и для радарных изображений используется единая нейросеть сегментации.

После дообучения значения показателей качества улучшаются. Отметим, что в данном случае сегментация оптических изображений в общем домене превосходит сегментацию в исходном домене по всем показателям, за исключением точности: по-видимому, необходимую информацию в рамках заданного множества классов удастся закодировать в 3 каналах, незначительно жертвуя при этом сходством изображений.

Значения показателя IoU по каждому из классов приведены в таблице 4.3.

Таблица 4.3 – Значения показателя IoU по классам

Тип данных	Класс									
	1	2	3	4	5	6	7	8	9	10
Исходные домены										
Оптика	29,2	51,3	46,8	9,9	25,7	89,5	68,9	30,1	75,4	21,8
Радар	21,6	40,1	35,9	9,7	19,7	82,1	32,1	22,6	66,7	18,7
Общий домен после дообучения										
Оптика	30,0	50,6	53,0	14,0	29,6	90,3	56,8	33,9	78,2	26,3
Радар	22,3	40,1	37,3	10,5	24,5	83,7	29,8	23,5	71,2	20,9

Из рассматриваемой таблицы видно, что значения показателя IoU выше при сегментации в рамках общего домена для всех классов в случае радарных

изображений и для большинства классов в случае оптических изображений: исключениями являются класс 2 (плотная застройка) и класс 7 (виноградники и сады).

Примеры изображений и результатов их сегментации приведены на рисунке 4.3.

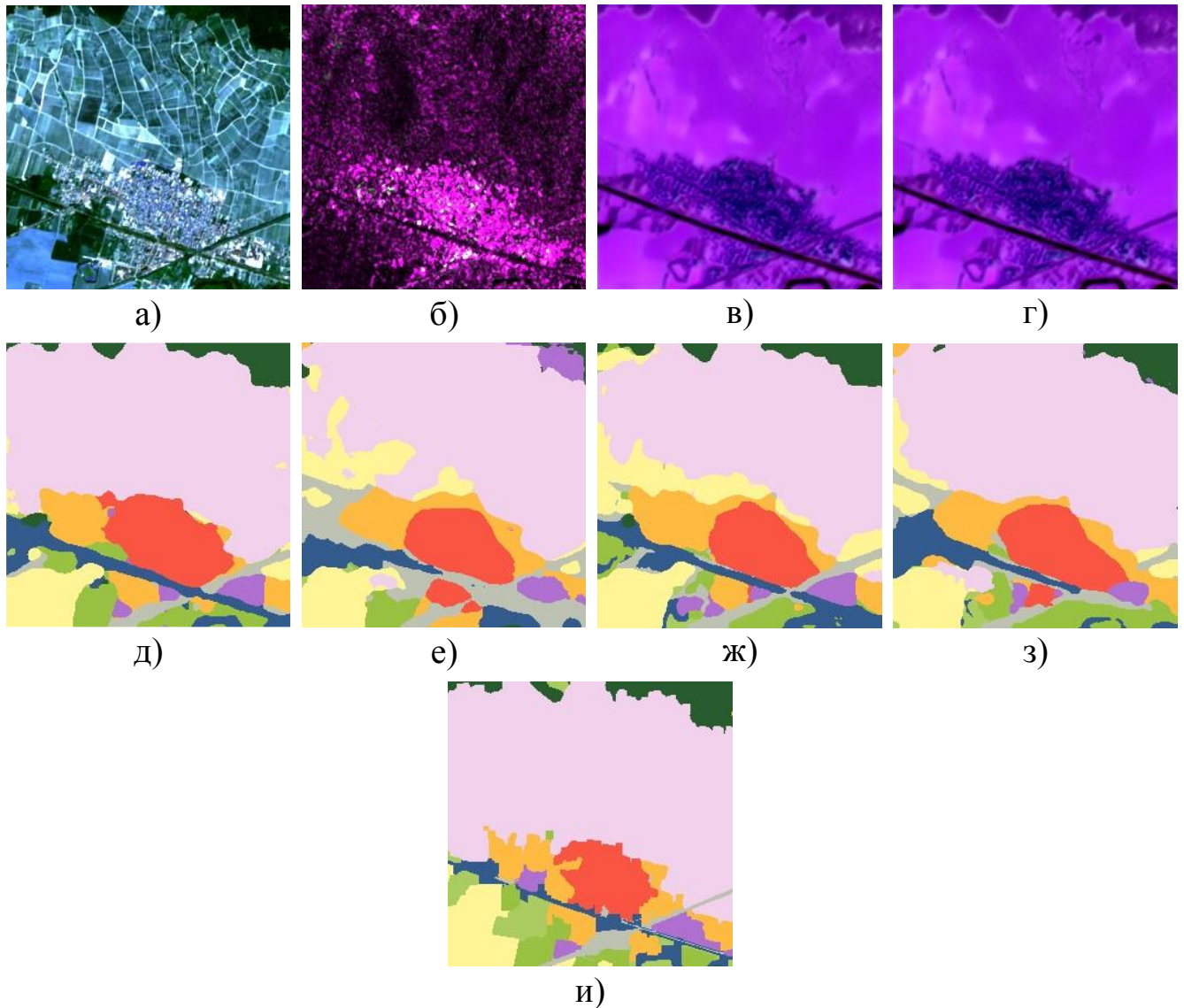


Рисунок 4.3 – Изображения и результаты их сегментации; а, б – изображения исходных доменов; в, г – изображения общего домена после дообучения; д, е – результаты сегментации изображения исходных доменов; ж, з – результаты сегментации изображений в общем домене; и – истинная маска сегментации

4.3 Алгоритм обнаружения изменений по данным разнородных изображений ДЗЗ

4.3.1 Постановка задачи

Задача обнаружения изменений по паре снимков ДЗЗ описывается следующим образом. Пусть даны 2 снимка $X \in \mathbb{R}^{W \times H \times K_X}$, $Y \in \mathbb{R}^{W \times H \times K_Y}$, снятые в разные моменты времени. Необходимо найти области, в которых произошло значимое изменение. Область изменения описывается маской $M \in \{0, 1\}^{W \times H}$, где 1 – наличие изменений, 0 – отсутствие изменений.

Понятие «значимого изменения» зависит от задачи. К примеру, относительно задачи классификации земной поверхности, значимым изменением может считаться изменение класса пикселя. В случае задачи мониторинга стихийных бедствий, под значимым изменением понимается область, затронутая бедствием: области пожаров, затопления или иных разрушений. Примерами незначимых изменений являются сезонные изменения или изменения вследствие отличий в условиях съемки.

В отличие от задачи классификации земной поверхности, рассмотренной в предыдущем разделе, задача обнаружения изменений сводится к задаче бинарной сегментации по данным пары снимков: истинная маска $M \in \{0, 1\}^{W \times H}$ является матрицей, состоящей из 0 и 1, оценка маски изменений $\hat{M} \in [0, 1]^{W \times H}$ является матрицей, элементы которой оценивают степень уверенности в наличии изменений для каждого из пикселей, а преобразование $F_T: \mathbb{R}^{W \times H \times K} \times \mathbb{R}^{W \times H \times K} \rightarrow [0, 1]^{W \times H}$ принимает сразу 2 снимка.

Для обнаружения изменений можно использовать предлагаемую технологию обработки изображений ДЗЗ, в рамках которой на первом этапе преобразовывать разнородные изображения к общему домену, а на втором этапе – выполнять бинарную сегментацию в рамках общего домена [154*]. Однако, в отличие от рассмотренной ранее задачи классификации земной поверхности разнородных

снимков (в рамках которой снимки обрабатывались независимо), задача обнаружения изменений решается на основе пары снимков, как следствие, функция $F_T(F_X(X), F_Y(Y))$ не может быть сведена к паре преобразований $F_X \circ F_T$ и $F_Y \circ F_T$. Тем не менее, как было ранее отмечено в разделе 4.1, применение общего домена позволяет выявлять изменения по паре изображений любых доменов, что является безусловным преимуществом: в стандартном случае $F_T(X', X)$, $F_T(Y', Y)$ и $F_T(Y, X)$ могут вовсе быть невалидными выражениями (к примеру, если $K_X \neq K_Y$, и F_T реализуется в виде ИНС, принимающей на вход тензор $\mathbb{R}^{W \times H \times (K_X + K_Y)}$, то в случае попытки вычисления $F_T(X', X)$ на входе окажется тензор $\mathbb{R}^{W \times H \times 2K_X}$, что приведет к ошибке).

Более формальное описание задачи можно представить следующим образом. Пусть $\mathcal{Z} = \{(X_0^i, X_1^i, Y_0^i, Y_1^i, M^i) | X_*^i \in \mathbb{R}^{W \times H \times K_X}, Y_*^i \in \mathbb{R}^{W \times H \times K_Y}, M^i \in \mathbb{M}^{W \times H}\}_{i=1}^N$ — обучающий набор кортежей, каждый из которых содержит изображения в исходных доменах до и после некоторого события и маску изменений. Из данного набора могут быть сформированы частные наборы $\mathcal{Z}_{XX} = \{X_0^i, X_1^i, M^i\}_{i=1}^N$, $\mathcal{Z}_{YY} = \{Y_0^i, Y_1^i, M^i\}_{i=1}^N$, $\mathcal{Z}_{YX} = \{Y_0^i, X_1^i, M^i\}_{i=1}^N$, $\mathcal{Z}_{XY} = \{X_0^i, Y_1^i, M^i\}_{i=1}^N$. Поскольку каждый из наборов должен обрабатываться своей комбинацией функций F_X/F_Y , введем обозначение $\mathcal{Z} = \{\{F_X, F_X, \mathcal{Z}_{XX}\}, \{F_X, F_Y, \mathcal{Z}_{XY}\}, \{F_Y, F_X, \mathcal{Z}_{YX}\}, \{F_Y, F_Y, \mathcal{Z}_{YY}\}\}$.

В случае задачи обнаружения изменений по снимкам некоторых заданных наперед доменов задача (4.1) может быть представлена в виде

$$F_T = \arg \min_{f_{cls} \in \mathcal{F}_{cls}} \mathbb{E}_{A, B, M \in \mathcal{Z}_{AB}} [L_T(f_T(A, B), M)], \quad (4.7)$$

где $\mathcal{Z}_{AB} \in \{\mathcal{Z}_{XX}, \mathcal{Z}_{YY}, \mathcal{Z}_{YX}, \mathcal{Z}_{XY}\}$, $L_T \equiv CE$ — перекрестная энтропия, взвешенная по частоте встречаемости классов.

В случае использования предлагаемой технологии, с учетом возможности произвольного выбора домена исходных снимков, задача может быть дополнена до вида

$$\mathbb{L}(f_T, f_A, f_B, Z_{AB}) = \mathbb{E}_{A,B,M \in Z_{AB}} [L_T(f_T(f_A(A), f_B(B)), M)],$$

$$F_T = \arg \min_{f_{cls} \in \mathcal{F}_{cls}} \sum_{F_A, F_B, Z_{AB} \in \mathcal{Z}} \mathbb{L}(f_T, F_A, F_B, Z_{AB}). \quad (4.8)$$

На первом этапе ИНС, реализующие преобразования F_X и F_Y обучаются на множестве пар изображений $Z_0 = \{X_0^i, Y_0^i\}_{i=1}^N$ (использование снимков до произошедшего события объясняется большей стабильностью содержимого снимков).

На втором этапе обучение ИНС, реализующей F_T , выполняется согласно (4.8).

Третий этап совместного обучения в рамках задачи обнаружения изменений не рассматривался.

4.3.2 Экспериментальные исследования

4.3.2.1. Набор данных

В качестве набора данных используется набор ImpactMesh-Flood [156], предназначенный для решения задачи обнаружения зон затопления.

По каждой территории, в которой происходило затопление, в данном наборе представлены временные последовательности снимков Sentinel-1, Sentinel-2 и карта высот. В оптимальном случае, по территории представлены 4 пары спутниковых снимков: пара снимков за месяц до затопления, пара снимков перед затоплением, пара снимков во время затопления и пара снимков после затопления. Стоит отметить, что все пар есть в наличии только в оптимальном случае: любой из снимков может быть пропущен. Кроме того, часть оптических снимков страдает от значительной облачности (что типично для зон затопления) – данные снимки были исключены из используемых выборок.

На этапе построения общего домена использовались снимки, полученные за месяц до затопления и перед затоплением, поскольку на данных снимках меньше

различий. Обучающая выборка состояла из 44 896 пар, валидационная – из 8 666 пар, тестовая – из 4 411 пар снимков.

На этапе обучения ИНС классификации использовались пары снимков до затопления/во время затопления. При обучении ИНС классификации в общем домене рассматривались все 4 комбинации снимков (оптика/радар, радар/оптика, радар/радар, оптика/оптика). В качестве альтернатив были обучены ИНС, выполняющие сегментацию в рамках исходных доменов в комбинациях оптика/оптика, оптика/радар и радар/радар. Обучающая выборка состояла из 4902 пар, валидационная – из 481 пар, тестовая – из 258 пар снимков.

4.3.2.2. Обучение ИНС

Обучение нейронных сетей выполнялось на вычислительной машине с ЦП AMD Ryzen 5600X с 64 ГБ ОЗУ и видеокартой Nvidia GeForce RTX 3090. В качестве библиотеки машинного обучения использовалась PyTorch 2.6. Предельное количество эпох равно 128. При обучении использовалась аугментация в виде случайных поворотов на 90/180/270° и отражений по горизонтали/вертикали.

В качестве алгоритма оптимизации использовался AdamW [157]. В течение первых 5 эпох темп обучения (англ. learning rate) линейно возрастал от начального до основного, в дальнейшем при отсутствии улучшения среднего значения функции потерь на валидационной выборке в течение более чем 5 эпох темп обучения уменьшался в 2 раза. Полная остановка обучения производилась при отсутствии улучшений в течение 10 эпох.

В качестве базовой архитектуры всех ИНС (и ИНС преобразования, и ИНС сегментации) используется U-Net [140] с сигмодой в качестве функции активации последнего слоя. В качестве кодирующей части в ИНС преобразования в общий домен используется сеть VGG-19 [156]. В качестве ИНС сегментации используется VGG-16.

На этапе обучения нейросетей, реализующих преобразования к общему домену, использовались следующие параметры:

- $\alpha = 16, \beta = 1, \gamma = 8$;
- начальный темп обучения: 10^{-7} ;
- основной темп обучения: 10^{-5} .

На этапе обучения нейросети сегментации использовались те же значения начального и конечного темпов обучения.

Этап совместного обучения в рамках решения данной задачи не применялся, поскольку обучение ИНС преобразования в общий домен требует использования снимков, снятых в близкие моменты времени, а ИНС обнаружения изменений – снятых в различные моменты времени, что не позволяет построить общий набор данных.

4.3.2.3. Результаты экспериментов

Поскольку задача выявления изменений сводится к задаче бинарной классификации, для оценки качества обнаружения изменений могут использоваться точность, полнота, F_1 и IoU , определенные в подразделе 4.2.2.3.

Средние значения показателей качества на тестовой выборке представлены в таблице 4.4. Примеры обнаружения зон затопления приведены на рисунке 4.4.

Таблица 4.4 – Показатели качества обнаружения зон затопления

Тип данных	<i>Precision</i>	<i>Recall</i>	F_1	<i>IoU</i>
Исходные домены				
Оптика	61,1	81,9	63,6	52,6
Радар	67,9	72,1	69,7	60,0
Общий домен				
Оптика	62,1	84,6	64,9	53,8
Оптика/радар	64,4	78,1	68,1	57,8
Радар/оптика	62,7	84,8	66,0	54,9
Радар	66,9	79,0	70,8	60,5

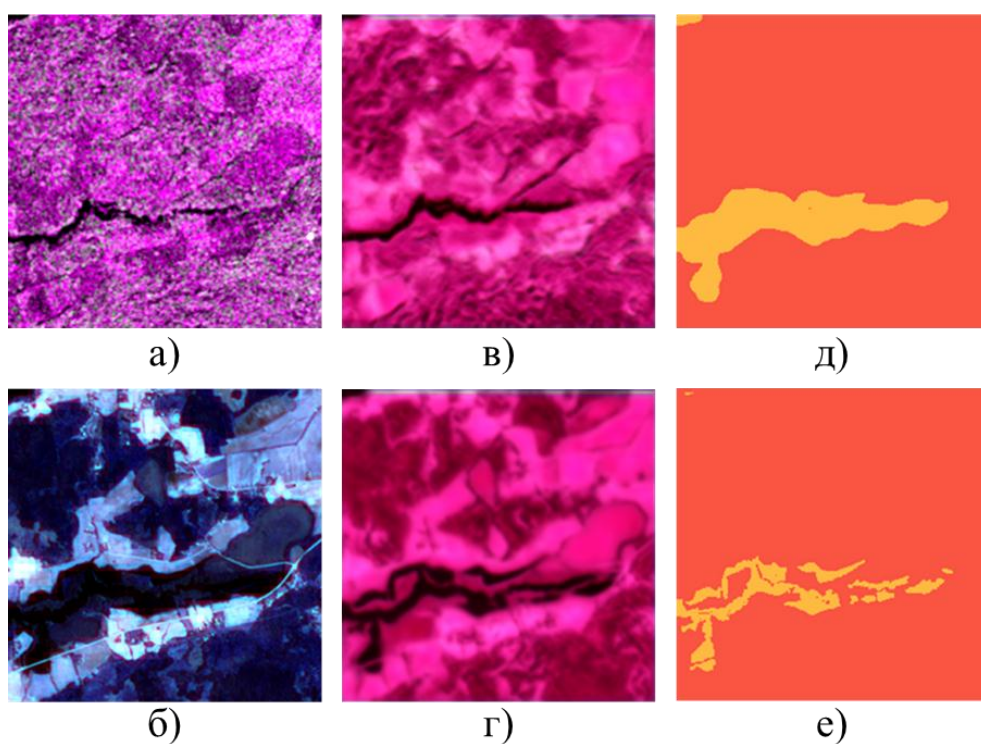


Рисунок 4.4 – Результаты обнаружения зоны затопления; а – радарное изображения до затопления, б – оптическое изображение во время затопления; в, г – образы изображений в общем домене; д – результат обнаружения; е – истинная маска зоны затопления

Полученные результаты демонстрируют успешное решение задачи обнаружения зон затопления с помощью предложенной технологии: по показателю *IoU* достигнуто улучшение с 60% до 60,5% при решении задачи по данным радарных снимков, и с 52,6% до 53,8% при решении задачи по данным оптических снимков.

При этом стоит отметить, что по частному показателю *Precision* лидирует радарный домен, общий домен лидирует по всем остальным показателям, включая агрегативные F_1 и *IoU*. Интересно, что оптический домен проигрывает радарному по всем показателям, за исключением *Recall*, несмотря на большее количество доступной информации. Можно предположить, что информация на снимках оптического домена избыточна для решения задачи, при этом на качестве обнаружения сказывается зависимость оптических снимков от погодных условий.

Лидерство радарного домена по показателю *Precision* в комбинации с проигрышем по показателю *Recall*, предположительно, указывает на то, что при преобразовании к общему домену часть водных областей, которые на изображениях оптического и радарного домена обладали несогласованными характеристиками (в силу неидеальной временной согласованности снимков), преобразовывались в области с некоторыми промежуточными характеристиками. Как следствие, на изображениях общего домена четко выделяемыми являются только однозначно распознаваемые водные поверхности. Утеря части таких неоднозначных поверхностей приводит к небольшому падению показателя *Precision*, но при этом увеличивает значение показателя *Recall*.

Стоит также отметить, что предлагаемая технология позволяет успешно решать задачу без ограничения на исходные домены снимков, при этом качество обнаружения зон затопления по комбинациям <оптический снимок до/радарный снимок после> и <радарный снимок до/оптический снимок после> оказалось выше по сравнению с обнаружением в оптическом домене (преимущество по показателю *IoU* на 5,3% и 2,3% соответственно).

4.4 Выводы и результаты четвертой главы

В четвертой главе предложена технология решения прикладных задач обработки совмещенных разнородных изображений ДЗЗ, использующая представление изображений в общем домене, и разработаны алгоритмы прикладной обработки изображений ДЗЗ в рамках данной технологии.

Проведены экспериментальные исследования по решению задач классификации земной поверхности и обнаружения зон затопления по данным оптических и радарных снимков. В задаче классификации земной поверхности предложенная технология продемонстрировала улучшение качества классификации земной поверхности: по показателю *IoU* достигнуто улучшение с 68,9% до 70,7% при решении задачи по данным радарных снимков, и с 78,3% до 79,1% при решении задачи по данным оптических снимков. В задаче обнаружения зон затопления предложенная технология также продемонстрировала лучшее качество решения: по показателю *IoU* достигнуто улучшение с 60% до 60,5% при решении задачи по данным радарных снимков, и с 52,6% до 53,8% при решении задачи по данным оптических снимков.

Показано, что предлагаемая технология позволяет успешно решать задачу без ограничения на исходные домены снимков.

Заключение

В диссертационной работе разработан метод нейросетевой обработки разнородных изображений ДЗЗ, позволяющий в рамках единой методологии – общего домена – и с улучшенными качественными показателями решать различные прикладные задачи, включая задачу пространственного совмещения разнородных снимков ДЗЗ, семантической сегментации/классификации и определения зон затопления по данным оптических и радарных снимков. Основные результаты диссертационной работы:

1) Проведен анализ текущего состояния исследований в области обработки разнородных снимков. Отмечено появление новых методов, сводящих обработку разнородных изображений к обработке однородных изображений с использованием генеративных ИНС, выполняющих преобразование изображений ДЗЗ одного из доменов (исходного домена) в другой (целевой) домен. Отмечен основной недостаток данного подхода, заключающийся в невозможности полного восстановления особенностей целевого домена по данным исходного домена. Установлена целесообразность разработки нового метода, обладающего преимуществами перед существующими методами. Сформулирована гипотеза, определяющая способ построения искомого решения на основе построения домена взаимной информации данных ДЗЗ – общего домена.

2) Предложена концепция общего домена, конструируемого по данным исходных доменов, включающая требования к общему домену: сохранение линейных размеров изображений при их преобразовании к общему домену; высокое сходство образов разнородных изображений одной территории, получаемых при их преобразовании в общий домен; невырожденность получаемых образов изображений. Разработан метод построения общего домена для разнородных изображений ДЗЗ, а также различные алгоритмы, конкретизирующие его для различных данных ДЗЗ и критериев. Проведены экспериментальные исследования по построению общего домена, получаемых как различными оптическими сенсорами, так и оптическим и радарным сенсорами. Для случая разнородных оптических снимков (разные оптические сенсоры) предложены

четыре алгоритма построения общего домена: аналитический и три нейросетевых. Продемонстрировано улучшение показателя *SSIM* с 0,38 (при сравнении изображений исходных доменов) до 0,87. Для случая построения общего домена оптических и радарных изображений предложено два алгоритма; достигнуто улучшение показателя *SSIM* с 0,03 до 0,75. Показано преимущество предлагаемого метода перед подходом на основе преобразования радарных изображений в оптический домен: предложенный метод превосходит по показателю *SSIM* преобразование ИНС *pix2pix* на 0,25; ИНС *LSGAN* – на 0,54.

3) Разработан алгоритм пространственного совмещения разнородных изображений ДЗЗ, использующий преобразование изображений в общий домен для устранения доменных различий между изображениями и учет данных геопривязки для увеличения точности совмещения изображений. Проведены экспериментальные исследования как по совмещению разнородных оптических изображений, так и по совмещению оптических и радарных изображений. При совмещении разнородных оптических изображений достигнуто снижение ошибки совмещения изображений с 0,74 пикселя при совмещении изображений в исходных доменах до 0,52 пикселей при использовании предлагаемого алгоритма. При совмещении оптических и радарных изображений достигнуто снижение ошибки совмещения изображений с 3,44 пикселей при совмещении изображений исходных доменов до 0,75 пикселей при использовании предлагаемого алгоритма; доля успешно совмещенных изображений возросла с 0,13 до 0,999. Показано преимущество разработанного алгоритма перед подходом на основе преобразования радарных изображений в оптический домен с последующим совмещением: предложенный алгоритм превосходит по точности совмещения преобразование ИНС *pix2pix* на 0,21 пикселя, по доле совмещенных изображений – на 0,019; преобразование ИНС *LSGAN* показало неудовлетворительные результаты.

Предложена технология решения прикладных задач обработки совмещенных разнородных изображений ДЗЗ, использующая представление изображений в общем домене, и разработаны алгоритмы прикладной обработки изображений ДЗЗ в рамках данной технологии. Проведены экспериментальные исследования по

решению задач классификации земной поверхности и обнаружения зон затопления по данным оптических и радарных снимков. В задаче классификации земной поверхности предложенная технология продемонстрировала улучшение качества классификации земной поверхности: по показателю *IoU* достигнуто улучшение с 68,9% до 70,7% при решении задачи по данным радарных снимков, и с 78,3% до 79,1% при решении задачи по данным оптических снимков. В задаче обнаружения зон затопления предложенная технология также продемонстрировала лучшее качество решения: по показателю *IoU* достигнуто улучшение с 60% до 60,5% при решении задачи по данным радарных снимков, и с 52,6% до 53,8% при решении задачи по данным оптических снимков. Показано, что предлагаемая технология позволяет успешно решать задачу без ограничения на исходные домены снимков.

Полученные результаты позволяют утверждать о *подтверждении высказанной в диссертационной работе гипотезы* (см. подраздел 1.4.2) о возможности построения единообразного способа решения конкретных прикладных задач обработки/анализа разнородных данных ДЗЗ с использованием общей (взаимной) содержащейся в них информации, определенной в настоящей работе как *общий домен*.

Решения, разработанные в рамках диссертационной работы, очевидно, не покрывают весь спектр задач, требующих решения при получении, передаче и использовании разнородных данных космического ДЗЗ. Дальнейшим этапом развития темы классификации земной поверхности по разнородным данным ДЗЗ с использованием общего домена является классификация по данным мультитемпоральных последовательностей космических снимков, при этом подход на основе общего домена позволит объединять в единую последовательность снимки, полученные от разных сенсоров. В перспективе предложенный подход на основе общего домена может быть использован для построения новых алгоритмов сжатия для хранения в базах данных космических снимков и мультिवременных покрытиях территорий.

Список литературы

- [1] Kazanskiy, N. A comprehensive review of remote sensing and artificial intelligence integration: advances, applications, and challenges / N. Kazanskiy, R. Khabibullin, A. Nikonorov, S. Khonina // *Sensors*. – 2025. – Vol. 25, no. 19. – Art. 5965.
- [2] Сойфер, В. А. Перспективные информационные технологии дистанционного зондирования земли / В. А. Сойфер, М. В. Гашников, Н. И. Глумов [и др.]; под редакцией В.А. Сойфера. – Самара: Новая техника, 2015. – 255 с. – ISBN 978-5-88940-138-4.
- [3] Mountrakis, G. Support vector machines in remote sensing: a review / G. Mountrakis, J. Im, C. Ogole // *ISPRS Journal of Photogrammetry and Remote Sensing*. – 2011. – Vol. 66, no. 3. – P. 247–259.
- [4] Friedl, M. A. Decision tree classification of land cover from remotely sensed data / M. A. Friedl, C. E. Brodley // *Remote Sensing of Environment*. – 1997. – Vol. 61, no. 3. – P. 399–409.
- [5] Wang, Q. An adaptive multi-splitting multivariate decision tree for multi-class classification applied on high-resolution and hyperspectral remote sensing images / Q. Wang, Z. Zheng, H. Lei, F. Wang, Z. Zhang, X. Zou, F. Nie // *Remote Sensing*. – 2026. – Vol. 18, no. 11. – Art. 1790.
- [6] Belgiu, M. Random forest in remote sensing: a review of applications and future directions / M. Belgiu, L. Drăguț // *ISPRS Journal of Photogrammetry and Remote Sensing*. – 2016. – T. 114. – С. 24–31.
- [7] Варламова, А. А. Информационная технология обработки данных ДЗЗ для оценки ареалов растений / А. А. Варламова, А. Ю. Денисова, В. В. Сергеев // *Компьютерная оптика*. – 2018. – Т. 42, № 5. – С. 864–876.
- [8] Ren, Z. Improved k-means and spectral matching for hyperspectral mineral mapping / Z. Ren, L. Sun, Q. Zhai // *International Journal of Applied Earth Observation and Geoinformation*. – 2020. – Т. 91. – С. 102154.
- [9] Gambardella, C. A principal components analysis-based method for the detection of cannabis plants using representation data by remote sensing / C.

Gambardella, R. Parente, A. Ciambrone, M. Casbarra // *Data*. – 2021. – T. 6, № 10. – C. 108.

[10] Myasnikov, E. V. Hyperspectral image segmentation using dimensionality reduction and classical segmentation approaches / E. V. Myasnikov // *Computer Optics*. – 2017. – Vol. 41, no. 4. – P. 564–572.

[11] Suesse, T. Spatial linear discriminant analysis approaches for remote-sensing classification / T. Suesse, A. Brenning, V. Grupp // *Spatial Statistics*. – 2023. – Vol. 57. – Art. 100775.

[12] Myasnikov, E. V. Nonlinear mapping methods with adjustable computational complexity for hyperspectral image analysis / E. V. Myasnikov // *Proceedings of SPIE – The International Society for Optical Engineering*: 8, Barcelona, 19–21 November 2015. – Barcelona, 2015. – P. 987508.

[13] Myasnikov, E. V. Analysis of approaches to feature space partitioning for nonlinear dimensionality reduction / E. V. Myasnikov // *Pattern Recognition and Image Analysis. Advances in Mathematical Theory and Applications*. – 2016. – Vol. 26, no. 3. – P. 474–482.

[14] Kattenborn, T. Review on Convolutional Neural Networks (CNN) in vegetation remote sensing / T. Kattenborn, J. Leitloff, F. Schiefer, S. Hinz // *ISPRS Journal of Photogrammetry and Remote Sensing*. – 2021. – Vol. 173. – P. 24–49.

[15] Fukushima, K. Neocognitron: A self-organizing neural network model for a mechanism of pattern recognition unaffected by shift in position / K. Fukushima // *Biological Cybernetics*. – 1980. – Vol. 36, no. 4. – P. 193–202.

[16] Hu, M. Hyperspectral anomaly change detection based on autoencoder / M. Hu, C. Wu, L. Zhang, B. Du // *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*. – 2021. – Vol. 14. – P. 3750–3762.

[17] Spiller, D. Wildfire detection using convolutional neural networks and PRISMA hyperspectral imagery: a spatial-spectral analysis / D. Spiller, A. Carbone, S. Amici [et al.] // *Remote Sensing*. – 2023. – Vol. 15, no. 19. – Art. 4855.

[18] Nemni, E. Fully convolutional neural network for rapid flood segmentation in synthetic aperture radar imagery / E. Nemni, J. Bullock, S. Belabbes, L. Bromley // Remote Sensing. – 2020. – Vol. 12, no. 16. – Art. 2532.

[19] Yao, Y. Ship detection in optical remote sensing images based on deep convolutional neural networks / Y. Yao, Z. Jiang, H. Zhang, D. Zhao, B. Cai // Journal of Applied Remote Sensing. – 2017. – Vol. 11, no. 4. – Art. 042611.

[20] Gui, S. Remote sensing object detection in the deep learning era—a review / S. Gui, S. Song, R. Qin, Y. Tang // Remote Sensing. – 2024. – Vol. 16, no. 2. – Art. 327.

[21] Liebel, L. Single-image super resolution for multispectral remote sensing data using convolutional neural networks / L. Liebel, M. Körner // The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences. – 2016. – Vol. 41. – P. 883–890.

[22] Mou, L. Deep recurrent neural networks for hyperspectral image classification / L. Mou, P. Ghamisi, X. X. Zhu // IEEE Transactions on Geoscience and Remote Sensing. – 2017. – Vol. 55, no. 7. – P. 3639–3655.

[23] Mugambi, G. Spatial temporal fusion based features for enhanced remote sensing change detection / G. Mugambi, R. Rimiru, M. Kimwele, C. Muriuki, C. Gakii, E. Mukoya // Scientific Reports. – 2025. – Vol. 15, no. 1. – Art. 34074.

[24] Wenger, R. Multimodal and multitemporal land use/land cover semantic segmentation on Sentinel-1 and Sentinel-2 imagery: an application on a MultiSenGE dataset / R. Wenger, A. Puissant, J. Weber, L. Idoumghar, G. Forestier // Remote Sensing. – 2023. – Vol. 15, no. 1. – Art. 151.

[25] Мясников, Е. В. Снижение размерности гиперспектральных данных с использованием нелинейных автокодировщиков / Е. В. Мясников // Информационные технологии и нанотехнологии (ИТНТ-2020) : Сборник трудов по материалам VI Международной конференции и молодежной школы. В 4-х томах, Самара, 26–29 мая 2020 года / Под редакцией В.В. Мясникова. Том 2. – Самара: Самарский национальный исследовательский университет имени академика С.П. Королева, 2020. – С. 168-174.

- [26] Kamod, T. G. Denoise auto-encoder based speckle reduction for RISAT-1 SAR imagery / T. G. Kamod, P. P. Rege, S. D. Kulkarni // 2021 8th International Conference on Signal Processing and Integrated Networks (SPIN). – IEEE, 2021. – P. 216–221.
- [27] Hu, M. Hyperspectral anomaly change detection based on autoencoder / M. Hu, C. Wu, L. Zhang, B. Du // IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing. – 2021. – Vol. 14. – P. 3750–3762.
- [28] Vaswani, A. Attention is all you need / A. Vaswani, N. Shazeer, N. Parmar, [et al.] // Advances in Neural Information Processing Systems. – 2017. – Vol. 30. – P. 6000–6010.
- [29] Wang, R. Transformers for remote sensing: a systematic review and analysis / R. Wang, L. Ma, G. He, B. A. Johnson, Z. Yan, M. Chang, Y. Liang // Sensors. – 2024. – Vol. 24, no. 11. – Art. 3495.
- [30] Chen, H. Remote sensing image change detection with transformers / H. Chen, Z. Qi, Z. Shi // IEEE Transactions on Geoscience and Remote Sensing. – 2021. – Vol. 60. – P. 1–14.
- [31] Chen, Y. PR-Deformable DETR: DETR for remote sensing object detection / Y. Chen, B. Liu, L. Yuan // IEEE Geoscience and Remote Sensing Letters. – 2024. – Vol. 21. – P. 1–5.
- [32] Lei, S. Transformer-based multistage enhancement for remote sensing image super-resolution / S. Lei, Z. Shi, W. Mo // IEEE Transactions on Geoscience and Remote Sensing. – 2021. – Vol. 60. – P. 1–11.
- [33] Asvadi, A. Multimodal vehicle detection: fusing 3D-LIDAR and color camera data / A. Asvadi, L. Garrote, C. Premebida, P. Peixoto, U. J. Nunes // Pattern Recognition Letters. – 2018. – Vol. 115. – P. 20–29.
- [34] Zhang, H. Image fusion meets deep learning: a survey and perspective / H. Zhang, H. Xu, X. Tian, J. Jiang, J. Ma // Information Fusion. – 2021. – Vol. 76. – P. 323–336.
- [35] Ghassemian, H. A review of remote sensing image fusion methods / H. Ghassemian // Information Fusion. – 2016. – Vol. 32. – P. 75–89.

[36] Li, J. Comparative analysis of pixel-level fusion algorithms and a new high-resolution dataset for SAR and optical image fusion / J. Li, J. Zhang, C. Yang, H. Liu, Y. Zhao, Y. Ye // *Remote Sensing*. – 2023. – Vol. 15, no. 23. – Art. 5514.

[37] Zhang, R. Optical and SAR image fusion: a review of theories, methods, and applications / R. Zhang, Y. Yang, Z. Li, P. Li, H. Wang // *Remote Sensing*. – 2026. – Vol. 18, no. 1. – Art. 73.

[38] Белов, А. М. Классификация изображений ДЗЗ с использованием алгоритма комплексирования данных различных сенсоров / А. М. Белов, А. Ю. Денисова // *Компьютерная оптика*. – 2020. – Т. 44, № 4. – С. 627–635.

[39] Liu, C. A review of optical and SAR image deep feature fusion in semantic segmentation / C. Liu, Y. Sun, Y. Xu, Z. Sun, X. Zhang, L. Lei, G. Kuang // *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*. – 2024. – Vol. 17. – P. 12910–12930.

[40] Dong, S. A multi-level feature fusion network for remote sensing image segmentation / S. Dong, Z. Chen // *Sensors*. – 2021. – Vol. 21, no. 4. – Art. 1267.

[41] Quan, Y. Learning SAR-optical cross modal features for land cover classification / Y. Quan, R. Zhang, J. Li, S. Ji, H. Guo, A. Yu // *Remote Sensing*. – 2024. – Vol. 16, no. 2. – Art. 431.

[42] Papadopoulos, S. Decision fusion at pixel level of multi-band data for land cover classification—a review / S. Papadopoulos, G. Koukiou, V. Anastassopoulos // *Journal of Imaging*. – 2024. – Vol. 10, no. 1. – Art. 15.

[43] Ahmed, M. Dual-modal approach for ship detection: fusing synthetic aperture radar and optical satellite imagery / M. Ahmed, N. El-Sheimy, H. Leung // *Sensors*. – 2025. – Vol. 25, no. 2. – Art. 329.

[44] Huo, C. When remote sensing meets foundation model: a survey and beyond / C. Huo, K. Chen, S. Zhang, Z. Wang, H. Yan, J. Shen [et al.] // *Remote Sensing*. – 2025. – Vol. 17, no. 2. – Art. 179.

[45] Tseng, G. Galileo: learning global & local features of many remote sensing modalities / G. Tseng, A. Fuller, M. Reil [et al.] // *arXiv*. – 2025. – arXiv:2502.09356.

[46] Astruc, G. Omnisat: self-supervised modality fusion for earth observation / G. Astruc, N. Gonthier, C. Mallet, L. Landrieu // European Conference on Computer Vision. – Cham : Springer Nature Switzerland, 2024. – P. 409–427.

[47] Astruc, G. Anysat: one earth observation model for many resolutions, scales, and modalities / G. Astruc, N. Gonthier, C. Mallet, L. Landrieu // 2025 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). – IEEE, 2025. – P. 19530–19540.

[48] Labatie, A. MAESTRO: masked autoencoders for multimodal, multitemporal, and multispectral earth observation data / A. Labatie, M. Vaccaro, N. Lardiere, A. Garioud, N. Gonthier // 2026 IEEE/CVF Winter Conference on Applications of Computer Vision (WACV). – IEEE, 2026. – P. 212–224.

[49] Jakubik, J. Terramind: large-scale generative multimodality for earth observation / J. Jakubik, F. Yang, B. Blumenstiel, E. Scheurer, R. Sedona, S. Maurogiovanni [et al.] // 2025 IEEE/CVF International Conference on Computer Vision (ICCV). – IEEE, 2025. – P. 7383–7394.

[50] Klaus, B. Robot vision / B. Klaus, P. Horn. – Cambridge, MA, USA : MIT Press, 1986.

[51] Wang, Y. An improved Normalized Cross Correlation algorithm for SAR image registration / Y. Wang, Q. Yu, W. Yu // 2012 IEEE International Geoscience and Remote Sensing Symposium. – IEEE, 2012. – P. 2086–2089.

[52] Chen, H. M. Mutual information-based image registration for remote sensing data / H. M. Chen, M. K. Arora, P. K. Varshney // International Journal of Remote Sensing. – 2003. – Vol. 24, no. 18. – P. 3701–3706.

[53] Voronov, S. V. Efficiency Analysis of Information Theoretic Measures in Image Registration / S. V. Voronov, A. G. Tashlinskii // Pattern Recognition and Image Analysis. Advances in Mathematical Theory and Applications. – 2016. – Vol. 26, No. 3. – P. 502–505.

[54] Ташлинский, А. Г. Псевдоградиентный алгоритм оценивания параметров привязки изображений на основе взаимной информации / А. Г.

Ташлинский, С. В. Воронов, А. В. Жукова // Радиотехника. – 2018. – № 6. – С. 14–19.

[55] Fischler, M. A. Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography / M. A. Fischler, R. C. Bolles // Communications of the ACM. – 1981. – Vol. 24, no. 6. – P. 381–395.

[56] Chum, O. Matching with PROSAC–progressive sample consensus / O. Chum, J. Matas // 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05). – IEEE, 2005. – Vol. 1. – P. 220–226.

[57] Lowe, D. G. Object recognition from local scale-invariant features / D. G. Lowe // Proceedings of the Seventh IEEE International Conference on Computer Vision. – IEEE, 1999. – Vol. 2. – P. 1150–1157.

[58] Ke, Y. PCA-SIFT: a more distinctive representation for local image descriptors / Y. Ke, R. Sukthankar // Proceedings of the 2004 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR 2004). – IEEE, 2004. – Vol. 2. – P. II–II.

[59] Dellinger, F. SAR-SIFT: a SIFT-like algorithm for SAR images / F. Dellinger, J. Delon, Y. Gousseau, J. Michel, F. Tupin // IEEE Transactions on Geoscience and Remote Sensing. – 2015. – Vol. 53, no. 1. – P. 453–466.

[60] Bay, H. SURF: speeded up robust features / H. Bay, T. Tuytelaars, L. Van Gool // European Conference on Computer Vision. – Berlin, Heidelberg: Springer Berlin Heidelberg, 2006. – P. 404–417.

[61] Rublee, E. ORB: an efficient alternative to SIFT or SURF / E. Rublee, V. Rabaud, K. Konolige, G. Bradski // 2011 International Conference on Computer Vision. – IEEE, 2011. – P. 2564–2571.

[62] Rosten, E. Machine learning for high-speed corner detection / E. Rosten, T. Drummond // European Conference on Computer Vision. – Berlin, Heidelberg: Springer Berlin Heidelberg, 2006. – P. 430–443.

[63] Yi, K. M. LIFT: learned invariant feature transform / K. M. Yi, E. Trulls, V. Lepetit, P. Fua // European Conference on Computer Vision. – Cham : Springer International Publishing, 2016. – P. 467–483.

- [64] Revaud, J. R2D2: repeatable and reliable detector and descriptor / J. Revaud, P. Weinzaepfel, C. De Souza, M. Humenberger // Proceedings of the 33rd International Conference on Neural Information Processing Systems. – Red Hook, NY, USA : Curran Associates Inc., 2019. – P. 12414–12424.
- [65] DeTone, D. SuperPoint: self-supervised interest point detection and description / D. DeTone, T. Malisiewicz, A. Rabinovich // 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW). – IEEE, 2018. – P. 337–349.
- [66] Tyszkiewicz, M. DISK: learning local features with policy gradient / M. Tyszkiewicz, P. Fua, E. Trulls // Advances in Neural Information Processing Systems. – 2020. – Vol. 33. – P. 14254–14265.
- [67] Zhao, X. ALIKED: a lighter keypoint and descriptor extraction network via deformable transformation / X. Zhao, X. Wu, W. Chen [et al.] // IEEE Transactions on Instrumentation and Measurement. – 2023. – Vol. 72. – P. 1–16.
- [68] Sun, J. LoFTR: detector-free local feature matching with transformers / J. Sun, Z. Shen, Y. Wang, H. Bao, X. Zhou // arXiv. – 2021. – arXiv:2104.00680.
- [69] Rocco, I. Efficient neighbourhood consensus networks via submanifold sparse convolutions / I. Rocco, R. Arandjelović, J. Sivic // European Conference on Computer Vision. – Cham: Springer International Publishing, 2020. – P. 605–621.
- [70] Liu, C. SIFT flow: dense correspondence across scenes and its applications / C. Liu, J. Yuen, A. Torralba // IEEE Transactions on Pattern Analysis and Machine Intelligence. – 2010. – Vol. 33, no. 5. – P. 978–994.
- [71] Edstedt, J. DKM: dense kernelized feature matching for geometry estimation / J. Edstedt, I. Athanasiadis, M. Wadenbäck, M. Felsberg // 2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). – IEEE, 2023. – P. 17765–17775.
- [72] Edstedt, J. RoMa: robust dense feature matching / J. Edstedt, Q. Sun, G. Bökman, M. Wadenbäck, M. Felsberg // 2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). – IEEE, 2024. – P. 19790–19800.

[73] Hel-Or, Y. Matching by tone mapping: photometric invariant template matching / Y. Hel-Or, H. Hel-Or, E. David // *IEEE Transactions on Pattern Analysis and Machine Intelligence*. – 2014. – Vol. 36, no. 2. – P. 317–330.

[74] Stumpf, A. Improved co-registration of Sentinel-2 and Landsat-8 imagery for Earth surface motion measurements / A. Stumpf, D. Michéa, J. P. Malet // *Remote Sensing*. – 2018. – Vol. 10, no. 2. – Art. 160.

[75] Rengarajan, R. Co-registration accuracy between Landsat-8 and Sentinel-2 orthorectified products / R. Rengarajan, M. Choate, M. N. Hasan, A. Denevan // *Remote Sensing of Environment*. – 2024. – Vol. 301. – Art. 113947.

[76] Chen, Z. An improved InSAR image co-registration method for pairs with relatively big distortions or large incoherent areas / Z. Chen, L. Zhang, G. Zhang // *Sensors*. – 2016. – Vol. 16, no. 9. – Art. 1519.

[77] Xiang, Y. OS-SIFT: a robust SIFT-like algorithm for high-resolution optical-to-SAR image registration in suburban areas / Y. Xiang, F. Wang, H. You // *IEEE Transactions on Geoscience and Remote Sensing*. – 2018. – Vol. 56, no. 6. – P. 3078–3090.

[78] Ma, W. Remote sensing image registration with modified SIFT and enhanced feature matching / W. Ma, Z. Wen, Y. Wu, L. Jiao, M. Gong, Y. Zheng, L. Liu // *IEEE Geoscience and Remote Sensing Letters*. – 2017. – Vol. 14, no. 1. – P. 3–7.

[79] Li, J. RIFT: multi-modal image matching based on radiation-variation insensitive feature transform / J. Li, Q. Hu, M. Ai // *IEEE Transactions on Image Processing*. – 2019. – Vol. 29. – P. 3296–3310.

[80] Li, J. RIFT2: speeding-up RIFT with a new rotation-invariance technique / J. Li, P. Shi, Q. Hu, Y. Zhang // *arXiv*. – 2023. – arXiv:2303.00319.

[81] Sommervold, O. A survey on SAR and optical satellite image registration / O. Sommervold, M. Gazzea, R. Arghandeh // *Remote Sensing*. – 2023. – Vol. 15, no. 3. – Art. 850.

[82] Li, Q. Matching between SAR images and optical images based on HOG descriptor / Q. Li, G. Qu, Z. Li // *IET International Radar Conference 2013*. – Stevenage, UK: IET, 2013. – P. 1–4.

[83] Ye, Y. Fast and robust matching for multimodal remote sensing image registration / Y. Ye, L. Bruzzone, J. Shan, F. Bovolo, Q. Zhu // *IEEE Transactions on Geoscience and Remote Sensing*. – 2019. – Vol. 57, no. 11. – P. 9059–9070.

[84] Ye, Y. Robust registration of multimodal remote sensing images based on structural similarity / Y. Ye, J. Shan, L. Bruzzone, L. Shen // *IEEE Transactions on Geoscience and Remote Sensing*. – 2017. – Vol. 55, no. 5. – P. 2941–2958.

[85] Harris, C. A combined corner and edge detector / C. Harris, M. Stephens // *Alvey Vision Conference*. – 1988. – Vol. 15, no. 50. – P. 147–151.

[86] Zhao, J. Intelligent matching method for heterogeneous remote sensing images based on style transfer / J. Zhao, D. Yang, Y. Li, P. Xiao, J. Yang // *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*. – 2022. – Vol. 15. – P. 6723–6731.

[87] Kang, Q. Robust optical and SAR image matching via attention-guided structural encoding and confidence-aware filtering / Q. Kang, J. Zhang, G. Huang, F. Liu // *Remote Sensing*. – 2025. – Vol. 17, no. 14. – Art. 2501.

[88] Chen, J. SAR and optical image registration based on deep learning with co-attention matching module / J. Chen, H. Xie, L. Zhang, J. Hu, H. Jiang, G. Wang // *Remote Sensing*. – 2023. – Vol. 15, no. 15. – Art. 3879.

[89] Liao, Y. Feature matching and position matching between optical and SAR with local deep feature descriptor / Y. Liao, Y. Di, H. Zhou, A. Li, J. Liu, M. Lu, Q. Duan // *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*. – 2021. – Vol. 15. – P. 448–462.

[90] Zhang, Y. PLISA: an Optical–SAR remote sensing image registration method based on pseudo-label learning and interactive spatial attention / Y. Zhang, R. Liu, Z. Zhang, L. Shi, L. Weng, L. Hu // *Remote Sensing*. – 2025. – Vol. 17, no. 21. – Art. 3571.

[91] Masi, G. Pansharpening by convolutional neural networks / G. Masi, D. Cozzolino, L. Verdoliva, G. Scarpa // *Remote Sensing*. – 2016. – Vol. 8, no. 7. – Art. 594.

[92] Коновалов, В. Ф. Единый нейросетевой метод повышения разрешения для разнородных цифровых изображений дистанционного зондирования Земли / В.

Ф. Коновалов, В. В. Мясников, В. В. Сергеев // Компьютерная оптика. – 2024. – Т. 48, № 6. – С. 944-955.

[93] Berra, E. F. Harmonized Landsat and Sentinel-2 Data with Google Earth Engine / E. F. Berra, D. C. Fontana, F. Yin, F. M. Breunig // Remote Sensing. – 2024. – Vol. 16, no. 15. – Art. 2695.

[94] Борисов, А. Н. Методы спектрального согласования разнородных снимков дистанционного зондирования земли / А. Н. Борисов, В. В. Мясников, В. В. Сергеев // Автометрия. – 2025. – Т. 61, № 4. – С. 20-27.

[95] Borisov, A. Methods of Spectral Matching for Remote Sensing Data / A. Borisov // Optical Memory and Neural Networks. – 2025. – Vol. 34, No. S2. – P. S221-S229.

[96] Gupta, V. Image style transfer using convolutional neural networks based on transfer learning / V. Gupta, R. Sadana, S. Moudgil // International Journal of Computational Systems Engineering. – 2019. – Vol. 5, no. 1. – P. 53–60.

[97] Zhu, J. Y. Toward multimodal image-to-image translation / J. Y. Zhu, R. Zhang, D. Pathak [et al.] // Proceedings of the 31st International Conference on Neural Information Processing Systems (NIPS'17). – Red Hook, NY, USA: Curran Associates Inc., 2017. – P. 465–476.

[98] Goodfellow, I. J. Generative adversarial nets / I. J. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, Y. Bengio // Proceedings of the 28th International Conference on Neural Information Processing Systems (NIPS'14). – Vol. 2. – Cambridge, MA, USA: MIT Press, 2014. – P. 2672–2680.

[99] Isola, P. Image-to-image translation with conditional adversarial networks / P. Isola, J. Y. Zhu, T. Zhou, A. A. Efros // 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). – IEEE, 2017. – P. 5967–5976.

[100] Zhu, J. Y. Unpaired image-to-image translation using cycle-consistent adversarial networks / J. Y. Zhu, T. Park, P. Isola, A. A. Efros // 2017 IEEE International Conference on Computer Vision (ICCV). – IEEE, 2017. – P. 2242–2251.

[101] Kim, J. U-GAT-IT: unsupervised generative attentional networks with adaptive layer-instance normalization for image-to-image translation / J. Kim, M. Kim, H. Kang, K. Lee // arXiv. – 2019. – arXiv:1907.10830.

[102] Sohl-Dickstein, J. Deep unsupervised learning using nonequilibrium thermodynamics / J. Sohl-Dickstein, E. Weiss, N. Maheswaranathan, S. Ganguli // International Conference on Machine Learning. – PMLR, 2015. – P. 2256–2265.

[103] Dhariwal, P. Diffusion models beat GANs on image synthesis / P. Dhariwal, A. Nichol // Advances in Neural Information Processing Systems. – 2021. – Vol. 34. – P. 8780–8794.

[104] Kim, S. InstaFormer: instance-aware image-to-image translation with transformer / S. Kim, J. Baek, J. Park, G. Kim, S. Kim // 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). – IEEE, 2022. – P. 18300–18310.

[105] Merkle, N. On the possibility of conditional adversarial networks for multi-sensor image matching / N. Merkle, P. Fischer, S. Auer, R. Müller // 2017 IEEE International Geoscience and Remote Sensing Symposium (IGARSS). – IEEE, 2017. – P. 2633–2636.

[106] Toriya, H. SAR2OPT: image alignment between multi-modal images using generative adversarial networks / H. Toriya, A. Dewan, I. Kitahara // IGARSS 2019 – 2019 IEEE International Geoscience and Remote Sensing Symposium. – IEEE, 2019. – P. 923–926.

[107] Huang, X. SAR and optical image registration method based on improved CycleGAN / X. Huang, L. Wen, J. Ding // 2019 6th Asia-Pacific Conference on Synthetic Aperture Radar (APSAR). – IEEE, 2019. – P. 1–6.

[108] Du, W. L. K-means clustering guided generative adversarial networks for SAR-optical image matching / W. L. Du, Y. Zhou, J. Zhao, X. Tian // IEEE Access. – 2020. – Vol. 8. – P. 217554–217572.

[109] Liaghat, A. OS-CycleGAN: modified CycleGAN-based descriptors for optical and SAR image matching / A. Liaghat, M. S. Helfroush, J. Norouzi, H. Danyali, J. M. Guerrero // IEEE Transactions on Geoscience and Remote Sensing. – 2025. – Vol. 63. – P. 1–18.

[110] Hu, X. GAN-based SAR and optical image translation for wildfire impact assessment using multi-source remote sensing data / X. Hu, P. Zhang, Y. Ban, M. Rahnemoonfar // *Remote Sensing of Environment*. – 2023. – Vol. 289. – Art. 113522.

[111] Amitrano, D. Burned area mapping using SAR and multispectral data integration via generative adversarial networks / D. Amitrano // *IEEE Access*. – 2025. – Vol. 13. – P. 136144–136159.

[112] Wu, B. Optical-to-SAR translation based on CDA-GAN for high-quality training sample generation for ship detection in SAR amplitude images / B. Wu, H. Wang, C. Zhang, J. Chen // *Remote Sensing*. – 2024. – Vol. 16, no. 16. – Art. 3001.

[113] Zhang, H. Seg-CycleGAN: SAR-to-optical image translation guided by a downstream task / H. Zhang, H. Li, J. Lin, Y. Zhang, J. Fan, H. Liu, K. Liu // *IEEE Geoscience and Remote Sensing Letters*. – 2025. – Vol. 22. – P. 1–5.

[114] Meng, Q. Closest spectral fit for removing clouds and cloud shadows / Q. Meng, B. E. Borders, C. J. Cieszewski, M. Madden // *Photogrammetric Engineering & Remote Sensing*. – 2009. – Vol. 75, no. 5. – P. 569–576.

[115] Zhu, X. A modified neighborhood similar pixel interpolator approach for removing thick clouds in Landsat images / X. Zhu, F. Gao, D. Liu, J. Chen // *IEEE Geoscience and Remote Sensing Letters*. – 2011. – Vol. 9, no. 3. – P. 521–525.

[116] Zhang, Q. Missing data reconstruction in remote sensing image with a unified spatial–temporal–spectral deep convolutional neural network / Q. Zhang, Q. Yuan, C. Zeng, X. Li, Y. Wei // *IEEE Transactions on Geoscience and Remote Sensing*. – 2018. – Vol. 56, no. 8. – P. 4274–4288.

[117] Enomoto, K. Filmy cloud removal on satellite imagery with multispectral conditional generative adversarial nets / K. Enomoto, K. Sakurada, W. Wang [et al.] // *2017 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*. – IEEE, 2017. – P. 1533–1541.

[118] Zhang, Q. U-SeqNet: learning spatiotemporal mapping relationships for multimodal multitemporal cloud removal / Q. Zhang, X. Liu, T. Peng [et al.] // *GIScience & Remote Sensing*. – 2024. – Vol. 61, no. 1. – Art. 2330185.

[119] Gao, F. On the blending of the Landsat and MODIS surface reflectance: predicting daily Landsat surface reflectance / F. Gao, J. Masek, M. Schwaller, F. Hall // IEEE Transactions on Geoscience and Remote Sensing. – 2006. – Vol. 44, no. 8. – P. 2207–2218.

[120] Eckardt, R. Removal of optically thick clouds from multi-spectral satellite images using multi-frequency SAR data / R. Eckardt, C. Berger, C. Thiel, C. Schmullius // Remote Sensing. – 2013. – Vol. 5, no. 6. – P. 2973–3006.

[121] Wang, M. IDF-CR: iterative diffusion process for divide-and-conquer cloud removal in remote-sensing images / M. Wang, Y. Song, P. Wei, X. Xian, Y. Shi, L. Lin // IEEE Transactions on Geoscience and Remote Sensing. – 2024. – Vol. 62. – P. 1–14.

[122] Zou, X. DiffCR: a fast conditional diffusion framework for cloud removal from optical satellite images / X. Zou, K. Li, J. Xing, Y. Zhang, S. Wang, L. Jin, P. Tao // IEEE Transactions on Geoscience and Remote Sensing. – 2024. – Vol. 62. – P. 1–14.

[123] Коновалов, В. Ф. Реконструкция фрагментов оптических снимков дистанционного зондирования Земли с использованием изображений радара с синтезированной апертурой и диффузионных нейронных сетей / В. Ф. Коновалов, В. В. Мясников, В. В. Сергеев // Автометрия. – 2025. – Т. 61, № 4. – С. 50-60.

[124] Xia, Y. CRformer: multi-modal data fusion to reconstruct cloud-free optical imagery / Y. Xia, W. He, Q. Huang, G. Yin, W. Liu, H. Zhang // International Journal of Applied Earth Observation and Geoinformation. – 2024. – Vol. 128. – Art. 103793.

[125] Ebel, P. UnCRtainTS: uncertainty quantification for cloud removal in optical satellite time series / P. Ebel, V. S. F. Garnot, M. Schmitt, J. D. Wegner, X. X. Zhu // 2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW). – IEEE, 2023. – P. 2086–2096.

[126] Zhang, J. Quality assessment of SAR-to-optical image translation / J. Zhang, J. Zhou, M. Li, H. Zhou, T. Yu // Remote Sensing. – 2020. – Vol. 12, no. 21. – Art. 3472.

[127] Пытьев, Ю.П. Методы морфологического анализа изображений / Ю.П. Пытьев, А.И. Чуличков. – М.: Физматлит, 2010. – 336 с.

[128] Визильтер, Ю. В. Теория и методы морфологического анализа изображений: диссертация на соискание ученой степени доктора физико-

математических наук: 05.13.17 – теоретические основы информатики / Ю.В. Визильтер. – Самара, 2008.

[129] Визильтер, Ю. В. Морфлеты: новый класс древовидных морфологических описаний формы изображений / Ю. В. Визильтер, В. С. Горбацевич, С. Ю. Желтов [и др.] // Компьютерная оптика. – 2015. – Т. 39, № 1. – С. 101–108.

[130] Ulyanov, D. Instance normalization: the missing ingredient for fast stylization / D. Ulyanov, A. Vedaldi, V. Lempitsky // arXiv. – 2016. – arXiv:1607.08022.

[131] Zhao, H. Loss functions for image restoration with neural networks / H. Zhao, O. Gallo, I. Frosio, J. Kautz // IEEE Transactions on Computational Imaging. – 2016. – Vol. 3, no. 1. – P. 47–57.

[132] Wang, Z. Structural similarity based image quality assessment / Z. Wang, A. C. Bovik, H. R. Sheikh // Digital Video Image Quality and Perceptual Coding. – 2005. – P. 225–242.

[133] 2021 IEEE GRSS Data Fusion Contest Track DSE [Электронный ресурс]. – URL: <https://www.grss-ieee.org/community/technical-committees/2021-ieee-grss-data-fusion-contest-track-dse/> (дата обращения: 10.09.2026).

[134] Studholme, C. An overlap invariant entropy measure of 3D medical image alignment / C. Studholme, D. L. G. Hill, D. J. Hawkes // Pattern Recognition. – 1999. – Vol. 32, no. 1. – P. 71–86.

[135] Maas, A. L. Rectifier nonlinearities improve neural network acoustic models / A. L. Maas, A. Y. Hannun, A. Y. Ng // Proceedings of the 30th International Conference on Machine Learning (ICML). – 2013. – Vol. 28, no. 3. – P. 1–6.

[136] Kingma, D. P. Adam: a method for stochastic optimization / D. P. Kingma, J. Ba // arXiv. – 2014. – arXiv:1412.6980.

[137] Борисов А.Н. Метод построения общего домена радарных и оптических изображений дистанционного зондирования Земли для их совмещения / А.Н. Борисов, В.В. Мясников, В.В. Сергеев // Компьютерная оптика. – 2026. – Т. 50, № 3. – С. 1980.

[138] Kramer, M. A. Autoassociative neural networks / M. A. Kramer // *Computers & Chemical Engineering*. – 1992. – Vol. 16, no. 4. – P. 313–328.

[139] Wenger, R. MultiSenGE: a multimodal and multitemporal benchmark dataset for land use/land cover remote sensing applications / R. Wenger, A. Puissant, J. Weber, L. Idoumghar, G. Forestier // *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*. – 2022. – Vol. 3. – P. 635–640.

[140] Ronneberger, O. U-Net: convolutional networks for biomedical image segmentation / O. Ronneberger, P. Fischer, T. Brox // *International Conference on Medical Image Computing and Computer-Assisted Intervention*. – Cham: Springer International Publishing, 2015. – P. 234–241.

[141] Tan, M. EfficientNet: rethinking model scaling for convolutional neural networks / M. Tan, Q. Le // *International Conference on Machine Learning*. – PMLR, 2019. – P. 6105–6114.

[142] Mao, X. Least squares generative adversarial networks / X. Mao, Q. Li, H. Xie, R. Y. Lau, Z. Wang, S. Paul Smolley // *Proceedings of the IEEE International Conference on Computer Vision*. – IEEE, 2017. – P. 2794–2802.

[143] Iakubovskii, P. Segmentation Models Pytorch [Электронный ресурс] / P. Iakubovskii. – URL: https://github.com/qubvel-org/segmentation_models.pytorch (дата обращения: 10.09.2026).

[144] Dozat, T. Incorporating Nesterov Momentum into Adam [Электронный ресурс] / T. Dozat. – URL: <https://openreview.net/forum?id=OM0jvwB8jIp57ZJjtNEZ> (дата обращения: 10.09.2026).

[145] Smith, L. N. Super-convergence: very fast training of neural networks using large learning rates / L. N. Smith, N. Topin // *Artificial Intelligence and Machine Learning for Multi-Domain Operations Applications*. – SPIE, 2019. – Vol. 11006. – P. 369–386.

[146] Zhu, J.-Y. CycleGAN and pix2pix in PyTorch [Электронный ресурс] / J.-Y. Zhu. – URL: <https://github.com/junyanz/pytorch-CycleGAN-and-pix2pix> (дата обращения: 10.09.2026).

[147] Heusel, M. GANs trained by a two time-scale update rule converge to a local Nash equilibrium / M. Heusel, H. Ramsauer, T. Unterthiner, B. Nessler, S. Hochreiter //

Proceedings of the 31st International Conference on Neural Information Processing Systems (NIPS'17). – Red Hook, NY, USA: Curran Associates Inc., 2017. – P. 6629–6640.

[148] Борисов А.Н. Метод автоматического совмещения разнородных цифровых изображений дистанционного зондирования Земли / А.Н. Борисов, В.В. Мясников, В.В. Сергеев // Компьютерная оптика. – 2024. – Т. 48. № 6. – С. 932-943.

[149] Edstedt, J. DeDoDe v2: analyzing and improving the DeDoDe keypoint detector / J. Edstedt, G. Bökman, Z. Zhao // 2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW). – IEEE, 2024. – P. 4245–4253.

[150] Ye, Y. Improving co-registration for Sentinel-1 SAR and Sentinel-2 optical images / Y. Ye, C. Yang, B. Zhu, L. Zhou, Y. He, H. Jia // Remote Sensing. – 2021. – Vol. 13, no. 5. – Art. 928.

[151] Trémas, T. L. Sentinel-2: presentation of the CAL/VAL commissioning phase / T. L. Trémas, C. Déchoz, S. Lacherade, J. Nosavan, B. Petrucci // Image and Signal Processing for Remote Sensing XXI. – SPIE, 2015. – Vol. 9643. – P. 94–106.

[152] Storey, J. Landsat 8 Operational Land Imager on-orbit geometric calibration and performance / J. Storey, M. Choate, K. Lee // Remote Sensing. – 2014. – Vol. 6, no. 11. – P. 11127–11152.

[153] Schubert, A. Sentinel-1A/B combined product geolocation accuracy / A. Schubert, N. Miranda, D. Geudtner, D. Small // Remote Sensing. – 2017. – Vol. 9, no. 6. – Art. 607.

[154] Борисов, А.Н. Технология обработки разнородных изображений дистанционного зондирования Земли на основе общего домена / А.Н. Борисов // LV Международная научно-практическая конференция "Актуальные проблемы науки и образования в условиях современных вызовов". – Москва, 2026. – С. 329-335.

[155] Simonyan, K. Very deep convolutional networks for large-scale image recognition / K. Simonyan, A. Zisserman // arXiv. – 2014. – arXiv:1409.1556.

[156] ImpactMesh-Flood [Электронный ресурс]. – URL: <https://huggingface.co/datasets/ibm-esa-geospatial/ImpactMesh-Flood> (дата обращения: 10.09.2026).

[157] Loshchilov, I. Decoupled weight decay regularization / I. Loshchilov, F. Hutter // arXiv. – 2017. – arXiv:1711.05101.

ПРИЛОЖЕНИЕ А

Акт использования результатов диссертационного исследования

УТВЕРЖДАЮ

Первый проректор – проректор
по науке Самарского университета

/А.И. Розенцвайг

» октябрь 2026 г.

АКТ

об использовании результатов диссертационного исследования
А.Н. Борисова

Настоящим подтверждаем, что результаты диссертационного исследования по теме «Метод общего домена и его применение в задачах обработки изображений дистанционного зондирования Земли», полученные в Научно-исследовательском институте акустики машин (НИИ-201) Самарского университета, использованы при выполнении научно-исследовательского проекта «Фундаментальные проблемы разработки аэрокосмических транспортных систем и управления в аэрокосмической технике для обеспечения связанности территории Российской Федерации» № 075-15-2024-558 (2024 – 2026 гг.) в рамках государственной программы Российской Федерации «Научно-технологическое развитие Российской Федерации».

Разработанные при выполнении диссертационного исследования методы и алгоритмы использованы при выполнении п. 4.9. «Создание комплекса методов и алгоритмов совмещения, повышения разрешения и реконструкции разнородных и/или разновременных данных ДЗЗ» и п. 4.27 «Создание комплекса методов и алгоритмов тематической интерпретации и понимания данных ДЗЗ с использованием разнородных и/или разновременных данных ДЗЗ» Технического задания на выполнение проекта.

Директор НИИ-201
д.т.н., проф.Ответственный исполнитель темы
д.т.н., проф.

А.Н. Крючков

Г.М. Макарянц

ПРИЛОЖЕНИЕ Б

Акт использования результатов диссертационного исследования
в АО «Самара-Информспутник»

УТВЕРЖДАЮ

Генеральный директор

АО «Самара-Информспутник»,

д.т.н., профессор,

В.В. Сергеев

«1» октября 2026 г.



АКТ

об использовании результатов диссертационного исследования

Борисова Алексея Николаевича

«Метод общего домена и его применение в задачах обработки изображений
дистанционного зондирования Земли»

Настоящим подтверждаем, что результаты диссертационного исследования Борисова А.Н. использованы при выполнении НИР по договору № 98/23 от 13.07.2023 с МИИГАиК (г. Москва) на выполнение НИОКР «Технологии автоматической геопривязки, совмещения и пространственно-временного оперативного анализа данных дистанционного зондирования Земли (ДЗЗ)»

В качестве сотрудника АО «Самара-Информспутник» Борисов А.Н. применил метод, использующий преобразование в общий домен, для разработки механизма пространственного совмещения изображений ДЗЗ.

Заместитель генерального директора,
к.т.н.

Н.И. Глумов

Ведущий инженер-математик,
д.т.н., доцент

А.А. Агафонов