

## ИСПОЛЬЗОВАНИЕ МЕТОДОВ ГЛУБОКОГО ОБУЧЕНИЯ ДЛЯ АНАЛИЗА ИЗОБРАЖЕНИЙ АЭРОКОСМИЧЕСКОГО МОНИТОРИНГА В ИНТЕРЕСАХ ОБНАРУЖЕНИЯ ЭЛЕМЕНТОВ НОВИЗНЫ НА МЕСТНОСТИ

© 2022 Р. Р. Отырба✉, М. А. Дрюченко, А. А. Сирота

*Воронежский государственный университет  
Университетская пл., 1, 394018 Воронеж, Российская Федерация*

**Аннотация.** Работа посвящена разработке математического и программного обеспечения в интересах обнаружения элементов новизны на аэрокосмических снимках городской, пригородной, лесистой местности и акваторий. Для достижения поставленной цели в ходе работы были исследованы традиционные и современные методы обнаружения изменений, а также популярные архитектуры сегментационных нейронных сетей, алгоритмы постобработки и сопоставления изображений. Основной идеей предлагаемого метода и реализуемых на его основе алгоритмов является применение постклассификационного подхода. Он базируется на оценке разности пары бинарных масок, получаемых в результате сегментации анализируемых разновременных изображений с использованием сегментационных нейронных сетей и специализированных алгоритмов постобработки. При этом предлагается реализовать обнаружение элементов новизны в условиях различий геометрических параметров сравниваемых изображений, которые большинство предложенных ранее методов и алгоритмов не учитывают. Такой подход позволяет использовать результирующие бинарные маски для качественного сопоставления нового и ранее отснятого снимка, коррекции перспективы и, в конечном итоге оценки их разности. Кроме этого, в отличие от большинства предшествующих работ, предлагается обнаруживать не только изменённые объекты, но и определять их классы и такие связанные топографические объекты, как: здания, дороги, деревья и водные объекты. Это предоставляет возможность использования метода в приложениях фотограмметрии. Ввиду отсутствия объективных количественных данных для оценки точности обнаружения было осуществлено визуальное тестирование. В результате был сделан вывод о том, что предложенный подход весьма хорошо работает лишь для обнаружения крупных изменений, поскольку реализованный метод обнаружения имеет высокую чувствительность к малейшим изменениям.

**Ключевые слова:** обнаружение элементов новизны, анализ аэрокосмических изображений, глубокие нейронные сети, семантическая сегментация, алгоритмы сопоставления изображений.

### ВВЕДЕНИЕ

Обнаружение изменений, или обнаружение элементов новизны (ОЭН) — это аналитический процесс, целью которого является выявление новых или исчезнувших объектов на изображениях, зарегистрированных в разные моменты времени для одной и той

же целевой области [1]. Иными словами, это процесс поиска измененных и неизменных пикселей, полученных в разное время изображений. Назначение каждому пикселю двоичной метки представляет собой основную цель системы ОЭН, где 1 означает, что область пикселя изменилась, а 0 представляет собой неизменную область.

Обнаружение изменений полезно и важно во многих сферах. Например, в сельскохозяйственном секторе для мониторинга обезлес-

✉ Отырба Ростислав Русланович  
e-mail: [rostislav.otyrba.97@gmail.com](mailto:rostislav.otyrba.97@gmail.com)



Контент доступен под лицензией Creative Commons Attribution 4.0 License.  
The content is available under Creative Commons Attribution 4.0 License.

ния, незаконной вырубки, оценки состояния полей и анализа последствий стихийных бедствий. В военной сфере для разведки перемещения вооруженных сил противника, района поля боя и оценки ущерба. В гражданской сфере для планирования развития городских территорий, обнаружения незаконных построек и так далее. [2, 3].

Классический ручной поиск таких изменений выполняется с использованием инструментальных методов и с непосредственным участием оператора-интерпретатора изображений. При частых изменениях и большом количестве одновременных снимков обнаружение таких изменений вручную будет трудоемкой и неточной задачей.

Обнаружение изменений является довольно важной и актуальной темой исследования и является привлекательной как для частных, так и государственных учреждений. Система ОЭН может значительно уменьшить затраты и повысить оперативность за счет автоматизации и может быть применено для решения различных прикладных задач [4].

К сожалению, задача обнаружения изменений не рассматривалась ранее комплексно в отечественной литературе. В связи с этим представляет интерес решение задачи ОЭН с реализацией расширенных возможностей по отношению к предложенным в зарубежных источниках подходам.

## 1. АНАЛИЗ ПРЕДШЕСТВУЮЩИХ РАБОТ

### 1.1. Традиционные методы и алгоритмы обнаружения изменений

Ранее было предложено множество методов и алгоритмов обнаружения изменений. В зависимости от предметной области традиционные методы обнаружения изменений делятся на две категории: обнаружения на основе пикселей и обнаружения на основе объектов.

В методах ОЭН на основе пикселей основной единицей анализа и сравнения являются спектральные и текстурные характеристики пикселей. Для сравнения изображений типично использовались простейшие математические методы [5]. Алгебраическая разница,

вычисляющая разницу изображений с почти одинаково радиометрическими характеристиками, использовалась, например, в работе [6] для анализа лесных областей. Нормирование изображений, вычисляющее соотношение между парой снимков, использовалась в работе [7] для анализа изменения городского и регионального землепользования. Регрессия изображения по методу наименьших квадратов для коррекции радиометрической составляющей рассматриваемого снимка в соответствии с эталонным и вычисления их итоговой разницы использовалась, например, в работе [8] для анализа изменений почвы и растительности. Векторный анализ изменений (CVA), представляющий разницу значений пикселей в виде разницы векторов признаков спектральных полос, использовался в работе [9] для обнаружения изменений лесных областей. Главная особенность перечисленных методов заключается в необходимости размещения пороговой границы, чтобы отделить измененные пиксели от неизмененных.

Следующий класс методов ОЭН относится к методам преобразования пространства признаков, например, метод главных компонент (РСА) [10]. Эти методы хорошо подходят для анализа лесной и городской местности и направлены на преобразование исходных изображений в подходящее анализируемое пространство признаков, уменьшая избыточность данных в предположении, что области изменений не сильно коррелированы.

Ещё одним классом методов является классификационные методы, которые подразделяются на методы постклассификационного сравнения и машинного обучения. Методы постклассификации неплохо подходят для анализа городской местности [11]. Суть методов заключается в том, что пара одновременных изображения независимо классифицируются и размечаются, и измененные области извлекаются путем прямого сравнения результатов классификации. Классификационные методы машинного обучения, такие как: методы опорных векторов (SVM), деревья принятия решений (Decision trees), ядерная регрессия (Kernel regression) и другие, предлагают, напротив, прямо класси-

фицировать и определять изменения, не имея никаких предположений о распределении данных. Эти методы, например, использовались в работе [12] для обнаружения лесных пожаров и наводнений.

Существует ещё один класс методов, основанных на решении задачи кластеризации. Например, K-means [13] или метод Fuzzy C-means [14] и другие, являющиеся методами обучения без учителя, использовались для анализа земной поверхности на наличие новизны. Они делят разновременные данные на кластеры изменений и кластеры без изменений, не требуя при этом размеченных данных.

Поскольку анализ на основе пикселей игнорирует пространственно-контекстную информацию, то это нередко приводит к результатам с большим количеством шума. Чтобы преодолеть эти недостатки, необходимо учитывать эту информацию для описания пространственных свойств. Для её моделирования в ряде работ было предложено использовать такие методы, как многослойные марковские случайные поля [15], модели и алгоритмы множеств уровня [16], условные случайные поля [17, 18] и модели гиперграфов [19].

Тем не менее с появлением и широким распространением изображений высокого разрешения использование методов обнаружения на основе пикселей стало нецелесообразным из-за их гиперчувствительности к мелким изменениям объектов изображения. Это создало новые проблемы для задачи ОЭН [20]. В связи с этим стали развиваться и предлагаться методы обнаружения изменений по объектам, базирующиеся на глубоком обучении. Их основная цель заключается в извлечении качественных признаков из изображений со сложными сценами и анализе сходства пары разновременных объектов с использованием этих признаков. И действительно, на сегодняшний день статистически подтверждено, что в последние годы наиболее предпочтительной методологией для анализа изображений дистанционного зондирования считаются алгоритмы глубокого обучения [21], в частности свёрточные нейронные сети, для которых существует огромное количество эффективных вариаций архитектуры [22].

## 1.2. Методы обнаружения изменений на основе глубокого обучения

Методы обнаружения изменений на основе глубокого обучения делятся на три категории: методы обнаружения на основе признаков (FB-DLCD), методы обнаружения на основе фрагментов изображения (PB-DLCD) и методы обнаружения на основе изображений (IB-DLCD).

### 1.2.1. Методы на основе признаков

Ранее было выполнено большое количество исследований методов обнаружения на основе признаков, суть которых состоит в использовании глубоких признаков напрямую из различных архитектур свёрточных нейронных сетей. Например, в работе [23] для обнаружения изменений на панорамных снимках улиц авторы предлагают делить изображения на грубые сетки и сравнивать признаки ячейки сетки для пар снимков из пятого слоя пулинга предварительно обученной сети VGG, совместно используя суперпиксельную сегментацию. В работе [24] для обнаружения изменений среди зданий каждый регион изображения подвергался суперпиксельной сегментации с различными масштабами, после чего они использовались в предварительно обученной VGG-16 на ImageNet. Извлечённые признаки разных уровней масштаба сети VGG конкатенировались и сравнивались на наличие несхожести с помощью евклидова расстояния.

Однако очевидно, что более качественные признаки могут быть извлечены, если обучать модели непосредственно на специально подготовленных данных. Например, в работе [25] для обнаружения изменений на аэрофотоснимках земной поверхности использовалась сиамская архитектура CNN с улучшенной функцией потерь триплет для изучения отличительных глубоких признаков между измененными и неизменными пикселями. Вычисление разницы осуществлялось на основе расчета евклидова расстояния глубоких признаков. В случае работы с разнородными данными из разных датчиков, как, например,

в работе [26], для их согласования используется условная генеративно-сопоставительная сеть (сGAN), так как предварительно обученные сети не работают корректно с такими признаками из-за различий в статистических свойствах изображений.

### **1.2.2. Методы на основе фрагментов и изображений**

Методы на основе фрагментов (PB-DLCD) делят входные изображения на фрагменты (или суперпиксели) фиксированного размера, которые используются в моделях глубокого обучения для непосредственного изучения зависимостей и изменений. Методы IB-DLCD на основе изображений отличаются лишь тем, что используют входные изображения целиком. Современные подходы стараются избегать прямого вычисления разницы из-за шумовых эффектов и использовать так называемые end-to-end архитектуры. Это такие архитектуры, которые принимают на вход пару изображений и дают в результате итоговое обнаружение. Они представляют собой полностью свёрточные нейронные сети (FCN), широко применяющиеся в семантической сегментации.

Существует ряд сравнительных исследований, проведённых на эталонном наборе данных CCD [27] от отечественных исследователей, состоящем из разновременных снимков высокого качества одних и тех же участков местности, выгруженных при помощи Google Earth Pro, а также на наборе данных LEVIR-CD работы [28]. При этом задача рассматривается как задача бинарной сегментации. Далее подробно рассматриваются эти методы и алгоритмы, начиная от самых первых моделей, заканчивая самыми новейшими эффективными подходами.

В самой работе [27] была предложена архитектура генеративной сопоставительной сети pix2pix [29], где в качестве генератора сети используется U-Net, а в качестве дискриминатора сеть PatchGAN. Единственным отличием является то, что в качестве входа в сеть поступает пара конкатенированных разновременных изображений.

В работе [30] авторы представили три архитектуры типа IB-DLCD для обнаружения изменений на спутниковых снимках: FC-EF, FC-Siam-conc и FC-Siam-diff. Архитектура FC-EF основана на сети U-Net, состоящая из четырёх кодирующих и декодирующих блока, и принимает на вход конкатенированную пару изображений. Архитектура FC-Siam-conc и FC-Siam-diff являются сиамскими расширениями FC-EF. Разница между архитектурами заключается только в том, как организованы пропускные соединения (skip connections) на этапе декодирования. Для FC-Siam-conc — это конкатенация пары соединений, а для FC-Siam-diff — это конкатенация абсолютной разницы соединений.

Развитием архитектуры FC-EF стала работа [31], предлагающая архитектуру FC-EF-Res типа IB-DLCD. Авторы предлагают интегрировать три FCNN с свёрточными блоками типа residual [32] в единую многозадачную сеть. Две сети имеют общие веса и сегментируют отдельно изображения для получения карты земного покрова. Третья сеть — это аналог FC-EF принимает на вход конкатенированную пару изображений и её задача в выводе карты изменений. Пропускные соединения сегментирующих сетей передаются в соответствующие декодирующие сети и декодирующую сеть третьей сети в виде разностных пропускных соединений как это реализовано в FC-Siam-diff.

В работе [33] авторы представили деконволюционную архитектуру CDNet типа IB-DLCD для задачи обнаружения изменений улиц. CDNet состоит из четырех блоков сжатия (Conv2D + BN + ReLU + MaxPool). И четырех блоков расширения (GuidedPooling + Conv2D + BN + ReLU). Последний слой представляет собой линейный оператор, за которым следует классификатор softmax.

В работе [34] была предложена модифицированная архитектура U-Net++ [35] типа PB-DLCD, которая называется UNet++\_MOSF. U-Net++ имеет один вход и с плотными пропускными соединениями (dense skip-connection) осуществляет получение различных разномасштабных семантических уровней представления признаков. Для улучше-

ния пространственных деталей авторы предложили модуль MSOF, который генерирует 4 различных выходов субдекодеров U-Net++, которые проходят через соответствующие сигмоидальные функции активации, а в конце результаты конкатенируются и проходят через финальную активацию.

В работе [36] авторы предлагают двухветвевую сеть IFN типа PB-DLCD, состоящую из 5 блоков VGG-16, обменивающаяся между собой извлечёнными соответствующими по размерам признаками, а также «Сети разностной дискриминации» (Difference discrimination network). Она является декодирующей сетью, состоящая из четырёх идентичных блоков: модуль пространственного внимания (SAM), повышающая дискретизация и модуль канального внимания (CAM) для фокусировки на релевантной информации по каналам и пространственной информации. Кроме того, чтобы гарантировать качественное обучение сети результаты каждого SAM модуля проходят через сигмоидальные функции активации.

В работе [37] предлагаемый метод использует сиамскую свёрточную нейронную сеть DASNet типа PB-DLCD, где в качестве кодирующих сетей выступают VGG16 или ResNet50, признаки которых поступают в модули двойного внимания (SAM и CAM). После чего авторы отображают результаты в пространство признаков и используют модификацию контрастной функции ошибки WDMC, которая измеряет расстояние между векторами пикселей. Затем происходит объединение векторов для изменённых областей, то есть для тех, которые имеют большее расстояние, после чего их подвергают повышающей билатеральной дискретизации.

В работе [38] авторы предлагают сиамскую сеть SNUNet-CD типа PB-DLCD, базирующуюся на архитектуре U-Net++, аналогично, как и в работе [34], где идентичные связанные кодирующие ветви состоят из 5 свёрточных блоков типа residual [32]. SNUNet-CD также, как и U-Net++ имеет четыре различных субдекодеров с выходами того же размера, как и входное изображение. Их данные используются в модуле канального внимания (CAM) идентичный по структуре как в работе [36],

но реализован в виде ансамбля и учитывает внутригрупповые и межгрупповые отношения результатов. Итоговый результат проходит через операцию свёртки и сигмоидальную функцию активации, получая итоговую карту изменений.

В совсем недавней работе [39], авторы предлагают архитектуру RDP-Net типа PB-DLCD, решающую проблему большинства архитектур, которые фокусируются лишь на глобальном контексте и отбрасывают локальную информацию посредством операций пулинга. Для этого используется слой разделения регионов (Region division layer), делящий свёрткой входные изображения на фрагменты. Далее следуют последовательно блоки ConvMixer (Depthwise convolution + Residual connection + Pointwise convolution). Каждый выход блока дополнительно поступает в слой слияния регионов (Region composition layer), объединяющий обратно разделённые фрагменты с помощью транспонированной свёртки, который в результате предоставляет классификационную карту на уровне пикселей. Далее полученные карты поступают в модуль внимания по глубине (Depth attention module), чтобы объединить несколько релевантных выходных каналов и получить результат обнаружения изменений.

## 2. ПРЕДЛАГАЕМЫЙ МЕТОД ОБНАРУЖЕНИЯ ИЗМЕНЕНИЙ

Основной идеей предлагаемого в данной работе метода ОЭН является применение постклассификационного подхода, базирующегося на определении разности пары бинарных масок, получаемых в результате сегментации анализируемых разновременных изображений с помощью глубоких нейронных сетей и специализированных алгоритмов постобработки. При этом предлагается реализовать ОЭН в условиях различия в геометрических аспектах (ракурс, высота и место съёмки) разновременных изображений, которые большинство предложенных ранее работ игнорируют. Для обеспечения инвариантности к вариации геометрических факторов съёмки изображений применяются алгоритмы сопо-

ставления изображений и GPS-метаданные. Таким образом, предлагаемый метод может быть применён к любой паре аэрокосмических снимков.

Кроме этого, в отличие от большинства предшествующих работ, предлагается обнаруживать не только изменённые объекты, но и определять их классы как топографические объекты: здания, дороги, деревья и водоёмы, что предоставляет возможность использования метода в приложениях фотограмметрии.

Общая схема обработки информации в соответствии с предлагаемым методом обнаружения имеет следующий вид (рис. 1).

1. На вход подаётся изображение (снимок) местности, которое анализируется на предмет содержания возможных изменений.

2. Выполняется предобработка изображения (формирование гистограммы изображений для качественной сегментации, сопоставления и его нарезка на фрагменты).

3. Производится сегментация нарезанных фрагментов ранее обученной глубокой нейронной сетью и обратная их сшивка.

4. Результирующая бинарная маска подвергается постобработке (очистка от шума, упрощение геометрии и т. п.).

5. Подбирается ранее отснятое изображение этой местности и его соответствующая готовая бинарная маска с целевой областью, используя GPS-метаданные.

6. Далее этот снимок и его маска сшиваются с прилегающими по координатам снимками и масками, чтобы увеличить целевую об-

ласть и учесть регион нового снимка с иным ракурсом или высотой съёмки.

7. Выполняется коррекция перспективы нового снимка по отношению к полученному ранее отснятому снимку посредством вычисления матрицы гомографии.

8. Выполняется вычисление разности двух бинарных масок, фильтрация шумов и отображение непосредственно изменений.

Основная причина использования постклассификационного подхода вместо end-to-end архитектур объясняется отсутствием на текущий момент обучающей мультимасочной выборки с разновременными изображениями, а также необходимостью использовать бинарные маски не только для вычисления их разности, но и для качественного сопоставления изображений.

## 2.1. Семантическая сегментация аэрофотоснимков и алгоритмы постобработки

При анализе аэрофотоснимков исследователи могут интересоваться не все компоненты изображения, а только некоторые области или объекты, обладающие определенными характеристиками, относящиеся к рассматриваемой задаче. Сегментация изображения — один из лучших методов решения этой проблемы, цель которой точно сегментировать каждый интересующий объект на изображении, указывая класс принадлежности и информацию о их расположении.

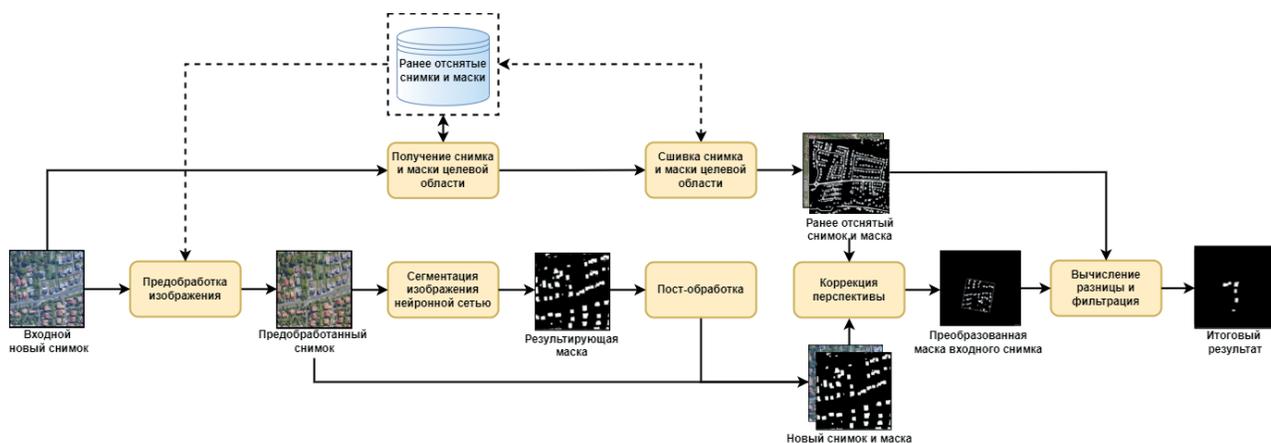


Рис. 1. Схема обработки информации предлагаемого метода ОЭН  
[Fig. 1. Information processing diagram of the proposed method for detecting changes]

В последние годы многие задачи семантической сегментации со сложными сценами решаются с использованием методов глубокого обучения, чаще всего свёрточных нейронных сетей, которые значительно превосходят другие подходы с точки зрения точности и эффективности. В данной работе предлагается использовать архитектуры сегментационных нейронных сетей, что позволяет использовать результирующие бинарные маски для качественного сопоставления нового и ранее отснятого снимка, коррекции перспективы и в конечном итоге для их разности и выявления элементов новизны.

### 2.1.1. Набор данных для обучения

Прежде чем приступать к реализации поставленной цели, было решено создать собственные наборы данных посредством ручной разметки снимков высокого разрешения (6792×4800) Google Earth Pro в онлайн сервисе CVAT, ввиду отсутствия открытых качественных наборов с множеством топологических объектов. Примеры созданных данных представлены на рис. 2. Для обучения сети была произведена нарезка снимков на фрагменты размером 512×512 с перекрытием на 100 пикселей для учёта контекста. Итоговый состав набора данных следующий:

- здания и дороги: 10 снимков (~2000 фрагментов);
- растительность и вода: 5 снимков (~1000 фрагментов).

Для лучшего качества обучения, была применена аугментация, которая происходит на лету во время обучения. В нее входили следующие преобразования: горизонтальные и вертикальные отражения, случайные повороты на 90 градусов, масштабирование с слу-

чайными поворотами, уменьшение контрастности (от 5 % до 20 %), изменение цветового тона и насыщенности (до 20 %).

Для корректного обучения растровые данные были нормализованы к диапазону от 0-1, а к размеченным данным было применено One-Hot кодирование. Все данные были разделены на обучающие и проверочные выборки (90 % и 10 % данных соответственно).

### 2.1.2. Архитектуры сегментационных нейронных сетей

Типичная архитектура сегментационной нейронной сети в целом рассматривается как совокупность кодирующей сети и следующей после неё декодирующей сети. Задача кодирующей сети заключается в извлечении признаков, которые часто представляют собой предварительно обученную классификационную нейронную сеть. Задача декодирующей сети состоит в семантическом проецировании извлекаемых признаков на пространство пикселей, чтобы получить итоговую классификационную карту.

На сегодняшний день существуют десятки архитектур сегментационных сетей. В наших исследованиях в плане сравнительного анализа были рассмотрены следующие.

1. Fully Convolutional Network (FCN) [40] — первая архитектура вида кодер-декодер, давшая толчок в развитии сетей семантической сегментации. FCN использует вместо полностью связанного классификационного слоя билинейную интерполяцию, восстанавливающая карты признаков до исходного размера входного изображения для получения классификационной карты.

2. U-Net [41] — архитектура, изначально разработанная для работы с медицинскими



Рис. 2. Примеры растровых и размеченных снимков  
[Fig. 2. Examples of raster and labeled images]

изображениями, основана на сети FCN и имеет симметрическую структуру кодирующей и декодирующей сети. Сеть оснащена пропускными соединениями, предоставляющие локальную информацию в декодирующую сеть.

3. SegNet [42] — архитектура, использующая в качестве кодирующей сети 13 слоёв VGG16, симметричную ей структуру декодера и пропускные соединения аналогично U-Net. Декодер использует индексы вычисленные на этапе понижения дискретизации функции максимального пулинга.

4. Pyramid Scene Parsing Network (PSPNet) [43] — архитектура, использующая в качестве кодирующей сети ResNet-101 и модуль пирамидального пулинга, который состоит из четырёх операций свёрток различного масштаба, что помогает уточнить общее глобальное представление признаков.

5. Feature Pyramid Network (FPN) [44] — использующая в качестве кодирующей сети ResNet-101. Декодирующая сеть для повышения дискретизации использует интерполяцию по ближайшим соседям. Использует информацию из кодирующей сети по латеральным соединениям путём поэлементного сложения. Дополнительно результат каждого сложения проходит через слой свёртки, и, таким образом, формируется четыре классификационные карты разных масштабов.

6. DeepLabV3+ [45] — архитектура, использующая в качестве кодирующей сети модификацию сети Xception с дополнительными разреженными свёртками. С учетом PSPNet в ней также был реализован модуль ASPP, использующий несколько разреженных свёрток разных масштабов и глобальный усредняющий пулинг для лучшего уточнения общего глобального представления признаков.

7. Dual Attention Network (DANet) [46] — архитектура, использующая в качестве кодирующей сети ResNet-101, а также применяет модули пространственного и канального внимания для взвешивания каналов и пикселей, чтобы сосредоточиться на более релевантной информации.

8. Attentional Class Feature Network (ACFNet) [47] — архитектура, использующая в качестве кодирующей сети обычную свёрточную сеть

для получения высокоуровневых признаков и грубой карты сегментации, которые поступают в модуль ACF, извлекающий глобальный контекст с категориальной точки зрения.

9. Spatial Object-Contextual Representations (OCRNet) [48] — архитектура использующая модуль OCR, который извлекает глобальный контекст с категориальной точки зрения и выполняет взвешенную агрегацию представлений областей объекта с весами, рассчитанными в соответствии с отношениями между пикселями и областями объекта.

10. Pix2Pix (cGAN) [29] — архитектура, представляющая собой совокупность двух соперничающих сетей: генератора и дискриминатора. Генератором является сеть типа U-Net, генерирующая нужный выход. Дискриминатор представляет собой классификационную свёрточную сеть PatchGAN и измеряет сходство между целевым изображением и выходным изображением генератора и пытается угадать, создано ли оно генератором.

### 2.1.3. Обучение и сравнительный анализ нейронных сетей

После создания набора данных был осуществлён сравнительный эксперимент всех рассмотренных ранее архитектур. Затем, после нахождения наилучшей модели, рассматривались и сравнивались классификационные свёрточные нейронные сети, предварительно обученные на наборе данных ImageNet, использовавшиеся в качестве кодирующих сетей. Были рассмотрены такие сети, как: VGG19, ResNet (101v2 и 152v2), DenseNet201 и EfficientNet (B3, B5 и B7).

В качестве оптимизатора использовался Момент Нестерова. Дополнительно, чтобы избежать увязания в процессе обучения, была реализована схема отжига скорости обучения (формула 1), где максимальная скорость равна 0.2, а минимальная 0.

$$\eta_t = \eta_{\min}^i + \frac{1}{2}(\eta_{\max}^i - \eta_{\min}^i) \cdot \left( 1 + \cos \left( \frac{T_{\text{cur}}}{T_i} \pi \right) \right), \quad (1)$$

где  $\eta_{\min}$  — минимальная скорость обучения,  $\eta_{\max}$  — максимальная скорость обучения,  $T_{\text{cur}}$  — количество выполненных эпох.

В качестве метрики используется средняя взвешенная степень пересечения изображений:

$$IoU = \frac{1}{N} \cdot \sum_{c=0}^C \omega(c) \cdot \frac{I(c)}{U(c)}, \quad (2)$$

где  $\omega(c)$  — весовой коэффициент класса,  $I(c)$  — область пересечения,  $U(c)$  — область объединения.

В качестве функции ошибки была использована комбинация взвешенной функции Дайса и Фокальной ошибки (формула 3), где функция Дайса — степень соответствия между предсказанием и истиной (формула 4), а фокальная ошибка — это модификация кросс-энтропии, занижающая вклад преобладающего класса и фокусируется на более сложных примерах (формула 5).

$$L = L_{Dice} + \lambda L_{Focal} \quad (3)$$

$$L_{Dice} = 1 - \frac{1}{N} \cdot \sum_{c=0}^C \omega(c) \cdot \frac{((1 + \beta) + TP_p(c))}{((1 + \beta) \cdot TP_p(c) + \beta^2 \cdot FN_p(c) + FP_p(c))} \quad (4)$$

$$L_{Focal} = -\frac{1}{N} \sum_{c=0}^C \sum_{n=1}^N g_n(c) \times (1 - p_n(c))^2 \cdot \log(p_n(c)), \quad (5)$$

где  $\lambda$  — компромисс между функциями (0.5),  $TP$  — положительный класс распознанный как положительный,  $FN$  — отрицательный класс распознанный как отрицательный,  $FP$  — отрицательный класс распознанный как положительный.

Результаты обучения архитектур на наборе данных «Здания и дороги» после 200 эпох обучения представлены в табл. 1 и на данных «Растительность и вода» в табл. 2.

Для набора данных «Здания и дороги» самыми лучшими оказались архитектуры DeepLabV3+ и OCRNet с кодирующей сетью DenseNet201. Для набора данных «Растительность и вода» было решено протестировать только лучшие модели, а также Pix2Pix.

В итоге было решено далее использовать DeepLabV3+ и OCRNet в качестве ансамбля, так как в результате экспериментов оказалось, что они хорошо дополняют друг друга.

Таблица 1. Результаты обучения на наборе данных «Здания и дороги»

[Table 1. Training results on the «Buildings and roads» dataset]

Модель	Loss	IoU
FCN	0.2993	0.8208
U-Net	0.3701	0.6571
SegNet	0.3547	0.6680
PSPNet	0.3575	0.6565
FPN	0.3126	0.7527
DeepLabV3+	0.2269	<b>0.8582</b>
DANet	0.2907	0.7636
ACFNet	0.2744	0.8010
OCRNet	0.2195	<b>0.8806</b>
Pix2Pix	0.2746	0.7621

Таблица 2. Результаты обучения на наборе данных «Растительность и вода»

[Table 2. Training results on the «Vegetation and water area» dataset]

Модель	Loss	IoU
DeepLabV3+	0.3047	<b>0.7576</b>
OCRNet	0.2972	<b>0.7789</b>
Pix2Pix	0.3186	0.7458

#### 2.1.4. Алгоритмы постобработки

На практике результаты сегментации часто могут содержать шум и нечёткость, которые необходимо обработать для получения качественного результата. Существуют некоторые эффективные алгоритмы постобработки результирующей сегментации, помогающие решить данную проблему.

1. Полносвязные условные случайные поля (Dense CRF) [49] — метод статистического моделирования, принимающий во внимание взаимосвязь между пикселями. При этом пиксели грубой бинарной маски являются набором возможных состояний. Затраты на присвоение состояния одному пикселю называются унарными затратами. Для моделирования связей между пикселями используются попарные затраты между всеми парами пикселей. Сумма унарных и попарных затрат на все пиксели носит название энергии,

значение которой минимизируется для получения качественного результата.

2. Рекуррентные условные случайные поля (RNN-CRF) [50] — метод статистического моделирования, являющийся модификацией Dense CRF, который используется в виде рекуррентного дополнительного обучаемого слоя в конце уже обученной модели нейронной сети с замороженными весами.

3. Сегментационная модель рафинирования CascadePSP [51] — архитектура типа кодер-декодер. Модель принимает на вход изображение и его бинарную маску с тремя различными уровнями детализации и выдаёт три выхода с различным масштабом из блоков декодирующей сети. В модели используется ResNet-101 в качестве кодирующей сети, пропускные соединения, а также пирамидальный пулинг, для уточнения общего глобального представления признаков.

### 2.1.5. Сравнительный анализ алгоритмов постобработки

В данной работе был осуществлён сравнительный эксперимент ранее рассмотренных алгоритмов постобработки. Кроме этого, была дополнительно реализована ручная обработка для устранения незначительных артефактов после постобработки и упрощения геометрии объектов, имеющая следующую последовательность шагов.

1. Метод пороговой обработки Оцу.
2. Медианная фильтрация.

3. Поиск контуров и очистка от лишних объектов.

4. Морфологические операции (открытие, закрытие, реконструкция).

5. Модификация алгоритма упрощения геометрии объектов Дугласа-Пекера, сохраняющий топологию объекта.

6. Медианная фильтрация в цикле.

На рис. 3 представлено визуальное сравнение алгоритмов постобработки. В табл. 3 представлена количественная оценка.

Таблица 3. Результаты постобработки  
[Table 3. Results of post-processing]

Модель	Loss	IoU	Время
DeepLabV3+	0.2166	0.8929	17 с.
OCRNet	0.2139	0.9102	20 с.
Ensemble	0.2125	0.9163	37 с.
DeepLabV3+ & RNN-CRF	0.2168	0.8992	312 с.
Ensemble & FC-CRF	0.2215	0.8791	125 с.
Ensemble & CascadePSP	0.2105	0.9248	221 с.
Ensemble & CascadePSP & Ручная обработка	0.2093	0.9301	258 с.

Таким образом, постобработка с CascadePSP даёт наиболее эффективный прирост к общему качеству сегментации, поскольку условные случайные поля придерживаются низкоуровневых цветовых границ без

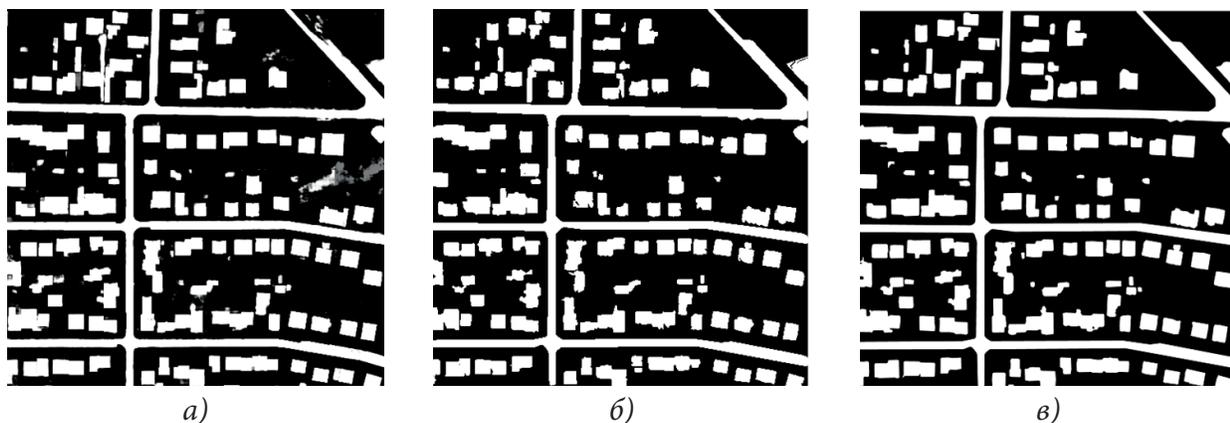


Рис. 3. Примеры постобработки: а) исходное изображение; б) RNN-CRF; в) CascadePSP  
[Fig. 3. Post-processing examples: a) original image; b) RNN-CRF; c) CascadePSP]

полного использования семантической информации высокого уровня и не всегда могут исправить большие области ошибок.

## 2.2. Обнаружение элементов новизны

### 2.2.1. Получение ранее отснятого снимка

Как отмечено ранее, предлагается реализовать метод ОЭН с инвариантностью к ракурсу и масштабу снятых снимков. Предполагается, что уже подготовлена база данных, содержащая ранее отснятые снимки и их предварительно обработанные бинарные маски целевой области. Для каждого нового снимка для получения ранее отснятого снимка целевой области используются GPS-метаданные, с помощью которых находится наиболее кратчайшее расстояние до координаты целевой области посредством вычисления геодезического расстояния. В основе этого лежит алгоритм Карнея, где используется эллипсоидальная модель Земли WGS-84, которая является наиболее точной в глобальном масштабе моделью. Для вычисления этого расстояния использовался пакет georu.

Когда ближайший ранее отснятый снимок найден, необходимо произвести сшивку ближайших снимков вокруг его области, в том числе соответствующих бинарных масок посредством вычисления и использования матрицы гомографии для того, чтобы увеличить целевую область и учесть все возможные регионы нового снимка с иным масштабом или углом.

Гомография представляет собой матрицу  $H$  размера  $3 \times 3$ , которая может преобразовывать плоскость  $A$  в плоскость  $B$  в трехмерном пространстве, путём отображения точек одного изображения в точки соответствия другого изображения. Для расчёта гомографии необходимо по крайней мере 4 и более таких точек соответствия. Для их нахождения используются:

- алгоритмы выявления локальных признаков изображения;
- методы сопоставления признаков;
- методы очистки выбросов.

Наиболее популярными алгоритмами выявления признаков являются следующие.

1. Oriented FAST and Rotated BRIEF (ORB) [52] — алгоритм, состоящий из детектора FAST и дескриптора BRIEF. FAST обнаруживает углы в каждом слое многомасштабной пирамиды, а угловатость точек оценивается с помощью показателя Harris Corner для получения качественных ключевых точек. Дескриптор BRIEF был модифицирован для устойчивости к поворотам.

2. KAZE [53] — алгоритм, детектор которого основан на детерминанте матрицы Гессе, нормализованном по масштабу, который вычисляется в различных масштабах. Максимумы отклика детектора фиксируются в виде характерных точек с использованием скользящего окна. Дескриптор для инвариантности к поворотам находит доминирующую ориентацию в круговой окрестности вокруг каждого обнаруженного признака.

3. Accelerated-KAZE (AKAZE) [54] — модификация алгоритма KAZE, детектор которого также основан на детерминанте матрицы Гессе, но улучшенный с точки зрения скорости вычисления. Инвариантность к поворотам улучшено с помощью фильтров Шарра. Максимумы откликов детектора в пространственных положениях фиксируются как ключевые точки. Дескриптор AKAZE основан на эффективном алгоритме Modified Local Difference Binary (MLDB).

4. Scale Invariant Feature Transform (SIFT) [55] — алгоритм, детектор которого основан на операторе разницы гауссианов DoG. Ключевые точки определяются путём поиска локального максимума, используя DoG различных масштабов. Дескриптор для описания каждой точки извлекает область размером  $16 \times 16$  и дополнительно сегментирует область на подблоки.

После описания признаков пары изображений, следует этап их сопоставления. Для этого существуют методы сопоставления признаков, например, FLANN Based Matcher с KNN, который ищет ближайших соседей и хорошо подходит для работы с крупными изображениями.

После проведения теста соотношения Дэвида Лоу и получения хороших совпадений, извлекаются соответственные ключевые точки для рассматриваемой пары изображений, которые используются уже для оценки матрицы гомографии. Поскольку при её оценке не все точки соответствуют перспективному преобразованию, рассматривались следующие три метода для хорошей аппроксимации и борьбы с выбросами: RANSAC, LMEDS и RHO. Эти методы основаны на формировании случайного подмножества соответствующих пар точек (по четыре пары каждая), оценке матрицы гомографии с использованием этого подмножества и алгоритма наименьших квадратов. Затем вычисляется качество полученной гомографии (для RANSAC количество выбросов, для LMEDS средняя ошибка повторного проецирования).

## 2.2.2. Сравнительный анализ алгоритмов выявления локальных признаков

При проведении исследований был осуществлён сравнительный эксперимент на проверку стабильности ранее рассмотренных алгоритмов выявления признаков на 5 парах разновременных снимков при различных сочетаниях таких факторов, как: шум, яркость, масштаб и угол. Алгоритм KAZE был исключён из-за медленных вычислений.

На рис. 4 представлены наилучшие совпадения ключевых точек на примере одной такой пары, используя SIFT — алгоритм, где красным цветом обозначены границы области сопоставляемого изображения. Результаты сравнительного эксперимента представлены в табл. 4.

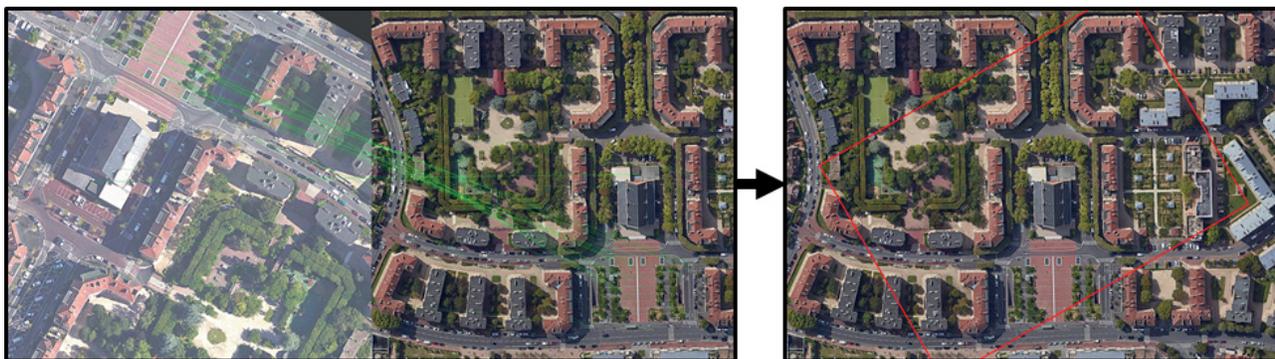


Рис. 4. Найденные совпадения на паре снимках и коррекция перспективы  
[Fig. 4. Found matches on a pair of images and perspective correction]

Таблица 4. Сравнение методов выявления локальных признаков изображения  
[Table 4. Comparison of methods for detecting local image features]

Метод	Время	Количество удачных совпадений при разных факторах					
		Исходный	Шум + Яркость	Масштаб (x1.4)	Угол (150°)	Масштаб + угол	Все факторы
AKAZE	49.7 с.	51/71	18/29	9/13	37/45	10/15	7/8
		47/48	11/14	8/11	25/26	6/11	0/0
		5/9	0/0	0/0	0/0	0/0	0/0
		46/58	11/15	26/30	42/45	20/24	5/8
		42/52	16/23	7/9	26/34	9/13	9/11
SIFT	55.1 с.	163/196	101/111	85/106	169/185	86/101	43/49
		117/131	82/89	34/54	102/120	47/55	44/54
		24/29	12/13	9/11	21/33	10/11	7/8
		190/204	20/22	83/89	152/156	91/100	26/32
		82/117	60/73	50/62	77/89	36/45	26/36
ORB	3.3 с.	<5	<5	<5	<5	<5	<5

В итоге этот алгоритм в сочетании с FLANN и методом очистки выбросов RHO оказался наиболее стабильным алгоритмом, показав самое хорошее количество удачных совпадений при разных факторах.

### 2.2.3. Обнаружение изменений

Обнаружение изменений осуществляется посредством анализа постклассификации пары бинарных изображений. После нахождения матрицы гомографии становится возможна реализация процесса сшивки ранее отснятых снимков и их соответствующих бинарных масок, что дает в результате расширенную целевую область.

Процесс сшивки представляет собой расчёт итогового размера изображения, операция переноса и конкатенация снимков.

Следующим этапом выполняется коррекция перспективы нового снимка по отношению к сшитому ранее отснятому снимку, используя матрицу гомографии, которая даёт в результате преобразованное изображение и его бинарную маску.

В итоге, имея две бинарные маски пары одновременных изображений, производится разность двух бинарных изображений и фильтрация возможных шумов, процесс которого аналогичен с ручной обработкой, представленный в разделе 2.1.3.

## 3. РЕЗУЛЬТАТЫ И ИХ ОБСУЖДЕНИЕ

### 3.1. Демонстрация на примере зданий и дорог

В качестве демонстрации работы предлагаемого метода обнаружения на примере зданий и дорог представлен следующий пример. На рис. 5 в центре представлен входной новый снимок, к которому с помощью GPS-метаданных подобран ранее отснятый снимок с целевой областью (снимок с зелёной границей) и который далее сшивается с ближайшими к нему остальными ранее отснятыми снимками. На рис. 6 для нового снимка производится сопоставление гистограмм по яркости и цветовой насыщенности, и после этого производится сегментация и постобработка топографических объектов отдельно. На рис. 7 производится коррекция перспективы, разность двух бинарных масок и фильтрация шумов, а на рис. 8 отображение изменений на исходном изображении.

### 3.2. Демонстрация на примере растительности и водных объектов

В качестве демонстрации работы предлагаемого метода обнаружения на примере растительности и водных объектов представлен следующий небольшой пример без этапа сшивки. На рис. 9 представлен новый снимок и ранее отснятый снимок, а также соответствующие их бинарные маски с постобработкой. На рис. 10 представлен результат

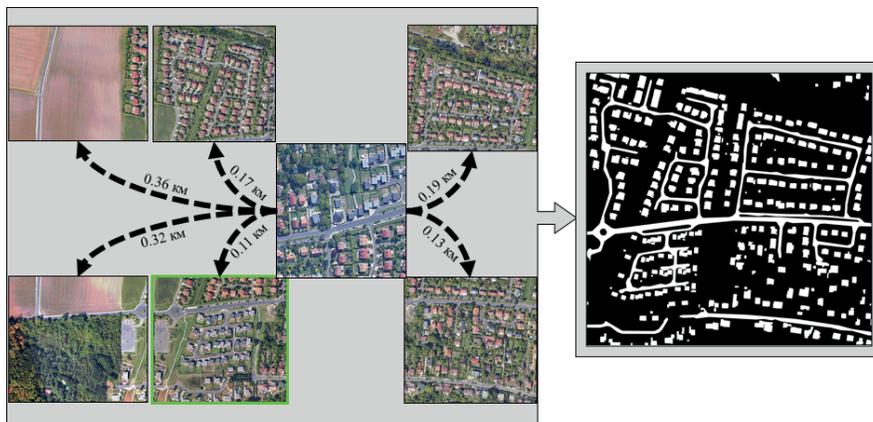


Рис. 5. Процесс получения ранее отснятого снимка  
[Fig. 5. The process of acquiring a previously captured image]

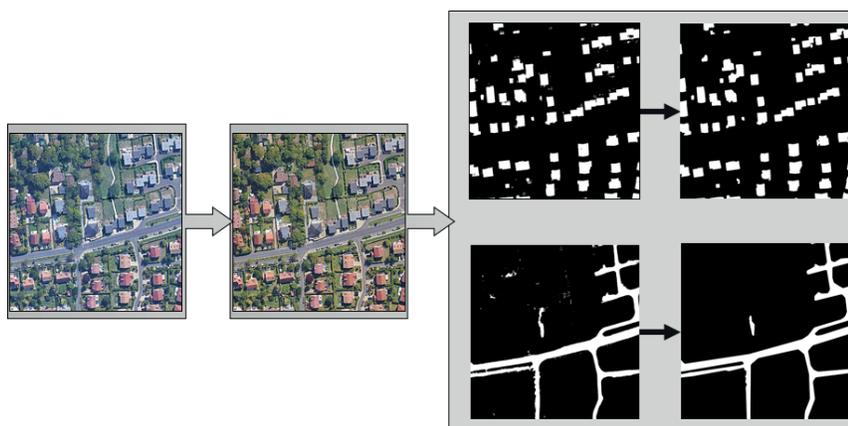


Рис. 6. Обработка входного нового снимка  
[Fig. 6. Processing the new input image]

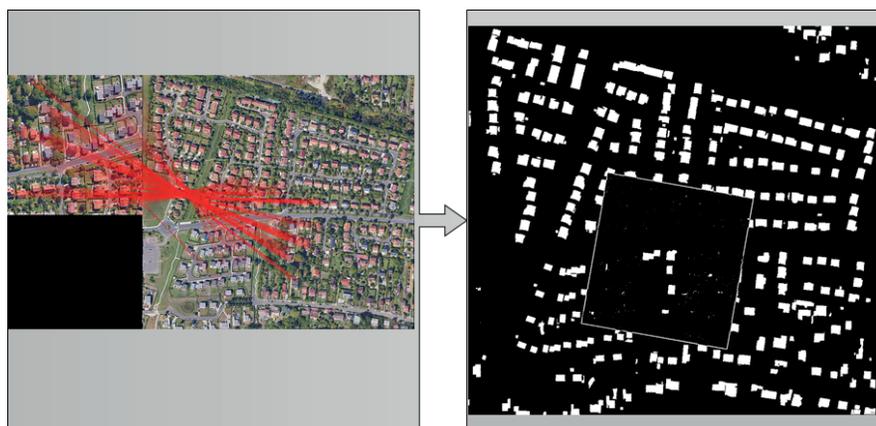


Рис. 7. Результат сопоставления и разницы двух масок  
[Fig. 7. The result of the matching and difference of the two masks]



Рис. 8. Обнаружение изменений: а) фильтрация; б) отображение изменений  
[Fig. 8. Change detection: a) filtering; b) displaying changes]

коррекции перспективы и разницы скорректированного нового снимка и ранее отснятого снимка. При вычислении разницы можно заметить, что в результате присутствует множество шумов, поскольку обученные сети реагируют даже на незначительные изменения.

На финальном этапе (рис. 11) происходит фильтрация бинарной маски и отображение изменений на исходный снимок. Несмотря на фильтрацию, в результате ещё присутствуют выбросы из-за не самой высокой точности обученных сетей. Дополнительно в итогов-

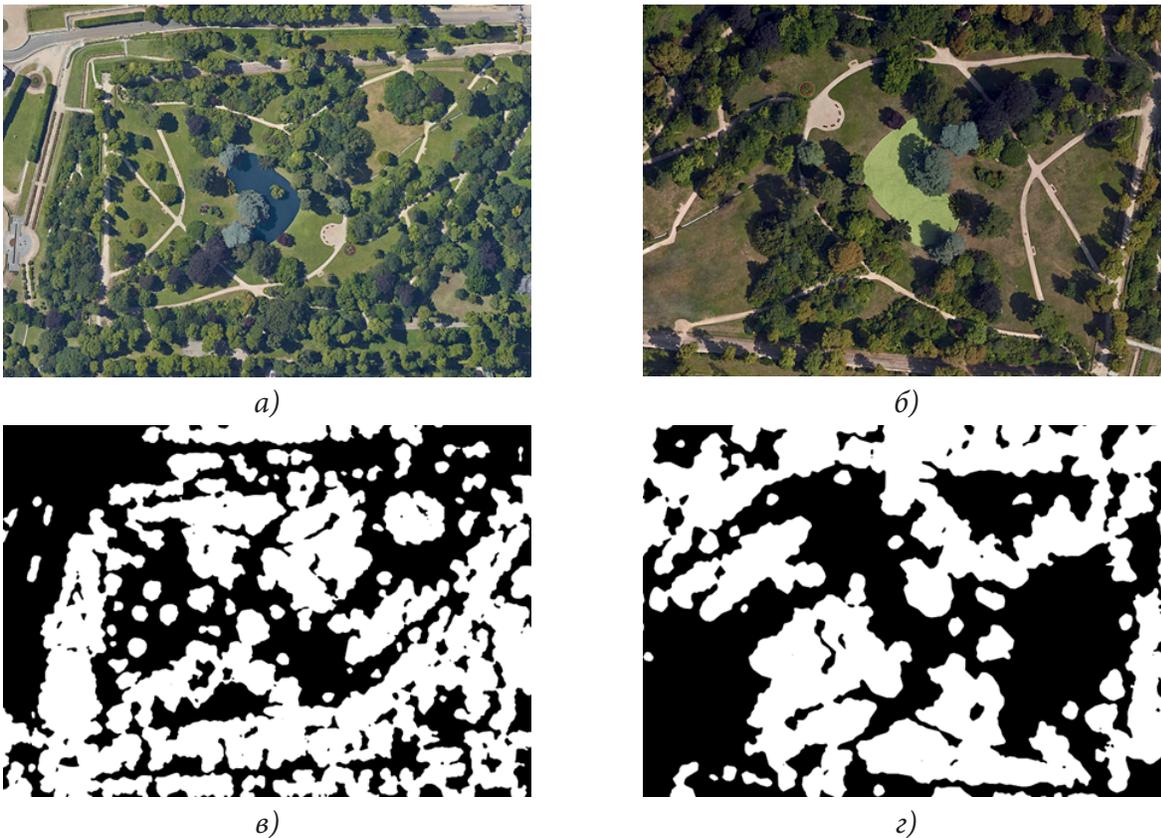


Рис. 9. Пример изображения с растительностью и водным объектом: а) ранее отснятый снимок; б) новый снимок; в) маска ранее отснятого снимка; г) маска нового снимка  
[Fig. 9. The example of the image with vegetation and a body of water: a) the previously captured image; b) the new picture; c) mask of the previously captured image; d) mask of the new image]

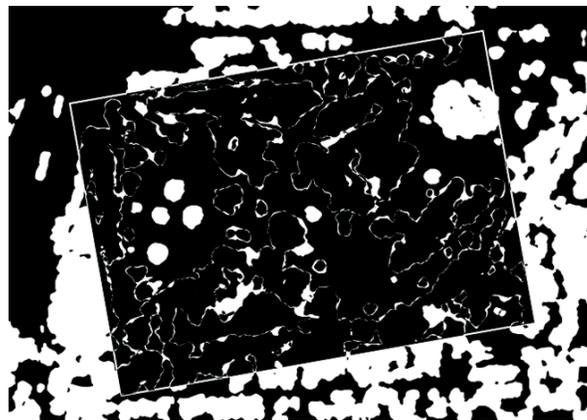


Рис. 10. Результат коррекции перспективы и разности масок  
[Fig. 10. Result of perspective correction and mask difference]

вом результате отображена площадь обнаруженных областей в квадратных метрах, с помощью которой можно приблизительно оценить, например, ущерб вырубке лесных насаждений или насколько высушился какой-либо водный объект. Высчитывается данная площадь на основе коэффициента пространственного разрешения отношения

метр на пиксель, определяющий длину пикселя в метрах, который всегда известен в спутниковых снимках.

### 3.3. Обсуждение результатов

На основе визуального анализа результатов тестирования можно сделать некоторые

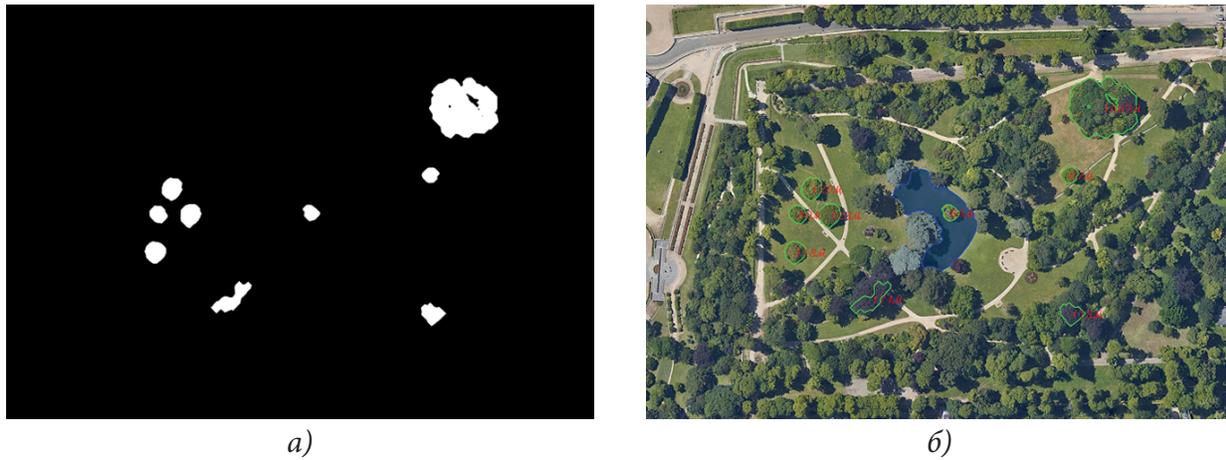


Рис. 11. Обнаружение изменений: а) результат фильтрации маски;  
б) итоговое обнаружение изменений

[Fig. 11. Change detection: a) the result of mask filtering; b) final detection of changes]

выводы. Предлагаемый метод обнаружения элементов новизны в целом успешно фиксирует изменения, однако обученные сети не совсем точны из-за недостаточно большого обучающего набора данных, вследствие чего могут происходить ложные срабатывания. Кроме этого, использование постклассификационного подхода показывает высокую чувствительность к малейшим изменениям, в том числе к элементам не представляющим интерес, что вынуждает использовать дополнительную фильтрацию. Это подвергает риску потери важных, но небольших изменений. Таким образом, предлагаемый подход хорошо работает для обнаружения относительно крупных изменений.

Чтобы исправить эти аспекты, в последующей работе планируется увеличить объем обучающих данных, чтобы обеспечить более точную сегментацию объектов. Кроме этого, необходимо создать выборку из разновременных изображений, чтобы можно было отказаться от постклассификации с её негативными эффектами, интегрировать в предлагаемый метод end-to-end архитектуру собственной разработки и количественно оценить работоспособность обнаружения.

## ЗАКЛЮЧЕНИЕ

Данная работа посвящена разработке тематического и программного обеспечения

в интересах создания системы для мониторинга элементов новизны на аэрокосмических снимках городской, пригородной, лесистой местности и акваторий. Особенностью предложенного подхода является реализация возможности сравнения изображений, отличающихся высотой и ракурсом съёмки.

В ходе работы были исследованы традиционные и современные методы обнаружения изменений, а также популярные архитектуры сегментационных нейронных сетей, алгоритмы постобработки результирующей сегментации, а также методы выявления локальных признаков изображения.

Были созданы и подготовлены два собственных набора данных: здания и дороги, густая растительность и водоёмы. Экспериментально исследованы архитектуры сегментационных нейронных сетей на эффективность, среди которых наиболее точными оказались архитектуры OCRNet и DeepLabV3+. Экспериментально исследованы алгоритмы постобработки, среди которых наилучшей по качеству показала себя сегментационная нейронная сеть Cascade PSP. Был также организован процесс получения обработки ранее отснятого снимка целевой области посредством сшивки изображений и процесс коррекции перспективы с помощью матрицы гомографии. Для её нахождения были исследованы алгоритмы выявления локальных признаков на стабильность, среди которых наилучшим оказался алгоритм SIFT.

В результате тестирования предложенного метода был сделан вывод о том, что подход хорошо работает лишь для обнаружения относительно крупных изменений, однако обладает высокой чувствительностью по отношению к малым изменениям, что может привести к возникновению ложноположительной диагностики.

## КОНФЛИКТ ИНТЕРЕСОВ

Авторы декларируют отсутствие явных и потенциальных конфликтов интересов, связанных с публикацией настоящей статьи.

## СПИСОК ЛИТЕРАТУРЫ

1. *Singh, A.* Review article digital change detection techniques using remotely-sensed data / A. Singh // International journal of remote sensing. – 1989. – V. 10, No 6. – P. 989–1003.

2. *Миртова, И. А.* Изучение динамики природных процессов и объектов по аэро- и космическим снимкам: учеб. пособие / И. А. Миртова – Москва : МИИГАиК, 2006. – 77 с.

3. *Chen, C.* Multi-decadal mangrove forest change detection and prediction in Honduras, Central America, with Landsat imagery and a Markov chain model / C. Chen [et al.] // Remote Sensing. – 2013. – V. 5. – P. 6408–6426. – DOI: 10.3390/RS5126408.

4. *Paparoditis, N.* Building Detection and Reconstruction from Mid- and High-Resolution Aerial Imagery / N. Paparoditis, M. Jordan, J. P. Cocosquez // Computer Vision And Image Understanding. – 1997. – V. 72. – P. 122–142. – DOI: 10.1006/CVIU.1998.0722.

5. *Bruzzone, L.* Automatic analysis of the difference image for unsupervised change detection / L. Bruzzone, D. F. Prieto // IEEE Transactions On Geoscience And Remote Sensing. – 2000. – V. 38. – P. 1171–1182. – DOI: 10.1109/36.843009.

6. *Сагалович, В. Н.* Обнаружение изменений на земной поверхности по разновременным мультиспектральным изображениям / В. Н. Сагалович, Э. Я. Фальков, Т. И. Царева // Исследование Земли из космоса. – 2008. – № 5. – С. 37–41.

7. *Todd, W.J.* Urban and regional land use change detected by using Landsat data / W. J. Todd // Journal research U.S. geological survey. – 1977. – V. 5, No 5. – P. 529–534.

8. *Jackson, R. D.* Spectral indices in N-Space / R. D. Jackson // Remote Sens Environ – 1983. – V. 13, No 5. – P. 409–421. – DOI: 10.1016/0034-4257(83)90010-X.

9. *Malila, W. A.* Change Vector Analysis: An Approach for Detecting Forest Changes with Landsat / W. A. Malila // IEEE Proceedings of the annual symposium on machine processing of remotely sensed data. – 1980. – P. 1565–1596.

10. *Плотникова, М. А.* Исследование методов выявления изменений территорий по разновременным космическим изображениям / М. А. Плотникова, Е. П. Хлебникова // Интерэкспо Гео-Сибирь. – 2020. – Т. 6, № 2. – С. 48–58.

11. *Хлебникова, Е. П.* Использование разновременных космических снимков при мониторинге городских территорий / Е. П. Хлебникова, М. А. Плотникова // Регулирование земельно-имущественных отношений в России: правовое и геопространственное обеспечение, оценка недвижимости, экология, технологические решения. – 2019. – Т. 1. – С. 248–252.

12. *Lupino, L. T.* Unsupervised image regression for heterogeneous change detection / L. T. Lupino, F. M. Bianchi, G. Moser, S. N. Anfinsen // IEEE Transactions on Geoscience and Remote Sensing. – 2019. – V. 50. – P. 9960–9975. – DOI: 10.1109/TGRS.2019.2930348.

13. *Мясников, Ф. С.* Анализ алгоритмов обнаружения изменений на космических снимках / Ф. С. Мясников // Передача, обработка, восприятие текстовой и графической информации : сб. тр. Междунар. науч.-практ. конф. Екатеринбург, 19–20 марта 2015 г. – Екатеринбург: УрФУ, 2015. – С. 66–70.

14. *Cui, B.* An Unsupervised SAR Change Detection Method Based on Stochastic Subspace Ensemble Learning / B. Cui [et al.] // Remote Sensing. – 2019. – V. 11, No 11. – P. 1314. – DOI: 10.3390/RS11111314.

15. *Benedek, C.* Change detection in optical aerial images by a multilayer conditional mixed

Markov model / C. Benedek, T. Szir'anyi // IEEE Transactions on Geoscience and Remote Sensing. – 2009. – V. 47, No 10. – P. 3416–3430. – DOI: 10.1109/TGRS.2009.2022633.

16. Bazi, Y. Unsupervised change detection in multispectral remotely sensed imagery with level set methods / Y. Bazi, F. Melgani, H. Al-Sharari // IEEE Transactions on Geoscience and Remote Sensing. – 2010. – V. 48, No 8. – P. 3178–3187. – DOI: 10.1109/TGRS.2010.2045506.

17. Cao, G. A new change detection method in high-resolution remote sensing images based on a conditional random field model / G. Cao, L. Zhou, Y. Li // Int. J. Remote Sens. – 2016. – V. 37, No 5. – P. 1173–1189. – DOI: 10.1080/01431161.2016.1148284.

18. Lv, P. Unsupervised change detection based on hybrid conditional random field model for high spatial resolution remote sensing imagery / P. Lv [et al.] // IEEE Trans. Geosci. Remote Sens. – 2018. – V. 56, No 7. – P. 4002–4015. – DOI: 10.1109/TGRS.2018.2819367.

19. Jian, P. A hypergraph-based context-sensitive representation technique for VHR remote-sensing image change detection / P. Jian, K. Chen, C. Zhang // Int. J. Remote Sens. – 2016. – V. 37, No 8 – P. 1814–1825. – DOI: 10.1080/2150704X.2016.1163744.

20. Chen, G. Object-based change detection / G. Chen, G. Hay, L. Carvalho, M. Wulder // International Journal of Remote Sensing. – 2012. – V. 33, No 4. – P. 4434–4457. – DOI: 10.1080/01431161.2011.648285.

21. Ma, L. Deep learning in remote sensing applications: A meta-analysis and review / L. Ma [et al.] // ISPRS Journal of Photogrammetry and Remote Sensing. – 2019. – V. 152. – P. 166–177. – DOI: 10.1016/J.ISPRSJPRS.2019.04.015.

22. Khelifi, L. Deep learning for change detection in remote sensing images: comprehensive review and meta-analysis / L. Khelifi, M. Mignotte // IEEE Access. – 2020. – V. 8. – P. 126385–126400. – DOI: 10.1109/ACCESS.2020.3008036.

23. Sakurada, K. Change Detection from a Street Image Pair using CNN Features and Superpixel Segmentation / K. Sakurada, T. Okatani // In Proceedings of the British Machine Vision Conference BMVC, Swansea, UK, 7–10 September 2015. – P. 61.1–61.12. – DOI: 10.5244/C.29.61.

24. El Amin, A. Zoom out CNNs features for optical remote sensing change detection / A. El Amin, Q. Liu, Y. Wang, // In Proceedings of the 2017 2nd International Conference on Image, Vision and Computing (ICIVC), Chengdu, China, 2–4 June 2017. – P. 812–817. – DOI: 10.1109/ICIVC.2017.7984667.

25. Zhang, M. Triplet-Based Semantic Relation Learning for Aerial Remote Sensing Image Change Detection / M. Zhang [et al.] // IEEE Geosci. Remote Sens. Lett. – 2019. – V. 16, No 2. – P. 266–270. – DOI: 10.1109/LGRS.2018.2869608

26. Niu, X. A Conditional Adversarial Network for Change Detection in Heterogeneous Images / X. Niu, M. Gong, T. Zhan, Y. Yang // IEEE Geoscience and Remote Sensing Letters. – 2019. V. 16, No 1. – P. 45–49. – DOI: 10.1109/LGRS.2018.2868704.

27. Lebedev, M. Change detection in remote sensing images using conditional adversarial networks / M. Lebedev, Y. Vizilter, O. Vygolov [et al.] // Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci. – 2018. – V. 42. – P. 565–571. – DOI: 10.5194/ISPRS-ARCHIVES-XLII-2-565-2018.

28. Hao, C. A Spatial-Temporal Attention-Based Method and a New Dataset for Remote Sensing Image Change Detection / C. Hao, Z. Shi // Remote Sensing. – 2020. – V. 12. – P. 1662. – DOI: 10.3390/RS12101662.

29. Isola, P. Image-to-Image Translation with Conditional Adversarial Networks / P. Isola [et al.] // IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, USA, 21–26 July 2017. – P. 5967–5976. – DOI: 10.1109/CVPR.2017.632.

30. Daudt, R. Fully Convolutional Siamese Networks for Change Detection / R. Daudt, B. Saux, A. Boulch // 25th IEEE International Conference on Image Processing (ICIP), Athens, Greece, 7–10 October 2018. – P. 4063–4067. – DOI: 10.1109/ICIP.2018.8451652

31. Daudt, R. High Resolution Semantic Change Detection [Электронный ресурс] / R. Daudt [et al.] // ArXiv. – 2018. – Режим доступа: <http://arxiv.org/abs/1810.08452>

32. He, K. Deep residual learning for image recognition / K. He, X. Zhang, S. Ren, [et al.] // IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA,

27–30 June 2016. – P. 770–778. – DOI: 10.1109/CVPR.2016.90.

33. *Alcantarilla, P.* Street-view change detection with deconvolutional networks / P. Alcantarilla [et al.] // *Autonomous Robots*. – 2018. – V. 42. – P. 1301–1322. – DOI: 10.1007/S10514-018-9734-5.

34. *Peng, D.* End-to-End Change Detection for High Resolution Satellite Images Using Improved UNet++ / D. Peng, Z. Yongjum, H. Guan // *Remote Sensing*. – 2019. – V. 11. – P. 1382. – DOI: 10.3390/RS11111382.

35. *Zhou, Z.* UNet++: A Nested U-Net Architecture for Medical Image Segmentation / Z. Zhou [et al.] // *Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support*. – 2018. – V. 11035. – P. 3–11. – DOI: 10.1007/978-3-030-00889-5\_1.

36. *Zhang, C.* A deeply supervised image fusion network for change detection in high resolution bi-temporal remote sensing images / C. Zhang, P. [et al.] // *ISPRS Journal of Photogrammetry and Remote Sensing* – 2020. – V. 166. – P. 183–200. – DOI: 10.1016/J.ISPRS-JPRS.2020.06.003.

37. *Chen, J.* DASNet: Dual Attentive Fully Convolutional Siamese Networks for Change Detection in High-Resolution Satellite Images / J. Chen [et al.] // *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*. – 2021. – V. 14. – P. 1194–1206. – DOI: 10.1109/JSTARS.2020.3037893.

38. *Fang, S.* SNUNet-CD: A Densely Connected Siamese Network for Change Detection of VHR Images / S. Fang [et al.] // *IEEE Geoscience and Remote Sensing Letters*. – 2022. – V. 19. – P. 1–5. – DOI: 10.1109/LRGS.2021.3056416.

39. *Chen, H.* RDP-Net: Region Detail Preserving Network for Change Detection [Электронный ресурс] / H. Chen [et al.] // *ArXiv*. – 2022. – Режим доступа: <http://arxiv.org/abs/1810.08452>.

40. *Long, J.* Fully convolutional networks for semantic segmentation / J. Long, E. Shelhamer, T. Darrell // *IEEE transactions on pattern analysis and machine intelligence*. – 2015. – V. 39, No 4. – P. 640–651. – DOI: 10.1109/CVPR.2015.7298965.

41. *Ronneberger, O.* U-Net: Convolutional Networks for Biomedical Image Segmentation /

O. Ronneberger, P. Fischer, T. Brox // *Medical Image Computing and Computer-Assisted Intervention*. – 2015. – P. 234–241. – DOI: 10.1007/978-3-319-24574-4\_28.

42. *Badrinarayanan, V.* SegNet: A Deep Convolutional Encoder-Decoder Architecture for Image Segmentation / V. Badrinarayanan, A. Kendall, R. Cipolla // *IEEE Transactions on Pattern Analysis and Machine Intelligence*. – 2017. – V. 39. – P. 2481–2495. – DOI: 10.1109/TPAMI.2016.2644615.

43. *Zhao, H.* Pyramid Scene Parsing Network / H. Zhao [et al.] // *The IEEE Conference on Computer Vision and Pattern Recognition*, Honolulu, HI, USA, 21–26 July 2017, IEEE, P. 6230–6239. – DOI: 10.1109/CVPR.2017.660.

44. *Lin, T.* Feature Pyramid Networks for Object Detection / T. Lin [et al.] // *The IEEE Conference on Computer Vision and Pattern Recognition*, Honolulu, HI, USA, 21–26 July 2017, IEEE, P. 936–944. – DOI: 10.1109/CVPR.2017.106.

45. *Chen, L.* Encoder-decoder with atrous separable convolution for semantic image segmentation / L. Chen // *Computer Vision – ECCV 2018*. – 2018. – P. 801–818. – DOI: 10.1007/978-3-030-01234-2\_49.

46. *Fu, J.* Dual attention network for scene segmentation / J. Fu [et al.] // *IEEE Conference on Computer Vision and Pattern Recognition*. – 2019. – P. 3146–3154. – DOI: 10.1109/CVPR.2019.00326.

47. *Zhang, F.* ACFNet: Attentional Class Feature Network for Semantic Segmentation / F. Zhang [et al.] // *2019 IEEE/CVF International Conference on Computer Vision*. – 2019. – P. 6797–6806. – DOI: 10.1109/ICCV.2019.00690.

48. *Yuan, Y.* Segmentation Transformer: Object-Contextual Representations for Semantic Segmentation / Y. Yuan [et. al] // *European Conference on Computer Vision*. – 2020. – P. 173–190. – DOI: 10.1007/978-3-030-58539-6\_11.

49. *Krähenbühl, P.* Efficient Inference in Fully Connected CRFs with Gaussian Edge Potentials / P. Krähenbühl, V. Koltun // *In Advances in neural information processing systems*. – 2011. – P. 109–117.

50. *Zheng, S.* Conditional Random Fields as Recurrent Neural Networks / S. Zheng [et al.] // *IEEE International Conference on Computer*

Vision (ICCV). – 2015. – P. 1529–1537. – DOI: 10.1109/ICCV.2015.179

51. *Cheng, H.* CascadePSP: Toward Class-Agnostic and Very High-Resolution Segmentation via Global and Local Refinement / H. Cheng [et al.] // Conference on Computer Vision and Pattern Recognition. – 2020. – P. 8887–8896. – DOI: 10.1109/CVPR42600.2020.00891.

52. *Rublee, E.* ORB: An efficient alternative to SIFT or SURF / E. Rublee [et al.] // IEEE International Conference on Computer Vision, Barcelona, Spain, 6–13 November 2011. – P. 2564–2571. – DOI: 10.1109/ICCV.2011.6126544.

53. *Alcantarilla, P.* Kaze Features / P. Alcantarilla, A. J. Davison, A. Bartoli // Proceedings

of the British Machine Vision Conference. – 2013. – V. 7577, No 6. – P. 13.1–13.11. – DOI: 10.1007/978-3-642-33783-3\_16

54. *Yang, X.* LDB: An ultra-fast feature for scalable Augmented Reality on mobile devices / X. Yang, K. T. Cheng // IEEE International Symposium on Mixed and Augmented Reality. – 2012. – P. 49–57. – DOI: 10.1109/ISMAR.2012.6402537.

55. *Lowe, D. G.* Object recognition from local scale-invariant features / D. G. Lowe // Proceedings of the Seventh IEEE International Conference on Computer Vision. – 1999. – V. 2. – P. 1150–1157. – DOI: 10.1109/ICCV.1999.79041

**Отырба Ростислав Русланович** — аспирант кафедры технологий обработки и защиты информации, факультета компьютерных наук, Воронежского государственного университета.

E-mail: rostislav.otyrba.97@gmail.com

ORCID iD: <https://orcid.org/0000-0002-0412-2465>

**Дрюченко Миахил Анатольевич** — канд. техн. наук, доц., доцент кафедры технологий обработки и защиты информации, факультета компьютерных наук, Воронежского государственного университета.

E-mail: m\_dryuchenko@mail.ru

ORCID iD: <https://orcid.org/0000-0001-8837-5875>

**Сирота Александр Анатольевич** — д-р техн. наук, проф., заведующий кафедрой технологий обработки и защиты информации, факультета компьютерных наук, Воронежского государственного университета.

E-mail: sir@cs.vsu.ru

ORCID iD: <https://orcid.org/0000-0002-5785-8513>

## ANALYSIS OF AEROSPACE MONITORING IMAGES FOR NOVELTY DETECTION ON THE GROUND USING DEEP LEARNING METHODS

© 2022 R. R. Otyrba✉, M. A. Dryuchenko, A. A. Sirota

Voronezh State University

1, Universitetskaya Square, 394018 Voronezh, Russian Federation

**Annotation.** The article is devoted to the development of mathematical software in order to detect elements of novelty in aerospace images of urban, suburban, wooded and water areas. To achieve this goal, in the course of the work, traditional and modern change detection methods, popular architectures of neural networks for segmentation, post-processing and image matching algorithms were explored. The main idea of the proposed method and the algorithms implemented on its basis is the use of the post-classification approach. It is based on the estimation of the difference between a pair of binary masks obtained as a result of segmentation of analyzed multi-temporal images using neural networks for segmentation, and specialized post-processing algorithms. The approach proposes to implement change detection in the conditions of differences in the geometric aspects of the compared images, which most of the previously proposed methods and algorithms have not taken into account. This approach makes it possible to use the resulting binary masks for qualitative matching of a new and previously captured image, perspective correction, and their difference, respectively. In addition, unlike most previous research, it is proposed to detect not only changed objects but also determine their classes and related topographic objects such as buildings, roads, trees, and water objects. It makes it possible to use the method in photogrammetry applications. Due to the lack of objective quantitative data to assess the accuracy of detection, visual testing was carried out. As a result, it was concluded that the proposed approach works very well only for detecting large changes since the implemented detection method is highly sensitive to the smallest changes.

**Keywords:** novelty detection, aerospace image analysis, deep neural networks, semantic segmentation, image matching algorithms.

### CONFLICT OF INTEREST

The authors declare the absence of obvious and potential conflicts of interest related to the publication of this article.

### REFERENCES

1. Singh A. (1989) Review Article Digital change detection techniques using remotely-sensed data. *International Journal of Remote Sensing*, 10(6). P. 989–1003.

2. Mitrova I. A. (2006) Izuchenie dinamiki prirodnyh processov i ob"ektov po aero- i kosmicheskim snimkam [Studying the dynamics of natural processes and objects from aerial and space images]. *Moscow, MIIGAiK*. 77 pp. (in Russian)

3. Chen C.-F., Son N., Chang N., Chen C.-R., Chang L., Valdez M., Centeno G., Thompson C. A. and Aceituno J. (2013) Multi-Decadal Mangrove Forest Change Detection and Prediction in Honduras, Central America, with Landsat Imagery and a Markov Chain Model. *Remote. Sens.* 5. P. 6408–6426. doi: 10.3390/rs5126408.

4. Paparoditis N., Cord M., Jordan M. and Cocquerez J.-P. (1998) Building Detection and Reconstruction from Mid- and High-Resolution Aerial Imagery. *Computer Vision and Image Understanding*. 72(2). P. 122–142, doi: 10.1006/cviu.1998.0722.

5. Bruzzone L. and Prieto D. F. (2000) Automatic analysis of the difference image for unsupervised change detection. *IEEE Transactions on Geoscience and Remote Sensing*. 38(3). P. 1171–1182, doi:10.1109/36.843009.

✉ Otyrba Rostislav R.  
e-mail: rostislav.otyrba.97@gmail.com

6. Sagalovich V. N., Fal'kov E. Ya. and Careva T. I. (2008) Obnaruzhenie izmenenij na zemnoj poverhnosti po raznovremennym mul'tispektral'nym izobrazheniyam [Detection of changes on the earth's surface from multi-temporal multispectral images]. *Issledovanie Zemli iz kosmosa*. (5). P. 37–41. (in Russian)
7. Todd W. J. (1977) Urban and regional land use change detected by using Landsat data. *Journal of Research of the U.S. Geological Survey*. 5(5). P. 529–534.
8. Jackson R. (1983) Spectral indices in N-Space. *Remote Sensing of Environment*. 13(5). P. 409–421. doi: 10.1016/0034-4257(83)90010-x.
9. Malila W. A. (1980) Change Vector Analysis: An Approach for Detecting Forest Changes with Landsat. *IEEE Proceedings of the annual symposium on machine processing of remotely sensed data*. P. 1565–1596.
10. Plotnikova M. A. and Hlebnikova E. P. (2020) Issledovanie metodov vyyavleniya izmenenij territorij po raznovremennym kosmicheskim izobrazheniyam [Investigation of methods for detecting changes in territories using multi-temporal space images]. *Interekspos Geo-Sibir'*. 6(2). P. 48–58.
11. Hlebnikova E. P. and Plotnikova M. A. (2019) Ispol'zovanie raznovremennykh kosmicheskikh snimkov pri monitoringe gorodskih territorij [Using multi-temporal space images in monitoring urban areas]. *Regulirovanie zemel'no-imushchestvennykh otnoshenij v Rossii: pravovoe i geoprostranstvennoe obespechenie, ocenka nedvizhimosti, ekologiya, tekhnologicheskie resheniya*. 1. P. 248–252. (in Russian)
12. Luppino L. T., Bianchi F. M., Moser G. and Anfinsen S. N. (2019). Unsupervised Image Regression for Heterogeneous Change Detection. *IEEE Transactions on Geoscience and Remote Sensing*. 57(12). P. 9960–9975. doi: 10.1109/tgrs.2019.2930348.
13. Myasnikov F. S. (2015) Analiz algoritmov obnaruzheniya izmenenij na kosmicheskikh snimkakh [Analysis of Algorithms for Detecting Changes in Satellite Images]. *Peredacha, obrabotka, vospriyatie tekstovoj i graficheskoy informacii [Transmission, processing, perception of text and graphic information]: Proceedings of the International scientific and practical conference*, March 19–20, Yekaterinburg, Russia. P. 66–70. (in Russian)
14. Cui B., Zhang Y., Yan L., Wei J. and Wu H. (2019) An Unsupervised SAR Change Detection Method Based on Stochastic Subspace Ensemble Learning. *Remote Sensing*. 11(11). P. 1314. doi: 10.3390/rs11111314.
15. Benedek C. and Sziranyi T. (2009) Change Detection in Optical Aerial Images by a Multilayer Conditional Mixed Markov Model. *IEEE Transactions on Geoscience and Remote Sensing*. 47(10). P. 3416–3430. doi: 10.1109/tgrs.2009.2022633.
16. Bazi Y., Melgani F. and Al-Sharari H. D. (2010) Unsupervised Change Detection in Multispectral Remotely Sensed Imagery With Level Set Methods. *IEEE Transactions on Geoscience and Remote Sensing*. 48(8). P. 3178–3187. doi: 10.1109/tgrs.2010.2045506.
17. Cao G., Zhou L. and Li Y. (2016) A new change-detection method in high-resolution remote sensing images based on a conditional random field model. *International Journal of Remote Sensing*. 37(5). P. 1173–1189. doi:10.1080/01431161.2016.1148284.
18. Lv P., Zhong Y., Zhao J. and Zhang L. (2018) Unsupervised Change Detection Based on Hybrid Conditional Random Field Model for High Spatial Resolution Remote Sensing Imagery. *IEEE Transactions on Geoscience and Remote Sensing*. 56(7). P. 4002–4015. doi: 10.1109/tgrs.2018.2819367.
19. Jian P., Chen K. and Zhang C. (2016) A hypergraph-based context-sensitive representation technique for VHR remote-sensing image change detection. *International Journal of Remote Sensing*. 37(8). P. 1814–1825. doi: 10.1080/2150704x.2016.1163744.
20. Chen G., Hay G., Carvalho L. and Wulder M. (2012) Object-based change detection. *International Journal of Remote Sensing*. 33(4). P. 4434–4457. doi: 10.1080/01431161.2011.648285.
21. Ma L., Liu Y., Zhang X., Ye Y., Yin G. and Johnson B. (2019) Deep learning in remote sensing applications: A meta-analysis and review. *ISPRS Journal of Photogrammetry and Remote Sensing*. 152. P. 166–177. doi: 10.1016/j.isprsjprs.2019.04.015.
22. Khelifi L. and Mignotte M. (2020) Deep learning for change detection in remote sensing

- images: comprehensive review and meta-analysis. *IEEE Access*. 8. P. 126385–126400. doi: 10.1109/access.2020.3008036.
23. Sakurada K. and Okatani T. (2015) Change detection from a street image pair using CNN features and superpixel segmentation. In: Xie, X., Jones, M. W., and Tam, G. K. L. (eds) *Proceedings of the British Machine Vision Conference 2015, 7–10 September, Swansea, UK, BMVA Press*. P. 61.1–61.12. doi: 10.5244/C.29.61.
24. El Amin A. M., Liu Q. and Wang Y. (2017) Zoom out CNNs features for optical remote sensing change detection. In *2017 2nd International Conference on Image, Vision and Computing (ICIVC), 2–4 June, Chengdu, China, IEEE*. P. 812–817. doi: 10.1109/icivc.2017.7984667.
25. Zhang M., Xu G., Chen K., Yan M. and Sun X. (2019). Triplet-Based Semantic Relation Learning for Aerial Remote Sensing Image Change Detection. *IEEE Geoscience and Remote Sensing Letters*. 16(2). P. 266–270. doi:10.1109/lgrs.2018.2869608.
26. Niu X., Gong M., Zhan T. and Yang Y. (2019). A Conditional Adversarial Network for Change Detection in Heterogeneous Images. *IEEE Geoscience and Remote Sensing Letters*. 16(1). P. 45–49. doi: 10.1109/lgrs.2018.2868704.
27. Lebedev M. A., Vizilter Y. V., Vygolov O. V., Knyaz V. A. and Rubis A. Y. (2018) Change detection in remote sensing images using conditional adversarial networks. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*. 2. P. 565–571. doi: 10.5194/isprs-archives-xlii-2-565-2018.
28. Chen H. and Shi Z. (2020) A Spatial-Temporal Attention-Based Method and a New Dataset for Remote Sensing Image Change Detection. *Remote Sensing*. 12(10). P. 1662. doi: 10.3390/rs12101662.
29. Isola P., Zhu J.-Y., Zhou T. and Efros A. A. (2017) Image-to-image translation with conditional adversarial networks. In *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 21–26 July, Honolulu, HI, USA, IEEE*. P. 5967–5976. doi: 10.1109/cvpr.2017.632.
30. Caye Daudt R., Le Saux B. and Boulch A. (2018) Fully convolutional Siamese networks for change detection. In *2018 25th IEEE International Conference on Image Processing (ICIP), 7–10 October, Athens, Greece, IEEE*. P. 4063–4067. doi: 10.1109/icip.2018.8451652
31. Daudt R. C., Saux B. L., Boulch A. and Gousseau Y. (2018) Multitask learning for large-scale semantic change detection, *arXiv* [Electronic resource]. Available at: <http://arxiv.org/abs/1810.08452>.
32. He K., Zhang X., Ren S. and Sun J. (2016) Deep residual learning for image recognition. In *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 27–30 June, Las Vegas, NV, USA, IEEE*. P. 770–778. doi: 10.1109/cvpr.2016.90.
33. Alcantarilla P. F., Stent S., Ros G., Arroyo R. and Gherardi R. (2018) Street-view change detection with deconvolutional networks. *Autonomous Robots*. 42(7). P. 1301–1322. doi: 10.1007/s10514-018-9734-5.
34. Peng D., Zhang Y. and Guan H. (2019). End-to-End Change Detection for High Resolution Satellite Images Using Improved UNet++. *Remote Sensing*. 11(11). P. 1382. doi: 10.3390/rs11111382.
35. Zhou Z., Rahman Siddiquee M. M., Tajbakhsh N. and Liang J. (2018) UNet++: A Nested U-Net Architecture for Medical Image Segmentation. *Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support*. 11035. P. 3–11. doi: 10.1007/978-3-030-00889-5\_1.
36. Zhang C., Yue P., Tapete D., Jiang L., Shang-guan B., Huang L. and Liu G. (2020) A deeply supervised image fusion network for change detection in high-resolution bi-temporal remote sensing images. *ISPRS Journal of Photogrammetry and Remote Sensing*. 166. P. 183–200. doi: 10.1016/j.isprsjprs.2020.06.003.
37. Chen J., Yuan Z., Peng J., Chen L., Huang H., Zhu J., Liu Y. and Li H. (2021) DASNet: Dual Attentive Fully Convolutional Siamese Networks for Change Detection in High-Resolution Satellite Images. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*. 14. P. 1194–1206. doi: 10.1109/jstars.2020.3037893.
38. Fang S., Li K., Shao J. and Li Z. (2022) SNUNet-CD: A Densely Connected Siamese Network for Change Detection of VHR Images. *IEEE Geoscience and Remote Sensing Letters*. 19. P. 1–5. doi: 10.1109/LGRS.2021.3056416.

39. Daudt R. C., Saux B. L., Boulch A. and Gousseau Y. (2019) Multitask Learning for Large-scale Semantic Change Detection. *arXiv* [Electronic resource]. Available at: <http://arxiv.org/abs/1810.08452>
40. Shelhamer E., Long J. and Darrell T. (2017) Fully convolutional networks for semantic segmentation. *IEEE transactions on pattern analysis and machine intelligence*. 39(4). P. 640–651. doi: 10.1109/tpami.2016.2572683.
41. Ronneberger O., Fischer P. and Brox T. (2015) U-Net: Convolutional Networks for Biomedical Image Segmentation. *Lecture Notes in Computer Science*. P. 234–241. doi: 10.1007/978-3-319-24574-4\_28.
42. Badrinarayanan V., Kendall A. and Cipolla R. (2017). SegNet: A Deep Convolutional Encoder-Decoder Architecture for Image Segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*. 39(12). P. 2481–2495. doi: 10.1109/tpami.2016.2644615.
43. Zhao H., Shi J., Qi X., Wang X. and Jia J. (2017) Pyramid scene parsing network. In *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 21–26 July, Honolulu, HI, USA, IEEE*. P. 6230–6239. doi: 10.1109/cvpr.2017.660.
44. Lin T.-Y., Dollar P., Girshick R., He K., Hariharan B. and Belongie S. (2017) Feature pyramid networks for object detection. In *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 21–26 July, Honolulu, HI, USA, IEEE*. P. 936–944. doi: 10.1109/cvpr.2017.106.
45. Chen L.-C., Zhu Y., Papandreou G., Schroff F. and Adam H. (2018) Encoder-Decoder with Atrous Separable Convolution for Semantic Image Segmentation. *Computer Vision – ECCV 2018*. P. 833–851. doi: 10.1007/978-3-030-01234-2\_49.
46. Fu J., Liu J., Tian H., Li Y., Bao Y., Fang Z. and Lu H. (2019) Dual Attention Network for Scene Segmentation. In *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 15–20 June, Long Beach, CA, USA, IEEE*. doi: 10.1109/cvpr.2019.00326.
47. Zhang F., Chen Y., Li Z., Hong Z., Liu J., Ma F., Han J. and Ding E. (2019) ACFNet: Attentional class feature network for semantic segmentation. In *2019 IEEE/CVF International Conference on Computer Vision (ICCV), 27 October–2 November, Seoul, Korea (South), IEEE*. P. 6797–6806. doi: 10.1109/iccv.2019.00690.
48. Yuan Y., Chen X. and Wang J. (2020) Object-contextual representations for semantic segmentation. In *Computer Vision – ECCV 2020. Cham: Springer International Publishing*. P. 173–190. doi: 10.1007/978-3-030-58539-6\_11.
49. Krähenbühl P. and Koltun V. (2011) Efficient Inference in Fully Connected CRFs with Gaussian Edge Potentials. In *Advances in neural information processing systems*. P. 109–117. doi: 10.48550/arxiv.1210.5644.
50. Zheng S., Jayasumana S., Romera-Paredes B., Vineet V., Su Z., Du D., Huang C. and Torr P.H.S. (2015) Conditional random fields as recurrent neural networks. In *2015 IEEE International Conference on Computer Vision (ICCV), 7–13 December, Santiago, Chile, IEEE*. P. 1529–1537. doi: 10.1109/iccv.2015.179
51. Cheng H. K., Chung J., Tai Y.-W. and Tang C.-K. (2020) CascadePSP: Toward class-agnostic and very high-resolution segmentation via global and local refinement. In *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 13–19 June, Seattle, WA, USA, IEEE*. P. 8887–8896. doi: 10.1109/cvpr42600.2020.00891.
52. Rublee E., Rabaud V., Konolige K. and Bradski G. (2011) ORB: An efficient alternative to SIFT or SURF. In *2011 International Conference on Computer Vision, 6–13 November, Barcelona, Spain, IEEE*. P. 2564–2571. doi: 10.1109/iccv.2011.6126544
53. Alcantarilla P. F., Bartoli A. and Davison A. J. (2012) KAZE Features. In *Computer Vision – ECCV 2012. Berlin, Heidelberg : Springer Berlin Heidelberg*. P. 214–227. doi: 10.1007/978-3-642-33783-3\_16
54. Yang X. and Cheng K.-T. (2012). LDB: An ultra-fast feature for scalable Augmented Reality on mobile devices. *2012 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*. P. 49–57. doi: 10.1109/ismar.2012.6402537.
55. Lowe D. G. (1999) Object recognition from local scale-invariant features. In *Proceedings of the Seventh IEEE International Conference on Computer Vision, 20–27 September, Kerkyra, Greece, IEEE*. P. 1150–1157. doi: 10.1109/iccv.1999.79041

**Otyrba Rostislav R.** — PhD student, Department of Information Security and Processing Technologies, Faculty of Computer Sciences, Voronezh State University.

E-mail: [rostislav.otyrba.97@gmail.com](mailto:rostislav.otyrba.97@gmail.com)

ORCID iD: <https://orcid.org/0000-0002-0412-2465>

**Dryuchenko Mikhail A.** — PhD in Technical Sciences, Associate Professor, Department of Information Security and Processing Technologies, Faculty of Computer Sciences, Voronezh State University.

E-mail: [m\\_dryuchenko@mail.ru](mailto:m_dryuchenko@mail.ru)

ORCID iD: <https://orcid.org/0000-0001-8837-5875>

**Sirota Alexander A.** — DSc in Technical Sciences, Head of the Department of Information Security and Processing Technologies, Faculty of Computer Sciences, Voronezh State University.

E-mail: [sir@cs.vsu.ru](mailto:sir@cs.vsu.ru)

ORCID iD: <https://orcid.org/0000-0002-5785-8513>