

Использование методов глубокого обучения для анализа изображений аэрокосмического мониторинга в интересах обнаружения элементов новизны на местности

Р. Р. Отырба, email: rostislav.otyrba.97@gmail.com

А. А. Сирота, email: sir@cs.vsu.ru

М. А. Дрюченко, email: m_dryuchenko@mail.ru

Воронежский государственный университет

***Аннотация.** На основе методов глубокого обучения, разработано математическое и программное обеспечение в интересах создания системы для обнаружения изменений на аэрокосмических снимках городской, пригородной местности, акваторий и лесных насаждений. В результате исследования был подготовлен собственный набор данных, рассмотрены и сравнены различные модели свёрточных нейронных сетей, алгоритмы пост-обработки результатов сегментации, также был организован процесс получения и сшивки ранее снятых снимков целевой области и процесс коррекции перспективы снимков с итоговым обнаружением изменений.*

***Ключевые слова:** Обнаружение элементов новизны, глубокие нейронные сети, семантическая сегментация, сопоставление изображений.*

Введение

Обнаружение изменений – это аналитический процесс, целью которого является обнаружения новых или исчезнувших объектов на изображениях, зарегистрированных в разные моменты времени для одной и той же целевой области. Обнаружение изменений полезно и важно во многих сферах, например, для анализа последствий катастроф, планирования развития городских территорий, обнаружения незаконной вырубki или построек и т.д. [1].

Классический ручной поиск элементов новизны выполняется с использованием инструментальных методов при непосредственном участии оператора-интерпретатора изображений, что является трудоёмким и неточным процессом при частых изменениях.

Известных алгоритмических и программных решений, касающихся обнаружения изменений на местности, существует не так много. В основном, эти методы слишком сильно полагаются на геопространственную информацию. Более того, они также ограничены геометрическими аспектами, такими как: одинаковой ориентацией, масштабом, координатами и калибровкой камеры. Дополнительно,

исследования, направленные на обнаружение изменений в городской и пригородной среде ограничены конкретными топографическими объектами, чаще всего зданиями, и не учитывают другие значимые объекты для мониторинга.

Поскольку на практике анализируемые изображения содержат огромное количество тематической информации разного характера, создание многофункциональной системы обнаружения элементов новизны является важной задачей. Целью данного исследования является разработка математического и программного обеспечения в интересах создания системы аэрокосмического мониторинга изменений на местности с использованием методов глубокого обучения.

1. Предлагаемый алгоритм функционирования системы

1. На вход подаётся изображение местности, которое анализируется на предмет содержания возможных изменений.
2. Выполняется предобработка изображения (сопоставление гистограммы изображений и нарезка снимка на фрагменты)
3. Производится циклическая сегментация нарезанных фрагментов ранее обученной глубокой нейронной сетью и обратная их сшивка.
4. Результирующее сегментированное изображение подвергается пост-обработке (очистка от шума, упрощение геометрии и т.п.).
5. Подбирается ранее отснятое изображение (снимок) этой местности и формируется соответствующая бинарная маска с целевой областью с использованием GPS метаданных.
6. Далее снимок и маска сшивается с ближайшими по координатам снимками.
7. Выполняется коррекция перспективы нового снимка по отношению к полученному ранее отснятому снимку, посредством вычисления и использования матрицы гомографии.
8. Выполняется вычисление разности двух бинарных масок, фильтрация шумов и отображение непосредственно изменений.

2. Подготовка набора данных для обучения нейронных сетей

Для реализации поставленной цели был создан собственный набор данных, посредством ручной разметки снимков высокого разрешения (6792x4800) Google Earth Pro в онлайн сервисе CVAT.

Для обучения сети была произведена нарезка снимков на фрагменты размером 512x512 с перекрытием на 100 пикселей для учёта контекста. Итоговый состав набора данных следующий:

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

Примеры размеченных данных представлены на рис.1.

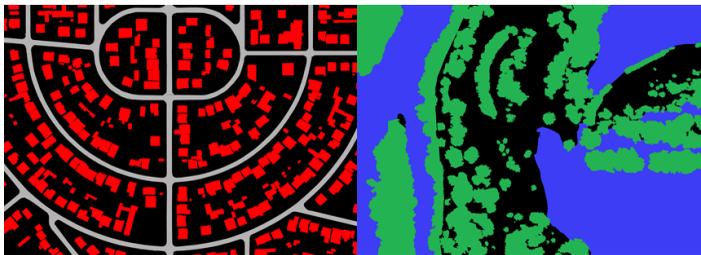


Рис. 1. Примеры размеченных снимков

Для лучшего качества обучения, была применена аугментация, которая происходит на лету во время обучения, это: горизонтальные и вертикальные отражения, случайные повороты на 90 градусов, масштабирование с случайными 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).

3. Обучение нейронных сетей

В качестве моделей нейронных сетей были рассмотрены 10 свёрточных нейронных сетей, а также условная генеративно-состязательная сеть pix2pix. После нахождения наилучшей модели, рассматривались такие кодирующие сети, извлекающие признаки как: VGG19, ResNet (101v2 и 152v2), DenseNet201 и EfficientNet (B3, B5 и B7). Нейронные сети, как уже упоминалось, использовались для сегментации анализируемых топографических объектов на целевых изображениях, что позволило использовать результирующую бинарную маску для качественного сопоставления свежего и ранее отснятого снимка, а также проводить коррекцию перспективы и последующей разности бинарных масок снимков для непосредственного выявления изменений на местности.

В качестве оптимизатора был выбран Момент Нестерова, который в отличие от адаптивных оптимизаторов, не так сильно переобучается, хорош в обобщающих способностях и поиске плоских локальных минимумов высокого качества, однако требует тщательной настройки скорости обучения [2].

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

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

где η_{\min} – минимальная скорость обучения, η_{\max} – максимальная скорость обучения, T_{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 $w(c)$ – 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 полученным результатом (формула 4), а фокальная ошибка – это модификация кросс-энтропии, занижающий вклад преобладающего класса и фокусируется на более сложных примерах (формула 5).

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

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

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

где λ – компромисс между функциями (0.5), TP – положительный класс распознанный как положительный, FN – отрицательный класс распознанный как отрицательный, FP – отрицательный класс распознанный как положительный.

Результаты обучений нейронных сетей на проверочных данных после 200 эпох обучения представлены в таблице 1. Для всех двух наборов данных самыми лучшими оказались архитектуры DeepLabV3+ и OCRNet с кодирующей сетью DenseNet201. Было решено использовать две модели как ансамбль, так как в результате экспериментов оказалось, что, они хорошо дополняют друг друга.

Таблица 1

Результаты на наборе данных «Здания и дороги»

Модель	Loss	IoU
FCN (DenseNet201)	0.29930	0.8208
U-Net (Classic)	0.37015	0.6571
SegNet (Classic)	0.35474	0.6680
PSPNet (DenseNet201)	0.35753	0.6565
FPN (DenseNet201)	0.31262	0.7527
DANet (DenseNet201)	0.29077	0.7636
OCRNet (DenseNet201)	0.21951	0.8806
ACFNet (DenseNet201)	0.27441	0.8010
DeepLabV3+ (DenseNet201)	0.22694	0.8582
Pix2Pix	0.27461	0.7621

Таблица 2

Результаты на наборе данных «Растительность и вода»

Модель	Loss	IoU
DeepLabV3+ (DenseNet201)	0.30472	0.7576
OCRNet (DenseNet201)	0.29729	0.7789
Pix2Pix	0.31861	0.7458

4. Постобработка результатов сегментации и коррекция перспективы

Результаты сегментирующих сетей необходимо обработать, так как часто они содержат шум и нечёткость. Для этого были рассмотрены следующие алгоритмы пост-обработки результирующей сегментации:

- Плотные условные случайные поля (Dense CRF);
- Рекуррентные условные случайные поля (RNN-CRF);
- Сегментационная модель рафинирования CascadePSP [3].

Для устранения незначительных артефактов после пост-обработки была дополнительно реализована следующая ручная обработка:

1. Метод пороговой обработки (метод Оцу);
2. Медианная фильтрация;
3. Поиск контуров и очистка от лишних объектов;
4. Морфологические операции (реконструкция, открытие, закрытие);
5. Модифицированный алгоритм упрощения геометрии Дугласа-Пекера, сохраняющий топологию объекта;
6. Медианная фильтрация в цикле.

В результате экспериментов, использование ансамбля моделей DeepLabV3+ и OCRNet в сочетании с сегментационной моделью пост-обработки CascadePSP и ручной обработки, дал наилучший прирост к общему качеству сегментации по показателю IoU: с 0.9163 до 0.9301.

Для получения ранее отснятого снимка целевой области использовались GPS метаданные снимков, с помощью которых находится наиболее кратчайшее расстояние до координаты целевой области, посредством вычисления геодезического расстояния, в которой используется эллипсоидальная модель Земли WGS-84. Когда ближайший ранее отснятый снимок найден, происходит сшивка вокруг его области, в том числе сшивка соответствующих масок, с помощью алгоритма масштабно-инвариантной трансформации признаков SIFT.

Для коррекции перспективы нового снимка по отношению к полученному ранее снимку производится расчёт матрицы гомографии, которая в процессе расчёта использует лучшие совпадения по ключевым признакам. Для их нахождения рассматривались:

1. Алгоритмы выявления локальных признаков изображения: SIFT, AKAZE, ORB;
2. Методы сопоставления признаков. Рассматривались два варианта: Brute-Force Matcher с перекрёстной проверкой и FLANN Based Matcher с тестом соотношения Дэвида Лоу [4];
3. Методы вычисления расстояния между дескрипторами: расстояние Хэмминга или L2 норма;
4. Методы вычисления гомографии: RANSAC, LMEDS, RHO.

Для нахождения наилучшего алгоритма, был произведён эксперимент на проверку стабильности данных алгоритмов на 5 парах снимков, зарегистрированных в разные моменты времени, при различных сочетаниях таких факторов как: шум и яркость, масштаб, угол. В результате, SIFT алгоритм в сочетании с FLANN и методом вычисления гомографии RHO оказался наиболее стабильным алгоритмом, показав самое хорошее количество удачных совпадений при разных факторах, поэтому, в итоге, выбор пал именно на него.

5. Демонстрация работы разработанной системы

На рис 3 в центре представлен входной свежий снимок, к которому с помощью GPS метаданных подобран один из ближайших ранее отснятых снимков с целевой областью (снимок с зелёной границей), который далее сшивается с ближайшими к нему остальными снимками. Для свежего снимка далее производится сопоставление гистограмм по яркости и цветовой насыщенности, для лучшей сегментации и дальнейшего сопоставления снимков. После этого производится сегментация и пост-обработка отдельно для топографических объектов.

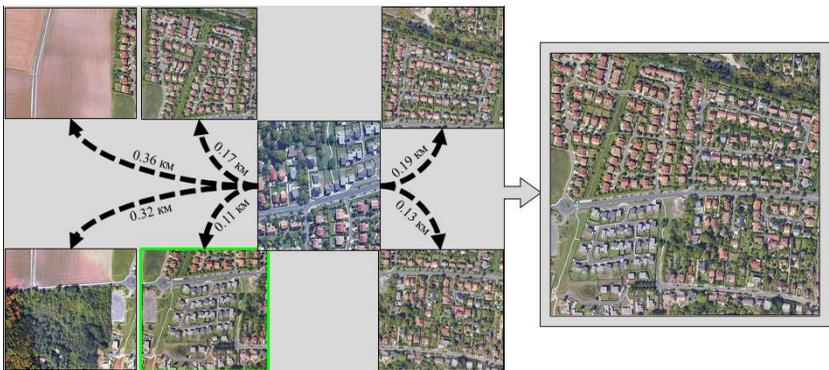


Рис. 2. Процесс шивки целевой области

В финальном этапе производится коррекция перспективы, разность двух бинарных масок, фильтрация шумов (рис. 4) и отображение изменений на исходный снимок (рис. 5).



Рис. 3. Результат сопоставления и разницы двух масок.



Рис. 4. Обнаруженные изменения

Заключение

Данная статья посвящена разработке математического и программного обеспечения системы для мониторинга элементов новизны на местности. В результате был подготовлен собственный набор данных, подготовлены модели глубоких нейронных сетей, проведена пост-обработка, организовано получение и обработка ранее отснятого снимка и процесс коррекции перспективы с итоговым обнаружением изменений.

Список литературы

1. Zhang, C. Detecting large-scale urban land cover changes from very high resolution remote sensing images using CNN-based classification / C. Zhang, S. Wei, S. Ji // ISPRS International Journal of Geo-Information – 2019. – Т. 8. – № 4. – С. 189.
2. Wilson, A. The marginal value of adaptive gradient methods in machine learning / A. Wilson, R. Roelofs, M. Stern, N Srebro, B Recht // In Advances in Neural Information Processing Systems – 2017. – С. 4148–4158.
3. Ho, C. CascadePSP: Toward Class-Agnostic and Very High-Resolution Segmentation via Global and Local Refinement / C., Ho, C. Jihoon, T. Yu-Wing, T. Chi-Keung // Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition – 2020. – С. 8890–8899.
4. Lowe, D. Distinctive Image Features from Scale-Invariant Keypoints / D. Lowe // International Journal of Computer Vision – 2004. – Т. 60. – № 2. – С. 91–110.