

Брызгин Георгий Константинович, студент,
«Московский авиационный институт
(национальный исследовательский институт)»,
г. Москва

МЕТОДИКА ИЗВЛЕЧЕНИЯ КОНТУРОВ ЗДАНИЙ ИЗ ОФФ-НАДИР МЕТОДАМИ ГЛУБОКОГО ОБУЧЕНИЯ

Аннотация: Извлечение контуров зданий с аэроснимков необходимо для точного картографирования городов с помощью аэрофотосъемки и технологий компьютерного зрения. Существующие подходы в основном предполагают, что крыша и контур здания хорошо совпадают, что может не соответствовать действительности на аэроснимках с ненадирным углом съемки, так как между ними часто существует значительное смещение. В данной работе была предложена архитектура глубокого обучения, в которой присутствует вектор смещения, т.е. задача извлечения контуров зданий на офф-надирных изображениях предстает в виде проблемы совместного предсказания на уровне экземпляров крыши здания и соответствующего вектора смещения от крыши к контуру. Таким образом, контур можно оценить, переместив предсказанную маску крыши согласно предсказанному вектору смещения.

Ключевые слова: компьютерное зрение, глубокое обучение, обработка спутниковых снимков, нейронные сети.

Введение

Автоматическое извлечение контуров зданий (АИКЗ) с аэроснимков городских сцен изучается уже десятилетиями и приносит пользу множеству задач геоматике и наблюдению Земли, таких как 3D моделирование городов, обнаружение изменений зданий и точное городское планирование [3] [4] [5] [6]

Ранние работы решали проблему АИКЗ, в основном используя структурные и визуальные особенности изображений для характеристики контуров зданий [7] [8] [9]. Однако эффективность этих методов часто ограничивается дискриминационными способностями поверхностных признаков. Формирование масок оснований зданий является одной из ключевых задач в области компьютерного зрения и геоинформатики. Маски оснований зданий представляют собой бинарные или многоуровневые карты, где каждый пиксель изображения отмечен как принадлежащий зданию или окружающей его территории. Такие карты широко используются в различных областях, включая городское планирование, строительство, мониторинг состояния объектов, создание 3D-карт и управление природными ресурсами.

Офф-надир (off-nadir) – это угол съемки, при котором камера или сенсор направлены не вертикально вниз (надир), а под углом к поверхности земли. Офф-надир съемка позволяет получить больше информации о структуре объектов, но также приводит к искажениям и изменению перспективы.

Автоматическое формирование масок оснований зданий имеет огромное значение в современном мире. Этот процесс позволяет значительно сократить время и затраты на ручную разметку изображений, а также увеличить точность и масштабируемость анализа инфраструктуры населенного пункта или выбранной местности. Автоматизированные методы могут быстро обрабатывать большие объемы данных, что особенно важно в условиях постоянного роста количества спутниковых снимков и аэрофотоснимков, которые помогают отслеживать изменения на местности [1] [2]



Извлечение контуров зданий с аэроснимков необходимо для точного картографирования городов с помощью фотограмметрических технологий компьютерного зрения. Существующие подходы в основном предполагают, что крыша и контур здания хорошо совпадают, что может не соответствовать действительности на аэроснимках с ненадирным углом съемки, так как между ними часто существует значительное смещение.

1. Существующие подходы

В последнее время решения на основе глубокого обучения [10] [11] [12] показали многообещающие результаты в АИКС, благодаря мощным возможностям глубоких моделей в обучении представлением. Однако методы АИКС, упомянутые выше, в основном фокусируются на изображениях с почти вертикальной съемкой, где проекции крыши и контура здания обычно хорошо совпадают, и проблема сводится к извлечению видимых крыш зданий.



а – оригинальный офф-надир снимок, б – модель здания

Рис. 1. Модель здания на офф-надир снимке

Лишь немногие из них могут работать с офф-надирными изображениями, которые часто получаются при большом угле зрения аэрофотосъемочной системы (например, больше 25 градусов). Пример офф-надирных изображений и модель здания в них проиллюстрированы на рисунке 1. Можно заметить, что для здания на офф-надирных изображениях часто наблюдается заметное смещение между проекциями крыши и контура. Более того, граница контура здания частично видна и сильно закрыта фасадом. Все это создает большие трудности для точного обнаружения контуров зданий на оффнадирных изображениях.

Несмотря на трудности, существует большой спрос на извлечение контуров зданий на уровне пикселей с ненадирных изображений, как, например, в задаче SpaceNet 4 [13] (соревнования). Контур крыши и здания часто совпадают для большинства одиночных зданий в городской среде, однако все так же остаются нерешенными следующие проблемы:

Геометрические искажения:

Объекты на изображениях, снятых под углом, выглядят деформированными. Трудности в определении точных границ объектов.

Окклюзия:

Части объектов могут быть скрытыми за другими объектами или элементами инфраструктуры. Сложность в создании полных масок оснований зданий.

Изменение перспективы:

Объекты на переднем плане кажутся больше, чем на заднем.

Необходимость учета перспективных изменений при создании масок



Целью настоящей работы является устранения влияния выше перечисленных проблем на результаты обработки снимков, а так же возможное замещение существующих методов, путем разработки методики извлечения данных из офф-надир снимков при помощи методов глубокого обучения нейронных сетей.

2. Методы исследований

Современные модели глубокого обучения значительно улучшили точность и устойчивость методов сегментации. Такие модели, как PANet [14],

Mask R-CNN [15] и HRNet-v2 [16], демонстрируют высокую производительность и адаптивность в различных условиях съемки. Они способны автоматически выделять признаки и обучаться на больших объемах данных, что делает их достаточно хорошими вариантами решения задач сегментации зданий. Однако они также сталкиваются с определенными вызовами:

Необходимость в большом количестве размеченных данных для тренировки.

Данные должны иметь хорошее разрешение (от 1 м. на пиксель), иметь качественную разметку зданий, а также полную информацию о угле съемки, месте, для дальнейшего синтеза дополнительных данных [17]. Высокие вычислительные затраты.

Трудности в интерпретации и объяснении результатов модели.

Mask R-CNN [15]: Расширение Faster R-CNN, добавляющее задачу сегментации к обнаружению объектов. Mask R-CNN позволяет одновременно обнаруживать и сегментировать объекты на изображении, что делает её мощным инструментом для анализа спутниковых снимков.

PANet [14]: Основное преимущество заключается в ее способности улучшать поток информации через всю сеть, что обеспечивает более точное определение границ и контуров объектов на изображениях. Данная архитектура отличается несколькими ключевыми моментами: дополнение пути снизу вверх (улучшение всей иерархии признаков, включая точные сигналы локализации из нижних слоев), адаптивное извлечение признаков (Эта технология связывает сетку признаков и все уровни признаков, позволяя полезной информации на каждом уровне эффективно использоваться в последующих подсетях предложений. Это обеспечивает точное использование релевантной информации при сегментации объектов.) и дополнительные ветви (захватывают различные виды для каждого предложения, улучшая предсказание масок).

HRNet-v2 [16]: Основные преимущества HRNet-v2: высокое разрешение (сохраняет высокое разрешение на всех уровнях сети, что позволяет лучше сохранять детали изображений и улучшает точность детектирования и сегментации), устойчивость к изменениям масштаба (благодаря своей структуре, лучше справляется с объектами различных размеров, что делает его идеальным для приложений, требующих высокой точности на разных масштабах), многоуровневая связность (архитектура использует многоуровневые связи, что позволяет обмениваться информацией между различными частями сети и улучшает общую производительность), гибкость и масштабируемость

Основной нерешенной проблемой являются геометрические искажения при офф-надире [18] [19] [20] [21].

3. Проблематика геометрических искажений при офф-надире

Когда съемка осуществляется при офф-надире, изображения объектов на земле испытывают значительные геометрические искажения. Эти искажения вызваны тем, что объекты, которые находятся дальше от центра изображения, кажутся деформированными из-за угла съемки. Это создает множество проблем при автоматическом формировании масок оснований зданий, большинство которых можно отнести к одной тематике: деформация объектов и изменение их размеров.



4. Предлагаемая методика решения

Вместо прямого извлечения контуров зданий предлагается решать проблему автоматического извлечения контуров зданий на ненадирных изображениях, одновременно изучая крышу здания на уровне экземпляра и соответствующий вектор смещения к контуру для каждого здания.

К типичным методам сегментации, таким как Mask R-CNN, можно применить offset head в целях предсказания вектора смещения "крыша-контур" для каждого здания. Контур здания затем оценивается путем перевода предсказанных масок крыш в соответствии с предсказанными векторами смещения. Поскольку крыши видны для большинства зданий, а векторы смещения фактически встроены в видимые структуры фасадов, данный метод может достичь лучшей производительности АИКЗ, чем методы, которые напрямую моделируют контуры зданий.

Замечая, что длина вектора смещения "крыша-контур" в основном зависит от высоты здания, в то время как угол вектора смещения чувствителен к углу обзора системы съемки, можно сделать вывод, что этот вектор может плохо сходиться при обучении offset head с ограниченным количеством обучающих выборок смещения.

Основная трудность извлечения границ для всех зданий – частичная видимость зданий, может быть решена путем обучения с учителем глубокой модели АИКЗ на наборе из N данных с метками уровня экземпляра D , то есть, $D = \{ (I_i, T_i); i = 1, \dots, N \}$, где T_i является соответствующей меткой зданий, содержащихся на изображении I_i в наборе данных D . В частности, T_i состоит из K контуров $\{f_i^j; j = 1, \dots, K\}$, причем каждый $f_i^j \in R^{2 \times M}$ представлен многоугольником с M вершинами.

Более того, поскольку контуры зданий частично видимы, а крыша всегда полностью видна, задача может быть преобразована в оценку крыши, связанной с вектором смещения к контуру для каждого здания. Таким образом, каждая метка контура f_{ij} в наборе данных D соответствует метке крыши r_{ji} дополнительной метке смещения $o_{ij} \in R^2$, которая представляет собой вектор смещения от крыши к контуру j -го здания на изображении I_i .

Поэтому аннотированные данные будут:
$$D = \left\{ \left(I_i, \{f_i^j, r_i^j, o_i^j\}_{j=1}^K \right); i = 1, \dots, N \right\}$$

Пример аннотации на уровне экземпляра здания на ненадирных изображениях показан на рисунке 1.

Проблему АИКЗ следует рассматривать как задачу оценки крыши здания на уровне экземпляра и одновременного предсказания её вектора смещения. Для изучения вектора смещения на уровне экземпляра для каждого здания, введем вектора смещения в рамках сегментации экземпляров сверху вниз, как показано на рисунке 1.

Для уточнения предсказания вектора смещения и уменьшения влияния шумов, возможна разработка и внедрение отдельного модуля увеличения смещения на уровне признаков.

Предполагаемая модель может решить проблему АИКЗ на офф-надир изображениях путем предсказания крыш и их связанных векторов смещения к контурам зданий с использованием набора данных D . Идея концептуально проста: добавление нового вектора смещения для изучения вектора смещения в рамках сегментации экземпляров сверху вниз, используемую для обучения крыш зданий. Идейно, предлагаемая модель может работать с многими предлагаемыми ранее моделями сегментации.

Общая предлагаемая архитектура иллюстрируется на рисунке 3. Для обучения модели требуются истинные метки ограничивающих прямоугольников зданий (B-Bbox) $\{b_i^j\}_{j=1}^K$,



маски крыш (R-Mask) $\{r_i^j\}_{j=1}^K$ и векторы смещения (Offset) $\{o_i^j\}_{j=1}^K$, где каждый $b_i^j \in R^4$ обозначен

как $\{(x_{ij}, y_{ij}), w_{ij}, h_{ij}\}$, а каждый $r_{ij} \in R^{2 \times M}$ представлен многоугольником из M

вершин. Эта информация может быть получена из меток $T_i = \{f_i^j, o_i^j\}_{j=1}^K$.

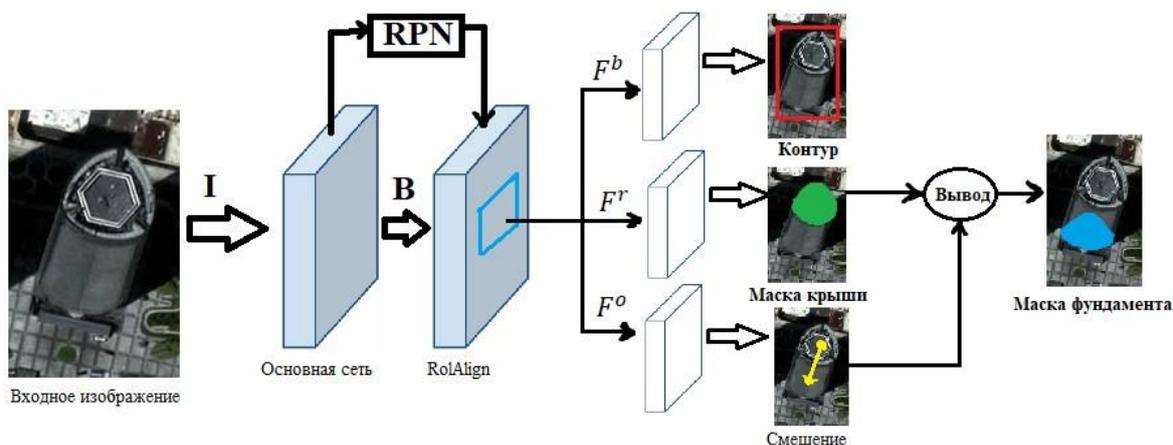


Рис. 2. Схема предлагаемой архитектуры (основанная на Mask R-CNN).

Основой служат ResNet с Feature Pyramid Network (FPN), а RPN обозначает Region Proposal Network. Слои RoIAlign создают карты признаков Fb , Fr и Fo для генерации ограничивающего прямоугольника здания, маски крыши и вектора смещения. Маска контура фундамента является финальной

На этапе обучения, используя набор данных с метками D , входное изображение I подается в базовую сеть, которая производит карту признаков базы B , как показано на рисунке 3. Затем Region Proposal Network (RPN) [22] использует карту признаков B в качестве входных данных для генерации предложений регионов P . Далее три слоя RoI Align [15] принимают каждое предложение $p_i = (x_i, y_i, w_i, h_i) \in P$ из P и B в качестве входных данных для вычисления карт признаков Fb , Fr и Fo , где p_i обозначает i -е предложение для здания, Fb , Fr и Fo являются картами признаков ограничивающего прямоугольника здания, маски крыши и вектора смещения соответственно в R-CNN head, mask head и offset head. R-CNN и mask heads наследуются от Mask R-CNN [15]. Для вектора смещения карта признаков Fo используется в качестве входных данных и состоит из нескольких сверточных и полностью связанных слоев. Обратите внимание, что истинные метки для RPN и R-CNN head являются ограничивающими прямоугольниками зданий, поскольку поле зрения сети в векторе смещения должно охватывать крышу и контур одновременно для регрессии вектора смещения, то есть информация о векторе смещения встроена в видимую структуру фасада здания.

Архитектура в конечном итоге получается путем минимизации совместной функции потерь: $L = LRPN + \alpha_1 LR-CNN + \alpha_2 LMask + \alpha_3 LOffset$, где

$LRPN$, $LR-CNN$, $LMask$ те же, что и в Mask R-CNN, то есть потери для RPN, RCNN и mask heads соответственно. $LOffset$ – это потеря для вектора смещения, где используется стандартная smooth L1 Loss. Для ускорения сходимости вектора смещения используются следующие функции кодирования: $\phi_x = \frac{o_x}{w^p}$, $\phi_y = \frac{o_y}{h^p}$, где w^p и h^p – ширина и высота соответствующего предложения $p \in P$, $[o_x, o_y]T$ – истинный вектор смещения, а $[\phi_x, \phi_y]T$ – закодированный вектор смещения для регрессии.



Во время стадии предсказания используется предсказанные векторы смещения для преобразования предсказанных масок крыш в контуры. Конкретно, предсказанная маска крыши сначала будет представлена многоугольником $r \in R^{2 \times M}$ с M вершинами с помощью алгоритма топологического структурного анализа [23]. Контурный многоугольник f в конечном итоге вычисляется путем смещения многоугольника крыши r с предсказанным смещением $\hat{d} = [\hat{d}_x, \hat{d}_y]^T$.

Новизна предлагаемого решения состоит в внедрении вектора смещения масок при использовании методов глубокого обучения НС, что, в отличие от известных способов, позволяет улучшить извлечение контуров здания при геометрических искажениях или окклюзиях.

Заключение

Проблема автоматического извлечения контуров зданий на оффнадирных изображениях актуальна из-за частичного перекрытия зданий и сложностей с точным определением их границ. Предлагаемая методика решает эту задачу, одновременно анализируя крыши зданий и соответствующие векторы смещения к контурам.

В данной работе была рассмотрена проблема извлечения контуров зданий на оффнадирных изображениях. Используя свойства, что контуры зданий частично видимы, а крыша полностью видна на оффнадирных изображениях, возможно декомпозировать проблему извлечения контуров зданий на задачу извлечения крыши здания с наложением регрессии вектора смещения "крыша-контур". Предлагаемый offset head можно легко применить к любому методу сегментации экземпляров сверху вниз, что делает данный подход более гибким и универсальным.

Кроме того, возможна интеграция модуля увеличения смещения на уровне признаков для уточнения предсказания вектора смещения, избегая значительных дополнительных вычислений, характерных для традиционного увеличения на основе изображений.

Список литературы:

1. Lei Ni, Chunlei Huo, Xin Zhang, Peng Wang, Luyang Zhang, Kangkang Guo and Zhixin Zhou NaGAN:Nadir-like Generative Adversarial Network for OffNadir Object Detection of Multi-View Remote Sensing Imagery // MDPI. 2022.
2. Mask-RCNN-for-Off-Nadir-Building-Detection // GitHub URL: <https://github.com/ashnair1/Mask-RCNN-for-Off-Nadir-Building-Detection> (дата обращения: 10.10.2024).
3. C.Benedek, X. Descombes, J. Zerubia, "Building development monitoring in multitemporal remotely sensed image pairs with stochastic birth-death dynamics," TPAMI, vol. 34, no. 1, pp. 33–50, 2011
4. J. Mahmud, T. Price, A. Bapat, J.-M. Frahm, "Boundary-aware 3d building reconstruction from a single overhead image," in CVPR, 2020, pp. 441–451
5. J. Yuan, "Learning building extraction in aerial scenes via convolutional network," TPAMI, vol. 40, no. 11, pp. 2793–2798, 2017
6. D. Marcos, D. Tuia, B. Kellenberger, L. Zhang, M. Bai, R. Liao, R. Urtasun, "Learning deep structured active contours end to-end," in CVPR, 2018, pp. 8877–8885
7. J. A. Shufelt, "Performance evaluation and analysis of monocular building extraction from aerial imagery," TPAMI, vol. 21, no. 4, pp. 311–326, 1999
8. M. Ortner, X. Descombes, J. Zerubia, "A marked point process of rectangles and segments for automatic analysis of digital elevation models," TPAMI, vol. 30, no. 1, pp. 105–119, 2007



9. J. Inglada, "Automatic recognition of man-made objects in high resolution optical remote sensing images by svm classification of geometric image features," *ISPRS J. Photogramm. Remote Sens.*, vol. 62, no. 3, pp. 236–248, 2007.
10. D. Cheng, R. Liao, S. Fidler, R. Urtasun, "Darnet: Deep active ray network for building segmentation," in *CVPR*, 2019, pp. 7431–7439.
11. Q. Zhu, C. Liao, H. Hu, X. Mei, H. Li, "Map-net: Multiple attending path neural network for building footprint extraction from remote sensed imagery," *IEEE Trans. Geosci. Remote Sensing*, pp. 1–13, 2020.
12. M. Li, F. Lafarge, R. Marlet, "Approximating shapes in images with lowcomplexity polygons," in *CVPR*, 2020, pp. 8633–8641.
13. SpaceNet 4: Off-Nadir Buildings // Spacenet.ai URL: <https://spacenet.ai/offnadir-building-detection/> (дата обращения: 10.09.2024).
14. S. Liu, L. Qi, H. Qin, J. Shi, and J. Jia, "Path aggregation network for instance segmentation," in *CVPR*, 2018, pp. 8759–8768.
15. K. He, G. Gkioxari, P. Dollar, R. Girshick, "Mask R-CNN," in *ICCV*, 2017, pp. 2961–2969
16. K. Sun, Y. Zhao, B. Jiang, T. Cheng, B. Xiao, D. Liu, Y. Mu, X. Wang, W. Liu, and J. Wang, "High-resolution representations for labeling pixels and regions," *CoRR*, vol. abs/1904.04514, 2019.
17. https://www.researchgate.net/publication/357087432_S2Looking_A_Satellite_Side-Looking_Dataset_for_Building_Change_Detection
18. Geometric Distortion in Imagery // Government Of Canada URL: <https://naturalresources.canada.ca/maps-tools-and-publications/satellite-imagery-elevationdata-and-air-photos/tutorial-fundamentals-remote-sensing/satellites-and-sensors/geometric-distortion-imagery/9401> (дата обращения: 08.08.2024).
19. Nicholas Weir, David Lindenbaum, Alexei Bastidas, Adam Van Etten, Sean McPherson, Jacob Shermeyer, Varun Kumar, Hanlin Tang SpaceNet MVOI: a Multi-View Overhead Imagery Dataset // *IEEE International Conference on Computer Vision. IEEE International Conference on Computer Vision. 2020.*
20. Chunyuan WANG, Ye ZHANG Yang WU, Yanfeng GU Highly Accurate Geometric Correction for Seriously Oblique Aero Remote Sensing Image Based on the Piecewise Polynomial Model // *Journal of Computational Information Systems. 2011.*
21. Chunyuan Wang, Xiang Liu, Xiaoli Zhao, Yongqi Wang An Effective Correction Method for Seriously Oblique Remote Sensing Images Based on Multi-View Simulation and a Piecewise Model // *MDPI. 2016.*
22. S. Ren, K. He, R. Girshick, J. Sun, "Faster R-CNN: Towards real-time object detection with region proposal networks," in *NeurIPS*, 2015, pp. 91–99.
23. S. Suzuki et al., "Topological structural analysis of digitized binary images by border following," *Computer Vision, Graphics, and Image Processing*, vol. 30, no. 1, pp. 32–46, 1985

