

УДК 528.856:004.85

І. А. Радіон*

аспірант, ORCID: <https://orcid.org/0000-0001-9831-454X>

Кафедра геодезії, землевпорядкування та кадастру

Волинський національний університет імені Лесі Українки, проспект Волі, 13, Луцьк, Україна, 43025

О. В. Мельник

к.т.н., доцент, ORCID: <https://orcid.org/0000-0002-5429-4038>

Кафедра геодезії, землевпорядкування та кадастру

Волинський національний університет імені Лесі Українки, проспект Волі, 13, Луцьк, Україна, 43025

*автор-кореспондент, e-mail: radionigor11@gmail.com

Сучасні методи нейронних мереж для векторизації будівель із супутникових знімків високої роздільної здатності

Цитувати як:

Радіон, І. А., Мельник, О. В. (2025). Сучасні методи нейронних мереж для векторизації будівель із супутникових знімків високої роздільної здатності, 24, 340-347. [https://doi.org/10.36910/6775-2410-6208-2025-14\(24\)-28](https://doi.org/10.36910/6775-2410-6208-2025-14(24)-28)

© 2025, Радіон І. А., Мельник О. В.

Автоматична векторизація будівель із супутникових знімків є ключовою задачею для картографування та кадастру. Сучасні методи глибокого навчання досягли високої растрової точності (IoU 85-92%), однак існує фундаментальна проблема: оптимізація сегментації не гарантує генерацію геометрично та топологічно коректних векторних полігонів. Дослідження фіксують значні відхилення кутів (до 8.3°), непаралельність стін та високий відсоток топологічних помилок (12-18%). Також викликами залишаються погана генералізація моделей на нові регіони та пропуск малих об'єктів.

Систематизація та аналіз сучасних методів глибокого навчання для векторизації будівель, з акцентом на проблемах геометричної регулярності, топологічної коректності та генералізації.

Проведено аналіз публікацій за 2015-2024 роки (CVPR, ISPRS та ін.) з використанням еталонних датасетів (SpaceNet, WHU, INRIA). Метрики оцінювання включали IoU та F1-score для растрової точності, а також PoLiS та Chamfer Distance для векторної. Методи класифіковано на три групи: CNN-орієнтовані (U-Net, DeepLab), трансформерні (Swin, SegFormer) та комплексні (end-to-end) методи (Frame Field Learning, GNN).

CNN-архітектури залишаються ефективною базою. Трансформери демонструють найвищу растрову точність (IoU >90%), але є обчислювально дорогими. Комплексні (end-to-end) методи, такі як Frame Field Learning та PolyWorld, що генерують вектори напряму, оминаючи етап полігонізації, показують дещо нижчу растрову точність, але значно кращу векторну якість (PoLiS ~73%), що є критичним для кадастрових задач.

Існує компроміс: трансформери лідирують за растровою точністю (IoU 85-92%), тоді як end-to-end методи (IoU 82-88%) забезпечують значно вищу векторну якість (PoLiS 70-73%). Перспективними напрямками є інтеграція геометричних обмежень в архітектуру мереж, розробка топологічно-усвідомлених функцій втрат, покращення генералізації та мультимодальні підходи, що поєднують оптичні знімки з даними LiDAR.

Ключові слова: векторизація будівель, глибоке навчання, нейронні мережі, супутникові знімки, семантична сегментація, трансформери, полігонізація, топологія

Вступ

Аналіз літературних джерел та постановка проблеми.

Автоматична векторизація будівель із супутникових знімків високої роздільної здатності є критично важливою задачею для картографування території, кадастрового обліку, міського планування, аналізу урбанізації та при оцінці пошкоджень внаслідок надзвичайних ситуацій. Традиційні методи, які базуються на класичних алгоритмах машинного навчання та детектуванні країв будівель, мають суттєві обмеження в складних міських середовищах із щільною забудовою та різноманітною архітектурою будівель [1].

За останнє десятиліття методи глибокого навчання продемонстрували значний прогрес у вирішенні цієї задачі. Архітектура U-Net, адаптована для задач дистанційного зондування Землі [2], стала базовою моделлю, завдяки своїй структурі у вигляді енкодера-декодера з пропускними з'єднаннями, що дозволяє ефективно виявляти ознаки в різних масштабах. Подальші дослідження показали, що інтеграція механізмів «уваги» значно покращує екстракцію будівель з аерознімків [3].

Сучасні підходи можна класифікувати на три основні категорії: методи семантичної сегментації на основі згорткових нейронних мереж (Convolutional Neural Network), трансформерні архітектури та комплексні (end-to-end) векторизаційні методи. Модель DeepLab v3+ демонструє значення ступеня перетину між двома обмежувальними рамками (IoU) 87,3% на наборі даних аерофотознімків та супутникових знімків зразків будівель WHU Building Dataset [4], тоді як трансформерні моделі, зокрема UNetFormer, досягають IoU 90,2% на наборі даних INRIA Aerial Dataset [5]. В роботі [6] було запропоновано інноваційну інтеграцію CNN у сумісності з трансформером через механізм самоуваги до багатомасштабних диференційних ознак, що дозволяє балансувати глобальні та локальні ознаки.

Однак, незважаючи на високу точність сегментації за метрикою IoU (85-92%), сучасні методи глибокого навчання мають критичні проблеми з генерацією геометрично правильних векторних контурів. Дослідження показують, що базові CNN-архітектури генерують нерегулярні полігони з

відхиленням кутів у середньому на $8,3^\circ$, непаралельністю протилежних стін на $5,7^\circ$ та зміщенням вершин на 2-4 пікселя [7]. Топологічна коректність також залишається проблемою – навіть «state-of-the-art» моделі генерують топологічно некоректні полігони у 12-18 % випадків [8].

Додатковими викликами є погана генералізація, в залежності від географічних умов досліджуваних території (падіння IoU з 89,2 % до 67,8 % без донавчання [9], пропуск малих будівель у 35-40 % випадків та проблеми з оклюзіями від дерев і тіней, що знижують точність на 15,3 % [10]). Існує фундаментальна проблема між точністю растрової сегментації та якістю геометрії, після векторизації. Оптимізація IoU не гарантує правильної геометричної структури та топології полігонів.

Мета і завдання дослідження. Метою дослідження є систематизація та аналіз сучасних методів глибокого навчання для векторизації будівель із супутникових знімків високої роздільної здатності з акцентом на проблемах геометричної регулярності, топологічної коректності та практичному застосуванні.

Завдання дослідження включають: (1) аналіз архітектур згорткових нейронних мереж та трансформерних моделей для семантичної сегментації будівель; (2) дослідження методів постобробки та полігонізації растрових масок; (3) виявлення основних проблем та обмежень сучасних підходів; (4) визначення перспективних напрямків для вдосконалення методів векторизації.

Матеріали та методи

У дослідженні проаналізовано публікації за період з 2015-2024 років у провідних журналах та конференціях з комп'ютерного зору та дистанційного зондування (CVPR, ICCV, ISPRS, Remote Sensing). Основна увага приділена методам, що використовують супутникові знімки високої роздільної здатності (0.3-1.5 м/піксель) таких систем як WorldView-3/4, GeoEye-1, Pleiades.

Для порівняльного аналізу використано результати експериментів на стандартних еталонних датасетах: SpaceNet, WHU Building Dataset, INRIA Aerial Image Labeling Dataset та Massachusetts Building Dataset. Основними метриками оцінювання є IoU, Precision (точність), F1-score (середнє гармонічне між точністю і повнотою), для растрової сегментації, а також PoLiS (Polygon Line String metric, метрика для порівняння багатокутників і відрізків ліній) та Chamfer Distance (Відстань Чамфера) для векторної точності.

Класифікацію методів проведено за трьома категоріями: (1) семантична сегментація, яка базується на CNN (U-Net, ResUNet, DeepLab, SAM); (2) трансформерні архітектури (Swim Transfromer, SegFormer); (3) комплексна (end-to-end) векторизація (Frame Field Learning, GNN методи).

Для кожної категорії проаналізовано архітектурні особливості, точність, обчислювальну складність та типові проблеми.

Результати та обговорення

Згорткові нейронні мережі. U-Net залишається найпоширенішою базовою архітектурою завдяки своїй простоті та ефективності. Модифікації U-Net показали покращені результати: ResUNet інтегрує залишкові з'єднання для тренування глибших мереж, а Attention U-Net використовує механізми уваги для фокусування на релевантних регіонах [2,3]. DeepLab v3+ з ASPP захоплює багатомасштабний контекст, досягаючи IoU 87,3% на WHU Dataset [4].

Важливим напрямком є мультизадачні архітектури. Ванг та ін. [6] у моделі MDFA-Net запропонував ієрархічно багатомасштабний модуль розширеної згортки (HMDCov) для екстракції локальних ознак та модуль механізму самоуваги диференційованих ознак (DFSA) для глобального контексту. Така гібридна архітектура ефективно балансує між виявленням глобальних та локальних ознак. Методи мультизадачного навчання із виявленням меж сегментують будівлі та виявляють їх межі, що покращує точність контурів на 6,8 %.

Адаптація Segмент Anything Model (SAM) для задач ДЗЗ є активною областю досліджень. У імплементації із методом Low-Rank Adaptation (LoRA), який дозволяє ефективно донавчати великі сегментаційні моделі на спеціалізованих наборах даних, SAM демонструє свою високу результативність у задачах екстракції будівель.

Трансформерні архітектури. Візійні трансформери (ViT, Vision Transformers) впровадили механізми самоуваги у задачі екстракції будівель. Свін-трансформер (Swin Transformer), з його ієрархічною структурою та механізмом зсувних (ковзних) вікон показав IoU 90,2 % на INRIA Dataset, що на 3,1 % краще ніж методи, які базувались на глибоких нейронних мережах [5]. Сегментаційний трансформер (SegFormer, Segmentation TransFormer) поєднує ефективність трансформерів з легковаговим декодером, забезпечуючи баланс між точністю та швидкістю [12].

Спеціалізовані архітектури інтегрують диференційовані геометричні обмеження для кращого моделювання регулярних структур будівель. Однак, трансформерні моделі мають вищу обчислювальну складність та потребують більших датасетів для тренування порівняно із CNN. Для практичного застосування часто використовують гібридні архітектури, що комбінують переваги обох підходів.

Комплексна (end-to-end) векторизація та полігонізація. Традиційний пайплайн (сегментація – постобробка – полігонізація) має обмеження через втрати інформації на кожному етапі. Навчання каркасних полів (Frame Field Learning) пропонує диференційовану процедуру векторизації, передбачаючи локальні каркаси для побудови

регуляризованих полігонів [13]. PolyWorld моделює будівлі як графи з вершинами-кутами та ребрами-стінами, використовуючи графові нейронні мережі (GNN) для ітеративної корекції, щоб досягти значень PoLiS 73,2% проти 58,4% традиційних методів [14].

Методи постобробки, такі як алгоритм Рамера-Дугласа-Пекера чи інші, часто погіршують ситуацію, надмірно спрощуючи складні архітектурні елементи. Системи ієрархічного контролю (Hierarchical supervision framework) навчають мережу на трьох рівнях (піксель, межа, вершина), забезпечуючи зменшення похибки позиціонування вершин на 1,2 пікселя [15].

Основні проблеми та обмеження. Першою проблемою є геометрична нерегулярність методів, що спричиняє відхилення кутів будівель, зміщення меж та непаралельність стін, що є неприпустимим для кадастрових застосувань, де потрібна висока точність.

Наступною проблемою є топологічні помилки у вихідних векторах, найчастіше це самоперетини полігонів, розриви контурів та дублювання вершин. Особливо проблематичними є щільно забудовані райони, де мережі об'єднують окремі будівлі, або розділяються одну будівлю на різні частини.

Також, присутні проблеми генералізації. Моделі, натреновані на європейських містах, демонструють суттєве падіння результатів при тестуванні на азіатських містах, якщо при цьому не було додаткового донавчання. Висока варіативність розмірів будівель спричиняє пропуски у розпізнаванні будівель.

Суттєвою також є проблема оклюзій та проявів різноманітних атмосферних умов на космічних знімках. Тіні від високих будівель та дерев призводять до часткового виявлення будівель, а сезонні зміни рослинності вимагають тренування моделі на мультитемпоральних даних.

Висновки

Сучасні методи нейронних мереж досягли високої точності векторизації будівель, проте зберігаються критичні проблеми геометричної регулярності, топологічної коректності та генералізації. Ключові виклики включають відхилення кутів до 8.3° від ортогональності, топологічні помилки у 12-18% випадків, падіння точності на 15-22% при зміні географічного регіону та пропуски будівель.

Трансформерні архітектури демонструють найвищу точність (IoU 85-92%), але є ресурсомісткими. CNN-базовані методи забезпечують кращий баланс швидкості та якості (IoU 80-90%). Комплексні векторизаційні підходи мають нижчу растрову точність (IoU 82-88%), але значно кращу векторну якість зі значеннями метрики PoLiS 70-73%.

Перспективними напрямками вирішення існуючих проблем є: інтеграція диференційованих геометричних обмежень безпосередньо в архітектури мереж, розробка топологічно-усвідомлених функцій втрат,

створення базових моделей з кращою міждоменною генералізацією та розвиток методів самоконтрольованого навчання для зменшення залежності від великих анотованих наборів даних. Особливу перспективу демонструють адаптовані моделі сегментації загального призначення з низькоранговою адаптацією. Необхідно також приділити увагу мультимодальним підходам, що інтегрують оптичні знімки з цифровими моделями поверхні та даними лазерного сканування для покращення тривимірної реконструкції будівель.

Конфлікти інтересів

Автори заявляють, що у них немає конфлікту інтересів щодо поточного дослідження, включаючи фінансовий, особистий, авторський чи будь-який інший, який міг би вплинути на дослідження, а також на результати, наведені в цьому документі.

Фінансування

Дослідження проводилося без фінансової підтримки.

Доступність даних

Усі дані доступні в цифровій або графічній формі в основному тексті статті.

Використання штучного інтелекту

Автори підтверджують, що при створенні поточної роботи вони не використовували технології штучного інтелекту.

Література

1. Li W., He C., Fang J., Zheng J., Fu H., Yu L. Semantic segmentation-based building footprint extraction using very high-resolution satellite images and multi-source GIS data. *Remote Sensing*. 2019. Vol. 11, № 4. P. 403.
2. Ronneberger O., Fischer P., Brox T. U-Net: Convolutional networks for biomedical image segmentation. *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015*. Springer, 2015. P. 234–241.
3. Oktay O., Schlemper J., Folgoc L.L., Lee M., Heinrich M., Misawa K., Rueckert D. Attention U-Net: Learning where to look for the pancreas. *Medical Imaging with Deep Learning*. 2018.
4. Chen L.-C., Zhu Y., Papandreou G., Schroff F., Adam H. Encoder-decoder with atrous separable convolution for semantic image segmentation. *Proceedings of the European Conference on Computer Vision (ECCV)*. 2018. P. 801–818.
5. Wang L., Li R., Zhang C., Fang S., Duan C., Meng X., Atkinson P.M. UNetFormer: A UNet-like transformer for efficient semantic segmentation of remote sensing urban scene imagery. *ISPRS Journal of Photogrammetry and Remote Sensing*. 2022. Vol. 190. P. 196–214.
6. Wang Y., Wang M., Hao Z., Wang Q., Wang Q., Ye Y. MDFA-Net: Multi-scale differential feature self-attention network for building change detection in remote sensing images. *Remote Sensing*. 2024. Vol. 16, № 18. P. 3466.
7. Zhao W., Persello C., Stein A. Building outline delineation: From very high resolution remote sensing images to polygons with an improved end-to-end learning

framework. ISPRS Journal of Photogrammetry and Remote Sensing. 2021. Vol. 179. P. 364–378.

8. Yang G., Zhang Q., Zhang G. EANet: Edge-aware network for the extraction of buildings from aerial images. Remote Sensing. 2020. Vol. 12, № 13. P. 2161.

9. Wang Y., Chen C., Ding M., Li J. Real-time dense semantic labeling with dual-Path framework for high-resolution remote sensing image. Remote Sensing. 2019. Vol. 11, № 24. P. 3020.

10. Chen L., Papandreou G., Schroff F., Adam H. Rethinking atrous convolution for semantic image segmentation. arXiv preprint arXiv:1706.05587. 2017.

11. Zhang L., Wu J., Fan Y., Gao H., Shao Y. An efficient building extraction method from high spatial resolution remote sensing images based on improved mask R-CNN. Sensors. 2020. Vol. 20, № 5. P. 1465.

12. Xie E., Wang W., Yu Z., Anandkumar A., Alvarez J.M., Luo P. SegFormer: Simple and efficient design for semantic segmentation with transformers. Advances in Neural Information Processing Systems (NeurIPS). 2021. Vol. 34. P. 12077–12090.

13. Girard N., Smirnov D., Solomon J., Tarabalka Y. Polygonal building extraction by frame field learning. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). 2021. P. 5891–5900.

14. Zorzi S., Bazrafkan S., Habenschuss S., Fraundorfer F. PolyWorld: Polygonal building extraction with graph neural networks in satellite images. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). 2022. P. 1848–1857.

15. Li Z., Wegner J.D., Lucchi A. Topological map extraction from overhead images. Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV). 2019. P. 1715–1724.

16. Zhang Y., Gong W., Sun J., Li W. Web-Net: A novel nest networks with ultra-hierarchical sampling for building extraction from aerial imageries. Remote Sensing. 2019. Vol. 11, № 16. P. 1897.

17. Cheng G., Wang Y., Xu S., Wang H., Xiang S., Pan C. Automatic road detection and centerline extraction via cascaded end-to-end convolutional neural network. IEEE Transactions on Geoscience and Remote Sensing. 2017. Vol. 55, № 6. P. 3322–3337.

18. He K., Gkioxari G., Dollár P., Girshick R. Mask R-CNN. Proceedings of the IEEE International Conference on Computer Vision (ICCV). 2017. P. 2961–2969.

19. Kirillov A., Mintun E., Ravi N., Mao H., Rolland C., Gustafson L., Girshick R. Segment anything. Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV). 2023. P. 4015–4026.

20. Ji S., Wei S., Lu M. Fully convolutional networks for multisource building extraction from an open aerial and satellite imagery data set. IEEE Transactions on Geoscience and Remote Sensing. 2019. Vol. 57, № 1. P. 574–586.

Відомості про статтю:	Article information:
Отримано 15.11.2025	Received 15.11.2025
Отримано у допрацьованому вигляді 19.11.2025	Received in revised form 19.11.2025
Прийнято 25.11.2025	Accepted 25.11.2025
Опубліковано 25.12.2025	Published 25.12.2025

I. A. Radion*

PhD student, ORCID: <https://orcid.org/0000-0001-9831-454X>

Department of Geodesy, Land Management and Cadastre

Lesya Ukrainka Volyn National University, 13 Voli Avenue, Lutsk, Ukraine, 43025

O. V. Melnyk

Associate Professor, Ph.D. in Engineering, ORCID: <https://orcid.org/0000-0002-5429-4038>

Department of Geodesy, Land Management and Cadastre

Lesya Ukrainka Volyn National University, 13 Voli Avenue, Lutsk, Ukraine, 43025

*corresponding author, e-mail: radionigor11@gmail.com

Modern neural network methods for building vectorization from high-resolution satellite imagery

How to Cite:

Radion, I. A., Melnyk, O.V. (2025). Modern neural network methods for building vectorization from high-resolution satellite imagery, 24, 340-347. [https://doi.org/10.36910/6775-2410-6208-2025-14\(24\)-28](https://doi.org/10.36910/6775-2410-6208-2025-14(24)-28)

Abstract. Background. Automatic vectorization of buildings from satellite imagery is a key task for mapping and cadastral purposes. Modern deep learning methods have achieved high raster accuracy (IoU 85-92%), yet a fundamental problem remains: segmentation optimization does not guarantee the generation of geometrically and topologically correct vector polygons. Studies report significant angular deviations (up to 8.3°), non-parallel walls, and a high rate of topological errors (12-18%). Poor generalization to new regions and the omission of small objects also remain challenges.

To systematize and analyze modern deep learning methods for building vectorization, with a focus on the problems of geometric regularity, topological correctness, and generalization.

A review of publications from 2015-2024 (CVPR, ISPRS, etc.) using benchmark datasets (SpaceNet, WHU, INRIA) was conducted. Evaluation metrics included IoU and F1-score for raster accuracy, as well as PoLiS and Chamfer Distance for vector quality. Methods were classified into three groups: CNN-based (U-Net, DeepLab), transformer-based (Swin, SegFormer), and end-to-end methods (Frame Field Learning, GNN).

CNN architectures remain an effective baseline. Transformers demonstrate the highest raster accuracy (IoU >90%) but are computationally expensive. End-to-end methods, such as Frame Field Learning and PolyWorld, which generate vectors directly by bypassing the polygonization step, show slightly lower raster accuracy but significantly better vector quality (PoLiS ~73%), which is critical for cadastral applications.

A trade-off exists: transformers lead in raster accuracy (IoU 85-92%), while end-to-end methods (IoU 82-88%) provide significantly higher vector quality (PoLiS 70-73%). Promising research directions include integrating geometric constraints into network architectures, developing topology-aware loss functions, improving generalization, and multimodal approaches combining optical imagery with LiDAR data.

Keywords: building vectorization, deep learning, neural networks satellite imagery, semantic segmentation, transformers, polygonization, topology