

# Аналіз існуючих методик автоматичного розпізнавання будівель і доріг на аеро– та супутникових знімках

Володимир Онищук, ас.<sup>1</sup> (ORCID: 0009–0008–2161–0971)

<sup>1</sup>Київський національний університет будівництва та архітектури, м. Київ, Україна

## АНОТАЦІЯ

Проведено аналіз сучасних методів автоматичного розпізнавання та семантичної сегментації будівель і дорожньої інфраструктури на аеро– та супутникових знімках. Розглянуто підходи на основі згорткових нейронних мереж (CNN), трансформерів і новітніх foundation–моделей. Проведено порівняння популярних архітектур таких як U–Net++, DeepLabv3+, HRNet, SegFormer, D–LinkNet за точністю сегментації та особливостями застосування до об’єктів забудови й транспортної мережі. Наведено ключові метрики оцінювання (mIoU, F1, APLS) та типові набори даних (SpaceNet, DeepGlobe, OpenEarthMap). За результатами аналізу укладено таблицю переваг і недоліків методів. Визначено практичне значення дослідження для завдань навігації, кадастру та оновлення картографічних матеріалів.

*Ключові слова:* семантична сегментація; будівлі; дорожня мережа; аерознімки; супутникові знімки; CNN; трансформери; U–Net++; DeepLabv3+; HRNet; SegFormer; D–LinkNet; mIoU; APLS; OpenEarthMap..

## 1. ВСТУП

Автоматичне виділення будівель і доріг на зображеннях з дронів та супутників є однією з ключових задач для геоінформаційних систем, міського планування та навігації. Традиційні методи, зокрема кластеризація та SVM, вимагали значних обсягів ручної роботи й не забезпечували належної точності. Натомість сучасні алгоритми глибокого навчання продемонстрували суттєвий прогрес у сегментації високороздільних знімків місцевості.

Важливим поштовхом до розвитку стали SpaceNet та DeepGlobe, спрямовані на автоматичне виділення будівель і дорожніх мереж [1]. SpaceNet запровадив спеціальну метрику Average Path Length Similarity (APLS) для оцінювання якості відтворення дорожнього графа, що доповнило стандартні показники точності сегментації. Основними метриками у цій сфері є mIoU (mean Intersection over Union) та F1–міра, які відображають загальну точність класифікації пікселів. Для дорожніх мереж критичною є саме APLS, оскільки вона враховує топологічну цілісність шляхової мережі, зіставляючи довжини найкоротших маршрутів у прогнозованому та еталонному графах.

Упродовж останніх років з’явилися масштабні відкриті набори даних – SpaceNet [1], DeepGlobe 2018 та новий глобальний датасет OpenEarthMap [4], що дали можливість навчати глибокі нейронні мережі для сегментації будівель і доріг на зображеннях з різних регіонів світу. У цьому контексті актуальним завданням є огляд новітніх методів та оцінка їхнього потенціалу для практичного застосування.

Метою роботи є здійснити системний аналіз сучасних методів автоматичного розпізнавання будівель і доріг на аеро– та супутникових знімках, визначити їхні переваги та недоліки й обґрунтувати доцільність практичного застосування.

## 2. ОГЛЯД СУЧАСНИХ МЕТОДІВ

Згорткові нейронні мережі (CNN) мають класичну основу сегментації зображень і залишаються ефективним інструментом для виділення будівель і дорожніх об’єктів. U–Net++, як розвиток U–Net із каскадними рівнями skip–зв’язків, щільно агрегує ознаки різних масштабів і завдяки

цьому підвищує точність на дрібних структурах; це робить модель доречною для детальної сегментації будівель у щільній міській забудові. DeepLabv3+ поєднує CNN–енкодер із модулем просторової пірамідальної вибірки (ASPP), що дає змогу захоплювати багатомасштабний контекст і покращувати сегментацію як великих будівель, так і протяжних дорожніх об’єктів. Водночас досягнення високих показників вимагає значних обчислювальних ресурсів. HRNet (High–Resolution Network), підтримуючи високу просторову роздільність представлень на всіх рівнях, точніше окреслює складні контури та кути будівель і надійніше виявляє вузькі відгалуження доріг, однак є ресурсоемною як на етапі навчання, так і під час інференсу. Для вузькоспеціалізованої задачі виділення дорожньої мережі запропоновано D–LinkNet – енкодер–декодерну CNN із каскадними дилатованими згортками, що забезпечує високу безперервність дорожнього графа; її обмеженням виступає менша універсальність у сценах із великою кількістю нецільових об’єктів.

На основі аналізу сучасних архітектур для задач сегментації наведено порівняння ключових методів: U–Net++, DeepLabv3+, HRNet, SegFormer та D–LinkNet.

Метод	Переваги	Недоліки
U–Net++	Детальне виділення дрібних об’єктів завдяки густим skip–зв’язкам.	Обмежений глобальний контекст (локальні властивості домінують).
DeepLabv3+	Ефективне багатомасштабне охоплення контексту (ASPP); Висока точність сегментації на рівні датасетів.	Високі вимоги до обчислювальних ресурсів. Не гарантує цілісності зв’язки структур.
HRNet	Збереження високої роздільної здатності ознак на всіх рівнях; Чітке окреслення меж будівель і доріг.	Великий розмір моделі, багато параметрів; Тривалий процес навчання, значні вимоги до пам’яті GPU.
SegFormer	Глобальне бачення сцени через механізм уваги трансформера; Компактний декодер, швидке зв’язування сегментації.	Необхідність великого обсягу даних для навчання трансформера; Обмежена інтерпретованість внутрішніх ознак моделі.
D–LinkNet	Спеціалізований для дорожньої мережі; Враховує контекст через дилатовані згортки.	Вузька спеціалізація; Архітектурні засади частково застарілі.

Рисунок 1. Порівняльна характеристика сучасних методів сегментації будівель і дорожніх об’єктів.

Потреба моделювати глобальні просторові залежності спричинила перехід до Transformer-архітектур, що природно враховують далекі взаємозв'язки між пікселями та підвищують узгодженість сегментації складних сцен. Показовий приклад SegFormer: компактний сегментатор із трансформерним енкодером і легким декодером, який на даних ДЗЗ часто не поступається CNN. Глобальне поле зору покращує сегментацію довгих звивистих доріг і множини розрізаних будівель. Недоліками є вищі вимоги до даних і обчислень тому в практичній реалізації використовують попередньо навчені енкодери MiT-B0–B5, що здешевлює донавчання. Інші підходи (SETR, Swin-UNet) застосовують у ДЗЗ, але менш усталені саме для класів «будівлі/дорogi». Наступний етап це масштабні foundation-моделі, попередньо навчені на колосальних масивах і здатні до широкої адаптації. У сегментації показова SAM (2023), навчена на понад мільярді масок, що виконує zero-shot-сегментацію за підказками точка, прямокутник. Для геопросторових сцен це дає швидку первинну розмітку будівель і доріг для подальшого уточнення або донавчання спеціалізованих мереж. Перевагами є універсальність і скорочення ручної розмітки, проте недоліки це висока потреба в пам'яті та доменна адаптація. На практиці foundation-моделі доречні як інструмент прискорення підготовки даних і отримання базових результатів із подальшим поліпшенням CNN- чи Transformer-архітектурами [2].

Отже, ландшафт методів сегментації для будівель і доріг – це спектр від перевірених CNN, через трансформери з глобальним контекстом, до універсальних foundation-моделей. Оптимальний вибір визначається компромісом між деталізацією й топологічною узгодженістю, обсягами даних і наявними ресурсами.

### 3. МЕТОДИЧНІ АСПЕКТИ ПІДГОТОВКИ ТА ВАЛІДАЦІЇ ДАНИХ

Якість результатів автоматичного розпізнавання будівель і доріг безпосередньо залежить від правильності формування вихідних даних. На етапі підготовки зображень до навчання моделей виконують кілька критичних процедур. По-перше, здійснюється уніфікація просторової роздільної здатності та спектральних каналів, що гарантує однакові умови обробки для різних джерел (спутникові серії, дрони, аерофотозйомка). По-друге, проводиться нормалізація значень пікселів та географічне вирівнювання для усунення відхилень у прив'язці до систем координат. Особливу увагу приділяють формуванню розмітки. Для будівель критично точне окреслення контурів, яке враховує складні форми дахів і тіні. Для доріг важливо зберегти топологічну неперервність мережі – розриви у масках навіть на кілька пікселів призводять до суттєвого падіння метрики APLS.

На практиці застосовують подвійне маркування із залученням незалежних експертів та консенсус-оцінку, що дозволяє мінімізувати помилки. Не менш важливим є етап валідації моделей. Стандартні метрики mIoU та F1 забезпечують оцінку точності на рівні пікселів, однак для транспортної мережі ключовим є показник APLS, який відображає топологічну цілісність графа. Доцільним є впровадження багаторівневої оцінки: від локальних ділянок (кварталів) до масштабів міста чи регіону. [2] Це дозволяє дослідити вплив морфологічних характеристик забудови на якість сегментації та підвищити надійність практичної

інтерпретації результатів. Таким чином, методична якість даних і коректність валідації виступають визначальними чинниками у застосуванні сучасних моделей. Вони забезпечують відтворюваність досліджень і роблять можливим масштабування підходів до національних кадастрових і навігаційних систем. [3]

### 4. ВИСНОВОК

Трансформерні архітектури та спеціалізовані CNN суттєво перевершують класичні підходи, але вимагають ретельної організації даних і оцінювання. Для будівель доцільні HRNet/U-Net++ з фокусом на точності контурів. Для доріг ефективні D-LinkNet разом із топологічною постобробкою та контролем APLS. SegFormer забезпечує кращу узагальнюваність у різних умовах зйомки. Запропоновано двоконтурну валідацію: mIoU/F1 для піксельної точності та APLS для топології.

Для стійкої експлуатації потрібні резерви якісно розмічених вибірок, резервні обчислювальні ресурси, політики контролю дрейфу даних, а також співпраця з муніципальними та кадастровими установами і спільнотами відкритих карт. Критична методична підготовка даних, уніфікація просторової роздільної здатності, нормалізація спектральних каналів, географічне вирівнювання, коректне формування розмітки. Багаторівнева валідація з урахуванням локальних і регіональних особливостей мінімізує помилки та підвищує відтворюваність.

### Список літератури

- [1] Van Etten A., Lindenbaum D., Bacastow T. M. SpaceNet: A Remote Sensing Dataset and Challenge Series. *arXiv preprint arXiv:1807.01232*, 2018. URL: [https://www.researchgate.net/publication/326171699\\_SpaceNet\\_A\\_Remote\\_Sensing\\_Dataset\\_and\\_Challenge\\_Series](https://www.researchgate.net/publication/326171699_SpaceNet_A_Remote_Sensing_Dataset_and_Challenge_Series)
- [2] 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. URL: <https://proceedings.neurips.cc/paper/2021/hash/64f1f27bf1b4ec22924fd0acb550c235-Abstract.html>
- [3] Kirillov A., Mintun E., Ravi N., Mao H., Rolland C., Gustafson L., Xiao T., Whitehead S., Berg A. C., Lo W.-Y., Dollár P., Girshick R. Segment Anything. *arXiv preprint arXiv:2304.02643*, 2023. DOI: <https://doi.org/10.48550/arXiv.2304.02643>
- [4] Xia J., Yokoya N., Adriano B., Broni-Bediako C. OpenEarthMap: A Benchmark Dataset for Global High-Resolution Land Cover Mapping. *Proceedings of IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)*, 2023. URL: [https://www.researchgate.net/publication/368310906\\_OpenEarthMap\\_A\\_Benchmark\\_Dataset\\_for\\_Global\\_High-Resolution\\_Land\\_Cover\\_Mapping](https://www.researchgate.net/publication/368310906_OpenEarthMap_A_Benchmark_Dataset_for_Global_High-Resolution_Land_Cover_Mapping)