

ПОРІВНЯННЯ АРХІТЕКТУР НЕЙРОННИХ МЕРЕЖ ДЛЯ ПРОСТОРОВОЇ ОРІЄНТАЦІЇ БЕЗПІЛОТНИХ ЛІТАЛЬНИХ АПАРАТІВ ТА НАЗЕМНИХ ДРОНІВ

Вступ. Забезпечення просторової орієнтації безпілотних літальних апаратів (БПЛА) у складних середовищах це ключове завдання автономної навігації. Традиційними методи локалізації:

- SLAM (Simultaneous Localization And Mapping) – алгоритмічна обчислювальна задача побудови і оновлення мапи невідомого оточення з одночасним відстежуванням місцеположення рухаючись по ньому [1].

- GNSS-IMU (Global Navigation Satellite System and Inertial Measurement Unit) – рішення з інтегрованим IMU дозволяють приймачу GNSS працювати у тунелях, у середині будівель, або за умов присутності природних чи штучних електронних завад.

Ці методи мають принципові недоліки, наприклад, у складних середовищах, зокрема, за відсутності системи глобального позиціонування (GPS), демонструють обмежену ефективність через накопичення похибок або недостатню здатність до узагальнення.

Деталізація проблематики. У реальному середовищі БПЛА стикаються з такими викликами:

- недостатня або зашумлена інформація з сенсорів (IMU, камери, GPS, лідари тощо);
- висока динаміка та нестабільність середовища (перешкоди, зміна освітлення, вітер);
- необхідність працювати у реальному часі та приймати рішення швидко;
- відсутність GPS (наприклад, у приміщеннях або під землею);
- обмеження обчислювальних ресурсів на борту дрона.

Тому більшої практичної ефективності набуває використання елементів комп'ютерного зору у вигляді глибоких нейронних мереж, які дозволяють вирішувати ряд основних проблем і задач для орієнтації об'єкта у просторі, зокрема:

- Визначення положення та орієнтації (Pose Estimation). Нейромережі можуть прогнозувати положення та орієнтацію дрона на основі зображень (наприклад, CNN + RNN), або комбінуючи вхідні дані з різних сенсорів (sensor fusion);

Досліджено порівняльну ефективність нейронних мереж сегментаційного типу (U-Net, DeepLabV3+) та архітектур об'єктного виявлення (YOLO, SSD, Faster R-CNN) у задачах просторової орієнтації безпілотних літальних апаратів (БПЛА). Розглянуто здатність моделей витягувати семантичні ознаки середовища, ідентифікувати реперні точки та динамічні об'єкти. Проведено експериментальне дослідження із застосуванням наборів даних UAVid та аерофотозйомки. Оцінено точність, швидкодію та ресурсні вимоги моделей.

Ключові слова: нейронні мережі, просторова орієнтація, БПЛА, наземні дрони, U-Net, YOLO, R-CNN.

– Локалізація та побудова карти (SLAM). Глибокі моделі дозволяють одночасно будувати карту середовища та локалізувати дрон, зменшуючи залежність від класичних алгоритмів (наприклад, ORB-SLAM);

– Аналіз середовища та планування шляху. Мережі можуть інтерпретувати сцени, виділяти перешкоди та приймати рішення про безпечний маршрут польоту (semantic segmentation, depth estimation, path planning);

– Прийняття рішень в умовах невизначеності. Нейромережі, використовуючи навчання з підкріпленням (reinforcement learning), дозволяють дронам самостійно знаходити оптимальні стратегії польоту в нових умовах. Використання нейронних мереж дозволяє ефективно вирішувати завдання оцінювання орієнтації, локалізації та побудови просторового представлення середовища шляхом моделювання складних нелінійних залежностей між вхідними сенсорними даними. Глибокі нейронні архітектури, зокрема згорткові та рекурентні мережі, демонструють здатність до узагальнення та стійкість до шумів, що робить їх перспективними для автономної навігації БПЛА в умовах часткової або ненадійної інформації. Зокрема, застосування методів глибокого навчання у завданнях оцінки просторової орієнтації дозволяє отримувати точні прогнози кутових параметрів (наприклад, кутів Ейлера або кватерніонів) на основі вхідних даних з камер, IMU або їх поєднання.

У цьому дослідженні порівнюються архітектури сегментаційного типу (які забезпечують піксельну класифікацію) з архітектурами для object detection (які ідентифікують об'єкти за координатами прямокутних обгорток). Мета даної роботи з'ясувати, який підхід більш інформативний та ефективний для задач просторової орієнтації.

Архітектури нейронних мереж. Сегментаційні мережі призначені для класифікації кожного пікселя вхідного зображення, що дозволяє точно визначити контури об'єктів, а також їхню форму, розташування та взаємне перекриття. Такі моделі широко застосовуються у задачах медичної діагностики, автономного керування, дистанційного зондування, навігації безпілотних апаратів тощо.

Один із найпоширеніших підходів це архітектура "енкодер-декодер", яка полягає у поетапному зменшенні просторової роздільності зображення за допомогою згортки для вилучення глибоких ознак (features), а згодом – поступовому відновленні його структури до початкового розміру з піксельною точністю класифікації. Приклад такої архітектури – U-Net [2], яка була запропонована для задач біомедичної сегментації, однак згодом знайшла широке застосування у комп'ютерному зорі. Вона складається з двох основних частин:

– Енкодер (скорочуюча частина) – послідовність згорткових та pooling-шарів, яка формує вискорівневе представлення вхідного зображення;

– Декодер (розширююча частина) – симетрична послідовність шарів апсемплінгу (наприклад, transposed convolution), які поступово відновлюють просторову роздільність;

– Skip-зв'язки – прямі з'єднання між відповідними шарами енкодера і декодера, що дозволяють передавати детальну локальну інформацію, втрачену при зменшенні роздільності. Ця симетрична структура у формі латинської літери "U" забезпечує ефективну передачу як глобального контексту, так і локальних деталей, що критично важливо для точної сегментації.

Мережі для виявлення об'єктів обмежувальними рамками.

До найвідоміших архітектур об'єктного виявлення належать YOLO [4], SSD [5], Faster R-CNN [6]. Вони здійснюють локалізацію об'єктів за координатами прямокутників (bounding boxes) та класифікацію кожного з них. Загальний підхід полягає у прогнозі множини детекцій: $D = \{(BB_i, c_i)\}_{i=1}^N$, де $BB_i = (x_i, y_i, w_i, h_i)$ – координати прямокутника, який виділяє об'єкт (bounding box), c_i – клас об'єкта, N – кількість об'єктів, виявлених на зображенні. YOLO, зокрема, виконує це за допомогою єдиної згорткової мережі, що робить його надзвичайно швидким. Для оцінки точності використовується

mean Average Precision (mAP), що інтегрує площу під кривою Precision-Recall, де Precision (точність) це метрика, що відповідає відношенню кількості об'єктів задектованих правильно до хибно виявлених об'єктів. Водночас Recall (повнота) відповідає на питання скільки з усіх реальних об'єктів, ми змогли задетектувати. В задачах об'єктної детекції ми змінюємо поріг впевненості (confidence threshold). Чим вищий поріг – тим більше непевних детектувань ми відкидаємо, відповідно зростає Precision, але падає Recall, візуально ця залежність показана у вигляді кривої на рис. 1.

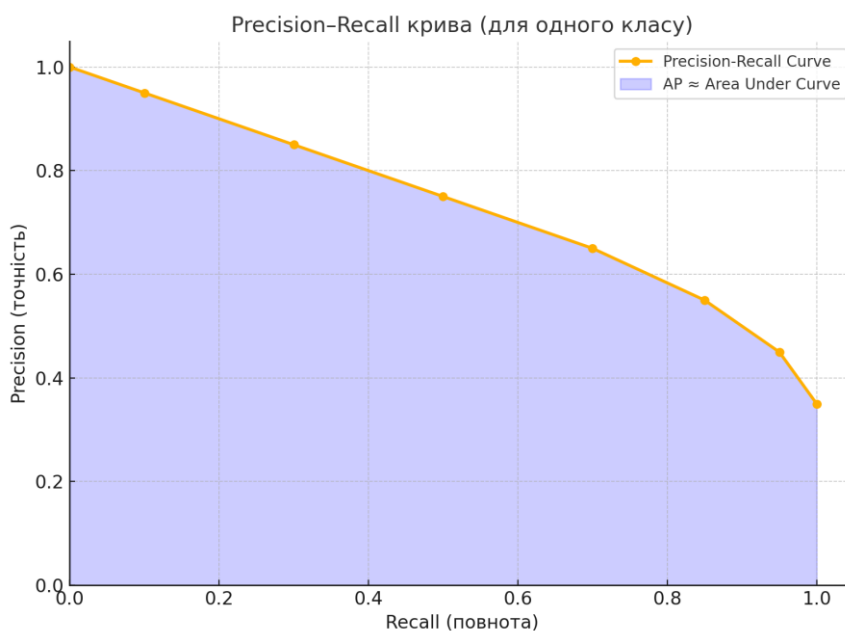


РИС. 1. Приклад precision-recall кривої для одного класу детектування

Площа під цією кривою – це Average Precision (AP). Якщо Precision висока → якщо модель каже "ось кіт", то це дійсно кіт (мало помилкових позитивів), але разом з тим Recall нижча → не всі коти знайдені (є пропущені).

Застосування до орієнтації БПЛА. Сегментація дозволяє детально відтворювати карту місцевості, розділяючи небо, будівлі, рослинність, дороги тощо. Це може використовуватись як джерело семантичних ознак у візуальному SLAM. Виявлення об'єктів – більш компактний варіант, який дозволяє локалізувати реперні точки: дерева, ліхтарі, машини тощо. Це забезпечує ключові прив'язки для орієнтації. Наприклад, точне виявлення дорожньої розмітки чи перехрестя може бути використано як якірні позиції при реконструкції траєкторії.

Експериментальна оцінка. У задачах автономної навігації БПЛА в умовах складного та нестандартного середовища – зокрема, у лісових масивах, гірських регіонах або серед щільної рослинності – використання методів семантичної або інстанційної сегментації має низку суттєвих переваг над традиційним об'єктним детектуванням. Об'єктна детекція (object detection) здебільшого оперує прямокутними обмежувальними рамками (bounding boxes), які лише умовно охоплюють об'єкти, не враховуючи їхню точну форму, просторову кривизну чи нерівномірні контури. Такий підхід може бути достатнім у структурованих середовищах (наприклад, у приміщеннях або на дорогах), проте у природному середовищі він втрачає ефективність: дерева з нерівними кронами, гілки, кущі, вузькі проходи та хаотичне розміщення перешкод потребують більш детального аналізу.

Сегментаційні мережі – зокрема, семантичні (які класифікують кожен піксель у складі певного класу) та інстанційні (які також розрізняють окремі екземпляри одного класу) – забезпечують піксельну точність розпізнавання об'єктів, що дозволяє формувати геометрично точні карти навколишнього простору. Це критично важливо для модулів планування траєкторії та ухвалення рішень у реальному часі, оскільки дозволяє точно оцінити не лише наявність, а й форму, щільність і розміщення перешкод. Наприклад, сегментація дозволяє БПЛА виявити прохід між двома стовбурами дерев [7] або визначити, що гілка знаходиться над або під безпечною траєкторією польоту, що неможливо зробити на основі лише обмежувальних рамок.

Крім того, сегментаційні моделі здатні класифікувати не лише об'єкти, а й типи поверхонь (наприклад, відкрите поле, кам'янистий ґрунт, вода), що надає додаткову інформацію для прийняття рішень щодо посадки, зміни маршруту або адаптації швидкості польоту. Таким чином, у динамічних і слабо структурованих середовищах сегментація дозволяє досягти більшої точності просторової орієнтації, адаптивності та безпеки навігації БПЛА, значно перевершуючи можливості класичного object detection. Більш детально порівняння нейронних мереж наведено в табл.1.

ТАБЛИЦЯ 1. Порівняльна таблиця сегментації та виділення об'єктів обмежувальних рамок для навігації БПЛА

Критерій	Object Detection	Семантична / Інстанційна Сегментація
Тип вихідних даних	Прямокутні bounding boxes	Піксельно-точні маски
Точність локалізації	Низька / середня	Висока
Форма об'єкта	Не враховується	Враховується точно
Виявлення тонких/довгих структур	Проблематичне (гілки, мотузки, трави)	Ефективне
Прохідність між об'єктами	Не визначається	Може бути оцінена по масках
Визначення типу поверхні	Неможливо	Можливо при семантичній сегментації
Планування траєкторії	Орієнтація на центри об'єктів	Можна аналізувати прохідні зони
Ефективність у складному середовищі	Обмежена, через можливе накладення обмежувальних рамок об'єктів	Висока
Використання в реальному часі	Швидше, але менш точне	Повільніше, але точніше (можна оптимізувати)
Складність реалізації	Простіше	Складніше (більше обчислень, даних)

Переваги сегментаційного виділення об'єктів (semantic/instance segmentation) перед звичайною бокс-ідентифікацією (bounding box detection) можна описати математично через такі метрики.

1. Точність локалізації.

Обмежувальні рамки (Bounding box далі BB) представляє об'єкт прямокутником:

$$BB = (x_{min}, y_{min}, x_{max}, y_{max}). \tag{1}$$

Сегментаційна маска (M) задає точну бінарну маску $M(x, y) \in \{0,1\}, \forall(x, y) \in$ зображення.

Точність локалізації можна виразити наступним чином:

$$IoU_{BB} = \frac{|BB_1 \cap BB_2|}{|BB_1 \cup BB_2|}, \quad IoU_{segmask} = \frac{|M_1 \cap M_2|}{|M_1 \cup M_2|}. \quad (2)$$

Оскільки прямокутні правильні форми не відповідають типовим реальним об'єктам розпізнавання [8] $IoU_{segmask} > IoU_{BB}$ тобто сегментація дає кращу геометричну відповідність реальній формі об'єкта. Відповідно з цього і впливає наступна перевага – краща точність оцінки площі, геометричних розмірів тощо. Що також важливо для оцінки дистанції до об'єкта.

2. Переваги для тонких / складних форм

Для об'єктів зі складною геометрією, наприклад: мотузки, людина з розведеними руками, рослини, будівлі зі шпилями. Обмежувальні рамки містять багато фонового зображення у тому числі можуть бути захоплені й інші об'єкти цього ж класу [9]. Сегментація дає множину пікселів:

$$P_{seg} = \{(x, y) \mid M(x, y) = 1\}. \quad (3)$$

Для них можна точно обчислювати:

- центроїд: $(x_c, y_c) = \left(\frac{1}{|P|} \sum_{(x,y) \in P} x, \frac{1}{|P|} \sum_{(x,y) \in P} y \right)$;
- орієнтацію об'єкта (через метод головних компонентів).

3. Зниження хибно позитивних та хибно негативних результатів.

При детекції об'єктів, особливо в густо заповнених сценах обмежувальні рамки можуть перетинатися, особливо при щільному розміщенні відповідно важко розділити об'єкти (instance confusion). Сегментаційний підхід дозволяє відрізнити пікселі, які належать до кожного екземпляра [10]. На рис. 2 показана різниця між детектуванням різними мережами.

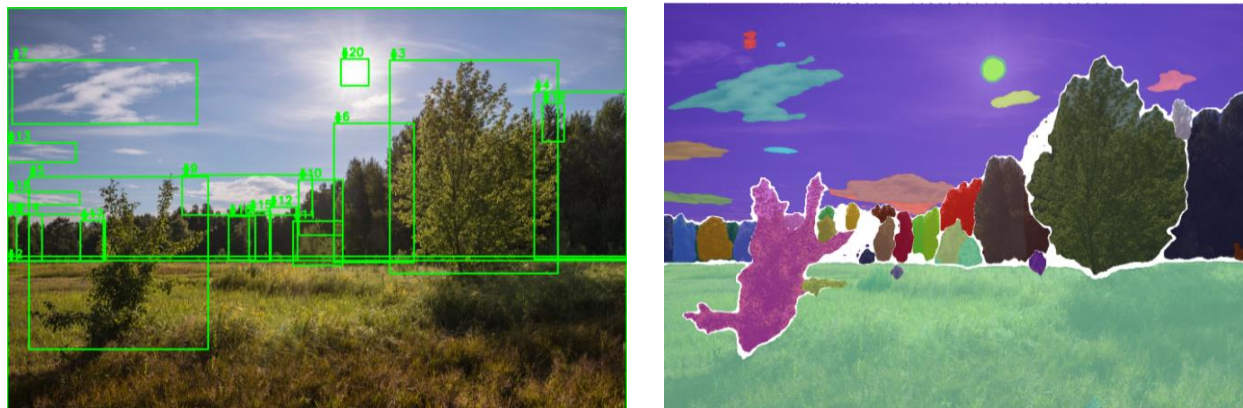


РИС. 2. Приклад детекції об'єктів – зліва детектор з обмежувальними рамками, справа – сегментаційний детектор

4. Покращена ефективність в задачах побудови траєкторії

Наприклад, для оцінки траєкторій сегментація дозволяє:

- побудувати мас-центричні траєкторії;
- оцінити поверхні зіткнення;
- створювати масштабовані маски для симуляцій.

Порівняння кількості параметрів нейронних мереж. Порівняння проводилось на наборі UAVid. Для сегментації оцінювались mIoU, для детекції – mAP. Також аналізувалась середня затримка обробки кадру при роздільності 640x360.

ТАБЛИЦЯ 2. Порівняльна таблиця нейронних мереж

Модель	Тип	mIoU / mAP	FPS	Параметри
U-Net	Сегментація	0.75 (mIoU)	19	7.8М
DeepLabV3+	Сегментація	0.79 (mIoU)	12	41М
YOLOv5s	Детектор	0.68 (mAP)	40	7.2М
SSD (Mobilenet)	Детектор	0.63 (mAP)	27	6.8М
Faster R-CNN	Детектор	0.75 (mAP)	8	42М

Висновки. Загалом, нейронні мережі виступають ефективним інструментом у розв'язанні задач просторової навігації БПЛА, забезпечуючи підвищену точність, надійність та адаптивність системи керування у реальних умовах експлуатації. Серед досліджених в рамках даної статті сегментаційні архітектури надають більше семантичної інформації, корисної для побудови карт і навігації. Проте вони важчі у виконанні. Об'єктні детектори витрачають менше часу на обробку відеозображення на системах з обмеженими ресурсами, хоча дають менш повну картину. У задачах орієнтації рекомендується комбінувати обидва підходи залежно від доступної обчислювальної потужності та вимог до точності. Доцільно досліджувати гібридні архітектури з багатовихідними системами (multitask learning).

Авторські внески: Сусленко О.Ю. – концепція дослідження, реалізація експериментів, аналіз результатів, підготовка тексту та написання та статті.

Наявність даних. Дані, які використовуються в даному дослідженні, доступні за посиланням <https://www.kaggle.com/datasets/dasmehdixtr/uavid-v1>.

Список літератури

- Jaulin L. A nonlinear set-membership approach for the localization and map building of an underwater robot using interval constraint propagation. *IEEE Transactions on Robotics*. 2009. Vol. 25, No. 1. P. 88–98. <https://doi.org/10.1109/TRO.2008.2010358>
- Ronneberger O., Fischer P., Brox T. U-Net: Convolutional Networks for Biomedical Image Segmentation. *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015*. Springer, Cham, 2015. P. 234–241. https://doi.org/10.1007/978-3-319-24574-4_28
- 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)*. Springer, 2018. P. 801–818. https://doi.org/10.1007/978-3-030-01234-2_49
- Redmon J., Divvala S., Girshick R., Farhadi A. You Only Look Once: Unified, Real-Time Object Detection. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. 2016. P. 779–788. <https://doi.org/10.1109/CVPR.2016.91>
- Liu W., Anguelov D., Erhan D., Szegedy C., Reed S., Fu C.-Y., Berg A. C. SSD: Single Shot MultiBox Detector. *Proceedings of the European Conference on Computer Vision (ECCV)*. Springer, 2016. P. 21–37. https://doi.org/10.1007/978-3-319-46448-0_2
- Ren S., He K., Girshick R., Sun J. Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks. *Advances in Neural Information Processing Systems (NeurIPS)*. 2015. P. 91–99. <https://arxiv.org/abs/1506.01497>
- Schiefer F., Kattenborn T., Frick A., Frey J., Schmidlein S. Mapping forest tree species in high-resolution UAV-based RGB imagery by means of convolutional neural networks. *ISPRS Journal of Photogrammetry and Remote Sensing*. 2020. Vol. 170. P. 205–215. <https://doi.org/10.1016/j.isprsjprs.2020.10.016>
- Xia G.-S., Bai X., Ding J., Zhu Z., Belongie S., Luo J., Datcu M., Pelillo M., Zhang L. DOTA: A Large-scale Dataset for Object Detection in Aerial Images. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. 2018. P. 3974–3983. <https://doi.org/10.1109/CVPR.2018.00418>
- Hu S., Liu J., Kang Z. DeepLabV3+/EfficientNet Hybrid Network for Complex Scene Segmentation and Safe Landing Zone Detection. *Sensors*. 2021. **21** (23). Article 8136. <https://doi.org/10.3390/s21238136>
- Anagnostis A., Zampogiannis K., et al. Deep Learning-based Semantic Segmentation for UAV-based Orchard and Forest Monitoring. *Remote Sensing*. 2021. **13** (12). Article 2364. <https://doi.org/10.3390/rs13122364>

Одержано 12.08.2025

Сусленко Олександр Юрійович,

аспірант Інституту кібернетики імені В.М. Глушкова НАН України, Київ.

alexandr.suslenko@gmail.com

MSC 68T05, 68U10

Oleksandr Suslenko

Comparison of Neural Network Architectures for Spatial Orientation of Unmanned Aerial Vehicles and Ground Drones

V.M. Glushkov Institute of Cybernetics of the NAS of Ukraine, Kyiv

Correspondence: alexandr.suslenko@gmail.com

Introduction. Ensuring spatial orientation of unmanned aerial vehicles (UAVs) and ground drones is one of the key tasks in modern robotics and autonomous systems. Traditional methods, such as SLAM or GNSS-IMU integration, have significant limitations in complex environments where sensor data is noisy or satellite signals are unavailable. In such cases, deep learning approaches, in particular neural network architectures for segmentation and object detection, are promising as they enable drones to interpret their surroundings at the semantic level.

Purpose. The purpose of this work is to conduct a comparative analysis of segmentation neural network architectures (U-Net, DeepLabV3+) and object detection architectures (YOLO, SSD, Faster R-CNN) regarding their applicability to solving the problem of spatial orientation of UAVs and ground drones in real time, as well as to identify the advantages and disadvantages of both approaches depending on the operating environment and computational resources.

Methods. The study employed an experimental approach using UAVid datasets and synthetic aerial imagery. To evaluate the effectiveness of the selected architectures, the following indicators were analyzed: accuracy (mIoU, mAP), speed (FPS), and resource requirements (number of parameters, computational complexity). Special attention was given to the use of segmentation models for constructing semantic maps of the environment and object detectors for localizing reference landmarks.

Results. The experiments demonstrated that segmentation architectures (U-Net, DeepLabV3+) provide more precise representation of object shapes and allow the construction of highly detailed maps, which is critical in environments with irregular obstacles (forests, mountainous areas). At the same time, object detectors (YOLO, SSD) showed significantly higher real-time performance, making them more suitable for systems with limited computational resources. Faster R-CNN achieved higher accuracy but lagged in processing speed. It was shown that segmentation models make it possible to estimate traversability between objects and classify surface types, tasks that are unattainable with traditional bounding box detectors.

Conclusions. Segmentation architectures provide drones with richer semantic information for spatial orientation but require higher computational resources and demonstrate lower inference speed. Object detection architectures are capable of real-time operation but at the cost of reduced environmental detail. A combined approach, applying both methods depending on navigation tasks and resource constraints, is considered promising. Future research should focus on developing hybrid multi-output models that combine the advantages of segmentation and detection.

Keywords: spatial orientation, unmanned aerial vehicles, neural networks, semantic segmentation, object detection, U-Net, YOLO, DeepLabV3+, Faster R-CNN.