

В. О. ВЛАСЕНКО, аспірант, НТУ "ХПІ"

ВИКОРИСТАННЯ МОДЕЛЕЙ ГЛИБОКОГО НАВЧАННЯ U-NET, DEEPLABV3+ ТА FEATURE PYRAMID NETWORK ДЛЯ СЕМАНТИЧНОЇ СЕГМЕНТАЦІЇ АЕРОФОТОЗНІМКІВ

У роботі представлено дослідження ефективності використання моделей глибокого навчання U-Net, DeepLabV3+ та Feature Pyramid Network (FPN) для семантичної сегментації аерофотознімків. Ця задача дозволяє автоматизувати аналіз великих обсягів даних у різних сферах людської діяльності (містобудуванні, екологічного моніторингу тощо). У процесі дослідження було підготовлено власні вхідні дані та проведено навчання кожної з трьох моделей. Для оцінювання експериментальної роботи було розраховано функцію втрат, точність, індекс Жаккара, F-міру та час прогнозування одного зображення, що дозволило порівняти якість сегментації. Результати показують, що кожна модель здатна ефективно розпізнавати основні об'єкти на аерофотознімках. Модель U-Net демонструє кращу швидкість при навчанні, DeepLabV3+ навчається довше, але показує кращі показники, навіть на невеликій кількості вхідних даних, а FPN забезпечує найкращий баланс одночасно між швидкістю та якістю. Результати можуть бути використані для подальших досліджень у сфері автоматизованого аналізу зображень. Іл.: 5. Табл.: 3. Бібліогр.: 13 назв.

Ключові слова: комп'ютерний зір; нейронні мережі; глибоке навчання; U-Net; DeepLabV3+; Feature Pyramid Network (FPN); семантична сегментація; розпізнавання об'єктів; просторовий аналіз.

Постановка проблеми. У сучасному світі, в умовах постійного розвитку технологій, аерофотознімки, отримані з дронів, починають відігравати ключову роль у системах моніторингу за навколишнім середовищем, у сферах картографування, планування територій тощо. Постійне збільшення обсягів просторових даних викликає потребу у створенні автоматизованих методів обробки та аналізу зображень. У даному контексті, семантична сегментація є однією з найважливіших задач комп'ютерного зору, тому що вона дозволяє автоматично розділити зображення на області, які відповідають конкретному об'єкту (будівлі, транспорту, водоймі, дорозі, рослинності тощо).

Аерофотознімки, як правило, мають складну візуальну структуру та складаються з різноманітних об'єктів. У таких випадках класичні алгоритми обробки зображень можуть демонструвати недостатню

точність та обмежену ефективність. Саме тому актуальність використання глибокого навчання постійно зростає.

Сучасні згорткові нейронні мережі, зокрема такі архітектури як U-Net, DeepLabV3+ та Feature Pyramid Network (FPN), що розроблені спеціально для сегментації, показують значно кращу ефективність у випадках розпізнавання об'єктів на зображеннях. Кожна з цих моделей має свої переваги та недоліки, тому важливо обирати оптимальну архітектуру для власних задач, зокрема, для аналізу об'єктів на аерофотознімках.

Аналіз останніх досліджень. Для обробки та аналізу аерофотознімків ключовим етапом може бути сегментація зображень, тому що з її допомогою можна автоматично виділити будь-який об'єкт на зображенні для його подальшого визначення (наприклад, дорога, будівля, рослинність, водойма тощо). Висока роздільна здатність аерофотознімків, отриманих з дронів, дозволяє робити детальний аналіз місцевості. У таких випадках важливо точно визначати межі подібних об'єктів, використовуючи методи сегментації.

Семантична сегментація аерофотознімків потребує використання глибоких згорткових нейронних мереж, що здатні до ефективного вилучення просторових ознак. Для досягнення такої мети найбільш поширеними моделями є архітектури U-Net, DeepLabV3+, FPN та різні ансамблеві чи комбіновані рішення.

Найвідомішою архітектурою для сегментації зображень є U-Net. Вона здобула свою популярність через використання симетричної структури "енкодер – декодер" з пропусками. Подібна структура надає можливість ефективно передавати як глобальні, так і локальні просторові ознаки [1, 2]. При аналізі аерофотознімків U-Net показує високу точність у розпізнаванні однорідних об'єктів (будівель, рослинності, доріг) незважаючи на обмежену кількість розмічених даних [3]. Також, U-Net активно модифікують та використовують з різними типами енкодерів, такими як ResNet чи DenseNet, або з модулями уваги [4].

Широкою популярністю серед науковців користується модель DeepLabV3+, основою якої є використання просторового усереднення та багаторівневого визначення ознак модулем Atrous Spatial Pyramid Pooling (ASPP). Модель DeepLabV3+ є критично важливою для аналізу аерофотознімків, тому що дозволяє розпізнавати контекст на різних масштабах, коли об'єкти можуть бути як великими (поля, водойми), так і дрібними (забудови, автомобілі) [5]. Так само, як і в моделі U-Net, у DeepLabV3+ використовується декодерна частина з високою роздільною здатністю для покращення якості розпізнавання об'єктів на межах [5, 6].

Серед інших ефективних архітектур для семантичної сегментації зображень виділяється модель FPN, яка дозволяє аналізувати високорівневу семантику з детальною просторовою інформацією на різних рівнях глибини. FPN часто використовують з енкодером типу ResNet, що показує високу ефективність обробки складних сцен із різними масштабами об'єктів [6, 9], а її комбінування із складнішими архітектурами може значно покращити якість класифікації пікселів за умов неоднорідного ландшафту [2].

У роботі [7] науковці пропонують використовувати архітектуру Pyramid Scene Parsing Network (PSPNet), яка акцентує увагу на багаторівневому контекстуальному аналізі сцени і дозволяє покращити глобальне розуміння зображення за допомогою побудови піраміди сцен. Використання PSPNet може бути особливо корисним для аналізу аерофотознімків, тому що ця модель чітко враховує глобальні та локальні контексти і демонструє свою ефективність на комплексних сценах. Використання Object Context Network (OCNet), також, знайшло своє застосування у випадках контекстної обробки зображень, тому що ця модель вводить концепцію «об'єктного контексту» для семантичної сегментації [8]. Модель OCNet влучно моделює зв'язки між пікселями на основі належності до об'єкта та може бути застосована для роботи з аерофотознімками, коли важливо відрізнити структури, що мають схожі кольорові характеристики, але належать до різних класів.

У сучасних дослідженнях особливу увагу приділяють комбінованим та ансамблевим методам, тому що вони поєднують переваги одразу декількох архітектур або об'єднують різні моделі одночасно. Науковці пропонують використовувати гібридні підходи, які поєднують U-Net з DeepLabV3+ [10, 11] або з трансформерами, за допомогою яких моделі адаптуються до умов освітлення та різних типів об'єктів. Інші дослідники реалізують підхід з мультимасштабним аналізом або з каскадними сегментаційними модулями [12].

Окрему увагу приділяють використанню трансформерів Vision Transformer та SegFormer у задачах сегментації. Такі моделі дозволяють охопити довготривалі просторові ознаки на глобальному рівні [13]. Основна перевага такого підходу полягає у моделюванні контекстних зв'язків між пікселями на великих зображеннях. Використання трансформерів може бути перспективним напрямком для покращення результатів при аналізі аерофотознімків.

Наукові дослідження показують, що найбільш ефективним є використання модифікованих архітектур нейронних мереж U-Net,

DeepLabV3+ та FPN або їх поєднання між собою чи з іншими методами, наприклад, механізмами уваги, мультимодальними входами, трансформерами тощо. Значною мірою ефективність кожної з моделей залежить від вхідних зображень, їх складності та характеристик.

Мета роботи. У роботі розглядається ефективність використання трьох популярних архітектур нейронних мереж U-Net, DeepLabV3+ та FPN для семантичної сегментації аерофотознімків. Мета дослідження полягає у проведенні кількісного та якісного порівняння результатів сегментації різних моделей, а також у визначенні найкращої архітектури, що демонструє найбільш збалансовані показники функції втрат, точності, індексу Жаккара, F-міри та часу прогнозування одного зображення. Отримані результати допоможуть визначити кращу модель для використання у подальших дослідженнях.

Основна частина. Основною задачею дослідження є порівняння та аналіз сучасних архітектур нейронних мереж U-Net, DeepLabV3+ та FPN, їх реалізація та навчання на підготовлених даних з аерофотознімками. Необхідно визначити, яка з моделей зможе забезпечити найкращі результати сегментації за допомогою оцінювання стандартних метрик якості.

Дані для навчання складаються з 50 аерофотознімків та 50 відповідних їм масок для сегментації розміром 512x512 пікселів (рис. 1). Вони включають в себе характерні типи об'єктів для аерофотоаналізу (дороги, будівлі, рослинність тощо). З метою забезпечення однакової роздільної здатності та коректного форматування, пари зображень було попередньо оброблено та масштабовано до розміру 256x256 пікселів.

Для збільшення кількості навчальної вибірки та покращення роботи моделей було застосовано стандартні методи аугментації зображень, такі як дзеркальне відображення по вертикалі та горизонталі, кут нахилу, змінення яскравості, контрастності, насиченості та відтінку. В результаті було отримано набір з 250 відповідних пар зображень.

Маски зображень було конвертовано у класові маски, де кожен піксель представляє певний клас за формулою:

$$class_mask(i, j) = class_index \quad (1)$$

де i та j відповідні індекси кожного пікселя.

Всього було використано 7 класів відповідних об'єктів і отримано відповідну кількість пікселів на масках без аугментації та з аугментацією. Отримані результати приведено у табл. 1. Для кращої обробки нейронними

мережами усі зображення було нормалізовано в діапазон $[0, 1]$ та зконвертовано в тензори, що використовуються для тренування моделі.

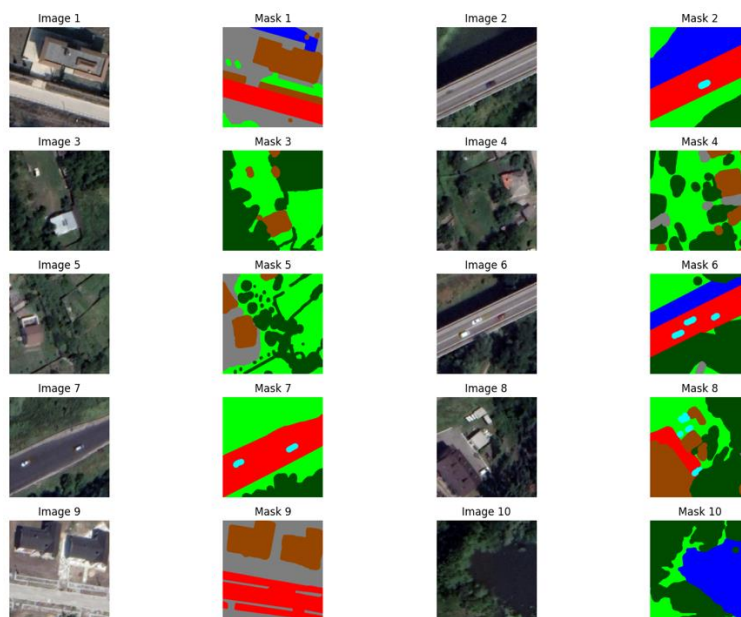


Рис. 1. Приклади пар зображень (аерофотознімків та відповідна маска)

Таблиця 1

Загальна кількість пікселів на масках зображень по класам

Номер класу	Об'єкти	Кількість пікселів (без аугментації)	Кількість пікселів (з аугментацією)
1	Будівлі	427318	2136590
2	Дороги	363790	1818950
3	Земля	414863	2074315
4	Трава	607481	3037405
5	Дерева	1289337	6446685
6	Водойми	126900	634500
7	Автомобілі	47111	235555

Для кількісної оцінки якості сегментації було використано такі метрики як точність (Accuracy), індекс Жаккара (IoU), F-міру (F1-Score) та час прогнозування одного зображення у секундах (Prediction Time), що особливо корисний у випадках визначення об'єктів у реальному часі.

Точність (Accuracy) – базова метрика для оцінки загальної коректності моделі, розраховує яка частка всіх передбачень є правильною.

Може бути недостатньою для оцінки роботи моделей у задачах сегментації. Розраховується за формулою:

$$\text{Accuracy} = \frac{\text{Правильні_прогнози}}{\text{Загальна_кількість_пікселів}} \quad (2)$$

Індекс Жаккара (IoU) показує наскільки добре передбачена маска збігається з реальною маскою. Часто використовується для задач сегментації, таких як виявлення об'єктів, сегментація медичних зображень тощо. Ця метрика дає чітке уявлення про якість сегментації. IoU розраховується за формулою:

$$\text{IoU} = \frac{|A \cap B|}{|A \cup B|} \quad (3)$$

де A – передбачувана маска, а B – справжня маска.

F-міра (F1-Score) показує наскільки добре модель знаходить об'єкти (наприклад, дороги, будівлі тощо), балансує між влучністю (precision) та повнотою (recall). Корисно використовувати, коли важлива якість сегментації за певним класом, особливо коли є незбалансованість класів (наприклад, дорога займає 10% зображення, а фон 90%). Розраховується за формулою:

$$F1 = \frac{2 * TP}{2 * TP + FP + FN} \quad (4)$$

де TP – true positives (правильно передбачені пікселі об'єкта), FP – false positives (модель помилилась і назвала фон об'єктом), FN – false negatives (модель не побачила частину об'єкта).

Розраховані метрики дозволяють комплексно оцінити глобальну відповідність сегментованого зображення з істинною розміткою та локальні помилки.

Навчання моделей проводилось окремо для кожної архітектури U-Net, DeepLabV3+ та FPN. Для кожної моделі було використано однакову конфігурацію: використання енкодера "ResNet34", кількість класів – 7, функція активації "softmax@", оптимізатор "Adam@" з початковою швидкістю навчання 0.0001 та кількість епох – 30.

Для розрахунку втрат при навчанні моделей було використано перехресну ентропію, яка обчислюється за формулою:

$$L(\bar{y}, y) = -\sum_i y_i \log(\bar{y}_i) \quad (5)$$

де y_i – передбачене значення, а \bar{y}_i – реальне значення.

Для якісного порівняння ефективності роботи моделей нейронних мереж, було проведено навчання як на оригінальному наборі з 50 зображень (рис. 2), так і на розширеному наборі з 250 аугментованих зображень (рис. 4). Моделі навчалися на одному і тому ж підготовленому наборі даних із фіксованим розділенням на тренувальну та тестову вибірку (20% від усіх зображень).

Після завершення навчання для кожної моделі було проведено оцінювання на тестовому наборі. Результати роботи моделей без аугментованих даних наведені у табл. 2, а з аугментованими – у табл. 3. Щоб впевнитись у коректності роботи навчених моделей було протестовано декілька зображень із тестової вибірки і отримано прогнозоване зображення для варіанту без аугментації (рис. 3) та з аугментацією (рис. 5).

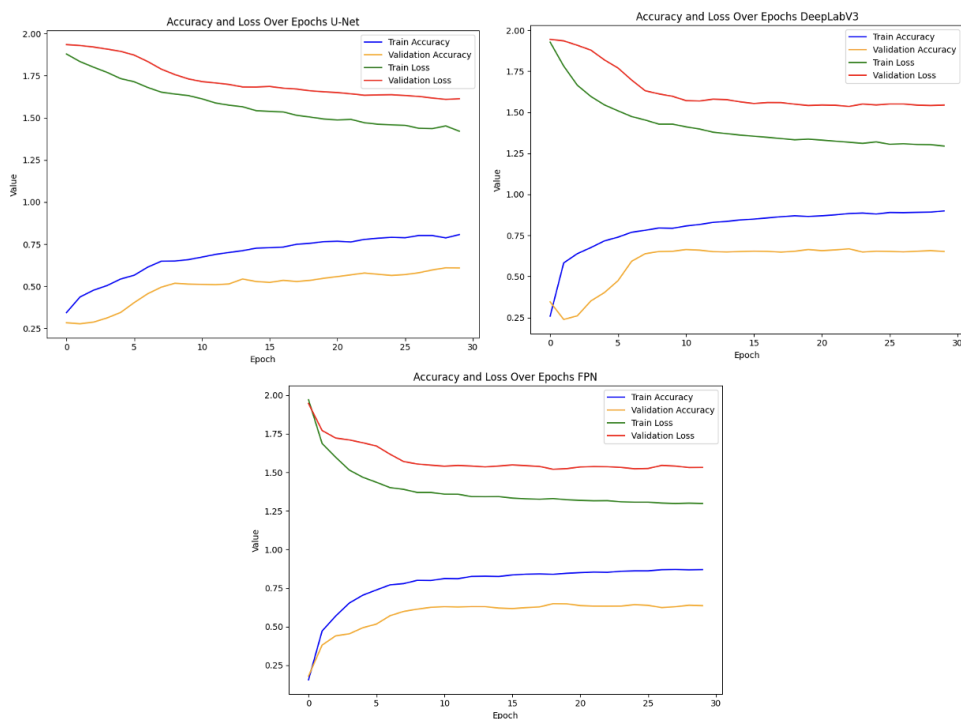


Рис. 2. Тренування моделей без аугментованих даних

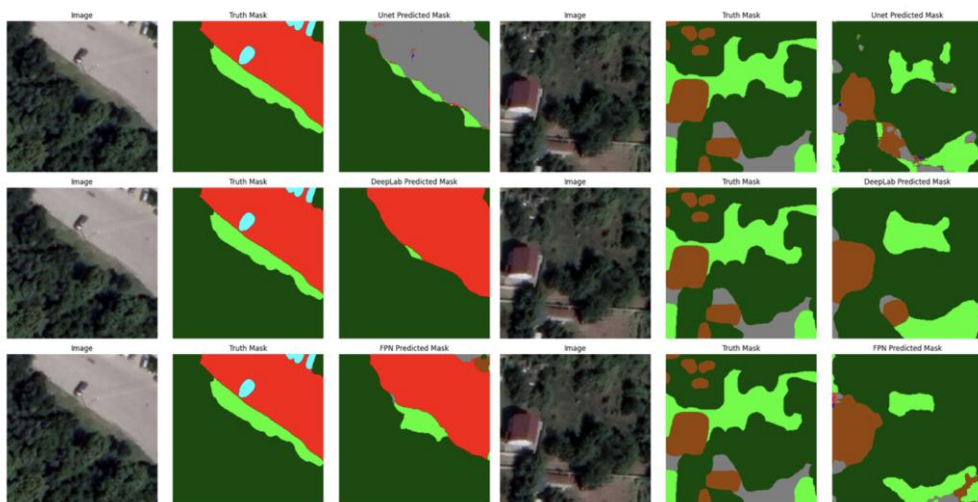


Рис. 3. Результати роботи моделей без аугментації на тестових даних (1 рядок – U-Net, 2 рядок – DeepLabV3+, 3 рядок – FPN)

Таблиця 2

Результати роботи моделей на тестових даних без аугментації (50 пар зображень)

Модель	Loss	Accuracy	IoU	F1-Score	Prediction Time
U-Net	1.6082	0.6083	0.2917	0.3907	0.1219
DeepLabV3+	1.5322	0.6524	0.3588	0.4767	0.3193
FPN	1.5243	0.6362	0.3408	0.4626	0.1055

Таблиця 3

Результати роботи моделей на тестових даних з аугментацією (250 пар зображень)

Модель	Loss	Accuracy	IoU	F1-Score	Prediction Time
U-Net	1.2960	0.8784	0.6611	0.7419	0.1393
DeepLabV3+	1.2778	0.8930	0.6817	0.7547	0.3162
FPN	1.2854	0.8803	0.7249	0.8321	0.1104

Проведені розрахунки та експерименти показали, що архітектура U-Net має найкращі значення по часу передбачення одного зображення, але найгірші показники функції втрат, точності, індексу Жаккара та F-міри. DeepLabV3+ показує кращі значення точності та непогані результати за іншими метриками, але FPN виявилась значно кращою за всіма іншими показниками.

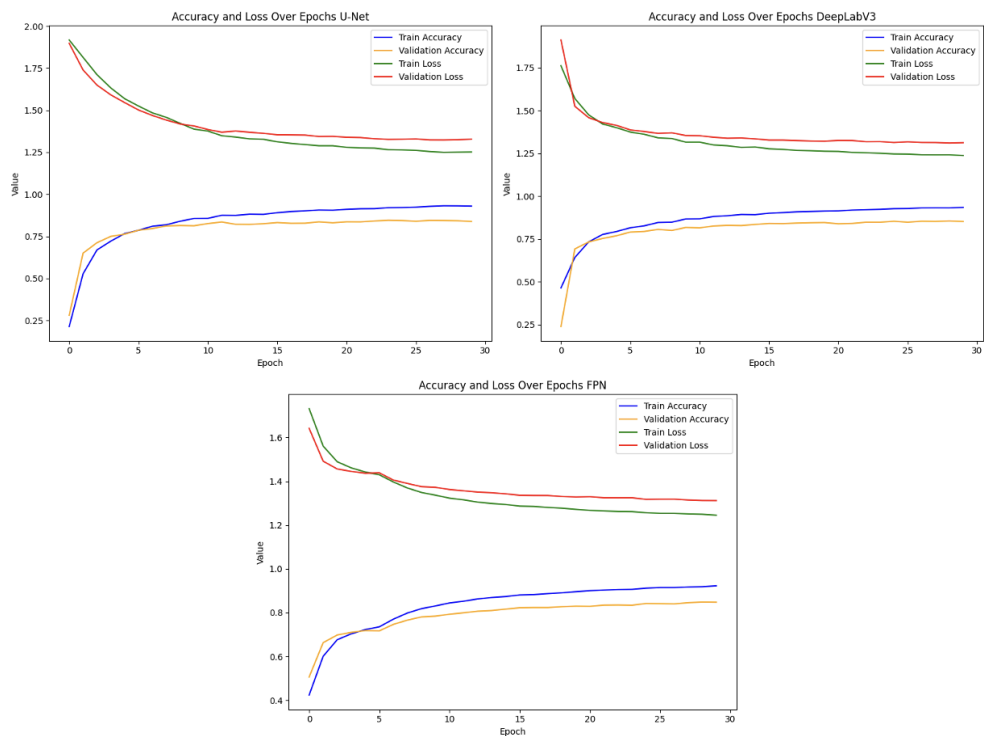


Рис. 4. Тренування моделей з аугментованими даними

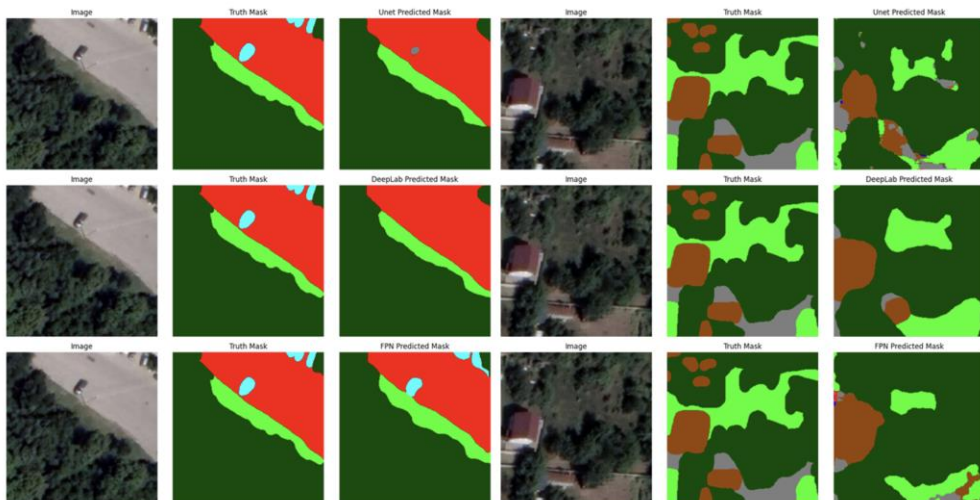


Рис. 5. Результати роботи моделей з аугментацією на тестових даних (1 рядок – U-Net, 2 рядок – DeepLabV3+, 3 рядок – FPN)

Таким чином, можна прийти до висновку, що серед розглянутих моделей найбільш ефективнішою, збалансованішою та якіснішою у

задачах сегментації аерофотознімків є Feature Pyramid Network. Саме тому вона є перспективною для подальшого використання в автоматизованих системах обробки зображень та визначенні об'єктів у різних сферах людської діяльності.

Висновки і перспективи подальших досліджень. Під час проведення дослідження було проаналізовано сучасні архітектури нейронних мереж U-Net, DeepLabV3+ та FPN. Результати проведених експериментів показали, що архітектура FPN демонструє найкращі показники якості сегментації у задачах розпізнавання об'єктів на аерофотознімках, бо забезпечує найкраще значення обчислюваних метрик.

Отримані результати підтверджують, що застосування глибоких згорткових нейронних мереж може ефективно впливати на сегментацію аерофотознімків, а також демонструють переваги певних архітектур у залежності від поставленої задачі. В подальшій роботі планується розширити набір тестових зображень, розробити ефективний метод для розпізнавання об'єктів використовуючи архітектуру FPN та адаптувати її до режиму реального часу з можливістю інтерактивного використання на мобільних чи планшетних пристроях.

Список літератури:

1. Ronneberger O. U-Net: Convolutional Networks for Biomedical Image Segmentation / O. Ronneberger, P. Fischer, T. Brox // *Lecture Notes in Computer Science*, 2015. – Vol. 9351. – P. 234–241.
2. Siddique N. U-Net and Its Variants for Medical Image Segmentation: A Review of Theory and Applications / N. Siddique, S. Paheding, C. P. Elkin, [et al.] // *IEEE Access*, 2021. – Vol. 9. – P. 82031–82057.
3. Gui S. Remote Sensing Object Detection in the Deep Learning Era—A Review / S. Gui, S. Song, R. Qin, Y. Tang // *Remote Sensing*, 2024. – Vol. 16, No. 2. – Article 327.
4. Wu X. UIU-Net: U-Net in U-Net for Infrared Small Object Detection / X. Wu, D. Hong, J. Chanussot // *IEEE Transactions on Image Processing*, 2023. – Vol. 32. – P. 364–376.
5. Liu L. A lightweight Deeplab V3+ network integrating deep transitive transfer learning and attention mechanism for burned area identification / L. Liu, Y. Guo, E. Chen, [et al.] // *Scientific Reports*, 2025. – Vol. 15. – Article 15969.
6. Ghiasi G. NAS-FPN: Learning Scalable Feature Pyramid Architecture for Object Detection / G. Ghiasi, T.-Y. Lin, Q. V. Le // *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2019. – P. 7029–7038.
7. Zhao H. Pyramid Scene Parsing Network / H. Zhao, J. Shi, X. Qi, X. Wang, J. Jia // *IEEE Conference on Computer Vision and Pattern Recognition*, 2017. – P. 6230–6239.
8. Yuan Y. Object-Contextual Representations for Semantic Segmentation / Y. Yuan, X. Chen, J. Wang // *Computer Vision – ECCV*. Cham : Springer International Publishing, 2020. – P. 173–190.

- 9.** Xu H. Auto-FPN: Automatic Network Architecture Adaptation for Object Detection Beyond Classification / H. Xu, L. Yao, Z. Li, [et al.] // IEEE/CVF International Conference on Computer Vision, 2019. – P. 6648–6657.
- 10.** Chen J. TransUNet: Rethinking the U-Net architecture design for medical image segmentation through the lens of transformers / J. Chen, J. Mei, X. Li, [et al.] // Medical Image Analysis, 2024. – Vol. 97. – Article 103280.
- 11.** Wang J. A multi-scale remote sensing semantic segmentation model with boundary enhancement based on UNetFormer / J. Wang, T. Chen, L. Zheng, [et al.] // Scientific Reports, 2025. – Vol. 15. – Article 14737.
- 12.** Gong Y. Effective Fusion Factor in FPN for Tiny Object Detection / Y. Gong, X. Yu, Y. Ding, [et al.] // IEEE Winter Conference on Applications of Computer Vision, 2021. – P. 1159–1167.
- 13.** Sun K. ECS-Net: Improving Weakly Supervised Semantic Segmentation by Using Connections Between Class Activation Maps / K. Sun, H. Shi, Z. Zhang [et al.] // IEEE/CVF International Conference on Computer Vision, 2021. – P. 7263–7272.

References:

- 1.** Ronneberger, O., Fischer, P., Brox, T., 2015. U-Net: Convolutional Networks for Biomedical Image Segmentation. Medical Image Computing and Computer-Assisted Intervention (MICCAI). Springer International Publishing, pp. 234–241. https://doi.org/10.1007/978-3-319-24574-4_28.
- 2.** Siddique, N., Paheding, S., Elkin, C.P., Devabhaktuni, V., 2021. U-Net and Its Variants for Medical Image Segmentation: A Review of Theory and Applications. IEEE Access 9, pp. 82031–82057. <https://doi.org/10.1109/ACCESS.2021.3086020>.
- 3.** Gui, S., Song, S., Qin, R., & Tang, Y., 2024. Remote Sensing Object Detection in the Deep Learning Era—A Review. Remote Sensing, 16(2), 327. <https://doi.org/10.3390/rs16020327>.
- 4.** Wu, X., Hong, D., Chanussot, J., 2023. UIU-Net: U-Net in U-Net for Infrared Small Object Detection. IEEE Transactions on Image Processing, pp. 364–376. <https://doi.org/10.1109/TIP.2022.3228497>.
- 5.** Liu, L., Guo, Y., Chen, E. et al, 2025. A lightweight Deeplab V3+ network integrating deep transitive transfer learning and attention mechanism for burned area identification. Sci Rep 15, 15969. <https://doi.org/10.1038/s41598-024-66060-7>.
- 6.** Ghiasi, G., Lin, T.-Y., Le, Q.V., 2019. NAS-FPN: Learning Scalable Feature Pyramid Architecture for Object Detection. IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), pp. 7029–7038. <https://doi.org/10.1109/CVPR.2019.00720>.
- 7.** Zhao, H., Shi, J., Qi, X., Wang, X., Jia, J., 2017. Pyramid Scene Parsing Network. IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 6230–6239, <https://doi.org/10.1109/CVPR.2017.660>.
- 8.** Yuan, Y., Chen, X., Wang, J., 2020. Object-Contextual Representations for Semantic Segmentation. Computer Vision – ECCV, Lecture Notes in Computer Science. Springer International Publishing, pp. 173–190. https://doi.org/10.1007/978-3-030-58539-6_11.
- 9.** Xu, H., Yao, L., Li, Z., Liang, X., Zhang, W., 2019. Auto-FPN: Automatic Network Architecture Adaptation for Object Detection Beyond Classification. IEEE/CVF International Conference on Computer Vision (ICCV), pp. 6648–6657. <https://doi.org/10.1109/ICCV.2019.00675>.
- 10.** Chen, J., Mei, J., Li, X., Lu, Y., Yu, Q., Wei, Q., Luo, X., Xie, Y., Adeli, E., Wang, Y., Lungren, M.P., Zhang, S., Xing, L., Lu, L., Yuille, A., Zhou, Y., 2024. TransUNet: Rethinking

the U-Net architecture design for medical image segmentation through the lens of transformers. *Medical Image Analysis* 97, 103280. <https://doi.org/10.1016/j.media.2024.103280>.

11. Wang, J., Chen, T., Zheng, L., Tie, J., Zhang, Y., Chen, P., Luo, Z., Song, Q., 2025. A multi-scale remote sensing semantic segmentation model with boundary enhancement based on UNetFormer. *Sci Rep* 15, 14737. <https://doi.org/10.1038/s41598-025-99663-9>.

12. Gong, Y., Yu, X., Ding, Y., Peng, X., Zhao, J., Han, Z., 2021. Effective Fusion Factor in FPN for Tiny Object Detection. *IEEE Winter Conference on Applications of Computer Vision (WACV)*, pp. 1159-1167. <https://doi.org/10.1109/WACV48630.2021.00120>.

13. Sun, K., Shi, H., Zhang, Z., Huang, Y., 2021. ECS-Net: Improving Weakly Supervised Semantic Segmentation by Using Connections Between Class Activation Maps. *IEEE/CVF International Conference on Computer Vision (ICCV)*, pp. 7263–7272. <https://doi.org/10.1109/ICCV48922.2021.00719>.

Поступила (received) 30.05.2025

Статтю представив канд. техн. наук, доц. каф. геометричного моделювання та комп'ютерної графіки Національного технічного університету "Харківський політехнічний інститут", Дашкевич Андрій Олександрович

Vlasenko Vitalii.

National Technical University "KhPI",

Str. Kyrpychova, 2, Kharkiv, Ukraine, 61002

Tel.: +38 (066) 271-54-72, e-mail: vlasenko.vitaliy99@gmail.com

ORCID ID: 0009-0008-6951-8496

УДК 004.8

Використання моделей глибокого навчання U-Net, DeepLabV3+ та Feature Pyramid Network для семантичної сегментації аерофотознімків / Власенко В.О. // Вісник НТУ "ХПІ". Серія: Інформатика та моделювання. – Харків: НТУ "ХПІ". – 2025. – № 2 (14). – С. 31 – 43.

У роботі представлено дослідження ефективності використання моделей глибокого навчання U-Net, DeepLabV3+ та Feature Pyramid Network (FPN) для семантичної сегментації аерофотознімків. Ця задача дозволяє автоматизувати аналіз великих обсягів даних у різних сферах людської діяльності (містобудуванні, екологічного моніторингу тощо). У процесі дослідження було підготовлено власні вхідні дані та проведено навчання кожної з трьох моделей. Для оцінювання експериментальної роботи було розраховано функцію втрат, точність, індекс Жаккара, F-міру та час прогнозування одного зображення, що дозволило порівняти якість сегментації. Результати показують, що кожна модель здатна ефективно розпізнавати основні об'єкти на аерофотознімках. Модель U-Net демонструє кращу швидкість при навчанні, DeepLabV3+ навчається довше, але показує кращі показники, навіть на невеликій кількості вхідних даних, а FPN забезпечує найкращий баланс одночасно між швидкістю та якістю. Результати можуть бути використані для подальших досліджень у сфері автоматизованого аналізу зображень. Іл.: 5. Табл.: 3. Бібліогр.: 13 назв.

Ключові слова: комп'ютерний зір; нейронні мережі; глибоке навчання; U-Net; DeepLabV3+; Feature Pyramid Network (FPN); семантична сегментація; розпізнавання об'єктів; просторовий аналіз.

UDC 004.8

Using deep learning models U-Net, DeepLabV3+ and Feature Pyramid Network for semantic segmentation of aerial images / Vlasenko V. // Herald of the National Technical University "KhPI". Series of "Informatics and Modeling". – Kharkov: NTU "KhPI". – 2025. – № 2 (14). – P. 31 – 43.

The paper presents a study of the effectiveness of using deep learning models U-Net, DeepLabV3+ and Feature Pyramid Network (FPN) for semantic segmentation of aerial images. This task allows to automate the analysis of large amounts of data in various areas of human activity (urban planning, environmental monitoring, etc.). In the process of the study, our own input data was prepared and models were trained. To evaluate the experimental work, the loss function, accuracy, Jaccard index, F-measure, and prediction time of one image were calculated, which allowed us to compare the quality of segmentation. The results show that each model is able to effectively recognize the main objects in aerial photographs. The U-Net model demonstrates better speed during training, DeepLabV3+ trains longer, but shows better performance, even on a small amount of input data, and FPN provides the best balance between speed and quality at the same time. The results can be used for further research in the field of automated image analysis. Figs.: 5. Tbls.: 3. Refs.: 13 titles.

Keywords: computer vision; neural networks; deep learning; U-Net; DeepLabV3+; Feature Pyramid Network (FPN); semantic segmentation; object recognition; spatial analysis.