

УДК 004.9

DOI <https://doi.org/10.32782/IT/2023-1-5>

### **Віта КАШТАН**

кандидат технічних наук, доцент, доцент кафедри інформаційних технологій та комп'ютерної інженерії, Національний технічний університет «Дніпровська політехніка», просп. Дмитра Яворницького, 19, м. Дніпро, Україна, 49005, [kashtan.v.yu@nmu.one](mailto:kashtan.v.yu@nmu.one)

ORCID: 0000-0002-0395-5895

Scopus Author ID: 57201902879

### **Володимир ГНАТУШЕНКО**

доктор технічних наук, професор, завідувач кафедри інформаційних технологій та комп'ютерної інженерії, Національний технічний університет «Дніпровська політехніка», просп. Дмитра Яворницького, 19, м. Дніпро, Україна, 49005, [hnatushenko.v.v@nmu.one](mailto:hnatushenko.v.v@nmu.one)

ORCID: 0000-0003-3140-3788

Scopus Author ID: 6505609275

### **Ірина УДОВИК**

кандидат технічних наук, доцент, деканка факультету інформаційних технологій, Національний технічний університет «Дніпровська політехніка», просп. Дмитра Яворницького, 19, м. Дніпро, Україна, 49005, [udovuk.i.m@nmu.one](mailto:udovuk.i.m@nmu.one)

ORCID: 0000-0002-5190-841X

Scopus Author ID: 55998874400

### **Ольга ШЕВЦОВА**

аспірантка кафедри програмного забезпечення комп'ютерних систем, Національний технічний університет «Дніпровська політехніка», просп. Дмитра Яворницького, 19, м. Дніпро, Україна, 49005, [shevtsova.o.s@nmu.one](mailto:shevtsova.o.s@nmu.one)

ORCID: 0000-0002-0148-5877

Scopus Author ID: 57220267804

**Бібліографічний опис статті:** Каштан, В., Гнатушенко, В., Удовик, І., Шевцова, О. (2023). Нейромережеве розпізнавання об'єктів забудови на аерофотознімках. *Information Technology: Computer Science, Software Engineering and Cyber Security*, 1, 30–39. doi: <https://doi.org/10.32782/IT/2023-1-5>

## **НЕЙРОМЕРЕЖЕВЕ РОЗПІЗНАВАННЯ ОБ'ЄКТІВ ЗАБУДОВИ НА АЕРОФОТОЗНІМКАХ**

Автоматизоване розпізнавання об'єктів забудови на аерофотознімках є однією з найбільш складних проблем у комп'ютерному зорі. Це пов'язано як з варіаціями зовнішнього вигляду будівель, так і з їх семантичною характеристикою в міських районах. У даній роботі запропоновано метод автоматизованого розпізнавання об'єктів забудови на цифрових аерофотознімках на основі 2D-CNN. Спочатку завантажуються та обробляються знімки високого просторового розрізнення, отримані з безпілотних літальних апаратів. Потім проводиться класифікація та сегментація зображення на основі архітектури нейронної мережі 2D-CNN з функцією *softmax* для вихідного шару та використовується випрямлений лінійний блок (*ReLU*) для решти шарів. Згорткові шари застосовують фільтри до всіх пікселів вхідного зображення, щоб отримати набір абстрактних особливостей високого рівня. Далі, для класифікації кожного пікселя зображення з БПЛА, була проведена сегментація даних, де поле сприйняття (*fov*) для кожного зображення вважається ковзним вікном розміру 3×3 вхідних даних. Це дозволяє визначити приналежність об'єкту до певного класу. Завершальним етапом є створення бінарної маски розпізнавання об'єктів забудови на основі функції втрат перехресної ентропії.

Тренування мережі проводилося на рівні пікселів, що дозволило підвищити точність ідентифікації об'єктів забудови та зменшити кількість неправильно класифікованих зон. Експериментальні результати показали значне покращення точності розпізнавання будівель у загальнодоступному наборі даних. Зокрема, метрики OA, AA та K покращилися на 2,6%, 5,6% та 3,2% відповідно для навчального набору даних і на 1,2%, 1,8% та 1,5% для тестового набору даних.

**Ключові слова:** згорткові нейронні мережі, сегментація зображень, розпізнавання, аерофотознімок, комп'ютерний зір.

### **Vita KASHTAN**

*Candidate of Technical Science, Associate Professor, Associate Professor of Department of Information Technology and Computer Engineering, Dnipro University of Technology, 19 Dmytra Yavornytskoho ave., Dnipro, Ukraine, 49005, kashtan.v.yu@nmu.one*

**ORCID:** 0000-0002-0395-5895

**Scopus Author ID:** 57201902879

### **Volodymyr HNATUSHENKO**

*Doctor of Technical Science, Professor, Head of Department of Information Technology and Computer Engineering, Dnipro University of Technology, 19 Dmytra Yavornytskoho ave., Dnipro, Ukraine, 49005, hnatushenko.v.v@nmu.one*

**ORCID:** 0000-0003-3140-3788

**Scopus Author ID:** 6505609275

### **Iryna UDOVYK**

*Candidate of Technical Science, Associate Professor, Dean of Information Technologies Department, Dnipro University of Technology, 19 Dmytra Yavornytskoho ave., Dnipro, Ukraine, 49005, udovyk.i.m@nmu.one*

**ORCID:** 0000-0002-5190-841X

**Scopus Author ID:** 55998874400

### **Oliha SHEVTSOVA**

*postgraduate of Department of Computer Systems Software, Dnipro University of Technology, 19 Dmytra Yavornytskoho ave., Dnipro, Ukraine, 49005, shevtsova.o.s@nmu.one*

**ORCID:** 0000-0002-0148-5877

**Scopus Author ID:** 57220267804

**To cite this article:** Kashtan, V., Hnatushenko, V., Udovyk, I., Shevtsova, O. (2023). Neiomerezheve rozpiznavannia ob'iektiv zabudovy na aerofotoznmkakh [Neural network recognition of building objects in aerial images]. *Information Technology: Computer Science, Software Engineering and Cyber Security*, 1, 30–39. doi: <https://doi.org/10.32782/IT/2023-1-5>

## **NEURAL NETWORK RECOGNITION OF BUILDING OBJECTS IN AERIAL IMAGES**

*Automated recognition of building objects based on aerial image data is a complex problem in computer vision. It is related to variations in the appearance of buildings and the semantic characteristics of scenes in urban areas. This paper proposes a method of automated recognition of building objects on digital aerial photographs based on 2D-CNN. The first step is to download and process high spatial resolution images from unmanned aerial vehicles. Then classification and segmentation of the image are performed based on 2D-CNN neural network architecture with softmax function for the output layer and rectified linear block (ReLU) for the remaining layers. Convolutional layers apply filters to all pixels of the input image to obtain a set of high-level abstract features. Data segmentation was performed to classify each pixel from the UAV images, where the field of view (fov) for each image is considered a sliding window of size 3×3 of the input data. It made it possible to determine whether the object of the researched scene belongs to a certain class. The last step is the binary mask creation of the building object recognition based on the cross-entropy loss function.*

*The neural network was trained at the pixel level, which allowed to increase the accuracy of identification of building objects and reduce the number of misclassified zones. Experimental results showed a significant improvement in the accuracy of building recognition in a publicly available dataset. Specifically, the OA, AA, and K metrics improved by 2.6%, 5.6%, and 3.2%, respectively, for the training dataset and by 1.2%, 1.8%, and 1.5% for the test dataset.*

**Key words:** convolutional neural networks, image segmentation, recognition, aerial image, computer vision.

**Вступ.** В останні роки у сфері аерофотозйомки популярним є використання безпілотних літальних апаратів (БПЛА) для збору даних над важкодоступними, віддаленими районами. Вони також відомі як дистанційно пілотовані літальні апарати (RPA) або дрони (багатомоторні, з фіксованими крилами, гібридні, тощо). Зображення, отримані за допомогою БПЛА, використовують

для баз даних географічної інформаційної системи, наборів даних для автоматизованого прийняття рішень, сільськогосподарського картографування, міського планування, землекористування, моніторингу й оцінки навколишнього середовища, в цивільних і військових цілях, тощо (Kalantar, 2019; Kalantar, 2017). Високе просторове розрізнення отриманих з БПЛА зобра-

жень (розмір пікселя від см до дм) дозволило створювати великі та детальні набори даних, які потрібно оброблювати. При обробці даних слід враховувати такі фактори як нестабільну поведінку літального апарату в повітряному просторі, так і дешеве бортове обладнання (часто на БПЛА встановлюють лише GPS-приймач без інерційної системи та не застосовують диференціальну корекцію).

**Літературний огляд.** Останнім часом було запропоновано багато методів та технологій на основі комп'ютерного зору для автоматичного розпізнавання об'єктів за будови. Ці методи базуються на емпіричних знаннях про будівлі та дозволяють виділяти такі характеристики, як кольори, текстури, краї, форми, тіні і т.д. Крім того, існують комплексні методи, які використовують зіставлення шаблонів, моделі активного контуру, математичну морфологію та інші методи. Однак, різноманітність типів будівель та складність фону роблять автоматичне розпізнавання об'єктів за будови на зображеннях надзвичайно складним завданням з високим ступенем складності.

Так, авторами Waleed A. et al. (Waleed, 2018, р. 1687–5265) запропоновано метод сегментації будівель на супутникових зображеннях з використанням архітектури U-Net, яка містить кодери ResNet і VGGNet. Запропонований метод автоматично розпізнає будівлі з точністю 84,9%. В роботі (Krówczyńska, 2020, р. 2–16) автори використовували геометричні ознаки об'єктів для ідентифікації дахів з різних матеріалів на основі кількох аерофотознімків (у природному кольорі (RGB) та інфрачервоному (CIR) композиціях) та генерували результати на основі інформації про стіни та тіні. Cheng et al. (Cheng, 2017, р. 1865–188) використовували математичну морфологію для класифікації сцени зображення дистанційного зондування шляхом інтеграції базової інформації, такої як спектральний вміст і контекст. В роботі (Avudaiamma, 2018, р. 1–6) авторами виконано виявлення та вилучення геометричних особливостей будівель у міській місцевості зображень дуже високого просторового розрізнення на основі колірної, спектральної, геометричної та фонові інформації. Але, вищезазначені методи покладаються на вибрані вручну функції та мають обмежену здатність до узагальнення.

В останні роки в методах автоматизованого вилучення будівель широко використовується глибоке навчання. Методи глибокого навчання уникають суб'єктивності вибору функцій вручну. Повністю згортова мережа від пікселя до пікселя (FCN), запропонована Лонгом та ін. (Long,

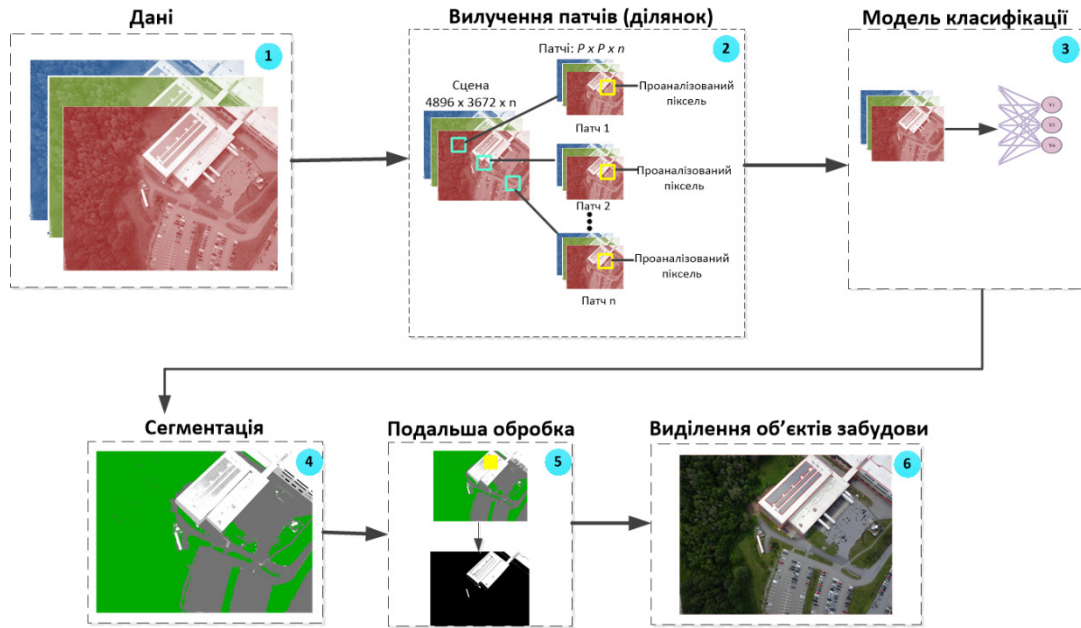
2015, р. 3431–3440) значно підвищила точність виділення об'єктів за будови.

Незважаючи на досягнення високої якості виявлення об'єктів, CNN має помітні недоліки (Liu, 2020, р. 261–31): навчання є багатоетапним і часто займає багато часу для реалізації та складні для оптимізації, так як кожен окремий етап потрібно навчати окремо. Юань (Yuan, 2017, р. 2793–2798) розробив глибоку згортову мережу, інтегровану з багатосаровою активацією для попиксельного виділення будівель. В роботі (Xu, 2018) запропоновано вдосконалену модель сегментації на основі FCN для виділення точних країв будівель, за допомогою ручних функцій і керованого фільтра. Хуан та ін. (Huang, 2019, р. 91–105) запропонували прогресивну мережу залишкового уточнення (GRRNet) для виділення будівель шляхом поєднання аерофотознімків високого просторового розрізнення та даних LiDAR. Незважаючи на те, що всі вищезазначені методи мають позитивні результати та розпізнають будівлі, деякі проблеми все ще потребують подальшого вивчення. Для розв'язання проблеми артефактів, можна використовувати методи додаткової обробки зображень, такі як методи згладжування зображень або методи залучення геометричної інформації. Деякі методи, такі як метод заснований на глибині, можуть використовувати додаткову інформацію про глибину об'єктів для покращення якості виділення будівель.

Метою роботи є розробка методу автоматизованого нейромережевого розпізнавання об'єктів за будови на цифрових аерофотознімках.

**Матеріали та методи.** Алгоритм методу представлений на рисунку 1 та складається з шести етапів. Першим етапом є завантаження зображень високого просторового розрізнення з безпілотних літальних апаратів (рис. 2). Збір даних проводився за допомогою БПЛА, що летів на висоті 353 м, за допомогою Canon PowerShot SX230 HS (5 мм). RGB-зображення складаються з трьох каналів (червоний, зелений, блакитний) з просторовим розрізненням приблизно 9,95 см/піксель, 4000 × 3000 пікселів.

Первинні зображення пропонується розділити на невеликі фрагменти для аналізу локальних деталей. Процес класифікації виконати моделлю DL, що дозволить нейронній мережі отримати більше інформації про границі зображення високого просторового розрізнення. Крім того, багато деталей втрачаються в процесі зменшення частоти дискретизації, тому пропонуємо виконати пост-обробку для уточнення даних після кожної пари згорток, щоб зменшити кількість неправильно класифікованих пікселів.



**Легенда:**  
 $n$  – кількість смуг зображення;  $P$  – розмір фрагмента.

**Рис. 1.** Алгоритм запропонованого методу



**Рис. 2.** Тестовий знімок з БПЛА

Етап вилучення патчів виконується з метою створення навчального набору даних для класифікації центральних пікселів. Спочатку створюються два набори даних, один для кожної послідовності. Для кожного пікселя зображення створюється патч розміром  $32 \times 32 \times n$  для 2D-CNN і  $8 \times 8 \times n$  для RNN. Ці патчі містять центральний піксель і деяку кількість сусідніх пікселів, що дозволяє аналізувати контекст пікселя. Далі кожен патч класифікується, щоб визначити, чи належить центральний піксель до інтересуючого нас класу. Для цього обчислюється частота пікселів в патчі зі значеннями класу, подібними до пікселя інтересу. Якщо патч має розмір більше 512 пікселів, то його

не обробляють. Після цього з кожного класу випадковим чином вибирається 1К патчів, що дозволяє збалансувати розподіл класів і зменшити розмір навчального набору. Результатом цього етапу є два навчальних набори даних, які містять 10К патчів для послідовності 1 та 9К патчів для послідовності 2, відповідно. Ці набори даних можуть бути використані для тренування моделі класифікації центральних пікселів. Потім, запропоновано використати геометричне перетворення (Kim, 2020, р. 3755):

$$\begin{bmatrix} x' \\ y' \\ 1 \end{bmatrix} = \begin{bmatrix} 1 & 0 & T_x \\ 0 & 1 & T_y \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \\ 1 \end{bmatrix} \quad (1)$$

де  $x'$  і  $y'$  – перетворені координати, а  $x$  і  $y$  – початкові координати пікселів;  $T_x$  і  $T_y$  – значення переходу по горизонталі та вертикалі.

Третім етапом є проведення класифікації зображення на основі згортової нейронної мережі. Як на точність, так і на збільшення кількості параметрів, що вивчаються, впливає глибина архітектури. Крім того, кількість параметрів має пряму залежність від технічних параметрів обчислювальної техніки. Тому, чим глибша архітектура, потрібно більша кількість різноманітних зразків для побудови надійних моделей і запобігання надмірному переповненню мережі. Для вирішення цієї проблеми пропонуємо використати 2D-CNN архітектуру нейронної

мережі. Зазначена архітектура складається з послідовності згорткових і об'єднуючих шарів, які в основному використовуються для навчання мережі на основі БПЛА даних (Yang, 2019, р. 6195–6211). Такий вид архітектури часто завершується повністю пов'язаними рівнями для прогнозування однієї мітки класу або набору ймовірностей класу (Rahman, 2020, р. 32–33). Згорткові шари застосовують фільтри до всіх пікселів вхідного зображення, щоб отримати набір абстрактних особливостей високого рівня; шари об'єднання зменшують кількість функцій – контроль надміру; і повністю підключені шари змінюють вихід у вектор із розміром, що дорівнює кількості класів (Rahman, 2020, р. 32–33). Для класифікації 2D-CNN зазвичай застосовує дві функції активації: softmax для вихідного шару та випрямлений лінійний блок (ReLU) для решти шарів. Softmax спрямований на масштабування виходів між нулем і одиницею. Це дозволить визначити приналежність об'єкту до певного класу. ReLU – це лінійна функція, яка безпосередньо виводить вхідні дані, якщо вони позитивні. В іншому випадку він виведе нуль. Розмір фільтра (fs) був обраний як  $fs = [2 \times 2, 4 \times 4, 8 \times 8]$ , що є першим значенням, яке дозволило досягти найвищої продуктивності. На рисунку 3 зображено графічне пояснення архітектури 2D-CNN.

Рівень об'єднання CNN (рис. 3). Операція об'єднання має на меті ідентифікувати подібні характеристики (значення, які статистично сильно корельовано) на основі вибору максимального значення в один патч (Mäyrgä, 2021, р. 112–322). Рівень об'єднання (операція

зменшення дискретизації) дозволяє зменшити просторовий розмір виходу, а також контролює перенавчання. Після цього кроку встановлюється наступний параметр для контролю надмірного навчання даних, що дозволяє покращити процес навчання.

Згортковий шар CNN. Класифікатор Softmax (функція активації) використовується для прогнозування ексклюзивних класів, а характеристики коду CNN були об'єднані в повністю щільний шар, щоб класифікувати кожен піксель зображення за найімовірнішою міткою. Класифікатор Softmax зважує відстані між даними перевірки та навчальними наборами даних із цього класу для ідентифікації категорії кожного пікселя (Каштан, 2019, с. 3–11).

На четвертому етапі запропонованого методу було проведено сегментацію даних. Після навчання DL-моделей отримані моделі використовуються для класифікації кожного пікселя із БПЛА зображень, де поле сприйняття (fov) для кожного зображення вважається ковзним вікном розміру  $3 \times 3$  вхідних даних. Цей процес завершується, коли всі пікселі класифіковано (Каштан, 2019, с. 3–11). Якщо клас має більше шести пікселів у вікні та він відрізняється від класу центрального пікселя та приймає мітку домінуючого класу; інакше значення мітки центрального пікселя зберігається. Такий алгоритм дозволяє підвищити точність ідентифікації об'єктів забудови та зменшити неправильно класифіковані зони. Результат сегментації наведено на рисунку 4.

На п'ятому етапі створюємо бінарну маску розпізнавання об'єктів забудови на основі функ-

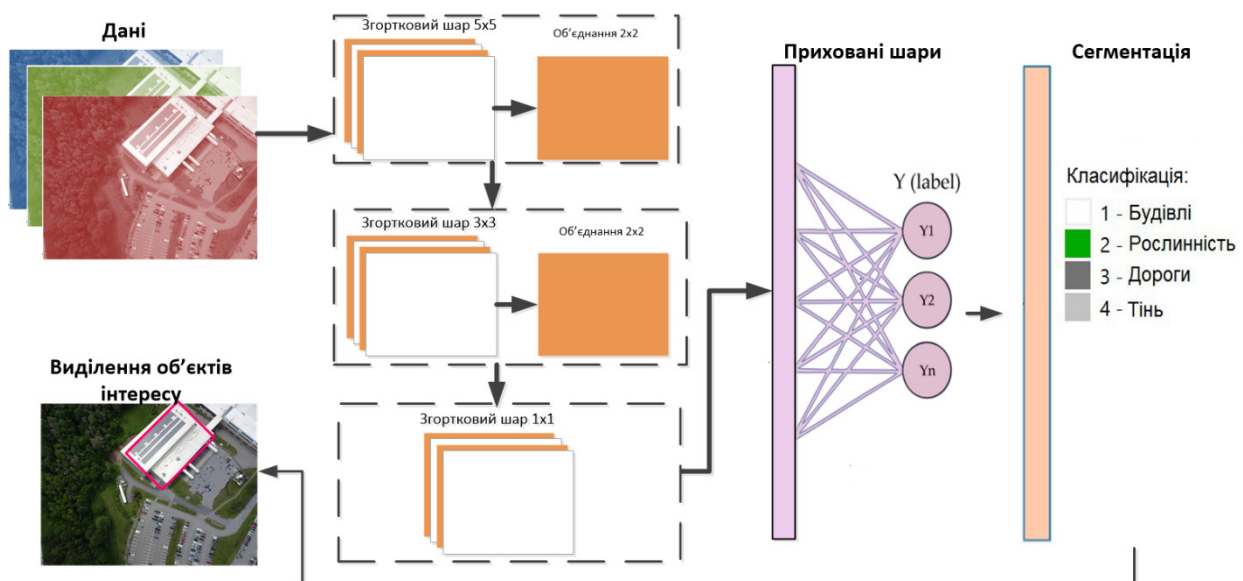
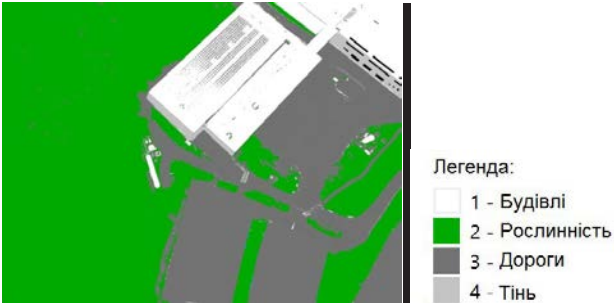


Рис. 3. Архітектура нейронної мережі для виділення об'єктів забудови

ції втрат перехресної ентропії. Багатоступенева маскова втрата  $L_M$  формулюється наступним чином (Yilmaz, 2020, р. 1–4):

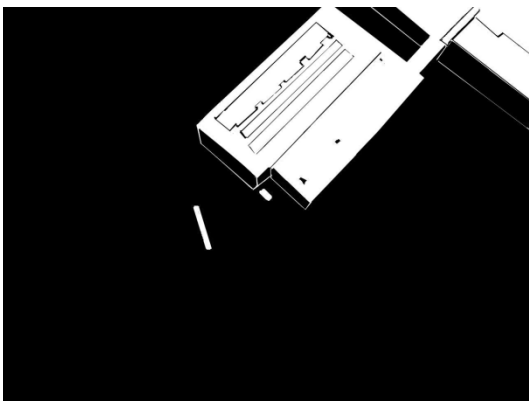
$$L_M = - \sum_{i,j} [y_{i,j} \log(\hat{y}_{i,j}) + (1 - y_{i,j}) \log(1 - \hat{y}_{i,j})] \quad (2)$$



**Рис. 4. Результат сегментації БПЛА зображення**

При побудові граничного прогнозу негативні приклади становлять більшість усіх зразків. У нашому експерименті ми вибираємо втрату Dice як функцію втрат для граничного прогнозування, що дозволяє зменшити ступінь навчання на простих негативних прикладах та підвищити значення F1-балу. Контурні втрати  $L_C$  визначаються наступним чином (Kim, 2020, р. 3755):

$$L_C = \frac{2 \sum_{i,j} y_{i,j} \cdot \hat{y}_{i,j}}{\sum_{i,j} y_{i,j}^2 + \sum_{i,j} \hat{y}_{i,j}^2} \quad (3)$$



**Рис. 5. Результат бінарної маски виділення об'єктів будови на БПЛА зображенні**

Запропонований метод розпізнавання об'єктів будови на основі загорткових мереж покращує точність сегментації на БПЛА зображеннях (рис. 6).

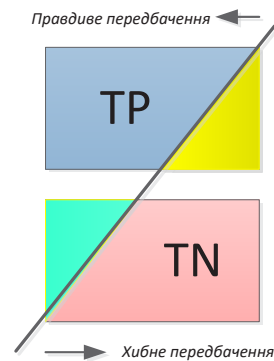
**Експерименти.** Нами проведено дослідження ефективності запропонованого в роботі методу та порівняння з існуючими підходами

для оцінки розпізнавання об'єктів будови. Було використано:

– матрицю помилок (confusion matrix), що дозволяє для кожного класу розрахувати кількість істинно позитивних передбачень (True Positive,  $TP$ ), істинно негативних (True Negative,  $TN$ ), хибно позитивних (False Positive,  $FP$ ) і хибно негативних передбачень (False Negative,  $FN$ ) (Mäyrä, 2021, р. 112–322). Схематично матриця помилок представлена на рисунку 7;



**Рис. 6. Результат неймережевого виділення об'єктів будови на БПЛА зображенні**



**Рис. 7. Графічне представлення матриці помилок**

– метрику intersection over union ( $IoU$ ) або коефіцієнт Жаккара (Rahman, 2020, р. 32–33):

$$IoU = \frac{|(G_d \cap G) \cap (P_d \cap P)|}{|(G_d \cap G) \cup (P_d \cap P)|}, \quad (4)$$

де  $G$  позначає двійкову маску зображення,  $P$  позначає прогнозовану двійкову маску,  $d$  позначає ширину пікселів граничної області, а  $G_d$  і  $P_d$  відносяться до наборів пікселів маски базової правди та прогнозованих пікселів маски відповідно, які знаходяться в межах задану відстань від кожного контуру;

– загальну точність ( $OA$ ), що вимірює відсоток загальної кількості правильно класифікованих

пікселів класу об'єктів забудови, на загальну кількість пікселів (Kim, 2020; Yang, 2019; Mäyrä, 2021):

$$OA = \frac{\sum D_{ij}}{N} \quad (5)$$

де  $D_{ij}$  – загальна кількість правильно класифікованих пікселів у рядку  $i$  та стовпці  $j$ ;

– середню точність (AA) (Kim, 2020; Mäyrä, 2021):

$$AA = \frac{\sum PA_m}{m}; \quad (6)$$

– точність для кожного класу (PA) (Yilmaz, 2020, р. 1-4):

$$PA = \frac{D_{ij}}{R_i} \quad (7)$$

де  $R_i$  – загальна кількість пікселів у рядку;

– індекс Каппа ( $K$ ) – є кількісним показником рівня узгодженості або точності в правильно класифікованих пікселях (Mäyrä, 2021). Значення карра 1 вказує на повну згоду, тоді як карра 0 вказує на випадкову згоду, еквівалентну ймовірності класифікації пікселів (Каштан, 2019, с. 3–11):

$$K = \frac{N \sum_{i,j=1}^m D_{ij} - \sum_{i,j=1}^m R_i \cdot C_j}{N^2 - \sum_{i,j=1}^m R_i \cdot C_j} \quad (8)$$

де  $m$  – кількість класів,  $D_{ij}$  – кількість правильно класифікованих пікселів у рядку  $i$  та стовпці  $j$ ,  $R_i$  – загальна кількість пікселів у рядку  $i$ ,  $C_j$  – загальна кількість пікселів у стовпці  $j$ ,  $N$  – загальна кількість пікселів.

Частина правильних відповідей є метрикою, що показує, покращення навчальних ітерацій продуктивності моделі (Yang, 2019, Mäyrä, 2021):

$$Accuracy = \frac{\sum_{i=1}^m C_{ii}}{\sum_{i=1}^m \sum_{j=1}^m C_{ij}} \quad (9)$$

де  $C_{ii}$  – правильна класифікація по діагоналі,  $m$  – кількість класів,  $C_{ij}$  – кількість разів, коли предмети класу  $i$  класифікувалися як клас  $j$  (неправильна класифікація) та  $\sum_{i=1}^m \sum_{j=1}^m C_{ij}$  – загальна кількість зразків, які були оцінені.

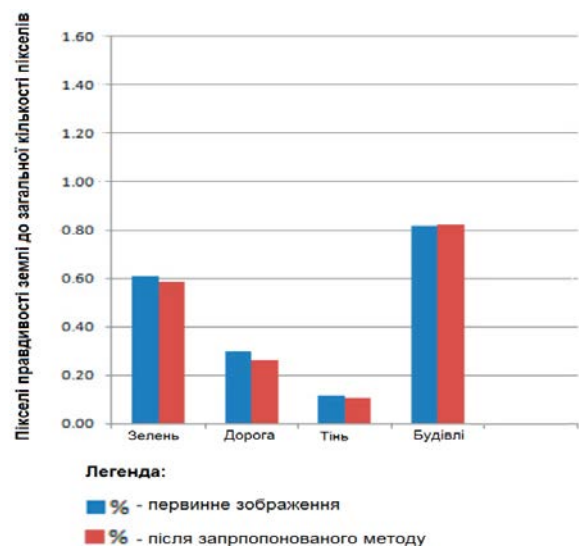
Процес класифікації передбачає узагальнення, існує певний рівень втрати інформації, що призводить до втрати повноти кінцевого результату (Cheng, 2018, р. 1865–188). Тому, пропонуємо розрахувати функцію втрат як:

$$втрати = 1 - Точність \quad (10)$$

**Результати.** Запропонований в роботі метод нейромережевого розпізнавання об'єктів забудови на БПЛА-зображеннях

реалізовано на основі структури глибокого навчання PyTorch на одному графічному процесорі Nvidia RTX 3090, на системі Intel Core-i3 8<sup>th</sup> Gen (2,90 ГГц) з 8 ГБ оперативною пам'яттю. Нейронна модель навчена за допомогою оптимізатора середньоквадратичного розповсюдження (RMSprop), і швидкість навчання ініціалізована на 0,0001 з розпадом ваги  $1 \times 10^{-8}$ . Загальна кількість епох встановлена на 150 з розміром 16. Під час процесу навчання на кожній ітерації використовується набір перевірки для обчислення індукованої похибки після завершення процесу навчання моделі. Використано стратегію ранньої зупинки, щоб запобігти перенавчанню моделі; зокрема, процедура навчання зупиняється автоматично, якщо значення IoU моделі в наборі перевірки не змінюється протягом 12 ітерацій. Аналогічно використано стратегію ReduceLRonPlateau, щоб зменшити швидкість навчання, коли продуктивність мережі на верифікаційному наборі даних не покращується протягом певної кількості епох. На етапі тестування тестовий набір даних використовується для оцінки запропонованої моделі, яка генерує бінарну маску зі значеннями 0 і 1.

На рисунку 8 зображено відсоток пікселів наборів даних для кожного класу земельного покриття для тестового зображення БПЛА та після запропонованого методу в роботі. Так, кількість пікселів для класів доріг з твердим покриттям в обох наборах даних становила приблизно 0,3%, тоді як відсотки для класу забудови були близькими до 0,8%, що було майже в 2,6 рази більше вибірки класу доріг з твердим покриттям.



**Рис. 8.** Відсоток і розподіл усіх GT до загальної кількості класифікованих пікселів для кожного класу земного покриття

Таблиця 1

**Ефективність методів класифікації**

Тип	Модель	OA	AA	K
Навчання	CNN з цифровою моделлю	<b>0.982</b>	<b>0.978</b>	<b>0.984</b>
	CNN без цифровою моделлю	0.956	0.936	0.949
Тестування	CNN з цифровою моделлю	<b>0.979</b>	<b>0.963</b>	<b>0.978</b>
	CNN без цифрової моделі	0.961	0.953	0.967

Таблиця 2

**Точність (PA) для кожного класу**

Клас	CNN	Запропонований метод
Рослинність	0.971	0.982
Дороги	0.987	0.994
Тіні	0.879	0.981
Будівлі	0.953	0.992

В таблиці 1 наведено результат методів класифікації CNN з цифровою та без цифровою моделями. Загальна точність (OA) має кращі результати на основі запропонованого методу в роботі (з 0,968 до 0,980), середня точність (AA) – з 0,952 до 0,970. І навпаки, найвище значення індексу Каппа (K) було виміряно (0,988) у даних навчання на основі реалізації загорткових нейронних мереж. Дослідження показали, що нейронна мережа без урахування цифрової моделі має нижчу продуктивність, ніж об'єднані набори даних з цифровою моделлю.

В таблиці 2 наведено результати точності для кожного розглянутого класу: зелень, дороги, тіні, будівлі. Ці результати свідчать про те, що модель на основі запропонованого в роботі методу змогла класифікувати майже всі класи з відносно високою точністю. Максимальна точність становила 1,0 для класу дороги та будівлі, тоді як мінімальна точність належала класу тіні (0,981).

**Обговорення.** Для дослідження розпізнавання границь об'єктів за будови пропонуємо проаналізувати метрику *IoU*. Результати наведено на рисунку 9. Порівняно з методом Baseline (базова лінія), запропонований метод забезпечує точне виділення контурів об'єктів за будови.

В таблиці 3 наведено результати дослідження для тестового набору даних. Значення *IoU* збільшується на 1,3%, вказуючи на те, що допоміжна інформація про контур може зробити границі більш регулярними та підвищити точність моделі. Результати, наведені в табл. 3, підтверджують кращу ефективність запропонованого методу порівняно з іншими методами оцінки розпізнавання об'єктів за будови.



Рис. 9. Результат метрики IoU

Таблиця 3

**Результати дослідження метрики IoU**

Метод	IoU (%)
Baseline	76,3
<b>Запропонований метод</b>	<b>81,6</b>

**Висновки.** У даній роботі пропонується метод автоматизованого розпізнавання об'єктів за будови на цифрових аерофотознімках з використанням глибокого навчання. Новий метод дозволяє точніше визначати границі будівель, завдяки використанню функції втрати багатозадачності для отримання додаткової інформації про межі. Порівняльні експерименти показали, що запропонований метод значно покращує точність розпізнавання будівель у загальнодоступному наборі даних. Зокрема, метрики OA, AA і K покращилися на 2,6%, 5,6% і 3,2% відповідно для навчального набору даних і на 1,2%, 1,8% і 1,5% для тестового набору даних. Використання запропонованого методу з урахуванням цифрової моделі успішно покращило PA як для тестових, так і для навчальних наборів даних.



## ЛІТЕРАТУРА:

1. Kalantar B., Halin A.A., Al-Najjar H.A.H., Mansor S., van Genderen J.L., Shafri H.Z.M., Zand, M. A Framework for Multiple Moving Objects Detection in Aerial Videos. *In Spatial Modeling in GIS and R for Earth and Environmental Sciences. Elsevier: Amsterdam, The Netherlands.* 2019. P. 573–588.
2. Kalantar B., Mansor S.B., Halin A.A., Shafri H.Z.M., Zand, M. Multiple moving object detection from UAV videos using trajectories of matched regional adjacency graphs. *IEEE Trans. Geosci. Remote Sens.* 2017. Vol. 55. P.5198–5213.
3. Kalantar B., Mansor S.B., Sameen M.I., Pradhan B., Shafri H.Z.M. Drone-based land-cover mapping using a fuzzy unordered rule induction algorithm integrated into object-based image analysis. *Int. J. Remote Sens.* 2017. Vol.38. P. 2535–2556.
4. Waleed A., Turky A., Basil D. Detecting Buildings and Nonbuildings from Satellite Images Using U-Net. *Computational Intelligence and Neuroscience*, 2018. P.1687-5265. DOI: 10.1155/2022/4831223.
5. Krówczyńska M., Raczko E., Staniszevska N., Wilk, E. Asbestos–cement roofing identification using remote sensing and convolutional neural networks (CNNs). *Remote Sens.* 2020. Vol. 12, P.2-16.
6. Cheng G., Han J., Lu X. Remote sensing image scene classification: Benchmark and state of the art. *Proceedings of the IEEE*, 2017. Vol.105 (10). P. 1865-188.
7. Avudaïamma R., Dayana S., Prabhu R. and Swarnalatha A. Automatic Building Extraction from VHR Satellite Image. *International Conference on Current Trends towards Converging Technologies (ICCTCT)*. 2018. P. 1-6, DOI: 10.1109/ICCTCT.2018.8551119.
8. Long J., Shelhamer E., Darrell T. Fully convolutional networks for semantic segmentation. *In Proceedings of the 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Boston, MA, USA.* 2015. P. 3431–3440.
9. Liu L., Ouyang W., Wang X., Fieguth W.P., Chen J., Liu X., Pietikinen M. Deep learning for generic object detection. *A survey. I International Journal of Computer Vision.* 2020. Vol. 128. P. 261–31.
10. Yuan J. Learning building extraction in aerial scenes with convolutional networks. *IEEE Trans. Pattern Anal. Mach. Intell.* 2017. P. 2793–2798.
11. Xu Y., Wu L., Xie Z., Chen Z. Building extraction in very high resolution remote sensing imagery using deep learning and guided filters. *Remote Sens.* 2018. Vol. 10. P. 144.
12. Huang J., Zhang X., Xin Q., Sun Y., Zhang P. Automatic building extraction from high-resolution aerial images and LiDAR data using gated residual refinement network. *ISPRS J. Photogramm. Remote Sens.* 2019. P. 91–105.
13. Kim E.K., Lee H., Kim J.Y., Kim S. Data Augmentation Method by Applying Color Perturbation of Inverse PSNR and Geometric Transformations for Object Recognition Based on Deep Learning. *Appl. Sci.* 2020. Vol. 10. P. 3755. DOI: 10.3390/app10113755
14. Yang J., Guo J., Yue H., Liu Z., Hu H., Li K. CDnet: CNN-Based Cloud Detection for Remote Sensing Imagery. *IEEE Trans. Geosci. Remote Sens.* 2019. Vol. 57. P. 6195–6211.
15. Rahman T., Chowdhury M.E., Khandakar A., Islam K.R., Islam K.F., Mahbub Z.B., Kadir M.A., Kashem S. Transfer Learning with Deep Convolutional Neural Network (CNN) for Pneumonia Detection using Chest X-ray. *Appl. Sci.* 2020. Vol. 10. P. 32-33.
16. Mäyrä J., Keski-Saari S., Kivinen S., Tanhuanpää T., Hurskainen P., Kullberg P., Poikolainen L., Viinikka A., Tuominen S., Kumpula T., et al. Tree species classification from airborne hyperspectral and LiDAR data using 3D convolutional neural networks. *Remote Sens. Environ.* 2021. Vol.256. P. 112-322.
17. Каштан В.Ю., Гнатушенко В.В. Контурна сегментація цифрових супутникових знімків з використанням особливих точок вейвлет-перетворення. *Системні технології. Регіональний міжвузівський збірник наукових робіт.* 2019. Випуск 1 (120). С.3 - 11.
18. Yilmaz I., Imamoğlu M., Özbülak G., Kahraman F., Aptoula E. Large Scale Crop Classification from Multi-temporal and Multi-spectral Satellite Images. *In Proceedings of the 2020 28th Signal Processing and Communications Applications Conference (SIU).* 2020. P. 1–4.
19. Cheng H., Lian D., Gao S., Geng Y. Evaluating Capability of Deep Neural Networks for Image Classification via Information Plane. *In Proceedings of the European Conference on Computer Vision (ECCV).* 2018. P. 168–182.

## REFERENCES:

1. Kalantar B., Halin A.A., Al-Najjar H.A.H., Mansor S., van Genderen J.L., Shafri H.Z.M., Zand, M. A Framework for Multiple Moving Objects Detection in Aerial Videos. *In Spatial Modeling in GIS and R for Earth and Environmental Sciences. Elsevier: Amsterdam, The Netherlands.* 2019. P. 573–588.

2. Kalantar B., Mansor S.B., Halin A.A., Shafri H.Z.M., Zand, M. Multiple moving object detection from UAV videos using trajectories of matched regional adjacency graphs. *IEEE Trans. Geosci. Remote Sens.* 2017. Vol. 55. P.5198–5213.
3. Kalantar B., Mansor S.B., Sameen M.I., Pradhan B., Shafri H.Z.M. Drone-based land-cover mapping using a fuzzy unordered rule induction algorithm integrated into object-based image analysis. *Int. J. Remote Sens.* 2017. Vol.38. P. 2535–2556.
4. Waleed A., Turkey A., Basil D. Detecting Buildings and Nonbuildings from Satellite Images Using U-Net. *Computational Intelligence and Neuroscience*, 2018. P.1687-5265. DOI: 10.1155/2022/4831223.
5. Krówczyńska M., Raczko E., Staniszevska N., Wilk, E. Asbestos–cement roofing identification using remote sensing and convolutional neural networks (CNNs). *Remote Sens.* 2020. Vol. 12, P.2-16.
6. Cheng G., Han J., Lu X. Remote sensing image scene classification: Benchmark and state of the art. *Proceedings of the IEEE*, 2017. Vol.105 (10). P. 1865-188.
7. Avudaamma R., Dayana S., Prabhu R. and Swarnalatha A. Automatic Building Extraction from VHR Satellite Image. *International Conference on Current Trends towards Converging Technologies (ICCTCT)*. 2018. P. 1-6, DOI: 10.1109/ICCTCT.2018.8551119.
8. Long J., Shelhamer E., Darrell T. Fully convolutional networks for semantic segmentation. *In Proceedings of the 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Boston, MA, USA.* 2015. P. 3431–3440.
9. Liu L., Ouyang W., Wang X., Fieguth W.P., Chen J., Liu X., Pietikinen M. Deep learning for generic object detection. *A survey. I International Journal of Computer Vision.* 2020. Vol. 128. P. 261–31.
10. Yuan J. Learning building extraction in aerial scenes with convolutional networks. *IEEE Trans. Pattern Anal. Mach. Intell.* 2017. P. 2793–2798.
11. Xu Y., Wu L., Xie Z., Chen Z. Building extraction in very high resolution remote sensing imagery using deep learning and guided filters. *Remote Sens.* 2018. Vol.10. P. 144.
12. Huang J., Zhang X., Xin Q., Sun Y., Zhang P. Automatic building extraction from high-resolution aerial images and LiDAR data using gated residual refinement network. *ISPRS J. Photogramm. Remote Sens.* 2019. P. 91–105.
13. Kim E.K., Lee H., Kim J.Y., Kim S. Data Augmentation Method by Applying Color Perturbation of Inverse PSNR and Geometric Transformations for Object Recognition Based on Deep Learning. *Appl. Sci.* 2020. Vol. 10. P. 3755. DOI: 10.3390/app10113755
14. Yang J., Guo J., Yue H., Liu Z., Hu H., Li K. CDnet: CNN-Based Cloud Detection for Remote Sensing Imagery. *IEEE Trans. Geosci. Remote Sens.* 2019. Vol. 57. P. 6195–6211.
15. Rahman T., Chowdhury M.E., Khandakar A., Islam K.R., Islam K.F., Mahbub Z.B., Kadir M.A., Kashem S. Transfer Learning with Deep Convolutional Neural Network (CNN) for Pneumonia Detection using Chest X-ray. *Appl. Sci.* 2020. Vol. 10. P. 32-33.
16. Mäyrä J., Keski-Saari S., Kivinen S., Tanhuanpää T., Hurskainen P., Kullberg P., Poikolainen L., Viinikka A., Tuominen S., Kumpula T., et al. Tree species classification from airborne hyperspectral and LiDAR data using 3D convolutional neural networks. *Remote Sens. Environ.* 2021. Vol.256. P. 112-322.
17. Kashtan V.Yu., Hnatushenko V.V. Konturna sehmentatsiia tsyfrovoykh suputnykovykh znimkiv z vykorystanniam osoblyvykh tochok veivlet-peretvorennia. *Systemni tekhnolohii. Rehionalnyi mizhvuzivskiy zbirnyk naukovykh robit.* 2019. Vypusk 1 (120). S.3 - 11.
18. Yilmaz I., Imamoğlu M., Özbulak G., Kahraman F., Aptoula E. Large Scale Crop Classification from Multi-temporal and Multi-spectral Satellite Images. *In Proceedings of the 2020 28th Signal Processing and Communications Applications Conference (SIU).* 2020. P. 1–4.
19. Cheng H., Lian D., Gao S., Geng Y. Evaluating Capability of Deep Neural Networks for Image Classification via Information Plane. *In Proceedings of the European Conference on Computer Vision (ECCV).* 2018. P. 168–182.