

КЛАСИФІКАЦІЯ ЗОБРАЖЕНЬ ДИСТАНЦІЙНОГО ЗОНДУВАННЯ НА ОСНОВІ БАГАТОПОРОВОГОЇ БІНАРИЗАЦІЇ

Б. П. Русин, О. А. Луцик, Р. Я. Косаревич, В. В. Корній

Фізико-механічний інститут ім. Г. В. Карпенка НАН України, Львів

E-mail: olutsyk@ipm.lviv.ua

Запропоновано новий підхід для виділення ознак під час класифікації та розпізнавання багатоспектральних зображень, отриманих засобами дистанційного зондування з використанням мультипорогової бінаризації. Встановлено, що ознаки багатоспектральних зображень можна виділити набором деякого числа бінарних характеристик, кожна з яких описує певні ознаки зображення. Виявлено залежність між розміром навчальної вибірки та точністю роботи запропонованого класифікатора, тренуваного на основі цієї вибірки. Розроблений підхід застосовано до тестових навчальних вибірок багатоспектральних зображень та наведено результати їх оцінювання. Встановлено, що час навчання моделі значно швидший, ніж нейронної мережі зі згортковими шарами. Це дає можливість використовувати запропонований підхід для моделей глибокого навчання в задачах розпізнавання та класифікації багатоспектральних зображень у реальному масштабі часу.

Ключові слова: багатоспектральні зображення, мультипорогова бінаризація, класифікація.

CLASSIFICATION OF REMOTE SENSING IMAGES BASED ON MULTI-THRESHOLD BINARIZATION

B. P. Rusyn, O. A. Lutsyk, R. Ya. Kosarevych, V. V. Korniy

H. V. Karpenko Physico-Mechanical Institute of the NAS of Ukraine, Lviv

In the field of remote sensing, the classification and recognition of multi-spectral images play a key role in understanding and monitoring our environment. This paper introduces an innovative method, which uses multi-threshold binarization for feature extraction in the classification and recognition of multi-spectral images. The core idea of this approach is to represent the intricate characteristics of multi-spectral images through a series of binary attributes, each of which captures specific image properties. One of the key findings of this research is the correlation between the size of the training dataset and the performance accuracy of the classifier trained on this dataset. This relationship shows the importance of having an appropriately sized training dataset to achieve high precision of classification. Moreover, the approach proposes a significant advantage in terms of efficiency compared to traditional convolutional neural network-based training processes. The reduced training time and faster operation of the method open up exciting possibilities for real-time recognition and classification of multi-spectral images. To validate the effectiveness of the approach, we conducted experiments on various test training datasets of multi-spectral images, and the results of these evaluations are presented. The findings highlight the promising potential of this method for a wide range of applications, including environmental monitoring, urban planning, and agricultural management, where accurate and real-time classification is essential. A novel approach to multi-spectral image classification, employing multi-threshold binarization for feature extraction is proposed. The method ability to represent image features as binary characteristics offers a fresh perspective in the field of remote sensing. The results of this research not only contribute to advancing the state of the art in multi-spectral image analysis but also provide a practical and efficient solution for real-time recognition and classification, aligning with the growing demands of various applications.

Keywords: multi-spectral image, multi-threshold binarization, classification.

Вступ. Сьогодні методи дистанційного зондування надзвичайно важливі, оскільки з їх допомогою розв'язують багато задач людської діяльності. Великі потоки даних обробляють, аналізують і на виході системи прийняття рішень от-

© Б. П. Русин, О. А. Луцик, Р. Я. Косаревич, В. В. Корній, 2023

римують експертні оцінки, прогнози та рішення. Усе це стає можливим через використання машинного навчання та моделей глибокого навчання.

З допомогою засобів штучного інтелекту вдалось вийти на новий рівень в оперативності прийняття рішень, якості та точності аналізу, дало можливість обробляти значно більші масиви даних [1]. Засобами штучного інтелекту можна в лічені секунди досягнути результату порівняно з класичним підходом. В багатьох випадках системи аналізу даних, отриманих методами дистанційного зондування, забезпечують режим роботи в реальному масштабі часу [2]. Їх особливістю є висока швидкодія, яку досягають оптимальністю алгоритмів та достатніми обчислювальними можливостями апаратного забезпечення. У результаті така система опрацьовує 15–20 кадрів на секунду, що сприймає людина як роботу в реальному часі. Такі часові рамки опрацювання інформації не є межею, оскільки існує чимало задач, де ставлять значно вищі вимоги до систем оброблення інформації і час оброблення може становити лічені мілісекунди [3].

Найскладніше виділити ознаки [4], від чого суттєво залежить робота класифікатора. Цей процес може займати до 80% обчислювальних ресурсів [5]. У праці [6] детально описано системи прийняття рішень на основі моделей глибокого навчання, які застосовують до зображень та відеопотоків. Виділяють ознаки з допомогою згорткових нейронних мереж. Процес оброблення та розпізнавання відбувається в режимі реального часу внаслідок оптимального використання згорткових та повнозв'язних шарів моделі. Найкращий результат одержано за багатошарової структури нейронної мережі. Недоліком запропонованої архітектури є вузька спрямованість моделі та відсутність простих методів масштабування.

Принципово іншим підходом виділення ознак є побудова моделей глибокого навчання на основі бінарних нейронних мереж [7], які використовують тільки двійкові ваги та функції активації, і ефективні у випадках, де основною вимогою є висока швидкодія на базі простішого апаратного забезпечення. Цей підхід добре зарекомендував себе в задачах сегментації, де важливо узагальнювати ознаки зображення. Застосовуючи бінарну згорткову мережу, можна досягнути задовільних показників точності за високої швидкодії, при цьому значно скоротити обсяги оперативної пам'яті, необхідної для навчання та роботи моделі [8].

У праці [9] описано моделі на основі згорткових нейронних мереж, які використовують у задачах прогнозування. Виділяють ознаки з допомогою згорткових шарів мережі. Такі моделі схильні до перенавчання, тобто до запам'ятовування вхідних даних, а не їх узагальнення. Щоб уникнути цього, пропонують вносити адитивний шум щоразу безпосередньо перед подаванням даних на згорткові шари. Також експериментально вибирають оптимальний розмір вибірки для навчання моделі.

Таким чином, залежно від поставленої задачі будують конкретну модель нейронної мережі [10], беручи до уваги структурні особливості, вимоги до обчислювальної складності та розмір навчальної вибірки. Кожну нейромережеву модель можна структурно розділити на частини для виділення ознак та класифікатора. Частина виділення ознак, як правило, складається зі згорткових шарів. Можливе використання LSTM шарів, що дає можливість поєднати виділення ознак з генераційною моделлю.

У задачах дистанційного зондування найчастіше використовують багатоспектральні зображення [11] – сукупність серії монохромних зображень однієї і тієї ж сцени, кожне з яких отримано різними давачами. Кожен канал цього зображення подають певною смугою частот оптичного діапазону. Таким чином, кольорове зображення RGB, що складається з червоного, зеленого та синього зображень, є частковим багатоспектральним зображенням, оскільки кожен з каналів формують давачем, чутливим до різних довжин хвиль. Супутники зазвичай отри-

мують кілька зображень із діапазонів частот у візуальному та невізуальному діапазонах. Для прикладу, Landsat 5 працює зі сімома смуговими зображеннями з довжинами хвильових діапазонів від 450 до 1250 нм. Тут багатоспектральні зображення підвищують інформативність видимого спектра та додатково поширюються на невидимі його області.

Існують спеціальні методи оброблення багатоспектральних зображень. У задачах їх сегментації [12] піксель, який належить до однієї конкретної області, в різних смугах створює вектор ознак з різними рівнями інтенсивності, що описує його розташування в k -вимірному просторі ознак. Найпростіший спосіб визначити клас – вибрати верхній і нижній пороги для кожної смуги, створюючи так m -вимірний гіперкуб у просторі ознак. Якщо вектор ознак пікселя збігається з місцем локалізації в кубі, то його класифікують як приналежний до цього класу. В багатьох задачах сегментацію зображень достатньо звести до бінаризації, а отже, спростити обчислення та уникнути невизначеності [13].

Основний недолік під час оброблення та розпізнавання багатоспектральних зображень є оброблення додаткових масивів даних, при цьому час обчислень і використовувана оперативна пам'ять значно збільшуються, що призводить до суттєвих обмежень, оскільки унеможливорює використання моделей глибокого навчання в реальному часі. Тому актуально опрацювати методи виділення ознак багатоспектральних зображень.

Виділення ознак багатоспектральних зображень. Під час побудови інформативних ознак важливо досягти деяких умов інваріантності, на основі яких класифікаційний алгоритм має функціонувати. Для цього використані методи пороговування, придатні для багатьох типів напівтонових та кольорових зображень. Встановлюючи глобальний поріг, можна перейти від зображення до бінарного подання, яке можна розцінювати як карту бінарних ознак. Так вдається кластеризувати цілі області зображення, надаючи значення 0 та 1, відповідно. В цьому випадку порогова бінаризація має вигляд

$$\begin{cases} B(x, y) = 0, & \text{якщо } f(x, y) < T; \\ B(x, y) = 1, & \text{якщо } f(x, y) \geq T, \end{cases} \quad (1)$$

де $B(x, y)$ – бінарне подання; $f(x, y)$ – зображення градації сірого; T – поріг бінаризації.

Для виділення ознак багатоспектральних зображень вживали багатопорогову бінаризацію, яка полягає в тому, що залежно від виставленого порога можна по-різному бінаризувати окремі канали зображення і за отриманою кількістю бінаризованих подань описати ознаки зображення, а отже, побудувати інформативний вектор ознак.

Спочатку попіксельно аналізували багатоспектральне зображення, враховуючи усі смуги, наведені у вигляді набору матриць. Визначали глобальне порогове значення, яке подали як середнє значення серед усього набору даних:

$$H_G = 1 / n_G \sum_{i=0}^{n_G} P_i, \quad (2)$$

де P_i – інтенсивність окремого пікселя, $n_G = x \times y \times m$ – загальна їх кількість з урахуванням всіх смуг m .

Також враховували верхній і нижній пороги, які визначали як максимальне та мінімальне значення пікселів зображення:

$$H_U = \max(I), \quad (3)$$

$$H_D = \min(I), \quad (4)$$

де H_U , H_D – верхній та нижній порогі, відповідно; I – напівтонове зображення.

Експериментально встановили, що для однієї смуги достатньо набору з 5–7 порогів та один глобальний поріг. За набору з $r = 7$ порогів найкраще працює еквідистантне взаємне розміщення.

Багатопороговою бінарizaцією можна подати окрему смугу багатоспектрального зображення I у вигляді набору матриць L , елементами яких є бінарні величини (рис. 1). При цьому ширина і висота матриці $L_m(x, y)$ повністю відповідають розміру зображення $I(x, y)$.

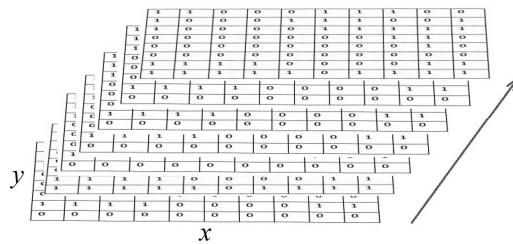


Рис. 1. Подання окремої смуги багатоспектрального зображення у вигляді набору зі семи матриць після багатопорогової бінарizaції.

Залежно від кількості смуг m багатоспектрального зображення мінятиметься кількість матриць L_m . Тому для їх об'єднання використовуємо поелементну виключну диз'юнкцію:

$$L = (L_1 \oplus L_2) \vee (L_3 \oplus L_4) \vee \dots (L_{m-1} \oplus L_m), \quad (5)$$

де L – результуюча бінарна матриця.

Якщо не передбачаємо інваріантності до поворотів, матрицю L подаємо безпосередньо на класифікатор. Це можуть бути вхідні дані для застосування як класичних підходів класифікації, так і моделей глибокого навчання на основі повнозв'язних шарів. У багатьох задачах розпізнавання зображень вимагають інваріантності до повороту, чого досягають різними підходами, зокрема, в методах глибокого навчання, – створюючи навчальну вибірку з повертанням зображень на деякий кут [14]. Так модель тренується класифікувати зображення під довільним кутом, що суттєво збільшує навчальну вибірку.

Сучасні системи дистанційного зондування обладнані великою кількістю давачів, які точно визначають висоту зображення, GPS прив'язку, орієнтацію в просторі. Для кожного зображення задають параметри орієнтації в кватерніонах Ω виразом

$$\Omega = a + bi + cj + dk, \quad (6)$$

де a, b, c, d – дійсні числа; i, j, k – уявні.

Для переходу від подання в кватерніонах в кути Ейлера використовують співвідношення

$$\varphi = \arctan\left(\frac{2(ab + cd)}{1 - 2(b^2 + c^2)}\right), \quad (7)$$

$$\theta = \arcsin(2(ac - db)), \quad (8)$$

$$\psi = \arctan\left(\frac{2(ac+bc)}{1-2(c^2+d^2)}\right), \quad (9)$$

де φ , θ , ψ – кути повороту відносно осей x , y , z , відповідно, які можна застосувати для повороту матриці ознак L_m та вирівнювання перспективи. Таким чином, кожна смуга багатоспектрального зображення $L_m(x, y)$ повертається на кут φ . Якщо поворот здійснюється відносно початку координат $(0, 0)$, то перехід відбувається згідно з виразами

$$x_2 = \cos(\varphi)x_1 + \sin(\varphi)y_1, \quad (10)$$

$$y_2 = -\sin(\varphi)x_1 + \cos(\varphi)y_1, \quad (11)$$

де x_2 , y_2 – координати після повороту.

Оскільки зображення можна отримувати під різними кутами до нормалі, то слід враховувати перетворення перспективи, використовуючи для цього матриці повороту R_θ , R_ψ відносно осей y та z .

Попередньою обробкою матриці ознак L_m досягають інваріантності до повороту та вирівнювання проекції, забезпечуючи так високу швидкодію під час використання моделей глибокого навчання.

Аналіз та порівняння результатів. Запропонована модель класифікації багатоспектральних зображень передбачає попереднє виділення ознак, яке базується на багатопороговій бінаризації їх смуг. Із кожної смуги виділяють певну кількість бінарних зображень для різних порогів бінаризації, об'єднують їх та формують вектор ознак. При цьому забезпечують інваріантність до повороту операторами повороту та вирівнюванням проекції і навчанням нейронної мережі на основі повнозв'язних шарів.

Оптимальну кількість порогів для смуг багатоспектрального зображення встановлюють експериментально. Зі залежностей на рис. 2. можна зробити висновки, що сім порогів добре працює у більшості випадків, оскільки їх збільшення не призводить до суттєвого росту ймовірності правильної класифікації.

Для порівняння властивостей запропонованого підходу з моделями RESNET та ансамблю CNN виконали експерименти, використовуючи бази даних багатоспектральних зображень, сформованих на основі дистанційного зондування Землі. Їх розбивають на навчальну, тестову та валідаційну вибірки у співвідношенні 60; 20 та 20%, окремо додають мітки реальних класів, які доступні на електронній сторінці університету Валенсії: база даних TipJul1 на основі даних Landsat має сім смуг, роздільна здатність кожної 169 на 169 пікселів, 5 класів; база даних Zurich на основі даних QuickBird має чотири смуги, роздільна здатність кожної 828 на 889 пікселів, 4 класи.

Отже, різні моделі глибокого навчання проявляють різні властивості класифікації, навіть для однієї навчальної вибірки (рис. 3). Запропонований підхід за малих розмірів вибірки дає вищі значення ймовірності правильної класифікації порівняно із моделями RESNET та ансамблем CNN. Це пов'язано з тим, що такі моделі для виділення ознак використовують згорткові шари, яким необхідна велика навчальна вибірка для тренування. Зокрема, за малих розмірів навчальної вибірки (200–800 зображень) запропонована модель дає кращі результати ймовірності правильної класифікації, проте дещо поступається за суттєво більших її розмірів (більше 1800 зображень).

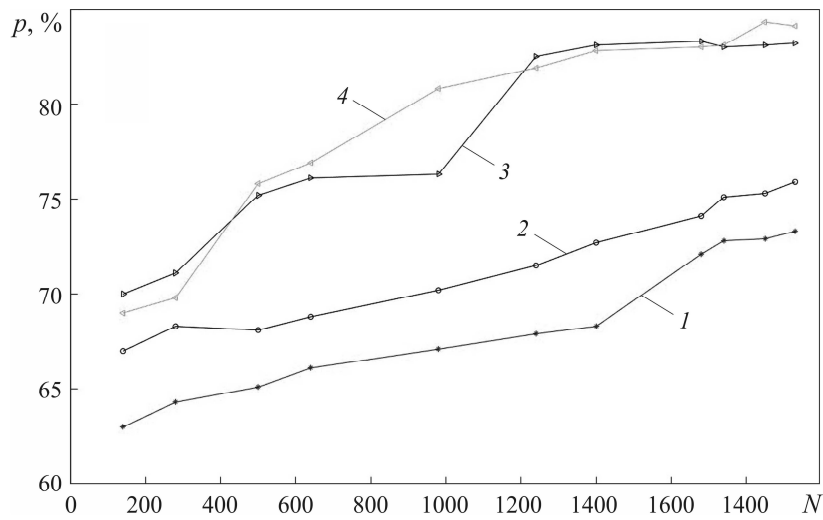


Рис. 2. Залежність ймовірності правильної класифікації від розміру навчальної вибірки бази Zurich для різної кількості порогів: 1 – 3 порогів; 2 – 5; 3 – 7; 4 – 9.

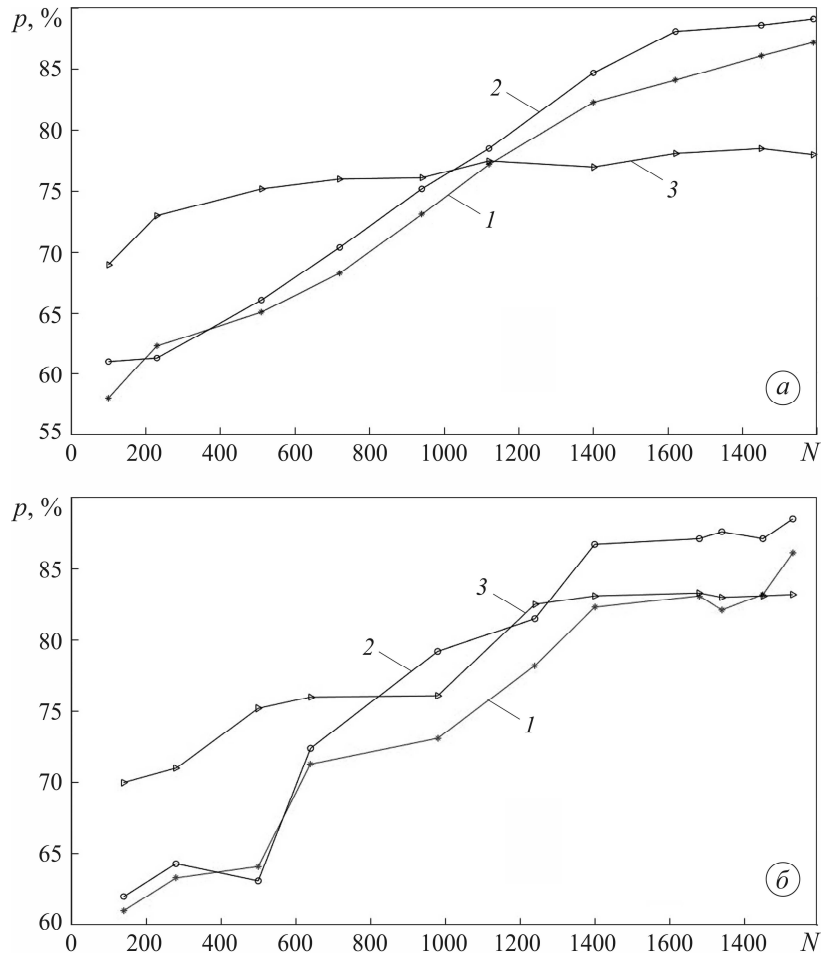


Рис. 3. Залежність ймовірності правильної класифікації від розміру навчальної вибірки баз TipJul1 (а) та Zurich (б): 1 – мережа типу RESNET; 2 – ансамбль з CNN мереж; 3 – запропонована модель.

Порівняємо різні моделі за часом навчання. Оскільки у запропонованій відсутні згорткові шари, її обчислювальна складність нижча проти RESNET та ансамблю CNN.

Час навчання моделей глибокого навчання залежно від кількості зображень навчальної вибірки для запропонованої моделі, RESNET та ансамблю CNN

Розмір навчальної вибірки	Час тренування запропонованої моделі, с	Час тренування RESNET, с	Час тренування ансамблю CNN, с
140	281	1316	2018
280	517	2874	4412
500	944	5307	8730
980	1708	9881	17021
1400	2394	15118	21218
1740	2579	18622	25743
1930	2891	20113	29805

Час тренування моделі RESNET та ансамблю CNN є співрозмірний, оскільки вони базуються на структурі зі згорткових шарів. Запропонований підхід із використанням виділення ознак мультипорогової бінаризації обчислювально простий, при цьому час навчання моделі в 5 разів коротший. Подібна закономірність зберігається і в оперативності класифікації. Час класифікації окремого багатоспектрального зображення становить 0,08 с. Для експериментів використовували персональний комп'ютер на базі процесора Intel Core I5 з частотою 3,6 ГГц та графічним процесором Nvidia GTX 1060.

ВИСНОВКИ

Методи глибокого навчання для розпізнавання зображень передбачають використання згорткових шарів, які є визначальними для формування вектора ознак. Але під час їх застосування для розпізнавання багатоспектральних зображень необхідно обробити велику кількість каналів у вигляді окремих матриць. Як відомо, згорткові шари нейронної мережі під час навчання найкраще описують усі особливості ознак зображень, але за великої навчальної вибірки та довгого тренування. Для роботи моделей глибокого навчання із багатоспектральними зображеннями необхідне апаратне забезпечення, що унеможливує їх застосування в реальному часі.

Запропонований метод виділення ознак зображень на основі багатопорогових методів бінаризації дає можливість значно зменшити обчислювальну складність виділення ознак багатоспектральних зображень. Також навчання моделі з виділенням ознак запропонованим методом значно швидше та потребує меншої навчальної вибірки порівняно із використанням згорткових шарів нейронних мереж. Швидкодія запропонованого підходу майже в 5 разів вища, ніж згорткових нейронних мереж, при цьому точність розпізнавання погіршується незначно.

1. Gong, C.; Xingxing, X.; Junwei, Han, L.; Gui-Song, X. Remote Sensing Image Scene Classification Meets Deep Learning: Challenges, Methods, Benchmarks, and Opportunities. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*. **2020**, 13, 3735–3756. <https://doi.org/10.1109/JSTARS.2020.3005403>
2. Mohammad, R. Real-time statistical image and video processing for remote sensing and surveillance applications. *Journal of Real-Time Image Processing*. **2021**, 18, 1435–1439. <https://doi.org/10.1007/s11554-021-01168-x>

3. Huang, X.; Banerjee, T.; Chen, K. Machine Learning based Video Processing for Real-time Near-Miss Detection. *International Conference VTITS*. **2020**, 169–179. <https://doi.org/10.5220/0009345400002550>
4. Mutlag, W.; Ali, S.; Mosad, Z.; Hussein, B. Feature Extraction Methods: A Review. *Journal of Physics*. **2020**, 1–10. <https://doi.org/10.1088/1742-6596/1591/1/012028>
5. Basavaiah, J.; Patil, C. Robust feature extraction and classification based automated human action recognition system for multiple datasets. *International Journal of Intelligent Engineering and Systems*. **2020**, 13(1), 13–24. <https://doi.org/10.22266/ijies2020.0229.02>
6. Xiao, Y.; Chen, J.; Wang, Y.; Cao, Z.; Zhou, J.; Bai, X. Action recognition for depth video using multi-view dynamic images. *Information Sciences*, **2019**, 480, 287–304. <https://doi.org/10.1016/j.ins.2018.12.050>
7. Zhuang, B.; Shen, C.; Tan, M.; Chen, P.; Liu, L.; Reid, I. Structured Binary Neural Networks for Image Recognition. *Computer Vision and Pattern Recognition*. 2019, 3.
8. Ding, R.; Liu, H.; Zhou, X. IE-Net: Information-Enhanced Binary Neural Networks for Accurate Classification. *Electronics*. 2022, 11. <https://doi.org/10.3390/electronics11060937>
9. Madaeni, F.; Chokmani, K.; Lhissou, R.; Homayouni, S.; Gauthier, Y. Convolutional neural network and long short-term memory models for ice-jam predictions. *The Cryosphere*. **2022**, 16, 1447–1468. <https://doi.org/10.5194/tc-16-1447-2022>
10. Rusyn, B.; Korniy, V.; Lutsyk, O.; Kosarevych, R. Deep Learning for Atmospheric Cloud Image Segmentation. *International Scientific and Practical Conference on Electronics and Information Technologies*. **2019**, 188–191. <https://doi.org/10.1109/ELIT.2019.8892285>
11. Feng, X.; He, L.; Cheng, Q.; Long, X.; Yuan, Y. Hyperspectral and Multispectral Remote Sensing Image Fusion Based on Endmember Spatial Information. *Remote Sensing*. **2020**, 12(6). <https://doi.org/10.3390/rs12061009>
12. Yuan, X.; Shi, J.; Gu, L.C. A review of deep learning methods for semantic segmentation of remote sensing imagery. *Expert Systems with Applications*. **2021**, 169. <https://doi.org/10.1016/j.eswa.2020.114417>
13. Farshi, T.; Demirci, R. Multilevel image thresholding with multimodal optimization. *Multimedia Tools and Applications*. **2021**, 80, 15273–15289. <https://doi.org/10.1007/s11042-020-10432-4>
14. Rusyn, B.; Lutsyk, O.; Kosarevych, R.; Maksymyuk, T.; Gazda, J. Features extraction from multi-spectral remote sensing images based on multi-threshold binarization. *Scientific Reports*. **2023**, 13. <https://doi.org/10.1038/s41598-023-46785-7>

Одержано 14.08.2023