

DOI: 10.34185/1991-7848.itmm.2025.01.095

**АВТОМАТИЧНА СЕМАНТИЧНА СЕГМЕНТАЦІЯ ЗОБРАЖЕНЬ SENTINEL-2:
ІНТЕГРАЦІЯ МЕТОДІВ КЛАСТЕРИЗАЦІЇ ТА ВЕЛИКИХ МОВНИХ МОДЕЛЕЙ
ДЛЯ ІНТЕРПРЕТАЦІЇ КЛАСТЕРІВ**

Гончаров О. Г.¹, Гнатушенко В. В.²

¹Український державний університет науки і технології, аспірант, Україна

²Український державний університет науки і технології, доктор технічних наук,
професор, Україна

Анотація. Семантична сегментація супутникових зображень, зокрема даних Sentinel-2, є важливим інструментом екологічного моніторингу та управління земними ресурсами. У роботі запропоновано метод автоматичної класифікації земного покриття без використання маркованих даних. Підхід ґрунтується на поєднанні кластеризації (K-Means, DBSCAN, автоенкодеру) та мультимодального маркування за допомогою великих мовно-візуальних моделей (GPT-4, Claude, Gemini 2.0 тощо). Інтерпретація кластерів здійснюється на основі візуалізацій, що автоматично аналізуються моделями. Методологія дозволяє досягти точності сегментації на рівні 85–90%, порівнянної з супервізованими підходами, та забезпечує інтерпретованість і масштабованість. Система також підтримує нормалізацію термінів і голосування кількох моделей для підвищення надійності. Результати валідуються з використанням карт ESA WorldCover. Підхід є перспективним для швидкого картографування в умовах обмежених ресурсів.

Ключові слова: семантична сегментація, кластеризація, Sentinel-2, земний покрив, мультимодальні моделі

Семантична сегментація супутникових зображень, зокрема даних Sentinel-2, є критично важливою для картографування земного покриття та екологічного моніторингу. Проте отримання анотацій на рівні пікселів для навчання моделей сегментації є трудомістким і дорогим процесом. У цій роботі пропонується новий підхід до автоматизації класифікації земного покриття шляхом інтеграції невизначених методів кластеризації з великими мультимодальними моделями (Anthropic Claude, OpenAI GPT, Google Gemini 2.0, QWEN 2.5-VL, Phi-4).

Мета дослідження — розробка методології для автоматичної сегментації та класифікації земного покриття на зображеннях Sentinel-2 без потреби у маркованих навчальних даних.

Методи дослідження: проведено огляд існуючих підходів до невизначеної сегментації зображень Sentinel-2 і кластеризації — від традиційних (K-Means, DBSCAN) до глибоких нейронних мереж (автоенкодера, U-Net). Запропонована методологія об'єднує кластеризацію супутникових зображень з автоматичним маркуванням кластерів за допомогою мультимодальних моделей. Для кожного кластера формується візуалізація, яка передається до моделей для інтерпретації класу земного покриву.

Алгоритм складається з трьох етапів: (1) підготовка даних Sentinel-2 L2A, включно з маскуванням хмар, вибором спектральних каналів (окрім атмосферних) та обчисленням вегетаційних індексів; (2) кластеризація пікселів методом K-Means, DBSCAN, автоенкодера з K-Means або попередньої сегментації з U-Net/SAM; (3) автоматичне маркування кластерів за допомогою мультимодальних моделей, які отримують зображення із виділеним кластером і визначають його категорію на основі просторового контексту та кольору.

Для узгодження відповідей моделей із заданими класами (вода, рослинність, міська забудова тощо) реалізовано механізм нормалізації термінів і голосування більшості при використанні кількох моделей. Для валідації використовуються карти ESA WorldCover. Аналізується вплив вибору кластеризаційної стратегії на якість сегментації.

Очікувані результати: комбінація автоенкодера з K-Means дозволить фіксувати тонкі спектральні відмінності, мультимодальні моделі — точно інтерпретувати класи, а ансамбль моделей — підвищити стабільність класифікації. Прогнозована точність сегментації основних класів — 85–90%, що порівнянно з результатами супервізованих методів.

Переваги: метод не вимагає навчальних даних, гнучкий щодо джерел зображень, не прив'язаний до регіону, забезпечує інтерпретованість кластерів і підвищену автоматизацію процесу класифікації. Методологія перспективна для швидкого картографування в умовах обмежених ресурсів або надзвичайних ситуацій.

Висновки. Запропоновано інноваційний підхід до семантичної сегментації зображень Sentinel-2 на основі кластеризації та мультимодального

маркування. У подальших дослідженнях планується: 1) провести експериментальну верифікацію; 2) оптимізувати візуалізацію кластерів; 3) додати зворотній зв'язок моделей; 4) перевірити можливість використання отриманих анотацій для навчання моделей типу U-Net.

ЛІТЕРАТУРА

- 1.Zhang, S., Wang, Q., Liu, J., & Xiong, H. (2024). ALPS: An Auto-Labeling and Pre-training Scheme for Remote Sensing Segmentation With Segment Anything Model. arXiv preprint. <https://doi.org/10.48550/arXiv.2406.10855>
- 2.Nagar, S., Farahbakhsh, E., Awange, J., & Chandra, R. (2024). Remote sensing framework for geological mapping via stacked autoencoders and clustering. *Advances in Space Research*, 74(10), 4502–4516. <https://doi.org/10.1016/j.asr.2024.09.013>
- 3.Osco, L. P., de Lemos, E. L., Gonçalves, W. N., Ramos, A. P. M., & Junior, J. M. (2023). The Potential of Visual ChatGPT For Remote Sensing. arXiv preprint. <https://doi.org/10.48550/arXiv.2304.13009>
- 4.Radford, A., Kim, J. W., Hallacy, C., Ramesh, A., Goh, G., Agarwal, S., Sastry, G., Askell, A., Mishkin, P., Clark, J., Krueger, G., & Sutskever, I. (2021). Learning Transferable Visual Models from Natural Language Supervision. arXiv preprint. <https://doi.org/10.48550/arXiv.2103.00020>
- 5.Duan, X., Huang, X., & Zhang, L. (2015). Unsupervised Global Urban Area Mapping via Automatic Labeling from ASTER and PALSAR Satellite Images. *Remote Sensing*, 7(2), 2171–2192. <https://doi.org/10.3390/rs70202171>

AUTOMATIC SEMANTIC SEGMENTATION OF SENTINEL-2 IMAGES: INTEGRATION OF CLUSTERING AND LARGE MULTIMODAL MODELS FOR CLUSTER INTERPRETATION

Honcharov O.H., Hnatushenko V.V.

Abstract. *Semantic segmentation of satellite imagery, particularly Sentinel-2 data, is crucial for environmental monitoring and land cover mapping. This paper presents an unsupervised method for land cover classification that eliminates the need for pixel-level annotations. The approach combines clustering techniques (K-Means, DBSCAN, autoencoders) with automated cluster labeling using large vision-language models (e.g., GPT-4, Claude, Gemini 2.0). Clusters are visualized and interpreted by these models based on spatial context and color. The methodology achieves segmentation accuracy of 85–90%, comparable to supervised methods, while ensuring interpretability and scalability. A majority voting mechanism and terminology normalization improve consistency across model outputs. Validation is performed using ESA WorldCover maps. The proposed approach is promising for rapid land cover mapping in resource-constrained or emergency situations.*

Keywords: semantic segmentation, clustering, Sentinel-2, land cover, multimodal models

REFERENCE

- 1.Zhang, S., Wang, Q., Liu, J., & Xiong, H. (2024). ALPS: An Auto-Labeling and Pre-training Scheme for Remote Sensing Segmentation With Segment Anything Model. arXiv preprint. <https://doi.org/10.48550/arXiv.2406.10855>
- 2.Nagar, S., Farahbakhsh, E., Awange, J., & Chandra, R. (2024). Remote sensing framework for geological mapping via stacked autoencoders and clustering. *Advances in Space Research*, 74(10), 4502–4516. <https://doi.org/10.1016/j.asr.2024.09.013>
- 3.Osco, L. P., de Lemos, E. L., Gonçalves, W. N., Ramos, A. P. M., & Junior, J. M. (2023). The Potential of Visual ChatGPT For Remote Sensing. arXiv preprint. <https://doi.org/10.48550/arXiv.2304.13009>
- 4.Radford, A., Kim, J. W., Hallacy, C., Ramesh, A., Goh, G., Agarwal, S., Sastry, G., Askell, A., Mishkin, P., Clark, J., Krueger, G., & Sutskever, I. (2021). Learning Transferable Visual Models from Natural Language Supervision. arXiv preprint. <https://doi.org/10.48550/arXiv.2103.00020>
- 5.Duan, X., Huang, X., & Zhang, L. (2015). Unsupervised Global Urban Area Mapping via Automatic Labeling from ASTER and PALSAR Satellite Images. *Remote Sensing*, 7(2), 2171–2192. <https://doi.org/10.3390/rs70202171>