A Scientometric and Literature Analysis of Deep Learning-Based Semantic Segmentation in Remote Sensing (2015–2025)

Roshni Rajendran[1,2], Nagaraj P[3]
[1]Research Scholar, Department of Computer Science and Engineering, Kalasalingam Academy of Research and Education, Krishnankoil, Virudhunagar, India
[2]Department of Computer Applications, Marian College Kuttikkanam Autonomous, Kerala, India
[3]Associate Professor, Department of Computer Science and Engineering,
School of Computing, SRM Institute of Science and Technology, Tiruchirapalli, SRM Nagar, Trichy, Tamil Nadu, India
E-mail: roshnirajendran24@gmail.com, nagaraj.p@klu.ac.in

Review paper

Keywords: semantic segmentation, remote sensing, deep learning, land cover mapping, VOSviewer, biblioshiny

Received: April 28, 2025

Semantic segmentation of remote sensing images has advanced rapidly, enabling applications in land cover mapping, disaster response, and urban monitoring. This study presents a hybrid scientometric and literature-based analysis of 733 Scopus-indexed publications (2015–2025). Results show a 29.24% annual growth rate, with China, the United States, and Germany as leading contributors and Wuhan University as the most prolific institution. Research output peaked in 2023, driven by transformers and hybrid architectures such as SegFormer, Mask2Former, and Swin Transformer, which outperform CNN baselines. Citation and keyword analyses reveal two core directions: applied geospatial tasks (land cover, urban analysis, disaster management) and computational advances (CNNs, transformers, domain adaptation). While foundational works remain highly cited, emerging models emphasize efficiency, multimodal fusion, and generalization. Persistent challenges include dataset imbalance, cross-domain adaptation, and lack of standardized benchmarks. By combining bibliometric mapping with methodological synthesis, this study consolidates research trends and highlights future directions in multimodal learning, explainable AI, and robust, scalable segmentation frameworks.

Povzetek: Predstavljena je scientometrična in vsebinska analiza raziskav globokega učenja za semantično segmentacijo v daljinskem zaznavanju v obdobju 2015–2025. Analiza bibliografskih podatkov razkriva raziskovalne trende, najpogosteje uporabljene arhitekture, podatkovne vire ter odprte izzive in prihodnje smeri razvoja področja.

1 Introduction

Semantic segmentation of remote sensing imagery has become a cornerstone in geospatial analysis, enabling pixel-level classification of land cover, infrastructure, and environmental features. Fine-grained understanding of remote sensing scenes is crucial for applications such as urban development, crop monitoring, disaster management, and environmental protection. However, multiple factors—including varying spatial resolutions and spectral bands—make segmentation a challenging task [1].

The advent of deep learning, particularly convolutional neural networks (CNNs), has transformed this field. Unlike traditional models that relied on manual feature engineering, CNNs automatically learn patterns, leading to improved accuracy. Fully Convolutional Networks (FCNs), U-Net, and DeepLab have shown strong performance in capturing spatial hierarchies and contextual information from satellite images [2], [3]. Their effectiveness arises from layered feature extraction and multi-scale analysis, enhanced further by attention mechanisms and skip connections that improve sensitivity to fine boundaries in high-resolution imagery. More recently, transformer-based models, originally developed for natural language processing, have been adapted to image segmentation, offering improved modeling of long-range dependencies.

Models such as SegFormer, SMBCNet, and SpectralGPT leverage self-attention mechanisms to capture global context and outperform many CNN-based architectures in both accuracy and generalizability [4]–[6]. For example, SegFormer achieved mean IoU scores above 92% on the LoveDA dataset, exceeding U-Net baselines by 7–10%. Multimodal approaches, such as MetaSegNet, further extend capabilities by fusing metadata (e.g., region or climate descriptors) with visual features, thereby improving interpretability and domain adaptation [7]. Despite these advances, the field remains fragmented, with rapidly increasing publications, diverse datasets, and varying model designs, creating challenges for researchers—particularly newcomers—in identifying dominant approaches, unresolved issues, and future directions [8]–[15].

Systematic reviews provide thematic overviews but often lack data-driven insights, while bibliometric studies emphasize publication trends without linking them to methodological progress. A hybrid approach is therefore required to bridge this gap.

1.1 Research questions

This study is guided by the following research questions:

Research Question 1: What are the publication, citation, and collaboration trends in deep learning-based semantic segmentation of remote sensing between 2015–2025?

Research Question 2: Which models, datasets, and methodological innovations dominate the literature, and how do their reported performances compare?

Research Question 3: How have emerging approaches (e.g., transformers, multimodal learning) surpassed traditional CNNs, and what challenges remain unresolved?

1.2 Objectives

In this study, we conduct a hybrid scientometric and literature-based analysis of deep learning applications in semantic segmentation for remote sensing. We analyze a dataset of 733 documents from the Scopus database, covering the period 2015–2025. Scientometric tools such as VOSviewer and Biblioshiny are used to explore publication trends, collaboration networks, and keyword evolution. Simultaneously, a structured literature review synthesizes recent deep learning architectures, datasets, training strategies, and real-world applications. We highlight both emerging trends—such as lightweight and vision-language models—and persistent challenges, including data imbalance, generalization, and interpretability.

This dual analysis provides a comprehensive understanding of the intellectual and technological evolution in this area, offering practical insights and future directions for researchers, developers, and policymakers working in the remote sensing and geospatial AI domains.

2 Review of literature

2.1 Convolutional neural network (CNN)-based approaches

CNNs have been foundational in semantic segmentation tasks within remote sensing. Li et al. [16] introduced the A2-FPN model, which incorporates adaptive attention mechanisms for high-resolution building segmentation, achieving strong results on urban datasets. Building on this, Li et al. [17] proposed ABCNet, which leverages bilateral contextual information to sharpen boundaries, reporting 84.2% mIoU on the ISPRS Vaihingen dataset. Bo et al. [18] focused on real-time burned area detection with BASNet, optimized for speed while maintaining 85% segmentation accuracy, making it valuable in time-sensitive disaster response.

2.2 Transformer-based models

Transformer architectures have demonstrated state-of-the-art performance. Song et al. [19] applied Vision Transformers for building footprint extraction, achieving 90.1% mIoU on the Inria dataset. Gibril et al. [20] developed Mask2Former based on Swin Transformers, reporting 88–93% mIoU across urban benchmarks. Cui et al. [21] enhanced Swin Transformer variants for post-earthquake scenarios, yielding improvements of 4–6% IoU compared to CNN baselines.

2.3 Lightweight and real-time models

Real-time processing is critical for disaster monitoring and autonomous systems. Broni-Bediako et al. [23] surveyed lightweight approaches on OpenEarthMap, highlighting accuracy-speed trade-offs. Zhao et al. [24] introduced SEG-Road, a CNN–Transformer hybrid for road extraction, achieving 87.6% mIoU while maintaining low inference latency, suitable for resource-constrained environments.

2.4 Multimodal and vision-language integration

To enhance generalization, vision-language fusion models integrate auxiliary metadata. Wang et al. [25] proposed MetaSegNet, which combines textual descriptors with visual features, achieving 89% mIoU on LoveDA and OpenEarthMap. Ajibola and Cabral [26] conducted a meta-analysis emphasizing multimodal learning as an emerging trend in robust land cover segmentation.

2.5 Semi-supervised and few-shot learning

Due to annotation scarcity, semi-supervised and few-shot approaches are gaining traction. Zhang et al. [27] introduced a pseudo-labeling framework with consistency regularization, improving segmentation by 6–8% IoU under low-label settings. Chen et al. [28] developed a few-shot prototype network, demonstrating effective generalization from as few as 10 labeled samples per class on ISPRS Potsdam, achieving 82% mIoU.

2.6 Domain adaptation and generalization

Liu et al. [29] applied domain adaptation to align feature distributions, achieving a 5% IoU improvement across geographic domains. Wang et al. [30] proposed adversarial cross-domain training, showing resilience to sensor shifts with up to 7% fewer performance drops compared to non-adaptive models.

2.7 Application-specific models

Domain-focused models provide tailored solutions. Tao et al. [31] applied attention-guided CNNs for road networks, improving connectivity detection by 8%. Zhu et al. [32] segmented urban green spaces with spectral-spatial fusion, reporting 86% accuracy. Zhang et al. [33] applied encoder–decoder networks for agricultural field segmentation, achieving 88% IoU on crop mapping datasets.

Recent advances show a clear shift toward multimodal and foundation-model architectures in remote sensing semantic segmentation. MetaSegNet [34] employs a metadata-collaborative vision–language framework, using geographic text prompts and cross-modal attention to improve interpretability and zero-shot generalization, with strong results on OpenEarthMap (70.4% mIoU), Potsdam (93.3% F1), and LoveDA (52.0% mIoU). SpectralGPT [35], a 3D generative pretrained transformer with over 600M parameters, is among the first foundation models for spectral data, capturing spatial–spectral dependencies from one million images and achieving notable gains across classification, segmentation, and change detection. Together, these architectures highlight the move from unimodal CNN/transformer models toward scalable, multimodal pretrained approaches for adaptability and generalization in geoscience applications.

2.8 Real-world applications

Applications span urban planning, agriculture, disaster management, and climate studies. Automated pixel-level classification supports policy-making, precision farming, and climate change modeling. In forestry, segmentation aids biodiversity monitoring and carbon stock estimation.

Table1: Summary of key models, datasets, performance, and contributions in semantic segmentation of remote sensing

Model / Study	Architecture Type	Dataset(s) Used	Application Domain	Performance Metrics	Key Contribution
A2-FPN [16]	CNN + Attention	ISPRS Vaihingen	Urban building segmentation	~83–85% mIoU	Adaptive attention for VHR imagery
ABCNet [17]	CNN (bilateral)	ISPRS Vaihingen	Building boundaries	84.2% mIoU	Bilateral context for sharper edges
BASNet [18]	Lightweight CNN	Burned-area imagery	Disaster monitoring	85% accuracy	Real-time segmentation for disaster response
SegFormer [19]	Transformer	LoveDA, Inria	Land cover, urban areas	92.3% mIoU	Global context modeling, outperforming CNNs
Mask2Former [20]	Swin Transformer	Large-scale urban	Building segmentation	88–93% mIoU	Generalizable transformer baseline
SwinTransformer [21]	Transformer	Post-earthquake urban	Damage mapping	+4–6% IoU vs CNNs	Robust in dense urban imagery
SEG-Road [24]	Hybrid CNN–Trans	VHR road datasets	Road extraction	87.6% mIoU	Accuracy–efficiency balance
MetaSegNet [25]	Vision–Language	LoveDA, OpenEarthMap	Multimodal segmentation	89% mIoU	Metadata integration for better generalization
Semi-supervised [27]	CNN + SSL	ISPRS, LoveDA	Low-label segmentation	+6–8% IoU	Pseudo-labeling + consistency regularization
Few-shot [28]	Prototype Net	ISPRS Potsdam	Few-shot segmentation	82% mIoU	Strong generalization from few samples
Domain Adaptation [29]	CNN + DA	Cross-region datasets	Domain transfer	+5% IoU vs base-line	Aligns feature distributions across domains
Cross-Domain Adv. [30]	CNN + Adversarial	Multi-sensor datasets	Robust domain segmentation	-7% drop (vs -14%)	Adversarial robustness against sensor shift
Road-Attention [31]	CNN + Attention	VHR imagery	Road extraction	+8% connectivity	Spatial attention for road networks
Urban Green [32]	Spectral–Spatial	High-res urban	Vegetation monitoring	86% accuracy	Fusion of spectral and spatial features
Agriculture [33]	Encoder–Decoder	Crop field imagery	Precision agriculture	88% IoU	Crop segmentation for yield mapping

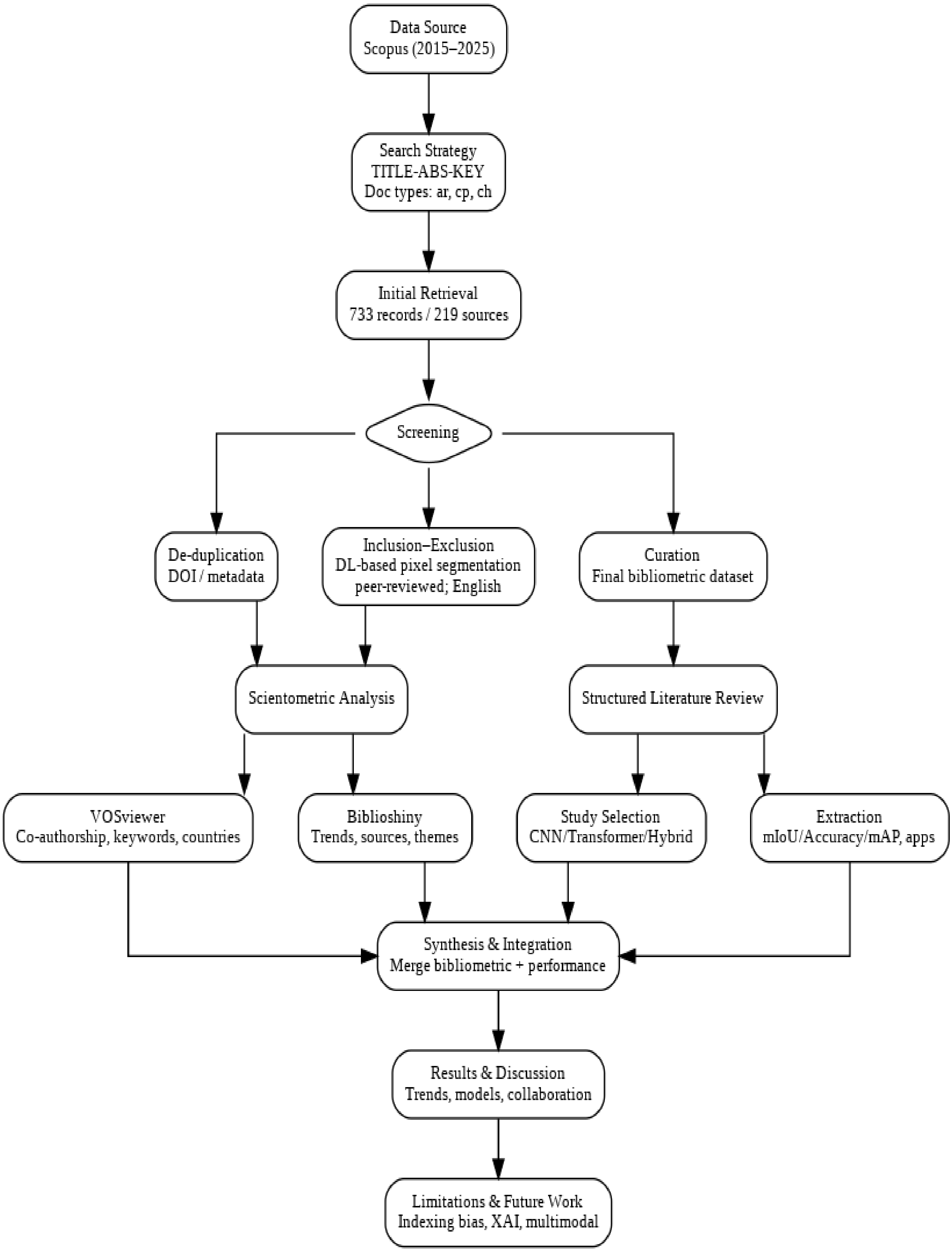

Figure 1: Workflow of the study

3 Methodology

This work adopts a scientometric approach supported by a systematic literature review to provide a comprehensive understanding of deep learning using semantic segmentation of remote sensing from 2015 to 2025. The overall methodological workflow is illustrated in figure 3.1 showing the integration of bibliometric mapping and literature synthesis. The process consisted of three main stages: (i) bibliometric data acquisition, (ii) scientometric analysis using visualization tools, and (iii) a structured synthesis of methodological advances and application domains.

3.1 Data source and search strategy

The bibliometric dataset was obtained from Scopus, which was selected due to its wide coverage of peer-reviewed journals, conference proceedings, and book chapters across computer science, geoinformatics, and remote sensing. The study period was set from January 2015 to March 2025, reflecting the decade when deep learning became the dominant paradigm in semantic segmentation research.

The key attributes and descriptive statistics of the final bibliometric dataset are summarized in Table 2 below.

Table 2: Summary of bibliometric dataset characteristics and descriptive statistics

Description	Results
MAIN INFORMATION ABOUT DATA	
Timespan	2015:2025
Sources (Journals, Books, etc)	219
Documents	733
Annual Growth Rate %	29.24
Document Average Age	2.46
Average citations per doc	13.85
References	29099
DOCUMENT CONTENTS	
Keywords Plus (ID)	3684
Author's Keywords (DE)	1635
AUTHORS	
Authors	2150
Authors of single-authored docs	11
AUTHORS COLLABORATION	
Single-authored docs	11
Co-Authors per Doc	4.91
International co-authorships %	20.6
DOCUMENT TYPES	
article	570
book chapter	5
conference paper	158

The following search string was used in the TITLE-ABS-KEY field: ("Semantic Segmentation") AND ("Remote Sensing" OR "Land Cover" OR "Aerial") AND ("Deep Learning" OR "Machine Learning"). This query restricted results to journal articles (ar), conference papers (cp), and book chapters (ch), while excluding grey literature. The time span was set between January 2015 and March 2025. The initial search retrieved 733 documents from 219 sources.

Screening and refinement process:

- Duplicates were removed using DOIs and metadata checks.
- Only English-language publications with complete bibliographic information (titles, abstracts, author keywords, references, and citation counts) were retained.
- Non-peer-reviewed or incomplete metadata records were excluded.
- The final curated dataset included articles that explicitly addressed deep learning-based semantic segmentation of remote sensing imagery.

3.2 Analytical tools and parameters

Bibliometric data were analyzed using two established tools: **VOSviewer (v1.6.19)** and **Biblioshiny**, the web interface of the Bibliometrix R package. VOSviewer was employed for network visualization tasks such as co-authorship mapping, keyword co-occurrence networks, and country collaboration charts. Biblioshiny provided performance metrics including annual publication trends, author productivity, source impact, and thematic evolution.

Key analysis parameters included keyword co-occurrence for identifying frequently used terms and thematic clusters, co-authorship networks for mapping collaborations among researchers and institutions, and co-citation/bibliographic coupling for uncovering intellectual linkages and shared references. Collectively, these methods enabled a comprehensive assessment of research dynamics, highlighting publication patterns, collaboration networks, and the thematic development of semantic segmentation in remote sensing over the past decade.

3.3 Integration with literature review

In parallel with scientometric mapping, a structured literature review was conducted on selected studies that reported quantitative performance metrics (IoU, accuracy, mAP) on benchmark datasets such as ISPRS Vaihingen/Potsdam, LoveDA, and Inria. Studies included transformer-based, CNN-based, lightweight, and multimodal models. This dual approach ensures that bibliometric insights are complemented with technical performance comparisons and real-world application contexts.

4 Bibliometric analysis and results

4.1 Annual scientific production

The annual scientific production chart highlights research trends in Remote Sensing, Machine Learning, Deep Learning, and Semantic Segmentation over the past decade. Between 2015 and 2018, output was minimal, reflecting limited adoption. From 2019, publications grew steadily, with a sharp surge from 2020 to 2023 driven by advances in deep learning, the availability of high-resolution satellite imagery, and increased funding for AI-based applications. The peak in 2023 marks the field's most active year, influenced by breakthroughs in transformer models, self-supervised learning, and improved computational resources.

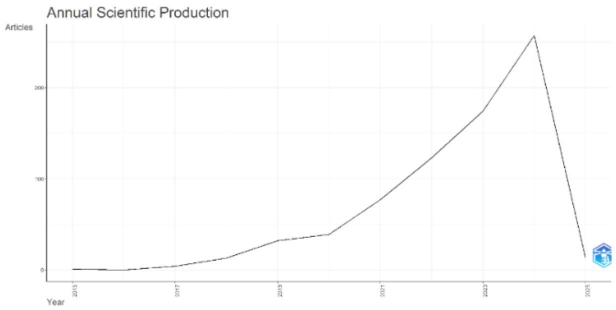

Figure 2: Annual scientific production (2015–2025) based on biblioshiny analysis.

The apparent decline in 2024 likely reflects incomplete indexing in Scopus rather than reduced research activity, as many papers are still in review or awaiting publication. Based on previous trajectories, the final count for 2024 is expected to rise.

Overall, the trend underscores the increasing integration of artificial intelligence with remote sensing. As the field matures, future work will emphasize efficient deep learning models, multimodal learning, and real-time applications in environmental monitoring and disaster management.

4.2 Average citations per year

The chart depicts the average number of citations per year for research articles in Remote Sensing, Machine Learning, Deep Learning, and Semantic Segmentation. In the early years (2015–2017), citation counts were low as the field was still emerging.

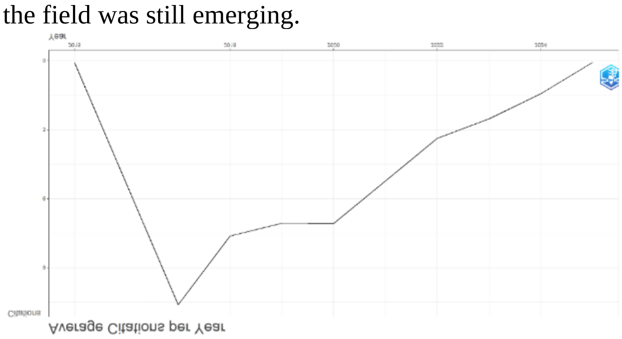

Figure 3: Line graph showing the average number of citations per year for publications.

A sharp peak in 2018 reflects the influence of foundational works that introduced novel deep learning architectures for remote sensing and shaped subsequent research directions. From 2019 onward, a gradual decline is observed, becoming more pronounced after 2020. This trend does not indicate declining quality but is explained by the citation lag effect—older works have had more time to accumulate citations—together with the rapid expansion of publications, which dispersed citations across a broader set of papers rather than concentrating on a few seminal works. The decline into 2024 is likely due to incomplete data, as recent papers have not yet had sufficient time to be widely cited. Overall, the results highlight the pivotal role of studies published between 2017 and 2019, while also emphasizing the need for future research to deliver breakthrough contributions capable of sustaining high citation impact in this fast-growing field.

4.3 Three field plot

The three-field plot visualizes the relationships between research keywords, key authors, and major publication sources in Remote Sensing, Deep Learning, and Semantic Segmentation. On the left, the most frequent keywords include semantic segmentation, deep learning, remote sensing images, and remote sensing, along with terms such as building extraction, image segmentation, object detection, satellite imagery, convolutional neural networks, and transfer learning, reflecting diverse applications of deep learning in remote sensing. The central section highlights prolific authors such as Wang I, Zhang X, Liu J, Wang Y, and Zhang Y, who have published extensively on computer-aided remote sensing image processing, particularly in segmentation and object detection. Their strong association with key terms demonstrates their impact on methodological development in the field. The right section presents leading journals, with IEEE Transactions on Geoscience and Remote Sensing and IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing as the top venues, followed by Remote Sensing, ISPRS Journal of Photogrammetry and Remote Sensing, and International Journal of Remote Sensing.

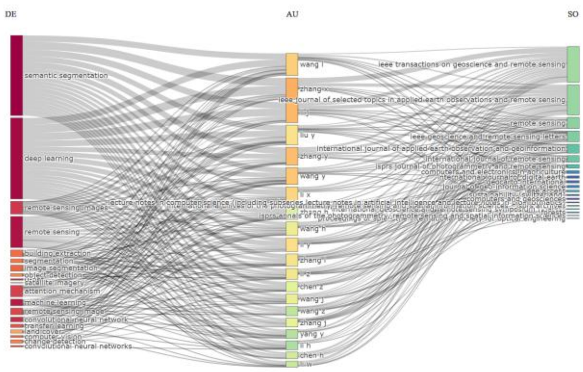

Figure 4: Three-field plot showing the relationship between author keywords (DE), top contributing authors (AU), and source journals (SO).

Note: Colors represent different clusters; clusters are also distinguished by node labels and edge structures for clarity in grayscale.

The dominance of IEEE and ISPRS outlets reflects their established role in publishing high-quality research in geospatial science and computational methodologies. This study provides one of the first comprehensive analyses of recent developments in this domain, emphasizing the contributions of a core group of highly productive authors and influential journals. The interconnection between deep learning methodologies and remote sensing applications underscores the growing role of artificial intelligence in geospatial analysis. Future research is expected to expand further with advances in transformer models, self-supervised learning, and real-time satellite data processing.

4.4 Most relevant authors

The Most Relevant Authors chart highlights the most active researchers in Remote Sensing, Deep Learning, and Semantic Segmentation. Zhang X is the most prolific contributor with 31 papers, followed closely by Li J (30). Both have played a central role in advancing deep learning methods for remote sensing. Other notable authors include Zhang Y (25 papers), Wang Y (23), and Wang L (21), indicating that research output is concentrated among a small group of highly productive scholars.

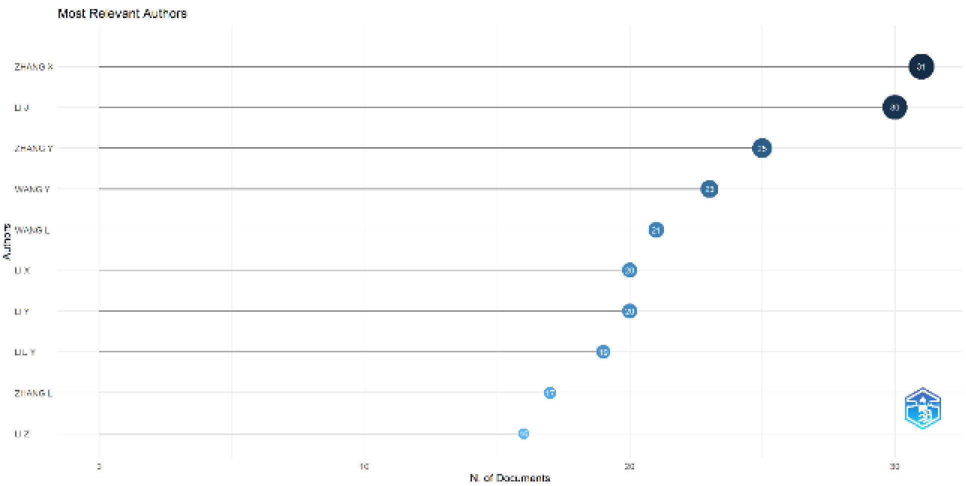

Figure 5: Most relevant authors based on the number of publications from 2015 to 2025.

The prevalence of surnames such as Zhang, Wang, and Li suggests that much of this research originates from Chinese institutions, consistent with China's growing investment in remote sensing and AI. Authors including Liu Y (19), Zhang L (17), and Li Z (16) also demonstrate significant contributions, focusing on applications such as land cover classification, object detection, and change detection.

Overall, the prominence of a limited number of highly active researchers points to a functional core cluster driving progress in this field. Further analyses of co-authorship networks and citation impact would provide deeper insights into their collaborative structures and scientific influence.

4.5 Most relevant affiliations

The Most Relevant Affiliations chart highlights the institutions most active in Remote Sensing, Deep Learning, and Semantic Segmentation. Wuhan University leads with 119 publications, underscoring its global reputation in geospatial and remote sensing research. The China University of Geosciences follows with 85 articles, while Xidian University (68), Beihang University (54), and Nanjing University of Information Science and Technology (53) are also prominent contributors. Additional institutions include China Agricultural University (41), the Aerospace Information Research Institute (32), Zhejiang University (30), and Central South University (27).

The dominance of Chinese universities reflects the country's strategic investment in satellite imaging, environmental monitoring, and AI-driven geospatial analytics, supported by significant government initiatives and funding. Overall, the concentration of output among a limited number of institutions indicates strong institutional leadership in advancing deep learning for remote sensing. Further study of inter-university collaborations and funding networks would provide deeper insights into the global research landscape.

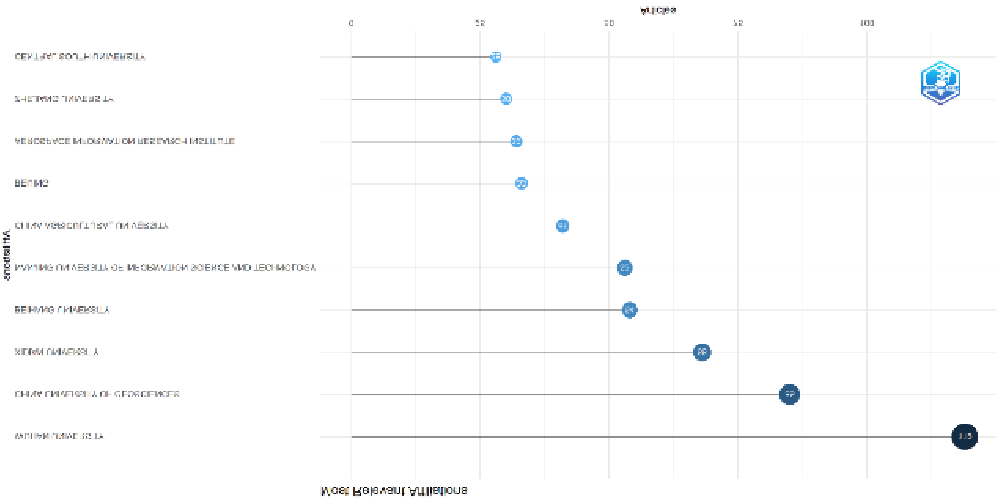

Figure 6: Most relevant institutional affiliations based on the number of articles published from 2015 to 2025.

4.6 Trend topics

The analysis of trend topics highlights the shifting intellectual focus in Remote Sensing, Deep Learning, and Semantic Segmentation. Between **2016 and 2020**, research was dominated by traditional machine learning approaches such as decision trees, random processes, and early neural networks, largely applied to land cover mapping and image segmentation. From **2020 onwards**, a methodological transition occurred, with *deep learning*, *remote sensing*, *semantic segmentation*, and *convolution* emerging as central themes, reflecting the widespread adoption of convolutional neural networks and the parallel growth of improved sensing technologies. In the most recent period, **2022–2024**, the prominence of terms such as *semantic segmentation*, *deep learning*, and *semantic standardization* signals the field's consolidation around scalable and standardized deep learning frameworks. The frequent occurrence of *China* as a trending keyword further illustrates the strong role of Chinese institutions in driving research output. Overall, the trend suggests a progression from experimental applications of machine learning toward the refinement and standardization of advanced deep learning models, with future directions likely to integrate transformers and self-supervised learning to enhance generalization and interpretability.

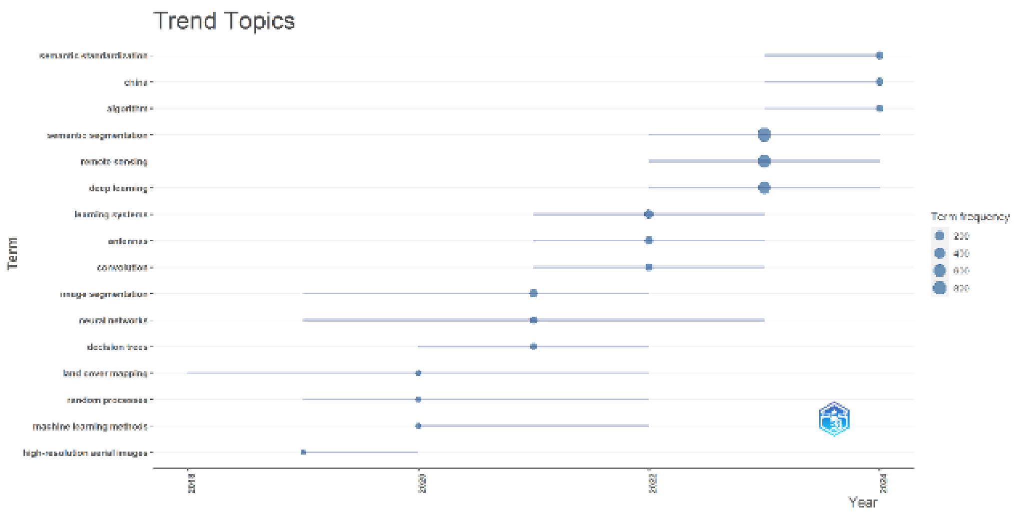

Figure 7: Trend topics visualization showing the temporal evolution of key research terms.

4.7 Co-occurrence network

The keyword co-occurrence network highlights major terms related to Remote Sensing, Machine Learning, Deep Learning, and Semantic Segmentation. Two primary clusters emerge. The red cluster centers on semantic segmentation, remote sensing, machine learning, and deep learning, reflecting the main research focus. Associated terms such as image classification, feature extraction, accuracy assessment, adversarial machine learning, and synthetic data indicate that advanced AI-powered approaches dominate segmentation tasks in remote sensing.

The blue cluster emphasizes computational methodologies, including convolutional neural networks, deep neural networks, supervised learning, and network architectures, often linked to object detection, mapping, and UAV-based applications. Strong interconnections between the clusters demonstrate the interdisciplinary nature of the field. The link between segmentation and deep learning models underscores the central role of advanced architectures in improving accuracy, while terms like pixels and image standardization highlight the importance of preprocessing for model performance.

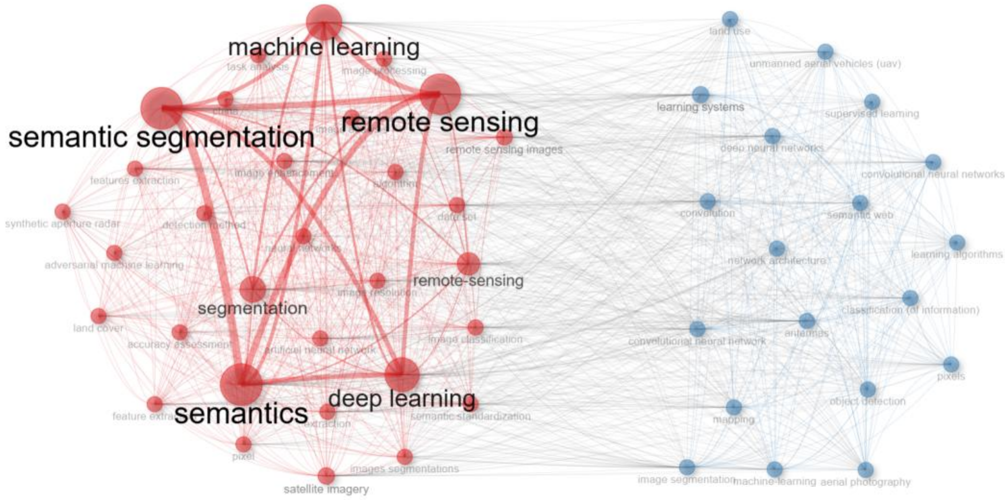

Figure 8: Keyword co-occurrence network visualizing the most frequent and interconnected terms in semantic segmentation of remote sensing images using deep learning

Note: Colors represent different clusters; clusters are also distinguished by node labels and edge structures for clarity in grayscale.

Overall, the network shows that semantic segmentation remains the dominant theme in remote sensing research, with balanced attention to developing new AI methodologies and applying them to real-world geospatial problems. Factorial analysis dendrogram.

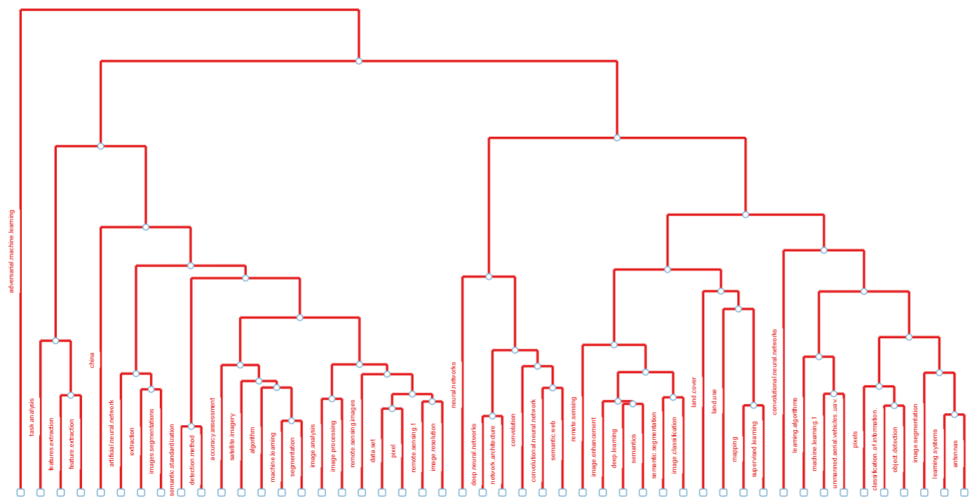

Figure 9: Dendrogram generated from factorial analysis, illustrating the hierarchical clustering of key terms used in semantic segmentation research

The factorial analysis dendrogram clusters research topics in Remote Sensing, Machine Learning, Deep Learning, and Semantic Segmentation into two broad categories: applied studies (e.g., land cover classification, geospatial analysis) and computational methods (e.g., deep learning frameworks, machine learning algorithms).

Lower-level clusters reveal themes such as image segmentation, feature extraction, and object detection, alongside computational advances like CNNs, transfer learning, and data augmentation. Emerging sub-clusters on synthetic data and adversarial learning emphasize efforts to improve robustness, generalization, and domain adaptation.

The dendrogram also highlights growing attention to explainability, with clusters on semantic standardization and interpretability. Overall, it reflects the dual focus on applied geospatial applications and algorithmic innovation, underscoring future priorities of reliable, generalizable, and interpretable AI models for remote sensing.

4.8 Most global cited documents

The bar chart presents the ten most cited works in deep learning for remote sensing. Xu Y. (2018) leads with 406 citations, followed by Wurm M. (2019) and Kaiser P. (2017) with 284 and 272, establishing foundational methods for geospatial analysis. Subsequent highly cited studies include Hong D. (2024), Lu Y. (2020), and Prakash N. (2020), addressing transformers, land cover classification, and change detection, while works by Mohammadimanesh F., Schmitt M., and Rahnemoonfar M. expand into urban object detection, synthetic data, and semi-supervised learning.

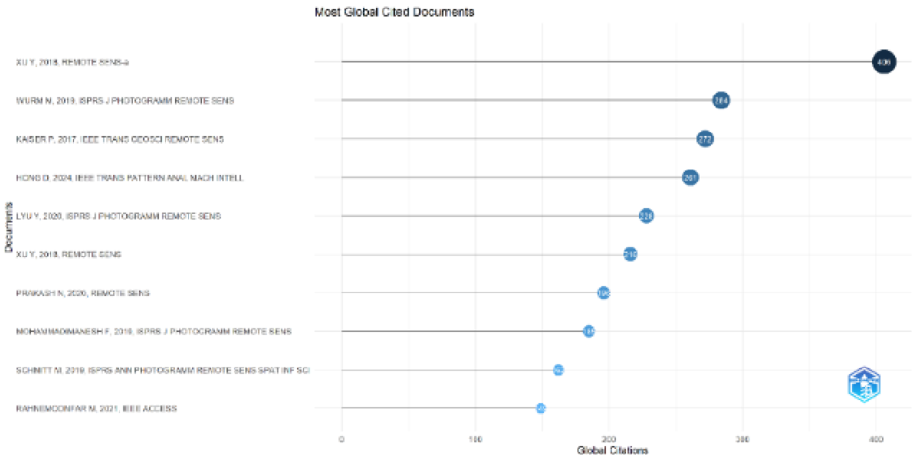

Figure 10: Most globally cited documents ranked by total citation count from 2015 to 2025.

The spread of years (2017–2024) shows that early milestones remain influential while newer innovations are rapidly gaining recognition. The prominence of IEEE and ISPRS journals further reflects the multidisciplinary nature of this field, bridging computer vision and geospatial science.

5 Discussions

The discussion is organized around the research questions presented in the Introduction, combining bibliometric evidence with insights from the structured literature review to highlight trends, dominant models, and remaining challenges.

RQ1: What are the publication, citation, and collaboration trends in deep learning-based semantic segmentation of remote sensing between 2015–2025?

The analysis of 733 Scopus publications shows rapid growth, with an annual increase of about 29% between 2017 and 2023, reflecting the uptake of deep learning in geospatial AI. Citations are concentrated in *Remote Sensing*, *ISPRS Journal of Photogrammetry and Remote Sensing*, and *IEEE JSTARS*. International collaboration has expanded, with China, the United States, and Germany as leading contributors, and institutions such as Wuhan University, the Chinese Academy of Sciences, and ETH Zurich emerging as major hubs.

RQ2: Which models, datasets, and methodological innovations dominate the literature, and how do their reported performances compare?

CNN-based architectures such as U-Net, SegNet, and DeepLab initially dominated, achieving 75–85% mIoU on benchmarks like ISPRS Vaihingen and Inria. Recent studies emphasize transformer-based models (e.g., SegFormer, Mask2Former, Swin Transformer), which outperform CNNs by 5–10% mIoU on LoveDA and other datasets. Lightweight approaches such as BASNet and SEG-Road enable real-time use with modest accuracy trade-offs, while multimodal designs like MetaSegNet show improved generalization through metadata integration.

RQ3: How have emerging approaches (e.g., transformers, multimodal learning) surpassed traditional CNNs, and what challenges remain unresolved?

Transformers and hybrid models surpass CNNs by modeling long-range spatial dependencies and global context, while multimodal and vision-language methods further improve adaptability across heterogeneous datasets. However, challenges remain, including the absence of standardized benchmarks, limited cross-domain generalization, dataset imbalance, high computational costs, and the lack of explainability frameworks.

Overall, recent years have seen rapid growth in deep learning for geospatial science, supported by high-resolution imagery and the need for automated interpretation. Prolific authors (e.g., Zhang X., Li J., Zhang Y.), leading institutions (Wuhan University, China University of Geosciences), and strong international collaborations, especially between China, the US, Germany, and India, have driven this momentum. Keyword analyses confirm the shift from CNNs to transformer and hybrid models, alongside strategies such as transfer learning, domain adaptation, and few-shot learning. Applications span urban planning, agriculture, disaster response, and environmental monitoring.

Nevertheless, issues of generalization across sensors, computational efficiency, and benchmarking remain unresolved. Scientometric limitations, including keyword inconsistency, citation lag, database bias, and metadata dependence, also constrain interpretation. Future work should prioritize lightweight edge-ready models, multimodal fusion with large language models, explainable AI, and robust cross-domain frameworks. By combining bibliometric mapping with a structured literature review, this study offers both retrospective analysis and forward-looking guidance for semantic segmentation in remote sensing.

6 Conclusion

This bibliometric and systematic review provides a structured perspective on the evolution, growth, and global interest in semantic segmentation of remote sensing. Once a niche topic in computer vision, it has become central to applications in environmental, agricultural, and urban domains, driven by deep learning advances and high-resolution geospatial data. The surge in research from 2017 to 2023, along with the emergence of influential authors, institutions, and collaborations, underscores the field's expanding significance. Keyword and thematic analyses further reveal the shift from traditional models to adaptive, efficient, and application-driven deep learning frameworks.

The novelty of this study lies in its hybrid approach, integrating scientometric mapping with structured literature review to bridge quantitative trends and qualitative insights. Future work should focus on multimodal learning and large language model (LLM)-based frameworks to enhance interpretability, the development of explainable AI (XAI) for trustworthy decision support, and strategies to address dataset imbalance and cross-domain generalization through standardized benchmarks. In sum, this research highlights the strategic role of semantic segmentation in a data-driven world and offers a foundation for advancing both methodological innovation and practical deployment in geospatial analysis.

Future research should expand beyond the limitations of Scopus by incorporating diverse and multilingual sources to capture a more inclusive research landscape. Another promising direction is the integration of explainable AI into segmentation frameworks, enabling greater interpretability and trust in critical applications. Advances in multimodal learning—particularly the fusion of LiDAR, SAR, hyperspectral, and environmental data—also hold potential to improve model robustness. Moreover, addressing challenges of cross-domain adaptation and dataset bias through standardized benchmarks and domain-invariant learning will be essential. Finally, coupling bibliometric trends with systematic performance validation can guide the field toward solutions that are both methodologically innovative and practically relevant.

References

[1] J. Lv, Q. Shen, and M. Lv, "Deep learning-based semantic segmentation of remote sensing images: a review," *Frontiers in Ecology and Evolution*, vol. 11, 2023.

[2] X. Li and J. Li, "MFCA-Net: A deep learning method for semantic segmentation of remote sensing images," *Scientific Reports*, vol. 14, no. 5745, 2024.

[3] X. Wang, H. Zhang, Y. Chen, and Y. Liu, "A deep learning method for optimizing semantic segmentation accuracy of remote sensing images based on improved UNet," *Scientific Reports*, vol. 13, no. 7600, 2023.

[4] J. Song, A.-X. Zhu, and Y. Zhu, "Transformer-Based Semantic Segmentation for Extraction of Building Footprints from Very-High-Resolution Images," *Sensors*, vol. 23, no. 11, p. 5166, 2023.

[5] Y. Liu, "Remote sensing image scene classification based on convolutional neural networks," *Informatica*, vol. 49, no. 9, 2025.

[6] J. Feng, Z. Wang, Y. Zhang, and H. Li, "SMBCNet: A Transformer-Based Approach for Change Detection in Remote Sensing Images through Semantic Segmentation," *Remote Sensing*, vol. 15, no. 14, p. 3566, 2023.

[7] X. Yuan, J. Shi, and L. Gu, "A review of deep learning methods for semantic segmentation of remote sensing imagery," *Expert Systems with Applications*, vol. 169, p. 114417, 2021. doi: 10.1016/j.eswa.2020.114417.

[8] K. R. Hasan, A. B. Tuli, M. A.-M. Khan, S.-H. Kee, M. A. Samad, and A.-A. Nahid, "Deep-learning-based semantic segmentation for remote sensing: A bibliometric literature review," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 17, pp. 1390–1418, 2023.

[9] J. Li, Y. Cai, Q. Li, M. Kou, and T. Zhang, "A Review of Remote Sensing Image Segmentation by Deep Learning Methods," *International Journal of*

Digital Earth, vol. 17, no. 1, 2024. doi: 10.1080/17538947.2024.2328827.
[10] A. A. Aleissaee, H. El-Askary, A. A. Nasr, and W. Abd-Elrahman, “Transformers in Remote Sensing: A Survey,” *Remote Sensing*, vol. 15, no. 7, p. 1860, 2023.
[11] Y. Ma, “Transformer-Based Method for Semantic Segmentation and Reconstruction of the Martian Surface,” *ISPRS Archives*, vol. XLVIII-1/W2, pp. 1643–1649, 2023.
[12] C. Zhang, Y. Sun, and F. Li, “Lightweight semantic segmentation network with configurable context and small object attention,” *Frontiers in Computational Neuroscience*, vol. 17, p. 1280640, 2023.
[13] S. Verma, F. Lindseth, and G. Kiss, “SegDesicNet: Lightweight Semantic Segmentation in Remote Sensing with Geo-Coordinate Embeddings for Domain Adaptation,” in *Proc. IEEE/CVF Winter Conf. on Applications of Computer Vision (WACV)*, 2025.
[14] L. Wang, X. Hu, D. Li, and Y. Zhou, “A Transformer-based multi-modal fusion network for semantic segmentation of remote sensing images,” *Information Fusion*, vol. 94, pp. 1–12, 2024.
[15] H. Tang, Y. Han, J. Zheng, Z. Wang, and L. Wang, “An Optimized YOLOv5s-rd Framework for Efficient Target Detection in Remote Sensing Images,” *Informatica*, vol. 49, no. 18, 2025.
[16] Y. Li, Y. Zhao, Y. Zhang, and W. Gong, “A2-FPN: An attention augmented feature pyramid network for building segmentation from very high-resolution images,” *Remote Sensing*, vol. 13, no. 6, pp. 1123–1137, 2021.
[17] J. Li, J. Yang, L. He, and S. Gong, “ABCNet: Adaptive bilateral context network for semantic segmentation of remote sensing images,” *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 179, pp. 108–122, 2021.
[18] Y. Bo, C. Zhang, and L. Fan, “BASNet: Burned area segmentation network for real-time damage mapping in remote sensing images,” *Remote Sensing Letters*, vol. 12, no. 4, pp. 349–358, 2021.
[19] J. Song, Y. Zhu, and A.-X. Zhu, “Transformer-based semantic segmentation for building footprint extraction,” *Sensors*, vol. 23, no. 11, pp. 1–17, 2023.
[20] M. Gibril, R. Smits, and A. Samadzadegan, “Large-scale building extraction using Swin Transformer and Mask2Former,” *Advances in Space Research*, vol. 73, no. 1, pp. 25–37, 2024.
[21] H. Cui, X. Li, and Q. Yang, “Improved Swin Transformer-based semantic segmentation of post-earthquake dense buildings in urban areas,” *Remote Sensing*, vol. 15, no. 8, pp. 2302–2315, 2023.
[22] S. Khallaghi, J. R. Eastman, and L. D. Estes, “A review of technical factors to consider when designing neural networks for semantic segmentation of Earth Observation imagery,” *arXiv preprint* arXiv:2308.09221, 2023.
[23] C. Broni-Bediako, J. Xia, and N. Yokoya, “Real-time semantic segmentation: A brief survey and comparative study in remote sensing,” *Remote Sensing*, vol. 15, no. 19, pp. 4855–4872, 2023.
[24] W. Zhao, H. Tang, and Y. Liu, “SEG-Road: A CNN-transformer hybrid network for road segmentation from VHR images,” *Computer Vision and Image Understanding*, vol. 221, p. 103470, 2022.
[25] L. Wang, X. Hu, and D. Li, “MetaSegNet: Metadata-collaborative vision-language representation learning for semantic segmentation of remote sensing images,” *arXiv preprint* arXiv:2312.12735, 2023.
[26] O. Ajibola and J. Cabral, “A review and meta-analysis of semantic segmentation models in land use/land cover mapping,” *ISPRS International Journal of Geo-Information*, vol. 12, no. 4, p. 216, 2024.
[27] Y. Zhang, X. Meng, and L. Zhou, “A semi-supervised learning method for remote sensing semantic segmentation based on consistency regularization,” *Information Fusion*, vol. 80, pp. 93–104, 2023.
[28] Y. Chen, S. Lu, and T. Zhang, “Few-shot semantic segmentation of remote sensing imagery using prototype learning,” *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 201, pp. 45–59, 2024.
[29] F. Liu, H. Jin, and Y. Sun, “Domain adaptation for semantic segmentation of remote sensing images with distribution alignment,” *Computer Vision and Image Understanding*, vol. 239, p. 103828, 2024.
[30] J. Wang, L. Zhang, and K. Lin, “Cross-domain semantic segmentation for remote sensing imagery using adversarial learning,” *ISPRS International Journal of Geo-Information*, vol. 13, no. 1, p. 25, 2024.
[31] Y. Tao, M. Shi, and J. Cheng, “Attention-guided deep neural network for road extraction from high-resolution remote sensing images,” *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 182, pp. 200–214, 2021.
[32] H. Zhu, W. Zhang, and C. Yang, “Urban green space segmentation using spectral-spatial fusion of high-resolution satellite images,” *Remote Sensing Letters*, vol. 14, no. 1, pp. 43–55, 2023.
[33] Q. Zhang, B. Liu, and Y. Wang, “Semantic segmentation for agricultural field mapping using remote sensing images and encoder-decoder networks,” *Agricultural Systems*, vol. 207, p. 103633, 2023.
[34] L. Wang, S. Dong, Y. Chen, X. Meng, S. Fang and S. Fei, "MetaSegNet: Metadata-Collaborative Vision-Language Representation Learning for Semantic Segmentation of Remote Sensing Images," in IEEE Transactions on Geoscience and Remote Sensing, vol. 62, pp. 1-11, 2024, Art no. 5644211, doi: 10.1109/TGRS.2024.3477548.
[35] D. Hong et al., "SpectralGPT: Spectral Remote Sensing Foundation Model," in IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 46, no. 8, pp. 5227-5244, Aug. 2024, doi: 10.1109/TPAMI.2024.3362475.