Deep-Learning-Based Semantic Segmentation for Remote Sensing: A Bibliometric Literature Review

Kazi Rakib Hasan, Anamika Biswas Tuli, Md. Al-Masrur Khan, Seong-Hoon Kee, Md Abdus Samad, *Member, IEEE*, and Abdullah-Al Nahid

***Abstract*—Deep learning (DL) has emerged as a powerful technique for a wide range of computer vision applications. Consequently, DL is also being adopted to process geospatial and remote sensing (RS) images. As these methods are sporadic over different studies, many review papers have also been published to gather the approaches and summarize the existing models in this field. However, a state-of-the-art review paper is still scarce in this field that will present a bibliometric analysis as well as a critical analysis of the recent works. Therefore, this article aims to spur the researchers with a bibliometric analysis to identify the current research trend. As a research sample, in total, 281 related papers were collected from the Web of Science source, and bibliometric analysis was accomplished using VOSviewer software. Among the collection of associated works from the database, 28 papers were selected according to the defined criteria for detailed analysis. Besides this, a few research questions were generated to extract necessary information from the literature for extracting the pros and cons of the selected works. DL techniques were applied in these works and achieved results. Furthermore, the papers were also categorized based on the addressed RS application domain.**

***Index Terms*—Bibliometric analysis, deep learning, remote sensing (RS), segmentation, VOSviewer.**

I. Introduction

REMOTE sensing (RS) technology is being used as a primary tool to analyze data and provide necessary information about various fields related to agriculture [1], environment monitoring [2], catastrophe risk management [3], urban planning [4], and so on. The data used in the RS technology are mainly comprised of RGB images, captured by unmanned aerial vehicles (UAVs) and hyperspectral images collected from satellites. Consequently, processing these remotely sensed images is one of the crucial steps for utilizing RS technology in the application of the abovementioned fields to find helpful information. The processing of the images involves several tasks like classification, change detection, image fusion, segmentation, etc. [5]. Hence, researchers from all over the world are devoting their efforts to developing RS methods for analyzing the images accurately and improving the performance of the tasks. Over the years, neural networks (NNs) have been employed by RS researchers.

However, with the advancement of machine learning (ML) classifiers, the RS researchers shifted their focus to the tree-based classifiers (e.g., random forest (RF) [10] and support vector machine (SVM) [9]), from the very basic NN algorithms. The SVM got more attention than the other methods due to its good results with less training data and high-dimensional data handling. On the other hand, the RF proved itself a lightweight algorithm with high accuracy (fewer parameters). In the last decade, deep learning (DL) technology has gained vast popularity among researchers by showing mesmerizing performance in different computer vision applications. The DL models can automatically decode necessary and relevant features from the images while handling different complex scenarios to solve the targeted computer vision problems. The remarkable success of DL algorithms in other fields has impelled the RS community to adopt the DL methods in RS technology [11]. It is found that since 2014, RS researchers have already started implementing DL models like convolutional neural networks (CNNs) [12], graph NNs [13], generative adversarial networks [14], etc., in various RS tasks. And these algorithms have produced significant outcomes in many RS applications including land cover classification [15], forest change detection, [16] semantic segmentation of urban scenes [17], etc.

Semantic segmentation is a widely adopted process nowadays for analyzing remotely sensed images among the DL-based image analysis problems in the RS domain [24]. Rather than predicting a single image, the semantic segmentation process predicts all the pixels of an image, giving detailed information about the embedded objects in an image. With the continuous advancement of the semantic segmentation methods, they are being employed to solve RS problems, including crop type analysis [18], building damage monitoring [19], urban village settlement [20], etc. As a result, already many outstanding technical research works have been published that focus on solving RS problems by utilizing semantic segmentation techniques [21], [22], [47]. Fig. 1 shows a few examples of the segmentation result on various RS fields.

However, there is still a scarcity of review papers that will pay attention to bibliometric analysis with critical extraction

Manuscript received 2 June 2023; revised 9 August 2023, 4 September 2023, and 21 September 2023; accepted 24 October 2023. Date of publication 30 October 2023; date of current version 14 December 2023. *(Corresponding authors: Abdullah-Al Nahid; Md Abdus Samad.)*

Kazi Rakib Hasan, Anamika Biswas Tuli, and Abdullah-Al Nahid are with the Electronics and Communication Engineering Discipline, Khulna University, Khulna 9208, Bangladesh (e-mail: rakib17.rh@gmail.com; tulianamikabiswas@gmail.com; nahid.ece.ku@gmail.com).

Md. Al-Masrur Khan and Seong-Hoon Kee are with the Department of ICT Integrated Ocean Smart Cities Engineering, Dong-A University, Busan 49315, South Korea (e-mail: almasrurkhan@donga.ac.kr; shkee@dau.ac.kr).

Md Abdus Samad is with the Department of Information and Communication Engineering, Yeungnam University, Gyeongsan-si 38541, South Korea (e-mail: masamad@yu.ac.kr).

Digital Object Identifier 10.1109/JSTARS.2023.3328315

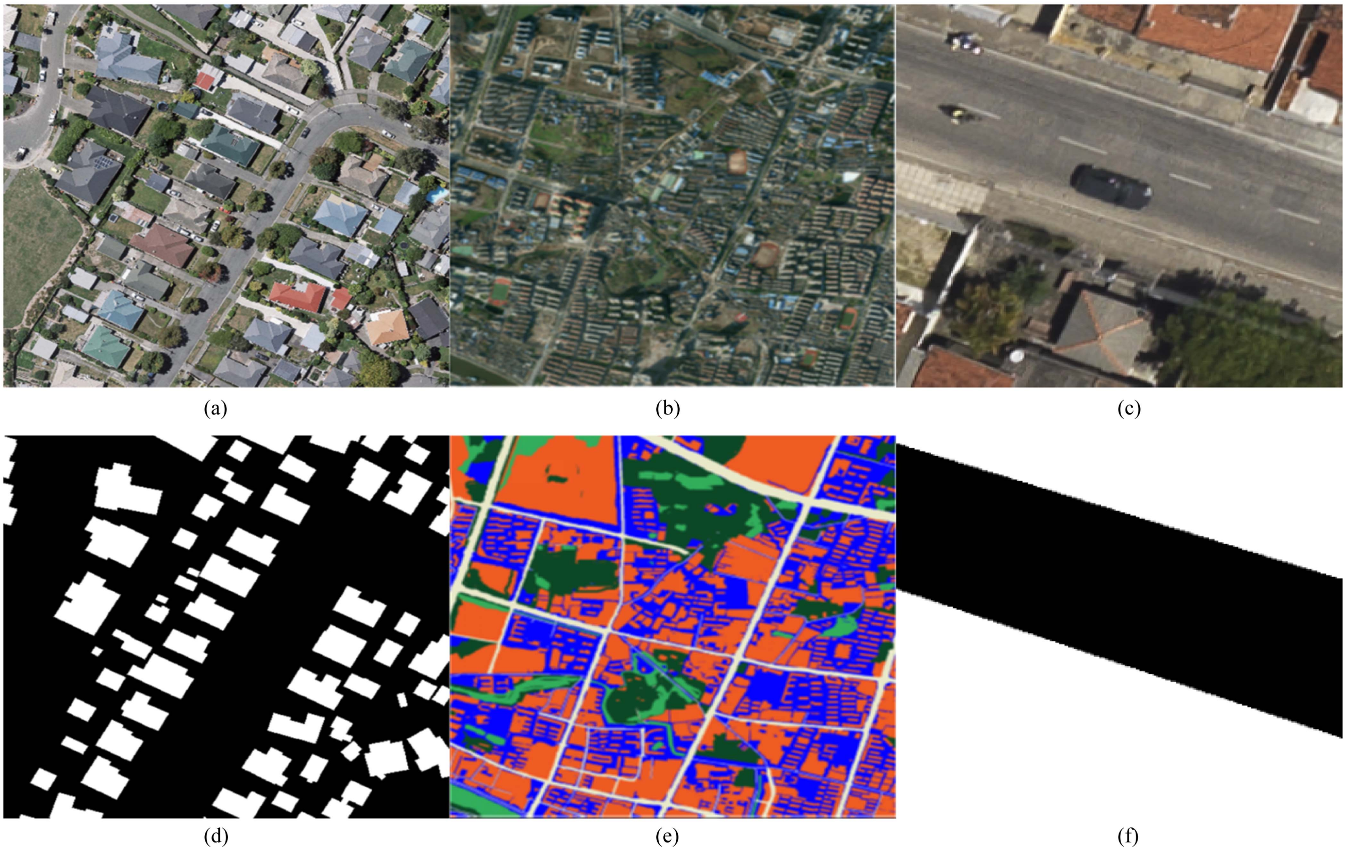

Fig. 1. Illustration of several segmentation examples on various RS images. These image are collected from [6], [7], and [8]. (a) Original image containing buildings. (b) Original image containing different scenes. (c) Original image containing road. (d) Segmented image after extracting buildings. (e) Segmented image after performing scene classification. (f) Segmented image after extracting road.

of particular RS domain knowledge. The bibliometric analysis will extract current research trends, influential articles, journals, countries, and pioneering authors in this arena. The critical analysis of the existing research works will filter out the specific application field and utilize DL architectures on the corresponding papers. Accordingly, considering this deficit and research scope in this work, we will present a review paper by analyzing a few influential documents published between 2015 and 2021, focusing on semantic segmentation in RS technology. The significant contributions of this review article are the following.

1) This article presents a bibliometric analysis of a few selected papers focused on the semantic segmentation of remotely sensed images. Data mining techniques were used to highlight the current research trend, important research terms, influential publications, journals, and collaboration patterns of this research field.
2) This article presents a comprehensive analysis of the selected papers to find out the specific application fields of segmentation techniques in RS.
3) This article discusses the DL architectures used in the corresponding papers and also provides a summary of the prominent papers by answering a few research questions in the form of a table.

The rest of this article is organized as follows. Section II provides a concise recap of the existing survey papers on the DL-based RS works. Section III provides the methodology of this work. Section IV depicts the scientometric analysis of a few papers from the utilized dataset of this work. Section V presents the critical analysis of the DL-based works. Section VI lists out the findings of the study, key challenges of this research domain, and research direction for future researchers. Finally, Section VII concludes this article.

II. Literature Review

DL has proven its potential to meet the challenging requirements of RS image processing for the last few years. Innovation and progression in this field are all moving at a very fast speed. In addition to the continual endeavor of upgrading the algorithms, researchers also enlisted the existing methodologies in a theme of the review paper to accelerate the research in this area. This section recapitulates a few leading aspects from previous papers and covers the landmarks of the neoteric articles that prove themselves a great inclusion to the research field. Table II summarizes the survey papers published in the domain of DL-based RS technology.

The first review paper that we examine in this work was published in June 2017 [23]. The authors presented the findings of a few papers and did a meta-analysis of image source sensors used in those papers. They also discovered that RF

TABLE I
PRESENTATION OF THE CLASSIFIED GROUPS AND THE ACCOMPANYING REVIEW PAPERS

Group	Reference
A (Traditional)	[24],[25],[28],[29],[34],[35]
B (Systematic)	[30],[31],[36]
C (Architectural)	[23],[27],[31],[33],[36]
D (Application-based)	[26],[28],[29],[32],[33]

exhibits the best performance in object-based categorization, and detection-based frameworks limit the application of fuzzy techniques. Moreover, they identified that the SVM method also enhances classification accuracy in supervised object-based image classification. Zhu et al. [32] showed that in the RS arena, DL opens up additional opportunities such as global trend monitoring and appraising the current resource conservation techniques. They also addressed the issues of DL in RS mechanics and recommended critical resources to identify the DL-based RS research field. Ming et al. [33] presented a survey paper in 2019 where they summarized the basic principles of RS along with its conventional applications. They illuminated the most important DL models in RS for land cover classification, target recognition, and change detection. They gave an incisive description of the advantages and disadvantages of these methods. They also explored the research progress, development history, and significant issues that researchers face in this field and future development strategies.

In 2021, Osco et al. [34] provided a systematic overview of the fundamentals of DL in UAV-based imagery. The classification and regression algorithms used in recent UAV-acquired data applications are also discussed in their paper. In 2021, Neupane et al. [29] published an article of review and meta-analysis in the *Remote Sensing* journal. Among the research papers that were analyzed and discussed in this work, this was the only review paper that issued statistical and bibliometric analyses. It also reviewed some articles that employed DL algorithms to classify urban images. In 2021, Yuan et al. [24] reviewed a few DL algorithms for semantic segmentation of RS imagery. The constitutive structures of different CNNs and their applications in the field of RS semantic segmentation were also discussed by the reviewer. They also reviewed DL model performance in different datasets and identified how the lack of proper trainable datasets affects the model performance. Zang et al. [26] used DL for land-use mapping (LUM) of high-spatial RS images in 2021. They summarized two LUM criteria that were utterly dependent on DL. Moreover, this article gives an idea of the productiveness of representative semantic segmentation and single object segmentation. According to some interrelationships and congruity, these works can be classified into four groups shown in Table I. All of the papers have extraordinary contributions to this augmented research field, but they have several limitations, such as these review papers did not systematically collect the articles [5], [24]. The review papers [25] and [33] did not provide proper statistical or bibliometric analysis of the papers except [29]. Moreover, most reviews did not classify the articles according to the application sectors [32]. Besides these, a few papers gave only an overview of the benchmark DL methods for utilizing in RS tasks; they did not analyze the papers individually based on the modified architecture and the integrated modules on the networks, which accelerate the performance of the tasks [23], [27]. Since there is still a significant literature gap and huge research opportunities in this domain, we will delineate the existing research paper systematically.

In this study, we will provide both statistical and bibliometric analyses, giving the new researchers a vast idea about research trends, important research terms, collaboration patterns, influential journals, authors, and articles in this field. Moreover, we will cluster the influential articles based on different application sectors, which will give a clear vision of the research topic to novice researchers. Furthermore, this article will present a critical analysis of a few prominent research articles based on their problem statement, methodology, and achieved results. During the analysis, we will mention the benchmark DL method. We will also discuss how the papers are modifying the benchmark DL architectures to develop new methods for getting better results in different complex scenarios. This will help the readers of this article to get familiarized with the modified DL methods and the fast-changing RS image semantic segmentation sector.

III. METHODOLOGY

This work has been designed using a mixed method to present a bibliometric and critical analysis of the papers, focusing on analyzing the remotely sensed images using DL technology. The purpose of utilizing a mixed method is to give readers an overall idea about the RS application fields so that new researchers can easily grasp the necessary knowledge. From bibliometric analysis, the readers can easily learn about the influential authors and journals in this research field. The critical analysis, on the other hand, gives an overview of the different DL architectures that are developed in this domain. Fig. 2 represents the overall methodology of this study.

1) *Data collection*: In this work, we searched for peer-reviewed papers in an online digital library named Web of Science (WoS) using a set of search strings for constructing a database on October 25, 2021. The keywords and Boolean operators were applied in the abstract of the publications. The search string we used in this work was in the following format: “Deep Learning” AND “Segmentation” AND “Hyperspectral OR Image” AND “Remote Sensing.”
After searching, we found within the search string that the first paper was published in 2015, in which DL was used for remotely sensed images. Therefore, we selected all the papers till the 25th of October and collected a total of 281 papers. Then, we generated a text file from the WoS, which contained the entire record and cited references from the 281 papers. Later, we used this text file as our dataset for performing bibliometric studies in the second stage of our work.
2) *Bibliometric analysis*: In this stage, we performed a bibliometric-based analysis of the WoS dataset. We split the process into two categories for conducting the

TABLE II
SUMMARY OF PREVIOUS REVIEW ON DL-BASED RS

Ref	Year of publication	Name of the Journal/Conference	Major contributions	Limitations
[23]^C	2017	*ISPRS Journal of Photogrammetry and Remote Sensing*	• Constituted a database by systematic analysis. • Found out that RF performs the best in object-based categorization and object-based framework limits the development of fuzzy technique. • Marked off two methods, DL and type-2 fuzzy technique, in which DL enhances classification accuracy in supervised object-based image classification. • Findings of the papers' meta-analysis are presented in detail.	• Didn't present any bibliometric analysis as well as did not analyze the papers individually.
[24]^A	2021	*Expert Systems with Applications*	• Discussed the constitutive architectures of NN. • Analyzed the recent developments in the domain of semantic segmentation of RS. • Discussed precisely the vital contraventions along with the overcomings of this research area.	• Didn't collect the papers in a systematic way. • They only discussed the DL architectures but didn't mention how the papers used these models or modified the architectures for better results.
[25]^A	2019	*ISPRS Journal of Photogrammetry and Remote Sensing*	• Presented a systematic analysis of DL applications in various fields of RS image technologies. • This overview covers practically every technology in the realm of remote sensing.	• The papers are not collected in a systematic way. • Didn't categorize the papers based on their corresponding image processing techniques.
[26]^D	2019	*ISPRS International Journal of Geo-Information*	• Provides a full-scale overview of how the UAV-based damage mapping process has progressed from simple expository overviews to more complicated textures and methods based on segmentation. • Concentrating on the findings of two recent European research initiatives (RECONAS and INACHUS), the work examines studies that have been conducted on the effectiveness of advanced mapping methodologies and image processing pipelines for first responses.	• The articles are not collected in a systematic way. • They did not mention any critical or statistical analysis.
[27]^C	2021	IEEE JOURNAL OF SELECTED TOPICS IN APPLIED EARTH OBSERVATIONS AND REMOTE SENSING	• This study offered a thorough examination of the Land-Use Mapping (LUM) method based on DL. • It Summarized DL-based approaches using two LUM criteria (one is supervised or semi-supervised or unsupervised learning and another one is pixel-based or object-based) • Provides an idea of the effectiveness of representative semantic segmentation and single object segmentation. • It reviews some mostly used basic DL techniques for LUM.	• Though they collected the dataset in a systematic way, however, didn't analyze the papers one by one based on their features or characteristics.
[28]^{A,D}	2020	*Remote Sensing*	• Discussed influential DL architectures for land cover classification of coastal wetlands. • They look into the possibility of DL architectures with a small amount of labeled training data. • Identified a DL model that is far better in the basis of inference speed and accuracy.	• They didn't present a systematic and bibliometric analysis of the papers. • They only discussed the DL architectures but didn't analyze individual papers.
[29]^{A,D}	2020	*Remote Sensing*	• Found out that mostly used data sources for research in oil control are satellites and microwave airborne. • Along with comparing a few models, they found that DL showed better accuracy than the other models. • Highlighted the issues interrupting the management process for oil spills.	• They didn't collect the papers in a systematic way and didn't present any critical analysis of the papers one by one.
[30]^B	2021	*Remote Sensing*	• The only paper that provided bibliometric and statistical analysis of the papers. • Reviewed some related papers which used DL methods for urban image classification. • Also found out the answers to some research problems after analyzing the papers. • Thoroughly inspected and discovered that DL surpasses standard methods in terms of accuracy as well as addresses some prior unsolved issues. • Not only analyzes the papers but also discussed future directions in this field.	• They discussed some DL models in remote sensing and gathered statistical information but didn't discuss elaborately the mentioned papers one by one.
[31]^{B,C}	2021	*Remote Sensing*	• This study gives a summary of the advanced attention mechanism and its affinity with various DL network models and RS-image processing based on DL. • Represents a structured review to define the current tendencies of research-related terms (publications, publishers, improved DL methods, data types used, etc.) • Found out the current limitations of research fields and also mentioned how to overcome those.	• They extracted the papers for review using the search engines of magnificent publication platforms and analyzed them statistically but didn't describe the papers one by one.
[32]^D	2021	*Remote Sensing*	• A detailed survey of the most recent breakthroughs in precision agriculture (PA) using UAV RS and edge intelligence is presented here. • This study explores the cloud computing and edge computing concepts for the UAV RS in PA. • The authors analyzed the relevant edge computing technique thoroughly for the first time in their work. • This study presents a comprehensive overview of UAV intelligent edge devices as well as the most recent developments in edge inference using model compression.	• Didn't provide any systematic review of the papers. • Didn't present any bibliometric analysis of the papers

TABLE II
(CONTINUED.)

Ref	Year of publication	Name of the Journal/Conference	Major contributions	Limitations
[33][C,D]	2017	*IEEE Geoscience and Remote Sensing Magazine*	• Showed that apart from addressing technological issues, DL in RS gives up new possibilities such as monitoring global trends or evaluating resource conservation measures. • They addressed the challenges of DL in RS technologies. • Provided the essential resources to enrich the research field of DL-based RS. • Highlighted the state-of-the-art developments in DL-based RS.	• Didn't gather the papers in a structured way. • Though they discussed the DL architectures elaborately, they didn't analyze or cluster the papers according to the application sectors.
[34][A]	2019	*International Journal of Geosciences*	• This study summarizes the fundamental principles as well as the most common RS applications. • Analyzes the research progress, development history, and the main problems which researchers are facing in this research field and further development supervision. • Exposes the major DL models for land cover classification, target detection, and change detection in RS.	• They didn't collect the papers in a systematic way. • Didn't mention any statistical or bibliometric analysis. • Didn't analyze the papers individually.
[35][A]	2021	2020 *International Journal of Applied Earth Observation and Geoinformation*	• Provides a wide overview of the foundations of DL in UAV-based imagery. • Explains the classification and regression algorithms employed in recent UAV-acquired data applications. • This work evaluates the available papers and assessed their properties in terms of applications, sensors, and procedures. • Also discussed how DL analyses UAV-based image data.	• They presented a systematic review but didn't show any bibliometric analysis. • They didn't look at each article individually to identify what features or characteristics they had.
[36][B,C]	2021	*ISPRS Journal of Photogrammetry and Remote Sensing*	• Discussed the constitutive architectures of CNNs • Presented a systematic analysis of DL methods in various fields of image segmentation technologies.	• Didn't analyze the papers individually • Didn't discuss the papers based on their corresponding image processing techniques.

[A]Traditional review papers; [B]Systematic review papers; [C]Review papers that focus on the DL architectures; [D]Review papers that focus on the RS application sectors.

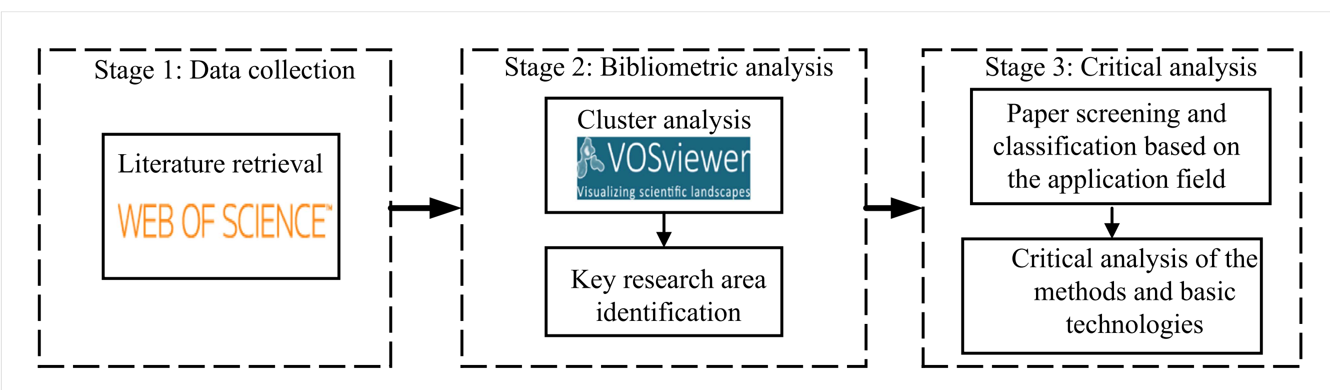

Fig. 2. Overview of research methodology.

bibliometric analysis: a) performance analysis and b) science mapping.

We discussed the emerging trend of publications and journal performance. We extracted the pioneer authors and the leading countries to examine the contribution of research constituents and the intellectual structure of the RS research field. After then, we utilized the VOSviewer software to present a science mapping analysis of our dataset to show the relatedness between research constituents. Moreover, we discussed the collaboration patterns in this research field using the science mapping technique and found the critical research terms in this domain.

3) *Critical analysis:* For conducting critical analysis, we have selected a subset of the initially collected papers from the WoS. The utilized dataset had an H-index value of 29. We then extracted these 29 papers to present and analyze them critically for having a better idea about the application sectors of RS and utilized DL architectures. However, when we screened all 29 papers, we found that one paper was a review paper. Excluding that paper, the 28 papers were briefly discussed in this section.

Finally, we critically analyzed the papers based on the problem statement and the utilized methodology of the papers. In this section, the papers were categorized into several groups based on the application sectors of the remotely sensed images. The applied DL methods and their modified architectures were briefly discussed in this section.

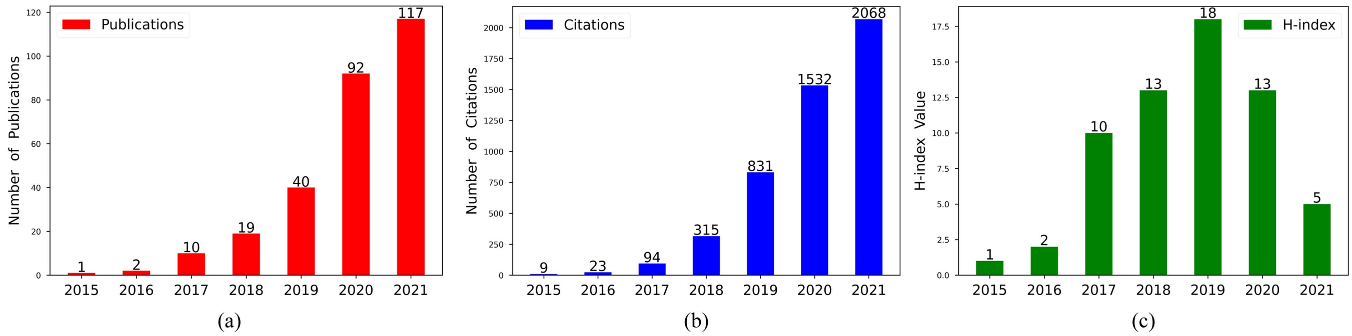

Fig. 3. Overview of annual statistics of the publications. (a) Number of publications over the years (2015–2021). (b) Number of received citations by the publications over the years (2015–2021). (c) H-index of the publications over the years (2015–2021).

IV. Scientometric Analysis

A. Overview of the Publications

1) Annual Analysis of the Publications: This section demonstrates the annual exploration of the publications of the elected 281 papers from 2015 to 2021. From Fig. 3(a), it can be observed that in the very beginning, research in this field started with only one and two articles in 2015 and 2016, respectively. The number of annual publications has started to increase significantly since 2017, five times more than the previous year. An upward tendency can be observed from the figure in the following years. Consequently, in 2021, 117 papers had been listed in the publications, which was 41.64% of the total published papers. Therefore, it can be comprehended that researchers became more and more interested in exploring this field.

Another exploration of this work is of annual citations gained by the publications. Fig. 3(b) presents the number of total attained citations per year. It is evident from the figure that the citation rate had been increasing from the beginning. From 2015 to 2021, papers achieved a total of 4849 citations. The figure also showed that in 2015, the attained citation number was 9, while in 2016 and 2017, it turned to 23 and 94, respectively. After that, the citation rate started to grow notably. In 2018, the publications were cited 315 times. During the following three years, the publications got 4449 citations, which was 90.90% of the total citations. The received citation numbers from 2019 to 2021 were 831, 1532, and 2086, respectively, which indicated an upward trend in the case of received citations. From this, it can be said that researchers are contributing frequently to this field. Therefore, it is expected that this research area will reign in the upcoming decades.

This work also analyzed the annual publications using another metric named H-index. The values of the H-index of the publications over the years from 2015 to 2021 are exhibited in Fig. 3(c). It clearly shows that the highest H-index was gained in 2018 and the value was 18, and the lowest H-index was 1 in 2015. Of the 281 papers, only 29 had at least 29 citations; thus, the final H-index of the elected publications for critical analysis was decided to be those 29 papers. From the figure, an upward tendency can be noticed from 2015 to 2019 in the value of the H-index. This can be contributed to the growing research interest of the researchers as a new research direction (DL) became prominent. And as time went by, the interest became saturated. By October 2021, the H-index for the year 2021 was 5, which was not a good statistic. However, it could be expected from the trend of the previous year's H-index that a significant number of citations might be obtained by the end of the year.

2) Most Cited Publications: This work extracted 281 papers that were the most commanding, creative, and popular among the researchers. From them, 15 top-cited papers were considered for statistical analysis in this section. Analysis revealed that these papers were cited 3053 times, about 62.04% of the total citations received by all publications. These papers were arranged in Table III with the following information: their title, the source journal that published the paper, the corresponding author's name and his country, the number of citations, and the average citation number per year.

The top-cited paper on the list titled "Road extraction by deep residual U-Net" was published in 2018 and earned 1749 citations. It was published in IEEE Geoscience and Remote Sensing Letters. The second paper in the list titled "Deep learning classification of land cover and crop types using remote sensing data" has 1438 citations. The average citation rate for the paper is 239.6, which is lower than the paper entitled "Deep learning in remote sensing applications: A meta-analysis and review." Though this is in the third position (1364 citations) in the list of top-cited papers, it holds the second-highest average citation rate (341). It was published in the *ISPRS Journal of Photogrammetry and Remote Sensing* in 2019. The least cited papers among these 15 were cited 162 times, and the average citation per year is 27. With a more profound analysis, it can be seen that all these 15 articles are from only four journals. Those are IEEE Geoscience and Remote Sensing Letters, *ISPRS Journal of Photogrammetry and Remote Sensing*, IEEE Transactions on Geoscience and Remote Sensing, and *Remote Sensing*. One more thing to note from this analysis is that the authors of 11 papers of these 15 are from China, which appraises that researchers from China are dedicating their huge endeavor to this research area.

B. Influential Journal, Authors, and Countries

In this section, we extracted and analyzed the influential journals, authors, and countries in the research domain of RS.

TABLE III
SUMMARY OF THE TOP-CITED PAPERS

Title	Journal	Corresponding author	Country of corresponding author	Publication year	Citation (Google Scholar)	Average citation per year
Road Extraction by Deep Residual U-Net [38]	IEEE Geoscience and Remote Sensing Letters	Qingjie Liu	Peoples R China	2018	1749	349.8
Deep Learning Classification of Land Cover and Crop Types Using Remote Sensing Data [37]	IEEE Geoscience and Remote Sensing Letters	Nataliia Kussul	Ukraine	2017	1438	239.6
Deep learning in remote sensing applications: A meta-analysis and review [39]	*ISPRS Journal of Photogrammetry and Remote Sensing*	Lei Ma	Peoples R China	2019	1364	341
A review of supervised object-based land-cover image classification [40]	*ISPRS Journal of Photogrammetry and Remote Sensing*	Lei Ma, Manchun Li	Peoples R China	2017	801	133.5
Fully Convolutional Networks for Multisource Building Extraction From an Open Aerial and Satellite Imagery Data Set [44]	IEEE Transactions on Geoscience and Remote Sensing	Shunping Ji	Peoples R China	2018	705	141
ResUNet-a: A deep learning framework for semantic segmentation of remotely sensed data [49]	*ISPRS Journal of Photogrammetry and Remote Sensing*	Foivos Diakogiannis I	Australia	2020	638	212.6
Compressed-Domain Ship Detection on Spaceborne Optical Image Using Deep Neural Network and Extreme Learning Machine [41]	IEEE Transactions on Geoscience and Remote Sensing	Chenwei Deng	Peoples R China	2014	435	48.33
End-to-End Change Detection for High Resolution Satellite Images Using Improved UNet plus [50]	*Remote Sensing*	Daifeng Peng	Peoples R China	2019	408	102
Building Extraction in Very High Resolution Remote Sensing Imagery Using Deep Learning and Guided Filters [43]	*Remote Sensing*	Xie Zhong	Peoples R China	2018	378	75.6
Classification for High Resolution Remote Sensing Imagery Using a Fully Convolutional Network [42]	*Remote Sensing*	Gang Fu	Peoples R China	2017	307	51.16
Segment-before-Detect: Vehicle Detection and Classification through Semantic Segmentation of Aerial Images [47]	*Remote Sensing*	Nicolas Audebert	France	2017	269	44.83
Semantic segmentation of slums in satellite images using transfer learning on fully convolutional neural networks [48]	*ISPRS Journal of Photogrammetry and Remote Sensing*	Michael Wurm	Germany	2019	258	64.5
Change detection based on deep feature representation and mapping transformation for multi-spatial-resolution remote sensing images [45]	*ISPRS Journal of Photogrammetry and Remote Sensing*	Maoguo Gong	Peoples R China	2016	255	36.42
Gated Convolutional Neural Network for Semantic Segmentation in High-Resolution Images [46]	*Remote Sensing*	Shiming Xiang	Peoples R China	2017	193	32.16
Superpixel-Based Difference Representation Learning for Change Detection in Multispectral Remote Sensing Images [51]	IEEE Transactions on Geoscience and Remote Sensing	Maoguo Gong	Peoples R China	2017	162	27

TABLE IV
SUMMARY OF THE MOST PRODUCTIVE JOURNALS

Journal name	Total publications	Total citations	Average citations	Impact factor	5 years Impact factor	Publisher	H-index	Quartile in Category
Remote Sensing	92	1200	13.04	4.848	5.353	MDPI	15	Q2
IEEE JOURNAL OF SELECTED TOPICS IN APPLIED EARTH OBSERVATIONS AND REMOTE SENSING	32	205	6.41	3.784	3.734	IEEE	7	Q2
IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING	18	624	34.67	5.60	6.086	IEEE	9	Q1
IEEE ACCESS	16	1163	72.69	3.367	3.671	IEEE	11	Q2
ISPRS *Journal of Photogrammetry and Remote Sensing*	16	1163	72.69	8.979	9.948	International Society for Photogrammetry and Remote Sensing	4	Q1
International Journal of Remote Sensing	13	108	8.31	2.976	3.266	Taylor & Francis	2	Q2
Information	9	70	7.78	2.899	2.971	MDPI	4	Q2
IEEE GEOSCIENCE AND REMOTE SENSING LETTERS	9	1026	114	3.966	3.986	IEEE	6	Q2
Journal of Applied Remote Sensing	7	44	6.29	1.53	1.565	Spie-Soc Photo-Optical Instrumentation Engineers	3	Q4
Applied Sciences-Basel	6	21	3.5	2.679	2.736	MDPI	2	Q3

From this, the new researchers can get a compact overview of this domain. By extracting the influential authors, it became clear which authors are contributing to the field of the RS domain, and the new researchers can learn which authors they should follow. From the influential journals, it became clear which journals a new researcher should publish their work to and which journals are more authoritative and of better quality. Readers can also get an overview of which journals they should follow to get more recent and relevant information in the field of RS. The productive countries in this domain provide an overview of development trends in the field of the RS domain. Also, researchers can get an idea about which countries would be interested in collaborating with their research.

1) Most Productive Journals: The selected 281 papers from this research are published in 64 different journals. In this section, this work delineates the ten most abundant publication sources in the field of image-based RS technology. Since 218 papers (77.58% of all papers) are published in these ten journals, it is possible to get an overall idea about the publication sources of this research field by analyzing these journals. The journals have been analyzed in Table IV based on their publication as well as their received citation numbers, their average citations, impact factor, five-year impact factor, publishers of the journals, and H-index.

Along with the table, for getting a better visualization of the productive journals, we provided a citation network of a few prominent journals in Fig. 4. The journal "*Remote Sensing*" by "MDPI" is in first place with 92 publications (see Table IV). This journal collected a total of 1200 citations over the years. From these, we can realize the productivity and popularity of this journal among researchers. Moreover, its impact and five-year impact factors are 4.848 and 5.353, respectively. The "IEEE JOURNAL OF SELECTED TOPICS IN APPLIED EARTH OBSERVATIONS AND REMOTE SENSING" has 32 publications and 205 citations. Publications of the following four journals on the list are 18, 16, 16, and 13, but the received citations of these journals are noticeably divergent. Total publication and total citation, along with an average citation, are the same for IEEE ACCESS and *ISPRS Journal of Photogrammetry and Remote Sensing*, and they are 16, 11.63, and 72.69. There is a noticeable journal from the IEEE, "IEEE GEOSCIENCE AND REMOTE SENSING LETTERS," with an average citation of 114. Though the journal has lesser publications, it has a good impact factor because of its higher citation rate.

Another fact to discuss in Table IV is the quartile in the category. A quartile is a rank-order classification that categorizes scientific journals and delineates the citation level associated with scientometric indicators. Along with citation, the journal is also categorized by impact factor and index number by Journal Citation Report and SCImago Journal Rank. These journals are ranked into four categories based on quartiles: quartile 1 (Q1) represents the top 25% of the journals, quartile 2 (Q2) includes ranking between top 25% and 50%, quartile 3 (Q3) corresponds to the 25% of the journals after Q2, and quartile 4 (Q4) includes the last 25% journals. Among these ten journals, "*Journal of Applied Remote Sensing*" is in the Q4 category in the quartile, and "*Applied Sciences-Basel*" is in Q3. "IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING" and "*ISPRS Journal of Photogrammetry and Remote Sensing*" are Q1 categories and the rest of the journals are Q2.

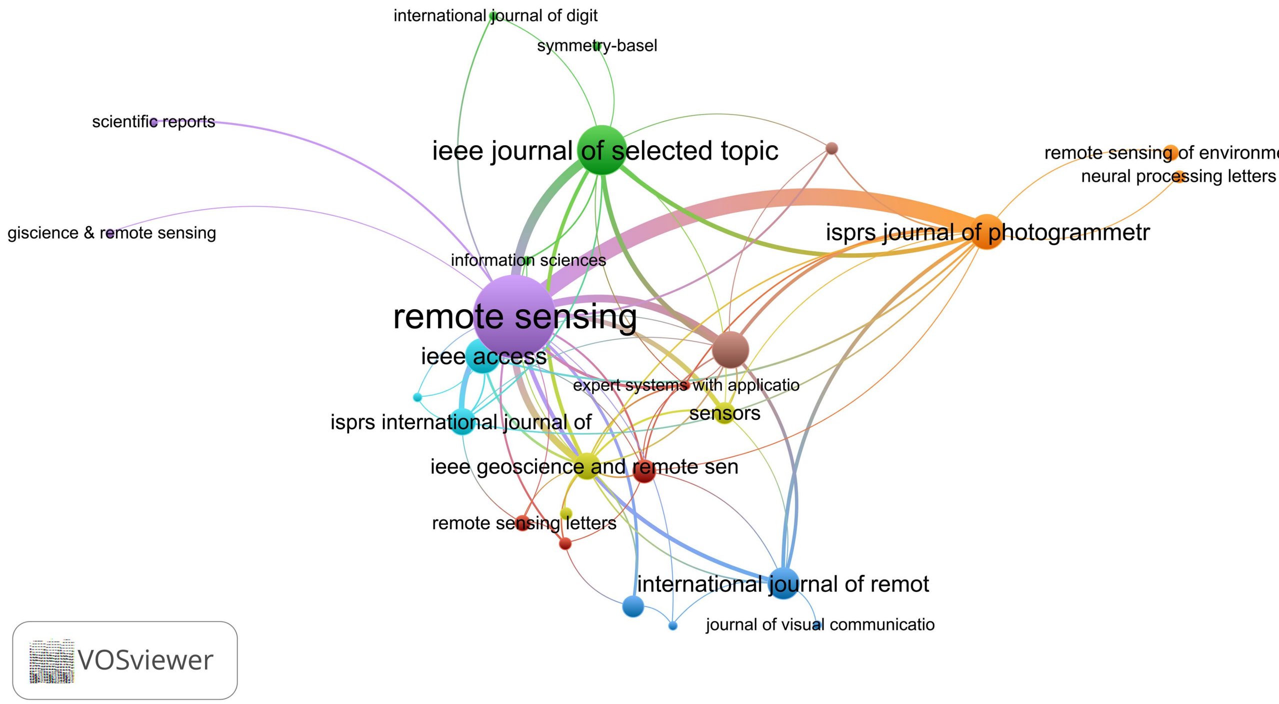

Fig. 4. Citation network of the prominent journals. (The size of the circle indicates the citation number of journals, and the thickness of the connected line shows their relative citation strength.)

This research has also analyzed the historical development of the top publication sources based on publications and citations. This scrutiny is presented in Table V, where P denotes the number of articles and C denotes the citations received by the paper. This table shows that this work collected data from 2015, but in 2015, "IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING" had only one publication and nine citations. This journal did not publish an article in 2016, but it gained 20 citations this year. The publication number is meager even in the next two years; in 2016, "IEEE JOURNAL OF SELECTED TOPICS IN APPLIED EARTH OBSERVATIONS AND REMOTE SENSING" and "*Journal of Applied Remote Sensing*" had one publication each. If we compare "IEEE JOURNAL OF SELECTED TOPICS IN APPLIED EARTH OBSERVATIONS AND REMOTE SENSING" with "IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING," the total publication is almost twice (32 and 18, respectively), but the received citations are much more for the second one. Considering these two journals for 2021, the first one got 84 citations from 19 publications, whereas the second one got 224 citations from seven publications. Thus, we realize that the number of publications does not reflect the popularity. The journal named "*Journal of Applied Remote Sensing*" is notable. It has published nine articles from 2017 to 2021. However, its received citations are 12, 65, 239, 352, and 358, respectively. It is noticed that articles by this journal are influential and popular among researchers. From the discussion, it is also noticed that the journals are following an upward trend in case of receiving citations; as a result, all of the journals received the maximum number of citations in 2021.

2) Most Productive Authors: The work describes the most productive authors in the research domain of image-based RS technology in this subsection. From the WoS database, this work found that 1189 authors are involved with these 281 papers. The top ten authors based on the highest publication numbers were specified. Table VI expounds these authors by their name, number of articles, total citations, citation rate, number of published papers as first author, H-index, and author's belonging country. Along with the table, for getting a better visualization of the influential authors, we provide the citation network of a few prominent authors in Fig. 5.

From the table, it can be seen that Xing Li has the highest publications of seven papers with 40 citations. However, four of his papers did not get any single citations; two papers got four and five citations. The paper entitled "Building-A-Nets: Robust building extraction from high-resolution remote sensing images with adversarial networks" received 31 citations, which are 77.5% of all his citations. He is the first author of four articles among these seven, and his H-index is 3. The following six authors Lei Ma, Shunping Ji, Xian Sun, Ying Zhang, Wei Liu, and Min Xia have the same publication numbers. They all have at least five publications each. Lei Ma gained 747 citations for only five publications. His papers are excellent in quality, significantly productive, and influential. This work got more in-depth observation that Lei Ma had written two review papers. These are "Deep learning in remote sensing applications: A meta-analysis and review" and "A review of supervised object-based land-cover image classification." These two papers received 379 and 348 citations, respectively. Therefore, 97.32% of his citations were gained from only these two review papers.

TABLE V
HISTORICAL DEVELOPMENT OF THE JOURNALS IN TERMS OF THE PUBLICATIONS AND CITATIONS

Journal name	2015		2016		2017		2018		2019		2020		2021	
	P	C	P	C	P	C	P	C	P	C	P	C	P	C
Remote Sensing	0	0	0	0	4	12	5	84	11	212	28	388	44	504
IEEE JOURNAL OF SELECTED TOPICS IN APPLIED EARTH OBSERVATIONS AND REMOTE SENSING	0	0	1	0	3	4	2	22	0	46	7	49	19	84
IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING	1	9	0	20	1	52	2	60	2	115	5	144	7	224
IEEE ACCESS	0	0	0	0	0	0	0	0	5	3	8	15	3	66
ISPRS *Journal of Photogrammetry and Remote Sensing*	0	0	1	3	1	14	0	79	5	160	4	411	5	494
International Journal of Remote Sensing	0	0	0	0	0	0	1	1	3	14	5	33	4	59
ISPRS International Journal of GEO-Information	0	0	0	0	0	0	1	0	1	5	5	27	2	37
IEEE GEOSCIENCE AND REMOTE SENSING LETTERS	0	0	0	0	0	0	0	2	3	7	2	15	2	20
Journal of Applied Remote Sensing	0	0	0	0	1	12	2	65	2	239	2	352	2	358
Applied Sciences-Basel	0	0	0	0	0	0	1	0	2	3	1	8	2	10

^{P}denotes the number of publications; ^{C}denotes the number of obtained citations.

TABLE VI
SUMMARY OF THE MOST PRODUCTIVE AUTHORS

Author's name	Total publications	Total citations	Average citations	As 1^{st} author	H-index	Country
Xing Li	7	40	5.71	4	3	China
Lei Ma	5	747	149.4	2	4	China
Shunping Ji	5	208	41.6	3	4	China
Xian Sun	5	31	6.2	3	2	China
Ying Zhang	5	19	3.8	0	2	China
Wei Liu	5	13	2.6	2	2	China
Min Xia	5	12	2.4	0	2	China
Abolfazl Abdollahi	4	32	8	4	3	Australia
Kun Fu	4	16	4	1	2	China
Jie Li	4	12	3	0	2	China

Shunping Ji got 208 citations for five publications with a citation rate of 41.6. He is the first author of three papers, with the H-index 4.

The rest of the authors, with five publications, attained 31, 19, 13, and 12 citations. Abolfazl Abdollahi, Kun Fu, and Jie Li have four publications each. Abolfazl received 32 citations for his four publications, and his average citation is eight. The total citation and average citation for Kun Fu and Jie Li are 16, 12 and 4, 3, respectively. There is an exciting fact that Kun Fu has only one paper as a first author, and Jie Li is not the first author even in a single paper, whereas Abolfazl Abdollahi is the first author of all of his publications. Again, he is in the third position considering the average citation. If we consider the countries the authors are from, it is seen that they all are from China except Abolfazl Abdollahi, who is from Australia. Thus, we can see that Chinese researchers are so far ahead of others in this research field of image-based RS technology.

3) Most Productive Countries: This work found the most productive countries in the DL-based semantic segmentation research domain. In this section, the work will discuss the countries' contributions in this field. According to the collected dataset from the WoS, authors from 58 countries put their contributions in the case of publishing papers. The geographical distribution of the publications is represented in Fig. 6. Here, we can observe that the lion's share of the contribution is from China (67.978%). The USA, Germany, Japan, and Australia are responsible for 9.608%, 6.049%, 4.626%, and 4.27% of the published articles. Moreover, the rest of 13.167% of the publications are from the other mentioned parts of Fig. 6. Along with the geographical distribution of the publications, this work describes the most productive countries in this research domain.

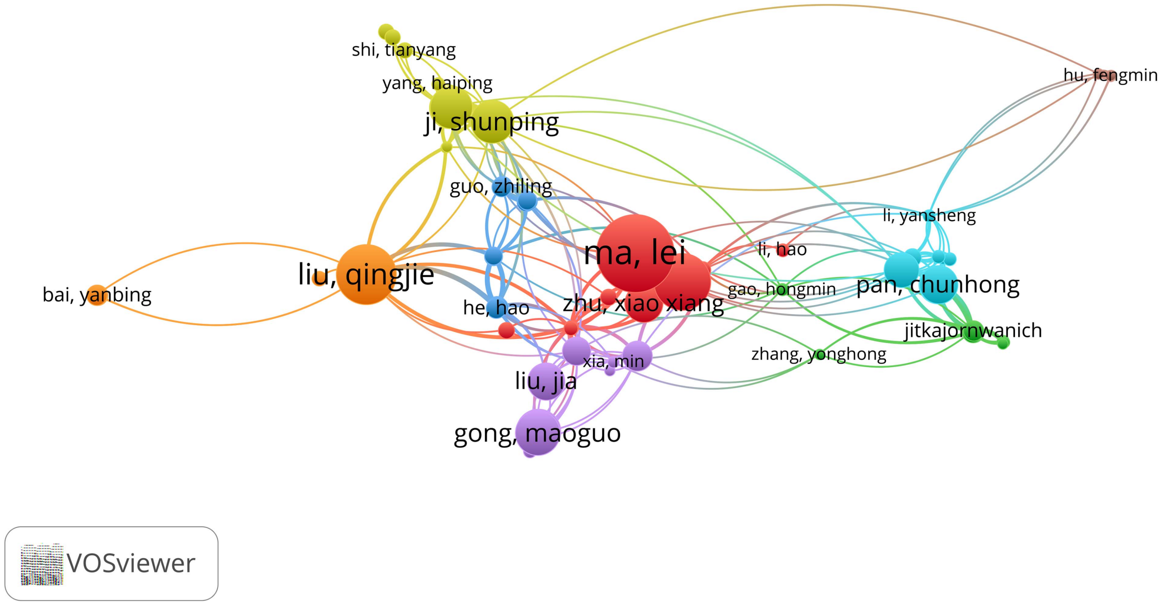

Fig. 5. Citation network of the influential authors. (The size of the circle indicates the citation number of the authors, and the thickness of the connected line indicates their relative citation strength.)

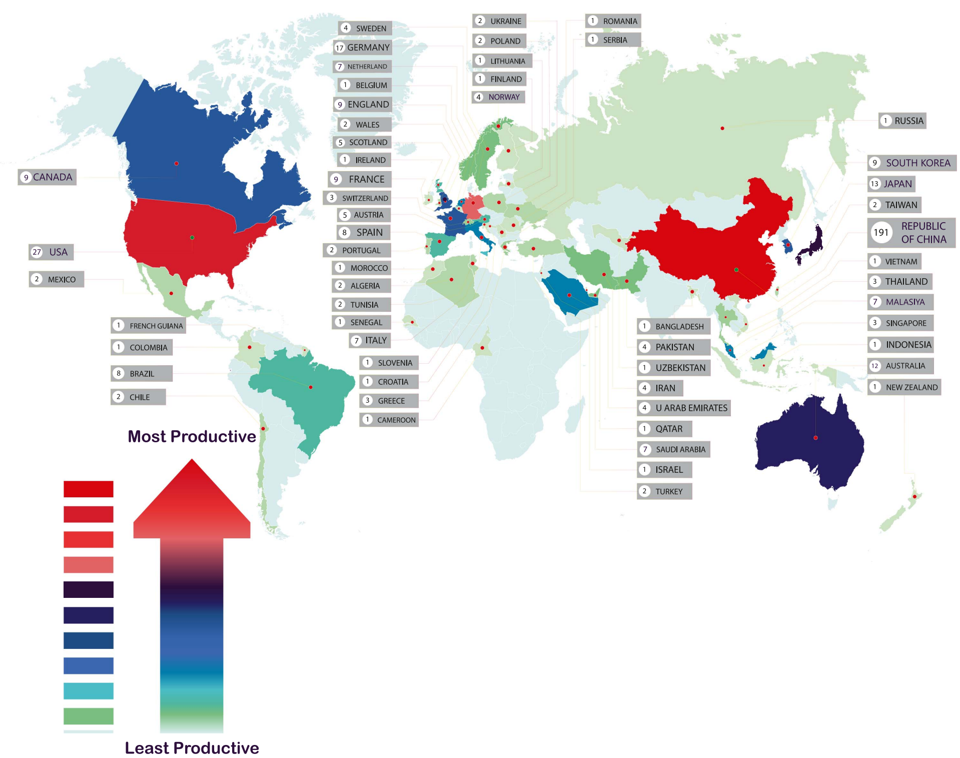

Fig. 6. Geographical distributions of the publications.

The top ten influential countries are summarized in Table VII by their name, full publications (TP), total citations (TC), average citation per year, the number of cited papers more than or equal to 30/20/10/5, and H-index. Moreover, for getting a better visualization of the productive countries, we provided a citation network of a few major countries in Fig. 7.

In the table of most productive countries, China is in the first position with 3400 citations for 191 publications. The H-index of China is 22, which is double compared to the second position holder country in the list. The USA claimed second place in the list with 27 publications and 1145 citations. The citation rate for the USA is 42.41. If we compare China and the USA, though

TABLE VII
SUMMARY OF THE MOST PRODUCTIVE COUNTRIES

Country	TP	TC	AC	≥30	≥20	≥10	≥5	H-index
China	191	3400	17.8	18	12	21	24	22
USA	27	1145	42.41	5	2	5	0	11
Germany	17	781	45.94	6	1	5	1	11
Japan	13	481	37	2	1	1	1	6
Australia	12	154	12.83	1	0	3	2	5
Canada	9	44	4.89	0	1	0	1	3
France	9	211	23.44	2	2	1	0	5
South Korea	9	48	5.33	0	0	2	1	4
England	9	32	3.56	0	0	1	2	3
India	8	33	4.13	0	1	0	1	2

^{TP}denotes total publications, ^{TC} denotes total citations, ^{AC} denotes average citations, and ≥30, ≥20, ≥10, and ≥5 denote citations more than or equal to 30, 20, 10, 5, respectively.

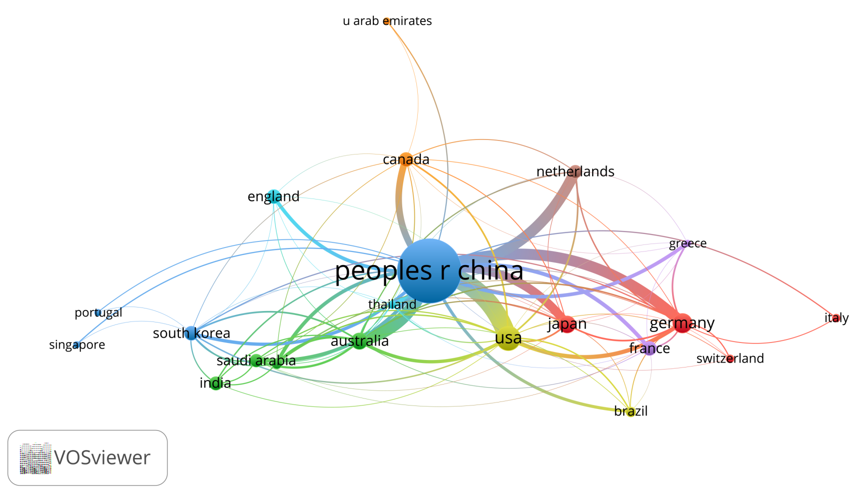

Fig. 7. Citation network of the influential countries. (The size of the circle indicates the citation number of the countries, and the thickness of the connected line indicates their relative citation strength.)

the number of articles and citations in China is far more than in the USA, the citation rate in the USA is more significant than two times that of China. Germany is in the third position with 17 publications. However, the citation rate for Germany (45.41) is the highest among all the countries. Its H-index (11) is the same as that for the USA. The following countries in the table are Japan and Australia with 13 and 12 publications, which denotes that the performances of these two countries are proceeding at an equal pace. The following four countries have an equal number of publications (9). However, Canada and South Korea gained 44 and 48 citations and 4.89 and 5.33 citation rates, respectively. France received 211 citations for nine publications, and its average citation is 23.44, which is the second-highest citation rate of the countries. It is even greater than China's rate, the top productive country with H-index-5. India is at the bottom of the list with eight publications and got 33 citations in this table. The rest of the countries not shown in the table also have outstanding contributions to DL based on semantic segmentation.

C. Science Mapping Analysis

1) Co-Citation Analysis: This work considered co-citation analysis as one of the techniques for science mapping analysis. Co-citation occurs when a pair of published articles, say x and y, are cited together in any other published document z. When the same pairs of papers are co-cited by many authors, clusters of RS research domains begin to form where each cluster has one common theme. With the help of single-link clustering and multidimensional scaling techniques, co-citation analysis can map the structure of different research areas [49]. In this section, this work will present a co-citation analysis based on cited sources and cited authors to show the relatedness between the journals and authors in terms of the research focus. Because if two sources or two authors are cited together, it is needless to say that they have a shared research interest.

This work set a threshold of at least 50 citations and found 34 sources among the 3052 sources that satisfied the threshold. Table VIII presents these sources with the total link strength of co-citation.

For better analysis of the sources, this article generated the scientific landscape of co-citation analysis of journals using the VOSviewer software. From Fig. 8 , it can be seen that the journals have been divided into a total of four clusters (pink, mint, violate, and black) with 561 links and total link strength of 190 489. Each node in Fig. 8 represents the corresponding journal. The bigger node has a higher weight. For the co-citation analysis, the total

TABLE VIII
CO-CITATION INDICES OF THE SOURCES

Cited Sources	Citations	Total link strength
IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING	999	46652
Remote Sensing	1203	44893
ISPRS *Journal of Photogrammetry and Remote Sensing*	773	37302
Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition	984	31176
Remote Sensing of Environment	516	24535
IEEE GEOSCIENCE AND REMOTE SENSING LETTERS	466	24430
IEEE JOURNAL OF SELECTED TOPICS IN APPLIED EARTH OBSERVATIONS AND REMOTE SENSING	383	21663
International Journal of Remote Sensing	390	17732
Lecture Notes in Computer Science	463	15812
IEEE TRANSACTIONS ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE	383	13236
International Conference on Computer Vision	262	9469
IEEE International Symposium on Geoscience and Remote Sensing	229	8512
International Journal of Applied Earth Observation and Geoinformation	115	8408
arXiv	205	7291
Photogrammetric Engineering and Remote Sensing	106	5673
Sensors	125	5473
IEEE ACCESS	132	5167
IEEE Geoscience and Remote Sensing Magazine	93	5049
Advances in Neural Information Processing Systems	129	4630
Journal of Applied Remote Sensing	77	4161
Neurocomputing	78	4107
Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition	94	3950
International Journal of Computer Vision	97	3798
Pattern Recognition	79	3794
IEEE TRANSACTIONS ON IMAGE PROCESSING	98	3426
Advances in Neurology	73	3065
PROCEEDINGS OF THE IEEE	64	3050
Remote Sensing Letters	54	2550
Proceedings of Machine Learning Research	64	2344
The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences	68	2256
Nature	59	2107
Conference on Computer Vision and Pattern Recognition (CVPR)	60	1891
ISPRS International Journal of Geo-Information	50	1871
Computers and Electronics in Agriculture	54	1505

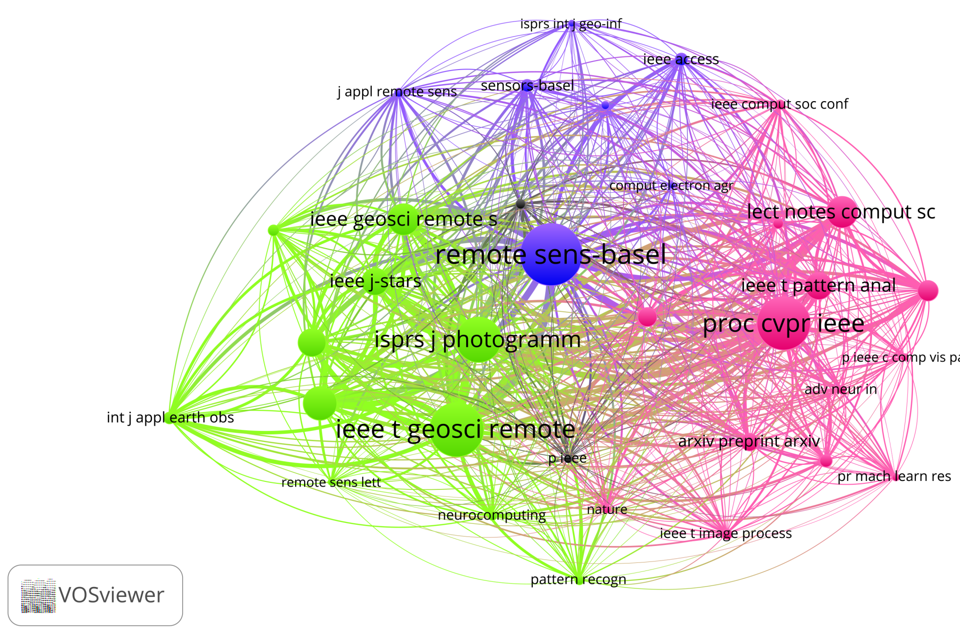

Fig. 8. Co-citation analysis of the sources. (The size of the circle indicates the total link strength of the publication sources, and the thickness of the connected line shows their relative strength.)

link strength has been selected as the weight by this work; that means the sources with higher weight have higher link strength. The connection line between two consecutive journals illustrates that these two sources have been cited in a publication. The frequency of the authors being cited together is proportional to the thickness of the line.

The interesting thing about the clusters is that all the sources of each cluster have been cited with the sources of all other clusters. From Fig. 8, it can be observed that the journals in the mint cluster have thicker links among themselves, which suggest similar interests among the journals. The prominent journal in the mint cluster and among all the clusters is "IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING," with 33 links and 46 652 total link strength. The journal has been cited with all the other journals; however, it has been cited maximum times (6159) along with the journal "*Remote Sensing*," which indicates the high relatedness of these two journals in the research field of DL-based semantic segmentation. "*Remote Sensing*" achieves the second position employing total link strength (44 893), placed in the violate cluster. Looking deeper into the clusters, we observe that the mint cluster has 11 sources (32%). The other influential sources in the mint cluster are "*ISPRS Journal of Photogrammetry and Remote Sensing*" (37 302), "*Remote Sensing of Environment*" (24 535), "IEEE GEOSCIENCE AND REMOTE SENSING LETTERS" (24 430), and "IEEE JOURNAL OF SELECTED TOPICS IN APPLIED EARTH OBSERVATIONS AND REMOTE SENSING" (21 663).

The pink cluster consists of 14 sources (41%), making it the largest cluster. The top influential source in this cluster is "Proceedings CVPR IEEE," with a total link strength of 31 176. This source has been cited together with the journals from other clusters such as "*Remote Sensing*" (3919, violate), "IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING" (3695, mint), and "*ISPRS Journal of Photogrammetry and Remote Sensing*" (3276, mint). This shows the highly similar research interests among the journals of different clusters. The other prominent journals in this cluster are: "Lecture Notes in Computer Science" (15 812), "IEEE TRANSACTIONS ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE" (13 236), and "IEEE International Conference on Computer Vision" (9469).

The violate cluster consists of seven sources (21%), among which "*Remote Sensing*" takes the lead with a total link strength of 44 893. It is also the most-cited journal (1203 citations), with thick links connecting the journal to all other clusters, making it highly influential in DL-based semantic segmentation research. "*Sensors*" (5473), "IEEE ACCESS" (5167), "*Journal of Applied Remote Sensing*" (4146), etc., are some of the other journals in this cluster. The black cluster is the most underpopulated of them all, having only two members: "*IEEE Geoscience and Remote Sensing Magazine*" (5049) and "PROCEEDINGS OF THE IEEE" (3050). The link strength between these two journals is also low (42). The sources of the black cluster have been cited more times with the sources of pink, mint, and violate clusters than the sources of the black cluster. It implies that though the sources of the black cluster have relatedness based on their research topic over the years, these sources have not been cited too many times together.

TABLE IX
CO-CITATION INDICES OF THE AUTHORS

Cited Authors	Citations	Total link strength
Liang-Chu Chen	156	1491
J Long	154	1346
Olaf Ronneberger	157	1221
Nicolas Audebert	87	1088
Vijay Badrinarayanan	118	1079
He Km	110	1004
Karen Simonyan	100	952
Matteo Maggiori	78	883
Alex Krizhevsky	93	861
Dimitrios Marmanis	67	812
Yann LeCun	75	669
Michele Volpi	41	581
Wz Zhao	46	574
Kaiming He	59	553
Volodymyr Mnih	53	551
Sergey Ioffe	54	513
Tung-Yen Lin	54	501
Ronald Kemker	42	482
Ross Girshick	44	458
Chendi Zhang	47	446
Christian Szegedy	44	445
Gordon Cheng	51	434
X Huang	48	432
Kaiming He	43	414
Durk Kingma	50	407
L. P. Zhang	40	391
Lei Ma	49	390
Zixuan Zhang	40	328

This work set a threshold of at least 40 citations. After analyzing the co-citation network of cited sources, this work analyzes the co-citation network of cited authors. A threshold of at least 40 citations was set to analyze further in-depth. Consequently, among the 6390 authors, 28 authors satisfied the threshold. Table IX presents these cited authors with the total link strength of co-citation.The co-citation network of the cited authors was generated using VOSviewer for deeper analysis (see Fig. 9). From Fig. 9, it can be seen that the authors have been divided into a total of three clusters (orange, azure, and brown) with 378 links and 9653 total link strength. Like the cited sources, in the case of cited authors, the total link strength has been selected as the weights of the nodes. If there is a connection between two authors, it indicates that the authors have been cited together in any other publications. The thicker the line, the more frequently the authors cited together. The leading author with the highest total link strength (1491) is "LC Chen" from the brown cluster. The author has been cited the maximum number of times (126) with "J Long," another member of the brown cluster, indicating similarity in their research topics. LC Chen also gets cited many times, along with authors like "Olaf Ronneberger" (119), "V Badrinarayanan" (108) from the brown cluster, and "N Audebert" (96) from the orange cluster, as well as "Kangmin He" (95) from the azure cluster. The brown cluster consists of eight authors in total. The other influential authors employing link strength are: "J Long" (1346), "O Ronneberger" (1221), "V Badrinarayanan"(1079), "V Mnih" (551), "S Ioffe" (513), "D Kingma" (407), and "Zx Zhang" (328). The orange cluster has 12 authors (43%), making it the largest cluster. The most prominent author in this cluster is "N Audebert" (1088).

2) Co-Authorship Analysis: Collaboration in research is essential to produce creative ideas and implement them more accessible and more sophisticatedly. One individual can find it too challenging to complete a research task. This work considered co-authorship analysis as one of the techniques for science mapping analysis. Co-authorship is a form of a coalition in which

TABLE X
CO-AUTHORSHIP INDICES OF THE COUNTRIES

Countries	Documents	Citations	Total link strength	Cluster
China	191	3374	40	Violet
USA	26	1139	17	Cyan
Germany	17	776	15	Brown
Japan	13	478	13	Violet
Australia	12	153	12	Cyan
The Netherlands	7	258	10	Brown
Saudi Arabia	7	41	8	Cyan
France	9	210	7	Brown
South Korea	9	48	7	Cyan
Italy	3	76	6	Brown
Canada	9	41	5	Violet
Switzerland	3	52	4	Brown
Singapore	2	274	3	Brown

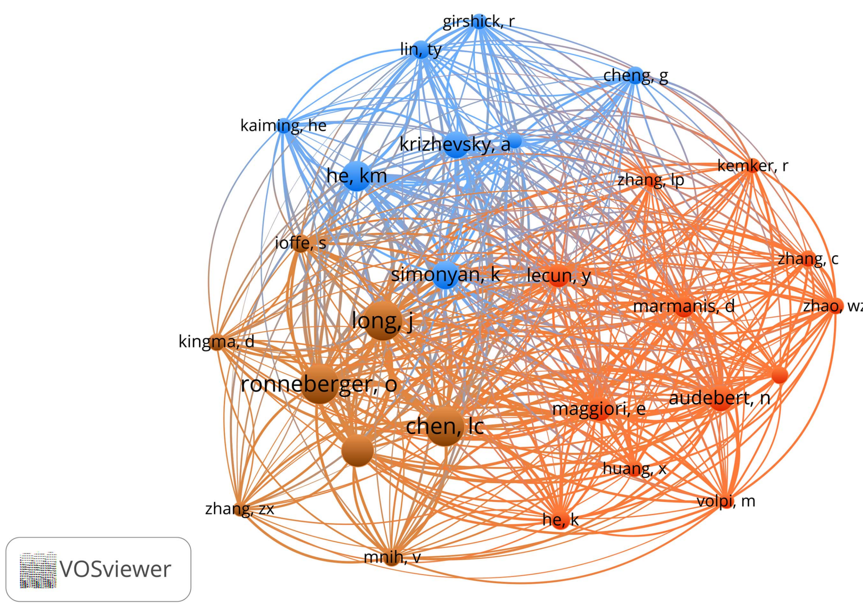

Fig. 9. Co-citation analysis of the authors. (The size of the circle indicates the total link strength of the authors, and the thickness of the connected line shows their relative strength.)

two or more researchers publish their works together on some topics. It is another technique that has been used in this work as bibliometric content. This work will present a co-authorship analysis by considering countries and organizations to show the collaboration pattern and relatedness among the authors of different countries and institutions in terms of the research focus. In the case of the co-authorship analysis of the countries, this work set the threshold of the maximum number of collaborating countries per document to ten as well as a minimum of two documents and 40 citations per country and found 14 countries among the 51 countries which satisfied the threshold. Due to poor link strength, Thailand was further omitted from this list of countries. Table X presents the countries with the total link strength.

To better analyze the scientific landscape of co-authorship analysis of countries, Fig. 10 was generated using the VOSviewer software. From Fig. 10, it can be seen that 13 countries have been divided into a total of three clusters (brown, cyan, and violet) with 32 links and 73 total link strengths, where the brown cluster holds the majority number of countries (six items). Germany (total link strength 15) and The Netherlands (total link strength 10) are the prominent members of this cluster. However, the most influential country among all employing total link strength (40) is China, a member of the violet cluster that only has three members. Therefore, despite being the smallest cluster, it holds a more significant position. The cyan cluster has four members, including the USA (total link strength 15)

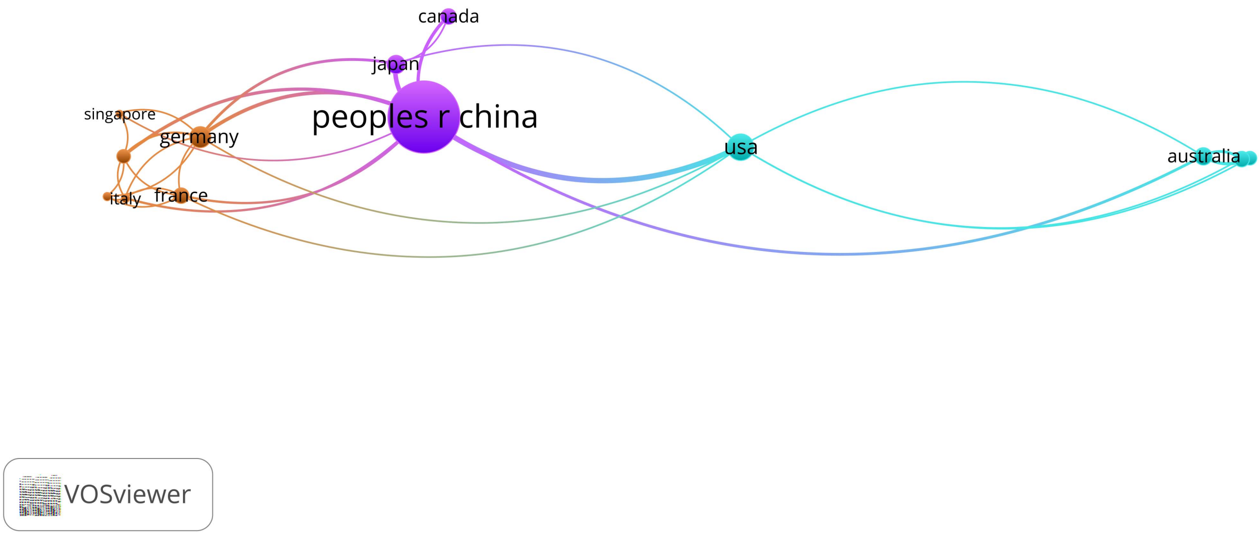

Fig. 10. Co-authorship analysis of the countries. (The size of the circle indicates the total link strength of countries, and the thickness of the connected line shows their relative strength.)

TABLE XI
CO-AUTHORSHIP INDICES OF THE TOP TEN ORGANIZATIONS

Organization	Documents	Citations	Total link strength	Cluster
Chinese Academy of Science	34	373	29	Violet
University of Chinese Academy of Science	21	277	23	Violet
Wuhan University	31	447	11	Cyan
Technical University of Munich	6	497	8	Brown
China University of Geosciences	8	160	6	Gold
German Aerospace Center	4	80	6	Brown
Southwest Jiaotong University	5	403	6	Brown
Nanjing University	4	744	5	Brown
German Aerospace Center (DLR)	4	559	4	Brown
Tokyo University	6	41	4	Gold

and Australia (total link strength 15). The different clusters of Fig. 10 each represent a comparatively similar research focus; the link strength among members like China and the USA (link strength 11) represents their collaboration frequency. This work selected organizations having at least three papers and 40 citations to analyze the research collaboration patterns of different organizations. This work found 14 organizations among the 422 organizations that satisfied the threshold. Table XI presents the top ten of these organizations with the total strength as well as the number of documents of different organizations.

For better analysis of the scientific landscape of co-authorship analysis of organizations, Fig. 11 was generated using the VOSviewer software. From Fig. 11, it can be seen that the 14 organizations have been divided into a total of five clusters (brown, cyan, violet, pink, and gold) with 23 links and 54 total link strengths. The brown and cyan clusters consist of four organizations each, and they reside in two corners of Fig. 11, signifying their vast differences in research focus. Top universities like the Technical University of Munich (total link strength 8) and Southwest Jiaotong University (total link strength 6) are members of the brown cluster. The most prominent organization employing citations (744), Nanjing University, is also a member of the brown cluster. On the other hand, one of the prominent members of these organizations, Wuhan University (total link strength 11), is a member of the cyan cluster. The most influential among these organizations through total link strength is the Chinese Academy of Science, which belongs to the violet cluster. Its high link strength (29) suggests the large number of research collaborations conducted by this institution. The gold cluster consists of only two members: the German Aerospace Center and Tokyo University. Finally, the sole member of the pink cluster is the China University of Geosciences.

3) Co-Occurrence Analysis of Keywords: This work considered co-occurrence analysis of keywords as one of the techniques for science mapping analysis. Co-occurrence is a correlation

TABLE XII
SUMMARY OF THE TOP TEN KEYWORDS

Rank Order	Keyword	Occurrences	Total link strength	Cluster
1	Deep Learning	115	480	Brown
2	Semantic Segmentation	89	390	Violet
3	Classification	82	393	Brown
4	Remote Sensing	77	420	Cyan
5	Image Segmentation	51	299	Cyan
6	Segmentation	39	176	Pink
7	Extraction	30	149	Violet
8	Semantics	28	207	Cyan
9	Feature Extraction	27	206	Cyan
10	Convolutional Neural Network	22	80	Gold

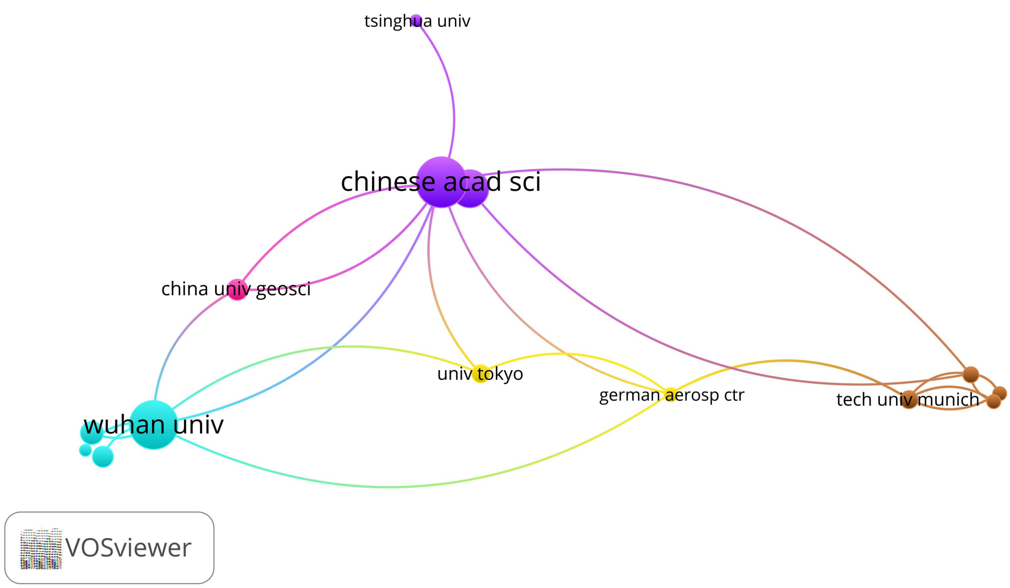

Fig. 11. Co-authorship analysis of the organizations. (The size of the circle indicates the total link strength of the organizations, and the thickness of the connected line shows their relative strength.)

in which a thing (i.e., a keyword) occurs with something else (i.e., another keyword). Here, a keyword refers to any noun or phrase that reflects or highlights the core concept of work. This work considered keywords that have occurred at least five times, and among the 1077 keywords, 89 were selected. To further analyze the co-occurrence of keywords, a scientific landscape was generated using VOSviewer software (see Fig. 12). The top ten of these keywords are listed in Table XII, utilizing total link strength. Fig. 12(a) depicts the 89 keywords that have been divided into a total of five clusters (brown, cyan, violet, pink, and gold) with 1319 links and 3045 total link strength. The occurrences of each selected keyword are presented by the weight of the circle, where the more prominent the circle, the more a keyword has co-occurred. Also, the distance between any two keywords demonstrates their relative strength and similarity in research topics. Fig. 12(b) depicts the most recent keyword occurrences among these keywords. The colors of keywords shown in Fig. 12(b) represent the time-varying keyword occurrences from 2019 (in dark blue) to 2021 (in yellow).

1) *Satellite-based technologies (Cluster 1)*: The leading keyword with the highest number of occurrences (115) and highest total link strength (480) is "deep learning" from the brown cluster. This keyword has 85 links, making it linked with almost all other keywords, suggesting its very high relation with the subject field. Keywords from the brown cluster such as "Deep learning" (115 occurrences, total link strength 480), "Classification" (82 occurrences, total link strength 393), "Convolutional neural network" (22 occurrences, total link strength 80), "Landsat 8" (six occurrences), "Sentinel-2" (seven occurrences),

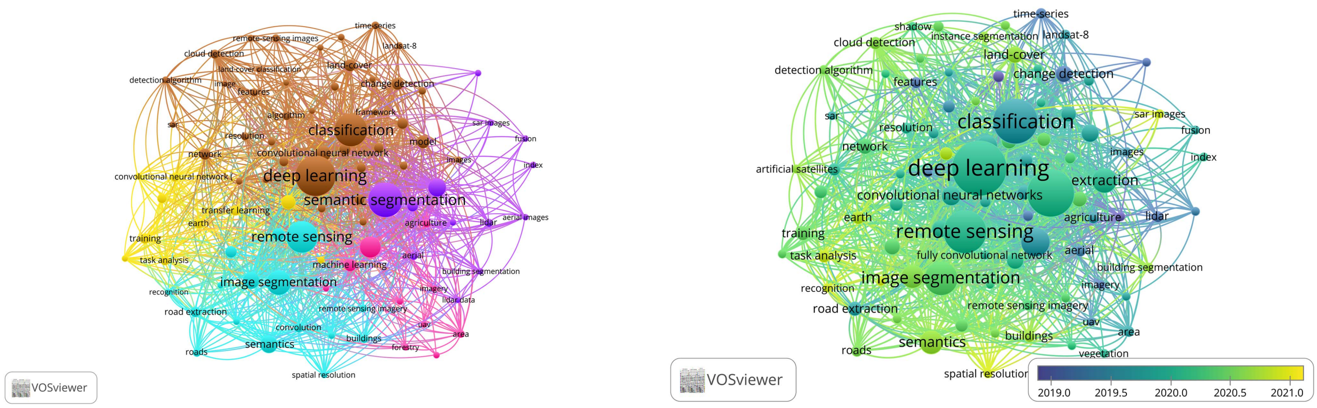

Fig. 12. Overview of keywords. (a) Co-occurrence analysis of the keywords. (The circles denote the number of occurrences of the keywords, and the thickness denotes the number of co-occurrence between the keywords.) (b) Co-occurrence analysis of the keywords shown in the scale of average publication year.

"Algorithm" (nine occurrences), "Remote sensing images" (seven occurrences), etc., suggest that the cluster focuses on "Deep-learning-based remote sensing technologies using satellite."

2) *Urban planning technologies (Cluster 2)*: In the cyan cluster (lower-left corner, 16 items), the leading keyword with the highest number of occurrences (77) and highest total link strength (420) is "Remote sensing" with total links of 78. Notable other keywords such as "Image segmentation" (51 occurrences, total link strength 299), "Semantics" (28 occurrences, total link strength 207) "Road extraction" (12 occurrences), "Building extraction" (16 occurrences), "Road" (eight occurrences), "Building" (eight occurrences), etc., focus on the primary domain of "Application of remote sensing technologies in the field of urban planning."

3) *Aerial image segmentation technologies (Cluster 3)*: In the violet cluster (middle right corner, 13 items), the most influential keyword with the highest number of occurrences (89) and highest total link strength (390) is "Semantic segmentation" with 80 links connecting this keyword with other keywords. Other essential keywords are "Extraction" (30 occurrences, total link strength 149), "Aerial" (eight occurrences), "Aerial images" (five occurrences), "Lidar" (eight occurrences), etc. By analyzing the keywords, the cluster items can be associated with the topic of "Application of semantic segmentation in remote sensing aerial technologies."

4) *Environment/crop detection technologies (Cluster 4)*: The leading keyword of the pink cluster (lower right corner, ten items) is "Segmentation" with the highest number of occurrences (39) and highest total link strength (176), which has 65 links connecting the keyword with other keywords. The keywords in this cluster like "Machine learning" (16 occurrences), "Agriculture" (eight occurrences), "Forest" (seven occurrences), "Forestry" (six occurrences), etc., suggest its association with the term: "Remote sensing-based environment, plant diversity, and crop monitoring."

5) *Ground/surface analysis technologies (Cluster 5)*: In the gold cluster, the keyword "Convolutional neural networks" holds a key position with 45 links (30 occurrences, total link strength 149) connecting the keyword with other important keywords. Some of the domain-specific important keywords are "Transfer learning" (11 occurrences), "earth" (ten occurrences), and "artificial satellites" (five occurrences), etc., which corresponds to the focus of "Aerial image-based remote sensing ground analysis technologies."

Overall, the five main clusters each represent relatively similar topics or a subfield of a field of RS, and by analyzing the central node circle, each of the clusters these similarities can be noticed.

V. Critical Analysis

By analyzing the influential papers in this field, this work would like to show how different segmentation applications of RS technologies are being improved using the help of DL techniques. As discussed in the methodology, after a laborious selection process, 28 papers were selected by the authors. These papers were further divided into nine groups based on RS applications, i.e., building extractions, change detection, road, slum, crop, or different land cover detection, etc. The authors then analyzed these papers based on their problem statement, methodologies, and results. A summary of these papers based on their dividing groups is given as follows.

A. Building Extraction

Ji et al. [21] presented a DL model named scale robust fully convolutional neural network (SR FCN). The model incorporated VGG-16 backbone architecture to identify building structures from RS imagery enhancing the segmentation result of building data from high-resolution aerial images. The authors proposed a network model which includes two atrous convolutions to be applied on the 1/4 and 1/8 scales. The uniqueness of this approach is that they applied atrous convolution on the concatenated multilayer features and thereby successfully scaled the network. They trained, validated, and tested their model using an aerial image dataset called the "WHU Building dataset" [41] consisting of approximately

187 000 buildings. Ji et al. [41] mentioned that the model showed a 94.4% precision and a 93.9% recall rate.

Ji et al. [41] introduced a novel U-Net backbone-based structure which was further modified by adding a siamese neural network (SiU-Net). The Siamese network creates a two-channel map from the original image tile and its downsampled counterpart, which corresponds to the two-channel labels. The resultant labels are then used for training the modified U-Net architecture. The model showed 93.8% precision and 93.9% recall rate. Ji et al. also employed radiometric augmentation on the images before training, which enhanced the prediction result of satellite images. The authors compared their methods with a few other classification techniques (such as U-Net) and showed that their model was more accurate than others. In another related work [40], Wu et al. introduced a modified U-net that assimilated residual neural network (ResNet) as the backbone (Res-U-Net). In this work, Res-U-Net was used to perform the classification of RS images with very high resolution. The feature extraction was done in four stages, each containing several residual blocks. The authors used the guided filter to fine-tune the achieved features by DL in this work. The authors trained, validated, and tested their model using the open-source Potsdam and Vaihingen [66] datasets and showed 97.71% (Vaihingen) and 96.91% (Potsdam) accuracy, respectively, which are moderately better than a few other classification techniques (including FCN, CNN+RF, and SegNet) comparatively.

Yi et al. [50] introduced DeepResUnet, a U-Net-based deep residual network comprising two subnetworks. A cascade downsampling and an upsampling network network were designed to reduce the learning parameters and reconstruct those features, respectively. In this work, the upsampling network used four upsampling layers instead of the max-pooling layer used in the downsampling network to reconstruct the feature maps. The authors trained and tested their model using open-source aerial images of Christchurch City, New Zealand [67], in which approximately 25 454 buildings were present. Yi et al. compared their model against six different existing DL models while obtaining a 0.9364 F1-score and. 9176 Kappa score. The numerical results proved their model performs better than all six models, especially U-Net. The only drawback of their approach is its poor time complexity.

B. Change Detection

Peng et al. [47] proposed a semantic segmentation technology, which is a U-Net++-based encoder–decoder architecture for change detection in earth surface imagery. The Multiple Side-Output Fusion strategy was used for deep supervision to capture finer spatial details. The open-source satellite image-pair datasets of different timelines and the same location obtained by Google Earth [68] were used to train, test, and evaluate the proposed model and obtained 89.54% precision and 87.11% recall rate, respectively, which is the best compared to the six different DL approaches mentioned in this work. However, it should be noted that in this work, the seasonal change was ignored, and the implementation of radiometric corrections, a necessary step in traditional change detection methods, was omitted from the design of the model as it was deemed unnecessary by the authors.

Gong et al. [48] designed a change detection framework based on DL for multispectral image pairs. The proposed method was hybrid in nature. The constitution of the model could primarily be divided into two parts. The core part is an unsupervised multispectral change detection framework for feature extraction and difference representation learning. It was noted as the main focus of this work. Another part of the work entailed a semisupervised classifier for the final processing and evaluation stage. Satellite images from four different regions taken by GF-1 [69], [76] and WorldView-2 [70], [76] satellites were used to train, test, and validate the model. The authors transformed the change detection task into a typical classification one through NN. They then used semisupervised-difference-representation-learning-based DL models, which incorporated stacked denoising autoencoders (SDAE) to solve the problem. In the validation part, the proposed methods were compared and evaluated together, which showed the strength of the novel framework and demonstrated its effectiveness.

In another related work, Gong et al. [42] introduced a unique approach to detect changes in multispatial images. This work incorporated a hybrid method for feature learning and change analysis. An unsupervised multilayered stacked autoencoders (SDAE) structure was implemented for multilevel feature learning of four different image datasets (two SAR and two SAT images). To evaluate results, both object-based (OBCD) and pixel-based change detection methods were examined, and OBCD was selected as a preliminary change map, which later helped in robust model construction, and different DL algorithms applied to validate this work confirmed this.

Yue et al. [51] proposed a DL-based image fusion network (IFN) to analyze the land surface change. The proposed IFN network utilizes a combination of bitemporal images for raw data image feature extraction. A difference discrimination network (DDN)-based backbone architecture was then used to analyze change detection on images from two different datasets. The model's performance was evaluated against DL-based change detection methods where IFN gains were measured using evaluation matrices such as F1-score (0.6733), precision (67.11%), recall (67.54%), and overall accuracy (88.86%).

Mou et al. [57] analyzed the results of the "2016 Data Fusion Contest" hosted by the "Image Analysis and Data Fusion Technical Committee" of the "IEEE Geoscience and Remote Sensing Society." The authors went over the methodology of the first two winning teams, their problem-solving methodologies, and the experimental validation results in great detail. The VHR pictures and video obtained from space by the Demos-2 satellite, which covers an urban and harbor area of the Vancouver, Canada region, were included in the open-source dataset of the data fusion challenge 2016 [71]. This paper discussed in detail the numerical assessment of the change detection results, which demonstrated the merits of the proposed methods.

Liebel and Körner [55] presented a CNN-architecture-based DL method called super-resolution CNN (SRCNN). The method was proposed for scaling the RS imagery from satellite images. The proposed network consisted of three inner convolutional layers of different kernel sizes (9 × 9, 1 × 1, and 5 × 5 px, respectively). The authors trained, validated, and tested their model using datasets that consisted of publicly available "Sentinel-2" multispectral satellite images and showed the ability to successfully scale the single-band satellite images from the multispectral satellite images [68], [76]. Researchers used different state-of-the-art evaluation matrices like structural similarity (SSIM) and PSNR to evaluate scaling results for different test datasets. They claimed to successfully scale the Sentinel-2 single-band images.

C. Scene Classification

Diakogiannis et al. [46] proposed "ResUNet-a," a novel DL architecture that is based on the U-Net basic architecture to verify the application of DL technologies on RS applications such as semantic segmentation. The proposed model consisted of residual blocks in which atrous convolution was employed and pyramid scene parsing pooling. Researchers also proposed a dice-loss-based function, which was termed "Tanimoto." The ISPRS Potsdam dataset [66] was used to classify different classes. The background, building, car, impervious surfaces, low vegetation, and tree were objects of classification in this experiment. Researchers evaluated the performances of different models based on ResUNet-a architectures and later compared them against eight different DL models to show the robustness of the model. One ResUNet-a-architecture-based model obtained an average F1-score of 92.9% and overall accuracy of 91.5%.

Wang et al. [43] discussed information entropy and the label error in the DL model feature maps and proposed a gated convolutional neural network (GCNN) method in RS imagery. The basic building blocks of GCN architecture consist of ResNet-101 as the backbone architecture for extracting features in the encoder part, followed by residual convolution modules (RCMs) as basic processing units, and lastly, entropy control module for feature fusion in the decoder, which is again followed by RCMs before obtaining the final prediction output. The softmax layer was used in the entropy control module (ECM) module as the classifier for the given feature map. The open-source ISPRS 2-D semantic labeling contest dataset [66] was used for the performance evaluation of the proposed method. Researchers used the method to again classify six different classes (background, building, car, impervious surfaces, low vegetation, and tree), which is similar to the experiment done by Diakogiannis et al. [46]. The mean F1-score of 85.2% and overall accuracy of 87.9% were obtained, which proved to be more effective than several other mainstream DL models such as FCN-8s, SegNet, Deeplab-v2, and RefineNet.

Fu et al. [39] introduced an FCN-based classification method for RS image classification. The authors employed a pretrained VGG-16 backbone network consisting of five downsampling layers for increasing convergence speed. Researchers later used atrous convolution to create dense feature maps. The model was trained and evaluated using the GF-2 [69] and IKONOS imagery datasets [76], and the mean precision score of 81%, mean recall of 78%, and mean Kappa score of 83% were obtained, which proved to be more effective than several other DL models such as SVM, patch-based CNN, and even FCN-8s.

Panboonyuen et al. [65] proposed an approach to improve the global convolutional network (GCN) for multiobject segmentation of aerial and satellite images by focusing on mainly three aspects: by varying ResNet backbones, by applying a channel attention block for weighing influential features, and by employing transfer learning (TL). The experimental models were trained and validated using the Landsat-8 satellite imagery dataset supplied by the Thailand government and the publicly available aerial imagery dataset from the ISPRS Vaihingen Challenge [66]. While being trained using the Landsat-8 datasets [76], the proposed system obtained an F1-score of 82.75% and mIoU of 71.78%. On the other hand, by using the ISPRS datasets [66], the proposed system obtained an F1-score of 79.42% and a mean IoU of 91.23%. The final result showed that the enhanced GCN model surpassed the benchmark encoder–decoder model by 17.48% and 2.48% on the Landsat-8 corpus and ISPRS corpus datasets, respectively.

D. Ice Wedge Polygon Detection

Zhang et al. [54] proposed an experimental study where they used a DL-based object instance segmentation method to precisely mark and classify ice wedge polygons in aerial orthoimagery. The authors used Mask R-CNN, a CNN-based model, to detect imagery objects and generate an instance segmentation mask. The building blocks of Mask R-CNN are ResNet as backbone, feature pyramid network, region of interest (RoI) generating network, RoI classifier, boundary box regressor, and FCN. The dataset used in this experiment consists of Nuiqsut [72], a Northern Alaska town area. The researchers detected approximately 79% of ice wedge polygons in study sites. The VHSR imagery pixel resolution was 0.15 m. The imagery had a pixel resolution of 0.25 m in this work.

E. Land Cover Classification/Crop Type Detection

Kussul et al. [36] experimented on procedures to determine DL methods to be applied to images to classify different land areas and types of crops from satellite imagery. The DL methods (1-D and 2-D CNNs) that had been discussed in this article were compared against RF and ENN. 1-D and 2-D CNN methods discussed in this work each had five CNNs in the ensemble. Each of them consisted of two convolution layers and max-pooling layers. In addition, two fully connected layers were added at the end. Datasets consisting of Landsat-8 and Sentinel-1 satellite images of the Kyiv region of Ukraine from different time frames [76], [76] were secured to be used in this work, and overall classification accuracies achieved for the ensemble of 1-D and 2-D CNNs were 93.5%, and 94.6%, respectively. Meanwhile, the overall accuracy observed for RF and ENN methods was 88.7% and 92.7%, respectively. Major crops such as wheat, maize, sunflower, soybeans, and sugar beet and land covers such as forest, grassland, and bare land were subjects of classification in this work.

Zhang et al. [53] proposed an experiment that mainly focused on dl methods such as stacked denoising autoencoder (SDA) or stacked autoencoder (SAE)-based classification approach for object-based land cover image classification. Two datasets that consisted of aerial images from Anhui and Tianjin provinces of China and contained three spectral bands, representing the red, green, and blue bands separately, were used in this experiment. The overall accuracy of 72% (SAE) and 73.5% (SDA) was obtained, which indicated an increase in performance of around 6% when compared to other ML methods such as Bayes, KNN, and different SVM models. This result shows the necessity of shifting from ML to DL methods.

Kattenborn et al. [52] presented a CNN-based DL model for spatial distribution and plant species segmentation from satellite or UAV imagery. The model had U-Net as the basic architecture. The datasets for two plant species, namely "*Pinus radiata*" and "*Ulex europaeus*," were obtained. They were further processed to extract the orthoimagery data using the Agisoft Photoscan. The program implemented Structure from Motion to generate the output and later obtain the image tiles. The CNN model was trained using the obtained image tiles. The authors claimed to have at least 84% accuracy using this approach but further validation and comparison among other DL methods are needed to prove the benefits of this approach.

F. Road Segmentation

Lan et al. [58] proposed a novel DL model global context-based dilated convolutional neural network (GC-DCNN) for road segmentation from RS imagery. The proposed method was mainly made up of three different parts: an encoder network, PPM, and a decoder network; a total of 21 convolution layers were used in the design of this method. The class imbalance issues were dealt with using the dice coefficient loss. In this work, the authors also presented two datasets; CasNet data composed of 224 different VHR images collected from Google Earth [68] and RoadTracer data [73], which contains 300 satellite images from 40 cities in six countries. The RoadTracer data were first used by Bastani et al. The dice loss function was used in the proposed model. To evaluate the work, four evaluation matrices (competence, correctness, quality score, and F1-score) were used. The proposed model was compared against six different CNNs (FCN, U-Net, etc.). The model needed a high number of model parameters and high runtime. Despite that, the results obtained by the proposed model showed 92.88% competence (COM), 95.40% correctness (COR), 88.87% quality score (Q), and 94.02% F1-score on the CasNet dataset. It clearly indicated the model's efficiency against several different DL baseline models. However, results from the RoadTracer dataset showed that the model had further room for improvement.

In [37], a DL NN-based model that combined residual learning and U-Net and was termed as DeepResUnet (a seven-level architecture) was proposed. The model was applied in semantic-segmentation-based road area extraction applications. The basic structure of the method could be divided into three parts, namely the encoding network, the bridge network, and the decoding network. A sigmoid operation was performed to obtain the final output. The Massachusetts roads dataset [74] was used to train, test, and evaluate the proposed model. The mean squared error (MSE) was used in this work as a loss function. Researchers observed the break-even point to be 91.87%, which is the best compared to the five different DL approaches (four CNN architecture based and one U-Net-based model) mentioned in this work.

In [63], DL models based on FCN architecture (FCN-8s) were examined to learn if this can be used for road segmentation applications. The first two FCN networks examined in this work had VGG-19 and ResNet backbone architecture, respectively. The DeepLabv3+ model had an Xception backbone architecture. The loss function used in this research was MSE. The researchers created their own dataset with the help of TerraSAR-X images [75], [76] and later identified and classified different types of road data (i.e., major roads, country roads, dirt paths, etc.) using Google Earth [68]. Three regions were selected (Lincoln, Kalisz, and Bonn) in the dataset to create and evaluate the FCN-8s model. The VGG-19-based model in particular obtained an IoU score of 45.46%, a precision score of 71.69%, and a recall score of 75.17% in the Lincoln area. Researchers later compared the model against U-Net where it outperformed in all regions. Against the DeepLabv3+, in the Bonn area, the model performed better (IoU 42.57%).

He et al. [64] proposed "ASPP-Integrated Encoder–Decoder Network," a novel DL architecture for road extraction. The U-Net architecture was used as a basic method. An atrous spatial pyramid pooling (ASPP) network processed the encoder-generated feature maps, which were then used as input for the decoder to generate the segmentation map. SSIM was used as the loss function and an evaluation method in this research. Here, researchers used the Massachusetts roads dataset [74] to train, test, and evaluate the proposed model. The achieved results were 83.5% F1-score and 0.893 SSIM value. Compared to the normal U-net, the proposed model showed 2.6% F1-score improvement and 0.18 SSIM improvement.

G. Ship Detection

Tang et al. [38] propose SDA-ELM, a DL-based approach that was modeled using a DL network and the extreme learning machine (ELM) algorithm. A SDA was used as the primary method. The model was built to develop a compressed-domain framework for ship detection. 4000 SPOT 5 panchromatic images [76] of 5-m resolution and 2000 $\times$ 2000 size (in pixels) were prepared to build the image dataset to compare the performances of four different DL-based models. Among them, 1600 training samples were extracted for feature learning. The research methodology was that the discrete wavelet transform was performed on the preprocessed image data to generate two different subband images (low- and high-frequency images). The bands were separately trained using two SDAs to obtain different band features. After that, ELM was performed to combine the learned features and obtain the feature vector of 100 dimensions. Researchers obtained an accuracy of 94.57% (MF-SVM), 91.63% (SA-SVM), 96.45% (SDA-SVM), and 97.58%

(SDA-ELM). Among the overall time cost of different methods, SDA-ELM achieved the lowest score of 2.6844 s.

Cheng et al. [62] proposed FusionNet. It is an edge-aware convolutional network based on encoder–decoder architecture for segmentation to parse an RS harbor image. The basic network architecture employed in this work is "SegNet." In addition, an edge network was introduced, which contained concatenation, convolution, and softmax layers to help generate the final prediction maps. The edge network was designed to work in parallel with the "SegNet" network to extract and combine the low- and high-frequency features. The images were parsed mainly into three objects, such as sea, land, and ship. In this work, a novel loss function based on a conditional random field model and termed LossFusionNet has been introduced as the edge-aware regularization. Two datasets, one containing mainly "Harbor images" and another one containing "Sea-Land images," were created, which contained a total of 520 images from Google Earth [68] to train, test, and validate the models. Researchers achieved an average F1-score of 97.43% and 99.36% using these datasets, respectively. Compared with the results of other ship detection methods, the proposed model's F1-score of 93.77% is a major improvement, which clearly indicates the usefulness of the proposed model.

H. Urban Regions Detection

Wurm et al. [45] analyzed TL capabilities of different FCN-based models to analyze slum mapping in various satellite images, mainly in QuickBird, Sentinel-2, and TerraSAR-X imagery datasets [68], [75], [76]. The VGG-19 backbone architecture-based model contained multiple convolution, ReLU activation, and upsampling layers. The input images had to be 224 $\times$ 224 pixels in size to train the models. The models were trained to identify four different classes namely: "Urban," "Vegetation," "Water," and "Slums." The results obtained by researchers showed that positive predictive value and sensitivity for QuickBird images [76] were up to 86% and 88%, respectively. For Sentinel-2 dataset [68], the results were observed up to 55% and 85% for positive predictive value and sensitivity, respectively. However, for TerraSAR-X image datasets [75], researchers found that the DL methods not only decreased the performance but also did not improve the results. Researchers found that the networks were unable to identify the urban structures from optical images and successfully apply the DL methods.

Albert et al. [60] analyzed the usage of CNN-based DL models to explore urban environments from satellite imagery. The DL models discussed in this work mainly consist of VGG-16 and ResNet-based backbone architectures. Researchers used the "Google Maps Static API" platform to obtain sat images and created a dataset consisting of approximately 140 000 images [77]. Information was collected from six different cities, namely Athens, Barcelona, Berlin, Budapest, Madrid, and Rome, to train and test this work. A subsection of this work contained training the model using one (or a set of) location datasets and testing using one or more different city datasets. The "Urban Atlas dataset" was used in this work to pretrain the classifier using ten different environment classes by analyzing the ground truth for land use distribution for three example cities (Budapest, Rome, and Barcelona) [77]. The analyzed results were used as references to develop the data sampling method for training data collection. In this research, results showed that the models could successfully classify green urban areas, densely populated urban areas, or water bodies. As airports, forests, agricultural lands, etc., tend to be similar in different cities, these can also be identified successfully. Despite this, the research showed that the models struggled to successfully identify the "high-," "medium-," and "low-"density urban environments.

I. Vehicle Detection

Audebert et al. [44] proposed a three-step segment-before-detect pipeline to perform vehicle classification in VHR RS data. Researchers at first implemented SegNet, a DL network based on VGG-16 backbone architecture for semantic segmentation. Basic SegNet architecture consists of the following layers: encoder, decoder, dense prediction, and softmax layers, to produce semantic maps. The ISPRS Potsdam [66], NZAM/ONERA [78], and Christchurch [67] data were used by researchers in this step to extract vehicles and filter small objects and create vehicle masks. At last, the ImageNet initialized CNN model was used for vehicle classification on the VEDAI dataset [79]. The researchers trained the CNN in four classes, namely, car, van, truck, and pick-up. Classification results obtained using the VEDAI dataset showed that the VGG-16-based architecture achieved approximately 89.7% overall accuracy. When applied to Potsdam and Christchurch data, the VGG-16-based classifier obtained 89% and 96% of overall accuracy, respectively. The authors claimed that the DL models used in this experiment achieved an average accuracy of 67% while applied on the ISPRS dataset and 80% while using the NZAM/ONERA dataset.

Mou and Zhu [59] proposed a boundary-aware ResNet (B-ResFCN) for semantic segmentation and vehicle detection from aerial imagery. Researchers used ResNet as the backbone architecture of the proposed model for its usage in vehicle instance segmentation tasks. The network outputs two homogeneous branches, which are used to create a segmentation mask and boundary map. The ISPRS Potsdam dataset [66] and a video dataset created by the researchers using a UAV were used in this research. To evaluate the results researchers used the Dice coefficient among other more common evaluation matrices such as F1-score, precision, and recall. The results were compared among six (UAV video data) to eight (ISPRS Potsdam) different FCN-based models to show the effectiveness of the model. Results showed that the instance-level Dice similarity coefficient varied from 77.5 to 79.39 for the UAV video data and from 77.72 to 93.80 in the case of the ISPRS Potsdam data. Researchers observed an increment in the Dice coefficient of 1.16% on the Potsdam dataset and 7.31% improvement on the UAV video dataset when compared to the ResFCN or ResNet-based model.

After summarizing the papers, some critical questions came forward such as the following.

Q1: Which DL method is used in an article?
Q2: Which backbone is used by the DL method?
Q3: Which framework is used to conduct the experiment?

Q4: Which datasets are used by an article?
Q5: Which loss function is taken into consideration by researchers?
Q6: Which optimizer is used in the experiment?
Q7: What performances are achieved by the researchers?

We summarized the findings of the questions in Table XIII.

VI. Findings, Key Challenges, and Future Research Direction

A. Findings of the Study

This study provides both the bibliometric analysis and the critical examination of several chosen papers within the field of DL-based semantic segmentation in RS and image processing. The bibliometric analysis aimed to identify research trends, influential authors, prominent journals, publications, countries, significant research terms, and collaboration patterns. We list our findings from the bibliometric analysis as follows.

1) The publication rate in the earlier years (before 2015) was too low. Research output in this area has increased since 2018. The number of published articles increased dramatically in 2020 and 2021 (117 papers were published in 2021).
2) China, the USA, and Germany are the top three countries with the most research output in the field.
3) IEEE Transactions on Geoscience and Remote Sensing, *Remote Sensing*, *ISPRS Journal of Photogrammetry and Remote Sensing*, and Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition are the most impactful and most cited journals.
4) Papers [5], [23], [36], [37], and [38] are the most cited papers in our data collection period.
5) The research organizations contributing most to the field are Chinese Academy of Science, Wuhan University, and Technical University of Munich.

In the critical section analysis of our work, we have categorized the papers according to their respective application domains. We present several significant findings as follows.

1) U-net was the most widely used DL backbone network, appearing in nine out of the 28 reviewed works. ResNet and VGG followed closely, each being utilized in eight works. The popularity of these networks indicates their effectiveness in semantic segmentation tasks in the RS domain.
2) TensorFlow emerged as the dominant choice among researchers, with a usage rate of 46.4%. MATLAB and Caffe were also utilized, although less frequently, showcasing their relevance in specific contexts.
3) Cross entropy, dice coefficient, and MSE loss functions were most popular among researchers. These loss functions play a vital role in training DL models and achieving accurate segmentation results.
4) Researchers predominantly employed Adam and stochastic gradient descent optimizer for model development. The Adam optimizer was favored slightly more, being used in 40% of the reviewed works.
5) Several researchers proposed modifications to the DL models to enhance their performance in various tasks. Some noteworthy approaches involve implementing a skip layer structure [64], adding gate mechanism [43], residual connections [46], ASPP [56], and designing multiscale network architecture [39].
6) Many works leveraged the existing DL models and experimented to analyze the TL capability of these models in various domains [45], [60], [65].

B. Key Challenges

In this section, we list several key challenges encountered within this research domain.

1) *Limited training data*: One of the major challenges in applying DL to RS is the availability of limited labeled training data. Obtaining high-quality labeled data for various RS applications can be time consuming and resource intensive.
2) *Class imbalance*: Imbalanced class distribution is a common issue in RS datasets, where certain classes may be underrepresented. This imbalance can lead to biased model performance and affect the accuracy of segmentation results.
3) *Standardization of evaluation criteria*: The use of diverse performance metrics by researchers highlights the need for standardized evaluation criteria within the field. This could enhance comparability and reliability in assessing model performances.
4) *Domain adaptation*: RS data may vary significantly across different regions, sensors, and environmental conditions. Adapting DL models to handle variations in data distributions and domain shifts is a challenge.
5) *Computational complexity*: DL models, particularly those with deep architectures, can be computationally intensive and require significant computational resources for training and inference. This can hinder the practical application of DL methods in RS tasks.
6) *Model interpretability*: DL models are often perceived as "black boxes" due to their complexity, making it challenging to interpret their decisions. It can be a concern, especially in critical applications such as plant species detection [52], change analysis [57], etc.

C. Future Research Direction

The previous sections listed the findings of our study and the key challenges encountered in this research domain. Moreover, along with these, we have also determined a set of research scopes and directions for future researchers by analyzing the extracted DL-based papers.

1) *Benchmarking studies*: Conducting comprehensive benchmarking studies is essential for evaluating the performance of different DL models and frameworks in the context of RS semantic segmentation. These studies should involve diverse datasets, encompassing various

TABLE XIII
SUMMARY OF DL TECHNIQUES FOR RS IMAGE ANALYSIS

Ref	Method	Backbone	Framework	Dataset	Application	Loss function	Optimizer	Performance (%)
[21]	SR FCN	VGG-16	TensorFlow	WHU Building dataset[69]	Building extraction	Cross Entropy	Adam	Satellite: Precision=79 %, Recall= 77 %, IoU=64 %; Aerial: Precision=94.4 %, Recall= 93.9 %, IoU=88.9 %
[44]	SiU-Net	Siamese U-Net	TensorFlow	WHU Building dataset[69]	Building extraction	-	Adam	Precision=93.8 %, Recall= 93.9 %, IoU=88.4 %
[43]	Res-U-Net	ResNet, U-Net	TensorFlow	ISPRS Potsdam and Vaihingen dataset [69]	Building extraction	Softmax cross entropy loss	Adam	OA=97.71% (Vaihingen), 96.91% (Potsdam)
[53]	DeepResUnet	U-Net, ResNet	Tensorflow	Christchurch City data	Building extraction	Cross entropy	Adam	F1 Score=93.64 %, Kappa score=91.76 %, Precision=94.01 %, Recall= 93.28 %, OA=97.09 %
[50]	Enhanced UNet++	UNet++	TensorFlow	Google Earth dataset	Change Detection	Binary cross entropy+Dice coefficient	Adam	F1 Score=87.56 %, Precision=89.54 %, Recall= 87.11 %, OA=96.73 %
[51]	SP+SDAE, SP+DBN	SDAE	Matlab	4 Imagery datasets takren by GF-1 and WorldView-2	Change Detection	Reconstruction error	-	-
[45]	MNN	SDAE	Matlab	2 SAR and 2 SAT images	Change Detection	reconstruction error	backpropagation algorithm and SGD	-
[54]	IFN	VGG-16, DDN	-	Google Earth dataset	Change Detection	Binary Cross-Entropy	-	F1 Score=67.33 %, Precision=67.11 %, Recall= 67.54 %, OA=88.86 %
[60]	-	VGG-16	-	DFC 2016 dataset	Earth Surface Change Detection	-	-	Average accuracy=90.5%, OA=96.5%, Kappa score= 0.9353
[58]	SRCNN	CNN	Caffe	Sentinel-2 dataset	Image Resolution Scaling	Euclidean loss	-	-
[49]	ResUNet-a	U-Net	MXNET	ISPRS Potsdam dataset [70]	Semantic Segmentation of Surface Objects	Tanimoto	Adam	F1 Score=92.9 %, OA=91.5 %
[46]	GSN	ResNet-101	Caffe	ISPRS 2D Semantic Labeling dataset [70]	Semantic Segmentation of Objects	Softmax Cross Entropy Loss	SGD	F1 Score=85.2 %, OA=87.9 %
[42]	CNN	VGG-16	-	GF-2 and IKONOS imagery datasets	Semantic Segmentation	Softmax Cross Entropy Loss	SGD	Kappa Score=83 %, Precision=81 %, Recall= 78 %
[68]	GCN	ResNet-50, ResNet-101, ResNet-152	Tensorflow-Slim	Landsat-8 and ISPRS Vaihingen imagery datasets	Semantic Segmentation of Aerial Images	cross entropy	-	-
[57]	Mask R-CNN	ResNet	Tensorflow	Nuiqsut imagery datasets	Aerial Image Object Instance Segmentation	Cross entropy loss, Smooth-L1 loss	-	-

TABLE XIII
(CONTINUED.)

Ref	Method	Backbone	Framework	Dataset	Application	Loss function	Optimizer	Performance (%)
[37]	1D and 2D CNNs	-	Tensorflow	Landsat-8 and Sentinel-1 satellite imagery datasets of Kyiv	Land and Crop Images Classification	Softmax cross entropy	AdaGrad	1D CNN OA=93.5 % , 2D CNN OA=93.5 %
[56]	SAE, SDAE	-	-	Anhui and Tianjin Aerial imagery datasets	Land Cover Classification	Mean squared error and Cross entropy	-	SAE OA=72 % , SDAE OA=73.5 %
[55]	CNN	U-Net	TensorFlow	4 UAV imagery datasets	Plant Segmentation and Classification	dice-coefficient	RMSprop	Accuracy=84 90 %
[61]	GC-DCNN	U-Net	Pytorch	CasNet and Roadtracer imagery datasets	Road Segmentation	dice-coefficient	Adam	COM=92.88 %, COR=95.40 %, Q Score=88.87 %, F1 Score=94.02 %
[38]	DeepResUnet	U-Net, ResNet	Keras	Massachusetts roads imagery datasets	Road extraction	Mean Squared Error	SGD	Breakeven point=91.87 %
[66]	FCN-8s, DeepLabv3+	VGG-19, Xception	Tensorflow	Created TerraSAR-X imagery datasets	Road Segmentation	Mean Squared Error	Adam	FCN-8s: IoU=45.46 % (Lincoln), 42.57 % (Bonn); DeepLabv3+: IoU=45.64 % (Lincoln), 40.91 % (Bonn)
[67]	ASPP-U-net-SSIM	U-Net	PyTorch	Massachusetts roads imagery datasets	Road Segmentation	SSIM, Cross entropy	Adam	F1 Score=83.5 %, Precision=87.1 %, Recall= 80.5 %
[51]	SDA-ELM	SDA	Matlab	SPOT 5 imagery datasets	Ship Detection	Traditional squared error+Cross entropy	SGD	Accuracy=97.58 %
[65]	FusionNet	-	Caffe	Harbor and Sea-Land imagery datasets	Ship Detection	Cross entropy loss, smooth loss, LossFusionNet	-	Average F1 score=97.43 %
[51]	FCNs	VGG-19	TensorFlow	QuickBird, Sentinel-2 and TerraSAR-X imagery datasets	Different Urban Regions Detection	Sparse softmax cross entropy loss	Adam	FCN-QuickBird: IoU: 77.02 %
[63]	CNN	ResNet-50, VGG-16	TensorFlow	Google Maps Static API imagery datasets	Different urban regions detection	Cross entropy	SGD+Adadelta	-
[47]	SegNet, CNN	VGG-16	Caffe	ISPRS Potsdam Dataset [70]	Vehicle Detection	-	SGD	-
[62]	B-ResFCN	ResNet	TensorFlow	ISPRS Potsdam Dataset [70]	Vehicle Detection	Sigmoid and Binary cross entropy loss	Nesterov Adam	OA=99.79 %, F1 score=93.44 %

RS applications and scene types. By comparing the accuracy, efficiency, and robustness of different models, researchers can identify best practices and gain insights into the strengths and weaknesses of various approaches. Standardized evaluation protocols and metrics are crucial to ensure the comparability of results and to establish a reference point for future advancements.

2) *Transfer learning*: Exploring the potential of TL offers a promising direction to leverage existing knowledge from related domains and apply it to RS tasks [2], [6], [45], [60], [65]. By pretraining DL models on large-scale datasets from domains like aerial images, satellite images, or ground label images for computer vision, researchers can capture generic features that are

transferrable to RS imagery. This process can significantly enhance model generalization and alleviate the data scarcity issue faced in RS. Fine-tuning the pretrained models on specific RS data allows for faster convergence during training and can lead to more accurate segmentation results.

3) *Uncertainty estimation*: Estimating uncertainty in DL-based semantic segmentation results is a crucial step toward building trust in the model outputs. Uncertainty measures provide insights into the confidence of the model's predictions, particularly in challenging or ambiguous scenarios. Some works give approximate estimations of uncertainty in the models [32]. However, most of the time, researchers fail to mention this in their work [17]. Researchers should investigate techniques to quantify uncertainty in segmentation maps. Understanding and visualizing uncertainty will be valuable for decision making in critical applications, where high confidence is required.
4) *Integration of contextual information*: Integrating contextual information into DL models can significantly improve the accuracy and robustness of semantic segmentation in RS. Spatial relationships, spectral characteristics, and temporal context play a vital role in interpreting RS data. Researchers should explore attention mechanisms and graph-based approaches to capture long-range dependencies and contextual cues effectively. By incorporating such information, DL models can better understand the complex relationships within the scene and produce more semantically meaningful segmentations.
5) *Domain-specific architectures*: Developing domain-specific DL architectures tailored to the unique characteristics of RS data is essential to unlock the full potential of semantic segmentation. RS data differ significantly from traditional images due to its high spatial resolution, multiple spectral bands, and the presence of various land cover types. Researchers modify the architectures that exploit these specific properties. Domain-specific architectures will allow DL models to better capture the intricacies of RS data and deliver more efficient and accurate segmentation outcomes.
6) *Explainability and interpretability*: Advancing research on interpretable DL models is critical to foster trust and confidence in the predictions made by the models. By enhancing model interpretability, researchers can understand the reasoning behind the model's decisions, making them more transparent and interpretable for end users and stakeholders. Techniques like attention maps, saliency analysis, and feature visualization can provide insights into the regions and features that influence the model's predictions.

VII. Discussions and Conclusion

In this article, the authors have provided a literature review of the recently published papers (2015–2021) which utilize DL-based semantic segmentation techniques to analyze remotely sensed images. DL has become a captivating part of the RS research field. DL is an essential consideration of modern RS research for making the RS image segmentation swifter and more precise. This domain has observed a significant surge in the prosperity curves of RS tasks. This study has conducted a bibliometric analysis of the prevailing papers published in the last five years to determine the current research practice in the RS field. Moreover, this study has identified potential future research avenues by highlighting the important analytical terms in this domain, allowing the following researchers to explore this research theme. In addition, this article has presented a thorough survey of a few scrutinized DL-based papers and also abridged many essential insights of the articles. Some intriguing research questions have been yielded as an annex to the main findings from the reviewed articles. This article has also supplied elegant responses to questions about the robustness and viability of an individual paper or DL technique in this research arena. Following the extracted outcomes of this article and preceding history, the authors of this work convoke at one point that modifying the benchmark DL models by integrating different modules is going to be a fascinating research topic in this field. Employing feasible outcomes and practical applications and modifying models can lead the DL practice in RS technologies in the near days. It will be a rational move for the researchers if they exaggerate their research to modify the DL models for better results in the RS field.

References

[1] S. Ghaffarian and M. Turker, "An improved cluster-based snake model for automatic agricultural field boundary extraction from high spatial resolution imagery," *Int. J. Remote Sens.*, vol. 40, no. 4, pp. 1217–1247, 2018.

[2] N. E. Zimmermann et al., "Modern remote sensing for environmental monitoring of landscape states and trajectories," in *A Changing World* (Landscape Series). New York, NY, USA: Springer, 2007, pp. 65–91.

[3] S. Ghaffarian and S. Emtehani, "Monitoring urban deprived areas with remote sensing and machine learning in case of disaster recovery," *Climate*, vol. 9, no. 4, 2021, Art. no. 58.

[4] M. M. Nielsen, "Remote Sensing for urban planning and management: The use of window-independent context segmentation to extract urban features in Stockholm," *Comput., Environ. Urban Syst.*, vol. 52, pp. 1–9, 2015.

[5] L. Ma, Y. Liu, X. Zhang, Y. Ye, G. Yin, and B. A. Johnson, "Deep learning in remote sensing applications: A meta-analysis and review," *ISPRS J. Photogrammetry Remote Sens.*, vol. 152, pp. 166–177, 2019.

[6] L. Luo, P. Li, and X. Yan, "Deep learning-based building extraction from remote sensing images: A comprehensive review," *Energies*, vol. 14, no. 23, 2021, Art. no. 7982.

[7] M. Alam, J.-F. Wang, C. Guangpei, L. V. Yunrong, and Y. Chen, "Convolutional neural network for the semantic segmentation of remote sensing images," *Mobile Netw. Appl.*, vol. 26, no. 1, pp. 200–215, 2021.

[8] R. S. Babu, B. Radhakrishnan, and L. P. Suresh, "Detection and extraction of roads from satellite images based on Laplacian of Gaussian operator," in *Proc. Int. Conf. Emerg. Technol.*, 2016, no. 1.

[9] M. A. Hearst, S. T. Dumais, E. Osuna, J. Platt, and B. Scholkopf, "Support vector machines," *IEEE Intell. Syst. Appl.*, vol. 13, no. 4, pp. 18–28, Jul./Aug. 1998, doi: 10.1109/5254.708428.

[10] L. Breiman, "Random forests," *Mach. Learn.*, vol. 45, pp. 5–32, Oct. 2001, doi: 10.1023/A:1010933404324.

[11] Q. Yuan et al., "Deep learning in environmental remote sensing: Achievements and challenges," *Remote Sens. Environ.*, vol. 241, 2020, Art. no. 111716.

[12] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," *Commun. ACM*, vol. 60, no. 6, pp. 84–90, 2017.

[13] L. Wu, P. Cui, J. Pei, L. Zhao, and L. Song, "Graph neural networks," *Graph Neural Netw.: Found., Front., Appl.*, pp. 27–37, 2022.

[14] X. Mao and Q. Li, "Generative adversarial networks (GANs)," in *Generative Adversarial Networks for Image Generation*. Singapore: Springer, 2021, pp. 1–7.

[15] H. Zhang, D. Roy, and V. Martins, "A novel deep convolutional neural network approach for large area satellite time series land cover classification," *Authorea*, Nov. 4, 2021. doi: 10.1002/essoar.10508610.1

[16] S. H. Khan, X. He, F. Porikli, and M. Bennamoun, "Forest change detection in incomplete satellite images with deep neural networks," *IEEE Trans. Geosci. Remote Sens.*, vol. 55, no. 9, pp. 5407–5423, Sep. 2017.

[17] G. Ros et al., "Semantic segmentation of urban scenes via domain adaptation of SYNTHIA," in *Domain Adaptation in Computer Vision Applications*. New York, NY, USA: Springer, 2017, pp. 227–241.

[18] Z. Du, J. Yang, C. Ou, and T. Zhang, "Smallholder crop area mapped with a semantic segmentation deep learning method," *Remote Sens.*, vol. 11, no. 7, 2019, Art. no. 888.

[19] V. Rashidian, L. G. Baise, and M. Koch, "Detecting collapsed buildings after a natural hazard on VHR optical satellite imagery using U-Net convolutional neural networks," in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, 2019, no. 9394.

[20] Z. Pan, J. Xu, Y. Guo, Y. Hu, and G. Wang, "Deep learning segmentation and classification for urban village using a worldview satellite image based on U-Net," *Remote Sens.*, vol. 12, no. 10, 2020, Art. no. 1574.

[21] S. Ji, S. Wei, and M. Lu, "A scale robust convolutional neural network for automatic building extraction from aerial and satellite imagery," *Int. J. Remote Sens.*, vol. 40, no. 9, pp. 3308–3322, 2018.

[22] Y. Bai, E. Mas, and S. Koshimura, "Towards operational satellite-based damage-mapping using U-Net convolutional network: A case study of 2011 Tohoku earthquake-tsunami," *Remote Sens.*, vol. 10, no. 10, 2018, Art. no. 1626.

[23] L. Ma, M. Li, X. Ma, L. Cheng, P. Du, and Y. Liu, "A review of supervised object-based land-cover image classification," *ISPRS J. Photogrammetry Remote Sens.*, vol. 130, pp. 277–293, 2017.

[24] X. Yuan, J. Shi, and L. Gu, "A review of deep learning methods for semantic segmentation of remote sensing imagery," *Expert Syst. Appl.*, vol. 169, 2021, Art. no. 114417.

[25] N. Kerle, F. Nex, M. Gerke, D. Duarte, and A. Vetrivel, "UAV-based structural damage mapping: A review," *ISPRS Int. J. Geo-Inf.*, vol. 9, no. 1, 2019, Art. no. 14.

[26] N. Zang, Y. Cao, Y. Wang, B. Huang, L. Zhang, and P. T. Mathiopoulos, "Land-use mapping for high-spatial resolution remote sensing image via deep learning: A review," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 14, pp. 5372–5391, 2021.

[27] M. Pashaei, H. Kamangir, M. J. Starek, and P. Tissot, "Review and evaluation of deep learning architectures for efficient land cover mapping with UAS hyper-spatial imagery: A case study over a wetland," *Remote Sens.*, vol. 12, no. 6, 2020, Art. no. 959.

[28] S. T. Yekeen and A.-L. Balogun, "Advances in remote sensing technology, machine learning and deep learning for marine oil spill detection, prediction and vulnerability assessment," *Remote Sens.*, vol. 12, no. 20, 2020, Art. no. 3416.

[29] B. Neupane, T. Horanont, and J. Aryal, "Deep learning-based semantic segmentation of urban features in satellite images: A review and meta-analysis," *Remote Sens.*, vol. 13, no. 4, 2021, Art. no. 808.

[30] S. Ghaffarian, J. Valente, M. Van Der Voort, and B. Tekinerdogan, "Effect of attention mechanism in deep learning-based remote sensing image processing: A systematic literature review," *Remote Sens.*, vol. 13, no. 15, 2021, Art. no. 2965.

[31] J. Liu, J. Xiang, Y. Jin, R. Liu, J. Yan, and L. Wang, "Boost precision agriculture with unmanned aerial vehicle remote sensing and edge intelligence: A survey," *Remote Sens.*, vol. 13, 2021, no. 21, Art. no. 4387.

[32] X. X. Zhu et al., "Deep learning in remote sensing: A comprehensive review and list of resources," *IEEE Geosci. Remote Sens. Mag.*, vol. 5, no. 4, pp. 8–36, Dec. 2017.

[33] M. Zhu, Y. He, and Q. He, "A review of researches on deep learning in remote sensing application," *Int. J. Geosci.*, vol. 10, no. 1, pp. 1–11, 2019.

[34] L. P. Osco et al., "A review on deep learning in UAV remote sensing," *Int. J. Appl. Earth Observ. Geoinf.*, vol. 102, 2021, Art. no. 102456.

[35] I. Kotaridis and M. Lazaridou, "Remote sensing image segmentation advances: A meta-analysis," *ISPRS J. Photogrammetry Remote Sens.*, vol. 173, pp. 309–322, 2021.

[36] N. Kussul, M. Lavreniuk, S. Skakun, and A. Shelestov, "Deep learning classification of land cover and crop types using remote sensing data," *IEEE Geosci. Remote Sens. Lett.*, vol. 14, no. 5, pp. 778–782, May 2017.

[37] Z. Zhang, Q. Liu, and Y. Wang, "Road extraction by deep residual U-Net," *IEEE Geosci. Remote Sens. Lett.*, vol. 15, no. 5, pp. 749–753, May 2018.

[38] J. Tang, C. Deng, G.-B. Huang, and B. Zhao, "Compressed-domain ship detection on spaceborne optical image using deep neural network and extreme learning machine," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 3, pp. 1174–1185, Mar. 2015.

[39] G. Fu, C. Liu, R. Zhou, T. Sun, and Q. Zhang, "Classification for high resolution remote sensing imagery using a fully convolutional network," *Remote Sens.*, vol. 9, no. 5, 2017, Art. no. 498.

[40] Y. Xu, L. Wu, Z. Xie, and Z. Chen, "Building extraction in very high resolution remote sensing imagery using deep learning and guided filters," *Remote Sens.*, vol. 10, no. 1, 2018, Art. no. 144.

[41] S. Ji, S. Wei, and M. Lu, "Fully convolutional networks for multisource building extraction from an open aerial and satellite imagery data set," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 1, pp. 574–586, Jan. 2019.

[42] P. Zhang, M. Gong, L. Su, J. Liu, and Z. Li, "Change detection based on deep feature representation and mapping transformation for multi-spatial-resolution remote sensing images," *ISPRS J. Photogrammetry Remote Sens.*, vol. 116, pp. 24–41, 2016.

[43] H. Wang, Y. Wang, Q. Zhang, S. Xiang, and C. Pan, "Gated convolutional neural network for semantic segmentation in high-resolution images," *Remote Sens.*, vol. 9, no. 5, 2017, Art. no. 446.

[44] N. Audebert, B. Le Saux, and S. Lefèvre, "Segment-before-detect: Vehicle detection and classification through semantic segmentation of aerial images," *Remote Sens.*, vol. 9, no. 4, 2017, Art. no. 368.

[45] M. Wurm, T. Stark, X. X. Zhu, M. Weigand, and H. Taubenböck, "Semantic segmentation of slums in satellite images using transfer learning on fully convolutional neural networks," *ISPRS J. Photogrammetry Remote Sens.*, vol. 150, pp. 59–69, 2019.

[46] F. I. Diakogiannis, F. Waldner, P. Caccetta, and C. Wu, "ResUNet-A: A deep learning framework for semantic segmentation of remotely sensed data," *ISPRS J. Photogrammetry Remote Sens.*, vol. 162, pp. 94–114, 2020.

[47] D. Peng, Y. Zhang, and H. Guan, "End-to-end change detection for high resolution satellite images using improved UNET++," *Remote Sens.*, vol. 11, no. 11, 2019, Art. no. 1382.

[48] M. Gong, T. Zhan, P. Zhang, and Q. Miao, "Superpixel-based difference representation learning for change detection in multispectral remote sensing images," *IEEE Trans. Geosci. Remote Sens.*, vol. 55, no. 5, pp. 2658–2673, May 2017.

[49] G. Surwase, A. Sagar, B. S. Kademani, and K. Bhanumurthy, "Co-citation analysis: An overview," in *Proc. Bosla Nat. Conf. Proc.*, 2011, pp. 179–185.

[50] Y. Yi, Z. Zhang, W. Zhang, C. Zhang, W. Li, and T. Zhao, "Semantic segmentation of urban buildings from VHR remote sensing imagery using a deep convolutional neural network," *Remote Sens.*, vol. 11, no. 15, 2019, Art. no. 1774.

[51] C. Zhang et al., "A deeply supervised image fusion network for change detection in high resolution bi-temporal remote sensing images," *ISPRS J. Photogrammetry Remote Sens.*, vol. 166, pp. 183–200, 2020.

[52] T. Kattenborn, J. Eichel, and F. E. Fassnacht, "Convolutional neural networks enable efficient, accurate and fine-grained segmentation of plant species and communities from high-resolution UAV imagery," *Sci. Rep.*, vol. 9, no. 1, 2019, Art. no. 17656.

[53] X. Zhang, G. Chen, W. Wang, Q. Wang, and F. Dai, "Object-based land-cover supervised classification for very-high-resolution UAV images using stacked denoising autoencoders," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 10, no. 7, pp. 3373–3385, Jul. 2017.

[54] W. Zhang, C. Witharana, A. Liljedahl, and M. Kanevskiy, "Deep convolutional neural networks for automated characterization of arctic ice-wedge polygons in very high spatial resolution aerial imagery," *Remote Sens.*, vol. 10, no. 9, 2018, Art. no. 1487.

[55] L. Liebel and M. Körner, "Single-image super resolution for multispectral remote sensing data using convolutional neural networks," *Int. Arch. Photogrammetry, Remote Sens. Spatial Inf. Sci.*, vol. XLI-B3, pp. 883–890, 2016.

[56] L. Mi and Z. Chen, "Superpixel-enhanced deep neural forest for remote sensing image semantic segmentation," *ISPRS J. Photogrammetry Remote Sens.*, vol. 159, pp. 140–152, 2020.

[57] L. Mou et al., "Multitemporal very high resolution from space: Outcome of the 2016 IEEE GRSS Data Fusion Contest," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 10, no. 8, pp. 3435–3447, Aug. 2017.

[58] M. Lan, Y. Zhang, L. Zhang, and B. Du, “Global context based automatic road segmentation via dilated convolutional neural network,” *Inf. Sci.*, vol. 535, pp. 156–171, 2020.
[59] L. Mou and X. X. Zhu, “Vehicle instance segmentation from aerial image and video using a multitask learning residual fully convolutional network,” *IEEE Trans. Geosci. Remote Sens.*, vol. 56, no. 11, pp. 6699–6711, Nov. 2018.
[60] A. Albert, J. Kaur, and M. C. Gonzalez, “Using convolutional networks and satellite imagery to identify patterns in urban environments at a large scale,” in *Proc. 23rd ACM SIGKDD Int. Conf. Knowl. Discov. Data Mining*, 2017, no. 1357.
[61] Y. Chen, R. Fan, X. Yang, J. Wang, and A. Latif, “Extraction of urban water bodies from high-resolution remote-sensing imagery using deep learning,” *Water*, vol. 10, no. 5, 2018, Art. no. 585.
[62] D. Cheng, G. Meng, S. Xiang, and C. Pan, “FusionNet: Edge aware deep convolutional networks for semantic segmentation of remote sensing harbor images,” *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 10, no. 12, pp. 5769–5783, Dec. 2017.
[63] C. Henry, S. M. Azimi, and N. Merkle, “Road segmentation in SAR satellite images with deep fully convolutional neural networks,” *IEEE Geosci. Remote Sens. Lett.*, vol. 15, no. 12, pp. 1867–1871, Dec. 2018.
[64] H. He, D. Yang, S. Wang, S. Wang, and Y. Li, “Road extraction by using atrous spatial pyramid pooling integrated encoder-decoder network and structural similarity loss,” *Remote Sens.*, vol. 11, no. 9, 2019, Art. no. 1015.
[65] T. Panboonyuen, K. Jitkajornwanich, S. Lawawirojwong, P. Srestasathiern, and P. Vateekul, “Semantic segmentation on remotely sensed images using an enhanced global convolutional network with channel attention and domain specific transfer learning,” *Remote Sens.*, vol. 11, no. 1, 2019, Art. no. 83.
[66] F. Rottensteiner et al., “The ISPRS benchmark on urban object classification and 3D building reconstruction,” *ISPRS Ann. Photogrammetry, Remote Sens. Spatial Inf. Sci.*, vol. I-3, pp. 293–298, 2012.
[67] “Christchurch city building footprints—Christchurch city building footprints—Data.govt.nz—Discover and use data.” Accessed: Jul. 8, 2022. [Online]. Available: https://catalogue.data.govt.nz/dataset/christchurch-city-building-footprints/resource/b77d1dfb-d310-47ce-a244-f09067ad1a73
[68] “Sentinel-2 datasets in Earth engine—Earth Engine Data catalog—Google developers,” Google. Accessed: Jul. 8, 2022. [Online]. Available: https://developers.google.com/earth-engine/datasets/catalog/sentinel-2
[69] S. Zhou, “Dataset of GF-1 satellite images (2017-2018),” National Tibetan Plateau Data Center. [Online]. Available: https://sjynp.tpdc.ac.cn/en/data/af865f0d-c0ff-4e3b-80a2-591a4f5adeeb/
[70] E. S. Agency, “Worldview-2,” Earth Online. Accessed: Jul. 8, 2022. [Online]. Available: https://earth.esa.int/eogateway/missions/worldview-2
[71] D. Tuia, G. Moser, B. L. Saux, R. Fabrizi, and S. Cowan, “Data Fusion Contest 2016 (DFC2016),” IEEE Dataport, Oct. 25, 2019, doi: 10.21227/1ky7-0r79.
[72] “Nuiqsut: LiDAR—North slope science catalog,” Dataset—North Slope Science Catalog, Sep. 29, 2017. Accessed: Jul. 8, 2022. [Online]. Available: https://catalog.northslopescience.org/sv/dataset/2591
[73] F. Bastani and S. Madden, “Beyond road extraction: A dataset for map update using aerial images,” in *Proc. IEEE/CVF Int. Conf. Comput. Vis.*, 2021, no. 11905.
[74] V. Mnih, “Machine learning for aerial image labeling,” Ph.D. dissertation, Dept. Comput. Sci., Univ. Toronto, Toronto, ON, Canada, 2013.
[75] “TerraSAR-X ESA archive,” *TerraSAR-X ESA archive—Earth Online*. Accessed: Jul. 8, 2022. [Online]. Available: https://earth.esa.int/eogateway/catalog/terrasar-x-esa-archive
[76] “Satellite images, satellite map” Satellite Imaging Corp. Accessed: Jul. 8, 2022. [Online]. Available: https://www.satimagingcorp.com/gallery/high-resolution/
[77] Google Maps Platform Documentation—Maps Static API—Google Developers. Accessed: Jul. 8, 2022. [Online]. Available: https://developers.google.com/maps/documentation/maps-static
[78] R.C. Daudt, B. Le Saux, A. Boulch, and Y. Gousseau, “Urban change detection for multispectral earth observation using convolutional neural networks,” in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, 2018 pp. 2115–2118.
[79] S. Razakarivony and F. Jurie, “Vehicle detection in aerial imagery: A small target detection benchmark,” *J. Vis. Commun. Image Represent.*, vol. 34, pp. 187–203, 2016.

Kazi Rakib Hasan received the B.Sc. degree in electronics and communication engineering from Khulna University, Khulna, Bangladesh, in 2021. He is currently working toward the Integrated Ph.D. degree in biomedical science with Kyungpook National University, Daegu, South Korea.

He was actively involved in research during his undergraduate studies under the guidance of Dr. Md. Abdul Awal. He is passionate about artificial intelligence, deep learning, and robotics. His research interests include exploring the applications of these cutting-edge engineering technologies within the realm of biomedical sciences.

Anamika Biswas Tuli received the B.Sc. degree in electronics and communication engineering from Khulna University, Khulna, Bangladesh, in 2023.

In 2023, she joined the Nanobiophotonics Lab, Pusan National University, Busan, South Korea. She is the author of one conference paper. She will continue her research in plasmon-based optical imaging and optical sensing techniques and in the development of optical medical devices. Her research interests include photonics, machine learning, and remote sensing.

Md. Al-Masrur Khan was born in Tangail, Bangladesh, in 1997. He received the B.S. degree in electronics and communication engineering from Khulna University, Khulna, Bangladesh, in 2020, and the M.S. degree in smart city engineering from the Department of ICT Integrated Ocean Smart Cities Engineering, Dong-A University, Busan, South Korea, in 2023.

He was a Research Assistant with the Advanced Multiphysics Structural Condition Evaluation Laboratory, Dong-A University, under the supervision of Dr. Seong-Hoon Kee. In 2023, he joined the Vehicle Autonomy and Intelligence Lab, Bloomington, IN, USA. He is interested in visual perception for navigating mobile robots. He is the author of six journals and two conference papers. His research interests include deep learning, semantic segmentation, mobile robots, and self-driving vehicles.

Mr. Khan was awarded the Brain Korea 21 (BK21) scholarship by the Korean Government.

Seong-Hoon Kee received the B.S. degree in architectural engineering and the M.S. degree in structural engineering from Hanyang University, Seoul, South Korea, in 2001 and 2005, respectively, and the Ph.D. degree in structural engineering from the Department of Civil, Architectural and Environmental Engineering, University of Texas, Austin, TX, USA, in 2011.

From 2011 to 2014, he was a Research Associate with Rutgers University, New Brunswick, NJ, USA. He is currently an Associate Professor and Chair of the Department of Architectural Engineering, and an Associate Director with the National Research Center for Disaster-free and Safe Ocean City Construction, Dong-A University, Busan, South, Korea, where he is also the Director of the Advanced Multiphysics Structural Condition Evaluation Laboratory. He is involved in the supervision of four Ph.D. students and eight master's degree students. He provides research leadership in embedded devices, sensors, and robotic vehicles, as well as instruments that incorporate sensing, actuation, control, wireless transmission, automated data acquisition and analysis, machine learning, deep learning for nondestructive testing, and structural health monitoring of concrete structures. He has authored more than 85 publications, including journals and conference papers. He has been involved in several research grants and has managed projects and delivered research solutions to various Korean agencies.

Dr. Kee has received several awards, including the ACI-James Instruments Student Award for Research on NDT of Concrete, for meritorious and innovative research in concrete and concrete materials using NDT methods, and to publicize the field of NDT and the work of ACI Committee 228 in 2011.

Md Abdus Samad (Member, IEEE) received the Ph.D. degree in information and communication engineering from Chosun University, Gwangju, South Korea, in 2022.

From 2013 to 2017, he was an Assistant Professor with the Department of Electronics and Telecommunication Engineering, International Islamic University Chittagong, Chattogram, Bangladesh. He is currently a Research Professor with the Department of Information and Communication Engineering, Yeungnam University, Gyeongsan, South Korea. His research interests include signal processing, antenna design, electromagnetic wave propagation, applications of artificial neural networks and deep learning, and millimeter-wave propagation by interference or atmospheric causes for 5G and beyond wireless networks.

Dr. Samad received the Prestigious Korean Government Scholarship for his doctoral study.

Abdullah-Al Nahid received the B.Sc. degree in electronics and communication engineering from Khulna University, Khulna, Bangladesh, in 2007, the M.Sc. degree in telecommunication engineering from the Institute for Telecommunication Research, University of South Australia, Mawson Lakes, SA, Australia, in 2012, and the Ph.D. degree in machine learning from Macquarie University, Sydney, NSW, Australia, in 2018.

He is currently with the Electronics and Communication Engineering Discipline, Khulna University. His research interests include machine learning, biomedical image processing, data classification, and smart grid.