Semantic Segmentation of Remote Sensing Images: Definition, Methods, Datasets and Applications

Vlatko Spasev(✉), Ivica Dimitrovski, Ivan Kitanovski, and Ivan Chorbev

Faculty of Computer Science and Engineering, Ss Cyril and Methodiuos University, Skopje, North Macedonia
{vlatko.spasev,ivica.dimitrovski,ivan.kitanovski, ivan.chorbev}@finki.ukim.mk

Abstract. Semantic segmentation of remote sensing images is a vital task in the field of remote sensing and computer vision. The goal is to produce a dense pixel-wise segmentation map of an image, where a specific class is assigned to each pixel, enabling detailed analysis and understanding of the Earth's surface. This paper provides an overview of semantic segmentation in remote sensing, starting with a definition of the task and its significance in extracting valuable information from remote sensing imagery. Various methods used for semantic segmentation in remote sensing are discussed, including traditional approaches such as region-based and pixel-based methods, as well as more recent deep learning-based techniques. Next, the paper delves into the available datasets for semantic segmentation of remote sensing images. Many available datasets are reviewed, highlighting their characteristics, including the number of images, image size, number of labels, spatial resolution, format and spectral bands. These datasets serve as valuable resources for training, evaluating, and benchmarking semantic segmentation algorithms in remote sensing applications. Furthermore, the paper highlights the broad range of applications enabled by semantic segmentation in remote sensing, including urban planning, land cover mapping, disaster management, environmental monitoring, and precision agriculture. Overall, this paper serves as a comprehensive guide to semantic segmentation of remote sensing images, providing insights into its definition, methods, available datasets and wide-ranging applications.

Keywords: Remote Sensing Images · Semantic Segmentation · Deep Learning · Earth Observation

1 Introduction

Remote sensing images refer to images captured from a distance by sensors or instruments mounted on satellites, aircraft, drones, or other platforms. These images are used to collect information about the Earth's surface, atmosphere,

Supported by Faculty of Computer Science and Engineering, Skopje, N. Macedonia.

M. Mihova and M. Jovanov (Eds.): ICT Innovations 2023, CCIS 1991, pp. 127–140, 2024.
https://doi.org/10.1007/978-3-031-54321-0_9

and other objects or phenomena without direct physical contact [44]. In recent years, the advent of sophisticated machine learning techniques, coupled with the abundance of remote sensing data, has paved the way for significant advancements in image analysis and interpretation [10,18].

The concept of semantic segmentation has made substantial strides [21]. Its application to remote sensing imagery spans various domains and has been a prominent research area for decades [47]. Operating at the forefront of computer vision, semantic segmentation equips machines with the capability to intricately understand and demarcate image content down to individual pixels. Unlike traditional object detection methods that label entire objects or regions within an image, semantic segmentation meticulously labels each pixel according to its associated object or class. This finer granularity of analysis endows us with a deeper understanding of the intricate spatial distribution of features within remote sensing images.

The implications of semantic segmentation within the realm of remote sensing are vast and profound. It finds application in a multitude of fields, including urban planning, agriculture, environmental monitoring, disaster management, forestry, and more [31]. Semantic segmentation enables the automated extraction of vital information from imagery, unraveling patterns and changes that might otherwise elude human perception. By unraveling the complex tapestry of pixels, semantic segmentation unveils insights that drive informed decision-making and facilitate holistic comprehension of the Earth's ever-evolving landscapes.

This paper makes significant contributions to the understanding and advancement of semantic segmentation within the context of remote sensing imagery. It provides a comprehensive and cohesive overview of the concept of semantic segmentation in the domain of remote sensing. It serves as an accessible introduction for both novice and seasoned researchers, offering a clear understanding of the underlying principles and significance of semantic segmentation in interpreting remote sensing data. A significant contribution of the paper lies in its exploration of various methods and techniques employed in semantic segmentation for remote sensing images. The paper conducts a thorough examination of datasets used in training and evaluating semantic segmentation models for remote sensing imagery. Highlighting various applications, the paper demonstrates the real-world implications of semantic segmentation within the realm of remote sensing.

In the reminder, we first explain the main characteristics of the remote sensing image data and then define the task of semantic segmentation, outlining the input data and the desired output. Various machine learning methods used for semantic segmentation are discussed, along with evaluation measures to assess the performance of trained models. The paper further summarizes and highlights different datasets available for training and evaluating semantic segmentation models for remote sensing data. Lastly, the paper explores the potential applications of semantic segmentation in the context of remote sensing data. It discusses how semantic segmentation can be employed in various domains such as urban planning, disaster management, environmental monitoring, precision agriculture, deforestation analysis, climate assessment, and water resource management. These applications showcase the broad range of benefits that semantic segmentation offers in understanding and analyzing remote sensing imagery.

2 Characteristics of Remote Sensing Image Data

Remote sensing images can be broadly categorized into two main types: aerial images and satellite images. Satellite images and aerial images are both valuable sources of remote sensing data, but they differ in how they are acquired and the characteristics of the imagery. Remote sensing data encompasses various aspects of information representation, including spectral, spatial, radiometric, and temporal resolutions. Spectral resolution involves the bandwidth and sampling rate employed for data capture. High spectral resolution signifies narrower bands of the spectrum, while low resolution indicates broader bands. These spectral bands span diverse wavelengths such as ultraviolet, visible, near-infrared, infrared, and microwave. Image sensors range from multi-spectral, covering numerous bands (e.g., Sentinel-2[1] with 12 bands), to hyper-spectral sensors like Hyperion (part of the EO-1 satellite), gathering thousands of spectral bands (0.4–2.5 μm) [34].

Spatial resolution refers to the Earth's surface area represented by each pixel in an image. Higher spatial resolutions (small pixel size) capture finer details, whereas lower resolutions (large pixel size) retain fewer details. Moderate Resolution Imaging Spectroradiometer (MODIS), for instance, observes most bands with a spatial resolution of 1 km, where each pixel signifies a 1 km $\times$ 1 km ground area [19]. Conversely, UAV-captured images can achieve highest spatial resolutions, even less than 1 cm pixel size [33].

Radiometric resolution defines the sensor's capability to record signals of varying strengths (dynamic range). A larger dynamic range enables the detection of intricate details in recordings. Landsat 7 records 8-bit images, discerning 256 distinct gray values of reflected energy, while Sentinel-2 boasts a 12-bit radiometric resolution (4095 gray values). Enhanced radiometric resolution facilitates the differentiation of subtle variations in ocean color, crucial for water quality assessment.

Temporal resolution denotes how often a satellite revisits a specific observation area. Polar-orbiting satellites exhibit varying temporal resolutions, ranging from 1 to 16 days (e.g., ten days for Sentinel-2). Temporal considerations are pivotal in monitoring changes within observation areas, encompassing aspects like land use alteration, deforestation, and mowing.

Satellite images are captured by sensors mounted on satellites orbiting the Earth. These satellites can be classified into different types, including optical, radar, and thermal satellites [44]. Satellites, with varying altitudes and predetermined orbits, capture images across large expanses at regular intervals. Ranging from a few square kilometers to entire continents, satellite images offer a global perspective, crucial for monitoring extensive phenomena and long-term changes.

[1] https://sentinel.esa.int/web/sentinel/missions.

Spatial resolution varies by sensor and platform. High-resolution satellites unveil details down to meters, while lower-resolution ones provide a broader view of Earth's surface. Temporal resolution hinges on satellite revisit periods, spanning days to weeks or months based on specifics. Atmospheric conditions like cloud cover and haze influence image quality, although some sensors counter these effects, while others, like radar sensors, remain unaffected.

Aerial images are captured from platforms that are closer to the Earth's surface, such as airplanes, helicopters, or drones [33]. These platforms are equipped with cameras or other sensors, allowing for the acquisition of images at specific locations and altitudes. Aerial images offer localized coverage and can be acquired over targeted areas of interest. They are particularly useful for capturing detailed imagery of specific regions, such as cities, construction sites, or natural landscapes. Aerial images generally have higher spatial resolution compared to satellite images. They can capture fine details, objects, and features with greater clarity and precision. The spatial resolution of aerial images can range from centimeters to a few meters, depending on the sensor and flight parameters. Aerial images can be acquired on-demand, allowing for more frequent revisit times compared to satellites. The temporal resolution of aerial images depends on factors like flight scheduling and availability of aircraft or drones. Aerial images are less affected by atmospheric conditions compared to satellite images. Being closer to the Earth's surface, they are captured under relatively clearer atmospheric conditions, resulting in improved image quality and reduced atmospheric interference.

Both satellite and aerial images have their advantages and are used in various applications [31]. Satellite images provide a global perspective and long-term monitoring capabilities, while aerial images offer higher spatial resolution and localized coverage for detailed analysis of specific areas of interest. The choice between satellite and aerial imagery depends on the specific requirements of the application, the desired level of detail, and the availability of data.

3 Definition of Semantic Segmentation

Semantic segmentation tasks focus on labeling each pixel of an image with a corresponding class of what the pixel represents. The goal is to produce a dense pixel-wise segmentation map of an image, where specific class is assigned to each pixel. The tasks of image semantic segmentation aim at the fine-grained identification of objects in an image. In contrast to object detection, which aims at coarser localization of the detected objects. Recently, more sophisticated extensions of the semantic segmentation task, referred as instance segmentation [15] and panoptic segmentation [20] have emerged. Instance segmentation takes into account different semantic types and focuses on delineating multiple objects present in an image. On the basis of instance segmentation, panoptic segmentation needs to detect and segment all objects in the image, including the background.

Semantic segmentation of remote sensing images is a fundamental task in the field of remote sensing and computer vision. The goal is to partition the image into meaningful regions, enabling detailed analysis and understanding of the Earth's surface. Figure 1 illustrates an example of semantic segmentation of remote sensing images in the context of buildings extraction.

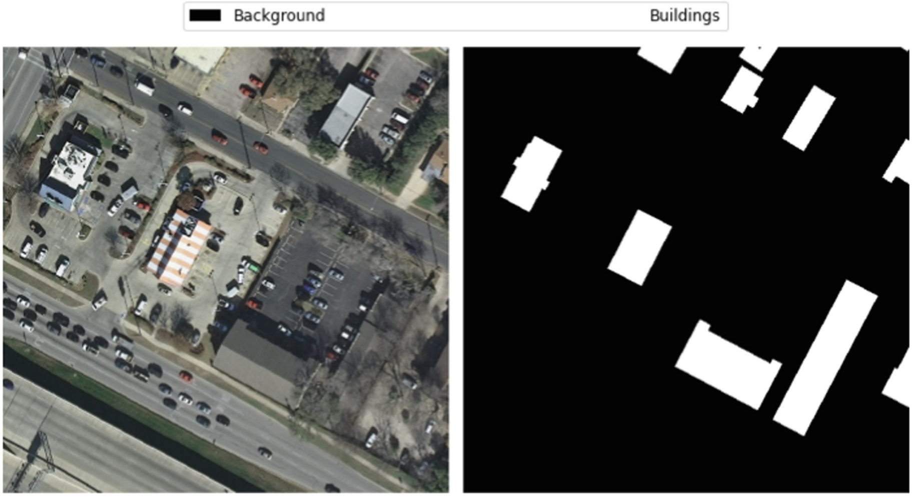

Fig. 1. Semantic segmentation of buildings in remote sensing imagery: A sample image (left) and its corresponding output (right) displaying prediction overlays. This example is sourced from the Massachusetts Buildings dataset [25].

With the continuous advancements in semantic segmentation techniques, they have found applications in addressing diverse and data-rich remote sensing problems [31]. These problems often involve complex and high-dimensional datasets, such as aerial and satellite images, which require accurate and detailed analysis. The semantic segmentation of remote sensing images plays an important role in many applications [31].

4 Methods for Semantic Segmentation

The task of semantic segmentation in remote sensing images has its unique challenges, due to the high resolution, complex spatial structures, diverse object scales, and the huge amounts of data.

Initially, traditional machine learning methods were the go-to solutions for this task, primarily grouped into two categories: pixel-based methods and region-based methods [46]. Both, pixel-based and region-based methods, relied heavily on handcrafted features and manual or heuristic threshold selection. Although sometimes effective, these methods often struggled with complex images with varying lighting conditions, textures, and scales. As a result, their performance could be inconsistent and their application limited compared to modern deep learning-based approaches [47].

The advent of deep learning brought forth a paradigm shift in semantic segmentation of remote sensing imagery. Deep learning models, unlike traditional machine learning techniques, can automatically learn hierarchical representations from raw data. This inherent ability enables them to detect complex patterns, handle high-dimensional data, and minimize the need for manual feature engineering, which is crucial for remote sensing imagery [21].

One of the most prominent deep learning models applied in the context of this task are the **Convolutional Neural Networks** (CNNs). CNNs are a class of deep learning models that excel at processing grid-like data, such as images [14].

Fully Convolutional Networks (FCNs) are an important development in the field of semantic segmentation. Unlike traditional CNNs that are confined to fixed-size inputs and outputs, FCNs are designed to handle inputs of any size and produce corresponding spatial outputs. This quality makes them particularly adept at pixel-level prediction tasks, a key requirement in semantic segmentation [22].

U-Net, initially designed for biomedical image segmentation, has proven to be highly effective in the semantic segmentation of remote sensing images as well. A unique mark of the U-Net architecture is its symmetric encoder-decoder structure. Its structure delivers detailed and accurate segmentation maps, even when working with relatively small datasets, a feature that has made U-Net particularly popular for tasks requiring precise localization [13].

Multi-Scale Contextual Models, represented by notable architectures such as DeepLabv3 [4] and Pyramid Scene Parsing Network (PSPNet) [48], introduce a novel approach to semantic segmentation that seeks to incorporate context information at varying scales. This is particularly beneficial in remote sensing image analysis where objects of interest often appear at different scales and densities.

Attention-based are pivotal in semantic segmentation of remote sensing images due to their capacity to allocate computational focus selectively [12]. They employ an attention mechanism, which uses attention maps, learned during the training process, to assign varying weights to different regions in the feature maps. This capability is particularly crucial for remote sensing images where some regions can be more relevant than others depending on the task at hand. The weights assigned by the attention mechanism help to amplify the influence of important regions and suppress the less important ones in the subsequent layers of the model, enhancing the ability to distinguish different land cover classes or physical objects within the imagery.

The Masked-attention Mask Transformer (Mask2Former) architecture falls in this category [6]. It is a novel architecture proficient in managing various image segmentation tasks such as panoptic, instance, or semantic segmentation. The design is built upon a simple meta architecture, which includes a backbone feature extractor, a pixel decoder, and a Transformer decoder. A main feature is the incorporation of masked attention within the Transformer decoder.

The evaluation of the semantic segmentation models in the context of remote sensing images is primarily done using *Pixel Accuracy*, *Mean Accuracy*, *Intersection over Union (IoU)* and *F1 Score (Dice Coefficient)*. Further details on the definition of the various metrics can be found in [31].

5 Datasets for Semantic Segmentation of Remote Sensing Images

There are numerous datasets specifically designed for semantic segmentation of remote sensing images that have been widely used for training and evaluating algorithms. Table 1 presents a summary of these datasets, including their source, type, number of images, image size, spatial resolution, format and provided bands. The datasets are with different spatial resolutions and sizes. The number of semantic labels in these datasets ranges from 2 to 20.

The datasets in the table exhibit a wide range of sources. While the majority of the datasets consist of aerial RGB images, there are also datasets that include multi-spectral data from specific satellite missions such as Sentinel-1, Sentinel-2, and Gaofen-2 [5]. Additionally, some datasets are sourced from platforms like Google Earth. It is worth noting that a significant portion of the datasets are obtained using unmanned aerial vehicles (UAVs), specifically drones, which provide high-resolution imagery for various applications.

Several datasets are specifically designed to support multi-class semantic segmentation tasks for accurate land cover mapping, such as SEN12MS [39], SemCity Toulouse [37], Christchurch Aerial Semantic Dataset (CASD) [2], DFC2022 [17], DLRSD [40] and Dubai's Satellite Imagery Dataset [28]. The LandCover.ai dataset, also known as Land Cover from Aerial Imagery, is specifically designed for the automatic mapping of buildings, woodlands, water, and roads from aerial images [3]. It contains a selection of aerial images taken over the area of Poland. The Inria Aerial Image Labeling dataset focuses on the task of semantic segmentation of aerial imagery [24] by providing ground truth data for two semantic classes: building and not building.

The Massachusetts Roads dataset consists of 1171 aerial images of the state of Massachusetts [25]. The target maps for the dataset are created by converting road centerlines obtained from the OpenStreetMap project into raster format. The labels are generated without any smoothing, using a line thickness of 7 pixels. The Massachusetts Buildings Dataset is composed of 151 aerial images capturing the Boston area [25]. The target maps for this dataset are generated by converting building footprints obtained from the OpenStreetMap project into raster format. GTA-V [49] is synthetic dataset for remote sensing image segmentation tailored for building extraction.

The Potsdam dataset [38] and Vaihingen dataset [38] are for urban semantic segmentation. These datasets are used in the 2D Semantic Labeling Contest. The datasets encompass five foreground classes: impervious surface, building, low vegetation, tree, car and one background class referred as clutter. The masks for these datasets are 3-channel geotiffs with unique RGB values for each class. The

Remote Sensing Land-Cover dataset for Domain Adaptive Semantic Segmentation (LoveDA) [45] encompasses two scenes (urban and rural) with significant challenges arise from the presence of multi-scale objects, intricate background elements, and uneven class distributions within the dataset. The Gaofen Image Dataset (GID-15) for semantic segmentation [42] contains 150 satellite images. The images are taken by the Gaofen-2 (GF-2) satellite over 60 cities in China. The images are organized into 15 semantic categories. The DeepGlobe Land Cover Classification Challenge dataset is designed for semantic segmentation tasks [9]. It includes high-resolution sub-meter satellite imagery for classifying land cover categories.

The Cloud Cover Segmentation Dataset was created through a crowdsourcing competition and subsequently validated by a team of expert annotators [35]. This dataset comprises Sentinel-2 satellite imagery along with corresponding cloud labels stored as GeoTiffs. The 95-Cloud dataset is an expansion of the previously released 38-Cloud dataset for cloud detection [27]. This binary classification allows for the precise identification and separation of cloudy areas within the imagery.

FloodNet, as described in [36], offers high-resolution UAS imagery with detailed semantic annotations specifically focusing on damage assessment after Hurricane Harvey. The dataset is captured using DJI Mavic Pro quadcopters, providing valuable information for flood damage analysis. The semantic annotations in FloodNet offer precise labeling for various classes, enabling accurate assessment of flood-related damages. ETCI2021 Flood Detection dataset [30] contains data from the flood event detection contest, organized by the NASA Interagency Implementation and Advanced Concepts Team. The primary objective of the dataset is to foster innovation in the detection of flood events and water bodies.

RIT-18 [18] dataset contains very-high resolution multispectral imagery collected by an unmanned aircraft system. The primary use of this dataset is for evaluating semantic segmentation frameworks designed for non-RGB remote sensing imagery. Several datasets, such as UAVid [23], Aeroscapes [32], DroneDeploy [11] and Semantic Drone Dataset [43] are specifically designed to enhance semantic understanding of urban scenes. These semantic segmentation datasets are with fine-resolution images obtained using UAVs.

6 Applications Enabled by Semantic Segmentation in Remote Sensing

With the ever-expanding number of aerial and satellite images, coupled with the constant development of modern deep-learning techniques, the application of semantic segmentation of aerial images takes significant contribution to wide range of applications, like urban planning, land cover mapping, disaster management, environmental monitoring, and precision agriculture.

Semantic segmentation of remote sensing images, can help urban planners to gain insights into various aspects of urban environments, including the identification of buildings, roads, vegetation, water bodies, and other infrastructure

Table 1. Properties of the different semantic segmentation datasets. The table contains information for the source, type, number of images, image size, spatial resolution, format and provided bands for the different datasets.

Name	Source	#Images	Image Size	Spatial Res.	#Labels	Format	Bands
LandCover.ai [3]	Aerial	10674	512 × 512	0.25–0.5 m	5	geo tif	RGB
Inria [24]	Aerial	360	5000 × 5000	0.3 m	2	geo tif	RGB
Massachusetts Roads [25]	Aerial	1171	1500 × 1500	1 m	2	tif	RGB
Massachusetts Buildings [25]	Aerial	151	1500 × 1500	1 m	2	tif, png	RGB
Vaihingen [38]	Aerial	33	2494 × 2064	0.09 m	6	geo tif	RG, NIR
Potsdam [38]	Aerial	38	6000 × 6000	0.05 m	6	geo tif	RGB, NIR
LoveDA [45]	Google Earth	5987	1024 × 1024	0.03 m	7	png	RGB
GID-15 [42]	Gaofen-2	150	6800 × 7200	3 m	15	tif	RGB
DeepGlobe Land Cover [9]	DigitalGlobe Vivid+	803	2448 × 2448	0.5 m	7	png, jpg	RGB
UAVid [23]	UAV	420	4096 × 2160	n/a	8	png	RGB
Cloud Cover Segmentation [35]	Sentinel-2	22728	512 × 512	10 m	2	geo tif	MSI
95-Cloud [27]	Landsat 8	43902	384 × 384	30 m	2	png	RGB, NIR
ETCI2021 Flood Detection [30]	Sentinel-1	66810	256 × 256	5–20 m	2	png	SAR
FloodNet [36]	UAV	2343	4000 × 3000	0.015 m	10	jpg, png	RGB
SEN12MS [39]	Sentinel-1/2, MODIS	180662	256 × 256	10 m	33	tif	SAR, MSI
CASD [2]	Aerial	4	4800 × 3600	0.1 m	4	tif	RGB
SemCity Toulouse [37]	Worldview-II	16	3504 × 3452	2 m	7	geo tif	MSI
DFC2022 [17]	Aerial	3981	2000 × 2000	0.5 m	16	geo tif	RGB
DLRSD [40]	Aerial	2100	256 × 256	0.305 m	17	tif, png	RGB
Dubai's Satellite Imagery Dataset [28]	Aerial	72	different sizes	n/a	6	jpg, png	RGB
GTA-V-SID [49]	Synthetic	121	500 × 500	1 m	2	png	RGB
RIT-18 [18]	UAV	3	different sizes	0.047 m	19	mat	MSI
Aeroscapes [32]	UAV	3269	1280 × 720	n/a	12	jpg, png	RGB
DroneDeploy [11]	UAV	55	11084 × 12326	0.1 m	7	tif, png	RGB
Semantic Drone Dataset [43]	UAV	600	6000 × 4000	n/a	20	png	RGB

elements [3,24]. Further more, it can helps in understanding the spatial distribution and patterns of the different land cover types, which is crucial for urban planning tasks such as infrastructure development and transportation planning. The semantic segmentation of remote sensing images can support the analysis of urban growth and change over time. The study presented in [41] introduces an approach that involves comparing segmented images taken at various time intervals. This allows urban planners to analyze the dynamics of urban development, monitor changes in land use, and evaluate the impact of urban planning policies and interventions.

During natural disasters such as floods, earthquakes, or wildfires, remote sensing images can be used to monitor the affected areas and detect changes over time [7,16], [?]. In [16], deep learning techniques including PSPNet, DeepLabV3, and U-Net are employed on the FloodNet dataset [36] to detect floods. The focus of the study is on identifying flooded roads and buildings, as well as distinguishing between natural water and flooded water.

By leveraging the spatial and temporal resolution of remote sensing images, semantic segmentation can also facilitate the monitoring and management of post-disaster recovery and reconstruction activities. In [7], the authors present an improved Swin transformer for semantic segmentation of post-earthquake dense buildings in urban areas. The method is used to identify damaged buildings, allowing emergency response teams to prioritize rescue and recovery efforts.

One of the primary applications of semantic segmentation in environmental monitoring is the detection and monitoring of deforestation. By segmenting satellite or aerial images, semantic segmentation algorithms can identify forested areas and distinguish them from cleared or degraded regions. In [1], the authors propose a specialized variant of the DeepLabv3+ architecture called DeepLab Change Detection (DLCD) for detecting deforestation in areas of significant forest loss.

Climate monitoring is another important aspect of environmental monitoring that benefits from semantic segmentation. By analyzing remote sensing images, semantic segmentation can identify and classify different climate-related features, such as clouds, aerosols, atmospheric conditions, and surface temperature variations [27,35]. This information aids in climate modeling, weather forecasting, and understanding the dynamics of climate change.

By segmenting remote sensing images, semantic segmentation can identify water bodies, such as rivers, lakes, and reservoirs, as well as detect changes in water levels and quality. Further more, Semantic segmentation aids in mapping and monitoring wetlands, coastal areas, and other ecologically sensitive water-based ecosystems. In [29], the authors utilize two semantic segmentation methods, namely DeepLabv3+ and SegNet, for the detection of water bodies. The objective of this detection is to estimate water levels and monitor temporal fluctuations in water levels.

Important aspect of precision agriculture is crop health monitoring. Semantic segmentation can detect and classify vegetation health indicators, such as areas affected by pests, diseases, nutrient deficiencies, or water stress. For example, in

[8] a high-resolution aerial imagery and UNet with a convolutional LSTM is used to accurately detect regions of the field showing nutrient deficiency stress. This approach minimizes resource wastage and optimizes crop health and yield. Water management is another critical component of precision agriculture. By analyzing remote sensing images, semantic segmentation can map and monitor soil moisture levels, water stress, and irrigation efficiency across agricultural fields. Furthermore, semantic segmentation aids in weed detection and management as in [26] where semantic segmentation is applied in two stages using UNet architecture. This approach promotes sustainable weed control practices and reduces the development of herbicide resistance.

7 Conclusion

In conclusion, semantic segmentation of remote sensing images is a powerful technique that enables the accurate classification and labeling of individual pixels or regions within an image, based on their semantic meaning. It plays a crucial role in various applications across different domains. The paper provides a comprehensive overview of semantic segmentation, starting with its definition and the underlying methods used for image analysis and classification. It discusses the importance of high-quality datasets for training and evaluation purposes, highlighting several notable datasets available for semantic segmentation of remote sensing images.

Furthermore, the paper explores the wide range of applications where semantic segmentation is utilized. It covers areas such as urban planning, disaster management, environmental monitoring, precision agriculture, deforestation, climate analysis, water management, and more. Each application benefits from the accurate and detailed understanding of land cover and object classification provided by semantic segmentation. The diverse datasets described in the paper, ranging from aerial and satellite imagery to UAV and synthetic data, reflect the breadth of sources used for remote sensing image analysis. The inclusion of different spectral bands and data formats demonstrates the flexibility and adaptability of semantic segmentation techniques.

Acknowledgements. This work was partially financed by the Faculty of Computer Science and Engineering at the Ss. Cyril and Methodius University in Skopje.

References

1. de Andrade, R.B., Mota, G.L.A., da Costa, G.A.O.P.: Deforestation detection in the amazon using deeplabv3+ semantic segmentation model variants. Remote Sens. **14**(19) (2022). https://doi.org/10.3390/rs14194694, https://www.mdpi.com/2072-4292/14/19/4694
2. Audebert, N., Saux, B.L., Lefèvre, S.: Segment-before-detect: vehicle detection and classification through semantic segmentation of aerial images. Remote. Sens. **9**, 368 (2017)

3. Boguszewski, A., Batorski, D., Ziemba-Jankowska, N., Dziedzic, T., Zambrzycka, A.: LandCover.ai: dataset for automatic mapping of buildings, woodlands, water and roads from aerial imagery. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops, pp. 1102–1110 (2021)
4. Chen, L.C., Papandreou, G., Kokkinos, I., Murphy, K., Yuille, A.L.: DeepLab: semantic image segmentation with deep convolutional nets, Atrous convolution, and fully connected CRFs. IEEE Trans. Pattern Anal. Mach. Intell. **40**(4), 834–848 (2017)
5. Chen, L., et al.: An introduction to the Chinese high-resolution earth observation system: Gaofen-1 7 civilian satellites. J. Remote Sens. **2022** (2022)
6. Cheng, B., Choudhuri, A., Misra, I., Kirillov, A., Girdhar, R., Schwing, A.G.: Mask2former for video instance segmentation. arXiv preprint arXiv:2112.10764 (2021)
7. Cui, L., Jing, X., Wang, Y., Huan, Y., Xu, Y., Zhang, Q.: Improved swin transformer-based semantic segmentation of postearthquake dense buildings in urban areas using remote sensing images. IEEE J. Sel. Top. Appl. Earth Observ. Remote Sens. **16**, 369–385 (2023). https://doi.org/10.1109/JSTARS.2022.3225150
8. Dadsetan, S., Rose, G.L., Hovakimyan, N., Hobbs, J.: Detection and prediction of nutrient deficiency stress using longitudinal aerial imagery. In: AAAI Conference on Artificial Intelligence (2020)
9. Demir, I., et al.: DeepGlobe 2018: a challenge to parse the earth through satellite images. In: The IEEE Conference on Computer Vision and Pattern Recognition (CVPR) Workshops (2018)
10. Dimitrovski, I., Kitanovski, I., Kocev, D., Simidjievski, N.: Current trends in deep learning for earth observation: an open-source benchmark arena for image classification. ISPRS J. Photogramm. Remote. Sens. **197**, 18–35 (2023)
11. DroneDeploy: Dronedeploy machine learning segmentation benchmark (2019). https://github.com/dronedeploy/dd-ml-segmentation-benchmark. Accessed 7 June 2023
12. Fu, J., et al.: Dual attention network for scene segmentation. In: 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), pp. 3141–3149 (2019)
13. Gu, X., Li, S., Ren, S., Zheng, H., Fan, C., Xu, H.: Adaptive enhanced swin transformer with u-net for remote sensing image segmentation. Comput. Electr. Eng. **102**, 108223 (2022)
14. Guo, Y., Liu, Y., Georgiou, T., Lew, M.S.: A review of semantic segmentation using deep neural networks. Int. J. Multimed. Inf. Retrieval **7**, 87–93 (2018)
15. Hafiz, A.M., Bhat, G.: A survey on instance segmentation: state of the art. Int. J. Multimed. Inf. Retrieval **9** (2020)
16. Hernández, D., Cecilia, J.M., Cano, J.C., Calafate, C.T.: Flood detection using real-time image segmentation from unmanned aerial vehicles on edge-computing platform. Remote Sens. **14**(1) (2022)
17. Hänsch, R., et al.: Data fusion contest 2022 (DFC2022) (2022). https://dx.doi.org/10.21227/rjv6-f516
18. Kemker, R., Salvaggio, C., Kanan, C.: Algorithms for semantic segmentation of multispectral remote sensing imagery using deep learning. ISPRS J. Photogram. Remote Sens. **145**, 60–77 (2018)
19. King, M., Herring, D.: Satellites | research (atmospheric science). In: Holton, J.R. (ed.) Encyclopedia of Atmospheric Sciences, pp. 2038–2047. Academic Press, Oxford (2003)

20. Kirillov, A., He, K., Girshick, R., Rother, C., Dollar, P.: Panoptic segmentation. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) (2019)
21. Lateef, F., Ruichek, Y.: Survey on semantic segmentation using deep learning techniques. Neurocomputing **338**, 321–348 (2019)
22. Long, J., Shelhamer, E., Darrell, T.: Fully convolutional networks for semantic segmentation. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 3431–3440 (2015)
23. Lyu, Y., Vosselman, G., Xia, G.S., Yilmaz, A., Yang, M.Y.: UAVid: a semantic segmentation dataset for UAV imagery. ISPRS J. Photogramm. Remote. Sens. **165**, 108–119 (2020)
24. Maggiori, E., Tarabalka, Y., Charpiat, G., Alliez, P.: Can semantic labeling methods generalize to any city? The INRIA aerial image labeling benchmark. In: IEEE International Geoscience and Remote Sensing Symposium (IGARSS) (2017)
25. Mnih, V.: Machine learning for aerial image labeling. Ph.D. thesis, University of Toronto (2013)
26. Moazzam, S.I., Khan, U.S., Qureshi, W.S., Nawaz, T., Kunwar, F.: Towards automated weed detection through two-stage semantic segmentation of tobacco and weed pixels in aerial imagery. Smart Agric. Technol. **4**, 100142 (2023)
27. Mohajerani, S., Saeedi, P.: Cloud-net+: a cloud segmentation CNN for landsat 8 remote sensing imagery optimized with filtered Jaccard loss function, vol. 2001.08768 (2020)
28. Mohammed Bin Rashid Space Center: Semantic segmentation dataset (2020). https://humansintheloop.org/resources/datasets/semantic-segmentation-dataset-2/. Accessed 7 June 2023
29. Muhadi, N.A., Abdullah, A.F., Bejo, S.K., Mahadi, M.R., Mijic, A.: Deep learning semantic segmentation for water level estimation using surveillance camera. Appl. Sci. **11**(20) (2021). https://doi.org/10.3390/app11209691, https://www.mdpi.com/2076-3417/11/20/9691
30. NASA Interagency Implementation and Advanced Concepts Team: ETCI 2021 competition on flood detection (2021). https://nasa-impact.github.io/etci2021/. Accessed 7 June 2023
31. Neupane, B., Horanont, T., Aryal, J.: Deep learning-based semantic segmentation of urban features in satellite images: a review and meta-analysis. Remote Sens. **13**(4), 808 (2021)
32. Nigam, I., Huang, C., Ramanan, D.: Ensemble knowledge transfer for semantic segmentation. In: 2018 IEEE Winter Conference on Applications of Computer Vision (WACV), pp. 1499–1508 (2018)
33. Osco, L.P., et al.: A review on deep learning in UAV remote sensing. Int. J. Appl. Earth Obs. Geoinf. **102**, 102456 (2021)
34. Pearlman, J., Barry, P., Segal, C., Shepanski, J., Beiso, D., Carman, S.: Hyperion, a space-based imaging spectrometer. IEEE Trans. Geosci. Remote Sens. **41**(6), 1160–1173 (2003)
35. Radiant Earth Foundation: Sentinel-2 cloud cover segmentation dataset (version 1) (2022). https://doi.org/10.34911/rdnt.hfq6m7. Accessed 7 June 2023
36. Rahnemoonfar, M., Chowdhury, T., Sarkar, A., Varshney, D., Yari, M., Murphy, R.R.: FloodNet: a high resolution aerial imagery dataset for post flood scene understanding. IEEE Access **9**, 89644–89654 (2021)
37. Roscher, R., Volpi, M., Mallet, C., Drees, L., Wegner, J.D.: Semcity toulouse: a benchmark for building instance segmentation in satellite images. ISPRS Ann. Photogram. Remote Sens. Spat. Inf. Sci. **V-5-2020**, 109–116 (2020)

38. Rottensteiner, F., et al.: The ISPRS benchmark on urban object classification and 3D building reconstruction. ISPRS Ann. Photogram. Remote Sens. Spat. Inf. Sci. **I-3** (2012)
39. Schmitt, M., Wu, Y.L.: Remote sensing image classification with the sen12ms dataset. In: ISPRS Ann. Photogram. Remote Sens. Spat. Inf. Sci. **V-2-2021**, 101–106 (2021)
40. Shao, Z., Yang, K., Zhou, W.: Performance evaluation of single-label and multi-label remote sensing image retrieval using a dense labeling dataset. Remote Sens. **10**(6), 964 (2018)
41. Toker, A., et al.: DynamicEarthNet: daily multi-spectral satellite dataset for semantic change segmentation. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), pp. 21158–21167 (2022)
42. Tong, X.Y., et al.: Land-cover classification with high-resolution remote sensing images using transferable deep models. Remote Sens. Environ. **237**, 111322 (2020)
43. TU Graz, Graz University of Technology : Semantic drone dataset (2020). http://dronedataset.icg.tugraz.at. Accessed 7 June 2023
44. Tupin, F., Inglada, J., Nicolas, J.M.: Remote Sensing Imagery. Wiley, New York (2014)
45. Wang, J., Zheng, Z., Ma, A., Lu, X., Zhong, Y.: LoveDA: a remote sensing land-cover dataset for domain adaptive semantic segmentation. In: Vanschoren, J., Yeung, S. (eds.) Proceedings of the Neural Information Processing Systems Track on Datasets and Benchmarks, vol. 1. Curran Associates, Inc. (2021)
46. Weinland, D., Ronfard, R., Boyer, E.: A survey of vision-based methods for action representation, segmentation and recognition. Comput. Vis. Image Underst. **115**(2), 224–241 (2011)
47. Yuan, X., Shi, J., Gu, L.: A review of deep learning methods for semantic segmentation of remote sensing imagery. Expert Syst. Appl. **169**, 114417 (2021)
48. Zhao, H., Shi, J., Qi, X., Wang, X., Jia, J.: Pyramid scene parsing network. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 2881–2890 (2017)
49. Zou, Z., Shi, T., Li, W., Zhang, Z., Shi, Z.: Do game data generalize well for remote sensing image segmentation? Remote Sens. **12**(2), 275 (2020)