

Article

Geographical Information System Enhancement Using Active Learning-Enhanced Semantic Segmentation

Sungkwan Youm¹ and Sunghyun Go^{2,*}¹ Department of Information & Communication Engineering, Wonkwang University, Iksan 54538, Republic of Korea; skyoum@gmail.com² Department of Computer Software Engineering, Wonkwang University, Iksan 54538, Republic of Korea

* Correspondence: saintgo7@wku.ac.kr; Tel.: +82-63-850-6748

Abstract: Images captured by drones are increasingly used in various fields, including geographic information management. This study evaluates a procedure that incorporates active learning semantic segmentation for verifying the building registration ledger. Several semantic segmentation techniques were evaluated to extract building information, with ResNet identified as the most effective method for accurately recognizing building roofs. Using active learning, the training data were refined by removing instances with low similarity, leading to improved network performance of the model. The procedure was demonstrated to identify discrepancies between the building information system and the inferred label images, as well as to detect labeling errors on a training dataset. Through this research, the geographic information system dataset is enhanced with minimal human oversight, offering significant potential for urban planning and building detection advancements.

Keywords: semantic segmentation; UAV; ResNet; map; building; active learning



Citation: Youm, S.; Go, S. Geographical Information System Enhancement Using Active Learning-Enhanced Semantic Segmentation. *Appl. Sci.* **2023**, *13*, 11254. <https://doi.org/10.3390/app132011254>

Academic Editors: Miguel Angel Maté-González and Julia Aramendi

Received: 5 September 2023

Revised: 6 October 2023

Accepted: 10 October 2023

Published: 13 October 2023



Copyright: © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Drones, or Unmanned Aerial Vehicles (UAVs), have transformed the landscape of aerial photography. Their ability to operate at lower altitudes, coupled with their ability to capture high-resolution imagery, presents a cost-effective alternative to traditional methods. This has led researchers and businesses alike to adopt drones for various applications. Given their adaptability, drones can operate both indoors and outdoors, and the integration of diverse camera systems allows them to navigate around obstacles with precision. Today, drones are widely used in surveillance, agriculture, forestry, wildlife monitoring, disaster management, and even facial recognition [1,2]. Their surge in popularity can be attributed to their potential role in fostering sustainable urban environments [3]. Globally, UAVs are being recognized as potent tools for remote sensing in agricultural and environmental sectors [4,5]. One of the emerging areas is the use of remote sensing data for tree identification and mapping, which is pivotal for forest management [6].

In the construction realm, drones are game changers. Their capabilities range from pre-construction surveys, ensuring detailed mapping of site topography and underground infrastructures, to real-time monitoring of construction activities. Moreover, drones are making high-altitude inspections safer, obviating the need for risk-prone traditional methods. By enabling precise surveying, drones are reducing both the duration and costs associated with construction processes [7–9]. Semantic segmentation, an advanced computer vision technique, enhances the efficacy of image analysis, especially for images captured by drones. This technique classifies image pixels into coherent segments or objects, finding applications across construction, agriculture, and disaster management domains [10,11]. Incorporating drone-captured images into building information systems is of paramount importance. Traditional aerial or satellite imagery, often plagued by lower resolution and outdated data, is being outpaced by drone imagery, making it the go-to solution for updating building information systems.

In the realm of semantic segmentation, algorithms like SegNet, DeepLab, and U-Net, which are derivatives of convolutional neural networks (CNNs), have been prominent. These models, with their foundation in CNNs, are adept at image classification tasks. For instance, U-Net employs an output feature map as its input, enabling it to classify intricate images more accurately. SegNet, tailored for image segmentation, enhances CNNs by preserving intricate information through decoding prior pooling layers [12]. DeepLab, an evolved version of SegNet, excels in processing high-resolution imagery, ensuring the clear classification of even the most complex images [13–15]. In this research, while our focus was on detecting roofs using semantic segmentation, deliberate choices were made not to use other discriminative techniques like edge detection. Our primary objective was to concentrate on the most suitable semantic segmentation technique using human-labeled training data.

This study underscores the significance of labeling and refining data, leveraging only the building information registered in the GIS (geographic information system). Our approach leans towards active learning, wherein automatically labeled data undergo refinement with minimal human intervention, resulting in a curated labeled training dataset [16]. Often, when delineating specific areas in satellite or drone images, the parcel data from the geographic information system is harnessed, considering any discrepancies or errors [17,18]. Manual labeling, while detailed, is labor-intensive and prone to errors. This necessitates the optimal use of GIS data. At times, discrepancies in images may arise due to inaccuracies in GIS building data or challenges in deciphering image data. In this study, the aim was to enhance the training dataset by gauging network similarity during the refinement phase.

Building detection using drone images necessitates cross-referencing with the building registration ledger [19–22]. When the extracted data align with the building registration ledger, updates are in order. Any building not documented in the ledger is earmarked for special administrative consideration. In scenarios where building coordinate information is awry, a robust system is imperative to rectify these variances. It is essential that the drone images are cross-verified with the building registration ledger to ensure data integrity. Moreover, the service interface should be visually cohesive with the database, mirroring the building registration ledger records.

In this research, methods that recognize buildings by earmarking roof edges, commonly integrated into many studies, were excluded. The emphasis was on data refinement. Edge detection was intentionally omitted. Moreover, it was recognized that drone images can be influenced by shadows based on the sun's orientation—a factor not considered. Hence, meticulous care is required when segmenting such regions. The primary objective was data refinement using GIS data, and there is optimism that the proposed method will elevate recognition in application development.

This study increases the accuracy of map data by identifying unregistered structures and refreshing the geographic information system with data on demolished buildings. This study presents a UAV-centric semantic image segmentation methodology to pinpoint building locations using geographic information. The paper unfolds as follows: Section 2 outlines the methodologies employed, ranging from the refinement of the Geographical Information System to the detailed processes of semantic segmentation for roof detection and geographic image processing techniques. Section 3 explains our experimental results, highlighting the optimal algorithm selection, the GIS dataset refinement, and the practical application of our findings on test images. Finally, Section 4 concludes with a summary of our significant discoveries and their broader impact on geographic information management.

2. Materials and Methods

2.1. Methods for Geographical Information System Refinement

The goal of this research was to investigate the building registration ledger on South Korean maps by following the building registration ledger inspection procedure illustrated in Figure 1. The process consisted of three steps: The process of finding the optimal

algorithm, the process of refining data using active learning, and investigating the building registration ledger.

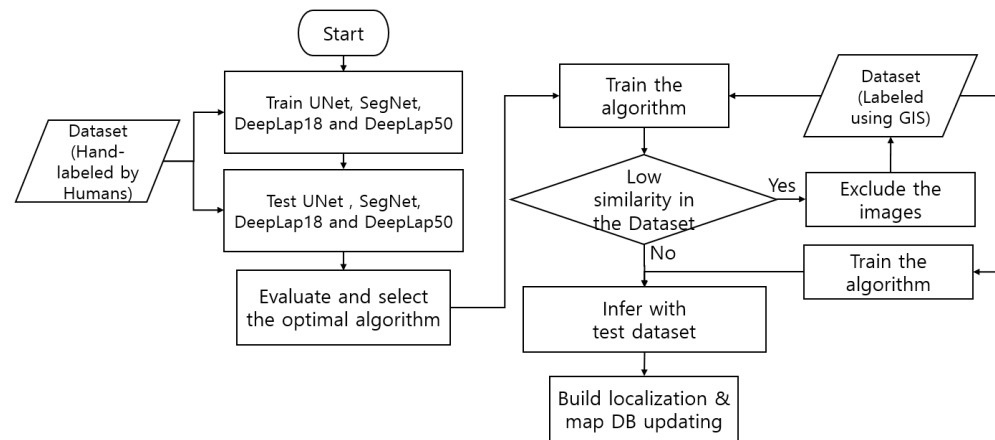


Figure 1. Proposed procedure for building registration ledger inspection using semantic segmentation.

As seen in the figure, the process progressed in the order of algorithm selection, data refinement, and checking the building registration ledger. In the process of finding the highest-performing algorithm for the semantic segmentation of drone images, the aim was to validate the performance of UNet, SegNet, DeepLab18, and DeepLab50. For accurate training and validation, manually labeled information was utilized. Once the algorithm was selected, a method was proposed to refine the dataset for training using this algorithm. Labeling every building for processing a vast range of images is challenging and difficult. However, the GIS of the building registration ledger can be used for automatic labeling, and there is a need to train and validate using these data. By measuring the training data similarity of the network trained with the algorithm and removing images with low similarity with human verification, the performance of the trained network can be improved. While the performance might be lower than when humans perform the labeling, it allows for the processing of a large amount of data without labeling. This process is referred to as active learning. After this process, if the IoU exceeded the threshold, the investigation results were recorded in the building database of the GIS system based on the building registration ledger.

2.2. Semantic Segmentation Algorithms for Roof

In this study, UNet, SegNet, DeepLab18, and DeepLab50 were evaluated as building detection methods for urban areas. Figures 2 and 3 show the network architecture used in this study. Various networks were employed to identify the one with the best performance, taking into account both accuracy and learning time. The input represents images captured by drones which have been resized to a format of $256 \times 256 \times 3$, indicating RGB channels. The output, on the other hand, consists of values in a 256×256 format, which determine whether a given pixel corresponds to a roof or not.

After capturing drone images with a resolution of $2560 \times 2560 \times 3$, they were georeferenced using RTK-based GPS systems to achieve centimeter-level accuracy in latitude and longitude information. These images were then labeled and divided into 10×10 segments corresponding to the labeled image. Four different networks, UNet, SegNet, DeepLab18, and DeepLab50, were applied to image learning in both urban and rural areas. The input network size was $256 \times 256 \times 3$, and the output network size was $256 \times 256 \times 1$. The input comprised a color image, and the output was divided into roof and non-roof areas using two classes.

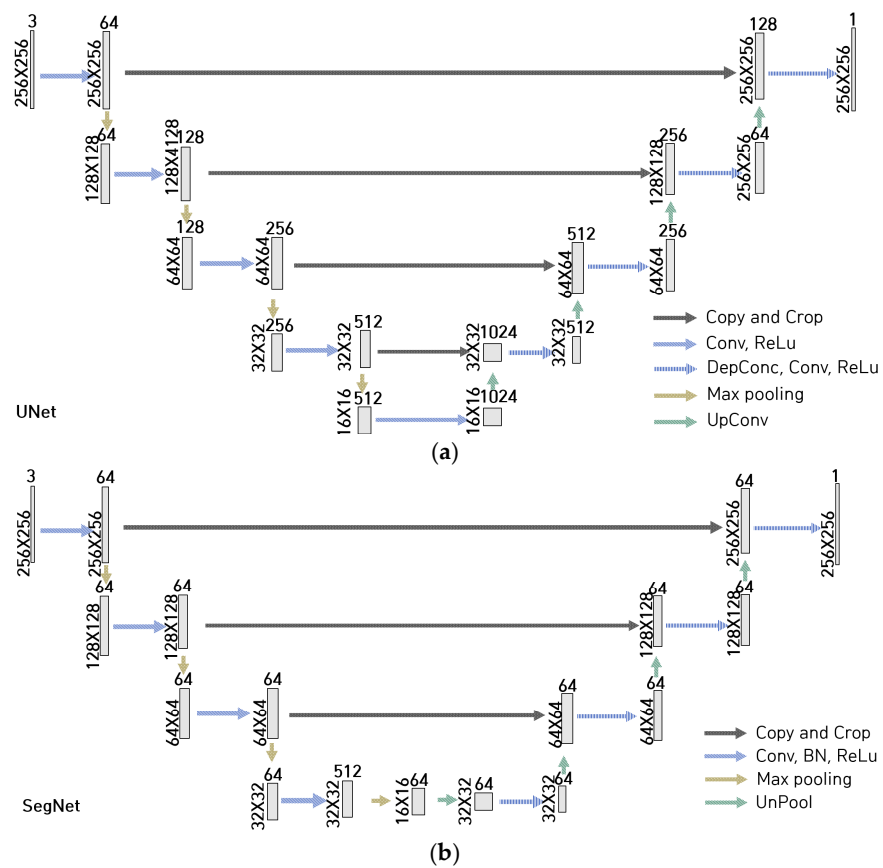


Figure 2. Illustration of neural networks: (a) Unet and (b) SegNet, which both have input dimensions of $256 \times 256 \times 3$ and output dimensions of 256×256 .

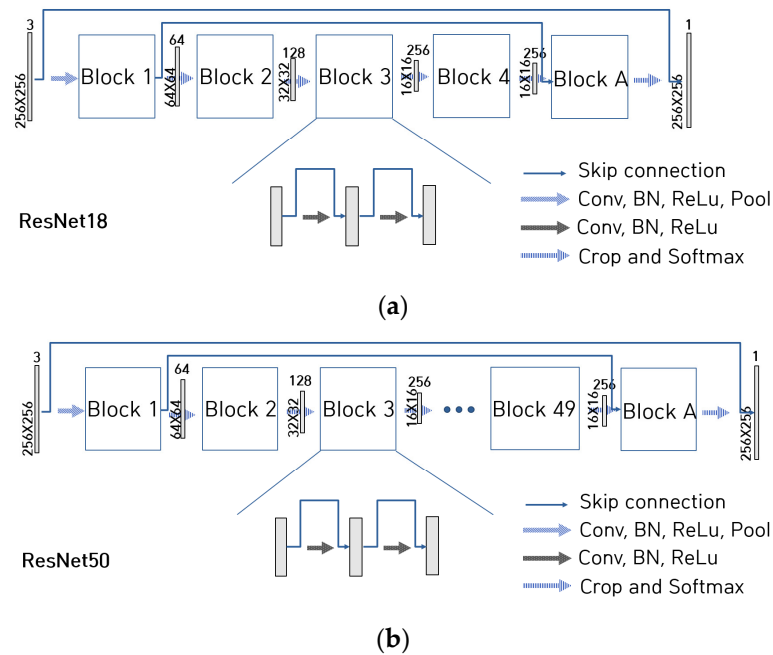


Figure 3. ResNet (a) 18 with input dimensions of $256 \times 256 \times 3$ and (b) 50, where only the input and output structures are modified.

The networks were tested with filter sizes of 32 and 3 and had two output classes representing the roof and non-roof regions. The encoder depth was set to 4, batch size to 10,

and learning rate to 1/1000. The performance of each network was evaluated to determine its suitability for utilizing existing networks.

Figure 2a,b show the architectures of UNet and SegNet, respectively, with $\times 256 \times 3$ color image input and a 2-class output. Figure 3a,b illustrate DeepLap18 and DeepLap50, which are similar, but with DeepLap50 having 49 block connections. The learning network used in this study was the same as that used in a previous study. I/O changed in the network model.

2.3. Geographic Image Processing and Latitude and Longitude Determination

When analyzing drone images, it is crucial to minimize coordinate system errors by accurately applying the images to maps. The coordinate system of the building map and drone image must be precise to ensure accurate analysis. The building's coordinate system was defined as 'Korea Central Belt 2010', with a possible error of approximately 10 cm. Meanwhile, the accuracy of the drone image could have errors of up to 50 cm using RTK GPS. The input test image resulted in a predicted output in two-pixel classes that could be used to locate and update a building's presence on the map. Before dividing the image into 10×10 labeling, the TIFF image's tag was read to calculate the coordinate information using the 'Korea Central Belt 2010' coordinate system. Once the location of the roof was identified through semantic segmentation, the corresponding TIFF Cartesian coordinate system values were determined. To calculate the position of an image pixel matrix in the Cartesian coordinate system, the latitude and longitude positions of the coordinate system, WGS84, projected onto the image must be determined. Therefore, it is important to work with these two coordinate systems in a unified manner.

The image labeling process involves assigning a roof or non-roof label to each pixel in the image, as illustrated in Figure 4. The sizes of the input and output images were $M = 256$ and $N = 256$, respectively, and they were in RGB format. The labeled images were used to train the network. When a test image is input into a trained network, the network outputs a predicted image showing which pixels are roofs and which are not.

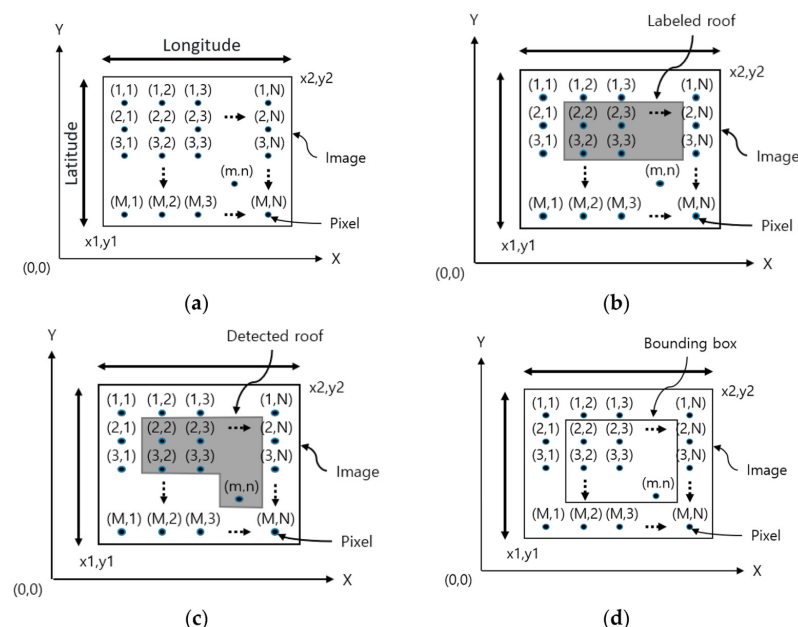


Figure 4. (a) Image matrix used for recognition, (b) ground truth showing labeled pixels, (c) detected roofs inferred by algorithms, and (d) bounding box made by detected pixels.

The image labeling process involves assigning a roof or non-roof label to each pixel in the image, as illustrated in Figure 4. Panel (a) shows the image with the corresponding coordinate system. This coordinate system is necessary because standard Cartesian and image pixel coordinates differ. Converting the image coordinates into geographical

coordinates requires an accurate coordinate system. Our study aimed to determine the geographical locations of the detected buildings, and hence, the detected pixel locations needed to be transformed into their corresponding geographical coordinates.

Figure 4b shows the image with ground truth labels, indicating which pixels represent roofs and which do not. Panel (c) presents the inferred roofs obtained using the proposed algorithm. Finally, in panel (d), the bounding boxes are generated based on the inferred roofs. The geographical coordinates of these bounding boxes were extracted and compared with those of the registered buildings in the cadastral map to validate their correspondence.

The input and output images had sizes of $M = 256$ and $N = 256$, respectively, and were represented in RGB format. These labeled images were used to train the network. Once a test image is fed into the trained network, the network predicts the location of the roof by outputting an image that indicates the pixels corresponding to the roof.

3. Results and Discussion

3.1. Selection of the Optimal Algorithm

The first procedure was conducted to select the best-performing algorithm. Drone images are typically prepared for this purpose, as shown in Figure 5. Jeju's urban area has a high concentration of buildings. When using the RTK GPS, the resulting images were highly accurate and large, with a precision of 5 cm. Figure 5a displays the entire region, with an image size of $500 \times 500 \text{ m}^2$. Figure 5b illustrates the image after it was segmented into 10×10 smaller images, resulting in a total of 100 images. In Figure 5c, the map was manually labeled by humans, with only the roofs highlighted in black, which were later categorized under the roof class. Figure 5d shows the corresponding 100 labels for the segmented images. The label images created manually in this manner had two classes: ground and roof. The aim was to find the best-performing algorithm through training and testing without validation, using the same options whenever possible. Therefore, separate validation was not conducted.

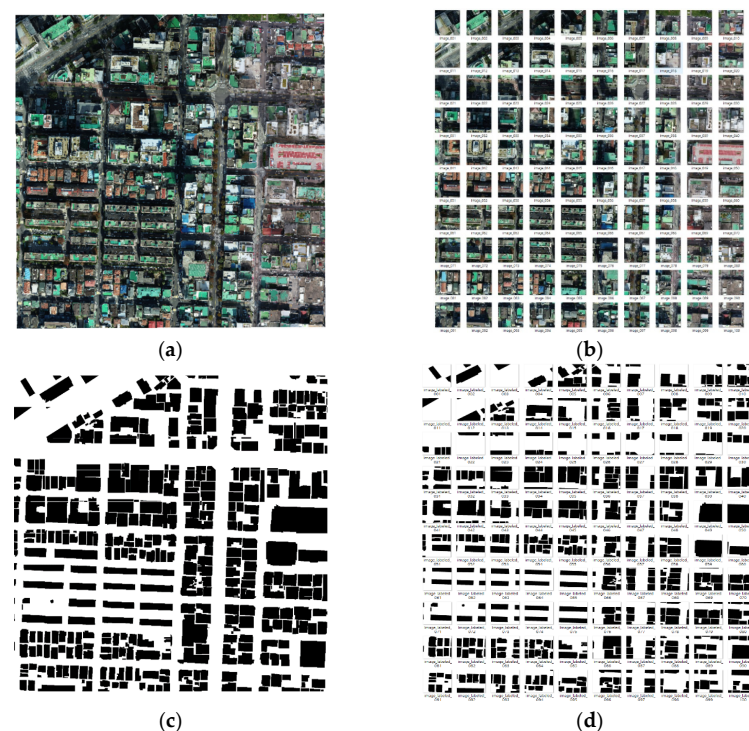


Figure 5. The data used for training: (a) drone-captured images divided into 100 images and (b) divided images with dimensions of $256 \times 256 \times 3$. (a) Labeled images corresponding to one image and (c,d) the divided labeled images.

Figure 6 presents the test images produced in a consistent manner. Figure 6a displays the entire test image. Figure 6b illustrates the 100 segmented images derived from the main image. Figure 6c depicts the comprehensive label image, while Figure 6d showcases the 100 individual images that were manually labeled. For testing purposes, segmented images were utilized. Notably, these images include the latitude and longitude details for each pixel.

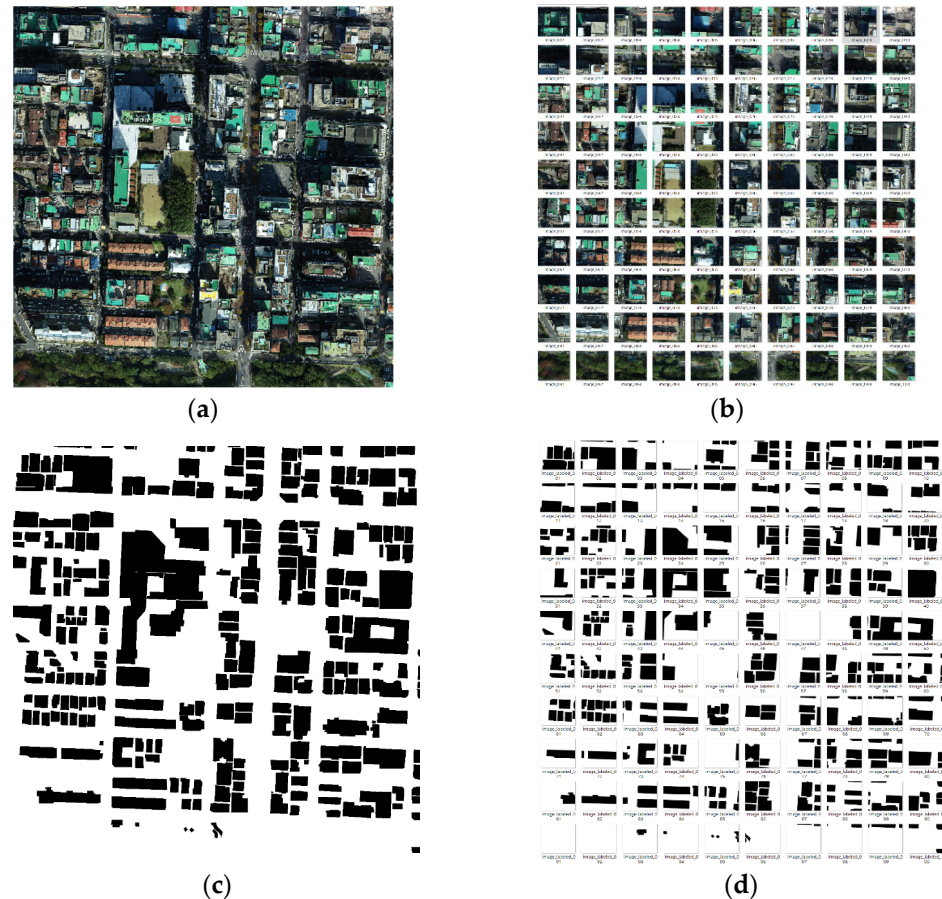


Figure 6. Data used for testing. (a) Drone-captured images divided into 100 images and (b) divided images with dimensions of $256 \times 256 \times 3$. (c) Labeled images corresponding to each divided image and (d) divided labeled images. Images contain latitude and longitude information for each pixel.

A program was developed incorporating algorithm verification, data refinement, and GIS DB update functionalities using the MATLAB 2022b library to facilitate the proposed process. To determine the most effective performance of Unet, SegNet, ResNet18, and ResNet50, tests were conducted on an Intel(R) Core(TM) i9-10900 and RTX 3070 8 Mega. For all algorithms, the number of epochs and batch size were set to 50 and 10, respectively. The learning time was measured and compared among the different algorithms to select the algorithm that required the least amount of time. The results are shown in Table 1. ResNet19 had the shortest learning time among all models.

Table 1. Elapsed time.

	UNet	SegNet	ResNet18	ResNet50
Training time	220.9	170.3	85.3	174.5

After measuring the learning time for U-Net, SegNet, ResNet18, and ResNet50, their performances were evaluated for roof detection accuracy. The total accuracy, mean accuracy,

mean IoU, and weighted IoU were measured and are listed in Table 2. Although ResNet50 exhibited an accuracy of over 80%, it required a long training time. Therefore, ResNet18 was chosen for this test because of its relatively high accuracy and short training time. The training was conducted using the same dataset and for the same number of epochs across all models. The respective layer counts for UNet, SegNet, ResNet18, and ResNet50 were 58, 59, 100, and 206. Their training speeds were recorded as 220.9, 170.3, 85.3, and 174.5 respectively. It seems that ResNet50, with its most complex layer structure, took the longest training time and achieved the highest accuracy. Given that accuracy is more crucial than time, memory, or resources in our context, ResNet50 appeared to be the most suitable algorithm.

Table 2. Performance results of the test network.

	Global Accuracy	Mean Accuracy	Mean IoU	Weighted IoU
UNet	0.77065	0.75067	0.60549	0.63206
SegNet	0.75993	0.70205	0.56705	0.60776
ResNet18	0.78701	0.75992	0.62369	0.65181
ResNet50	0.82275	0.80698	0.67971	0.70237

These metrics were used to evaluate how well the model classified the two classes: roof and background. As shown in Table 3, when the performance metrics were calculated, the Global Accuracy was approximately 82.3%, indicating that approximately 82.3% of the pixels were correctly classified. The *Mean Accuracy*

$$Global\ Accuracy = \frac{\sum TP}{\sum (TP + FP + FN)}$$

was approximately 80.7%, representing the average accuracy

$$Mean\ Accuracy = \frac{1}{2} \sum \frac{TP}{TP + FP}$$

across classes, which was approximately 80.7%. Mean Intersection over Union (*IoU*)

$$Mean\ IoU = \frac{1}{2} \sum \frac{TP}{TP + FP + FN}$$

was approximately 68.0%, demonstrating how well the model separated the classes. The *Weighted IoU*

$$Weighted\ IoU = \frac{1}{2} \sum \frac{(TP + FN) \frac{TP}{TP + FP + FN}}{TP + FP} \quad (1)$$

was approximately 70.2%, calculated by assigning weights to each class based on their size and then calculating the *IoU*.

Table 3. Confusion matrix of ResNet50.

Confusion Matrix		Expected	
		Roof	Background
Labeled	Roof	0.8615	0.1385
	Background	0.24754	0.75246

Table 3 presents the confusion matrix for ResNet50. Figure 7 displays the misclassified regions, highlighting the areas where incorrect predictions occurred. Specifically, Figure 7a represents false negatives, and Figure 7b illustrates false positives, both drawn to identify their distinct characteristics.

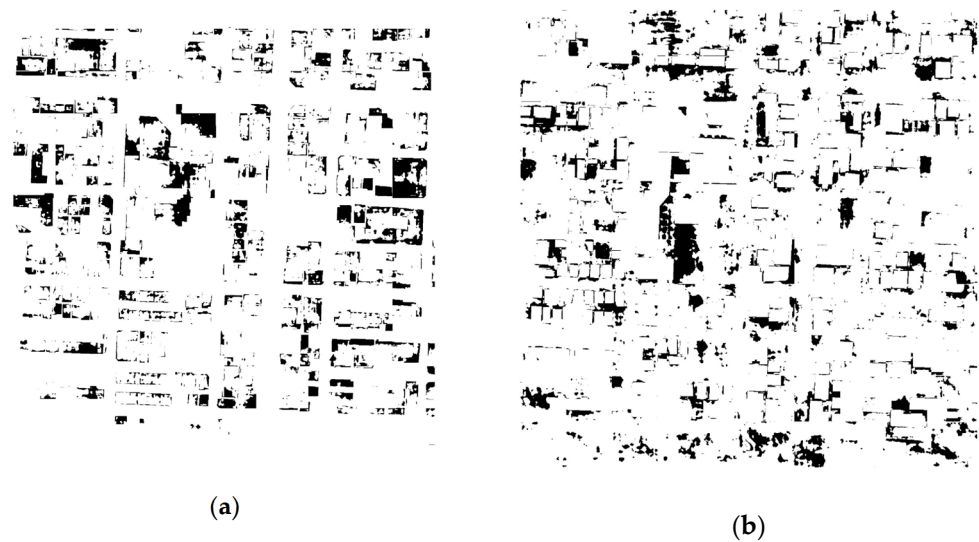


Figure 7. Misclassified regions of ResNet50: (a) false negatives and (b) false positives.

Upon reviewing the results of the initial test in the first test file, ResNet detected roofs more accurately, as shown in Figure 8. The results indicated that SegNet did not consider morphological aspects, whereas ResNet did consider the morphological model. Due to the high rainfall in the area, green waterproofing materials are commonly used, resulting in predominantly green roofs. While ResNet correctly identified green roofs, there was no significant difference compared to ResNet18. However, SegNet was unsuitable for this study because it did not consider morphological aspects.

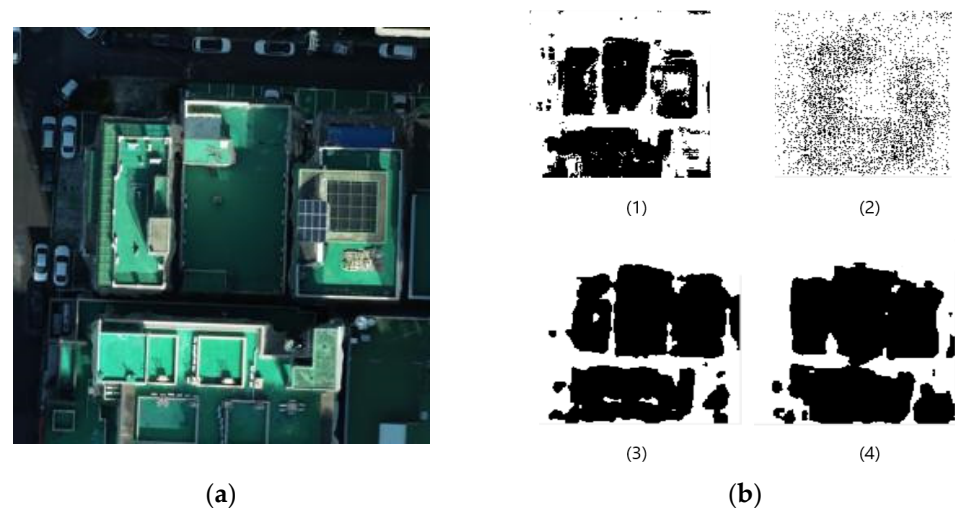


Figure 8. (a) Actual site image and (b) results of (1) Unet, (2) SegNet, (3) Resnet18, (4) and Resnet50, respectively.

3.2. Refinement of the GIS-Labeled Dataset

The intention was to conduct data refinement using Resnet50, which had the best performance. Manually labeling roofs is time-consuming, so there is a need to automatically label using the GIS of the building registration ledger. As mentioned earlier, once the top-performing algorithm was selected, the aim was to refine the GIS-labeled training dataset using this algorithm. After undergoing this refinement process, the performance of the neural network could be enhanced. The training was conducted using GIS labeling from two urban areas. To test the trained network, a third city was selected for evaluation. Figure 9 displays the study's focus on urban areas. Figure 9a,b present the targeted urban regions, while the corresponding GIS system labeling information is illustrated in

Figures 9c and 9d, respectively. Furthermore, with the aim of refining the data, a validation procedure that adjusts separate options was not established. The primary objective of the training here was data refinement. As seen in the figure, by using the GIS, the building registry contained latitude and longitude coordinates, allowing it to be utilized as labeling data that matched the coordinates in the image.

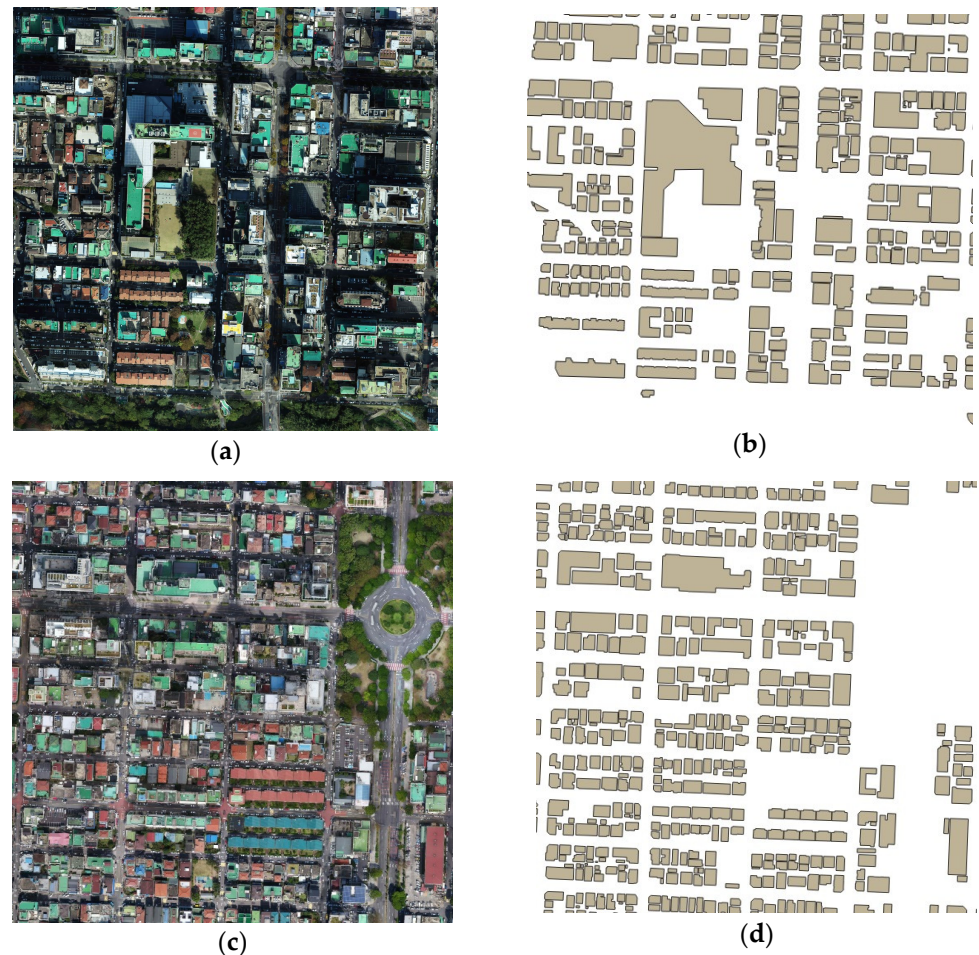


Figure 9. For data refinement, images in (a,c) were divided into 100 images and used as input for the training network, while (b,d) were used as the target for the network with 100 images.

Figure 10 displays the images used for training. Figure 10a is the initial training image with its corresponding label shown in Figure 10b, and the similarity score for the network regarding this image is 0.2944. Similarly, Figure 10c presents another training image with its label illustrated in Figure 10d, and it has a similarity score of 0.4541. By removing these two images and measuring the test results, improved performance was observed when comparing before and after refinement. The table displays performance metrics before and after manual refinement. Direct human intervention identified and removed 10 instances with the lowest similarity out of the initial 200 data points. By excluding these potential sources of error, improved results were achieved. Such hands-on refinement is vital to curtail error propagation, ensuring the model's robustness in real-world applications. Through this refinement process, it was confirmed that performance can be enhanced by cleansing the input data. In this manner, it was verified that the network performance can be improved, as shown in Table 4, by training and refining using the GIS building information system without manual labeling by humans. However, compared to manually labeled data, the performance using ResNet50 was found to be inferior. Yet, there is a

significant advantage in being able to utilize GIS information without manual labeling. Regarding *IoU*, it is

$$IoU = \frac{TP}{TP + FP + FN}$$

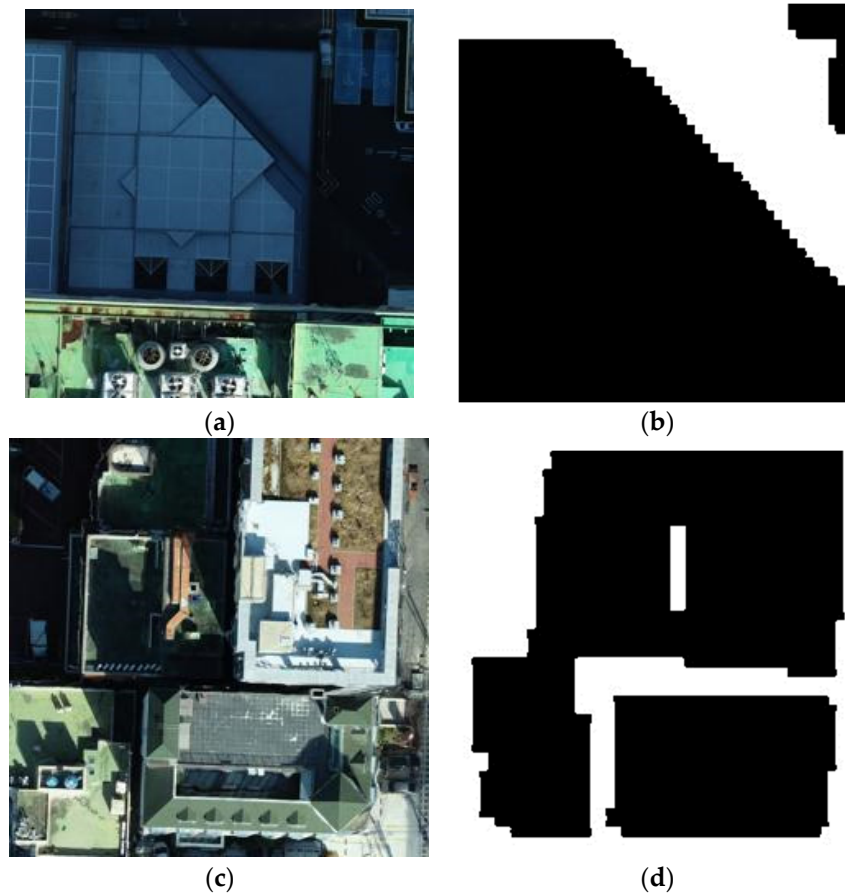


Figure 10. We identified images from the GIS-based labeled training data with low IoU values, where (a) visually appears like a rectangular building in (b), and (c) has a labeling error that makes it appear like a walkway in (d).

Table 4. Performance results of the refinement on the training dataset.

	Global Accuracy	Mean Accuracy	MeanIoU	WeightedIoU
Before refinement	0.7841	0.77669	0.6392	0.64405
After refinement	0.78896	0.78383	0.64729	0.6514

3.3. Inference of Test Images and Validation of GIS Application Implementation

Figure 11 illustrates the process of verifying the accuracy of building information. In Figure 11a, the building is located within the image. After identifying its position, Figure 11b showcases the roof detected by Resnet18, represented as a rectangle. Using this detected area, its location on the map in Figure 11c is determined and compared with the building information system. The verification results of this comparison are displayed in Figure 11d. The comparison method checks for a match between the roof value of the building information system and the recognized value; if it matches the IOU threshold, the database containing the building information is updated. In this study, a match of 0.6 was set as the threshold. Additionally, the area enclosed by the rectangle was checked for inclusion in the database using the Contain function of PostGIS, and the database was updated accordingly.

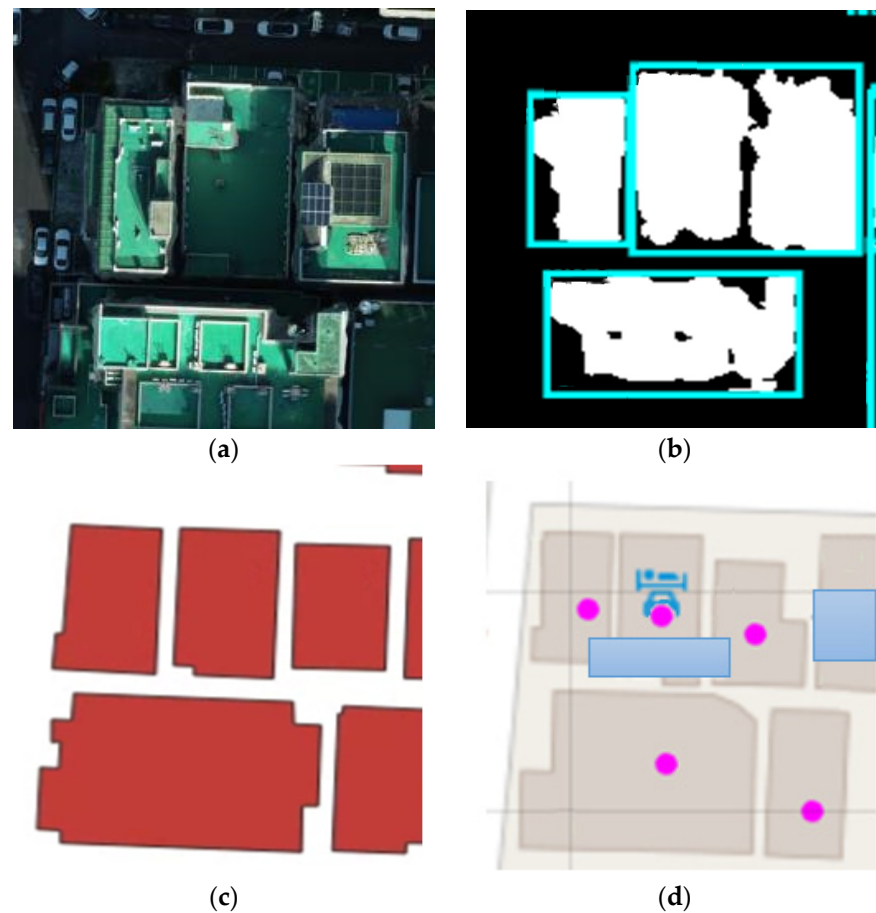


Figure 11. The result is from verifying the registered buildings through the proposed system: (a) a photo taken from a drone, (b) a photo predicting the roof using ResNet50, where the blue square in the image represents the building area, (c) a map showing building information from the geographic information system, and (d) results of marking them on the actual geographic information system with a pink dot indicating buildings confirmed in the ledger.

Methods for recognizing buildings from satellite or aerial imagery have been extensively researched [23,24]. In such building recognition tasks, various techniques are employed, including the incorporation of GIS information as input and the use of edge detection [25,26]. Generalizing these methods for application across different regions is challenging due to the diverse roof structures in different areas. For instance, some regions might require waterproofing for building roofs, while others might maintain a consistent style due to sunlight exposure. Furthermore, images captured by drones have high resolution, allowing for more accurate roof detection using precise models. The significance of this study lies in selecting an appropriate model for drone-captured images, refining the data, and verifying the buildings in the GIS.

4. Conclusions

This study aimed to inspect the building registration ledger, proposing and testing a procedure for its verification. Human-labeled training data were leveraged to evaluate multiple semantic segmentation methods, with the aim of identifying the most effective technique. Our findings indicated that ResNet50, due to its superior accuracy, emerged as the best candidate. To further refine and expand our GIS-labeled training dataset, the active learning approach was incorporated. This approach pinpointed training data with low similarity, enabling their refinement and, consequently, the generation of a more robust training set. After training the network with this optimized dataset, the model was employed to scrutinize the registered buildings on maps. This procedure was designed

to confirm the presence or absence of buildings, aligning with the records of the building registration ledger. Through active learning and the meticulous selection of semantic segmentation methods, our methodology offers a promising avenue for maintaining accurate and up-to-date building information systems and maps. Furthermore, the Contain function of PostGIS was utilized to ascertain that the recognized area was accurately incorporated into the database.

Author Contributions: Conceptualization, S.Y. and S.G.; methodology, S.Y.; software, S.Y.; validation, S.G.; formal analysis, S.Y.; investigation, S.Y.; resources, S.Y.; data curation, S.Y.; writing—original draft preparation, S.Y.; writing—review and editing, S.Y.; visualization, S.Y.; supervision, S.Y.; project administration, S.Y.; funding acquisition, S.G. All authors have read and agreed to the published version of the manuscript.

Funding: This paper was supported the Wonkwang University in 2021.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data presented in this study are available upon request from the corresponding author.

Acknowledgments: The authors thank the anonymous reviewers and editors for their insightful comments and suggestions.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Farooq, M.S.; Riaz, S.; Abid, A.; Abid, K.; Naeem, M.A. A Survey on the Role of IoT in Agriculture for the Implementation of Smart Farming. *IEEE Access* **2022**, *10*, 53374–53397. [\[CrossRef\]](#)
2. Sun, Y.; Zhi, X.; Han, H.; Jiang, S.; Shi, T.; Gong, J.; Zhang, W. Enhancing UAV Detection in Surveillance Camera Videos through Spatiotemporal Information and Optical Flow. *Sensors* **2023**, *23*, 6037. [\[CrossRef\]](#) [\[PubMed\]](#)
3. Kentsch, S.; Lopez Caceres, M.L.; Serrano, D.; Roure, F.; Diez, Y. Computer Vision and Deep Learning Techniques for the Analysis of Drone-Acquired Forest Images, a Transfer Learning Study. *Remote Sens.* **2020**, *12*, 1287. [\[CrossRef\]](#)
4. Nevavuori, P.; Narra, N.; Linna, P.; Lipping, T. Crop Yield Prediction Using Multitemporal UAV Data and Spatio-Temporal Deep Learning Models. *Remote Sens.* **2020**, *12*, 4000. [\[CrossRef\]](#)
5. Onishi, M.; Ise, T. Explainable identification and mapping of trees using UAV RGB image and deep learning. *Sci. Rep.* **2021**, *11*, 903. [\[CrossRef\]](#)
6. Ecke, S.; Dempewolf, J.; Frey, J.; Schwaller, A.; Endres, E.; Klemmt, H.-J.; Tiede, D.; Seifert, T. UAV-Based Forest Health Monitoring: A Systematic Review. *Remote Sens.* **2022**, *14*, 3205. [\[CrossRef\]](#)
7. Oh, S.; Ham, S.; Lee, S. Drone-Assisted Image Processing Scheme using Frame-Based Location Identification for Crack and Energy Loss Detection in Building Envelopes. *Energies* **2021**, *14*, 6359. [\[CrossRef\]](#)
8. He, Y.; Ma, W.; Ma, Z.; Fu, W.; Chen, C.; Yang, C.-F.; Liu, Z. Using Unmanned Aerial Vehicle Remote Sensing and a Monitoring Information System to Enhance the Management of Unauthorized Structures. *Appl. Sci.* **2019**, *9*, 4954. [\[CrossRef\]](#)
9. Diez, Y.; Kentsch, S.; Fukuda, M.; Caceres, M.L.L.; Moritake, K.; Cabezas, M. Deep Learning in Forestry Using UAV-Acquired RGB Data: A Practical Review. *Remote Sens.* **2021**, *13*, 2837. [\[CrossRef\]](#)
10. Lin, X.; Zhang, J. Object-based morphological building index for building extraction from high resolution remote sensing imagery. *Acta Geod. Cartogr. Sin.* **2017**, *46*, 724–733.
11. Gundu, S.; Syed, H. Vision-Based HAR in UAV Videos Using Histograms and Deep Learning Techniques. *Sensors* **2023**, *23*, 2569. [\[CrossRef\]](#)
12. Munawar, H.S.; Ullah, F.; Heravi, A.; Thaheem, M.J.; Maqsoom, A. Inspecting Buildings Using Drones and Computer Vision: A Machine Learning Approach to Detect Cracks and Damages. *Drones* **2022**, *6*, 5. [\[CrossRef\]](#)
13. Xu, H.; Song, J.; Zhu, Y. Evaluation and Comparison of Semantic Segmentation Networks for Rice Identification Based on Sentinel-2 Imagery. *Remote Sens.* **2023**, *15*, 1499. [\[CrossRef\]](#)
14. Ronneberger, O.; Fischer, P.; Brox, T. U-Net: Convolutional Networks for Biomedical Image Segmentation. In *Medical Image Computing and Computer-Assisted Intervention—MICCAI 2015*; Navab, N., Hornegger, J., Wells, W., Frangi, A., Eds.; Springer: Munich, Germany, 2015; pp. 234–241.
15. Badrinarayanan, V.; Kendall, A.; Cipolla, R. Segnet: A deep convolutional encoder-decoder architecture for image segmentation. *IEEE Trans. Pattern Anal. Mach. Intell.* **2017**, *39*, 2481–2495. [\[CrossRef\]](#)
16. Tharwat, A.; Schenck, W. A Survey on Active Learning: State-of-the-Art, Practical Challenges and Research Directions. *Mathematics* **2023**, *11*, 820. [\[CrossRef\]](#)

17. Chen, L.-C.; Papandreou, G.; Kokkinos, I.; Murphy, K.; Yuille, A.L. Semantic Image Segmentation with Deep Convolutional Nets and Fully Connected CRFs. *arXiv* **2014**, arXiv:1412.7062. Available online: <https://arxiv.org/abs/1412.7062> (accessed on 1 August 2023).
18. Wang, S.; Ding, L.; Chen, Z.; Dou, A. A Rapid UAV Image Georeference Algorithm Developed for Emergency Response. *J. Sens.* **2018**, *2018*, 8617843. [[CrossRef](#)]
19. Ballesteros, J.R.; Sanchez-Torres, G.; Branch-Bedoya, J.W. A GIS Pipeline to Produce GeoAI Datasets from Drone Overhead Imagery. *ISPRS Int. J. Geo-Inf.* **2022**, *11*, 508. [[CrossRef](#)]
20. Liu, X.; Li, R.; Wang, Y.; Nielsen, P.S. SEGSys: A mapping system for segmentation analysis in energy. *arXiv* **2012**, arXiv:2012.06446.
21. Dong, R.; Li, W.; Fu, H.; Gan, L.; Yu, L.; Zheng, J.; Xia, M. Oil palm plantation mapping from high-resolution remote sensing images using deep learning. *Int. J. Remote Sens.* **2020**, *41*, 2022–2046. [[CrossRef](#)]
22. Ballesteros, J.R.; Sanchez-Torres, G.; Branch-Bedoya, J.W. HAGDAVS: Height-Augmented Geo-Located Dataset for Detection and Semantic Segmentation of Vehicles in Drone Aerial Orthomosaics. *Data* **2022**, *7*, 50. [[CrossRef](#)]
23. Audebert, N.; Le Saux, B.; Lefèvre, S. Joint Learning from Earth Observation and OpenStreetMap Data to Get Faster Better Semantic Maps. In Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), Honolulu, HI, USA, 21–26 July 2017; pp. 1552–1560.
24. Krapf, S.; Bogenrieder, L.; Netzler, F.; Balke, G.; Lienkamp, M. RID—Roof Information Dataset for Computer Vision-Based Photovoltaic Potential Assessment. *Remote Sens.* **2022**, *14*, 2299. [[CrossRef](#)]
25. Xia, L.; Zhang, X.; Zhang, J.; Yang, H.; Chen, T. Building Extraction from Very-High-Resolution Remote Sensing Images Using Semi-Supervised Semantic Edge Detection. *Remote Sens.* **2021**, *13*, 2187. [[CrossRef](#)]
26. Lu, T.; Ming, D.; Lin, X.; Hong, Z.; Bai, X.; Fang, J. Detecting Building Edges from High Spatial Resolution Remote Sensing Imagery Using Richer Convolution Features Network. *Remote Sens.* **2018**, *10*, 1496. [[CrossRef](#)]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.