Indonesian Journal of Electrical Engineering and Computer Science
Vol. 42, No. 1, April 2026, pp. 115∼122
ISSN: 2502-4752, DOI: 10.11591/ijeecs.v42.i1.pp115-122

Dilated residual U-Net for vegetation detection from high resolution drone aerial imagery

Mgs. M. Luthfi Ramadhan[1], Rizal Maulana[2], Lalu Syamsul Khalid[2]
[1]Department of Informatics Engineering, Faculty of Computer Science, Universitas Sriwijaya, Palembang, Indonesia
[2]Department of Computer Science, Faculty of Computer Science, Universitas Indonesia, Depok, Indonesia

Article Info

Article history:

Received Jul 11, 2025
Revised Feb 10, 2026
Accepted Mar 4, 2026

Keywords:

Computer vision
Deep learning
Machine learning
Remote sensing
Semantic segmentation

ABSTRACT

Vegetation plays a vital role in regulating air quality and mitigating climate change by converting carbon dioxide into oxygen. However, ongoing human activity continues to degrade vegetation ecosystems, necessitating scalable and accurate monitoring methods. Traditional field-based statistical approaches are often costly and inefficient. This study proposes a deep learning model, dilated residual U-Net, for semantic segmentation of vegetation from drone-acquired aerial imagery. The model incorporates residual connections to reduce information loss and dilated convolutions to enhance receptive field coverage without increasing computational cost. Experiments conducted on the DroneDeploy Segmentation dataset demonstrate that the proposed model achieves a Dice coefficient of 0.4451 with an inference speed of 0.0675 seconds per image, outperforming baseline U-Net and Residual U-Net models. These results highlight the potential of lightweight, CNN-based architectures for environmental monitoring in resource-constrained settings.

Corresponding Author:

Mgs. M. Luthfi Ramadhan
Department of Informatics Engineering, Faculty of Computer Science, Universitas Sriwijaya
Palembang, Indonesia
Email: luthfiramadhan@unsri.ac.id

1. INTRODUCTION

Vegetation plays an important role in preventing air pollution, as it converts carbon into oxygen, which is crucial for humans [1]–[5]. The amount of carbon dioxide is a major problem for humans, as carbon dioxide gas causes global warming [6], [7]. Despite the benefits provided by vegetation, the vegetation ecosystem is often destroyed by humans for certain purposes [8]. This damage phenomenon is usually studied by forestry experts to minimize vegetation damage [9]. Direct field studies can be costly and time-consuming, therefore remote sensing techniques can be applied as an alternative to minimize costs and time in studying vegetation damage phenomena [10], [11]. One of the most commonly used remote sensing techniques for vegetation detection is the normalized difference vegetation index (NDVI), which analyzes vegetation greenness and health from satellite imagery, using red and near-infrared light reflectance [12]. However, this technique cannot distinguish between trees, grass, and shrubs. NDVI requires near infrared (NIR), which is sometimes difficult to obtain [13]–[15].

Therefore, a better method than NDVI is needed for vegetation detection. With the massive development of technology, vegetation detection can be done using artificial intelligence, with semantic segmentation being the most commonly used example. Semantic segmentation is one of the tasks in computer vision that aims to label or classify each pixel in an image [16]. Semantic segmentation in remote sensing images is very useful, particularly for monitoring in the field of agriculture to increase productivity and environmental protection. Thus, semantic segmentation can be used to divide regions in an image [17], [18].

Convolutional neural networks (CNNs) have become the state-of-the-art approach in computer vision tasks, including semantic segmentation. One early CNN-based method for segmentation is the fully convolutional network (FCN), which, despite its effectiveness, is computationally expensive and suffers from prolonged inference times. U-Net, originally developed for biomedical image segmentation, offers improved performance but presents challenges such as training instability due to its depth and a tendency toward vanishing gradients [16], [19], [20]. For these reasons, a study by [21] proposed incorporating residual connections into both the encoder and decoder paths of U-Net, thus enhancing gradient flow and preserving contextual information. A similar approach was also done by [22], who improved the Dice coefficient from 0.772 to 0.808 in nasopharyngeal cancer segmentation by integrating residual mechanisms.

Meanwhile, a study by [16] introduced dilated convolution blocks into the LinkNet architecture for aerial image segmentation, achieving an accuracy score of 86.1%. The use of dilated convolutions has also proven effective in biomedical applications, where [23] demonstrated improvements over classical U-Net by 2% to 14% across five different medical imaging modalities. These studies highlight the individual strengths of residual and dilated components for semantic segmentation tasks.

Drawing on these insights, we propose a novel architecture called dilated residual U-Net, which integrates both residual connections and dilated convolution blocks within the standard U-Net architecture. The residual connections aid in mitigating information loss and vanishing gradients, while the dilated convolutions expand the receptive field without increasing the kernel size or computational cost [24], [25]. In contrast to prior studies that introduce residual connections or dilated convolutions as independent architectural enhancements [16], [22]–[24] this work integrates both mechanisms within a single U-Net architecture and evaluates their combined effect on vegetation segmentation from RGB drone imagery. Unlike approaches that rely on multispectral data or vegetation indices, the proposed model operates solely on RGB inputs and targets resource-constrained deployment scenarios. This positioning allows a direct assessment of how architectural refinements alone can improve segmentation robustness under realistic data limitations.

The rest of this paper is organized as follows. Section 2 provides related works from previous studies. Section 3 details our proposed method. Section 4 discusses our experiment result. Section 5 concludes our paper.

2. RELATED WORKS

U-Net is a CNN that was originally developed for medical image segmentation. Its architecture is based on the FCN, but it is specifically modified to perform well with limited training data while maintaining high segmentation accuracy. The U-Net structure resembles a ”U” shape and consists of two main parts: the contracting path, which is famously known as the encoder, and the expansive path, which is famously known as the decoder. The contracting path follows a typical CNN architecture, with multiple convolutional layers, ReLU activation functions, and max pooling layers. This path progressively reduces spatial dimensions while capturing increasingly abstract feature representations. Conversely, the expansive path reconstructs the spatial resolution of the input image through a sequence of up-convolution (transposed convolution) and concatenation with corresponding feature maps from the contracting path. This combination of high-resolution spatial features and learned abstract features allows the network to localize more precisely. Instead of pooling operations, the expansive path uses upsampling layers to restore resolution, enabling the network to propagate contextual information to finer-grained output layers and ultimately achieve more accurate segmentation results [26].

The application of U-Net in tree segmentation conducted by [27] has demonstrated U-Net’s ability to detect the presence or absence of trees in an image. It is claimed that U-Net can absorb strong visual information from the local and spatial structures of high-resolution images. The U-Net used in their study employed a basic U-Net architecture with validation using Monte Carlo. In the process of splitting the dataset into training and testing data, repeated resplitting was performed to produce prediction results with accuracy within an interval.

Aerial imagery datasets generally contain a lot of information in each image. Information loss can be a problem in semantic segmentation with datasets that are rich in information. To address the issue of information loss, a study by [16] introduced dilated blocks to the LinkNet architecture with a ResNet50 backbone. The dilated block is implemented at the bottleneck of LinkNet. The purpose of adding the dilated block is to enlarge the receptive fields of a kernel without needing to increase the size of the kernel [16], [28], [29]. The architecture of the dilated block, which is an element-wise addition of various convolution layers with different dilation rates.

Zhang *et al.* [21] focused on road extraction from aerial images. They have it that the U-Net has a weakness, which is information loss in the U-Net layers. To minimize this issue, they proposed a new approach by adding residual connections to the U-Net architecture. The residual connection aims to minimize information loss and reduce the number of parameters in the model. They concluded that their proposed model was able to achieve better segmentation results.

3. PROPOSED METHOD

This study proposes an enhanced semantic segmentation model derived from the U-Net architecture, aimed at addressing common limitations associated with deep convolutional networks. An overview of the architectures and their detailed building blocks is illustrated in Figure 1. While U-Net has demonstrated strong performance in various segmentation tasks, its high number of layers often leads to information loss as spatial features are progressively downsampled, as shown in Figure 1(a). For this reason, we integrate residual connections into each convolutional block within both the encoder and decoder, as shown in Figure 1(b). The full layout of the proposed architecture is illustrated in Figure 1(c). Each convolutional block is based on the convRelu structure, whose detailed blocks are presented in Figure 1(d). These residual connections enable the preservation of critical low-level features by allowing gradient flow across non-adjacent layers, as illustrated in Figure 1(e), thus alleviating the vanishing gradient problem commonly encountered in deep networks [30]–[34]. To further improve efficiency without compromising segmentation accuracy, a dilated convolutional block was employed in the bottleneck block of the network. Dilated convolutions allow for a larger receptive field without increasing the number of parameters or computational cost, thus enabling the model to capture contextual information while preserving fine spatial details [35]. The integration of the dilated bottleneck block is illustrated in Figure 1(f).

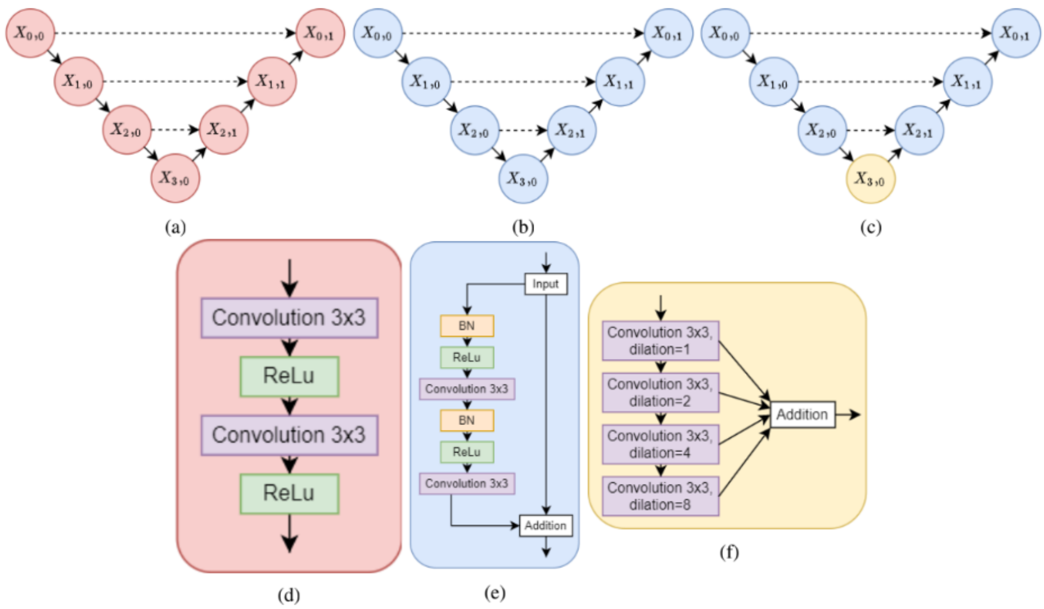

Figure 1. Overview of the architectures and their building blocks: (a) the architecture of U-Net, (b) residual U-Net, (c) and dilated residual U-Net, (d) the detailed blocks of convRelu, (e) residual connection, and (f) dilated

To evaluate the effectiveness of the proposed model, a comparative experiment was conducted against two baseline architectures, namely the standard U-Net and a residual U-Net. All models were trained and tested using the publicly available DroneDeploy segmentation dataset, which comprises 38 high-resolution aerial images with a spatial resolution of 10 cm/pixel. The primary focus of this study was vegetation detection, and thus, the dataset was preprocessed to include only two classes, vegetation and non-vegetation (a combination of all other classes). From the total dataset, 21 images were used for training, 5 for validation, and 12 for testing. Given the limited amount of data, data augmentation techniques were applied to improve the generalization capability of the models. Specifically, each training image was augmented through horizontal flipping and rotation by 53 degrees, resulting in a more diverse training set. Due to the very high spatial resolution, each image was subdivided into multiple 304×304 segments. This resulted in 4,238 training instances and 801 validation instances. Applying the same subdivision strategy, the 12 test images correspond to approximately 2,172 testing instances.

Model training was conducted using the Adam optimizer with an initial learning rate of 0.001 to minimize the binary cross-entropy loss function. The training process spanned 50 epochs, with a batch size of 12. To enhance convergence and prevent stagnation in poor local minima, a learning rate scheduling strategy was employed using ReduceLROnPlateau, with a patience of 2 epochs and a decay factor of 0.5. This adaptive learning rate adjustment ensured continued improvement in model performance during training by lowering the learning rate when validation loss plateaued.

4. RESULTS AND DISCUSSION

To assess the performance of the proposed model, this study employed three evaluation metrics, the dice coefficient, intersection over union (IoU), and inference time (measured in seconds). The dice coefficient is a widely used similarity measure for evaluating the overlap between predicted segmentation masks and ground truth labels. A higher dice score indicates a greater degree of spatial correspondence, reflecting more accurate segmentation. This metric is particularly effective in binary segmentation tasks where class imbalance may be significant. In addition to the dice coefficient, computational efficiency was quantified through measurement of the inference speed, defined as the time required for the model to generate predictions for a given input image. The comparative performance of U-Net, dilated U-Net, residual U-Net, and the proposed dilated residual U-Net is provided in Table **??**.

Table 1. Experimental results

Method	Dice	IoU	Inference Speed
Dilated residual U-Net (proposed)	**0.4451±0.0307**	**0.2863 ± 0.0284**	0.0675
Dilated U-Net	0.4316 ± 0.0305	0.2752 ± 0.028	0.0535
Residual U-Net [18]	0.4262 ± 0.0359	0.2708 ± 0.028	0.0713
U-Net [23]	0.4127 ± 0.0305	0.2600 ± 0.0276	**0.0484**

It can be inferred that the dilated residual U-Net demonstrates superior segmentation accuracy, achieving the highest dice coefficient among all tested models. This improvement can be attributed to the combined effect of residual connections, which preserve multi-scale contextual features, and dilated convolutions, which expand the receptive field. However, this increase in performance comes with a trade-off. The inclusion of additional modules, specifically, residual blocks and dilated convolution layers, leads to a higher number of parameters relative to the standard U-Net architecture. As a result, the inference speed of the dilated residual U-Net is slightly slower than that of the U-Net. Nonetheless, when compared to the residual U-Net, which shares similar structural complexity but lacks the efficiency of dilated operations in the bottleneck, the proposed model achieves a faster inference time.

Despite the overall effectiveness of the models, several challenges were observed during the evaluation, particularly related to class imbalance and annotation inconsistencies in the dataset. All models have a tendency to misclassify grassy or bushy areas adjacent to vegetation regions. This behavior can be attributed to the imbalanced class distribution in the training data, where the vegetation class is underrepresented compared to non-vegetation. Such an imbalance limits the model's capability to generalize vegetation features, especially in ambiguous or transitional regions. A representative example of this issue is presented in Figure 2, which highlights the difficulty of correctly classifying the grass in a football field. In the ground truth as shown in Figures 2(a)-(f), the grass in the field is labeled as non-vegetation, despite its visual similarity to typical vegetation, particularly in terms of green color intensity and texture. This variance poses a challenge for the model, which tends to associate green-colored areas with vegetation, leading to over-segmentation in regions that are contextually ambiguous.

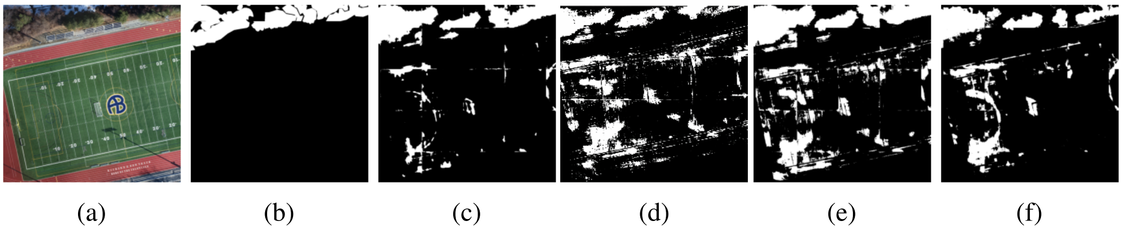

Figure 2. Predictions result of various methods: (a) input image, (b) ground truth, (c) U-Net, (d) dilated U-Net, (e) residual U-Net, and (f) dilated residual U-Net

In addition to class imbalance, the dataset also contains labeling errors in the ground truth masks, which further complicates model training and evaluation, as shown in Figure 3. For instance, in some test images, clearly visible trees in the input image are excluded from the vegetation class in the ground truth labels, as illustrated in Figure 3(a) and 3(b). The results produced by the U-Net, dilated U-Net, and residual U-Net are presented in Figures 3(c)–3(e), respectively. Despite these limitations, the proposed dilated residual U-Net demonstrates a notable ability to correctly infer the presence of vegetation in mislabeled regions, as shown in Figure 3(f).

Although the evaluation metrics may appear modest, they should be interpreted in the context of the dataset characteristics and task definition. The DroneDeploy segmentation dataset exhibits substantial class imbalance, ambiguous visual boundaries between vegetation and non-vegetation, and inconsistencies in ground truth annotations, all of which inherently limit achievable overlap-based metrics. Similar challenges have been reported in prior RGB-based vegetation and land-cover segmentation studies, where dice scores below 0.5 are common when fine-grained vegetation boundaries and noisy labels are present [13]. Within this constrained setting, the proposed dilated residual U-Net consistently outperforms all baseline architectures, indicating that the observed improvement reflects a meaningful relative gain.

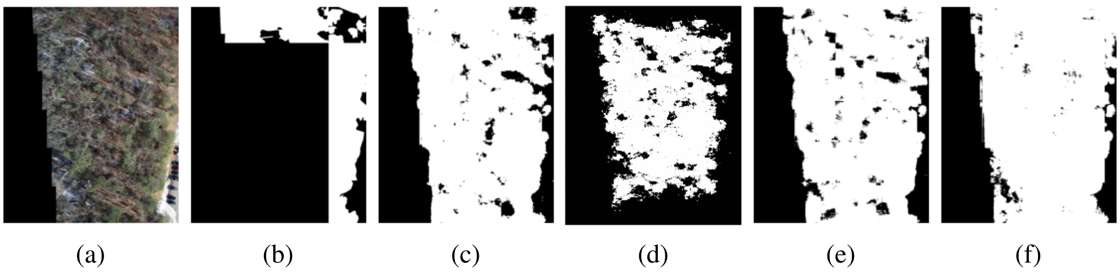

Figure 3. Predictions result of various methods with missing ground truth: (a) input image, (b) ground truth, (c) U-Net, (d) dilated U-Net, (e) residual U-Net, and (f) dilated residual U-Net

5. CONCLUSION

In this study, we proposed a novel deep learning architecture, namely dilated residual U-Net, which integrates residual connections and dilated convolutions into the U-Net framework to enhance semantic segmentation performance for vegetation detection. The residual connections help mitigate information loss by preserving low-level features across layers, while the dilated convolutions allow the model to capture wider contextual information without increasing the number of parameters excessively. Experimental results showed that this architecture achieved a higher dice coefficient score than standard U-Net, dilated U-Net, and residual U-Net, though at the cost of a modest increase in inference time due to the added module. Despite some inconsistencies in the ground truth annotations, such as unlabeled vegetation areas, the proposed model successfully identified these regions, demonstrating strong generalization capabilities. However, challenges remain in distinguishing grassy or bushy areas from true vegetation, primarily due to class imbalance and the limited semantic cues in RGB imagery.

Future work may focus on exploring class-rebalancing techniques and advanced loss functions, and incorporating auxiliary data such as elevation maps or multispectral imagery to improve semantic segmentation capability. Additionally, evaluating the model's performance across diverse environments and integrating domain adaptation methods could enhance its robustness for real world application.

Overall, this study demonstrates that modifying a U-Net architecture with both residual connections and dilated convolutions yields consistent segmentation gains for vegetation detection in high-resolution drone imagery, even under imperfect annotations and limited training data. While absolute performance remains constrained by dataset ambiguity and class imbalance, the proposed dilated residual U-Net offers a practical accuracy–efficiency trade-off, making it suitable for real-world vegetation monitoring scenarios where computational resources and annotation quality are limited.

ACKNOWLEDGMENTS

We would like to thank the Research and Innovation Services (RIS) Laboratory, Department of Computer Science, Sriwijaya University, for providing the necessary infrastructure.

FUNDING INFORMATION

The authors declare no funding involved.

AUTHOR CONTRIBUTIONS STATEMENT

The individual contributions of all authors are presented in Table 5. in accordance with the CRediT contributor roles framework.

Name of Author	C	M	So	Va	Fo	I	R	D	O	E	Vi	Su	P	Fu
Mgs. M. Luthfi Ramadhan	✓	✓	✓		✓	✓	✓		✓	✓			✓	
Rizal Maulana	✓	✓	✓		✓	✓	✓	✓	✓		✓			
Lalu Syamsul Khalid	✓	✓	✓	✓		✓	✓			✓	✓			

C : **C**onceptualization
M : **M**ethodology
So : **So**ftware
Va : **Va**lidation
Fo : **Fo**rmal Analysis
I : **I**nvestigation
R : **R**esources
D : **D**ata Curation
O : Writing - **O**riginal Draft
E : Writing - Review & **E**diting
Vi : **Vi**sualization
Su : **Su**pervision
P : **P**roject Administration
Fu : **Fu**nding Acquisition

CONFLICT OF INTEREST STATEMENT

The authors declare no conflict of interest.

DATA AVAILABILITY

This research make use of public a data,set namely the DroneDeploy Segmentation data which can be downloaded from: https://github.com/dronedeploy/dd-ml-segmentation-benchmark.

REFERENCES

[1] B. L. Boyle, W. Franklin, A. Burton, and R. E. Gullison, "Vegetation quality assessment: A sampling-based loss-gain accounting framework for native, disturbed and reclaimed vegetation," *Ecological Indicators*, vol. 158, p. 111510, 2024, [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1470160X23016527.

[2] D. Wen, X. Huang, H. Liu, W. Liao, and L. Zhang, "Semantic classification of urban trees using very high resolution satellite imagery," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 10, no. 4, pp. 1413–1424, 2017.

[3] K. Natarajan *et al.,* "Screening urban tree species for air quality enhancement: A review," *Asian Journal of Soil Science and Plant Nutrition*, vol. 10, no. 3, p. 511–518, Aug. 2024, [Online]. Available: https://journalajsspn.com/index.php/AJSSPN/article/view/363.

[4] X. Ren and X. Jiang, "Can carbon sequestration by terrestrial vegetation improve air quality? evidence from china," *Polish Journal of Environmental Studies*, 2025, doi: 10.15244/pjoes/195772.

[5] X. Ren and X. Jiang, "Can carbon sequestration by terrestrial vegetation improve air quality? evidence from china," *Polish Journal of Environmental Studies*, 2025, doi: 10.15244/pjoes/195772.

[6] C. J. Bradshaw *et al.,* "Brave new green world – consequences of a carbon economy for the conservation of australian biodiversity," *Biological Conservation*, vol. 161, pp. 71–90, 2013, [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0006320713000608.

[7] N. M. Haddad *et al.,* "Habitat fragmentation and its lasting impact on earth's ecosystems," *Science Advances*, vol. 1, no. 2, p. e1500052, 2015, [Online]. Available: https://www.science.org/doi/abs/10.1126/sciadv.1500052.

[8] S.-H. Lee, K.-J. Han, K. Lee, K.-J. Lee, K.-Y. Oh, and M.-J. Lee, "Classification of landscape affected by deforestation using high-resolution remote sensing data and deep-learning techniques," *Remote Sensing*, vol. 12, no. 20, 2020, [Online]. Available: https://www.mdpi.com/2072-4292/12/20/3372.

[9] D. O. Fuller, "Tropical forest monitoring and remote sensing: A new era of transparency in forest governance?" *Singapore Journal of Tropical Geography*, vol. 27, no. 1, pp. 15–29, 2006, [Online]. Available: https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1467-9493.2006.00237.x.

[10] M. Dalponte, L. Bruzzone, and D. Gianelle, "Fusion of hyperspectral and lidar remote sensing data for classification of complex forest areas," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 46, no. 5, pp. 1416–1427, 2008.

[11] M. M. L. Ramadhan, G. Jati, and W. Jatmiko, "Building damage assessment using feature concatenated siamese neural network," *IEEE Access*, vol. 12, pp. 19 100–19 116, 2024.

[12] A. Nițu, C. Florea, M. Ivanovici, and A. Racoviteanu, "Ndvi and beyond: Vegetation indices as features for crop recognition and segmentation in hyperspectral data," *Sensors*, vol. 25, no. 12, 2025, [Online]. Available: https://www.mdpi.com/1424-8220/25/12/3817.

[13] B. Ayhan and C. Kwan, "Tree, shrub, and grass classification using only rgb images," *Remote Sensing*, vol. 12, no. 8, 2020, [Online]. Available: https://www.mdpi.com/2072-4292/12/8/1333.

[14] L. Wang, Y. Duan, L. Zhang, T. U. Rehman, D. Ma, and J. Jin, "Precise estimation of ndvi with a simple nir sensitive rgb camera and machine learning methods for corn plants," *Sensors*, vol. 20, no. 11, 2020, [Online]. Available: https://www.mdpi.com/1424-8220/20/11/3208.

[15] P. Soglia, P. Gonzalez, I. Azpiroz, U. Otamendi, M. Quartulli, and S. Salata, "Estimating ndvi from sar images using conditional generative adversarial networks," in *IGARSS 2023 - 2023 IEEE International Geoscience and Remote Sensing Symposium*, 2023, pp. 5222–5225.

[16] Q. Zhu, Y. Zheng, Y. Jiang, and J. Yang, "Efficient multi-class semantic segmentation of high resolution aerial imagery with dilated linknet," in *IGARSS 2019 - 2019 IEEE International Geoscience and Remote Sensing Symposium*, 2019, pp. 1065–1068.

[17] G. Bilgin, S. Erturk, and T. Yildirim, "Segmentation of hyperspectral images via subtractive clustering and cluster validation using one-class support vector machines," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 49, no. 8, pp. 2936–2944, 2011.

[18] E. Maggiori, Y. Tarabalka, G. Charpiat, and P. Alliez, "Fully convolutional neural networks for remote sensing image classification," in *2016 IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*, 2016, pp. 5071–5074.

[19] A. F. I. Osman and N. M. Tamam, "Contrast-enhanced mri synthesis using dense-dilated residual convolutions based 3d network toward elimination of gadolinium in neuro-oncology," *Journal of Applied Clinical Medical Physics*, vol. 24, no. 12, p. e14120, 2023, doi: 10.1002/acm2.14120.

[20] M. A. Ortega-Ruíz, C. Karabağ, E. Roman-Rangel, and C. C. Reyes-Aldasoro, "Drd-unet, a unet-like architecture for multi-class breast cancer semantic segmentation," *IEEE Access*, vol. 12, pp. 40 412–40 424, 2024.

[21] Z. Zhang, Q. Liu, and Y. Wang, "Road extraction by deep residual u-net," *IEEE Geoscience and Remote Sensing Letters*, vol. 15, no. 5, pp. 749–753, 2018.

[22] Y. Zeng *et al.,* "Dctr u-net: automatic segmentation algorithm for medical images of nasopharyngeal cancer in the context of deep learning," *Frontiers in Oncology*, vol. 13, 2023, [Online]. Available: https://www.frontiersin.org/journals/oncology/articles/10.3389/fonc.2023.1190075

[23] N. R. Gudhe, H. Behravan, M. Sudah, H. Okuma, R. Vanninen, V.-M. Kosma, and A. Mannermaa, "Multi-level dilated residual network for biomedical image segmentation," *Scientific Reports*, vol. 11, no. 1, p. 14105, Jul 2021, doi: 10.1038/s41598-021-93169-w.

[24] D. Saadati, O. N. Manzari, and S. Mirzakuchaki, "Dilated-unet: A fast and accurate medical image segmentation approach using a dilated transformer and u-net architecture," *ArXiv*, vol. abs/2304.11450, 2023. [Online]. Available: https://api.semanticscholar.org/CorpusID:258298561.

[25] X. Jia, Y. Peng, J. Li, Y. Xin, B. Ge, and S. Liu, "Pyramid dilated convolutional neural network for image denoising," *Journal of Electronic Imaging*, vol. 31, no. 2, p. 023024, 2022, doi: 10.1117/1.JEI.31.2.023024.

[26] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," 2015. [Online]. Available: https://arxiv.org/abs/1505.04597.

[27] N. Flood, F. Watson, and L. Collett, "Using a u-net convolutional neural network to map woody vegetation extent from high resolution satellite imagery across queensland, australia," *International Journal of Applied Earth Observation and Geoinformation*, vol. 82, p. 101897, 2019. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0303243419302041.

[28] Z. Ivan, M. Konstantin, M. Nikolay, and M. Evgeniy, "Analysis of hardware-implemented u-net–like convolutional neural networks," in *Tools and Methods of Program Analysis*, R. Yavorskiy, A. R. Cavalli, and A. Kalenkova, Eds. Cham: Springer Nature Switzerland, 2024, pp. 52–63.

[29] A. Roy, A. Paul, P. Roy, R. Bhattacharyya, and A. Bhowmick, "Aerounet3+: An attention-aware unet3+ architecture for enhanced semantic segmentation of uav aerial imagery," in *2025 IEEE Guwahati Subsection Conference (GCON)*, 2025, pp. 1–8.

[30] D. Alhajim, K. Ansari-Asl, G. Akbarizadeh, and M. N. Soorki, "Improved lung nodule segmentation with a squeeze excitation dilated attention based residual unet," *Scientific Reports*, vol. 15, no. 1, p. 3770, Jan 2025, doi: 10.1038/s41598-025-85199-5.

[31] M. Mubashar, H. Ali, C. Grönlund, and S. Azmat, "R2u++: a multiscale recurrent residual u-net with dense skip connections for medical image segmentation," *Neural Computing and Applications*, vol. 34, no. 20, pp. 17 723–17 739, Oct 2022, doi: 10.1007/s00521-022-07419-7.

[32] D.-W. Kim, G. Jang, and H.-J. Kim, "Development of cnn-based semantic segmentation algorithm for crop classification of korean major upland crops using nia ai hub," *IEEE Access*, vol. 13, pp. 8425–8438, 2025.

[33] I. Ulku, "Reslmffnet: a real-time semantic segmentation network for precision agriculture," *Journal of Real-Time Image Processing*, vol. 21, no. 4, p. 101, May 2024, doi: 10.1007/s11554-024-01474-0.

[34] S. Zhang, J. Yue, X. Wang, H. Feng, Y. Liu, and M. Shu, "Segmentation and fractional coverage estimation of soil, illuminated vegetation, and shaded vegetation in corn canopy images using ccsnet and uav remote sensing," *Agriculture*, vol. 15, no. 12, 2025, [Online]. Available: https://www.mdpi.com/2077-0472/15/12/1309.

[35] S. P. G and P. N, "Texture image classification with dilated convolution layers," in *2023 International Conference on Wireless Communications Signal Processing and Networking (WiSPNET)*, 2023, pp. 1–5.

BIOGRAPHIES OF AUTHORS

Mgs. M. Luthfi Ramadhan is a lecturer in the Faculty of Computer Science at the Universitas Sriwijaya. He received his master degree from the Faculty of Computer Science at the Universitas Indonesia. He received his bachelor degree from the Faculty of Computer Science at the Universitas Sriwijaya. His research interests are in deep learning, pattern recognition, and computer vision. He can be contacted at email: luthfiramadhan@unsri.ac.id.

Rizal Maulana is a graduate student in Computer Science at Universitas Indonesia. He earned his Bachelor's degree in Informatics Engineering from Universitas Brawijaya, where he was actively involved in various extracurricular activities. His academic focus centers on Biomedical Image Analysis, Deep Learning, and Computer Vision. During his graduate studies, he published a conference paper in the field of Biomedical Image Analysis and served as a teaching assistant for the Advanced Machine Learning course for three consecutive semesters. He has also delivered invited talks on biomedical image segmentation, with a specific focus on brain MRI analysis. He can be contacted at: rizal.maulana01@ui.ac.id.

Lalu Syamsul Khalid holds a Master's degree in Computer Science from Universitas Indonesia and a Bachelor's degree in Informatics Engineering from Universitas Mataram. His academic and research interests include deep learning, computer vision, and remote sensing. He can be contacted at: lalu.syamsul@ui.ac.id.