U-Net Based Vegetation Segmentation for Urban Green Space (UGS) Regulation in Yogyakarta City

Delfia Nur Anrianti Putri
Computer Science Undergraduate Program
Faculty Of Mathematics And Natural Science
Sleman, Indonesia
delfia.nur2004@mail.ugm.ac.id

Raden Bagus Muhammad AdryanPutra Adhy Wijaya
Computer Science Undergraduate Program
Faculty Of Mathematics And Natural Science
Sleman, Indonesia
raden.bag2003@mail.ugm.ac.id

Nailfaaz
Computer Science Undergraduate Program
Faculty Of Mathematics And Natural Science
Sleman, Indonesia
nailfaaz@mail.ugm.ac.id

Wahyono
Department of Computer Science and Electronics
Faculty Of Mathematics And Natural Science
Sleman, Indonesia
wahyo@ugm.ac.id

***Abstract*—Urban Green Space (UGS) plays a vital role in maintaining the balance of urban ecosystems. In Yogyakarta City, urban development and an increase in population accompanied by a decrease and insufficiency in the proportion of UGS correlate with rising urban temperatures. Monitoring the availability of UGS can serve as a foundation for decision-making in managing the city's UGS. With advancements in technology and computing in the field of Computer Vision (CV), it's possible to recognize UGS from the geometric patterns of satellite images. This study proposes a vegetation segmentation method using satellite images and a Deep Learning (DL) method with the U-Net architecture to estimate the availability of UGS. Evaluation results show that the image dataset without preprocessing with the Red Green Blue (RGB) color channel gives an average Intersection over Union (IoU) of 80.74%. These results are better than data that underwent grayscaling and contrast stretching preprocessing. Based on a case study conducted in three sub-districts in the Yogyakarta region, the constructed U-Net model gave an average IoU evaluation of 60.88%. These results demonstrate the model's capability to provide a reference in calculating the UGS area at the sub-district level, which can then be utilized by the Yogyakarta City government to analyze the availability of UGS more efficiently. With this analysis, the government can make strategic decisions to manage UGS and create a sustainable urban environment.**

Keywords—Land Cover Classification, Green Space, Remote Sensing, Semantic Segmentation, Computer Vision

I. Introduction

Urban Green Space (UGS) supports various Sustainable Development Goals (SDGs). UGS can reduce the impact of climate change through the absorption of carbon dioxide [1] and reduce the Urban Heat Island (UHI) effect. In urban areas, UGS also plays a crucial role in maintaining biodiversity [2] and improving the health and well-being of the community [3]. Therefore, to realize a sustainable urban area, an adequate availability of UGS is required.

In Yogyakarta City, the availability of UGS still does not meet the standards based on Law Number 26 of 2007. In Article 29 paragraph (2) of this law, it is stipulated that the minimum proportion of UGS to maintain the balance of the city's ecosystem is 30% of the area's total size. In 2014, based on a field survey by [4], the total UGS area scattered across Yogyakarta City, as shown in Fig. 1, is 584.45 hectares or 17.78% of the city's area. Subsequently, in 2016, this availability decreased to 11.13% of the area's size. This percentage has continued to decline, recorded at 9.28% in 2019 [5] and 8.06% in 2022 [6].

The availability of UGS over time should be monitored because the impacts of changes in UGS conditions can be significantly observed in urban areas. The decrease in UGS availability in Yogyakarta City correlates with the rise in temperature, as marked by the UHI index based on [7]. This temperature increase is also in line with urban development, population growth, and greenhouse gas production that aren't accompanied by an increase in UGS. By knowing the extent of available UGS, the government can take actions in managing UGS availability, for example by purchasing land from the public, building city parks, organizing reforestation programs, and providing greening assistance to the community [8].

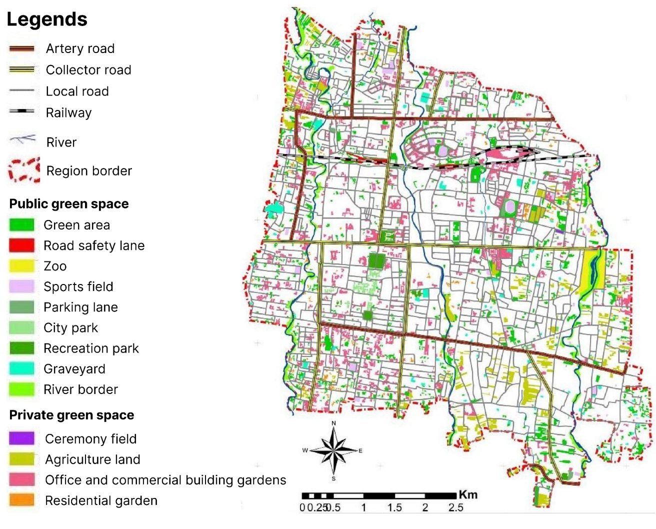


Fig. 1. Distribution of UGS in Yogyakarta City in 2014 based on [4].

Various methods can be employed to monitor the development of UGS availability. Data collection through field surveys and questionnaires provides detailed information [9], but this process is time-consuming and costly. To enhance efficiency, segmentation can be carried out on vegetated areas using satellite images. By comparing satellite images taken at different times, changes in the area and quality of urban green spaces can be observed [10].

Efficiency can further be improved by using Artificial Intelligence (AI), specifically through Computer Vision (CV), Machine Learning (ML), and Deep Learning (DL) methods. In this study, a solution is proposed to estimate the availability of city-level UGS by segmenting the vegetation of urban areas using a DL method with the U-Net architecture [11] on satellite images at the sub-district level. The accumulated UGS area in each sub-district can be utilized by the Yogyakarta City government in making decisions to manage or develop UGS, thereby fostering a sustainable urban environment.

In this study, we propose a novel approach to estimate the availability of UGS at the sub-districts level by segmenting urban area vegetation using a deep learning method with the U-Net architecture on satellite images. This application of deep learning in the context of UGS is innovative, particularly in its use of the U-Net architecture for vegetation segmentation, which has not been extensively explored. Our research contributes significantly to the field by offering a more accurate, efficient, and cost-effective method of monitoring UGS compared to traditional field surveys. The application of this model in Yogyakarta City demonstrates its potential impact on urban planning, providing a scalable and replicable solution for sustainable urban development. Furthermore, this research aids in the understanding of UGS dynamics in relation to urban development and climate change, offering valuable insights for policy-making and the achievement of Sustainable Development Goals. By enhancing the ease of monitoring UGS, our study supports governmental decision-making in urban planning and area development, contributing to the broader scientific community's efforts in fostering sustainable urban environments

II. Literature Review

Similar research using deep learning methods for segmentation has been conducted previously. In study [12], a comparison between ML (Machine Learning) and DL methods was made for vegetation segmentation on drone imagery data. From this study, the DL method produced an accuracy that was 4% better than ML. Research [13] also demonstrated that road infrastructure segmentation using DL achieved an accuracy performance of 97.7%.

Additionally, a study using the DeepGlobe dataset was conducted in research [14]. In that study, the AD-LinkNet and D-LinkNet34 models were proposed for road and land cover segmentation. Both architectures are developments from LinkNet, which primarily differs from U-Net in its use of residual modules to replace convolutions and a feature synthesis mechanism based on addition. The AD-LinkNet model yielded an Intersection over Union (IoU) evaluation of 64.73% for road segmentation and 47.86% for land cover segmentation.

Research using a similar dataset was also conducted by [21]. In this research, the GLNet model was applied to segment land cover. This architecture introduces a Global Local Network (GLNet) to perform segmentation combined with deep feature map sharing that can utilize local branches and global branches to perform segmentation. The local branch segmentation will enhance the segmentation results performed by the global branch. The GLNet achieved IoU performance of 71.60% for the land cover segmentation case.

A different research was conducted by study [22] by applying MagNet to perform segmentation. The model has a main core, namely segmentation model and refinement model at each processing stage. The refinement model has input in the form of cumulative previous segmentation results and segmentation results at that stage. The purpose of the refinement module is to use the last segmented image to refine the previous segmentation results at specific locations determined based on the uncertainty of the two segmentation map estimates. The MagNet model achieved a mean IoU performance of 72.96% in the case of land cover segmentation on DeepGlobe data.

III. Methods

Based on the flowchart in Fig. 2, this research is conducted in stages: data acquisition, data preprocessing, model training, model evaluation, and case study. The model training process is carried out on a computational device with specifications: CPU Intel(R) Core(TM) i7-6700 CPU @ 3.40GHz with 32 GB RAM, integrated with the GPU NVIDIA GeForce GTX 980Ti from the Intelligent Systems Laboratory, Department of Compute and Electronics Science, Faculty of Mathematics and Natural Sciences, Gadjah Mada University.

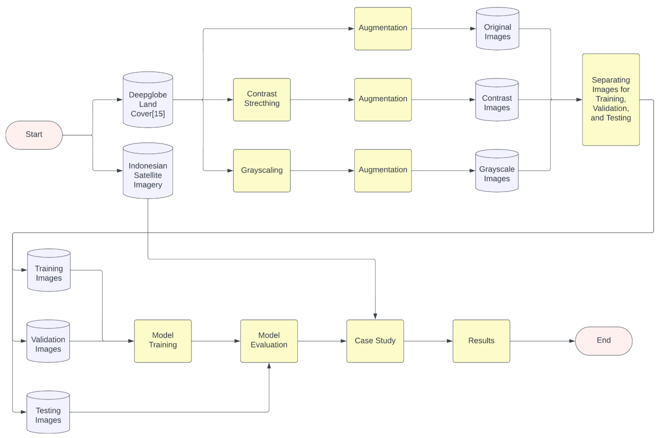


Fig. 2. Flowchart of the Research Method Conducted.

A. Data Acquisition

The dataset used for creating the segmentation model in this research is the DeepGlobe Land Cover Classification Dataset [15]. This dataset consists of 803 Red Green Blue (RGB) satellite images sized 2,448×2,448 pixels with a pixel resolution of 50 cm, along with corresponding land cover maps with samples shown in Fig. 3. The land cover labels or annotations provide information about the type of land cover, such as grassland areas (magenta), water bodies (dark blue), non-forest green areas and agricultural regions (yellow), built-up areas (cyan), forested regions (green), non-vegetative areas (black), and unidentified areas (white). Subsequently, an aggregation was performed on areas categorized as grasslands and other green areas with forest categories since the segmentation is focused on obtaining the extent of UGS. The legend for the built-up category was also modified from cyan to magenta. Details of the land use map after aggregation can be found in Fig. 4.

For the case study, satellite images and their labels were acquired from three sub-districts in Yogyakarta City, namely Bumijo, Kotabaru, and Suryatmajan, as listed in Table I. Data was obtained through screenshots on Google Earth with a camera altitude of 2,457 m. By cropping and converting the scale, images of 720×720 pixels in size with a resolution of 170 cm were obtained, in which this scale is equivalent to the pixel resolution of the training data. The annotation process or ground truth labeling of these satellite images was carried out using the Figma software, supervised by professionals in this field.

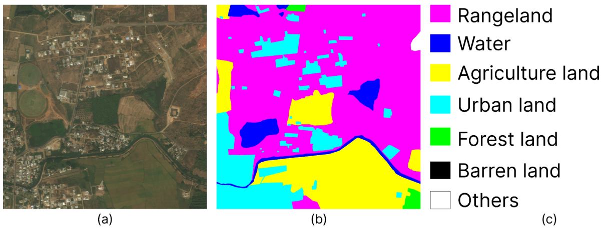


Fig. 3. DeepGlobe Land Cover Classification Dataset sample, (a) Satellite Image (b) Land Use Map (c) Land Cover Legend.

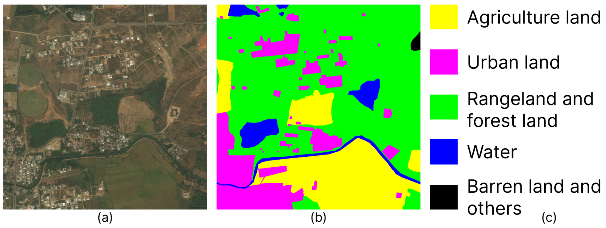
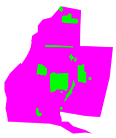
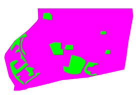


Fig. 4. Sample DeepGlobe Land Cover Classification Dataset after Aggregation and Label Modification, (a) Satellite Image (b) Land Use Map (c) Land Cover Legend.

TABLE I. Case Study Satellite Images and Land Cover Map

Map Category	**Sub-district**		
	Bumijo	***Kotabaru***	***Suryatmajan***
Satellite Imagery			
Land Cover Label			

B. Data Pre-Processing

The preprocessing methods employed in this research are color space conversion and contrast stretching [16], as illustrated in Fig. 5. Then, to add variation to the data, augmentation methods were used, consisting of rotations by 90°, 180°, and 270°, horizontal and vertical flips, as well as combinations of rotation and flips. Augmentation was applied to both the satellite images and the segmentation masks that serve as their labels. This augmentation process resulted in a total of 9,636 image and label data pairs. The dataset was divided into training/validation/testing sets, each consisting of 6,937/964/1,735 images with distribution splits of 72%/10%/18% respectively.

Fig. 5. Pre-Processing Methods Used on the Dataset, (a) RGB (b) RGB + Contrast Stretching (c) Grayscaling.

To determine the most suitable preprocessing method for segmentation, three distinct models were constructed. Each was trained using augmented data that underwent different preprocessing processes, as shown in Table II.

TABLE II. Model Type Based on Pre-processing Method

Model	**Pre-processing Methods**
1	Augmentation on RGB images
2	Augmentation and contrast stretching on RGB images
3	Augmentation and grayscaling

C. Modeling

The model used for semantic segmentation on the map is U-Net [11]. The input to the model is an image size 128×128, with channels adjusted based on the preprocessing method. 1 channel is used for grayscale images and 3 channels for RGB images. The U-Net architecture consists of an encoder and a decoder. The encoder comprises 5 convolution blocks producing feature maps, which are then utilized to generate a new image in the decoder layer, consisting of 4 upsampling blocks. Both blocks are interconnected by skip connections to prevent information loss during encoding. The model outputs a segmented image of size 128×128×5, where 5 indicates the number of prediction classes. An illustration of the used architecture can be found in Fig. 6.

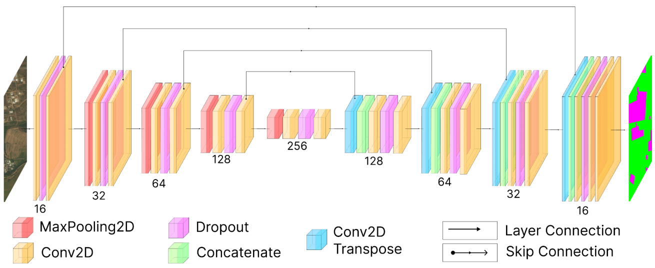


Fig. 6. U-Net Segmentation Model Architecture.

The entire model was trained using a batch size parameter of 16 for 2,000 epochs. This parameter was chosen based on previous experiments, where a lesser number of epochs still allowed for accuracy improvement, and stagnation occurred at higher epochs. Additionally, the use of a batch size either above or below this value resulted in extreme oscillations in loss and accuracy metrics. The Adam optimizer [17] was also used with a learning rate of 0.001 to prevent extreme oscillations in finding a convergence point.

D. Validation and Evaluation

Model validation was conducted to ensure that the model was neither overfitting nor underfitting and that it could generalize well to new data. Validation was carried out on the validation subset for all three models using the accuracy metric [18] and the categorical cross-entropy loss function [19]. The results from the model validation were utilized to enhance the model's performance during the training process through backpropagation.

Subsequently, evaluation was conducted to test the model using new data. The evaluation process was carried out on the three trained models using different preprocessing methods, applying the suitable preprocessing methods on the test data subset. This evaluation process involved calculating the IoU (Intersection over Union) [18], which is the proportion of the area predicted by the model that intersects with the actual labeled area. The IoU for each image was calculated and then averaged to obtain the mean IoU. A high average IoU indicates good performance.

E. Case Study

The case study was conducted to observe the model's performance on the Yogyakarta City map by performing vegetation segmentation inference on several sub-districts in Yogyakarta City, namely Kotabaru, Bumijo, and Suryatmajan sub-districts. Before inference, the dimensions of the images were adjusted to match the model's input size. The inference process was conducted on satellite image data for the three sub-districts using the U-Net model with the preprocessing method that provided the most optimal results during the evaluation process. This was followed by the removal of areas outside the sub-districts and aggregation of labels, where all UGS areas were changed to green and non-UGS areas to magenta. Subsequently, an evaluation was conducted by calculating the IoU (Intersection over Union) between the predicted images and the ground truth. Then, the availability of UGS was calculated based on segmentation results by finding the ratio of vegetation area pixels to the sub-district area pixels and multiplying it by the total area of the sub-district.

IV. Results and Discussion

A. Optimal Pre-Processing

To obtain the most optimal model for the Indonesian region, an analysis was conducted on the model evaluation results on the test data subset using the mean IoU metric and a visual analysis of the segmentation results on the Indonesian region images. The evaluation conducted on the three models produced mean IoU results of 80.74% for Model 1, 78.26% for Model 2, and 78.54% for Model 3.

TABLE III. Model Evaluation Results on the Test Data Subset

Model	Pre-processing Methods	Mean IoU
1	Augmentation on RGB images [11]	80.74%
2	Augmentation and contrast stretching on RGB images	78.26%
3	Augmentation and grayscaling	78.54%

Based on the visual analysis, there are clear differences in the predictions of each model as shown in Table III and Fig. 7. Predictions from the model with augmentations on the original image visually resemble the ground truth more closely compared to the other models. The preprocessing method of contrast stretching delivered evaluations below the original images because there is a significant color difference with the images used to train the model.

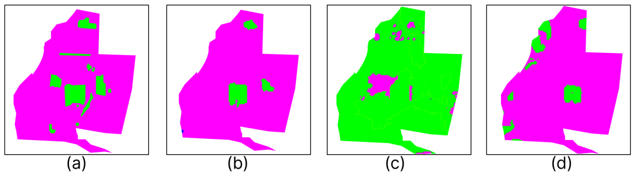


Fig. 7. Model Prediction Results on Bumijo Sub-district Image, (a) Ground Truth (b) RGB with Augmentation (c) RGB + Contrast Stretching (d) Grayscaling.

Furthermore, Model 1 exhibited better performance compared to the D-LinkNet34 and AD-LinkNet models constructed in the study [14], GLNet model constructed in the study [21] and MagNet model constructed in the study [22], with the IoU evaluation values listed in Table IV. That research also used a similar dataset in performing land cover segmentation. This evaluation indicates that the constructed model outperforms those from previous research. One underlying reason for this is that the U-Net model we designed is adaptive to the limitations of the amount of data, more so than the other four models, as evidenced by its widespread use in medical image segmentation with limited data [20]. In research [21], there was no augmentation process and that is the reason our model performs better. Nevertheless, there are differences in the data acquisition methods. The classes used in that data were not aggregated in the research [14], [21] and [22]. This could also be a factor influencing the evaluation results.

TABLE IV. Performance Comparison with Other Architectures

Architectures	Mean IoU
D-LinkNet34 [14]	47.51%
AD-LinkNet [14]	47.86%
GLNet [21]	71.60%
MagNet [22]	72.96%
U-Net [11]	80.74%

B. Model Training

In this section, we review the training process results for Model 1, which provided the most optimal evaluation outcome. The change in the loss of Model 1 during the training, as presented in Fig. 8, is divided into three phases. In the first phase, which is the initial 50 epochs, the model is adjusting to the training data, resulting in a significant reduction in both training and validation loss. The rate of reduction slows down in the subsequent phase, from after the 50th epoch up to the 250th epoch, indicating the beginning of convergence. In the following phase, the relatively minor changes signify that convergence has been achieved, where the loss values hover around 0.3 for validation and 0.1 for training.

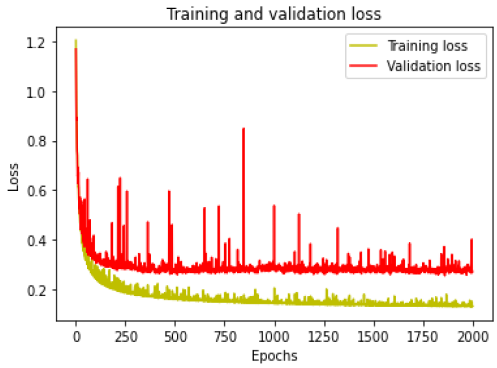


Fig. 8. Training and Validation Loss During Training Phase.

C. *Case Study Analysis*

The evaluation was conducted by comparing prediction results with the ground truth, as shown in Table V. The ground truth and satellite images of the three sub-districts refer to Table I. Applying the model to satellite images of the three sub-districts in Yogyakarta City yielded IoU evaluation results above 50%. The IoUs for the segmentation of Bumijo, Kotabaru, and Suryatmajan sub-districts are 69.92%, 56.74%, and 56.00%, respectively. This shows that vegetation segmentation using our method has achieved good visual accuracy and alignment with the actual vegetation boundaries. However, misclassified pixels still exist, particularly in challenging scenarios such as densely vegetated areas or regions with complex interclass boundaries.

Based on the segmentation results for each sub-district, a conversion was made from the pixel count indicating UGS to the area of UGS and the percentage of UGS area, which is displayed in Table VI. These conversion results indicate that the UGS area in Bumijo sub-district is 4.27 hectares, in Kotabaru sub-district is 6.63 hectares, and in Suryatmajan sub-district is 6.76 hectares.

TABLE V. Vegetation Segmentation Results in the Case Study

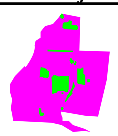
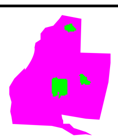
Map Category	**Sub-district**		
	Bumijo	***Kotabaru***	***Suryatmajan***
Land Cover Label			
Prediction Result			

TABLE VI. UGS Area Calculation and Evaluation in the Case Study

Sub-district	**Mean IoU**	**UGS area**	**% UGS**
Bumijo	69,92%	4,27 Ha	5,34%
Kotabaru	56,74%	6,63 Ha	9,30%
Suryatmajan	56,00%	6,76 Ha	24,16%

The evaluation shows that the constructed U-Net model is capable of distinguishing areas classified as UGS in each of those sub-districts. These results indicate the effectiveness of our approach in accurately delineating vegetation areas. Given that the visual characteristics of land cover tend to be similar across regions in Indonesia, especially for green areas and residential categories, the model can be applied to satellite image data across the country. This indicates the potential for using the model to efficiently calculate the availability of UGS in other areas by leveraging available satellite image data.

V. Conclusion

The evaluation results indicate that training with original RGB satellite images performs better than with grayscale images and RGB with contrast stretching preprocessing, achieving an average IoU of 80.74%. The performance of the constructed U-Net model achieved an average IoU value of 60.88% on the satellite images of Yogyakarta City (specifically in the sub-districts of Kotabaru, Bumijo, and Suryatmajan). It can be concluded that this research successfully developed a U-Net model to estimate the UGS area from training dataset satellite images as well as satellite images at the sub-district level in Yogyakarta City. By accumulating these UGS area estimates, the model has the potential for application at the city level and even broader regions. Such accumulations can assist regional leaders in making decisions related to UGS management. Future research could implement transfer learning with satellite image training data from regions across Indonesia, so that the constructed model can better adapt to the geometric patterns of Indonesian regions.

Acknowledgment

We express our gratitude to Dr.Sc. Sanjiwana Arjasakusuma, S.Si., M.GIS. and Dr. Eng. Guruh Samodra, S.Si., M.Sc. from the Faculty of Geography, Gadjah Mada University, for their supervision in creating the land cover map to conduct the case study in Yogyakarta City.

References

[1] S. Bi, F. Dai, M. Chen, and S. Xu, "A new framework for analysis of the morphological spatial patterns of urban green space to reduce PM2.5 pollution: A case study in Wuhan, China," *Sustainable Cities and Society*, vol. 82, p. 103900, Jul. 2022, doi: 10.1016/j.scs.2022.103900.

[2] R. W. F. Cameron *et al.*, "Where the wild things are! Do urban green spaces with greater avian biodiversity promote more positive emotions in humans?," *Urban Ecosyst*, vol. 23, no. 2, pp. 301–317, Apr. 2020, doi: 10.1007/s11252-020-00929-z.

[3] M. Jabbar, M. M. Yusoff, and A. Shafie, "Assessing the role of urban green spaces for human well-being: a systematic review," *GeoJournal*, vol. 87, no. 5, pp. 4405–4423, Oct. 2022, doi: 10.1007/s10708-021-10474-7.

[4] A. Ratnasari, S. R. P. Sitorus, and B. Tjahjono, "PERENCANAAN KOTA HIJAU YOGYAKARTA BERDASARKAN PENGGUNAAN LAHAN DAN KECUKUPAN RTH," *TATALOKA*, vol. 17, no. 4, pp. 196–208, Nov. 2015, doi: 10.14710/tataloka.17.4.196-208.

[5] Triyono, A. Maryono, C. Fandeli, and P. Setyono, "Reliability analysis of water supply based on green open space (case study of Yogyakarta city)," *AIP Conference Proceedings*, vol. 2202, no. 1, p. 020115, Dec. 2019, doi: 10.1063/1.5141728.

[6] "SIPSN - Sistem Informasi Pengelolaan Sampah Nasional." https://sipsn.menlhk.go.id/sipsn/public/rth (accessed Jun. 29, 2023).

[7] S. Fathurrohmah and A. C. Kurniati, "Dinamika Urban Heat Island di Kawasan Perkotaan Yogyakarta," *ReTII*, pp. 619–623, Nov. 2022.

[8] R. Kumalasari, "Strategi Perluasan Ruang Terbuka Hijau Publik di Kota Yogyakarta," Universitas Gadjah Mada, 2020. Accessed: Jun. 30, 2023. [Online]. Available: https://etd.repository.ugm.ac.id/penelitian/detail/187087

[9] D. A. K. Sari, L. F. Widyawati, and D. Pramesti, “The availability and role of urban green space in South Jakarta,” *IOP Conf. Ser.: Earth Environ. Sci.*, vol. 447, no. 1, p. 012055, Feb. 2020, doi: 10.1088/1755-1315/447/1/012055.

[10] C. Huang *et al.*, “Mapping the maximum extents of urban green spaces in 1039 cities using dense satellite images,” *Environ. Res. Lett.*, vol. 16, no. 6, p. 064072, Jun. 2021, doi: 10.1088/1748-9326/ac03dc.

[11] O. Ronneberger, P. Fischer, and T. Brox, “U-Net: Convolutional Networks for Biomedical Image Segmentation,” in *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015*, N. Navab, J. Hornegger, W. M. Wells, and A. F. Frangi, Eds., in Lecture Notes in Computer Science. Cham: Springer International Publishing, 2015, pp. 234–241. doi: 10.1007/978-3-319-24574-4_28.

[12] S. Bhatnagar, L. Gill, and B. Ghosh, “Drone Image Segmentation Using Machine and Deep Learning for Mapping Raised Bog Vegetation Communities,” *Remote Sensing*, vol. 12, no. 16, p. 2602, Jan. 2020, doi: 10.3390/rs12162602.

[13] N. Y. Q. Abderrahim, S. Abderrahim, and A. Rida, “Road Segmentation using U-Net architecture,” in *2020 IEEE International conference of Moroccan Geomatics (Morgeo)*, May 2020, pp. 1–4. doi: 10.1109/Morgeo49228.2020.9121887.

[14] M. Wu, C. Zhang, J. Liu, L. Zhou, and X. Li, “Towards Accurate High Resolution Satellite Image Semantic Segmentation,” *IEEE Access*, vol. 7, pp. 55609–55619, 2019, doi: 10.1109/ACCESS.2019.2913442.

[15] I. Demir *et al.*, “DeepGlobe 2018: A Challenge to Parse the Earth through Satellite Images,” presented at the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), IEEE Computer Society, Jun. 2018, pp. 172–17209. doi: 10.1109/CVPRW.2018.00031.

[16] S. Yelmanov and Y. Romanyshyn, “A New Approach to Image Enhancement by Non-Linear Contrast Stretching,” in *2020 IEEE Third International Conference on Data Stream Mining & Processing (DSMP)*, Aug. 2020, pp. 178–184. doi: 10.1109/DSMP47368.2020.9204129.

[17] S. Bock and M. Weiß, “A Proof of Local Convergence for the Adam Optimizer,” in *2019 International Joint Conference on Neural Networks (IJCNN)*, Jul. 2019, pp. 1–8. doi: 10.1109/IJCNN.2019.8852239.

[18] W. Yan, C. Chen, and D. Zhang, “U-Net-based medical image segmentation algorithm,” in *2021 13th International Conference on Wireless Communications and Signal Processing (WCSP)*, Oct. 2021, pp. 1–5. doi: 10.1109/WCSP52459.2021.9613447.

[19] A. H. Mostafa, H. Abdel-Galil, and M. Belal, “Ensemble Model-based Weighted Categorical Cross-entropy Loss for Facial Expression Recognition,” in *2021 Tenth International Conference on Intelligent Computing and Information Systems (ICICIS)*, Dec. 2021, pp. 165–171. doi: 10.1109/ICICIS52592.2021.9694244.

[20] N. Siddique, S. Paheding, C. P. Elkin, and V. Devabhaktuni, “U-Net and Its Variants for Medical Image Segmentation: A Review of Theory and Applications,” *IEEE Access*, vol. 9, pp. 82031–82057, 2021, doi: 10.1109/ACCESS.2021.3086020.

[21] W. Chen, Z. Jiang, Z. Wang, K. Cui and X. Qian, "Collaborative Global-Local Networks for Memory-Efficient Segmentation of Ultra-High Resolution Images," in *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, Long Beach, CA, USA, 2019, pp. 8916-8925, doi: 10.1109/CVPR.2019.00913.

[22] C. Huynh, A. T. Tran, K. Luu and M. Hoai, "Progressive Semantic Segmentation," in *2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, Nashville, TN, USA, 2021, pp. 16750-16759, doi: 10.1109/CVPR46437.2021.01648.