



Classification of tea plantation using orthomosaics stitching maps from aerial images based on CNN

Andri Agustav Wirabudi^{1,*}, Nurwan Reza Fachrurrozi²

^{1,2}Departement of Telecommunication Engineering, Institut Teknologi Telkom Jakarta

¹Departement of Intelligent Media Engineering, Hanbat National University

^{1,2}Jl. Daan Mogot, KM. 11, Jakarta Barat 11710, Indonesia

¹125 Dongseo-daero, Yuseong-gu, Daejeon, South Korea

*Corresponding email: andriagustav@ittelkom-jkt.ac.id

Received 15 December 2022, Revised 13 January 2023, Accepted 31 January 2023

Abstract — Indonesia is the largest tea-producing country in the world. With the large area of tea plantations, it is difficult for the planters to monitor and estimate harvest time. Therefore, drone technology, or unmanned aerial vehicle, is needed. Image results obtained by drones can help estimate the harvest period that occurs. Tea leaf classification using the orthomosaics stitching imagery from aerial photographs based on the convolutional neural network. In this study, we took a sample of 20 hectares. The image classification was grouped based on 5 types: ready-to-harvest tea leaves, medium tea leaves, old tea leaves, and environmental conditions impacting crop yields. The success of image recognition is shown in the error matrix data by testing 123 random points spread across the map, where 113 random points were identified with an average accuracy of 91.87%. This value is certainly very good and exceeds the specified success threshold of 75%.

Keywords – convolutional neural networks, tea plantation mapping, unmanned aerial vehicle

Copyright ©2023 JURNAL INFOTEL
All rights reserved.

I. INTRODUCTION

Indonesia is one of the largest tea-producing countries in the world, with total tea production reaching more than 100,000 metric tons per year, with a plantation area of 112,308 ha (hectare) in 2020. Such an area is still an obstacle when the harvest season arrives. This is because planters in Indonesia still use manual methods in harvesting the tea they get, which has an impact on the estimated number of harvests that can be obtained and many of the trees that lack nutrients and monitoring from planters, which results in a decrease in the number of trees each year.

In previous research, we managed to make a monitoring drone that is useful for taking aerial images to be mapped. Still, our previous research only focused on the map quality of the drones we tested. As the largest tea-producing country in the world, with a very large area, tools are needed to help monitor the tea plantation area as a whole. Unmanned Aerial Vehicle (UAV) washing was chosen as the solution for the monitoring

process. Calculating the optimal flight path is necessary to produce good-quality images and influence power consumption. The algorithms proposed in this study are Dynamic Programming and Kruskal's Algorithm. The application of these two network algorithms is expected to find the optimal path in aerial photography. Experimental results show that the algorithm produces the optimal path, and the power consumption is more efficient than conventional lines. Image data obtained while monitoring tea plantations produce high-quality images, with an accuracy of each map above 90% and an error assumption below 5%. Currently, image classification is widely used in the field of remote sensing to monitor objects, land contours, mineral resources, and many others. This classification method is even used before planning a project to be worked on.

The purpose of this study was to classify images based on maps generated by drones in previous studies [1]. The method used is Convolutional Neural Networks (CNN) to identify the parameters seen, namely

the condition of the tea leaves, estimated yields obtained, and monitoring areas where no trees are caused by tree death. In this study, we took a sample of 20 ha. CNN will be used as a method to identify each image sample on the map used. The CNN takes advantage of the convolution process by moving the convolution kernel (filter) of a certain size according to the object samples used in this study. Parameters classified are young tea leaves, middle tea leaves, old tea leaves, roads, and damaged areas. After the sample is obtained, the system will calculate the appropriate information based on the similarity of the sample given and classify it according to the group [2], [3].

For a better understanding, the rest of this paper is organized as follows. Section II discusses the related work and the proposed method, while the experimental results are presented in section III. Finally, the conclusion is shown in section IV.

II. RESEARCH METHOD

This section discusses related work, proposed method, image sampling with UAV, orthomosaics stitching maps, sample classification, and convolutional neural network.

A. Related Work

Our previous research investigated the workings of drones to take pictures in the air. Still, in this study, we only focused on the flight path of the drone and the parameters measured, namely the power consumption of the drone and the flight path of taking pictures. At the time of sampling images in the air, the results obtained from the study were only limited to an ordinary map, as shown in Fig. 3. Therefore, in this follow-up study, we tried to develop in terms of images that had been obtained during previous research by focusing on image classification and image stitching.

B. Proposed Method

To get perfect results, research will utilize the image of the UAV as the basis of the map, which will later be combined or stitched together images.

The mapped image results undergo an image classification process to determine the information in the map and classify it by type. The area of the tea plantation tested was 20 ha, and the threshold of the images captured by the UAV was 5% of the total captured area of about 1 ha. If an image produced by a UAV has more than 5% error, it does not meet the requirements for classification. This error bound is determined based on the capture performance of the drone camera when captured in Fig. 2. The resolution used is 4096 x 2160 for each image captured by the drone.

Classifying tea plantation areas and estimating the yields obtained are the aims of this study. Fig. 1 shows the classification process that took place during this

study. Breaks this into five steps to convert drone imagery to base map format. Then proceed to the first process of stitching the image onto the map using the orthomosaic method. The second stage is image classification based on the type of information contained in the card. The third stage is multiplying the image with a mask or filter using convolution calculations. The fourth step is identifying the classification results obtained and calculating the areas found. Finally, the last stage is Stage 5, which makes the classification results of the card's class.

C. Image Sampling with UAV

The images were taken using a quadcopter UAV with a drone height of 80m. This height is determined based on the calculation of power consumption in previous research. The UAV takes each sample from the image determined by the program [4]. The results obtained from each experiment have a resolution of 4096 x 2160 for each image. This experiment was carried out 10 times by looking at the map results obtained, as shown in Table 1. The area that can be taken by the UAV is $75 m^2 \times 75 m^2$, as can be seen in Fig. 2.

Table 1. Total Image Results Obtained by UAV Altitude 80M Previous Research

1	2	3	4	5	6	7	8	9	10
126	122	118	118	125	127	128	126	122	127

D. Orthomosaics Stitching Maps

Orthomosaics is a method of image stitching that utilizes aerial photographs made of mosaics. The mosaic itself is a photo arranged based on the flying direction of the drone with a path that does not form a loop and has been calculated beforehand. The results of the photos obtained have been corrected in terms of geometry with a uniform scale in all areas and a fixed distortion value [5]–[7].

The image data was obtained as part of the camera capture from the UAV according to the flight pattern used. In general, the stitching process is divided into several stages. Judging from (1), the first step is to use the image pixels taken by the UAV by equating the shooting motion of the UAV so that the images will be arranged in order. Then, the second step is to determine the spatial transformation and simultaneously wrap the target image into the reference image [8].

The last step is to cut and blend the images in an arrangement that matches the order they were taken, as shown in Fig. 2. Our image merging process only focuses on the information achieved when it has become a map with high-quality resolution and perfect precision. The map that has been generated will be calculated by CNN for the next stages of finding the classification of tea leaves. The results of the 10 experiments that have been carried out are selected maps that have errors that do not exceed 5%, as shown in Fig. 3. In the image stitching phase, we have to sort

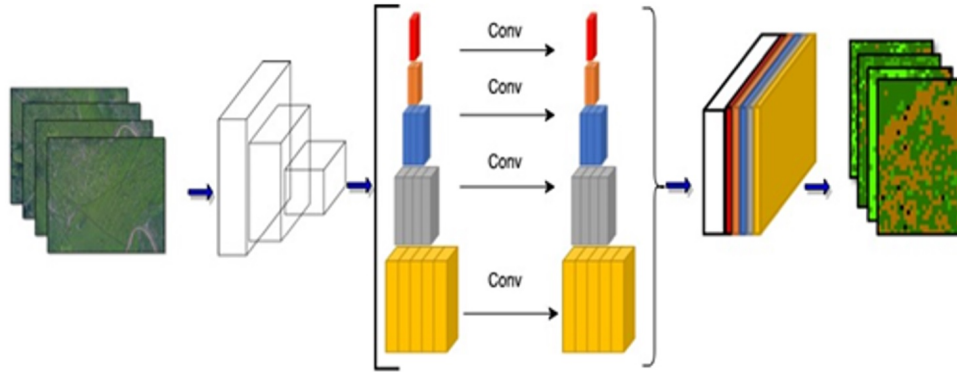


Fig. 1. Diagram classification process.

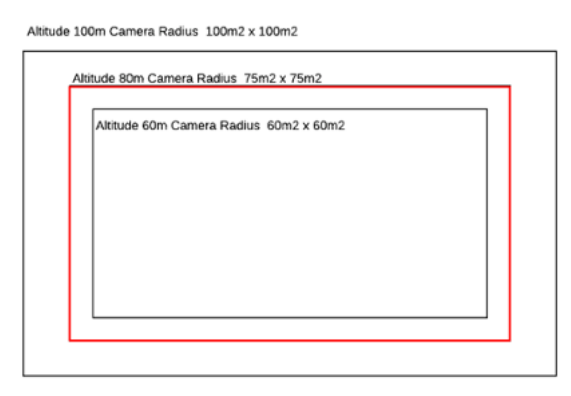


Fig. 2. Camera radius.

the pictures according to the order in which they were shot while in the air to get the required matching point. It is known that the input images and the match point $x_i = (x_i, y_i)T, y_i = (x'_i, y'_i)T, i = 1, \dots, N$ homographic transformation between two images $y = h(x)$ can be represented as (1) and (2):

$$h_x(x) = \frac{h_1x + h_2y + h_3}{h_7x + h_8y + h_9} \tag{1}$$

$$h_y(x) = \frac{h_4x + h_5y + h_6}{h_{10}x + h_{11}y + h_{12}} \tag{2}$$

where h_x and h_y : is a sequence of images from the drone that is stitched into a map using orthomosaic, namely stitching images based on the sequence of images taken by the drone starting from $h_1x + h_2x \dots n$

E. Sample Classification

After the process of stitching the image into a map, the next step is to classify the image in order to separate the information contained in the map. This process has several stages, including taking samples from each part of the map; and grouping the sample results according to the information that has been determined. Such as ready-to-harvest tea leaves, middle tea leaves, old tea leaves, and damaged plantation areas [9], [10].

The classification used is the guided classification Maximum Likelihood Classification (MLC). In classification using MLC, this practice area should be used to see the statistical characteristics of each classified category. The classification process with MLC is based on calculating the probability density for each land cover category. Probability calculation, also known as probability, aims to find pixels of a class, which can be described using (3):

$$P(i|x) = \frac{(P(x|i)P(i))}{(P(x))} \tag{3}$$

where $P(i|x)$ is the conditional probability of a class, which is calculated provided that the vector x is unconditional, while $P(x|i)$ is the conditional probability of the vector x , which is calculated by the unconditional vector, $P(i)$ itself means the probability of class I emerging from an image, and $P(x)$ is the probability vector x .

Region of interest (ROI) is a way of taking samples from map results that have been done with the previous image stitching process. This stage is carried out by looking at the results of visual image interpretation, namely high-resolution maps, and monitoring the results of the ground check survey. This process is carried out to determine the amount of information from the conditions to be classified. One point of the ground check can represent one type of classification information. The number of sample polygons taken in this study is at least 3 polygon samples, and in each polygon, there are at least 9 image pixels with similarity and color uniformity. If the color is different, it will impact the classification process [11], [12]. Based on the results of the ground check, five different types of information were obtained, as shown in Table 2.

F. Convolutional Neural Network (CNN)

In this study, CNN utilizes the convolution process by moving a convolution filter into the map. After that, this approach generates new representative information

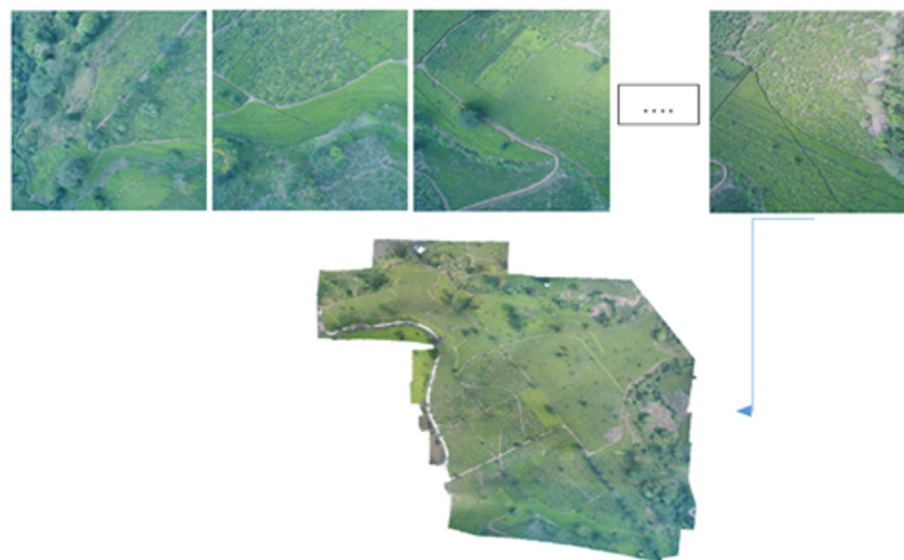
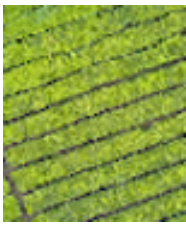





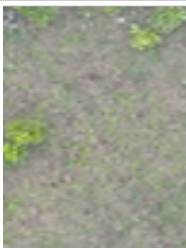

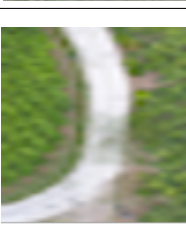



Fig. 3. Illustration of stitching images.

Table 2. Ground Check and Type of Information

No	On Maps	Ground Check	Type Information
1			Harvest time tea
2			Middle leaf tea
3			Old leaf tea
4			Ground
5			Road

by multiplying the part of the image with the filter used and the sample from the classification [15], [16].

Each sample image obtained at the sample classification stage is used as input to produce a classification object. Whenever the object's position appears on the map, it will be marked as the same unit. This process is carried out for all parts of each image that has been obtained in Table 11. Thus, each part of the image in the map will have the same multiplier factor, or in the context of a neural network, referred to as weight sharing. If there is something that looks interesting or has something in common with the classification sample, it will be marked as an object of interest.

After that, the results of the selected image will be carried out in a down-sampling process to reduce the size of the array by taking the largest pixel value from each kernel so that even though it reduces the number of parameters, important information from each part is still taken. After the down-sampling process is complete, the next step is to make predictions for each decision from the values in the array by matching the samples taken with the results that have been calculated using CNN. If there is a similarity between the sample and the calculation results, a different colour group will be assigned [13], [14].

Table 3 shows the overall probability results of each sample on the map that has been approached through this CNN from the results that show the similarity of the data with the sample used.

Table 3. Result Convolution for Each Sample

No	Class	Sample Value	Overall Probability (pixel)
1	Harvest Time Tea	16	2216130
2	Middle Leaf Tea	17	2531164
3	Old Leaf Tea	27	920589
4	Ground	44	77365
5	Road	55	5076824

III. RESULT

This section discusses identification result and confusion matrix.

A. Identification Result

At this stage, the results of the experiments that have been carried out will be discussed. The results obtained during the research showed that there was an approach using the classification method to separate information between leaves. The resulting image is made into a map with guided classification to identify the information in the map. Fig. 4 shows a map with an area of 20 ha that has been successfully stitched using the orthomosaics stitching method, namely, image stitching by utilizing geometric information from images taken by the UAV. From the map results, the information needed is classified according to the sample in Table 2. The classification results in Table 3 are similar to the samples taken starting from Harvest

time tea. Convolution with 16 samples can produce 2216130 identical pixels; Middle leaf tea, with 17 samples, yields 2531164 identical pixels; Old leaf tea, with 27 samples, can produce 920589 identical pixels; land or areas where no trees grow, with 44 samples resulting in 77365 identical pixels; and road tea, where 55 samples yield 5076824 identical pixels.



Fig. 4. Result map after the stitching process.

This guided classification uses map results of stitching data as a basic reference point to determine the classes to be classified [17], [18]. The training area that has been created has 5 classes and types of information that will be classified. The map of the classification results using the MLC method is shown in Fig. 5. The identification of the land cover area of this MLC method can be seen in Table 4.

Table 4. Result Identification

No	Class	Area (Ha)	Percentage %
1	Harvest Time Tea	4.41	21
2	Middle Leaf Tea	4.83	23
3	Old Leaf Tea	9.87	47
4	Ground	1.6	8
5	Road	0.21	1

The results of identification using the MLC method showed that the area of the garden that was successfully classified was 47% or 9.87 ha for old tea leaves, 23% or 4.83 ha for middle tea leaves, and 21% or 4.41 ha for ready-to-harvest tea leaves of the total value and the percentage of vacant or not covered with tea trees is 9% or 1.89 ha. where the total death of tea trees is 8% or 1.6 ha, as shown in Fig. 6. The map results that have been classified in Fig. 5, when viewed from the image of the uneven distribution of trees, cause the tree area to become untidy and poorly organized. This will impact the harvest that can be obtained in the following season.

B. Confusion Matrix

This confusion matrix is the final stage in the classification to determine the accuracy of the classification data that has been made. This accuracy test takes samples from the probabilities encountered when the convolution process occurs, as shown in Table 3 [18], [19].

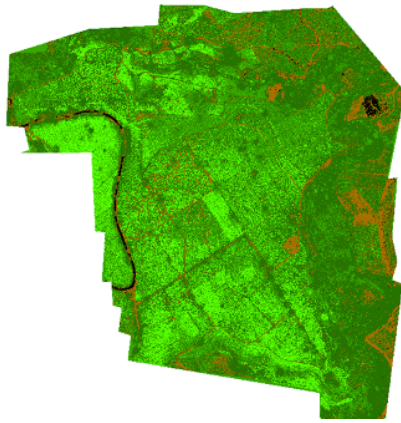


Fig. 5. Result map classification.

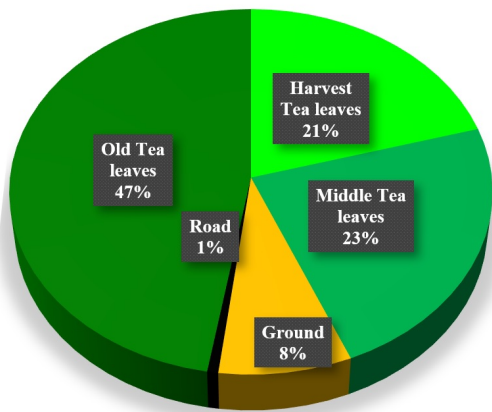


Fig. 6. Result chart classification.

Based on the observations that have been made, we use random points to test the accuracy of the classification data that has been carried out, it is known that 123 test points are spread on the classification map, and several points are the same as the results of the classification that occurred and as for points that are not the same as classification result and make this an error. Errors during the classification process are due to the types of information that have similar colors and hues, making it difficult to distinguish the same or different information. Errors or errors that occur are then entered into the error matrix.

Based on Table 5, the test with a random point of 123 successfully detected 113 points, it can be seen that the accuracy of the classification data for each class has 16 test points that have succeeded in detecting harvest time conditions on class 1 tea leaves, in Middle tea class 2 there are 21 test points with the success of 19 and error of 2 points, Old tea Class 5 there were 36 points that were successfully detected and 6 points of error out of a total of 42 points, testing of vacant land or ground class 3 30 times of testing were 29 of them managed to detect and error of 1 point and the last one is detecting road as many as 14 points which have been successfully detected by 13 points and failure by 1 point. When viewed from these results, the largest number of errors is in detecting old tea leaves. This error is caused due to the similarity of the pixel sample

with the surrounding area, which results in incorrect detection [20]–[22].

Table 5. Error Matrix

No	Class	Class					Ground Truth
		1	2	3	4	5	
1	Harvest Tea Time	16	0	0	0	1	17
2	Middle Leaf Tea	0	19	0	0	5	24
3	Ground	0	0	29	1	0	30
4	Road	0	0	0	13	0	13
5	Old Leaf Tea	0	2	1	0	36	39
Total		16	21	30	14	42	123

In Table 6, the confusion matrix can provide important information to determine the accuracy value of the data that has been obtained. The value of the overall accuracy is the value that will be used as a feasible and inappropriate decision value for a map that has been calcified with a specified threshold of at least 75% [23], [24].

Table 6. Result Identification

No	Class	Accuracy Sample
1	Harvest time	100.0 %
2	Middle leaf tea	90.5 %
3	Ground	96.7 %
4	Road	92.9 %
5	Old leaf tea	85.7 %
Total		91,87%

The accuracy value obtained using this method is 91.87% in Table 6. This value is obtained from the average number of classification results presentations from each class. The results of this accuracy show a map image that meets the requirements and with an accuracy level above 75% of the specified limit, close to the results according to real conditions.

IV. CONCLUSION

This research develops a method of stitching aerial images from drones based on image classification to predict. Utilizing image stitching can make it easier to monitor land better and estimate the harvest period with the classification method. Basic images that have been geometrically stitched using the orthomosaics method by stitching images according to the direction of drone flight that has been determined in previous studies. There are 5 types of information classified as harvest time, middle leaf tea, ground tea, road, and old leaf tea. The success of image identification is indicated by the data in the error matrix by testing from 123 random points spread over the map with the success of detecting 113 random points and an average accuracy rate of 91.87%, of course, this value is very good and exceeds the specified threshold value of 75%. The error that occurs in using this method is that it cannot distinguish similar pixel colors, so the detection is wrong. In addition, the image stitching method using the orthomosaics method has succeeded in performing image stitching which can be applied well in classification using the CNN approach.

ACKNOWLEDGEMENT

We would like to thank the Directorate of Research and Community Service of IT Telkom Jakarta for funding the research.

REFERENCES

- [1] A. A. Wirabudi, R. Munadi, A. Rusdinar, D. Rohdiana, and D. H. Lee, "Design autonomous drone control for monitoring tea plantation using dynamic programming and kruskal algorithm," 2019 *IEEE International Conference on Signals and Systems (ICSigSys)*, 2019.
- [2] S. Yeşilmen and B. Tatar, "Efficiency of convolutional neural networks (CNN) based image classification for monitoring construction related activities: A case study on aggregate mining for concrete production," *Case Studies in Construction Materials*, 2022, doi: 10.1016/j.cscm.2022.e01372.
- [3] X. Zhao, L. Gao, Z. Chen, B. Zhang, and W. Liao, "CNN-based large scale landsat image classification," 2018. [Online]. Available: <http://www.geodata.cn>
- [4] N. Kerle, F. Nex, M. Gerke, D. Duarte, and A. Vetrivel, "UAV-based structural damage mapping: A review," *ISPRS international journal of geo-information* 9.1, 2019.
- [5] Q. Xu, J. Chen, L. Luo, W. Gong, and Y. Wang, "UAV image stitching based on mesh-guided deformation and ground constraint," *IEEE J Sel Top Appl Earth Obs Remote Sens*, vol. 14, pp. 4465–4475, 2021, doi: 10.1109/JSTARS.2021.3061505.
- [6] T. Hinzmann, J. L. Schönberger, M. Pollefeys, and R. Siegwart, "Mapping on the fly: Real-time 3D dense reconstruction, digital surface map and incremental orthomosaics generation for unmanned aerial vehicles," in *Field and Service Robotics. Springer Proceedings in Advanced Robotics*, Cham, 2018.
- [7] M. L. Cheng, M. Matsuoka, W. Liu, and F. Yamazaki, "Near-real-time gradually expanding 3D land surface reconstruction in disaster areas by sequential drone imagery," *Automation in Construction*, 2022, doi: 10.1016/j.autcon.2021.104105.
- [8] S. Tanathong, W. A. P. Smith, and S. Remde, "SurfaceView: Seamless and tile-Based orthomosaics using millions of street-level images from vehicle-mounted cameras," *IEEE Transactions on Intelligent Transportation Systems*, vol. 23, no. 4, pp. 3482–3497, 2022, doi: 10.1109/TITS.2020.3036928.
- [9] S. Minaee, Y. Boykov, F. Porikli, A. Plaza, and N. Kehtarnavaz, "Image segmentation using deep learning: A Survey," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 44, 2021, doi: 10.1109/tpami.2021.3059968.
- [10] X. Zhao, L. Gao, Z. Chen, B. Zhang, W. Liao, and X. Yang, "An Entropy and MRF model-based CNN for large-scale landsat image classification," *IEEE Geoscience and Remote Sensing Letters*, vol. 16, no. 7, pp. 1145–1149, 2019, doi: 10.1109/LGRS.2019.2890996.
- [11] Institute of Electrical and Electronics Engineers and IEEE Geoscience and Remote Sensing Society, 2019 IEEE International Geoscience & Remote Sensing Symposium: proceedings: July 28–August 2, 2019, Yokohama, Japan.
- [12] J. Liu, Y. Chen, and K. Liu, "Exploiting the ground-truth: An adversarial imitation based knowledge distillation approach for event detection," *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 33, no. 1, 2019.
- [13] Institute of Electrical and Electronics Engineers, 2017 *IEEE 2nd International Conference on Signal and Image Processing*, ICSIP: August 4-6, 2017, Singapore.
- [14] Z. Tang, M. Li, and X. Wang, "Mapping tea plantations from VHR images using OBIA and convolutional neural networks," *Remote Sensing*, vol. 12, no. 18, 2020.
- [15] G. Proskura, I. Vasilyeva, and V. Lukin, "Analysis of improvement of noisy multichannel image controlled pixel-by-pixel classification by post-classification processing," in *Proceedings-15th International Conference on Advanced Trends in Radioelectronics, Telecommunications and Computer Engineering*, TCSET 2020, Feb. 2020, pp. 525–530. doi: 10.1109/TCSET49122.2020.235488.
- [16] L. Zhang, Z. Gong, Q. Wang, D. Jin, and X. Wang, "Wetland mapping of Yellow River Delta wetlands based on multi-feature optimization of Sentinel-2 images," *J. Remote Sens*, vol. 23, no. 2, pp. 313–326, 2019.
- [17] D. J. Ford, "UAV imagery for tree species classification in Hawai'i: A Comparison of MLC, RF, and CNN supervised classification," Master Thesis, Diss. University of Hawai'i at Manoa, 2020.
- [18] M. A. Boateng, N. E. N. Sey, A. A. Amproche, and M. K. Domfeh, "Instance segmentation scheme for roofs in rural areas based on Mask R-CNN," *The Egyptian Journal of Remote Sensing and Space Science*, vol. 25, no. 2, pp. 569–577, 2022.
- [19] V. Kafedziski, S. Pecov, and D. Tanevski, "Detection and classification of land mines from ground penetrating radar data using faster R-CNN," in *2018 26th Telecommunications Forum (TELFOR)*, Belgrade, Serbia, 2018.
- [20] C. Zhang, L. Wang, S. Cheng, and Y. Li, "SwinSUNet: Pure transformer network for remote sensing image change detection," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 60, 2022, doi: 10.1109/tgrs.2022.3160007.
- [21] H. Jiang, M. Peng, Y. Zhong, H. Xie, Z. Hao, J. Lin, X. Ma, and X. Hu, "A Survey on deep learning-based change detection from high-resolution remote sensing images," *Remote Sensing*, vol. 14, no. 7, 2022, doi: 10.3390/rs14071552.
- [22] R. Liu, M. Kuffer, and C. Persello, "The temporal dynamics of slums employing a CNN-based change detection approach," *Remote Sensing*, vol. 11, no. 23, 2844, 2019.
- [23] J. Ou, X. Guo, M. Zhu, and W. Lou, "Autonomous quadrotor obstacle avoidance based on dueling double deep recurrent Q-learning with monocular vision," *Neurocomputing*, vol. 441, pp. 300–310, 2021, doi: 10.1016/j.neucom.2021.02.017.
- [24] X. Dai, Y. Mao, T. Huang, N. Qin, D. Huang, and Y. Li, "Automatic obstacle avoidance of quadrotor UAV via CNN-based learning," *Neurocomputing*, vol. 402, pp. 346–358, 2020.