Intelligent agricultural landscape identification system

Ching-Ju Chen
Department of Electrical Engineering, National Yunlin University of Science and Technology, Yunlin, Taiwan

Yu-Cheng Chen
Department of Electrical Engineering, National Yunlin University of Science and Technology, Yunlin, Taiwan

Jing-Yao Lin
Department of Electrical Engineering, National Yunlin University of Science and Technology, Yunlin, Taiwan

Rung-Tsung Chen
Taiwan Biodiversity Research Institute, Nantou, Taiwan

Candera Wijaya
Agricultural Engineering Research Center, Taoyuan, Taiwan

E-mail: cjchen@yuntech.edu.tw, M11212080@yuntech.edu.tw, M11212086@yuntech.edu.tw, tsung2358@gmail.com, rbhung@aerc.org.tw

Abstract

In recent years, Taiwan's economic development and land restructuring have decreased agricultural land area. The decrease in agricultural land area will affect the food supply for human beings and reduce the habitat of wild animals, destroying biodiversity. With the development of drone technology, the survey of farmland ecology no longer requires a large workforce to visit the site for inspection and analysis. This paper proposes a recognition system based on the Semantic segmentation method to classify agricultural landscapes, watersheds, and habitats automatically. We use two models for training: U-Net with the VGG16 model and U-Net with the Resnet50 model. The experiments in this paper show that the U-Net with VGG16 model and the U-Net with Resnet50 model applied to semantic segmentation of farmland landscape images have their classification categories. Still, some categories may be misclassified due to the similarity of the features, such as grassland, fallow field, and upland fields. This paper suggests that in the future, the number of data sets and the diversity of samples.

Keywords: Semantic Segmentation, Artificial Intelligence, Image Recognition, Biodiversity, Deep Learning

1. Introduction

Agricultural land serves not only as a source of human food but also as a habitat for wildlife and a provider of ecological services [1]. Reducing, overusing, or mismanaging agricultural land disrupts biodiversity. Agricultural biodiversity, a critical component of conservation, was integrated into the United Nations' Sustainable Development Goals (SDGs) in 2015.

This study proposes using UAVs to capture aerial images and establish an identification system for agricultural landscapes, watersheds, and habitats. This approach enables real-time biodiversity monitoring and supports adequate wildlife habitat protection and management, aligning with and advancing the SDGs for biodiversity conservation.

Recent advancements in artificial intelligence (AI) and deep learning have expanded the use of image recognition technologies. This study utilizes VGGNet, ResNet, and the U-Net semantic segmentation algorithm, as described below:

1. VGGNet
   The Visual Geometry Group [2] proposed in 2014 that VGGNet simplifies network design using 3x3 convolutional kernels and includes versions like VGG16 and VGG19, denoting the number of convolutional and fully connected layers. VGGNet demonstrated strong performance in the ImageNet competition.

2. ResNet
   Developed by Kaiming He et al. [3] in 2015, ResNet introduces Residual Blocks and Skip Connections to enable deeper networks while addressing model degradation. Popular versions include ResNet50 and ResNet101, indicating the number of layers.

3. Image Segmentation
   Image segmentation assigns pixels to specific categories, making it suitable for complex farmland landscapes like upland fields, freshwater fields, and grasslands. This study employs the U-Net model, introduced by Ronneberger et al. [4] in 2015, featuring

an encoder-decoder architecture with skip connections to retain spatial details. Initially designed for medical imaging, U-Net has been adapted for various tasks, including aerial image segmentation, with variations like Attention U-Net. [5]

2. Research methods

In this study, different artificial intelligence algorithms based on semantic segmentation are collected, among which the model of U-Net architecture, whose structure is suitable for dealing with complex scenes and high-resolution aerial images, is also applied to aerial photos. The relevant data of the study also proves the applicability of the U-Net architecture to aerial photos. Therefore, this study proposes to build a U-Net architecture-based drone aerial image recognition system for farmland scenes.

2.1. *Input image*

In this study, we used 10cm/pixel high-resolution aerial photos provided by the Agricultural Engineering Research Center (AERC), and selected samples suitable for the identification of farmland landscapes from the provided images, as shown in Fig.1 and Fig.2 below, which should cover diverse farmland landscapes such as farmland and water.

Fig.1. Map Number 94202015_20200605

Fig.2. Map Number 94202004_20200712

2.2. *Classification category definition*

Because of the complexity and diversity of Taiwan's agricultural landscapes, a total of 21 categories were classified based on the actual context of the agrarian landscapes in the images, which were Freshwater field, Upland Field, Fallow fields, Land grading, Mixed forest, Grasslands, Bare land, Building, Orchard, (Ecological pond), Water pool, Fishery farm, Road, Farm road, Solar panel, Duck farm, Irrigation and drainage system, Unused land, Agricultural production facilities, No data.

2.3. *Image Labeling*

In this study, we use the open-source labeling software Labelme to mark 80 images with the size of 2048x2048, and the file (*.json) generated after labeling is processed to create the corresponding mask.

2.4. *Model training*

This study employs two backbone networks, U-Net+VGG16 and U-Net+ResNet50, using 800 images (560 for training, 120 for validation, and 120 for testing). The models were trained with a batch size of 1, 200 iterations, a learning rate of 0.00005, and Cosine Annealing for learning rate decay. Mixed precision training was applied due to GPU memory limitations, with Adam optimizer selected for better performance over SGD. The focal loss function was used to address class imbalance.

The training was conducted on a Z790 Pro CPU, a Gigabyte RTX4080 16GB GPU, and 128GB of RAM, using PyCharm (version 2023.1) and Python (version 3.11).

2.5. *Model Evaluation Metrics*

In this study, we use standard evaluation metrics of semantic segmentation, including mIoU, mPA, mRecall, and mPrecision, as an essential reference to measure the performance of the model, which are introduced as follows

(1) mIoU(mean Intersection over Union)

Calculate the intersection and concurrency ratio between the predicted and real regions and then average the IoU for all categories. The calculation formula is Eq.(1).

$$mIoU = \frac{1}{N}\sum_{i=1}^{N}\frac{TPi}{TPi + FPi + FNi} \quad (1)$$

(2) mPA(mean Pixel Accuracy)

Calculate the ratio of the number of pixels correctly predicted by the model to the total number of pixels, and then take the average value of Pixel Accuracy for all categories. The calculation formula is Eq.(2).

$$mPA = \frac{1}{N}\sum_{i=1}^{N}\frac{TPi}{TPi+FNi} \quad (2)$$

(3) mRecall

Calculate the proportion of True Positive pixels correctly recognized by the model, averaged over all classes of Recall, the calculation formula is as Eq.(3).

$$mRecall = \frac{1}{N}\sum_{i=1}^{N}\frac{TPi}{TPi + FNi} \quad (3)$$

(4) mPrecision

Calculate the proportion of pixels predicted by the model to be Positive that is Positive, then average the Precision values across all categories; the calculation formula is as Eq.(4).

$$mPrecision = \frac{1}{N}\sum_{i=1}^{N}\frac{TPi}{TPi+FPi} \quad (4)$$

2.6. *Model Prediction*

When the model is trained, the weights will be accessed and a new image will be imported. In the prediction, the image size needs to be the same as the training image size, which needs to be cut into several 2048x2048 images.

Then the prediction will be done for each 2048x2048 image, and the final image will be synthesized with the same size as the original imported image.

3. Experimental results

3.1. *Training results*

We calculate various evaluation metrics based on the test set's and training set's prediction results to assess the model's performance, serving as an important basis for subsequent parameter adjustments. As shown in Table. 1, U-Net+VGG16 achieves a mIoU of 70.94, mPA of 79.69%, mRecall of 79.69%, and mPrecision of 81.09%. In comparison, U-Net+ResNet50 achieves an mIoU of 57.71, mPA of 68.21%, mRecall of 68.21%, and mPrecision of 73.29%. This indicates that U-Net+VGG16 outperforms U-Net+ResNet50 in all evaluation metrics. However, the mPA, mRecall, and mPrecision metrics of U-Net+ResNet50 demonstrate its potential in agricultural land scene recognition.

Table. 1. Training results

	VGG16	Resnet50
mIoU	70.94	57.71
mPA	79.69	68.21
mRecall	79.69	68.21
mPrecision	81.09	73.29

3.2. *Prediction results*

In terms of prediction, we would like to compare the prediction results of U-Net+VGG16 and U-Net+Resnet50 to further verify the performance of the two models in the same scenario and to visualize the prediction effect of the two models as well as their ability to handle boundaries, etc. We would like to understand further the advantages and disadvantages of the two models and their applicability. To further understand the strengths and weaknesses of the two models and their applicability, the following is an example of a localized map of the predictions in image number 94202006c.

In Fig. 3, both the U-Net+VGG16 and U-Net+ResNet50 methods demonstrate strong performance in accurately identifying solar panels, indicating the robustness of these models in detecting this specific category. However, both methods need help classifying unused land, often mislabeling it as bare land, farm roads, fishery farms, or other similar categories.

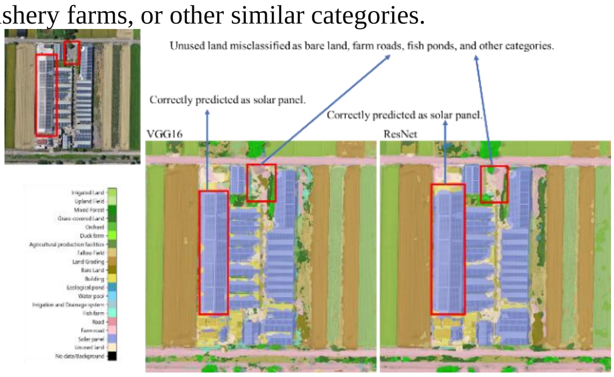


Fig. 3. Comparison of the partial map with the partial forecast map for map number 94202006c

3.3. *Calculate area*

In semantic segmentation, if the image resolution is known, then the number of pixels in each category can be used to calculate the size of the area covered by each category with the following formula: Eq. (5).

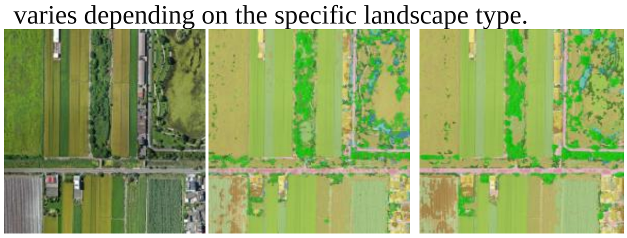
$$Area = \frac{Pixel\ count \times Area\ per\ pixel}{100{,}000{,}000} \quad (5)$$

Model prediction accuracy directly impacts area calculations, with incorrect predictions leading to inaccuracies. To evaluate this, we calculate the error between the predicted areas of U-Net+VGG16 and U-Net+ResNet50 and the actual area using Eq. (6), measuring the accuracy of both models.

$$\text{Error}(\%) = \frac{\text{Predicted area} - \text{Actual area}}{\text{Actual area}} \times 100\% \quad (6)$$

Fig. 4 illustrates the ground image and the prediction results of two models: U-Net+VGG16 (Fig. 4 a) and U-Net+ResNet50 (Fig. 4 b).

Table.2 compares the error values between the actual and predicted areas for various land types. The results show that U-Net+VGG16 provides more accurate predictions for freshwater fields, upland fields, land grading, mixed forests, bare lands, ecological ponds, roads, farm roads, irrigation systems, and unused lands. In contrast, U-Net+ResNet50 achieves greater accuracy for fallow fields, grasslands, buildings, orchards, fish farms, and agricultural production facilities. This indicates that the prediction accuracy of each model varies depending on the specific landscape type.



Actual image (a)U-Net+VGG16 (b)U-Net+Resnet50

Fig. 4. Actual image and Prediction result

Table.2.Comparison of actual area and predicted area

	Actual Area	U-Net +VGG16	Error	U-Net +Resnet50	Error
Freshwater field	3.58	3.68	2.79%	4.17	16.48%
Upland Field	2.58	2.53	1.94%	2.3	10.85%
Fallow field	3.65	5.17	41.64%	5.06	38.63%
Land grading	0	0.49		0.68	
Mixed forest	1.08	1.18	9.26%	1.47	36.11%
Grasslands	1.27	0.6	52.76%	0.64	49.61%
Bare land	0.04	0.03	25%	0.01	75%
Building	0.77	0.51	33.77%	0.57	25.97%
Orchard	0.06	0.76	1166.67%	0.4	566.67%
Ecological pond	1.70	0.25	85.29%	0.13	92.35%
Water pool	0	0	0%	0	0%
Fishery farm	0	0.21		0.12	
Road	0.49	0.13	73.47%	0.08	83.67%
Farm road	0.20	0.69	245%	0.73	265%
Solar panel	0	0	0%	0	0%

Duck farm	0	0	0%	0	0%
Irrigation and drainage system	0.66	0.13	80.3%	0.07	89.39%
Unused land	0.58	0.29	50%	0.02	96.55%
No data	0	0	0%	0	0%
Agricultural production facilities	0	0.01		0	0%

4. Conclusion

This study utilized U-Net+VGG16 and U-Net+ResNet50 models to develop an AI system for farmland landscape recognition. Five hundred sixty images were used for training, 120 for validation, and 120 for testing. U-Net+VGG16 achieved better performance with mIoU of 70.94%, mPA of 79.69%, mRecall of 79.69%, and mPrecision of 81.09%, compared to U-Net+ResNet50, which had mIoU of 57.71%, mPA of 68.21%, mRecall of 68.21%, and mPrecision of 73.29%.

In predictions, U-Net+VGG16 performed better in recognizing freshwater fields, upland fields, land grading, mixed forests, bare lands, ponds, roads, irrigation systems, and unused lands. Meanwhile, U-Net+ResNet50 identified fallow fields, grasslands, buildings, orchards, and agricultural facilities. Both models have strengths suitable for different segmentation tasks, but misjudgments due to similar landscape features remain challenging. Improvements in combining categories and increasing the diversity of image samples could improve the overall recognition accuracy of agricultural landscape segmentation.

Acknowledgment

This research was supported by the National Science and Technology Council, Taiwan (NSTC 113-2635-M-224-001-). Special thanks to the Agricultural Engineering Research Center (AERC) for providing aerial images and the Taiwan Biodiversity Research Institute for their expertise and resources, which greatly enriched the study.

References

1. Power Alison G., "Ecosystem services and agriculture: tradeoffs and synergies Phil." Trans. R. Soc. B3652959–2971,2010, doi: http://doi.org/10.1098/rstb.2010.0143
2. Karen Simonyan, Andrew Zisserman, "Very Deep Convolutional Networks for Large-Scale Image Recognition," 2014.,arXiv:1409.1556.
3. Kaiming He, Xiangyu Zhang, Shaoqing Ren, Jian Sun, "Deep Residual Learning for Image Recognition," arXiv:1512.03385,2015.
4. Olaf Ronneberger, Philipp Fischer, Thomas Brox, "U-Net: Convolutional Networks for Biomedical Image Segmentation," 2015, arXiv:1505.04597.
5. N. Gonthina, S. Adunuri, R. Mateti, S. S. Allampalli and L. V. Narasimha Prasad, "Accurate Semantic Segmentation of Aerial Imagery Using Attention Res U-Net Architecture," 2024 International Conference on Emerging Smart Computing and Informatics (ESCI), Pune, India, 2024, pp. 1-5, doi: 10.1109/ESCI59607.2024.10497287.

Authors Introduction

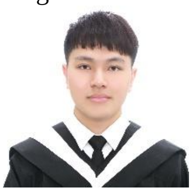
Ching-Ju Chen

She received a Ph.D. in engineering science from National Cheng Kung University, Taiwan, in 2011. She is an Assistant Professor at the Department of Electrical Engineering, National Yunlin University, Taiwan. Her expertise includes digital image processing, AI, AIoT, UAV system integration, GIS, GPS, and RS.

Yu-Cheng Chen

He is pursuing a master's in Electrical Engineering at Yunlin National University of Science and Technology, Yunlin, Taiwan. His research interests include deep learning applications in aerial imaging and image recognition.

Jing-Yao Lin



He is pursuing a master's in Electrical Engineering at Yunlin National University of Science and Technology, Yunlin, Taiwan. His research interests include deep learning applications, GIS, and image processing.

Rung-Tsung Chen

He received a Ph.D. in fisheries science from National Taiwan University, Taiwan, in 2009. He is an Assistant Research Fellow with the Wild Animals Division at Taiwan Biodiversity Research Institute. He majored in the studies of organic agriculture's aquatic ecology and biodiversity.

Candera Wijaya

He earned a master's in civil engineering from National Central University. He focused on spatial information at the Agricultural Engineering Research Center, assisting the Ministry of Agriculture in promoting the use of drones in agriculture, forestry, fisheries, and animal husbandry.