ISSN 2616-5775 Vol. 7, Issue 11: 59-66, DOI: 10.25236/AJCIS.2024.071108

Semantic Segmentation Method for Sugarcane Planting Fields

Dawei Zhang[1,a], Zhiguang Zeng[2,b*]

[1]School of Electronic Information Engineering, Beihai Vocational College, Xizang Street, Beihai, China
[2]Academic Affairs Office, Beihai University of Art And Design, 1 Xinshiji Avenue, Yinhai District, Beihai, Guangxi, China
[a] zhangdawei@bhzyxy.edu.cn, [b]zbrlwl@163.com
**Corresponding author*

Abstract: *Sugar industry is the core industry of economic crops in Guangxi. At present, the actual sugar production in Guangxi accounts for more than 60% of the national total, and it is the backbone of the national sugar industry. At present, in order to support the healthy and stable development of Guangxi's sugar industry, the state has made financial subsidies for the promotion of sugarcane varieties in the whole region. Farmers can apply for national subsidies based on the actual plot contours measured by longitude and latitude. In this paper, a semantic segmentation-based sugar planting plot recognition method is proposed to realize the automatic recognition and accurate classification of sugar planting plots by combining remote sensing images and deep learning technology. Experimental results show that the proposed method has high accuracy and robustness in the recognition of sugar planting plots.*

Keywords: *Semantic Segmentation,Sugarcane Planting Fields*

1. Introduction

In the context of global agricultural development, sugar, as one of the important cash crops, has a wide range of planting and application value in many countries and regions. Traditional methods of measuring sugar plots usually require a lot of manual labor and time, are inefficient, and have problems such as inaccurate measurement and repeated measurement. With the advancement of technology, remote sensing data has become an important means of identifying and monitoring sugar plots. By analyzing the feature information in remote sensing images, sugar plots can be quickly and accurately identified, monitored and evaluated, so as to help relevant departments quickly understand the number and distribution of sugar plots, which is of great significance for sugar production management, resource optimization and environmental protection. Semantic segmentation is a computer vision task that aims to classify objects in visual images and help computers understand visual information. The segmentation map is essentially a reconstruction of the original image, in which each pixel has been color-coded by semantic class to create a segmentation mask. With the improvement of computing power and the success of deep learning algorithms, machines are performing better and better in identifying objects. Feature extraction algorithms can automatically extract feature information related to target objects from remote sensing images to support the identification and monitoring of sugar-growing plots. It can achieve accurate identification, health monitoring and optimized management of sugar-growing plots, and provide a scientific decision-making basis for sugar production.

2. Related Works

At present, there have been many research works on crop identification and monitoring based on remote sensing data at home and abroad. Some scholars have used multispectral remote sensing data and machine learning algorithms to identify and classify sugar plantations. However, due to the growth characteristics of sugarcane fields and the complexity of soil conditions, there are problems such as low recognition accuracy and high missed detection rate. In order to improve the accuracy of remote sensing image semantic segmentation, Jiang Wenwen [1] et al. proposed an improved U-Net multi-scale feature fusion remote sensing image semantic segmentation network, which effectively improved the

ISSN 2616-5775 Vol. 7, Issue 11: 59-66, DOI: 10.25236/AJCIS.2024.071108

semantic segmentation accuracy of remote sensing images. Li Linjuan [2] et al. took Taiyuan City, Shanxi Province as the research area, made a high-resolution remote sensing land cover dataset of Taiyuan City, and proposed a semantic segmentation model guided by cross-layer detail perception and group attention for high-resolution remote sensing image analysis, solving the problem of poor segmentation effect on multi-scale objects in complex backgrounds and discontinuous boundaries of segmented areas. Huan Hai [3] et al. proposed a global guided multi-feature fusion network (GGMNet) for extracting roads in remote sensing images to solve the problem of low recognition efficiency caused by high inter-class similarity between buildings and roads in remote sensing images and the presence of shadows and occlusions. Li Xuqing [4] proposed a CAHRNet (change attention high-resolution Net) semantic segmentation model to solve the problems of complex objects interfering with winter wheat recognition, resulting in low recognition accuracy and blurred boundary segmentation.

3. Manuscript Preparation

3.1. Experimental environment

This experiment used an NVIDIA RTX3080 GPU with 10GB video memory. The server was equipped with an Intel Core i7 9750 processor with 64GB RAM, a 2TB SSD storage device, and was configured with relevant machine learning libraries. The details are shown in Table 1.

Table 1: Experimental environment

Processor	Intel Core i7 9750
GPU	RTX 3080Ti
RAM	64G
OS	Ubuntu 22.04 LTS
Language	Python
Vision library	MMEngine, MMSegmentation
IDE	Pycharm

3.2. MMEngine

MMEngine is a general training framework for deep learning model training, which aims to improve the development efficiency and maintainability of machine learning projects. It is designed for model training and evaluation in computer vision tasks, but also has the flexibility to be applicable to other tasks. MMEngine provides a modular design, a high-level interface that simplifies the training process, and support for distributed training. It provides flexible training and inference support for deep learning models based on PyTorch.

3.3. Dataset

This study uses a high-resolution remote sensing image dataset publicly available from an institution. This dataset is divided into two parts: the converted remote sensing images (bitmaps) of the sugar planting plots in Guangxi and the annotated semantic segmentation maps.

Deep learning models usually require input image data in standard bitmap formats (such as PNG, JPEG, BMP, etc.) so that they can be processed and loaded using existing image processing libraries. The original format of remote sensing images may be a specialized format for a specific field (such as BSQ, BIL, BIP, and HDF). Although these formats can save more metadata or accuracy information, they may not be directly readable by deep learning frameworks. The dataset used in this experiment has preprocessed the remote sensing images (as shown in Figure 1). In addition, the dataset also provides annotation information corresponding to the remote sensing images (as shown in Figure 2). By adding labels to the data, the model can recognize and learn patterns and features in the data, which is conducive to training and segmentation.

ISSN 2616-5775 Vol. 7, Issue 11: 59-66, DOI: 10.25236/AJCIS.2024.071108

Figure 1: Remote sensing image

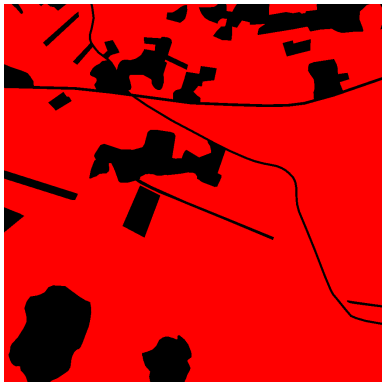

Figure 2:Semantic segmentation map

3.4. Feature extraction

Feature extraction in deep learning automatically extracts features from raw data through a multi-layer neural network structure, gradually extracting from low-level features to high-level features, so that the model can capture key patterns and structures in the data. Compared with traditional manual feature extraction, feature extraction in deep learning is more flexible, automated, and can handle complex high-dimensional data. This powerful feature extraction capability is a key factor in deep learning's outstanding performance in various tasks, as shown in Figure 3.

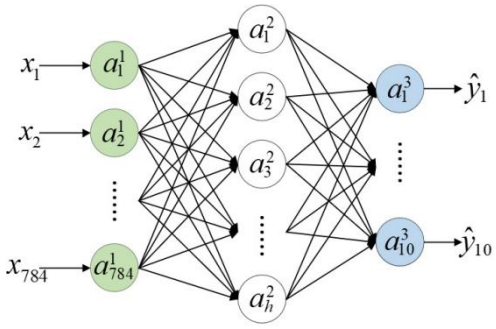

Figure 3: Feature extraction of deep learning network

3.5. Model building

3.5.1. Configure file

MMEngine organizes models, datasets, training hyperparameters, and other content through configuration files. A semantic segmentation configuration file mainly includes model configuration, data configuration, training configuration, and log configuration. The model configuration defines the model architecture used for semantic segmentation, such as using segmentation models such as U-Net, DeepLabV3, and SegFormer; the data configuration is used to configure the dataset path and preprocessing methods for training, validation, and testing, such as normalization and data

ISSN 2616-5775 Vol. 7, Issue 11: 59-66, DOI: 10.25236/AJCIS.2024.071108

enhancement; the training configuration is used to set hyperparameters such as learning rate, optimizer, loss function, and number of training rounds; the log configuration defines the model training log, checkpoint saving method, etc.

3.5.2. Data preprocessing

The remote sensing images provided in this experiment are 4k resolution bitmaps, which need to be cropped before loading. In this experiment, they are divided into 1080P images. Considering that the original data set has only 196 images (784 after cropping), the data set is relatively small, so a variety of data enhancement strategies are used, including random cropping, rotation, flipping and other methods, to improve the generalization ability and robustness of the model.

3.5.3. Model training

MMEngine training uses the Runner mechanism to uniformly manage the entire training process. When training is started through the train.py script, the system automatically calls each component (model, data, optimizer, etc.) according to the configuration file. During the training process, the training progress can be viewed through the log, including indicators such as loss and accuracy. The model weights will also be saved at the specified checkpoint. The principle of the MMEngine executor is shown in Figure 4.

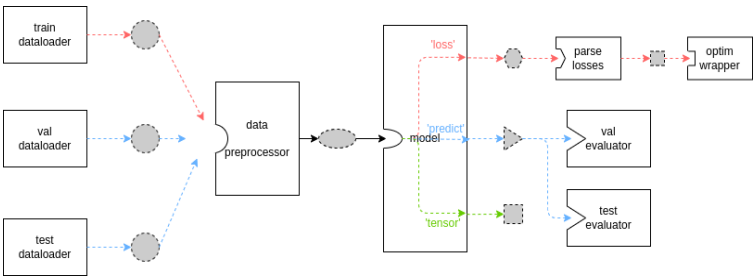

Figure 4:basic dataflow

3.5.4. Model Evaluation

After training is completed, you can evaluate the model on the validation set to view the semantic segmentation performance of the model. During the evaluation, load the optimal checkpoint for inference and generate segmentation results. Common evaluation indicators of semantic segmentation models include intersection over union (IoU), accuracy, precision, recall, F-score, Dice coefficient, etc. In image segmentation tasks, these evaluation functions have different functions.

IOU is an indicator that measures the degree of overlap between the predicted result and the true label. It is particularly commonly used in object detection and image segmentation tasks. It calculates the ratio of the intersection and union of the predicted area and the actual area. The value of IOU is between 0 and 1, 1 means that the prediction is completely correct, and 0 means there is no overlap. It is generally believed that the higher the IOU value, the closer the segmentation result of the model is to the actual situation. The principle is shown in Figure 4. mIOU is the result of averaging the IOU of multiple categories and is often used in multi-category image segmentation tasks. It can more comprehensively reflect the segmentation effect of the model, especially in multi-category image segmentation tasks, and is a more comprehensive indicator, as shown in Figure 5.

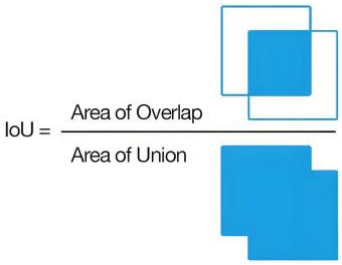

Figure 5: IoU

Confusion Matrix,as shown in Table 2, is a tool commonly used in machine learning, data mining and statistical analysis to evaluate the performance of classification models. It shows the matching between the model prediction results and the true labels in an intuitive way, and is particularly suitable for binary and multi-classification tasks. Through the confusion matrix, you can clearly see the

ISSN 2616-5775 Vol. 7, Issue 11: 59-66, DOI: 10.25236/AJCIS.2024.071108

classification between the model prediction samples and the actual samples. TP (true positive), TN (true negative), FP (false positive) and FN (false negative) are usually used to represent different classification results. Their meanings are as follows:

- TP (True Positive): The number of samples correctly predicted as positive by the model;
- TN (True Negative): The number of samples correctly predicted as negative by the model;
- FP (False Positive): The number of samples incorrectly predicted as positive by the model;
- FN (False Negative): The number of samples incorrectly predicted as negative by the model.

Through the confusion matrix, a variety of commonly used classification model evaluation indicators can be derived, such as Accuracy, Precision, Recall, etc.

Table 2: Confusion Matrix

	True Positive	True Negative
Predict Positive	True Positive, TP	False Positive, FP
Predict Negative	False Negative, FN	True Negative, TN

Accuracy refers to the proportion of samples that are correctly predicted by the classifier to all samples. The accuracy can be expressed as shown in Equ 1.

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \tag{1}$$

Precision refers to the proportion of samples predicted by the model to be positive that are actually positive (for FP), as shown in Equ 2.

$$\text{Precision} = \frac{TP}{TP + FP} \tag{2}$$

Recall, also known as sensitivity, indicates the proportion of all actual positive samples that are correctly identified as positive (for FN), as shown in Equ 3.

$$\text{Re}\,call = \frac{TP}{TP + FN} \tag{3}$$

The F score is the harmonic mean of precision and recall, which is used to comprehensively measure model performance. Its expression is shown in Equ 4.

$$\text{F} - Score = 2 \times \frac{\text{Pr}\,ecision \times \text{Re}\,call}{\text{Pr}\,ecision + \text{Re}\,call} \tag{4}$$

Dice coefficient, also known as Sørensen-Dice coefficient, is a statistical tool used to measure the similarity between two sample sets. It is often used in binary classification tasks or image segmentation tasks. It evaluates similarity by calculating the degree of overlap between two sets. The value range is 0 to 1. The larger the value, the more similar the two sets are. The calculation formula of Dice coefficient is shown in Equ 5.

$$Dice = \frac{2 \times |A \cap B|}{|A| + |B|} \tag{5}$$

4. Experimental analysis

4.1. Small target sensitivity problem

Whether from the perspective of absolute scale or relative scale, small targets have the problem of low resolution compared to large and medium-sized targets. Low-resolution small targets have little visual information, making it difficult to extract discriminative features, and are easily disturbed by environmental factors, which makes it difficult for the detection model to accurately locate and identify small targets. Although the resolution of remote sensing satellite images is relatively large, they cannot accurately display small mountain roads and other green plants at the pixel level, especially when

ISSN 2616-5775 Vol. 7, Issue 11: 59-66, DOI: 10.25236/AJCIS.2024.071108

mountain paths are above the classification boundary, which is the key point that causes the model to lose points when performing binary classification. Some scholars have proposed adding a separate branch to CNN (convolutional neural network) to learn edge information, but its training cost is relatively high.

4.2. Sparse targets and overlapping mixed problems

In the task of semantic segmentation, dealing with sparse targets (i.e. targets that account for a small proportion of the image and are unevenly distributed) is a difficult point. Sparse targets are usually easily ignored or misclassified by the model because of their small number of pixels in the image. Since sparse targets account for a small proportion, the model is more likely to be biased towards categories with a large proportion during training, resulting in imbalance problems in the training process. As shown in Figure 6 and Figure 7, when performing semantic segmentation, some similar green plants (number) are inaccurately classified due to the imbalance of the data set. The current method to solve this problem is mainly to increase the loss weight for the sparse target category so that the model pays more attention to these targets during optimization. Commonly used weighting methods include weighted cross entropy and focal loss. This study uses the former.

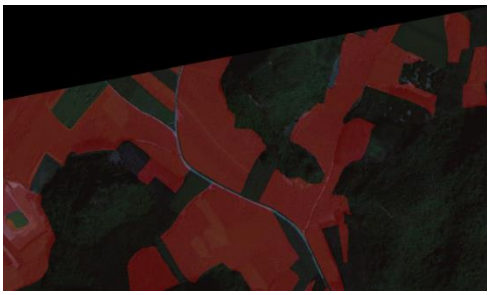

Figure 6:Verification Image

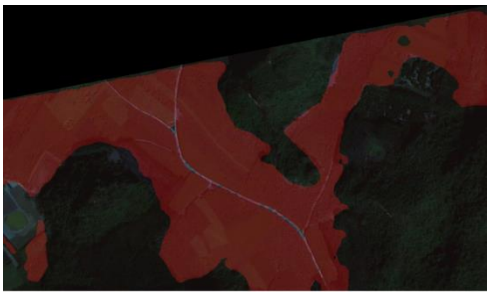

Figure 7:Predict Image

4.3. Performance indicators

mIoU is an important indicator for measuring semantic segmentation. It is the average value of the intersection over union (IoU) of all categories. It can well measure the segmentation performance of the model in each category and reflect the overall performance through the average value. The mIoU shown in Figure 8 is 0.74, which means that the entire model has good performance in the classification of sugarcane fields and other crops.

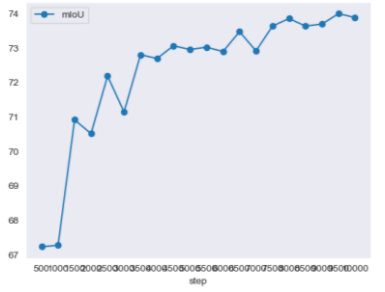

Figure 8: mIoU

ISSN 2616-5775 Vol. 7, Issue 11: 59-66, DOI: 10.25236/AJCIS.2024.071108

After the model converges, aACC and mACC are used to evaluate the overall classification accuracy and classification performance difference of the model. As shown in Figure 9 and Figure 10, the overall classification accuracy of the model is 0.86, and the overall performance of the model in pixel classification is 0.85, indicating that the model performs well in terms of overall classification accuracy.

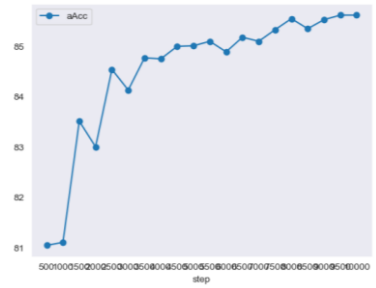

Figure 9: aACC

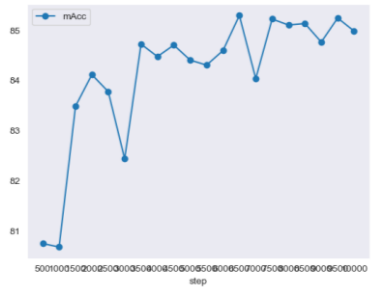

Figure 10: mACC

The average precision, average F-score and average of the entire model are measured by the three indicators of mPrecision, mF-score and mRecall to evaluate the performance of the model under the above three indicators (as shown in Figure 11,Figure 12 and Figure 13), which are 0.85, 0.85 and 0.85 respectively. This shows that the semantic segmentation model of this study performs balanced under the above three indicators and has good accuracy and balance.

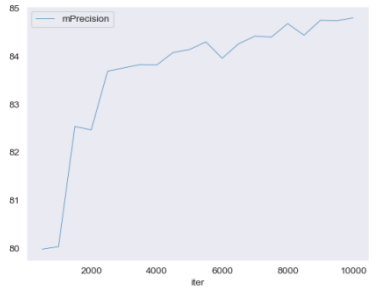

Figure 11:mPrecission

Figure 12:mF-score

ISSN 2616-5775 Vol. 7, Issue 11: 59-66, DOI: 10.25236/AJCIS.2024.071108

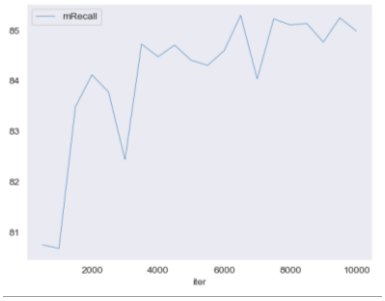

Figure 13:mRecall

The Dice coefficient is introduced to evaluate the prediction ability of the model for small and sparse targets. The experimental results show that the Dice coefficient is 0.85, as shown in Figure 14, which indicates that the model has good generalization ability for small and sparse targets.

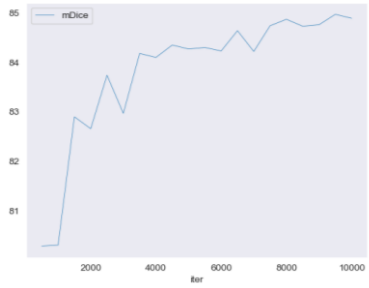

Figure 14:mDice

5. Conclusions

This study successfully developed an efficient algorithm that combines machine learning methods to semantically segment sugar plantations in remote sensing images, effectively achieving accurate positioning and boundary extraction of sugarcane fields, and providing data support for relevant departments to issue agricultural subsidies. Although this study has achieved certain results, there are still certain limitations in dealing with target overlap and crop sparseness in complex scenes. To improve the classification ability of this model for sparse targets, it is necessary to optimize it from multiple levels such as data, model structure, loss function, contextual information, and post-processing. In certain cases, the use of weighted loss functions, multi-scale feature extraction, attention mechanisms, and contextual information aggregation can also be combined with strategies such as partition training, local refinement, and data enhancement to further improve the accuracy of the model.

Acknowledgements

2023 Guangxi University Young and Middle-aged Teachers' Basic Scientific Research Ability Improvement Project "Research on Semantic Segmentation of Sugar Planting Plots Based on Remote Sensing Images" (2023KY1493)

References

[1] Jiang W, Xia Y.Improved U-Net multi-scale feature fusion remote sensing image semantic segmentation network[J]. Computer Science, 2024: 1-10.
[2] Li L, He Y, Xie G.Remote sensing image semantic segmentation guided by cross-layer detail perception and group attention[J]. Journal of Image and Graphics, 2024, 29(05): 1277-1290.
[3] Huan H, Sheng Y, Gu C. Global guidance multi-feature fusion network based on remote sensing image road extraction[J]. Journal of Zhejiang University (Engineering Science), 2024, 58(04): 696-707.
[4] Li X, Wu D, Wang Y. Semantic segmentation method of winter wheat in remote sensing images based on improved HRNet[J]. Transactions of the Chinese Society of Agricultural Engineering, 2024, 40(03): 193-200.