Postprint: Research on Remote Sensing Extraction of Terraced Fields Using an Improved DeepLab v3+ Model

Authors: Zhang Jun, Yuyan Chen, Qin Zhenyu, Zhang Mengyao, Zhang Jun

Date: 2024-08-30T00:00:00+00:00

Abstract

[目的与意义] As a key element of agricultural production, terraced field area estimation is crucial for agricultural policy formulation, land planning, and resource management. To address the challenge that complex terrain conditions and planting environments hinder automated terraced field extraction using traditional remote sensing data and monitoring methods, this study explores a method for accurately extracting terraced field area from high-resolution remote sensing imagery using deep learning technology.

[方法] A semantic segmentation dataset was constructed using Gaofen-6 images of terraced fields during the fallow period, and an improved DeepLab v3+ model was proposed. This model employs the lightweight MobileNet v2 network as the backbone. To simultaneously capture local details and global context, the Atrous Spatial Pyramid Pooling (ASPP) module is replaced with a Multi-scale Feature Fusion (MSFF) module, which utilizes a cascaded pattern of atrous convolutions with progressively increasing dilation rates to mitigate information loss. Additionally, coordinate attention mechanisms are applied to both shallow and deep features to enhance the network's target learning capability.

[结果与讨论] The combination of red, green, and near-infrared bands achieved optimal performance in terraced field extraction accuracy and effectiveness. Compared with the original DeepLab v3+ network, the precision, recall, F1-score, and Intersection over Union (IoU) metrics improved by 4.62%, 2.61%, 3.81%, and 2.81%, respectively. Furthermore, compared with UNet and the original DeepLab v3+, the improved DeepLab v3+ demonstrates superior performance in terms of parameter count and floating-point operations. Its parameter count is only 28.6% of UNet and 19.5% of the original DeepLab v3+, while its floating-point operations are only one-fifth of those of UNet and DeepLab v3+. This not only enhances computational efficiency but also makes the improved model more suitable for resource-constrained environments or those with lower computational capabilities.

[结论] Deep learning exhibits high accuracy in terraced field recognition from high-resolution remote sensing imagery, providing a valuable reference for refined monitoring and management of terraced fields.

Full Text

Remote Sensing Extraction Method of Terraced Fields Based on Improved DeepLab v3+

ZHANG Jun[1], CHEN Yuyan[1], QIN Zhenyu[2], ZHANG Mengyao[1], ZHANG Jun[1*]

[1]School of Earth Sciences, Yunnan University, Kunming, China
[2]Institute of International Rivers and Eco-security, Yunnan University, Kunming, China

Abstract

[**Objective**] Terraced fields are critical components of agricultural production, and accurate area estimation is essential for agricultural policy formulation, land planning, and resource management. Traditional remote sensing data and monitoring methods struggle with automated terraced field extraction due to complex terrain conditions and planting environments. This study explores a method using deep learning technology to accurately extract terraced field areas from high-resolution remote sensing imagery. [**Methods**] A semantic segmentation dataset was constructed using Gaofen-6 imagery of terraced fields during the fallow period. An improved DeepLab v3+ model was proposed that employs the lightweight MobileNet v2 as the backbone network. To simultaneously preserve local details and global context, a Multi-scale Feature Fusion (MSFF) module replaced the Atrous Spatial Pyramid Pooling (ASPP) module, utilizing a cascaded pattern of atrous convolutions with progressively increasing dilation rates to mitigate information loss. Additionally, coordinate attention mechanisms were applied to both shallow and deep features to enhance target learning. [**Results and Discussion**] The combination of red, green, and near-infrared bands yielded the best accuracy and performance for terraced field extraction. Compared to the original DeepLab v3+ network, precision, recall, F-score, and Intersection over Union (IoU) improved by 2.61%, 4.62%, 3.81%, and 2.81%, respectively. Furthermore, the improved model demonstrated superior performance in terms of parameter count and floating-point operations, with only 19.5% of the parameters and 28.6% of the floating-point operations of the original DeepLab v3+. This not only improves computational efficiency but also makes the enhanced model more suitable for resource-constrained or low-computing-power environments. [**Conclusion**] Deep learning achieves high accuracy in terraced field recognition from high-resolution remote sensing im-

agery, providing valuable reference for refined terraced field monitoring and management.

Keywords: terraced field extraction; remote sensing; convolutional neural network; Gaofen-6 satellite; DeepLab v3+

Introduction

Terraced fields are critical components of agricultural production, and accurate area estimation is essential for agricultural policy formulation, land planning, and resource management. Terraced field monitoring constitutes a vital component of dynamic soil and water conservation monitoring and evaluation, as determining the scale and spatial distribution of terraced fields provides a foundation for their management and maintenance [1]. Satellite remote sensing technology enables comprehensive, all-weather monitoring of surface features and geographical phenomena [2] and has been widely applied in urban planning, target recognition, and land cover mapping [3-5], providing reliable technical support for terraced field extraction.

Traditional terraced field extraction methods primarily utilize unique texture, spectral, and geometric features for classification, with main techniques including texture spectrum analysis [6], object-oriented approaches [7,8], and shallow machine learning [9,10]. However, as image resolution continues to improve, semantic information such as boundaries and spatial layouts of different ground objects becomes increasingly rich, which adds complexity to image information [11]. In high-resolution imagery, complex terrain environments and the "same spectrum, different objects" phenomenon challenge the discriminative capability of traditional methods among terraced fields, cultivated land, and bare land, leading to interpretation issues such as mixed land features [12] and salt-and-pepper effects [13]. Moreover, traditional methods focusing only on shallow features struggle to effectively utilize detailed characteristics in high-resolution remote sensing imagery, resulting in fragmented and dispersed extracted terraced fields that fail to meet the precision and automation requirements of smart agriculture [14,15].

In recent years, deep learning has demonstrated outstanding performance in semantic segmentation [16], with its automatic deep feature learning approach providing a viable solution for accurate extraction of high-resolution terraced field imagery. Convolutional Neural Network (CNN)-based semantic segmentation models have become the preferred method for terraced field extraction due to their exceptional image analysis capabilities [17]. For example, Wang et al. [18] achieved pixel-level intelligent extraction of terraced fields through an improved UNet deep learning model. Yu et al. [19] utilized deep transfer learning strategies to improve extraction accuracy with small sample datasets. Liu Dongjie [20] enhanced the robustness of deep learning models for terraced field recognition by combining spectral and topographic features. Zhao et al. [21] employed the EfficientNet v2 backbone network for feature extraction in terraced field extraction tasks and introduced the Convolutional Block Attention Module (CBAM) to improve the DeepLab v3+ network, successfully balancing accuracy and speed for ultra-high-resolution UAV image-based terraced field extraction. The classic semantic segmentation model DeepLab v3+ improves segmentation performance through encoder-decoder structures and depthwise separable convolutions that fully consider both shallow and deep semantic information. However, the original DeepLab v3+ model suffers from complex structure and large parameter count, requiring substantial computational resources during training and inference.

This study proposes a lightweight remote sensing image semantic segmentation method through improvements to the DeepLab v3+ model. The method adopts the lightweight MobileNet v2 as the backbone network to reduce model parameters. To enhance multi-scale feature extraction and avoid information loss, the original atrous spatial pyramid pooling module is replaced with a multi-scale feature fusion module. Building upon this, coordinate attention mechanisms are applied to both shallow and deep features to strengthen the network's spatial position learning.

1.1 Study Area

As shown in [Figure 1: see original paper], the study area, Yuanyang County, is located in southern Yunnan Province, China, within the Honghe Hani and Yi Autonomous Prefecture, on the southern bank of the Ailao Mountains and the Honghe River. The geographical coordinates range from 102°27'~103°13'E, 22°49'~23°19'N, covering an area of 2,212.32 km^2 with 14 townships under its jurisdiction. The terrain features layered mountains and crisscrossing gullies with significant elevation differences, with the lowest altitude at 164 m and the highest at 2,939.6 m. The Hani terraced fields in Yuanyang County have a cultivation history of over 1,300 years, with the maximum number of terrace levels reaching 3,700. Individual terraced fields range from less than 1 m^2 to as large as 1,000 m^2 [22]. The rich variety of terraced field morphologies can represent typical mountainous terraced regions in China, meeting the research requirements for county-level automated terraced field extraction. Precise extraction of terraced field information in Yuanyang County provides fundamental data for soil and water conservation monitoring in the region.

The color, texture, and morphological characteristics of terraced fields vary seasonally. From October to April of the following year, terraced fields in Yuanyang County enter the fallow period. During this time, water is diverted through ditches to irrigate the terraces, ensuring sufficient water for the cultivation period and presenting a typical water-storage morphology.

1.2 Data Sources

This study uses Gaofen-6 (GF-6) satellite imagery provided by the Yunnan Provincial High-Resolution Earth Observation Center as the data source to meet high spatial resolution requirements. The GF-6 satellite is equipped with a 2 m panchromatic and 8 m multispectral high-resolution camera, featuring four bands: red, green, blue, and near-infrared (NIR).

Cloud-free GF-6 satellite imagery from March 30, 2021 (image ID: GF6_{{{PMS}}{{E102}}}.8{N23}.2_{{{20210330}}_{{L1A1120093056}}}}) was used, which completely covers the study area. During this period, terraced fields in the study area were in the fallow stage, with flat surfaces and sparse vegetation after irrigation, creating significant spectral differences from other land features and facilitating remote sensing identification and extraction. Additionally, a Digital Elevation Model (DEM) for elevation and slope analysis was obtained from the Geospatial Data Cloud platform (https://www.gscloud.cn/search) with a spatial resolution of 30 m.

1.3 Dataset Construction

Most publicly available remote sensing segmentation datasets do not include a terraced field category. Therefore, this study constructed a terraced field dataset using GF-6 imagery. The process of constructing the GF-6 deep learning terraced field segmentation dataset includes four key steps: data preprocessing, sample annotation, sample cropping, and dataset partitioning with training set augmentation, as shown in [Figure 2: see original paper]. First, to meet the requirements for multispectral data spatial resolution and quality, a series of preprocessing steps were performed, including radiometric calibration, atmospheric correction, and orthorectification of GF-6 multispectral imagery, as well as radiometric calibration and orthorectification of panchromatic imagery. Then, the NNDiffuse Pan Sharpening [23] tool was used to fuse the 8 m resolution multispectral imagery with the 2 m resolution panchromatic imagery to obtain 2 m resolution multispectral images.

Terraced field conditions vary with terrain characteristics, with slope, elevation, and climate being the most significant influencing factors. To ensure that terraced fields in the training samples exhibit different morphologies, 14 typical regions were selected as training sample areas based on the topographical distribution characteristics of Yuanyang County (FIGURE:1), with training samples accounting for 8.9% of the entire county area. Using ArcGIS software, terraced field samples in the 14 regions were visually interpreted and vectorized. The vector data was then converted to raster data to complete label annotation. Since CNN predictions rely on contextual information features, classification accuracy depends on the positions of various objects in the input image; objects near image edges may lack complete context and be misclassified. To mitigate this effect, a sliding window of 256 pixels with a stride of 192 pixels in each direction was used when cropping images and labels, thereby changing the positions of terraced fields in the images. Additionally, to increase sample diversity, data augmentation was applied to the training and validation sets, including random rotations of 90°, 180°, 270°, and horizontal and vertical mirroring operations to expand the training sample size, ultimately yielding 14,760 training images and 3,690 validation images.

2 Research Methods

Deep learning technology can automatically learn advanced feature representations from raw imagery, better adapting to the complexity and diversity of terraced fields in high-resolution remote sensing data. Addressing the characteristics of terraced fields, this study optimized the DeepLab v3+ network and weighted the binary cross-entropy loss function to improve model performance.

2.1 Improved DeepLab v3+

Since its proposal, the DeepLab v3+ network has been widely used for high-precision image segmentation due to its excellent capabilities [24,25]. In the encoder, the DeepLab v3+ model uses Xception as the backbone network to extract shallow and deep features, with deep features input to the ASPP module [26]. The ASPP module consists of four convolutional layers with dilation rates of 1, 6, 12, and 18, plus a global average pooling operation. In the decoder, a 1×1 convolutional layer adjusts the channel numbers of twice-compressed low-level features, which are then concatenated with high-level feature maps upsampled by a factor of 4. After stacking, features are refined through 3×3 convolution. Finally, linear interpolation upsampling produces predictions at the original image resolution.

This study proposes improvements to the classic DeepLab v3+ network model, as shown in [Figure 3: see original paper]. In the encoder section, lightweight MobileNet v2 [27] replaces Xception as the semantic segmentation model's backbone network. Two shallow features are extracted from the 4th and 7th layers of the MobileNet v2 network, with coordinate attention (CA) [28] mechanisms applied to enhance lower-level semantic information. Additionally, while the original DeepLab v3+ network uses the Atrous Spatial Pyramid Pooling (ASPP) module to enhance deep features, dilated convolution's discrete sampling tends to ignore dependencies between continuous points at large dilation rates, causing local information loss and affecting prediction results. To simultaneously preserve local details and global context, this study employs the MSFF module to replace the ASPP module, utilizing a cascaded pattern of atrous convolutions with progressively increasing dilation rates to improve information loss. In the decoder section, the 7th layer features with CA attention are adjusted and upsampled to match the size of the 4th layer features. Then, consistent with the original model, deep features with CA attention are concatenated with shallow features. Finally, through 3×3 convolution and upsampling operations, the image is restored to its original size.

2.1.1 Feature Extraction Network Compared to the original Xception, MobileNet v2 introduces inverted residual modules based on depthwise separable convolutions and linear bottleneck layers, significantly reducing model parameters and enabling faster network convergence [29]. This study further improves MobileNet v2 to reduce model parameters and simplify the model. Specifically, the first 8 layers of the MobileNet v2 network are used with a downsampling factor of 3. Meanwhile, the stride of layers 5 and 7 is changed from the original 2 to 1, and the 3×3 standard convolution in layer 7 is replaced with an atrous convolution with a dilation rate of 4. The specific network structure is shown in .

2.1.2 Coordinate Attention Module In convolutional neural networks, attention mechanisms such as SENet [30] and CBAM [31] are widely used. However, SENet only focuses on channel dimension information without considering spatial dimension information, while CBAM, although integrating channel and spatial dimension information, cannot solve long-distance dependency problems in the spatial dimension. The CA mechanism is a lightweight attention mechanism that simultaneously considers channel and spatial dimensions and can solve long-range dependency issues. Its key idea is to use coordinate information as part of the input, allowing the model to achieve cross-channel information acquisition to extract features more accurately. In terraced field semantic segmentation tasks, CA pays greater attention to the position of each pixel in the image, enabling the model to better understand the spatial structure of terraced fields and improving the recognition of terraced field boundaries, shapes, and positions.

2.1.3 Multi-scale Feature Fusion Module The MSFF module achieves multi-scale feature fusion through atrous convolutions with different dilation rates and pooling operations to improve network performance. As shown in [Figure 3: see original paper], the module processes through four parallel branch network structures: the first branch uses three atrous convolutions with dilation rates of 1, 2, and 3 to obtain feature information at smaller scales; the second branch uses three atrous convolutions with dilation rates of 1, 6, and 12 to further expand the scale of feature information; the third and fourth branches use Average Pooling and Max Pooling to obtain global and local information, respectively, followed by upsampling to restore the input image size. Finally, a 1×1 convolutional layer further fuses the combined features and adjusts the number of output feature maps. Introducing this multi-scale feature fusion module between the encoder and decoder helps improve network performance in terraced field extraction tasks, ensuring the network can better handle multi-scale terraced terrain features and reduce information loss.

Compared to the ASPP module, the MSFF module contains only four branches. In each branch, instead of using a single atrous convolution, atrous convolutions with progressively increasing dilation rates are cascaded. This design aims to reduce parameters while expanding the model's receptive field to capture information at different scales, ensuring each branch can effectively extract more multi-scale features. Additionally, the other two branches contain 2$×$2 Average Pooling and Max Pooling layers, which, when combined in parallel, can slow information loss and better preserve diversity information from the original feature maps.

2.2 Loss Function

Different land feature types in high-resolution remote sensing images occupy different proportions. Extreme imbalance between target and background affects segmentation network performance. Therefore, to reduce the impact of large feature category proportion differences on model feature classification accuracy, this study adopts an improved weighted binary cross-entropy loss function to address classification object proportion imbalance. The calculation is shown in equation (1):

$$L_{BCE} = W_1 \times [-y \log(p(y))] + W_2 \times [-(1-y)\log(1-p(y))] \quad (1)$$

Where: y is the binary label 0 or 1; $p(y)$ is the probability that the output belongs to label y; W_1 and W_2 are weight coefficients, implemented in the experiment by calculating the ratio of target and background pixels to total pixels across all training samples.

2.3 Evaluation Metrics

Accuracy evaluation is a crucial component for describing model reliability. This study calculates model accuracy using the confusion matrix method, analyzing the relationships between true positives (TP), true negatives (TN), false positives (FP), and false negatives (FN) in terraced field extraction results. Precision, recall, F1-Score, and IoU are adopted as criteria for model evaluation, with formulas shown in .

3.1 Experimental Setup

To effectively train the deep learning model, experiments were conducted on a computer with an I9-12900k 16-core 24-thread CPU, 64 GB memory, and a GV-N3090GAMING GPU. The software environment consisted of Windows 11 Professional 64-bit operating system, Anaconda 3 (64-bit) for environment configuration, Python 3.7 installed in the environment, and TensorFlow 2.4.0 as the backend deep learning framework. During experiments, batch size was set to 8, initial epochs to 200, learning rate to 0.003, with Adam as the optimizer and multi-step dynamic learning rate adjustment. To prevent overfitting, training stops and the optimal model is saved when training loss and IoU show no improvement for 10 consecutive epochs.

3.2 Impact of Different Band Combinations on Terraced Field Extraction

Previous studies have shown that the near-infrared band significantly influences deep learning models in recognizing cultivated land [32]. This study selected a 5 km × 5 km concentrated terraced field region to explore model performance under different band combinations. The reference terraced field area in this region is 976 hm^2, with 79 reference vector parcels. presents accuracy metric comparisons for RGB, NirRG, and NirRGB band combinations in the test region. In the RGB band combination, precision achieved the highest value but recall was relatively low, resulting in the lowest F1-Score and IoU. The NirRGB band combination showed more balanced performance across all metrics, with noticeable improvements in recall and IoU compared to RGB, demonstrating the promotional effect of the near-infrared band on terraced field recognition. The NirRG band combination maintained high performance across all metrics, with IoU reaching the relatively highest value, indicating better overlap between model predictions and ground truth and further confirming the promotional effect of the near-infrared band. The comparison shows that the model trained with NirRG band combination performed best overall, achieving precision, recall, F1-Score, and IoU of 90.11%, 90.22%, 90.17%, and 82.10%, respectively.

To further analyze the practical impact of band combinations on terraced field classification results, the study statistically analyzed the predicted parcel count and total area under different band combinations in the test region. The RGB band combination generated the highest number of parcels, exceeding 700, while NirRG and NirRGB combinations produced 228 and 326 parcels, respectively. In terms of total area, RGB, NirRG, and NirRGB band combinations yielded areas of 1,015 hm^2, 964 hm^2, and 928 hm^2, respectively, with NirRG showing only a 12 hm^2 error from the reference area.

To validate extraction results, visual comparisons of different band combination results overlaid on imagery are presented in [Figure 4: see original paper]. Overall, all three band combinations could roughly extract terraced field extents. However, at the local detail level, RGB combination results showed higher fragmentation, with large fragmented slope farmlands mistakenly identified as terraced fields on the left side. The NirRGB combination exhibited more obvious errors, particularly in terraced fields near building areas. In contrast, NirRG extraction results were more complete, with relatively better extraction of terraced and slope farmlands.

3.3 Terraced Field Extraction Results and Analysis

After training the network with the proposed method, the optimal trained model was used to predict terraced fields across the entire Yuanyang County. Prediction results were stitched together using a quarter-overlap approach between adjacent images to obtain the terraced field extraction results for Yuanyang County, as shown in [Figure 5: see original paper]. The results indicate that central Yuanyang County is the main concentrated distribution area for terraced fields, featuring contiguous terraced fields covering tens of thousands of mu (such as Qingkou, Quanfuzhuang, Malizhai, and Zhulu scenic terraced areas). Large-area terraced field distributions were also observed in the southern and eastern regions, though more scattered compared to the central area, mostly distributed near residential points. Almost no terraced fields were found along the northern river valley area, likely due to the steep terrain being unsuitable for terraced agriculture. [Figure 5: see original paper]a–[Figure 5: see original paper]e show test results for four typical terraced field regions in Yuanyang County. In [Figure 5: see original paper]b and [Figure 5: see original paper]c, some minor noise phenomena exist within the predicted terraced fields, slightly affecting model discrimination accuracy, but the model shows relatively accurate judgment of vegetation within terraced fields. [Figure 5: see original paper]d and [Figure 5: see original paper]e demonstrate certain accuracy in distinguishing between slope farmlands and terraced fields, showing sensitivity to different terrain features. Overall, the model exhibits high accuracy in large-area terraced field extraction tests, highlighting its good overall performance for large-scale terraced field extraction while also indicating room for further improvement in areas with complex vegetation and significant topographic changes.

To further analyze the distribution of terraced fields in Yuanyang County, slope was divided into six grades according to the Ministry of Water Resources' "Standards for Classification and Gradation of Soil Erosion" (SL190—2007) [33] ([Figure 6: see original paper]), and the distribution of terraced fields across different slopes was statistically analyzed. shows that the total predicted terraced field area is 15,562.18 hm^2. The documented Hani terraced field area in Yuanyang County is approximately 13,000 hm^2, and adding scattered non-Hani terraced fields in southern Yuanyang County roughly matches the actual terraced field area. According to , most terraced fields are distributed on slopes between 8° and 25°, accounting for 84.97% of the total terraced area. Areas with slopes less than 5° and greater than 35° account for only 1.65%, indicating that terrain conditions with excessively high or low slopes may be unsuitable for terraced agriculture. Terraced field distribution is primarily concentrated in moderate slope ranges.

Using 500 m elevation gradient intervals, six grades were established to statistically analyze the relationship between terraced field spatial distribution and elevation in Yuanyang County ([Figure 7: see original paper]). According to , terraced field elevation distribution varies significantly within Yuanyang County, with the vast majority distributed between 1,000 m and 2,000 m, occupying 95.02% of the total terraced area. The elevation gradient of 1,000–1,500 m contains the most terraced fields, accounting for 69.57% of the total area. Almost no terraced fields exist in areas below 500 m or above 2,000 m. This distribution pattern may be influenced by geographical conditions and climatic factors, holding important implications for local agricultural planning and land use decisions.

3.4 Comparison with Other Algorithms

To validate the effectiveness of the improved DeepLab v3+ model for terraced field extraction, the improved lightweight DeepLab v3+ model based on the MobileNet v2 backbone was compared with UNet, PSPNet, and the original DeepLab v3+ model under unchanged training parameters. According to , the improved DeepLab v3+ network achieved precision of 93.93%, recall of 92.08%, F1-Score of 93.17%, and IoU of 83.21%. Compared to the original DeepLab v3+ network, the four metrics improved by 4.62%, 2.61%, 3.81%, and 2.81%, respectively. Compared to PSPNet and UNet, precision improved by 7.72% and 3.49%, recall by 8.01% and 1.59%, F1-Score by 7.96% and 2.71%, and IoU by 4.73% and 3.52%, respectively.

As shown in [Figure 8: see original paper], all four models successfully extracted terraced fields, but the proposed method yielded superior results. PSPNet extraction results exhibited extensive adhesion phenomena, failing to effectively extract small non-terraced areas within terraced fields. UNet and DeepLab v3+ significantly improved adhesion issues, with more accurate overall extraction results. Although DeepLab v3+ better focused on terraced field features at different positions, excessive attention to local features caused more pronounced burring at edge areas. In contrast, the improved DeepLab v3+ network improved adhesion problems while maintaining edge clarity without introducing boundary burrs.

statistically compares parameters, floating-point operations (FLOPs), and optimal model epochs (OME) among the algorithms. The improved DeepLab v3+ network has 8 M parameters, only 28.6% of UNet and 19.5% of the original DeepLab v3+. In terms of FLOPs, the improved DeepLab v3+ network shows smaller values compared to UNet and DeepLab v3+, indicating significantly reduced model complexity and more friendly computational resource requirements for practical applications. This is also reflected in the epochs required to achieve optimal models: the improved DeepLab v3+ network requires only 108 epochs, while UNet and DeepLab v3+ require 115 and 128 epochs, respectively, demonstrating more efficient training. Compared to lightweight PSPNet, the improved DeepLab v3+ network does not show obvious advantages in the three metrics in . However, comprehensive accuracy analysis reveals that the improved DeepLab v3+ network demonstrates absolute overall superiority. In practical scenarios, model parameters and FLOPs are not the only factors affecting performance; network structure and training strategies also play key roles in comprehensive performance.

4.1 Discussion

In terraced field semantic segmentation tasks, the UNet model has a simple structure, and its skip connections facilitate multi-resolution feature fusion, achieving good results with small datasets [34]. However, due to its fixed receptive field, it cannot completely extract detailed features when facing complex land features. Compared to UNet and DeepLab v3+, PSPNet can be considered a lightweight semantic segmentation model [35], but as convolution depth increases, the model may reach a performance bottleneck, making it difficult to further improve segmentation accuracy. The DeepLab v3+ model introduces the ASPP module to simultaneously capture multi-scale contextual information, helping to improve target understanding and segmentation accuracy [36]. Since terraced fields have multi-scale spatial features, the DeepLab v3+ model can more comprehensively understand details and structures in terraced field images. However, the relatively large parameter count and high computational complexity of the DeepLab v3+ model pose challenges for practical applications, making it difficult to efficiently deploy and operate in resource-constrained environments. Lightweight DeepLab v3+ models have become an important research direction in deep learning to ensure effective segmentation of complex scenes like terraced field images while maintaining lightweight characteristics.

Model extraction accuracy is influenced not only by model structure but also by research objects and dataset processing. This study used fallow-period GF-6 terraced field imagery as the data source and selected the NirRG band combination for model training, achieving high accuracy in the specific scenario. However, these measures have limitations. First, terraced fields exhibit obvious temporal characteristics with crop growth states. Models trained on single-season data may not adapt well to such temporal variations, losing the ability to effectively capture annual changes in terraced fields. Second, while the NirRG band combination ensured extraction accuracy and terraced field integrity, obtaining high-resolution imagery with near-infrared bands is not easy; high-resolution imagery with only RGB bands is more common in practical applications. This significantly reduces the applicability of the model trained in this study when NIR band information is lacking, limiting its potential for broader application scenarios.

4.2 Conclusion

This study proposes an improved DeepLab v3+ model that replaces the backbone network with the lightweight MobileNet v2, introduces the MSFF module to replace the original ASPP module, and applies CA mechanisms to both shallow and deep features to strengthen spatial position learning. Conducting a county-level study on terraced field extraction in Yuanyang County, Honghe Hani and Yi Autonomous Prefecture, Yunnan Province, yielded satisfactory identification results. The main conclusions are as follows:

1) The near-infrared band significantly promotes the model's ability to learn terraced field features. Through band combination comparison, the NirRG band combination achieved the highest overall recognition performance and precision metrics for terraced fields.

2) Compared to PSPNet, UNet, and the original DeepLab v3+, the proposed model demonstrates higher accuracy and better performance on

the terraced field dataset. The improved DeepLab v3+ model also shows superior efficiency in terms of total parameters, floating-point operations, and epochs required to achieve optimal model performance compared to UNet and the original DeepLab v3+.

Conflict of Interest Statement: This study has no conflicts of interest to declare.

References

[1] ZHANG Y C, YANG H L, XIN Z B, et al. Extraction of small watershed terraces in the hilly areas of loess plateau through UAV images with object-oriented approach[J]. Journal of soil and water conservation, 2023, 37(3): 139-146.

[2] LI D R. Towards photogrammetry and remote sensing: Status and future development[J]. Geomatics and information science of Wuhan university, 2000, 25(1): 1-6.

[3] ZHANG H W, ZHANG W F, JIANG Z J, et al. GUS-YOLO remote sensing target detection algorithm introducing context information and Attention Gate[J]. Journal of frontiers of computer science and technology, 2024, 18(2): 453-464.

[4] SHI S S, DOU Y Y, CHEN Y Q, et al. Remote sensing monitoring based analysis of the spatio-temporal changing characteristics of regional urban expansion and urban land cover in China's coastal zones[J]. Remote sensing for natural resources, 2022, 34(4): 76-86.

[5] TIAN Z H, CHANG P, HE X H, et al. Land cover classification of high resolution remote sensing images based on CNN-GCN[J]. Science of surveying and mapping, 2023, 48(6): 59-72.

[6] ZHAO J Y, LAI G Y. Enhancement and extraction of small-scale terrace texture information for high-resolution remote sensing image[J]. Jiangxi science, 2020, 38(2): 263-268.

[7] DANG T M, MU X M, SUN W Y, et al. Review of quickly discriminating approaches of terrace information based on high resolution remote sensing images[J]. Yellow river, 2017, 39(3): 85-89, 94.

[8] LI M H, SHI Y, MA Y Q, et al. Terrace information extraction in loess hilly-gully region landscape based on object-oriented classification method[J]. Geomatics & spatial information technology, 2019, 42(5): 50-54.

[9] WU A, YUAN L, QI F, et al. Identification and evaluation of terracing measure types in hilly areas based on random forest[J]. Journal of Shandong agricultural university (natural science edition), 2023, 54(4): 582-594.

[10] DENG C X, ZHANG G Y, LIU Y J, et al. Advantages and disadvantages of terracing: A comprehensive review[J]. International soil and water conservation

research, 2021, 9(3): 344-359.

[11] ZHAO W Z, DU S H. Learning multiscale and deep representations for classifying remotely sensed imagery[J]. ISPRS journal of photogrammetry and remote sensing, 2016, 113: 155-165.

[12] JAWAK S D, DEVLIYAL P, LUIS A J. A comprehensive review on pixel oriented and object oriented methods for information extraction from remotely sensed satellite images with a special emphasis on cryospheric applications[J]. Advances in remote sensing, 2015, 4(3): 177-195.

[13] GHAMISI P, COUCEIRO M S, BENEDIKTSSON J A. Classification of hyperspectral images with binary fractional order Darwinian PSO and random forests[C]// Proc SPIE 8892, image and signal processing for remote sensing. Washington, D.C., USA: SPIE, 2013, 8892: 215-222.

[14] LIU X Y, YANG S T, WANG F G, et al. Analysis on sediment yield reduced by current terrace and shrubs-herbs-arbor vegetation in the loess plateau[J]. Journal of hydraulic engineering, 2014, 45(11): 1293-1300.

[15] XIONG L Y, TANG G A, YANG X, et al. Geomorphology-oriented digital terrain analysis: Progress and perspectives[J]. Journal of geographical sciences, 2021, 31(3): 337-354.

[16] HINTON G E, SALAKHUTDINOV R R. Reducing the dimensionality of data with neural networks[J]. Science, 2006, 313(5786): 504-507.

[17] ZHOU J, LI M M, WANG X Q, et al. Extraction of farming terraces using object-based convolutional neural networks from very high resolution satellite images[J]. Remote sensing information, 2022, 37(2): 138-144.

[18] WANG Y N, KONG X B, GUO K, et al. Intelligent extraction of terracing using the ASPP ArrU-net deep learning model for soil and water conservation on the loess plateau[J]. Agriculture, 2023, 13(7): 1283.

[19] YU M G, RUI X P, XIE W Y, et al. Research on automatic identification method of terraces on the loess plateau based on deep transfer learning[J]. Remote sensing, 2022, 14(10): ID 2446.

[20] LIU D J. Study on terraced field extraction with a deep learning method combined with both spectral and topographic features[D]. Lanzhou: Lanzhou University, 2022.

[21] ZHAO Y L, CAI D M, LYU X J, et al. Terraced field extraction in UAV imagery using improved DeepLab v3+ network[C]// 2023 8th International Conference on Intelligent Computing and Signal Processing (ICSP). Piscataway, New Jersey, USA: IEEE, 2023: 854-859.

[22] LIU J, LIU C J, JIAO Y M, et al. Study on the spatial distribution rules and variation of natural factors of hani rice terrace in Yuanyang county based on GIS spatial data[J]. Research of soil and water conservation, 2020, 27(2): 337-343.

[23] SUN W H, CHEN B, MESSINGER D. Nearest-neighbor diffusion-based pan-sharpening algorithm for spectral images[J]. Optical engineering, 2014, 53(1): ID 013107.

[24] WANG C S, DU P F, WU H R, et al. A cucumber leaf disease severity classification method based on the fusion of DeepLab v3+ and U-Net[J]. Computers and electronics in agriculture, 2021, 189: ID 106373.

[25] AZAD R, ASADI-AGHBOLAGHI M, FATHY M, et al. Attention DeepLab v3+: Multi-level context attention mechanism for skin lesion segmentation[C]// BARTOLI A, FUSIELLO A. European Conference on Computer Vision. Berlin, German: Springer, 2020: 251-266.

[26] ZHANG D Y, DING Y, CHEN P F, et al. Automatic extraction of wheat lodging area based on transfer learning method and deeplab v3+ network[J]. Computers and electronics in agriculture, 2020, 179: ID 105845.

[27] SANDLER M, HOWARD A, ZHU M L, et al. MobileNet V2: Inverted residuals and linear bottlenecks[C]// 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway, New Jersey, USA: IEEE, 2018: 4510-4520.

[28] HOU Q B, ZHOU D Q, FENG J S. Coordinate attention for efficient mobile network design[C]// 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway, New Jersey, USA: IEEE, 2021: 13713-13722.

[29] LI W, LIU K. Confidence-aware object detection based on MobileNet v2 for autonomous driving[J]. Sensors, 2021, 21(7): ID 2380.

[30] HU J, SHEN L, SUN G. Squeeze-and-excitation networks[C]// 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway, New Jersey, USA: IEEE, 2018: 7132-7141.

[31] WOO S, PARK J, LEE J Y, et al. CBAM: convolutional block attention module[C]// European Conference on Computer Vision. Berlin, German: Springer, 2018: 3-19.

[32] LIU Z Z, LI N, WANG L J, et al. A multi-angle comprehensive solution based on deep learning to extract cultivated land information from high-resolution remote sensing images[J]. Ecological indicators, 2022, 141: ID 108961.

[33] Ministry of Water Resources of the People's Republic of China. Standards for classification and gradation of soil erosion: SL 190—2007[S]. Beijing: China water & power press, 2008.

[34] RONNEBERGER O, FISCHER P, BROX T. U-net: Convolutional networks for biomedical image segmentation[C]// International Conference on Medical Image Computing and Computer-Assisted Intervention. Berlin, German: Springer, 2015: 234-241.

[35] ZHAO H S, SHI J P, QI X J, et al. Pyramid scene parsing network[C]// 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway, New Jersey, USA: IEEE, 2017.

[36] CHEN L C, ZHU Y K, PAPANDREOU G, et al. Encoder-decoder with atrous separable convolution for semantic image segmentation[C]// European Conference on Computer Vision. Berlin, German: Springer, 2018: 833-851.

Foundation items: State Administration of Science, Technology and Industry for National Defense Gaofen Special Yunnan Provincial Government Comprehensive Management of Deep Application and Large-Scale Industrialization Demonstration Projects (89-Y50G31-9001-22/23); Yunnan University Graduate Research Innovation Fund (KC-22222840)

Biography: ZHANG Jun, research direction: remote sensing information extraction. E-mail: zhjun@mail.ynu.edu.cn
Corresponding author: ZHANG Jun, Ph.D., Associate Professor, research direction: remote sensing application and GIS development. E-mail: zhjun@ynu.edu.cn