Postprint: Precise Apple Orchard Extraction in the Loess Plateau Using an Improved Linknet Network

Authors: Zhang Zhibo, Zhao Xining, Gao Xiaodong, Zhang Li, Yang Menghao

Date: 2023-02-17T00:00:00+00:00

Abstract

Over the past two decades, the apple planting area in the Loess Plateau has increased rapidly, exerting significant impacts on regional eco-hydrology and socioeconomic development. However, orchard plots in this region are small and the scenes are complex, with only county/city-scale statistical data available, and no actual spatial distribution information of apple orchards. To this end, this study established a professional dataset of low-altitude remote sensing images from unmanned aerial vehicles (UAVs). By integrating transfer learning and deep learning methods, the ResNet34 residual neural network was migrated to the Linknet network to obtain the R_{34}*{Linknet} network. The R*{34}*{Linknet} network and five commonly used deep learning semantic segmentation models (SegNet, FCN*{8s}, DeeplabV3+, UNet, and Linknet) were applied to extract the spatial distribution of apple orchards in the Loess Plateau. The best-performing model was R_{34}*{Linknet}, achieving an F1 score (harmonic mean) of 87.1%, Pixel Accuracy (PA) of 92.3%, Mean Intersection over Union (MIoU) of 81.2%, Frequency Weighted Intersection over Union (FWIoU) of 85.7%, and Mean Pixel Accuracy (MPA) of 89.6% on the test set. By combining the Atrous Spatial Pyramid Pooling (ASPP) structure with the R*{34}*{Linknet} network to expand the network's receptive field, the R*{34}*{Linknet}*{ASPP} network was obtained; then the ASPP structure was improved to obtain the R_{34}*{Linknet}*{ASPP}+ network. Comparing the performance of the three networks, the optimal performer was R_{34}*{Linknet}*{ASPP}+, achieving an F1 score of 86.3%, PA of 94.7%, MIoU of 82.7%, FWIoU of 89.0%, and MPA of 92.3% on the test set. Using R_{34}*{Linknet}*{ASPP}+ to extract apple orchard areas in Wangdonggou, Changwu County and Tongji Village, Baishui County, the accuracies were 94.22% and 95.66%, respectively. The R_{34}*{Linknet}*{ASPP}+ method proposed in this study extracts apple orchards more accurately, with better performance at orchard plot edges, and

can serve as technical support and a theoretical basis for research on spatial distribution mapping of apple orchards in the Loess Plateau.

Full Text

Preamble

Title: Accurate Extraction of Apple Orchard on the Loess Plateau Based on Improved Linknet Network

Authors: ZHANG Zhibo[1], ZHAO Xining[2]*, GAO Xiaodong[2], ZHANG Li[1], YANG Menghao[2]

Affiliations: 1. College of Water Resources and Architectural Engineering, Northwest A&F University, Yangling, Shaanxi, China 2. Institute of Soil and Water Conservation, Chinese Academy of Sciences & Ministry of Water Resources, Yangling, Shaanxi, China

Abstract: The rapid expansion of apple cultivation area on the Loess Plateau in recent years has significantly impacted regional eco-hydrology and socio-economic development. However, orchard plots in this region are small and spatially complex, with only county- or city-scale statistical data available, lacking actual spatial distribution information. To address this gap, this study established a professional dataset of low-altitude UAV remote sensing images. By integrating transfer learning and deep learning methods, the residual neural network ResNet34 was migrated to the Linknet network to obtain R_{34}*{Linknet}. When applied to apple orchard extraction on the Loess Plateau alongside five commonly used deep learning semantic segmentation models (SegNet, FCN*{8s}, DeeplabV3+, UNet, and Linknet), the best-performing model was R_{34}*{Linknet}, achieving an F1 score of 87.1%, pixel accuracy (PA) of 92.3%, mean intersection over union (MIoU) of 81.2%, frequency weighted intersection over union (FWIoU) of 85.7%, and mean pixel accuracy (MPA) of 89.6%. The Atrous Spatial Pyramid Pooling (ASPP) structure was then combined with the R*{34}*{Linknet}* network to expand the receptive field, yielding the *R*{34}*{Linknet}*{ASPP} network. Subsequently, the ASPP structure was improved to develop the R_{34}*{Linknet}*{ASPP}+ network. Comparative analysis of the three networks showed that R_{34}*{Linknet}*{ASPP}+ achieved the best performance, with F1 of 86.3%, PA of 94.7%, MIoU of 82.7%, FWIoU of 89.0%, and MPA of 92.3% on the test set. When applied to extract apple orchard areas in Wangdonggou, Changwu County and Tongji Village, Baishui County, the R_{34}*{Linknet}*{ASPP}+ method achieved extraction accuracies of 94.22% and 95.66%, respectively. The proposed method extracts apple orchards more accurately with better performance at plot boundaries, providing technical support and theoretical basis for mapping the spatial distribution of apple orchards on the Loess Plateau.

Keywords: UAV remote sensing; apple orchard extraction; deep learning;

Loess Plateau; transfer learning; residual neural network; semantic segmentation

Citation: ZHANG Zhibo, ZHAO Xining, GAO Xiaodong, ZHANG Li, YANG Menghao. Accurate extraction of apple orchard on the loess plateau based on improved linknet network[J]. Smart Agriculture, 2022, 4(3): 95-107. (in Chinese with English abstract)

1 Introduction

Over the 20 years since the implementation of the Grain-for-Green Project, apple cultivation on the Loess Plateau has developed rapidly, establishing the region as China' s and even the world' s largest high-quality apple production base. In 2018, for example, the apple cultivation area and yield on the Loess Plateau were 2.11 times and 1.47 times those of the Bohai Bay production area, respectively [1]. While the rapid development of the apple industry has greatly increased farmers' income and contributed to poverty alleviation and rural revitalization, it has inevitably altered regional eco-hydrological processes, causing negative water cycle effects [2]. Rational allocation of apple orchard distribution patterns to achieve water-adapted development has become key to sustainable development of the apple industry on the Loess Plateau. However, the complex and variable terrain of the Loess Plateau results in fragmented spatial patterns of apple orchards, making it essential to find a rapid and accurate method for obtaining regional apple orchard spatial distribution information as a prerequisite for water-adapted development.

Traditional methods for obtaining apple orchard planting area information on the Loess Plateau have relied mainly on statistics from local administrative units, hierarchical reporting, or proportional sampling surveys [3, 4]. These methods are not only resource-intensive but also fail to obtain accurate orchard planting areas at watershed and regional scales. With the rapid development of UAV and satellite remote sensing platforms, traditional machine learning methods (such as support vector machines, random forests, and maximum likelihood classification) combined with remote sensing technology [5-8] have been widely used to extract apple orchard spatial distribution information. However, these methods suffer from insufficient extraction accuracy and efficiency. Deep learning has become a research hotspot in artificial intelligence, and convolutional neural networks, as an important branch of deep learning, have achieved significant results in image classification [9, 10] and semantic segmentation [11, 12]. Classification networks use fully connected layers in their last three layers, which are one-dimensional vectors that lose two-dimensional information. In contrast, segmentation networks convert these layers into multi-channel convolutional layers with 1$×$1 convolution kernels, replacing fully connected layers with fully convolutional layers. Furthermore, semantic segmentation using Fully Convolutional Networks (FCN) performs pixel-wise classification, enabling precise segmentation of remote sensing images and demonstrating clear advantages over traditional methods and convolutional neural networks for vegetation extraction [13].

Building upon FCN, Olaf et al. [14] modified and expanded the network to create U-Net, which achieves precise segmentation results with limited training data [15]. SegNet consists mainly of symmetric encoder-decoder structures with additional output layers [16]. The backbone of FCN, U-Net, and SegNet networks is VGG16 from the Visual Geometry Group [17]. However, when networks reach a certain depth, degradation problems occur where performance becomes worse than shallow networks. Introducing ResNet into image segmentation [18] effectively solves this degradation problem and improves extraction accuracy [19]. To address the conflict between feature map size and receptive field, the Deeplab series introduced the Atrous Spatial Pyramid Pooling (ASPP) structure [20-22]. However, Deeplab series networks have too many layers, while the LinkNet network proposed by Chaurasia et al. [23] has too few layers and lacks an ASPP structure, resulting in suboptimal segmentation performance [24].

Although satellite remote sensing remains the primary means for large-scale crop classification, it suffers from high costs, long revisit cycles, low spatial resolution, and weather dependency, limiting its real-time performance and accuracy. UAV remote sensing offers advantages of high spatial resolution, short cycles, high flexibility, and minimal cloud and weather impacts, compensating for the limitations of traditional satellite remote sensing and becoming the main approach for small-area agricultural remote sensing data acquisition. In recent years, deep learning semantic segmentation algorithms that eliminate manual feature selection and fully utilize the ultra-high resolution characteristics of UAV imagery [25, 26] have provided new approaches for UAV image segmentation and classification. Therefore, this study addresses the challenges of small orchard plots and fragmented spatial distribution patterns on the Loess Plateau [27, 28] by employing UAV-acquired apple orchard remote sensing imagery, migrating ResNet34 to the Linknet network to construct an R_{34}_{Linknet} network, and combining it with ASPP for apple orchard extraction from UAV imagery on the Loess Plateau. Experimental results demonstrate the effectiveness of the proposed method, providing technical support for clarifying the distribution patterns of apple orchards on the Loess Plateau.

---

2 Methods

2.1 Study Area and Dataset Construction

The study area is located in the main apple cultivation zone of the Loess Plateau. This region has a typical continental monsoon climate, with average temperatures in the coldest month below 5.0°C, average temperatures exceeding 10°C for no more than five months, and average annual precipitation less than 470 mm [29, 30]. Twenty-eight apple orchard sampling sites with different stand ages, varieties, and management practices were selected across 12 cities/counties in the Loess Plateau region. The distribution of sampling points is shown in Table 1 , with data collection conducted from late May 2020 to early October 2020.

The flight platform was a DJI Mavic 2 Pro quadcopter UAV system, featuring a compact, foldable, and easy-to-operate design with a self-weight of only 907 g. The maximum horizontal flight speed is 72 km/h, with a maximum flight time of 31 minutes. The gimbal is three-axis stabilized with a pitch angle range of -90° to +30°. The imaging sensor is a Hasselblad 1-inch CMOS visible light RGB camera with 20 million effective pixels and a field of view of approximately 77°, with 64 GB onboard storage. During acquisition, weather conditions were sunny with ground wind speeds less than level 2, meeting aerial photography requirements. Images were captured at flight altitudes of 80-120 m, yielding 300 apple orchard images. After preprocessing, image pixels were 256$\times$480. Python' s open-source software Labelme was used for manual annotation, with pixels classified into two categories: apple orchard (ID 0) and background (ID 1). The collected images included various orchard scenes and representative interference factors, such as different light intensities, shaded and sunny slopes of mountains, background vegetation interference, varying orchard growth conditions, and changes in apple bagging status, representing multiple characteristics of apple orchards on the Loess Plateau.

To better utilize the training dataset, this study employed 9-fold cross-validation for dataset partitioning. All images were randomly divided into 9 subsets of 100 images each, with each subset selected from original and augmented images at a 4:6 ratio. Each subset contained various orchard scenes, including intercropped orchards, mountain orchards, orchards of different ages, and orchards under complex plant backgrounds. In each iteration, one subset was used as the test set while the remaining eight served as training sets, cycling through all 9 subsets to ensure every image was used for both training and testing. This produced 9 evaluation models, with the mean performance across the 9 iterations serving as the final evaluation result.

To expand the dataset, geometric and saturation transformations were applied. Saturation transformations (0.1, 0.5, 1.5), image rotation, zero-padding with scaling, and cropping with scaling were used to augment the original 300 images, generating 600 new images. Combined with the original 300 images, the total dataset comprised 900 images.

2.2 Overall Apple Orchard Extraction Scheme

Figure 1 [Figure 1: see original paper] illustrates the overall extraction scheme for apple orchards in this study. UAV imagery is processed through an R_{34}*{Linknet} network enhanced with ASPP, where ASPP is positioned between the encoder and decoder of the R*{34}_{Linknet} network. ASPP can enlarge the network' s receptive field and improve edge segmentation performance without introducing additional parameters. During training, the predicted map output by the decoder is compared with the ground truth label through a loss function until the minimum value is reached. If not at the minimum, backpropagation adjusts the parameters to obtain the final prediction map.

2.3 R_{34}_{Linknet} and ASPP

The Linknet network introduces ResNet into a U-shaped fully convolutional neural network architecture to achieve pixel-level classification. The original LinkNet uses ResNet18 as its encoder, enabling low-power semantic segmentation on mobile devices, but suffers from low accuracy and weak representation capability. This study migrated ResNet34, ResNet50, ResNet101, and ResNet152 to LinkNet as encoders. Experimental results on our dataset showed that ResNet34 performed slightly better than ResNet50, ResNet101, and ResNet152, while having a simpler network structure and fewer parameters. Therefore, ResNet34 was selected as the encoder for Linknet, constructing R_{34}_{Linknet} to enhance overall network performance and accuracy.

The R_{34}_{Linknet} network consists of two parts: the encoder and the decoder. The encoder begins with an initial block that performs convolution on the input image using a 7$×7*kernelwithstride*2, *followedbya*3×$3 max pooling layer with stride 2. The subsequent portion comprises four encoding layers. The decoder consists of four decoding layers, each containing one deconvolution layer and two convolution layers.

During R_{34}*{Linknet} network training, 3×H×W UAV images pass through the first convolution layer with stride 2, 7$×$7 kernel size, and 64 channels, producing output pixel dimensions as calculated in equation (1). The channel count becomes 64. The calculation method is shown in equation (1), where X_t^j represents the j-th feature map output at layer t, f is the ReLU activation function, n is the number of convolution kernels, X*{t-1}^i is the i-th channel image from layer t-1, E_t^{ij} is the convolution kernel at layer t, denotes the convolution operation, and B_t^j is the bias for the j-th feature map after the convolution kernel at layer t.

To enhance model robustness, reduce parameter count, and prevent overfitting, a max pooling layer is added after each convolution layer with stride 2 and size 3$×$3. After pooling, image dimensions become H/4 × W/4 while channel count remains 64.

In the encoder, the four encoding layers output feature maps with channel numbers of 64, 128, 256, and 512, expanding to 8 times the pre-encoding size. In the decoder, the four decoding layers input feature maps with channel numbers of 512, 256, 128, and 64. After passing through the decoder, feature map dimensions and channel counts are restored to match the pre-encoding size H × W, with channel number becoming 1, producing the final apple orchard extraction prediction map.

Compared with traditional convolution algorithms, atrous convolution can expand the network' s receptive field and more accurately locate targets without increasing parameters or computational cost, enabling better capture of multi-scale contextual information [31]. Feature maps obtained through atrous convolution maintain the same dimensions as input feature maps, but each output neuron possesses a larger receptive field, allowing acquisition of more detailed information while reducing resolution loss. Atrous convolution can set different dilation rates by inserting zeros into the convolution kernel to expand its size. Different dilation rates result in different network receptive fields, enabling better acquisition of multi-scale contextual information.

The receptive field determines the input layer region size corresponding to an element in a layer' s output result. Generally, larger receptive fields yield better performance than smaller ones. The input image' s receptive field is defined as 1 (RF_1 = 1), calculated as shown in equation (2), where RF_{n+1} is the receptive field of the (n+1)-th layer feature map, RF_n is the receptive field of the n-th layer feature map, kernel_{size} is the convolution kernel size, and stride is the convolution kernel stride.

Assuming the original feature is feat0 with a 3$×3*convolutionkernel*, *atrousconvolutionwithrate*0*firstgenerate*
thesecondatrousconvolutionkernelsizeequalsthereceptivefieldofasinglepixelinthethirdatrousconvolution, *m*
receptive field. The calculation of convolution kernel size corresponding to atrous rate is shown in equation (3), where K_{new} is the new convolution kernel size, k_{ori} is the original convolution kernel size, and rate is the atrous rate.

ASPP consists of four atrous convolutions with different rates and a feature fusion layer. Based on the same Input Feature Map, four parallel atrous convolutions are applied with rates set to r = {6, 12, 18, 24} and kernel size 3$×$3. The results from different convolutional layers are then fused through pixel-wise addition.

ASPP+ improves upon ASPP by adding BN layers after each atrous convolution and replacing standard convolutions with Deep Separable Convolution (DSC), with different atrous rates. ASPP+ 主要包括以下几部分: (1) a
1$×1*convolutionlayerandthree*3×3*atrousconvolutionswithrates*6, 12, 18*andBNlayers*; (2)*globalaveragepooli*
levelfeatures, *followedby*1×1*convolutionandbilinearinterpolationtorestoreoriginalsize*; (3)*thefourfeaturem*
convolution for fusion to obtain a new 256-channel feature map.

2.4 Evaluation Metrics

2.4.1 Orchard Extraction Performance Metrics Pixel Accuracy (PA), Frequency Weighted Intersection over Union (FWIoU), Mean Intersection over Union (MIoU), and Mean Pixel Accuracy (MPA) were used as evaluation metrics. Apple orchard extraction was treated as a semantic segmentation problem, with apple orchard pixels assigned value 0 and background pixels assigned value 1.

The F1 score for apple orchard extraction is calculated as shown in equation (4), where P and R represent precision and recall for the apple orchard class, respectively.

Pixel accuracy is calculated as shown in equation (5). Frequency weighted intersection over union is calculated as shown in equation (6). Mean intersection over union is calculated as shown in equation (7). Mean pixel accuracy is calculated as shown in equation (8). In these equations, k represents the number of target categories, with k+1 total categories (including target and background); i and j represent category indices; p_{ii} represents correctly classified pixels; p_{ij} and p_{ji} represent misclassified pixels.

2.4.2 Area Accuracy Evaluation Two study areas, Wangdonggou in Changwu County and Tongji Village in Baishui County, were selected for area accuracy evaluation of the semantic segmentation models. Visual interpretation identified apple orchard areas of 139.41 hm^2 and 44.97 hm^2 in these regions, respectively. Model-extracted areas were compared against these visually interpreted areas. This study proposes a method based on Python' s open-source libraries PIL and OpenCV to calculate apple orchard areas from semantic segmentation results. First, RGB images are converted to grayscale using PIL and OpenCV. Then, numpy.where broadcasting is used to count pixels belonging to the "apple orchard" class, enabling calculation of the extracted orchard area. Finally, relative accuracy compared to visually interpreted areas is computed and analyzed, as shown in equations (9) and (10). In equation (9), A_i represents the area of a certain class (hm^2), Pr represents the pixel ratio of that class, and A represents the total area of the region (hm^2). In equation (10), Pre represents area accuracy, A_1 is the apple orchard area extracted by different semantic segmentation methods (hm^2), and A_0 is the visually interpreted apple orchard area (hm^2).

2.5 Experimental Parameter Settings

The hardware platform consisted of an NVIDIA GeForce RTX 2080S (8G) GPU, 64GB RAM, and an i7-9700k CPU. The PyTorch framework was used to build the semantic segmentation networks. The initial learning rate was set to 1$\times$10^{-4}, withtheAdamoptimizerand30trainingepochs.Topreventexcessivelylargelearningratesinlatertrainingstagesfromcausing
4}$, and momentum was set to 0.9.

---

3 Results

3.1 Comparison of R_{34}_{Linknet} with Multiple Segmentation Algorithms

To validate the effectiveness of R_{34}*{Linknet}, comparative experiments were conducted with multiple semantic segmentation algorithms including Linknet,* *SegNet, FCN*{8s}, UNet, and DeeplabV3+. Parameter initialization used the He_{uniform} method, sampling from a uniform distribution in the range [-limit, limit], where limit = sqrt(6 / fan_{in}) and fan_{in} is the number of input units in the weight tensor. All networks used NLLLoss as the loss function.

The loss curves and MIoU curves during training for different models are shown in Figure 2 [Figure 2: see original paper], with performance on the test set presented in Table 2 . The results indicate that R_{34}*{Linknet} demonstrates advantages over other semantic segmentation networks in terms of training loss convergence speed and test set metrics. Its loss curve converges quickly, reaching a minimum value of 0.004 after the final iteration. On the test set, MIoU improved by 13.6%, 1.2%, 5.4%, 7.4%, and 3.1% compared to SegNet, FCN*{8s}, DeeplabV3+, UNet, and the original Linknet, respectively. Pixel accuracy improved by 6.3%, 4.9%, 2.4%, 4.0%, and 0.6%, respectively.

Figure 3 [Figure 3: see original paper] shows the segmentation results of various networks on the test set. SegNet erroneously extracted many other vegetation types, bare land, and field roads, with frequent omissions in narrow plots, producing relatively chaotic extraction results. DeeplabV3+ and UNet showed significant improvement over SegNet but still misextracted bare land and other vegetation. Linknet achieved relatively better apple orchard extraction results, reducing interference from bare land and other vegetation, but performed poorly on plot boundary details and failed to completely avoid field road effects. Compared to other networks, R_{34}*{Linknet} extracted apple orchards with less influence from field roads and other vegetation, showing substantial improvement in plot boundary details. The extraction accuracies for the apple orchard class were 86.3%, 87.2%, 89.5%, 88.7%, 92.0%, and 92.8% for SegNet, FCN*{8s}, DeeplabV3+, UNet, original Linknet, and R_{34}*{Linknet}, respectively, with R*{34}_{Linknet} achieving the highest accuracy.

3.2 Comparison Before and After Adding ASPP and ASPP+ to R_{34}_{Linknet}

To validate the effectiveness of ASPP and ASPP+ for apple orchard extraction from UAV remote sensing imagery, these modules were added to R_{34}*{Linknet} to create R*{34}*{Linknet}*{ASPP} and R_{34}*{Linknet}*{ASPP}+ networks, with comparative analysis performed before and after addition.

The training loss curves and MIoU curves are shown in Figure 4 [Figure 4: see original paper], with evaluation metrics presented in Table 3 . After adding ASPP, MIoU improved by 2.1% and pixel accuracy by 1.1%. After adding ASPP+, MIoU and pixel accuracy further improved by 2.2% and 1.3%, respectively. Figure 4 demonstrates that ASPP effectively improves network MIoU without affecting convergence speed, validating the effectiveness of ASPP and ASPP+.

The training loss curves in Figures 2 and 4 show that loss values decrease while MIoU values increase with training iterations. Table 2 and Table 3 indicate that the proposed improved algorithm R_{34}*{Linknet}*{ASPP}+ achieves the best performance across all metrics on the test set. Figure 5 [Figure 5: see original paper] compares extraction results of various networks on the test set, showing that R_{34}*{Linknet}*{ASPP} and R_{34}*{Linknet}*{ASPP}+ produce smoother and more accurate orchard plot edges compared to R_{34}*{Linknet}. Calculations on Figure 5(d) reveal that the total number of apple orchard pixels extracted by R*{34}*{Linknet}*{ASPP} and R_{34}*{Linknet}*{ASPP}+ is closer to the ground truth than that extracted by R_{34}_{Linknet}.

The improved deep learning semantic segmentation algorithm R_{34}*{Linknet}*{ASPP}+ proposed in this study demonstrates excellent extraction performance under various complex backgrounds. As shown in the segmentation result comparisons in Figures 3 and 5, the high accuracy is attributed to migrating ResNet34 [21] to the Linknet network and adding the ASPP+ [24] structure between the encoder and decoder. Figure 5 shows that R_{34}*{Linknet}*{ASPP}+ achieves the best boundary extraction accuracy, indicating that the model learns effective orchard features during training, enabling boundary pixels to connect tightly during prediction. Figure 5(a) shows input images containing wheat at the heading stage with similar color and texture to surrounding orchards; Figures 5(b) and 5(c) contain summer corn at maturity stage, post-harvest wheat stubble, and bare land with similar characteristics; Figure 6(d) shows orchards covered with reflective films appearing bright. Through learning from labeled images in complex backgrounds, the model becomes more suitable for real-world environments and demonstrates stronger robustness.

3.3 Application of R_{34}*{Linknet}*{ASPP}+

Based on the analysis in Sections 3.1 and 3.2, R_{34}*{Linknet}*{ASPP}+ demonstrates optimal performance across all metrics. To further validate its practical segmentation effectiveness, this model was applied to extract the spatial distribution of apple orchards in Wangdonggou, Changwu County and Tongji Village, Baishui County, with area extraction accuracy analyzed.

PhotoScan 1.4.5 software was used to mosaic aerial images from Wangdonggou and Tongji Village [32, 33]. Figure 6 [Figure 6: see original paper] shows the geographical locations, complete orthophotos, and high-altitude partial aerial photographs of these two regions, with complete orthophoto areas of 8.3 km^2 and 1.5 km^2, respectively. Data collection times were June 21-22, 2021 for Wangdonggou, Changwu County and June 23, 2021 for Tongji Village, Baishui County.

Tables 4 and 5 present the apple orchard area extraction accuracies of various models for these two regions. The results show that the same model achieves similar apple orchard area extraction accuracies in both Wangdonggou and Tongji Village due to similar planting structures and ground object types in the orthophotos. In both locations, the three improved semantic segmentation methods (R_{34}*{Linknet}, R*{34}*{Linknet}*{ASPP}, and R_{34}*{Linknet}*{ASPP}+) consistently outperform the other five methods. Among the three improved algorithms proposed in this study, R_{34}*{Linknet}*{ASPP}+ achieves the highest accuracy, with apple orchard area extraction accuracies of 94.22% in Wangdonggou and 95.46% in Tongji Village. In Wangdonggou, this represents improvements of 1.21% and 0.58% over R_{34}*{Linknet} and R*{34}*{Linknet}*{ASPP}, respectively. In Tongji Village, the improvements are 1.70% and 0.90%, respectively.

Figure 7 [Figure 7: see original paper] shows the spatial distribution of apple orchards in Wangdonggou, Changwu County and Tongji Village, Baishui County extracted using the R_{34}*{Linknet}*{ASPP}+ semantic segmentation method.

---

4 Conclusion

This study proposes a more efficient and accurate extraction method for apple orchard distribution information on the Loess Plateau, ensuring extraction accuracy and achieving better performance at orchard plot boundaries.

1. Based on the characteristics of UAV imagery on the Loess Plateau, a professional dataset was created for apple orchard recognition from low-altitude remote sensing images in this region, containing various orchard scenes and representative interference factors.

2. By integrating transfer learning and deep learning methods, the residual neural network ResNet34 was migrated to the Linknet network to obtain the R_{34}*{Linknet} network. When applied to apple orchard spatial distribution extraction on the Loess Plateau alongside five commonly used deep learning semantic segmentation models (SegNet, FCN*{8s}, DeeplabV3+, UNet, and Linknet), the best-performing model was R_{34}_{Linknet}, achieving an F1 score of 87.1%, pixel accuracy (PA) of 92.3%, mean intersection over union (MIoU) of 81.2%, frequency weighted intersection over union (FWIoU) of 85.7%, and mean pixel accuracy (MPA) of 89.6% on the test set.

3. The Atrous Spatial Pyramid Pooling (ASPP) structure was combined with the R_{34}*{Linknet} network to expand the receptive field, yielding the R*{34}*{Linknet}*{ASPP} network. The ASPP structure was then improved to develop the R_{34}*{Linknet}*{ASPP}+ network. Comparative analysis of the three networks showed that R_{34}*{Linknet}*{ASPP}+ achieved the best performance, with F1 of 86.3%, PA of 94.7%, MIoU of 82.7%, FWIoU of 89.0%, and MPA of 92.3% on the test set. When applied to extract apple orchard areas in Wangdonggou, Changwu County and Tongji Village, Baishui County, the extraction accuracies were 94.22% and 95.66%, respectively. In Wangdonggou, this represents improvements of 1.21% and 0.58% over R_{34}*{Linknet} and R*{34}*{Linknet}*{ASPP}, respectively. In Tongji Village, the improvements are 1.70% and 0.90%,

respectively.

The proposed R_{34}*{Linknet}*{ASPP}+ method can accurately extract apple orchards with better edge treatment, providing technical support and theoretical basis for research on spatial distribution mapping of apple orchards on the Loess Plateau.

References

[1] ZHOU J, ZHAO D, CHEN Y, et al. Analysis of apple producing area changes in China[J]. Journal of Fruit Science, 2021, 38(3): 372-384.

[2] GAO X, ZHAO X, WU P, et al. The economic-environmental trade-off of growing apple trees in the drylands of China: A conceptual framework for sustainable intensification[J]. Journal of Cleaner Production, 2021(3): 126-497.

[3] KANG L, LEI Y, ZHENG L, et al. Vegetation classification based on MODIS data and the accuracy evaluation in the pixel scale[J]. Remote Sensing Technology and Application, 2007(3): 361-366.

[4] WU M, YANG L, YU B, et al. Mapping crops acreages based on remote sensing and sampling investigation by multivariate probability proportional to size[J]. Transactions of the CSAE, 2014, 30(2): 146-152.

[5] LIU J. Research on apple orchards information extraction of Fufeng County based on Landsat8 remote sensing image[D]. Yangling: Northwest A&F University, 2015.

[6] KUMAR A, SINGH K, LAL B, et al. Mapping of apple orchards using remote sensing techniques in cold desert of Himachal Pradesh, India[J]. Journal of the Indian Society of Remote Sensing, 2008, 36(4): 387-392.

[7] YUAN H, MA R, LUO J. Mapping orchards on plain terrains using multi-temporal medium-resolution satellite imagery[J]. Applied Engineering in Agriculture, 2015, 31: 351-362.

[8] XU H, LIU C, WANG J, et al. Study on extraction of citrus orchard in Gannan region based on Google Earth Engine platform[J]. Journal of Geo-information Science, 2018, 20(3): 396-404.

[9] WAN Y, ZHONG X, LIU H, et al. Survey of application of convolutional neural network in classification of hyperspectral images[J]. Computer Engineering and Applications, 2021, 57(4): 1-10.

[10] BERA S, SHRIVASTAVA V K. Analysis of various optimizers on deep convolutional neural network model in the application of hyperspectral remote sensing image classification[J]. International Journal of Remote Sensing, 2020, 41: 2664-2683.

[11] MAJEED Y, ZHANG J, ZHANG X, et al. Deep learning based segmentation for automated training of apple trees on trellis wires[J]. Computers and Electronics in Agriculture, 2020, 170: ID 170.

[12] JEONGEUN K, JEAHWI S, SUKWOO L, et al. An intelligent spraying system with deep learning-based semantic segmentation of fruit trees in orchards[C]// IEEE International Conference on Robotics and Automation (ICRA). Piscataway, New York, USA: IEEE, 2020: 3923-3929.

[13] LONG J, SHELHAMER E, DARRELL T. Fully convolutional networks for semantic segmentation[J]. IEEE Transactions on Pattern Analysis & Machine Intelligence, 2016, 39: 640-651.

[14] OLAF R, PHILIPP F, THOMAS B. U-Net: convolutional networks for biomedical image segmentation[C]// International Conference on Medical Image Computing and Computer-Assisted Intervention. Piscataway, New York, USA: IEEE, 2015: 234-241.

[15] MA J, DENG Y, MA Z. Review of deep learning segmentation methods for CT images of liver tumors[J]. Journal of Image and Graphics, 2020, 25(10): 2024-2036.

[16] ZHANG K, JI J, JIANG L, et al. The segmentation of driving regions on unstructured road based on SegNet architecture[J]. Journal of Chongqing University, 2020, 43(3): 79-87.

[17] SIMONYAN K, ZISSERMAN A. Very deep convolutional networks for large-scale image recognition[J/OL]. arXiv: 1409.1556, 2014.

[18] HE K, ZHANG X, REN S, et al. Identity mappings in deep residual networks[J/OL]. arXiv: 1603. 05027, 2016.

[19] ANAGNOSTIS A, TAGARAKIS A C, ASIMINARI G, et al. A deep learning approach for anthracnose infected trees classification in walnut orchards[J]. Computers and Electronics in Agriculture, 2021, 182: ID 182.

[20] PAPANDREOU G, CHEN L, KEVIN P, et al. Weakly- and semi-supervised learning of a deep convolutional network for semantic image segmentation[C]// IEEE International Conference on Computer Vision. Piscataway, New York, USA: IEEE, 2015: 1742-1750.

[21] CHEN L, PAPANDREOU G, KOKKINOS I. DeepLab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected CRFs[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2016, 40: 834-848.

[22] CHEN L, PAPANDREOU G, SCHROFF F, et al. Rethinking atrous convolution for semantic image segmentation[J/OL]. arXiv:1706.05587v3, 2017.

[23] CHAURASIA A, CULURCIELLO E. LinkNet: exploiting encoder representations for efficient semantic segmentation[C]// IEEE Visual Communications and Image Processing. Piscataway, New York, USA: IEEE, 2017: 1-4.

[24] DENG W, MOU Y, KASHIWA T, et al. Vision based pixel-level bridge structural damage detection using a link ASPP network[J]. Automation in Construction, 2020, 110: ID 102973.

[25] YU Y, LI H, WANG J, et al. A multilayer pyramid network based on learning for vehicle logo recognition[J]. IEEE Transactions on Intelligent Transportation Systems, 2021, 22(5): 3123-3134.

[26] MO J, LAN Y, YANG D, et al. Deep learning-based instance segmentation method of litchi canopy from UAV-acquired images[J]. Remote Sensing, 2021, 13(19): ID 3919.

[27] WANG A. Apple tree identification and evapotranspiration process simulation on the Loess Plateau[D]. Yangling: Northwest A&F University, 2020.

[28] WANG P. Nutrient input, soil nutrient contents and their spatial distribution characteristics in the main apple producing areas of the Loess Plateau[D]. Yangling: Northwest A&F University, 2021.

[29] YAN L. Characteristics of temperature and precipitation on the Loess Plateau from 1961 to 2014[J]. Journal of Earth Environment, 2015, (5): 276-282.

[30] WU Q, ZHANG B, MA B, et al. Impact of climate warming on winter wheat planting in the Loess Plateau[J]. Ecology and Environmental Sciences, 2017, 26: 429-436.

[31] CHEN L, ZHU Y, PAPANDREOU G, et al. Encoder-decoder with atrous separable convolution for semantic image segmentation[J]. Computer vision-ECCV, 2018, 11211: 833-851.

[32] CHEN P, XU X. A comparison of photogrammetric software packages for mosaicking unmanned aerial vehicle (UAV) images in agricultural application[J]. Acta Agronomica Sinica, 2020, 46(7): 1112-1119.

[33] WANG H. Application of PhotoScan in UAV remote sensing image data processing[J]. Geomatics & Spatial Information Technology, 2017, 40(5): 109-111.

Note: Figure translations are in progress. See original paper for figures.