



Article

Combining Object-Oriented and Deep Learning Methods to Estimate Photosynthetic and Non-Photosynthetic Vegetation Cover in the Desert from Unmanned Aerial Vehicle Images with Consideration of Shadows

Jie He ¹ , Du Lyu ^{2,3} , Liang He ¹, Yujie Zhang ¹, Xiaoming Xu ⁴, Haijie Yi ², Qilong Tian ² , Baoyuan Liu ¹ and Xiaoping Zhang ^{1,2,*}

¹ State Key Laboratory of Soil Erosion and Dryland Farming on the Loess Plateau, Institute of Soil and Water Conservation, Northwest A&F University, Xianyang 712100, China

² Institute of Soil and Water Conservation, Chinese Academy of Sciences and Ministry of Water Resources, Xianyang 712100, China

³ Shaanxi Satellite Application Center for Natural resources, Xi'an 710002, China

⁴ College of Urban, Rural Planning and Architectural Engineering, Shangluo University, Shangluo 726000, China

* Correspondence: zhangxp@ms.iswc.ac.cn



Citation: He, J.; Lyu, D.; He, L.; Zhang, Y.; Xu, X.; Yi, H.; Tian, Q.; Liu, B.; Zhang, X. Combining Object-Oriented and Deep Learning Methods to Estimate Photosynthetic and Non-Photosynthetic Vegetation Cover in the Desert from Unmanned Aerial Vehicle Images with Consideration of Shadows. *Remote Sens.* **2023**, *15*, 105. <https://doi.org/10.3390/rs15010105>

Academic Editors: Kuo-Chin Fan, Yang-Lang Chang, Toshifumi Moriyama and Ying-Nong Chen

Received: 9 November 2022

Revised: 20 December 2022

Accepted: 20 December 2022

Published: 25 December 2022



Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

Abstract: Soil erosion is a global environmental problem. The rapid monitoring of the coverage changes in and spatial patterns of photosynthetic vegetation (PV) and non-photosynthetic vegetation (NPV) at regional scales can help improve the accuracy of soil erosion evaluations. Three deep learning semantic segmentation models, DeepLabV3+, PSPNet, and U-Net, are often used to extract features from unmanned aerial vehicle (UAV) images; however, their extraction processes are highly dependent on the assignment of massive data labels, which greatly limits their applicability. At the same time, numerous shadows are present in UAV images. It is not clear whether the shaded features can be further classified, nor how much accuracy can be achieved. This study took the Mu Us Desert in northern China as an example with which to explore the feasibility and efficiency of shadow-sensitive PV/NPV classification using the three models. Using the object-oriented classification technique alongside manual correction, 728 labels were produced for deep learning PV/NPV semantic segmentation. ResNet 50 was selected as the backbone network with which to train the sample data. Three models were used in the study; the overall accuracy (OA), the kappa coefficient, and the orthogonal statistic were applied to evaluate their accuracy and efficiency. The results showed that, for six characteristics, the three models achieved OAs of 88.3–91.9% and kappa coefficients of 0.81–0.87. The DeepLabV3+ model was superior, and its accuracy for PV and bare soil (BS) under light conditions exceeded 95%; for the three categories of PV/NPV/BS, it achieved an OA of 94.3% and a kappa coefficient of 0.90, performing slightly better (by ~2.6% (OA) and ~0.05 (kappa coefficient)) than the other two models. The DeepLabV3+ model and corresponding labels were tested in other sites for the same types of features: it achieved OAs of 93.9–95.9% and kappa coefficients of 0.88–0.92. Compared with traditional machine learning methods, such as random forest, the proposed method not only offers a marked improvement in classification accuracy but also realizes the semiautomatic extraction of PV/NPV areas. The results will be useful for land-use planning and land resource management in the areas.

Keywords: unmanned aerial vehicle (UAV) images; PV/NPV classification; shaded features; deep learning semantic segmentation; object-oriented technique; desert

1. Introduction

Vegetation is a critical component of terrestrial ecosystems and an important link between material cycles and energy flows in the soil, hydrosphere, and atmosphere [1].

Depending on its activity characteristics, vegetation can be divided into photosynthetic vegetation (PV) (i.e., plant bodies with chlorophyll tissues that allow photosynthesis, especially green leaves) and non-photosynthetic vegetation (NPV) (i.e., plant bodies that do not contain chlorophyll or do not have photosynthetic functions due to reduced chlorophyll, such as branches, stems, litter, and dry leaves) [2]. Among them, surface litter, as an important material that affects the energy cycle and ecosystem material cycle, is also included as NPV in this study. The accurate estimation of the fractional cover of PV (f_{PV}), NPV (f_{NPV}), and bare soil (BS) (f_{BS}) is of considerable importance for estimating ecosystem service functions (e.g., biomass, carbon sources/sinks, and water as well as soil conservation) in addition to improving the precision of related models; it has become a contemporary important issue in the field of ecological environment research [3–6].

Estimates of f_{PV} and f_{NPV} are typically retrieved by field surveys and remote sensing [7]. In recent years, low-altitude unmanned aerial vehicle (UAV) remote sensing technology, which offers the rapid and accurate extraction of vegetation information [8], has been widely applied in forest resource surveys [9], biomass estimation [10,11], vegetation cover estimation [12], leaf area index monitoring [13], precision agriculture, and more [14,15]. Traditional visual interpretation is highly accurate but time-consuming and labor-intensive, requiring experienced interpreters. Object-based image analysis (OBIA) can improve image classification accuracy and efficiencies by implementing multiple types of information (e.g., spectrum, texture, and shape) [16].

Numerous studies have combined OBIA with random forest (RF) [17], support vector machine (SVM) [18], k-nearest neighbors (KNN) [19], and other machine learning algorithms for unmanned vegetation cover estimations in wetland, sandy, urban, and other subsurface land classes, achieving an overall accuracy of 85 to 90% [12,20]. However, traditional machine learning algorithms have a strong reliance on the human–machine interaction process, lack intelligence and automation capabilities, and are less migratable.

In recent years, following the development of computer vision and artificial intelligence, deep learning convolutional neural networks (CNNs) have been widely applied in various image processing applications, owing to their powerful ability to automatically extract features. CNNs are trained by mass data labeling to build deep neural networks, and they automatically extract features closely related to the target task through the application of a loss function; these models offer strong robustness as well as migration capabilities [21] and have emerged as a hot topic of research in remote sensing. CNNs are primarily applied in the surveillance of weeds [22], the identification of plant pests and diseases [23], the estimation of crop yields [24], and the classification of precision crops [25].

Deep learning semantic segmentation algorithms offer strong extraction capabilities for spatial features; they can ensure the rapid and accurate classification of remote sensing images [26]. More advanced semantic segmentation models have been developed via the addition of fully convolutional neural networks (FCNs); current examples include U-Net [27], SegNet [28], PSPNet [29,30], DeepLab [31,32], and more.

An increasing number of studies have combined UAV data with deep learning semantic segmentation models to perform vegetation extraction [21,33,34]. However, in desert areas vegetation is difficult to classify due to the small leaf sizes and discrete spatial distributions. Currently, most research on desert vegetation based on UAVs focuses on the estimation of f_{PV} , and few studies have considered the estimation of f_{NPV} , which occupies a large ecological niche.

When estimating UAV PV/NPV cover using deep learning semantic segmentation, it is extremely important to obtain image data labels quickly and accurately. The deep semi-supervised learning methods based on object-oriented random forest are highly accurate, easy to operate, and very popular; however, applications for PV/NPV classification using UAV images remain rare. Vegetation has a three-dimensional structure, and large areas of shadow masking and obscuration occur when the observation angle of the UAV sensor does not coincide with the direct sunlight direction [35]. Features present under high levels of shadow occlusion are misclassified, which seriously affects the accuracy of feature

classification [12,36]. The current shadow processing approaches for remote sensing images often apply spectral recovery; however, this changes the information for non-shaded areas and reduces the accuracy of the source image [37]. Commonly used models and algorithms for de-shadowing (e.g., histogram matching and the linear stretching of gray-scale images) involve more parameters, are complicated to operate, and are unsuitable for the fast extraction of PV/NPV from UAV images [38]. Whether the features in shadow areas of UAV images can be further refined and classified must also be studied in depth.

The Chinese Mu Us Sandland is located in the arid and semi-arid desert vegetation zone; it presents with few plant species, low vegetation cover, a simple structure, and fragile ecology. The accurate and rapid estimation of its vegetation cover characteristics has become a pressing problem [39,40]. In this study, we take the Mu Us Sandland as an example, use DJI Phantom 4 Pro UAV RGB aerial photographs as the database, and incorporate shaded object classification; we then explore the feasibility and efficiency of combining this classification with object-oriented techniques for PV/NPV cover estimation using three recently emerging deep learning semantic segmentation models: DeepLabV3+, PSPNet, and U-Net. The research proceeds as follows:

1. We evaluate the feasibility of object-oriented detection combined with deep learning semantic segmentation for the estimation of desert-area f_{PV} and f_{NPV} from UAV high-spatial-resolution visible light images.
2. We compare three deep learning semantic segmentation methods as well as three machine learning algorithms for the estimation of f_{PV} and f_{NPV} , considering shaded feature classification.
3. We develop a semiautomatic labeling method based on object-oriented random forest for PV/NPV data preprocessing.
4. We apply the results to vegetation monitoring in the same type of area to verify the generalizability of the model.

This study will provide a theoretical basis for automated and accurate vegetation surveyance in desert areas and provide scientific as well as technological support for the evaluation and formulation of relevant ecological protection policies.

2. Research Region and Data

2.1. Study Area

The Mu Us Sandland is located south of the Ikezhao League in Inner Mongolia and northwest of Yulin City in Shaanxi Province; it covers an area of 32,100 km², with an annual precipitation decreasing from 400 mm in the southeast to ~250 mm in the northwest. The vegetation coverage in this area is 40–50%, and a wide variety of sandy and hardy plants grow on the fixed and semi-fixed sand dunes, primarily *Artemisia scoparia* Waldst. et Kit., *Caragana Korshinskii* Kom., *Salix cheilophila* Schneid., and *Populus* L. [41,42]. In this study, a sample plot including trees, shrubs, and grasses was selected in a typical sandy area of Balasu Town, Yuyang District, Yulin City, as shown in Figure 1.

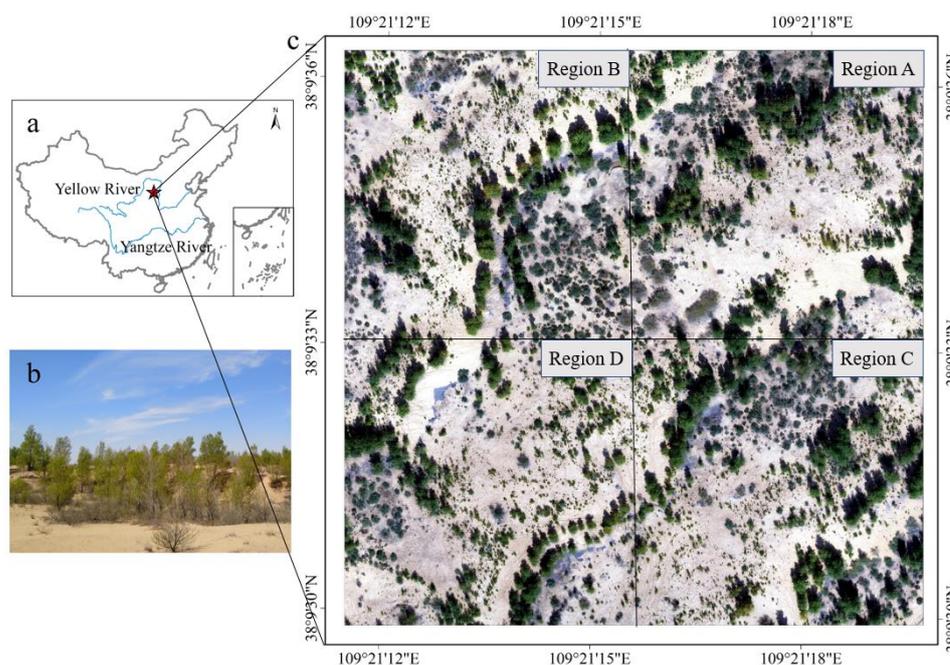


Figure 1. (a) Location, (b) photograph of the study area, and (c) the UAV orthographic mosaic image of the study area and four sub-regions (Regions A, B, C, and D were used as alternative location samples and did not intersect).

2.2. UAV Image Stitching and Processing

The quadrotor DJI Phantom 4 Pro UAV, which is lightweight, flexible, and offers a stable shooting platform, was used to acquire low-altitude image datasets. The UAV platform was equipped with a complementary metal oxide semiconductor digital camera with a resolution of 1920×1080 pixels and a field of view of 84° , covering three visible wavelengths: R (red), G (green), and B (blue).

Remote sensing images of the UAV were acquired at 11:30–12:00 a.m. on 13 July and 24 September 2019 under clear weather conditions, without clouds or wind. The flight height of the UAV was ~ 50 m, the camera angle was -90° , the flight speed was 1.2 m/s, the direction and side overlaps were 80%, the number of main routes was nine, the flight area was 4×10^4 m², the flight time was ~ 15 min, and a total of 700 original aerial images were obtained in the two time periods and at four sample sites, with an image resolution of 0.015 m (Table 1).

Table 1. UAV flight parameters.

Item	Basic Parameters
Flight height/m	50
Flight elevation/m	1200
Flight speed/(m/s)	1.2
Camera angle/ $^\circ$	-90
Heading overlap/%	80
Parallel overlap/%	80
Flight duration/min	15
Image Resolution/m	0.015
Number of aerial images	350
Aerial survey area/m ²	40,000

In this study, the images were orthorectified and stitched using Pix4Dmapper UAV image stitching software. First, the filtered UAV images were imported, and the geographic location information was added. Second, the eponymous feature points were automatically

encrypted by an aerial triangulation model. The geometric correction of the images was then obtained by combining the above eponymous feature points with ground control point data. Finally, the orthophoto and digital surface model were finally generated, as shown in Figure 1. The R-, G-, and B-band transformations were used to enhance the image and minimize the influences of shadows upon vegetation estimation. The data from September Region A were selected to train the model, September Regions B, C, and D were used as alternative location samples for the same period, July Region A was used as an alternative set of samples for the same location, and July Regions B, C, and D were used as alternative location and period samples to verify the migration capabilities of the model. The four regions did not intersect (see Figure 1).

3. Methods

3.1. PV/NPV Estimation Process

An object-oriented and deep-learning-based PV/NPV estimation architecture was proposed, as shown in Figure 2. This architecture contained four layers: UAV image preprocessing, labeled dataset construction, semantic segmentation model training, and PV/NPV classification accuracy evaluation. The experiment used the deep learning module in ArcGIS Pro, Pytorch, as the development framework, a GeForce GTX 1080 Ti as the hardware configuration, 11 Gb of video memory, and 128 Gb of RAM; this setup could quickly train and test the semantic segmentation model. The parameters of the network training model were set as follows: number of channels: 3; number of categories: 6; batch size: 4; epoch: 100; base learning rate: 1E-3; learning strategy: ploy; power: 0.9; and gradient descent method: stochastic gradient descent.

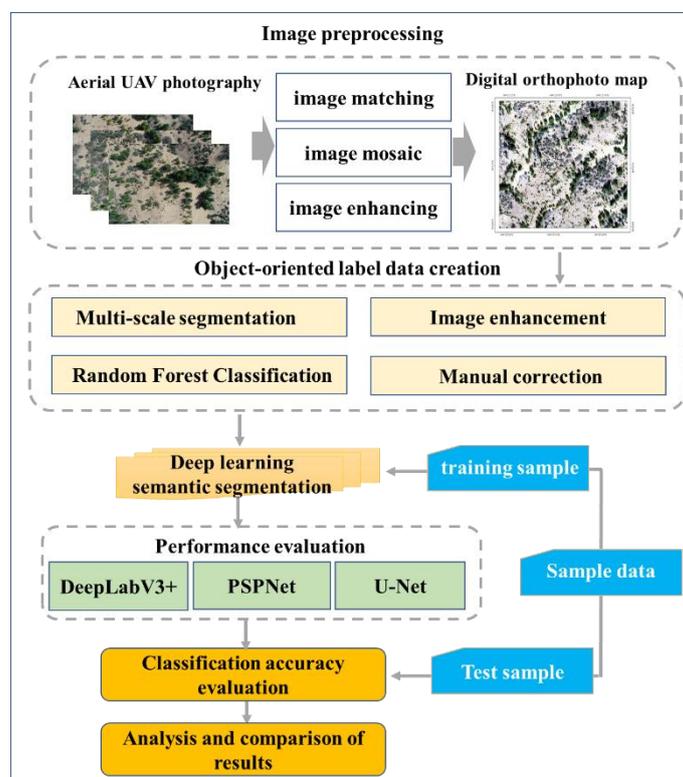


Figure 2. PV/NPV estimation architecture.

To investigate the advantages and disadvantages of deep learning for the estimation of PV/NPV in desert areas from high-resolution UAV remote sensing images, this study used three typical semantic segmentation networks, DeepLabV3+, PSPNet, and U-Net, to conduct comparison experiments and select the most suitable candidate for migration

testing. To demonstrate the advantages and disadvantages of the selected model in UAV PV/NPV estimation, three typical machine learning methods, RF, SVM, and KNN, were compared against the optimized deep learning model. The three typical machine learning experiments were based on ENVI and eCognition software.

3.2. Object-Oriented Label Data Creation

The first stage of feature extraction in a deep learning semantic segmentation model is to annotate the image data at the pixel level. The traditional annotation method is manual visual interpretation, which is simple but time-consuming and labor-intensive. We propose an object-oriented classification method supplemented with manual correction for data annotation; this includes three steps: (1) image segmentation, (2) segmented image classification, and (3) the manual correction of the classification result.

3.2.1. Feature Classification

To maximally mitigate the impacts of shading upon the feature extraction accuracy, the features in this study were classified into six feature types: unshaded PV (PV1), shaded PV (PV2), unshaded NPV (NPV1), shaded NPV (NPV2), unshaded BS (BS1), and shaded BS (BS2). The unshaded PV appeared bright green in the image; the unshaded NPV was distributed around the PV and appeared dark gray in the image, with a rough texture and a tridimensional appearance; and the unshaded BS was distributed in successive patches and appeared bright yellow. The shaded PV, NPV, and BS were dark green, dark gray, and dark yellow, respectively.

3.2.2. Multi-Resolution Segmentation

The object-oriented classification was based on image segmentation, in which image elements of the same object are given the same attribute meaning after segmentation. The principle of multi-resolution segmentation (MRS) is to quantitatively evaluate intra- and inter-object heterogeneity via bottom-up traversal and merging from a single pixel, and then to fully optimize the image to aggregate the elements at different scales [32]. MRS consists of three core parameters: scale parameter (SP), shape, and compactness. SP determines the maximum overall heterogeneity allowed for the resulting object and can vary the size of the image segmentation object; it is the most important parameter in MRS and has a large impact on the classification accuracy. The shape defines the size of the structural contrast between the uniformity of the shape and the uniformity of the spectral value. The compactness is used to optimize the tightness or smoothness of the structural contrast of the object in the image. The results obtained using different segmentation parameters are shown in Figure 3.

To determine the parameter combinations suitable for PV–NPV estimation in the study area, the shape and compactness factors were set to 0.1, 0.3, 0.5, 0.7, and 0.9; the traversal method was adopted to compare different parameter combinations. When the shape was 0.1 and the compactness was 0.3, the internal homogeneity of the object was higher, the segmentation of the PV and NPV edge was clearer, and the estimation performance was strong (Figure 4b). Based on the combinations of parameters of the optimal homogeneity criterion, this study performed MRS for parameters in the range of 1 to 400 (increment: 10) to determine the optimal parameters of the segmentation scale. The results showed that, for PV/NPV features in sandy land, the optimal segmentation parameters were shape: 0.1; compactness: 0.3; and SP: 61. The segmentation results could finely portray the spatial distribution patterns of PV and NPV.

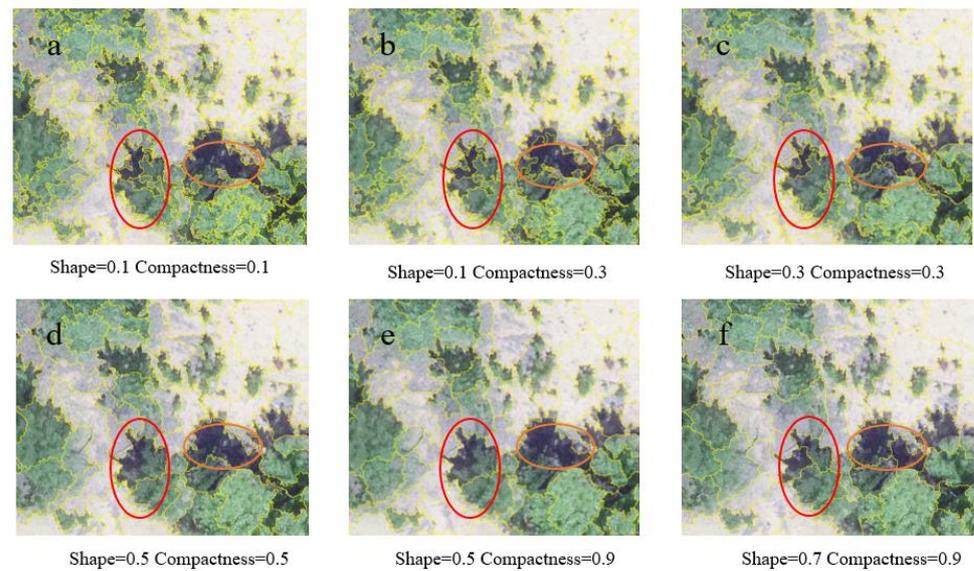


Figure 3. Effect of image segmentation for different segmentation parameters under the object-oriented method. (a) Segmentation effect with shape factor and compactness factor set to 0.1 and 0.1; (b) Segmentation effect with shape factor and compactness factor set to 0.1 and 0.3; (c) Segmentation effect with shape factor and compactness factor set to 0.3 and 0.3; (d) Segmentation effect with shape factor and compactness factor set to 0.5 and 0.5; (e) Segmentation effect with shape factor and compactness factor set to 0.5 and 0.9; (f) Segmentation effect with shape factor and compactness factor set to 0.7 and 0.9.

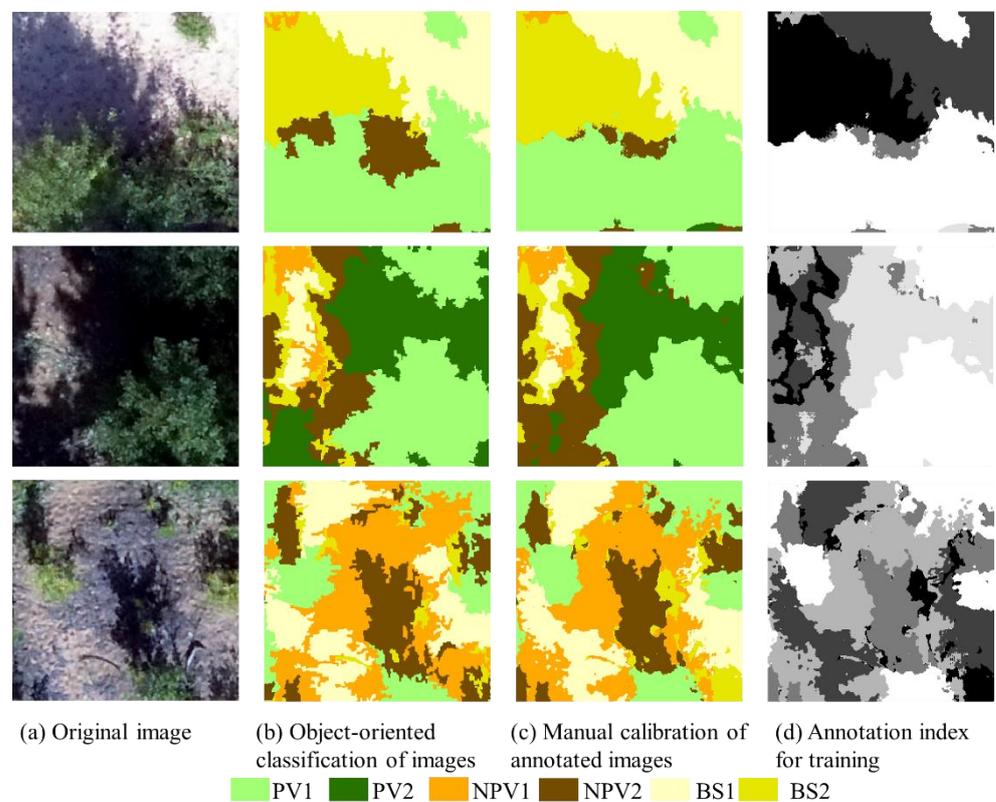


Figure 4. Object-oriented approach (combined with manual correction) for data annotation.

3.2.3. Object-Oriented Random Forest Classification Algorithm

We defined six feature classes: PV1, PV2, NPV1, NPV2, BS1, and BS2. For each class, an appropriate number (~2/3) of samples was selected as training samples to calculate the shape, spectrum, and other characteristic values of the objects; we then established a suitable feature space for the object to guide classification using the parameters of the feature space. However, too many classification features can lead to redundancies, resulting in problems such as increased computation, reduced classification efficiency, and even reduced accuracy. After comparison, the optimal feature space was selected to include 15 features that offer the largest differentiability between different classes; these were as follows: visible vegetation index, normalized green–blue difference index, red–green ratio index, normalized green–red difference index, overgreen index, difference vegetation index, spectral features, RGB raw band mean, brightness, max.dff, texture features, gray level cooccurrence matrix (GLCM) mean, GLCM homogeneity, GLCM ang.2nd moment, GLCM dissimilarity, and GLCM correlation. The random forest algorithm was then used to classify the images into features.

3.2.4. Manual Correction of Classification Results

The random forest algorithm can suffer from misclassifications and omissions when classifying features; these need to be corrected by manual visual interpretation. Manual correction requires the direct assignment of attributes to polygon units or the changing of attributes based on the results of manual visual interpretation. After manual correction was completed, the classification result map was converted into an index map (Figure 4) for subsequent training of the semantic segmentation model. Considering the limitations of the experimental equipment and the network structure of the model, the regional images were segmented into 512×512 pixel image sets with a cut step of 256. To expand the dataset (to alleviate the overfitting and enhance the generalizability of the model), data augmentation was performed; this not only increases the number of datasets but also enhances the stability and generalizability of the network. The data-augmented training set contained 728 images, the validation set contained 548 images, and the test set contained 180 images.

3.3. Deep Learning Semantic Segmentation Methods

Currently, numerous network models are available for deep semantic segmentation, including both fully and weakly supervised image semantic segmentation methods; however, the performances of most weakly supervised methods still lag behind those of the former. Therefore, in this study fully supervised image semantic segmentation methods were chosen. Three of the more advanced network models, DeepLabV3+, PSPNet, and U-Net, were selected to segment UAV images.

3.3.1. DeepLabV3+

DeeplabV3+ is the fourth-generation model of the Deeplab series. DeepLabV3+ incorporates an atrous spatial pyramid pooling (ASPP) structure and encoder–decoder method (commonly used for semantic segmentation) to fuse multi-scale information; it not only fully exploits multi-scale contextual information but can also determine the boundary of an object by reconstructing the empty information of an image.

In the encoder–decoder architecture, multiple dilated convolution modules with different dilation rates are designed to fuse multi-scale features without increasing the number of parameters, and the dilated convolution increases the receptive field without information loss, giving each convolution output a larger range of information. DeepLabV3+ applies deep separable convolution to ASPP and a decoder to replace all max pooling operations, thereby reducing the number of parameters and improving the model speed.

The details of DeepLabV3+ model structure could be seen in ref [31]. The proponent study used ResNet50 as the skeleton network in the encoding part to initially extract low-level features from UAV images. The ASPP module then performed a 1×1 convolution; a

3×3 convolution with expansion rates of 6, 12, and 18; and global average pooling before the fusion of the feature maps. The fused feature maps were inputted into a 256-channel 1×1 convolution layer to obtain high-level feature maps, and these maps were inputted into the decoder. In the decoding part, the low-level feature map obtained from Xception was first downsampled using a 1×1 convolution; it was then fused with the high-level feature map up-sampled via four-fold bilinear interpolation. Finally, a 3×3 convolution operation was performed and then restored to the original map size via four-fold bilinear interpolation to obtain the segmentation prediction map. In this study, the activation functions used in DeepLabV3+ were all ReLU functions.

3.3.2. PSPNet

PSPNet is a pyramid scene parsing network originally proposed by Zhao et al. [29]; it is a widely cited semantic segmentation network. This network is characterized by the inclusion of a pyramid pooling module (PPM) and atrous convolution, which can fully exploit the global scene to aggregate contextual information between different label categories [30].

The details of the PSPNet model structure could be seen in ref [29]. To maintain the weight of global features, the UAV image sample set was first inputted into a ResNet50 skeleton network [43] to extract feature maps; the PPM then module divided the feature layer into 1×1 , 2×2 , 3×3 , and 6×6 regions, and it performed averaging pooling operations for each in order to extract multi-scale features. Next, the multi-scale features were fed into a 512-channel 1×1 convolutional layer for parallel processing; this reduced the depth of the multi-scale feature map to one-fourth that of the original. Finally, the original feature map was fused with the compressed multi-scale feature map; a 3×3 convolution was then performed to output the classification result map.

3.3.3. U-Net

The U-Net network model was proposed by Ronneberger et al. in 2015. It is an improved end-to-end network model structure based on the FCN framework, follows the encoder–decoder structure, and exploits the characteristics of jump networks; this allows it to fuse high-level semantic information with shallow-level features, making full use of contextual and detailed information to obtain a more accurate feature map [27]. The model primarily consists of two components: feature extraction (i.e., down-sampling) and an up-sampling component. These form a symmetric U-shaped structure, and the model is therefore called the U-Net network.

The details of the U-Net model structure could be seen in ref [27]. The left-hand side is the feature extraction component; we inputted here the UAV image sample set and extracted feature information via convolution and pooling operations. The right-hand side shows the up-sampling component, where the network restores the feature map of the image to the original image dimension via up-sampling. After each up-sampling step, the feature maps of the same dimension (previously down-sampled) are fused, and the missing edge information generated by the pooling operation is continuously recovered. The U-Net feature extraction network in this study consisted of five down-sampling blocks, each comprising two 3×3 convolutional layers and one 2×2 pooling layer. The up-sampling network also included five up-sampling blocks with identical structures. Each up-sampling block consisted of two 3×3 convolutional layers and one 2×2 deconvolutional layer; each up-sampling doubled the image size and halved the number of channels.

3.4. Accuracy Assessment Metrics and Strategies

To quantitatively examine the extraction results and evaluate the advantages and disadvantages of each classification method, we used confusion matrices and the overall accuracy (OA) score to evaluate the classification results. The confusion matrix calculates and compares the location and classification of each surface real image element with the corresponding location and classification image element in the classified image. When

calculating the confusion matrix, different classification accuracy metrics were calculated further, including the OA, kappa coefficient, user accuracy (UA), and producer accuracy (PA).

The OA reflects the relationship between the number of correctly classified elements and the total number of elements in the classification results; it was used to evaluate the precision of the classification results from a global perspective. The kappa coefficient can be used to evaluate the consistency and credibility of multi-classification results for remote sensing images. UA refers to the ratio between the total number of pixels correctly classified into a certain class and the total number of pixels classified into that class by the classifier (i.e., the ratio between classification results and user-defined objects). PA expresses the ratio between the number of image elements in the whole image correctly classified into a certain class by the classifier and the total number of true references of that class. The indicators were calculated as follows:

$$OA = \frac{\sum_{i=1}^r X_{ii}}{N_{all}} \times 100\%, \quad (1)$$

$$kappa = \frac{N \sum_{i=1}^r X_{ii} - \sum_{i=1}^r (X_{i+} X_{+i})}{N^2 - \sum_{i=1}^r (X_{i+} X_{+i})}, \quad (2)$$

$$UA = \frac{X_{ii}}{X_{+i}} \times 100\%, \quad (3)$$

$$PA = \frac{X_{ii}}{X_{i+}} \times 100\%, \quad (4)$$

where X_{ii} is the number of correctly classified samples, X_{i+} is the number of real samples in category i , X_{+i} is the number of samples predicted to be in category i , N_{all} is the total number of samples, and r is the total number of categories.

The size of the whole-view test image was large, and it was difficult to evaluate it pixel-by-pixel; therefore, we adopted a random method for accuracy evaluation. One thousand sample points were randomly selected in a 10,000 m² test area actually surveyed via UAV orthophotos; these were taken as a reference, and the feature types of the sample points were visually inspected and recorded as image interpretation accuracy verification samples. Within the 1000 sample points, the percentages of feature types known by visual identification were as follows: PV1 (25%); PV2 (4%); NPV1 (9%); NPV2 (4%); BS1 (50%); and BS2 (8%), which could be representative of the area.

4. Results

4.1. Comparison of Semantic Segmentation Network Results

Using a semi-automatic annotation method combining object-oriented techniques and manual correction (Section 3.2), the vegetation estimation results of the UAV images were obtained; samples are shown in Figure 5. It can be seen that DeepLabV3+, PSPNet, and U-Net all performed well in terms of estimation results.

An accuracy comparison of deep learning semantic segmentations using DeepLabV3+, PSPNet, and U-Net is shown in Table 2. From the quantitative accuracy analysis, DeepLabV3+ was found to be the best among the three methods, with an OA of up to 91.9% and a kappa coefficient of 0.87 (2% and 0.03, respectively, higher than PSPnet and 3.6% and 0.06, respectively, higher than U-Net). DeepLabV3+'s unique null space convolutional pooling pyramid increases the receptive field; thus, it can more effectively utilize the feature information of vegetation and improve the model estimation performance.

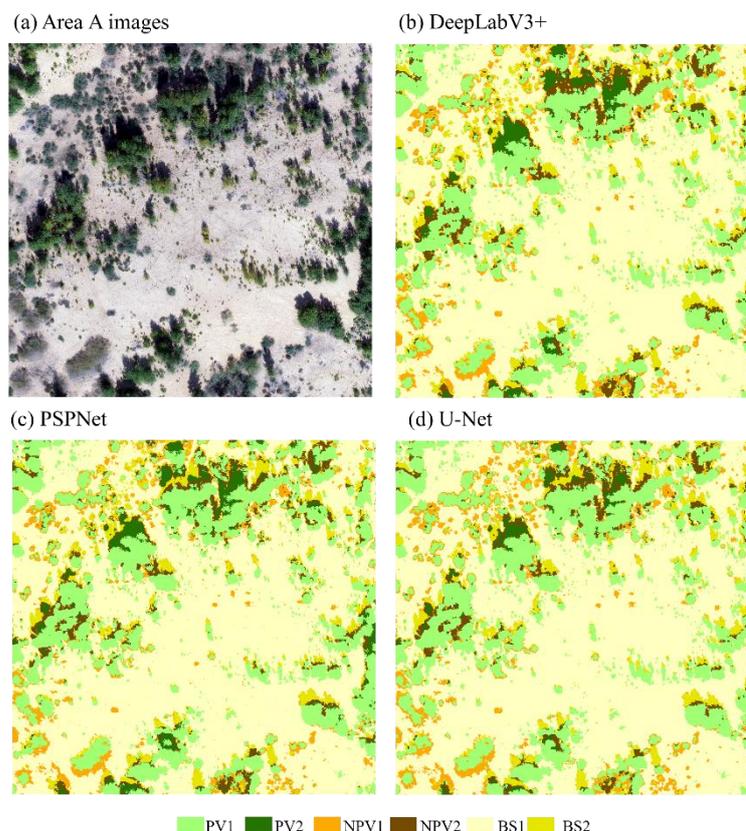


Figure 5. Region A (September) estimation results under different semantic segmentation models. (a) Original images; (b) Estimation results by DeepLabV3+; (c) Estimation results by PSPNet; (d) Estimation results by U-Net.

Table 2. Accuracy of the results under different semantic segmentation networks.

Classification Category	DeepLabV3+		PSPNet		U-Net	
	Producer Accuracy	User Accuracy	Producer Accuracy	User Accuracy	Producer Accuracy	User Accuracy
PV1	95.2	92.6	94.8	94.1	94.8	89.1
PV2	60.0	92.3	72.5	70.7	32.5	81.3
NPV1	78.3	82.3	61.4	87.9	53.0	89.8
NPV2	64.5	66.7	58.0	39.1	61.3	44.2
BS1	98.7	96.1	98.8	95.5	99.2	91.6
BS2	78.2	80.3	66.7	77.6	71.8	86.2
Overall accuracy (%)	91.9		89.9		88.3	
Kappa coefficient	0.87		0.84		0.81	

The accuracy of feature classification varies between network models. The class of features classified with the highest precision was BS1, with both PA and UA exceeding 90–95%. The next was PV1, with both PA and UA exceeding 90%. As for NPV1, the effect of each model was reduced, and DeepLabV3+ had a higher PA (78.3%) than PSPNet and U-Net. All three models achieved lower accuracies for PV2, NPV2, and BS2 estimations under shadow conditions; these may be attributable to insufficient training samples (owing to the small number of these types of features in the study area). Nevertheless, DeepLabV3+ still obtained a high PA for shaded PV2, NPV2, and BS2 estimations (60%, 64.5%, and 78.2%, respectively).

The comparison shows that the three types of semantic segmentation networks achieved relatively low estimation accuracies for PV2, NPV2, and BS2 under shadow conditions; this is attributable to the fact that fewer actual objects of this type are present in the study area and that fewer samples are available for training. In the case of a small number of training samples, DeepLabV3+ obtained higher mapping accuracies of 60%, 64.5%, and 78.2% for PV2, NPV2, and BS2 estimations under shadow conditions, respectively.

A comparison of the details of the semantic segmentation results for U-Net, PSPNet, and DeepLabV3+ deep learning is shown in Figure 6. The results showed that DeepLabV3+ outperformed PSPNet and U-Net; not only did it better retain the detailed information for various types of objects and obtained the subsumed feature edges, but it could also accurately estimate the sandy land classes (in particular, it could distinguish between PV2 and NPV2 under shading), reduce the confusion between NPV1 and NPV2, and render the overall estimation more comprehensive. The PSPNet segmentation effect was slightly inferior to DeepLabV3+, with NPV1 misclassified as PV2. U-Net performed poorly in this region, with a large number of misclassified features under shadows and poor differentiation between similar features.

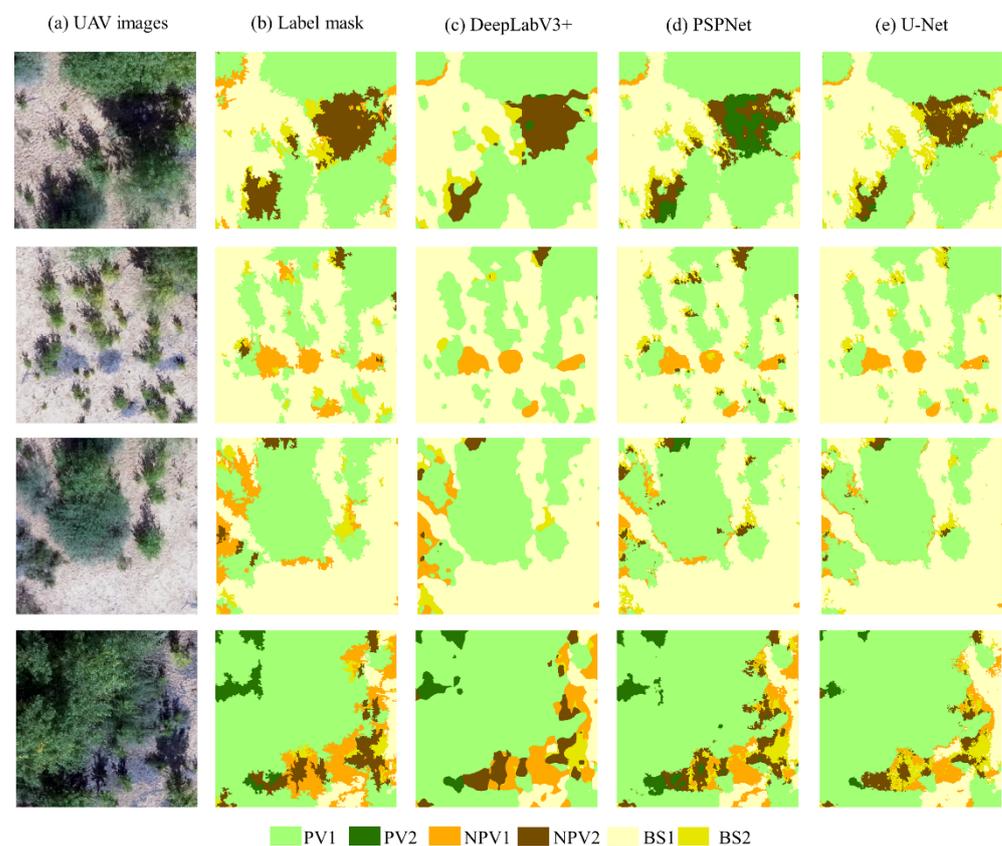


Figure 6. Estimation results under different semantic segmentation models for the test set data. (a) Original images; (b) Label masks; (c) Estimation results by DeepLabV3+; (d) Estimation results by PSPNet; (e) Estimation results by U-Net.

It can be seen that the DeepLabV3+ deep learning semantic segmentation network had the strongest applicability and practicality in the estimation of PV and NPV from UAV images. The features were well-distinguished, the objects could be estimated completely with minimal fragmentation, and most of the objects had clear edges that were generally consistent with the field situation of the study area.

4.2. Comparison of the Results Using Typical Machine Learning Methods

To further verify the advantages of DeepLabV3+, we compared its classification results (Figure 7) and accuracy against those of three machine learning classification algorithms: RF, SVM, and KNN (Table 3). Each accuracy index for DeepLabV3+ exceeded those of RF, SVM, and KNN, with an OA of up to 91.9% and a kappa coefficient of up to 0.87 (1.5% and 0.02, respectively, higher than RF; 1.6% and 0.02, respectively, higher than SVM; and 4.2% and 0.06, respectively, higher than SVM). The accuracy of RF (OA = 90.4%, kappa = 0.85) exceeded that of the SVM and KNN methods.

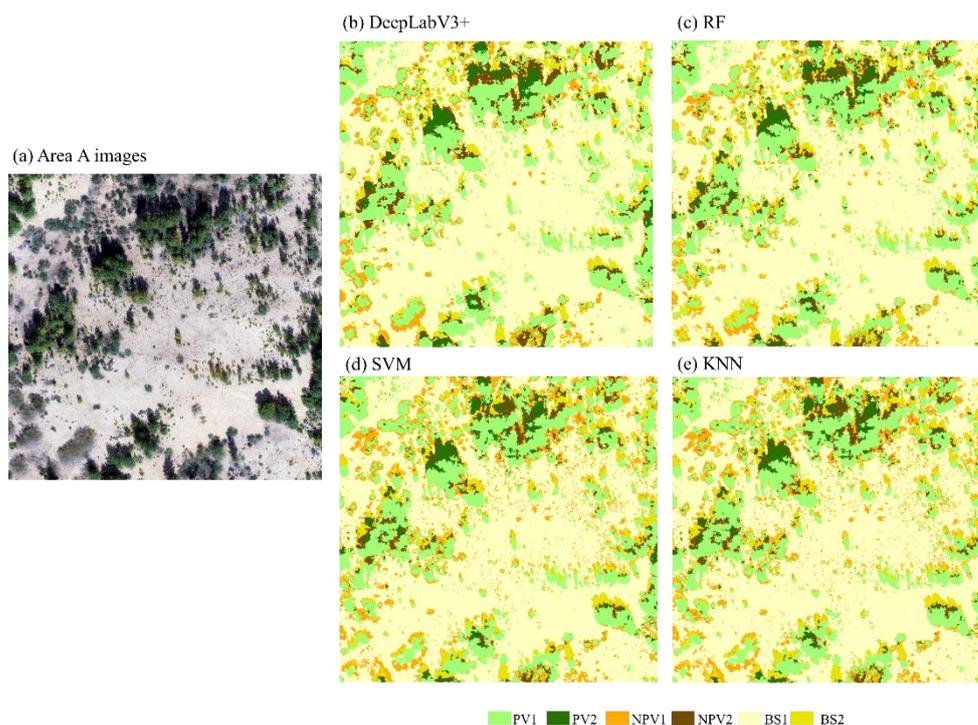


Figure 7. Identification results of the deep learning and machine learning methods. (a) Original images; (b) Estimation results by DeepLabV3+; (c) Estimation results by RF; (d) Estimation results by SVM; (e) Estimation results by KNN.

Table 3. Comparison of deep learning and machine learning methods for object classification accuracy at all locations.

Classification Category	DeepLabV3+		RF		SVM		KNN	
	Producer Accuracy	User Accuracy						
PV1	95.2	92.6	93.2	97.1	93.6	93.2	90.8	93.1
PV2	60.0	92.3	72.5	85.9	72.5	93.5	52.5	95.5
NPV1	78.3	82.3	73.5	81.3	74.7	64.5	63.9	65.4
NPV2	64.5	66.7	64.5	48.8	58.1	58.0	58.1	47.4
BS1	98.7	96.1	99.0	95.3	96.7	97.2	98.1	94.4
BS2	78.2	80.3	61.5	66.7	75.6	77.6	62.8	64.5
Overall accuracy (%)	91.9		90.4		90.3		87.7	
Kappa coefficient	0.87		0.85		0.85		0.81	

The above results further demonstrate the advantages of DeepLabV3+ for estimating PV and NPV in sandy land environments. The estimation effect of object-oriented machine learning was inferior to that of DeepLabV3+ deep learning, possibly due to its failure to take into account deeper semantic information of the target, although it considered the

spectral characteristics, texture characteristics, optical vegetation indices, and geometric characteristics of the image.

All indicators of the estimation accuracy of DeepLabV3+ exceeded those of RF, and the above results further reflect the advantages of the semantic segmentation network in the estimation of sandy PV and NPV. The estimation efficiency of object-oriented machine learning was inferior to that of deep learning DeepLabV3+ because it failed to consider the deeper semantic information of the target, although it considered the spectral features, texture features, optical vegetation indices, and geometric features of the image.

The accuracy of feature classification also varies between methods. The most accurately classified feature class was bare soil (BS1), with both PA and UA exceeding 94–95% for each classifier. This was followed by photosynthetic vegetation (PV1), with both PA and UA exceeding 90% for each classifier; meanwhile, for unshaded non-photosynthetic vegetation (NPV1), each network model was generally effective, although DeepLabV3+ had a higher PA (78.3%) than RF, SVM, and KNN.

The comparison shows that DeepLabV3+, RF, SVM, and KNN all achieve lower estimation accuracies for PV2, NPV2, and BS2 under shadow conditions; this is due to the fact that there are fewer actual objects of this type in the study area and therefore fewer samples available for training. In the case of a small number of training samples, DeepLabV3+ obtained a high mapping accuracy for the estimation of PV2, NPV2, and BS2 under shadow conditions, comparable to the RF classification results.

A detailed comparison of the semantic segmentation results for DeepLabV3+, RF, SVM, and KNN is shown in Figure 8. The results show that DeepLabV3+ outperforms RF, SVM, and KNN; it not only better retains the detailed information for various types of object and obtains the subsumed feature edges, but also accurately estimates the sandy land classes (in particular, it can distinguish between PV2 and NPV2 under shading conditions), reduces the confusion between NPV1 and NPV2, and renders the overall estimation more comprehensive. The RF and SVM segmentation results are slightly lower than those of KNN; all methods perform poorly in this region, with a large number of misclassified objects under shadow conditions and poor differentiation between similar objects.

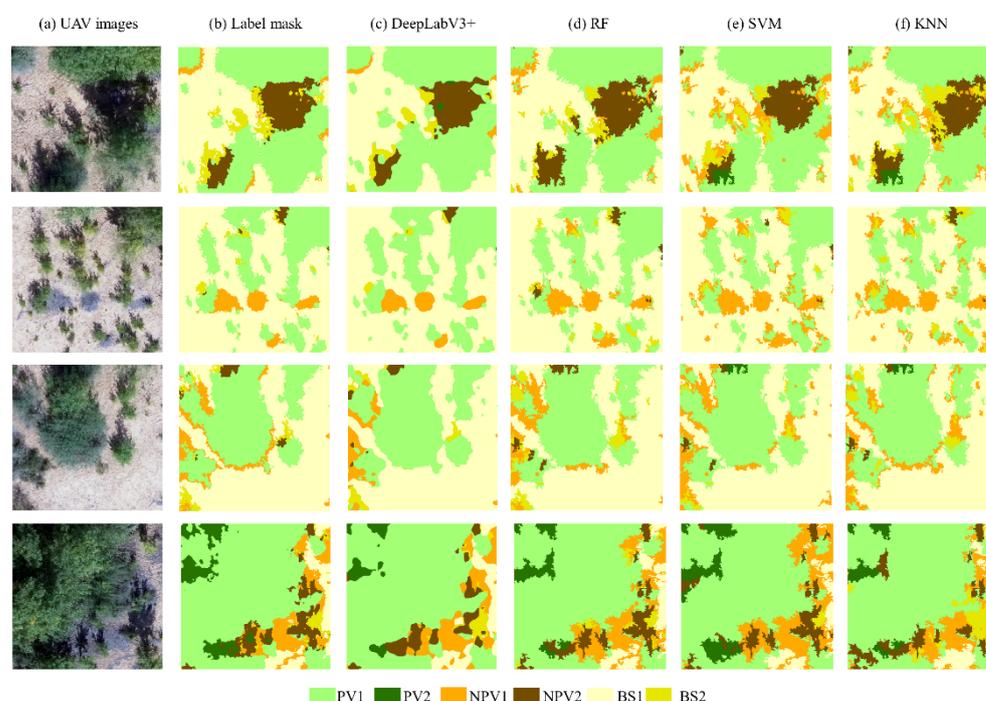


Figure 8. Comparison of detailed estimation results between deep learning and machine learning methods. (a) Original images; (b) Label masks; (c) Estimation results by DeepLabV3+; (d) Estimation results by RF; (e) Estimation results by SVM; (f) Estimation results by KNN.

It can be seen that the DeepLabV3+ network has the strongest applicability and the optimal practical estimation effect for PV and NPV. The features were well-distinguished, the objects could be estimated completely with less fragmentation, most of the objects had clear edges, and the overall effect was more consistent with the field situation in the study area. The unsatisfactory accuracy of the model in certain cases was attributable to the lack of relevant samples; in these cases, the model failed to learn enough semantic information.

The total classification accuracies of PV, NPV, and BS obtained by different deep learning semantic segmentation methods were compared (Table 4). Compared with the PSPNet and U-Net classification methods, the OAs for DeepLabV3+ classification were improved by 2.6% and 2.2%, respectively, and the kappa coefficients were improved by 0.05 and 0.05, respectively. The PAs of PV, NPV, and BS feature classification reached 93.8%, 78.1%, and 97.6%, respectively, and the UAs reached 96.1%, 81.7%, and 95.7%, respectively.

Table 4. Classification accuracy of PV, NPV, and BS under different classification algorithms.

Classification Category	DeepLabV3+		PSPNet		U-Net		RF		SVM		KNN	
	Producer Accuracy	User Accuracy										
PV	93.8	96.1	93.8	92.9	92.4	95.1	91.8	97.1	92.1	94.7	86.2	94.0
NPV	78.1	81.7	64.0	70.2	60.5	75.0	71.1	69.8	72.8	65.3	63.1	60.5
BS	97.7	95.7	95.9	94.9	97.9	93.3	96.1	93.9	94.9	95.7	94.9	92.0
Overall accuracy (%)	94.3		91.7		92.1		92.0		91.6		88.8	
Kappa coefficient	0.90		0.85		0.85		0.85		0.85		0.79	

The OAs of RF, SVM, and KNN were 92.0%, 91.6%, and 88.8%, respectively; the PAs of PV, NPV, and BS classification for RF reached 91.8%, 71%, and 96.1%, respectively; and the UAs reached 97.1%, 69.8%, and 93.9%, respectively. KNN classification was poor; for PV, NPV, and BS, its PAs were 86.3%, 63.2%, and 94.9%, respectively, and its UAs reached 94.0%, 60.5%, and 92.0%, respectively.

DeepLabV3+ obtained the highest accuracy in the PV/NPV estimation process on sandy land, higher than the other two types of semantic segmentation networks and the three types of typical machine learning algorithms.

4.3. Model Generalizability Verification

The DeepLabV3+ semantic segmentation model has been shown to be more suitable for estimating PV/NPV in desert areas. However, another advantage of the deep learning semantic segmentation method is that it can automatically and highly accurately estimate vegetation from UAV remote sensing images of different regions using the established model. In this study, we took the optimal network model, DeepLabV3+, established for the Region A comparison experiment and selected:

- i. Regions B, C, and D in September as alternative location sample areas in the same period;
- ii. Area A in July as an alternative period sample region at the same location;
- iii. Regions B, C, and D in July as both alternative location and period sample areas. With these, the classification accuracy was verified to validate the migration capabilities of the DeepLabV3+ model.

Figure 9 and Table 5 show that the DeepLabV3+ model can be seen to obtain better results in all three case class generality tests, with overall accuracies of up to 93–95% and kappa coefficients of up to 0.88–0.92. For PV and BS land classes, the DeepLabV3+ model achieved a higher estimation accuracy in all three case classes, with PAs of up to 95–99%. For the NPV land class, the DeepLabV3+ model estimated PAs of 80% in different areas but the same month; this exceeded the estimation accuracy for different locations across different periods, which achieved a PA of 72%. Therefore, the optimal network model, the DeepLabV3+ model, established using the Region A comparison experiment can very effectively estimate (i) different location sample areas in the same period, (ii) different

period sample areas at the same location, and (iii) different location sample areas across different periods. DeepLabV3+ proved to be suitable for the estimation of PV/NPV in desert areas.

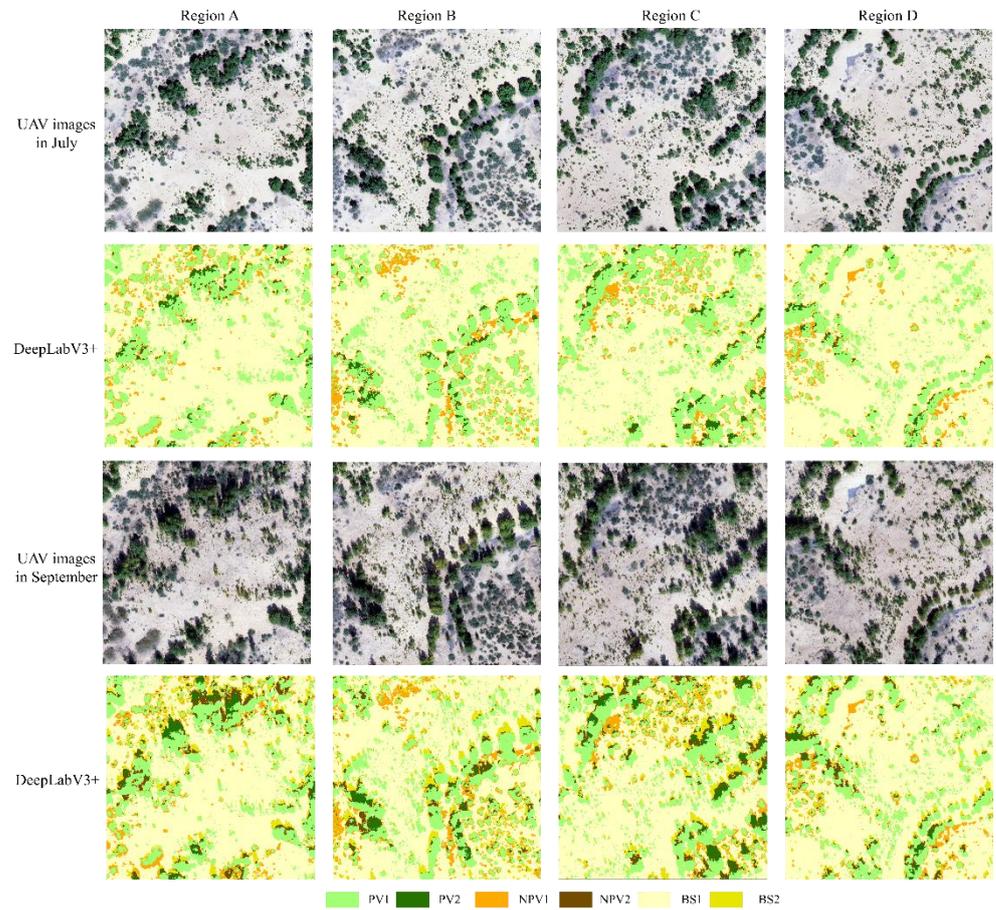


Figure 9. Comparison of PV–NPV estimation results by region for July and September, indicating model migration capability(The data from September Region A were selected to train the model, September Regions B, C, and D, and July Regions A, B, C, and D were used to verify the migration capabilities of the model).

Table 5. PV/NPV estimation accuracy of the DeepLabV3+ model migrated to September and July for each region.

Date	Type	Region A		Region B		Region C		Region D	
		Producer Accuracy	User Accuracy						
September	PV	93.8	96.1	99.06	94.9	91.4	98.6	98.7	95.8
	NPV	78.1	81.7	79.79	92.0	78.3	75.5	77.9	79.8
	BS	97.7	95.7	97.57	95.4	98.3	95.4	97.2	98.4
	Overall accuracy (%)	94.3		94.7		94.1		95.9	
	Kappa coefficient	0.9		0.91		0.89		0.92	
July	PV	92.8	96.2	93.5	96.5	90.3	97.2	94.2	96.4
	NPV	66.7	81.7	68.3	72.7	72.0	76.6	68.3	83.6
	BS	98.9	94.7	98.2	96.0	98.7	94.5	98.7	95.2
	Overall accuracy (%)	94.2		94.3		93.9		94.8	
	Kappa coefficient	0.89		0.89		0.88		0.90	

5. Discussion

5.1. Classification of Shaded Objects

Due to lighting and atmospheric scattering (amongst other reasons), large numbers of shadows are often present in UAV images [35]. Particularly for our study area, which is located in sandy land, the weather is generally sunny. It is necessary to investigate the optimal method when estimating PV–NPV coverage under shadow when light exists.

There are few UAV-based studies that discuss the effects of shadows on their classification results. De Sá et al. [43] found that shadows greatly reduced the model accuracy of detecting Acacia species under sunny conditions, while images collected under diffuse light conditions caused by clouds showed a significant improvement in classification accuracy because of the reduction in cast shadows. Moreover, Zhang et al. [13] found that cloudy conditions also reduce the separability of spectrally similar classes. However, Ishida et al. [44] proposed that incorporating shadows into the training samples improved classification performance. It is clear that no consistent conclusion was obtained about the effect of shadow on classification when using UAV images [12,36].

The above remote sensing image processes for shadows are unsuitable for high-resolution UAV image shadow processing. Shadowed areas show dark images, owing to the lack of direct solar light and low spectral values in all bands. However, the difference in the reflection of light scattered by various types of features allows the images to capture the types of features in shadow regions to some extent.

In this study, we applied object-oriented and deep learning methods to high-spatial-resolution UAV images to classify their features into six types: unshaded photosynthetic and non-photosynthetic vegetation (PV1, NPV1), shaded photosynthetic and non-photosynthetic vegetation (PV2, NPV2), and bare soil with and without shading (BS2, BS1). The results show that DeepLabV3+, PSPNet, and U-Net all have good estimation results. Not only do they retain the detailed information of all types of features and obtain clearer image feature edges, they also accurately estimate PV and NPV under shadow conditions, reducing the influence of shadows on NPV classification. The shadowed PV/NPV-BS image elements still reflect the spectral and textural characteristics specific to each feature type. Therefore, by exploiting these characteristics, PV/NPV estimation in shaded areas is achieved by adopting a direct reclassification method for shaded objects, and a high classification accuracy is obtained.

5.2. Advantages and Disadvantages of Object-Oriented Deep Learning Semantic Segmentation; Its Prospects

Numerous scholars estimate vegetation cover in study areas based on UAV images [45]. Typical previous machine learning algorithms (e.g., RF, SVM, and KNN), when combined with object orientation for vegetation classification and estimation, usually rely on complex feature parameter estimation technologies; hence, the models constructed are only suitable for a specific research area, and the obtained parameters have poor migration capabilities. Wang et al. [46] propose a new method to classify woody vegetation as well as herbaceous vegetation and calculate their FVC based on the high-resolution orthomosaic generated from UAV images by the machine learning algorithm of classification and regression tree (CART). Yue et al. [47] evaluate the use of broadband remote sensing, the triangular space method, and the random forest (RF) technique to estimate and map the FVC, CRC, and BS of cropland in which SM-CRM changes dramatically.

DeepLabV3+, PSPNet, and U-Net, the deep learning semantic segmentation algorithms adopted in this study, apply relatively mature computer vision technologies and rules. Whether deep learning methods may improve accuracy when applied to the estimation of photosynthetic and non-photosynthetic vegetation in desert areas requires further research to verify. In contrast, deep learning semantic segmentation algorithms automatically estimate the deep features most relevant to the target task by running iterative convolutions according to a loss function, without requiring human input. However, the establishment of an image segmentation model requires a large amount of labeled data for

early training; in particular, when the study area is large and the types of ground objects are complex, the manual annotation workload is too large [48]. Therefore, CNN-based semantic segmentation models will increase the technical and computational complexity to some extent [49].

The desert area PV/NPV estimation and classification proposed in this study applies the deep learning method for a small area, flat terrain, simple ground objects, and a standard computer. Using object-oriented technology and the multi-scale segmentation of UAV tag images, this study first applied the random forest classification method (combined with manual correction) to construct a preliminary tag dataset, which simplified the redundant and labor-intensive manual labeling process. The deep learning method was then used to further conduct sample training, feature extraction, and model construction. Finally, the built model was automatically estimated from PV/NPV. From the perspective of applications, the DeepLabV3+, PSPNet, and U-Net models all achieved good estimation accuracies (overall accuracy > 90%, Table 4) by using the deep learning semantic segmentation algorithm to mark a small number of images and save time. In particular, DeepLabV3+ achieved the highest overall accuracy (94.3%). Ayhan et al. also confirmed that DeepLabV3+ had a significantly higher accuracy than other methods to classify three vegetation land covers, which were tree, shrub, and grass, using only three-band-color (RGB) images [50]. Our results further developed the work carried out by Ayhan. In recent years, a series of studies have shown that deep learning methods of convolutional neural networks (CNNs) can help to efficiently capture spatial patterns of the area, enabling the extraction of a wide range of vegetation features from remotely sensed images. Yang et al. [51] proposed an efficient convolutional neural network (CNN) architecture to learn important features related to rice yield from images remotely detected at a low altitude. Egli et al. [52] proposed a novel CNN based on low-cost UAV RGB images using a tree species classification method, achieving 92% validation results based on spatially and temporally independent data.

Moreover, after testing, the model established by DeepLabV3+ also achieved good accuracy when estimating ground objects of the same type (overall accuracy > 94%, Table 5). This occurs because DeepLabV3+ introduces void convolution, which can increase the receptive field without changing the size of the feature map; as a result, each convolution output contains a wider range of information. In addition, DeepLabV3+ spatial pyramid pooling (ASPP) captures multi-scale information to optimize extraction results.

In fact, DeepLabV3+, PSPNet, and U-Net each have their own limitations in terms of classification. The classification algorithm should be selected according to the spectral features, texture features, and required accuracy given the remote sensing images, computer configuration, and time investment. Under the premise of ensuring accuracy, the classification efficiency is maximally improved, the classification process is simplified, and the classification efficiency is improved.

To summarize, this paper demonstrates that deep learning semantic segmentation combined with object-oriented learning represents a rapid method for estimating and monitoring PV/NPV vegetation. In this study, considering the computer performance, 728 image label data training models generated by object-oriented combined with manual correction are investigated, and the results basically reach the requirement of classification accuracy in the area. In later stages we will increase the number of labels to check if the accuracy of ground object estimation can be improved or not. The PV/NPV varies considerably in different seasons. Subsequent studies will identify PV/NPV by combining multi-temporal and multi-height UAV image data to provide technical support and a theoretical basis for the application of UAV remote sensing technology in the research on and protection of desert areas. The focus of this study was to verify the applicability of the model. In the future, we will expand the scope of the research to study different types of ground objects and to consider forest, grassland, farmland, and other types of land.

6. Conclusions

This study adopts aerial UAV photography as the data source, considers the classification of ground objects under shadows, and introduces object-oriented classification (supplemented with manual correction) to make data labels during the early stages of classification; this effectively simplifies the labeling process and improves the feasibility of deep learning semantic segmentation for desert region class estimation tasks. Furthermore, the feasibility and efficiency of combining three different deep learning semantic segmentation models (U-Net, PSPNet, and DeepLabV3+) with object-oriented technology for the estimation of PV/NPV coverage are discussed. The main results are as follows:

1. The application of deep learning semantic segmentation models combined with object-oriented techniques simplifies the PV–NPV estimation process of UAV visible images without reducing the classification accuracy.
2. The accuracy of the DeepLabV3+ model is higher than that of the U-Net and PSPNet models.
3. The estimation experiments for different time periods of the same ground class confirm that this method has better generalization ability.
4. Compared with three typical machine learning methods, RF, SVM, and KNN, the DeepLabV3+ method can achieve accurate and fast automatic PV–NPV estimation.

In summary, deep learning semantic segmentation is a cost-effective scheme suitable for vegetation systems in desert or similar areas; we hope that this study can provide a useful technical reference for land-use planning and land resource management.

Author Contributions: Conceptualization, J.H. and X.Z.; methodology, J.H. and D.L.; software, J.H. and L.H.; validation, J.H. and Y.Z.; formal analysis, J.H. and X.Z.; investigation, J.H. and X.X.; resources, J.H.; data curation, H.Y. and Q.T.; writing—original draft preparation, J.H.; writing—review and editing, J.H. and X.Z.; visualization, X.Z. and B.L.; supervision, J.H.; project administration, X.Z. and B.L.; funding acquisition, X.Z. and B.L. All authors have read and agreed to the published version of the manuscript.

Funding: This research was jointly supported by the National Key R&D Program of China (grant no. 2022YFE0115300) and the National Natural Science Foundation of China (grant no. 41877083).

Acknowledgments: We would like to thank the reviewers and editors for their valuable comments and suggestions. We are very grateful to Professor Jose Alfonso Gomez from the Institute for Sustainable Agriculture (CSIC), Professor Josef Krasa, Professor Tomas Dostal, and Professor Tomas Laburda from the Czech Technical University for their help in correcting and improving the language in the article. We are very grateful to Jinweiguo, Fan Xue, Tao Huang, Qi Cao, Jufeng Wang, Yadong Zou, Haojia Wang, Miaoqian Wang, Zefeng An, Yichen Wang, Kaiyang Yu, and Weinan Sun from the Northwest Agriculture and Forestry University, Yangling, China, for their help in the process of data collection and article modification.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Schimel, D.S. Terrestrial biogeochemical cycles: Global estimates with remote sensing. *Remote Sens. Environ.* **1995**, *51*, 49–56. [[CrossRef](#)]
2. Feng, Q.; Zhao, W.; Ding, J.; Fang, X.; Zhang, X. Estimation of the cover and management factor based on stratified coverage and remote sensing indices: A case study in the Loess Plateau of China. *J. Soils Sediments* **2018**, *18*, 775–790. [[CrossRef](#)]
3. Kim, Y.; Yang, Z.; Cohen, W.B.; Pflugmacher, D.; Lauer, C.L.; Vankat, J.L. Distinguishing between live and dead standing tree biomass on the North Rim of Grand Canyon National Park, USA using small-footprint lidar data. *Remote Sens. Environ.* **2009**, *113*, 2499–2510. [[CrossRef](#)]
4. Newnham, G.J.; Verbesselt, J.; Grant, I.F.; Anderson, S.A. Relative Greenness Index for assessing curing of grassland fuel. *Remote Sens. Environ.* **2011**, *115*, 1456–1463. [[CrossRef](#)]
5. Ouyang, D.; Yuan, N.; Sheu, L.; Lau, G.; Chen, C.; Lai, C.J. Community Health Education at Student-Run Clinics Leads to Sustained Improvement in Patients' Hepatitis B Knowledge. *J. Community Health* **2012**, *38*, 471–479. [[CrossRef](#)]
6. Guan, K.; Wood, E.; Caylor, K. Multi-sensor derivation of regional vegetation fractional cover in Africa. *Remote Sens. Environ.* **2012**, *124*, 653–665. [[CrossRef](#)]

7. Okin, G.S.; Clarke, K.D.; Lewis, M.M. Comparison of methods for estimation of absolute vegetation and soil fractional cover using MODIS normalized BRDF-adjusted reflectance data. *Remote Sens. Environ.* **2013**, *130*, 266–279. [[CrossRef](#)]
8. Yao, H.; Qin, R.; Chen, X. Unmanned Aerial Vehicle for Remote Sensing Applications—A Review. *Remote Sens.* **2019**, *11*, 1443. [[CrossRef](#)]
9. Vivar-Vivar, E.D.; Pompa-García, M.; Martínez-Rivas, J.A.; Mora-Tembre, L.A. UAV-Based Characterization of Tree-Attributes and Multispectral Indices in an Uneven-Aged Mixed Conifer-Broadleaf Forest. *Remote Sens.* **2022**, *14*, 2775. [[CrossRef](#)]
10. Zheng, C.; Abd-Elrahman, A.; Whitaker, V.; Dalid, C. Prediction of Strawberry Dry Biomass from UAV Multispectral Imagery Using Multiple Machine Learning Methods. *Remote Sens.* **2022**, *14*, 4511. [[CrossRef](#)]
11. Fu, B.; Liu, M.; He, H.; Lan, F.; He, X.; Liu, L.; Huang, L.; Fan, D.; Zhao, M.; Jia, Z. Comparison of optimized object-based RF-DT algorithm and SegNet algorithm for classifying Karst wetland vegetation communities using ultra-high spatial resolution UAV data. *Int. J. Appl. Earth Obs. Geoinf. ITC J.* **2021**, *104*, 102553. [[CrossRef](#)]
12. Xie, L.; Meng, X.; Zhao, X.; Fu, L.; Sharma, R.P.; Sun, H. Estimating Fractional Vegetation Cover Changes in Desert Regions Using RGB Data. *Remote Sens.* **2022**, *14*, 3833. [[CrossRef](#)]
13. Zhang, Y.; Ta, N.; Guo, S.; Chen, Q.; Zhao, L.; Li, F.; Chang, Q. Combining Spectral and Textural Information from UAV RGB Images for Leaf Area Index Monitoring in Kiwifruit Orchard. *Remote Sens.* **2022**, *14*, 1063. [[CrossRef](#)]
14. Sa, I.; Popović, M.; Khanna, R.; Chen, Z.; Lottes, P.; Liebisch, F.; Nieto, J.; Stachniss, C.; Walter, A.; Siegwart, R. WeedMap: A Large-Scale Semantic Weed Mapping Framework Using Aerial Multispectral Imaging and Deep Neural Network for Precision Farming. *Remote Sens.* **2018**, *10*, 1423. [[CrossRef](#)]
15. Wan, L.; Zhu, J.; Du, X.; Zhang, J.; Han, X.; Zhou, W.; Li, X.; Liu, J.; Liang, F.; He, Y.; et al. A model for phenotyping crop fractional vegetation cover using imagery from unmanned aerial vehicles. *J. Exp. Bot.* **2021**, *72*, 4691–4707. [[CrossRef](#)] [[PubMed](#)]
16. Mishra, N.B.; Crews, K.A. Mapping vegetation morphology types in a dry savanna ecosystem: Integrating hierarchical object-based image analysis with Random Forest. *Int. J. Remote Sens.* **2014**, *35*, 1175–1198. [[CrossRef](#)]
17. Harris, J.; Grunsky, E. Predictive lithological mapping of Canada’s North using Random Forest classification applied to geophysical and geochemical data. *Comput. Geosci.* **2015**, *80*, 9–25. [[CrossRef](#)]
18. Petropoulos, G.P.; Kalaitzidis, C.; Vadrevu, K.P. Support vector machines and object-based classification for obtaining land-use/cover cartography from Hyperion hyperspectral imagery. *Comput. Geosci.* **2012**, *41*, 99–107. [[CrossRef](#)]
19. Weinberger, K.Q.; Saul, L.K. Distance Metric Learning for Large Margin Nearest Neighbor Classification. *J. Mach. Learn. Res.* **2009**, *10*, 207–244.
20. Feng, Q.; Liu, J.; Gong, J. UAV Remote Sensing for Urban Vegetation Mapping Using Random Forest and Texture Analysis. *Remote Sens.* **2015**, *7*, 1074–1094. [[CrossRef](#)]
21. Kattenborn, T.; Leitloff, J.; Schiefer, F.; Hinz, S. Review on Convolutional Neural Networks (CNN) in vegetation remote sensing. *ISPRS J. Photogramm. Remote Sens.* **2021**, *173*, 24–49. [[CrossRef](#)]
22. Hasan, A.S.M.M.; Sohel, F.; Diepeveen, D.; Laga, H.; Jones, M.G. A survey of deep learning techniques for weed detection from images. *Comput. Electron. Agric.* **2021**, *184*, 106067. [[CrossRef](#)]
23. Barbedo, J.G.A. Plant disease identification from individual lesions and spots using deep learning. *Biosyst. Eng.* **2019**, *180*, 96–107. [[CrossRef](#)]
24. Koirala, A.; Walsh, K.B.; Wang, Z.; McCarthy, C. Deep learning—Method overview and review of use for fruit detection and yield estimation. *Comput. Electron. Agric.* **2019**, *162*, 219–234. [[CrossRef](#)]
25. Zhong, Y.; Hu, X.; Luo, C.; Wang, X.; Zhao, J.; Zhang, L. WHU-Hi: UAV-borne hyperspectral with high spatial resolution (H2) benchmark datasets and classifier for precise crop identification based on deep convolutional neural network with CRF. *Remote Sens. Environ.* **2020**, *250*, 112012. [[CrossRef](#)]
26. Hu, F.; Xia, G.-S.; Hu, J.; Zhang, L. Transferring Deep Convolutional Neural Networks for the Scene Classification of High-Resolution Remote Sensing Imagery. *Remote Sens.* **2015**, *7*, 14680–14707. [[CrossRef](#)]
27. Ronneberger, O.; Fischer, P.; Brox, T. U-Net: Convolutional networks for biomedical image segmentation. In *Medical Image Computing and Computer-Assisted Intervention 2015*; Navab, N., Hornegger, J., Wells, W.M., Frangi, A.F., Eds.; Springer International Publishing: Cham, Switzerland, 2015; pp. 234–241. [[CrossRef](#)]
28. Badrinarayanan, V.; Kendall, A.; Cipolla, R. Segnet: A deep convolutional encoder-decoder architecture for image segmentation. *IEEE Trans. Pattern Anal. Mach. Intell.* **2017**, *39*, 2481–2495. [[CrossRef](#)]
29. Zhao, H.; Shi, J.; Qi, X.; Wang, X.; Jia, J. Pyramid scene parsing network. In Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017; pp. 2881–2890.
30. Fang, H.; Lafarge, F. Pyramid scene parsing network in 3D: Improving semantic segmentation of point clouds with multi-scale contextual information. *ISPRS J. Photogramm. Remote Sens.* **2019**, *154*, 246–258. [[CrossRef](#)]
31. Chen, L.-C.; Papandreou, G.; Kokkinos, I.; Murphy, K.; Yuille, A.L. DeepLab: Semantic Image Segmentation with Deep Convolutional Nets, Atrous Convolution, and Fully Connected CRFs. *IEEE Trans. Pattern Anal. Mach. Intell.* **2018**, *40*, 834–848. [[CrossRef](#)]
32. Liu, C.; Chen, L.-C.; Schroff, F.; Adam, H.; Hua, W.; Yuille, A.L.; Fei-Fei, L. Auto-DeepLab: Hierarchical Neural Architecture Search for Semantic Image Segmentation. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Long Beach, CA, USA, 18 June 2019; pp. 82–92. [[CrossRef](#)]

33. Fricker, G.A.; Ventura, J.D.; Wolf, J.A.; North, M.P.; Davis, F.W.; Franklin, J. A Convolutional Neural Network Classifier Identifies Tree Species in Mixed-Conifer Forest from Hyperspectral Imagery. *Remote Sens.* **2019**, *11*, 2326. [[CrossRef](#)]
34. Wagner, F.H.; Sanchez, A.; Tarabalka, Y.; Lotte, R.G.; Ferreira, M.P.; Aidar, M.P.; Gloor, E.; Phillips, O.L.; Aragao, L.E. Using the U-net convolutional network to map forest types and disturbance in the Atlantic rainforest with very high resolution images. *Remote Sens. Ecol. Conserv.* **2019**, *5*, 360–375. [[CrossRef](#)]
35. Zhou, K.; Deng, X.; Yao, X.; Tian, Y.; Cao, W.; Zhu, Y.; Ustin, S.L.; Cheng, T. Assessing the Spectral Properties of Sunlit and Shaded Components in Rice Canopies with Near-Ground Imaging Spectroscopy Data. *Sensors* **2017**, *17*, 578. [[CrossRef](#)] [[PubMed](#)]
36. Ji, C.C.; Jia, Y.H.; Li, X.S.; Wang, J.Y. Research on linear and nonlinear spectral mixture models for estimating vegetation fractional cover of nitraria bushes. *Natl. Remote Sens. Bull.* **2016**, *20*, 1402–1412. [[CrossRef](#)]
37. Guo, J.H.; Tian, Q.J.; Wu, Y.Z. Study on Multispectral Detecting Shadow Areas and A Theoretical Model of Removing Shadows from Remote Sensing Images. *Natl. Remote Sens. Bulletin.* **2006**, 151–159. [[CrossRef](#)]
38. Jiao, J.N.; Shi, J.; Tian, Q.J.; Gao, L.; Xu, N.X. Research on multispectral-image-based NDVI shadow-effect-eliminating model. *Natl. Remote Sens. Bulletin.* **2020**, *24*, 53–66. [[CrossRef](#)]
39. Lin, M.; Hou, L.; Qi, Z.; Wan, L. Impacts of climate change and human activities on vegetation NDVI in China's Mu Us Sandy Land during 2000–2019. *Ecol. Indic.* **2022**, *142*, 109164. [[CrossRef](#)]
40. Zheng, Y.; Dong, L.; Xia, Q.; Liang, C.; Wang, L.; Shao, Y. Effects of revegetation on climate in the Mu Us Sandy Land of China. *Sci. Total Environ.* **2020**, *739*, 139958. [[CrossRef](#)]
41. Lü, D.; Liu, B.Y.; He, L.; Zhang, X.P.; Cheng, Z.; He, J. Sentinel-2A data-derived estimation of photosynthetic and non-photosynthetic vegetation cover over the loess plateau. *China Environ. Sci.* **2022**, *42*, 4323–4332. [[CrossRef](#)]
42. Deng, H.; Shu, S.G.; Song, Y.Q.; Xing, F.L. Changes in the southern boundary of the distribution of drifting sand in the Mawusu Sands since the Ming Dynasty. *Chin. Sci. Bull.* **2007**, *21*, 2556–2563. [[CrossRef](#)]
43. De Sá, N.C.; Castro, P.; Carvalho, S.; Marchante, E.; López-Núñez, F.A.; Marchante, H. Mapping the Flowering of an Invasive Plant Using Unmanned Aerial Vehicles: Is There Potential for Biocontrol Monitoring? *Front. Plant Sci.* **2018**, *9*, 293. [[CrossRef](#)]
44. Ishida, T.; Kurihara, J.; Viray, F.A.; Namuco, S.B.; Paringit, E.C.; Perez, G.J.; Takahashi, Y.; Marciano, J.J. A novel approach for vegetation classification using UAV-based hyperspectral imaging. *Comput. Electron. Agric.* **2018**, *144*, 80–85. [[CrossRef](#)]
45. Yang, M.-D.; Tseng, H.-H.; Hsu, Y.-C.; Tsai, H.P. Semantic Segmentation Using Deep Learning with Vegetation Indices for Rice Lodging Identification in Multi-date UAV Visible Images. *Remote Sens.* **2020**, *12*, 633. [[CrossRef](#)]
46. Wang, H.; Han, D.; Mu, Y.; Jiang, L.; Yao, X.; Bai, Y.; Lu, Q.; Wang, F. Landscape-level vegetation classification and fractional woody and herbaceous vegetation cover estimation over the dryland ecosystems by unmanned aerial vehicle platform. *Agric. For. Meteorol.* **2019**, *278*, 107665. [[CrossRef](#)]
47. Yue, J.; Tian, Q. Estimating fractional cover of crop, crop residue, and soil in cropland using broadband remote sensing data and machine learning. *Int. J. Appl. Earth Obs. Geoinf. ITC J.* **2020**, *89*, 102089. [[CrossRef](#)]
48. He, K.; Zhang, X.; Ren, S.; Sun, J. Spatial Pyramid Pooling in Deep Convolutional Networks for Visual Recognition. *IEEE Trans. Pattern Anal. Mach. Intell.* **2015**, *37*, 1904–1916. [[CrossRef](#)]
49. Martins, V.S.; Kaleita, A.L.; Gelder, B.K. Digital mapping of structural conservation practices in the Midwest U.S. croplands: Implementation and preliminary analysis. *Sci. Total Environ.* **2021**, *772*, 145191. [[CrossRef](#)]
50. Ayhan, B.; Kwan, C. Tree, Shrub, and Grass Classification Using Only RGB Images. *Remote Sens.* **2020**, *12*, 1333. [[CrossRef](#)]
51. Yang, Q.; Shi, L.; Han, J.; Zha, Y.; Zhu, P. Deep convolutional neural networks for rice grain yield estimation at the ripening stage using UAV-based remotely sensed images. *Field Crop. Res.* **2019**, *235*, 142–153. [[CrossRef](#)]
52. Egli, S.; Höpke, M. CNN-Based Tree Species Classification Using High Resolution RGB Image Data from Automated UAV Observations. *Remote Sens.* **2020**, *12*, 3892. [[CrossRef](#)]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.