J. For. Res. (2022) 33:923–936
https://doi.org/10.1007/s11676-021-01375-z

ORIGINAL PAPER

Land cover classification in a mixed forest-grassland ecosystem using LResU-net and UAV imagery

Chong Zhang[1] · Li Zhang[1] · Bessie Y. J. Zhang[4] · Jingqian Sun[1] · Shikui Dong[2] · Xueyan Wang[1] · Yaxin Li[1] · Jian Xu[3] · Wenkai Chu[3] · Yanwei Dong[3] · Pei Wang[1]

Received: 17 April 2021 / Accepted: 30 June 2021 / Published online: 23 July 2021

Abstract Using an unmanned aerial vehicle (UAV) paired with image semantic segmentation to classify land cover within natural vegetation can promote the development of forest and grassland field. Semantic segmentation normally excels in medical and building classification, but its usefulness in mixed forest-grassland ecosystems in semi-arid to semi-humid climates is unknown. This study proposes a new semantic segmentation network of LResU-net in which residual convolution unit (RCU) and loop convolution unit (LCU) are added to the U-net framework to classify images of different land covers generated by UAV high resolution. The selected model enhanced classification accuracy by increasing gradient mapping via RCU and modifying the size of convolution layers via LCU as well as reducing convolution kernels. To achieve this objective, a group of orthophotos were taken at an altitude of 260 m for testing in a natural forest-grassland ecosystem of Keyouqianqi, Inner Mongolia, China, and compared the results with those of three other network models (U-net, ResU-net and LU-net). The results show that both the highest kappa coefficient (0.86) and the highest overall accuracy (93.7%) resulted from LResU-net, and the value of most land covers provided by the producer's and user's accuracy generated in LResU-net exceeded 0.85. The pixel-area ratio approach was used to calculate the real areas of 10 different land covers where grasslands were 67.3%. The analysis of the effect of RCU and LCU on the model training performance indicates that the time of each epoch was shortened from U-net (358 s) to LResU-net (282 s). In addition, in order to classify areas that are not distinguishable, unclassified areas were defined and their impact on classification. LResU-net generated significantly more accurate results than the other three models and was regarded as the most appropriate approach to classify land cover in mixed forest-grassland ecosystems.

Keywords UAV images · Semantic segmentation · LResU-net · Land cover classification

Introduction

As one of the world's largest renewable natural resources, mixed forest-grassland resources directly affect the development of agriculture, forestry and other industries (Langley et al. 2001; Ma et al. 2010). According to Scurlock et al. (2002) and Dong et al. (2017b), mixed forest-grassland ecosystems are approximately 3.2 billion hectares, accounting for 40% of the total land area. Using remote sensing to classify land cover in a mixed forest-grassland ecosystem can

Project funding: The work was supported by the Fundamental Research Funds for the Central Universities (NO.2021ZY92), major program of State Administration of Forestry and Grassland "Study on the assessment technologies for ecologically restoring the degraded grasslands" (20,200,507).

The online version is available at http://www.springerlink.com

Corresponding editor: Tao Xu

✉ Li Zhang
zhang_li@bjfu.edu.cn

[1] College of Science, Beijing Forestry University, Beijing 100083, People's Republic of China

[2] College of Grassland Science, Beijing Forestry University, Beijing 100083, People's Republic of China

[3] Xing'an League Grassland Workstation, Inner Mongolia Autonomous Region 137400, People's Republic of China

[4] Mathematical and Computational Science Department, Stanford University, Stanford, CA 94305, USA

provide detailed grassland and woodland information over large areas (Fang Fanget al. 2010).

In the past decade, UAV image analysis technology has been widely applied to the identification and classification in forest and grassland resource surveys (Chen 2019). It has increasingly become an opportunity to attach high resolution cameras (Huseyin et al. 2019), LiDAR (Yang et al. 2020), thermal infrared (Crusiol et al. 2019) and hyperspectral cameras (Clark et al. 2018) on UAV to better collect field information for land classification. Christian and Christiane (2014) compared forest point cloud data collected from UAV images and airborne LiDAR and concluded that more information was captured through UAV image data. Zhang et al. (2020a) used aerial hyperspectral images to classify tree species on forest farms in China, and obtained an accuracy of 93.1%. However, hyperspectral imaging may be limited when used on grassland areas with low-level color contrast as it creates a large amount of redundant data (Grigorieva et al. 2020). In addition, wind has considerable influence on LiDAR data which leads to noise and ghost points around the detected targets (Yun et al. 2016; Xu et al. 2018). Therefore, using UAV high resolution cameras is one of the most preferred methods to classify land cover in a mixed forest-grassland ecosystem.

Traditional segmentation methods of remote sensing involve pixel-based segmentation (Bhadoria et al. 2020) object-based analysis (José et al. 2013), and random forest segmentation (Fei et al. 2015). The analysis of pixel-based segmentation aims only at the color information among pixels, ignoring the semantic information of the classified objects, giving a poor performance in multi-object classification (Zhang et al. 2020c). Numerous researchers have studied forestry classification algorithms based on a combination of object-based analysis, random forest and manual feature extraction. Ke et al. (2010) applied an object-based approach to evaluate the synergism in high spatial resolution multispectral imagery and low-posting-density LiDAR data for forest species classification. Random forest segmentation was applied to classify tree species using satellite images of temperate forests in Austria, and the overall accuracy was 82% (Immitzer et al. 2012). In practice, the above approaches require extensive manual marking which contributes to a waste of human resources for high accuracy feature extraction (Wolf and Bochum 2013; Dalponte et al. 2015).

With the development of deep learning and convolutional neural networks (CNN) (Zhang et al. 2020b; Lou et al. 2021), numerous semantic segmentation algorithms exist for automatic classification (Fu and Qu 2018; Braga et al. 2020). U-Net (Ronneberger et al. 2015), is a semantic segmentation model based on a fully convolutional network and was initially used for biomedical image segmentation (Dong et al. 2017a; Rad et al. 2020). In comparison to other deep learning networks such as fully convolutional networks(FCN) (Long et al. 2015) and Densenet (Huang et al. 2017), U-Net has the overwhelming advantage of overall accuracy using a small number of data sets (Liu et al. 2020). In this context, U-net was used to extract complex terrain features to classify hills and ridges of the Loess Plateau in China (Li et al. 2020a). Due to unsurpassed reliability and excellent segmentation quality, some researchers have applied U-net to train hyperspectral satellite images and obtain the distribution of trees in the Sahara and Sahel regions of West Africa (Brandt et al. 2020).

Numerous studies have indicated that defects occur during U-net's feature extraction process (Freudenberg et al. 2019; Cao and Zhang 2020; Li et al. 2020b). Since U-net's down-sampling depends on a stack modules of Conv-BN-ReLU (CBR), this may cause extraction scales to vary at different depths, leading to more exaggerated classification errors (Cicek et al. 2016). In an effort to correct the defects, the following improvements have been made:

(1) Replace CBR modules with RCU of ResNet (He et al. 2016). ResNet directly connects the encoder and the decoder in the sample and has a capability to prevent the loss of the encoded information within different layers (Zahangir et al. 2017). For example, a building-extraction algorithm based on ResNet's remote imaging demonstrated an outstanding performance in an urban setting (Xu et al. 2018).

(2) Add LCU to down-sampling in feature extraction. LU-net is a combination of U-net and LCU, where the number of convolution at each layer of the network was increased while the convolutional dimension was shorten. Alom et al. (2018) proposed a recurrent convolutional neural network based on U-net structure that exhibited superior performance on skin cancer segmentation tasks.

However, the above methods performed well in the form of binary classification for medical and urban building domains, but the ability to classify land cover in a complex forest and grassland ecosystem remains a major challenge. As such, in this study, applying the improved U-net model to achieve accurate land cover classification in a mixed forest-grassland ecosystem is proposed. The objectives of this study include the following: (1) To propose an LResU-net model applicable to land cover classification based on U-net framework as well as a combination of RCU and LCU; (2) To evaluate the classification accuracy of U-net, ResU-net, LU-net and LResU-net in a mixed forest-grassland ecosystem; and, (3) To calculate the actual areas of various land covers using the best model of this study.

Materials and methods

Study area

The study area (Fig. 1) is located near the green Lv Shui Animal Breeding Farm of Horqin, Xing’an League, Inner Mongolia Autonomous Region at 46°42′51″ N and 120°30′1″ E with an altitude of 230–300 m. The local climate is mid-temperate, semi-arid continental monsoon within an average annual temperature of 13 °C, rainfall of 420 mm, and humidity of 18%. The area consists of forests, grasslands and cultivated lands, and provides a variety of land covers such as natural grasslands, trees, roads, rivers, and buildings.

Field survey and acquisition of UAV image data

The field investigation, September 23rd to 28th, 2020 was near the Lv Shui Animal Breeding Farm. And involved the determination of land cover classes and UAV image data collection. The river bed has been eroded over many years and so some land cover is not identifiable. Aerial images were taken by the DJI Mavic 2 Pro drone equipped with Suha’s one-inch 20-megapixel CMOS sensor (Table 1). The flight airspace was 1210 m × 600 m at an altitude of 260 m in which the overlap of flight paths was 85% and a side overlap of 80%. A total of 798 photos was produced.

Data preprocessing

The usual way to obtain an orthophoto map is three dimensional (3D) reconstruction over the entire study area. The steps of 3D reconstruction are (Fig. 2). First, applying the structure from motion (SfM) algorithm achieved detection and matching of the feature points to obtain the sparse point clouds. Second, based on sparse point clouds, the dense point cloud was systematically acquired using

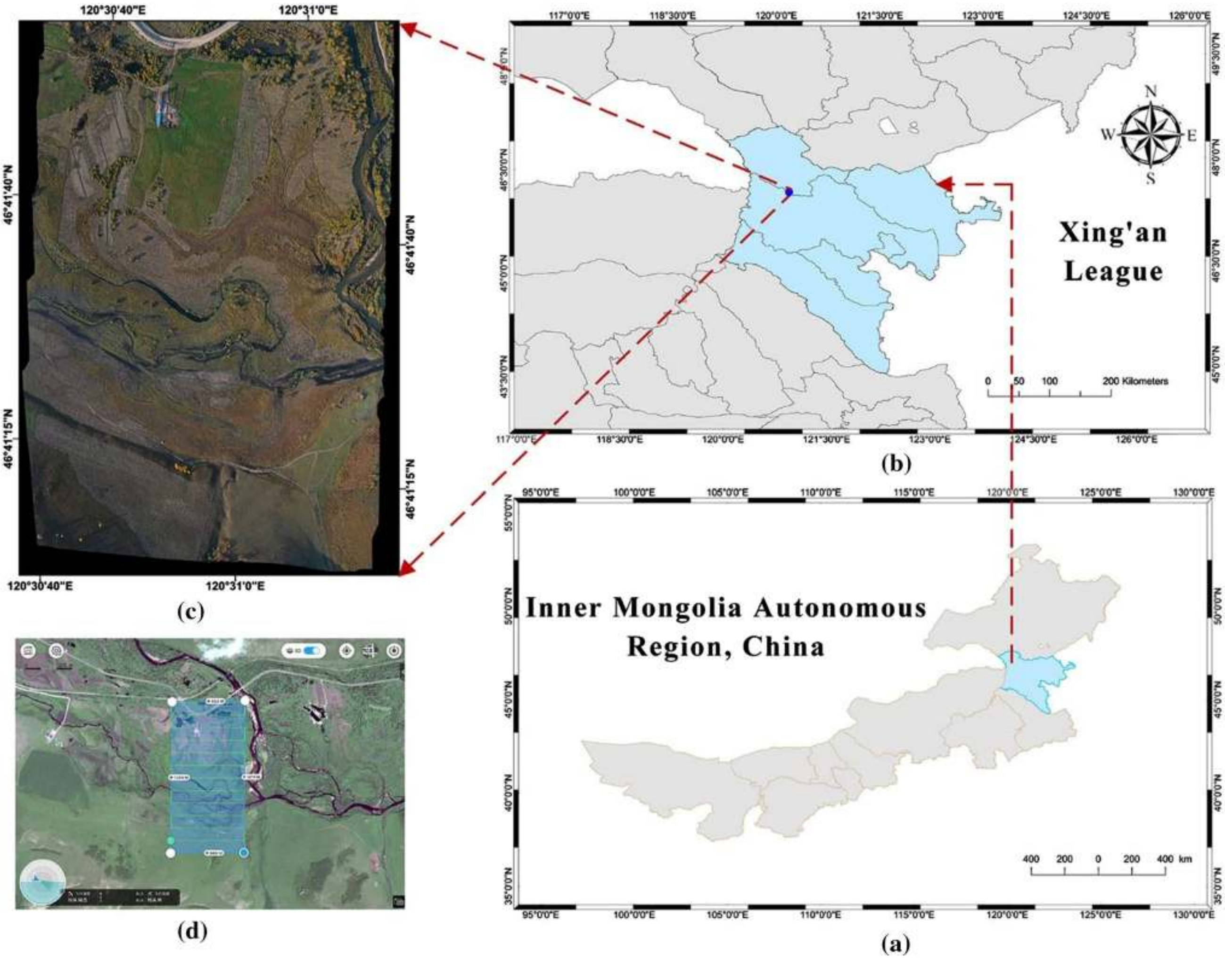

Fig. 1 **a** and **b** Study area: natural grasslands near Xing’an League, Inner Mongolia Autonomous Region; **c** synthesis orthophoto using UAV image; **d** UAV flight path generated from satellite planning route

Table 1 DJI Mavic 2 Pro UAV flight parameters

	Size	322 mm × 242 mm × 84 mm
	Takeoff weight	907 g
	Longest flight time	31-min
	Hover accuracy	V: ±0.1 m H: ±0.3 m
	Maximum flight speed	72 km h^{-1}
	Maximum cruising range	18 km
	Maximum wind resistance level	level 5 wind
	Image sensor	effective pixels 20 million

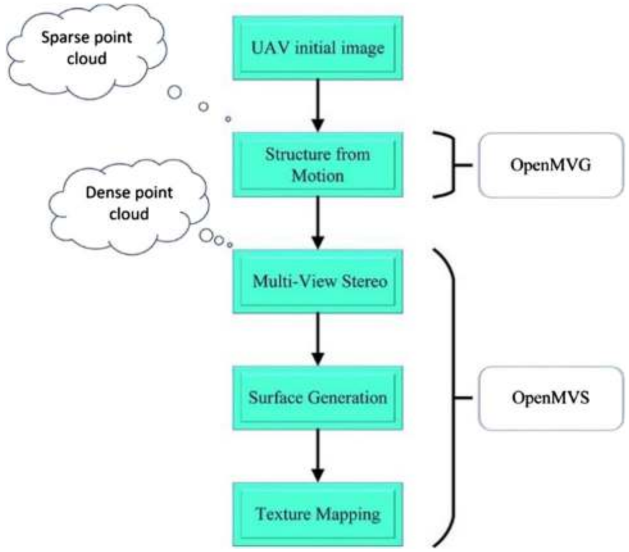

Fig. 2 3D reconstruction workflow based on OpenMVG library and OpenMVS library

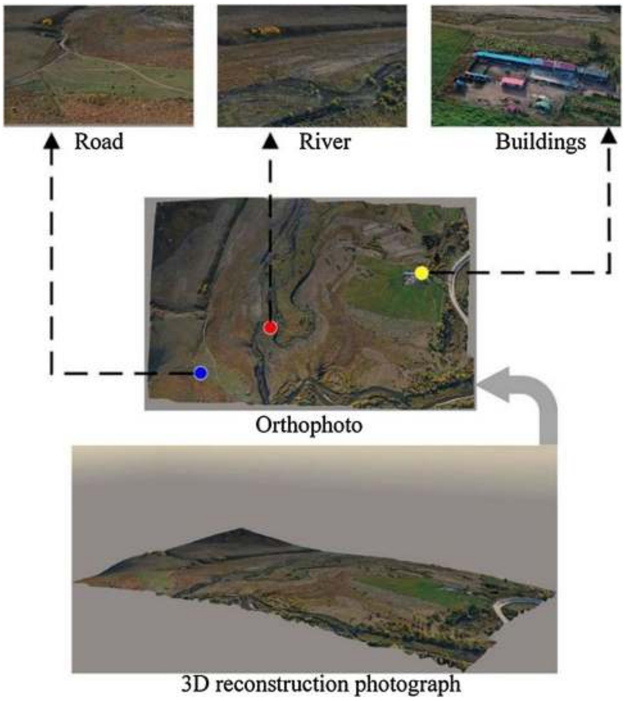

Fig. 3 An orthophoto derived from the 3D imagery using the Context Capture platform; details of roads, rivers and buildings are displayed in high resolution 3D imagery

multi-view stereo (MVS) algorithm. Third, a 3D imagery (Fig. 3) was generated in a way of surface reconstruction and texture mapping upon dense point cloud.

During the reconstruction process, open multiple view geometry (OpenMVG) library was y used for the SfM algorithm to get sparse point clouds. The subsequent procedures, including MVS, surface reconstruction and texture mapping, were implemented with the open multiple view stereo (OpenMVS) library. Finally, the 3D reconstructed model was compressed to an orthophoto on the Context Capture platform. The complete orthophoto with a pixel resolution of 20,167 × 13,534 was used to establish a high resolution data set for land cover classification.

Production of data sets

To prevent data loss, the data sets were produced by overlapping and cutting the entire orthophoto. The orthophoto with resolutions of 20,167 × 13,534 pixels, was first reshaped into an original dataset in which the image resolution was modified to 1024 × 1024 pixels. Based on the ratio of 6:2:2, the data sets were then divided into training, validation and test sets to ensure the mutual independence in data and to maintain the robustness of the model. For expending the data sets as well as reducing the performance requirements to graphics processing unit (GPU), images of 128 × 128 pixel were obtained from the original data sets by cutting with step of 64. The training, validation and test set were assigned to13145, 4598 and 4596 images (Table 2).

In the field investigation, some complex land covers were difficult to define as a category, such as the mixed landscape of swamps and the eroded lands around rivers. However, since the image grid is very large, it was difficult to label the whole image without any gaps; therefore, unclassified areas were defined as one of the label categories. According to visual interpretation, ten different categories of land cover were recognized using different colors as image classification objectives (Fig. 4). The original orthophoto and labeled image were then respectively divided into samples and objectives in training set, verification set and test set by Photoshop software.

LResU-net network

The backbone of LResU-net (Fig. 5) is a combination of sampling characteristics in ResU-net and LU-net. On the one hand, due to encoder layer of U-net being relatively shallow, LCU was added to down-sampling in the feature extraction. In comparison to U-net, LCU increases model depth as well as achieve the improvement of sample details during feature extraction. On the other hand, advances in the closed-loop feedback mapping function of RCU effectively avoided the problem of gradient overflow and disappearance, i.e., when the network's loss rate reached the lowest value, ResU-net ensures that the network of the next layer still works in the most optimal state.

At the same time, the number of convolution kernels was modified from 64 → 128 → 256 → 512 → 1024 to 32 → 64 → 128 → 256 → 512 to decrease the overall kernel count to 50% of U-net in the whole training process. According to previous studies (Liang and Hu 2015; Alom et al. 2018), when the loop step of the LCU was 3, the feature

Table 2 Land cover classification classes and related data sets

Number	Class	Label RGB	Number of images		
			Training	Validation	Test
1	River	(255,0,0)	2352	784	753
2	Road	(0,0,255)	1283	4561	4498
3	Grassland	(0,255,128)	3360	1120	1120
4	Harvested grassland	(128,128,128)	2688	896	896
5	Tree	(0,200,0)	1686	560	555
6	Building	(128,128,0)	112	64	53
7	Crop	(0,255,255)	109	51	59
8	Harvested crop	(70,255,200)	327	163	147
9	Background	(64,0,0)	1294	402	405
10	Unclassified area	(0,0,0)	321	108	100
	Total	–	13,145	4598	4596

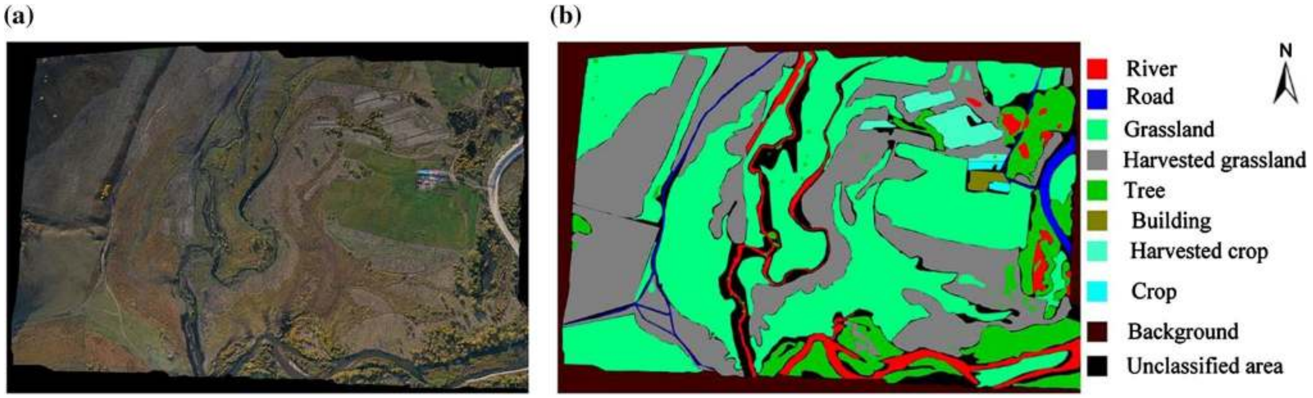

Fig. 4 **a** Map of the drone's orthophoto; **b** Map of ten different category labels

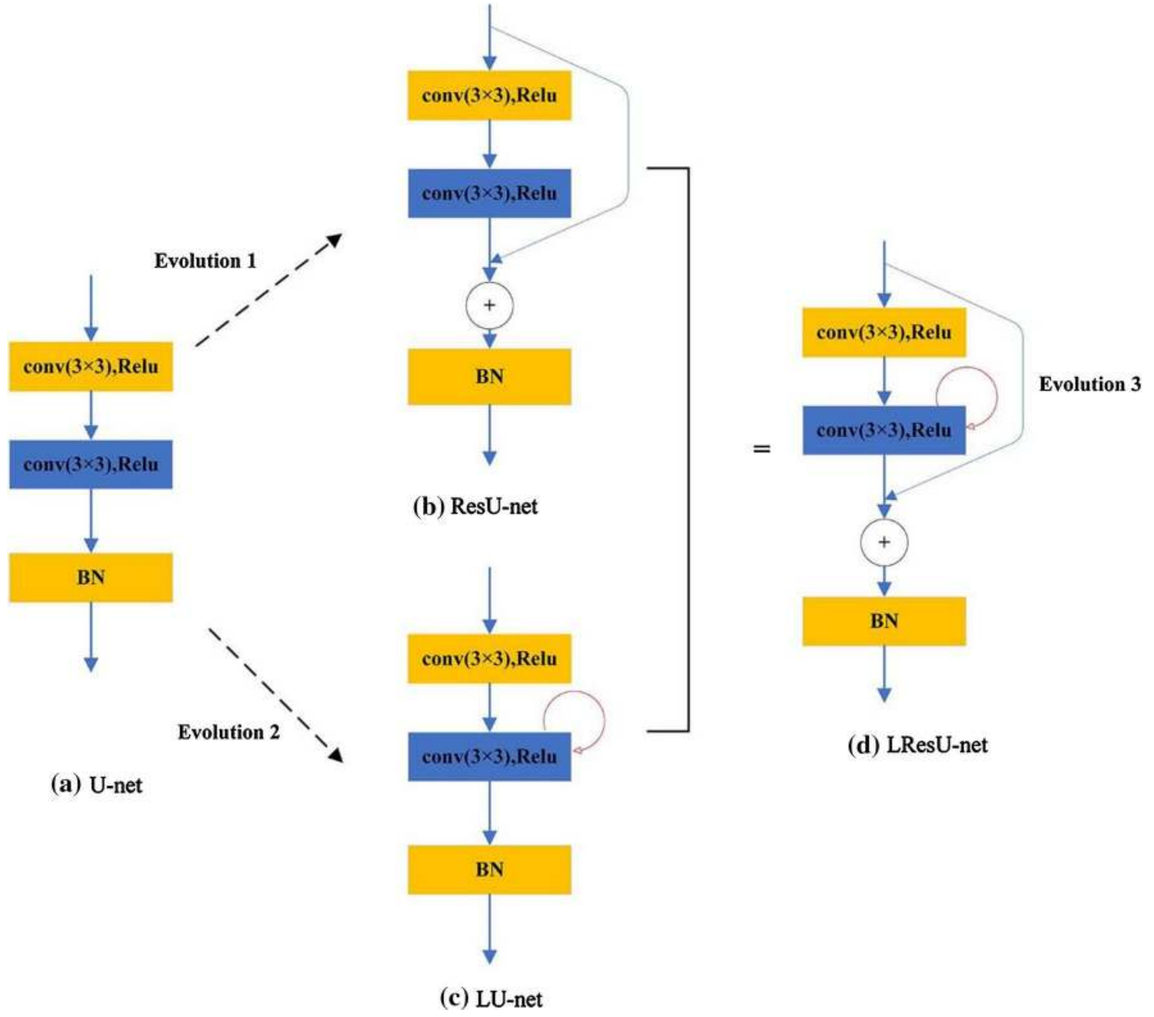

Fig. 5 **a** Backbone of the U-net feature extraction structure was a common CBR module (conv->BN->ReLU); **b** backbone of the ResU-net feature extraction structure with RCU added; **c** backbone of the LU-net feature extraction structure with LCU added; **d** backbone of the LResU-net feature extraction structure with RCU and LCU added

extraction effect and training time were optimal; the entire LResU-net structure is shown in Fig. 6.

Comparison with U-net

There are three differences of LResU-net from U-net:

1) The original feature extraction backbone CBR has been abandoned and replaced with RCU of Resnet.
2) Aiming at the feature extraction structure encoding–decoding, LCU was added to the network. Meanwhile, the number of convolutions at each layer can be changed quantitatively in accordance with the difficulty of feature extraction.
3) The total convolution kernels was shorten by half in the training progress.

Three advantages of LResU-net compared with U-Net as follows.

1) Using the modified network solved the problem of gradient overflow and gradient disappearance.
2) Reducing the rate of misjudgment of image segmentation with low color contrast by applying an improvement of accuracy in detailed feature extraction.
3) The training time was shortened in approaches of optimizing network parameters and reducing redundant convolution kernels.

Loss function and accuracy evaluation index

Loss function estimated the inconsistency between the classification data of the model and the reference data during network training progress. The pixel set and the category set,

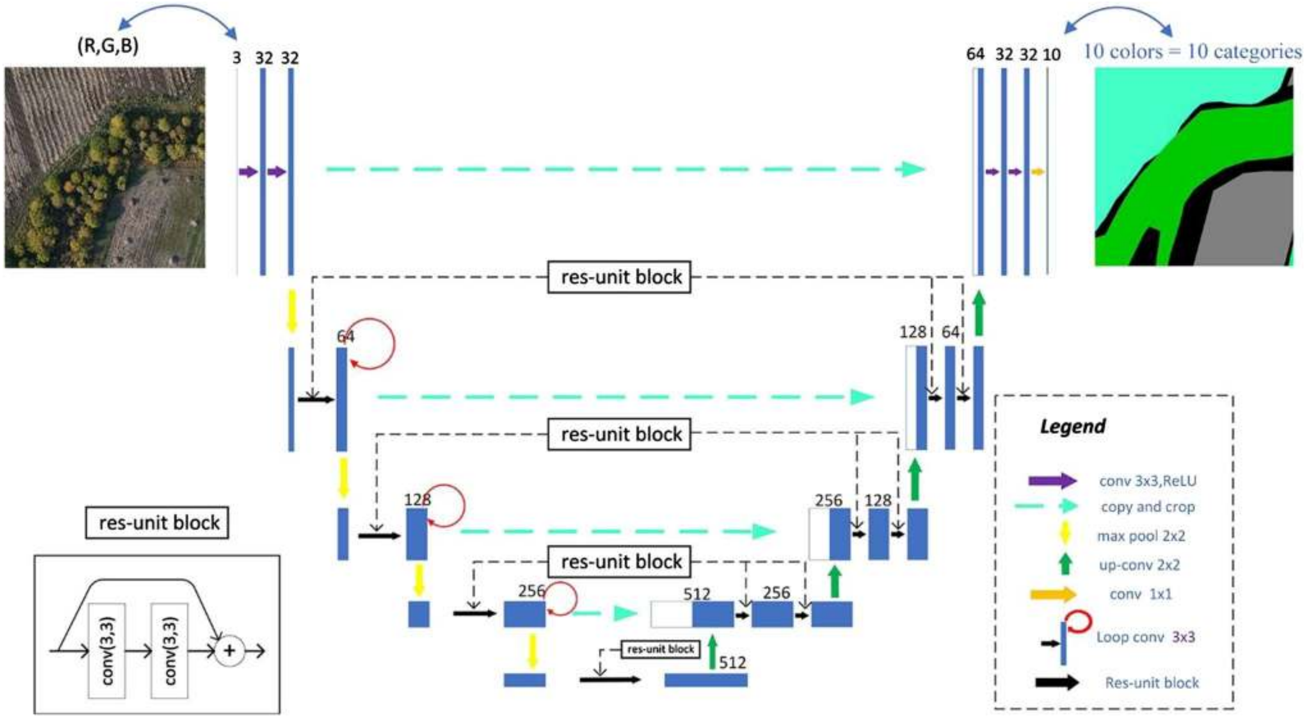

Fig. 6 LResU-net model structure with the number of convolution kernels halved

respectively, are defined as $i\{1,2, \ldots,N\}$ and $c\{1,2\ldots,M\}$, the image set becomes $y_c^i\{c = 1, 2...M\}$. After feature extraction, the probability of different categories of pixels becomes a M-dimensional tensors $p_c^i\{c = 1, 2...M\}$ and the extent of [0, 1], thereby, resulting in a multi-category, cross-entropy loss function:

$$CELoss = \frac{1}{N}\sum_{i=1}^{N}\sum_{c=1}^{M} y_c^i \log\left(p_c^i\right) \tag{1}$$

In this study, producer's and user's accuracy (Story and Congalton 1986; Olofsson et al. 2013; Shao et al. 2019), as well as a kappa coefficient, were used to evaluate classification accuracy.

Network training

To facilitate a performance comparison among different networks, four different models were used to train and predict, U-net, ResU-net, LU-net, and LResU-net. The learning rate was adjusted to 1 × e^{-4} and the batch size was 32. A total of 60 epoch with 18,000 steps allowed the model accuracy to reach the maximum in the training process. For the software platform, tensorflow-gpu 1.15 and keras 2.3.1 based on a Linux operating system was used as the learning framework, and all code was written by python. For the hardware platform-(Table 3), an Intel Xeon E5-2650 processor, a Nvidia GTX-1070 GPU, a 2 T ROM and 4 of 8 GB RAM were used to train and test.

Table 3 Computer hardware attributes

Hardware	Configuration
CPU	E5-2650
GPU	GTX1070 GPU
ROM	2 T SSD
RAM	32 GB RAM

Large-scale remote sensing imagery prediction and real area calculation

Given that memory overflow may be caused if the entire orthophoto is directly inputted to the model to predict, all images were cropped into a group of 128 × 128 image slices. After prediction of the slices, a composite imagery was spliced using these images in accordance with the order they were cropped. However, the splice approach of clipping-prediction-splicing can result in obvious segmentation edges. An alternate method of clipping overlapping images and ignoring edges (Wang et al. 2020) may mitigate this, i.e., apply the area ratio between the ignored edge image and the stitched image to calculate the overlapping area size among the image slices. Of real area calculation, the ratio of UAV real flight area and the pixel area was used to calculate the

Table 4 Classification evaluation coefficients of four models with unclassified areas

Model type	Kappa coefficient		Overall (%)	
	Training set	Test set	Training set	Test set
U-net	0.48	0.46	68.48	65.60
ResU-net	0.56	0.55	73.37	71.64
LU-net	0.54	0.53	70.83	69.18
LResU-net	0.64	0.63	80.27	80.03

Table 5 Classification evaluation coefficients of four models without unclassified area

Model type	Kappa coefficient		Overall (%)	
	Training set	Test set	Training set	Test set
U-net	0.66	0.60	88.25	82.33
ResU-net	0.80	0.79	91.96	91.03
LU-net	0.81	0.73	92.58	90.29
LResU-net	0.88	0.86	94.57	93.74

area of ten different land covers where the flight region was 70.54 ha.

Results

Accuracy of land cover classification using different networks

The reference data were derived from the pixel area of each land cover in the labeled orthophoto, and the classified data from the predicted pixel area of each land cover using different models.

Kappa coefficient and overall accuracy

The first step is to make an accurate assessment of the different models for image classification, including U-net, ResU-net, LU-net, and LResU-net. Table 4 and Table 5, respectively, show kappa coefficient and classification accuracy with and without undefined areas in the whole data sets. The undefined area has a strong impact on accuracy assessment because the unclassified area was predicted in other land covers.

On the basis of the separate data analysis in Table 4, the raise of the kappa coefficient and overall accuracy generated by ResU-net and LU-net indicates that both RCU and LCU played a positive role in modifying U-net. At the same time, the accuracy assessment of LResU-net was obviously higher, which also reflects a positive effect on the combination of RCU and LCU.

When not involving the unclassified areas, the variation tendency of Table 4 and Table 5 are consistent, i.e., the kappa coefficient and overall accuracy had an improvement to different extents on the modified model adding RCU and LCU. As expected, the optimum performance of accuracy assessment of LResU-net (kappa coefficient = 0.86, overall accuracy = 93.7%) was found in a test set, which is attributed to the advancement of ResU-net and LU-net.

Producer's and user's accuracy derived from LResU-net

The producer's and user's accuracy obtained from the LResU-net model is presented in Table 6. For most categories, both demonstrate highly favorable results. For example, trees occupy the highest value (producer's = 0.98 and user's = 0.93), and the harvested crop second (producer's = 0.94 and user's = 0.91). However, there are obvious differences between producer's and user's accuracy among some categories, including harvested grassland (producer's = 0.43 and user's = 0.99), and river (producer's = 0.85 and user's = 0.96). Such results are due to some undefined areas that were classified to above classes.

Land cover classification results from different networks

Figure 7 shows the graphs of land cover classification based on the four different network models. Compared with the results of U-net (Fig. 7a), noise and misjudgment rate of ResU-net (Fig. 7b) were slightly reduced, and the capacity of building classification significantly strengthened. Similarly, LU-net (Fig. 7c) was superior to U-net in overall classification performance. Even with some noise in the harvested grassland, the classification capacity for road, river, and building was better than that of U-net. As for the results of LResU-net (Fig. 7d), it exceeded others in classification performance, especially noise suppression from the grassland, harvested grassland, tree, and river.

The real area of various land covers

The outcome of various land covers is presented in Table 7. Regardless of unclassified area, the differences of the classification and the reference data for various land covers was insignificant. According to the results of the unclassified area, grassland (38.7%, area = 27.3 ha) is the largest proportion of the area, followed by harvested grassland (28.7%, area = 20.3 ha). The entire grassland area accounted for 67.4% of the study area. The smallest area was buildings

Table 6 Population error matrix involving producer's and user's accuracy

Classified data	Reference data											
	Ri	Ro	Ha	Bu	Gl	Tr	Hg	Bg	Cr	Ua	Total	User's accuracy
Ri	4.924	0.003	0.000	0.007	0.007	0.000	0.009	0.000	0.008	0.170	5.127	0.96
Ro	0.137	20.003	0.016	0.111	0.028	0.030	0.055	0.045	0.535	2.202	23.161	0.86
Ha	0.002	0.000	1.726	0.002	0.030	0.000	0.000	0.000	0.000	0.128	1.888	0.91
Bu	0.243	0.016	0.005	13.491	0.012	0.002	0.016	0.000	0.169	1.148	15.102	0.89
Gl	0.031	0.000	0.006	0.000	0.600	0.003	0.002	0.000	0.000	0.146	0.788	0.76
Tr	0.000	0.000	0.001	0.000	0.019	2.832	0.000	0.000	0.009	0.172	3.034	0.93
Hg	0.000	0.000	0.000	0.000	0.000	0.000	7.701	0.000	0.000	0.047	7.748	0.99
Bg	0.000	0.000	0.000	0.000	0.007	0.000	0.000	0.609	0.002	0.031	0.649	0.94
Cr	0.011	0.019	0.001	0.027	0.009	0.011	0.012	0.000	16.038	1.273	17.401	0.92
Ua	0.423	0.145	0.083	0.257	0.053	0.010	10.267	0.020	0.422	8.122	19.802	0.41
Total	5.771	20.185	1.837	13.895	0.766	2.887	18.063	0.675	17.181	13.439	94.700	–
Producer's accuracy	0.85	0.99	0.94	0.97	0.78	0.98	0.43	0.90	0.93	0.60	–	–

Ri, Ro, Ha, Bu, Gl, Tr, Hg, Bg, Cr and Ua represent, respectively: River, Road, Building, Grassland, Tree, Harvested grassland, Background, Crop and Unclassified area; the measurement unit of classified and reference data is 1.000×10^6 pixel

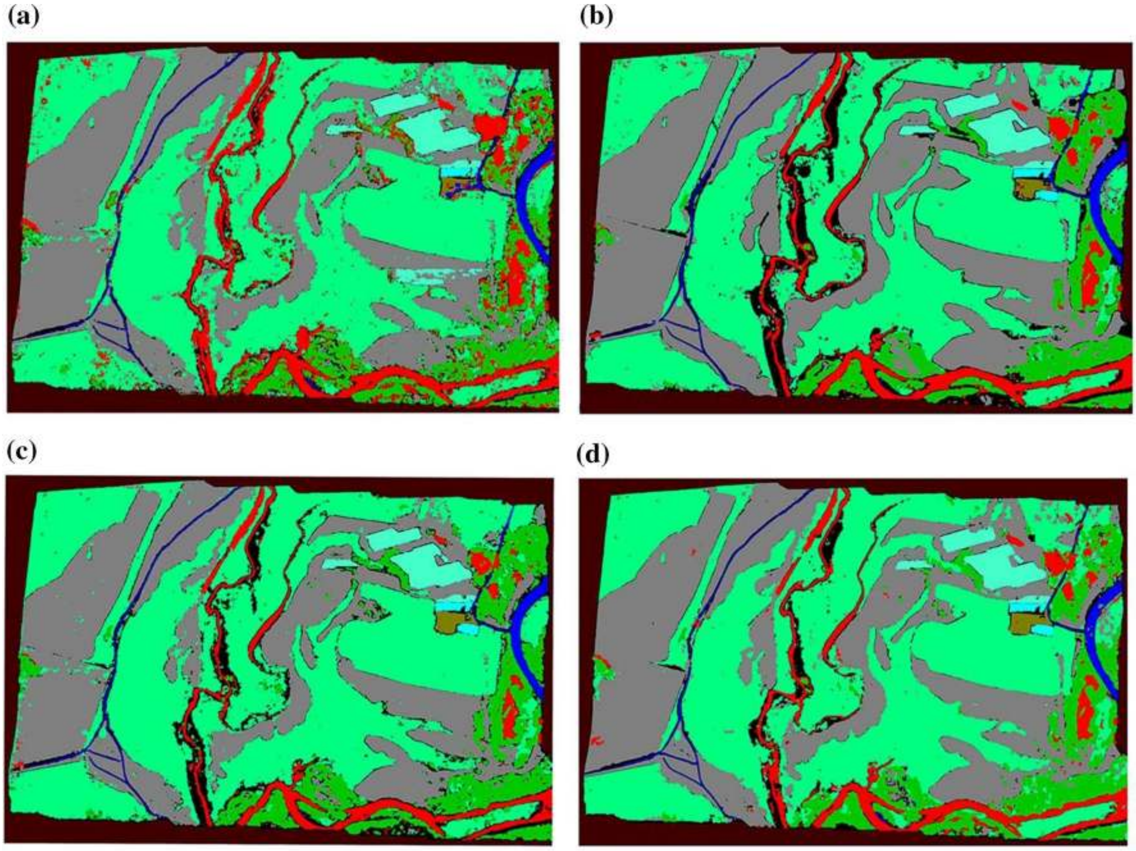

Fig. 7 **a** U-net model prediction result; **b** ResU-net model prediction result; **c** LU-net model prediction result; **d** LResU-net model prediction result

(0.2%, area = 4.8 ha). The proportion of forest area was 8.0%, which was average among all land covers.

Table 7 Areas of each land cover in the reference data and the classified data

Class	Reference data		Classified data without unclassified area		Classified data with unclassified area	
	Area (ha)	Proportion (%)	Area (ha)	Proportion (%)	Area (ha)	Proportion (%)
River	2.68	3.8	2.70	3.8	3.52	4.9
Road	0.78	1.1	0.73	1.0	0.73	1.0
Grassland	25.46	36.1	25.40	36.0	27.30	38.7
Harvested grassland	19.66	27.9	19.70	27.9	20.27	28.7
Tree	5.62	8.0	5.64	7.9	5.67	8.0
Building	0.20	0.3	0.23	0.3	0.23	0.3
Crop	0.16	0.2	0.16	0.2	0.17	0.2
Harvested crop	1.16	1.6	1.19	1.6	1.19	1.7
Background	8.11	11.5	8.16	11.5	8.16	11.5
Unclassified area	6.67	9.5	–	–	3.28	4.7
Total	70.54	100	63.93	90.6	70.54	100

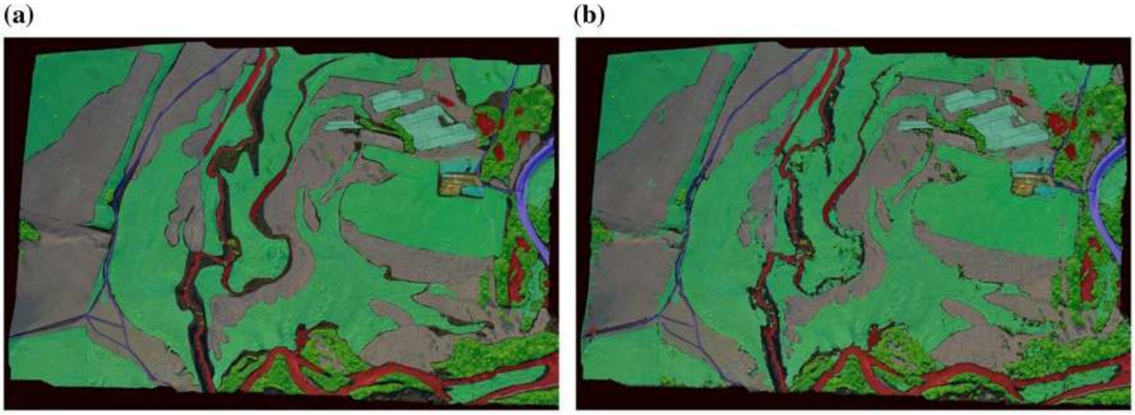

Fig. 8 **a** Orthophoto map fused with label image; **b** orthophoto map fused with LResU-net model prediction image

Discussion

Effect of unclassified areas on classification results

Unclassified areas will change the attribute of land cover. According to Fig. 8 and Table 7, about 50% of unclassified areas was likely to have the same reference data in label, which indicated that some unclassified areas are subject to distinctive features and attributes, for example, swamp. Amid other unclassified areas, the regions attached fissures were predicted to be correct land cover, but other areas were grassland, harvested grassland and forest. It is attributed to the same or similar features between the unclassified and above classes in LResU-net's vision.

Performance of RCU and LCU on the model training process

Table 8 Four different model parameters and training time in each epoch

Type	Model parameters (million)	Time in each epoch s^{-1}
U-net	31.05	358
ResU-net	73.38	512
LU-net	10.91	177
LResU-net	25.11	282

In line with the Table 8, the total parameters and training time in ResU-net were greater than that of U-net, which resulted from an addition of RCU. This was consistent with previous studies on improvements of U-net (Alom et al. 2018; Rad et al. 2020). In contrast, as the convolutional dimension decreased, LU-net significantly reduced parameters and time. As RCU and LCU are combined, the number of parameters and training time in LResU-net (25.11 million, 282 s) were slightly lower than U-net (31.05 million, 358 s).

Curves of accuracy and loss (Fig. 9) show the overall error between the predicted data and the reference data

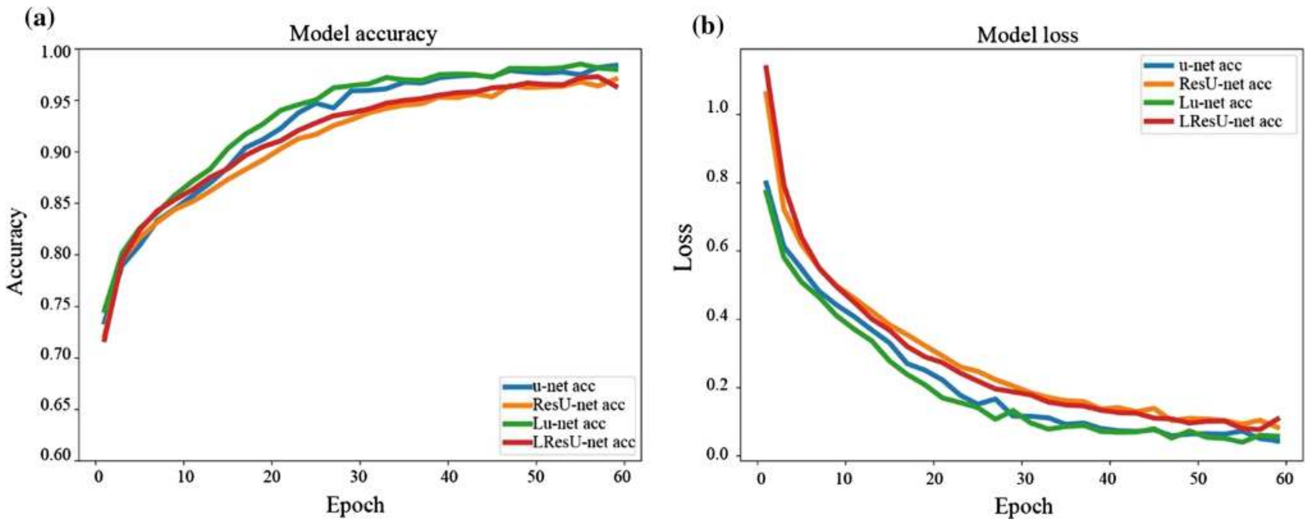

Fig. 9 **a** Training accuracy curves of four different models; **b** training loss curves of four different models

during the training process. When training is near the 55th epoch, accuracy and loss tend to be steady and hardly need the supplement of more epochs. Thus, all training stopped at the 60th epoch. In addition, it can be seen that ResU-net provided the fastest convergence rate in comparison with the other networks, which is attributed to the decline of the encoded loss in different layers.

Comparison of user's accuracy on train and test sets using different network models

Figure 10a, b shows that the user's accuracy of ResU-net and LU-net had a similar improvement in grasslands, crops, and buildings, which did not include unclassified areas. However, it is not as precise as the classification of harvested crops and river, indicating that the modification based separately on RCU or LCU still had some defects on classification in mixed forest-grassland ecosystems. At the same time, LRes-Unet produced the highest user's accuracy, proving the positive effect of the combination of RCU and LCU on classification of land cover. Figure 10, d, proves that the above statement was still valid for unclassified areas and also recognizes the influence of unclassified area on each classification based on the four models.

Effect of background area on overall accuracy

Because of a few areas, the background effects had been ignored in previous studies (Cao and Zhang 2020; Zhang et al. 2020c). In this study, the impacts of background on classification can be analyzed by the producer's and user's accuracy (Table 6). The accuracy (producer's = 0.90, user's = 0.94) of background area was higher than that of other classes, which led to a false improvement of overall accuracy. However, the results in Fig. 10 exhibit drastically different user's accuracy of background area based on U-net and ResU-net model. This difference was linked to the effect of LCU which can deepen the depth of image feature points and further improve classification accuracy.

Failure classification

Figure 11 illustrates the error of classifying trees under shadows which is the most common classification failure in data sets. Environment problems from sampling and image mosaics were the main factors deteriorating the classification performance. Under low light or shadow conditions, the image features of some land cover change and further weaken similarities with other land covers in color level. In addition, the orthophoto obtained from the 3D reconstructed model may produce a blurry edge for images (Skabek et al. 2020), which is likely to destroy classification.

Conclusions

Classifying land cover in a mixed forest-grassland ecosystem is a significant use of remote sensing technology, particularly from unmanned aerial vehicles (UAV) to manage forests and grasslands. This study presents a new method, LResU-net, to do land cover classification based on U-net, residual convolution and loop convolution network. On the basis of U-net, it adds RCU and LCU on U-net approach to improve the model and reduces the number of parameters and training time. Compared with other networks (U-net, ResU-net, LU-net), LResU-net has higher Kappa coefficients and greater accuracy in the entire data sets. The analysis of producer's and user's

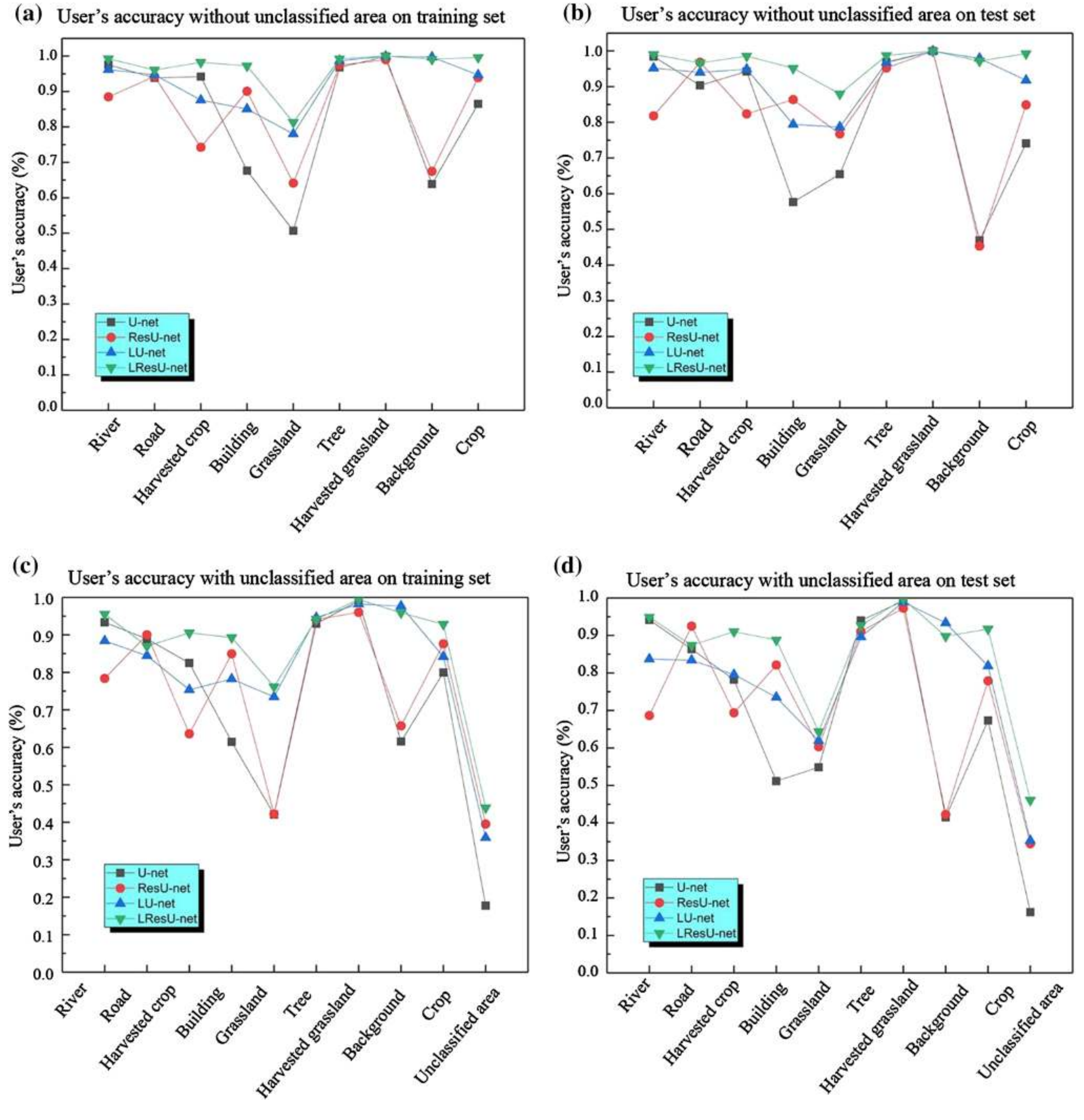

Fig. 10 **a** User's accuracy without unclassified area on training set using four network models; **b** user's accuracy without unclassified area on test set using four network models; **c** user's accuracy with unclassified area on training set using four network models; **d** user's accuracy with unclassified area on test set using four network models

accuracy indicates that LResU-net had the favorable performance in various land covers. The result of classification was affected by unclassified areas, and a solution to some unclassified lands was found. The area of various land covers, which can be used for statistics and analysis of landform was calculated. However, this study does not include height data and future research should use the 3D reconstructed model to study height data of land cover classification.

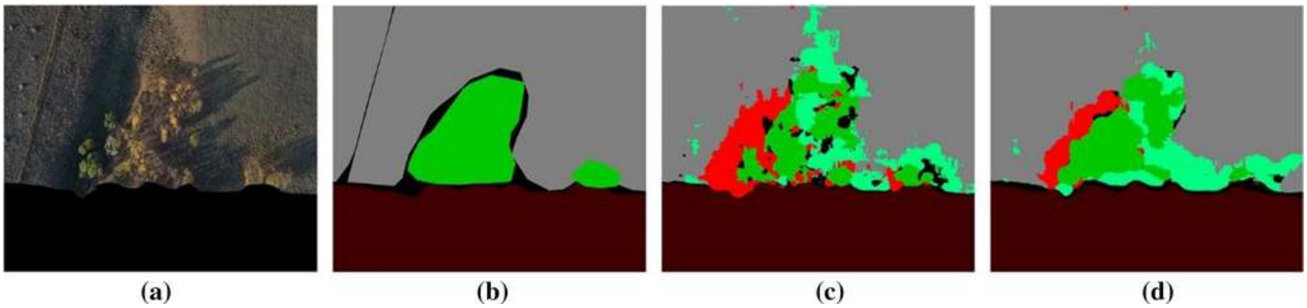

Fig. 11 **a** UAV remote sensing image; **b** label image; **c** results of U-net classification; **d** results of LResU-net classification

References

Alom MZ, Hasan M, Yakopcic C, Taha TM, Asari VK (2018) Recurrent residual convolutional neural network based on u-net (r2u-net) for medical image segmentation. Dissertation, Cornell University. https://arxiv.org/abs/1802.06955

Bhadoria P, Agrawal S, Pandey R (2020) Image segmentation techniques for remote sensing satellite images. IOP Conf Ser Mater Sci Eng 993:012050. https://doi.org/10.1088/1757-899X/993/1/012050

Braga JRG, Peripato V, Dalagnol R, Ferreira MP, Tarabalka Y, Aragão LEOC, Campos VHF, Shiguemori EH, Wagner FH (2020) Tree crown delineation algorithm based on a convolutional neural network. Remote Sens 12:1288. https://doi.org/10.3390/rs12081288

Brandt M, Tucker CJ, Kariryaa A, Rasmussen K, Abel C, Small J, Chave J, Rasmussen LV, Hiernaux P, Diouf AA, Kergoat L, Mertz O, Lgel C, Gieseke F, Schöning J, Li S, Melocik K, Meyer J, Sinno S, Romero E, Glennie E, Montagu A, Dendoncker M, Fensholt R (2020) An unexpectedly large count of trees in the West African Sahara and Sahel. Nature 587:78–82. https://doi.org/10.1038/s41586-020-2824-5

Cao KL, Zhang XL (2020) An improved ResU-net model for tree species classification using airborne high-resolution images. Remote Sens 12:1128. https://doi.org/10.3390/rs12071128

Chen X (2019) Application of UAV digital photogrammetry technology in marine topographic surveying and mapping. J Coastal Res 93:674. https://doi.org/10.2112/SI93-092.1

Christian T, Christiane S (2014) Impact of tree species on magnitude of PALSAR interferometric coherence over Siberian forest at frozen and unfrozen conditions. Remote Sens 6(2):1124–1136. https://doi.org/10.3390/rs6021124

Cicek O, Abdulkadir A, Lienkamp SS, Brox T, Ronneberger O (2016) 3D u-net: learning dense volumetric segmentation from sparse annotation. In: Ourselin S, Joskowicz L, Sabuncu M, Unal G, Wells W (eds) Medical image computing and computer-assisted intervention–MICCAI 2016. Lecture notes in computer science. Springer, Cham. https://doi.org/10.1007/978-3-319-46723-8_49

Clark ML, Buck-Diaz J, Evens J (2018) Mapping of forest alliances with simulated multi-seasonal hyperspectral satellite imagery. Remote Sens Environ 210:490–507. https://doi.org/10.1016/j.rse.2018.03.021

Crusiol LGT, Nanni MR, Furlanetto RH, Sibaldelli RNR, Cezar E, Mertz-Henning LM, Nepomuceno AL, Neumaier N, Farias JRB (2019) UAV-based thermal imaging in the assessment of water status of soybean plants. Int J Remote Sens 1(23):3243–3265. https://doi.org/10.1080/01431161.2019.1673914

Dalponte M, Ene LT, Marconcini M, Gobakken T, Næsset E (2015) Semi-supervised SVM for individual tree crown species classification. ISPRS J Photogramm Remote Sens 110:77–87. https://doi.org/10.1016/j.isprsjprs.2015.10.010

Dong H, Yang G, Liu FD, Mo YH, Guo YK (2017a) Automatic Brain Tumor Detection and Segmentation Using U-Net Based Fully Convolutional Networks. In: Valdés HM, González-Castro V (eds) Medical image understanding and analysis. MIUA 2017 Communications in computer and information science. Springer, Cham. https://doi.org/10.1007/978-3-319-60964-5_44

Dong SK, Wolf AW, Lassoie JP, Liu SL, Long RJ, Yi SL, Jasra AW, Phuntsho K (2017b) Bridging the gaps between science and policy for the sustainable management of rangeland resources in the developing world. Bioscience 67:566–663. https://doi.org/10.1093/biosci/bix042

Fang JY, Yang YH, Ma WH, Mohammat A, Shen HH (2010) Ecosystem carbon stocks and their changes in china's grasslands. Sci China Life Sci 53(007):757–765. https://doi.org/10.1007/s11427-010-4029-x

Fei XY, Wang T, Wei XL (2015) Coastal wetland classification based on multi-scale image segmentation using high spatial rs images. Remote Sens Technol Appl 30(2):298–303. https://doi.org/10.11873/j.issn.1004-0323.2015.2.0298

Freudenberg M, Nölke N, Agostini A, Urban K, Wörgötter F, Kleinn C (2019) Large scale palm tree detection in high resolution satellite images using u-net. Remote Sens 11(3):312. https://doi.org/10.3390/rs11030312

Fu XM, Qu H (2018) research on semantic segmentation of high-resolution remote sensing image based on full convolutional neural network. In: International Symposium on Antennas, Propagation and EM Theory. Hangzhou, China. https://doi.org/10.1109/ISAPE.2018.8634106

Grigorieva O, Brovkina O, Saidov A (2020) An original method for tree species classification using multitemporal multispectral and hyperspectral satellite data. Silva Fennica. 54(2):10143. https://doi.org/10.14214/sf.10143

He KM, Zhang Xy, Ren SQ, Sun J (2016) Deep residual learning for image recognition. In: IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, USA, 27–30. https://doi.org/10.1109/CVPR.2016.90

Huang G, Liu Z, LaurensV, Maaten L, Weinberger KQ (2017) Densely connected convolutional networks. In: IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, USA, 21–26. https://arxiv.org/abs/1608.06993

Huseyin Y, Akgul M, Coban S, Gulci S (2019) Determination and accuracy analysis of individual tree crown parameters using UAV based imagery and OBIA techniques. Measurement 145:651–664. https://doi.org/10.1016/j.measurement.2019.05.092

Immitzer M, Atzberger C, Koukal T (2012) Tree species classification with random forest using very high spatial resolution 8-band worldview-2 satellite data. Remote Sens 4(9):2661–2693. https://doi.org/10.3390/rs4092661

José MP, Torres-Sánchez J, Castro ALD, Kelly M, Francisca LG (2013) Weed mapping in early-season maize fields using object-based analysis of unmanned aerial vehicle (UAV) images. PLoS ONE 8(10):77151. https://doi.org/10.1371/journal.pone.0077151

Ke Y, Quackenbush LJ, Im J (2010) Synergistic use of quickbird multispectral imagery and LIDAR data for object-based forest species classification. Remote Sens Environ 114(6):1141–1154. https://doi.org/10.1016/j.rse.2010.01.002

Langley SK, Cheshire HM, Humes KS (2001) A comparison of single date and multitemporal satellite image classifications in a semi-arid grassland. J Arid Environ 49(2):401–411. https://doi.org/10.1006/jare.2000.0771

Li S, Xiong LY, Tang GA, Strobl J (2020a) Deep learning-based approach for landform classification from integrated data sources of digital elevation model and imagery. Geomorphology 354:107045. https://doi.org/10.1016/j.geomorph.2020.107045

Li XL, Song WH, Gao DZ, Gao W, Wang HZ (2020b) Training a u-net based on a random mode-coupling matrix model to recover acoustic interference striations. J Acoust Soc Am 147(4):363–369. https://doi.org/10.1121/10.0001125

Liang M, Hu XL (2015) Recurrent convolutional neural network for object recognition. IEEE Conference on Computer Vision & Pattern Recognition, Boston, USA, 7–12. https://doi.org/10.1109/CVPR.2015.7298958

Liu LL, Cheng JH, Quan Q, Wu FX, Wang YP, Wang JX (2020) A survey on u-shaped networks in medical image segmentations. Neurocomputing 409:244–258. https://doi.org/10.1016/j.neucom.2020.05.070

Long J, Shelhamer E, Darrell T (2015) Fully convolutional networks for semantic segmentation. IEEE Trans Pattern Anal Mach Intell 39(4):640–651. https://doi.org/10.1109/CVPR.2015.7298965

Lou XW, Huang YX, Fang LM, Huang SQ, Gao HL, Yang LB, Weng YH, Hung IK (2021) Measuring loblolly pine crowns with drone imagery through deep learning. J for Res. https://doi.org/10.1007/s11676-021-01328-6

Ma WH, Fang JY, Yang YH, Mohammat A (2010) Biomass carbon stocks and their changes in northern china's grasslands during 1982–2006. Sci China Life Sci 53(7):841–850. https://doi.org/10.1007/s11427-010-4020-6

Olofsson P, Foody GM, Stehman SV, Woodcock CE (2013) Making better use of accuracy data in land change studies: estimating accuracy and area and quantifying uncertainty using stratified estimation. Remote Sens Environ 129:122–131. https://doi.org/10.1016/j.rse.2012.10.031

Rad RM, Saeedi P, Au J, Havelock J (2020) Trophectoderm segmentation in human embryo images via inceptioned u-net. Med Image Anal 62:101612. https://doi.org/10.1016/j.media.2019.101612

Ronneberger O, Fischer P, Brox T (2015) convolutional networks for biomedical image segmentation. In: Navab N, Hornegger J, Wells W, Frangi A (eds) Medical image computing and computer assisted intervention MICCAI 2015. Lecture notes in computer science. Springer, Cham. https://doi.org/10.1007/978-3-319-24574-4_28

Scurlock J, Johnson K, Olson RJ (2002) Estimating net primary productivity from grassland biomass dynamics measurements. Glob Change Biol 8(8):736–753. https://doi.org/10.1046/j.1365-2486.2002.00512.x

Shao GF, Tang LN, Liao JF (2019) Overselling overall map accuracy misinforms about research reliability. Landscape Ecol 34:2487–2492. https://doi.org/10.1007/s10980-019-00916-6

Skabek K, Łabędź P, Ozimek P (2020) Improvement and unification of input images for photogrammetric reconstruction. Comput Assist Method Eng Sci 26(3–4):153–162

Story M, Congalton RG (1986) Accuracy assessment: a user's perspective. Photogramm Eng Remote Sens 52:397–399. https://doi.org/10.1016/0031-0182(86)90068-4

Wang ZQ, Zhou Y, Wang SX, Wang FT, Xu ZY (2020) House building extraction from high resolution remote sensing image based on IEU-Net. J Remote Sens 12(1):133. https://doi.org/10.11834/jrs.20200042

Wolf N, Bochum B (2013) Object features for pixel-based classi cation of urban areas comparing different machine learning algorithms. Photogrammetrie Fernerkundung Geoinformation 3:149–161. https://doi.org/10.1127/1432-8364/2013/0166

Xu YY, Wu L, Xie Z, Chen ZL (2018) Building extraction in very high resolution remote sensing imagery using deep learning and guided filters. Remote Sens 10(1):144. https://doi.org/10.3390/rs10010144

Yang JT, Kang ZZ, Cheng S, Yang Z, Akwensi PH (2020) An individual tree segmentation method based on watershed algorithm and 3d spatial distribution analysis from airborne LiDAR point clouds. IEEE J Sel Top Appl Earth Obs Remote Sens 13:1055–1067. https://doi.org/10.1109/JSTARS.2020.2979369

Yun T, Feng A, Li WZ, Sun Y, Cao L, Xue LF (2016) A novel approach for retrieving tree leaf area from ground-based lidar. Remote Sens 8(11):942. https://doi.org/10.3390/rs8110942

Zahangir AM, Mahmudul H, Chris Y, Tarek MT, Asari VK (2017) Improved inception-residual convolutional neural network for object recognition. Neural Comput Appl 32:79–293. https://doi.org/10.1007/s00521-018-3627-6

Zhang B, Zhao L, Zhang XL (2020a) Three-dimensional convolutional neural network model for tree species classification using airborne hyperspectral images. Remote Sens Environ 247:111938. https://doi.org/10.1016/j.rse.2020.111938

Zhang C, Xia K, Feng HL, Yang YH, Du XC (2020b) Tree species classification using deep learning and rgb optical images obtained by an unmanned aerial vehicle. J for Res. https://doi.org/10.1007/s11676-020-01245-0

Zhang CX, Yue P, Tapete D, Shangguan BY, Wang M, Wu ZY (2020c) A multi-level context-guided classification method with object-based convolutional neural network for land cover classification using very high resolution remote sensing images. Int J Appl Earth Obs Geoinf 88:102086. https://doi.org/10.1016/j.jag.2020.102086