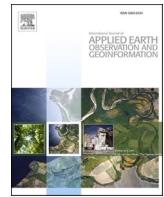




Contents lists available at ScienceDirect

International Journal of Applied Earth Observations and Geoinformation

journal homepage: www.elsevier.com/locate/jag



Automated delineation of agricultural field boundaries from Sentinel-2 images using recurrent residual U-Net



Huanxue Zhang^{a,*}, Mingxu Liu^b, Yuji Wang^a, Jiali Shang^c, Xiangliang Liu^a, Bin Li^a, Aiqi Song^d, Qiangzi Li^{e,*}

^a College of Geography and Environment, Shandong Normal University, Jinan 250300, China

^b Shandong Survey and Design Institute of Water Conservancy, Jinan 250013, China

^c Ottawa Research and Development Centre, Agriculture and Agri-Food Canada, Ottawa, ON, Canada

^d Shandong Zhengyuan Geophysical Information Technology CO. LTD, Jinan 250101, China

^e Aerospace Information Research Institute, Chinese Academy of Sciences, Beijing 100094, China

ARTICLE INFO

Keywords:

Agriculture
Automated delineation
Field boundary
Recurrent residual U-Net
Sentinel-2
Remote sensing

ABSTRACT

Delineation of agricultural fields is desirable for operational monitoring of agricultural production and is essential to support food security. Due to large within-class variance of pixel values and small inter-class difference, automated field delineation remains to be a challenging task. In this study, a strategy is proposed to effectively address this issue. Firstly, a framework was developed using the Canny operator connected with the Watershed segmentation algorithm (CW) to quickly label the training dataset, which minimizes the workload of dataset generation in comparison with the commonly used manual vectorization. Secondly, a CW-trained deep semantic segmentation network, recurrent residual U-Net, was selected to mine the low level and deep semantic features. Finally, a boundary connecting method (to integrate fragmented boundaries) was used to generate the agricultural field boundary. The proposed methods are tested over smallholder agricultural landscape in Heilongjiang province, China, using Sentinel-2 imagery. Compared with the U-Net (overall accuracy (OA) 82.18%), the residual U-Net (ResU-Net, with OA 85.78%), traditional object-based image analysis (OBIA, with OA about 82%), and the existing 10-m resolution global land cover map (FROM-GLC10), the proposed method shows an improved performance (OA 89.28%, and Kappa 0.85). The successful application of the proposed method suggests that the recurrent residual U-Net model has great universality in agricultural field boundary extraction, and the automated technique has the potential of being applied to other regions.

1. Introduction

With the global population exceeding seven billion, food security has become a global concern (Tollefson 2011). Agriculture is considered as one of the driving factors of climate change, accounting for 30–35% of the global greenhouse gas emissions. This poses a major global challenge: maintain food security while reducing the footprint of agricultural activities (Karamura et al., 2013). It requires the development of an accurate agricultural monitoring system that offers timely information of crop conditions, make accurate estimates of crop yields, and provide information for decision-making (Matton et al., 2015). The scope and location of farmland are the basic data requirements of the agricultural monitoring systems.

Remotely sensed observations can now cover almost every part of the

Earth's surface (Lefsky et al., 2002) and have been an important tool for acquiring and determining the spatial distribution of farmlands at local, regional, and global scales (Gong et al., 2019; Massey et al., 2018; Phalke and Ozdogan 2018; Xiong et al., 2017). Compared with traditional field surveys, satellite-based methods can greatly reduce costs and increase efficiency. This provides the possibility to systematically map agricultural resources over large geographic areas. It is well known that using object-based methods to classify crops produces better results than pixel-based methods (Belgiu and Csillik 2018; Pena et al., 2014). First, the classification unit of object-based method (image objects) reduces the within-class variance of pixel values. Second, a large set of features (not only spectral information but also spatial information, i.e, the shape of an object, the contextual information of an object with respect to its neighbours) can be derived to characterize the objects, which can

* Corresponding authors.

E-mail addresses: zhanghuanxue@sdu.edu.cn (H. Zhang), liqz@aircas.ac.cn (Q. Li).

potentially improve classification accuracy due to the increase of complementary information (Liu and Xia 2010). However, despite many advantages of the object-based approach, its inherent limitations such as over- or under-segmentation (Moller et al., 2007) need to be addressed. Therefore, reliable techniques for automated field boundary delineation is urgently needed (Belgiu and Csillik 2018).

In recent literatures, a number of remote sensing approaches for agricultural field delineation have been explored (Persello et al., 2019; Waldner and Diakogiannis 2020). Most of these methods can be divided into two categories: edge-based and region-based methods. The edge-based methods focus on the field boundary detection using spatial gradient derived from changing pixel values in the imagery by various filters such as Scharr, Sobel, and Canny operators (Graesser and Ramankutty 2017; Turker and Kok 2013; Yan and Roy 2014). In contrast, the region-based method groups adjacent pixels into objects according to a predefined uniformity criterion (Evans et al., 2002; Garcia-Pedrero et al., 2017; Mueller et al., 2004). However, a number of issues are associated with these methods. For edge-based algorithms, the over sensitivity to high-frequency noise often produces false edges, and the arbitrary parameterization could extract incomplete boundaries. For region-based algorithms, finding the optimal segmentation parameters is a process of trial and error. For example, in the case of poor parameterization, the segmented results may lead to over segmentation for fields with high internal variability and under segmentation for small adjacent fields (Belgiu and Csillik 2018). The arbitrary parameterization related to edge-based algorithms and the over-segment and under-segment fields caused by region-based algorithms have limited the successful implementation of these two algorithms to some extent (Chen et al., 2015; Fan et al., 2001).

To solve the problem of false edge produced by edge-based algorithms and to overcome the difficulty in finding optimal parameters for region-based algorithms, many solutions have been proposed. For instance, post-processing and local adjustment of thresholds (Cheng and Liu 2020), merging adjacent objects using machine learning methods (Garcia-Pedrero et al., 2017), and combining edge-based and region-based methods to form a hybrid method (Rydberg and Borgefors 2001). Chen et al., (2015) adopted a three-step hybrid method to segment images. Given the advantage of their method over the traditional region-based approach, the manual selection process of segmentation settings poses a limitation for its wide acceptance. The most up-to-date global map of field size distribution was manually digitized (Lesiv et al., 2019). In China, the parcels of land and resources survey still heavily rely on manually digitized field boundaries with the aid of visual interpretation. Hence, there is an urgent need to develop a universal and effective method that requires minimal manual preprocessing.

In recent years, contour (boundary) detection based on convolutional neural networks (CNN) has emerged, which can automatically learn representative discriminative features from the training datasets, such as original pixel value to edges, local shapes, and complex texture patterns (Bergado et al., 2016; Farabet et al., 2013; Marmanis et al., 2016; Szegedy et al., 2015). Several CNN methods have achieved superb performance on various classification or segmentation tasks including scene classification (Cheng et al., 2018), land-cover or land-use classification (Bergado et al., 2016; Bergado et al., 2018), object localization and detection (Chen et al., 2016; Long et al., 2017), digital terrain extraction (Gevaert et al., 2018), road extraction (Cheng et al., 2017), crop type identification (Ji et al., 2018), river and water body extraction (Isikdogan et al., 2018), and urban mapping (Diakogiannis et al., 2019). The CNNs have opened up a new avenue for boundary extraction because they do not need to extract features manually, and their architectures are highly adaptable (Laukamp et al., 2020).

In the field of edge detection, various structures and training strategies of convolutional networks have been studied. Xie and Tu (2015) proposed an edge detection algorithm that combines Fully Convolutional Network (FCN) with multiple-side outputs for a deeply-supervised training. Bertasius et al., (2015) used advanced object-related features

generated by pre-trained CNN to identify contours, and the results showed that without any feature engineering, multi-scale deep learning methods have achieved the best results in contour detection. Marmanis et al. (2016) developed a deep CNN-based semantic segmentation ensemble model that extracted cleanly boundaries between areas of different land cover types. Masoud et al. (2020) designed a multiple dilation FCN (MD-FCN) for agricultural field boundary delineation using Sentinel-2 images and developed a novel super-resolution semantic contour detection network (SRC-Net) to improve the mapping resolution. The success of the aforementioned examples suggests that deep learning technology is a suitable approach for field boundary delineation from satellite images.

However, training such a CNN is very hard due to the requirement of very large labeled training datasets. Large remotely sensed datasets with labels are not readily available. One way to solve this problem is to employ a pre-trained network and then fine-tune it using the target datasets (Long et al., 2015). In the case of a small sample size, some machine learning models perform better than deep learning models. Another way is to employ an extensive data augmentation (Li et al., 2019; Ronneberger et al., 2015) or semi-supervised learning using unlabeled data (Li et al., 2019). However, these two methods are not suitable for boundary extraction because they are better at identifying land cover objects than extracting small and high-precision boundaries. For target boundary detection, because the training dataset is usually obtained manually, it can be laborious and time consuming to generate a large number of labeled data (Li et al., 2019; Scott et al., 2017). Therefore, it is necessary to develop an automated or semi-automated method to collect the training dataset.

This study evaluates the robustness of a generalized deep CNN-based approach to identify and delineate agricultural fields using Sentinel-2 images. Specifically, the training dataset for the CNN can be automatically derived by combing the Canny operator and the Watershed algorithm (CW). The main contributions of this paper are three-fold: (a) proposed a robust agricultural field delineation strategy, an automated technique based on a deep recurrent residual U-Net and boundary connection method, to delineate field boundaries; (b) developed a fast training-dataset generation strategy to meet the requirement of labeling a very large dataset for training the CNN model. The automated CW-based method is more accurate and efficient; (c) conducted extensive experimental analyses. The proposed strategies were evaluated, and the results were compared with what obtained using conventional OBIA method and universally acknowledged cropland products.

2. Study area and data

2.1. Study area

Heilongjiang Province is located in northeast China, with a total area of 473,000 square kilometers. It is China's largest commercial grain base, and the terrain is flat and open, which is suitable for large-scale crop planting. It is a temperate continental monsoon climate, with an average temperature of 21–22 °C in July. The area has sufficient sunshine, abundant rainfall, and fertile soil, making it suitable for the cultivation of spring wheat, spring corn, rice, soybean, and other crops. Spring wheat is usually sown in mid-April and harvested in mid-August; while spring corn, rice, and soybeans are sown from the end of April to early May and harvested in early October.

Four study areas including Yangming county in Mudanjiang city (YM, Study area 1), Huachuan county in Jiamusi city (HC, Study area 2), Kedong county in Qiqihar city (KD, Study area 3), and Wangkui county in Suihua city (WK, Study area 4) were chosen for this study (Fig. 1). YM is located in a mountainous area with fragmented croplands, and the other three counties are mainly distributed in plain areas. The HC and WK are characterized by smallholder farms with complex crop mix including corn, cotton, and soybean in the summer growing season, while the crop fields in KD are in general large and regular in shape.

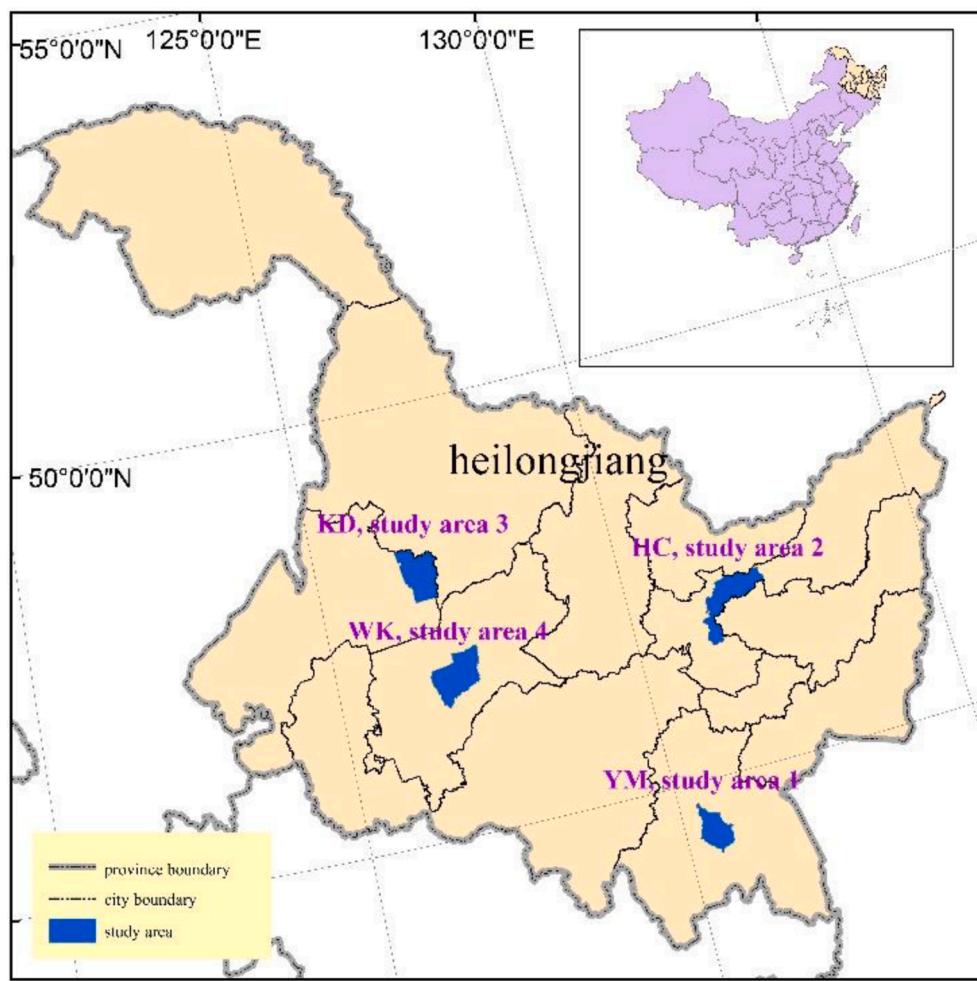


Fig. 1. Map of the four study areas in Heilongjiang province, China.

2.2. Remotely sensed data and preprocessing

Given that the main crops in Heilongjiang Province are usually harvested between mid-September and October, seven cloud-free sentinel-2 images covering the four study areas were collected between mid-July and mid-September (Table 1). The Sentinel-2 constellation provides spatial resolutions of 10 m, 20 m, and 60 m in 13 bands with a 5-day revisit time. In this study, nine bands were used with a resolution of 10 m and 20 m for field boundary extraction (Table 2).

The Sentinel-2 Level-1C data were obtained and processed using the Sen2Cor processor (<http://step.esa.int/main/third-party-plugins-2/sen2cor/>) to convert the top of atmosphere (TOA) reflectance to the bottom of atmosphere (BOA) reflectance. Subsequently, the output were resampled to a spatial resolution of 10 m using the bilinear interpolation method using the Sentinel Application Platform (SNAP) (<http://step.esa.int/main/third-party-plugins-2/sen2cor/>)

Table 1
Cloud-free Sentinel-2 images per study area.

Study sites	Area (km ²)	Pixels (10 m resolution)	Number of images	Acquisition dates
Study area 1 (YM)	1345	4164 × 6352	1	20190707
Study area 2 (HC)	2228	9042 × 8002	2	20190831, 20190915
Study area 3 (KD)	2083	5058 × 6710	2	20190911, 20190911
Study area 4 (WK)	2314	6496 × 6974	2	20190911, 20190911

Table 2

Characteristics of multi-spectral Sentinel-2 images used in the study.

Sentinel-2 bands	Wavelength (nm)	Central Wavelength (nm)	Resolution (m)	description in the study
1 (Coastal)	433–453	443	60	not used
2 (Blue)	458–523	490	10	used
3 (Green)	543–578	560	10	used
4 (Red)	650–680	665	10	used
5 (Red Edge)	698–713	705	20	used
6 (Red Edge)	733–748	740	20	used
7 (Red Edge)	773–793	783	20	used
8 (NIR)	785–900	842	10	used
8A (Narrow-NIR)	855–875	865	20	not used
9 (Water vapour)	935–955	945	60	not used
10 (SWIR-Cirrus)	1360–1390	1375	60	not used
11 (SWIR-1)	1565–1655	1610	20	used
12 (SWIR-2)	2100–2280	2190	20	used

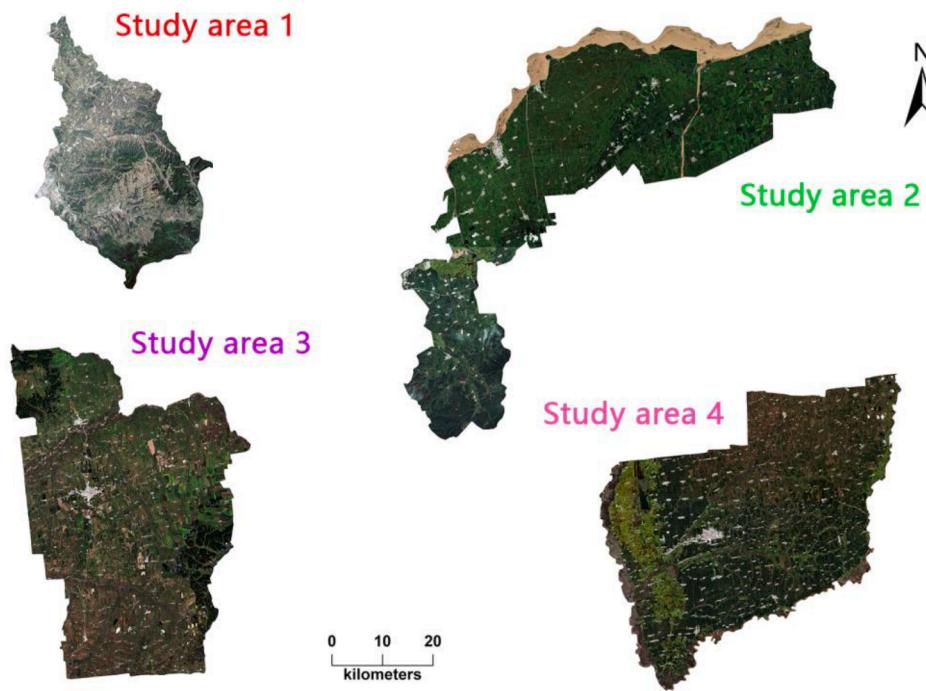


Fig. 2. Four study areas (locations shown in Fig. 1) with Sentinel-2 true color images acquired in July 07, August 31, September 11, and September 11, in 2019, respectively.

.int/main/download/). The images were then subset according to the county administrative boundaries of the four study areas (Fig. 2).

3. Methods

In the proposed framework, we exploited the CNNs with strong generalization ability to delineate agricultural field boundaries from Sentinel-2 images. It includes the following steps.

1. Automatically collect training samples for all study areas by the CW-based methods (Section 3.1);
2. Train the model for agricultural field boundary delineation by feeding the training samples into the modified recurrent residual U-Net to learn representative and discriminative features;
3. Connect the fragmented boundaries by band merging and edge tracking algorithm; and
4. Generate complete field boundary layers using the trained model and assess the accuracy.

3.1. CW-based dataset generation

We generated a total of 400 labeled subset images in the four study areas, each subset with 256 by 256 pixels, 350 of which were used for training, and the remaining 50 were used for model testing. Labeling of the training dataset was done using the following four automated processing steps:

(1) Canny-based edge detection

Many algorithms such as Sobel, Prewitt, Robert, Scharr, and Canny have been used to detect and enhance edges (Mittal et al., 2019). These edge detection algorithms generate a gray scale image representing the spatial gradient that can be used to detect edges in the image. A larger value represents a larger spatial gradient in original image space, hence likely the edge of a patch of land cover. In our study, the Canny operator (Canny 1986) was selected due to its superior performance compared

with other edge detection algorithms (Watkins and van Niekerk 2019).

The Canny operator involves a multi-stage process which can be summarized as: initial image smoothing using a Gaussian filter followed by gradient calculation, non-maximum suppression, and hysteresis threshold segmentation. In our study, the Canny algorithm was implemented in Python using the Scikit image Library. The nine Sentinel-2 bands as listed in Table 2 were used as input for edge detection, resulting in nine edge images for each training image. The nine edge layers were then aggregated by calculating the mean value per pixel of each band layer. At the end, a multi-band composite (aggregated or layer-stacked) edge image was generated.

(2) Watershed segmentation

In this step, the watershed segmentation algorithms (WSA) was applied to the aggregated edge images obtained from the previous step. The WSA has been frequently used in agricultural applications (Lebourgeois et al., 2017; Pena et al., 2014). The WSA is a region-based algorithm suitable for a single gradient image (Li et al., 2010). The objective of the algorithm is to segment a region into sub-regions using the local maxima as separation boundaries. The algorithm uses the local minimum value as the seed point. As the value increases, the object expands outward until it reaches the boundary of another object, and the expanding stops. However, the WSA is prone to over-segmentation and often produces a large number of small objects; therefore, some form of region merging needs to be performed during or after the watershed segmentation operation (Bleau and Leon 2000).

When using the WSA, the height threshold is used to control when to merge adjacent objects. This situation usually occurs when the maximum pixel value of the object is lower than the specified height threshold. Considering that our ultimate goal is to develop an automated system for agricultural field boundary delineation, including automatically labeled training dataset, we replaced the local thresholding process using Jenk's algorithm, which was implemented in the Scikit image Python package, to automatically find the most suitable height threshold. To obtain the height threshold, we first applied a Gaussian filter (using a 9×9 pixels moving window) to the aggregated edge

image (obtained based on the Canny operator) and then derived a generalized edge image by calculating the weighted average of each pixel as the middle pixel value of the moving window. Finally, the height threshold was obtained from the standard deviation of the generalized image. Furthermore, putting this threshold into the WSA results in a set of image objects (segments) with the boundaries coinciding with spectral transitions (edges).

(3) Cropland extraction

The image objects (segments) obtained by the WSA are for all the land cover types of the whole image, including cultivated land and non-cultivated land. To reduce the impact of non-cultivated land cover types, we extracted the planting area of the cultivated land for each training image based on the original Sentinel-2 images. Vegetation indices, especially the Normalized Difference Vegetation Index (NDVI), have shown to be strongly correlated with crop growth conditions during the growing season (Huete 1988; Maselli and Rembold 2001; Zhang et al., 2015). In our study, we extracted four NDVI-based features from each pixel of the original Sentinel-2 image, namely the maximum NDVI value (max), the minimum NDVI value (min), the range of NDVI values (ran), and the standard deviation (std) in one object. Using field survey points as training samples, a regression tree (CART) algorithm (Loh 2011) was used to generate the decision tree for separating (classifying) cultivated and uncultivated land cover types based on the four features. Finally, the image objects obtained by WSA were masked using the extracted cultivated land to get the agricultural field boundary results.

Table 3 lists the rules used for each study area. YM and KD require one or two rules, while HC and WK require four to five rules. A commonality among the regions is the use of the max NDVI feature, the std, and the min NDVI. In the final extraction result, the image objects that do not meet the rules are excluded from further consideration.

(4) Accuracy assessment

A reference field boundary dataset is used to evaluate the accuracy of the training dataset. The dataset consists of 200 fields, which are manually digitized from Google Earth images. Because the acquisition dates of the Google Earth images are different from that of the Sentinel-2, the digitized field boundary was superimposed on the Sentinel-2 images, and the inconsistent visually interpreted reference boundaries were removed. The details of accuracy assessment are described in **Section 3.4**.

3.2. Deep recurrent residual U-Net architecture

In semantic segmentation, it is very important to use both low-level detailed features and high-level semantic information together to achieve improved segmentation results (Long et al., 2015). In this study, we adopted the recurrent residual U-Net structure originated from the R2U-Net (Alom et al., 2020). The R2U-Net was originally designed to segment medical images by taking advantages of three recently developed CNN, the U-Net model (Ronneberger et al., 2015), the deep residual model (He et al., 2016), and the recurrent CNN (RCNN) (Liang and Hu 2015).

Table 3

Rule sets per study area.

Study area	Rule sets	
	#1	#2
YM	max NDVI > 0.63 & std NDVI < 0.15	
HC	max NDVI > 0.68 & min NDVI < =0.19	max NDVI > 0.51 & std NDVI < 0.11 & min NDVI > 0.19
KD	min NDVI > 0.49 & std NDVI > 0.05	
WK	max NDVI > 0.61 & std NDVI < =0.17	max NDVI > 0.60 & std NDVI > 0.05 & min NDVI > 0.45

According to Alom et al. (2020), the recurrent residual convolutional operations work well in object recognition when combined with the Inception neural network.

Fig. 3 shows the four model architectures that are relevant to our study, the U-Net with basic forward convolution layers (**Fig. 3a**) (Ronneberger et al., 2015), the U-Net model with forward convolutional layers and residual unit, i.e. residual U-Net (ResU-Net), (**Fig. 3b**) (Milletari et al., 2016; Zhang et al., 2018b), the U-Net model with recurrent convolution layers and residual unit, i.e. R2U-Net, (**Fig. 3c**) (Alom et al., 2019), and the R2U-Net model architecture that is adopted in this study (**Fig. 3d**).

Specifically, the R2U-Net architecture adopted in this study is different from the original U-Net (Ronneberger et al., 2015), ResU-Net (Zhang et al., 2018b), and R2U-Net (Alom et al., 2019) in the following aspects:

- The R2U-Net adopted in this study, has kept the convolutional encoding and decoding units of original U-Net model. However, compared with the basic forward convolutional layers in U-Net, the recurrent convolutional layers (RCLs) with residual units are used in the R2U-Net, which can help the development of a more efficient deeper model. In addition, the R2U-Net model only uses concatenation operations and removed the cropping and copying units used in the U-Net model.
- Compared with the U-Net, a batch normalization operation is adopted in this work as illustrated in **Fig. 3a,d**. Batch normalization is a technique that can accelerate deep network training by alleviating the problem of internal covariate shift while training a significant deep neural network (Ioffe and Szegedy 2015). Studies have shown that batch normalization can significantly reduce the number of iterations towards convergence and improve the final performance (Kampffmeyer et al., 2016).
- Compared with ResU-Net (Zhang et al., 2018b), the R2U-Net model used in this study added the RCLs and included the efficient feature accumulation method with different time steps in the RCLs, as shown in **Fig. 3b,d**.
- Compared with the R2U-Net (Alom et al., 2019), the max pooling layer was replaced by a convolutional layer with a stride of 2 (**Fig. 3c, d**). This modification was incorporated because the convolutional layer with increased stride is better than the maximum pool in terms of multiple image recognition benchmarks (Springenberg et al., 2015).

To evaluate the performance of R2U-Net adopted in the study on agricultural field boundary extraction, the U-Net and ResU-Net models were selected as the contrast experiments. A 3×3 convolutional kernel was used as the encoding unit (for U-Net and ResU-Net) followed by a batch normalization layer (for ResU-Net). For down-sampling, a 2×2 max-pooling layer was used between convolutional blocks; subsequently, a 1×1 convolutional layer was used. In the decoding unit, each convolutional block consists of a transpose layer, two convolutional layers, and a concatenation layer. At the end, 1×1 convolutional kernels and a sigmoid activation function were used, resulting in a single output feature map.

The parameter setup of the R2U-Net is the same as the ResU-Net, except for the RCLs. In the encoding unit, each block consists of two or three RCLs, and 3×3 convolutional kernels with a stride of 2 replaces the max-pooling layer. Additionally, three feature fusion methods (addition, concatenation, addition and concatenation) were compared using an element-wise addition operation between encoding and decoding units. The results show that the performance of concatenation operation series operation is the best; hence, the concatenation operation was used among the features in the model.

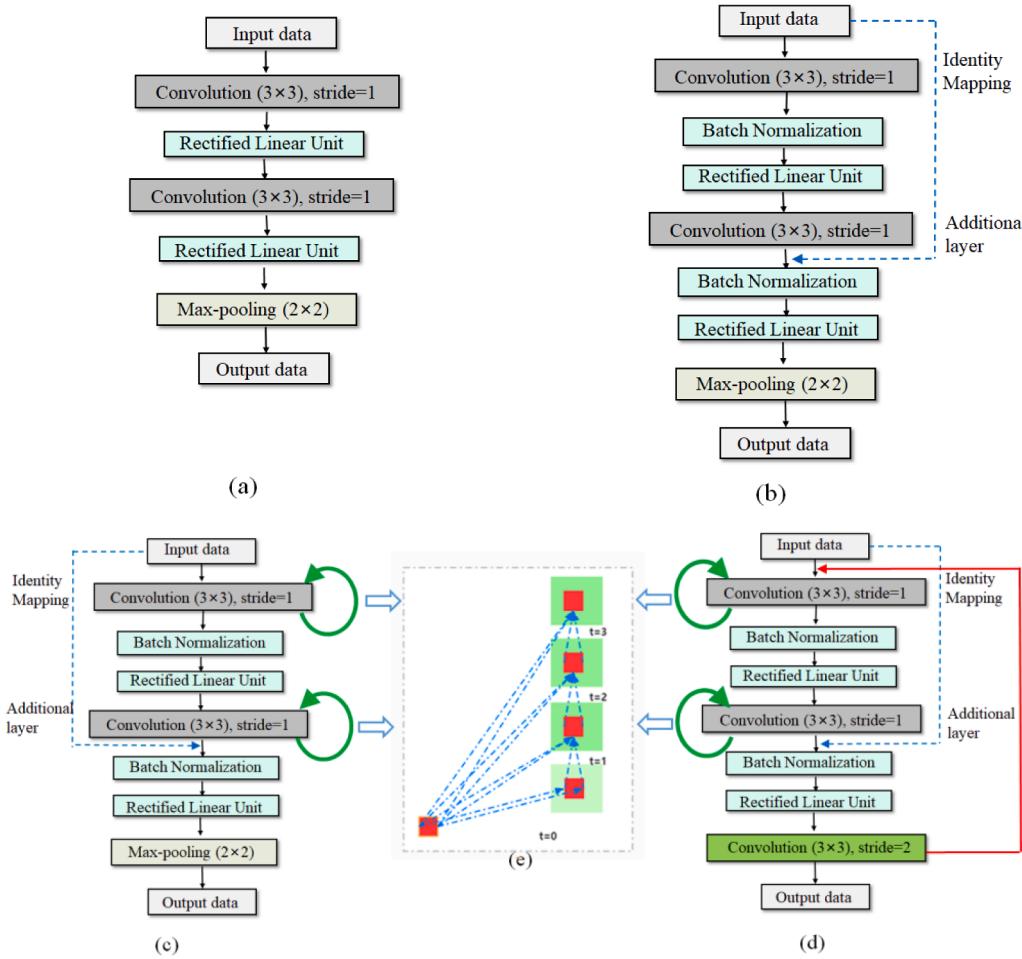


Fig. 3. Different variants of the convolutional and recurrent convolutional units in the study. (a) the U-Net (Ronneberger et al., 2015), (b) the ResU-Net (Zhang et al., 2018b), (c) the R2U-Net (Alom et al., 2019), (d) R2U-Net adopted in the study, and (e) the pictorial representation of the unfolded recurrent convolutional layer (RCL) with the time step $t = 3$, refers to one forward convolutional layer followed by three RCLs. The blue arrows in (e) represent the convolutional operation with respect to different time steps. The red and green rectangles indicate the kernels and the feature maps for each layers. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

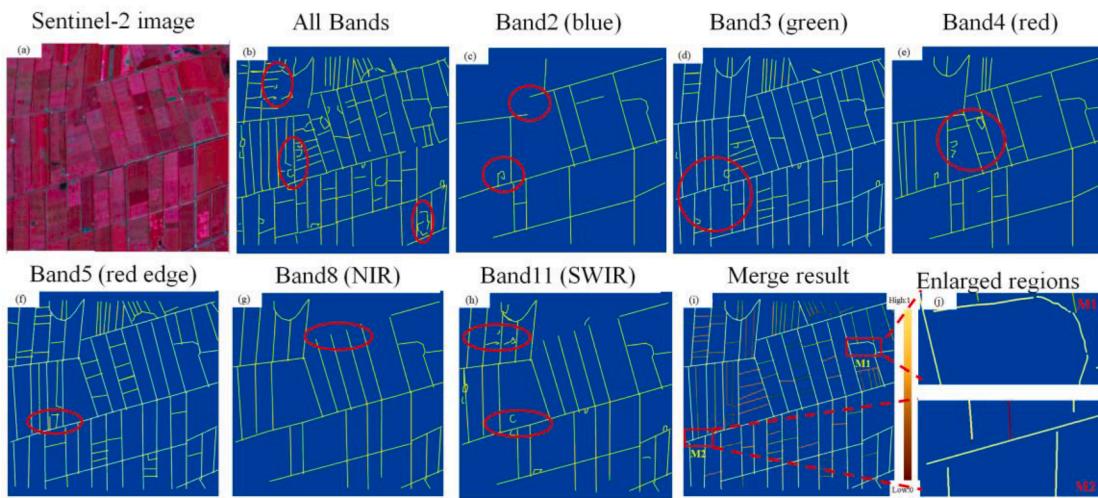


Fig. 4. An example showing the fragmented boundary strength by calculating the frequency of each boundary from one test in KD, with (a) the standard false colour Sentinel-2 image, (b) all bands results, (c-h) each band results, (i) the boundary strength result, and (j) two enlarged regions from image (i). The red circles mark the locations of fragmented boundaries. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

3.3. Boundary connection and field generation

The contour detector based on the R2U-Net results in fragmented contours, which do not partition the image into closed segments. As an illustrative example, Fig. 4 (b), shows the extracted boundary using all

image bands from the KD test tiles. In order to solve this problem, we adopted an individual band fusion method. Firstly, each band of the multi-spectral Sentinel-2 images was ingested into the trained R2U-Net model one at a time, resulting in nine agricultural field boundary layers for the same area, as shown in Fig. 4. Due to high similarity of the three

red-edge bands and the two SWIR bands, the Fig. 4 only shows one of them, namely Band 5 and Band 11, respectively. And then the frequency of each boundary was calculated and normalized to 0–1.

Secondly, the layers were fused according to a preset threshold. In the study, the optimum thresholds were automatically investigated by a receiver operating characteristics (ROC) curve. ROC is a graph showing the degree of a true prediction in contrast to a false prediction. The ROC graph describes the relative trade-off between the true positive rate (TPR, plotted in the Y axis) and the false positive rate (FPR, plotted in the X axis) (Narudin et al. 2014). In our study, we selected the reference fields to calculate the TPR and FPR, and changed the threshold value from 0.1 to 1.0 to generate the ROC space. In the ROC space, each threshold value creates a diverse point, and the optimal cutoff point is where TPR is high and FPR is low (Narudin et al. 2014). In Fig. 5, the optimal cutoff point and the different threshold values for different study areas are illustrated in ROC curve. From the results, 0.9, 0.8, 0.7, and 0.8 were selected as the optimum thresholds for the four study areas, respectively.

Finally, the fragmented boundaries were reserved if the boundary strength is more than the threshold, and the boundary strength outputs (Fig. 4(i)) were obtained. However, some of the generated boundaries are not closed. As an illustrative example (regions of M1 and M2 in Fig. 4(j)), we can always recover that many boundary pieces remain disconnected from other pieces on the different or same boundary by gaps. To address this issue, we adopted the boundary connecting method developed by Persello et al. (2019). Firstly, we applied the Oriented Watershed Transform (OWT) to construct the finest set of regions, i.e., an over-segmentation from an oriented contour signal. Here, the threshold selection of OWT followed the method introduced in Section 3.1. Secondly, the gaps are closed by iteratively merging the OWT-based contours and the boundary strength results based on their common boundary. The two-step technique leads to the final agricultural field boundaries.

From Fig. 4, almost all the individual bands produce fragmented boundaries and failed to divide the image into closed areas, especially in areas with highly similar spectral features between adjacent fields. The red edge band and green band achieves a relatively better result, followed by the NIR band, and other bands.

3.4. Accuracy assessment

In the study, two types of methods were used to evaluate the accuracy of both training dataset and final agricultural field boundary maps, described as follows.

3.4.1. Area-based method

Area-based metrics are calculated using a confusion matrix. The matrix consists of 10,000 pixels, of which 5000 “boundary” pixels are randomly selected within 10 m of the reference boundary, and another 5000 “non-boundary” pixels are randomly selected at a distance of more than 10 m from the reference boundary.

The accuracy indicators include overall accuracy (OA), Kappa coefficient (K), commission error (CE), omission error (OE), and precision-recall (PR). The OE provides an indication of the accuracy of the extracted boundary along the reference field boundary, where a high error indicates that the segmentation does not describe the field boundary well. CE represents a false boundary in the field, usually caused by excessive segmentation in the field.

The PR, introduced in (Martin et al., 2004; Martin et al., 2002), is a framework computing Precision (P), Recall (R), and F1-score (or F-measure). For the agricultural field boundary map, the PR measure is calculated based on four terms which are true positive, true negative, false positive, and false negative. P is the number of true positive predicted pixels divided by the number of all positive pixels returned by the network, measuring how close the detected boundaries are to the reference boundaries. R is the number of true positive predicted pixels divided by all pixels that should have been identified as positive, indicating the proportion of reference boundaries correctly detected. F1-Score (F) is a harmonic average of P and R, ranging from 0 to 1 and tending to be at best when it approaches 1 (Masoud et al., 2020).

$$F = 2 \frac{PR}{P + R} \quad (1)$$

3.4.2. Edge-based method

Edge-based accuracy is evaluated by four indicators (over-segmentation rate, under-segmentation rate, position offset, and eccentricity), which reflect the shape, size, and shifts of the extracted fields relative to the target or reference fields, respectively. The overlapping regions between the reference fields and the delineated fields were extracted first, then the four metrics were calculated: (1) the relative area of an overlapping region to a reference field (the over-segmentation rate, RA_{or}); (2) the relative area of an overlapping region to an extracted field (the under-segmentation rate, RA_{os}); (3) the position discrepancy of the extracted field (s) to a reference field (the location shift, D_{sr}) calculated as the average distance between the centroids of the extracted fields and the centroids of the reference fields; and (4) the absolute differences in shape (eccentricity factor, ε):

$$RA_{or}\% = \frac{1}{n} \sum_{i=1}^n \frac{A_o(i)}{A_r} \times 100 \quad (2)$$

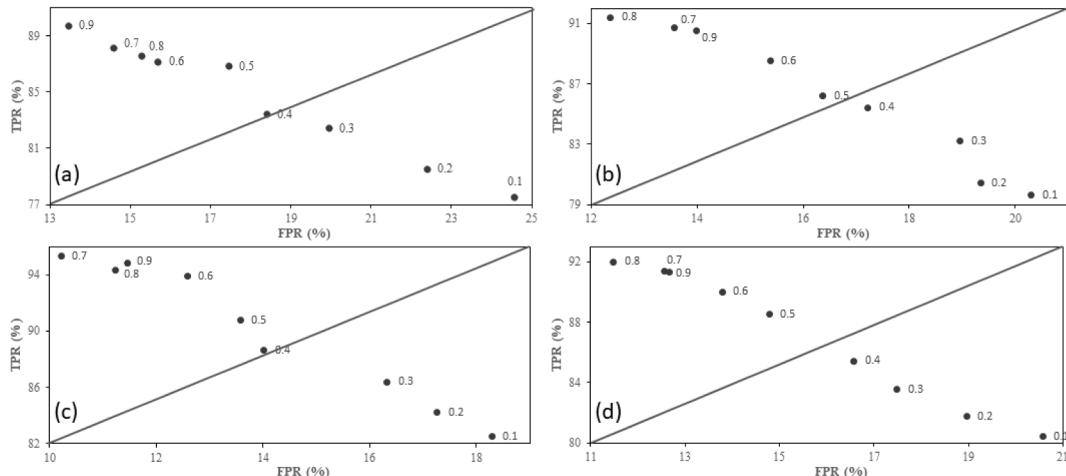


Fig. 5. Optimal cutoff point and the different threshold values for the four study areas. (a) YM, (b) HC, (c) KD, and (d) WK.

$$RA_{os}\% = \frac{1}{n} \sum_{i=1}^n \frac{A_o(i)}{A_s(i)} \times 100 \quad (3)$$

$$D_{sr} = \frac{1}{n} \sum_{i=1}^n \sqrt{(X_s(i) - X_r)^2 + (Y_s(i) - Y_r)^2} \quad (4)$$

$$\epsilon = \|\text{Eccentricity}_{\text{or}} - \text{Eccentricity}_{\text{os}}\| \quad (5)$$

where n represents the number of fields resulted from the R2U-Net model, $A_o(i)$ is the area of the i_{th} overlapping region associated with a reference field, A_r is the area of the reference field, $A_s(i)$ is the area of the i_{th} extracted field, D_{sr} is the average distance between the centroids of the extracted fields and the centroid of the reference field, representing the position accuracy of the delineated fields; $X_s(i)$ and $Y_s(i)$ are the coordinates of the centroid of the i_{th} extracted field, and X_r and Y_r are the coordinates of the centroid of the reference field. RA_{or} and RA_{os} were used to evaluate the topological similarity between the extracted fields and the reference fields, providing rate values ranging from 0 to 100: the closer the value is to 100, the less prone to over- and under-segmentation. The mean of RA_{or} , RA_{os} , and D_{sr} represent the overall segmentation quality (Zhang et al., 2018a). The eccentricity indicates how much the shape of a field deviates from a circle (Eccentricity = 0) (Persello and Bruzzone 2010).

3.5. Comparison with conventional edge detection

In this research, the representative object-based image analysis (OBIA) method was selected as the reference to evaluate the classification performance of the proposed method. One of the widely used image segmentation methods in OBIA is the multi-resolution Fractal Net Evolution Approach (FNEA) embedded in the commercial Definiens Professional software (Baatz and Schape 2000). The FNEA is a bottom-up region growing algorithm and is usually defined using both spectral variance and object geometry. It starts with a one-pixel object and subsequently merges adjacent paired objects into a larger one with the smallest increase of heterogeneity until the heterogeneity exceeds a user-defined threshold (such as the scale parameter). The image segmentation in this study was performed using the eCognition Developer (Version 9).

The segmentation algorithm requires several user-defined parameters: 1) the weights associated with the input image layers, 2) a ratio defining the contributions of spectral and spatial heterogeneities, 3) a ratio defining the contributions of the compactness and smoothness of the segmentation objects, and 4) the scale parameter that defines the heterogeneity threshold used to merge objects. The scale parameter influences the spatial scale of image analysis by controlling the size of the resultant objects. Currently, there is no automatic and objective method to determine the optimal scale parameter, which is usually determined through a trial-and-error approach by visual interpretation of the input images (Ke et al., 2010; Li et al., 2015; Myint et al., 2011). In the study, a shape parameter of 0.3, a compactness parameter of 0.5, and a scale of 70 were determined through trial and error.

4. Results and analysis

4.1. CW-based results and quality assessment

Table 4 shows that the combination of Canny and Watershed edge detection algorithm (CW) produced the mean OA of 93.76%, and the highest OA (96.61%), RA_{or} (95.45), and F (0.90) were recorded in KD. The lowest OA (91.29%) was obtained in YM. When OEs are considered, the CW scenarios recorded errors ranging from 15.81% (KD) to 22.13% (HC), with a mean error of 18.49%. Regarding the CE, Experiments 4 and 3 produced the lowest errors (0.92% and 1.23% respectively), and the highest CE (1.91%) was recorded in YM. In addition, the standard deviations produced by our method for OA (2.19%), Kappa (0.03%), and CE (0.48%) suggest a high consistency. All the results indicate that the datasets can be used to train the CNN model.

Fig. 6 visually shows the CW results in a detailed area in the four study areas, and the red circles highlight some examples where the experiments were unable to delineate the adjacent fields individually. From the figures, KD produced clearer field delineation than other areas (about more 5.3% accurate from **Table 3**), because of the relatively regular and large fields of the study area (the third row of **Fig. 6**). In study area 3 (KD), the larger areas making up those individual fields were successfully identified and delineated. The red circles in the first row illustrates the inability of the Sentinel-2 imagery to effectively delineate small and irregular fields in the mountainous area (YM) due to its fragmented landscape, and the second and the fourth rows show the missing of the smaller fields in plain areas, because of a high degree of homogeneity between adjacent fields. However, except for the red circles, the fields of the four study areas were almost all successfully delineated in **Fig. 6**. The high mean OA (93.76%) and low CE (1.48%) of the CW workflow suggest that it works well for most of the fields included in the reference set.

4.2. Analysis of field boundary delineation

4.2.1. Result of agricultural field boundary delineation

Table 5 illustrates the area-based and edge-based accuracies obtained over the four study areas using U-Net, ResU-Net, and R2U-Net, where R2U-Net obtained the best performance (mean OA 89.28%, Kappa 0.85, and F 0.83) among the three methods for all study areas. It is apparent that ResU-Net (mean OA 85.78%, Kappa 0.81, and F 0.77) can achieve better field mapping results than U-Net (mean OA 82.18%, Kappa 0.77, and F 0.74). R2U-Net still presents best accuracies on both area based metrics and edge based metrics.

For the area-based metrics (**Table 5**), the results of R2U-Net show that study area 3 (KD) produced the best OA (92.61%) and F (0.85) partly because fields in this area are more regular in shape. Study area 4 (WK) has the second best OA (89.21%) and F (0.83). This OA is, however, not significantly higher than the OA (88.73%) in HC. In contrast, the OAs of these three experiments (HC, KD and WK) are significantly higher than the accuracy in YM (R2U-Net 86.56%); YM produced the lowest OA for all three models due to its mountainous landscape. When OEs are considered, the R2U-Net in KD produced the lowest errors (OE 14.21%) compared with the others. In contrast, the OEs for HC and WK are significantly higher than the value in YM and range from 16.89% to

Table 4
Accuracy metrics of the training datasets.

Study area	Area_based metrics						Edge_based metrics				
	OA(%)	OE(%)	CE(%)	K	P	R	F	RAor(%)	RAos(%)	Dsr(m)	ϵ
YM	91.29	19.31	1.91	0.87	0.84	0.85	0.84	90.34	92.52	43.16	0.91
HC	93.34	22.13	1.85	0.89	0.86	0.85	0.84	93.24	93.45	40.23	0.92
KD	96.61	15.81	1.23	0.93	0.87	0.87	0.90	95.45	94.67	37.56	0.94
WK	93.78	16.71	0.92	0.92	0.84	0.84	0.85	92.28	93.02	39.85	0.92
Mean	93.76	18.49	1.48	0.90	0.85	0.85	0.86	92.83	93.42	40.20	0.92

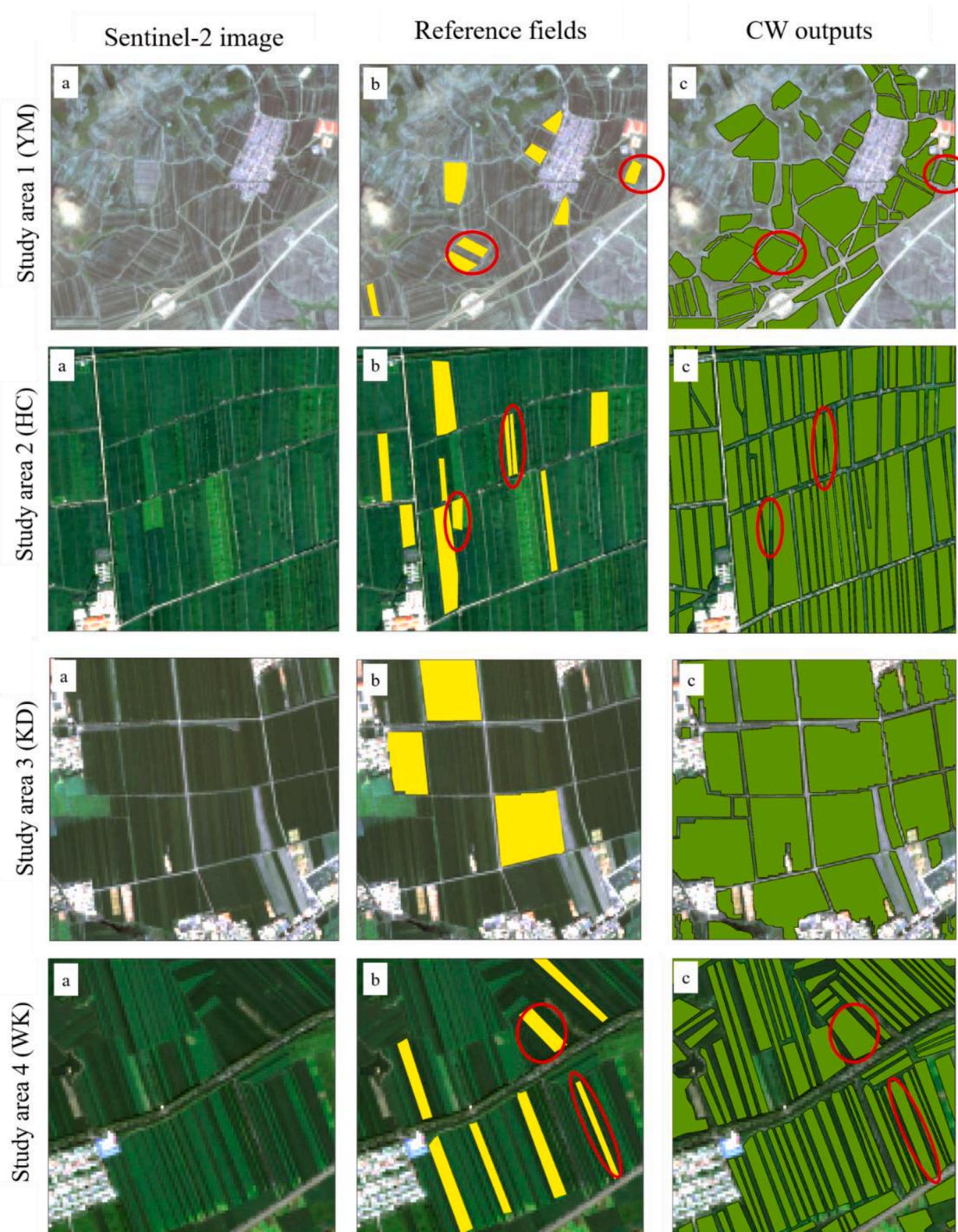


Fig. 6. CW-based datasets results for the four study areas. First column (a): RBG composite of Sentinel-2 imagery. Second column (b): the reference fields. Third column (c): the CW outputs. The red circles mark some examples where the experiments were unable to delineate the individual adjacent fields. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

19.32% for R2U-Net. This clearly demonstrates that the R2U-Net identified field boundaries much more accurately than the experiments involving U-Net and ResU-Net. Regarding CE, the R2U-Net experiment still produced the lowest error in KD (1.03%).

For the edge-based metrics, results reveal that the field boundaries of the R2U-Net experiments match the reference boundaries more closely than those of the U-Net and ResU-Net experiments. Specifically, the RA_{or} , RA_{os} , and ϵ of the R2U-Net experiment are 92.54%, 93.43%, and 0.91 in KD respectively, which is significantly higher than any of the

other techniques.

Fig. 7 shows the final field boundary maps of the four study areas, with several typical subsets for different landscapes. **Fig. 7(a)** shows the final boundary delineation results using the R2U-Net model of the four counties. **Fig. 7(b)** shows more details as S1 and S2 in each study area, displaying the original standard false color images and comparison results. The yellow circles highlight some examples where the three models performed a lot differently. From the figures, in general, the R2U-Net algorithm achieved good visual results in all four areas,

Table 5

Area-based and edge-based metric results for the four study areas of U-Net, ResU-Net and R2U-Net.

Experiment name		Area based metrics						Edge based metrics				
		OA(%)	OE(%)	CE(%)	K	P	R	F	RAor(%)	RAos(%)	Dsr(m)	ϵ
YM	U-Net	76.19	32.78	4.01	0.73	0.70	0.71	0.69	78.45	79.65	71.07	0.75
	ResU-Net	80.43	29.56	3.78	0.78	0.75	0.73	0.71	81.07	82.78	69.67	0.78
	R2U-Net	86.56	23.45	2.92	0.81	0.82	0.80	0.80	88.93	89.62	65.32	0.83
HC	U-Net	82.67	26.56	3.12	0.76	0.73	0.76	0.73	82.14	83.47	65.78	0.80
	ResU-Net	86.12	22.78	2.67	0.80	0.77	0.79	0.76	85.42	86.99	59.67	0.83
	R2U-Net	88.73	16.89	2.01	0.84	0.86	0.82	0.82	91.76	92.04	55.67	0.88
KD	U-Net	86.72	21.12	1.99	0.84	0.81	0.82	0.80	88.62	89.64	53.78	0.84
	ResU-Net	89.73	19.76	1.67	0.85	0.84	0.82	0.82	90.87	91.67	50.23	0.87
	R2U-Net	92.61	14.21	1.03	0.89	0.86	0.87	0.85	92.54	93.43	42.32	0.91
WK	U-Net	83.12	27.17	3.45	0.76	0.75	0.78	0.75	83.54	84.93	61.72	0.81
	ResU-Net	86.82	23.56	2.78	0.81	0.79	0.79	0.78	86.87	87.56	58.92	0.84
	R2U-Net	89.21	19.32	1.98	0.84	0.81	0.83	0.83	90.65	91.39	53.78	0.89
mean	U-Net	82.18	26.91	3.14	0.77	0.75	0.77	0.74	83.19	84.42	63.09	0.80
	ResU-Net	85.78	23.92	2.73	0.81	0.79	0.79	0.77	86.06	87.25	59.62	0.83
	R2U-Net	89.28	18.47	1.99	0.85	0.84	0.83	0.83	90.97	91.62	54.27	0.88

including large average field size (in KD and HC) and fragmented and heterogeneous landscapes (in YM, S2 of WK), which is consistent with **Table 5** (mean OA of 89.28% compared with 85.78% and 82.18% from ResU-Net and U-Net respectively). From the yellow circles, results show that obvious boundaries could be delineated using the U-Net model; however, the small and irregular fields are missing due to the homogeneity between adjacent fields (such as in S2 of YM and HC). Compared with the U-Net, although the detailed boundaries could be delineated by ResU-Net (such as in S1 of HC, S2 of KD), the areas with no plants could not be effectively characterized (in S2 of WK). However, R2U-Net could perform well regardless of fragmentation (YM), homogeneity (HC), and whether or not plants were growing in the cropland areas (WK).

From the detailed results of the R2U-Net (**Fig. 7**), large fields were distinguished from others such as built-up, rivers and roads for plain areas (in HC, KD and WK), and boundaries for fragmented cropland with regular distribution can be precisely located for YM. The area of irregularly distributed cropland areas were extracted, and forest with similar spectral features to cropland was well discriminated (in S1 of WK). In mountainous areas, small villages and large forest areas could be filtered out (in study area 1, YM). In WK, cropland boundaries without plants (fallow) was effectively detected. The proposed R2U-Net was able to successfully delineate croplands with various texture and spectral characteristics.

4.2.2. Effects of the CNN models

In this study, CNN models were tested and compared over four study areas, R2U-Net (adopted in this study), ResU-Net, and U-Net. It should be noted that we applied the hyper parameter setting as mentioned in **Section 3.2** to train the three CNN models.

The loss curves during the training process and accuracy assessment of the different CNN models are presented in **Fig. 8**. Clearly, the accuracies in study area 1 were the lowest among all the models. The accuracies in study area 3 are not much superior to study area 2 and 4, suggesting that the models perform differently in plain and mountainous areas. Regardless of study areas, the results of the R2U-Net is better than that of the ResU-Net and U-Net, demonstrating that it is an effective strategy to combine low-level features with high-level semantic features to generate fine segmentation results.

4.3. Comparison with OBIA method and current cropland products

In the study, we selected 5×5 km boxes in YM and KD (**Fig. 9**) to replace the entire study area for experiments, considering the complexity and difficulty of object-based classification for large-scale field extraction. The YM and KD represented typical complex fields and regular fields, respectively.

From the perspective of qualitative and quantitative analysis, we

compared individually the performance of R2U-Net and OBIA methods in both plain and mountainous areas (as displayed in **Table 6** and the second and third column of **Fig. 9**). It should be noted that OBIA and R2U-Net are inner-area model for its test area in KD whereas OBIA and R2U-Net are inter-area model in YM. In general, regardless of R2U-Net or OBIA, the inner-area model had more accurate classification performance than the inter-area model. And R2U-Net is significantly superior to OBIA, corroborating the advantages of CNN in extracting field boundaries using Sentinel-2 images. According to the results of inter-area model, the more undesirable visual effects and misclassifications were observed for OBIA approach in contrast to R2U-Net. In particular, the inter-area model of R2U-Net (OA 86.56%, and F 0.80 in study area 1) achieved a higher accuracy compared with inner-area model of OBIA (OA 78.23%, and F 0.71). This shows that R2U-Net possesses a strong spatiotemporal transferability and generalization capability for data variation. Among all the R2U-Net experiments, the model used for the whole study areas (mean OA 89.28%, **Table 5**) realized better classification results than inter-area models (OA 86.56%), which suggests that the R2U-Net can generalize various characteristics of fields through large amount of training samples and demonstrates the applicability of the R2U-Net to large-scale field boundary delineation.

Secondly, we also compared our results with the existing cropland maps generated from FROM-GLC10, i.e. 10 m resolution global land cover maps with ten classes for 2017 from Sentinel-2 data ([Gong et al., 2019](#)) (the last column of **Fig. 9**). It can be seen from the figures that the proposed approach produced broadly similar distribution patterns of cropland with FROM-GLC10, but more accurate boundary characteristics with finer spatial details. In particular, the proposed approach cannot only extract more details within regular field zones, but can also detect small scattered cropland parcels in fragmented areas. For example, KD shows the proposed approach not only distinguished cropland distribution from others in plain areas, but also delineated the detailed boundaries, cannot be distinguished by FROM-GLC10. Similarly, YM shows that fragmented small cropland areas can be successfully extracted by the proposed approach, whereas some water areas were misclassified as cropland by FROM-GLC10.

From the results, we found that the existing global datasets associated with cropland are limited by insufficient spatial resolution to properly represent areas with small field parcel size, and their less-than-ideal accuracies hamper the application for parcels at local scale. Therefore, it is necessary to develop automated and accurate cropland boundary extraction method especially at regional and local scales.

5. Discussion

In this study, training dataset was automatically labeled (with a mean overall accuracy of 93.76%) using Canny operation and watershed

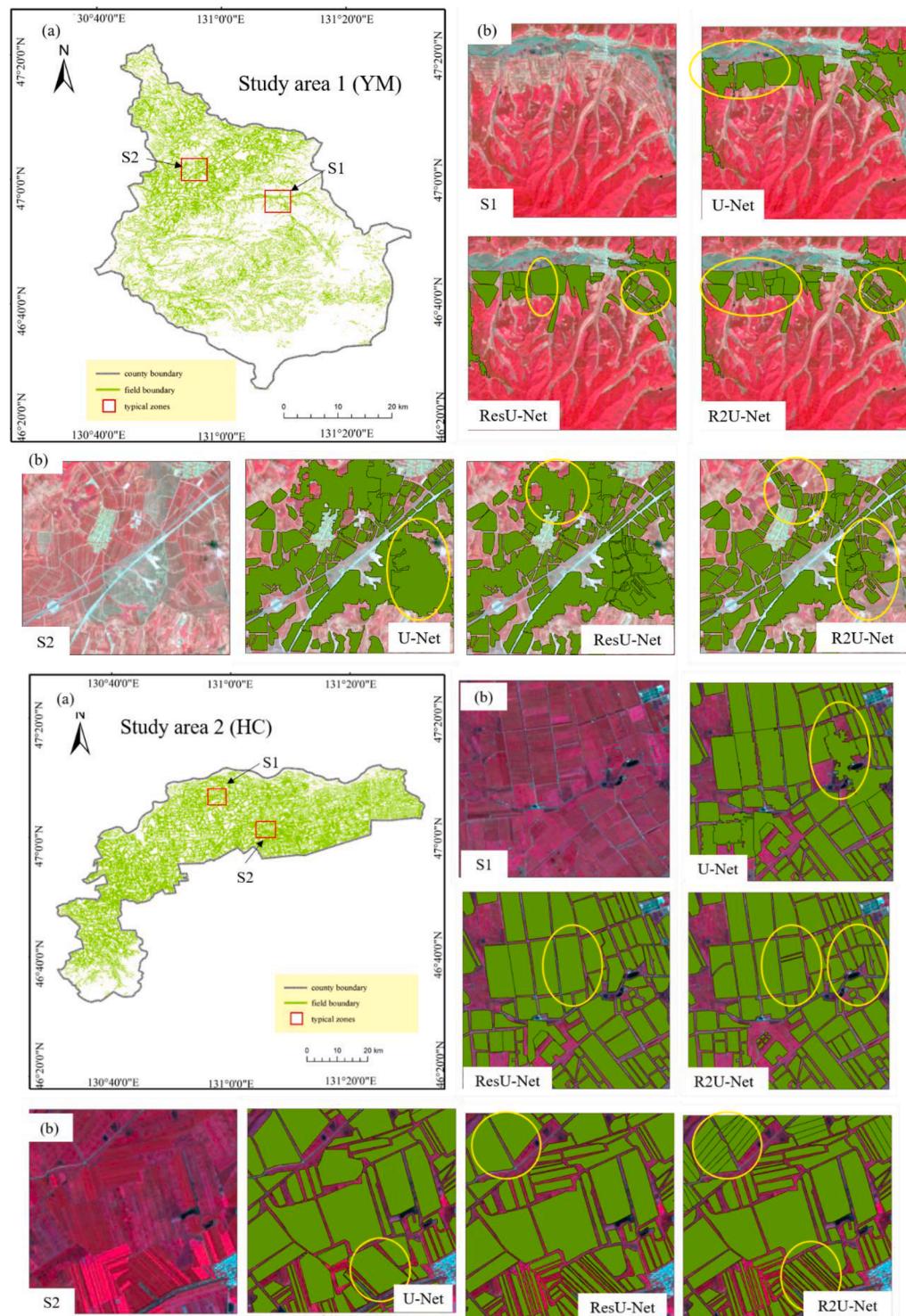
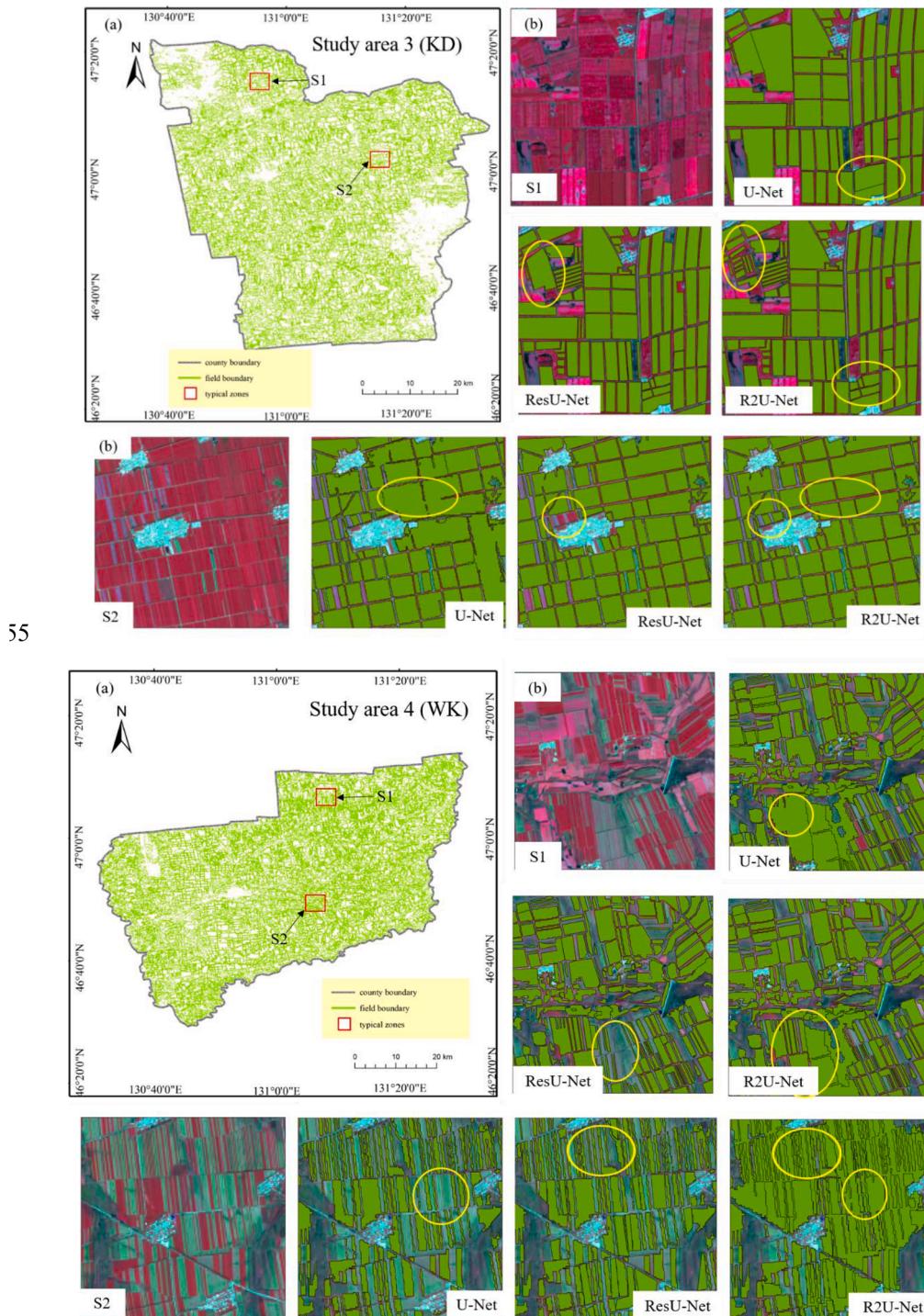


Fig. 7. Results of field boundary delineation for the four study areas of U-Net, ResU-Net and R2U-Net: (a) field boundaries using R2U-Net model in all four study areas. (b) 5 × 5 km typical subsets marked in each study area as S1 and S2, with original standard false color sentinel-2 images, and delineating results using U-Net, ResU-Net, and R2U-Net, respectively. The yellow circles highlight example locations where the performance was different of the three models. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

segmentation algorithm, and agricultural field boundary was delineated using trained R2U-Net and boundary connection methods (the mean overall accuracy of 89.28%). The Canny-watershed-based method was chosen to collect the training dataset, but the training dataset was not used to obtain the final boundary results of the study area. Table 7

explained the reason, which showed the results using CW-based method and R2U-Net method in study area 1(YM) and study area 3 (KD). Results show that for mountain areas (YM), the accuracy is almost the same between the R2U-Net model (OA of 86.56%) and CW approach (OA of 89.59%), but the computation time is different when using the R2U-Net



55

Fig. 7. (continued).

(27 min in the whole study area) and CW approach (53 min). For plain areas (KD), the accuracy is obviously lower using the CW approach (OA of 86.78%) than R2U-Net (OA of 92.61%). This is because the deviation, caused by regions with great field density and the similar crop growth period and spectral characteristics in KD (5058×6710 pixels), when automatically obtaining the threshold of watershed algorithm. As a result, some of the small adjacent fields were not delineated when using the CW approach at large scales. However, when using the CW approach in the training datasets (256×256 pixels), the OA reached 96.61%, suggests that the CW approach could perform well when the extracted

agricultural field boundaries are in a small plain area. Meanwhile, the computation time is 40 min and 68 min for the trained R2U-Net model and CW approach, respectively.

This paper presents an automated agricultural field boundary delineation method using Sentinel-2 images (with 10 m or 20 m spatial resolution) without having to rely on human interventions. However, the Sentinel-2 image's spatial resolution is not fine enough to accurately delineate fields with small and fragmented parcel distributions, which is commonly occur in mountainous areas (e.g. study area 1, YM). Within complex landscapes, it has found that some fields had weak boundaries

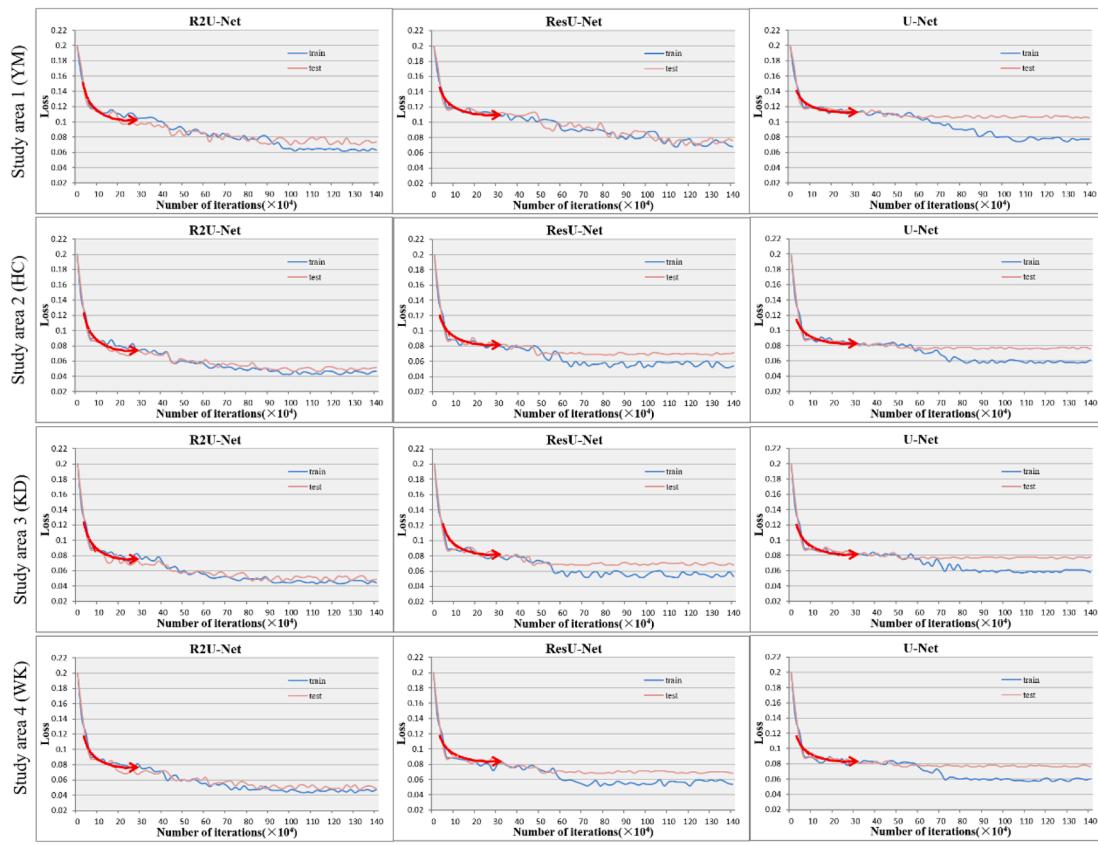


Fig. 8. Loss curves of the different CNN models during the training process in the four study areas.

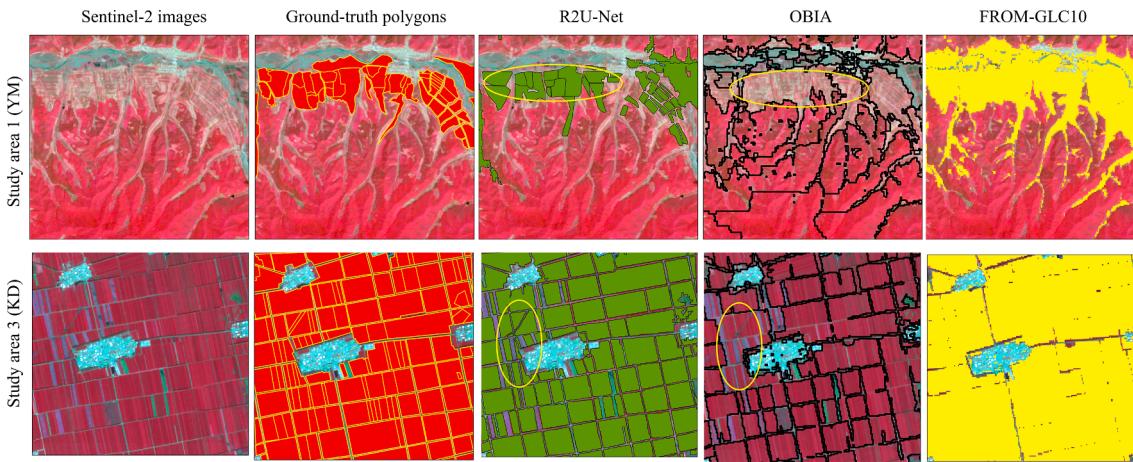


Fig. 9. Sentinel-2 standard false color images in YM and KD (the 5 × 5 km box example), ground-truth polygons, two field boundary maps (the map generated by R2U-Net model and the map generated by the OBIA), and FROM-GLC10 product.

Table 6
Accuracy comparison of R2U-Net model and the OBIA.

Experiment name		Area_based metrics							Edge_based metrics			
		OA(%)	OE(%)	CE(%)	K	P	R	F	RAor(%)	RAos(%)	Dsr(m)	ϵ
study area 1	R2U-Net	86.56	23.45	2.92	0.81	0.82	0.80	0.80	88.93	89.62	65.32	0.83
	OBIA	78.23	32.67	4.34	0.74	0.70	0.72	0.71	77.67	78.98	70.64	0.76
study area 3	R2U-Net	92.61	14.21	1.03	0.89	0.86	0.87	0.85	92.54	93.43	42.32	0.91
	OBIA	85.78	21.78	2.01	0.83	0.79	0.76	0.77	85.67	85.76	60.78	0.83

Table 7

Accuracy comparison of R2U-Net model and the CW-based method in study area 1 and study area 3.

Experiment name		Area_based metrics						Edge_based metrics				Computational time (min/study area)	
		OA (%)	OE (%)	CE (%)	K	P	R	F	RAor (%)	RAos (%)	Dsr (m)	ϵ	
study area 1	R2U-Net	86.56	23.45	2.92	0.81	0.82	0.80	0.80	88.93	89.62	65.32	0.83	27
	CW-based	85.95	25.42	3.05	0.79	0.82	0.81	0.76	88.04	88.35	67.43	0.81	53
study area 3	R2U-Net	92.61	14.21	1.03	0.89	0.86	0.87	0.85	92.54	93.43	42.32	0.91	40
	CW-based	86.78	25.34	3.42	0.80	0.78	0.77	0.77	87.32	86.67	56.12	0.85	68

due to the homogeneity between adjacent fields, making field extraction problematic. With the implementation of the Earth observation sharing policy of increasing number of countries, more and more satellite image time series can be obtained for free, such as high spatial resolution (HR) images (such as Gaofen 2 (GF2), Ziyuan 3 (ZY3), etc.), highly-detailed agricultural field mapping can likely solve the problem of small-scale farmland parcels.

Another common limitation in field boundary delineation is the use of single-date imagery. When different crop types have similar spectral and structural characteristics at certain phenological growth stage, finding the edge of the field is a challenge (Pittman et al., 2010) (e.g. study area 4, WK). This situation is even more complicated when the same crops are planted on adjacent fields. Therefore, acquiring higher resolution images throughout the growing season is also necessary for consistent and accurate field boundary extraction, although this is difficult at a large scale.

The proposed method achieved good performance and efficiency for agricultural field boundary delineation based on Sentinel-2 images. However, accuracy imbalance occurred between heterogeneous and homogeneous zone. Therefore, future research should take into consideration of the terrain and landscape, separate heterogeneous and homogeneous regions according to certain standards, and conduct individual model training and field extraction. Additionally, this study used the CW-based method to establish training dataset. Although this is effective and fast in a large-scale data environment, how to further improve the accuracy of the training data is still a problem. This problem might be solved by adding existing land cover and arable land product information on the basis of existing methods in order to extract training samples in future research.

6. Conclusion and recommendations

In this study, a new automated method for agricultural field boundary delineation was proposed. The performances of the proposed method were evaluated over four sites in northeastern China. The final boundary maps showed an overall accuracy of 89.28% and a Kappa coefficient 0.85 using CW-trained R2U-Net model. Compared with traditional object-based classification method and existing 10-m resolution global land cover map (FROM-GLC10), the proposed method shows good performance. Since only single-date image is used, this may limit the accuracy of the mapping, especially in mountain and complex cropland areas. It is recommended to apply the proposed method to multi-temporal remotely sensed data in the future research, which maybe lead to more accurate field boundaries. The successful application of the proposed method in agricultural field delineation in Northeast China shows that the method has great application potential.

The most impressive feature of the automated method is that it reduces the workload for labeled datasets in model training using a CW-based method, thereby significantly reduce the cost of agricultural field extraction. This is especially important for fast field boundary delineation at regional or larger scales using high resolution remote sensing data. Considering the data characteristics, the simple and robust R2U-Net algorithm, which provides deep meaningful features, was

proven to have good application potential for deep semantic segmentation tasks.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

The authors are very grateful for the financial support provided by National Natural Science Foundation of China (Grant No. 42001364, No. 42071403, and No. 41901356), Natural Science Foundation of Shandong Province, China (Grant No. ZR2018BD006), National College students' Innovative Entrepreneurial Training Plan Program (Grant No. 202110445034), and funding supported by China Postdoctoral Science Foundation (Grant No. 2018M632707). The authors would also like to thank the editors and anonymous reviewers of the journal for their thorough review and insightful comments that have helped improving the quality of the paper.

References

- Alom, M.Z., Hasan, M., Yakopcic, C., Taha, T.M., Asari, V.K., 2020. Improved inception-residual convolutional neural network for object recognition. *Neural Comput. Appl.* 32 (1), 279–293. <https://doi.org/10.1007/s00521-018-3627-6>.
- Alom, M.Z., Yakopcic, C., Hasan, M., Taha, T.M., Asari, V.K., 2019. Recurrent residual U-Net for medical image segmentation. *J. Med. Imag.* 6 (01), 1. <https://doi.org/10.1117/1.JMI.6.1.014006>.
- Baatz, M., Schape, A., 2000. Multiresolution segmentation: an optimization approach for high quality multi-scale image segmentation. *Proc. Angew. Geogr. Informationsverarb.* 12, 12–23.
- Belgiu, M., Csillik, O., 2018. Sentinel-2 cropland mapping using pixel-based and object-based time-weighted dynamic time warping analysis. *Remote Sens. Environ.* 204, 509–523. <https://doi.org/10.1016/j.rse.2017.10.005>.
- Bergado, J.R., Persello, C., Govaert, C., 2016. A deep learning approach to the classification of sub-decimetre resolution aerial images. *IEEE Int. Geosci. Remote Sens. Symp.* 1516–1519. <http://doi.org/10.1109/IGARSS.2016.7729387>.
- Bergado, J.R., Persello, C., Stein, A., 2018. Recurrent Multiresolution Convolutional Networks for VHR Image Classification. *IEEE Trans. Geosci. Remote Sens.* 56 (11), 6361–6374. <https://doi.org/10.1109/tgrs.2018.2837357>.
- Bertasius, G., Shi, J., Torresani, L., 2015. DeepEdge: A Multi-Scale Bifurcated Deep Network for Top-Down Contour Detection. *IEEE Conf. Comput. Vis. Pattern Recognit.* 4380–4389. <http://doi.org/10.1109/CVPR.2015.7299067>.
- Bleau, A., Leon, L.J., 2000. Watershed-based segmentation and region merging. *Comput. Vis. Image Und.* 77 (3), 317–370. <https://doi.org/10.1006/cviu.1999.0822>.
- Canny, J., 1986. A computational approach to edge detection. *IEEE T Pattern Anal.* 8 (6), 679–698. <https://doi.org/10.1109/tpami.1986.4767851>.
- Chen, B., Qiu, F., Wu, B., Du, H., 2015. Image Segmentation Based on Constrained Spectral Variance Difference and Edge Penalty. *Remote Sens.* 7 (5), 5980–6004. <https://doi.org/10.3390/rs70505980>.
- Chen, S., Wang, H., Xu, F., Jin, Y.-Q., 2016. Target Classification Using the Deep Convolutional Networks for SAR Images. *IEEE T. Geosci Remote Sens.* 54 (8), 4806–4817. <https://doi.org/10.1109/TGRS.2016.2551720>.
- Cheng, G., Wang, Y., Xu, S., Wang, H., Xiang, S., Pan, C., 2017. Automatic Road Detection and Centerline Extraction via Cascaded End-to-End Convolutional Neural Network. *IEEE T. Geosci Remote Sens.* 55 (6), 3322–3337. <https://doi.org/10.1109/TGRS.2017.2669341>.
- Cheng, G., Yang, C., Yao, X., Guo, L., Han, J., 2018. When Deep Learning Meets Metric Learning: Remote Sensing Image Scene Classification via Learning Discriminative CNNs. *IEEE T. Geosci Remote Sens.* 56 (5), 2811–2821. <https://doi.org/10.1109/tgrs.2017.2783902>.

- Cheng, X., Liu, H.S., 2020. A Novel Post-Processing Method Based on a Weighted Composite Filter for Enhancing Semantic Segmentation Results. *Sensors-Basel.* 20 (19), 13. <https://doi.org/10.3390/s20195500>.
- Diakogiannis, F.I., Waldner, F., Caccetta, P., Wu, C., 2019. ResUNet-a: a deep learning framework for semantic segmentation of remotely sensed data. *ISPRS J. Photogramm.* 162, 94–114. <https://doi.org/10.1016/j.isprsjprs.2020.01.013>.
- Evans, C., Jones, R., Svalbe, I., Berman, M., 2002. Segmenting multispectral landsat TM images into field units. *IEEE T. Geosci. Remote Sens.* 40 (5), 1054–1064. <https://doi.org/10.1109/tgrs.2002.1010893>.
- Fan, J.P., Yau, D.K.Y., Elmagarmid, A.K., Aref, W.G., 2001. Automatic image segmentation by integrating color-edge extraction and seeded region growing. *IEEE T Image Process.* 10 (10), 1454–1466. <https://doi.org/10.1109/83.951532>.
- Farabet, C., Couprie, C., Najman, L., LeCun, Y., 2013. Learning Hierarchical Features for Scene Labeling. *IEEE T Pattern Anal.* 35 (8), 1915–1929. <https://doi.org/10.1109/tpami.2012.231>.
- García-Pedrero, A., Gonzalo-Martín, C., Lillo-Saavedra, M., 2017. A machine learning approach for agricultural parcel delineation through agglomerative segmentation. *Int. J. Remote Sens.* 38 (7), 1809–1819. <https://doi.org/10.1080/01431161.2016.1278312>.
- Gevaert, C.M., Persello, C., Nex, F., Vosselman, G., 2018. A deep learning approach to DTM extraction from imagery using rule-based training labels. *ISPRS J. Photogramm.* 142, 106–123. <https://doi.org/10.1016/J.ISPRSJPRS.2018.06.001>.
- Gong, P., Liu, H., Zhang, M., Li, C., Wang, J., Huang, H., Clinton, N., Ji, L., Li, W., Bai, Y., Chen, B., Xu, B., Zhu, Z., Yuan, C., Ping Suen, H., Guo, J., Xu, N., Li, W., Zhao, Y., Yang, J., Yu, C., Wang, X., Fu, H., Yu, L., Dronova, I., Hui, F., Cheng, X., Shi, X., Xiao, F., Liu, Q., Song, L., 2019. Stable classification with limited sample: transferring a 30-m resolution sample set collected in 2015 to mapping 10-m resolution global land cover in 2017. *Science Bulletin. Sci. Bull.* 64 (6), 370–373. <https://doi.org/10.1016/J.JSCB.2019.03.002>.
- Graesser, J., Ramankutty, N., 2017. Detection of cropland field parcels from Landsat imagery. *Remote Sens. Environ.* 201, 165–180. <https://doi.org/10.1016/j.rse.2017.08.027>.
- He, K., Zhang, X., Ren, S., Sun, J., 2016. Deep Residual Learning for Image Recognition. *IEEE Conf. Comput. Vis. Pattern Recognit.* 770–778. <https://doi.org/10.1109/CVPR.2016.90>.
- Huete, A.R., 1988. A soil-adjusted vegetation index (SAVI). *Remote Sens. Environ.* 25 (3), 295–309. [https://doi.org/10.1016/0034-4257\(88\)90106-X](https://doi.org/10.1016/0034-4257(88)90106-X).
- Ioffe, S., Szegedy, C., 2015. Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift. *Proc. 32nd Int. Conf. Mach. Learn.* 37, 448–456.
- Isikdогan, F., Bovik, A., Passalacqua, P., 2018. Learning a River Network Extractor Using an Adaptive Loss Function. *IEEE Geosci. Remote S.* 15 (6), 813–817. <https://doi.org/10.1109/LGRS.2018.2811754>.
- Ji, S., Zhang, C., Xu, A., Shi, Y., Duan, Y., 2018. 3D Convolutional Neural Networks for Crop Classification with Multi-Temporal Remote Sensing Images. *Remote Sens.* 10 (1), 75–91. <https://doi.org/10.3390/rs10010075>.
- Kamppfmeyeier, M.C., Salberg, A., Jenssen, R., 2016. Semantic Segmentation of Small Objects and Modeling of Uncertainty in Urban Remote Sensing Images Using Deep Convolutional Neural Networks. *IEEE Conf. Comput. Vis. Pattern Recognition Workshops* 1, 680–688. <https://doi.org/10.1109/CVPRW.2016.90>.
- Karamura, E.B., Jogo, W., Rietveld, A., Ochola, D., Staver, C., Tinazaera, W., Karamura, D., A., Kubiriba, J., Weise, S., 2013. Effectiveness of agro-ecological intensification practices in managing pests in smallholder banana systems in East and Central Africa. *Int. ISHS-ProMusa Symp. Bananas Plant.* (986), 119–126. <https://doi.org/10.17660/ActaHortic.2013.986.10>.
- Ke, Y., Quackenbush, L.J., 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>.
- Laukamp, K.R., Pennig, L., Thiele, F., Reimer, R., Görtz, L., Shakirin, G., Zopfs, D., Timmer, M., Perkuhn, M., Borggrefe, J., 2020. Automated Meningioma Segmentation in Multiparametric MRI Comparable Effectiveness of a Deep Learning Model and Manual Segmentation. *Clin. Neuroradiol.* 31 (2), 357–366. <https://doi.org/10.1007/s00062-020-00884-4>.
- Lebourgéois, V., Dupuy, S., Vintrou, E., Ameline, M., Butler, S., Begue, A., 2017. A Combined Random Forest and OBIA Classification Scheme for Mapping Smallholder Agriculture at Different Nomenclature Levels Using Multisource Data (Simulated Sentinel-2 Time Series, VHRS and DEM). *Remote Sens.* 9 (3), 259–278. <https://doi.org/10.3390/rs9030259>.
- Lefsky, M.A., Cohen, W.B., Parker, G.G., Harding, D.J., 2002. Lidar remote sensing for ecosystem studies. *BioSci.* 52 (1), 19–30. [https://doi.org/10.1641/0006-3568\(2002\)052\[019:Lrsfes\]2.0.Co;2](https://doi.org/10.1641/0006-3568(2002)052[019:Lrsfes]2.0.Co;2).
- Lesiv, M., Laso Bayas, J.C., See, L., Duerauer, M., Dahlia, D., Durando, N., Hazarika, R., Kumar Sahariah, P., Vakolyuk, M., Blyshchyk, V., Bilous, A., Perez-Hoyos, A., Gengler, S., Prestele, R., Bilous, S., Akhtar, I.u.H., Singha, K., Choudhury, S.B., Chetri, T., Malek, Ž., Bungnamei, K., Saikia, A., Sahariah, D., Narzary, W., Danylo, O., Sturm, T., Karner, M., McCallum, I., Schepaschenko, D., Molchanova, E., Fraisl, D., Moorthy, I., Fritz, S., 2019. Estimating the global distribution of field size using crowdsourcing. *Glob. Change Biol.* 25 (1), 174–186. <https://doi.org/10.1111/gcb.14492>.
- Li, D.R., Zhang, G.F., Wu, Z.C., Yi, L.N., 2010. An Edge Embedded Marker-Based Watershed Algorithm for High Spatial Resolution Remote Sensing Image Segmentation. *IEEE T Image Process.* 19 (10), 2781–2787. <https://doi.org/10.1109/tip.2010.2049528>.
- Li, Y., Jiang, Y., Gao, L., Fan, Y., 2015. Fast mutual modulation fusion for multi-sensor images. *Optik.* 126 (1), 107–111. <https://doi.org/10.1016/j.jle.2014.08.136>.
- Li, Z., Ko, B., Choi, H.-J., 2019. Naive semi-supervised deep learning using pseudo-label. *Peer Peer Netw Appl.* 12 (5), 1358–1368. <https://doi.org/10.1007/s12083-018-0702-9>.
- Liang, M., Hu, X., 2015. Recurrent convolutional neural network for object recognition. *IEEE Conf. Comput. Vis. Pattern Recognit.* 3367–3375. <http://doi.org/10.1109/CVPR.2015.7298958>.
- Liu, D., Xia, F., 2010. Assessing object-based classification: advantages and limitations. *Remote Sens. Lett.* 1 (4), 187–194. <https://doi.org/10.1080/01431161003743173>.
- Loh, W.-Y., 2011. Classification and regression trees. *Wires Data Min Knowl.* 1 (1), 14–23. <https://doi.org/10.1002/widm.8>.
- Shelhamer, E., Long, J., Darrell, T., 2015. Fully Convolutional Networks for Semantic Segmentation. *IEEE T Pattern Anal.* 39 (4), 640–651. <https://doi.org/10.1109/TPAMI.2016.2572683>.
- Long, Y., Gong, Y., Xiao, Z., Liu, Q., 2017. Accurate Object Localization in Remote Sensing Images Based on Convolutional Neural Networks. *IEEE T Geosci. Remote Sens. Environ.* 55 (5), 2486–2498. <https://doi.org/10.1109/TGRS.2016.2645610>.
- Marmanis, D., Datcu, M., Esch, T., Stilla, U., 2016. Deep Learning Earth Observation Classification Using ImageNet Pretrained Networks. *IEEE Geosci. Remote Sens. Lett.* 13 (1), 105–109. <https://doi.org/10.1109/lgrs.2015.2499239>.
- Martin, D.R., Fowlkes, C.C., Malik, J., 2004. Learning to detect natural image boundaries using local brightness, color, and texture cues. *IEEE T Pattern Anal.* 26 (5), 530–549. <https://doi.org/10.1109/tpami.2004.1273918>.
- Martin, D.R., Malik, J., Patterson, D.A., 2002. An empirical approach to grouping and segmentation.
- Maselli, F., Rembold, F., 2001. Analysis of GAC NDVI data for cropland identification and yield forecasting in Mediterranean African countries. *Photogramm. Eng. Remote Sensing.* 67 (5), 593–602.
- Masoud, K.M., Persello, C., Tolpekin, V.A., 2020. Delineation of Agricultural Field Boundaries from Sentinel-2 Images Using a Novel Super-Resolution Contour Detector Based on Fully Convolutional Networks. *Remote Sens.* 12 (1), 59–74. <https://doi.org/10.3390/rs12010059>.
- Massey, R., Sankey, T.T., Yadav, K., Congalton, R.G., Tilton, J.C., 2018. Integrating cloud-based workflows in continental-scale cropland extent classification. *Remote Sens. Environ.* 219, 162–179. <https://doi.org/10.1016/j.rse.2018.10.013>.
- Matton, N., Canto, G., Waldner, F., Valero, S., Morin, D., Inglada, J., Arias, M., Bontemps, S., Koetz, B., Defourny, P., 2015. An Automated Method for Annual Cropland Mapping along the Season for Various Globally-Distributed Agrosystems Using High Spatial and Temporal Resolution Time Series. *Remote Sens.* 7 (10), 13208–13232. <https://doi.org/10.3390/rs71013208>.
- Milletari, F., Navab, N., Ahmadi, S.-A., 2016. V-Net: Fully Convolutional Neural Networks for Volumetric Medical Image Segmentation. *Fourth International Conference on 3D Vision.* 565–571. <http://doi.org/10.1109/3DV.2016.79>.
- Mittal, M., Verma, A., Kaur, I., Kaur, B., Sharma, M., Goyal, L.M., Roy, S., Kim, T.-H., 2019. An Efficient Edge Detection Approach to Provide Better Edge Connectivity for Image Analysis. *IEEE Access* 7, 33240–33255. <https://doi.org/10.1109/access.2019.2902579>.
- Möller, M., Lymburner, L., Volk, M., 2007. The comparison index: A tool for assessing the accuracy of image segmentation. *Int. J. Appl. Earth Obs.* 9 (3), 311–321. <https://doi.org/10.1016/j.jag.2006.10.002>.
- Mueller, M., Segl, K., Kaufmann, H., 2004. Edge- and region-based segmentation technique for the extraction of large, man-made objects in high-resolution satellite imagery. *Pattern Recogn.* 37 (8), 1619–1628. <https://doi.org/10.1016/j.patcog.2004.03.001>.
- Myint, S.W., Gober, P., Brazel, A., Grossman-Clarke, S., Weng, Q., 2011. Per-pixel vs. object-based classification of urban land cover extraction using high spatial resolution imagery. *Remote Sens. Environ.* 115 (5), 1145–1161. <https://doi.org/10.1016/j.rse.2010.12.017>.
- Narudin, F.A., Feizollah, A., Anuar, N.B., Gani, A., 2014. Evaluation of machine learning classifiers for mobile malware detection. *Soft. Comput.* 20 (1), 343–357. <https://doi.org/10.1007/s00500-014-1511-6>.
- Peña, J., Gutiérrez, P., Hernández-Martínez, C., Six, J., Plant, R., López-Granados, F., 2014. Object-Based Image Classification of Summer Crops with Machine Learning Methods. *Remote Sens.* 6 (6), 5019–5041. <https://doi.org/10.3390/rs6065019>.
- Persello, C., Bruzzone, L., 2010. A Novel Protocol for Accuracy Assessment in Classification of Very High Resolution Images. *IEEE T. Geosci. Remote Sens.* 48 (3), 1232–1244. <https://doi.org/10.1109/tgrs.2009.2029570>.
- Persello, C., Tolpekin, V.A., Bergado, J.R., de By, R.A., 2019. Delineation of agricultural fields in smallholder farms from satellite images using fully convolutional networks and combinatorial grouping. *Remote Sens. Environ.* 231, 111253. <https://doi.org/10.1016/j.rse.2019.111253>.
- Phalke, A.R., Özdogan, M., 2018. Large area cropland extent mapping with Landsat data and a generalized classifier. *Remote Sens. Environ.* 219, 180–195. <https://doi.org/10.1016/j.rse.2018.09.025>.
- Pittman, K., Hansen, M.C., Becker-Reshef, I., Potapov, P.V., Justice, C.O., 2010. Estimating Global Cropland Extent with Multi-year MODIS Data. *Remote Sens.* 2 (7), 1844–1863. <https://doi.org/10.3390/rs2071844>.
- Ronneberger, O., Fischer, P., Brox, T., 2015. U-Net: Convolutional Networks for Biomedical Image Segmentation. *Int. Conf. Med. Image Comput. Comput.-Assisted Intervention.* 3 (3), 234–241. https://doi.org/10.1007/978-3-319-24574-4_28.
- Rydberg, A., Borgefors, G., 2001. Integrated method for boundary delineation of agricultural fields in multispectral satellite images. *IEEE T. Geosci. Remote Sens.* 39 (11), 2514–2520. <https://doi.org/10.1109/36.964989>.
- Scott, G.J., England, M.R., Starms, W.A., Marcum, R.A., Davis, C.H., 2017. Training Deep Convolutional Neural Networks for Land-Cover Classification of High-Resolution Imagery. *IEEE Geosci. Remote S.* 14 (4), 549–553. <https://doi.org/10.1109/lgrs.2017.2657778>.

- Springenberg, J.T., Dosovitskiy, A., Brox, T., Riedmiller, M.A., 2015. Striving for Simplicity: The All Convolutional Net. Int. Conf. Learn. Representations. 2. <https://arxiv.org/abs/1412.6806v2>.
- Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Rabinovich, A., 2015. Going deeper with convolutions. IEEE Conf. Comput. Vis. Pattern Recognit. 7 (12), 1–9. <https://doi.org/10.1109/CVPR.2015.7298594>.
- Tollefson, J., 2011. Seven billion and counting. Nature 478 (7369), 300. <https://doi.org/10.1038/478300a>.
- Turker, M., Kok, E.H., 2013. Field-based sub-boundary extraction from remote sensing imagery using perceptual grouping. ISPRS J. Photogramm. 79, 106–121. <https://doi.org/10.1016/j.isprsjprs.2013.02.009>.
- Waldner, F., Diakogiannis, F.I., 2020. Deep learning on edge: Extracting field boundaries from satellite images with a convolutional neural network. Remote Sens. Environ. 245, 111741. <https://doi.org/10.1016/j.rse.2020.111741>.
- Watkins, B., van Niekerk, A., 2019. A comparison of object-based image analysis approaches for field boundary delineation using multi-temporal Sentinel-2 imagery. Comput. Electron. Agric. 158, 294–302. <https://doi.org/10.1016/j.compag.2019.02.009>.
- Xie, S., Tu, Z., 2015. Holistically-Nested Edge Detection. IEEE Int. Conf. Comput. Vis. 125, 3–18. <https://doi.org/10.1109/ICCV.2015.164>.
- Xiong, J., Thenkabail, P.S., Gumma, M.K., Teluguntla, P., Poehnelt, J., Congalton, R.G., Yadav, K., Thau, D., 2017. Automated cropland mapping of continental Africa using Google Earth Engine cloud computing. ISPRS J. Photogramm. 126, 225–244. <https://doi.org/10.1016/j.isprsjprs.2017.01.019>.
- Yan, L., Roy, D.P., 2014. Automated crop field extraction from multi-temporal Web Enabled Landsat Data. Remote Sens. Environ. 144, 42–64. <https://doi.org/10.1016/j.rse.2014.01.006>.
- Zhang, H., Cao, X., Li, Q., Zhang, M., Zheng, X., 2015. Research on Crop Identification Using Multi-temporal NDVI HJ Images. Remote Sens. Technol. Appl. 30 (2), 304–311. <https://doi.org/10.11873/J.ISSN.1004-0323.2015.2.0304>.
- Zhang, H., Li, Q., Liu, J., Du, X., Dong, T., McNairn, H., Champagne, C., Liu, M., Shang, J., 2018a. Object-based crop classification using multi-temporal SPOT-5 imagery and textural features with a Random Forest classifier. Geocarto Int. 33 (10), 1017–1035. <https://doi.org/10.1080/10106049.2017.1333533>.
- Zhang, Z., Liu, Q., Wang, Y., 2018b. Road Extraction by Deep Residual U-Net. IEEE Geosci. Remote S. 15 (5), 749–753. <https://doi.org/10.1109/lgrs.2018.2802944>.