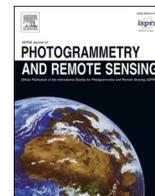


Contents lists available at [ScienceDirect](https://www.sciencedirect.com)

ISPRS Journal of Photogrammetry and Remote Sensing

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

Using a semantic edge-aware multi-task neural network to delineate agricultural parcels from remote sensing images

Mengmeng Li^{a,*}, Jiang Long^a, Alfred Stein^b, Xiaoqin Wang^a^a Key Lab of Spatial Data Mining & Information Sharing of Ministry of Education, Academy of Digital China (Fujian), Fuzhou University, 350108 Fuzhou, China^b Faculty of Geo-Information Science and Earth Observation (ITC), University of Twente, Enschede, The Netherlands

ARTICLE INFO

Keywords:

Agricultural parcel delineation
SEANet
Multi-task neural networks
Semantic edge-aware detection
Uncertainty weighted loss

ABSTRACT

This paper presents a semantic edge-aware multi-task neural network (SEANet) to obtain closed boundaries when delineating agricultural parcels from remote sensing images. It derives closed boundaries from remote sensing images and improves conventional semantic segmentation methods for the extraction of small and irregular agricultural parcels. SEANet integrates three correlated tasks: mask prediction, edge prediction, and distance map estimation. Related features learned from these tasks improve the generalizability of the network. We regard boundary extraction as an edge detection task and extract rich semantic edge features at multiple levels to improve the geometric accuracy of parcel delineation. Moreover, we develop a new multi-task loss that considers the uncertainty of different tasks. We conducted experiments on three high-resolution Gaofen-2 images in Shandong, Xinjiang, and Sichuan provinces, China, and on two medium-resolution Sentinel-2 images from Denmark and the Netherlands. Results showed that our method produced a better layout of agricultural parcels, with higher attribute and geometric accuracy than the existing ResUNet, ResUNet-a, R2UNet, and BsiNet methods on the Shandong and Denmark datasets. The total extraction errors of the parcels produced by our method were 0.214, 0.127, 0.176, 0.211, and 0.184 for the five datasets, respectively. Our method also obtains closed boundaries by one single segmentation, leading to superiority as compared with existing multi-task networks. We showed that it could be applied to images with different spatial resolutions for parcel delineation. Finally, our method trained on the Xinjiang dataset could be successfully transferred to the Shandong dataset with different dates and landscapes. Similarly, we obtained satisfactory results when transferring from the Denmark dataset to the Netherlands dataset. We conclude that SEANet is an accurate, robust, and transferable method for various areas and different remote sensing images. The codes of our model are available at https://github.com/long123524/SEANet_torch.

1. Introduction

Agricultural parcels serve as basic units for conducting agricultural practices and applications (Belgiu and Csillik 2018; Kocur-Bera 2019). Many regions and countries have established regional and national databases of agricultural parcels (Boryan et al. 2011; Teluguntla et al. 2018). Conventional collection of information on agricultural parcels mainly relies on field surveys and manually digitizing remote sensing images. Such approaches are time-consuming and labour-intensive. It is thus challenging to acquire updated information regarding agricultural parcels for large areas. Nowadays, a vast amount of remote sensing images covering various spatial, spectral, and temporal resolutions, is available to the public, promoting the demand for automatically

deriving agricultural parcels from these images (Jong et al. 2022; Persello et al. 2019; Tang et al. 2020).

In the past, many studies have been conducted on agricultural parcel delineation using remote sensing images (Masoud et al. 2020; Stein et al. 2016; Xu et al. 2019; Zhang et al. 2021). Delineating the boundaries of agricultural parcels has long been considered an edge detection problem, for which the Sobel and Canny edge filters are suited (Robb et al. 2020; Watkins and Van Niekerk 2019a). Edges thus detected are sensitive to high-frequency noise, resulting for instance in false edges. Post-processing, such as edge smoothing and connection, can improve the quality of detected edges (Rydberg and Borgefors 1999). Recent studies investigated using advanced edge detectors for parcel extraction. For example, Crommelinck et al. (2017) used a hierarchical edge detector

* Corresponding author.

E-mail addresses: mli@fzu.edu.cn (M. Li), 205527028@fzu.edu.cn (J. Long), a.stein@utwente.nl (A. Stein), wangxq@fzu.edu.cn (X. Wang).<https://doi.org/10.1016/j.isprsjprs.2023.04.019>

Received 8 November 2022; Received in revised form 10 April 2023; Accepted 21 April 2023

Available online 5 May 2023

0924-2716/© 2023 International Society for Photogrammetry and Remote Sensing, Inc. (ISPRS). Published by Elsevier B.V. All rights reserved.

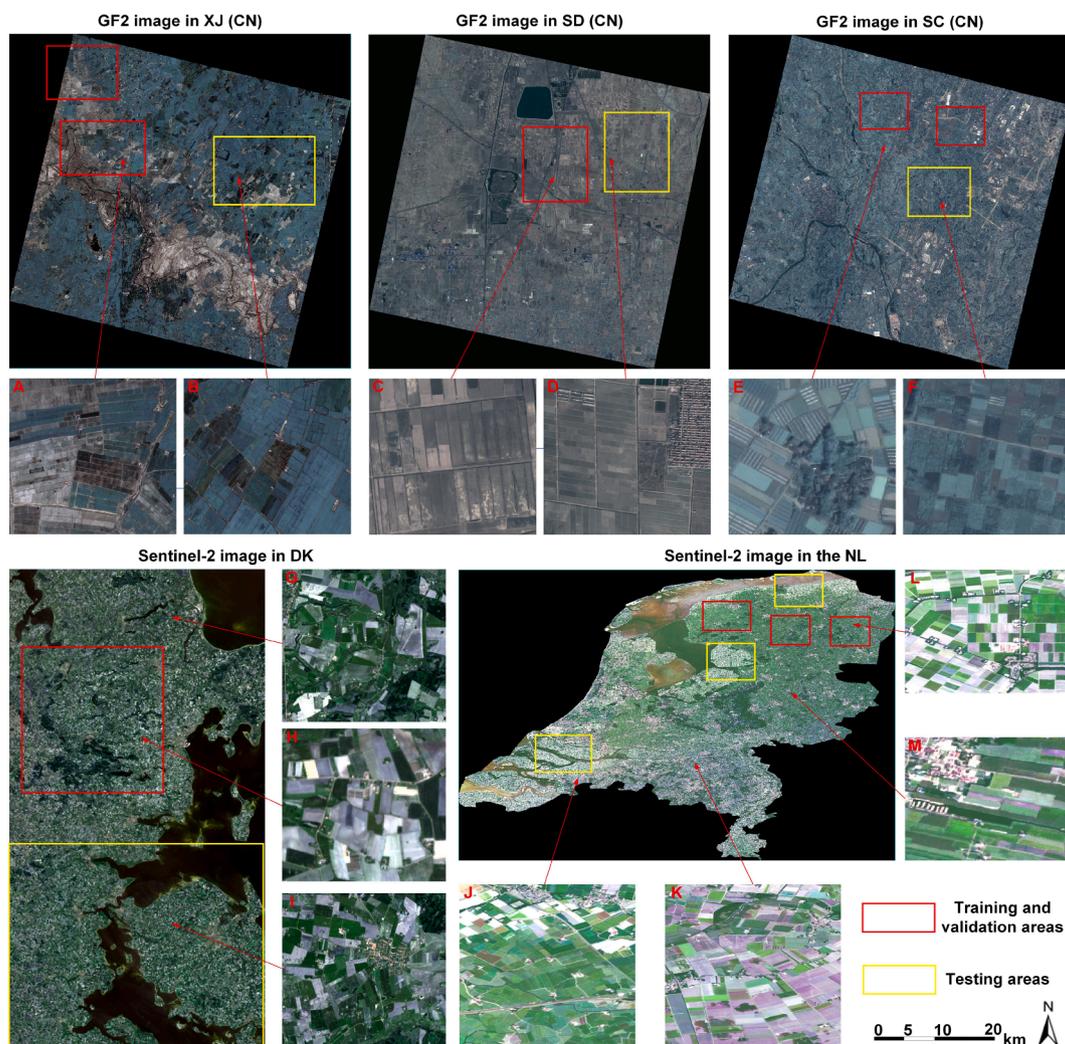


Fig. 1. Overview of the study areas including three areas in Shandong (SD), Xinjiang (XJ), and Sichuan (SC) provinces of China (CN), one area in Denmark (DK), and one area in the Netherlands (NL). A-M refers to agricultural parcel examples.

based upon the globalized probability of boundary to extract land parcels from fine-resolution unmanned aerial vehicle (UAV) images. These edge detection methods highlight the edge characteristics regarding high-frequency singles in intensity, color, and texture. So far, high-level semantic features of the region of interest enclosed by field boundaries are insufficiently exploited. These methods also tend to obtain unclosed boundaries for detected image regions, leading to a low degree of accuracy in extracting agricultural parcels.

An agricultural parcel identified from a remote sensing image, refers to an image region. To make full use of high-level features of image regions, object-based image analysis methods have been preferably used. They merge a group of adjacent pixels to form a homogenous image object, corresponding to image segmentation (Belgiu and Drăguț 2014). Many features regarding the spectral, textural, geometrical, and spatial properties of image objects can be extracted for identifying agricultural parcels, for which machine learning algorithms like random forest and support vector machine are commonly used (Cai et al. 2022; Garcia-Pedrero et al. 2017; Lebourgeois et al. 2017; Li et al. 2022). Lebourgeois et al. (2017) conducted a multi-resolution segmentation to obtain image objects from high-resolution images and used a random forest to extract agricultural parcels based upon spectral and textural features. A challenge in using object-based methods is to obtain suitable image objects by means of adequate image segmentation, as poor segmentation may increase the uncertainties of information extraction based upon objects (Hay et al. 2003). Moreover, typical errors generated

by edge-based (e.g., unclosed boundaries) and object-based methods (e.g., imperfect image objects) limit the use of these methods for parcel delineation over large areas (Watkins and Van Niekerk 2019b). In addition, hybrid methods combining edge detection and image segmentation have been investigated for delineating agricultural parcels (Cheng et al. 2020; Rydberg and Borgfors 2001; Watkins and Van Niekerk 2019b). Rydberg and Borgfors (2001) combined a gradient detector with clustering methods to delineate parcel boundaries. Mueller et al. (2004) combined a region-growing algorithm to merge ragged edges to improve the boundaries of delineated parcels.

Recently, deep neural networks have been increasingly used to extract agricultural parcels from remote sensing images own to their powerful abilities in extracting high-level semantic features, particularly using convolutional neural networks (CNNs) (Lu et al. 2022; Taravat et al. 2021; Zhang et al. 2020). Using a CNN for parcel extraction typically refers to semantic segmentation, which distinguishes pixels assigned to agricultural parcels from others. For instance, Persello et al. (2019) used a popular encoder-decoder model to extract parcel boundaries from high-resolution satellite images. Zhang et al. (2021) used a modified R2UNet to extract parcels from Sentinel-2 satellite images. Potlapally et al. (2019) used a double-branch network to extract agricultural parcels. Furthermore, multi-task neural networks that learn multiple correlated tasks through a joint representation have been used to improve parcel extraction (Waldner and Diakogiannis 2020; Waldner et al. 2021). Waldner and Diakogiannis (2020) developed a multi-task

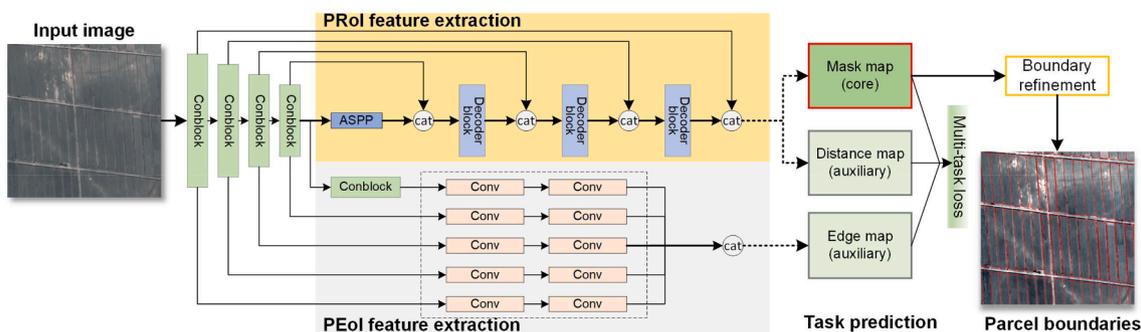


Fig. 2. Workflow of the proposed multi-task model for agricultural parcel delineation from remote sensing images. It consists of two feature extraction modules regarding parcel region of interest (PРоI) and parcel edge of interest (PЕoI). Parcel extraction refers to the prediction of mask map regularized by two auxiliary tasks, i.e., prediction of distance and edge maps. Parcel boundaries are further improved using a boundary refinement operation.

network (ResUNet-a) integrating a core task with two auxiliary tasks to optimize feature learning for agricultural parcels. Waldner et al. (2021) proposed a ResUNet-a variant, called FracTAL ResUNet, based upon fractal attention, to further improve parcel delineation for large areas. Long et al. (2022) developed a lightweight multi-task network (BsiNet) using a single encoder-decoder, achieving state-of-the-art extraction accuracy and efficiency. Nevertheless, an unsolved issue is that parcels extracted by existing methods inevitably have either unclosed boundaries, incomplete boundaries, or both. Previous studies have justified the use of edge information for parcel extraction, but they mainly extract edge information at the global (context) level, failing to integrate local-level edge information (Long et al. 2022). Relevant works unifying edge detection and semantic segmentation for enhancing information extraction using remote sensing images include Cheng et al. (2016), Marmanis et al. (2018), and Heidler et al. (2021). For instance, Heidler et al. (2021) combined edge detection and semantic segmentation to extract deep edge features to identify coastlines from Sentinel-1 images in the Antarctic. To the best of our knowledge, no previous studies have been carried out to consider local and global multi-level semantic edge information for agricultural parcel delineation.

This paper aims to develop a semantic segmentation network for agricultural parcel delineation from remote sensing images, with particular attention to extracting parcels with regularized and closed boundaries. We build a semantic edge-aware multi-task neural network, called SEANet, to enhance the extraction of local and global features regarding the edge and of thematic information regarding agricultural parcels. Compared with conventional CNNs, SEANet fully uses semantic edge features at both the local and global levels, improving the geometric accuracy of agricultural parcel delineation. The novelty of this study lies in the following:

- 1) It presents a semantic edge-aware multi-task neural network to effectively obtain closed-boundary agricultural parcels from remote sensing images;
- 2) It develops a multi-task loss function considering the uncertainty of different tasks into account to effectively balance the weights among various learning tasks;
- 3) It shows the transferability of the network to agricultural areas with similar or dissimilar agricultural parcel characteristics, even across different dates and regions.

The rest of the paper is structured as follows: Section 2 presents our study areas and datasets, and section 3 describes the proposed method. Experimental results and related analysis are given in section 4, while section 5 contains the discussion, and is followed by conclusions in section 6.

2. Study areas and datasets

We chose five study areas, i.e., three areas in Shandong, Xinjiang, and Sichuan provinces in China, one area in Denmark, and one area in the Netherlands, to evaluate our method (Fig. 1). The Shandong study area is situated in north of China, covering an area of 712.66 km². Most parcels in this area are regular, whereas the sizes vary from 3204 to 300000 m². For this area, we acquired a whole scene of Chinese Gaofen-2 (GF-2) satellite image on 20 December 2021, on which the agricultural land was observed in a fallow period. This image shows evident spectral heterogeneity. The Xinjiang study area is situated in northwestern China, with a whole scene of GF-2 image covering 772.72 km² acquired on 27 July 2021. The image shows that agricultural parcels have large sizes and regular shapes, and most of the parcels were planted with various crops, challenging the delineation of individual parcels. The Sichuan study area is situated in southwestern China, covering an area of 890.32 km². We also acquired a whole scene of GF-2 image for the Sichuan area on 21 March 2021. In contrast with the Xinjiang and Shandong areas, agricultural parcels in this area characterized by smallholder farms have small sizes and irregular shapes, forming fragmented patterns. It has been recognized as a challenging task to effectively delineate smallholder farm parcels from satellite images (Long et al. 2022). To preprocess the GF-2 images, we conducted *ortho*-rectification and Gram-Schmidt pan-sharpening, followed by a resampling operation to 1 m spatial resolution, for experiment convenience. The processed GF-2 images have a size of 27087 × 26310, 28206 × 27377, and 30764 × 28940 pixels for the Shandong, Xinjiang, and Sichuan areas, respectively. Moreover, we collected two medium-resolution Sentinel-2 datasets, with a spatial resolution of 10 m, one in Denmark and the other in the Netherlands. The Denmark dataset is publicly available from the European Union Land Parcel Identification System (LPIS) (<https://collections.eurodatacube.com/>). It comprises two cloudless Sentinel-2 images on 8 May 2016 with a true-color composite (10 m), located in the eastern part of Denmark (Rieke 2017), with 10982 × 20978 pixels covering an area of 20900 km². In this big dataset, agricultural parcels are found with a huge variety of sizes and shapes. The Netherlands dataset consists of five cloudless Sentinel-2 images with a true-color composite (10 m) downloaded from Google Earth Engine (GEE), covering nearly the whole country (i.e., 41528 km²) with 42845 × 31580 pixels. Each cloudless image was fused by GEE (Ghorbanian et al. 2020) using images ranging from 1 May to 1 October 2020. These images cover areas more than 64800 km², and are therefore adequate to evaluate the effectiveness of our method in various agricultural scenarios.

The ground-truth agricultural parcels for the GF-2 datasets were generated by manual delineation, while the ground-truth parcels for the Denmark and the Netherlands Sentinel-2 datasets were obtained from LPIS and from <https://www.pdok.nl/>, respectively. To train and test our network, subset areas were used, as shown in red and yellow boxes in

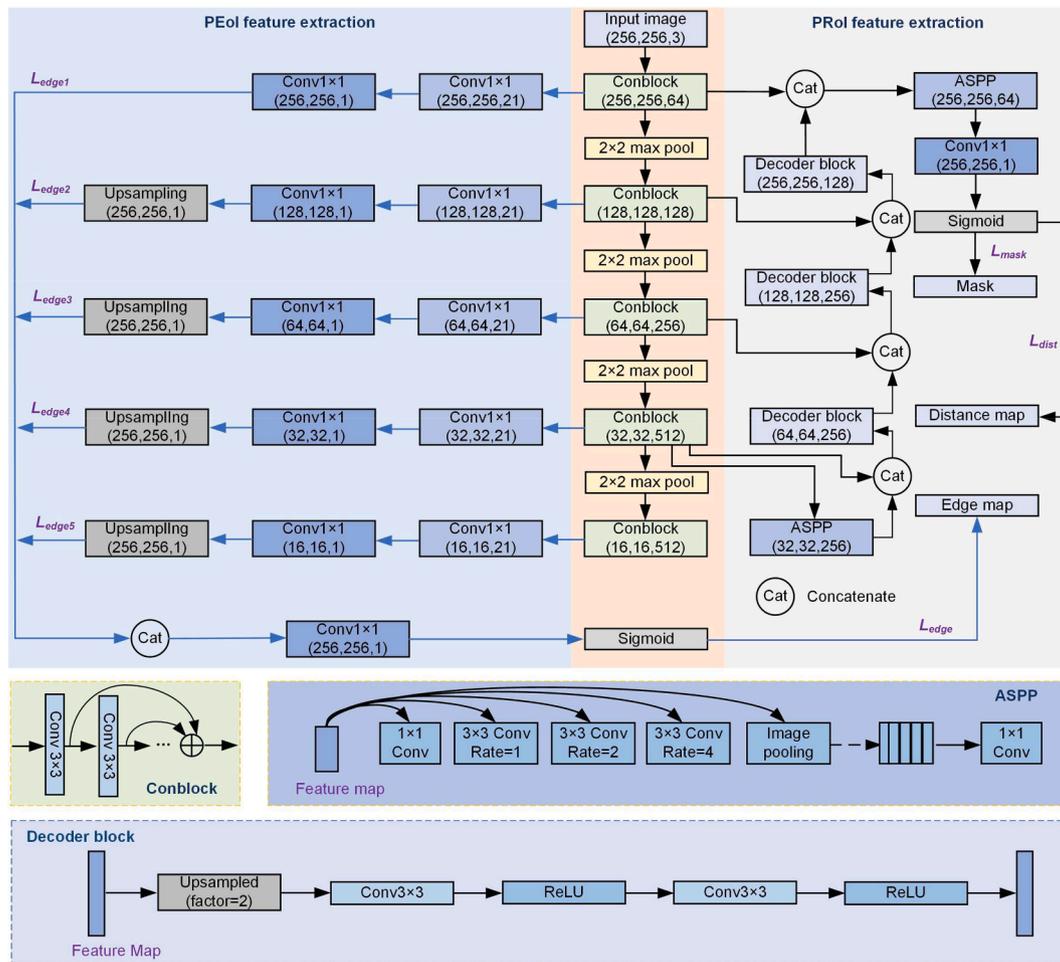


Fig. 3. Architecture of the proposed SEANet with two main components, corresponding PRoI and PEoI feature extractions. The numbers in parentheses refer to the height, width, and dimensionality of output channels.

Moreover, training and validation areas were partitioned into training and validation sets using a 0.8:0.2 random split.

3. Methods

Fig. 2 illustrates the workflow of our method for agricultural parcel delineation from remote sensing images. First, image pre-processing including *ortho*-rectification and pansharpening was conducted for input images. Next, we built a semantic edge-aware multi-task neural network (SEANet) with three tasks, corresponding to predicting mask, distance, and edge feature maps (Fig. 2). The mask map prediction extracted deep semantic features regarding the parcel region of interest (PRoI). We incorporated an atrous spatial pyramid pooling (ASPP) module to derive the context information of PRoI at multi-scales to enhance the identification of parcels with various shapes and sizes. The distance map prediction extracted coarse-grained geometric information to constrain the feature learning of mask task. The edge map prediction extracted multi-level edge features regarding the parcel edge of interest (PEoI). To improve the applicability of the proposed method for different areas, we developed a multi-task loss considering the uncertainty of different tasks. Last, we conducted a refinement operation based upon morphological thinning and Douglas-Peucker algorithms to further improve the extracted parcel boundaries.

3.1. Semantic edge-aware multi-task neural network (SEANet)

Most semantic segmentation networks for parcel delineation focus

on extracting high-level image features regarding PRoI, while few integrate features of the corresponding PEoI, and even less consider the integration of edge features at multiple levels. It is thus difficult to obtain closed and complete boundaries in one segmentation processing step using those networks, particularly in agricultural areas with parcels of various shapes and irregular boundaries. To deal with this challenge, we developed an edge-aware semantic segmentation network in a multi-task learning framework to jointly learn multi-level features of both the PRoI and PEoI.

The architecture of our network is illustrated in Fig. 3. It consists of two main components, corresponding to PRoI and PEoI feature extractions. We used an encoder-decoder structure for PRoI feature extraction. An encoder was used to learn hidden representations of agricultural parcel features, while a decoder further extracted high-level image features associated with a specific task (e.g., mask or distance map prediction). More specifically, the encoder, based upon a VGG16 backbone pre-trained on ImageNet, was comprised of multiple Conblocks. Each Conblock stacks a series of 3 × 3 convolutions, and was used to obtain rich local and global context features at multiple levels. Then, a 2 × 2 max-pooling was applied for reducing feature dimensions. The risk of gradient disappearing and exploding increases when using a deeper neural network, while a shallow network fails to learn adequate features (He et al. 2016). Therefore, this study constructed four Conblocks to extract high-level parcel features. Moreover, we added a multi-scale feature fusion module, i.e., an ASPP, at the end of the fourth Conblock (Fig. 3). The ASPP was comprised of a 1 × 1 convolution, three atrous convolutions using different dilated rates, and a 2 × 2 average pooling,

to enlarge the receptive field of convolution's kernel size to capture the context information of PROI at a large neighborhood (Zhao et al. 2017). The use of ASPP learned distinct features with respect to PROI, and thus improved the identification of parcels with varied sizes. It is different from using an attention module that focuses on learning features relevant to the target object (Vaswani et al. 2017). The decoder part is composed of three blocks. Each block first conducts an upsampling by a factor of 2, and then uses two 3×3 convolutions and a rectified linear unit (ReLU) to refine the encoded features. Likewise, an ASPP module, followed by a 1×1 convolution and a Sigmoid function, predicts the mask map.

In a multi-task network, using distance information (to parcel mask boundaries) improves semantic segmentation in the geometric aspect (Diakogiannis et al. 2020; Long et al. 2022; Ma et al. 2020; Murugesan et al. 2019). Likewise, we also formulated a sub-task regarding distance prediction based upon quasi-Euclidean distance transformation (Long et al. 2022). Here the distance prediction was treated as an auxiliary task, by which coarse-grained boundary information was exploited to constrain the feature extraction regarding the core task, i.e., mask prediction.

To further incorporate fine-grained and multi-level boundary information into SEANet, we constructed an edge extraction module, corresponding to PEOI feature extraction. In this module, we made full use of the image features derived from Conblocks. Yu et al. (2017) stated that decoder structures might be unsuited to extracting edge semantic features. Here we constructed a stack of convolution blocks, with 1×1 convolutions and upsampling operations, to derive edge feature maps at multiple levels (Fig. 3). Moreover, inspired by Liu et al. (2019), we added an extra Conblock to extract richer edge features regarding agricultural parcels. The derived multi-level edge maps were then fused by a concatenation operation, followed by a 1×1 convolution. Construction of the edge map by the PEOI module was treated as another auxiliary task of SEANet.

3.2. Construction of a multi-task loss considering uncertainties

3.2.1. Loss functions for different tasks

Mask prediction refers to a classification task. In this study, we combined a binary cross-entropy (BCE) l_{bce} and a dice loss l_{dice} to deal with class imbalance and training instability (Diakogiannis et al. 2020; Milletari et al. 2016). Given a ground-truth mask image I_M with N pixels $x_M^i (i = 1, \dots, N)$, let y_M^i be the label value of pixel $x_M^i \in \{0, 1\}$, and let \hat{y}_M^i be the predicted mask label after a Sigmoid function. The BCE loss l_{bce} is then defined as:

$$l_{bce} = -\frac{1}{N} \sum_{i=1}^N (y_M^i \log \hat{y}_M^i + (1 - y_M^i) \log (1 - \hat{y}_M^i)) \quad (1)$$

Let Y_M and \hat{Y}_M be the label vectors of the ground-truth mask image I_M and the corresponding predicted mask image \hat{I}_M , $|Y_M|$ and $|\hat{Y}_M|$ be the norm of Y_M and \hat{Y}_M , and $|Y_M \cap \hat{Y}_M|$ be their intersection norm. The dice loss l_{dice} is then defined as:

$$l_{dice} = 1 - \frac{2|Y_M \cap \hat{Y}_M| + \epsilon}{|Y_M| + |\hat{Y}_M| + \epsilon} \quad (2)$$

where ϵ is a small value equal to 10^{-5} for avoiding zero value in the denominator of l_{dice} .

Based upon l_{bce} and l_{dice} losses, we obtained the mask prediction loss l_M :

$$l_M = 0.5 \cdot l_{bce} + l_{dice}. \quad (3)$$

Distance prediction refers to a regression task. We used the mean square error (MSE) to obtain the distance loss l_D based upon the ground-truth distance value y_D^i that computes the shortest distance of pixel x_D^i to

the mask boundary of ground-truth image I_B , and the predicted distance value \hat{y}_D^i being the shortest distance to the mask boundary of the predicted image. We obtained l_D by:

$$l_D = \frac{1}{N} \sum_{i=1}^N (y_D^i - \hat{y}_D^i)^2. \quad (4)$$

Edge prediction refers to an edge classification task. The numbers of pixels on an image belonging to the edge and non-edge classes are usually highly imbalanced. Such imbalance tends to guide a model to learn features that fail to effectively represent the edge class with a smaller number of positive samples. We improved the BCE loss by introducing two coefficients α and β to weigh the losses of edge (positive) and non-edge (negative) samples Y_E^+ and Y_E^- (Liu et al. 2019). Let $y_E^i \in \{0, 1\}$ be the label of an edge pixel x_E^i in the ground-truth boundary image I_B , and \hat{y}_E^i be the predicted edge label. Thus, the edge loss l_E^i of each edge pixel is derived by:

$$l_E^i = \begin{cases} -\alpha \cdot \log(1 - \hat{y}_E^i) & \text{if } y_E^i = 0, \\ -\beta \cdot \log \hat{y}_E^i & \text{if } y_E^i = 1. \end{cases} \quad (5)$$

with coefficients α and β defined as:

$$\alpha = \gamma \frac{|Y^+|}{|Y^+ + Y^-|} \quad (6)$$

$$\beta = \frac{|Y^-|}{|Y^+ + Y^-|} \quad (7)$$

where γ is a parameter balancing the amount of positive and negative samples, which depends on training data and equals 1.1 as recommended by (Liu et al. 2019).

We formed the total edge loss l_E by aggregating the individual loss of edge extraction after an upsampling layer at multi-levels $l_{E,k}^i (k = 1, \dots, 5)$ and their fused loss \tilde{l}_E^i :

$$l_E = \frac{1}{N} \sum_{i=1}^N \left(\sum_{k=1}^5 l_{E,k}^i + \tilde{l}_E^i \right) \quad (8)$$

where N indicates the number of pixels of the edge image I_E .

3.2.2. Fusing multi-task losses considering uncertainties

To fuse multi-task losses, traditional methods usually combine these losses linearly to form a total loss (Long et al. 2022; Waldner and Diakogiannis 2020). It is, however, hard to set suitable weights for the losses of different tasks. A common practice is to do this by trial-and-error. In this paper, we formulated a multi-task loss function to adaptively adjust the weights of different losses considering uncertainty associated with different tasks, namely task-dependent uncertainty. The task-dependent uncertainty is also referred to as homoscedastic uncertainty (Kendall et al. 2018). It keeps constant for input data, but varies between tasks. The weighting of the losses between different tasks can be achieved by modeling these uncertainties.

Following Kendall and Gal (2017), we formulated task-dependent uncertainties as probabilities over the outputs of a neural network. Let $f^W(x)$ be the output of the network with weight W given input x , and $p(y|f^W(x))$ be the model likelihood. For a regression task (i.e., distance prediction), the model likelihood was defined as a Gaussian function with mean and variance determined by the model output with an observation noise σ , respectively:

$$p(y|f^W(x)) = \mathcal{N}(f^W(x), \sigma^2) \quad (9)$$

For a classification task (e.g., mask prediction), a Softmax function was used to squash the model output. The classification is considered following a Boltzmann distribution in Kendall et al. (2018):

$$p(y|f^W(x)) = \text{Softmax}(f^W(x), \sigma^2). \quad (10)$$

For the model with two outputs, i.e., a regression output y_1 and a classification output y_2 , we obtained a joint probability:

$$p(y_1, y_2 | f^W(x)) = p(y_1 | f^W(x)) \cdot p(y_2 | f^W(x)) \\ = \mathcal{N}(y_1; f^W(x), \sigma_1^2) \cdot \text{Softmax}(y_2; f^W(x), \sigma_2^2). \quad (11)$$

Inference of the joint probability was transformed to obtain a minimization objective, namely loss $L(W, \sigma_1, \sigma_2)$ defined by:

$$L(W, \sigma_1, \sigma_2) = -\log p(y_1, y_2 = c | f^W(x)) \\ = -\log \mathcal{N}(y_1; f^W(x), \sigma_1^2) \cdot \text{Softmax}(y_2 = c; f^W(x), \sigma_2^2) \\ \approx \frac{1}{2\sigma_1^2} L_1(W) + \frac{1}{\sigma_2^2} L_2(W) + \log \sigma_1 \sigma_2 \quad (12)$$

where $L_1(W) = \|y_1 - f^W(x)\|^2$ refers to the Euclidean loss of y_1 , $L_2(W) = -\log[\text{Softmax}(y_2, f^W(x))]$ is the cross-entropy loss of y_2 , and c is the entity of the vector $f^W(x)$. We considered noise parameters σ_1 and σ_2 as relative weights balancing the loss of $L_1(W)$ and $L_2(W)$ given the input data. Increasing the value of σ_1 (or σ_2) leads to a decrease of the weight of $L_1(W)$ (or $L_2(W)$). Moreover, $\log \sigma_1 \sigma_2$ acts as a regularizer avoiding the noise σ_1 or σ_2 increasing too much during model training. In the training process, we replaced $\log \sigma_1 \sigma_2$ with $\log(1 + \sigma_1^2) + \log(1 + \sigma_2^2)$ to maintain positive regularization values (Liebel and Körner 2018).

The edge map derived from the edge prediction task was fused by several fallow edge cues at multiple levels. These fallow edge cues may introduce additional observation noise, leading to a relatively small weight for the edge prediction task if using the uncertainty weighting method above. To formulate the total multi-task loss L_{total} , we therefore added the edge loss L_E to the mask and distance losses considering uncertainty weighting $L(W, \sigma_1, \sigma_2)$:

$$L_{total} = L(W, \sigma_1, \sigma_2) + L_E \\ = \frac{1}{2\sigma_1^2} L_1(W) + \frac{1}{\sigma_2^2} L_2(W) + \log(1 + \sigma_1^2) + \log(1 + \sigma_2^2) + L_E. \quad (13)$$

3.3. Boundary refinement

We performed a boundary refinement to further improve the boundaries of extracted parcels. We first conducted a morphological thinning operation to remove small boundaries, and then used the Douglas-Peucker method (with default parameter settings) to regularize extracted boundaries (Wei et al. 2019).

3.4. Comparison with existing methods for agricultural parcels delineation

To evaluate the performance of our method, we compared it with four existing methods for agricultural parcel delineation, corresponding to two single-task networks R2UNet (Zhang et al. 2021), ResUNet (Taravat et al. 2021; Zhang et al. 2018), and two multi-task networks, i.e., ResUNet-a (Waldner and Diakogiannis 2020) and BsiNet (Long et al. 2022). Moreover, we created a SEANet variant that formulated the multi-task loss considering uncertainty weighting for all three tasks, labeled as SEANet_{MDE}. These five methods are described below.

R2UNet is an encoder-decoder network with recurrent residual blocks. It improves feature representation by recurrently accumulating semantic features of multi-scales, and has been recently used to delineate agricultural parcels from Sentinel-2 satellite images (Zhang et al. 2021).

ResUNet is a U-shaped network with residual blocks that retains a good segmentation performance while reducing the number of parameters (Zhang et al. 2018).

ResUNet-a is an early multi-task neural network that has been

successfully used to extract agricultural parcels from Sentinel-2 satellite images (Waldner and Diakogiannis 2020).

BsiNet is a recently proposed multi-task network based upon Psi-Net that has reached state-of-the-art accuracy in agricultural parcel delineation from high-resolution satellite images (Long et al. 2022).

SEANet_{MDE} is a variant of the proposed SEANet, with the difference that it formulates a multi-task loss L_{total} considering uncertainty weighting for all three tasks (Kendall et al. 2018).

3.5. Accuracy assessment and performance evaluation

We used both pixel-based attribute and object-based geometric measures to assess the accuracy of extracted agricultural parcels. The pixel-based attribute measures refer to precision (P), recall (R), overall accuracy (OA), Intersection over Union (IoU), and F1-score. The P, R, IoU, and OA measures are defined as follows:

$$P = \frac{TP}{TP + FP} \quad (14)$$

$$R = \frac{TP}{TP + FN} \quad (15)$$

$$IoU = \frac{TP}{TP + FP + FN} \quad (16)$$

$$OA = \frac{TP + TN}{TP + FN + FP + TN} \quad (17)$$

where TP , TN , FP , and FN indicate true positive, true negative, false positive, and false negative, respectively. TP and TN indicate the number of pixels correctly identified as agricultural parcels and non-agricultural parcels, while FN and FP indicate the number of pixels mis-identified as non-agricultural parcels (i.e., omissions) and agricultural parcels (i.e., mistakes). The F1 score is derived from P and R:

$$F1 = 2 \times \frac{P \times R}{P + R} \quad (18)$$

The object-based geometric measures are based upon under-segmentation and over-segmentation error measures (Persello and Bruzzone 2009). Suppose O_i is the reference object that overlaps the largest area with the classified object S_i , $i = 1, 2, \dots, m$, where m is the number of classified objects. Let $\text{area}(S_i)$ and $\text{area}(O_i)$ be the area of S_i and the O_i , respectively, and $\text{area}(S_i \cap O_i)$ be their overlapping area. We obtained an over-classification $OC(S_i)$ error and an under-classification $UC(S_i)$ error:

$$OC(S_i) = 1 - \frac{\text{area}(S_i \cap O_i)}{\text{area}(O_i)}, \quad (19)$$

$$UC(S_i) = 1 - \frac{\text{area}(S_i \cap O_i)}{\text{area}(S_i)}. \quad (20)$$

A total-classification $TC(S_i)$ error was then obtained based upon $OC(S_i)$ and $UC(S_i)$:

$$TC(S_i) = \sqrt{\frac{OC(S_i)^2 + UC(S_i)^2}{2}}. \quad (21)$$

Moreover, we derived three global error measures, corresponding to GOC, GUC, and GTC, based upon $OC(S_i)$, $UC(S_i)$, and $TC(S_i)$ measures:

$$GOC = \sum_{i=1}^m \left(OC(S_i) \times \frac{\text{area}(S_i)}{\sum_{i=1}^m \text{area}(S_i)} \right), \quad (22)$$

$$GUC = \sum_{i=1}^m \left(UC(S_i) \times \frac{\text{area}(S_i)}{\sum_{i=1}^m \text{area}(S_i)} \right), \quad (23)$$

Table 1

Evaluation of extracted results by using SEANet and SEANet without PEoI module (i.e., SEANet_{noPEoI}) on the Shandong (SD) GF-2 and Denmark (DK) Sentinel-2 images.

Methods	IoU (%)	F1-score (%)	GOC	GUC	GTC
SEANet (SD)	87.34	93.15	0.115	0.218	0.214
SEANet _{noPEoI} (SD)	80.06	87.49	0.092	0.304	0.259
SEANet (DK)	76.24	85.84	0.226	0.130	0.211
SEANet _{noPEoI} (DK)	74.52	84.04	0.264	0.114	0.225

$$GTC = \sum_{i=1}^m \left(TC(S_i) \times \frac{\text{area}(S_i)}{\sum_{i=1}^m \text{area}(S_i)} \right). \quad (24)$$

3.6. Implementation details

We cropped the satellite images of the study areas and the corresponding ground-truth images into a set of smaller image tiles with a size of 256×256 pixels, with an overlap rate of 25% according to experience. To improve the generalizability of SEANet, we conducted data augmentation including horizontal and vertical flips, mix-up, and color jittering. Moreover, we discarded images having agricultural pixels less than 20% to reduce the proportion of negative samples. In total, we obtained 5676, 5714, 4251, 5706, and 7254 agricultural parcel samples

in Shandong, Xinjiang, Sichuan, Denmark, and the Netherlands training and validation areas, respectively. We then randomly split these samples into training and validation datasets according to 0.8:0.2. The testing samples were collected from the test areas of five study areas (see yellow boxes in Fig. 1), including 437, 748, 346, 1441, and 1041 samples for Shandong, Xinjiang, Sichuan, Denmark, and the Netherlands datasets. We used the Apollo optimizer with initial learning rates of 10^{-2} and weight decay of 10^{-4} to train mask prediction and distance map estimation tasks, while the Stochastic Gradient Descent (SGD) optimizer with initial learning rates of 10^{-8} and weight decay of 10^{-4} to train edge prediction task. We also used a weight initialization strategy referred to by Liu et al. (2019) to initial the parameters of the proposed model. We followed the recommended loss and recipes from existing studies to train BsiNet, ResUNet-a, ResUNet, and R2UNet. These experiments using a NVIDIA GeForce RTX 3090 GPU and trained with 100 epochs using a batch size of 8.

4. Results

4.1. Ablation study for the PEoI module

We conducted an ablation study to evaluate the effectiveness of the proposed PEoI module. We removed the PEoI module from the proposed SEANet (labeled as SEANet_{noPEoI}), and compared it with the original

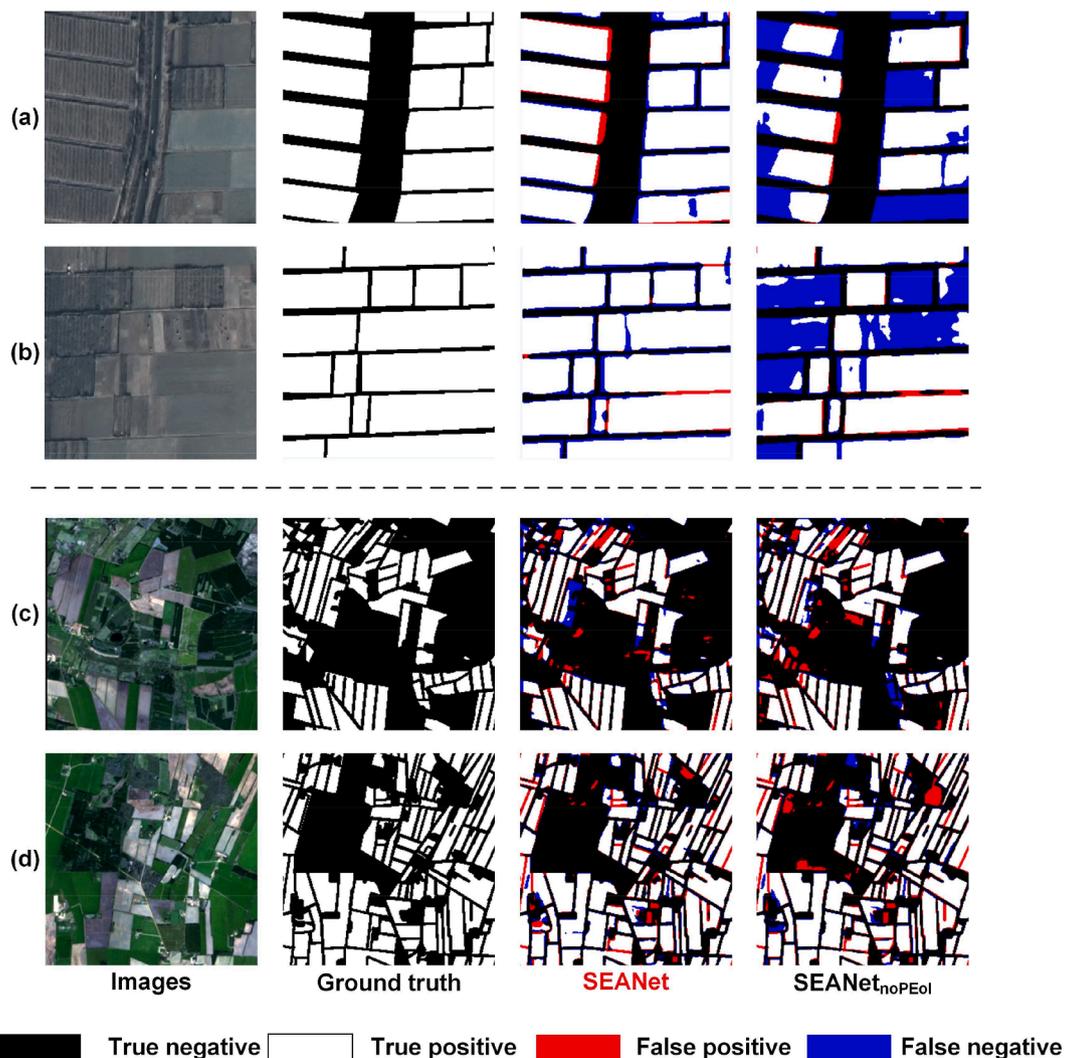


Fig. 4. Examples of agricultural parcels delineated by SEANet and SEANet without PEoI (i.e., SEANet_{noPEoI}). (a-b) and (c-d) refer to examples from the Shandong GF-2 and Denmark Sentinel-2 images, respectively.

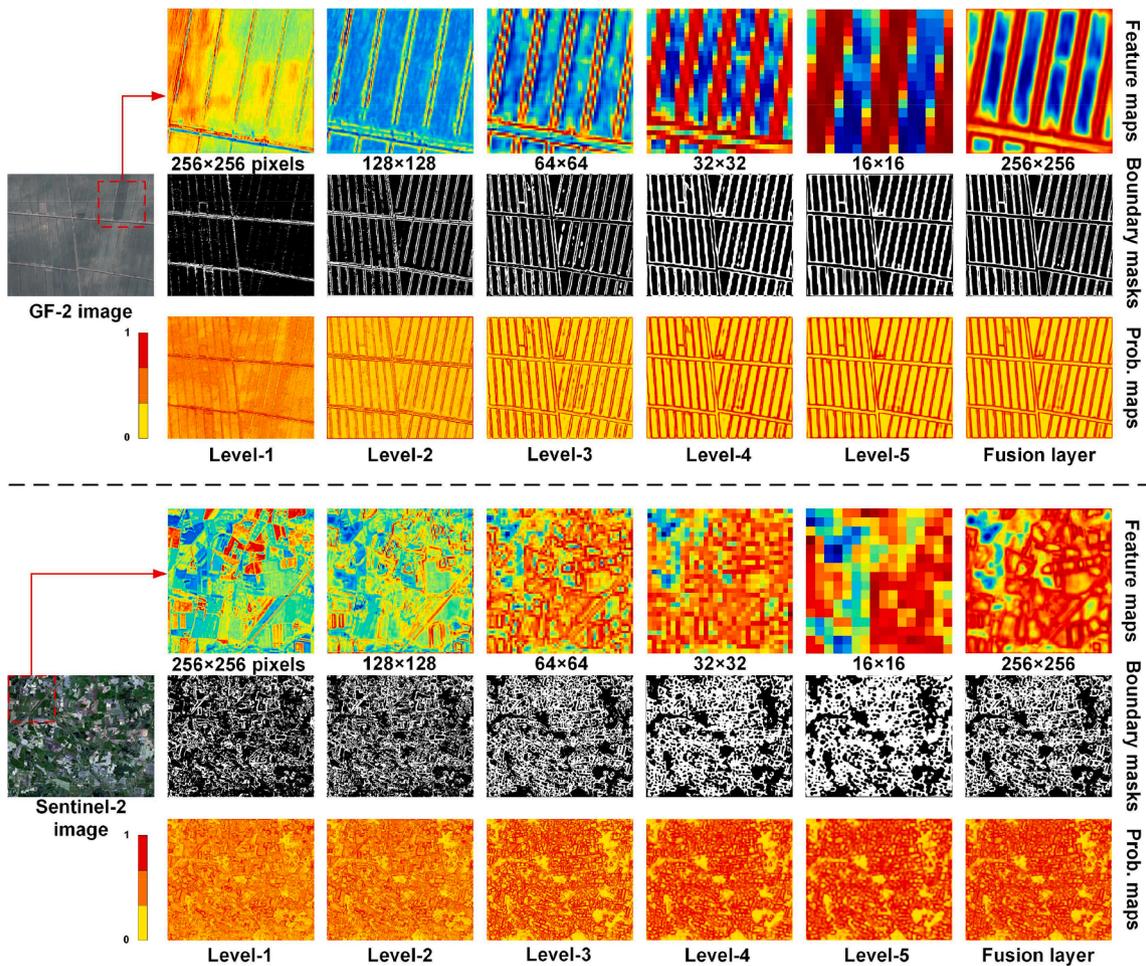


Fig. 5. Demonstration of extracted parcel boundary masks and feature maps using SEANet at different levels on the Shandong GF-2 (top) and Denmark Sentinel-2 images (bottom). Prob. refers to probability.

SEANet for parcel delineation using the Shandong GF-2 and Denmark Sentinel-2 images. Table 1 compares the accuracies of extracted results by SEANet and SEANet_{noPEoI} on the two images. This table shows that both the attribute and geometric accuracies of the extracted results by SEANet_{noPEoI} (without PEoI module) are evidently lower than that of the original SEANet on the two images, implying the effectiveness of the proposed PEoI module. Fig. 4 displays the delineated parcels by SEANet and SEANet_{noPEoI} on the two images. This figure shows that both false positives and false negatives were evidently reduced in the extraction results when using the proposed PEoI module. Last, we visualized the learned edge features and extracted boundary masks using SEANet (Fig. 5). The results indicate that the proposed PEoI module retained well-edge semantics at different levels.

4.2. Agricultural parcel delineation on the Shandong GF-2 image

Fig. 6 shows the results of agricultural parcel delineation by the proposed method (i.e., SEANet) on the high-resolution Shandong GF-2 image. This figure shows that the agricultural parcels obtained by the proposed method have a high agreement with the ground truth for both densely arranged and regular agricultural parcels. Benefitting from the powerful extraction of edge semantics, the obtained parcel instances have clear boundaries and little noise. We compared our method with five different methods in the same testing area of the Shandong dataset (i.e., the yellow box in Fig. 1) using pixel-based attribute measures (Table 2). Results show that SEANet obtained the highest accuracy for precision (P), F1-score, and mean intersection over union (IoU)

measures. It outperforms the-state-of-art BsiNet and ResUNet-a models by up to 2.98% and 5.9% for the IoU measure. By contrast, SEANet_{MDE}, with a multi-task loss incorporating the uncertainties of all tasks, failed to further improve the accuracy of parcel delineation.

We next visualized the extraction errors in Fig. 7. This figure clearly shows that our method achieved the lowest segmentation errors than others, corresponding to the fewest mistakes and omissions. It performs better than the recent BsiNet, which is the second-best method, implying that incorporating multi-level edge semantic information further improves parcel delineation. Comparing with single-task networks (i.e., ResUNet and R2UNet), multi-task neural networks (i.e., SEANet, SEANet_{MDE}, BsiNet, and ResUNet-a) generally obtained better results.

We further compared the extraction results by our method with other five methods using geometric measures (Table 3). This table shows that our method produced the best results with the lowest GTC error of 0.214. Results showed the superiority of our method in capturing robust edge information, and thus reducing geometric errors. We also see that ResUNet-a produced the highest GTC error of 0.396. This implies that ResUNet-a, developed based on coarse-resolution Sentinel-2 images, may be unsuitable for extracting parcels from high-resolution remote sensing images. Moreover, ResUNet yielded the highest GUC error of 0.303, while R2UNet produced the highest GOC error of 0.481. This implies that conventional convolution neural networks, e.g., those from UNet family, still have difficulty in accurately capturing the location and shape of parcels, leading to higher segmentation errors.

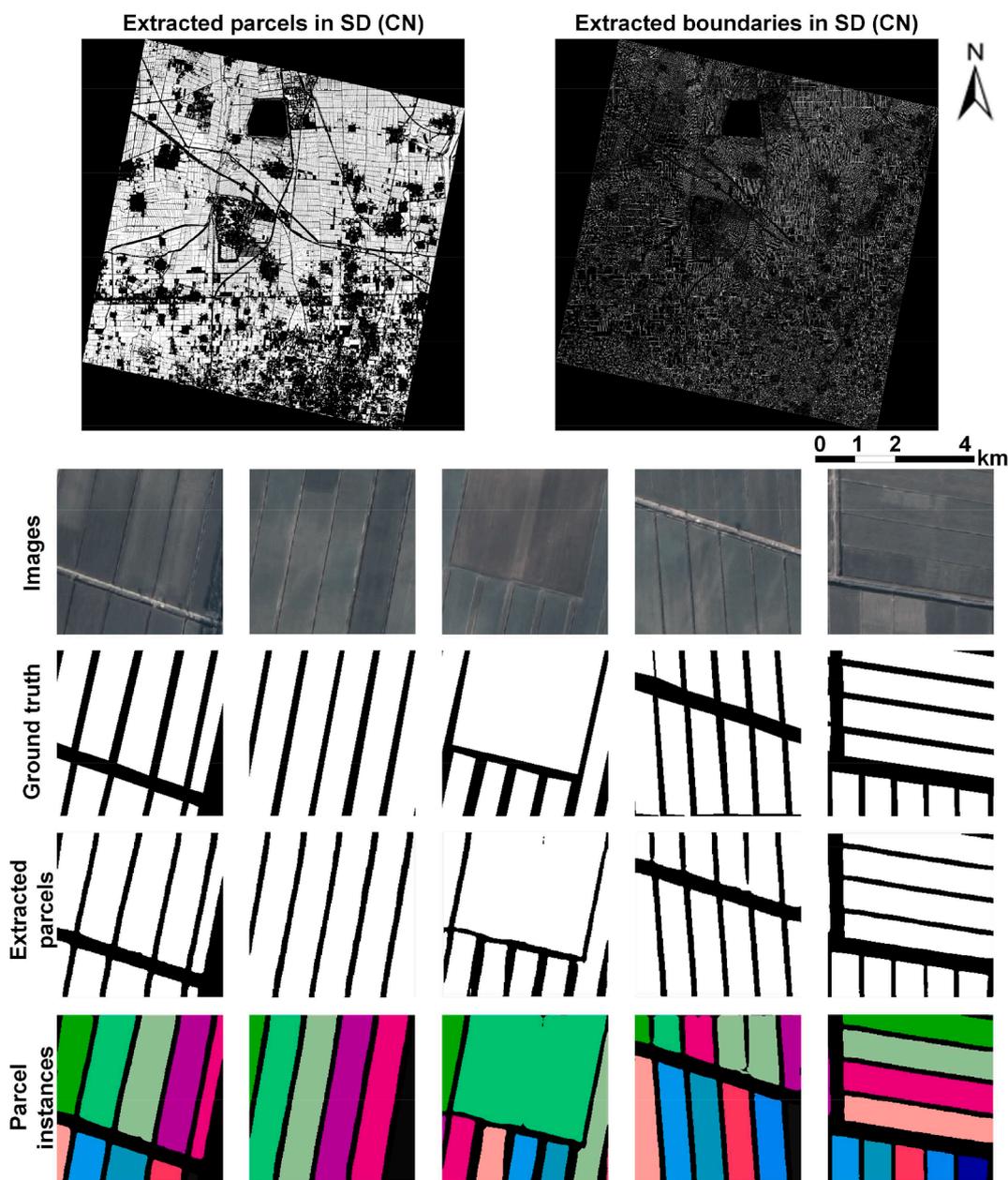


Fig. 6. Extracted agricultural parcels by SEANet on the Shandong GF-2 image, China (CN).

Table 2

Evaluation of extracted results by different methods on the Shandong GF-2 image using pixel-based attribute measures, i.e., precision (P), recall (R), overall accuracy (OA), F1-score, mean intersection over union (IoU).

Methods	P (%)	R (%)	OA (%)	F1-score (%)	IoU (%)
SEANet	97.74	88.98	91.54	93.15	87.34
SEANet _{MDE}	89.87	92.20	91.67	91.02	85.03
BsiNet	92.33	89.25	91.55	90.76	84.36
ResUNet-a	95.49	82.94	88.53	88.77	81.44
ResUNet	96.77	81.96	87.76	88.75	80.70
R2UNet	93.26	77.28	81.98	84.52	73.47

4.3. Boundary refinement of delineated parcels

To further improve the boundaries of extracted parcels, we applied morphological thinning and Douglas-Peucker operations to refine the extracted boundaries. The obtained results can be stored in a vector

format such as shapefile. We compared the results of the refined SEANet with those without using boundary refinement in a Shandong testing area (Fig. 8). This figure shows that the refinement removed small edges, and regularized ragged and incompleted boundaries to form smooth and closed boundaries (yellow boxes in Fig. 8). We also evaluated these delineation results using both attribute and geometric measures (Table 4). Clearly, compared with SEANet, the refined SEANet further improved delineation accuracy in both attribute and geometric aspects.

To make fair comparisons between our method and other existing methods, we directly applied SEANet, without boundary refinement, to delineate parcels on more datasets in the following experiments.

4.4. Agricultural parcel delineation on the Denmark Sentinel-2 image

To validate the effectiveness of SEANet on medium-resolution satellite images, we conducted experiments on the Denmark Sentinel-2 image. Fig. 9 displays the extraction results. This figure shows that our method obtains excellent results on the Denmark dataset, in which

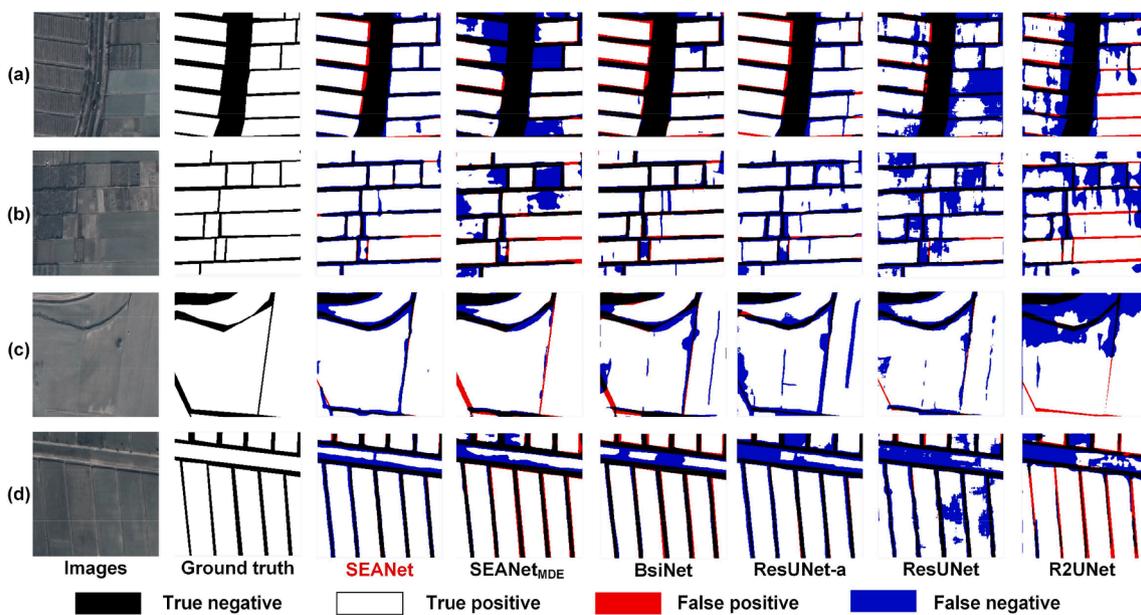


Fig. 7. Examples of delineated parcels using different methods on the Shandong GF-2 image.

Table 3

Evaluation of extracted results by different methods on the Shandong GF-2 image using object-based geometric measures.

Methods	GOC	GUC	GTC
SEANet	0.115	0.218	0.214
BsiNet	0.040	0.302	0.229
SEANet _{MDE}	0.217	0.180	0.253
ResUNet	0.107	0.303	0.265
R2UNet	0.481	0.093	0.372
ResUNet-a	0.359	0.266	0.396

Table 4

Evaluation of extracted boundaries by different methods on the Shandong GF-2 images using two pixel-based attribute measures and three object-based measures, corresponding to OA and F1-score, and GUC, GOC, and GTC.

Methods	OA (%)	F1-score (%)	GOC	GUC	GTC
Refined SEANet	88.14	79.12	0.255	0.296	0.302
SEANet	87.90	75.46	0.401	0.229	0.346
BsiNet	86.86	69.29	0.239	0.789	0.593
ResUNet-a	81.84	56.80	0.396	0.790	0.641
SEANet _{MDE}	78.68	59.80	0.855	0.392	0.701

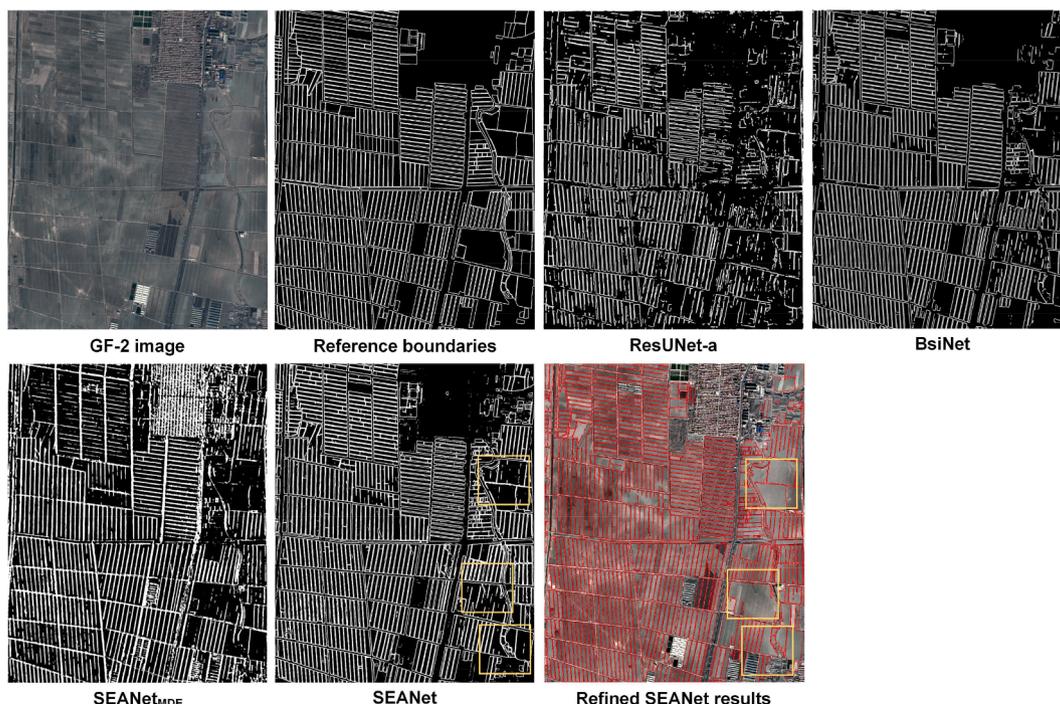


Fig. 8. Examples of parcel boundaries extracted by different methods in a Shandong testing area.

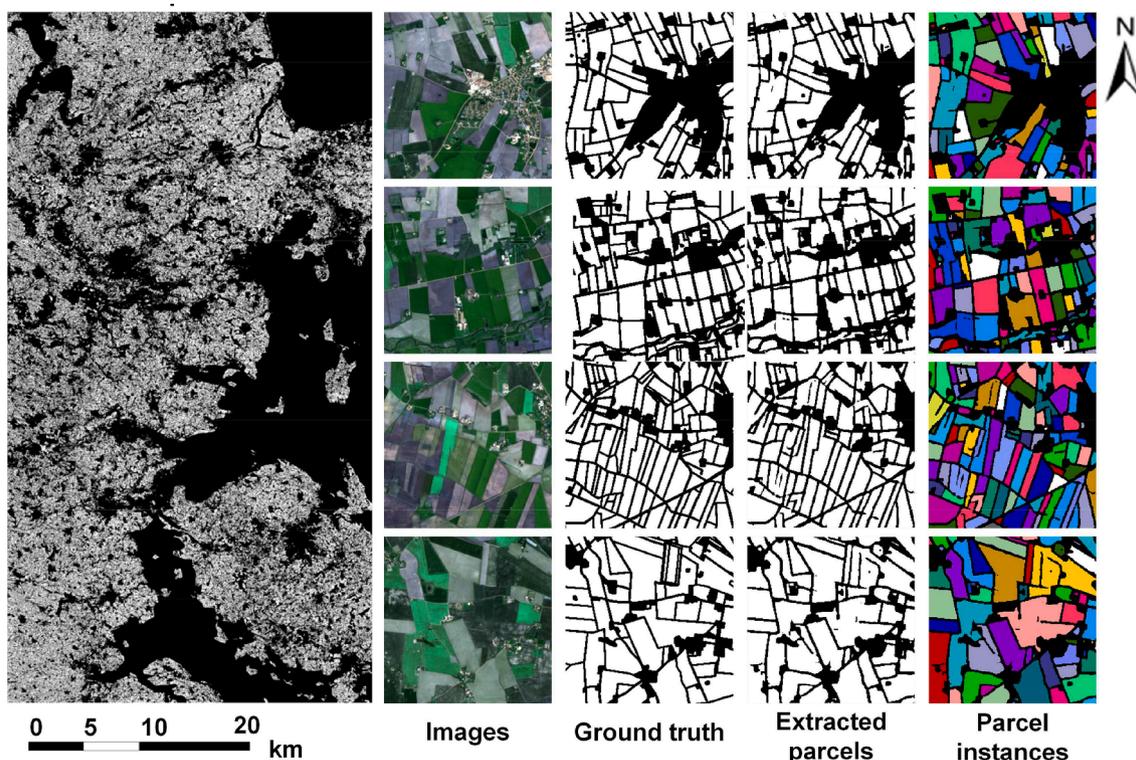


Fig. 9. Extracted agricultural parcels by SEANet on the Denmark (DK) Sentinel-2 image.

Table 5
Evaluation of extracted results by different methods on the Denmark Sentinel-2 image using pixel-based attribute measures.

Methods	P (%)	R (%)	OA (%)	F1-score (%)	IoU (%)
SEANet	84.58	87.13	90.32	85.84	76.24
ResUNet-a	85.09	83.61	91.74	84.34	75.53
ResUNet	85.01	85.04	90.17	85.02	75.37
BsiNet	83.58	85.57	89.32	84.56	74.86
SEANet _{MDE}	83.32	85.74	89.52	84.52	74.36
R2UNet	62.28	90.76	75.09	73.87	59.53

parcels vary a lot in size and shape. It derived the agricultural parcels retaining consistent shapes and closed boundaries, implying the effectiveness of our method to medium-resolution images (i.e., Sentinel-2 images).

We then compared our method with other methods using pixel-based accuracy measures on the Denmark dataset (Table 5). This table shows that our method obtained the highest F1-score of 85.84% and IoU of 76.24% compared with other methods. It improved by 0.71% and 1.5% than the recent ResUNet-a, which was developed based upon Sentinel-2 images, for IoU and F1-score measures, respectively. It also performs better than the state-of-the-art BsiNet by 1.28% and 1.38% for F1-score

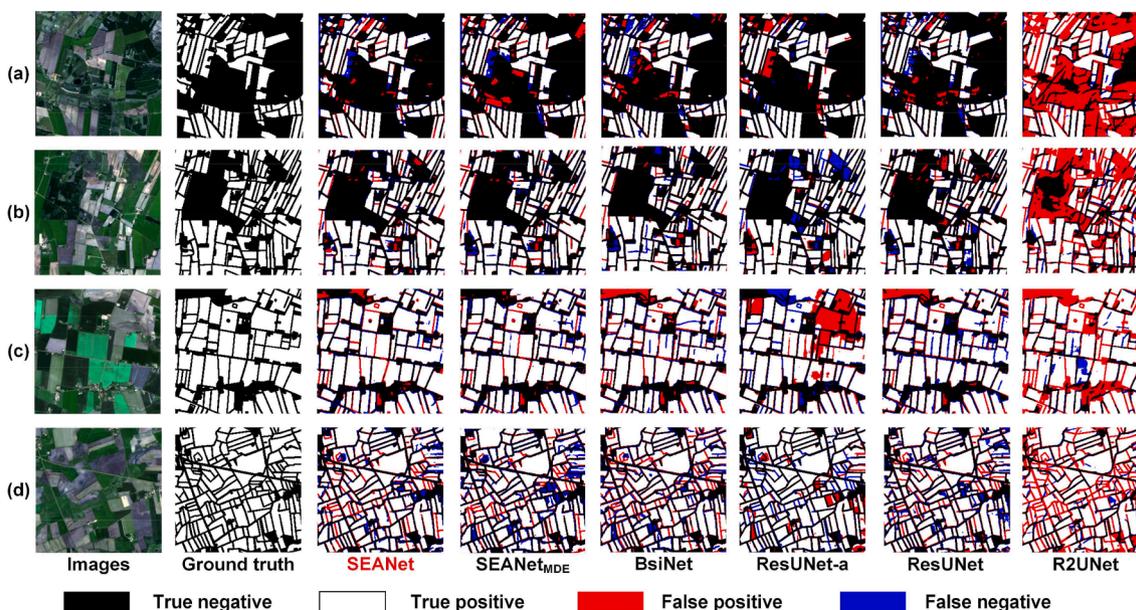


Fig. 10. Examples of agricultural parcels delineated by different methods on the Denmark dataset.

Table 6

Evaluation of extracted results by different methods on the Denmark Sentinel-2 image using object-based geometric measures.

Methods	GOC	GUC	GTC
SEANet	0.226	0.130	0.211
SEANet _{MDE}	0.229	0.135	0.215
ResUNet-a	0.212	0.143	0.215
BsiNet	0.222	0.148	0.216
ResUNet	0.235	0.139	0.217
R2UNet	0.910	0.263	0.679

Table 7

Estimates of the number of model parameters and computation load for different models. FLOPs represents floating-point operations per second and 1 M equals 10^6 .

Methods	Image size	FLOPs	Parameters
SEANet	256×256	208.14G	28.75 M
ResUNet	256×256	80.82G	13.04 M
ResUNet-a	256×256	70.90G	131.47 M
R2UNet	256×256	196.13G	39.09 M
BsiNet	256×256	13.30G	7.84 M

and IoU measures, as well as its variant SEANet_{MDE}. These results show that our method can effectively handle spectral heterogeneity by giving focus on discrepant features related to parcels. Fig. 10 visualizes extraction errors by these methods on subset images of the Denmark dataset. This figure shows that the parcels obtained by our method have the highest agreement with the ground truth, corresponding to the fewest false positives and false negatives.

We further compared our method with other methods using geometric measures on the Denmark dataset (Table 6). This table shows that the parcels delineated by our method have the lowest GUC of 0.130 and GTC of 0.211 than other methods. Moreover, multi-task networks, i.e., SEANet, SEANet_{MDE}, BsiNet, and ResUNet-a, have lower GTC errors than the adopted single-task networks, i.e., ResUNet and R2UNet. The results further strengthen that incorporating auxiliary tasks related to parcel geometry can effectively constrain the extraction of parcel features to form well-delineated boundaries.

4.5. Training plots and computation load of different models

We first assessed the computational efficiency of the proposed model, and compared it with existing methods (Table 7). This table indicates that the proposed SEANet (i.e., 28.75 M) has a smaller number of parameters than R2UNet (i.e., 39.09 M) and ResUNet-a (i.e., 131.47 M). Next, we visualized the training plots for the validation score of different models on the Shandong and Denmark datasets (Fig. 11). The figure

shows that SEANet showed more competitive performance and higher validation accuracy than others, particularly on the Denmark dataset.

4.6. Agricultural parcel delineation on the Xinjiang and Sichuan GF-2 images

To evaluate the generalizability of our methods for different agricultural settings, we applied the proposed method to delineate agricultural parcels from GF-2 images in Xinjiang and Sichuan areas. The results of parcel delineation are displayed in Fig. 12. This figure shows that our method produced parcels with excellent boundaries, e.g., clearly distinguishable between adjacent parcels. The results further imply that our method can be applied to different agricultural scenarios. Table 8 gives the accuracy assessment of the extracted parcels in these two areas. Both extractions have OAs larger than 90%. In addition, the extraction in the Xinjiang area has a higher accuracy than that in Sichuan, because many parcels in Sichuan are found with small sizes and irregular shapes, leading to more difficulties in parcel delineation.

4.7. Agricultural parcel delineation on the Netherlands Sentinel-2 image

We further applied our method to extract agricultural parcels from a Sentinel-2 image covering almost the whole Netherlands (Fig. 13). This figure shows that our method obtains well-delineated parcels on such a big dataset. It accurately identified parcel boundaries for fields with various crops. Noticeably, our method effectively discarded irrelevant ground features like buildings, even with a small size. These results suggest that our method is highly capable of learning features related to parcels, while ignoring irrelevant objects. Next, we evaluated the extracted results using OA, F1-score, GOC, GUC, and GTC measures. Although this Sentinel-2 image has high spectral heterogeneity because of seasonal differences, our method still derived satisfied results with an OA of 87.75%, F1-score of 90.22%, GOC of 0.143, GUC of 0.157, and GTC of 0.184, even using a small training set.

4.8. Transferability testing

To test the transferability of our method to high-resolution images (i.e., GF-2 image) across different areas, we applied the SEANet trained on the Xinjiang GF-2 image to delineate agricultural parcels on the Shandong GF-2 image, in which these areas have regular agricultural parcels and flat terrain. The delineation results are shown in Fig. 14. Clearly, our method successfully eliminated the interference of water and building in the transfer processing, even for images with obvious spectral differences. For this experiment, the results obtained by transferring SEANet trained on the Xinjiang area to the Shandong area (XJ-SD) were close to that directly trained on the Shandong areas (i.e., SD_{dt}) in terms of OA, F1-score, and GTC measures (Table 9). The results further indicate that

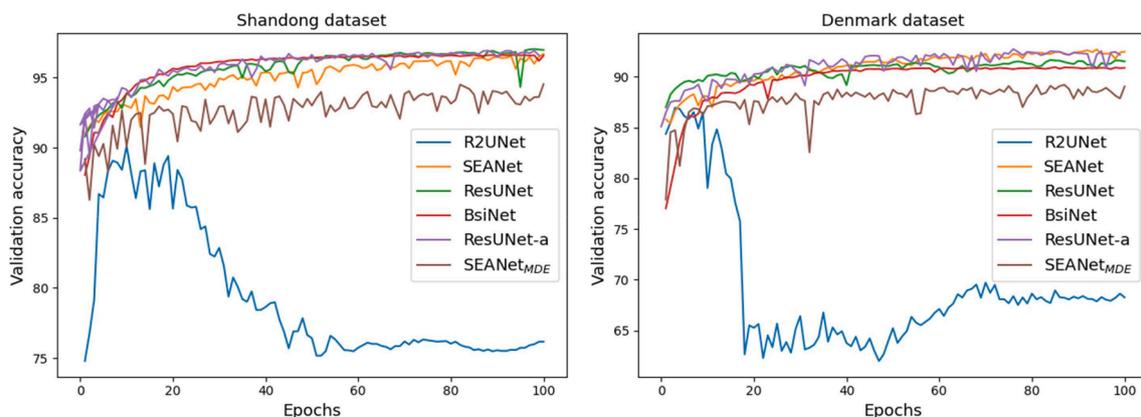


Fig. 11. The validation accuracy of different methods on the Shandong and Denmark datasets.

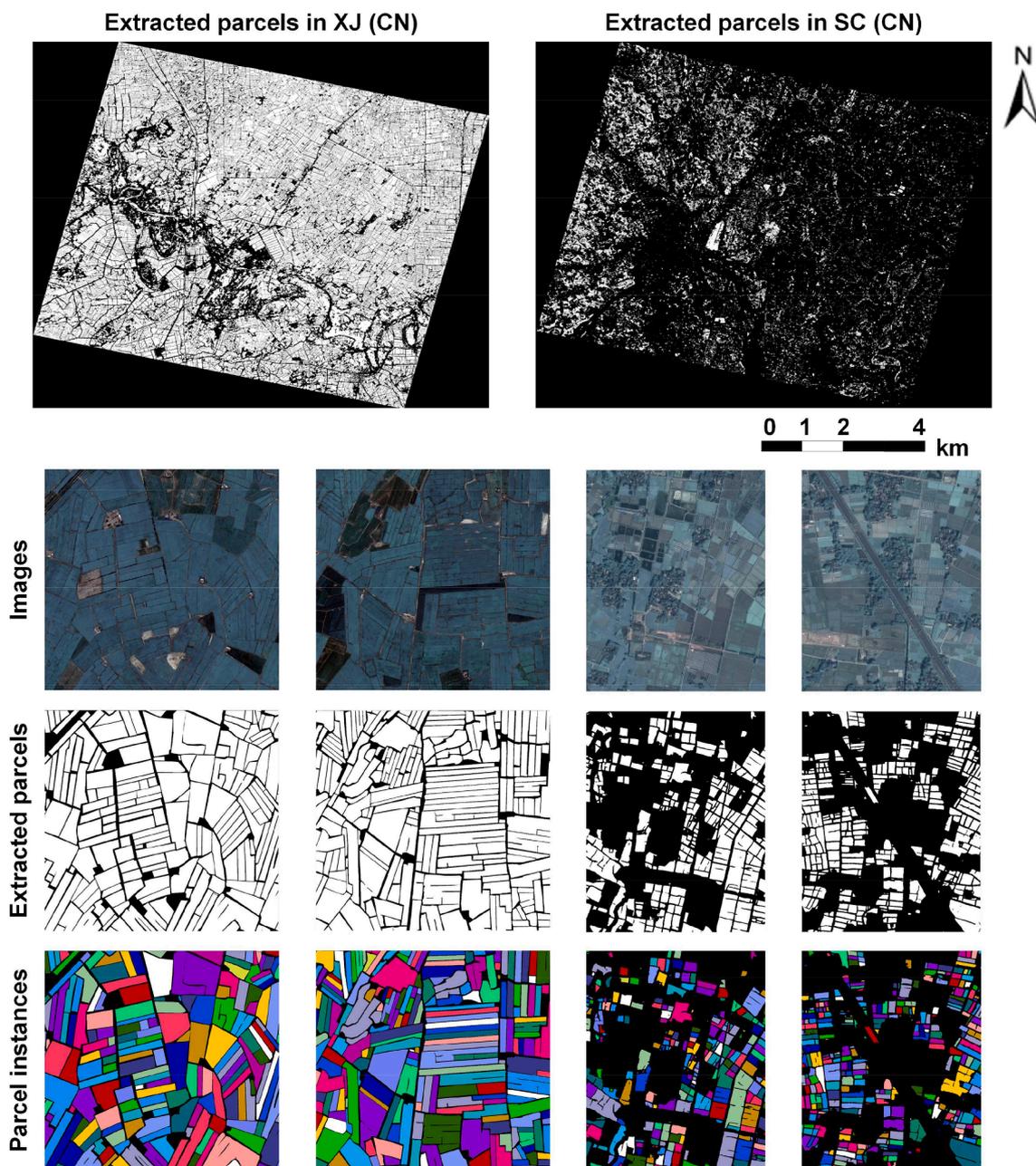


Fig. 12. Extracted agricultural parcels by SEANet on the Xinjiang (XJ) and Sichuan (SC) GF-2 images, China (CN).

Table 8

Evaluation of extracted results by SEANet on the Xinjiang and Sichuan GF-2 images using three pixel-based attribute measures and three object-based measures, corresponding to OA, F1-score, and IoU, and GUC, GOC, and GTC.

Study areas	OA (%)	F1-score (%)	IoU (%)	GOC	GUC	GTC
Xinjiang	94.75	96.76	93.75	0.129	0.067	0.127
Sichuan	92.85	84.67	75.73	0.163	0.118	0.176

our method has a strong transferability for areas with similar agricultural parcel characteristics, even across different landscapes and acquisition dates.

We further tested the transferability of our method to medium-resolution images (i.e., Sentinel-2 images). We used the SEANet trained on the Demark Sentinel-2 image to extract agricultural parcels on the Netherlands Sentinel-2 image. Fig. 15 displays the delineated

results of agricultural parcels. Clearly, our method accurately identified single agricultural parcels with dense arrangements and irregular shapes, while abandoning irrelevant objects.

We also evaluated the delineation results using attribute and geometric measures (Table 9). Agricultural parcels in the Denmark and Netherlands areas are highly different in their shapes and sizes. Moreover, the Netherlands dataset was comprised of five Sentinel-2 images across different dates and seasons, challenging transferability testing. Nonetheless, as shown in geometric measures, our method still obtained a high accuracy close to that directly trained on the Netherlands area (NL_{d1}), suggesting that our method has a high potential to be transferred to areas with different agricultural parcel characteristics.

5. Discussion

In this study, we developed a novel method to delineate agricultural parcels in various agricultural areas from both high- and medium-

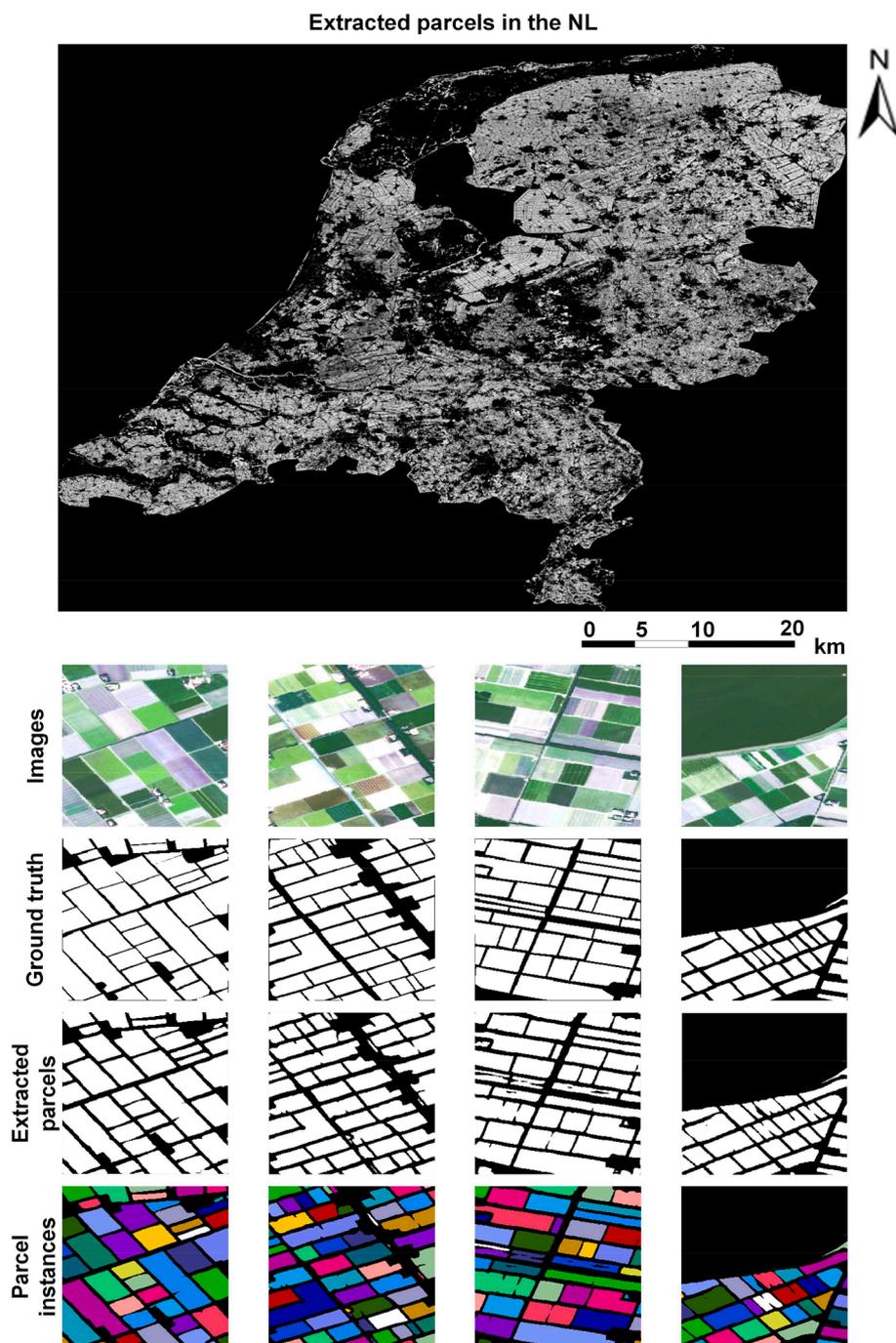


Fig. 13. Extracted agricultural parcels by SEANet on the Netherlands (NL) Sentinel-2 image.

resolution satellite images. More concretely, we considered the agricultural parcel delineation task as a multi-task segmentation problem and construct a new semantic edge-aware multi-task neural network, i. e., SEANet, for the segmentation task. It consists of three components corresponding to mask, edge, and distance map predictions. Different from previous research about mask and distance prediction tasks that use encoder-decoder architecture to extract semantic features, we used a dense multi-level architecture to capture rich edge semantics. By doing so, we obtained well-delineated parcels with regularized and closed boundaries by one segmentation. Moreover, we formulated a multi-task loss considering the uncertainty of different tasks to improve the generalizability of our methods for different areas. We conducted extensive experiments and analyses using five different datasets including both high-resolution and medium-resolution satellite images,

i. e., Shandong, Xinjiang, and Sichuan GF-2 images, and Denmark and the Netherlands Sentinel-2 images. Our results show that the proposed method extracted satisfactory parcel boundaries for all study areas. Moreover, it outperformed five other methods, corresponding to ResUNet, ResUNet-a, R2UNet, BsiNet, and a variant of the proposed model, i. e., SEANet_{MDE}, in terms of both attribute and geometric measures. Furthermore, our methods have a high potential to be transferred into different parcel areas, even across different dates and landscapes.

The proposed multi-task neural network comprises a core task (i. e., mask prediction) and two axillary tasks (i. e., edge prediction and distance map estimation). The axillary tasks provide extra meaningful information regarding parcel geometry, constraining mask prediction to produce regularized boundaries. Adding the axillary tasks further improves the generalizability of SEANet for different agricultural settings.

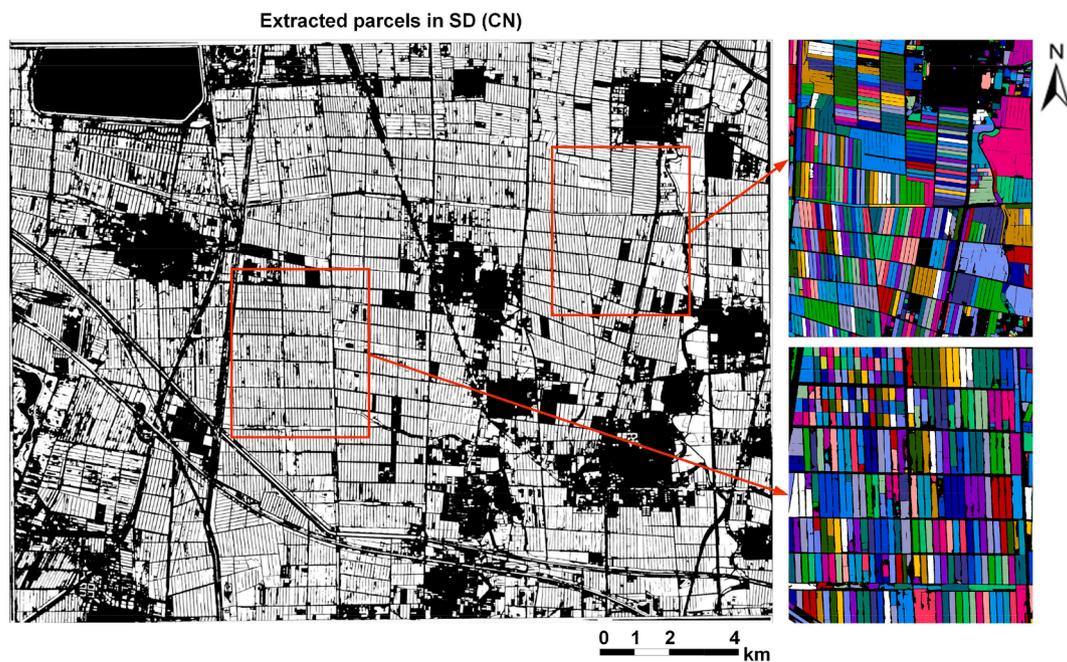


Fig. 14. Extracted results of agricultural parcels by transferring SEANet trained on the Xinjiang GF-2 image to the Shandong GF-2 image.

Table 9

Evaluation of extracted results by transferring SEANet trained on the Xinjiang and Denmark areas to the Shandong (XJ-SD) and the Netherlands (DK-NL) areas, respectively, using both attribute and geometric measures, corresponding to OA, F1-score, GOC, GUC, and GTC.

Methods	OA (%)	F1-score (%)	GOC	GUC	GTC
SD _{dt}	91.54	92.77	0.115	0.218	0.214
XJ-SD	92.48	91.71	0.271	0.145	0.266
NL _{dt}	87.75	90.22	0.143	0.157	0.184
DK-NL	81.73	83.69	0.112	0.223	0.210

Specifically, the proposed method achieved excellent performance (overall accuracy greater than 85%) in all study areas, and obtained the highest attribute and geometric accuracy than existing methods. In this study, our method used a VGG16 backbone. This can be replaced by other state-of-art backbones, like by one based upon a self-attention transformer, which has been proven to outperform convolution neural networks (CNNs) in extracting long-range global context cues (Strudel et al. 2021; Zheng et al. 2021). In the future, more auxiliary tasks can be investigated to further improve extraction accuracy.

The edge prediction task extracts edge semantics at different levels. It extracts both local and global edge semantics related to spectral and spatial context information, facilitating to deriving closed boundaries. We found that existing multi-task network methods like ResUNet-a had difficulty in obtaining closed parcel boundaries, even on the Shandong dataset with regular and large-size parcels. We concluded that conventional CNNs insufficiently use local and global edge information at different levels. Compared with the state-of-the-art BsiNet, our method obtained closed and clear boundaries with a higher accuracy, even for irregular parcels (Fig. 8). These results suggest that our method effectively learns features of the parcel edge of interest while abandoning edge-irrelevant features. Nevertheless, we are also aware that unclosed and incomplete boundaries are still exist for some challenging cases, where true parcel boundaries are vague on remote sensing images or parcels are surrounded with complex land covers. In this study, we used a refinement operation based upon morphological thinning and Douglas-Peucker algorithms to further improve parcel boundaries. Future research can be conducted to investigate more advanced

refinement operations (Waldner and Diakogiannis, 2020).

We formulated a multi-task loss considering uncertainties associated with different tasks to adaptively adjust the weights of mask prediction and distance map estimation. The advantage of using this multi-task loss is that it can automatically assign weights to mask and distance prediction tasks. Nevertheless, we formulate the final multi-task loss without using the uncertainties associated with the edge prediction task. Although the edge prediction task can be seen as a classification task, our study shows that it produces a loss that is hard to be treated as a Boltzmann distribution as the mask prediction task. Because the edge prediction task has a high imbalance of negative and positive samples, leading to a complicated probability distribution. Moreover, the edge prediction task considers multi-level follow edge features, involving more observation noise compared with other tasks like mask prediction and distance map estimation. However, the uncertainty weighting method tends to assign a smaller weight for the task with a larger noise. Our experiments also justified this assumption (Fig. 8). Results showed that SEANet_{MDE}, incorporating the loss of edge task-dependent uncertainty failed to perform better than SEANet in terms of the attribute and geometric accuracy measures. To further improve parcel delineation results, future studies can be conducted to explore more sophisticated methods to rigorously model the probability distribution of the edge prediction task.

We conducted extensive experiments in five study areas with different parcel characteristics using both high- and medium-resolution remote sensing images. Our method was successfully applied to all these datasets. In general, the accuracy of extracted parcels in agricultural areas with flat patterns and regular farm fields (e.g., the Shandong area) is higher than those areas (e.g., the Sichuan area) where parcels are characterized by smallholder farms with small sizes and irregular shapes. Our results provide quite positive signals that the proposed method highly suits for parcel extraction using remote sensing images from various sensors. Moreover, our method was successfully transferred to different areas (Figs. 14 and 15) and with little interference by other objects, even using a small training set (see Section 3.6). It is also highly capable of reducing the effects of the discrepancy on time, landscapes, spectral, and parcel characteristics. In this study, the samples used to train our model (and other existing methods) were collected manually. Sample collection is always labor-intensive work for applying

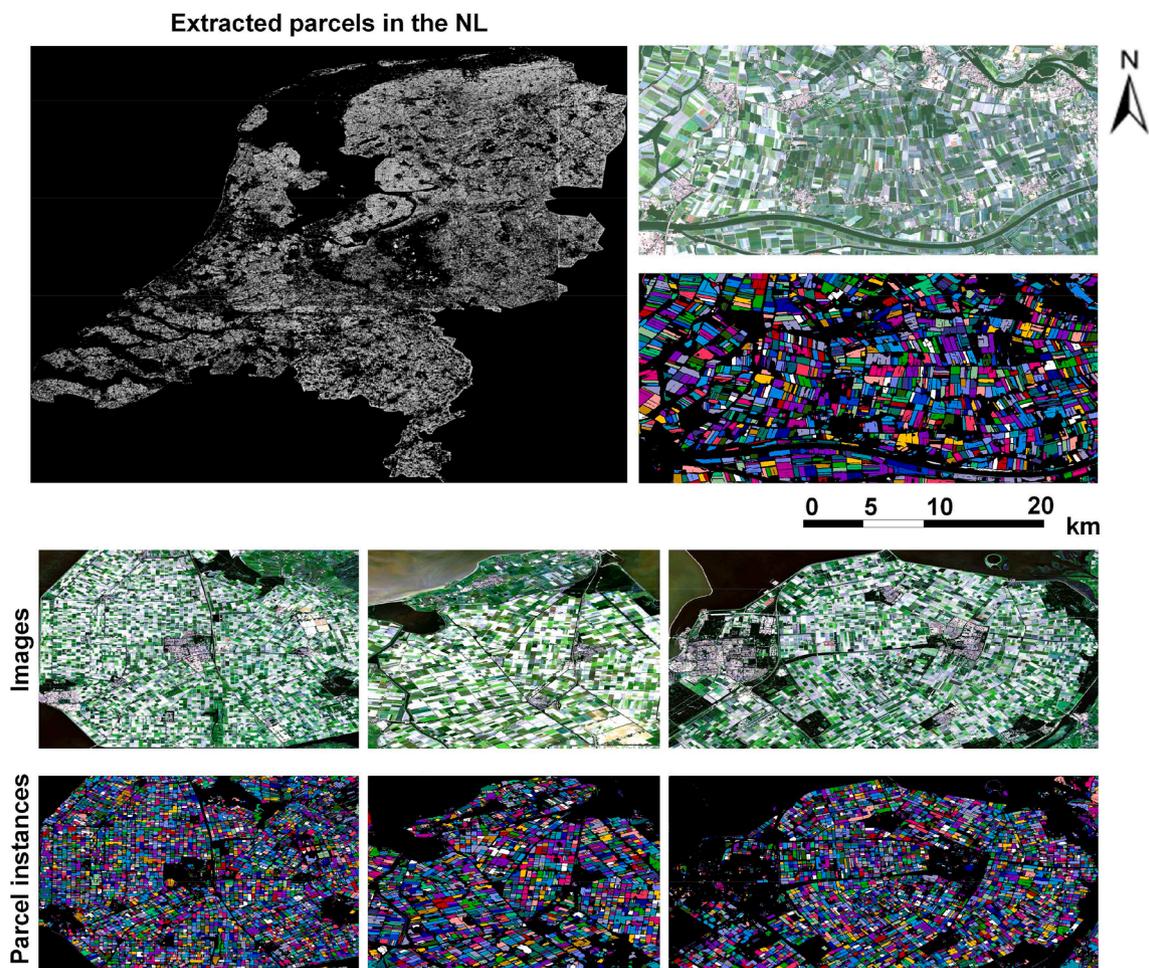


Fig. 15. Results of transferring SEANet trained on the Denmark Sentinel-2 image to extract parcels from the Netherlands Sentinel-2 image.

deep neural networks. In the future, we may consider using few-shot learning such as generative adversarial networks (GANs) to reduce the cost of manual labeling, and to further enhance the applicability of the proposed method (Cao and Huang 2022; Jong et al. 2022).

6. Conclusions

This paper presents a semantic edge-aware multi-task neural network SEANet that delineates agricultural parcels from remote sensing images. It incorporates local and global edge semantics at multiple levels to enhance edge feature extraction, and to obtain closed parcel boundaries by one segmentation processing step. We formulated a multi-task loss to improve parcel extraction. A refinement method further improved the geometric accuracy of the extracted boundaries, obtaining vector-formatted outputs. We conducted experiments on five different datasets, corresponding to Shandong, Xinjiang, and Sichuan GF-2 images in China, and Denmark and the Netherlands Sentinel-2 images. Experimental results allowed us to conclude that:

- 1) SEANet improved the extraction of agricultural parcels from remote sensing images. It performed better than state-of-the-art methods (namely, ResUNet, ResUNet-a, R2UNet, and BsiNet) in both attribute and geometric accuracies;
- 2) Incorporating the edge detection module effectively extracted relevant information at the local and global levels. Extracted parcels were evidently refined with multi-level edge features, leading to parcels with regularized and closed boundaries;

- 3) We provided an effective multi-task loss considering task-dependent uncertainties to automatically balance the weights of different tasks.

Hence, SEANet is effective for delineating agricultural parcels with various shapes and sizes from remote sensing images of both high- and medium-resolutions. It is well transferable to areas with different agricultural landscapes, and provides a promising solution for agricultural parcel delineation using remote sensing images at various mapping scales.

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.

Acknowledgment

This work was supported by the National Natural Science Foundation of China (No. 42001283) and the Science and Technology Program of Fujian Province of China (No. 2022N0019).

References

- 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.
- Belgiu, M., Drăguț, L., 2014. Comparing supervised and unsupervised multiresolution segmentation approaches for extracting buildings from very high resolution imagery. *ISPRS J. Photogramm. Remote Sens.* 96, 67–75.

- Boryan, C., Yang, Z., Mueller, R., Craig, M., 2011. Monitoring US agriculture: the US department of agriculture, national agricultural statistics service, cropland data layer program. *Geocarto Int.* 26, 341–358.
- Cai, Z., Hu, Q., Zhang, X., Yang, J., Wei, H., He, Z., Song, Q., Wang, C., Yin, G., Xu, B., 2022. An adaptive image segmentation method with automatic selection of optimal scale for extracting cropland parcels in smallholder farming systems. *Remote Sens. (Basel)* 14, 3067.
- Cao, Y., Huang, X., 2022. A coarse-to-fine weakly supervised learning method for green plastic cover segmentation using high-resolution remote sensing images. *ISPRS J. Photogramm. Remote Sens.* 188, 157–176.
- Cheng, T., Ji, X., Yang, G., Zheng, H., Ma, J., Yao, X., Zhu, Y., Cao, W., 2020. DESTIN: A new method for delineating the boundaries of crop fields by fusing spatial and temporal information from WorldView and Planet satellite imagery. *Comput. Electron. Agric.* 178, 105787.
- Cheng, D., Meng, G., Cheng, G., Pan, C., 2016. SeNet: Structured edge network for sea-land segmentation. *IEEE Geosci. Remote Sens. Lett.* 14, 247–251.
- Crommelinck, S., Bennett, R., Gerke, M., Yang, M.Y., Vosselman, G., 2017. Contour detection for UAV-based cadastral mapping. *Remote Sens. (Basel)* 9, 171.
- Diakogiannis, F.I., Waldner, F., Caccetta, P., Wu, C., 2020. ResUNet-a: a deep learning framework for semantic segmentation of remotely sensed data. *ISPRS J. Photogramm. Remote Sens.* 162, 94–114.
- Garcia-Pedrero, A., Gonzalo-Martin, C., Lillo-Saavedra, M., 2017. A machine learning approach for agricultural parcel delineation through agglomerative segmentation. *Int. J. Remote Sens.* 38, 1809–1819.
- Ghorbanian, A., Kakooei, M., Amani, M., Mahdavi, S., Mohammadzadeh, A., Hasanlou, M., 2020. Improved land cover map of Iran using Sentinel imagery within Google Earth Engine and a novel automatic workflow for land cover classification using migrated training samples. *ISPRS J. Photogramm. Remote Sens.* 167, 276–288.
- Hay, G.J., Blaschke, T., Marceau, D.J., Bouchard, A., 2003. A comparison of three image-object methods for the multiscale analysis of landscape structure. *ISPRS J. Photogramm. Remote Sens.* 57, 327–345.
- He, K., Zhang, X., Ren, S., Sun, J., 2016. Deep residual learning for image recognition. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 770–778.
- Heidler, K., Mou, L., Baumhoer, C., Dietz, A., Zhu, X.X., 2021. HED-UNet: combined segmentation and edge detection for monitoring the Antarctic coastline. *IEEE Trans. Geosci. Remote Sens.* 60, 1–14.
- Jong, M., Guan, K., Wang, S., Huang, Y., Peng, B., 2022. Improving field boundary delineation in ResUNets via adversarial deep learning. *Int. J. Appl. Earth Obs. Geoinf.* 112, 102877.
- Kendall, A., Gal, Y., Cipolla, R., 2018. Multi-task learning using uncertainty to weigh losses for scene geometry and semantics. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 7482–7491.
- Kendall, A., Gal, Y., 2017. What uncertainties do we need in Bayesian deep learning for computer vision? *Adv. Neural Inf. Process. Syst.* 30.
- Kocur-Bera, K., 2019. Data compatibility between the Land and Building Cadaster (LBC) and the Land Parcel Identification System (LPIS) in the context of area-based payments: a case study in the Polish Region of Warmia and Mazury. *Land Use Policy* 80, 370–379.
- Lebourgeois, V., Dupuy, S., Vintrou, É., Ameline, M., Butler, S., Bégue, 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. (Basel)* 9, 259.
- Li, T., Johansen, K., McCabe, M.F., 2022. A machine learning approach for identifying and delineating agricultural fields and their multi-temporal dynamics using three decades of Landsat data. *ISPRS J. Photogramm. Remote Sens.* 186, 83–101.
- Liebel, L., Körner, M., 2018. Auxiliary tasks in multi-task learning. *arXiv preprint arXiv: 1805.06334*.
- Liu, Y., Cheng, M.-M., Hu, X., Bian, J.-W., Zhang, L., Bai, X., Tang, J., 2019. Richer convolutional features for edge detection. *IEEE Trans. Pattern Anal. Mach. Intell.* 41, 1939–1946.
- Long, J., Li, M., Wang, X., Stein, A., 2022. Delineation of agricultural fields using multi-task BsiNet from high-resolution satellite images. *Int. J. Appl. Earth Obs. Geoinf.* 112, 102871.
- Lu, R., Wang, N., Zhang, Y., Lin, Y., Wu, W., Shi, Z., 2022. Extraction of agricultural fields via DASNet with dual attention mechanism and multi-scale feature fusion in South Xinjiang, China. *Remote Sensing* 14, 2253.
- Ma, J., Wei, Z., Zhang, Y., Wang, Y., Lv, R., Zhu, C., Gaoxiang, C., Liu, J., Peng, C., Wang, L., 2020. How distance transform maps boost segmentation CNNs: an empirical study. In: *Medical Imaging with Deep Learning*. PMLR, pp. 479–492.
- Marmanis, D., Schindler, K., Wegner, J.D., Galliani, S., Datcu, M., Stilla, U., 2018. Classification with an edge: improving semantic image segmentation with boundary detection. *ISPRS J. Photogramm. Remote Sens.* 135, 158–172.
- 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. (Basel)* 12, 59.
- Milletari, F., Navab, N., Ahmadi, S.-A., 2016. V-net: fully convolutional neural networks for volumetric medical image segmentation. In: *2016 Fourth International Conference on 3D Vision (3DV)*. Ieee, pp. 565–571.
- 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, 1619–1628.
- Murugesan, B., Sarveswaran, K., Shankaranarayana, S.M., Ram, K., Joseph, J., Sivaprakasam, M., 2019. Psi-Net: shape and boundary aware joint multi-task deep network for medical image segmentation. In: *2019 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*. IEEE, pp. 7223–7226.
- Persello, C., Bruzzone, L., 2009. A novel protocol for accuracy assessment in classification of very high resolution images. *IEEE Trans. Geosci. Remote Sens.* 48, 1232–1244.
- Persello, C., Tolpekin, V., Bergado, J., de By, R., 2019. Delineation of agricultural fields in smallholder farms from satellite images using fully convolutional networks and combinatorial grouping. *Remote Sens. Environ.* 231, 111253.
- Potlupally, A., Chowdary, P.S.R., Shekhar, S.R., Mishra, N., Madhuri, C.S.V.D., Prasad, A., 2019. Instance segmentation in remote sensing imagery using deep convolutional neural networks. In: *2019 International Conference on contemporary Computing and Informatics (IC3I)*. IEEE, pp. 117–120.
- Rieke, C., 2017. Deep learning for instance segmentation of agricultural fields. Available from: <https://github.com/chrieke/InstanceSegmentation_Sentinel2/> (accessed 21 May 2022).
- Robb, C., Hardy, A., Doonan, J.H., Brook, J., 2020. Semi-automated field plot segmentation from UAS imagery for experimental agriculture. *Front. Plant Sci.* 11, 591886.
- Rydberg, A., Borgefors, G., 1999. Extracting multispectral edges in satellite images over agricultural fields. In: *Proceedings 10th International Conference on Image Analysis and Processing*. IEEE, pp. 786–791.
- Rydberg, A., Borgefors, G., 2001. Integrated method for boundary delineation of agricultural fields in multispectral satellite images. *IEEE Trans. Geosci. Remote Sens.* 39, 2514–2520.
- Stein, A., Esfahani, A.G., Abkar, A.-A., 2016. Opportunities of using random sets to model uncertainties in agricultural field boundaries observed from remote sensing images. *Asian J. Agric. Extension Econ. Sociol.* 8, 1–11.
- Strudel, R., Garcia, R., Laptev, I., Schmid, C., 2021. Segmenter: transformer for semantic segmentation. In: *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 7262–7272.
- Tang, Z., Li, M., Wang, X., 2020. Mapping tea plantations from VHR images using OBIA and convolutional neural networks. *Remote Sens. (Basel)* 12, 2935.
- Taravat, A., Wagner, M.P., Bonifacio, R., Petit, D., 2021. Advanced fully convolutional networks for agricultural field boundary detection. *Remote Sens. (Basel)* 13, 722.
- Teluguntla, P., Thenkabail, P.S., Oliphant, A., Xiong, J., Gumma, M.K., Congalton, R.G., Yadav, K., Huete, A., 2018. A 30-m landsat-derived cropland extent product of Australia and China using random forest machine learning algorithm on Google Earth Engine cloud computing platform. *ISPRS J. Photogramm. Remote Sens.* 144, 325–340.
- Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A.N., Kaiser, Ł., Polosukhin, I., 2017. Attention is all you need. *Adv. Neural Inf. Process. Syst.* 30.
- 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.
- Waldner, F., Diakogiannis, F.I., Batchelor, K., Ciccosto-Camp, M., Cooper-Williams, E., Herrmann, C., Mata, G., Toovey, A., 2021. Detect, consolidate, delineate: scalable mapping of field boundaries using satellite images. *Remote Sens. (Basel)* 13, 2197.
- Watkins, B., Van Niekerk, A., 2019a. Automating field boundary delineation with multi-temporal Sentinel-2 imagery. *Comput. Electron. Agric.* 167, 105078.
- Watkins, B., Van Niekerk, A., 2019b. A comparison of object-based image analysis approaches for field boundary delineation using multi-temporal Sentinel-2 imagery. *Comput. Electron. Agric.* 158, 294–302.
- Wei, S., Ji, S., Lu, M., 2019. Toward automatic building footprint delineation from aerial images using CNN and regularization. *IEEE Trans. Geosci. Remote Sens.* 58, 2178–2189.
- Xu, L., Ming, D., Zhou, W., Bao, H., Chen, Y., Ling, X., 2019. Farmland extraction from high spatial resolution remote sensing images based on stratified scale pre-estimation. *Remote Sens. (Basel)* 11, 108.
- Yu, Z., Feng, C., Liu, M.-Y., Ramalingam, S., 2017. Casenet: deep category-aware semantic edge detection. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 5964–5973.
- Zhang, Z., Liu, Q., Wang, Y., 2018. Road extraction by deep residual u-net. *IEEE Geosci. Remote Sens. Lett.* 15, 749–753.
- Zhang, H., Liu, M., Wang, Y., Shang, J., Liu, X., Li, B., Song, A., Li, Q., 2021. Automated delineation of agricultural field boundaries from Sentinel-2 images using recurrent residual U-Net. *Int. J. Appl. Earth Obs. Geoinf.* 105, 102557.
- Zhang, D., Pan, Y., Zhang, J., Hu, T., Zhao, J., Li, N., Chen, Q., 2020. A generalized approach based on convolutional neural networks for large area cropland mapping at very high resolution. *Remote Sens. Environ.* 247, 111912.
- Zhao, H., Shi, J., Qi, X., Wang, X., Jia, J., 2017. Pyramid scene parsing network. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 2881–2890.
- Zheng, S., Lu, J., Zhao, H., Zhu, X., Luo, Z., Wang, Y., Fu, Y., Feng, J., Xiang, T., Torr, P. H., 2021. Rethinking semantic segmentation from a sequence-to-sequence perspective with transformers. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 6881–6890.