Computers and Electronics in Agriculture 203 (2022) 107473

Contents lists available at ScienceDirect

Computers and Electronics in Agriculture

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

A deep learning method for individual arable field (IAF) extraction with cross-domain adversarial capability

Shuaijun Liu [a], Licong Liu [a], Fei Xu [b], Jin Chen [a], Yuhen Yuan [a], Xuehong Chen [a,*]

[a] State Key Laboratory of Remote Sensing Science, Institute of Remote Sensing Science and Engineering, Faculty of Geographical Science, Beijing Normal University, Beijing, China
[b] Department of Land Surveying and Geo-Informatics, The Hong Kong Polytechnic University, Hong Kong, China

ARTICLE INFO

Keywords:
Individual arable field
Gaofen-2
Semantic segmentation
Deep learning
Domain transfer
Generalization

ABSTRACT

The monitoring of agricultural fields from remote sensing imagery allows for the effective management of agricultural resources at large spatial scales. Deep learning-based methods have shown great potential in the extraction of individual arable fields (IAF) from high-resolution imagery; however, accurate boundary localization and collection of adequate training samples remain a challenge. This study proposes a novel deep learning network (FieldSeg-DA) to extract IAFs from Chinese high-resolution satellite imagery (Gaofen-2). FieldSeg-DA adopts a parallel network structure consisting of two branch networks (UNet and DeepNetV3+) to independently extract the boundary and extent of IAFs. A post-processing module, connecting boundaries and filling field (CB-FF), was used to integrate the extracted boundary and extent and promote the integrity of each extracted IAF. The parallel network coupled with CB-FF improved the boundary and extent accuracy of IAF extraction. Moreover, we used fine-grain adversarial domain adaption (FADA) in the training stage to promote the transferability of the trained network from source domains with labeled samples to target domains without any labeled samples. We found that FieldSeg-DA outperformed IAFRes, a state-of-the-art method for IAF extraction, in both the source and target domains, with improvements of 0.016 and 0.069 in the F1-score, respectively. Therefore, the proposed FieldSeg-DA method has the potential to extract IAFs accurately across diverse farming areas without training samples.

1. Introduction

An individual arable field (IAF)—also known as a crop plot or crop parcel—is the basic spatial unit of crop farming that is demarcated by ridges, paths, and ditches, and is usually planted with one crop per growing season (Persello et al., 2019; Yan and Roy, 2014). The spatial information (location and area) of IAFs is not only relevant to determining ownership of croplands but is also required when planning cooperative farming activities, such as irrigation, fertilization, and harvesting. Moreover, it is the basic unit of crop mapping (Matton et al., 2015; Waldner and Diakogiannis 2020).

Although the traditional method of obtaining the spatial distribution of IAFs is by field survey, it is ineffective in analyzing large areas (Blaes et al., 2005; De Wit and Clevers, 2004). Remote sensing techniques, used to generate detailed geometric information for large areas from satellite images, have emerged as efficient alternatives (De Wit and Clevers 2004). With the increasing availability of high-resolution satellite imagery, various remote sensing-based methods have been developed for the analysis of IAFs. Visual interpretation is a common practical approach for remote sensing techniques. However, at regional or national scales, this approach is impractical (Garcia-Pedrero et al., 2017; Rahman et al., 2019; Turker and Kok, 2013), which is why researchers have opted for the automatic extraction of IAFs from remotely sensed images.

Automatic IAF extraction, as a typical application of image segmentation techniques, has received increasing attention recently (Belgiu and Csillik 2018). Similar to image segmentation, automatic IAF extraction methods can be classified into graphical operator- and deep learning-based approaches based on different feature extractors.

Graphical operator based approaches are based on edge or region detection. Edge detection methods, such as the Canny operator, χ algorithm, and Gaussian edge filter, use convolution kernels to identify linear ground objects (e.g., ridges, paths, and ditches) (Xi and Zhang 2012; Belgiu and Csillik 2018; Liu et al., 2010; Verrelst et al., 2014). The

* Corresponding author.
E-mail address: chenxuehong@bnu.edu.cn (X. Chen).

https://doi.org/10.1016/j.compag.2022.107473
Received 2 August 2022; Received in revised form 8 October 2022; Accepted 29 October 2022
Available online 21 November 2022

role of the kernel is to distinguish between linear ground objects and other neighboring objects. The region detection method uses the homogeneous spectral features of an image to identify the spatial extent of the IAFs (Segl and Kaufmann 2001; Da Costa et al., 2007). These texture-based heterogeneity characteristics derived from high-resolution remotely sensed images were used to identify the IAFs. Unfortunately, image noise heavily influences the application of most methods, resulting in incomplete or fragmented boundaries (Yan et al., 2015; Zhou et al., 1989). More importantly, the features and parameters associated with model performance rely heavily on experience, which may not be consistent across users and study regions, thus reducing the robustness of the algorithms (Aquino et al., 2010).

Deep learning-based methods reduce this reliance on expert knowledge owing to their automated feature extraction capabilities. These methods employ convolutional neural networks (CNN)—a set of convolutional operators—to obtain spatial and spectral features (also called dense features) at different spatial scales for semantic segmentation of IAFs (Long et al., 2015; Huang et al., 2020; Persello et al., 2019; Gao et al., 2020; Rabbi et al., 2020; Meyer et al., 2020; Waldner and Diakogiannis, 2020; Ma et al., 2019). For example, Huang et al. (2020) used a UNet model with a "U-shaped" encoder-decoder structure. Persello et al. (2019) proposed a contour delineation technique based on deep, fully convolutional networks to extract IAFs of various sizes. Gao et al. (2020) proposed a multi-scale feature fusion module robust UNet (MMUUNet), an improved UNet model that eliminates adhesion in the extracted IAFs. Considering that the CNN-based approach is not sufficiently accurate to extract the boundary of IAFs, Rabbi et al. (2020) used super-resolution generative adversarial networks to upsample remotely sensed images for edge enhancement and improved performance of semantic segmentation. Meyer et al. (2020) proposed a mask R-CNN-based model for delineating IAF boundaries over large areas. Yang et al. (2020) extracted IAF boundaries by training a "Snakes" algorithm that considers geometric and topological constraints. Waldner and Diakogiannis (2020) proposed a multitasking network called IAFRes to extract both boundaries and extents with a post-processing module combining Otsu's thresholding algorithm and the watershed segmentation approach. However, there could be a trade-off between rich context information extraction and accurate boundary localization for a single network (Ma et al., 2019; Peng et al., 2022). Therefore, simultaneously improving the extent and boundary accuracy of IAFs remains challenging.

Another pressing challenge of deep learning-based IAF extraction methods is the transferability of CNN models across study areas, also known as the "domain transfer problem" (Yoo et al., 2016). In general, the transferability of CNN models depends on the similarity in data distribution between an area with model training (source domain) and another area without (target domain) (Li et al., 2020; Matasci et al., 2015). The differences between areas, such as size, shape, and spectral characteristics, may decrease the accuracy of the trained semantic segmentation network in recognizing the objects of interest in non-trained areas. An effective technique to address the issue is so-called "fine-tuning," which adapts the parameters of a pre-trained model from the source domain to the target domain with only a small number of samples in the target domain (Liu et al., 2018; Nogueira et al., 2017). However, the fine-tuning technique cannot be applied to study areas without labeled samples. The zero-shot domain adaptation techniques were developed for tasks without target domain samples. These techniques align the image features extracted from the target domain to those of the source domain in terms of statistical distribution through adversarial learning, thereby guaranteeing transferability in trained models (Duan et al., 2012; Matasci et al., 2015; Yoo et al., 2016). Recently, category information was used in domain transfer to improve the alignment of conditional class distributions, thus enhancing the transferability of classification tasks. Although these techniques have shown some success in many computer vision tasks, they have not been tested in IAF extraction for different study areas.

This study aims to propose an IAF extraction deep learning network to address the two challenges mentioned above: difficulty in simultaneously improving the extent and boundary extraction accuracy of IAFs and the limited transferability of CNN models for IAF extraction across different study areas. In the novel network, two independent networks were coupled in a parallel network structure to identify IAF extent and boundary simultaneously; and an adversarial learning-based domain adaption technique was introduced to guarantee the transferability of the proposed network.

2. Method

This study proposed a field segmentation network named FieldSeg-DA to automatically extract IAFs from Chinese Gaofen-2 (Gaofen-2) images. The technical flow of the proposed FieldSeg-DA is illustrated in Fig. 1. Firstly, FieldSeg-DA integrated two branch networks, UNet (Persello et al., 2019) and DeepLabV3+ (Yan and Roy 2014), to identify the IAF extent and boundary, respectively, with an extent-boundary parallel network (EBPN) structure. Subsequently, a typical domain adaption method, fine-grain adversarial domain adaption (FADA), was employed in the training stage to guarantee the robust performance of both branch networks over the target domain without labeled samples to enhance the transferability of the deep learning networks trained from the source domain. Finally, a post-processing procedure consisting of several morphological image operators, called connecting boundaries and filling field (CB-FF), was proposed to eliminate the fragmented extents and boundaries and consistently integrate the extent and boundary extracted by the two branch networks.

2.1. IAF extent and boundary extraction with the parallel deep convolutional neural network

2.1.1. UNet and DeepLabV3+

DeepLabV3+ and UNet are semantic segmentation networks that provide dense pixel-based predictions. These networks typically operate with an encoder-decoder structure to maintain consistent dimensions between the input and output (Fig. 2). Notably, as the number of convolutions increases, encoder-derived features are obtained from a larger receptive field; thus, features with few convolutions (low-level convolution features) represent the image texture at the local scale and features with more convolutions (high-level convolution features) represent abstract context over a larger spatial neighborhood.

Besides the shared network characteristics mentioned above, DeepLabV3+ and UNet have different preferences in identifying "face" and "line" objects, respectively (Hosseinpour et al., 2022; Mosinska et al., 2018; Peng et al., 2022). UNet delivers low-level convolutional features directly to the decoder with a "skip connection" (brown lines in Fig. 2a), which enables the effective recognition of line-like objects (e.g., cropland boundaries). In contrast, DeepLabV3+ employs limited "skip connections" and introduces a dilated convolution kernel to enlarge the receptive field; thus, its applications are more focused on high-level convolutional features and recognition of face-like objects (e.g., the spatial extent of a single IAF). We used UNet to estimate the IAF boundary by training it with samples of a 1-pixel buffer of the IAF boundary vector; thus, UNet was focused more on boundary connectivity than extent area. On the other hand, DeepLabV3+ was trained with the "face" object (i.e., IAF extent), with a greater focus on IAF location than boundary accuracy.

2.1.2. Training and inference

Owing to limited GPU memory size, we created subsets of the images with dimensions of 256 × 256 pixels as input for UNet and DeepLabV3+ in the training stage. To avoid degradation in the segmentation accuracy of pixels near the image edges, we used overlapping inference windows with a stride of 128 pixels. This is a commonly used approach for large land cover classification that guarantees the incorporation of neighborhood information for each pixel (Waldner et al., 2017). We replaced

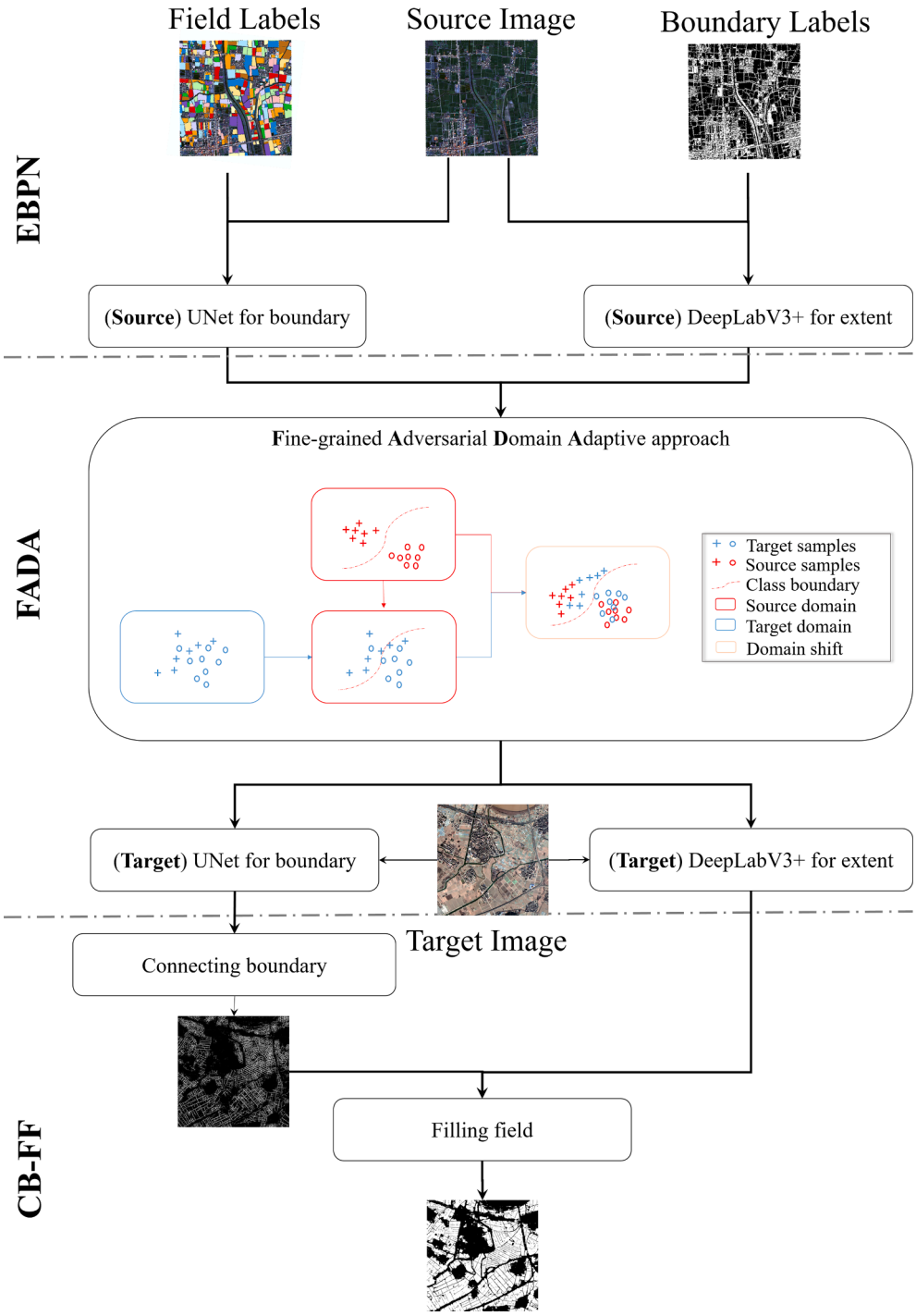

Fig. 1. Workflow of FieldSeg-DA analysis.

the cross-entropy loss function in the original UNet and DeepLabV3+ models with the Tonimoto loss function (Diakogiannis et al., 2020) to address sample imbalance, which is especially important for boundary extraction.

2.2. Fine-grained domain adaptation

We used the DA framework (Yoo et al., 2016)—a typical zero-shot domain adaptation technique—to simultaneously train a feature extractor and discriminator, which learn the discriminative features for the classification task from the source domain and align the indiscriminative features between the source and target domains, respectively. However, the original DA framework lacks categorical semantic information, which limits the expressiveness of the DA model. Therefore, we used an improved version of the DA framework, FADA (Wang et al., 2020), which aligns features according to category information, in the training of both the UNet and DeepLabV3+ models to guarantee robust extent and boundary extraction performance in the target domain (Fig. 3).

2.2.1. Domain adaptation framework

Labeled sample data in the source domain and unlabeled sample data in the target domain can be tagged as follows:

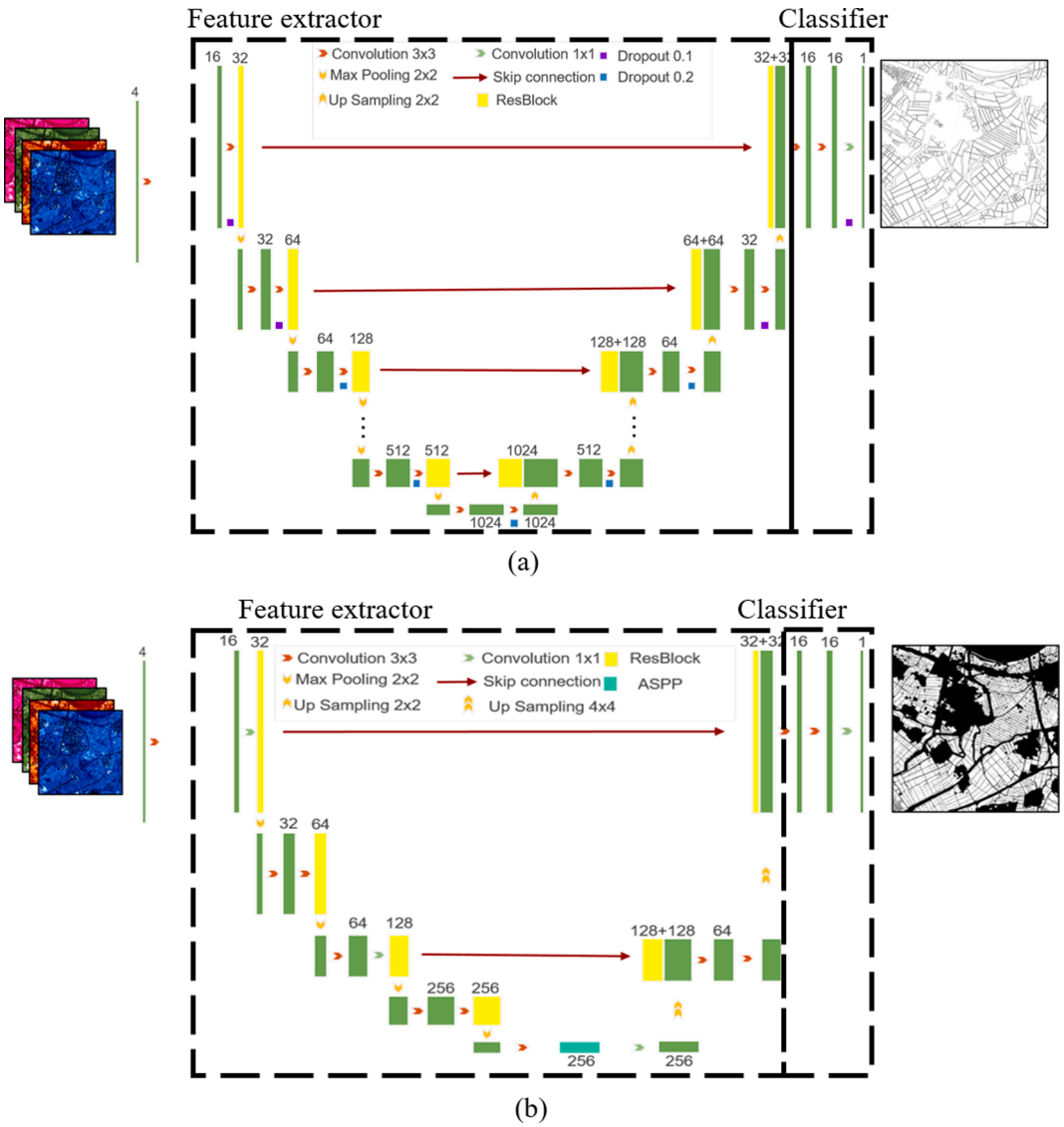

Fig. 2. The parameters of the UNet (a) and DeepLabV3+ (b) models for satellite image segmentation.

$$Source = \{Img(S), L(S)\} \tag{1}$$

$$Target = \{Img(T)\} \tag{2}$$

where *Img(S)* and *Img(T)* represent the image samples (256 × 256 pixels) within the source domain *S* and target domain *T*, respectively, and *L(S)* denotes the class labels (i.e., "field" and "non-field") assigned to the pixels in *Img(S)*.

DA uses the output features of the target and source domains from the feature extractors (***F***) of the network as the discriminator (***D***) input. The discriminator indicates the probability of features being extracted from the target or source domains. A cross-entropy loss function (Eq. 3) based on the probability is formulated to train the discriminator as follows:

$$L_{\mathrm{D}} = -\sum_{i=1}^{n_s}(1-d)\log P(d=0|f_i) - \sum_{j=1}^{n_t} d\log P(d=1|f_j) \tag{3}$$

where f_i and f_j are the features extracted by ***F*** from source samples *i* and target sample *j*. $P(d=0)$ and $P(d=1)$ are the predicted probabilities of belonging to the source (0) and target (1) domains, respectively, based on the output by the discriminator. The loss function represents the discriminator's ability to determine features belonging to the source or target domain, with smaller values indicating higher domain discriminating accuracy. After training the discriminator, the feature extractor is trained by minimizing the combined loss function L_{fea} of segmentation loss (L_{seg}) and adversarial loss (L_{adv}) to obtain domain features that are discriminated by segmentation tasks but not by the discriminator ***D***:

$$L_{\mathrm{adv}} = -\sum_{i=1}^{n_s}\log P(d=0|f_i) \tag{4}$$

$$L_{fea} = L_{seg} + \lambda L_{adv} \tag{5}$$

where L_{seg} is the cross-entropy used to minimize the segmentation difference between the ground truth and prediction, and λ is the empirical weight used to balance segmentation and adversarial losses. Therefore, by alternately minimizing the loss functions (Eq. (3) and Eq. (5)), the modified feature extractors would produce features from the target domains that are similar to those extracted from the source domains.

2.2.2. FADA learning framework

The alignment of features in the same class between different domains is vital for improving classification accuracy. The FADA learning framework was improved from the DA framework by replacing the domain discriminator with a class-domain discriminator (Wang et al.,

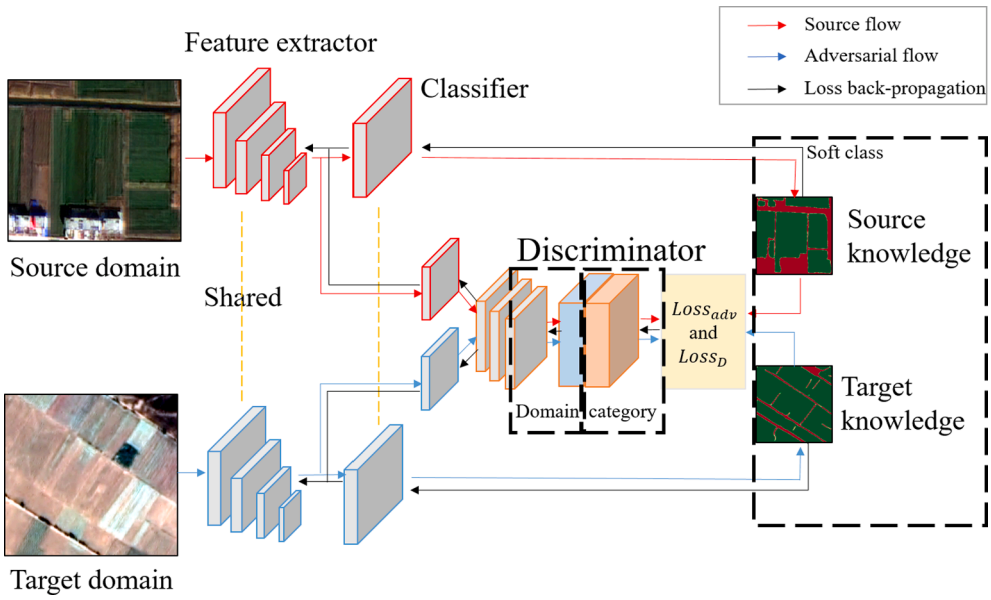

Fig. 3. Overview of the fine-grained adversarial domain adaptive framework (FADA) for field extraction.

2020). The loss function of the class-domain discriminator is formulated by additionally considering class label:

$$L_{\mathrm{D}} = -\sum_{i=1}^{n_s} a_i \log P(d=0, c=0|f_i) - \sum_{i=1}^{n_s} a_i \log P(d=0, c=1|f_i) - \sum_{j=1}^{n_s} a_j \log P(d=1, c=0|f_j) - \sum_{j=1}^{n_s} a_j \log P(d=1, c=1|f_j) \quad (6)$$

where d denotes the domain label (as in Eq. 3), and c is the field label extracted from classifier $\boldsymbol{C}$ ($c = 0$ refers to the IAF extent or boundary, and $c = 1$ refers to the non-extent or non-boundary). P is the discriminator's predicted probability of belonging to a different category of different domains, and a_i and a_j are the probabilities of the class soft label for the source domain's i_{th} sample and the target domain's j_{th} sample, respectively. Here, the probability was produced by classifier $\boldsymbol{C}$ trained from the source domain data because of the lack of ground-truth labels in the target domain.

The feature loss function L_{fea} was minimized to obtain the features for the segmentation task. Here, adversarial loss (L_{adv}) is different from that in Eq. 4 and encourages $\boldsymbol{F}$ to generate domain and class invariant features by the class information as follows:

$$L_{\mathrm{adv}} = -\sum_{i=1}^{n_s} a_i \log P(d=0, c=0|f_i) - \sum_{i=1}^{n_s} a_i \log P(d=0, c=1|f_i) \quad (7)$$

2.2.3. Training and inference

The FADA is implemented by alternatively optimizing feature extractor $\boldsymbol{F}$, classifier $\boldsymbol{C}$, and discriminator $\boldsymbol{D}$ in two steps until the Training converges.

In step one, the parameters of the already-trained feature extractor $\boldsymbol{F}$ and claissifier $\boldsymbol{C}$were fixed, and the output soft labels in the target and source domains were then used to train the discriminator by minimizing the discrimination loss (Eq. 6). In step two, the parameters of discriminator $\boldsymbol{D}$ were fixed, and the source domain data with labels and the target domain data without labels were used to train the feature extractor $\boldsymbol{F}$ and classifier $\boldsymbol{C}$ by minimizing both segmentation and adversarial loss (Eq. 7). During the inference process, discriminator $\boldsymbol{D}$ in the FADA was removed, and only the trained feature extractors $\boldsymbol{F}$ and classifier $\boldsymbol{C}$ were used to predict the segmentation results.

2.3. Connecting boundaries and filling field (CB-FF)

Owing to the inevitable misidentification of IAF extent and boundary by UNet and DeepLabV3+, the extracted boundary and extent were often too fragmented and aggregative, respectively, and thus incompatible. We used CB-FF to polish the extracted boundary and extent outputs (Fig. 4).

The first step of CB-FF is connecting boundaries (CB) with breakpoints. We used a sequence of morphological operations, "skeletonization" (Zhang and Suen 1984), " dilation", "skeletonization" and "dilation" operations, to connect most of the small breakpoints in the extracted boundaries (Fig. 4c). The first "skeletonization" operation reduced the extracted boundary regions into a centerline with one-pixel width, which helps to exhibit the breakpoints. The following "dilation" operation with a 5 × 5 square kernel connected the boundary's breakpoint to its nearest boundary with a distance of less than 5 pixels. However, the thickness of the boundaries became uneven after the "dilation" operation (Fig. 4d). Thus, the second "skeletonize" operation was executed again to extract centerlines with connected breakpoints (Fig. 4e). As the IAF boundaries (ridges, paths, and ditches) are commonly wider than one pixel in high-resolution images, a "dilation" operation with a 3 × 3 square kernel was executed to dilate the too thin centerlines (Fig. 4 f). The dilated boundary image was then merged with the original UNet produced boundary to retain as many of the boundary pixels as possible and create a polished boundary.

The second step involves fusing the boundary and extent images with a filling-field (FF) operation. The extent image was firstly clipped using the polished boundary (Fig. 4i); then, a morphological "closing" operation (i.e., to dilate an image and then erode the dilated image) with a 3 × 3 square kernel was used to fill each clipped IAF patch (Fig. 4j). Finally, the closed IAF patches were merged again to generate complete IAF map (Fig. 4k). With the CB-FF operation, we effectively eliminated the over-fragmented boundaries and split the over-aggregative extents in the original IAF boundary and extent images produced by UNet and DeeplabV3 +.

3. Study areas

We selected four farming areas (Fig. 5) around the cities of Funan (FN), Chengdu (CD), Heihe (HH), and Ruian (RA), China, which are characterized by different climates, crop types, and landscapes,

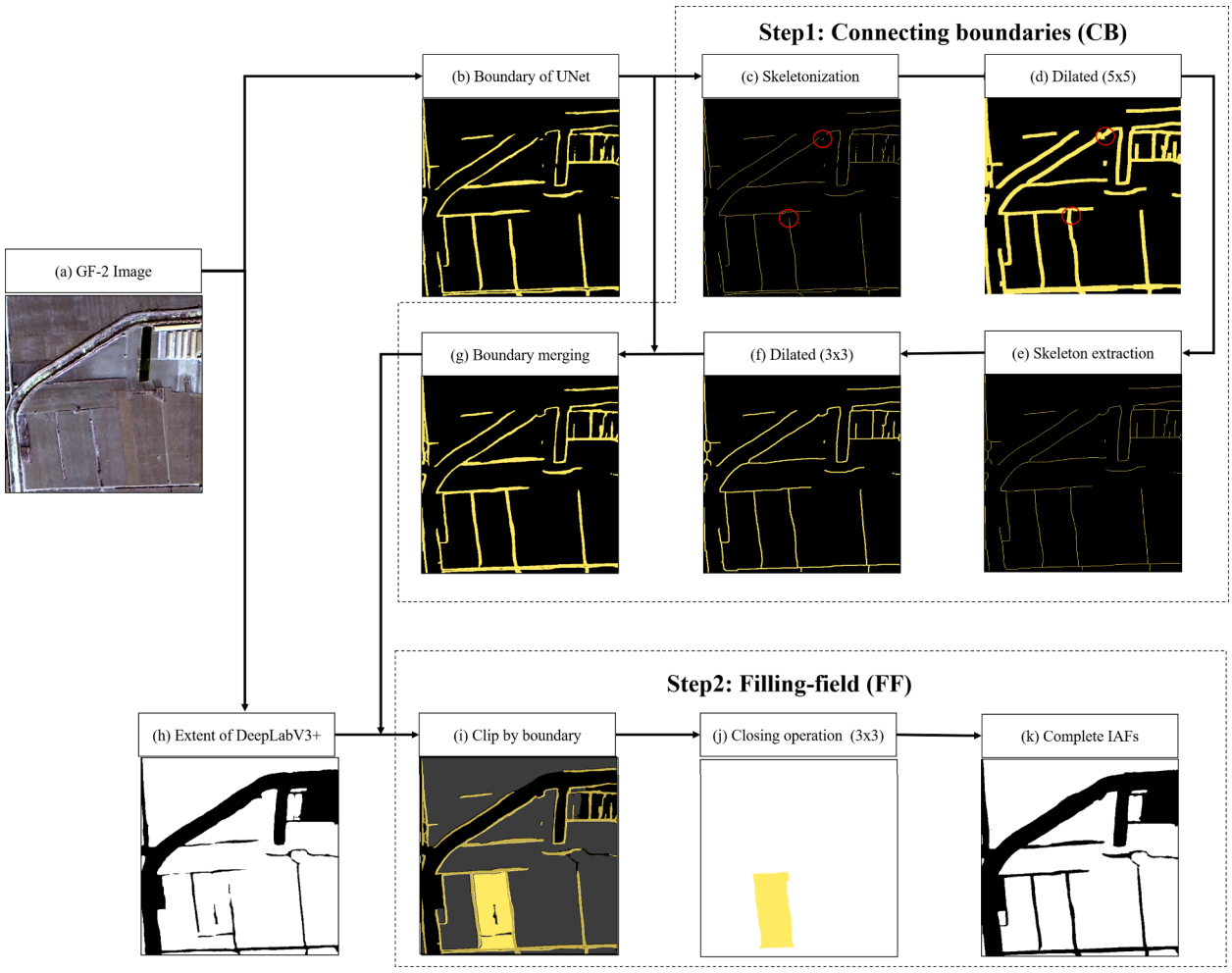

Fig. 4. Workflow of CB-FF post-processing module. The red circles represent examples of breakpoints connecting in the extracted boundary; the yellow patch represents an example of Filling-field for one IAF patch.

including the Koppen–Geiger climate type Cfa (humid subtropical climate, FN and RA), Dwb (monsoon-influenced warm summer humid continental climate, CD), and Dwc (monsoon-influenced subarctic climate, HH). These climates are associated with different crop types (e.g., maize, cotton, and soybeans) and cropping cycles. Landscape heterogeneity between the sites caused considerable variation in the size and shape of their respective IAFs. For instance, RA farmlands exhibited more fragmentation than those in CD, FN, and HH, which was related to complex river networks and mountainous terrain. We selected FN as the source domain based on the adequate number of labeled samples at this site. The transferability of the trained model was therefore assessed using CD, HH, and RA as the target domains.

3.1. Datasets

3.1.1. Images and preprocessing

We acquired four Gaofen-2 imageries in the four experimental areas. The four images were respectively captured in different seasons because of the limited revisiting frequency for the high-resolution data (Table 1). To obtain surface reflectance images, we performed orthorectification, radiometric calibration, and atmospheric correction (Table 1), which can reduce the radiometric inconsistency caused by varying topographical and atmospheric conditions. To maximize the use of spectral and spatial information from Gaofen-2 image data, the Gram-Schmidt Adaptive (GSA) was used to fuse the coarse spatial resolution (4 m) multispectral band image with the corresponding high spatial resolution (1 m) panchromatic band image. The surface reflectance images acquired had four multispectral bands at a spatial resolution of 1 m (see Table 2).

3.1.2. Training and validation samples

We digitized the sharpened Gaofen-2 images by marking the image pixels with IAF boundaries and extents (Fig. 6). The labels of the extents were created by manually digitizing all subareas. We applied a buffer of 1 m to the labels of the extent to create single-line vector data as boundary labels. These vector polygons were converted into a binary raster at a resolution of 1 m (Fig. 7). The source domain images and labels were split into small patches (256 × 256 pixels), and two-thirds of the patches were assigned to the training and validation sets, and the remaining one-third was assigned to the testing sets. For the target domain images, one-third of the patches randomly selected from the unlabeled dataset was used to adjust the training model by FADA, and the patches with labels were used for testing.

3.1.3. Data augmentation

The CNN can integrate spatial information but does not consider rotation invariance (Goodfellow et al., 2016). To model rotationally invariant features and enhance their generalizability across areas, we augmented the training samples with geometric transformations, including horizontal, vertical, and diagonal flips.

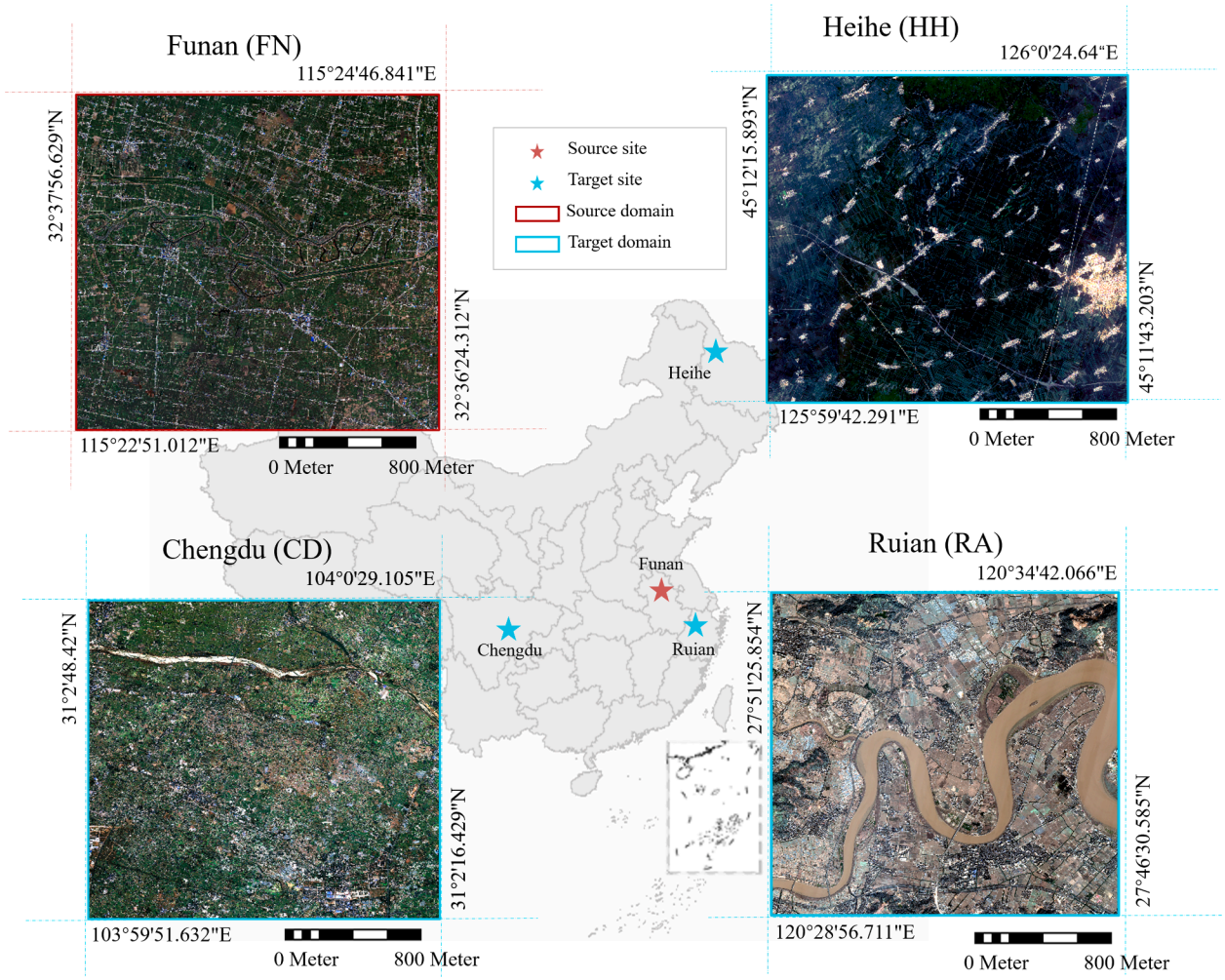

Fig. 5. Geographic locations of the experimental areas and the respective Gaofen-2 images. The red star represents the location of the source domain, and the blue stars represent the locations of the target domains. All images use red-greenblue as RGB.

Table 1
Description of the main sites.

Domain	Site	Location	Climate type	Main characteristics	Main crops	Acquisition date
Source	Funan (FN)	32°36 N, 115°24E	Arid to humid continental climate	Mostly flat	Soybean, maize, and wheat	2020/7/29
Target	Chengdu (CD)	31°2′N, 104°0′E	Monsoon climate	Flat plains to mountainous	Wheat, barley, canola, and legume crops	2018/4/21
	Heihe (HH)	45°11′N, 126°0′E	Temperate climate	Flat	Rice, wheat, sunflower, and soybean	2017/7/31
	Ruian (RA)	27°46′N, 120°34′E	Subtropical monsoon climate	Gently undulating landscape	Rice, maize, barley, potatoes, and soybean	2021/1/17

Table 2
The acquisition times and band settings of the Gaofen-2 images.

	Band number	Spectral Range (μm)	Resolution (m)
Panchromatic	1	0.45–0.90	1
Multispectral	2	0.45–0.52	4
	3	0.52–0.59	
	4	0.63–0.69	
	5	0.77–0.89	

3.2. Accuracy assessment

We assessed the accuracy of IAF segmentation using the "precision-recall" framework (Martin et al., 2004). Precision and recall are derived from a confusion matrix with four elements of "true positive" *(TF)*, "false positive" (Asefpour Vakilian and Saradjian), "false negative" *(FN)*, and "true negative" *(TN)*:

$$Precision = \frac{TP}{TP + FP} \tag{7}$$

$$Recall = \frac{TP}{TP + FN} \tag{8}$$

The F_1-*score* is commonly used to evaluate classification accuracy as

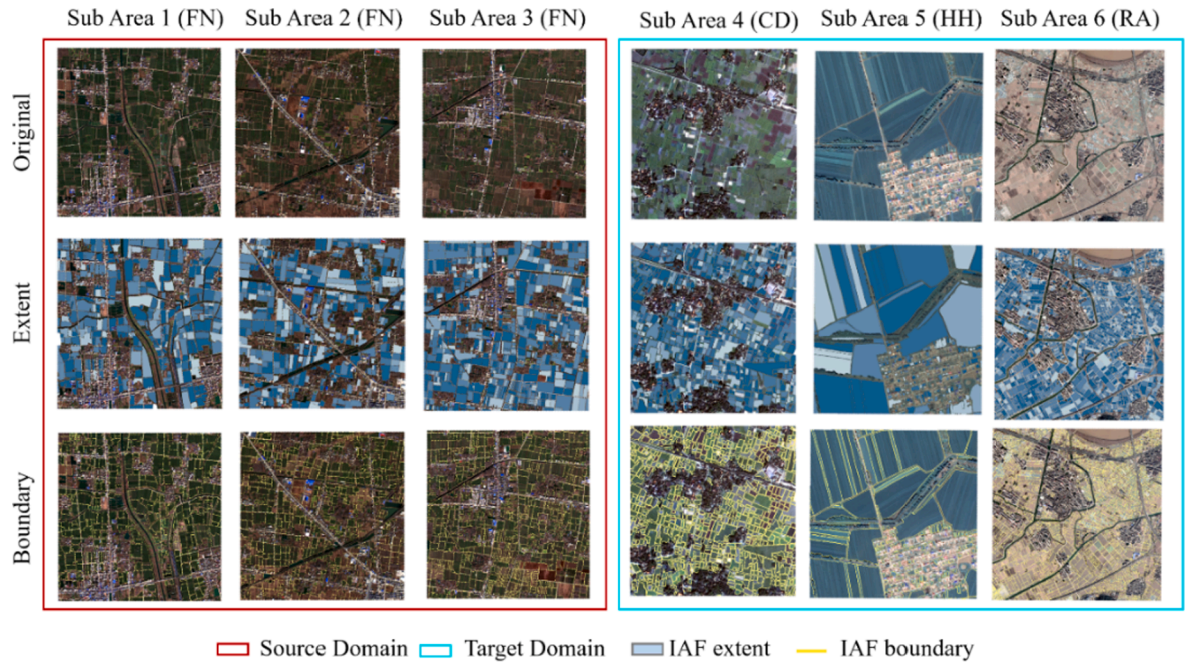

Fig. 6. Ground truth data of the extents and boundaries from the training and test images sampled at different sites (the red, green, and blue bands are the RGB composite in these images).

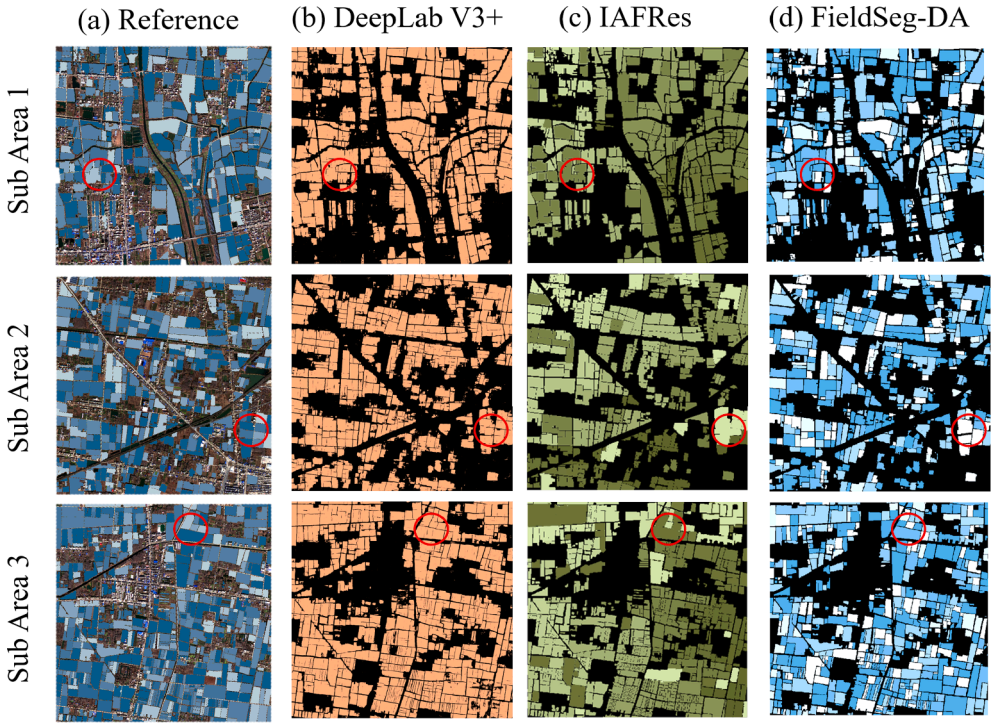

Fig. 7. IAF extraction in three subareas of Ruian (RA, Subareas 1–3) based on (a) ground truth labels (reference), (b). DeepLabV3+ (orange), (c) IAFRes (green), and (d) FieldSeg-DA (blue). The different colors within each map are used to illustrate different IAF patches.

follows:

$$F_1 \text{ - score} = 2 \times \frac{Precision \times Recall}{Precision + Recall} \tag{9}$$

We also used the Matthews correlation coefficient *(MCC)* (Matthews, 1975)—an indicator representing both "Precision" and "Recall" information—to evaluate the IAF extraction accuracy as follows:

$$MCC = \frac{TP \times TN - FP \times FN}{\sqrt{(TP + FP) \times (TP + FN) \times (TN + FN) \times (TN + FP)}} \tag{10}$$

We used *MCC* to select an optimal IAF segmentation model by assessing the training and validation losses of different models in the validation stage. We used the eccentricity coefficient (ε, Eq. 11) to evaluate the differences in IAF geometric characteristics (size and shape) between the extracted and reference IAF objects (Persello and Bruzzone

2009):

$$\varepsilon = \frac{1}{N}\sum_{i=0}^{N}\left\|Eccentricity_{ip}/Eccentricity_{ir}\right\| \quad (11)$$

where $Eccentricity_r$ relates to the reference instance, and $Eccentricity_p$ relates to the prediction results.

4. Results

4.1. Results of source domain images

The IAF extents extracted using FieldSeg-DA (Fig. 7d) were generally consistent with the ground truth label (Fig. 7a). We compared our results with those generated by IAFRes, a state-of-the-art method for IAF extraction, and DeepLabV3+, the branch network used for extracting IAF extent in our proposed model. Compared to FieldSeg-DA, DeepLabV3+ and IAFRes tended to mistakenly merge multiple distinct IAFs (red circle in Fig. 7c-d) because the boundaries were not well recognized. The FieldSeg-DA model also achieved higher accuracy metrics (i.e., *F_1-score*, *MCC*, and *Eccentricity*) than IAFRes and DeepLabV3+ in all three subareas (Table 3).

The introduction of parallel network structure (EBPN) and CB-FF post-processing are essential enhancements to the FieldSeg-DA model over previous networks. To further demonstrate the effectiveness of the parallel network structure, we specifically analyzed how these modules improve boundary accuracy. The boundary of the IAF extent extracted by DeepLabV3+ (Fig. 8b and Table 4), which was obtained using a buffer function from the extracted extent, omitted many boundaries compared to the ground truth boundary label (Fig. 8a). The UNet branch, which was specifically trained for extracting boundaries, better recognized most IAF boundaries but showed fragmented results (Fig. 8c). After CB-FF processing, the extracted boundaries showed the highest consistency with the ground truth labels (Fig. 8d). The quantitative accuracy also illustrates the improvement of EBPN and CB-FF for the boundary extraction. The FieldSeg-DA model generated boundary achieved the highest accuracy, followed by the UNet branch and DeeplabV3+ branch (Table 4). These results illustrate that the EBPN and CB-FF modules improved boundary identification and thus IAF extraction.

4.2. Results of target domain images

Regarding the target domain images without labeled samples, FieldSeg-DA still produced an IAF extent with good consistency to the reference map (Fig. 9a and d). In contrast, IAFRes omitted many IAF extents (red circle in Fig. 9b). The FieldSeg-DA model showed an improvement of nearly 9% for all three accuracy metrics compared with the IAFRes model (Table 5). The difference in accuracy between the FieldSeg-DA and IAFRes method was even more pronounced in the target than the source domain because of the differences in spectral and morphological characteristics of the IAFs.

To validate the effect of the FADA module, we compared the performance of the FieldSeg-DA model to one without the FADA module (FieldSeg) in the target domain. Although it performed slightly better than IAFRes, the FieldSeg model performed much worse than the FieldSeg-DA. These results confirm the role of the FADA module in enhancing the transferability of the FieldSeg-DA model.

Table 3
Assessment of the IAF accuracy in the source domain using different methods.

Area		Sub Area1	Sub Area2	Sub Area3	Overall
F1-score	DeepLabV3+	0.8639	0.8314	0.8416	0.8456
	IAFRes	0.8741	0.8966	0.8832	0.8846
	FieldSeg-DA	0.9041	0.907	0.8916	0.9009
MCC	**DeepLabV3+**	0.8782	0.8627	0.8543	0.8651
	IAFRes	0.8537	0.8621	0.8194	0.8451
	FieldSeg-DA	0.8933	0.8864	0.8603	0.88
Eccentricity (ε)	**DeepLabV3+**	0.6927	0.7352	0.7123	0.7134
	IAFRes	0.8317	0.8286	0.8287	0.8297
	FieldSeg-DA	0.8511	0.8737	0.8668	0.8639

5. Discussion

5.1. The superiority of the parallel model structure

The parallel structure in FieldSeg-DA allowed the network to simultaneously estimate the extents and boundaries of IAFs, which is considered fundamental for IAF extraction (Waldner and Diakogiannis 2020). IAFRes uses a multitask head model, which could increase training difficulty because of variation in the classified features. In comparison, parallel networks in the FieldSeg-DA model were trained independently, allowing different feature extractors for the two tasks. This parallel structure also allows for flexibility in the use of alternative semantic segmentation models. We compared the accuracies and efficiencies of several popular semantic segmentation models and found that FPN model had the highest final segmentation accuracy (F_1-score = 0.9209) for extent extraction among all available networks (Table 6). DeepLabV3+ ranked second, though it performed more efficiently than the FPN. Therefore, we propose that DeepLabV3+ and FPN model are the optimal extent extraction branches, depending on the user's need for accuracy and efficiency.

Compared with the alternatives for boundary extraction, the UNet model achieved the second-best final accuracy and highest efficiency (1.32 min, Table 7). Although Manet achieved the highest accuracy on the test dataset (F_1-score of 0.9158), the difference between the training and validation loss was relatively large (0.0485), suggesting a high "generalization gap" (Hoffer et al., 2017). Therefore, UNet is the recommended boundary extraction branch owing to its excellent computational efficiency and segmentation accuracy.

5.2. Effectiveness of domain adaptation

FADA is an essential part of the FieldSeg-DA model, which improves the extraction of classification features from the feature extractor of FieldSeg-DA in the target domain. To illustrate the efficacy of this approach, we used the T-SNE method to show the distribution of features in a 2-D space (Van der Maaten and Hinton 2008). The extent (boundary) and non-extent (non-boundary) feature extracted by FieldSeg without FADA exhibited poor separability (Fig. 10). These two types of features extracted from the CD and RA regions almost overlapped in the feature space, leading to poor performance (F_1-score: 0.68) in extracting the IAF. In contrast, FieldSeg-DA adaptively adjusts the feature extractor, allowing for good separability between these features in the target domain, similar to that in the source domain (Fig. 10). Thus, the FieldSeg-DA model performed robustly in extracting IAFs over varying target domains without additional training samples, which effectively enhances its applicability in diverse regions.

6. Perspectives and limitations

Our results confirmed the effectiveness of the FADA module in enhancing the transferability of FieldSeg-DA to the target domain without additional training samples. However, the performance of FieldSeg-DA degraded from the source domain to the target domain (Tables 3 and 5). The degradation of FieldSeg-DA is different in different regions, depending on the feature similarity between the target and the source domains (Fig. 10). The feature similarity could be related to the differences in the agricultural characteristics (e.g., crop system, climate

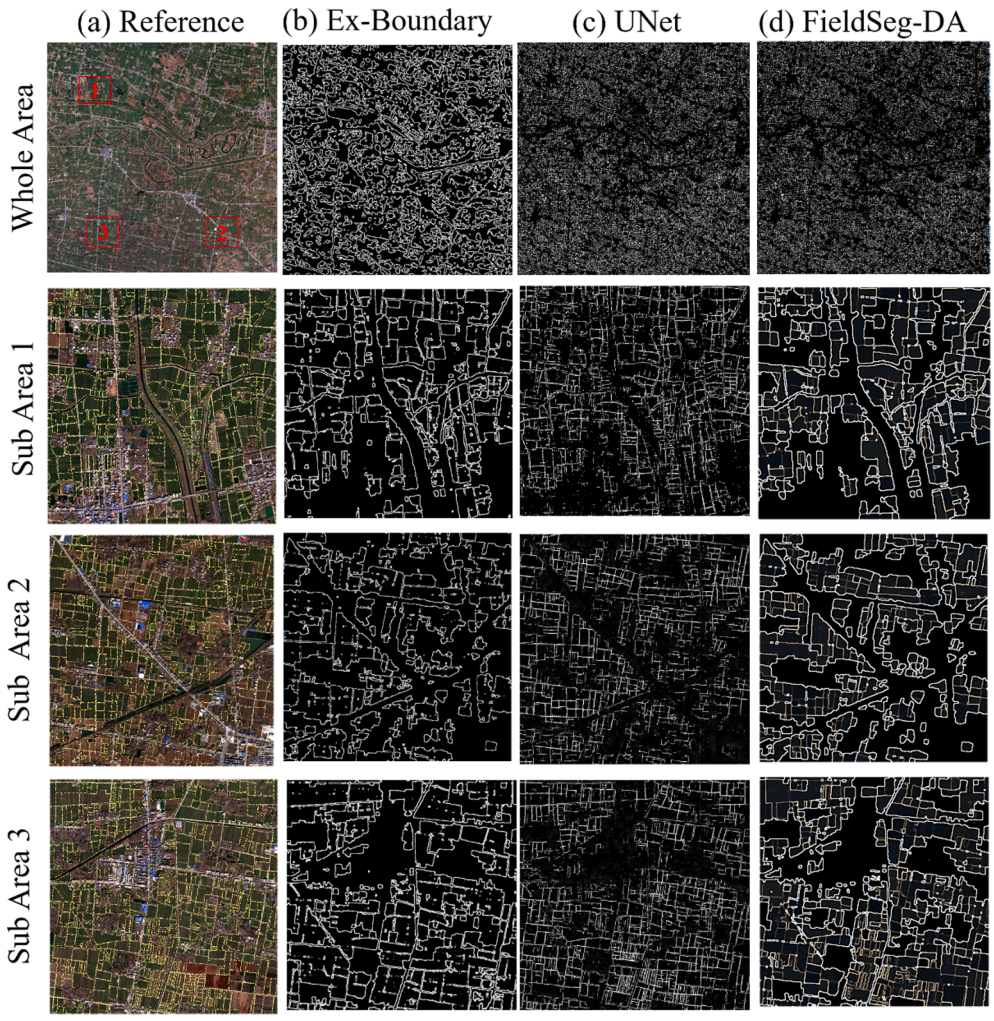

Fig. 8. IAF extraction in Ruian (RA). (a) Actual Gaofen-2 image; (b) IAF boundaries obtained by the 1-pixel buffer function of IAF extent produced by DeepLabV3+; (c) IAF boundary predicted by UNet; (d) final IAF boundary generated by FieldSeg-DA. The three red rectangles in the top-left inset indicate the locations of the three subareas (Subareas 1–3) of the source domain.

Table 4
Assessment of the boundary accuracy of source domain experiments.

Area		Sub Area1	Sub Area2	Sub Area3	Overall
F_1-score	Ex-Boundary	0.7717	0.7879	0.6287	0.7372
	UNet	0.7834	0.7962	0.6913	0.757
	FieldSeg-DA	0.8152	0.8303	0.8164	0.8206
MCC	Ex-Boundary	0.7109	0.6917	0.6609	0.6891
	UNet	0.7296	0.6915	0.6661	0.6957
	FieldSeg-DA	0.7881	0.8144	0.7659	0.7895
Eccentricity (e)	Ex-Boundary	0. 3144	0.3569	0.2839	0.3226
	UNet	0.3174	0.3691	0.2959	0.3275
	FieldSeg-DA	0.4383	0.4724	0.4342	0.4483

type, topography, IAF size, and shape), as well as the difference in the image acquisition season (Table. 1) between the target and the source domains. Particularly, the difference in the acquisition season between target and source domains was the most important factor affecting the model generalizability in our experiment. As shown in Table 5, all the compared models achieved the lowest accuracy in the RA site because the image in the RA site was acquired in the winter season, whereas the image in the source domain (FN site) was acquired in the summer season (Table 1).

Supervised domain alignment with additional training samples may improve the efficiency of FADA and the overall performance of the FieldSeg-DA model. Although precise samples are difficult to obtain, rough samples with relatively large uncertainties may be helpful for feature alignment. For example, recently developed land cover products at coarser resolution, such as GlobeLand30 (Chen et al., 2015), ESRI Global Land Cover (Karra et al., 2021), Dynamic World Land Cover (Brown et al., 2022), could provide rough samples for supervised domain alignment and, improve the transferability of FieldSeg-DA.

Another shortcoming of FieldSeg-DA is the CB-FF post-processing module. As a traditional image processing method, a sequence of morphological operations are designed based on empirical experience, of which the operational sequence and relevant parameters (e.g., kernel size) might need careful adjusting according to the different target domains. In future work, establishing an end-to-end framework for combing the two branch networks without empirical post-processing would be crucial to reducing the reliance on expert interactions.

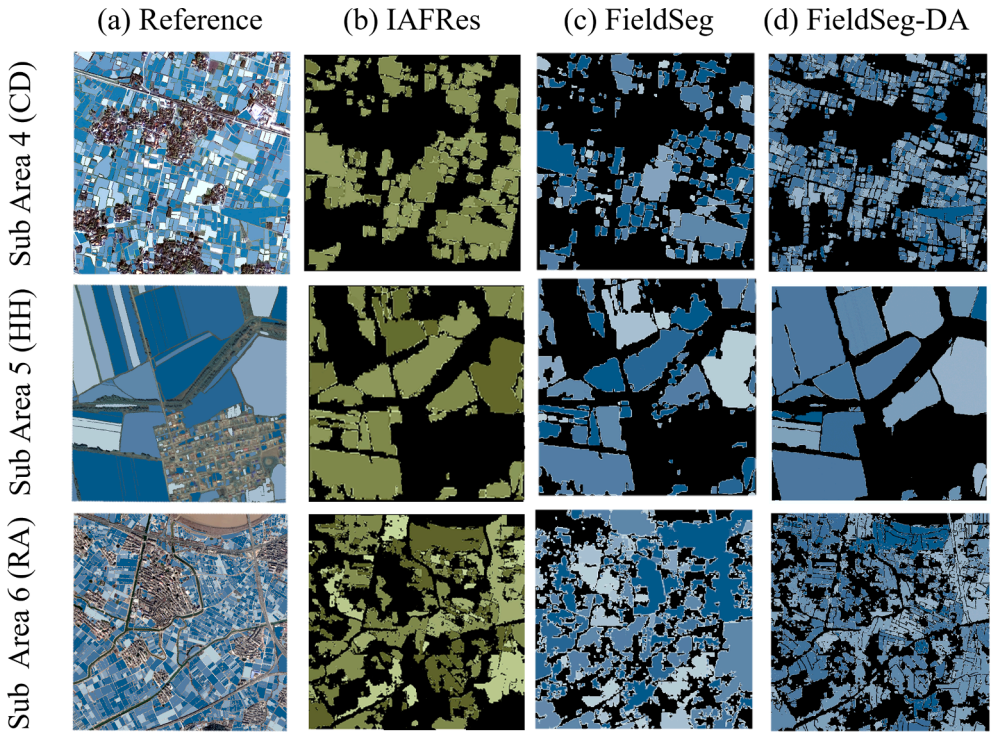

Fig. 9. IAF extraction from the target domains using the IAFRes, FieldSeg (red dashed box), and FieldSeg-DA model (blue dashed box).

Table 5
Assessment of the accuracy of target domain experiments.

Training mode		CD	HH	RA	Overall
F_1-score	**IAFRes**	0.7034	0.6783	0.6109	0.6642
	FieldSeg	0.7233	0.6853	0.6371	0.6819
	FieldSeg-DA	0.8012	0.7643	0.6881	0.7512
MCC	**IAFRes**	0.6854	0.6387	0.5529	0.6257
	FieldSeg	0.696	0.6419	0.5761	0.638
	FieldSeg-DA	0.7712	0.7188	0.626	0.7054
Eccentricity (ε)	**IAFRes**	0.7132	0.6652	0.5392	0.6392
	FieldSeg	0.7546	0.7061	0.5496	0.6701
	FieldSeg-DA	0.8224	0.7757	0.5962	0.7314

Table 6
Assessment of IAF extent extraction between different semantic segmentation models.

Training mode	Training loss	Validation loss	Test F_1-score	Efficiency Execution Time per scene (min)
UNet	0.1083	0.1237	0.8287	**1.4300**
UNet++	0.1224	0.1176	0.7979	1.5100
Manet	0.1342	0.1271	0.8287	2.1200
LinkNet	0.1052	0.0981	0.8772	1.8700
FPN	0.1203	0.1344	**0.9209**	2.3100
PSPNet	0.1464	0.1659	0.8268	2.5600
DeepLabV3	0.1031	0.1197	0.8679	1.7300
DeepLabV3+	**0.0972**	**0.1034**	0.8926	1.8700

7. Conclusions

This study proposes a novel FieldSeg-DA network for automatically extracting IAFs from high-spatial-resolution satellite imagery. We present the problem as semantic segmentation and domain-shift tasks. The EBPN parallel network structure and CB-FF post-processing module

Table 7
Assessment of IAF boundary extraction between different semantic segmentation models.

Training mode	Training loss	Validation loss	Test F_1-score	Efficiency Execution Time per scene (min)
UNet	0.1095	0.131	**0.9149**	**1.320**
UNet++	0.1118	0.1326	0.9117	1.570
Manet	**0.0801**	**0.1285**	**0.9158**	2.310
LinkNet	0.0947	0.1344	0.9091	2.110
FPN	0.1119	0.1369	0.9009	1.820
PSPNet	0.1017	0.1339	0.9086	3.120
DeepLabV3	0.0979	0.1326	0.9099	2.170
DeepLabV3+	0.0912	0.1316	0.9117	2.340

improved the prediction of the IAF extent and boundary by combining different branch networks with different preferences in identifying "face" and "line" objects in remote sensing images. FADA enhanced domain transferability between study areas without labeled samples. The proposed FieldSeg-DA model outperformed other state-of-the-art methods in both the source and target domains, confirming its effectiveness in extracting IAFs across diverse areas without training samples. Finally, owing to the flexibility of replacing the branch network in the FieldSeg-DA, the performance of the model should continue to improve with the development of better semantic segmentation models.

CRediT authorship contribution statement

Shuaijun Liu: Conceptualization, Methodology, Software, Formal analysis, Investigation, Writing – original draft, Writing – review & editing, Visualization. **Licong Liu:** Conceptualization, Investigation, Methodology. **Fei Xu:** Conceptualization, Investigation, Methodology. **Jin Chen:** Conceptualization, Methodology, Resources, Writing – review & editing, Project administration, Funding acquisition. **Yuhen Yuan:** Formal analysis, Investigation. **Xuehong Chen:** Conceptualization, Methodology, Resources, Writing – review & editing, Project administration, Funding acquisition.

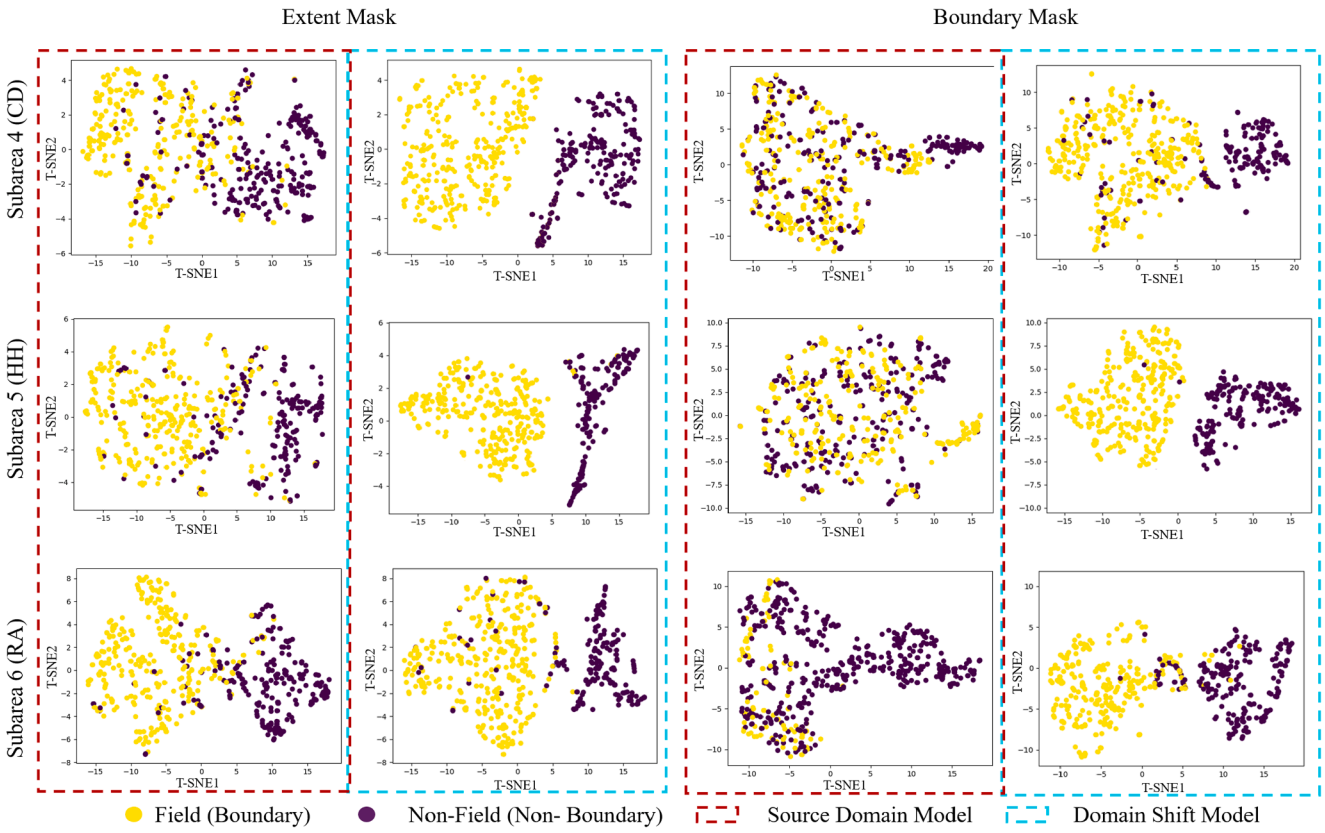

Fig. 10. Features extracted using t-SNE for different target domains (Van der Maaten and Hinton 2008).

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.

Data availability

Data will be made available on request.

Acknowledgement

This study is supported by National Natural Science Foundation of China (No. 41871224).

References

Aquino, A., Gegúndez-Arias, M.E., Marín, D., 2010. Detecting the optic disc boundary in digital fundus images using morphological, edge detection, and feature extraction techniques. IEEE Trans. Med. Imaging 29, 1860–1869.
Asefpour Vakilian, A., Saradjian, M.R., 2022. An object-based sparse representation model for spatiotemporal image fusion. Sci Rep 12, 5021.
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.
Blaes, X., Vanhalle, L., Defourny, P., 2005. Efficiency of crop identification based on optical and SAR image time series. Remote Sens. Environ. 96, 352–365.
Brown, C.F., Brumby, S.P., Guzder-Williams, B., Birch, T., Hyde, S.B., Mazzariello, J., Czerwinski, W., Pasquarella, V.J., Haertel, R., Ilyushchenko, S., 2022. Dynamic World, Near real-time global 10 m land use land cover mapping. Sci. Data 9, 1–17.
Chen, J., Chen, J., Liao, A., Cao, X., Chen, L., Chen, X., He, C., Han, G., Peng, S., Lu, M., 2015. Global land cover mapping at 30 m resolution: A POK-based operational approach. ISPRS J. Photogramm. Remote Sens. 103, 7–27.
Da Costa, J.P., Michelet, F., Germain, C., Lavialle, O., Grenier, G., 2007. Delineation of vine parcels by segmentation of high resolution remote sensed images. Precis. Agric. 8, 95–110.
De Wit, A., Clevers, J., 2004. Efficiency and accuracy of per-field classification for operational crop mapping. Int. J. Remote Sens. 25, 4091–4112.
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.
Duan, L., Tsang, I.W., Xu, D., 2012. Domain transfer multiple kernel learning. IEEE Trans. Pattern Anal. Mach. Intell. 34, 465–479.
Gao, X., Liu, L., Gong, H., 2020. MMUU-Net: A Robust and Effective Network for Farmland Segmentation of Satellite Imagery. In: Journal of Physics: Conference Series, IOP Publishing, p. 012189.
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 (7), 1809–1819.
Goodfellow, I., Bengio, Y., Courville, A., 2016. Deep learning. MIT press.
Hoffer, E., Hubara, I., Soudry, D., 2017. Train longer, generalize better: closing the generalization gap in large batch training of neural networks. Advances in neural information processing systems, 30.
Hosseinpour, H., Samadzadegan, F., Javan, F.D.D., 2022. A Novel Boundary Loss Function in Deep Convolutional Networks to Improve the Buildings Extraction from High-Resolution Remote Sensing Images. IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.
Huang, H., Lin, L., Tong, R., Hu, H., Zhang, Q., Iwamoto, Y., Han, X., Chen, Y.-W., Wu, J., 2020. Unet 3+: A full-scale connected unet for medical image segmentation. In: ICASSP 2020-2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). IEEE, pp. 1055–1059.
Karra, K., Kontgis, C., Statman-Weil, Z., Mazzariello, J.C., Mathis, M., Brumby, S.P., 2021. In: Global land use/land cover with Sentinel 2 and deep learning. IEEE, pp. 4704–4707.
Li, X., Luo, M., Ji, S., Zhang, L., Lu, M., 2020. Evaluating generative adversarial networks based image-level domain transfer for multi-source remote sensing image segmentation and object detection. Int. J. Remote Sens. 41, 7343–7367.
Liu, X., Chi, M., Zhang, Y., Qin, Y., 2018. In: Classifying high resolution remote sensing images by fine-tuned VGG deep networks. IEEE, pp. 7137–7140.
Liu, Y., Dobrinsky, A., Yakobson, B.I., 2010. Graphene edge from armchair to zigzag: the origins of nanotube chirality? Phys. Rev. Lett. 105 (23), 235502.
Long, J., Shelhamer, E., Darrell, T., 2015. Fully convolutional networks for semantic segmentation. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 3431–3440.
Ma, L., Liu, Y., Zhang, X., Ye, Y., Yin, G., Johnson, B.A., 2019. Deep learning in remote sensing applications: A meta-analysis and review. ISPRS J. Photogramm. Remote Sens. 152, 166–177.
Martin, D.R., Fowlkes, C.C., Malik, J., 2004. Learning to detect natural image boundaries using local brightness, color, and texture cues. IEEE Trans. Pattern Anal. Mach. Intell. 26, 530–549.

Matasci, G., Volpi, M., Kanevski, M., Bruzzone, L., Tuia, D., 2015. Semisupervised transfer component analysis for domain adaptation in remote sensing image classification. IEEE Trans. Geosci. Remote Sens. 53, 3550–3564.
Matthews, B.W., 1975. Comparison of the predicted and observed secondary structure of T4 phage lysozyme. Biochim. Biophys. Acta (BBA)-Protein Struct. 405, 442–451.
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 Sensing 7, 13208–13232.
Meyer, L., Lemarchand, F., Sidiropoulos, P., 2020. A deep learning architecture for batch-mode fully automated field boundary detection. Int. Arch. Photogrammetry, Remote Sens. Spatial Inform. Sci. 43, 1009–1016.
Mosinska, A., Marquez-Neila, P., Koziński, M., Fua, P., 2018. Beyond the pixel-wise loss for topology-aware delineation. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 3136–3145.
Nogueira, K., Penatti, O.A.B., dos Santos, J.A., 2017. Towards better exploiting convolutional neural networks for remote sensing scene classification. Pattern Recogn. 61, 539–556.
Peng, L., Chen, X., Chen, J., Zhao, W., Cao, X., 2022. Understanding the Role of Receptive Field of Convolutional Neural Network for Cloud Detection in Landsat 8 OLI Imagery. IEEE Trans. Geosci. Remote Sens. 60, 1–17.
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.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.
Rabbi, J., Ray, N., Schubert, M., Chowdhury, S., Chao, D., 2020. Small-object detection in remote sensing images with end-to-end edge-enhanced GAN and object detector network. Remote Sensing 12, 1432.
Rahman, M.S., Mohiuddin, H., Kafy, A.A., Sheel, P.K., Di, L., 2019. Classification of cities in Bangladesh based on remote sensing derived spatial characteristics. J. Urban Manag. 8 (2), 206–224.
Segl, K., Kaufmann, H., 2001. Detection of small objects from high-resolution panchromatic satellite imagery based on supervised image segmentation. IEEE Trans. Geosci. Remote Sens. 39, 2080–2083.
Turker, M., Kok, E.H., 2013. Field-based sub-boundary extraction from remote sensing imagery using perceptual grouping. ISPRS J. Photogramm. Remote Sens. 79, 106–121.
Van der Maaten, L., Hinton, G., 2008. Visualizing data using t-SNE. J. Mach. Learn. Res. 9.
Verrelst, J., Rivera, J.P., Leonenko, G., Alonso, L., Moreno, J., 2014. Optimizing LUT-Based RTM Inversion for Semiautomatic Mapping of Crop Biophysical Parameters from Sentinel-2 and -3 Data: Role of Cost Functions. IEEE Trans. Geosci. Remote Sens. 52, 257–269.
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.
Waldner, F., Hansen, M.C., Potapov, P.V., Löw, F., Newby, T., Ferreira, S., Defourny, P., 2017. National-scale cropland mapping based on spectral-temporal features and outdated land cover information. PLoS ONE 12, e0181911.
Wang, H., Shen, T., Zhang, W., Duan, L.-Y., Mei, T., 2020. In: Classes matter: A fine-grained adversarial approach to cross-domain semantic segmentation. Springer, pp. 642–659.
Xi, J., Zhang, J.-Z., 2012. In: Edge detection from remote sensing images based on Canny operator and Hough transform. Springer, pp. 807–814.
Yan, J.-Y., Lv, Q.-T., Chen, M.-C., Deng, Z., Qi, G., Zhang, K., Liu, Z.-D., Wang, J., Liu, Y., 2015. Identification and extraction of geological structure information based on multi-scale edge detection of gravity and magnetic fields: An example of the Tongling ore concentration area. Chin. J. Geophys. 58, 4450–4464.
Yan, L., Roy, D.P., 2014. Automated crop field extraction from multi-temporal Web Enabled Landsat Data. Remote Sens. Environ. 144, 42–64.
Yang, G., Zhang, Q., Zhang, G., 2020. EANet: Edge-aware network for the extraction of buildings from aerial images. Remote Sensing 12, 2161.
Yoo, D., Kim, N., Park, S., Paek, A.S., Kweon, I.S., 2016. In: Pixel-level domain transfer. Springer, pp. 517–532.
Zhang, T.Y., Suen, C.Y., 1984. A fast parallel algorithm for thinning digital patterns. Commun. ACM 27, 236–239.
Zhou, Y.-T., Venkateswar, V., Chellappa, R., 1989. Edge detection and linear feature extraction using a 2-D random field model. IEEE Trans. Pattern Anal. Mach. Intell. 11, 84–95.