

## Original papers

# FieldSeg-DA2.0: Further enhancing the spatiotemporal transferability of an individual arable field (IAF) extraction network using multisource remote sensing and land cover data

Chun Tian<sup>a,b</sup>, Xuehong Chen<sup>a,b,\*</sup>, Jin Chen<sup>a,b</sup>, Ruyin Cao<sup>c</sup>, Shuaijun Liu<sup>a,b</sup>

<sup>a</sup> State Key Laboratory of Remote Sensing Science, Institute of Remote Sensing Science and Engineering, Faculty of Geographical Science, Beijing Normal University, Beijing 100875, China

<sup>b</sup> Beijing Engineering Research Center for Global Land Remote Sensing Products, Institute of Remote Sensing Science and Engineering, Faculty of Geographical Science, Beijing Normal University, Beijing 100875, China

<sup>c</sup> School of Resources and Environment, University of Electronic Science and Technology of China, Chengdu, Sichuan 611731, China



## ARTICLE INFO

## Keywords:

Individual arable field  
Spatiotemporal transferability  
Spatiotemporal fusion  
Adversarial domain adaptation  
Land cover prior

## ABSTRACT

Deep learning has become the leading technique for precisely extracting individual arable fields (IAFs) from high-resolution remote sensing images. Maintaining transferability remains a major concern for deep learning methodologies due to the significant cost of acquiring labelled samples for network training. FieldSeg-DA, introduced by Liu et al. (2022), enhanced network transferability by introducing a finely tuned adversarial domain adaptation module (FADA). An improved version of the FieldSeg-DA framework, FieldSeg-DA2.0, which further enhances transferability through incorporating multisource remote sensing and land cover data is introduced in this paper. First, we introduce a spatiotemporal fusion module, U-LSTM, to extract the IAF extent by merging textural information from the high-resolution image (Gaofen-2) and phenological information from the coarse time-series data (Sentinel-2). Incorporating phenological information mitigates the risk of overfitting associated with the high-resolution imagery acquired in specific seasons, thereby improving the temporal transferability of FieldSeg-DA2.0 compared to FieldSeg-DA. Second, we introduce a novel fine-grained adversarial domain adaptation module with ancillary data (FADA-A) to enhance spatial transferability. FADA-A incorporates prior knowledge from the Dynamic World (DW) land cover dataset to guide adversarial training in the standard FADA, thereby enhancing the robustness of domain adaptation across diverse geographic regions. We evaluate the performance of FieldSeg-DA2.0 through various spatial and temporal transfer experiments utilizing GaoFen-2 and Sentinel-2 data. The results illustrate that the cross-domain performance of FieldSeg-DA2.0 is significantly better than that of the original FieldSeg-DA, highlighting its robust spatiotemporal transferability. FieldSeg-DA2.0 can accurately delineate IAFs across varied regions and seasons without requiring additional training samples, illustrating its considerable potential for large-scale IAF extraction.

## 1. Introduction

An individual farm arable field (IAF), also referred to as a crop parcel, is a land patch used for crop cultivation, where either a single crop is planted or several crops are regularly intercropped during each growing season. IAFs are often delineated by ridges, paths, and ditches, serving as fundamental spatial units for agricultural activities (Liu et al., 2022; Persello et al., 2019). Precise information about IAFs not only assists in formulating agricultural land management, such as land

ownership determination, water and irrigation allocation (Luo et al., 2021), and agricultural disaster insurance (Aung et al., 2020), but also facilitates the crop cultivation area estimation, crop yield prediction, and crop classification (Waldner and Diakogiannis, 2020).

Remote sensing imagery is an important data source for IAF extraction. Efforts have been made to automatically extract IAFs from high-resolution satellite images. Traditionally, empirical graphical operators (e.g., Sobel, Canny) and image segmentation techniques (e.g., watershed segmentation, multiresolution segmentation (Wen et al.,

\* Corresponding author at: State Key Laboratory of Remote Sensing Science, Institute of Remote Sensing Science and Engineering, Faculty of Geographical Science, Beijing Normal University, Beijing 100875, China.

E-mail address: [chenxuehong@bnu.edu.cn](mailto:chenxuehong@bnu.edu.cn) (X. Chen).

<https://doi.org/10.1016/j.compag.2024.109050>

Received 10 December 2023; Received in revised form 8 May 2024; Accepted 13 May 2024

Available online 18 May 2024

0168-1699/© 2024 Elsevier B.V. All rights reserved, including those for text and data mining, AI training, and similar technologies.

2022)) have been used to identify IAF boundaries and parcels, respectively (Da Costa et al., 2007; Mueller et al., 2004). However, these methods explore only shallow features, failing to exploit rich spatial context in high-resolution images and thus having limited accuracy in IAF extraction. In recent years, deep neural networks, as an effective technology for exploring spatial and spectral features across multiple scales (Ma et al., 2024; Yan et al., 2023), have also shown great potential for extracting IAFs from high-resolution satellite imagery (Aung et al., 2020; Fare Garnot and Landrieu, 2021; Li et al., 2023; Liu et al., 2022; Long et al., 2022; Matton et al., 2015; Persello et al., 2019; Sun et al.,

2022; Waldner and Diakogiannis, 2020). Although deep segmentation networks can be directly applied for IAF extent extraction by treating the extraction as a pixelwise classification task of cropland and non-cropland (Cai et al., 2023b; Zhang et al., 2020), these networks frequently exhibit adhesion issues due to the lack of attention on the IAF boundaries. Thus, the delineation of IAF boundaries has received increasing attention. Various deep segmentation models, e.g., FCN (Long et al., 2015), SegNet (Badrinarayanan et al., 2017), and UNet, have been applied for IAF boundary identification (Jong et al., 2022; Masoud et al., 2019; Persello et al., 2019; H. Zhang et al., 2021).

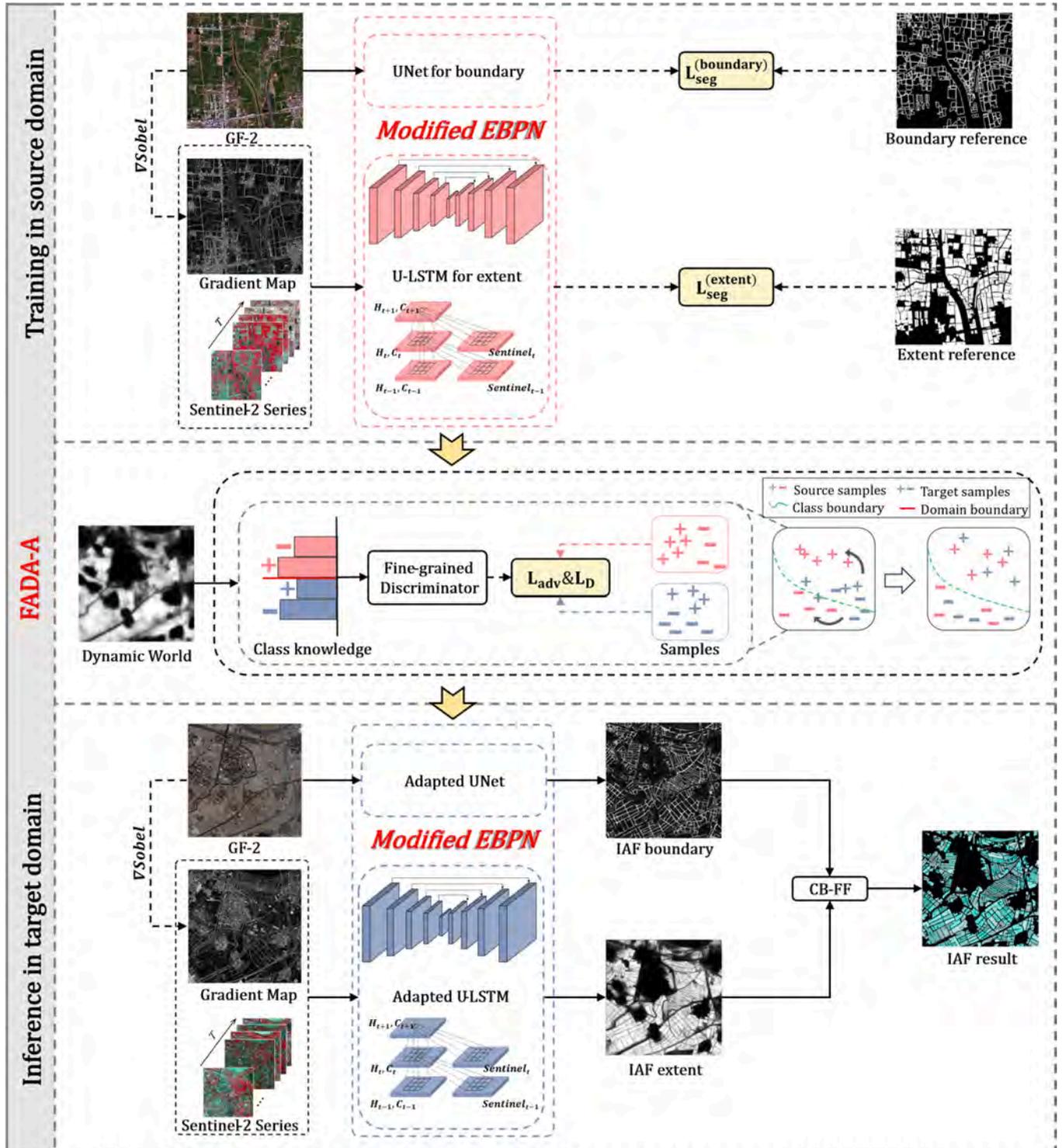


Fig. 1. Workflow of FieldSeg-DA2.0.

Nonetheless, direct outputs generated by deep learning models are often overly fragmented, requiring additional postprocessing to generate closed segments (Persello et al., 2019; H. Zhang et al., 2021). A multitask network is another approach to extracting distinct IAFs that comprehensively considers the boundary and extent information. For instance, Aung et al. (2020) trained two independent UNet branches to extract the IAF boundary and extent simultaneously. Liu et al. (2022) applied UNet and DeeplabV3 + for identifying IAF boundaries and extents, respectively, considering their different preferences for identifying “line” and “face” objects. Waldner et al. (2020) trained ResUNet-a for IAF extraction with a combined loss of four related tasks, i.e., IAF extent identification, IAF boundary identification, estimation of distance to boundary and image reconstruction. As the multi-task strategy showed promising performance in IAF extraction, it has become the mainstream approach in recent studies (Li et al., 2023; Long et al., 2022).

However, training deep learning models requires many labelled samples, which are often difficult to collect in sufficient numbers in practice. Domain Adaptation (DA), i.e., preventing performance degradation when a model trained in the source domain with sufficient samples is applied to the target domain with insufficient samples, is a practical approach to reducing the need for training samples (J. Peng et al., 2022). The transfer problem in IAF extraction mainly arises from differences between the spatial locations and data acquisition time of the source and target domains. For example, a model trained in a plain area may not accurately recognize terrace IAFs in a hilly area due to significant differences in the IAF morphology. A model trained with satellite imagery acquired in the growing season may perform poorly on imagery acquired in the non-growing season due to the distinct spectral differences in different seasons.

Several transfer learning techniques have been introduced in IAF extraction models to address spatial transfer. Fine-tuning is a straightforward and effective transfer technique that has demonstrated excellent performance in transferring IAF extraction models across different study areas (Jong et al., 2022; Wang et al., 2022). It slightly adjusts the model parameters pretrained from the source domain and thus needs only a small number of samples from the target domain. However, fine-tuning is not suitable for study areas without labelled samples. Unsupervised Domain Adaptation (UDA) is another type of transfer learning technology which does not require additional annotations in the target domain (Mirza et al., 2022; Wang et al., 2020; Xiao and Zhang, 2021; P. Zhang et al., 2021). UDA focus on reducing the distribution mismatch by optimizing feature divergence (Ge et al., 2023; P. Zhang et al., 2021) or adopting adversarial training to generate domain-invariant features (Cicek and Soatto, 2019; Long et al., n.d.; Wang et al., 2020) at different feature levels. For example, Zhang et al. (2021) proposed a novel method leveraging representative prototypes and feature distances to achieve remarkable performance improvements in UDA for segmentation. Mirza et al. (2022) proposed Dynamic Unsupervised Adaptation (DUA), using batch normalization adjustments for strong performance gains with minimal unlabelled data, addressing continuous adaptation challenges in segmentation tasks. Wang et al. (2020) proposed a fine-grain adversarial domain adaption (FADA) approach, which aligns features in the source and target domains at the class level by introducing a fine-grained domain discriminator and shows good performance in the cross-domain semantic segmentation task. Recently, Liu et al. (2022) adopted the FADA approach in their IAF extraction model (FieldSeg-DA), achieving excellent transferability across multiple target domains. However, FADA requires prior class information in the target domain to start the training process. This information is commonly accessed from the pseudo-labels generated by a classifier trained in the source domain. However, the inaccuracy of the pseudo-labels may result in a negative transfer of FADA if the IAF morphology and spectral characteristics of the source and target regions significantly differ (Wang et al., 2019).

Temporal transfer generally includes cross-season and cross-year transfers. Although both can raise critical issues for vegetation or crop classification (Capliez et al., 2023; Wang et al., 2023), the cross-year

transfer issue is less serious for IAF extraction. This is because the phenological differences between IAF and other land cover types are much more obvious than those between different vegetation or crop types, resulting in discriminative features that remain relatively stable across years. For the problem of cross-seasonal transfer, the introduction of time-series data is an effective solution. Time series data contain remote sensing observations in all seasons, so there are no differences in the acquisition seasons between the source and target domains. Aung et al. (2020) utilized spatiotemporal convolutional networks to encode time-series data from Sentinel-2 for IAF extraction. Garnot et al. (2021) developed an IAF extraction method using a self-attention-based image sequence encoding network for time-series analysis, and they emphasized that leveraging the phenological information from Sentinel-2 time series can better address the complex spatiotemporal patterns of IAFs. Song et al. (2023) demonstrated that using time series Sentinel-2 images with a 10-metre spatial resolution resulted in better performance in extracting fragmented IAFs than using a single-temporal image. However, time-series data often exhibit low spatial resolutions, which does not satisfy the need for clear IAF boundaries, especially on the fragmented landscape in China (Persello et al., 2019).

To mitigate these spatiotemporal transfer issues, an improved version of the FieldSeg-DA network (Liu et al., 2022), named FieldSeg-DA2.0, is proposed in this study. FieldSeg-DA2.0 includes a spatiotemporal fusion module and an improved FADA module with ancillary data (FADA-A). The spatiotemporal fusion module incorporates the textural information from the high spatial resolution image and the phenological information from the coarse time-series data, enhancing temporal transferability while retaining spatial details. FADA-A introduces the existing coarse-resolution (10 m) land cover data DW (Brown et al., 2022) as ancillary data to guide domain-adversarial training, which enhances the robustness of the transfer learning across distinct spatial regions. The proposed FieldSeg-DA2.0 is then tested in diverse areas with different IAF landscapes by using Gaofen-2 (GF2) data and the Sentinel-2 time series.

## 2. Method

FieldSeg-DA2.0 is developed based on the framework of FieldSeg-DA proposed by Liu et al. (2022). FieldSeg-DA involves three main stages: an extent-boundary parallel network (EBPN) combined with DeeplabV3 + and UNet for identifying the IAF extent and boundary independently; an FADA domain adaptation module for enhancing the transferability of the network; and a connecting boundaries and filling field (CB-FF) module for consistently integrating the identified IAF extent and boundary. In this study, we further enhance the transferability of FieldSeg-DA in two ways (Fig. 1). First, we replace DeeplabV3 + with a novel U-LSTM network in the EBPN to better extract the IAF extent. U-LSTM fuses spatial textural information from high-resolution images and phenological information from coarse time-series data to enhance the model's temporal transferability. Second, during the domain adaptation stage, we introduce additional ancillary data, the DW (DW) land cover data product (Brown et al., 2022), to serve as prior class information to guide the FADA training process. Despite the coarse spatial resolution (10 m), the DW provides a more stable land cover class prior across different areas than the pseudo-labels obtained from the source-domain classifier. Thus, FADA with ancillary data (FADA-A) is expected to achieve better spatial transferability than standard FADA without ancillary data.

### 2.1. EBPN for IAF extent and boundary extraction

In the proposed FieldSeg-DA2.0, the EBPN consists of two independent branch networks, U-LSTM and UNet, which are employed the IAF extent and IAF boundary, respectively.

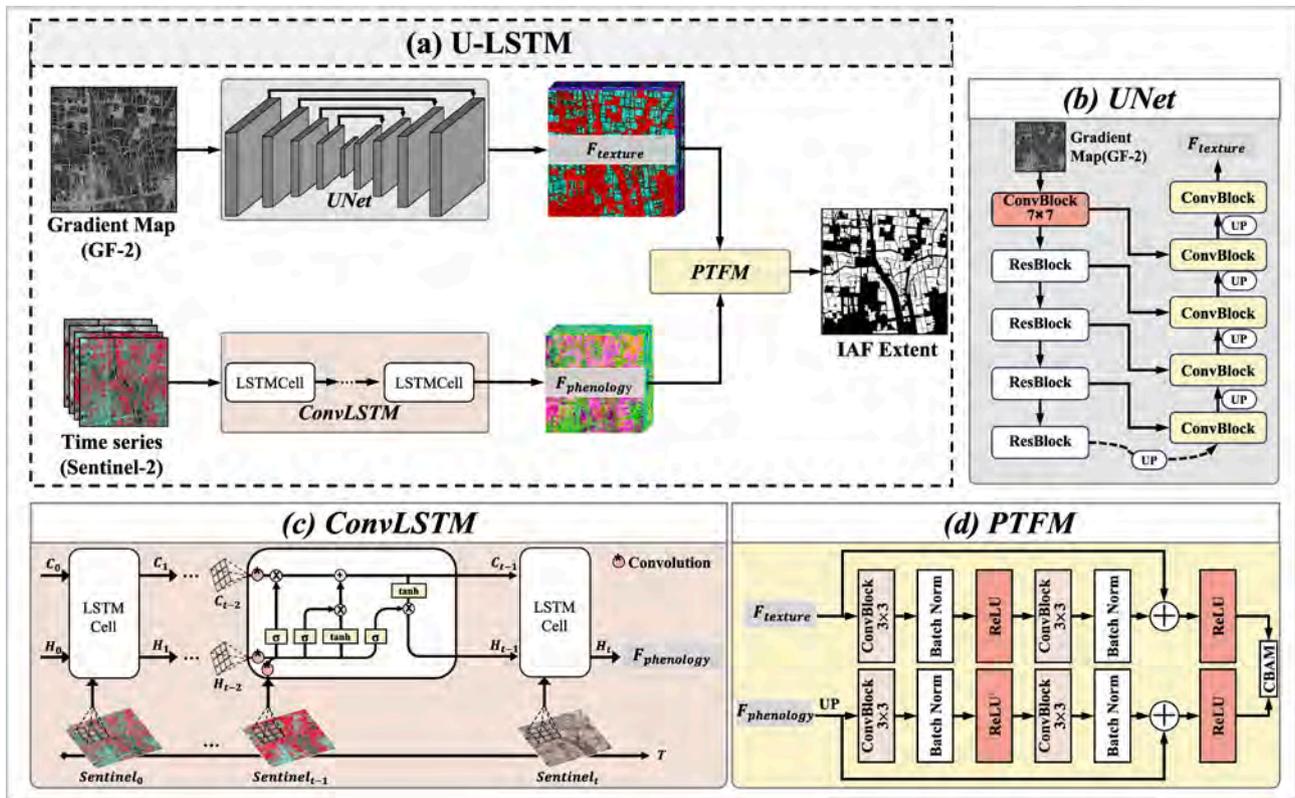


Fig. 2. (a) The main architecture of U-LSTM; (b) UNet used for extracting the textural information from the GF-2 gradient map. (c) ConvLSTM used for extracting the phenological information from the Sentinel-2 time series. (d) PTFM used for fusing the textural and phenological information.

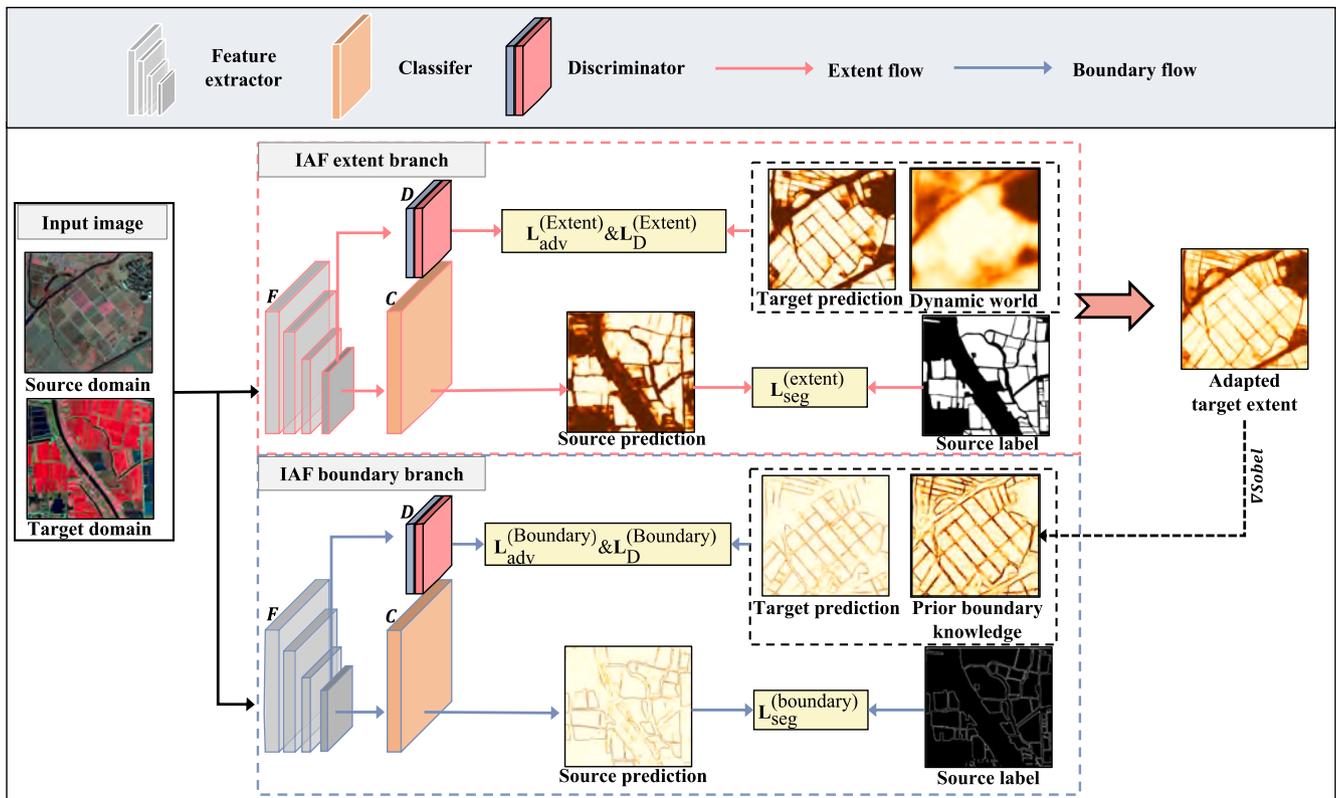


Fig. 3. Workflow of FADA-A.

### 2.1.1. U-LSTM for IAF extent extraction

We designed a U-LSTM to replace the DeeplabV3 + in the original EBPN in FieldSeg-DA for extracting IAF extent in this study. U-LSTM aims to combine the textural information from high spatial resolution images and the phenological information from coarse time series by incorporating UNet (Ronneberger et al., 2015) and ConvLSTM (Shi et al., 2015) networks (Fig. 2), which helps to mitigate the overfitting risk of the network on training data acquired in specific seasons.

First, UNet (Fig. 2b) is used to extract textural features from the gradient map of high-resolution images, considering that the skip-connection structure in UNet helps preserve spatial details in high-resolution images (L. Peng et al., 2022). The gradient map obtained by the Sobel operator from the multispectral image, rather than the multispectral image itself, is used as input. This is because the gradient map generally corresponds to high-frequency textural information (Yang et al., 2020; Zhou et al., 2022), which is insensitive to the colour variation induced by phenological changes across different seasons. Second, ConvLSTM (Fig. 2c) is employed to explore the phenological features from the Sentinel-2 time series due to its proven effectiveness in crop classification (Luo et al., 2023) and IAF segmentation (Cai et al., 2023a). Finally, a phenological-textural fusion module (PTFM) (Fig. 2d) is designed to fuse both the textural features generated from UNet and the phenological features generated from ConvLSTM. First, the phenological feature image is resampled to match the size of the textural feature image, and residual convolution is then applied to both types of features. Additionally, a convolutional block attention module (CBAM) (Woo et al., 2018) is employed to extract pertinent features for identifying IAF extents.

Through fusing textural features from high-resolution images and phenological features from coarse time series, U-LSTM demonstrates strong transferability when applied to high-resolution images acquired in different seasons and retains detailed spatial information when identifying the IAF extents.

### 2.1.2. UNet for IAF boundary extraction

We employ the UNet branch to extract IAF boundaries from high-resolution images, as in the original FieldSeg-DA model (Liu et al., 2022). Coarse time series data are not used as input in this model, as IAF boundaries tend to remain relatively stable across different seasons. Furthermore, we refrain from utilizing gradient maps alone, as relying solely on gradient map information for IAF boundary extraction proves inadequate.

## 2.2. Fine-grained domain adaptation with ancillary data (FADA-A)

Due to the difference between the source and target domains, domain adaptation technologies were proposed to further adapt the model pretrained in the source domain, which is a commonly used strategy in previous studies (Mirza et al., 2022; Wang et al., 2020; P. Zhang et al., 2021). In FieldSeg-DA 2.0, we propose FADA-A, an improved version of the standard FADA module that incorporates prior class information derived from ancillary data to further enhance the robustness of domain adaptation training (Fig. 3). When FieldSeg-DA2.0 is used in a target area distinct from the training areas, both branch networks (U-LSTM and UNet) are retrained with FADA-A to better adapt the model to the target area.

### 2.2.1. FADA

FADA is a domain adaptation technique based on class-level feature alignment. It achieves a better transferring performance in the image semantic segmentation task than the traditional adversarial domain adaptation technique (Wang et al., 2020). Thus, FADA was incorporated into the original FieldSeg-DA to enhance model transferability (Liu et al., 2022).

FADA consists of a segmentation network (G) and a discriminator (D). G is used to predict the segmentation result, i.e., the IAF extent or

boundary in this study, which can be divided into a feature extractor (F) and a binary classifier (C). D is used to distinguish the F derived features from the source and target domains. Then, F-derived features can be aligned at the class level by alternatively optimizing G and D in two steps.

In step one, D is trained to distinguish features from the source and target domains at the class level by minimizing the following loss function (Eq. (1)).

$$\min_{\mathbf{D}} L_{\mathbf{D}} = -\sum_{i=1}^{n_s} p_i \log P(d=0, c=k|f_i) - \sum_{j=1}^{n_t} p_j \log P(d=1, c=k|f_j) \quad (1)$$

where  $f_i$  and  $f_j$  are the features extracted by F on samples from the source and target domains;  $d$  denotes the domain code, where 0 refers to the source domain and 1 refers to the target domain;  $k$  denotes the class code, where 1 refers to the IAF extent or boundary class and 0 refers to the background class; and  $P(d, c|f)$  is the probability output from D.  $n_s$  and  $n_t$  are the numbers of samples from the source and target domains, respectively.  $p_i$  and  $p_j$  are the class soft labels (i.e., probability of class) for source sample  $i$  and target sample  $j$ . In this training step, only the parameters in D are updated, and the parameters in G are fixed.

In step two, G is trained to learn the domain invariant features for the segmentation task by minimizing a loss function combined with the segmentation loss ( $L_{\text{seg}}$ ) and the adversarial loss ( $L_{\text{adv}}$ ) (Eq. (2))

$$\min_{\mathbf{F}, \mathbf{G}} (L_{\text{seg}} + \lambda L_{\text{adv}}) \quad (2)$$

where  $\lambda$  is the empirical weight balancing segmentation and adversarial loss. The segmentation loss represents the difference between the prediction and the ground truth, which guides G to learn the knowledge of the segmentation task. The Tanimoto format loss is used here because of its superiority in addressing the sample imbalance issue (Diakogiannis et al., 2020):

$$L_{\text{seg}} = \frac{\sum_{i=1}^{n_s} p_i y_i}{\sum_{i=1}^{n_s} (p_i^2 + y_i^2) - \sum_{i=1}^{n_s} (p_i y_i)} \quad (3)$$

where  $p_i$  represents the prediction probability that the source domain sample  $x_i^{(s)}$  belongs to the positive class (the IAF extent or boundary in this study), and  $y_i$  refers to the corresponding ground truth label. The adversarial loss  $L_{\text{adv}}$  is a cross-entropy function denoting the difference between the features from the source and target domains; this function forces F to generate domain-aligned features.

$$L_{\text{adv}} = -\sum_{j=1}^{n_t} p_j \log P(d=0, c=k|f_j) \quad (4)$$

In this training step, only the parameters in G are updated, and the parameters in D are fixed.

Through the iterative optimization of both Equations (1) and (2), FADA achieves class-level feature alignment and enhances the transferability of the segmentation model.

### 2.2.2. Class prior knowledge from ancillary data

FADA achieves feature alignment at the class level by incorporating the class information in the source and target domains, i.e.,  $p_i$  and  $p_j$  in Eq. (1) and (4). However,  $p_j$  is unknown due to the absence of ground-truth labels in the target domain. As recommended by Wang et al. (2020), in standard FADA,  $p_j$  is estimated by classifier C trained from the source domain data (i.e., pseudo label). This strategy is also used in the original FieldSeg-DA. However, there is a risk of negative transfer when classifier C trained in the source domain produces unreliable pseudo-labels.

Unlike in general image segmentation tasks, prior information is available for IAF extraction through various global land cover datasets

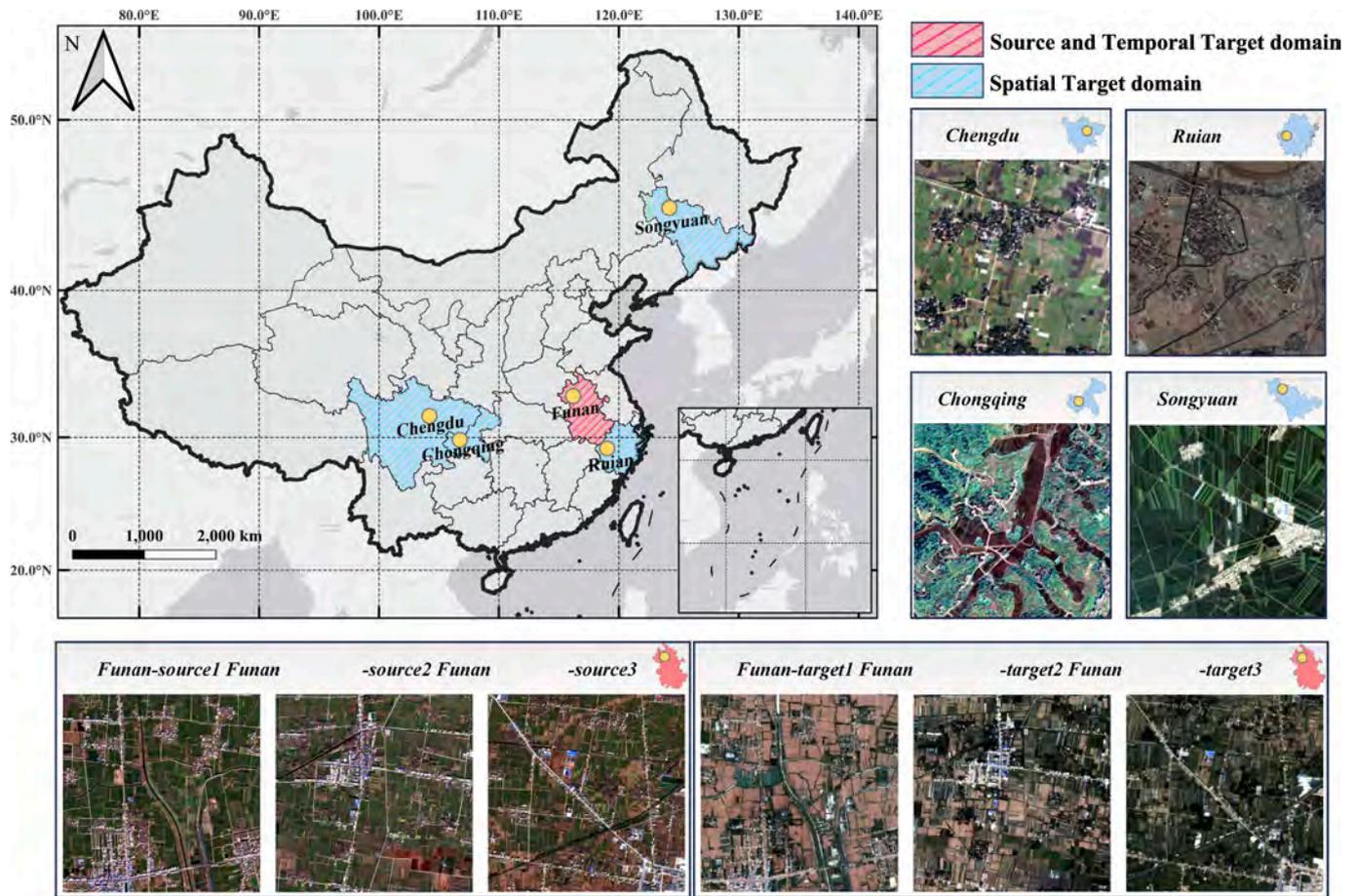


Fig. 4. Geographic locations of the experimental areas and the corresponding GF-2 images. The red region represents the location of the source and temporal target domain, and the blue region represents the locations of the spatial target domains.

produced by the remote sensing community. Most global land cover datasets, e.g., DW (Brown et al., 2022), ESA (Zanaga et al., 2022), GLC-FCS (X. Zhang et al., 2021), and GlobeLand30 (Chen et al., 2021), contain the cropland layer, which has semantic information closely related with IAF. Despite the coarse spatial resolution (10–30 m), these ancillary datasets could contain valuable class information for guiding FADA training in the IAF extraction network. DW data are chosen as the ancillary data in this study for their unique ability to provide globally consistent, near real-time land use and land cover classification information (Brown et al., 2022).

Regarding the domain adaptation of U-LSTM for IAF extent extraction, the pseudo-label information (i.e.,  $p_j$  in Eq. (1) and (4)) is directly

replaced with the cropland probability derived from the DW because of the similar semantic information between IAF extent and cropland land cover. Here, the cropland probability is computed using the mean of the top 10th percentile values of the DW cropland probability throughout one year, considering that cropland is often misclassified as other land cover types during the nongrowing season in the DW. For the domain adaptation of the branch network for IAF boundary extraction, the information provided by DW cannot be used because the 10-metre resolution image is too coarse to provide accurate IAF boundary information. Thus, the prior IAF boundary information is accessed from the IAF extent probability image predicted by the domain-adapted U-LSTM. The Sobel operator is applied to the predicted IAF extent

Table 1  
Summary of climate and crop conditions in five study areas and description of corresponding single-date GF-2 data.

Domain	Site	Location	Climate type	Main characteristics	Main crops	Acquisition date
Source	Funan-source1	32°36'N,115°24'E	Arid to humid continental climate	Mostly flat	Soybean, maize, and wheat	2020/7/29 (growing Season)
	Funan-source2					
	Funan-source3					
Target (Temporal)	Funan-target1	32°36'N,115°24'E	Arid to humid continental climate	Mostly flat	Soybean, maize, and wheat	2019/6/23 (non-growing season)
	Funan-target2					
	Funan-target3					
Target (Spatial)	Chengdu	31°2'N,104°0'E	Monsoon climate	Flat plains to mountainous	Wheat, barley, canola, and legume crops	2018/4/24
	Ruian	27°46'N,120°34'E	Subtropical monsoon climate	Gently undulating landscape	Rice, maize, barley, potatoes, and soybean	2021/1/17
	Chongqing	30°11'N,107°19'E	Warm temperate climate	Mountainous	Rice, maize, Oilseeds and peanuts	2021/5/9
	Songyuan	45°2'N,125°52'E	Humid continental climate	Vast plains in most areas	Soybean, maize, wheat and sorghum	2022/9/11

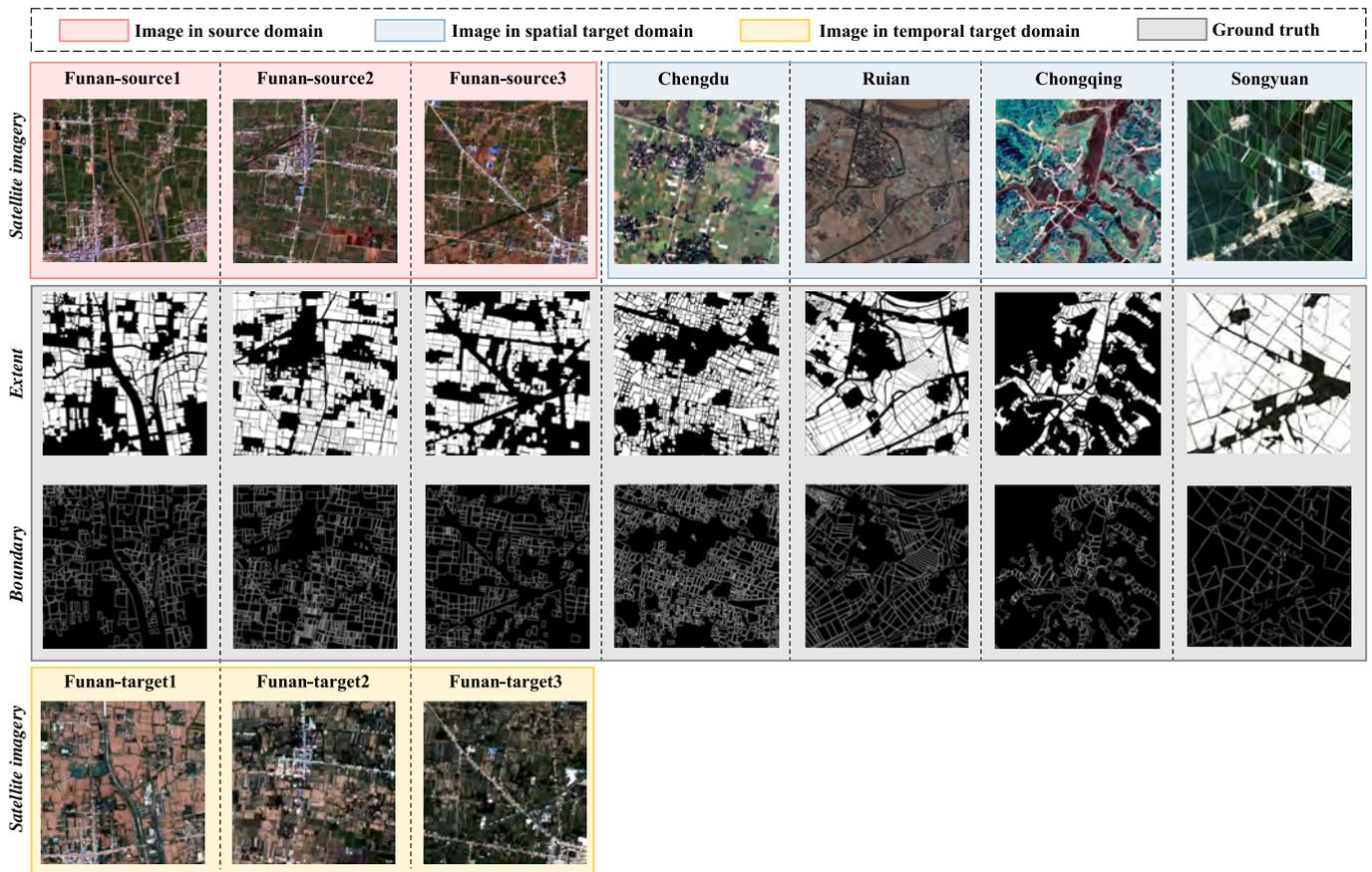


Fig. 5. Ground truth data from the extents and boundaries of training and test images sampled at different locations (RGB composites).

probability image to enhance the boundary information. Then, the output Sobel image is stretched into 0–1 (linear 1 %) and used to replace the pseudo-label information of the IAF boundary in standard FADA. Through the guidance of the DW, the proposed FADA-A is expected to further enhance the spatial transferability of the IAF extraction network.

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

Similar to the postprocess module in the original FieldSeg-DA (Liu et al., 2022), CB-FF is used to integrate the predicted IAF extent and boundary results in our proposed method. First, a series of morphological operations, including skeletonization, dilation, and re-skeletonization, is used to connect boundaries with breakpoints in the predicted boundary result. Second, a filling-field (FF) operation is applied to fuse the predicted boundary and extent images. In this operation, the predicted IAF extent image is clipped with the predicted boundary and then filled by a morphological “closing” operation. CB-FF effectively eliminates over-fragmented boundaries and resolves overly aggregated extents in the IAF boundary and extent images predicted by UNet and U-LSTM.

## 3. Experiments

### 3.1. Study areas

To assess the performance of our proposed FieldSeg-DA2.0 in IAF extraction, we conducted experiments in five distinct municipal districts in China, i.e., Funan, Chengdu, Ruian, Chongqing and Songyuan (Fig. 4). These regions showcase a diverse range of climates, crop types and IAF morphologies (Table 1). For example, IAFs in Ruian featured a higher degree of complexity than the other regions due to intricate river

networks and mountainous terrain. Conversely, IAFs in Songyuan show a larger size and more homogeneous pattern than the other four regions. Due to the adequate number of labelled samples, images acquired in Funan during the growing season (three subimages denoted as Funan-source1, Funan-source2, and Funan-source3) were selected as the training samples. Then, the temporal transferability was evaluated using test GF-2 images acquired in Funan during the non-growing season (three subimages denoted as Funan-target1, Funan-target2, and Funan-target3) when the wheat had been harvested. The spatial transferability was evaluated by using test images acquired in different areas, including Chengdu, Ruian, Chongqing and Songyuan.

### 3.2. Data and preprocessing

For the five experimental regions, we selected ten subimages of Gaofen-2 (GF-2) as high-resolution image input for IAF extraction networks. Their information is presented in Table 1. To obtain surface reflectance images, we performed orthorectification, radiometric calibration, and atmospheric correction. These steps reduce radiometric inconsistencies caused by variations in terrain and atmospheric conditions. To best utilize the spectral and spatial information from the GF-2 imagery, we employed the Gram-Schmidt Adaptive (GSA) method to fuse the coarse spatial resolution (4 m) multispectral bands with the corresponding high spatial resolution (1 m) panchromatic band image. The resulting surface reflectance images consist of four multispectral bands with a spatial resolution of 1 m.

Additionally, each GF-2 image was paired with a time series of Sentinel-2 images. We selected Sentinel-2 images that were composited every 10 days in the same year as the GF-2 images. The time series consisted of 36 frames with four bands (red, green, blue and near-infrared) at a resolution of 10 m.

Table 2

Assessment of the IAF accuracy in the .

Area		Funan-source1	Funan-source2	Funan-source3	Average
F1-score	DeeplabV3+	88.40 %	88.71 %	81.91 %	86.34 %
	FieldSeg	90.41 %	89.16 %	90.70 %	90.09 %
	FieldSeg2.0	<b>90.62 %</b>	<b>91.06 %</b>	<b>90.30 %</b>	<b>90.66 %</b>
mIoU	DeeplabV3+	80.02 %	77.46 %	70.43 %	75.97 %
	FieldSeg	89.13 %	86.01 %	88.64 %	87.93 %
	FieldSeg2.0	<b>89.93 %</b>	<b>88.84 %</b>	<b>89.91 %</b>	<b>89.56 %</b>
GTC	DeeplabV3+	53.01 %	61.01 %	63.02 %	59.01 %
	FieldSeg	17.77 %	16.10 %	13.12 %	15.66 %
	FieldSeg2.0	<b>12.21 %</b>	<b>11.28 %</b>	<b>9.02 %</b>	<b>10.84 %</b>

Source domain using different methods

The ground truth labels of the IAF boundary and extent were acquired through manual digitization of GF-2 images (Fig. 5).

### 3.3. Experimental setup

To evaluate the spatiotemporal transferability of FieldSeg-DA2.0, we compared it with two benchmark models, FieldSeg-DA (Liu et al., 2022) and DeeplabV3+ (Chen et al., 2018). We chose FieldSeg-DA for comparison because it is the predecessor of our model and achieves state-of-the-art performance in IAF extraction. Additionally, we chose DeeplabV3+ due to its widespread use in various image segmentation tasks. Notably, both of the compared models were designed primarily for single high-resolution image input without requiring Sentinel-2 time series. Additionally, we compared two other practical UDA methods, namely ProDA (P. Zhang et al., 2021) and DUA (Mirza et al., 2022), which replaced FADA-A respectively and combined with FieldSeg2.0 to validate the generality of our framework.

**Comparative experiments in the source domain.** We conducted cross-validation experiments in the source domain, encompassing sub-images of Funan-source1, Funan-source2, and Funan-source3. This enabled us to compare the performance of FieldSeg2.0 (FieldSeg-DA2.0 without FADA-A) with the performances of other IAF extraction methods, including DeeplabV3+ and FieldSeg (FieldSeg-DA without FADA). Domain adaptation modules were not utilized for all models in the source domain.

**Temporal transfer experiments.** To assess the temporal transfer capability of the proposed model, we conducted an experiment with FieldSeg2.0, FieldSeg-DA, FieldSeg, and DeeplabV3+ in the temporal target domain (Funan-target1, Funan-target2, and Funan-target3). FieldSeg2.0 did not utilize FADA-A because the introduced U-LSTM module mitigates the need for temporal domain adaptation.

**Spatial transfer experiments.** To evaluate the spatial transfer capability, we conducted an experiment with FieldSeg-DA2.0, FieldSeg-DA, FieldSeg2.0-ProDA, FieldSeg2.0-DUA and DeeplabV3+ in different spatial target domains, including subimages in Chengdu, Ruian, Chongqing and Songyuan.

**Ablation experiments.** We performed ablation experiments to analyse the contribution of individual components of our proposed method to the overall performance. ULSTM and FADA-A were individually removed from the model, and their contributions to temporal and spatial transferability were validated in temporal transfer domains and spatial transfer domains. Furthermore, we established a baseline by removing both modules, which serves as a reference point for evaluating the impact of removing specific components.

### 3.4. Implementation details

The GF-2 imagery and its corresponding IAF labels were divided into  $256 \times 256$  patches due to the limited GPU memory size. The corresponding Sentinel-2 data were clipped based on the coordinates of the GF-2 patch and resampled to a size of  $32 \times 32$  for convenient processing. To monitor the model's performance, 10 % of the training samples were

set aside as a validation set. The training process ends when the validation accuracy does not improve over 100 consecutive epochs. To augment the training data, we introduced geometric transformations to the patches, including horizontal, vertical, and diagonal flips, enabling the model to acquire rotation-invariant features. We initialized our learning rate at  $1e-4$  and employed the Adam optimizer with betas of 0.9 and 0.999. For the training of FADA-A and FADA, we initialized the discriminator's learning rate at  $1e-3$  and conducted training for 4,000 epochs. All experiments were executed utilizing the PyTorch framework on Ubuntu systems equipped with NVIDIA GeForce 3090 24 GB GPUs.

### 3.5. Accuracy assessment

We used both pixel-based attribute and object-based geometric measurements to assess the accuracy of IAF extraction. The pixel-based attribute measurements include the Precision, Recall, F1-score and the mean intersection over union (mIoU). These are calculated as follows:

$$\text{Precision} = \frac{TP}{TP + FP} \quad (5)$$

$$\text{Recall} = \frac{TP}{TP + FN} \quad (6)$$

$$\text{F1-score} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \quad (7)$$

$$\text{mIoU} = \left( \frac{TP}{TP + FP + FN} + \frac{TN}{TN + FP + FN} \right) / 2 \quad (8)$$

where TP refers to the "true positive", TN refers to "true negative", FP refers to "false positive" and FN refers to "false negative".

The global total classification (GTC) index (Li et al., 2023) developed based on under segmentation and over segmentation error measurements (Persello and Bruzzone, 2010) is applied as an object-based evaluation metric. Let  $D_i (i = 1, \dots, m)$  be a predicted IAF and  $O_j (j = 1, \dots, n)$  be the reference IAF. Let  $area(D_i)$  and  $area(O_j)$  be the areas of  $D_i$  and  $O_j$ , respectively, and  $area(D_i \cap O_j)$  be their overlapping area. The region  $D_x = \underset{v \in D_i}{\operatorname{argmax}} (area(D_i \cap O_j))$ . The local-error over

classification  $OC_j(O_j, D_x)$  error and an under classification  $UC_j(O_j, D_x)$  error can be calculated as:

$$OC_j(O_j, D_x) = 1 - \frac{area(O_j \cap D_x)}{area(O_j)} \quad (9)$$

$$UC_j(O_j, D_x) = 1 - \frac{area(O_j \cap D_x)}{area(D_x)} \quad (10)$$

Based on this, GTC was introduced, with lower GTC values indicating higher accuracy in IAF extraction:

$$TC_j(O_j, D_x) = \sqrt{\frac{OC_j(O_j, D_x)^2 + UC_j(O_j, D_x)^2}{2}} \quad (11)$$

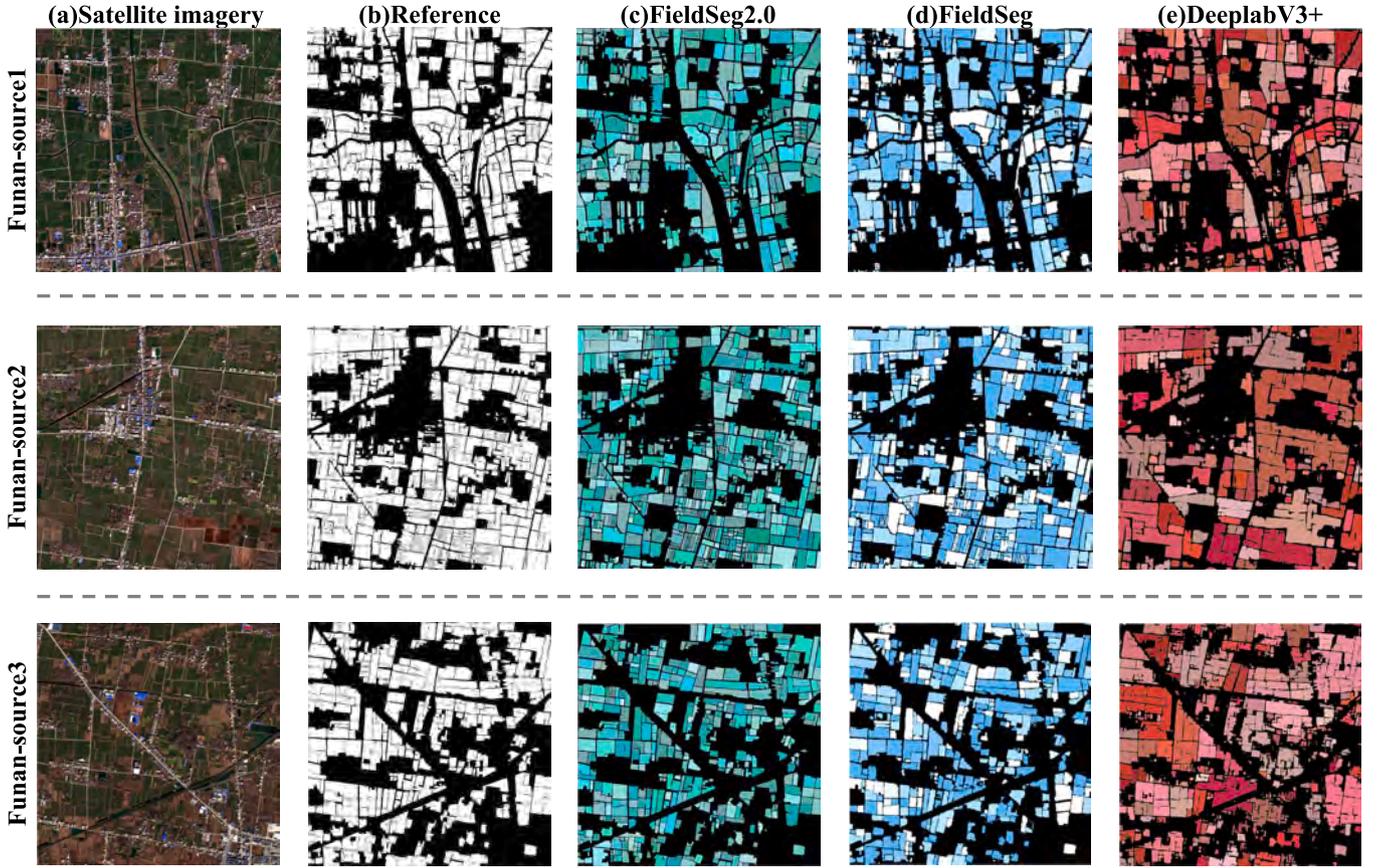


Fig. 6. Extracted IAFs with different methods for GF-2 images acquired in the source domain. (a) refers to the GF-2 image. (b) refers to the reference IAF. (c)– (e) are the IAFs extraction results from different methods.

Table 3

Assessment of the IAF accuracy in the temporal target domain using different methods.

Area		Funan-target1	Funan-target2	Funan-target3	Average
F1-score	DeeplabV3+	79.71 %	81.19 %	69.71 %	76.87 %
	FieldSeg	80.27 %	79.66 %	71.24 %	77.06 %
	FieldSeg-DA	80.94 %	81.75 %	75.27 %	79.32 %
	FieldSeg2.0	84.94 %	86.74 %	80.12 %	83.93 %
mIoU	DeeplabV3+	64.57 %	61.29 %	43.46 %	56.44 %
	FieldSeg	66.04 %	59.03 %	48.68 %	57.92 %
	FieldSeg-DA	65.43 %	65.65 %	57.66 %	62.91 %
	FieldSeg2.0	73.55 %	71.69 %	66.19 %	70.48 %
GTC	DeeplabV3+	75.01 %	71.02 %	73.02 %	73.02 %
	FieldSeg	52.74 %	51.08 %	56.62 %	53.48 %
	FieldSeg-DA	39.44 %	42.08 %	48.72 %	43.41 %
	FieldSeg2.0	34.12 %	33.3 %	31.21 %	32.88 %

$$GTC = \sum_{j=1}^n \left( TC(O_j, D_x) \times \frac{\text{area}(O_j)}{\sum_{j=1}^n \text{area}(O_j)} \right) \quad (12)$$

## 4. Results

### 4.1. Results of the source domain

Table 2 showcases the accuracy metrics of IAF extraction performances for different IAF extraction methods in the source domain experiments. Among all the methods, FieldSeg2.0 performs best in all three metrics. And the utilization of gradient maps also results in superior performance on the GTC metric compared to other methods. FieldSeg also achieves high accuracy values, with only slightly lower values than FieldSeg2.0. In contrast, DeeplabV3 + significantly

underperforms. In particular, the GTC of DeeplabV3 + is nearly five times that of FieldSeg2.0, indicating a large boundary recognition error for DeeplabV3 +. As shown in Fig. 6, IAFs extracted by DeeplabV3 + are excessively aggregated due to a severe omission of the IAF boundaries, while the IAFs extracted by FieldSeg2.0 and FieldSeg both show align well with the ground truth label.

### 4.2. Results of the temporal transfer experiment

Table 3 presents the accuracies of different IAF extraction algorithms in the temporal transfer experiment. Overall, the accuracies in the target domain (Table 3) are lower than the accuracies in the source domain (Table 2) for all compared methods compared. Nonetheless, the accuracy of FieldSeg2.0 decreases the least from the source domain to the target domain, thereby clearly outperforming other methods. FieldSeg-

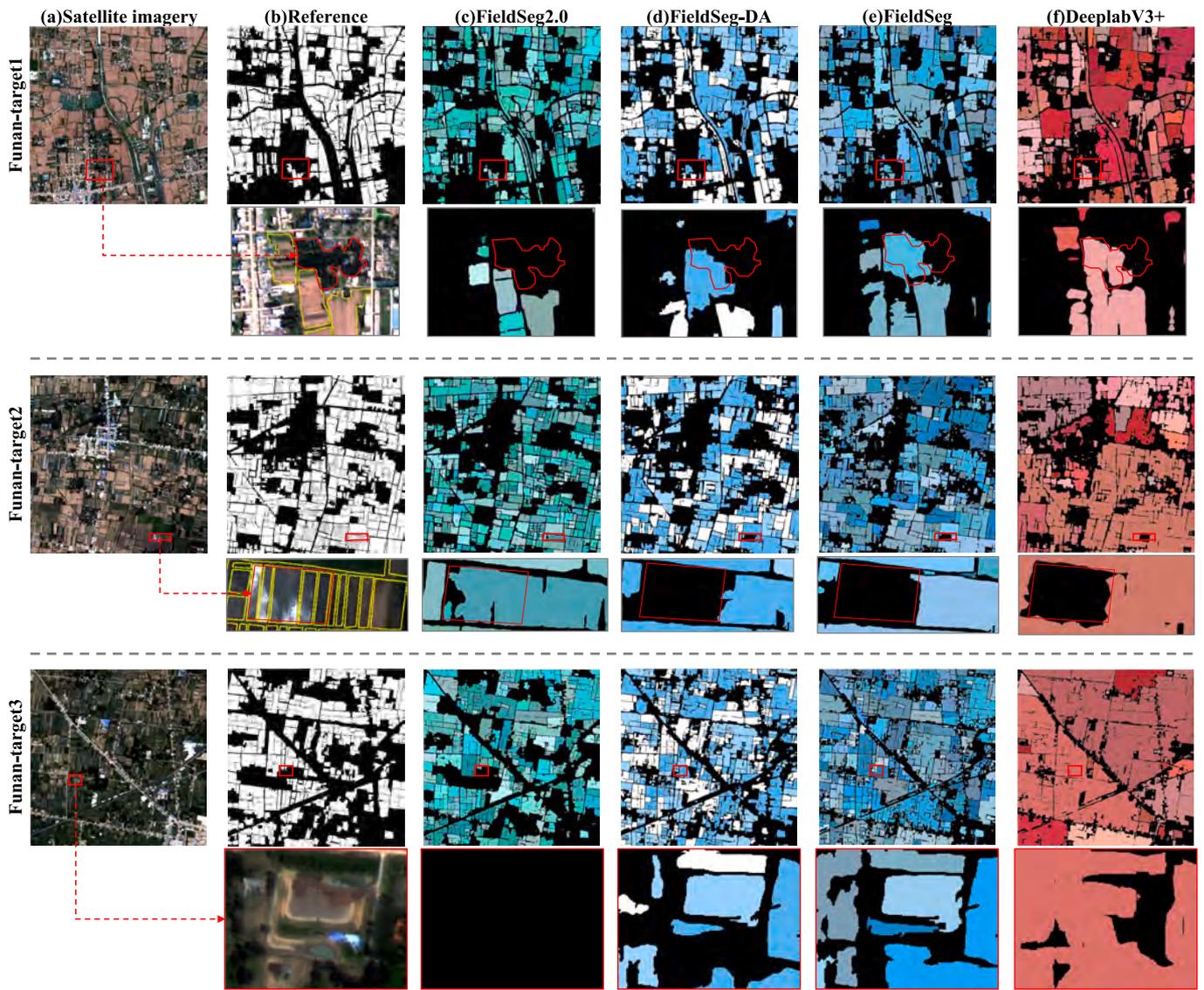
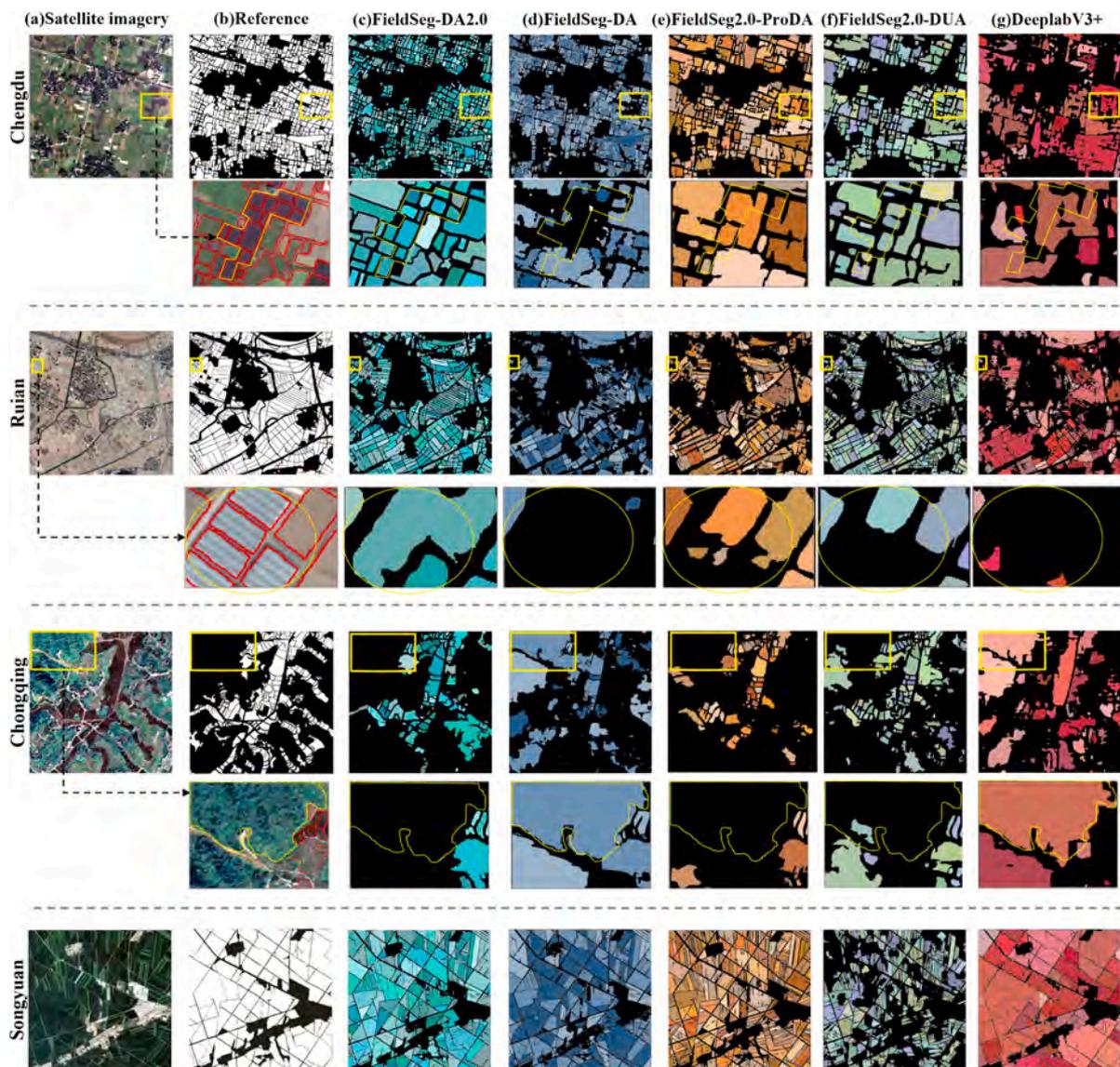


Fig. 7. Extracted IAFs using different methods for GF-2 images acquired in the temporal target domain. The red highlighted areas are confusing feature areas that require attention, while the yellow lines denote the reference IAF. (a) contains the GF-2 images. (b) contains the reference IAF. (c)- (f) are the IAFs extraction results from different methods.

Table 4  
Assessment of the IAF accuracy in the spatial target domain using different methods.

Area		Chengdu	Ruian	Chongqing	Songyuan	Average
F1-score	DeeplabV3+	65.39 %	68.05 %	27.20 %	95.28 %	63.98 %
	FieldSeg-DA	80.12 %	68.81 %	25.64 %	95.94 %	67.62 %
	FieldSeg2.0-ProDA	80.88 %	86.25 %	64.09 %	96.08 %	81.83 %
	FieldSeg2.0-DUA	81.01 %	84.21 %	63.87 %	74.11 %	75.80 %
	FieldSeg-DA2.0	<b>81.77 %</b>	<b>88.27 %</b>	<b>66.52 %</b>	<b>96.28 %</b>	<b>83.21 %</b>
mIoU	DeeplabV3+	53.38 %	55.59 %	33.42 %	79.53 %	55.48 %
	FieldSeg-DA	77.22 %	63.32 %	30.35 %	81.67 %	63.14 %
	FieldSeg2.0-ProDA	77.69 %	72.24 %	62.97 %	84.21 %	74.28 %
	FieldSeg2.0-DUA	77.71 %	70.94 %	62.19 %	69.72 %	70.14 %
	FieldSeg-DA2.0	<b>78.95 %</b>	<b>74.59 %</b>	<b>65.21 %</b>	<b>84.58 %</b>	<b>75.83 %</b>
GTC	DeeplabV3+	98.89 %	73.01 %	93.69 %	71.77 %	85.18 %
	FieldSeg-DA	34.62 %	68.81 %	90.01 %	61.07 %	63.80 %
	FieldSeg2.0-ProDA	33.32 %	60.37 %	84.11 %	44.07 %	55.47 %
	FieldSeg2.0-DUA	32.14 %	65.41 %	75.27 %	61.01 %	58.46 %
	FieldSeg-DA2.0	<b>29.2 %</b>	<b>42.21 %</b>	<b>53.41 %</b>	<b>37.08 %</b>	<b>40.70 %</b>



**Fig. 8.** IAF extracted with different methods in different spatial target domains. The yellow highlighted areas are confusing feature areas that require attention, while the red lines denote the reference IAF. (a) contains the GF-2 images. (b) contains the reference IAF. (c)–(g) are the IAFs extraction results from different methods.

DA ranks second in this experiment, indicating that the FADA module also benefits transfer learning across different temporal domains to some extent. FieldSeg and DeeplabV3 + achieve the lowest accuracy due to their lack of domain adaptation modules.

A detailed comparison of IAF extraction errors from different methods is shown in Fig. 7. Overall, FieldSeg2.0 effectively distinguishes IAF and non-IAF by incorporating phenological features throughout the year. In contrast, other methods often confuse IAF with other land cover types due to the significant spectral change in IAF from the source domain (growing season) to the target domain (non-growing season). For instance, in Funan-target1 and Funan-target3, some forest and bare land patches are misclassified as IAF, while in Funan-target2, the paddy fields are misclassified as non-IAFs. In summary, FieldSeg2.0 maintains robust performance in the temporal transferring experiments. This performance is attributed to its ability to capture dynamic phenological features throughout a year rather than focusing on the instantaneous features in a single image.

#### 4.3. Results of the spatial transfer experiment

Table 4 presents the accuracies of different IAF extraction models in

the spatial transfer experiment. FieldSeg-DA2.0 still achieves the best accuracies among all the methods, obviously outperforming FieldSeg-DA, FieldSeg2.0-ProDA, FieldSeg2.0-DUA and DeeplabV3 + . The pixel-based accuracy metric (**F1-score** and **mIoU**) and the object-based metric (**GTC**) values achieved by FieldSeg-DA2.0 are approximately 10 % and 20 %, respectively, higher than those of FieldSeg-DA. The pixel-based accuracy metrics of FieldSeg2.0-ProDA and FieldSeg2.0-DUA exceed those of FieldSeg-DA, illustrating the robust generality of our proposed framework, which can benefit from any novel UDA methods. However, these two UDA methods exhibit slightly worse domain adaptation performance compared to FADA-A. Additionally, DeeplabV3 + without using UDA method achieves the worst performance with the lowest accuracies.

We further visually examine the detailed difference in IAF extraction for the three compared methods (Fig. 8). The IAFs in Chengdu exhibit more diverse colours and fragmented patterns than the source domain; thus, FieldSeg-DA and DeepLabV3 + miss some purple IAFs and bare IAFs that do not exist in the source domain, and they also generate overly large IAF sizes, which obviously deviate from the ground truth (Fig. 8d). In comparison, FieldSeg-DA2.0, FieldSeg2.0-ProDA, and FieldSeg2.0-DUA yield more reasonable results. FieldSeg-DA2.0, in

**Table 5**  
Ablation experiments in the temporal target domain.

ULSTM	FADA-A	F1-score (average)	mIoU (average)	GTC (average)
–	–	77.05 %	57.91 %	50.93 %
–	✓	80.76 %	67.33 %	38.77 %
✓	–	83.93 %	70.48 %	32.88 %
✓	✓	<b>85.09 %</b>	<b>72.11 %</b>	<b>32.53 %</b>

**Table 6**  
Ablation experiments in the spatial target domain.

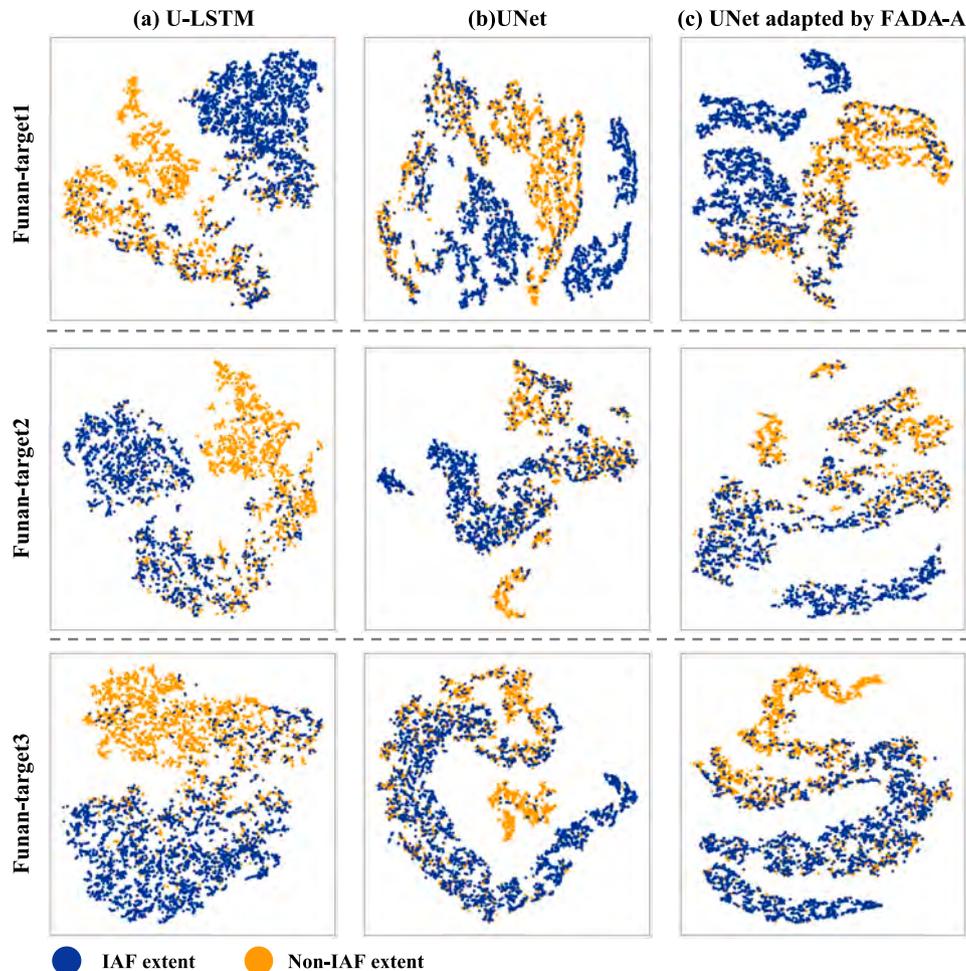
ULSTM	FADA-A	F1-score (average)	mIoU (average)	GTC (average)
–	–	61.53 %	55.72 %	81.84 %
–	✓	78.98 %	69.54 %	67.21 %
✓	–	75.02 %	66.53 %	71.18 %
✓	✓	<b>83.21 %</b>	<b>75.83 %</b>	<b>41.61 %</b>

particular, demonstrates the closest alignment with ground truth. In Ruian, there was a new IAF type, greenhouse, which is not present in the source domain. Both FieldSeg-DA and DeepLabV3 + fail to recognize this IAF type because its textures and colours greatly differ from those of other IAFs, while FieldSeg-DA2.0 correctly identifies them despite some adhesion issues. FieldSeg2.0-ProDA and FieldSeg2.0-DUA miss some greenhouses and also encounter adhesion issues. In Chongqing, the majority of IAFs take the form of terraced, irregularly shaped, and densely distributed features. Both FieldSeg-DA and DeepLabV3 + struggle to achieve satisfactory results in this region, exhibiting serious

confusion with forests. In contrast, FieldSeg-DA2.0, FieldSeg2.0-ProDA and FieldSeg2.0-DUA achieves much more acceptable results. But FieldSeg-DA2.0 has better adaptation performance. In Songyuan, the IAFs exhibit similar characteristics with the source domain, all methods except FieldSeg2.0-DUA generate reasonable results.

#### 4.4. Results of ablation experiments

Ablation experiments were conducted in both temporal and spatial transferring experiments to assess the influence of ULSTM and FADA-A on temporal and spatial transferability. For comparison, we removed the FADA-A module from FieldSeg-DA2.0 and replaced the ULSTM module with UNet to establish a baseline model. Then, ULSTM and FADA-A are respectively introduced into the baseline model to evaluate their contributions on model transferability. The results of each ablation experiment are presented and compared with the baseline model (Table 5 and Table 6). For the temporal transferring experiment, introducing ULSTM module (replacing corresponding UNet in the baseline model with ULSTM) resulted in 6.88 % increase in F1-score, a 12.57 % increase in mIoU, and an 18.05 decrease in GTC over the baseline model. Introducing FADA-A module also improves the accuracy of base-line model, yielding improvements of 3.71 % in F1-score, 9.42 % in mIoU, and 11.76 % in GTC. It indicates that both ULSTM and FADA-A contribute in enhancing the temporal transferability, with ULSTM making a larger contribution. For the spatial transferring experiment, introducing ULSTM module led to improvements of 13.49 % in F1-score, 10.81 % in mIoU, and 10.66 % in GTC over the baseline model. Introducing FADA-A



**Fig. 9.** Feature visualization using t-SNE for different temporal target domains. (a) depicts features from U-LSTM, which takes a GF-2 gradient map and Sentinel-2 time series as input. (b) depicts features from UNet, which takes a single-date GF-2 image as input. (c) depicts the combined FADA-A using (b) as a basis.

Table 7

The IAF extraction accuracy of various U-LSTMs employing either FADA or FADA-A in the spatial target domain.

Area		Chengdu	Ruian	Chongqing	Songyuan	Average
F1-score	FieldSeg2.0	69.80	78.98	54.74	96.57	75.02
	FieldSeg-DA2.0 with FADA	80.90	85.35	64.07	96.07	81.60
	FieldSeg-DA2.0	<b>81.77</b>	<b>88.27</b>	<b>66.52</b>	<b>96.28</b>	<b>83.21</b>
mIoU	FieldSeg2.0	57.15	65.01	59.41	84.55	66.53
	FieldSeg-DA2.0 with FADA	77.69	71.48	62.79	84.55	70.88
	FieldSeg-DA2.0	<b>78.95</b>	<b>74.59</b>	<b>65.21</b>	<b>84.58</b>	<b>75.83</b>
GTC	FieldSeg2.0	58.85	65.01	89.69	47.77	71.18
	FieldSeg-DA2.0 with FADA	32.72	65.85	87.11	45.02	61.89
	FieldSeg-DA2.0	<b>29.20</b>	<b>42.21</b>	<b>53.41</b>	<b>37.08</b>	<b>41.61</b>

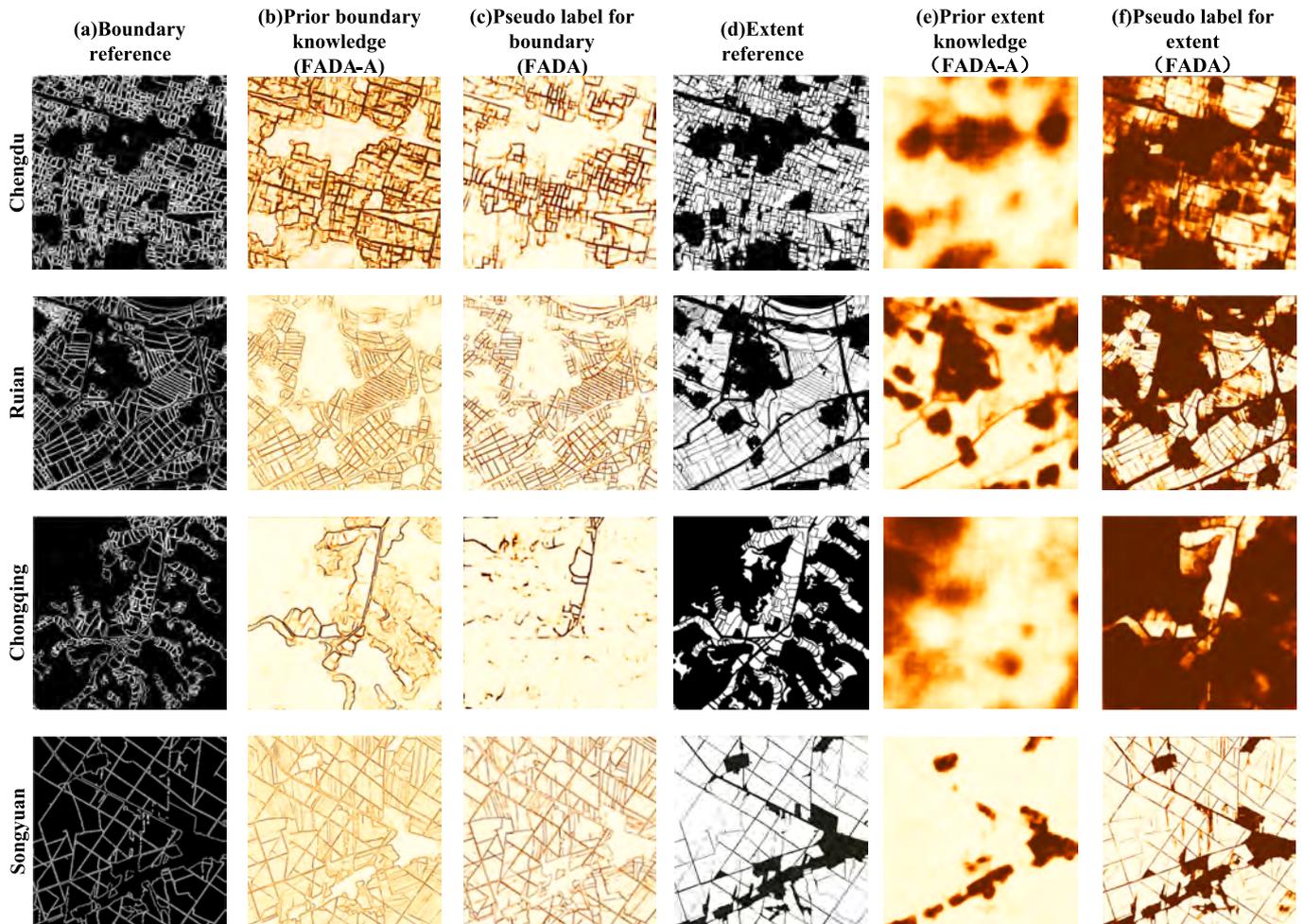


Fig. 10. Comparison of prior information in different domain adaptation models. (a) (d) depict the reference IAF. (b) (e) depict the prior knowledge guiding the FADA-A model, with (e) provided by DW data and (b) obtained through the Sobel operator on the adapted IAF extent probability maps. (c) (f) depict prior knowledge guiding the FADA model, with both directly provided by classifiers.

module into the baseline method resulted in a 17.45 % increase in F1-score, a 13.82 % increase in mIoU, and a 14.63 % decrease in GTC. It indicates that FADA-A contributes more in enhancing the spatial transferability than ULSTM. And the full model with both FADA-A and ULSTM module achieves the best results in both temporal and spatial transferring experiments. These results demonstrate that U-LSTM and FADA-A mainly contribute on enhancing temporal and spatial transferability respectively, which is consistent with our design objective.

## 5. Discussion

### 5.1. Superiority of the U-LSTM module in enhancing temporal transferability

To understand how U-LSTM improves temporal transferability, we employ the t-distributed stochastic neighbour embedding (t-SNE) algorithm to visualize feature representations obtained through U-LSTM, UNet adapted by FADA-A, and UNet trained in the source domain (Fig. 9). 5,000 IAF samples and 5,000 non-IAF samples are randomly selected in the temporal target domain (Funan-target1, Funan-target2, and Funan-target3) for training t-SNE. The results demonstrate that U-LSTM exhibits the highest capability in distinguishing IAF and non-IAF

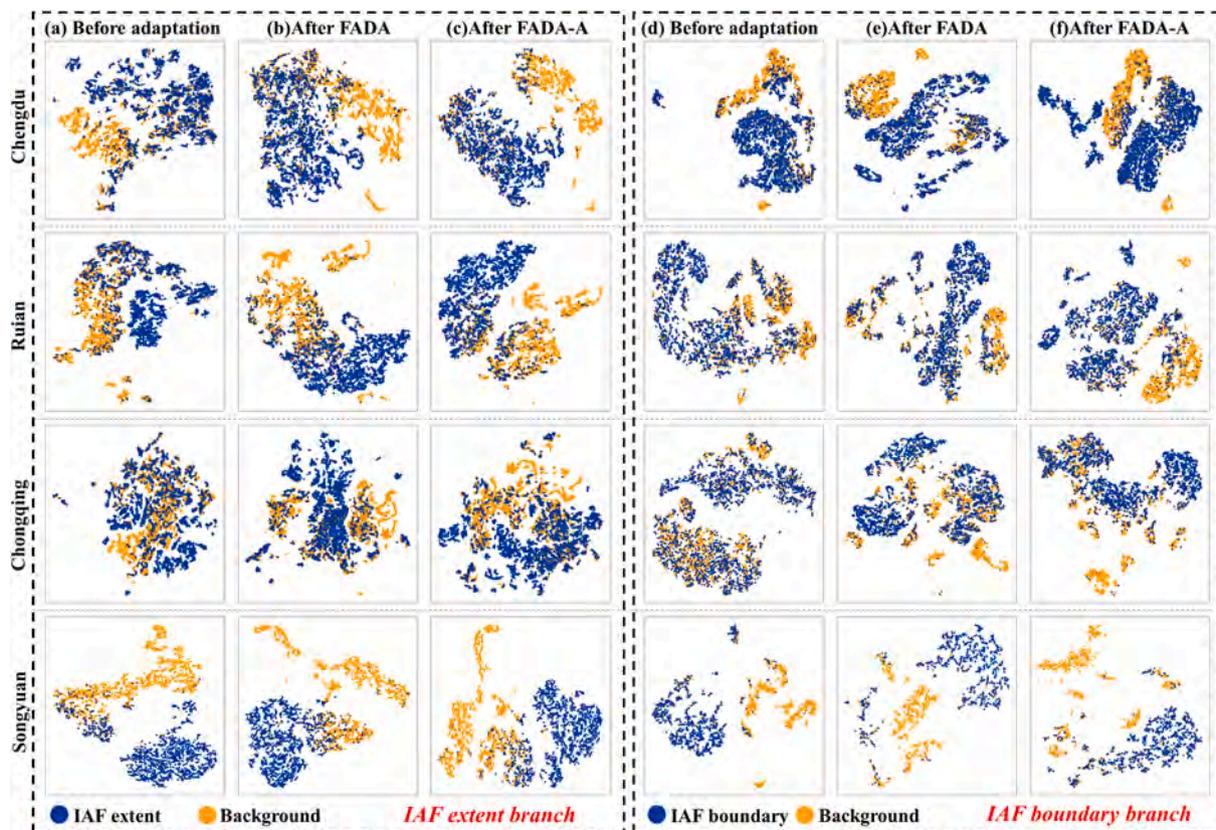


Fig. 11. Visualization of features using t-SNE for various domain adaptation methods in the spatial target domain. (a)- (c) represent feature visualization for different domain adaptation methods in the IAF extent branch, and (d)- (f) depict feature visualization for different domain adaptation methods in the IAF boundary branch. Blue points represent IAF extent or boundary samples, and yellow points represent non-IAF extent or non-IAF boundary samples (background samples).

regions in the feature space. In comparison, the abstract features extracted by UNet adapted by FADA-A and UNet trained from the source domain display weaker and the weakest separability, respectively, between IAF and non-IAF. These results indicate the ability of U-LSTM to generate robust features across temporal target domains. This can be attributed to the usage of the phenological features in the Sentinel-2 time series.

5.2. Superiority of FADA-A in enhancing the spatial transferability

To assess the effectiveness of FADA-A in enhancing transferability for the IAF extraction model, we conducted spatial transfer experiments comparing standard FieldSeg-DA2.0 with two degraded versions, FieldSeg-DA2.0 with FADA (where FADA-A is degraded to FADA) and FieldSeg-2.0 (where FADA-A is excluded) in spatial transferring experiments. FieldSeg-DA2.0 and FieldSeg-DA2.0 with FADA achieved significantly higher accuracies than FieldSeg2.0 without domain adaptation (Table 7), demonstrating the necessity of domain adaptation. FieldSeg-DA2.0 outperforms FieldSeg-DA2.0 with FADA in all four regions, particularly in Ruian and Chongqing. This is a result of the presence of “new” land cover types (e.g., rivers, greenhouses and forests) in these two areas that are absent from the source domain, leading to inaccuracies in the pseudo-labels generated by the source domain model (Fig. 10c and f) and thus negatively impacting the FADA. In contrast, these “new” land cover types are not new for the DW. Moreover, in contrast to the pseudo-labels generated by the source domain model (Fig. 10c and f), the DW does not overlook IAFs that significantly differ from the source domain features; instead, it comprehensively covers IAFs with various features in new areas (Fig. 10e), which is crucial for domain adaptation techniques. Consequently, FADA-A based on the DW can capture the common features of IAFs in diverse regions, rather than

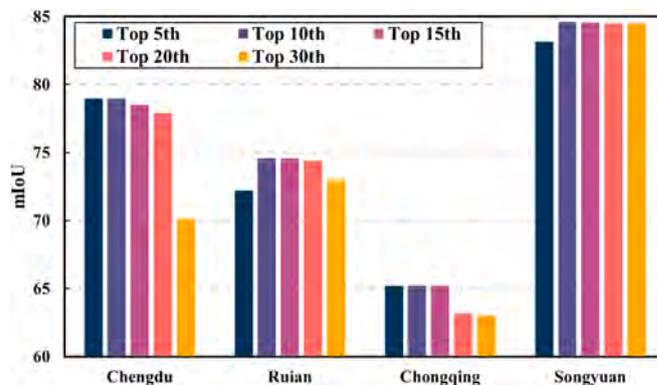


Fig. 12. Comparison of the mIoU of FieldSeg-DA2.0 by setting different synthesis percentile of DW. In the legend, the numbers represent percentiles of the composite. For instance, the number 10 signifies selecting the mean of the top 10th percentile values of the DW cropland probability for each sample over one year.

blindly relying on features learned by the source domain model. Furthermore, the DW also contributes to more complete delineation of IAF boundaries (Fig. 10b), facilitating domain adaptation of IAF boundary extraction models.

To further illustrate the feature alignment performance of FADA and FADA-A, we employed t-SNE to visualize the two-dimensional feature distributions before and after domain adaptation (Fig. 11). We randomly selected 5000 samples for both positive and negative classes. Overall, both FADA-A and FADA helped networks generate more distinct features between the positive and negative classes than FieldSeg2.0 without domain adaptation. However, FADA-A achieved better class

discrimination in the feature space than FADA, especially in Chongqing and Ruian. These results indicate that FADA-A is better at aligning features across different spatial domains than FADA.

### 5.3. Sensitivity analysis of the synthesis percentile of DW

To derive class prior knowledge to guide FADA-A training, we compute a synthetic cropland probability for each sample by averaging the DW cropland probability above the top 10th percentile values throughout one year. As the top 10th of synthesis percentile is an empirical setting, we further explore how different synthesis percentile values affect the accuracy of FieldSeg-DA2.0 (Fig. 12). The results show that unreasonable synthesis selections can impact the final accuracy of IAF extraction. Too large synthesis percentile would more probably include DW data in non-growing season, leading to inaccurate estimation of prior cropland probability. Conversely, a percentile that is too small ensures the exclusion of non-growing season data but is highly sensitive to outliers in the DW time-series. Based on our experiment, the setting of top 10th percentile helps to produce the most stable IAF extraction accuracy across all regions, thus is recommended for most applications.

## 6. Conclusion

In this article, we developed a novel IAF extraction network, FieldSeg-DA2.0, aiming to enhance the spatiotemporal transferability of the original FieldSeg-DA. FieldSeg-DA2.0 employs the main framework of FieldSeg-DA but utilizes two novel modules, U-LSTM for IAF extent extraction and FADA-A for domain adaptation. U-LSTM integrates incorporated features in the GF-2 image and phenological features in the Sentinel-2 time series, which enhances temporal transferability without sacrificing spatial details. The FADA-A module is an improvement of FADA that introduces the prior information derived from the coarse-resolution land cover product to enhance feature alignment across different spatial domains. The experimental results in different target domains confirm the greater spatial and temporal transferability of FieldSeg-DA2.0 compared to the original FieldSeg-DA. In conclusion, FieldSeg-DA2.0 demonstrates robust performance in extracting IAF across diverse scenarios due to its effective integration of multiple remote sensing and land cover data.

### CRedit authorship contribution statement

**Chun Tian:** Writing – review & editing, Writing – original draft, Visualization, Supervision, Software, Methodology, Formal analysis, Data curation, Conceptualization. **Xuehong Chen:** Writing – review & editing, Writing – original draft, Supervision, Methodology, Funding acquisition, Conceptualization. **Jin Chen:** Writing – review & editing, Methodology. **Ruyin Cao:** Writing – review & editing. **Shuaijun Liu:** Writing – review & editing, Validation.

### 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

The code of FieldSeg-DA2.0 is at <https://github.com/ChunTianBNU/FieldSeg-DA2.0>.

### Acknowledgements

This study was supported by the National Key Research and Development Program of China (No. 2022YFD2001101).

## References

- Aung, H.L., Uzcent, B., Burke, M., Lobell, D., Ermon, S., 2020. Farm Parcel Delineation Using Spatio-temporal Convolutional Networks, in: 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW). In: Presented at the 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), IEEE, Seattle, WA, USA, pp. 340–349. <https://doi.org/10.1109/CVPRW50498.2020.00046>.
- Badrinarayanan, V., Kendall, A., Cipolla, R., 2017. SegNet: A Deep Convolutional Encoder-Decoder Architecture for Image Segmentation. *IEEE Trans. Pattern Anal. Mach. Intell.* 39, 2481–2495. <https://doi.org/10.1109/TPAMI.2016.2644615>.
- 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., Schwehr, K., Weisse, M., Stolle, F., Hanson, C., Guinan, O., Moore, R., Tait, A.M., 2022. Dynamic World, Near real-time global 10 m land use land cover mapping. *Sci. Data* 9, 251. <https://doi.org/10.1038/s41597-022-01307-4>.
- Cai, Z., Hu, Q., Zhang, X., Yang, J., Wei, H., Wang, J., Zeng, Y., Yin, G., Li, W., You, L., Xu, B., Shi, Z., 2023a. Improving agricultural field parcel delineation with a dual branch spatiotemporal fusion network by integrating multimodal satellite data. *ISPRS J. Photogrammetry and Remote Sens.* 205, 34–49. <https://doi.org/10.1016/j.isprsjprs.2023.09.021>.
- Cai, Z., Wei, H., Hu, Q., Zhou, W., Zhang, X., Jin, W., Wang, L., Yu, S., Wang, Z., Xu, B., Shi, Z., 2023b. Learning spectral-spatial representations from VHR images for fine-scale crop type mapping: A case study of rice-crayfish field extraction in South China. *ISPRS J. Photogrammetry and Remote Sens.* 199, 28–39. <https://doi.org/10.1016/j.isprsjprs.2023.03.019>.
- Capliez, E., Ienco, D., Gaetano, R., Baghdadi, N., Salah, A.H., 2023. Temporal-Domain Adaptation for Satellite Image Time-Series Land-Cover Mapping With Adversarial Learning and Spatially Aware Self-Training. *IEEE J. Sel. Top. Appl. Earth Observations Remote Sens.* 16, 3645–3675. <https://doi.org/10.1109/JSTARS.2023.3263755>.
- Chen, J., Chen, L., Chen, F., Ban, Y., Li, S., Han, G., Tong, X., Liu, C., Stamenova, V., Stamenov, S., 2021. Collaborative validation of GlobeLand30: Methodology and practices. *Geo-Spatial Information Science* 24, 134–144. <https://doi.org/10.1080/10095020.2021.1894906>.
- Chen, L.-C., Zhu, Y., Papandreou, G., Schroff, F., Adam, H., 2018. Encoder-Decoder with Atrous Separable Convolution for Semantic Image Segmentation. In: Ferrari, V., Hebert, M., Sminchisescu, C., Weiss, Y. (Eds.), *Computer Vision – ECCV 2018*, Lecture Notes in Computer Science. Springer International Publishing, Cham, pp. 833–851. [https://doi.org/10.1007/978-3-030-01234-2\\_49](https://doi.org/10.1007/978-3-030-01234-2_49).
- Cicek, S., Soatto, S., 2019. Unsupervised Domain Adaptation via Regularized Conditional Alignment, in: 2019 IEEE/CVF International Conference on Computer Vision (ICCV). Presented at the 2019 IEEE/CVF International Conference on Computer Vision (ICCV), IEEE, Seoul, Korea (South), pp. 1416–1425. Doi: 10.1109/ICCV.2019.00150.
- Da Costa, J.P., Michelet, F., Germain, C., Laviolle, O., Grenier, G., 2007. Delineation of vine parcels by segmentation of high resolution remote sensed images. *Precis. Agric.* 8, 95–110. <https://doi.org/10.1007/s11119-007-9031-3>.
- 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. <https://doi.org/10.1016/j.isprsjprs.2020.01.013>.
- Fare Garnot, V.S., Landrieu, L., 2021. Panoptic Segmentation of Satellite Image Time Series with Convolutional Temporal Attention Networks, in: 2021 IEEE/CVF International Conference on Computer Vision (ICCV). Presented at the 2021 IEEE/CVF International Conference on Computer Vision (ICCV), IEEE, Montreal, QC, Canada, pp. 4852–4861. Doi: 10.1109/ICCV48922.2021.00483.
- Ge, P., Ren, C.-X., Xu, X.-L., Yan, H., 2023. Unsupervised Domain Adaptation via Deep Conditional Adaptation Network. *Pattern Recogn.* 134, 109088. <https://doi.org/10.1016/j.patcog.2022.109088>.
- 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. <https://doi.org/10.1016/j.jag.2022.102877>.
- Li, M., Long, J., Stein, A., Wang, X., 2023. Using a semantic edge-aware multi-task neural network to delineate agricultural parcels from remote sensing images. *ISPRS J. Photogramm. Remote Sens.* 200, 24–40. <https://doi.org/10.1016/j.isprsjprs.2023.04.019>.
- Liu, S., Liu, L., Xu, F., Chen, J., Yuan, Y., Chen, X., 2022. A deep learning method for individual arable field (IAF) extraction with cross-domain adversarial capability. *Comput. Electron. Agric.* 203, 107473. <https://doi.org/10.1016/j.compag.2022.107473>.
- Long, J., Shelhamer, E., Darrell, T., 2015. Fully Convolutional Networks for Semantic Segmentation.
- 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. <https://doi.org/10.1016/j.jag.2022.102871>.
- Long, M., Cao, Z., Wang, J., Jordan, M.I., n.d. Conditional Adversarial Domain Adaptation.
- Luo, C., Li, H., Zhang, J., Wang, Y., 2023. OB-ConvLSTM: A sequential remote sensing crop classification model with OBIA and ConvLSTM models, in: 2023 11th International Conference on Agro-Geoinformatics (Agro-Geoinformatics). Presented at the 2023 11th International Conference on Agro-Geoinformatics (Agro-Geoinformatics), IEEE, Wuhan, China, pp. 1–6. Doi: 10.1109/Agro-Geoinformatics59224.2023.10233296.
- Luo, B., Zhang, F., Liu, X., Pan, Q., Guo, P., 2021. Managing Agricultural Water Considering Water Allocation Priority Based on Remote Sensing Data. *Remote Sens. (Basel)* 13, 1536. <https://doi.org/10.3390/rs13081536>.

- Ma, Y., Chen, S., Ermon, S., Lobell, D.B., 2024. Transfer learning in environmental remote sensing. *Remote Sens. Environ.* 301, 113924 <https://doi.org/10.1016/j.rse.2023.113924>.
- Masoud, K.M., Persello, C., Tolpekin, V.A., 2019. 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. <https://doi.org/10.3390/rs12010059>.
- Matton, N., Canto, G., Waldner, F., Valero, S., Morin, D., Inglada, J., Arias, M., Bontemps, S., Koetz, B., Defourny, P., 2015. An Automated Method for Annual Cropland Mapping along the Season for Various Globally-Distributed Agrosystems Using High Spatial and Temporal Resolution Time Series. *Remote Sens. (Basel)* 7, 13208–13232. <https://doi.org/10.3390/rs71013208>.
- Mirza, M.J., Micorek, J., Possegger, H., Bischof, H., 2022. The Norm Must Go On: Dynamic Unsupervised Domain Adaptation by Normalization, in: 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Presented at the 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), IEEE, New Orleans, LA, USA, pp. 14745–14755. Doi: 10.1109/CVPR52688.2022.01435.
- 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. <https://doi.org/10.1016/j.patcog.2004.03.001>.
- Peng, L., Chen, X., Chen, J., Zhao, W., Cao, X., 2022b. Understanding the Role of Receptive Field of Convolutional Neural Network for Cloud Detection in Landsat 8 OLI Imagery. *IEEE Trans. Geosci. Remote Sensing* 60, 1–17. <https://doi.org/10.1109/TGRS.2022.3150083>.
- Peng, J., Huang, Y., Sun, W., Chen, N., Ning, Y., Du, Q., 2022a. Domain Adaptation in Remote Sensing Image Classification: A Survey. *IEEE J Sel. Top. Appl. Earth Observations Remote Sensing* 15, 9842–9859. <https://doi.org/10.1109/JSTARS.2022.3220875>.
- Persello, C., Bruzzone, L., 2010. A Novel Protocol for Accuracy Assessment in Classification of Very High Resolution Images. *IEEE Trans. Geosci. Remote Sensing* 48, 1232–1244. <https://doi.org/10.1109/TGRS.2009.2029570>.
- Persello, C., Tolpekin, V.A., Bergado, J.R., De By, R.A., 2019. Delineation of agricultural fields in smallholder farms from satellite images using fully convolutional networks and combinatorial grouping. *Remote Sens. Environ.* 231, 111253 <https://doi.org/10.1016/j.rse.2019.111253>.
- Ronneberger, O., Fischer, P., Brox, T., 2015. U-Net: Convolutional Networks for Biomedical Image Segmentation. In: Navab, N., Hornegger, J., Wells, W.M., Frangi, A.F. (Eds.), *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015*, Lecture Notes in Computer Science. Springer International Publishing, Cham, pp. 234–241. [https://doi.org/10.1007/978-3-319-24574-4\\_28](https://doi.org/10.1007/978-3-319-24574-4_28).
- Shi, X., Chen, Z., Wang, H., Yeung, D.-Y., Wong, W., Woo, W., 2015. Convolutional LSTM Network: A Machine Learning Approach for Precipitation Nowcasting. *Advances in neural information processing systems* 28.
- Song, W., Wang, C., Dong, T., Wang, Z., Wang, C., Mu, X., Zhang, H., 2023. Hierarchical extraction of cropland boundaries using Sentinel-2 time-series data in fragmented agricultural landscapes. *Comput. Electron. Agric.* 212, 108097 <https://doi.org/10.1016/j.compag.2023.108097>.
- Sun, W., Sheng, W., Zhou, R., Zhu, Y., Chen, A., Zhao, S., Zhang, Q., 2022. Deep edge enhancement-based semantic segmentation network for farmland segmentation with satellite imagery. *Comput. Electron. Agric.* 202, 107273 <https://doi.org/10.1016/j.compag.2022.107273>.
- Waldner, F., Diakogiannis, F.I., 2020. Deep learning on edge: Extracting field boundaries from satellite images with a convolutional neural network. *Remote Sens. Environ.* 245, 111741 <https://doi.org/10.1016/j.rse.2020.111741>.
- Wang, Z., Dai, Z., Poczos, B., Carbonell, J., 2019. Characterizing and Avoiding Negative Transfer, in: 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Presented at the 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), IEEE, Long Beach, CA, USA, pp. 11285–11294. Doi: 10.1109/CVPR.2019.01155.
- Wang, H., Shen, T., Zhang, W., Duan, L.-Y., Mei, T., 2020. Classes Matter: A Fine-Grained Adversarial Approach to Cross-Domain Semantic Segmentation, in: Vedaldi, A., Bischof, H., Brox, T., Frahm, J.-M. (Eds.), *Computer Vision – ECCV 2020*, Lecture Notes in Computer Science. Springer International Publishing, Cham, pp. 642–659. Doi: 10.1007/978-3-030-58568-6\_38.
- Wang, Y., Feng, L., Zhang, Z., Tian, F., 2023. An unsupervised domain adaptation deep learning method for spatial and temporal transferable crop type mapping using Sentinel-2 imagery. *ISPRS J. Photogramm. Remote Sens.* 199, 102–117. <https://doi.org/10.1016/j.isprsjprs.2023.04.002>.
- Wang, S., Waldner, F., Lobell, D.B., 2022. Unlocking Large-Scale Crop Field Delineation in Smallholder Farming Systems with Transfer Learning and Weak Supervision. *Remote Sens. (Basel)* 14, 5738. <https://doi.org/10.3390/rs14225738>.
- Wen, C., Lu, M., Bi, Y., Zhang, S., Xue, B., Zhang, M., Zhou, Q., Wu, W., 2022. An Object-Based Genetic Programming Approach for Cropland Field Extraction. *Remote Sens. (Basel)* 14, 1275. <https://doi.org/10.3390/rs14051275>.
- Woo, S., Park, J., Lee, J.-Y., Kweon, I.S., 2018. CBAM: Convolutional Block Attention Module, in: Ferrari, V., Hebert, M., Sminchisescu, C., Weiss, Y. (Eds.), *Computer Vision – ECCV 2018*, Lecture Notes in Computer Science. Springer International Publishing, Cham, pp. 3–19. Doi: 10.1007/978-3-030-01234-2\_1.
- Xiao, N., Zhang, L., 2021. Dynamic Weighted Learning for Unsupervised Domain Adaptation, in: 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Presented at the 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), IEEE, Nashville, TN, USA, pp. 15237–15246. Doi: 10.1109/CVPR46437.2021.01499.
- Yan, X., Shen, L., Pan, J., Wang, J., Chen, C., Li, Z., 2023. ALNet: Auxiliary Learning-Based Network for Weakly Supervised Building Extraction From High-Resolution Remote Sensing Images. *IEEE Trans. Geosci. Remote Sensing* 61, 1–16. <https://doi.org/10.1109/TGRS.2023.3314465>.
- Yang, Y., Lao, D., Sundaramoorthi, G., Soatto, S., 2020. Phase Consistent Ecological Domain Adaptation, in: 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Presented at the 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), IEEE, Seattle, WA, USA, pp. 9008–9017. Doi: 10.1109/CVPR42600.2020.00903.
- Zanaga, D., Van De Kerchove, R., Daems, D., De Keersmaecker, W., Brockmann, C., Kirches, G., Wevers, J., Cartus, O., Santoro, M., Fritz, S., Lesiv, M., Herold, M., Tsandbazar, N.-E., Xu, P., Ramoino, F., Arino, O., 2022. ESA WorldCover 10 m 2021 v200. Doi: 10.5281/zenodo.7254221.
- Zhang, P., Zhang, B., Zhang, T., Chen, D., Wang, Y., Wen, F., 2021. Prototypical Pseudo Label Denoising and Target Structure Learning for Domain Adaptive Semantic Segmentation, in: 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Presented at the 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), IEEE, Nashville, TN, USA, pp. 12409–12419. Doi: 10.1109/CVPR46437.2021.01223.
- Zhang, H., Liu, M., Wang, Y., Shang, J., Liu, X., Li, B., Song, A., Li, Q., 2021a. Automated delineation of agricultural field boundaries from Sentinel-2 images using recurrent residual U-Net. *Int. J. Appl. Earth Obs. Geoinf.* 105, 102557 <https://doi.org/10.1016/j.jag.2021.102557>.
- Zhang, X., Liu, L., Chen, X., Gao, Y., Xie, S., Mi, J., 2021c. GLC\_FCS30: global land-cover product with fine classification system at 30 m using time-series Landsat imagery. *Earth Syst. Sci. Data* 13, 2753–2776. <https://doi.org/10.5194/essd-13-2753-2021>.
- 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 <https://doi.org/10.1016/j.rse.2020.111912>.
- Zhou, Z., Qi, L., Shi, Y., 2022. Generalizable Medical Image Segmentation via Random Amplitude Mixup and Domain-Specific Image Restoration, in: Avidan, S., Brostow, G., Cissé, M., Farinella, G.M., Hassner, T. (Eds.), *Computer Vision – ECCV 2022*, Lecture Notes in Computer Science. Springer Nature Switzerland, Cham, pp. 420–436. Doi: 10.1007/978-3-031-19803-8\_25.