IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

IEEE GEOSCIENCE AND REMOTE SENSING SOCIETY SECTION

Received 3 July 2024, accepted 28 July 2024, date of publication 1 August 2024, date of current version 23 September 2024.

Digital Object Identifier 10.1109/ACCESS.2024.3436620

RESEARCH ARTICLE

Cross Domain Early Crop Mapping Using CropSTGAN

YIQUN WANG, HUI HUANG, AND RADU STATE
Services and Data Management Research Group (SEDAN), Interdisciplinary Centre for Security, Reliability and Trust (SnT), University of Luxembourg, 1359 Luxembourg City, Luxembourg

Corresponding author: Hui Huang (hui.huang@uni.lu)

ABSTRACT Driven by abundant satellite imagery, machine learning-based approaches have recently been promoted to generate high-resolution crop cultivation maps to support many agricultural applications. One of the major challenges faced by these approaches is the limited availability of ground truth labels. In the absence of ground truth, existing work usually adopts the ''direct transfer strategy'' that trains a classifier using historical labels collected from other regions and then applies the trained model to the target region. Unfortunately, the spectral features of crops exhibit inter-region and inter-annual variability due to changes in soil composition, climate conditions, and crop progress, the resultant models perform poorly on new and unseen regions or years. Despite recent efforts, such as the application of the deep adaptation neural network (DANN) model structure in the deep adaptation crop classification network (DACCN), to tackle the above cross-domain challenges, their effectiveness diminishes significantly when there is a large dissimilarity between the source and target regions. This paper introduces the Crop Mapping Spectral-temporal Generative Adversarial Neural Network (CropSTGAN), a novel solution for cross-domain challenges, that doesn't require target domain labels. CropSTGAN learns to transform the target domain's spectral features to those of the source domain, effectively bridging large dissimilarities. Additionally, it employs an identity loss to maintain the intrinsic local structure of the data. Comprehensive experiments across various regions and years demonstrate the benefits and effectiveness of the proposed approach. In experiments, CropSTGAN is benchmarked against various state-of-the-art (SOTA) methods. Notably, CropSTGAN significantly outperforms these methods in scenarios with large data distribution dissimilarities between the target and source domains.

INDEX TERMS Early crop mapping, multispectral image data, cross domain, domain adaptation, CropSTGAN, cropland data layer.

I. INTRODUCTION

Early crop mapping, i.e. determining the cultivation regions of crops before their harvest season, is the fundamental building block for agricultural planning, resource allocation, crop insurance, risk management [1], [2] and many other agricultural applications. Since the spectral features of vegetation are determined by their structure, leaf biochemistry and phonological stages [3], time-series analysis on multi-spectral images captured by satellites is the dominant approach for land cover classifications. With the rapid growth in data volume and increasing accessibility of satellite imagery, deep learning (DL) based remote sensing has been promoted in recent years to produce high-resolution, high-accuracy crop cultivation maps [4]. This class of approaches relies on a large amount of ground truth data, also known as labels, to train and validate the classification model. The ground truth can be obtained from surveys as first-hand labels [5], [6], or use public datasets, such as the United States Department of Agriculture (USDA)'s Cropland Data Layer (CDL), as weak labels for model training [7]. Different DL architectures, such as convolutional neural networks (CNNs), Temporal Convolutional Neural Network (TempCNN) [8], deep autoencoders and recurrent neural networks with long

The associate editor coordinating the review of this manuscript and approving it for publication was Wei He.

short-term memory (LSTM), are explored for crop mapping tasks [9], [10], [11]. The results suggest DL approaches outperform conventional support vector machine (SVM) and tree-based models in providing semantic information on the input images.

Unfortunately, obtaining appropriate ground truth for an arbitrary region is challenging. Public crop-type ground truth datasets, such as Cropland Data Layer (CDL) produced by National Agricultural Statistics Service (NASS), provide reliable and timely references for model training. However, these datasets are only available for a few countries and are usually released after harvest season. The collecting process can be costly, labour-intensive, and sometimes unfeasible, especially in underdeveloped countries. In the absence of ground truth, existing work usually adopts the ''direct transfer strategy'' that trains a classifier first using available labelled data for other regions and then applies the trained model to the target region [12], [13]. Nevertheless, spectral features of crops have both inter-region variability and inter-annual variability due to changes in soil composition, climate conditions, and crop progress [14]. These variability collectively contribute to the distribution shift between the training data (source domain) and the test data (target domain), called the cross-domain issue. Consequently, the direct transfer strategy often leads to poor performance on new and unseen regions and years as it compromises an implicit assumption of the machine learning-based crop mapping approaches: the labelled training data from the source domain and the data from the target region are independent and identically distributed (i.i.d), or at least come from similar distributions.

To address this cross-domain issue, including cross-region and cross-year issues, various methods have been developed aiming to enhance the model's ability on unseen domains. One effective approach involves training the model by incorporating multi-year data and the respective phonological metrics as the major inputs [15]. This method helps the model to capture the temporal variations in spectral patterns for target crops caused by changing environmental and climate conditions. Their methodology augments the model's proficiency in generalizing across diverse years within a singular region for which multi-annual ground truth data exists. However, the predicament of cross-regional issue remains unaddressed.

To mitigate the challenges associated with both inter-regional and inter-annual cross-domain issues, methodologies such as Domain Adversarial Neural Networks (DANN) [16] and their derivatives have been deployed. These approaches endeavour to delineate invariant features from data across both target and source domains, leveraging these features for enhanced accuracy in crop classification mappings. However, these methodologies hinge on the presupposition that the data distributions between the target and source domains exhibit relatively minor disparities. Furthermore, all these methodologies operate under the assumption that the crop types are identical in both the source and target domains.

To address this domain shift challenge encountered in early crop mapping endeavors under relevantly substantial discrepancies in data distribution across the domains, this paper introduces the Crop Mapping Spectral-temporal Generative Adversarial Neural Network (CropSTGAN) framework. The system's primary goal is to identify a specified crop variety, such as corn, within a target area lacking labelled data. To fulfil this aim, the CropSTGAN framework employs an unsupervised domain adaptation (UDA) strategy, a technique for adapting a model from a source domain with labelled data to an unlabelled target domain. The innovative method learns a function that transforms the spectral characteristics features of the target area to the source domain while retaining their local structure. For instance, the spectral features of corn in the target domain may diverge from those in the source domain. The CropSTGAN framework first transforms the target domain's spectral features to resemble those of the source domain, thereby minimizing the discrepancies in feature patterns between the transformed target data and the original source data, all the while maintaining the distinguishability of corn and other land cover types within the target region. This process enables the straightforward application of a crop mapping model, trained within the source domain, to the target domain without diminishing its precision.

From the highest level, the proposed CropSTGAN system consists of three key components: the pre-processor, the CropSTGAN domain mapper, and the TempCNN crop mapper. The pre-processor module employs linear interpolation to fill gaps due to cloud coverage in the Multi-Spectral Images (MSI), ensuring a complete time series. The CropSTGAN domain mapper, a modification of the CycleGAN architecture [17], is designed with a specific structure to capture the temporal features from the time-series MSI data and transform them from the target domain to the source domain. The CropSTGAN consists of two generator networks and two discriminator networks. The generators learn to transform data points from one domain to another domain, while the discriminators distinguish between the transformed data points provided by the generators and the original data points. Finally, the TempCNN crop mapper, a CNN-based model shared structure with the CropSTGAN discriminators, can be directly applied to the transformed target data to accurately determine the cultivation locations of the specified target crop on the target domain. To evaluate the distinct architecture of the CropSTGAN domain mapper, comparative experiments were undertaken utilizing a simpler, analogous structure named Crop Mapping Temporal Generative Adversarial Neural Network (CropTGAN), alongside benchmarking against several state-of-the-art (SOTA) methodologies, including TempCNN and STDAN. The empirical findings corroborated that the CropSTGAN architecture enhances crop mapping accuracy under the relevantly substantial discrepancies in data distribution across the domains.

Our contributions can be summarized as follows:

- Propose the CropSTGAN framework, consisting of a pre-processor, a CropSTGAN domain mapper, and a TempCNN crop mapper, to address the cross-domain issue due to the inter-region and inter-year variations in remote sensing-based early crop mapping under substantial discrepancies in data distribution across the domains.
- Design the CropSTGAN domain mapper to capture the temporal and spectral features from the time-series MSI data and learn to transform the target domain data into the source domain.
- Conduct cross-region and cross-year experiments in the study areas from the USA and China to evaluate the CropSTGAN framework. The results demonstrate superior performance compared to CropTGAN and several SOTA methods confirming the effectiveness and accuracy of the CropSTGAN framework for cross-domain early crop mapping.
- Conduct the comprehensive series of experiments based on CropSTGAN for cross-region analysis using MSI data collected at intervals of 50, 60, 70, 80, and 90 days starting from May 1st. The results indicate that longer time-series lengths yield better metrics, with results stabilizing from 80 days and the best performance observed at 90 days.

The rest of this paper is organised as follows. Section II describes related works. Section III presents the data and study areas. The methodology is described in Section IV. Experiment setup and results are presented in Section V. The discussion is presented in Section VI. Finally, Section VII and Section VIII provide conclusions and future works.

II. RELATED WORKS

Remote sensing-based crop-type mapping has undergone significant advancements with the adoption of machine learning methods. Traditionally, techniques such as SVM and random forest (RF) have been widely utilized for crop classification using remote sensing data [18], [19], [20]. However, the advent of deep learning has brought about a revolution in this field by leveraging their ability to automatically extract meaningful representations from data.

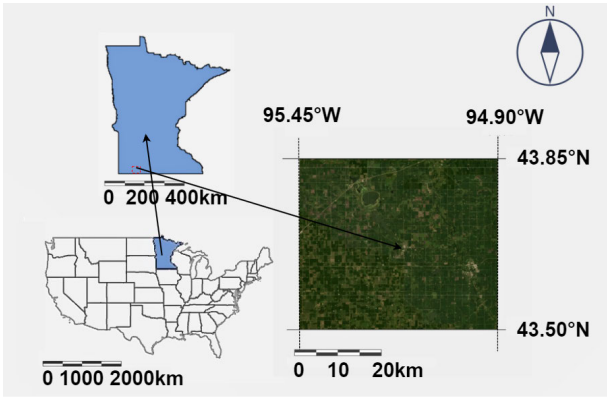
Deep learning models, such as LSTM networks and CNNs, have exhibited impressive performance in crop type mapping using remote sensing imagery [9], [11], [21]. They excel at capturing temporal and spatial characteristics, enabling accurate classification of various crop types. By training on labelled crop datasets, these models can learn complex relationships between spectral, spatial, and temporal features, leading to improved accuracy in distinguishing crop types.

To locate the target crop in the target domain without ground truth data, direct transfer techniques have been introduced, allowing the knowledge learned from a source domain, where labelled samples are abundant, to be directly applied to the target domain [12], [22]. It involves training models on regions with abundant labelled samples and directly applying them to other regions or years. This approach assumes that the knowledge learned from the source domain is applicable to the target domain. While direct transfer can be a simple and effective method, it may encounter a cross-domain problem when it comes to effectively capturing and comprehending the intricate relationships and variations in crop patterns across different regions and years.

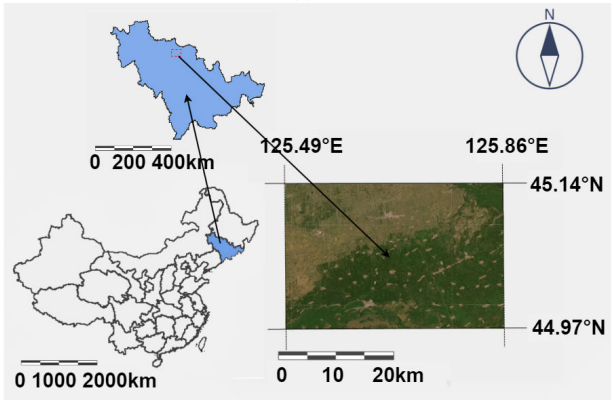
To address the cross-year issue, training the model with multi-year crop data proves to be a valuable adaptation method. Recent studies, including [15], emphasize the importance of incorporating multi-year crop data from a specific region as it enhances the model's understanding of temporal patterns, variations, and trends within that geographic context. This approach enhances the accuracy and reliability of crop classification by considering inter-annual climate variations and capturing the unique characteristics of the region. Leveraging region-specific multi-year data enables the model to become more robust to phenology shifts, ensuring consistent performance across different years. Training on multi-year crop data from the same region serves as an effective adaptation technique, resulting in more precise and dependable crop mapping outcomes within a particular geographic area. A key drawback of the multi-year crop data training method is its high reliance on labelled data. To effectively train deep learning models, a large number of accurately labelled samples are needed. However, collecting such a vast amount of labelled data from diverse regions and years is a challenging task. The process is time-consuming, costly, and often impractical due to the effort and resources required. Moreover, it fails to address the issue of the cross-region issue.

To tackle the challenges of cross-year and cross-region adaptability, current methods adopt two main strategies. From a sample perspective, fine-tuning pre-trained models with a few high-quality target domain samples is a common practice. This method adapts the original model to the new data distribution. For example, [23] improved deep models for nationwide land cover classification using high-confidence pseudo-labels. Similarly, [24] involved annotating new target domain samples to refine RF classifiers via active learning. However, this technique often necessitates labelling samples, which is not feasible for large-scale studies.

From a feature perspective, in the field of UDA, earlier studies have developed DANN-based methods to align samples from various regions into a unified feature space, minimizing disparities in deep features. For instance, [25] introduced the STDAN, a novel unsupervised domain adaptation framework for crop type classification. In the STDAN study [25], two case studies comprising four areas from Gangwon Province and Gyeongsang Province, Korea were selected. These areas, while not geographically distant, exhibit slight environmental differences. Moreover, the PAN [26] and the DACCN [27] extended the loss function using the Maximum Mean Discrepancy (MMD) and the



(a)



(b)

FIGURE 1. The study areas. (a) The study area in jackson county, minnesota, USA. (b) The study area, jilin province, china.

Multiple Kernel variant of Maximum Mean Discrepancy (MK-MMD), achieving improved accuracy compared to CNN and LSTM methods without domain adaptation. Similarly, the PAN study [27] chose three areas from Sichuan Province, Hubei Province, and Anhui Province in China, all positioned at similar longitudes, also leading to minor environmental variations Wang et al. The DACCN study [26], addressing cross-country issues for the first time in crop mapping experiments, selected sub-areas from Heilongjiang Province and Jilin Province, China as target domains, with four sub-areas from different states in the USA as source domains. This cross-country setup introduced significantly larger environmental variations compared to the previous studies. Additionally, the study included cross-year experiments in the USA. Notably, DACCN's performance metrics were significantly lower for cross-country issues than for cross-year experiments, indicating a reduction in efficiency due to substantial discrepancies in data distribution across countries. In other words, although these approaches adeptly tackle the issue of missing labels and diminish the detrimental impacts of domain shifts, their efficacy is constrained by relevantly substantial discrepancies in data distribution across the domains.

TABLE 1. The environmental conditions of the source domain and the target domains. "SA" represents the Study Area. "T" represents the yearly average temperature. "P" stands for the average hourly precipitation. "E" signifies the average hourly evaporation. "R" indicates the surface net solar radiation.

SA	Year	T(K)	P(mm/h)	E(mm/h)	R(kJ/m^2)
Jackson	2019	289.02	2.60	-1.01	4688.12
Jackson	2020	289.93	1.48	-1.11	5118.46
Jackson	2021	290.76	1.06	-1.06	5245.53
SA in China	2019	290.03	1.99	-1.99	12542.54

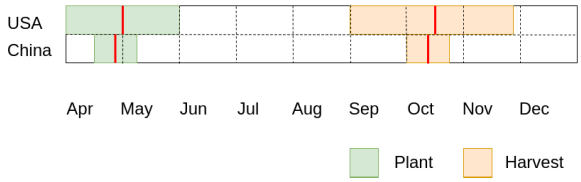


FIGURE 2. The crop calendar for corn. The red lines represent the average dates.

This paper introduces the CropSTGAN framework, designed to tackle the challenge of cross-domain early crop mapping without the need for labelled data from the target domain, even when there are significant differences between the data distributions of the target and source domains.

III. DATA AND STUDY AREAS

A. STUDY AREA

This paper chose corn as the example crop to demonstrate the cross-domain capability of the proposed approach. The objective is to map corn cultivation locations in a target domain using models trained by labelled data from a source domain. The target and source domains could differ in their geographic locations (i.e., cross-region) or be in the same region in different years (i.e., cross-year). Note that, the proposed approach can be applied to other types of crops. Toward this end, the study areas include the two regions: Jackon County of the USA, and one study area from Jilin Province, China. For a visual depiction of the geographical locations of these areas, please refer to Figure 1.

The study areas are characterized by unique environmental conditions such as temperature, precipitation, elevation, and solar radiation, which also vary over years. The 5th major atmospheric reanalysis (ERA5) produced by ECMWF [28], is utilized in our work with a resolution of $11132m \times 11132m$ from 2019 to 2021 as the climate data which is comprised of air and soil temperature, precipitation, Evaporation, and solar radiation. Table 1 presents the environmental metrics for these areas. Additionally, the corn cropping schedules differ among the regions, as depicted in Figure 2. These variations influence corn growth, resulting in distinct MSI feature patterns of the same crop across the different areas and time periods. Figures 2a, 2b, and 2c show the average NDVI values curves for the early growth stages of corn and other crops in Jackson County from 2019 to 2021. Figure 2d presents the curves in the study area of China. The corn

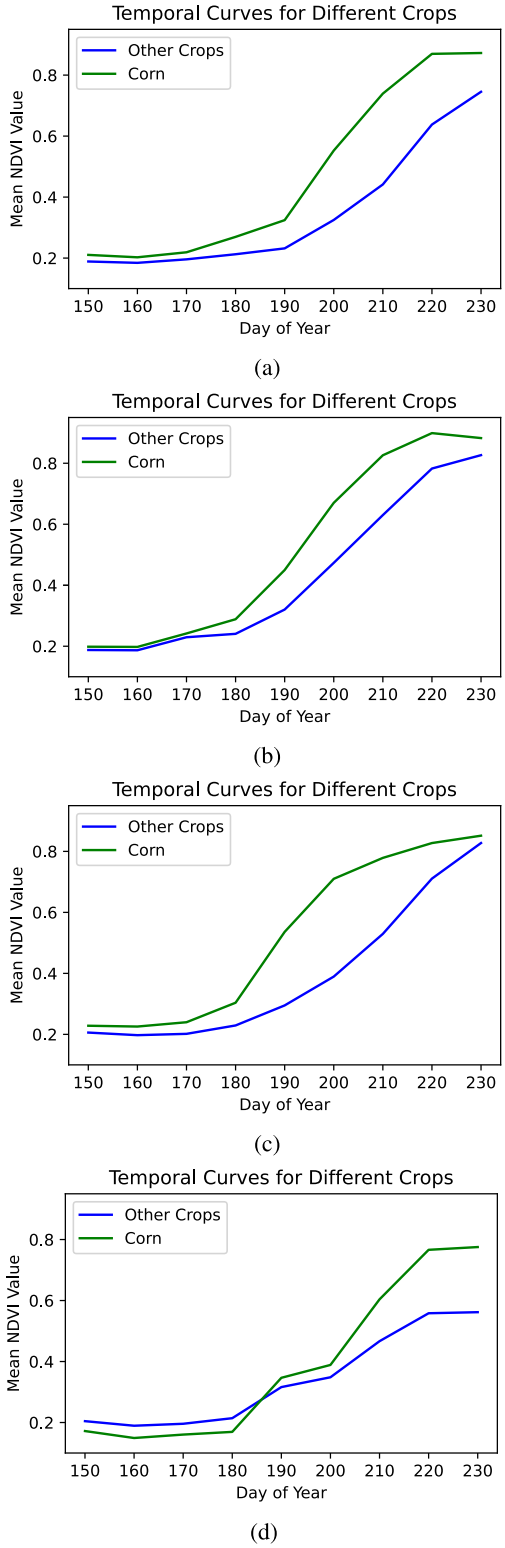


FIGURE 3. The Average Time-series NDVI of Corn for the Source Domain and Target Domain. (a) Jackson County, 2019. (b) Jackson County, 2020. (c) Jackson County, 2021. (d) The study area of China, 2019.

growth NDVI patterns in Jackson County vary across the different years. Additionally, a clear distinction is evident between the NDVI values in Jackson County and those in the Chinese study area.

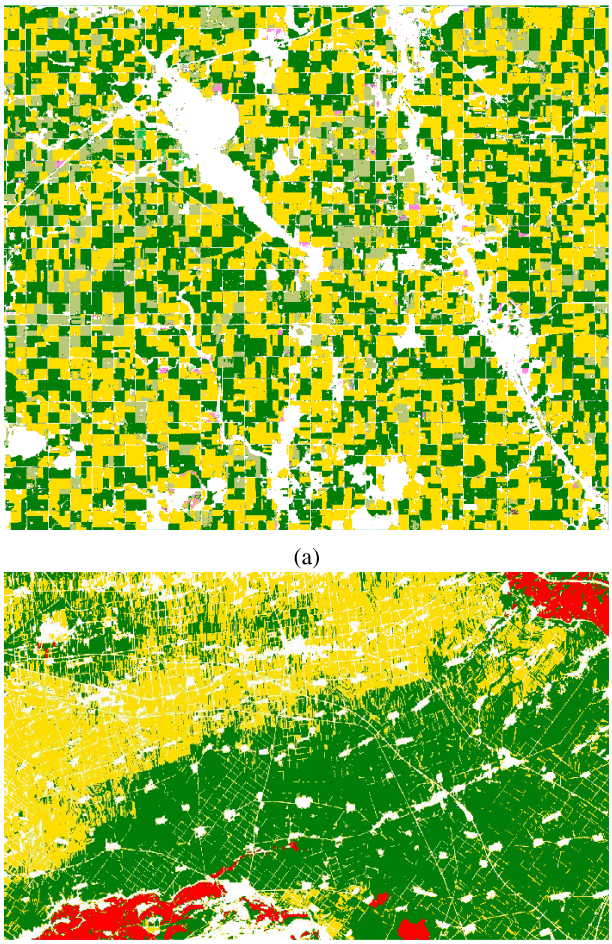


FIGURE 4. Ground Truth Data: Green represents corn, yellow represents soybean, red represents rice, light olive represents fallow cropland, pink represents alfalfa, and white represents non-cropland. (a) The USA Study Area 2019, (b) China Study Area 2019.

B. REFERENCE DATA

The CDL is used as the ground truth for the source domain. The CDL [7], a crop-specific land cover raster map dataset available for the entire conterminous U.S. land area (CONUS) at 30 m resolution provided by the USDA, regularly provides information on the annual temporal and spatial distribution of corn, as well as the area dedicated to its cultivation. As shown in Figure 4a, the main crop types in the study areas are corn and soybean, with alfalfa being a minor crop, cultivated in smaller proportions. Additionally, there are idle cropland.

For the target domain, official ground reference data for China is currently unavailable. Fortunately, [6] published maps of corn, soybean, and rice with a spatial resolution of 10 meters for Northeast China from 2017 to 2019. As a result, the 2019 crop maps from [6] are regarded as the ground truth

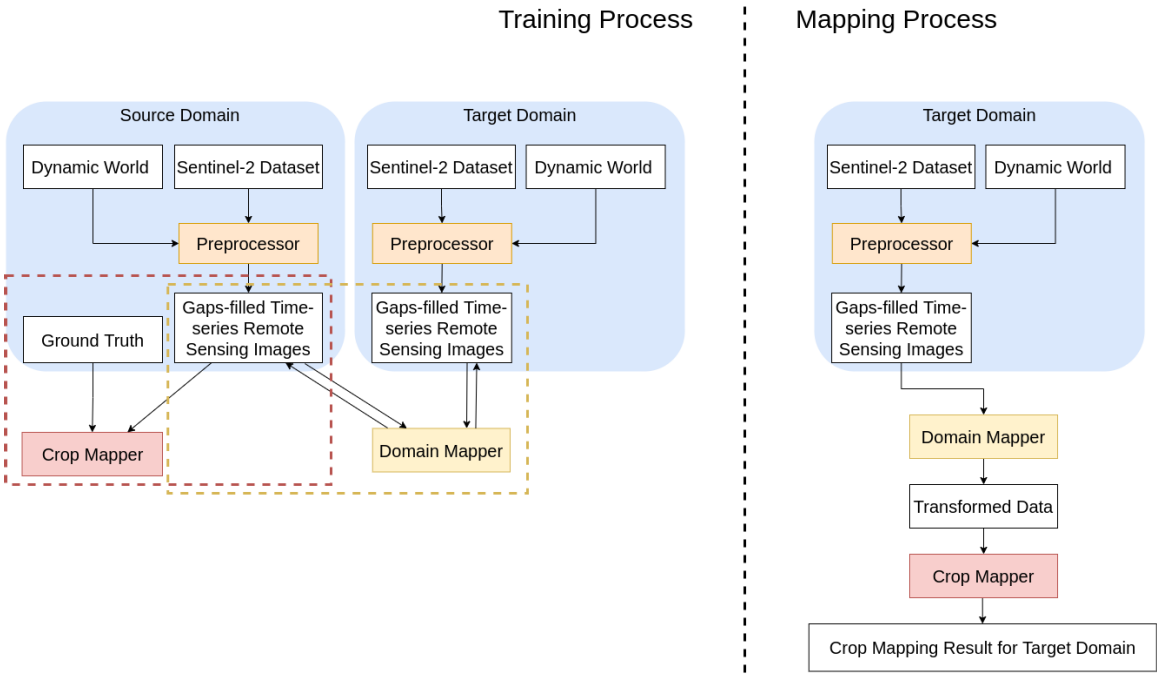


FIGURE 5. The architecture of the CropSTGAN framework.

reference for the study area in China. As shown in Figure 4b, the main crop types in the study areas are corn, soybean, and rice.

The ground truth data used are either from publicly available datasets or previous research efforts. Consequently, during model comparison, the evaluation is based on determining which model aligns more closely with previous methods rather than accurately reflecting reality. This is a significant shortcoming in model evaluation within the crop mapping research field. A more accurate evaluation method would involve using manually labeled data, but this process is extremely labor-intensive and time-consuming, making it impractical to undertake. However, the ground truth data from the USDA CDL and the 2019 ground truth data for China both achieve classification accuracies of approximately 85% to 95% for major crop-specific land cover categories. These datasets are widely used in crop mapping research papers for evaluation purposes, and the impact of this accuracy on the evaluation process is not particularly pronounced or severe.

The remote sensing data in this work are MSI images captured by the Sentinel-2 satellites, which have been widely used for many agricultural applications in the community [9], [29]. Sentinel-2 provides high-resolution MSI images (up to 10m) with a revisit time of 5 days, allowing for frequent monitoring of crop growth and changes. Its wide spectral coverage, including visible, near-infrared, and shortwave infrared bands, enables accurate assessment of vegetation health, crop type identification, and mapping. In this study, 10 bands, including B2 (Blue), B3 (Green), B4 (Red), B5 (Red Edge 1), B6 (Red Edge 2), B7 (Red Edge 3), B8 (Near-Infrared), B8A (Red Edge 4), B11 (Shortwave Infrared 1) and B12 (Shortwave Infrared 2), are used to map the target crop in early seasons.

As described, our primary objective is to locate the specific crop (corn) with cross-domain problems at an earlier growth stage. As a result, the time series remote sensing data should start after the planting period and conclude before the onset of the crop harvest period. Referring to Figure 2, it can be observed that the corn harvesting season typically starts in early September in the USA and begins in early October in China. Planting, on the other hand, starts at the beginning of April in the USA, and in mid-April in Jilin Province, China. As a result, the date of our annual collection of remote sensing data spans from May 1st to July 30th, encompassing a total of nine-time points with a ten-day observation window.

Additionally, this work employs the Dynamic World dataset [30]. It is a high-resolution 10m near-real-time (NRT) Land Use/Land Cover (LULC) dataset and features class probabilities and label data for nine distinct categories, including cropland. It is utilized to identify and select areas of cropland, enabling us to maintain a concentrated analysis solely within these regions.

C. REMOTE SENSING DATA

IV. METHODOLOGY

A. PROBLEM STATEMENT

Let $\mathbf{X}$ denote the time series remote sensing input data and $\mathbf{Y}$ denote the ground truth labels. Each sample $\mathbf{x}$ can be expressed as a temporal form $[\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_t]$, where $\mathbf{x}_i$ represents input at time i. $\mathbf{x}_i$ can be further expanded as

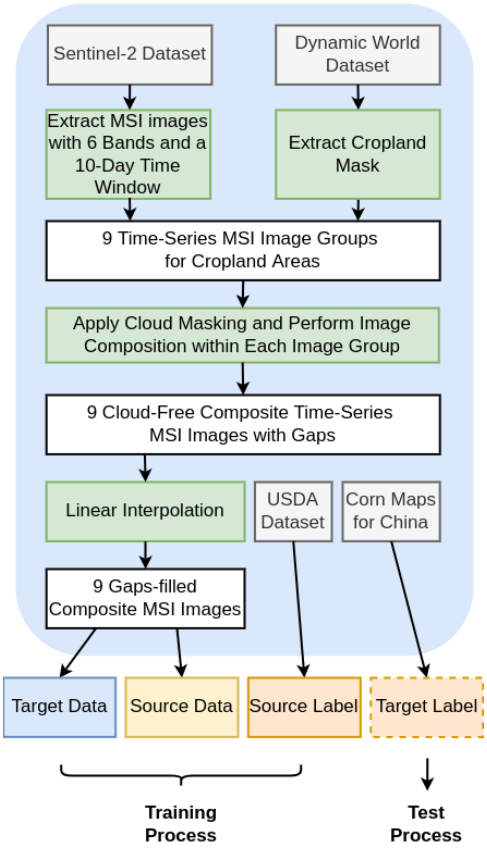


FIGURE 6. The Architecture of the Pre-processor.

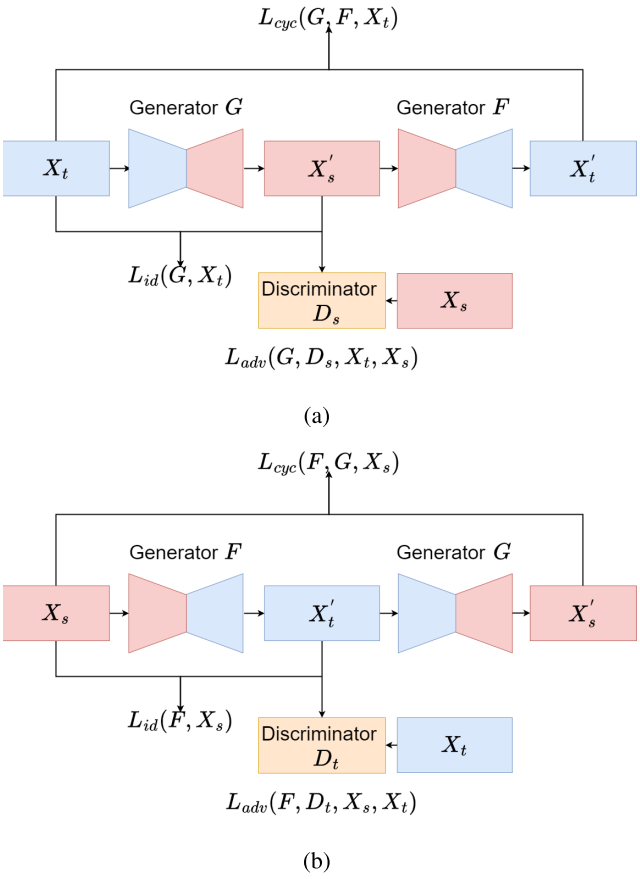


FIGURE 7. The CropSTGAN Domain Mapper Structure. (a) Cycle 1: The training process involves data flowing from the target domain to the source domain and then back to the target domain. The notation $\mathbf{X}_s'$ represents the transformed target domain data, while $\mathbf{X}_t'$ denotes the data transformed back to the target domain. (b) Cycle 2: The training process involves data flowing from the source domain to the target domain and then back to the source domain. The notation $\mathbf{X}_t'$ represents the transformed source domain data, while $\mathbf{X}_s'$ denotes the data transformed back to the source domain.

$[\mathbf{x}_{i1}, \mathbf{x}_{i2}, \ldots, \mathbf{x}_{ib}]$, containing multi-spectral bands information from band 1 to band b. Each ground truth label, denoted as y, is represented as a binary number, which corresponds to the categories of corn and other crops. Let $\mathbf{X}_t$ denote the time series input target data, $\mathbf{X}_s$ denote the time series input source data, and $\mathbf{Y}_s$ denote the GT for $\mathbf{X}_s$. Each sample $\mathbf{x}_s$ has a corresponding $\mathbf{y}_s$.

Our objective is to identify target crops to get target labels $\mathbf{Y}_t$ in the target domain during their early growth stages, utilizing labelled source domain data ($\mathbf{X}_s$, $\mathbf{Y}_s$) and unlabelled target domain data ($\mathbf{X}_t$). This approach addresses challenges related to the cross-domain issue and the large data distribution discrepancies issue.

B. SYSTEM OVERVIEW

The architecture of our CropSTGAN framework is depicted in Figure 5, showcasing both the training and mapping dataflows. This system includes three main components: the pre-processor, the CropSTGAN domain mapper, and the TempCNN crop mapper. During training, the pre-processor enhances MSI images quality by filling gaps through linear interpolation, ensuring complete MSI image series $\mathbf{X}_s$ and $\mathbf{X}_t$ for both source and target domains. The CropSTGAN domain mapper, trained with randomly sampled data from these MSI images, enables the transfer of time-series remote sensing data between domains: $\mathbf{X}_t \rightarrow \mathbf{X}_s$ and $\mathbf{X}_s \rightarrow \mathbf{X}_t$. The TempCNN crop mapper is then trained on labelled source domain data ($\mathbf{X}_s$, $\mathbf{Y}_s$) to identify the target crop in the source domain at an early stage.

For mapping, the process begins with the pre-processor supplying the target domain MSI image series $\mathbf{X}_t$. These images are then transformed by the CropSTGAN domain mapper from the target to the source domain as $\mathbf{X}_s'$. The trained TempCNN crop mapper uses these transformed images $\mathbf{X}_s'$ to locate the target crop $\mathbf{Y}_t$ in the target domain. This system design ensures accurate and reliable early-stage crop mapping in the target domain, particularly useful when labelled data is scarce or nonexistent, thereby overcoming the challenges of cross-domain mapping.

C. PRE-PROCESSOR

The pre-processor aims at providing complete time-series MSI data by filling gaps between MSI images due to cloud

TABLE 2. The Generator Structure. Each encoder layer consists of an instance normalization layer, an activation layer and a pooling layer. Each decoder layer consists of an instance normalization layer, an activation layer and an upsampling layer.

Layer	Input Shape	Output Shape	Activation
Input	9×10	9×10	-
Encoder 1	9×10	9×1×8	LeakyReLU
Encoder 2	9×1×4	7×1×8	LeakyReLU
Encoder 3	7×1×8	5×1×16	LeakyReLU
Encoder 4	5×1×16	3×1×32	LeakyReLU
Decoder 4	3×1×32	5×1×16	LeakyReLU
Decoder 3	5×1×16	7×1×8	LeakyReLU
Decoder 2	7×1×8	9×1×4	LeakyReLU
Decoder 1	9×1×4	9×10×4	LeakyReLU
Output	9×10×4	9×10	ReLU

cover, atmospheric interference, or sensor limitations. The processing pipeline is outlined in Figure 6.

The procedure illustrated in the figure begins with the collection of MSI images covering the entire targeted study areas, sourced from the Sentinel-2 Dataset at consistent 10-day intervals. The primary objective is to detect the target crop at an early growth stage, addressing cross-domain challenges. Therefore, the selected time series of remote sensing data spans from May 1st to July 30th, starting post-planting and concluding pre-harvest, resulting in nine image sets. To refine the focus on agricultural lands, the method employs a cropland mask from the Dynamic World dataset during the crop's growing season, adjusting it to a uniform 30-meter resolution. This step ensures the exclusion of non-agricultural areas, concentrating the analysis on crop areas within the nine groups of MSI images. For each set, cloud detection is performed using Sentinel Hub's tool, filtering out cloud-covered areas. The process averages the values of cloud-free images to produce a composite MSI image with a clear 30-meter resolution. This approach yields nine high-quality, cloud-free time-series MSI images, each comprising 10 bands and maintaining a 30-meter spatial resolution. Finally, the ground truth data are reprojected into the same projection system of the MSI images with 30-m resolution. Subsequently, the MSI data are aligned with the corresponding ground truth labels, and the training data along with the matched labels are extracted.

D. CropSTGAN DOMAIN MAPPER

1) THE GENERATORS AND DISCRIMINATORS STRUCTURES

The domain mapper is referred as CropSTGAN, which stands for Crop Mapping Spectral-temporal Generative Adversarial Network. CropSTGAN domain mapper serves as the key component in our framework for transforming time-series remote sensing data points between different domains. As shown in Figure 7, the domain mapper consists of two generator networks, G and F, and two discriminator networks, D_X and D_Y. In the training process, the generators and discriminators play distinct roles. The generators' primary task is to transform the data from one domain to another: $G : \mathbf{X}_t \rightarrow \mathbf{X}'_s$ and $F : \mathbf{X}_s \rightarrow \mathbf{X}'_t$, while the discriminator's main objective is to differentiate between real and transformed data. As the training progresses, both networks improve iteratively until an equilibrium is reached where the generator generates highly realistic data and the discriminator cannot reliably distinguish between real and fake samples. For example, when the discriminator D_s can not distinguish whether the transformed data $\mathbf{X}'_s$ generated by the generator G is from the source domain or not, it means that the generator G has been well-trained.

TABLE 3. The discriminator/crop mapper structure. The last layer of the discriminator uses ReLU as its activation function, while the classifier employs softmax.

Layer	Input Shape	Output Shape	Activation
Input	9×10	9×10	-
Conv 1	9×10	9×1×8	LeakyReLU
Conv 2	9×1×8	7×1×8	LeakyReLU
Conv 3	7×1×8	5×1×8	LeakyReLU
Conv 4	5×1×8	3×1×8	LeakyReLU
Flatten	3×1×8	24	-
FC1	24	4	ReLU
Output	4	1	ReLU or Softmax

TABLE 4. The generator structure of CropTGAN. Each encoder layer consists of an instance normalization layer, an activation layer and a pooling layer. Each decoder layer consists of an instance normalization layer, an activation layer and an upsampling layer.

Layer	Input Shape	Output Shape	Activation
Input	9×10	9×10	-
Encoder 1	9×10	9×10×4	LeakyReLU
Encoder 2	9×10×4	7×10×8	LeakyReLU
Encoder 3	7×10×8	5×10×16	LeakyReLU
Encoder 4	5×10×16	3×10×32	LeakyReLU
Decoder 4	3×10×32	5×10×16	LeakyReLU
Decoder 3	5×10×16	7×10×8	LeakyReLU
Decoder 2	7×10×8	9×10×4	LeakyReLU
Decoder 1	9×10×4	9×10×4	LeakyReLU
Output	9×10×4	9×10	ReLU

TABLE 5. The discriminator/crop mapper structure of CropTGAN. The last layer of the Discriminator uses ReLU as its activation function, while the classifier employs softmax.

Layer	Input Shape	Output Shape	Activation
Input	9×10	9×10	-
Conv 1	9×10	9×10×8	LeakyReLU
Conv 2	9×10×8	7×10×8	LeakyReLU
Conv 3	7×10×8	5×10×8	LeakyReLU
Conv 4	5×10×8	3×10×8	LeakyReLU
Flatten	3×10×8	240	-
FC1	240	4	ReLU
Output	4	1	ReLU or Softmax

The generator and discriminator structures are depicted in Table 2 and Table 3, respectively. Each generator consists of four encoders and four decoders. Its first encoder layer, equipped with 4 filters with a 3 kernel size, captures temporal features across all 10 spectral bands. Subsequent encoders use the same kernel size but increase filter counts to 8, 16, and 32, respectively. The decoders, designed to mirror the encoders, reverse the encoding process, thereby reconstructing the original input features.

The discriminator shares a similar design with the generator's encoder, with the distinction that its convolution layers uniformly utilize 8 filters. It then incorporates a flatten layer and a fully connected layer, with dimensions of 24 and 4, respectively. Its final layer is a fully connected layer with a single output, aimed at distinguishing between the transformed target domain data and the source domain data.

To assess the distinctive architecture of the CropSTGAN domain mapper, we conducted comparative experiments with a simpler yet analogous model named CropTGAN. This model features encoders with identical filter counts and sizes as those in CropSTGAN. The primary distinction between the generators and discriminators of the two models lies in the configuration of the first layers. Specifically, the CropTGAN generator and discriminator independently extract features from the temporal dimension on each band in their respective first encoder and convolutional layers. The details of the generators and discriminators are documented in Table 4 and Table 5.

2) LOSSES

As shown in Figure 7, there are two cycle training processes:

- In the first cycle training process (Cycle 1), data from the target domain is first transformed to the source domain and then reverted back to the target domain: $G : \mathbf{X}_t \rightarrow \mathbf{X}'_s \rightarrow \mathbf{X}'_t$ and
- In the second cycle training process (Cycle 2), data from the source domain is first transformed to the target domain and then reverted back to the source domain: $F : \mathbf{X}_s \rightarrow \mathbf{X}'_t \rightarrow \mathbf{X}'_s$

To ensure the creation of effective generators, the losses for both training processes are established as follows:

a: ADVERSARIAL LOSS

This loss function is inspired by GANs and encourages the generator to produce transformed data from one domain that is indistinguishable from real data in the other domain. It is computed by the discriminator network, which aims to classify the generated data as fake while the generator aims to fool the discriminator by generating realistic data (source domain data).

$$\mathcal{L}_{\text{adv1}}(G, D_s, \mathbf{X}_t, \mathbf{X}_s) = \mathbb{E}_{\mathbf{x}_s \sim p_{\text{data}}(\mathbf{x}_s)}[\log D_s(\mathbf{x}_s)] + \mathbb{E}_{\mathbf{x}_t \sim p_{\text{data}}(\mathbf{x}_t)}[1 - \log D_s(G(\mathbf{x}_t))] \quad (1)$$

$$\mathcal{L}_{\text{adv2}}(F, D_t, \mathbf{X}_s, \mathbf{X}_t) = \mathbb{E}_{\mathbf{x}_t \sim p_{\text{data}}(\mathbf{x}_t)}[\log D_t(\mathbf{x}_t)] + \mathbb{E}_{\mathbf{x}_s \sim p_{\text{data}}(\mathbf{x}_s)}[1 - \log D_t(G(\mathbf{x}_s))] \quad (2)$$

b: CYCLE CONSISTENCY LOSS

The cycle consistency loss ensures that the transformed data from one domain to another domain and back to the original domain is consistent with the original data. It measures the difference between the original input data and the data reconstructed after going through both generator mappings. Through the minimization of this loss, the preservation of crucial features during the transformation process is enforced.

$$\mathcal{L}_{\text{cyc1}}(G, F, \mathbf{X}_t) = \mathbb{E}_{\mathbf{x}_t \sim p_{\text{data}}(\mathbf{x}_t)}[||\mathbf{x}_t - F(G(\mathbf{x}_t))||_1] \quad (3)$$

$$\mathcal{L}_{\text{cyc2}}(F, G, \mathbf{X}_s) = \mathbb{E}_{\mathbf{x}_s \sim p_{\text{data}}(\mathbf{x}_s)}[||\mathbf{x}_s - G(F(\mathbf{x}_s))||_1] \quad (4)$$

c: IDENTITY LOSS

The identity loss encourages the generator to preserve the identity of the input data. It computes the difference between the generator output and the input data. The objective of minimizing this loss is to ensure that the generator does not make unnecessary alterations to the data and maintains its essential characteristics.

$$\mathcal{L}_{\text{id1}}(G, \mathbf{X}_t) = \mathbb{E}_{\mathbf{x}_t \sim p_{\text{data}}(\mathbf{x}_t)}[||G(\mathbf{x}_t) - \mathbf{x}_t||_1] \quad (5)$$

$$\mathcal{L}_{\text{id2}}(F, \mathbf{X}_s) = \mathbb{E}_{\mathbf{x}_s \sim p_{\text{data}}(\mathbf{x}_s)}[||F(\mathbf{x}_s) - \mathbf{x}_s||_1] \quad (6)$$

d: TOTAL LOSS

The total loss is a combination of these losses, weighted by respective coefficients:

$$\begin{aligned}\mathcal{L}_{\text{total}} = {} & \mathcal{L}_{\text{adv1}}(G, D_s, \mathbf{X}_t, \mathbf{X}_s) \\ & + \alpha \mathcal{L}_{\text{cyc1}}(G, F, \mathbf{X}_t) \\ & + \beta \mathcal{L}_{\text{id1}}(G, \mathbf{X}_t) \\ & \mathcal{L}_{\text{adv2}}(F, D_t, \mathbf{X}_s, \mathbf{X}_t) \\ & + \gamma \mathcal{L}_{\text{cyc2}}(F, G, \mathbf{X}_s) \\ & + \delta \mathcal{L}_{\text{id2}}(F, \mathbf{X}_s),\end{aligned} \quad (7)$$

where α, β, γ, δ are the weight parameters:

$$\alpha = \frac{\mathcal{L}_{\text{adv1}}}{\mathcal{L}_{\text{cyc1}}}, \beta = \frac{\mathcal{L}_{\text{adv1}}}{\mathcal{L}_{\text{id1}}}, \gamma \quad = \frac{\mathcal{L}_{\text{adv2}}}{\mathcal{L}_{\text{cyc2}}}, \delta = \frac{\mathcal{L}_{\text{adv2}}}{\mathcal{L}_{\text{id2}}} \quad (8)$$

Our aim is to maintain the training gradients of different loss types close to a 1:1:1 ratio. The total loss can be expressed as a minimax function:

$$\min_{G,F} \max_{D_t, D_s} \mathcal{L}_{\text{total}}(G, F, D_t, D_s, \mathbf{X}_t, \mathbf{X}_s). \quad (9)$$

where the generator seeks to minimize the loss while the discriminator aims to maximize it to achieve high-quality time-series MSI data transformation while maintaining consistency and identity preservation.

e: TRAINING STOP CRITERIA BASED ON THE TOTAL LOSS

The total loss is employed as an evaluation metric. After the GropSTGAN network has reached a stable state for the training process after several training epochs, the model is selected associated with the smallest total loss. This particular model is then designated as our well-trained model. This approach allows us to tackle the absence of paired datasets for evaluation in CropSTGAN and obtain an objective measure of the model's performance.

TABLE 6. Cross-year experiment metrics comparison: Targeting Jackson County 2020. The best metrics are indicated in bold, while the second-best metrics of the baseline methods are underlined.

	CropSTGAN			CropTGAN			STDAN			TempCNN		
	OA	F1	Kappa	OA	F1	Kappa	OA	F1	Kappa	OA	F1	Kappa
S1	0.9602	0.9613	0.9205	0.9477	0.9503	0.8952	0.9651	0.9670	0.9299	0.9688	0.9704	0.9376
S2	0.9619	0.9630	0.9239	0.9346	0.9403	0.8681	0.9550	0.9530	0.9122	-	-	-
S3	0.9641	0.9657	0.9281	0.9440	0.9477	0.8874	0.9549	0.9582	0.9093	-	-	-
Avg	<u>0.9621</u>	<u>0.9633</u>	<u>0.9242</u>	0.9421	0.9461	0.8836	0.9583	0.9594	0.9171	**0.9688**	**0.9704**	**0.9376**

E. TempCNN CROP MAPPER

The TempCNN crop mapper denoted as C, structured similarly to the CropSTGAN domain mapper's discriminators D_x and D_y as outlined in Table 3, is trained solely with labelled source domain data ($\mathbf{X}_s$, $\mathbf{Y}_s$). The binary cross-entropy loss:

$$\mathcal{L}_{\text{class}}(C) = -\sum_{s} \left(y_s \log(C(\mathbf{x}_s)) + (1 - y_s) \log(1 - C(\mathbf{x}_s))\right), \tag{10}$$

is used as the training loss for the TempCNN crop mapper. Its purpose is to locate the target crop within the source domain. However, applying this mapper directly in the target domain often leads to subpar results due to discrepancies in the variation of data distribution between the domains. To address this issue, the target domain's MSI data must first be transformed to the source domain using the domain mapper's generator. As illustrated in the right portion of Figure 5, this transformed data is then processed by the TempCNN crop mapper to generate mapping results:

$$\mathbf{Y}_t = C(G(\mathbf{X}_t)) \tag{11}$$

This integrated approach ensures accurate and reliable early crop mapping in the target domain, even in the absence of the target domain ground truth data.

V. EXPERIMENT SETUP AND RESULTS

A. EXPERIMENT SETUP

This work designs two sets of experiments to comprehensively evaluate the cross-domain performance of the proposed CropSTGAN method:

- The first set of experiments investigates the cross-year scenario, where Jackson County from 2020 to 2021 is considered as the target domain, and Jackson County in 2019 serves as the source domain. The discrepancies in data distribution between these domains are relatively slight.
- The second series of experiments explores cross-region scenarios, incorporating a study area in China and Jackson County. When Jackson County 2019 serves as the source domain, the target domain is the study area in China 2019. Conversely, when the study area in China 2019 is the source domain, Jackson County 2019 becomes the target domain. The discrepancies in data distribution across the domains are relatively large.

To evaluate the distinct architecture of the CropSTGAN domain mapper, comparative experiments were undertaken utilizing a simpler, analogous structure named CropTGAN. Within this framework, both the discriminators' and generators' encoders share the same filter numbers and sizes as those in CropSTGAN but focus solely on extracting features from the temporal dimension. Additionally, we benchmarked the CropSTGAN framework against various SOTA methods, including TempCNN and STDAN.

Furthermore, a series of experiments based on CropSTGAN were conducted for the cross-region analysis using MSI data collected at intervals of 50, 60, 70, 80, and 90 days starting from May 1st or the 150th day of the year, in order to determine the impact of different time-series input data lengths on the results of our method. Notably, with each reduction of two data points, one pair of encoder and decoder structures is removed from the domain mapper generator, and one convolutional layer is removed from both the domain mapper discriminator and the crop mapper.

B. TRAINING SETUP

In each experiment, all data points from both the source and target domains are used. The TempCNN is trained using 70% of labelled data from the source domain. The remaining 30% is divided equally into validation and test datasets. To train the CropSTGAN model, all the sampled data from the target domain and the source domain are used. During training, a batch size of 256 is used. All methods' training persisted until the completion of 500 epochs or upon convergence, as determined by an early stopping criterion set at 50 epochs. For optimization, the Adaptive Moment Estimation (Adam) optimizer is employed with a learning rate of 0.005 and an exponential decay rate of 0.9 for the first moment estimates.

Every method was repeatedly trained on each subset from scratch 3 times with the same training configuration. Notably, the TempCNN model was only trained once and served as the crop mapper for the CropSTGAN, resulting in a single TempCNN outcome for each test. Moreover, in the CropSTGAN and CropTGAN framework training processes, it's worth noting that the initial coefficient values of the total loss function in the experiments are set to 10. The initial coefficient values in our total loss function are chosen to ensure the stability and effectiveness of the CropSTGAN and CropTGAN training processes. Setting the initial values to 5 or 20 yields similar final experimental results, but it slows down the convergence process of the training loss function.

C. EVALUATION METRICS

To assess the performance of the binary corn crop mapping, the following metrics are used:

TABLE 7. Cross-year experiment metrics comparison: Targeting jackson county 2021. The best metrics are indicated in bold, while the second-best metrics of the baseline methods are underlined.

	CropSTGAN			CropTGAN			STDAN			TempCNN		
	OA	F1	Kappa	OA	F1	Kappa	OA	F1	Kappa	OA	F1	Kappa
S1	0.9608	0.9616	0.9216	0.9581	0.9573	0.9163	0.9597	0.9604	0.9195	0.9338	0.9387	0.8669
S2	0.9701	0.9701	0.9401	0.9396	0.9396	0.8792	0.9673	0.9685	0.9345	-	-	-
S3	0.9660	0.9663	0.9320	0.9523	0.9534	0.9047	0.9677	0.9688	0.9354	-	-	-
Avg	**0.9656**	**0.9660**	**0.9312**	0.9500	0.9501	0.9001	0.9649	0.9659	0.9298	0.9338	0.9387	0.8669

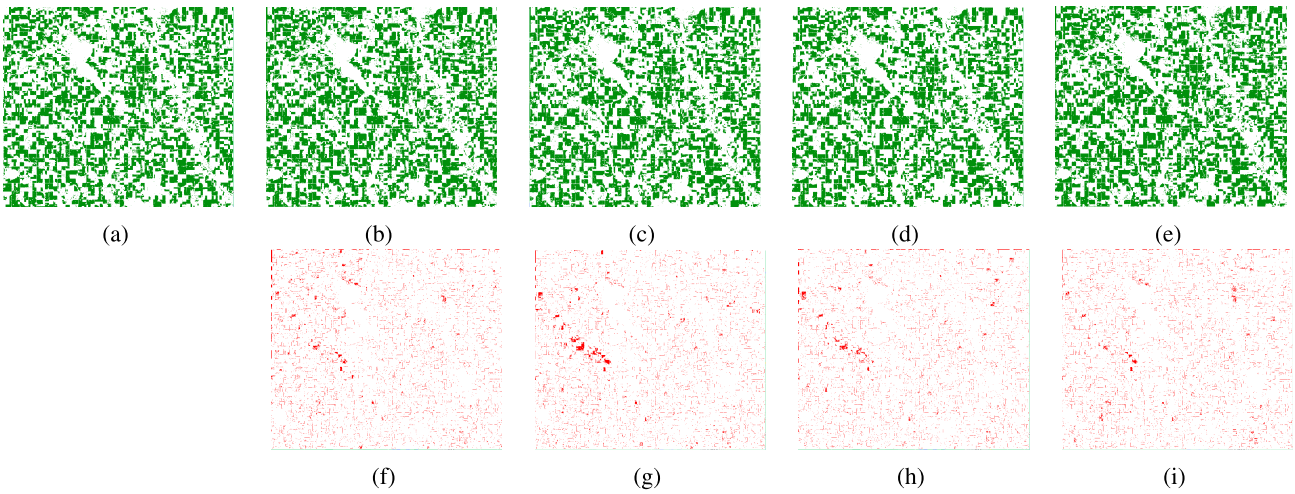


FIGURE 8. Corn Crop Mapping Results Comparison with Jackson County 2020 as the Target Domain. (a) displays the ground truth. The crop mapping results are depicted in (b) for CropSTGAN, (c) for CropTGAN, (d) for STDAN, and (e) for TempCNN. In this visualization, green denotes corn, and white represents other types. The corresponding error images are illustrated in panels (f) and (i) for CropSTGAN, CropTGAN, STDAN, and TempCNN, respectively. Red highlights the misclassified pixels, and white represents correctly classified pixels.

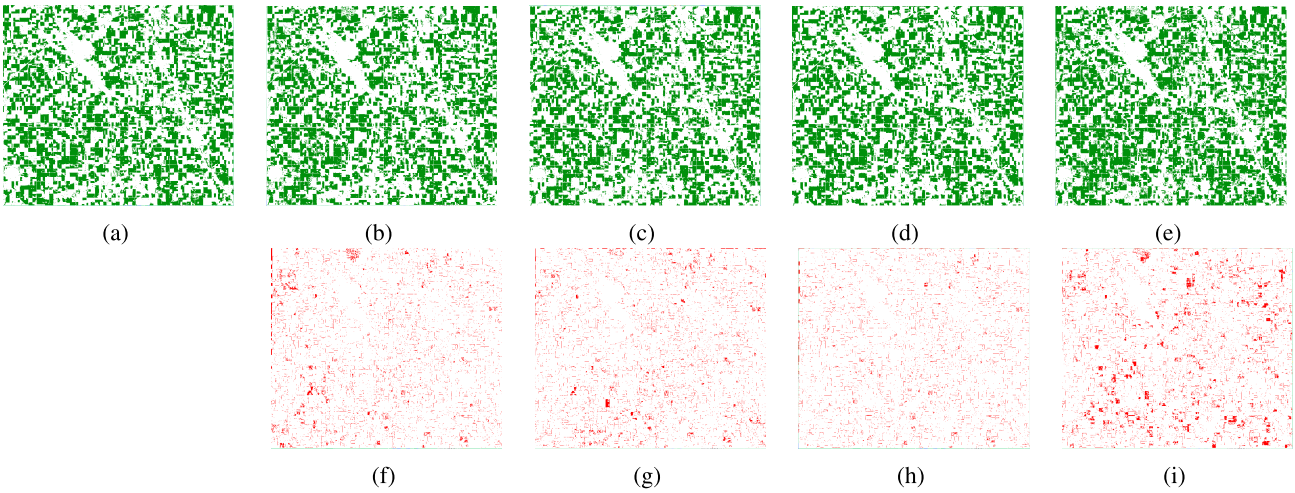


FIGURE 9. Corn crop mapping results comparison with jackson county 2021 as the target domain.

TABLE 8. Cross-region experiment metrics comparison: Targeting china 2019. The best metrics are indicated in bold, while the second-best metrics of the baseline methods are underlined.

	CropSTGAN			CropTGAN			STDAN			TempCNN		
	OA	F1	Kappa	OA	F1	Kappa	OA	F1	Kappa	OA	F1	Kappa
S1	0.8790	0.8901	0.7557	0.8539	0.8744	0.7019	0.8416	0.8613	0.6771	0.8448	0.8657	0.6826
S2	0.8935	0.9033	0.7847	0.8457	0.8675	0.6850	0.8353	0.8576	0.6630	-	-	-
S3	0.8842	0.8968	0.7654	0.8663	0.8755	0.7311	0.8339	0.8583	0.6590	-	-	-
Avg	**0.8856**	**0.8967**	**0.7686**	0.8553	0.8725	0.7060	0.8369	0.8591	0.6664	0.8448	0.8657	0.6826

1) OVERALL ACCURACY (OA)

represents the proportion of all correctly classified items to the total number of items in the dataset.

2) F1 SCORE

is a single metric that combines precision and recall to provide an overall measure of a model's accuracy in classification

TABLE 9. Cross-region experiment metrics comparison: Targeting jackson county 2019. The best metrics are indicated in bold, while the second-best metrics of the baseline methods are underlined.

	CropSTGAN			CropTGAN			STDAN			TempCNN		
	OA	F1	Kappa	OA	F1	Kappa	OA	F1	Kappa	OA	F1	Kappa
S1	0.8790	0.8901	0.7557	0.8409	0.8585	0.6783	0.8156	0.8364	0.6270	0.8465	0.8620	0.6900
S2	0.8743	0.8802	0.7485	0.8428	0.8594	0.6812	0.8383	0.8549	0.6735	-	-	-
S3	0.8687	0.8799	0.7351	0.8361	0.8505	0.6690	0.8431	0.8496	0.6856	-	-	-
Avg	**0.8740**	**0.8834**	**0.7464**	0.8399	0.8561	0.6762	0.8323	0.8470	0.6620	0.8465	0.8620	0.6900

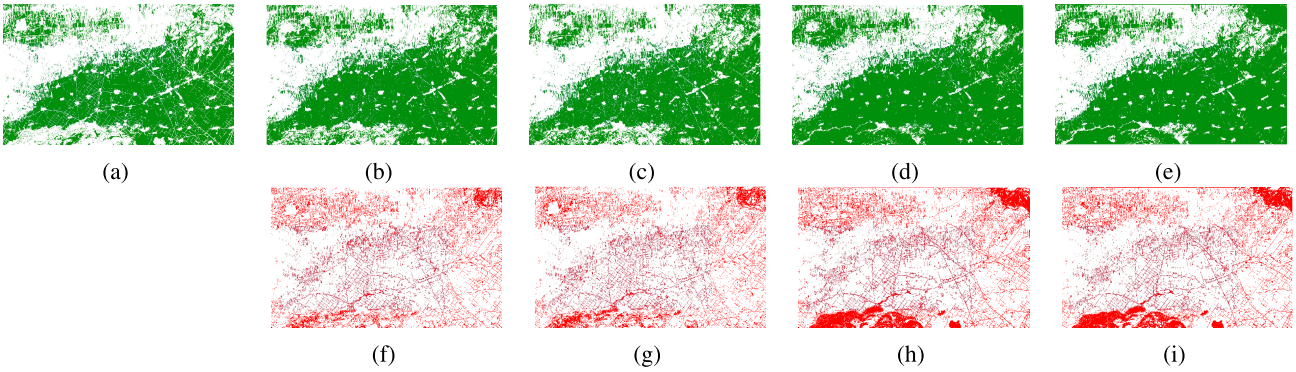


FIGURE 10. Corn crop mapping results comparison with the study area of china 2019 as the target domain.

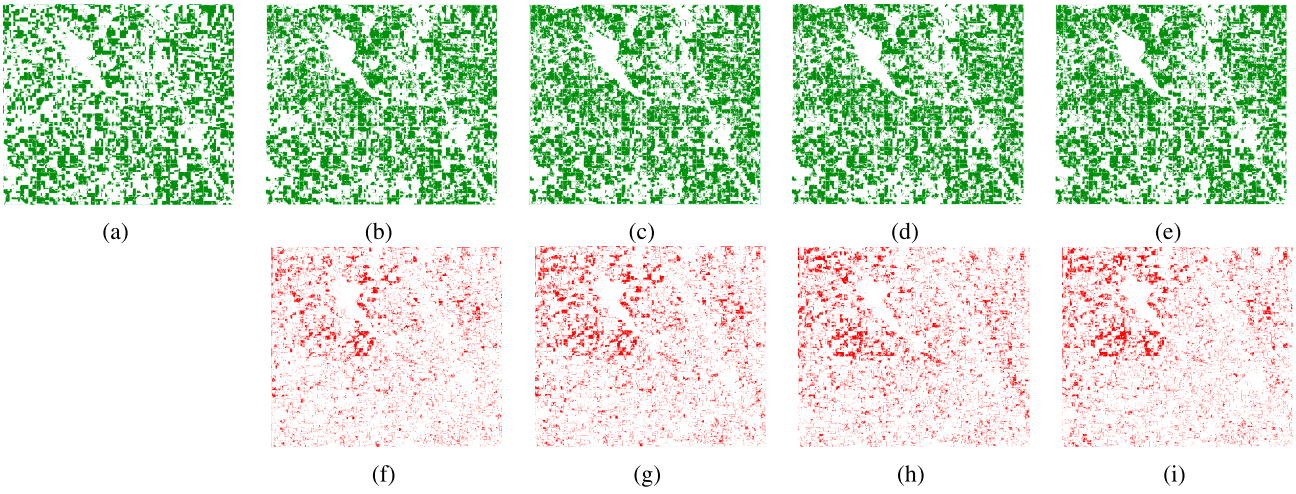


FIGURE 11. Corn crop mapping results comparison with jackson county 2019 as the target domain.

TABLE 10. Early crop mapping results with different time-series length inputs using CropSTGAN. "1" denotes the first series of cross-region experiments, where the study area in China is used as the target domain. "2" denotes the second series of cross-region experiments, where the study area in the USA is used as the target domain.

	190 (day of year)			200			210			220			230		
	OA	F1	Kappa	OA	F1	Kappa	OA	F1	Kappa	OA	F1	Kappa	OA	F1	Kappa
1	0.6823	0.6704	0.3663	0.7710	0.7819	0.5195	0.8830	0.8817	0.7600	0.8739	0.8850	0.7467	0.8856	0.8967	0.7686
2	0.5947	0.6359	0.1815	0.7655	0.7818	0.4268	0.7615	0.7819	0.5195	0.8619	0.8638	0.7243	0.8740	0.8834	0.7464

tasks. It balances the trade-off between correctly identifying positive instances and minimizing false positives and false negatives.

3) THE KAPPA COEFFICIENT

measures the agreement between the predicted and observed classifications, taking into account the agreement that would occur by chance alone.

D. RESULTS

Tables 6 and 7 present the results for the first set of cross-year experiments. Tables 8 and 9 present the results for the second set of cross-region experiments. Across two sets of experiments, our CropSTGAN method demonstrates the highest average metrics, except for the cross-year experiment targeting Jackson County 2020.

In the first set of experiments targeting Jackson County 2020, the CropSTGAN achieved remarkable results, securing

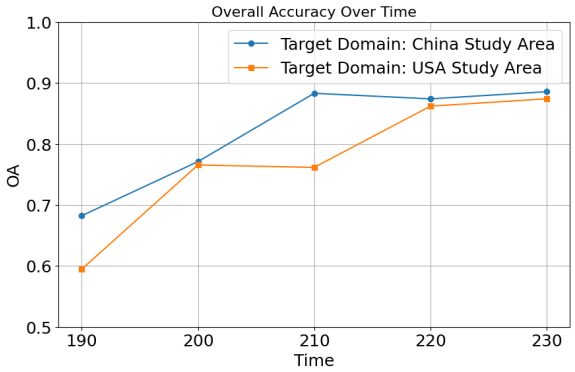


FIGURE 12. The mapping overall accuracy with different time-series length input using CropSTGAN.

second place with an average OA of 96.21%, an F1 score of 96.33%, and a Kappa coefficient of 92.42%. These results represent a slight improvement over STDAN. Additionally, CropSTGAN outperformed CropTGAN, underscoring the effectiveness of its unique structure, with increases of +2.00% in OA, +1.72% in F1, and +4.06% in Kappa. However, it was slightly outpaced by the TempCNN, which took first place with an OA of 96.88%, an F1 score of 97.04%, and a Kappa of 93.76%. When Jackson County 2021 plays as the target domain, CropSTGAN significantly surpasses TempCNN in performance, with an increase of +3.180% in OA, +2.73% in F1 Score, and +6.43% in Kappa Coefficient. CropSTGAN also outperforms STDAN, leading by a slight margin. Meanwhile, CropTGAN ranks third. Figures 8 and 9 display the visualization of cross-year experimental results and their error images using CropSTGAN, CropTGAN, STDAN, and TempCNN. In the second set of experiments, CropSTGAN outperformed all, securing the top spot in both tests. Specifically, when targeting the study area of China, CropSTGAN achieved an OA of 88.56%, an F1 score of 89.67%, and a Kappa coefficient of 76.86%. These metrics are markedly higher than those of TempCNN, showing increases of 4.08% in OA, 3.10% in F1, and 8.60% in Kappa. Conversely, with China as the source domain and Jackson County 2019 as the target domain, CropSTGAN reached an OA of 87.40%, an F1 score of 88.34%, and a Kappa coefficient of 74.64%, surpassing TempCNN by 2.75% in OA, 2.14% in F1, and 5.64% in Kappa. Additionally, CropTGAN outperformed TempCNN in the first cross-region experiment but underperformed in the second. Interestingly, STDAN performed worse than TempCNN in both experiments. Figures 8 and 9 present the results and error images from cross-region experiments.

Furthermore, Table 10 presents the cross-region experiment results of early crop mapping using different lengths of time-series MSI data. As shown in Figure 12, in both experiments, longer time-series lengths yield better result metrics. The results begin to stabilize from a length of 80 days, spanning from the 150th day of the year to the 220th day. The best metrics are observed for the 90-day length, from the 150th day of the year to the 230th day, which is the longest duration tested in these experiments.

E. THE t-SNE VISUALIZATION

To assess the effectiveness of our CropSTGAN method in addressing the domain shift problem, t-distributed stochastic neighbour embedding (t-SNE) [31] is utilized for the visualization to analyze the distribution of the target data, transformed target data, and source domain data of the target crop. By using t-SNE, the data points are projected into a two-dimensional space while preserving their local relationships. Figure 13 and 14 show the t-SNE visualization that illustrates the distribution of corn data points for the cross-year and cross-region experiments.

In the visualization, the orange points represent MSI data points of corn cropland sourced from the target domain, offering a glimpse into the data distribution within that domain. The green points denote the transformed corn cropland MSI data points from the target domain to the source domain, employing our proposed method. Finally, the blue points indicate the original corn cropland MSI data points extracted from the source domain.

Upon analyzing the t-SNE visualization, it is evident that the distribution of corn cropland MSI data points between the source and target domains differs. This disparity highlights the presence of a domain shift, which poses challenges for accurate crop mapping under the target domains. However, the application of our CropSTGAN domain mapper resulted in an improvement in the similarity between the data distribution of the transformed target domain data (green points) and the source domain data (blue points), compared to the similarity between the original target domain data (orange points) and the source domain data (blue points). This resemblance enables the TempCNN crop mapper, trained on the source domain, to effectively process the transformed remote sensing data obtained from the target domain. It coincides with our excellent crop mapping results for these years and this county. For example, in the cross-year experiments, the corn data distribution between the source domain (2019) and target domain (2020) is significantly more similar than that between 2019 and 2021. This explains why the direct method TempCNN performs better for 2020 (OA is 96.88%) than for 2021 (OA is 93.38%). Conversely, the transformed target data provided by our CropSTGAN domain mapper maintains a similar distribution to the source data for both 2020 and 2021. Consequently, the final results are also close to each other, with OAs of 96.21% and 96.56%, respectively. Furthermore, the target data distribution in the cross-region experiments between China and the USA is more disparate than in the cross-year experiments, leading to significantly worse results when using the direct method. Although the transformed target data distribution is closer to the source data, it is not as close as in the cross-year experiments. Consequently, CropSTGAN achieves better

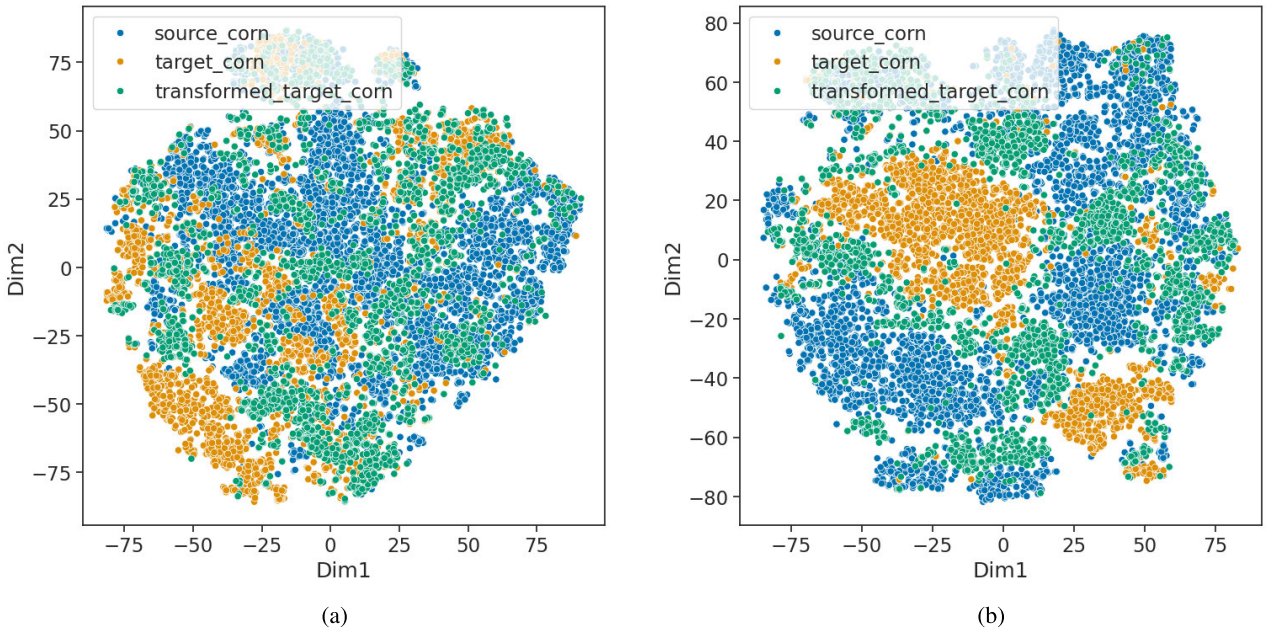


FIGURE 13. The t-SNE visualization of corn data points for the cross-year experiments: Comparison between target domain data, transformed target domain data, and source domain data. (a) Jackson county 2020 as the target domain. (b) jackson county 2021 as the target domain.

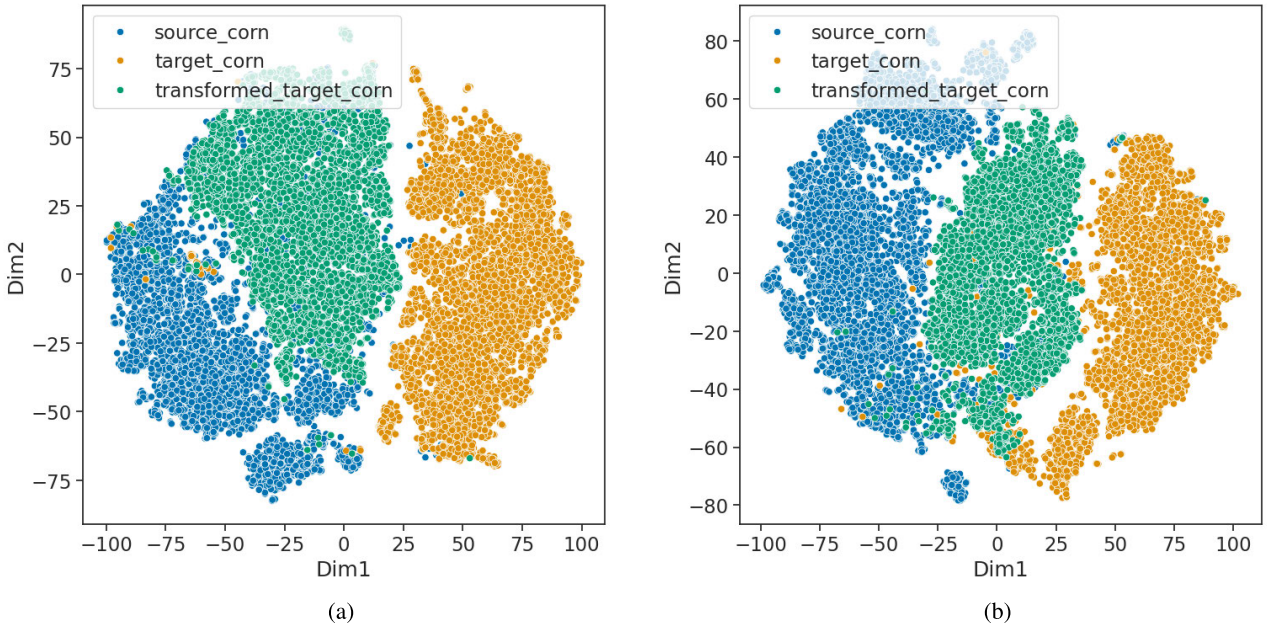


FIGURE 14. The t-SNE visualization of corn data points for the cross-domain experiments: Comparison between target domain data, transformed target domain data, and source domain data. (a) The study area of china as the target domain. (b) Jackson county 2019 as the target domain.

results than the direct method, but not as good as the results in the cross-year experiments.

VI. DISCUSSION

A. ANALYSIS OF CROP MAPPING RESULTS

In the cross-year experiments, CropSTGAN and STGAN showed similar performance, indicating that CropSTGAN effectively addresses the inter-annual cross-domain challenge, comparable to the state-of-the-art (SOTA) method. In cross-region experiments, CropSTGAN achieved better results than STGAN. As evident from Figure 13 and 14, the MSI data distribution discrepancy is more pronounced in cross-region experiments than in cross-year ones. Under the significant differences in data distribution, CropSTGAN outperforms STGAN, benefiting from the identity loss.

Moreover, CropSTGAN outperformed CropTGAN in all experiments, highlighting the effectiveness of the CropSTGAN domain mapper structure. Additionally, when there is a

significant difference in data distribution between the source and target domains, the results from both CropSTGAN and CropTGAN were significantly better than those obtained by directly applying a CNN-based crop classifier, trained on the source domain, to the target domain.

B. ADVANTAGES OF CROPSTGAN

Most existing crop mapping related studies rely heavily on a large number of local labelled data for modeling and making predictions, and thus tedious and costly sample collection needs to be carried out extensively and frequently. The sharing of collected labelled samples is an effective way to address the dilemma of ground truth sampling. However, due to the differences in climate conditions across regions and years, trained crop classification models may lose their validity when applied to new domains. Therefore, the CropSTGAN was developed to address the distribution discrepancy existed between the source and target domains, that is, cross-domain issue.

In order to address this issue, most SOTA methods, like STDAN, strive to extract invariant features across target and source domains to tackle cross-domain challenges. However, their effectiveness is often limited by significant differences in data distribution between these domains. In contrast, our CropSTGAN demonstrates superior performance in scenarios with large data distribution disparities, as evidenced by our experimental results. For instance, in the cross-year experiments, the corn data distribution between the source and target domains is markedly more congruent compared to the cross-region experiments, as illustrated in Figures 1 and 2. This alignment elucidates why the direct method, TempCNN, performs substantially better in cross-year experiments (96.88% OA for 2020, 93.38% for 2021) compared to cross-region experiments, where the target domain is the China study area with 84.48% OA, and the USA study area with 84.65% OA. In cross-year experiments, CropSTGAN and STDAN yield comparable metrics. However, in cross-region experiments, CropSTGAN achieves significantly higher metrics, with 88.56% and 87.40% OA, compared to STDAN, which attains 83.69% and 83.23% OA. Notably, STDAN performs worse than the direct method TempCNN in these scenarios, which has 84.48% and 84.65% OA.

C. LIMITATIONS OF CROPSTGAN

However, it is important to acknowledge the limitations of our work. One limitation of our CropSTGAN work is the underlying assumption that the primary crops in the target domain and the source domain are consistent. This assumption, which implies uniformity in the dominant crop types, may not always hold true in practice, particularly when considering diverse agricultural practices across different regions. In future research, addressing this limitation and developing methods that can accommodate variations in primary crop types across domains will be a valuable direction for enhancing the robustness and applicability of our approach.

VII. CONCLUSION

In conclusion, we introduced the CropSTGAN framework, which integrates a pre-processor, a domain mapper, and a TempCNN crop mapper, specifically designed to overcome the challenges of cross-domain early crop mapping caused by inter-region and inter-year variations. Notably, the framework is versatile, enabling not only crop mapping but also the classification of various land cover types. The CropSTGAN domain mapper is designed to extract both temporal and spectral features from time-series MSI data, effectively transforming target domain data to the source domain. The TempCNN crop mapper, trained by the labelled source domain data, takes the transformed target domain data as input to locate the target crop for the target domain.

Our comprehensive evaluation, conducted across various regions in the USA and China and spanning different years, demonstrates the CropSTGAN framework's superior performance. It outperforms several baseline and SOTA methods, including TempCNN and STDAN, thereby validating its effectiveness and accuracy in cross-domain early crop mapping scenarios, even with large data distribution disparities between the target domain and source domain.

VIII. FUTURE WORKS

In this work, our experiment focused solely on the classification of corn. In the future, we will extend our method to multiple crop mapping by replacing the binary crop mapper with a multi-crop mapper. Additionally, the limited revisit time of MSI images constrains the performance of the crop mapping classifier. For large-scale regions, there may be missing image pairs between the source domain and target domain, further restricting the applicability of our method. To address these challenges, we plan to integrate time-series MSI data from multiple datasets to fill gaps and enhance revisit times. This approach will enable us to obtain more stable temporal images and better evaluate our method.

REFERENCES

[1] F. Waldner, S. Fritz, A. Di Gregorio, and P. Defourny, "Mapping priorities to focus cropland mapping activities: Fitness assessment of existing global, regional and national cropland maps," *Remote Sens.*, vol. 7, no. 6, pp. 7959–7986, Jun. 2015.

[2] C. Singha and K. C. Swain, "Land suitability evaluation criteria for agricultural crop selection: A review," *Agricult. Rev.*, vol. 37, no. 2, pp. 125–132, Jun. 2016.

[3] J. Xue and B. Su, "Significant remote sensing vegetation indices: A review of developments and applications," *J. Sensors*, vol. 2017, pp. 1–17, Aug. 2017.

[4] A. Joshi, B. Pradhan, S. Gite, and S. Chakraborty, "Remote-sensing data and deep-learning techniques in crop mapping and yield prediction: A systematic review," *Remote Sens.*, vol. 15, no. 8, p. 2014, Apr. 2023.

[5] L. Zhang, Z. Liu, D. Liu, Q. Xiong, N. Yang, T. Ren, C. Zhang, X. Zhang, and S. Li, "Crop mapping based on historical samples and new training samples generation in Heilongjiang province, China," *Sustainability*, vol. 11, no. 18, p. 5052, Sep. 2019.

[6] N. You, J. Dong, J. Huang, G. Du, G. Zhang, Y. He, T. Yang, Y. Di, and X. Xiao, "The 10-m crop type maps in northeast China during 2017–2019," *Sci. Data*, vol. 8, no. 1, p. 41, Feb. 2021.

[7] C. Boryan, Z. Yang, R. Mueller, and M. Craig, "Monitoring U.S. agriculture: The U.S. department of agriculture, national agricultural statistics service, cropland data layer program," *Geocarto Int.*, vol. 26, no. 5, pp. 341–358, Aug. 2011.
[8] C. Pelletier, G. Webb, and F. Petitjean, "Temporal convolutional neural network for the classification of satellite image time series," *Remote Sens.*, vol. 11, no. 5, p. 523, Mar. 2019.
[9] Y. Wang, Z. Zhang, L. Feng, Y. Ma, and Q. Du, "A new attention-based CNN approach for crop mapping using time series Sentinel-2 images," *Comput. Electron. Agricult.*, vol. 184, May 2021, Art. no. 106090.
[10] M. Hamidi, A. Safari, and S. Homayouni, "An auto-encoder based classifier for crop mapping from multitemporal multispectral imagery," *Int. J. Remote Sens.*, vol. 42, no. 3, pp. 986–1016, Feb. 2021.
[11] H. Crisóstomo de Castro Filho, O. Abílio de Carvalho, O. L. Ferreira de Carvalho, P. Pozzobon de Bem, R. dos Santos de Moura, A. Olino de Albuquerque, C. Rosa Silva, P. H. Guimarães Ferreira, R. F. Guimarães, and R. A. T. Gomes, "Rice crop detection using LSTM, bi-LSTM, and machine learning models from Sentinel-1 time series," *Remote Sens.*, vol. 12, no. 16, p. 2655, Aug. 2020.
[12] P. Hao, L. Di, C. Zhang, and L. Guo, "Transfer learning for crop classification with cropland data layer data (CDL) as training samples," *Sci. Total Environ.*, vol. 733, Sep. 2020, Art. no. 138869.
[13] S. Ge, J. Zhang, Y. Pan, Z. Yang, and S. Zhu, "Transferable deep learning model based on the phenological matching principle for mapping crop extent," *Int. J. Appl. Earth Observ. Geoinformation*, vol. 102, Oct. 2021, Art. no. 102451.
[14] V. S. Konduri, J. Kumar, W. W. Hargrove, F. M. Hoffman, and A. R. Ganguly, "Mapping crops within the growing season across the United States," *Remote Sens. Environ.*, vol. 251, Dec. 2020, Art. no. 112048.
[15] L. Zhong, P. Gong, and G. S. Biging, "Efficient corn and soybean mapping with temporal extendability: A multi-year experiment using Landsat imagery," *Remote Sens. Environ.*, vol. 140, pp. 1–13, Jan. 2014.
[16] H. Ajakan, P. Germain, H. Larochelle, F. Laviolette, and M. Marchand, "Domain-adversarial neural networks," 2014, *arXiv:1412.4446*.
[17] J.-Y. Zhu, T. Park, P. Isola, and A. A. Efros, "Unpaired image-to-image translation using cycle-consistent adversarial networks," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Oct. 2017, pp. 2242–2251.
[18] B. Zheng, S. W. Myint, P. S. Thenkabail, and R. M. Aggarwal, "A support vector machine to identify irrigated crop types using time-series Landsat NDVI data," *Int. J. Appl. Earth Observ. Geoinformation*, vol. 34, pp. 103–112, Feb. 2015.
[19] P. Hao, Y. Zhan, L. Wang, Z. Niu, and M. Shakir, "Feature selection of time series MODIS data for early crop classification using random forest: A case study in kansas, USA," *Remote Sens.*, vol. 7, no. 5, pp. 5347–5369, Apr. 2015.
[20] R. Saini and S. K. Ghosh, "Crop classification on single date Sentinel-2 imagery using random forest and suppor vector machine," *Int. Arch. Photogramm., Remote Sens. Spatial Inf. Sci.*, vol. 5, pp. 683–688, Nov. 2018.
[21] Z. Sun, L. Di, and H. Fang, "Using long short-term memory recurrent neural network in land cover classification on Landsat and cropland data layer time series," *Int. J. Remote Sens.*, vol. 40, no. 2, pp. 593–614, Jan. 2019.
[22] A. Nowakowski, J. Mrziglod, D. Spiller, R. Bonifacio, I. Ferrari, P. P. Mathieu, M. Garcia-Herranz, and D.-H. Kim, "Crop type mapping by using transfer learning," *Int. J. Appl. Earth Observ. Geoinf.*, vol. 98, Jun. 2021, Art. no. 102313.
[23] X.-Y. Tong, G.-S. Xia, Q. Lu, H. Shen, S. Li, S. You, and L. Zhang, "Land-cover classification with high-resolution remote sensing images using transferable deep models," *Remote Sens. Environ.*, vol. 237, Feb. 2020, Art. no. 111322.
[24] Y. Hamrouni, E. Paillassa, V. Chéret, C. Monteil, and D. Sheeren, "From local to global: A transfer learning-based approach for mapping poplar plantations at national scale using Sentinel-2," *ISPRS J. Photogramm. Remote Sens.*, vol. 171, pp. 76–100, Jan. 2021.
[25] G.-H. Kwak and N.-W. Park, "Unsupervised domain adaptation with adversarial self-training for crop classification using remote sensing images," *Remote Sens.*, vol. 14, no. 18, p. 4639, Sep. 2022.
[26] Y. Wang, L. Feng, Z. Zhang, and F. Tian, "An unsupervised domain adaptation deep learning method for spatial and temporal transferable crop type mapping using Sentinel-2 imagery," *ISPRS J. Photogramm. Remote Sens.*, vol. 199, pp. 102–117, May 2023.
[27] Z. Wang, H. Zhang, W. He, and L. Zhang, "Phenology alignment network: A novel framework for cross-regional time series crop classification," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. Workshops*, Jun. 2021, pp. 2934–2943.
[28] H. Hersbach et al., "The ERA5 global reanalysis," *Quart. J. Roy. Meteorological Soc.*, vol. 146, no. 730, pp. 1999–2049, 2020.
[29] L. Blickensdörfer, M. Schwieder, D. Pflugmacher, C. Nendel, S. Erasmi, and P. Hostert, "Mapping of crop types and crop sequences with combined time series of Sentinel-1, Sentinel-2 and Landsat 8 data for Germany," *Remote Sens. Environ.*, vol. 269, Feb. 2022, Art. no. 112831.
[30] C. F. Brown, S. P. Brumby, B. Guzder-Williams, T. Birch, S. B. Hyde, J. Mazzariello, W. Czerwinski, V. J. Pasquarella, R. Haertel, S. Ilyushchenko, K. Schwehr, M. Weisse, F. Stolle, C. Hanson, O. Guinan, R. Moore, and A. M. Tait, "Dynamic world, near real-time global 10 m land use land cover mapping," *Sci. Data*, vol. 9, no. 1, p. 251, Jun. 2022.
[31] L. Van der Maaten and G. Hinton, "Visualizing data using t-SNE," *J. Mach. Learn. Res.*, vol. 9, no. 11, pp. 1–20, 2008.

YIQUN WANG received the M.Sc. degree from Karlsruhe Institute of Technology, Karlsruhe, Germany, in 2020. He is currently pursuing the Ph.D. degree with the Interdisciplinary Centre for Security, Reliability and Trust, University of Luxembourg, Luxembourg City, Luxembourg. His research interests include remote sensing, computer vision, and deep learning.

HUI HUANG received the M.Sc. degree in computing science from the University of Glasgow, Glasgow, U.K., in 2013, and the Ph.D. degree from the University of New South Wales, Sydney, NSW, Australia, in 2018. He is currently a Research Associate with the Interdisciplinary Centre for Security, Reliability and Trust, University of Luxembourg, Luxembourg City, Luxembourg. His research interests include V2X communications, autonomous driving, and intelligent transportation systems.

RADU STATE received the M.Sc. degree from Johns Hopkins University, Baltimore, MD, USA, and the Ph.D. and HDR degrees from the University of Lorraine, Nancy, France. He was a Professor at the University of Lorraine, and a Senior Researcher at INRIA, Nancy. He is currently a Professor with the Interdisciplinary Centre for Security, Reliability and Trust, Luxembourg. He has authored more than 100 papers. His research interest includes network and system security and management.