UNIVERSITÉ DU LUXEMBOURG

PhD-FSTM-2024-104
Faculty of Science, Technology and Medicine

Dissertation

Defence held on 9 December 2024 in Luxembourg

to obtain the degree of

Docteur de l'Université du Luxembourg en Informatique

by

Yiqun WANG
Born on 15th December 1994 in Shandong, China

Cross Domain Early Crop Mapping based on Time-series Remote Sensing Data

Dissertation Defense Committee

Dr. STATE, Radu, Dissertation Supervisor
Professor, University of Luxembourg, Luxembourg

Dr. FRANK, Raphaël, Chairman
Professor, University of Luxembourg, Luxembourg

Dr. LAHMADI, Abdelkader, Vice-Chairman
Professor, LORIA France, France

Dr. WAGNER, Marc, Member
Data Analytics and AI Lead, Kerry Group, Luxembourg

Dr. LAGRAA, Sofiane, Member
Team Lead Security Innovation, Fujitsu Luxembourg, Luxembourg

Abstract

This dissertation presents a comprehensive study of advanced remote sensing methodologies for early crop mapping, employing innovative machine learning techniques to address significant challenges in agricultural monitoring. The research encapsulates three main studies: ECMDCM (Early Crop Mapping using Dynamic Clustering Method), CropSTGAN (Crop Spectral-temporal Generative Adversarial Neural Network), and MultiCropGAN (Multiple Crop Mapping Generative Adversarial Neural Network), each contributing uniquely to the field of precision agriculture.

The ECMDCM introduces a novel dynamic clustering approach using time-series NDVI and EVI data to enhance the accuracy of early crop mapping across the continental United States. By optimizing ecoregion delineations through the elbow and silhouette methods and employing Kmeans++ for clustering, this method demonstrates significant improvements over traditional static clustering techniques, offering a more dynamic and precise mapping of crop types.

The CropSTGAN framework addresses the challenges of cross-domain variability in remote sensing-based crop mapping. It incorporates a domain mapper that effectively aligns temporal and spectral features across different geographic and temporal scales, facilitating robust model performance even in the presence of significant data distribution discrepancies. This framework has been validated across diverse regions and years, showcasing superior accuracy and adaptability in comparison to conventional approaches.

Lastly, the MultiCropGAN framework is developed to tackle domain shift and label space discrepancies, which are prevalent in global agricultural settings. By incorporating identity losses into the generator's loss function, MultiCropGAN ensures the preservation of essential characteristics in the data, enhancing the authenticity and accuracy of crop type classification. Extensive testing across various North American regions highlights its effectiveness, particularly in handling divergent label spaces, thereby improving the reliability and applicability of crop mapping techniques.

Together, these studies not only demonstrate the potential of generative adversarial networks and dynamic clustering in remote sensing but also pave the way for future innovations in agricultural monitoring. This thesis aims to contribute to the enhancement of global food security strategies through improved crop monitoring and management, underlining the critical role of advanced remote sensing technologies in the future of agriculture.

Acknowledgements

I would like to express my deepest gratitude to SEDAN group, Kerry Group Luxembourg, and my advisors, Prof. Dr. STATE, Radu, Prof. Dr. FRANK, Raphaël, Dr. HUANG, Hui, and Dr. Marc, Wagner, for their unwavering support, guidance, and mentorship throughout the duration of this research project. Their expertise, encouragement, and invaluable insights have been instrumental in shaping the course of my academic journey.

I would also like to extend my heartfelt appreciation to my family for their endless love, encouragement, and sacrifices. Their unwavering belief in my abilities has been a constant source of motivation, and I am profoundly grateful for their unyielding support.

Additionally, I want to thank all my colleagues, friends, and fellow researchers who provided valuable assistance, feedback, and inspiration during this research endeavor. Your contributions have enriched the quality of this work and have been immensely valuable.

Contents

List of Figures

List of Tables

1 Introduction

In this chapter, we begin by highlighting the significance of early crop mapping. We then discuss the limitations and challenges faced by researchers in addressing early crop mapping over extensive regions, particularly when dealing with issues related to missing labels. Finally, we outline the contributions of this dissertation and provide a roadmap for the subsequent sections.

Contents

1.1 Motivation

Early crop mapping plays a crucial role in improving agricultural productivity, optimizing resource allocation, and enhancing decision-making efficiency. This process is especially critical as it provides agricultural producers with predictive yield data, helping decision-makers to adjust strategies timely within the crop growth cycle to respond to the ever-changing market and environmental conditions.

Firstly, early mapping of crop planting areas can significantly refine the management of crops. By analyzing data collected from multi-temporal satellite images, agricultural producers can monitor crop growth conditions, promptly identify areas affected by pests and diseases, and take swift actions to minimize losses. Additionally, this process can optimize irrigation and fertilization plans by more precisely allocating resources to increase crop yield and quality.

Secondly, early crop mapping is particularly important for countries and regions facing difficulties in data acquisition. In these areas, traditional methods of agricultural data collection are often costly and inefficient. The application of satellite remote sensing technology provides an effective solution, allowing these regions to acquire real-time, high-precision agricultural data at lower costs, thus compensating for the lack of ground data.

Furthermore, as the impact of global climate change on agriculture increases, early crop mapping helps agricultural producers better adapt to the challenges brought by climate change. By monitoring the growth status of crops and changes in environmental conditions, producers can foresee potential risks, such as droughts or floods, earlier and adjust their crop planting strategies and harvest timings accordingly to mitigate the adverse effects of climate change.

Although remote sensing technologies have many advantages in crop mapping, enhancing the accuracy and applicability of these technologies remains a challenge. The interpretation and utilization of remote sensing data require complex algorithms and substantial computing resources, and without ground truth data, verifying and calibrating the accuracy of these data becomes even more difficult. Therefore, developing new machine learning models and algorithms that can improve the accuracy and reliability of crop mapping under various geographic and environmental conditions is a significant research direction in the remote sensing field.

Through continuous technological innovation and methodological improvements, early crop mapping technologies can better serve global agricultural production, help countries enhance food security levels, and promote sustainable development in global agriculture. This holds significant practical significance not only for agricultural producers and decision-makers but also plays a strategic role in promoting stability and development in the global food system.

1.2 Research Challenges and Questions

1.2.1 Early Crop Mapping in a Large Region

Early crop mapping on a large scale involves leveraging advanced remote sensing technologies, which have grown in popularity due to their non-invasiveness and ability to quickly gather extensive data. Utilizing multispectral imagery (MSI) captured by satellites such as Landsat-8 [2, 3, 4], Sentinel-2 [4, 5], and MODIS [3, 6, 7] has been transformative in providing insights into crop health, vegetation indices, and land cover classification across vast areas. These technologies enable the assessment of crop vigor, the identification of stress factors, and the mapping of crop types, thereby facilitating more informed agricultural decision-making.

The Normalized Difference Vegetation Index (NDVI) [8, 6, 9, 10, 11, 12, 7] and the Enhanced Vegetation Index (EVI) [8, 10] are particularly crucial in this context. NDVI helps quantify the greenness and density of vegetation, reflecting the health of crops by measuring the difference between near-infrared and red spectral bands. EVI further refines these measurements by incorporating additional spectral bands to reduce atmospheric and canopy background noise, thus enhancing the accuracy of vegetation dynamics characterization. These indices, derived from time-series data from Landsat-8 and Sentinel-2, support various classification methods, such as decision trees, for crop identification, proving to be highly effective in multiple studies.

However, the task of accurately mapping crops at a large scale, such as across the entire Continental United States (CONUS), presents significant challenges. The heterogeneous nature of crop-growing environments complicates this process. This heterogeneity stems from complex interactions among various factors like soil type, climate, and topography, each significantly influencing crop mapping success. For instance, different soil types across regions can affect vegetation growth and the spectral reflectance captured by remote sensing data. Similarly, climate variables such as temperature, precipitation, and solar radiation, as well as topographic features like elevation and slope, vary considerably across large regions, impacting crop phenology and productivity.

To address these complexities, the concept of ecoregions has been utilized. Ecoregions are defined as geographically distinct areas that exhibit unique ecological patterns and processes, providing a valuable framework for contextualizing crop mapping on a large scale. This approach involves analyzing environmental factors such as soil type, climate, and topography to divide a large area into multiple ecoregions, each characterized by unique crop growth dynamics. Crop mapping efforts can then be tailored to the specific conditions of each ecoregion, thereby improving the accuracy of the mappings. Methods like ecoregion clustering have shown to enhance mapping accuracy compared to broader national or state-level mappings by considering the specific ecological characteristics of each region.

The application of such methodologies not only enhances the accuracy of crop mapping across large regions but also addresses the challenge of inter-year variability. Since environmental conditions can vary significantly from year to year, static ecoregion clustering methods may not suffice. Adapting crop mapping strategies to accommodate year-to-year climatic variations within each ecoregion is crucial for maintaining the precision and reliability of crop mapping efforts over time.

1.2.2 Early Crop Mapping in the Region or Year without Ground Truth Datasets

Early crop mapping in regions or years without ground truth datasets presents a significant challenge due to the dependency of deep learning (DL) models on extensive labeled data for training and validation. These models utilize multi-spectral images captured by satellites to analyze the spectral features of vegetation, which are influenced by the structure, leaf biochemistry, and phenological stages of the crops. The accuracy of such models hinges on the availability of high-quality ground truth data, which serves as a benchmark for training and assessing the performance of the classification algorithms.

Unfortunately, in many regions, particularly in underdeveloped countries, obtaining this kind of ground truth data is challenging. It is often costly, labor-intensive, and sometimes entirely unfeasible. Public crop-type ground truth datasets, such as the United States Department of Agriculture's (USDA) Cropland Data Layer (CDL) [13], provide reliable references but are restricted to a few countries and are typically released post-harvest, limiting their utility for early crop mapping needs.

The scarcity of ground truth data often forces researchers to adopt a "direct transfer strategy," where a model trained on data from one region is applied to another without appropriate retraining. This strategy, however, struggles with the variability in spectral features across different regions and years, driven by changes in soil composition, climate conditions, and crop progress. Such variability leads to a distribution shift between the training data (source domain) and the application data (target domain), which often results in poor model performance on new and unseen data, challenging the assumption that data across domains are independent and identically distributed (i.i.d).

To tackle these cross-domain issues, including those that span both regions and years, several adaptive strategies have been developed. These strategies aim to enhance the model's robustness to changes in the data landscape without relying on concurrent ground truth data. Techniques such as Domain Adversarial Neural Networks (DANN) and other related methods like Self-Training with Domain Adversarial Network (STDAN) [14], Phenology Alignment Network (PAN) [15], and Deep Adaptation Crop Classification Network (DACCN) [16] have been explored. These methods focus on extracting invariant features from the data across both the source and target domains, leveraging these features to improve the accuracy of crop classification maps.

However, these advanced methodologies assume that the differences between data distributions in various domains are minimal. While they are effective in mitigating the adverse effects of domain shifts and the absence of labels, their success is often limited by substantial discrepancies in data distribution across the domains. This highlights the need for further research into methods that can handle significant variability in agricultural data and improve the generalizability of crop mapping models across different geographical and temporal contexts.

By addressing these challenges, the field can move towards more accurate and timely crop mapping in regions lacking ground truth data, significantly enhancing global agricultural monitoring and management capabilities.

1.3 Contributions

In this chapter, we provide a summary of the contributions of this dissertation as follows:

- **ECMDCM [17]:** Early Crop Mapping using Dynamic Clustering Method. We introduce a novel approach to mapping target crops, specifically soybean and corn, earlier than the harvest period across the United States. Utilizing time-series NDVI and EVI data, we implement a dynamic ecoregion clustering method. This method is validated through the use of both the elbow method and the silhouette method to determine the optimal number of ecoregions. Following this, we employ the Kmeans++ method to train our ecoregion clustering model. This model enables us to generate ecoregion maps for the years 2013 to 2022, comprehensively covering the entire cropland region within the Continental United States (CONUS). Our results demonstrate that the dynamic clustering method achieves significantly higher mapping accuracy compared to traditional static clustering methods, showcasing its effectiveness in enhancing the precision of early crop mapping in large-scale agricultural settings.
- **CropSTGAN [18]:** Crop Spectral-temporal Generative Adversarial Neural Network. We propose the CropSTGAN framework to address the challenges of cross-domain variability in remote sensing-based early crop mapping due to inter-region and inter-year variations. The framework comprises three key components: a pre-processor, a CropSTGAN domain mapper, and a TempCNN crop mapper. The CropSTGAN domain mapper is specifically designed to capture the temporal and spectral features from the time-series multispectral imagery (MSI) data. It learns to transform data from the target domain to closely resemble the source domain, facilitating the application of models across different agricultural contexts. We conducted extensive experiments to evaluate the effectiveness of the CropSTGAN framework across different regions and years, focusing on study areas in the USA and China. These experiments compared the CropSTGAN's performance against several baseline methods, including CropTGAN, TempCNN [19], and STDAN. The results clearly demonstrate that the CropSTGAN framework outperforms these methods, confirming its superior performance and accuracy in tackling cross-domain challenges in early crop mapping. This validation underscores the CropSTGAN's potential to enhance the precision and applicability of remote sensing technologies in agricultural monitoring on a global scale.
- **MultiCropGAN [20]:** Multiple Crop Mapping Generative Adversarial Neural Network. We introduce the MultiCropGAN framework designed to address the domain shift issue with label space discrepancies in early crop mapping tasks. MultiCropGAN is comprised of three main components: a generator, a discriminator, and a classifier. This comprehensive architecture is aimed at enhancing the adaptability of crop mapping models to various agricultural contexts, especially where label inconsistencies exist between different geographical domains. To improve the integrity of the mapping data, we propose incorporating identity losses into the generator's loss function. This modification ensures that the generator refrains from making unnecessary alterations to the data, thereby preserving its essential characteristics. By doing so, the generator helps

maintain the authenticity and relevance of the geographical features inherent in the original data, which is critical for accurate crop type classification. To evaluate the effectiveness of the MultiCropGAN, we conducted experiments across diverse agricultural regions in the USA and Canada. These experiments involved a comparative analysis where MultiCropGAN was pitted against various established methods, including CropGAN, STDAN, DACCN, TempCNN, and RF [21, 17]. The results from these studies demonstrate that MultiCropGAN outperforms these baseline methods, showing superior classification metrics, particularly in scenarios involving divergent label spaces in the target and source domains. This outcome underscores the potential of MultiCropGAN to significantly enhance the accuracy and reliability of early crop mapping across different regions and conditions.

1.4 Roadmap

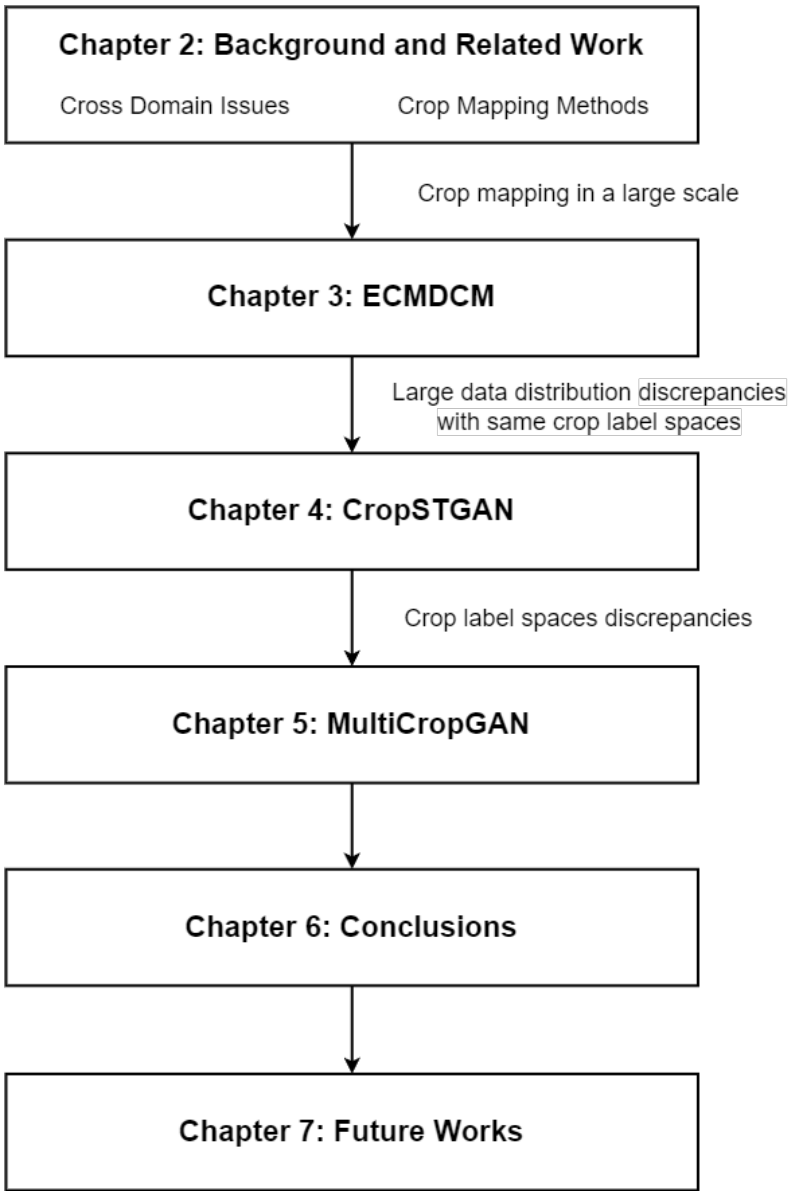


Figure 1.1: The Roadmap of this dissertation.

The roadmap of this dissertation is illustrated in Figure 1.1. Chapter 2 introduces the background of early crop mapping, and related work, followed by a summary of literature addressing crop mapping faced the large-scale regions and crop mapping with cross-domain issues. Chapter 3 presents a new method for early crop mapping for the entire conterminous USA (CONUS) land area using the Normalized Difference Vegetation Index (NDVI) and Enhanced Vegetation Index (EVI) data with a dynamic ecoregion clustering approach. This approach enables the consideration of the year-to-year climate variations that significantly impact crop growth, enhancing the accuracy of our early crop mapping process for a large-scale region. However, obtaining appropriate ground truth for an arbitrary region is challenging. Chapter 4 introduces the CropSTGAN, a novel solution for cross-domain challenges with large data distribution discrepancies, that doesn't require target domain labels. However, this methodology works on the assumption that the crop-type label spaces between

the target and source domains are similar. Building upon this study, Chapter 5 introduces the MultiCropGAN to address this label space discrepancies issue. Following, Chapter 6 concludes the dissertation, and Chapter 7 discusses future work.

2 Background and Related Works

This chapter introduces the background of early crop mapping, and related work, followed by a summary of literature addressing crop mapping faced the large-scale regions and crop mapping with cross-domain issues.

Contents

2.1 Multispectral Image (MSI), Vegetation Indices, and Satellite MSI Datasets

2.1.1 MSI

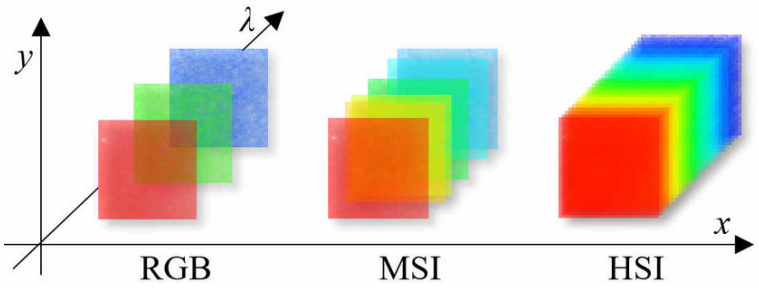


Figure 2.1: Difference between RGB imaging, multispectral imaging (MSI) and hyperspectral imaging (HSI). [1]

Remote sensing technology has revolutionized the way we observe and understand the Earth's surface, ecosystems, and atmospheres from a distance, primarily through the use of satellites or aircraft. Central to this field is the use of multispectral imaging (MSI), a technique that captures image data at specific wavelengths across the electromagnetic spectrum.

As shown in Figure 2.1, compared to an RGB image which includes 3 bands, a multispectral image is a collection of a few image layers of the same scene, each of them acquired at a particular wavelength band. Furthermore, unlike the human eye, which can only see visible light, multispectral cameras record data from both the visible light spectrum and other bands, including infrared and ultraviolet, providing a richer and more detailed view of an area than conventional photography.

One of the most impactful applications of multispectral imaging is in agricultural sciences, particularly in crop mapping. Crop mapping involves the classification and spatial representation of crop types over a large area, which is essential for monitoring crop health, predicting yields, and managing agricultural inputs more effectively. MSI technology is uniquely suited for this task because different crops reflect and absorb light differently at various wavelengths, including those beyond the visible spectrum. For instance, healthy vegetation strongly reflects near-infrared light and absorbs more visible light, a characteristic that can be distinctively captured through MSI.

2.1.2 Vegetation Indices

Vegetation absorbs solar radiation in different bands, which lie in various frequency ranges and wavelengths, and reflects different percentages of it back into the atmosphere. The percentage of reflected radiation in specific bands, such as near-infrared (NIR), red (RED), and blue (BLUE), varies with plant health and environmental stress. As a result, vegetation indices can effectively reflect the health and vigor of crops.

Common Vegetation Indices:

- NDVI (Normalized Difference Vegetation Index)
- EVI (Enhanced Vegetation Index)

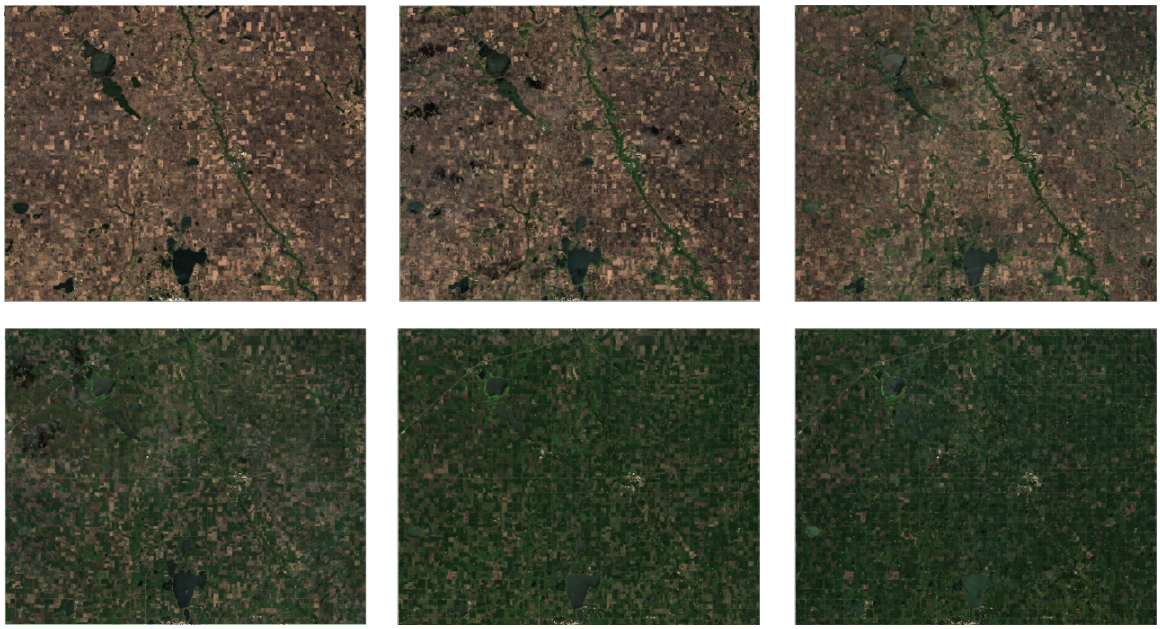


Figure 2.2: RGB images reconstructed from the red, green, and blue bands from the Sentinel-2 MSI images.

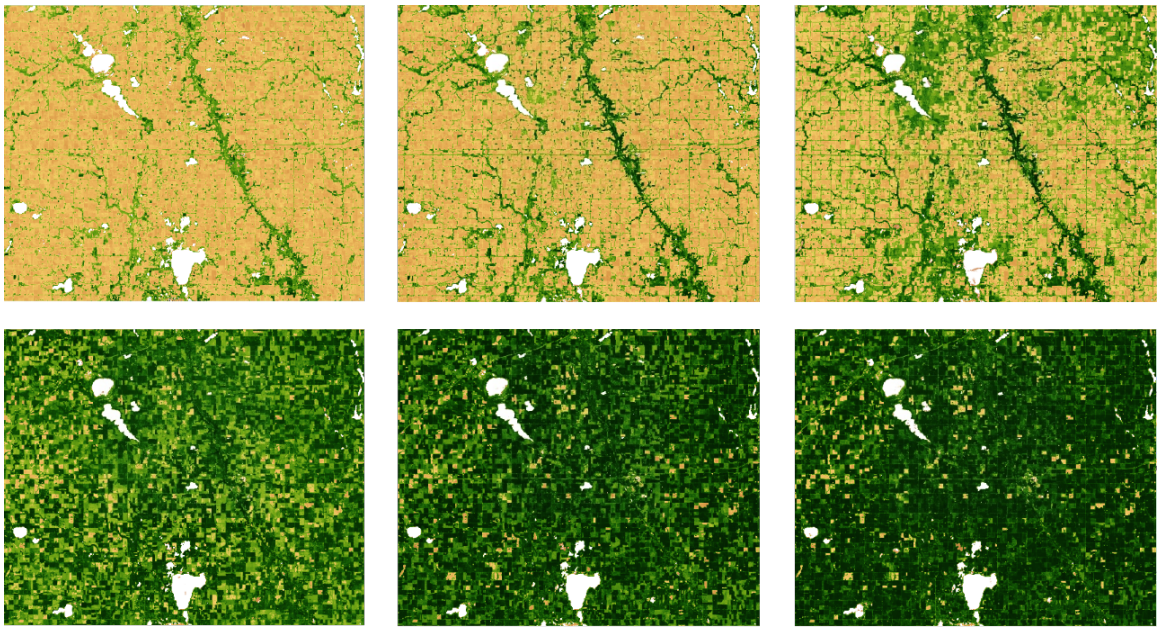


Figure 2.3: Corresponding NDVI images.

The most commonly used vegetation index is undoubtedly the NDVI (Normalized Difference Vegetation Index). NDVI describes the vigor level of vegetation and is calculated as the ratio between the difference and the sum of the reflected radiations in the near-infrared and red bands. The formula is:

$$\mathrm{NDVI} = \frac{\mathrm{NIR} - \mathrm{RED}}{\mathrm{NIR} + \mathrm{RED}} \tag{2.1}$$

Vegetation absorbs red-light energy to fuel photosynthesis and create chlorophyll. Therefore, a healthy plant with more chlorophyll will absorb more red-light energy and reflect more near-infrared energy than an unhealthy plant. This phenomenon

makes time-series NDVI a valuable tool for monitoring crop growth and forecasting yields.

For instance, Figure 2.2 displays RGB images reconstructed from the red, green, and blue bands of the Sentinel-2 MSI dataset, capturing the growth stages of crops such as soybean and corn from April to July in a study area in the USA. Additionally, Figure 2.3 presents the corresponding NDVI images, where green indicates high NDVI values and yellow indicates low values. As can be observed, over time, the color transition from yellow to green reflects the increasing NDVI values as the crops progress through their growth stages.

The Enhanced Vegetation Index (EVI) is another important vegetation index, designed to optimize the vegetation signal with improved sensitivity in high biomass regions and better accounting for atmospheric conditions. EVI is calculated using the NIR, RED, and BLUE bands as follows:

$$\text{EVI} = G \cdot \frac{\text{NIR} - \text{RED}}{\text{NIR} + C_1 \cdot \text{RED} - C_2 \cdot \text{BLUE} + L} \tag{2.2}$$

where: G is the gain factor, C_1 and C_2 are the coefficients for the aerosol resistance term, L is the canopy background adjustment term. Typically, the values for these constants are: $G = 2.5$, $C_1 = 6$, $C_2 = 7.5$, and $L = 1$.

EVI provides improved sensitivity over high biomass regions and better correction for atmospheric influences, making it a robust index for vegetation monitoring.

In conclusion, both NDVI and EVI are powerful tools for assessing vegetation health. NDVI is widely used due to its simplicity and effectiveness in providing insights into plant health and vigor. EVI, with its enhanced sensitivity and improved atmospheric correction, offers a more refined analysis, especially useful in regions with high biomass. Together, these indices play a crucial role in remote sensing applications, enabling accurate monitoring and management of agricultural and natural ecosystems.

2.1.3 MSI datasets

This chapter also provides a detailed overview of four frequently used MSI datasets: Sentinel-2, Landsat 7, Landsat 8, and MODIS, highlighting their resolution, spectral bands, revisit times, and significance.

- **Sentinel-2**, part of the European Space Agency's (ESA) Copernicus program, consists of two satellites: Sentinel-2A, launched in 2015, and Sentinel-2B, launched in 2017. These satellites provide high-resolution optical imagery and collectively cover the globe every five days.
  The key features of Sentinel-2 include its spatial resolution, which varies across different spectral bands. Specifically, the spatial resolution is 10 meters for Bands 2 (Blue), 3 (Green), 4 (Red), and 8 (Near Infrared - NIR). For Bands 5, 6, 7 (Red Edge), 8A (NIR), 11, and 12 (Shortwave Infrared - SWIR), the spatial resolution is 20 meters. Bands 1 (Coastal Aerosol), 9 (Water Vapor), and 10 (SWIR - Cirrus) have a spatial resolution of 60 meters. In total, Sentinel-2 encompasses 13 spectral bands ranging from the visible to the shortwave infrared spectrum. The revisit time for Sentinel-2, when both satellites are combined, is every five days at the equator.
  Sentinel-2's high spatial resolution and frequent revisit time make it particularly valuable for agricultural monitoring, vegetation analysis, and land cover

classification. Furthermore, its open access policy encourages extensive research and operational use, making it a critical resource for scientists and practitioners in various fields.

- **Landsat 7**, launched by NASA in 1999, continues the legacy of the Landsat program, providing critical data for Earth's surface monitoring. It has several key features that make it a valuable resource for researchers. The spatial resolution varies by band: the panchromatic band has a resolution of 15 meters, the multispectral bands (Bands 1-5 and 7) have a resolution of 30 meters, and the thermal infrared band (Band 6) has a resolution of 60 meters. Landsat 7 captures data in 8 spectral bands, including blue, green, red, near infrared (NIR), shortwave infrared (SWIR), and thermal infrared (TIR). The satellite revisits the same location on Earth every 16 days, providing a consistent time series of data.

  Landsat 7 is invaluable for historical comparisons and trend analysis due to its long-term data record, which spans over two decades. This extensive dataset allows researchers to study changes in land use and land cover, monitor forestry conditions, and conduct comprehensive environmental assessments. The satellite's ability to provide detailed and consistent imagery over time makes it a cornerstone for ongoing research and monitoring efforts.
- **Landsat 8**, launched in 2013, is the most recent addition to the Landsat series, providing improved data quality and additional spectral bands. This satellite offers significant enhancements over its predecessors, particularly in terms of spatial resolution and spectral coverage. The spatial resolution includes 15 meters for the panchromatic band (Band 8), 30 meters for the multispectral bands (Bands 1-7, 9), and 100 meters for the thermal infrared bands (Bands 10, 11), which are resampled to 30 meters.

  Landsat 8 features 11 spectral bands, including new coastal/aerosol and cirrus bands, in addition to the standard visible, near-infrared (NIR), shortwave infrared (SWIR), and thermal infrared (TIR) bands. This broad spectral coverage allows for more detailed and accurate observations of the Earth's surface.

  The satellite revisits each location on Earth every 16 days, providing frequent and regular data that is essential for monitoring changes over time.

  Landsat 8's improved sensors and additional bands enhance its utility for various applications, including water quality monitoring, vegetation health assessment, and urban mapping. Its continuity with previous Landsat missions supports long-term environmental studies, enabling researchers to compare current data with historical records to analyze trends and changes in the Earth's surface.
- **MODIS (Moderate Resolution Imaging Spectroradiometer)**, aboard NASA's Terra (launched in 1999) and Aqua (launched in 2002) satellites, provides comprehensive data covering the entire globe every 1-2 days.

  MODIS offers a range of key features that make it an invaluable tool for various research applications. Its spatial resolution varies across different bands: 250 meters for Bands 1-2 (Red, NIR), 500 meters for Bands 3-7 (Blue, Green, NIR, SWIR), and 1 kilometer for Bands 8-36 (ranging from visible to thermal infrared). This spectral diversity, encompassing 36 bands, allows for detailed analysis of the Earth's surface and atmosphere. The revisit time for MODIS

is daily at the equator, thanks to the combined capabilities of the Terra and Aqua satellites.

The importance of MODIS in research is profound. Its frequent global coverage and extensive range of data products make it indispensable for climate studies, large-scale environmental monitoring, and disaster response. MODIS data supports research on vegetation dynamics, land surface temperature, and atmospheric studies, providing critical insights into environmental changes and aiding in effective management and policy-making.

Sentinel-2, Landsat 7, Landsat 8, and MODIS are among the most frequently used MSI datasets in remote sensing research due to their unique capabilities and extensive applications. Sentinel-2's high spatial resolution and frequent revisit time, combined with the long-term continuity and free access of the Landsat program, provide robust data for detailed and long-term environmental studies. MODIS's global coverage and extensive data products complement these datasets by offering insights into large-scale and dynamic processes. Together, these datasets form a comprehensive foundation for a wide range of remote sensing applications and research initiatives.

By analyzing MSI data, scientists and agricultural managers can identify different crop types and assess their condition throughout the growing season. This capability is crucial for precision agriculture, where the goal is to optimize field-level management concerning crop farming practices. Information derived from MSI can help in making decisions on irrigation, pest and disease control, and harvesting, ultimately leading to increased efficiency and sustainability in agricultural production.

Furthermore, the ability of MSI to monitor changes over time supports the dynamic management of crops. This temporal analysis can detect anomalies in crop development, potentially indicating issues such as water stress, nutrient deficiencies, or infestations before they can cause significant damage. Regular monitoring through MSI thus helps in implementing timely interventions, ensuring crop health and productivity.

Technological advancements in remote sensing have led to improvements in the spatial, spectral, and temporal resolution of multispectral images, allowing for more precise and detailed data collection. This progression enhances our ability to manage natural resources, respond to environmental changes, and conduct scientific research across disciplines.

In sum, MSI represents a pivotal development in remote sensing, offering comprehensive insights that are vital for environmental monitoring, management, and the sustainable development of our planet. The application of MSI in crop mapping is a testament to its potential in enhancing agricultural productivity and sustainability, making it a critical tool in the arsenal of modern agronomy.

2.2 Problems in Crop Mapping using Time-Series MSI images

Crop mapping using time-series multi-spectral imaging (MSI) data offers significant potential for monitoring agricultural landscapes, assessing crop health, and managing agricultural resources. However, this approach also presents several challenges that researchers and practitioners must address to ensure accurate and reliable results. These challenges include issues related to crop location, cloud coverage, crop calendar and crop rotations, and the handling of big data. Each of these problems is detailed below.

2.2.1 Crop Location

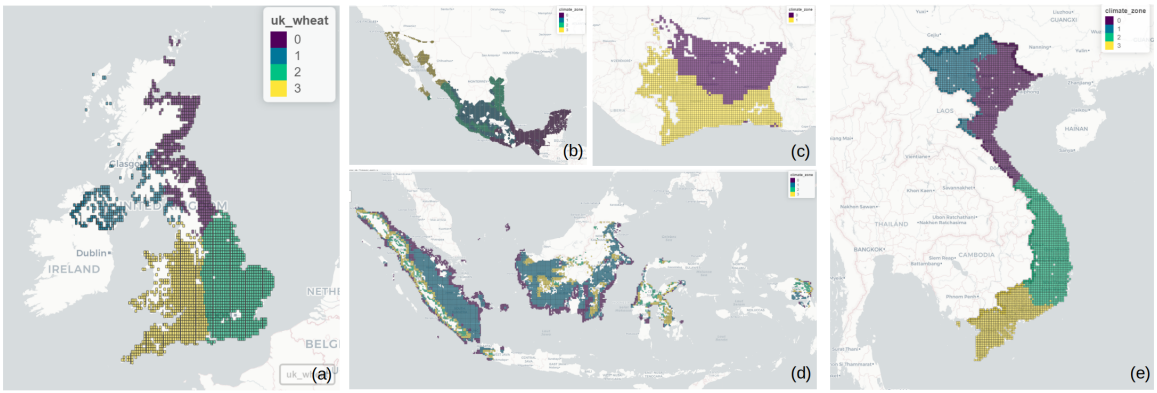


Figure 2.4: The Geographic Polygons with Climate Zones Clustering. (a) United Kingdom, (b) Mexico, (c) Ivory Coast, (d) Indonesia, (e) Vietnam.

When attempting to identify and map specific crops over large areas, having prior knowledge about the general distribution of crops can be crucial. However, this prerequisite can pose significant challenges:

- Lack of Prior Information: In many cases, detailed and up-to-date information about the main crops in a location is often unclear. This absence of data can make it challenging to pinpoint regions where specific crops are cultivated, complicating the early stages of crop mapping.
- Spatial Variability: Crop types can vary significantly within a large region. Even within known agricultural zones, the exact boundaries of different crops can change from year to year due to factors such as market demands, weather conditions, and farming practices.
- Climate Variability: In large regions, weather and climate conditions can differ significantly. These variations can lead to differences in crop growth characteristics, even for the same crop, making it more complex to accurately map and monitor crop development across diverse environments.

2.2.2 Cloud Coverage

Cloud cover is a significant obstacle in the use of MSI images for crop mapping. Clouds can obscure the view of the Earth's surface, leading to incomplete and unreliable data. In many regions, especially tropical and temperate zones, cloud cover can be frequent and persistent, reducing the number of usable images. This problem is exacerbated during the rainy season, which often coincides with critical

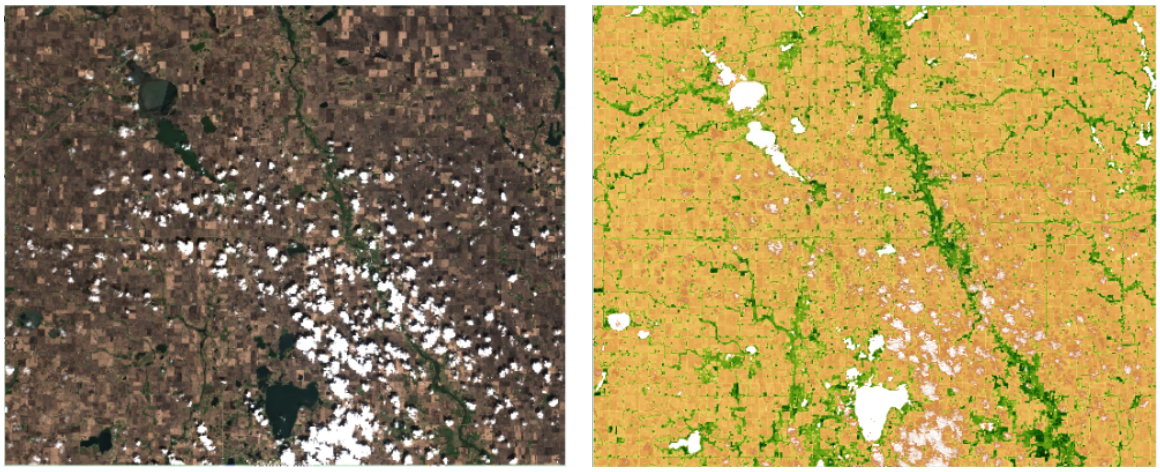


Figure 2.5: Cloud Coverage. Left: The RGB Image. Right: The NDVI Image.

periods of crop growth. Cloud cover can create gaps in the time-series data, making it challenging to track the development and growth stages of crops accurately. For instance, Figure 2.5 presents the RGB and NDVI images of the study area in the USA, where the primary crops are soybean and corn. A significant portion of the cropland is obscured by clouds, and since MSI cannot penetrate clouds, the white gaps in the NDVI image are caused by this obstruction. These gaps can lead to misinterpretations and errors in crop classification and monitoring. Additionally, even when clouds do not directly cover the target area, the shadows they cast can degrade image quality, introducing further challenges in data analysis.

2.2.3 Crop Calendar and Crop Rotations

The timing of planting, growing, and harvesting crops (crop calendar) and the practice of crop rotation (alternating crops in the same field across different seasons or years) add layers of complexity to crop mapping. Crop calendars can vary significantly between regions and even within the same region over different years. These variations depend on factors such as climate, soil conditions, and farming practices. Even for the same crop, differences in planting dates result in variations in growth stages, so when monitoring the Vegetation Index at the same time period, their patterns may not align. This variability requires adaptive algorithms that can account for changes in crop types and growth stages over time.

Moreover, accurately mapping crops requires precise knowledge of the local crop calendar, which can be difficult to obtain and maintain. While crop rotation practices benefit soil health and pest management, they also complicate the crop mapping process. Different crops have distinct spectral signatures, and rotating crops means that spectral data collected in one season may not be applicable to the next.

Additionally, the timing of satellite passes may not always coincide with critical phenological stages, potentially missing key growth periods and impacting the accuracy of crop classification and yield predictions.

2.2.4 Big Data

Handling the sheer volume of data generated by high-resolution MSI images over large areas introduces several challenges. High-resolution images, such as those with 30-meter spatial resolution, generate vast amounts of data, especially when covering

extensive areas like entire countries or continents. For example, mapping the entire USA with such resolution involves processing and storing petabytes of data. Storing large volumes of MSI data requires significant infrastructure, including high-capacity storage solutions and efficient data management systems. This requirement can be a substantial financial and logistical burden for many organizations. Processing time-series data for large-scale crop mapping demands substantial computational resources. Tasks such as image pre-processing, classification, and analysis can be computationally intensive, necessitating powerful hardware and optimized algorithms. Integrating MSI data with other datasets (e.g., weather data, soil maps, and agricultural statistics) to improve the accuracy of crop mapping further increases the complexity and volume of data to be handled. Efficient data integration and management strategies are essential to derive meaningful insights from these diverse data sources.

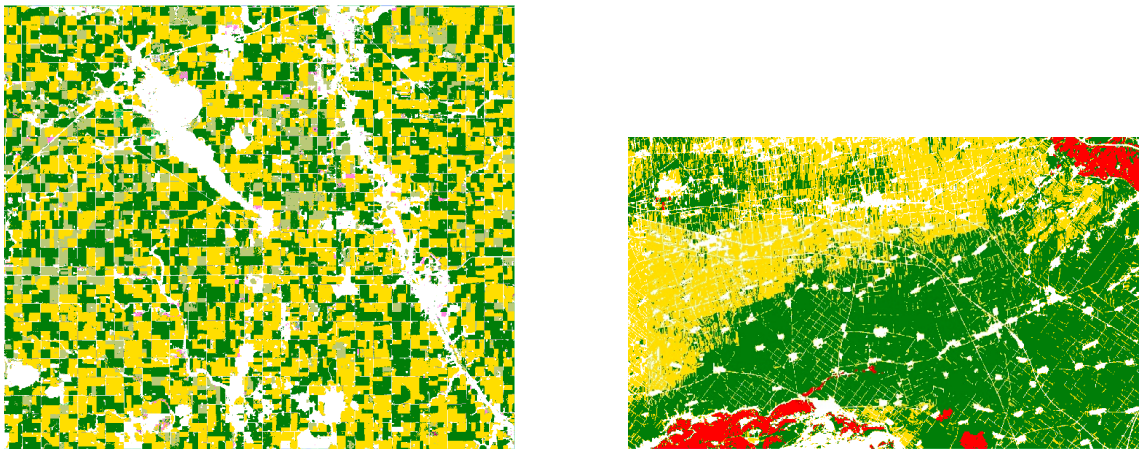
2.2.5 Cross Domain Issue in Crop Mapping

Early crop mapping, which entails identifying the areas where crops are cultivated prior to the harvest season, serves as a critical foundation for agricultural planning, resource allocation, crop insurance, and risk management [22, 23]. The identification process is largely driven by the spectral characteristics of vegetation, which are influenced by plant structure, leaf biochemistry, and phenological stages [24]. As such, time-series analysis of multispectral satellite images has become the primary method for classifying land cover. The surge in satellite data availability and advancements in technology have led to the adoption of deep learning (DL) for producing detailed and accurate maps of crop cultivation [25]. These sophisticated DL methods depend heavily on extensive ground truth data for training and validating models, which can be sourced directly from field surveys or indirectly from public datasets like the USDA's Cropland Data Layer (CDL), though the latter are often considered weak labels [26, 27, 13].

However, acquiring suitable ground truth data for any given region poses significant challenges. While public datasets like the CDL offer dependable and timely data for training purposes, they are limited to a few countries and typically become available post-harvest, complicating timely application. The process of gathering these data is not only expensive and labor-intensive but can also be impractical, particularly in less developed regions. In the absence of local ground truth data, researchers often resort to a "direct transfer strategy", utilizing labeled data from other regions to train models before applying them in new areas [28, 29]. However, spectral features of crops have both inter-region variability and inter-annual variability due to changes in soil composition, climate conditions, and crop progress [30]. These differences lead to what is known as the cross-domain issue, including the cross-year issue and the cross-region issue, where there is a shift in data distribution between the training data (source domain) and the application data (target domain). This shift can severely degrade the performance of DL models in new and untested regions and years, challenging the fundamental machine learning assumption that training and application data are independent and identically distributed (i.i.d) or at least similarly distributed.

2.2.5.1 Cross-Year Issue

The cross-year issue deals with the variability in crop spectral signatures from year to year, which can result from a myriad of factors including meteorological conditions, changes in crop management practices, and phenological shifts. This



(a) (b)

Figure 2.6: 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.

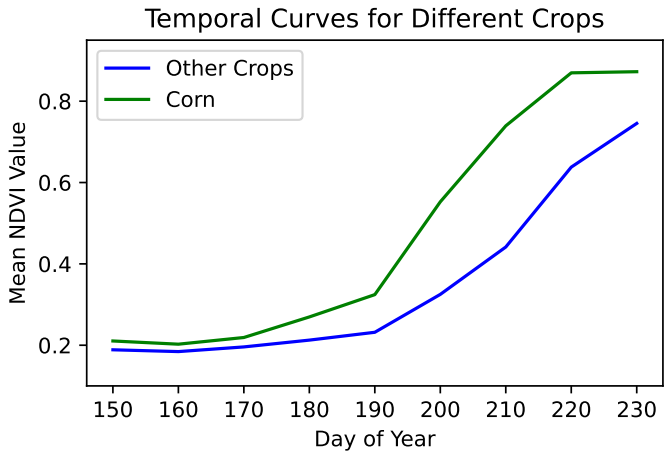
temporal variation can lead to shifts in data distribution, posing challenges for models trained on data from specific years. Models that fail to generalize well across different years can produce inaccurate crop type predictions and yield estimates, which can mislead resource distribution and crop management strategies.

For instance, Figure 2.6a shows the ground truth data (labels) for the study area shown in Figure 2.5. Green represents corn, yellow represents soybean, red represents rice, light olive represents fallow cropland, pink represents alfalfa, and white represents non-cropland. The temporal NDVI curves in Figure 2.7 display the Average Time-series NDVI (Normalized Difference Vegetation Index) of corn and other crops across two different years (2019 and 2020) in this study area, highlighting how NDVI values change over the growing season. 2.7a represents the data from 2019, while Figure 2.7b reflects data from 2020.

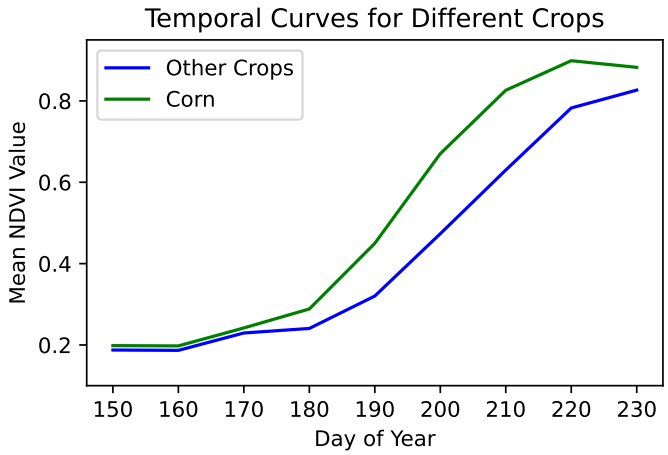
These plots demonstrate the typical seasonal growth patterns of corn, where the NDVI increases significantly between Day 160 and Day 210 of the year as the crop matures, reaching its peak between Day 210 and 220, before stabilizing or slightly declining. Other crops, however, consistently show lower NDVI values, indicating less dense or less healthy vegetation compared to corn.

The cross-year issue arises due to significant variability in the spectral signatures of crops like corn, which are influenced by external factors such as changing meteorological conditions, crop management practices, and phenological stages. These temporal variations can alter the distribution of spectral data from one year to the next, creating challenges for models trained on specific-year data to generalize effectively across different years.

In this figure, the differences between the NDVI curves from 2019 and 2020 reflect how these factors can shift spectral signatures and impact the overall data distribution. Models that are not adapted to handle such variability may struggle to make accurate crop type predictions or yield estimates, potentially leading to suboptimal decisions in resource allocation or crop management strategies. Thus, handling cross-year variability is critical for improving the robustness of models used



(a)



(b)

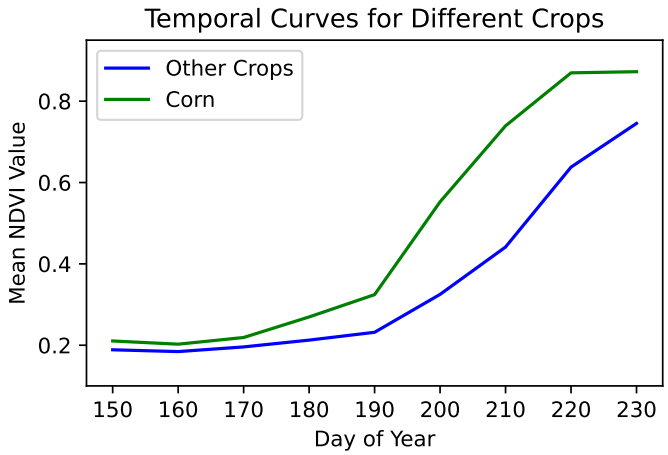
Figure 2.7: The Average Time-series NDVI of Corn for the Source Domain and Target Domain. (a) Jackson County, 2019. (b) Jackson County, 2020.

in crop mapping.

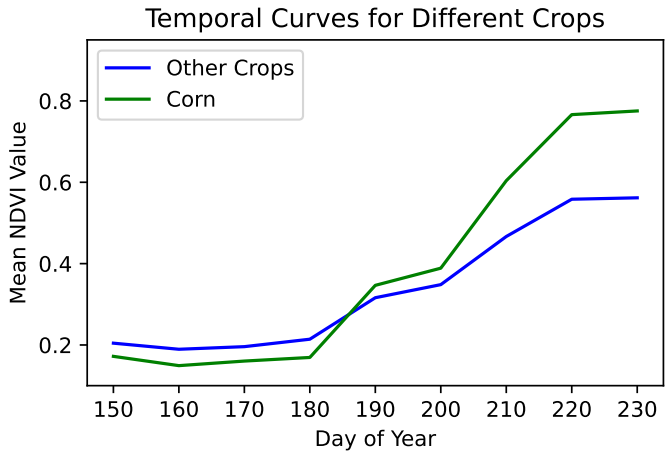
2.2.5.2 Cross-Region Issue

Figure 2.8 illustrates the Average Time-series NDVI of Corn for both the Source Domain (Jackson County, 2019) and the Target Domain (a study area in China, 2019). Subfigures 2.8a and 2.8b depict the temporal variations in NDVI for corn and other crops over a period of time (Day of Year), highlighting that crops exhibit a higher NDVI curve in Jackson county than in the study area of China in 2019.

The cross-region issue is evident when comparing these figures, which underscores the challenges faced when applying models trained in one geographic area to another. Differences in soil types, climate conditions, and farming practices between regions, such as between Jackson County and the study area in China, affect the spectral signatures of crops like corn. These variations lead to a shift in data distribution, posing a significant challenge for models. The cross-region variation, particularly in crop NDVI trends, tends to be more pronounced than cross-year differences, further complicating model transferability and performance across different regions.



(a)



(b)

Figure 2.8: The Average Time-series NDVI of Corn for the Source Domain and Target Domain. (a) Jackson County, 2019. (b) The study area of China, 2019.

2.2.5.3 Cross-Region Issue with Label Spaces Discrepancies

Adding complexity to the cross-region issue is the problem of label space discrepancies. This occurs when the categories of crops present in the target region do not match those in the source region used for training. These discrepancies can be due to different regional crop classifications, the presence of unique local crops not included in the source dataset, or different stages of crop maturity at the time of image capture.

For instance, Figure 2.6 illustrates the primary crop types in the study areas of the USA and China. While both regions include corn and soybean, rice is exclusive to the China study area. Addressing this aspect of the cross-domain issue requires not only domain adaptation techniques but also innovative approaches in model training to accommodate varying label spaces.

2.3 Platforms

Satellite image processing platforms play a crucial role in various fields, including environmental monitoring, urban planning, disaster management, and especially agriculture. These platforms provide access to vast amounts of remote sensing data and powerful tools for analyzing this data. This Section provides an overview of four prominent satellite image processing platforms: Google Earth Engine (GEE), Sentinel Hub, Earthdata (NASA Worldview), and USGS Earth Explorer.

- **Google Earth Engine (GEE)** is a cloud-based platform designed for planetary-scale environmental data analysis. It hosts a vast public data archive that includes petabytes of satellite imagery and geospatial datasets, making it an invaluable resource for researchers and practitioners. One of its key features is its massive data catalog, which provides access to a wide range of datasets from various satellite missions such as Landsat, Sentinel, MODIS, and more. Additionally, Google Earth Engine offers powerful processing capabilities, allowing users to perform complex analyses using JavaScript or Python APIs while leveraging Google's robust cloud infrastructure. This platform is also highly scalable, capable of efficiently processing large datasets, making it suitable for global-scale studies. The interactive interface, particularly the Code Editor, offers an engaging environment for writing scripts, visualizing data, and sharing results, enhancing the overall user experience and facilitating collaboration.
- **Sentinel Hub** is an innovative cloud-based platform developed by Sinergise for processing and analyzing satellite data. It provides easy access to data from a variety of satellite missions, including Sentinel-1, Sentinel-2, and Landsat-8. One of its key features is its wide data accessibility, which offers seamless access to multiple satellite data sources. This allows users to obtain comprehensive and diverse datasets for their analysis.
  The platform also supports customizable processing, enabling users to create custom scripts to process and visualize data according to their specific needs. This flexibility ensures that users can tailor their data analysis to suit their particular requirements. Additionally, Sentinel Hub provides robust APIs and web services that facilitate integration with other applications and services. This enhances the platform's flexibility and usability, allowing it to be incorporated into various workflows and systems.
  Another significant feature of Sentinel Hub is the EO Browser, an intuitive online tool for exploring and analyzing satellite data. The EO Browser allows users to interact with and analyze satellite imagery without the need for coding, making it accessible to a broader range of users, including those without a technical background. This combination of features makes Sentinel Hub a powerful and versatile tool for accessing, processing, and analyzing satellite data.
- **Earthdata (NASA Worldview)**, managed by NASA, provides access to a wealth of Earth observation data. The Worldview tool is a user-friendly online application for visualizing satellite imagery and other geospatial data in near real-time.
  One of the key features of Earthdata is its near real-time data access, which enables timely analysis and decision-making. This capability is crucial for applications such as disaster response, environmental monitoring, and agricultural

management, where up-to-date information is essential.
Another important aspect of Earthdata is its comprehensive data coverage. The platform includes data from a variety of NASA missions, including MODIS (Moderate Resolution Imaging Spectroradiometer) and VIIRS (Visible Infrared Imaging Radiometer Suite), among others. This extensive dataset allows users to perform in-depth analyses across a broad range of scientific and practical applications.
Moreover, Earthdata's free and open access policy ensures that all data is freely available to the public. This openness supports the principles of open science and research, allowing scientists, educators, policymakers, and the general public to utilize the data without restrictions.
The Worldview tool within Earthdata offers robust visualization tools that enable users to overlay various data layers, customize visualizations, and export images for reports and presentations. This functionality enhances the user's ability to interpret and communicate the data effectively, making it a valuable resource for both scientific research and practical applications.

- **USGS Earth Explorer** is a web-based application provided by the United States Geological Survey (USGS) for searching, downloading, and analyzing satellite imagery and geospatial data. This platform hosts an extensive data repository that includes a vast archive of remote sensing data from missions such as Landsat, Sentinel, and others. The advanced search capabilities of Earth Explorer allow users to find data based on specific locations, time frames, and other criteria, making it easier to pinpoint relevant datasets. Users benefit from free access to a wide range of datasets, which can be freely downloaded for further analysis. Additionally, Earth Explorer provides basic tools for data visualization and analysis, which are suitable for preliminary investigations and initial data exploration.

These satellite image processing platforms offer unique strengths and capabilities, making them indispensable tools for a wide range of applications. Google Earth Engine excels in large-scale data analysis with its powerful cloud infrastructure, Sentinel Hub provides flexible and customizable processing solutions, Earthdata (NASA Worldview) offers near real-time data access with an emphasis on usability, and USGS Earth Explorer serves as a comprehensive data repository with robust search functionalities. Together, they represent the cutting edge of remote sensing and geospatial data analysis, supporting research, decision-making, and innovation across various fields.

2.4 Crop Mapping Using Machine Learning Methods

Machine learning is a subset of artificial intelligence (AI) that involves developing algorithms and statistical models that enable computers to perform specific tasks without explicit instructions. By learning from data, these algorithms can make predictions, classify data, and uncover hidden patterns. Machine learning can be broadly categorized into traditional machine learning and deep learning (DL).

Traditional machine learning encompasses a range of algorithms that often require structured data and rely heavily on feature engineering, where humans manually select the attributes that best represent the underlying patterns. Methods like Support Vector Machines (SVM), Decision Trees (DTs), and Random Forests (RF) are prominent examples in this category. SVM aims to find the optimal hyperplane that separates different classes by maximizing the margin between them, while DTs operate by splitting the data into branches based on decision rules derived from the features. RF, an ensemble technique, improve prediction accuracy by constructing multiple decision trees and averaging their outcomes to mitigate the risk of overfitting.

In contrast, deep learning is inspired by the structure and function of the human brain and focuses on building neural networks that automatically learn hierarchical representations from large amounts of data. Convolutional Neural Networks (CNNs), Recurrent Neural Networks (RNNs), and Long Short-Term Memory networks (LSTM) are some of the widely used architectures in this domain. CNNs are particularly well-suited for image recognition tasks as they can capture spatial hierarchies by applying filters to extract features like edges and textures. RNNs, with their ability to handle sequential data, are often applied to time-series analysis or natural language processing, though they struggle with long-term dependencies. LSTMs, an advanced version of RNNs, are specifically designed to address this limitation, making them highly effective for tasks like speech recognition and sequence prediction.

To generate precise crop cultivation maps from these images, various supervised DL methods, including CNNs [31, 32], Temporal Convolutional Neural Network (TempCNN) [19], and LSTM [33, 34], have been explored. Notably, the findings indicate that DL approaches outperform conventional techniques like SVM [35], DTs [36, 37], and RF [17, 21]. In this section, the foundational knowledge of these methods will be introduced. Furthermore, Generative Adversarial Networks (GANs) will be discussed, as they play a crucial role in my work to address the cross-domain issues previously mentioned.

2.4.1 Traditional Machine Learning Methods

Crop mapping using remote sensing data has long relied on traditional machine learning techniques, such as SVM and RF. These methods have been extensively employed for their effectiveness in handling the spectral and spatial variability inherent in satellite imagery [38, 39, 40]. SVM and RF classifiers work by learning decision boundaries and ensemble predictions from labelled data, respectively. Although they provide a solid foundation for crop classification, their reliance on hand-engineered features and limited ability to capture complex non-linear relationships restrict their performance in dynamic and varied agricultural landscapes.

2.4.1.1 Support Vector Machines

SVM is a supervised machine learning algorithm used for classification and regression tasks, although it is more commonly applied to classification problems.

The primary objective of SVM is to find a hyperplane that best separates the data points of different classes in a high-dimensional space. SVM is known for its effectiveness in high-dimensional spaces and its robustness when the number of dimensions exceeds the number of samples.

The principle of SVM is to find the optimal hyperplane that maximizes the margin between the closest data points of different classes, known as support vectors. The margin is defined as the distance between the hyperplane and the nearest data points from either class.

In the simplest case, when the data is linearly separable, SVM finds a linear hyperplane to separate the classes. For data that is not linearly separable, SVM uses a technique called the kernel trick to transform the data into a higher-dimensional space where it becomes linearly separable. Common kernel functions include the polynomial kernel, radial basis function (RBF) kernel, and sigmoid kernel.

Mathematical Formulation

The optimization problem for a linear SVM can be expressed as:

$$\text{maximize} \quad \frac{2}{||\mathbf{w}||} \tag{2.3}$$

subject to the constraints:

$$y_i(\mathbf{w} \cdot \mathbf{x}_i + b) \geq 1 \quad \text{for all } i \tag{2.4}$$

, where: $\mathbf{w}$ is the weight vector, b is the bias term, $\mathbf{x}_i$ represents the feature vectors of the training data, y_i represents the class labels, which are either $+1$ or -1.

The goal is to maximize the margin, which is equivalent to minimizing $|\mathbf{w}|$, subject to the constraint that all data points are correctly classified. This can be formulated as a convex optimization problem:

$$\text{minimize} \quad \frac{1}{2}||\mathbf{w}||^2 \tag{2.5}$$

subject to:

$$y_i(\mathbf{w} \cdot \mathbf{x}_i + b) \geq 1 \quad \text{for all } i \tag{2.6}$$

In cases where the data is not linearly separable, the optimization problem is modified to allow some misclassification, introducing slack variables ξ_i:

$$\text{minimize} \quad \frac{1}{2}||\mathbf{w}||^2 + C\sum_{i=1}^{n}\xi_i \tag{2.7}$$

subject to:

$$y_i(\mathbf{w} \cdot \mathbf{x}_i + b) \geq 1 - \xi_i \quad \text{and} \quad \xi_i \geq 0 \quad \text{for all } i \tag{2.8}$$

, where C is a regularization parameter that controls the trade-off between maximizing the margin and minimizing the classification error.

For non-linear SVM, the kernel trick is applied to transform the input data into a higher-dimensional space using a kernel function $K(\mathbf{x}_i, \mathbf{x}_j)$. The optimization problem then becomes:

$$\text{minimize} \quad \frac{1}{2}||\mathbf{w}||^2 + C\sum_{i=1}^{n}\xi_i \tag{2.9}$$

subject to:

$$y_i(\sum_{j=1}^{n}\alpha_j y_j K(\mathbf{x}_i, \mathbf{x}_j) + b) \geq 1 - \xi_i \quad \text{and} \quad \xi_i \geq 0 \quad \text{for all } i \tag{2.10}$$

, where α_j are the Lagrange multipliers.

2.4.1.2 Random Forests

RF is a supervised machine learning algorithm that is widely used for classification and regression tasks. It is an ensemble learning method, which means it constructs multiple decision trees during training and combines their outputs to improve the overall performance and robustness of the model. The primary advantage of Random Forest is its ability to handle large datasets with higher dimensionality and its robustness to overfitting.

The principle of Random Forest involves building a multitude of decision trees and aggregating their results. Each tree in the forest is trained on a random subset of the data (with replacement, known as bootstrap sampling), and at each split in the tree, a random subset of features is considered for splitting. This randomness helps in creating trees that are less correlated and thus improves the overall generalization of the model.

Mathematical Formulation

For a dataset with n samples, a bootstrap sample is created by randomly sampling n samples with replacement. This process is repeated to create multiple bootstrap samples for training individual trees. For each node in the tree, a random subset of m features is selected from the total p features. The best split is determined based on this subset of features. For a given input $\mathbf{x}$, the prediction for classification is:

$$\hat{y} = \text{majority\_vote}(T_1(\mathbf{x}), T_2(\mathbf{x}), \ldots, T_B(\mathbf{x})) \tag{2.11}$$

For regression, the prediction is:

$$\hat{y} = \frac{1}{B}\sum_{b=1}^{B} T_b(\mathbf{x}) \tag{2.12}$$

RF is a powerful and versatile machine learning algorithm that excels in both classification and regression tasks. Its ability to handle high-dimensional data, robustness to overfitting, and ease of implementation make it a popular choice for many applications. By aggregating the predictions of multiple decision trees, Random Forest can provide accurate and reliable results even when dealing with complex and noisy datasets.

2.4.2 Deep Learning Methods

The introduction of deep learning has significantly enhanced the capabilities of crop mapping models. Techniques such as CNNs and LSTM are now at the forefront due to their superior ability to autonomously extract and learn from the temporal and spatial characteristics present in remote sensing data [31, 34, 41]. These models excel in integrating and analyzing vast amounts of data, allowing

for nuanced understanding and classification of various crop types. By training on multi-year, region-specific crop data, deep learning models can adapt to annual variations and regional specificities, thereby improving the accuracy and robustness of crop classifications.

To further address the challenges of adapting to new regions and years without ample labelled data, techniques like direct transfer and unsupervised domain adaptation have been implemented. Direct transfer involves applying knowledge gained from one domain to another, assuming similar data distributions [28, 42]. However, this method often struggles with cross-domain issues due to significant differences in environmental and agricultural conditions.

Unsupervised Domain Adaptation (UDA) techniques, such as Domain Adversarial Neural Networks (DANN), have been developed to bridge the gap between different regions by aligning deep feature distributions. This approach minimizes the domain discrepancies through adaptations like the Maximum Mean Discrepancy (MMD) and its variants, enhancing the model's generalization capabilities across diverse settings [14, 15, 16]. These adaptations help mitigate the impact of lacking matched, labelled datasets across regions, making deep learning models more flexible and applicable in global agricultural monitoring tasks.

2.4.2.1 Artificial Neural Networks

ANNs are a fundamental component of deep learning, modelled loosely after the human brain. ANNs consist of layers of interconnected nodes, known as neurons, where each connection has an associated weight. The network learns by adjusting these weights based on the error in predictions. ANNs are versatile and can be used for various tasks, including classification, regression, and time-series forecasting.

The basic structure of an ANN includes an input layer, one or more hidden layers, and an output layer. Each neuron in a layer is connected to every neuron in the subsequent layer, and these connections are represented by weights. The information flows from the input layer through the hidden layers and finally to the output layer, where a decision or prediction is made.

Mathematical Formulation

The output of a neuron is calculated as a weighted sum of its inputs, passed through an activation function. For a neuron j in a hidden layer, the output z_j can be expressed as:

$$z_j = \sigma\left(\sum_{i=1}^{n} \mathbf{w}_{ij} x_i + b_j\right) \tag{2.13}$$

where: $\mathbf{w}_{ij}$ represents the weight between the i-th input neuron and the j-th hidden neuron, x_i is the input to the neuron, b_j is the bias term for the j-th hidden neuron, σ is the activation function. The activation function σ introduces non-linearity into the network, allowing it to model complex relationships. Common activation functions include the sigmoid function, the hyperbolic tangent function (tanh), and the rectified linear unit (ReLU). The ReLU function is often used in modern neural networks and is defined as:

$$\sigma(z) = \max(0, z) \tag{2.14}$$

The output layer produces the final prediction of the network. For a classification problem with m classes, the output of the network can be expressed as:

$$\hat{y}_k = \frac{\exp(z_k)}{\sum_{j=1}^{m} \exp(z_j)} \tag{2.15}$$

, where: $\hat{y}_k$ is the predicted probability of class k, z_k is the output of the neuron corresponding to class k. This function is known as the softmax function and is commonly used in the output layer of a neural network for multi-class classification tasks.

Learning Process

The learning process in an ANN involves adjusting the weights $\mathbf{w}_{ij}$ to minimize the error between the predicted output $\hat{y}$ and the actual target value y. This error is quantified using a loss function $L(y, \hat{y})$, which measures the discrepancy between the predicted and actual values. A commonly used loss function for classification problems is the cross-entropy loss, defined as:

$$L(\hat{y}, y) = -\sum_{k=1}^{m} y_k \log(\hat{y}_k) \tag{2.16}$$

, where: y_k is the true label, represented as a one-hot encoded vector. The goal is to minimize this loss function by iteratively adjusting the weights using a process called backpropagation, combined with an optimization algorithm like gradient descent. During backpropagation, the gradient of the loss function with respect to each weight is calculated, and the weights are updated in the opposite direction of the gradient:

$$\mathbf{w}ij \leftarrow \mathbf{w}ij - \eta \frac{\partial L}{\partial \mathbf{w}_{ij}} \tag{2.17}$$

, where: η is the learning rate, a hyperparameter that controls the step size of the weight updates. This process is repeated over many iterations, called epochs, until the loss converges to a minimum, resulting in a trained neural network capable of making accurate predictions.

2.4.2.2 Convolutional Neural Networks

CNNs are a class of deep learning models specifically designed for processing structured grid data, such as images. CNNs have revolutionized the field of computer vision by automating the process of feature extraction and allowing for the hierarchical learning of features from raw pixel values to complex patterns. Unlike traditional machine learning methods, CNNs do not require manual feature engineering, making them particularly effective for image recognition, object detection, and segmentation tasks.

Principle of CNN

The core idea behind CNNs is the use of convolutional layers to extract spatial features from the input data. These layers apply a set of filters (also known as kernels) to the input image, generating feature maps that capture different aspects of the image, such as edges, textures, and shapes. The key operations in a CNN include convolution, activation, pooling, and fully connected layers.

- **Convolution Operation:**

  The convolution operation involves sliding a filter $\mathbf{w}$ over the input image $\mathbf{X}$ to produce a feature map $\mathbf{F}$. The filter $\mathbf{w}$ is a small matrix of learnable parameters, and the convolution is performed as follows:

  $$\mathbf{F}(i,j) = \sum_{m=1}^{M} \sum_{n=1}^{N} \mathbf{w}(m,n) \cdot \mathbf{X}(i+m-1, j+n-1) \tag{2.18}$$

  , where $\mathbf{F}(i,j)$ is the value of the feature map at position (i,j), $\mathbf{w}(m,n)$ represents the weights of the filter at position (m,n), and $M \times N$ is the size of the filter. The output feature map $\mathbf{F}$ retains the spatial relationships within the input image while highlighting certain features.

- **Activation Function:**

  After convolution, an activation function is applied to introduce non-linearity into the model. The most commonly used activation function in CNNs is the Rectified Linear Unit (ReLU), defined as:

  $$\text{ReLU}(x) = \max(0, x) \tag{2.19}$$

  , where x is the input to the activation function. The ReLU function sets all negative values to zero while keeping positive values unchanged, helping the network learn complex patterns.

- **Pooling Operation:**

  Pooling is a down-sampling operation that reduces the spatial dimensions of the feature maps, thereby decreasing the computational load and controlling overfitting. The most common pooling operation is max-pooling, where the maximum value within a pooling window is selected:

  $$\mathbf{P}(i,j) = \max_{(m,n) \in \text{pool window}} \mathbf{F}(i+m-1, j+n-1) \tag{2.20}$$

  , where $\mathbf{P}(i,j)$ is the pooled feature map value at position (i,j), and pool window denotes the pooling window size.

- **Fully Connected Layers:**

  After several convolutional and pooling layers, the high-level features extracted by the CNN are flattened into a vector and passed through fully connected layers. These layers perform the final classification based on the features learned during the convolutional stages. The output of the fully connected layer can be represented as:

  $$\mathbf{y} = \sigma(\mathbf{W} \cdot \mathbf{z} + \mathbf{b}) \tag{2.21}$$

  , where $\mathbf{y}$ is the output vector (e.g., class probabilities), $\mathbf{W}$ is the weight matrix, $\mathbf{z}$ is the input feature vector from the previous layer, $\mathbf{b}$ is the bias vector, and σ is the activation function (e.g., softmax for classification).

Mathematical Formulation

The overall CNN model can be described as a sequence of operations that involve the convolution of the input with a set of filters, followed by the application of an activation function, pooling, and finally, classification using fully connected layers.

For an input image $\mathbf{X}$ and a set of filters $\mathbf{W} = \mathbf{w}_1, \mathbf{w}_2, \ldots, \mathbf{w}_k$, the feature maps $\mathbf{F}_k$ are computed as:

$$\mathbf{F}_k = \text{ReLU}(\mathbf{X} * \mathbf{w}_k + \mathbf{b}_k) \tag{2.22}$$

, where $*$ denotes the convolution operation, $\mathbf{b}_k$ is the bias term for the k-th filter, and ReLU is the activation function applied element-wise.

The pooled feature maps $\mathbf{P}_k$ are obtained through the pooling operation:

$$\mathbf{P}_k = \text{MaxPooling}(\mathbf{F}_k) \tag{2.23}$$

Finally, the pooled features are flattened and passed through fully connected layers to produce the output:

$$\mathbf{y} = \text{softmax}(\mathbf{W}_{\text{fc}} \cdot \text{flatten}(\mathbf{P}) + \mathbf{b}_{\text{fc}}) \tag{2.24}$$

, where $\mathbf{W}_{\text{fc}}$ and $\mathbf{b}_{\text{fc}}$ are the weights and biases of the fully connected layer, and $\text{flatten}(\mathbf{P})$ represents the flattened feature maps.

CNNs are a powerful class of deep learning models that have achieved state-of-the-art performance in various computer vision tasks. By leveraging convolutional layers, activation functions, pooling operations, and fully connected layers, CNNs are able to automatically learn hierarchical features from raw input data. This ability to extract and utilize complex patterns makes CNNs a cornerstone in the field of image recognition, object detection, and beyond.

2.4.2.3 Recurrent Neural Networks

RNNs are a class of neural networks designed specifically for processing sequential data. Unlike traditional feedforward neural networks, RNNs have connections that form directed cycles, allowing them to maintain a hidden state that can capture information about previous inputs. This ability to retain memory of past data makes RNNs particularly well-suited for tasks involving sequences, such as time series prediction, natural language processing, and speech recognition.

However, standard RNNs suffer from the vanishing gradient problem, which makes it difficult for them to learn long-term dependencies. To address this issue, the Long Short-Term Memory (LSTM) network, a special type of RNN, was developed. LSTMs are designed to better capture long-term dependencies by introducing a more complex structure in their recurrent connections.

LSTM Architecture

An LSTM network consists of a series of LSTM cells, each of which contains three key components: an input gate, a forget gate, and an output gate. These gates control the flow of information into, through, and out of the cell, enabling the network to maintain and update a cell state that carries information across many time steps. The key equations governing an LSTM cell are as follows:

- **Forget Gate:** Determines which part of the previous cell state should be discarded.

$$f_t = \sigma(\mathbf{W}_f \cdot [\mathbf{h}_{t-1}, \mathbf{x}_t] + \mathbf{b}_f) \tag{2.25}$$

- **Input Gate:** Decides which values will be updated in the cell state.

$$i_t = \sigma(\mathbf{W}_i \cdot [\mathbf{h}_{t-1}, \mathbf{x}_t] + \mathbf{b}_i) \tag{2.26}$$

- **Candidate Cell State:** Generates the new candidate values for updating the cell state.
$$\tilde{\mathbf{C}}_t = \tanh(\mathbf{W}_C \cdot [\mathbf{h}_{t-1}, \mathbf{x}_t] + \mathbf{b}_C) \tag{2.27}$$
- **Cell State Update:** Updates the cell state using the forget and input gates.
$$\mathbf{C}_t = f_t \cdot \mathbf{C}_{t-1} + i_t \cdot \tilde{\mathbf{C}}_t \tag{2.28}$$
- **Output Gate:** Determines the output of the current cell and the hidden state for the next time step.
$$o_t = \sigma(\mathbf{W}_o \cdot [\mathbf{h}_{t-1}, \mathbf{x}_t] + \mathbf{b}_o) \tag{2.29}$$
- **Hidden State Update:** Computes the new hidden state based on the updated cell state.
$$\mathbf{h}_t = o_t \cdot \tanh(\mathbf{C}_t) \tag{2.30}$$
, where: f_t, i_t, and o_t are the forget, input, and output gates, respectively. $\mathbf{W}_f$, $\mathbf{W}_i$, $\mathbf{W}_C$, and $\mathbf{W}_o$ are weight matrices. $\mathbf{b}_f$, $\mathbf{b}_i$, $\mathbf{b}_C$, and $\mathbf{b}_o$ are bias vectors. $\mathbf{h}_{t-1}$ is the hidden state from the previous time step. $\mathbf{x}_t$ is the input at the current time step. $\mathbf{C}_t$ is the cell state at the current time step. $\tilde{\mathbf{C}}_t$ is the candidate cell state. σ is the sigmoid activation function, and tanh is the hyperbolic tangent activation function.

LSTMs provide a powerful solution to the limitations of standard RNNs by effectively capturing long-term dependencies in sequential data. The introduction of gates allows LSTM networks to regulate the flow of information, selectively forgetting, updating, and outputting parts of the cell state. This architecture has made LSTMs a foundational tool in deep learning for tasks that involve sequential data, such as language modeling, translation, and speech recognition. Through the careful manipulation of information at each time step, LSTMs have enabled significant advances in the processing and understanding of sequences.

2.4.2.4 Generative Adversarial Networks

GANs consist of two neural networks, a generator and a discriminator, that are trained simultaneously through adversarial processes. The core idea behind GANs is to generate new data samples that are indistinguishable from real data, which has led to their widespread use in applications such as image generation, video synthesis, and data augmentation.

The GAN framework involves two primary components: the generator, which tries to create realistic data samples, and the discriminator, which attempts to distinguish between real and generated (fake) data. These two components are set up in a game-theoretic scenario, where the generator aims to fool the discriminator, and the discriminator strives to accurately identify real versus fake data.

Mathematical Formulation

The objective of GANs can be formulated as a minimax optimization problem. The generator, denoted by G, maps a random noise vector $\mathbf{z}$ from a prior distribution $p_{\mathbf{z}}(\mathbf{z})$ to the data space. The discriminator, denoted by D, outputs a probability that a given sample $\mathbf{x}$ is real (i.e., from the true data distribution $p_{\text{data}}(\mathbf{x})$) rather than generated by G.

The minimax game between the generator and discriminator is expressed as:

$$\min_G \max_D \mathbb{E}_{\mathbf{x}\sim p_{\text{data}}(\mathbf{x})}[\log D(\mathbf{x})] + \mathbb{E}_{\mathbf{z}\sim p_{\mathbf{z}}(\mathbf{z})}[\log(1 - D(G(\mathbf{z})))] \tag{2.31}$$

In this equation, $D(\mathbf{x})$ represents the discriminator's estimate of the probability that $\mathbf{x}$ is a real sample, while $G(\mathbf{z})$ denotes the generator's output when given a noise vector $\mathbf{z}$. The generator aims to minimize the term $\log(1 - D(G(\mathbf{z})))$, which corresponds to the discriminator identifying the generated sample as fake. Simultaneously, the discriminator seeks to maximize its accuracy in classifying real versus fake samples.

The training process involves iteratively updating the parameters of G and D. First, the discriminator is trained to maximize the probability of correctly classifying real and fake samples. Then, the generator is trained to produce samples that are more likely to fool the discriminator. This process is repeated until the generator produces samples that the discriminator can no longer distinguish from real data.

3 Early Crop Mapping Using Dynamic Ecoregion Clustering: A USA-wide Study

Mapping target crops earlier than the harvest period is an essential task for improving agricultural productivity and decision-making. This chapter presents a new method for early crop mapping for the entire conterminous USA (CONUS) land area using the Normalized Difference Vegetation Index (NDVI) and Enhanced Vegetation Index (EVI) data with a dynamic ecoregion clustering approach. Ecoregions, geographically distinct areas with unique ecological patterns and processes, provide a valuable framework for large-scale crop mapping. We conducted our dynamic ecoregion clustering by analyzing soil, climate, elevation, and slope data. This analysis facilitated the division of the cropland area within the CONUS into distinct ecoregions. Unlike static ecoregion clustering, which generates a single ecoregion map that remains unchanged over time, our dynamic ecoregion approach produces a unique ecoregion map for each year. This dynamic approach enables us to consider the year-to-year climate variations that significantly impact crop growth, enhancing the accuracy of our crop mapping process. Subsequently, a Random Forest classifier was employed to train individual models for each ecoregion. These models were trained using the time-series MODIS (Moderate Resolution Imaging Spectroradiometer) 250-m NDVI and EVI data retrieved from Google Earth Engine, covering the crop growth periods spanning from 2013 to 2017, and evaluated from 2018 to 2022. Ground truth data was sourced from the US Department of Agriculture's (USDA) Cropland Data Layer (CDL) products. The evaluation results showed that the dynamic clustering method achieved higher accuracy than the static clustering method in early crop mapping in the entire CONUS. This study's findings can be helpful for improving crop management and decision-making for agricultural activities by providing early and accurate crop mapping.

This chapter is based on the work published in the following research paper:

- **Wang, Yiqun**, Hui Huang, and Radu State. "Early Crop Mapping Using Dynamic Ecoregion Clustering: A USA-Wide Study." Remote Sensing 15.20 (2023): 4962.

Contents

3.1 Overview

Mapping target crops earlier than the harvest period is an essential task for improving agricultural productivity and decision-making. Early crop mapping provides valuable information for crop management, such as predicting yield [43], monitoring crop growth [44, 45], and identifying areas with high production potential [46]. In recent years, the application of remote sensing techniques to early crop mapping has gained widespread popularity, owing to its inherent advantages of non-invasiveness and rapid data acquisition. By utilizing multispectral imagery (MSI) captured by satellites like Landsat-8 [2, 3, 4], Sentinel-2 [4, 5], and MODIS [3, 6, 7], valuable insights into crop health, vegetation indices, and land cover classification can be obtained. These datasets play a pivotal role in assessing crop vigour, identifying stress factors, and mapping crop types, and ultimately facilitating more informed agricultural decision-making.

Among the remote sensing techniques used in early crop mapping, the NDVI and EVI have emerged as widely employed indicators for monitoring vegetation growth and identifying different crop types. NDVI quantifies the greenness of vegetation based on the difference between near-infrared and red spectral bands, allowing for the detection of vegetation density and health. Similarly, EVI incorporates additional spectral bands to mitigate the influence of atmospheric and canopy background effects, providing enhanced accuracy in characterizing vegetation dynamics. Numerous studies have been conducted to develop and improve early crop mapping methods using NDVI and EVI data. For instance, both the harmonized time-series NDVI and EVI from Landsat-8 and Sentinel-2 data are used with the decision tree method for crop identification [8]. A time series analysis of MODIS NDVI data is utilized to map crop types for the US central great plains [6]. They applied an unsupervised classification method (ISODATA) to the 15-date NDVI time series to produce the crop or non-crop map. Similarly, [47] used a multi-temporal MODIS NDVI approach to map soybean in the USA for 2015. Furthermore, [48] proves that the general crop maps produced using the MODIS EVI and NDVI data both had very high overall (97.0%) and class-specific user's and producer's accuracies (ranging from 95% to 100%) for a case study for southwest Kansas.

However, accurately mapping crops at a large scale, such as for the entire CONUS, is challenging due to the heterogeneous nature of the crop-growing environment. This challenge arises from the intricate interactions among soil type, climate, and topography, which significantly influence the success of crop mapping [49, 50, 51, 11, 12, 7]. The crop-growing environment displays substantial spatial and temporal heterogeneity. Soil types vary across regions, impacting vegetation growth and the reflectance captured by remote sensing data [50, 12, 7]. Additionally, climate conditions, including temperature, precipitation, evaporation, wind speed, rainfall, and solar radiation, exhibit considerable variations, affecting crop phenology and productivity [49, 50, 51, 11, 12, 7]. Moreover, topographic features, such as slope, aspect, and elevation, play a vital role in determining crop growth patterns by influencing factors like solar radiation distribution, water availability, and wind patterns [50, 12, 7]. The failure to account for these influential factors in the crop mapping process can lead to inaccurate results, particularly when applied to large-scale areas. To address this issue, the concept of ecoregions was introduced. Ecoregions, which are geographically distinct areas with unique ecological patterns

and processes, provide a valuable framework for large-scale crop mapping. By integrating knowledge of ecological characteristics within each ecoregion, such as climate, geology, and topography, it becomes possible to identify suitable crop-growing areas and predict crop distribution based on shared environmental characteristics. By leveraging the ecoregional context and incorporating factors like climate, soil types, topography, crop mapping efforts can be tailored to effectively account for the heterogeneity of the crop-growing environment, leading to improved mapping accuracy at a large scale. Therefore, ecoregion clustering methods have been developed to address this issue [52]. And the static ecoregion clustering method [53] was used for crop mapping by [7]. The ecoregion clustering method involves analyzing environmental factors, such as soil type, climate, and topography, to divide the study area into multiple ecoregions. Each ecoregion has its own unique characteristics that affect crop growth, and these characteristics are taken into account during classification. This approach has been shown to improve crop mapping accuracy compared to country and state-level mapping [7].

Nevertheless, aside from the large-scale problem, the inter-year challenge is also of great significance for precise crop mapping. The climate variation between different years within a given region is anticipated to significantly influence the patterns of Vegetation Indices (VIs) related to crop growth and, consequently, crop mapping outcomes. The static ecoregion clustering method only provides a single ecoregion map for the CONUS. They proceed based on the assumption that the ecoregion remains unchanged over different years.

In order to address this limitation, our study considers the fluctuations in climate and introduces a dynamic ecoregion clustering approach. In this chapter, we present a novel approach for mapping target crops (soybean and corn) earlier than the harvest period in the CONUS using time-series MODIS 250-m NDVI and EVI data from Google Earth Engine with a dynamic ecoregion clustering method. The dynamic ecoregion clustering method analyzes sand, climate, elevation, and slope data to divide the entire cropland areas of the CONUS into multiple ecoregions. It presents a clear advantage over the static approach as it produces a unique ecoregion map for each year in the region of interest. This is particularly important considering the year-to-year variations in climate, which directly impact crop growth and vegetation indices patterns. A Random Forest classifier was then trained for each different ecoregion to classify the target crop. The results showed that the dynamic clustering method achieved significantly higher accuracy than the static clustering method. Specifically, for soybean mapping in 2018 in the entire CONUS, we observed an increase in User's Accuracy from 59.04% to 62.74%, representing a substantial improvement of 3.7%.

Our contributions can be summarised as follows:

- We propose a novel approach for mapping target crops (soybean and corn) earlier than the harvest period in the USA using time-series NDVI and EVI data with a dynamic ecoregion clustering method.
- We employ both the elbow method and silhouette method to ascertain the optimal number of ecoregions. Subsequently, we train an ecoregion clustering model using the Kmeans++ method, which allows us to generate ecoregion maps spanning the years 2013 to 2022, covering the entire cropland region within the CONUS.
- We demonstrate significantly higher mapping accuracy using the dynamic

clustering method compared to the static clustering method.

3.2 Related Works

Various classification algorithms, including unsupervised methods like k-means clustering and supervised methods like decision trees and deep learning approaches, have been utilized for crop classification at small scales using Landsat-8 and Sentinel-2A data. One study [50] focuses on utilizing the Random Forest algorithm for mapping and predicting rice yield through analysis of Sentinel-2 satellite data. Another research [8] effort concentrates on creating a high-resolution crop intensity mapping methodology by integrating data from Landsat-8 and Sentinel-2 satellites using a random forest algorithm. Furthermore, a hybrid deep-learning architecture called CerealNet [9] has been introduced for the specific purpose of cereal crop mapping, utilizing Sentinel-2 time-series data. However, [9] has limitations in its scope, as it specifically examines a research region characterized by a hot Mediterranean climate with dry summers. This region selection addresses a common challenge posed by cloud cover in time-series data analysis of Landsat-8 and Sentinel-2 images. Cloud cover plays a crucial role in crop mapping by impacting the availability and quality of remote sensing data. The presence of clouds obstructs satellite imagery, leading to the loss or concealment of vital information about the Earth's surface. Consequently, these cloud-induced data gaps undermine the accuracy and reliability of crop mapping results. The negative impact of cloud cover on crop mapping becomes even more pronounced when considering large-scale applications, such as mapping crops across the entire CONUS. The expansive coverage of such regions makes them more susceptible to varying cloud cover patterns, resulting in extensive areas with missing or incomplete data. This increases the risk of biased or inaccurate crop classification, making it challenging to obtain a comprehensive and reliable understanding of crop distribution and dynamics at a broader scale.

In order to overcome this limitation of Landsat-8 and Sentinel-2 images, some works focus on using time-series MODIS data for early crop mapping. MODIS data can overcome the limitations imposed by cloud cover on crop mapping due to its unique capabilities. Its moderate spatial resolution allows for wider coverage and reduces the impact of cloud cover, while its frequent revisit time ensures more opportunities to capture cloud-free images, enabling a more consistent and reliable monitoring of crop patterns at regional or global scales. Previous research [54, 10] has shown that individual major crops, such as corn and soybeans, can be mapped accurately as early as July and August using MODIS dataset.

However, these studies were also limited in scope, focusing on counties, states, or groups of states. A crop mapping model that is limited in scope cannot be directly applied to a large-scale setting due to spatial differences. The presence of spatial heterogeneity, encompassing diverse factors like soil types, climate, topography, and other environmental elements, significantly impacts crop growth patterns and diminishes the model's accuracy outside of its intended region. Overcoming this challenge necessitates the integration of spatially explicit information into the method, effectively capturing and using the spatial variations and encompassing the diverse conditions present across the target area. There are some works [51, 11], that use Growing-degree-day (GDD), which is a valuable metric that quantifies the accumulated heat necessary for the growth and development of vegetation. Its significance lies in its crop-specific nature, as the magnitude of GDD required during different growing stages varies for each crop. This characteristic makes GDD a

Table 3.1: The Methods of Related Work.

Factors taken into consideration	**Method**	**Dataset**	**Vegetation Indices**	**Research Region**
Soil, Climate, and Topography Conditions	[50]	Sentinel-2	NDVI, EVI, LSWI, RGVI2	Shanwei, Guangzhou(China)
-	[8]	Landsat-8 and Sentinel-2	NDVI, EVI	Four subregions in Kansas(USA), Gaoteng(South Africa), Punjab(India), Shandong(China)
-	[9]	Sentinel-2	Spectral bands, NDVI	Gharb plain(northwest of Morocco)
-	[54]	MODIS	Phenological and other variables	22 States in the USA
-	[10]	MODIS	NDVI, GCVI, EVI, LSWI	Champaign County, Illinois(USA)
GDD, Precipitation	[51]	Landsat-7	Spectral bands	Doniphan County, Kansas(USA)
GDD	[11]	MODIS	NDVI	Kansas (USA)
Precipitation, Soil conditions (within administrative demarcations)	[12]	MODIS	NDVI	The entire CONUS
Soil, Climate, and Topography Conditions	[7]	MODIS	NDVI	The entire CONUS

valuable tool in crop classification, as it provides insights into the progression and timing of crop growth. By considering the crop-specific GDD thresholds for various stages, it becomes possible to leverage this metric for accurate and effective crop classification and monitoring. Other studies have performed crop classifications at smaller administrative units such as Agriculture Statistics Counties and Districts [55], states [54], or Agroecological zones [12]. Nonetheless, these methodologies either disregard fluctuations in precipitation and soil characteristics or are conducted within administrative or political demarcations that lack relevance to crop phenology. Alternatively, they encompass areas of such magnitude that they fail to encompass the nuanced phenological fluctuations driven by climatic variations.

An ideal approach would be to model regions based on environmental variables that reflect crop growing conditions and are of small size, created using quantitative analytical methods that are both empirical and reproducible. Multivariate Geographic Clustering (MGC) algorithms [52] and Multivariate spatio-temporal clustering (MSTC) [53, 7] have been successfully used to create ecoregions that exist within similar combinations of ecologically relevant conditions such as temperature, precipitation, soil, and topographic properties on a map.

However, an identified limitation of the study [7] is that they assume that the ecoregion boundaries remain constant, which may not accurately represent the true variability. In addition, they solely present a single 500-ecoregions map directly utilizing MSTC without incorporating any decision-making process to determine the optimal number of ecoregions. In order to overcome these limitations, our proposed method introduces a dynamic ecoregion map that adjusts the boundaries based on climate data specific to each year's crop growing season. Additionally, we employ the elbow method to ascertain the optimal number of ecoregions. This dynamic approach ensures a more precise and up-to-date understanding of crop phenology across diverse regions by adapting the ecoregion boundaries to reflect the evolving environmental conditions for each growing season. Ultimately, in conjunction with the crop classifier, this approach enhances the accuracy of early crop mapping results.

3.3 Data and methods

In this section, we present the study area and data in Section 5.3.1, the system overview in Section 4.4.2, the development of the dynamic ecoregions in Section 3.3.3, and the training and evaluation methods of the crop classification model in Section 3.3.4.

3.3.1 Study area and data

In this chapter, we focus on mapping soybean and corn across the entire cropland area of the CONUS.

To achieve this, we split the cropland area into several ecoregions using soil, climate, elevation, and slope data as environmental data. The cropland area can be precisely delineated by the cultivated layer within the CDL, a crop-specific land cover raster map dataset available for the entire CONUS at 30 m resolution provided by the USDA [13]. As shown in Table 3.2, our soil data with a resolution of 250m*250m was obtained from the ISRIC SoilGrids Dataset [56] and included parameters such as bulk density, clay particle proportion, total nitrogen, soil pH, sand particle proportion, silt particle proportion, and soil organic carbon content for 2015. We utilize ERA5, the 5th major atmospheric reanalysis produced by ECMWF [57], with a resolution of 11132m*11132m from 2013 to 2022 as the climate data which is comprised of air and soil temperature, precipitation, evaporation, wind speed, solar radiation, runoff, and soil water volume. We also incorporated elevation data with a resolution of 231.92m*231.92m from the GMTED2010 Dataset [58] for 2010 and calculated the slope based on the elevation.

Furthermore, we use remote sensing data as the classification input data. We extract time-series MODIS 250-meter NDVI and EVI data from Google Earth Engine [59]. The data is captured at a 16-day interval, covering the growing seasons between 2013 and 2022. Our objective is to locate the target crops, namely corn and soybeans, at an earlier stage in the entire CONUS. Corn harvest commences on September 1st, while soybean harvest begins on October 1st. Consequently, each year, we collect VIs data from April 1st to mid-July, encompassing a total of seven-time points. This approach accounts for the 16-day temporal frequency of the MODIS data and aligns with the vegetation growth patterns during this period. The CDL data was used as the ground truth for the classification training and evaluation process. Our crop classifiers were trained over 2013–2017 and applied to the period 2018–2022 as test years.

3.3.2 System Overview

The system architecture of our early crop mapping system with dynamic ecoregion clustering method is shown in Figure 4.6. In summary, our proposed ecoregion clustering algorithm fetches and standardizes soil, climate, elevation, and slope data from various sources to build a clustering model using the K-means++ method [60]. This model is trained exclusively on target cropland regions from 2013, resulting in a well-trained ecoregion clusterer. Using this clusterer, we are able to partition the complete cropland area expanse into multiple distinct ecoregions, a process repeated annually within the timeframe spanning 2013 to 2022. With the provided ecoregion maps, we create a specific early crop mapping classifier using the Random Forest method for the same ecoregion. These classifiers utilize time-series 250m-resolution MODIS NDVI and EVI data from 2013-2017 to predict early crop mapping for

Table 3.2: The Metrics for Ecoregion Clustering.

Conditions	**Metrics**	**Resolution (m)**	**Dataset**
Soil Conditions	pH, Soil organic carbon content, Bulk density, Sand content, Silt content, Clay content, Total nitrogen, Organic carbon density	250	ISRIC SoilGrids Dataset
Climate Conditions	Temperature of air at 2m above the surface of land, Amount of solar radiation, Total evaporation, Total precipitation, Eastward component of the 10m wind, Northward component of the 10m wind.	11132	ERA5-Land Dataset
Topography Conditions	Elevation, Slope	231.92	GMTED2010 Dataset

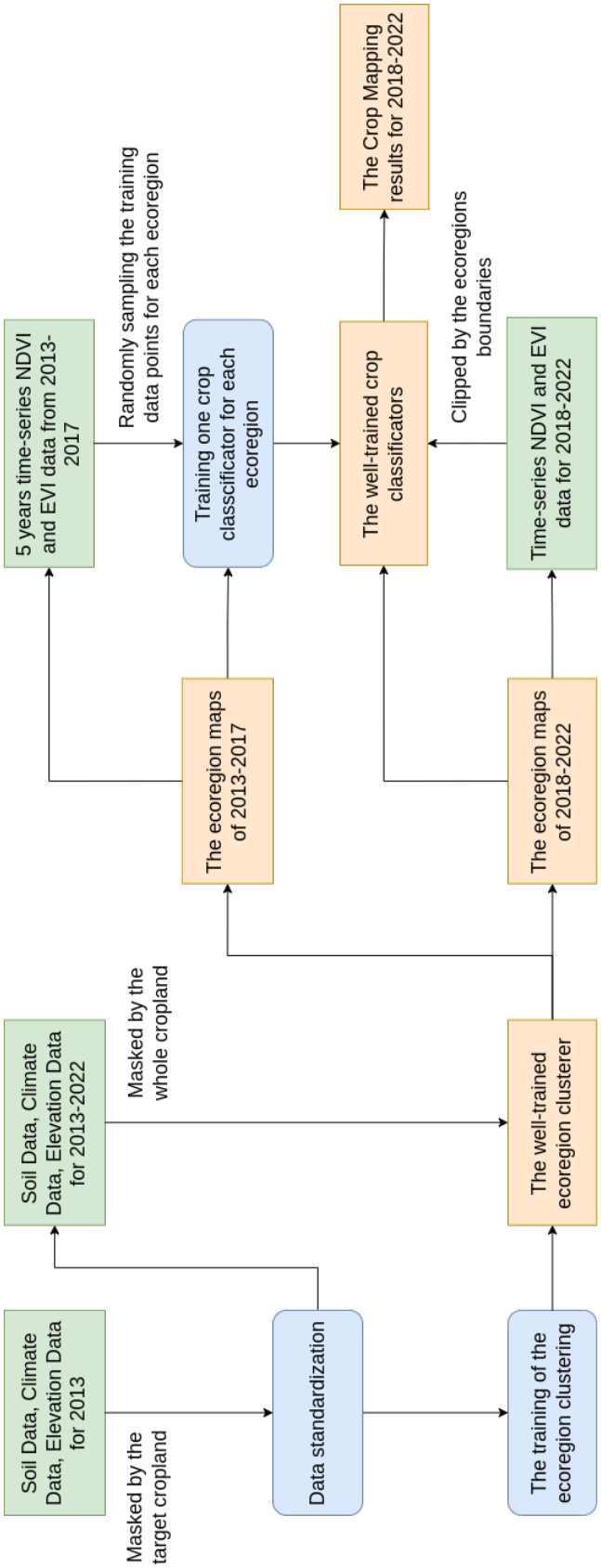


Figure 3.1: The system architecture comprises a dynamic ecoregion clustering method and a specific crop classifier for each ecoregion. Inputs to the system are represented by green blocks, system outputs are represented by orange blocks, and system processes are depicted in blue blocks.

2018-2022, incorporating dynamic ecoregion maps for each year. We conducted a random sampling of NDVI and EVI data points within each ecoregion, spanning from April 1st to mid-July, to gather our training data. This timeframe aligns with our objective of early locating target crops such as corn and soybean, as they are typically harvested starting from September 1st. We merge the training data points from multiple training years to train a dedicated crop mapping classifier for each ecoregion. Finally, we applied each classifier to the corresponding ecoregion and mosaiced the crop mapping results from all ecoregions to obtain the final crop mapping outcome.

3.3.3 Development of dynamic ecoregions

Our study focuses on accurately mapping the growing area of soybean and corn crops before their harvest periods throughout the entire cropland area of the CONUS. To achieve this goal, we utilize a dynamic ecoregion clustering method that relies on soil, climate, elevation, and slope data from various sources. Due to the absence of multi-year soil and topography data, our assumption is that soil quality, elevation, and slope conditions remain relatively stable over time, while climate conditions exhibit variability. Various methods exist for ecoregion clustering, with hierarchical clustering [61] being one such approach known for accommodating datasets with nested or hierarchical structures. However, for our specific scenario, we prefer to adopt a simpler clustering method. This decision is driven by the intention to maintain simplicity in our current approach while preserving compatibility for future integration of multi-year soil and topography data. Consequently, we have chosen a direct clustering method for ecoregion clustering, with the expectation that it will facilitate a smoother transition when more comprehensive data becomes available in the future.

Each year, we initiate the process by reprojecting the data into a 10000-meter resolution. In the case of soil data, we compute the average metric value by consolidating values from various depths. When dealing with climate data, we determine both the mean and variance for each metric within the specified crop growth period. Notably, when it comes to wind speed, we combine the eastward and northward components to create a total wind speed, discarding the directional information. Then we randomly sample 10000 points from the target crop region as the training data. Subsequently, we employ the Principal Component Analysis (PCA) method to reduce the data dimensions from 20 (8 for soil conditions, 10 for climate conditions, and 2 for topography conditions) down to 5. We trained our dynamic ecoregion clustering model using the K-means++ method and the pre-processed training data from 2013 and applied it to the entire cropland area, as determined by the cultivated layer of CDL, for the years 2013-2022.

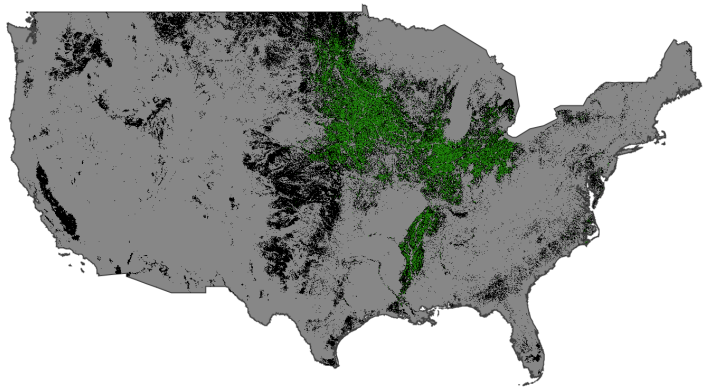
This allows us to identify regions with similar environmental characteristics, which is essential for accurate early crop mapping. In order to determine the optimal number of clusters, we utilize the elbow method to calculate the within-cluster sum of squares (WCSS) value using Equation (3.1):

$$WCSS = \sum_{i=1}^{n} \sum_{j=1}^{m_i} (x_{ij} - c_i)^2 \tag{3.1}$$

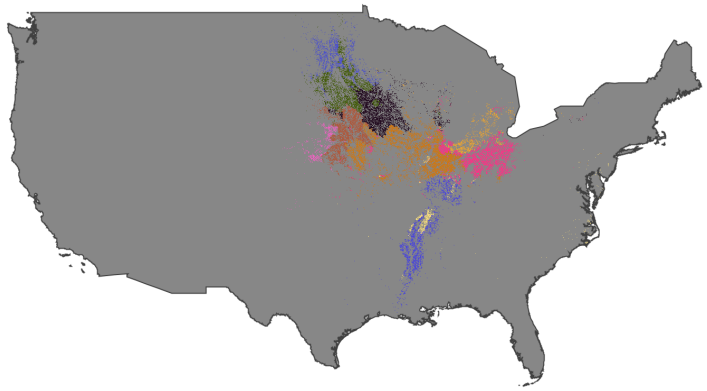
,where n denotes the total number of ecoregion clusters, m_i denotes the number of pixels contained within cluster i, x_{ij} denotes the vector representing each pixel within cluster i, and c_i represents the centroid vector of cluster i. This allows us

to strike a balance between cluster granularity and computational efficiency. If the number of clusters is too small, the environmental similarities may not be unique enough for each cluster, compromising the accuracy of our results. Conversely, if the number of clusters is too large, it may result in excessive computational cost, hindering the practical application of our approach. By finding the optimal number of clusters, we are able to maximize the accuracy of our approach while ensuring its computational feasibility.

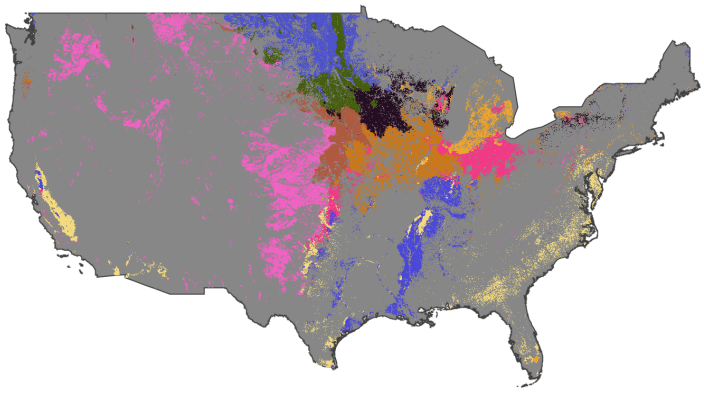
The resultant ecoregion maps with different ecoregion numbers, were then used in our crop classification training and testing processes. For instance, Figure 3.2a displays the entire cropland area of CONUS for the year 2013, as obtained from the CDL layer. We restrict our analysis to the green region that corresponds to the soybean cropland region. We determine the optimal number of ecoregions as 10, identified through the WCSS value. Subsequently, we employ our well-trained ecoregion clustering model to the entire cropland area to generate the ecoregion map for 2013, which is shown in Figure 3.2c. The ecoregion maps from 2014 to 2022 for soybean mapping are also shown in Figure 3.3.



(a)

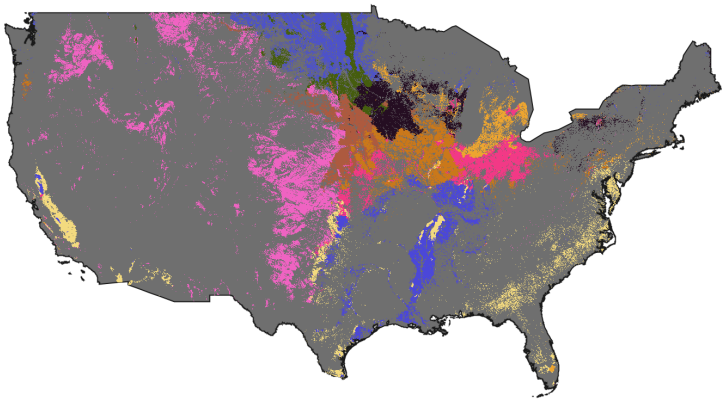


(b)

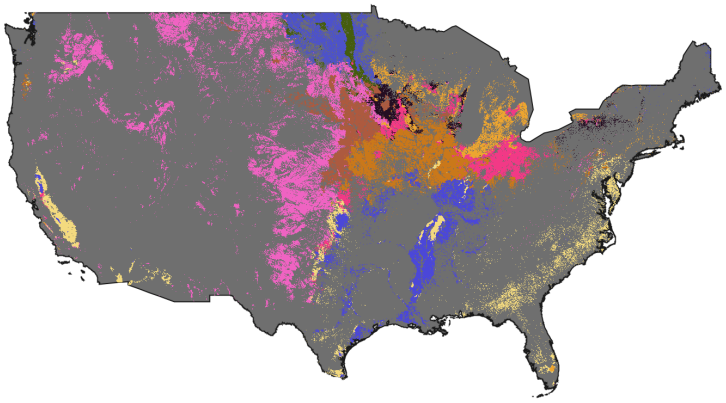


(c)

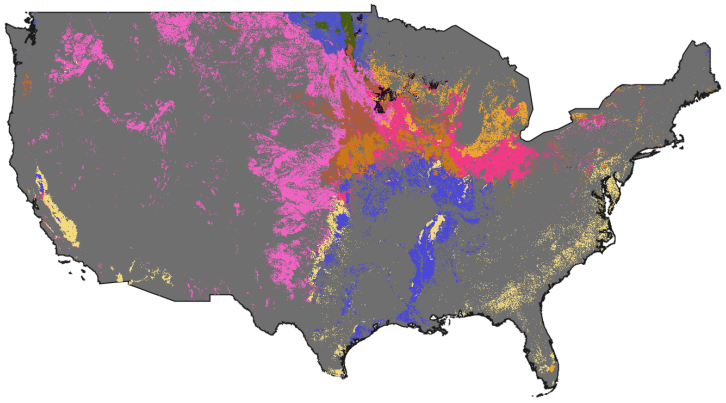
Figure 3.2: The ecoregion map with 10 clusters of 2013 for soybean mapping. **(a)** shows the soybean cropland region filtered out by the CDL cultivated layer of 2013, where the green color represents the soybean cropland, and the black color represents the non-soybean cropland. **(b)** displays the ecoregion map of the soybean cropland region, where each color represents a specific ecoregion. **(c)** shows the ecoregion map of the entire cropland region, where each color represents a specific ecoregion.



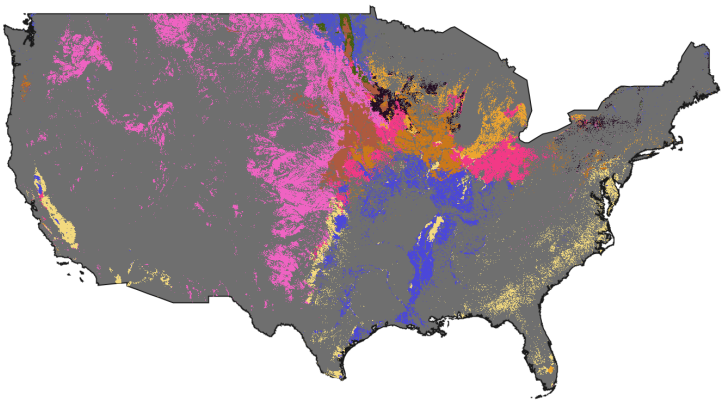
(a) The ecoregion map of 2014



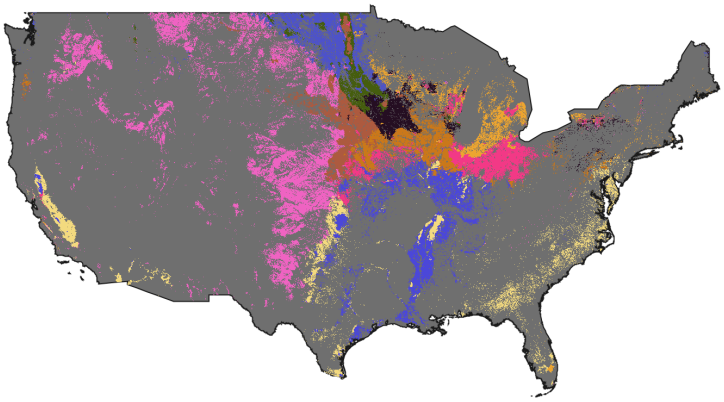
(b) The ecoregion map of 2015



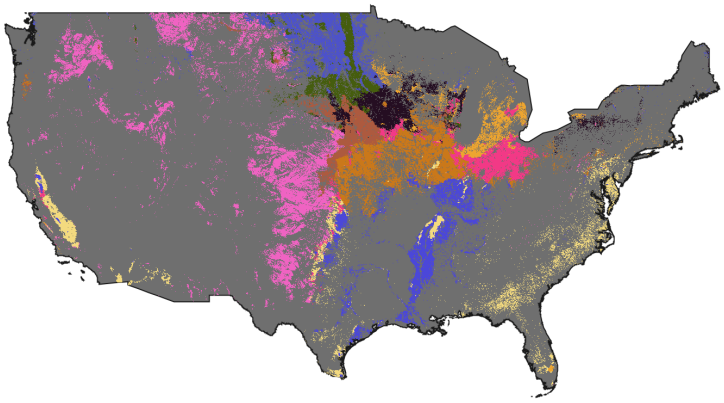
(c) The ecoregion map of 2016



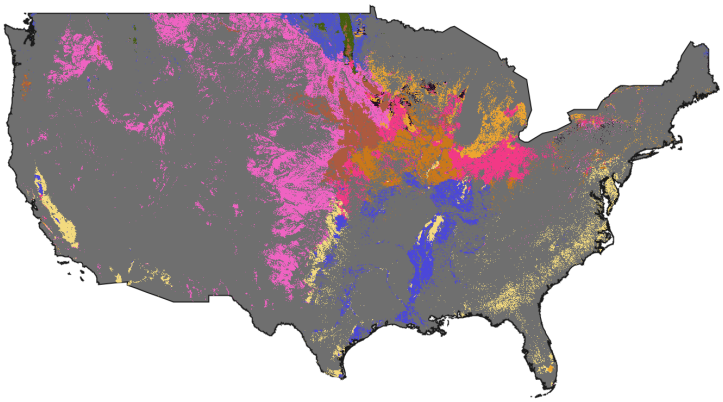
(d) The ecoregion map of 2017



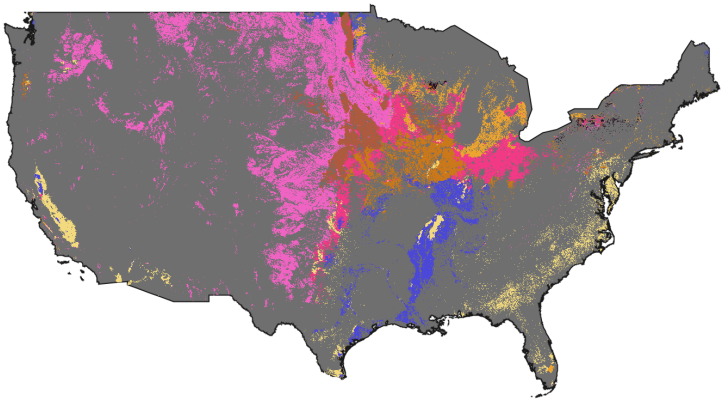
(e) The ecoregion map of 2018



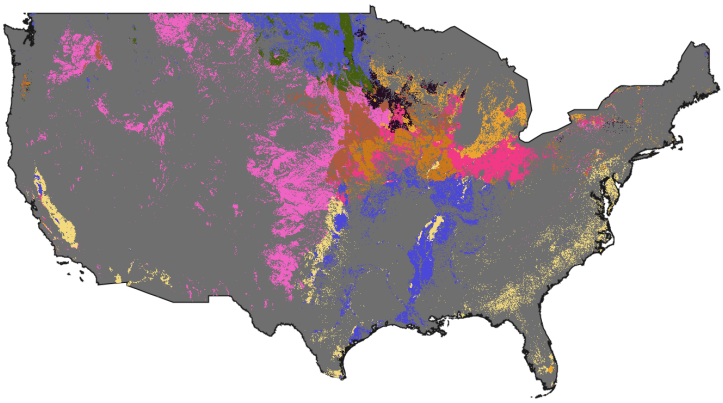
(f) The ecoregion map of 2019



(g) The ecoregion map of 2020



(h) The ecoregion map of 2021



(i) The ecoregion map of 2022

Figure 3.3: The dynamic ecoregions maps with 10 clusters for 2014-2022.

Table 3.3: The MMD values between the ecoregion 1 and each ecoregion.

ER	1	2	3	4	5	6	7	8	9	10
MMD (10^{-3})	0.13	222.88	95.96	114.54	29.14	56.31	186.52	52.58	80.69	92.94

To illustrate the variation in time-series vegetation indices of our target crop, we present the time-series average NDVI and EVI curves of soybean for each ecoregion based on ten clusters in 2013 in Figure 3.4. It is evident that the patterns of these features exhibit significant differences across different ecoregions. To demonstrate distinctions among the distributions of VIs data across various ecoregions, we employ the Maximum Mean Discrepancy (MMD) as a measure. MMD allows us to measure the discrepancy between the distributions of VIs from different ecoregions, providing a quantitative measure of the similarity. Specifically, we compute MMD values between ecoregion 1 and all other ecoregions in our analysis. To ensure statistical robustness, we perform the following steps:

- We randomly sample 20,000 data points from ecoregion 1.
- For ecoregions 2-10, we also draw random samples of 10,000 data points each.
- For ecoregion 1, we calculate the MMD by comparing two subsets of 10,000 data points each, where the first subset consists of the initial 10,000 data points, and the second subset comprises the last 10,000 data points.
- For the other ecoregions (ecoregions 2-10), we compute the MMD by comparing the initial 10,000 data points from ecoregion 1 with the 10,000 data points from each of the other ecoregions.

The resulting MMD values are summarized in the Table 3.3. The MMD within ecoregion 1 is notably lower than the MMD between ecoregion 1 and the other ecoregions, providing clear evidence of significant variations in VIs data distribution among different ecoregions.

Moreover, to highlight the varying patterns of VIs within a specific region across disparate ecoregions over different years, we have designated two distinct areas, which changed the ecoregion types between 2014 and 2021, as illustrated in Figure 3.5. The distribution of soybean VIs data is then presented using t-distributed stochastic neighbour embedding (t-SNE) [62] in Figure 3.6. Evident disparities in these distributions between the years 2014 and 2021 are discernible, underscoring the fluctuating nature of soybean growth-related VIs data within these designated regions.

3.3.4 Crop classification model

Once the ecoregions are established, a specific early crop mapping classifier can be trained for each different ecoregion based on the cropland region, which is filtered out by the cultivated layer of the CDL. Random Forest classification models have become a popular tool for mapping land cover due to their flexibility in handling nonlinear relationships between input features and class membership, and their intuitive decision rules. During training, the tree grows by recursively partitioning the data into less heterogeneous groups until the desired level of accuracy or purity is achieved. The final model is constructed when all leaf nodes are generated, with each leaf node representing either a pure class or a mixture of two classes, determined

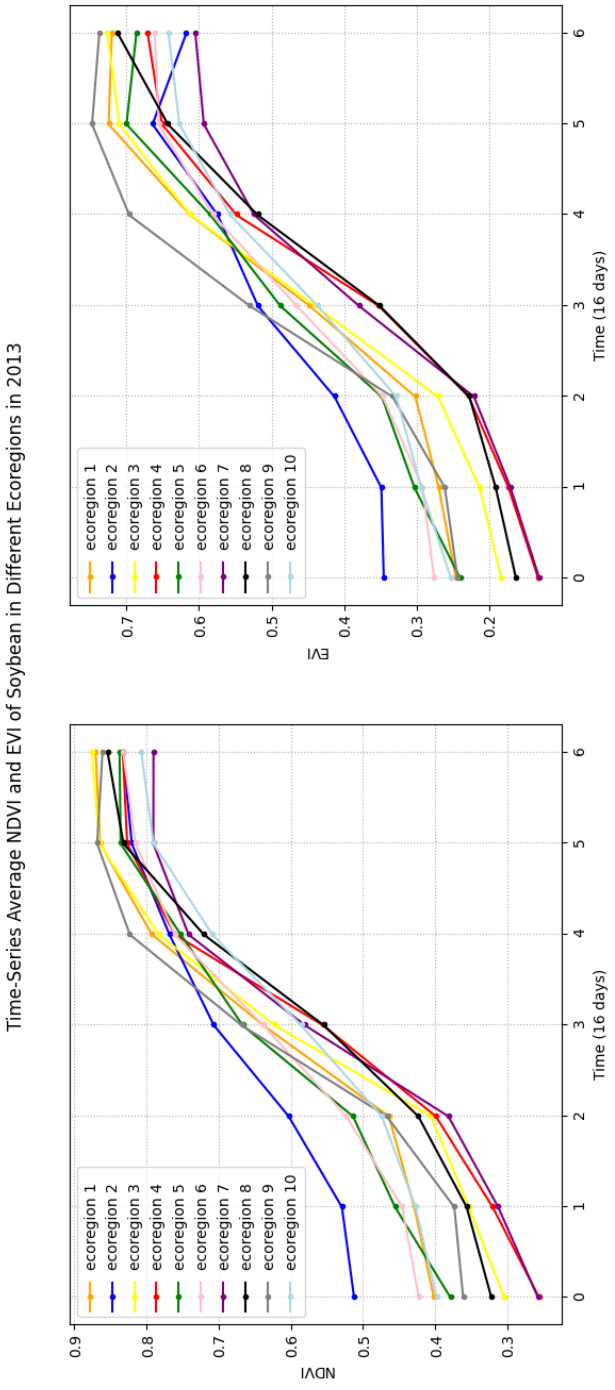
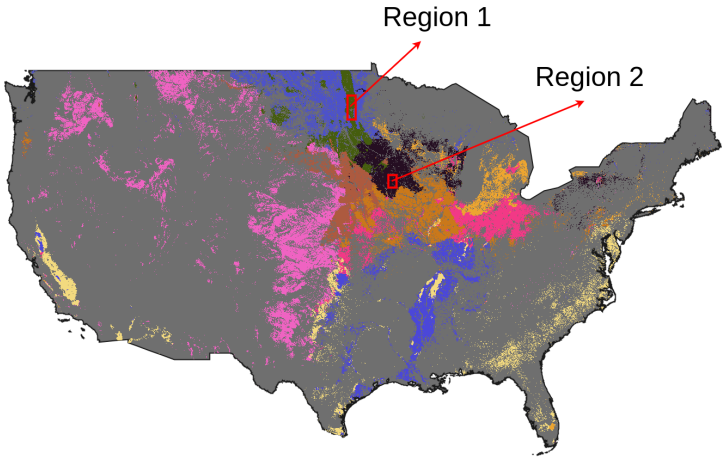
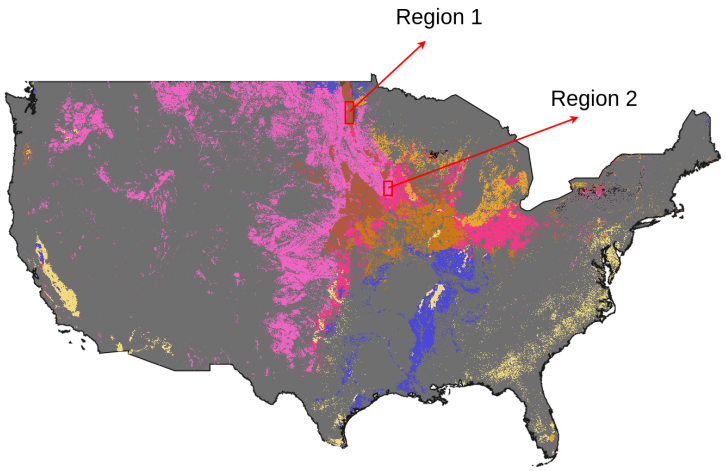


Figure 3.4: Time-Series Average NDVI and EVI of Soybean in Different Ecoregions in 2013. Left for NDVI. Right for EVI.

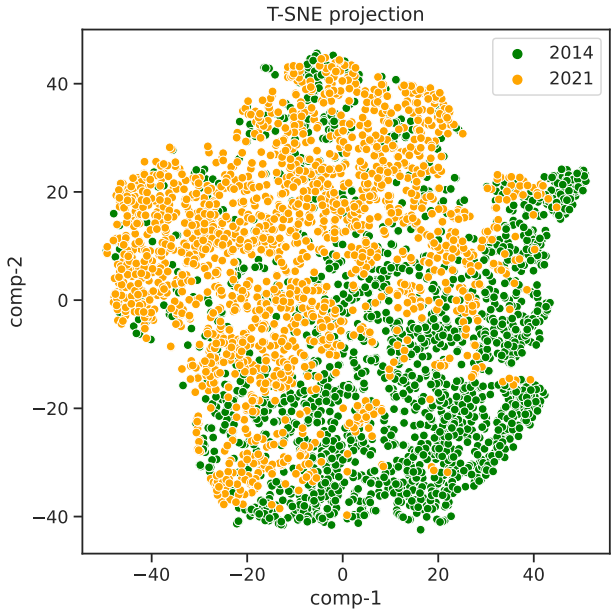


(a)

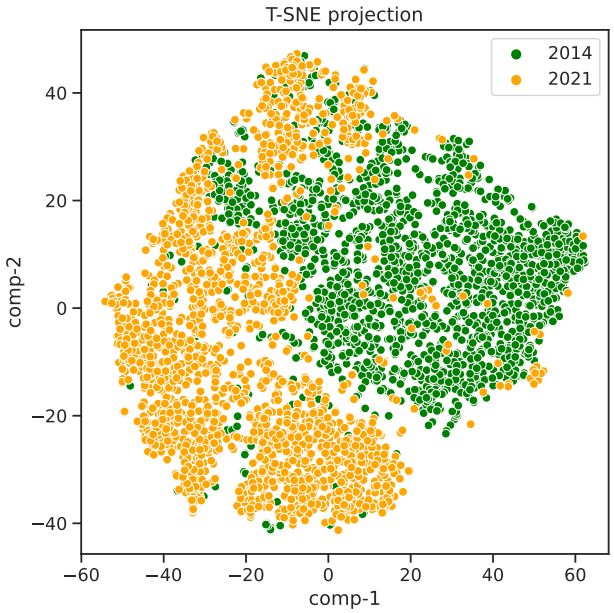


(b)

Figure 3.5: The Sample Regions. **(a)** shows the sample regions with the 2014 ecoregion map. **(b)** shows the sample regions with the 2021 ecoregion map.



(a)



(b)

Figure 3.6: The soybean VIs data distribution. The x- and y-axes represent the features following dimension reduction using PCA. **(a)** shows the soybean VIs data distribution in region 1. **(b)** shows the soybean VIs data distribution in region 2.

by the proportion of training pixels in the node. However, these models are prone to overfitting, which can lead to poor generalization performance. To address this issue, we applied a bagging procedure and trained a random forest consisting of 50 classification tree models. Using a binary target-crop and non-target-crop training dataset, we locate our target crop across the corresponding ecoregion with the well-trained classifier. This approach significantly improved the stability and prediction accuracy of the model.

Since the CDL with the cultivated layer has been available only since 2013, the model was trained and validated in the years 2013–2017 and tested independently in the years 2018–2022.

Model training within the training period:

As per the USDA crop calendar, the seeding for corn and soybean typically begins after April 1st. Hence, we extracted time-series MODIS 250-m NDVI and EVI images with a 16-day temporal frequency from April 1st until the middle of July for 2013-2017. For each year, we compose these VIs images into one image as the composed image. Each composed image comprised seven NDVI bands and seven EVI bands. For each training year, we randomly selected 10,000 training sample points from the composed image for each different ecoregion, with half for target-crop points and half for non-target-crop points. Then we merge all the training sample points for the same ecoregion to train the ecoregion-specific crop classification model.

Model evaluation within the test period:

In order to evaluate our crop classification model, we first extracted the VIs images and provided the composed images for the years 2018-2022. To assess the effectiveness of our approach, we designed an experiment to compare the results of using a dynamic ecoregions map versus a static one. For the static map, we sample the training points from the 2013-2017 period and classify the target crops for 2018-2022 only using the ecoregion map from 2013, as shown in Figure 3.2c, with fixed boundaries for each ecoregion. In contrast, for the dynamic map, we sample the training points from the 2013-2017 period based on the ecoregion maps for 2013-2017, as shown in Figures 3.2c, 3.3a, 3.3b, 3.3c, 3.3d, and classify the target crops for 2018-2022 based on each year's specific ecoregion map, as shown in Figures 3.3e, 3.3f, 3.3g, 3.3h, 3.3i. The following results show that in most cases the dynamic model outperformed the static model. These findings demonstrate the benefits of using dynamic ecoregion maps in crop classification and highlight the importance of accounting for climate fluctuations when developing land cover maps.

Evaluation metrics:

To evaluate the accuracy of our classification, we used three metrics: Producer's Accuracy, User's Accuracy and Overall Accuracy, which are defined in equations (3.2), (3.3) and (3.4):

$$Producer's \quad Accuracy = \frac{TP}{TP + FN} \times 100\% \tag{3.2}$$

$$User's \quad Accuracy = \frac{TP}{TP + FP} \times 100\% \tag{3.3}$$

$$Overall \quad Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \times 100\% \tag{3.4}$$

, where TP, TN, FP, and FN refer to the numerical values in a confusion matrix that correspond to true positive (correctly predicted positive cases), true negative (correctly predicted negative cases), false positive (incorrectly predicted positive cases), and false negative (incorrectly predicted negative cases) outcomes, respectively.

The Producer's Accuracy represents the map's accuracy from the map producer's perspective, indicating the probability that a ground feature is correctly classified by the map. On the other hand, the User's Accuracy represents the map's reliability from the user's perspective, i.e., the probability that a feature on the map is actually present on the ground. The Overall Accuracy typically quantifies the fraction of all CDL pixels, encompassing both target crop and non-target crop pixels, that our crop classification method correctly identifies and maps.

3.4 Experiments and Results

3.4.1 Dynamic ecoregion maps

Determine the elbow point (cluster number):

Rather than directly dividing the CONUS into 500 clusters as in [7], we employed both the elbow method and silhouette method to determine the optimal number of clusters for our approach. This allowed us to balance cluster granularity with computational efficiency, ensuring both accuracy and practicality. Specifically, we calculated the environmental dissimilarities for different cluster numbers in the training year 2013, as shown in Figure 3.7 and 3.8. As depicted in Figure 3.7, we found that ten clusters were optimal, while the dissimilarity significantly decreased with a noticeable reduction starting at ten clusters for both soybean and corn cropland regions. Further insights gleaned from Figure 3.8 indicate that, for corn, a configuration of two clusters attains the highest silhouette score, while for soybean, this optimal number becomes to five.

As a result, to further evaluate the effectiveness of our approach, we conducted experiments with different cluster numbers, including two, five, and ten, and compare our resulting metrics with those of a static clustering method. By comparing these results, we can confirm the superiority of our dynamic clustering method and demonstrate its capability to identify the target crops accurately. Additionally, we prove that the mapping result with ten ecoregion clusters has the highest accuracy in most cases in the following experiments.

The visualization of dynamic ecoregion maps:

As depicted in Figure 3.2a, the green area corresponds to the soybean cropland region, whereas the black region represents non-soybean cropland from the year 2013. Notably, we only train the ecoregion clusterer using data from the target cropland region. The resulting ecoregion cluster map for the soybean cropland region is displayed in Figure 3.2b, while the ecoregion cluster map for the entire cropland region, encompassing both soybean and non-soybean croplands, is illustrated in Figure 3.2c.

Moreover, in Figure 3.3, we present the dynamic ecoregion maps for the years 2014-2022 for the soybean mapping task. Notably, the ecoregions in the north and middle of the CONUS exhibit significant changes over the years and especially in 2021, while the ecoregions in the southern region experience relatively less variation over time.

3.4.2 Comparison of mapping result metrics across the continental United States

Our experiment serves as a compelling validation of the effectiveness of our approach, as it demonstrates significantly improved accuracy for soybean and corn mapping across the CONUS in most cases when we compare our method with a static one.

3.4.2.1 Metrics for soybean mapping

In Table 3.4, we present a detailed comparison of the Producer's Accuracy and the User's Accuracy between our dynamic ecoregion method and the static ecoregion method. Notably, we observe consistent trends in the User's Accuracy across different

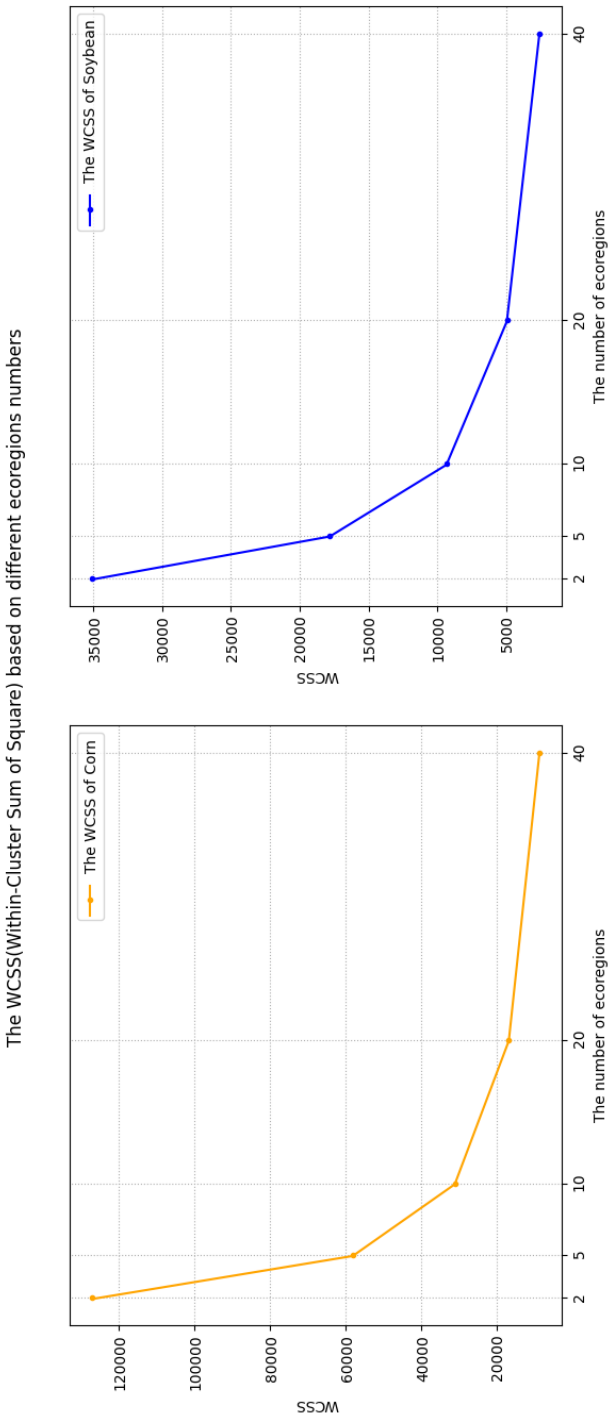


Figure 3.7: Within-cluster sum of squares (WCSS) of the environmental dissimilarity based on the pro-processed environmental data for different numbers of ecoregions in the corn cropland region (left) and soybean cropland region (right), during ecoregion clustering training in 2013.

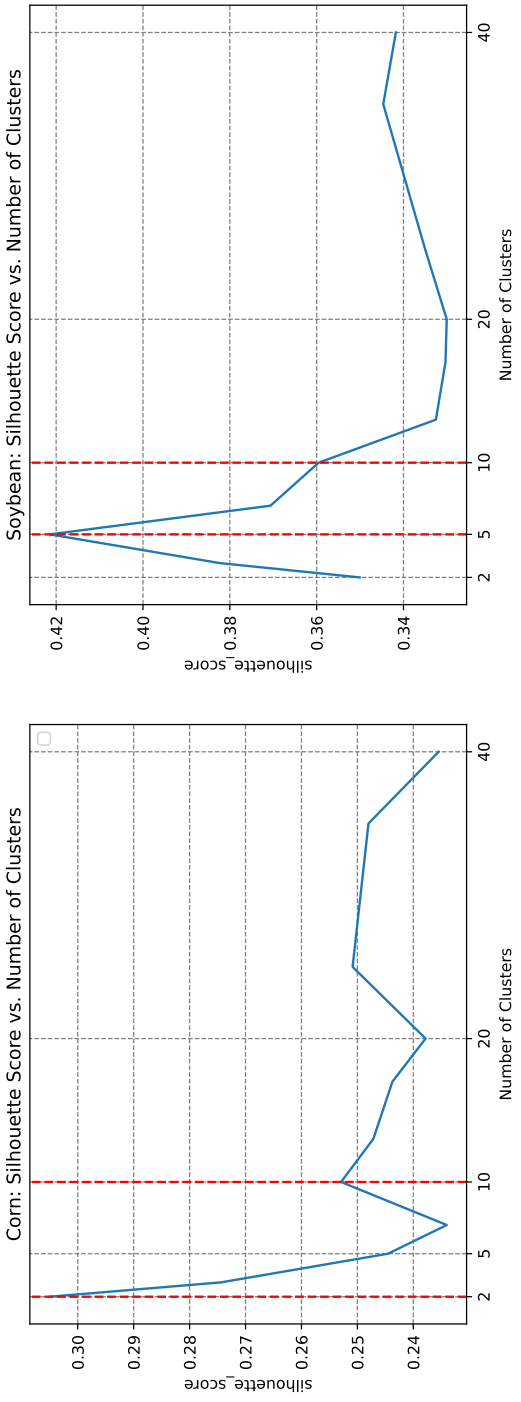


Figure 3.8: The silhouette score based on the pro-processed environmental data for different numbers of ecoregions in the corn cropland region (left) and soybean cropland region (right), during ecoregion clustering training in 2013.

Table 3.4: The experiment metrics for soybean early crop mapping based on different numbers of ecoregions.

Test year	Ecoregion number	Producer's Accuracy		User's Accuracy	
		dynamic	static	dynamic	static
2018	1	-	61.32	-	62.47
	2	60.61	59.74	67.52	67.01
	5	**62.74**	59.04	**67.96**	67.76
	10	61.40	58.85	67.94	67.50
2019	1	-	82.39	-	45.58
	2	83.97	83.86	48.35	48.06
	5	83.61	82.94	49.96	48.94
	10	**84.81**	83.85	**50.17**	49.35
2020	1	-	78.65	-	51.97
	2	78.30	77.89	56.09	56.05
	5	76.40	76.18	56.42	56.30
	10	**78.84**	76.92	56.47	**56.70**
2021	1	-	70.06	-	59.27
	2	70.67	70.04	**64.47**	63.36
	5	71.41	69.47	63.86	63.48
	10	**73.33**	69.87	63.94	63.54
2022	1	-	78.10	-	54.63
	2	79.83	78.79	57.55	56.99
	5	80.30	77.58	56.42	57.46
	10	**80.66**	78.80	**58.38**	58.21

years. With the utilization of ten different ecoregions, our dynamic ecoregion method achieves the highest Producer's Accuracy for soybean mapping in all years except for 2018. In that specific year, our method employing five different ecoregions yields the highest Producer's Accuracy. Furthermore, our analysis reveals that the most significant improvements in Producer's Accuracy using our dynamic ecoregion method occurred in 2018 and 2021. Specifically, in 2018, we observed an increase in Producer's Accuracy from 59.04% to 62.74%, representing a substantial improvement of 3.7%. Similarly, in 2021, our dynamic ecoregion method yielded an impressive improvement of 3.46%, increasing User's Accuracy from 69.87% to 73.33%. These findings emphasize the efficacy of our approach in accurately identifying and early mapping soybean crops across diverse environmental conditions.

Regarding the User's Accuracy, our dynamic ecoregion method with two, five and ten different ecoregions outperforms the static ecoregion method in most years. However, in 2020, the static ecoregion map exhibited better performance in terms of User's Accuracy. This exceptional result suggests that the static ecoregion map might offer advantages under specific circumstances, but overall, our dynamic ecoregion method consistently demonstrates superior accuracy. The mapping result with ten ecoregion clusters has the highest accuracy in most cases in the following experiments.

Table 3.5: The experiment metrics for corn early crop mapping only using NDVI. Dynamic and static method with 10 ecoregions. Normal presents the early crop mapping method without ecoregions

Test year	**Producer's Accuracy**			**User's Accuracy**		
	dynamic	static	normal	dynamic	static	normal
2018	57.94	54.79	**59.25**	**64.88**	64.63	58.76
2019	**84.06**	83.29	82.21	**48.38**	47.63	43.55
2020	79.29	77.15	**79.51**	53.02	**53.19**	48.56
2021	72.78	**77.15**	70.82	**58.66**	53.19	54.36
2022	**78.82**	75.92	76.82	55.88	**56.21**	52.30

Additionally, we conducted an experiment to compare early crop mapping results using both NDVI and EVI against using only NDVI. Figure 3.5 displays the Producer's Accuracy and User's Accuracy for the method exclusively utilizing NDVI. Notably, except for the year 2020, the Producer's Accuracy when employing NDVI and EVI is consistently higher than when using only NDVI. In most instances, the NDVI and EVI combination yields an approximately 2% improvement. Furthermore, when considering User's Accuracy across all years, the NDVI and EVI approach consistently outperforms the exclusive use of NDVI. Particularly, in 2021, the NDVI and EVI method achieves a remarkable 5.81% increase in accuracy compared to the NDVI-only approach. Consequently, in this study, we adopt the combined use of NDVI and EVI for our early crop mapping models.

Figure 3.9 presents our soybean early mapping result for the CONUS in 2021. To demonstrate the accuracy of our approach, we selected a sample test region and compared our soybean mapping result with 250m resolution to the soybean ground truth with 30m resolution from the CDL layer. Our result exhibits a broad-level spatial agreement with the CDL, indicating the effectiveness of our approach. However, due to the coarser resolution of MODIS products, our mapping result lacks sharpness and accuracy along the field boundaries. Furthermore, Figure 3.10a illustrates the pixel-wise overall accuracy using the dynamic method, where the colour scale indicates the accuracy, with greener colours representing higher overall accuracy. Figure 3.10 highlights the regions where the pixel-wise overall accuracy from the dynamic method outperforms the static method. It is evident that the dynamic method exhibits notably higher overall accuracy than the static method, particularly within regions where there has been a shift in ecoregion types between 2013 and 2021. This contrast is particularly pronounced in the northern part of the CONUS.

3.4.2.2 Metrics for corn mapping

Given the successful performance of our soybean mapping approach using ten different ecoregions and computational efficiency considerations, we extend our experiment to map corn crops using the same approach. As shown in Table 3.6, our dynamic ecoregion mapping method demonstrates similar levels of Producer's Accuracy across different years, except for 2021, where we observe a notable improvement from 82.04% to 83.83%, representing a 1.79% increase over the static method. In

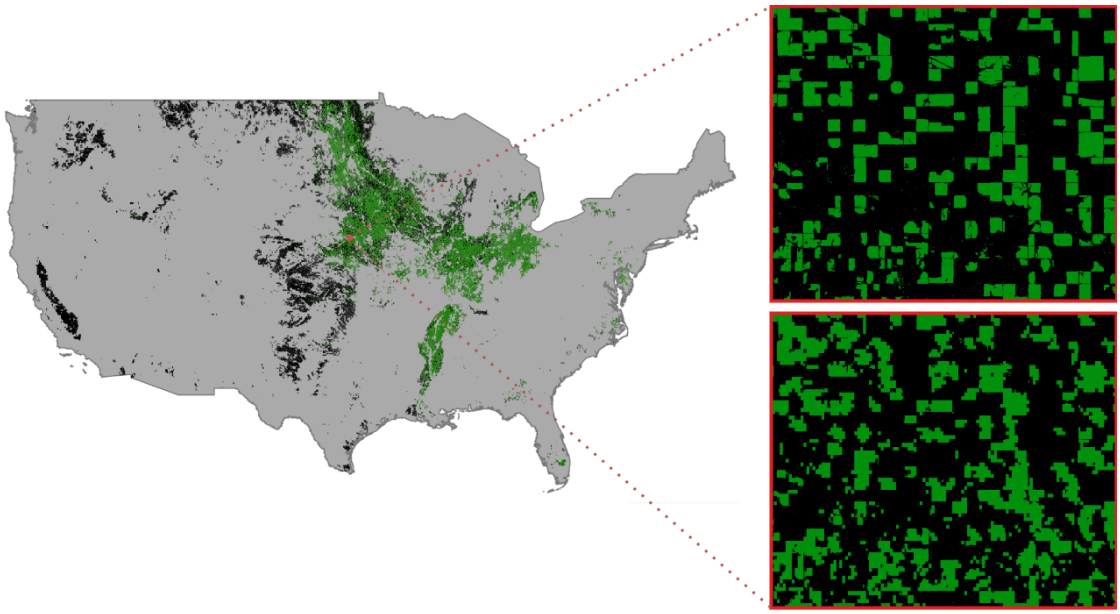
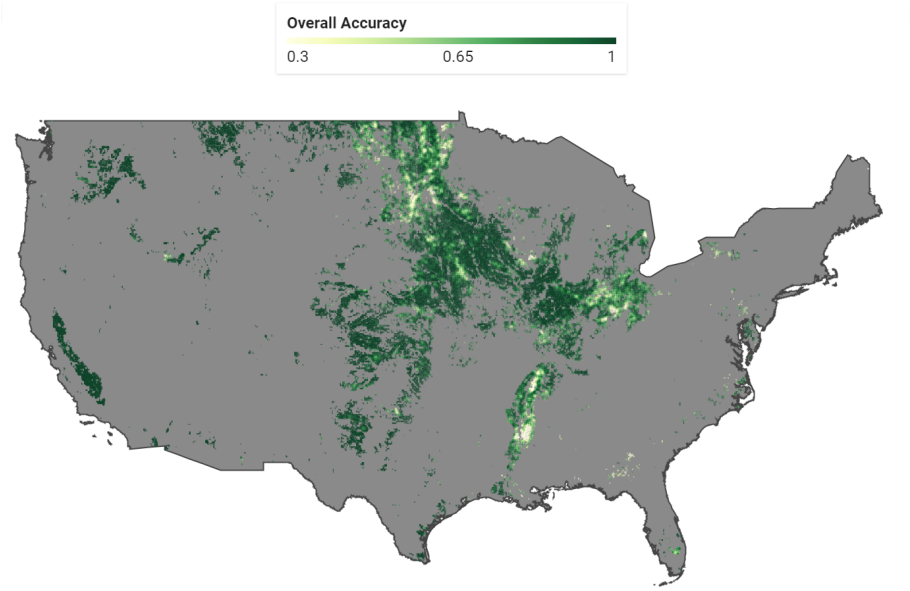


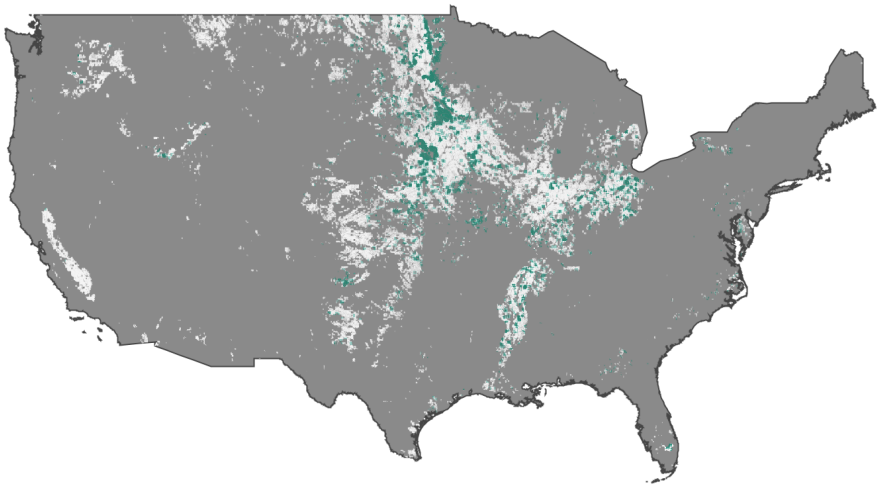
Figure 3.9: The soybean cropping result for the entire conterminous U.S. land area in 2021 (left). The ground truth of soybean cropland in a sample test region at 30-m resolution from the CDL layer (Right Top), and our soybean mapping result in the sample test region at 250-m resolution (Right Bottom)

Table 3.6: The experiment metrics for corn early crop mapping based on ten ecoregions.

Test year	**Producer's Accuracy**		**User's Accuracy**	
	dynamic	static	dynamic	static
2018	**85.70**	85.58	**62.51**	62.46
2019	60.14	**60.36**	**77.72**	76.62
2020	**77.77**	77.38	**68.81**	68.39
2021	**83.83**	82.04	65.55	**67.21**
2022	70.76	**70.87**	**73.31**	72.49



(a)



(b)

Figure 3.10: Soybean mapping pixel-wise Overall Accuracy for 2021. **(a)** illustrates the pixel-wise overall accuracy using the dynamic method, where the colour scale indicates the accuracy, with greener colours representing higher overall accuracy. **(b)** highlights the regions where the pixel-wise overall accuracy from the dynamic method outperforms the static method.

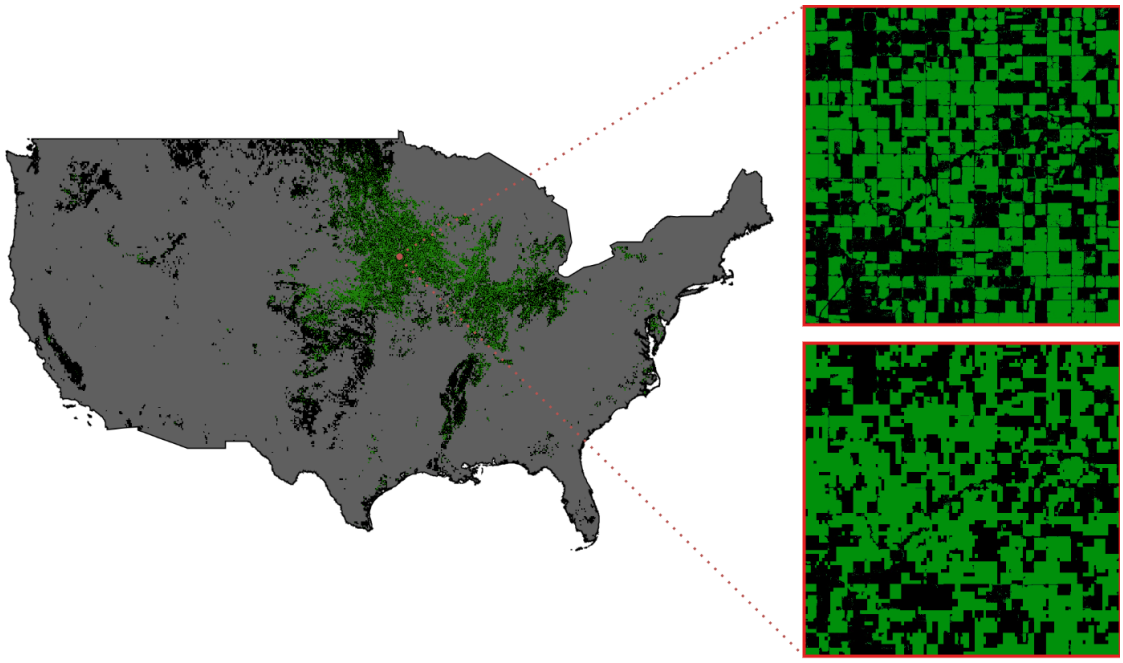


Figure 3.11: The corn cropping result for the entire conterminous U.S. land area in 2021 (left). The ground truth of soybean cropland in a sample test region at 30-m resolution from the CDL layer (Right Top), and our corn mapping result in the sample test region at 250-m resolution (Right Bottom)

Table 3.7: The metrics of corn mapping based on ecoregion-changed fields.

year	Producer's Accuracy		User's Accuracy		Overall Accuracy	
	dynamic	static	dynamic	static	dynamic	static
2018	85.53	**86.51**	**63.29**	61.75	**71.32**	70.02
2019	61.45	**64.71**	**77.94**	71.55	**77.47**	75.42
2020	**73.37**	72.96	**64.97**	60.49	**74.98**	71.76
2021	**87.59**	80.05	59.17	**65.72**	69.93	**72.83**
2022	73.09	**75.03**	**70.28**	65.21	**73.31**	71.66

terms of User's Accuracy, our dynamic approach outperforms the static method, with the greatest improvement observed in 2019 from 76.62% to 77.72%. Overall, these results demonstrate the effectiveness of our dynamic ecoregion mapping approach for mapping corn crops, with improvements in both Producer's Accuracy and User's Accuracy compared to the static method. Similar to the soybean mapping result, Figure 3.9 presents our corn mapping results for the CONUS in 2021.

To confirm our dynamic approach's superiority over the static method, we conducted another experiment focusing solely on the ecoregion-changing regions in the corn-growing area. As shown in Table 3.7, our dynamic method's User's Accuracy outperformed the static method in almost all cases, except for 2021. For Producer's Accuracy, both methods showed improvement or decline in different years. Therefore, we calculated the overall accuracy, which showed that the dynamic method in most cases outperformed the static method, with the largest improvement observed in 2020 from 71.76% to 74.98%. These results confirm the effectiveness of our dynamic ecoregion mapping approach for corn crops, especially in regions undergoing rapid

ecoregion changes.

3.5 Discussion

Our study demonstrates that our approach, incorporating dynamic ecoregion clustering and random forest classification, yields markedly higher accuracy compared to the static clustering method in the context of early crop mapping for target crops such as soybean and corn across the entire cropland region.

We employed the elbow method and silohuette method to ascertain the optimal number of ecoregion clusters. This approach allowed us to strike a balance between accuracy and computational efficiency, enabling us to achieve satisfactory early crop mapping results while conserving computational resources. However, there exists a limitation: when there is significant climate variability affecting the vegetation index (VIs) patterns of the target crops, but not substantial enough to alter the ecoregion boundaries, our system may struggle to address this scenario effectively. As illustrated in Figure 3.10a, the overall accuracy is notably low in the Southern USA. This issue becomes apparent when examining Figure 3.2c and 3.3, where, for the Southern USA region (depicted in blue), the ecoregion remains unchanged from 2013 to 2022. Consequently, our method fails to account for VIs pattern fluctuations in this region during this time period.

In our investigations, we observe that our dynamic ecoregion method, employing two, five, and ten distinct ecoregions, typically outperforms the static ecoregion method across most years in early soybean mapping. Particularly, the ecoregion configuration with ten divisions yields the most accurate outcomes. However, an interesting anomaly arises in the year 2020, where the static ecoregion map exhibits superior User's Accuracy performance. This divergence implies that the static ecoregion map could possess advantages in specific scenarios, although overall, our dynamic ecoregion method generally exhibits enhanced accuracy. Notably, a comparison of pixel-wise overall accuracy between the dynamic and static methods for the soybean mapping underscores that the dynamic approach significantly outperforms the static approach in regions where shifts in ecoregion types between 2013 and 2021 have occurred. This contrast is particularly evident in the northern CONUS, confirming the enhanced robustness of our method against climate fluctuations between different years.

Interestingly, we find that the dynamic method outperforms the static method significantly in soybean mapping, while the improvement is less pronounced in corn mapping. Although the exact cause has not been explored in this chapter, it is plausible that climate fluctuations exert a more substantial impact on soybean growing than on corn growing.

In conclusion, our proposed approach not only attests to its efficacy in terms of soybean and corn mapping but also highlights the benefits of dynamic ecoregion clustering in coping with the intricate influences of climate fluctuations on crop mapping accuracy. Furthermore, our result map provides early insights into crop conditions, significantly preceding the CDL release. This timeliness is pivotal for making informed decisions early in the growing season. While the USDA unveils its CDL layer in January or February of the following year, our results are accessible by mid-July of the same year, offering information approximately 7 months ahead of the CDL. In addition, producing and updating a 250m map proves to be more resource-efficient when compared to the maintenance of a nationwide 30m map, particularly for research and monitoring purposes. However, the lower resolution of

MODIS VIs data causes reduced accuracy in mapping target crops, when contrasted with Sentinel-2 and Landsat data. This decrease in resolution results in the loss of finer details and subtleties within the agricultural landscape, thereby presenting challenges in distinguishing between various crop types and detecting smaller-scale changes. It's important to note that this limitation constitutes a constraint within our work.

3.6 Conclusion

This chapter introduces an innovative approach for early crop mapping across the entire land area of the CONUS by utilizing NDVI and EVI data combined with a dynamic ecoregion clustering technique. Unlike static ecoregion clustering, which generates a single unchanging ecoregion map, our dynamic approach results in a unique ecoregion map for each year. This dynamic strategy enables us to incorporate the year-to-year climate variations that significantly influence crop growth, thereby heightening the precision of our crop mapping process.

With the ecoregion maps for 2013-2022 established by the dynamic ecoregion clustering, a specific early crop mapping classifier can be trained for each different ecoregion. We used a bagging procedure and trained a random forest consisting of 50 classification tree models to locate our target crop for each ecoregion separately across the entire cropland region of the CONUS. The model was trained and validated in the years 2013-2017 and tested independently in the years 2018-2022. The results showed that the dynamic clustering method achieved significantly higher accuracy than the static clustering method. Our method has significant implications for forecasting crop yield and food production for countries.

In our future research endeavors, we intend to broaden the scope of our work by applying the dynamic ecoregion method to estimate target crop yields. This represents a crucial evolution of our current approach, as it will enable us to not only identify and map crops within distinct ecological regions but also predict and quantify their potential yields. By harnessing the power of this method, we aim to provide valuable insights into agricultural productivity, aiding farmers, policymakers, and researchers in making informed decisions.

4 Cross Domain Early Crop Mapping using CropSTGAN

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 chapter 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.

This chapter is based on the work published in the following research paper:

- **Wang, Yiqun**, Hui Huang, and Radu State. "Cross Domain Early Crop Mapping using CropSTGAN." IEEE Access (2024).

Contents

4.1 Overview

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[22, 23] and many other agricultural applications. Since the spectral features of vegetation are determined by their structure, leaf biochemistry and phonological stages [24], 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 [25]. 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[26, 27], or use public datasets, such as the United States Department of Agriculture (USDA)'s Cropland Data Layer (CDL), as weak labels for model training [13]. Different DL architectures, such as convolutional neural networks (CNNs), Temporal Convolutional Neural Network (TempCNN) [19], deep autoencoders and recurrent neural networks with long short-term memory (LSTM), are explored for crop mapping tasks [31, 32, 34]. 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 [28, 29]. 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 [30]. 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 [51]. 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) [63] 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 chapter 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 [64], 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.

4.2 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 [38, 39, 40]. 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 [31, 34, 41]. 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 [28, 42]. 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.

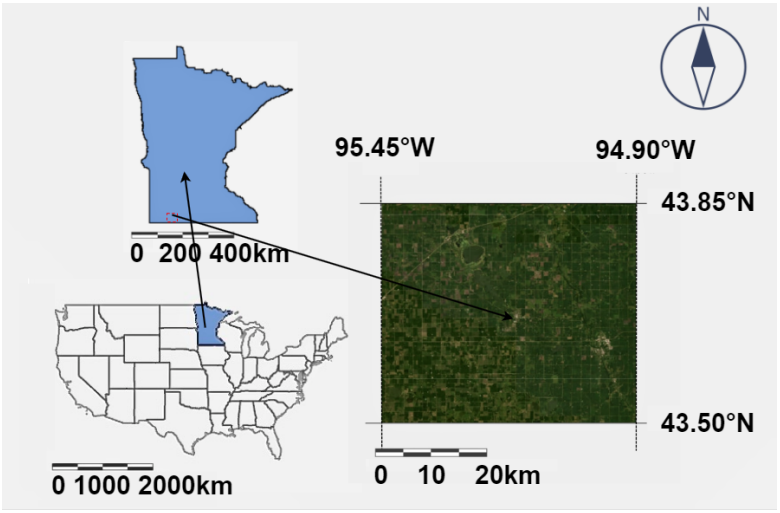
To address the cross-year issue, training the model with multi-year crop data proves to be a valuable adaptation method. Recent studies, including [51], emphasize the importance of incorporating multi-year crop data from a specific region as it enhances the model's understanding of temporal patterns, 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, [65] improved deep models for nationwide land cover classification using high-confidence pseudo-labels. Similarly, [66] involved annotating new target domain samples to refine RF classifiers via active learning. However, this technique often necessitates

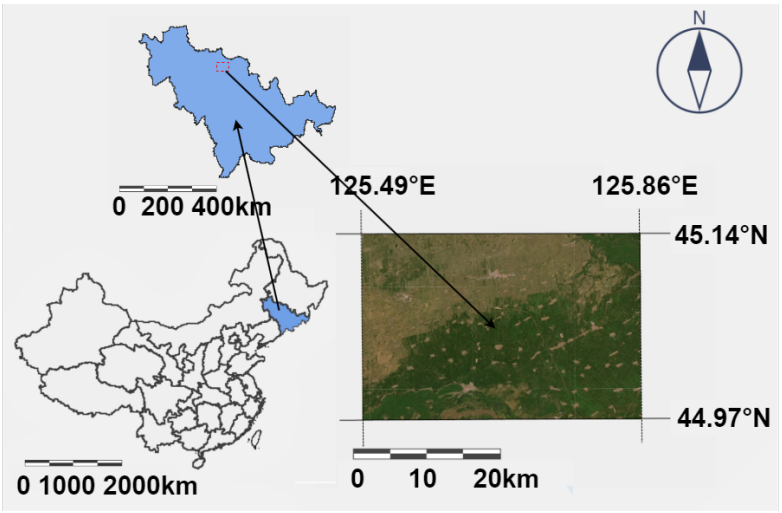
labelling samples, which is not feasible for large-scale studies.

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, [14] introduced the STDAN, a novel unsupervised domain adaptation framework for crop type classification. In the STDAN study [14], 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 [15] and the DACCN [16] extended the loss function using the Maximum Mean Discrepancy (MMD) and the 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 [15] 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, 2021). The DACCN study [16], 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.

This chapter 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.



(a)



(b)

Figure 4.1: The Study Areas. (a) The Study Area in Jackson County, Minnesota, USA. (b) The Study Area, Jilin Province, China.

4.3 Data and Study Areas

4.3.1 Study area

This chapter 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 4.1.

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 [57], is utilized

Table 4.1: 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

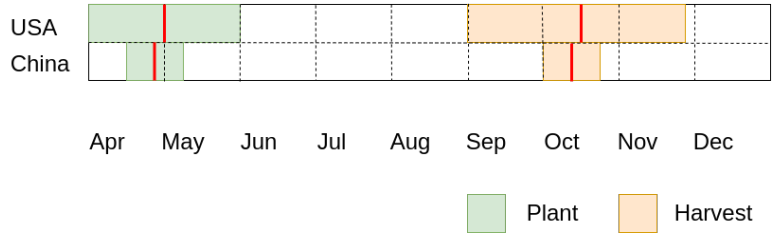


Figure 4.2: The Crop Calendar for Corn. The red lines represent the average dates.

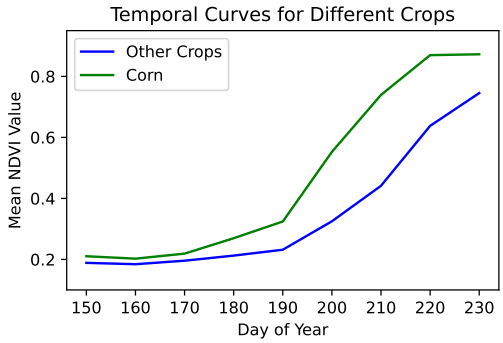
in our work with a resolution of $11132m \times 11132m$ from 2019 to 2021 as the climate data which is comprised of air and soil temperature, precipitation, Evaporation, and solar radiation. Table 5.1 presents the environmental metrics for these areas. Additionally, the corn cropping schedules differ among the regions, as depicted in Figure 4.2. 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 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.

4.3.2 Reference Data

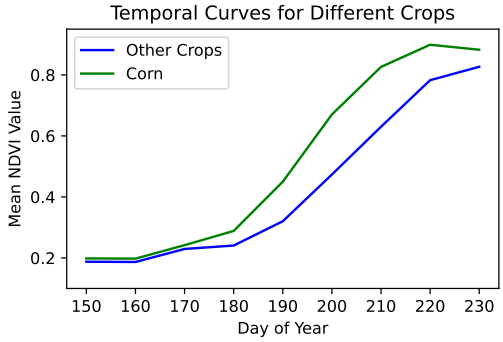
The CDL is used as the ground truth for the source domain. The CDL [13], a crop-specific land cover raster map dataset available for the entire conterminous U.S. land area (CONUS) at 30 m resolution provided by the USDA, regularly provides information on the annual temporal and spatial distribution of corn, as well as the area dedicated to its cultivation. As shown in Figure 4.5a, 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, [27] 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 [27] are regarded as the ground truth reference for the study area in China. As shown in Figure 4.5b, the main crop types in the study areas are corn, soybean, and rice.

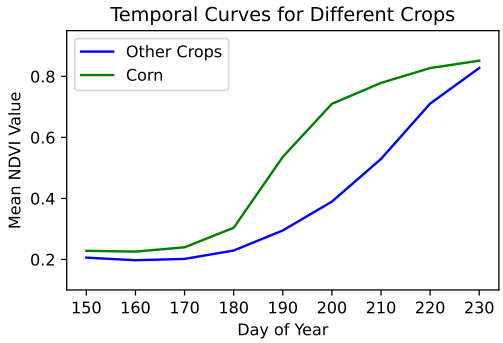
The ground truth data used are either from publicly available datasets or previous research efforts. Consequently, during model comparison, the evaluation is based on determining which model aligns more closely with previous methods rather than



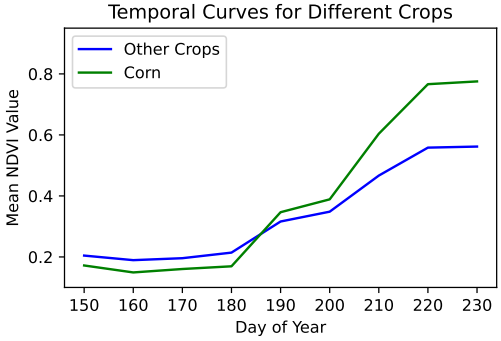
(a)



(b)



(c)



(d)

Figure 4.3: 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.

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.

4.3.3 Remote Sensing Data

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 [31, 67]. 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 4.2, 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 [68]. 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.

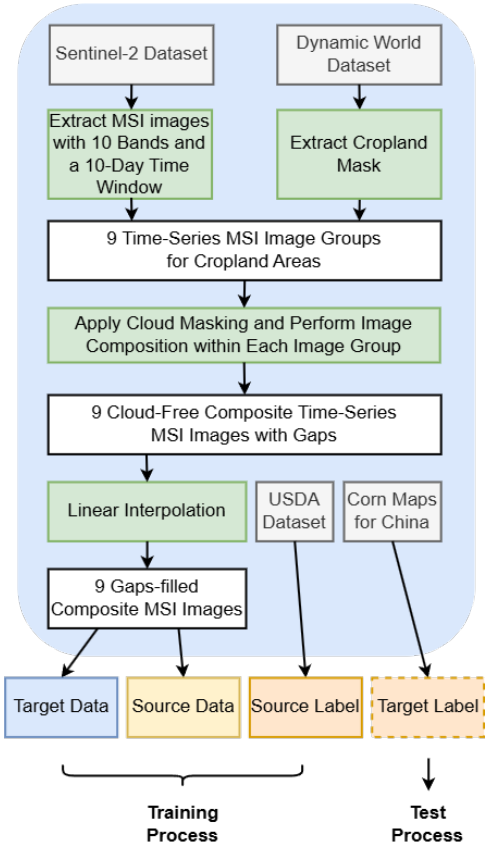


Figure 4.4: The Architecture of the Pre-processor.

4.4 Methodology

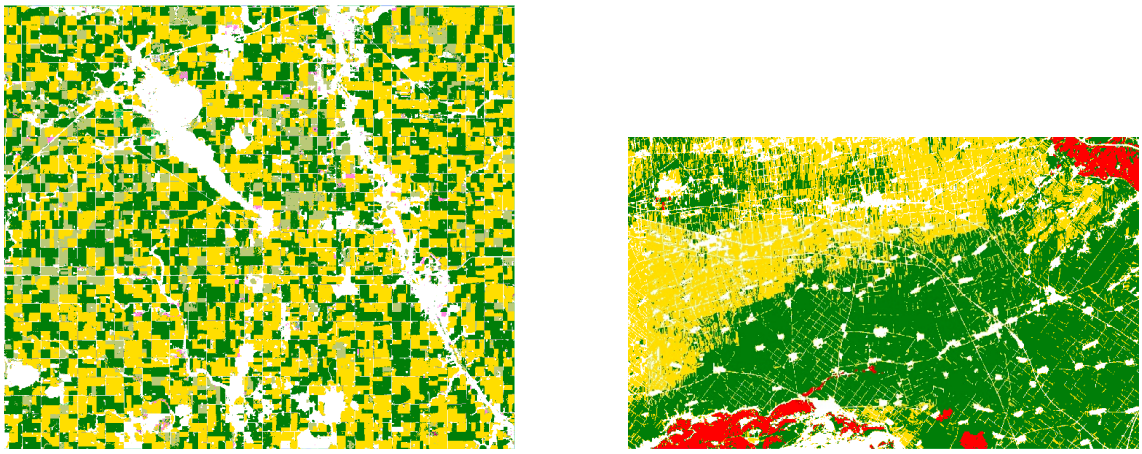
4.4.1 Problem Statement

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 $[\mathbf{x}_1, \mathbf{x}_2, ..., \mathbf{x}_t]$, where $\mathbf{x}_i$ represents input at time i. $\mathbf{x}_i$ can be further expanded as $[\mathbf{x}_{i1}, \mathbf{x}_{i2}, ..., \mathbf{x}_{ib}]$, 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.

4.4.2 System Overview

The architecture of our CropSTGAN framework is depicted in Figure 4.6, 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,



(a) (b)

Figure 4.5: 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.

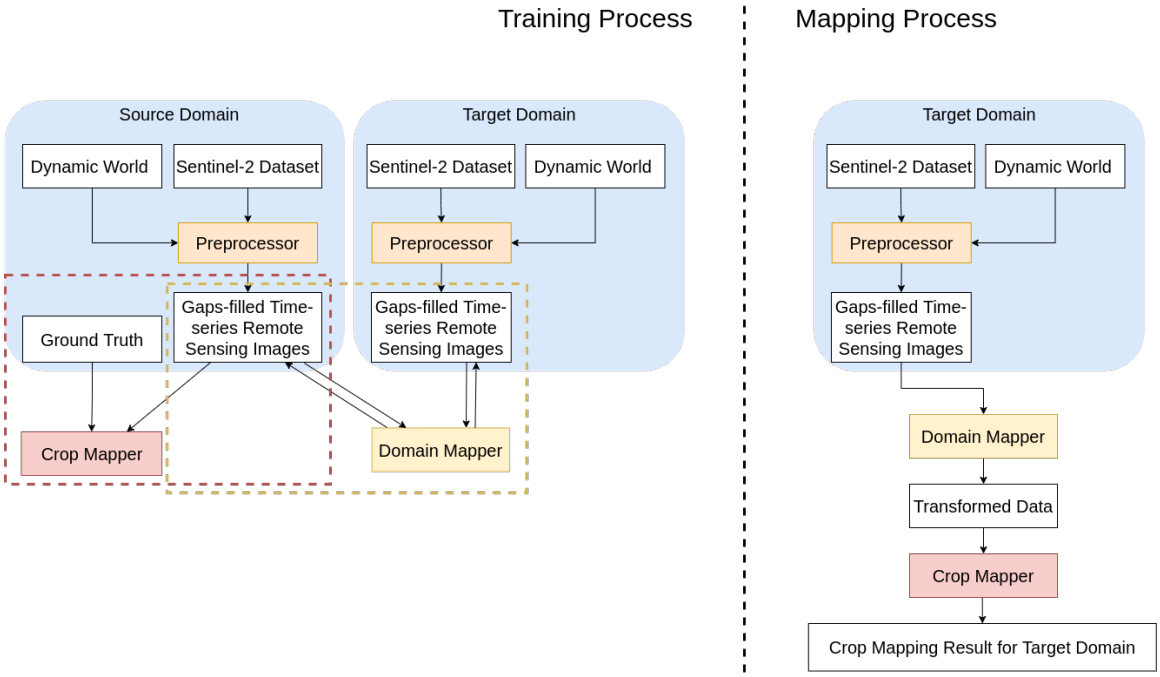


Figure 4.6: The Architecture of the CropSTGAN Framework.

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

Table 4.2: 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

mapping in the target domain, particularly useful when labelled data is scarce or nonexistent, thereby overcoming the challenges of cross-domain mapping.

4.4.3 Pre-processor

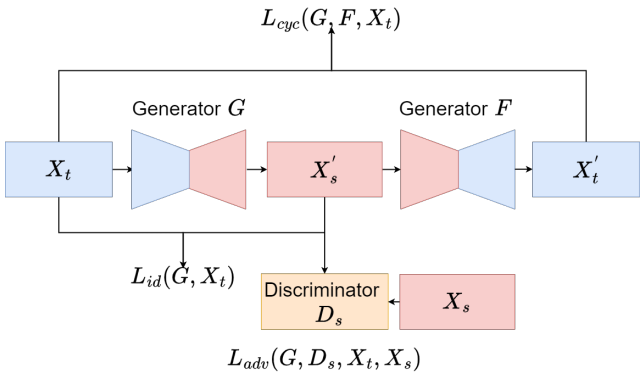
The pre-processor aims at providing complete time-series MSI data by filling gaps between MSI images due to cloud cover, atmospheric interference, or sensor limitations. The processing pipeline is outlined in Figure 4.4.

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.

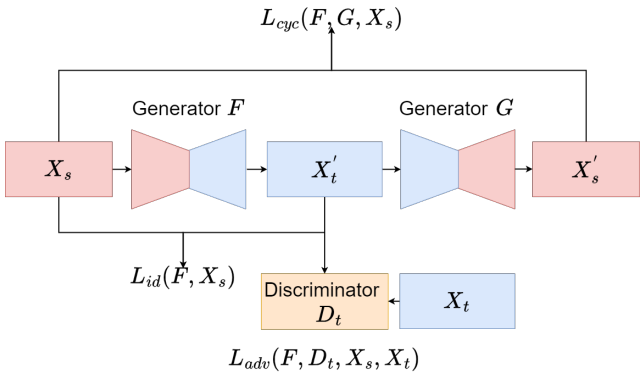
4.4.4 CropSTGAN Domain Mapper

4.4.4.1 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



(a)



(b)

Figure 4.7: 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.

Table 4.3: 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

sensing data points between different domains. As shown in Figure 4.7, the domain mapper consists of two generator networks, G and F, and two discriminator networks,

Table 4.4: 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 4.5: 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

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.

The generator and discriminator structures are depicted in Table 4.2 and Table 4.3, 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 4.4 and Table 4.5.

4.4.4.2 Losses

As shown in Figure 4.7, 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:

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).

$$\begin{aligned}\mathcal{L}_{\text{adv1}}(G, D_s, \mathbf{X}_t, \mathbf{X}_s) =& \mathbb{E}_{\mathbf{x}_s \sim p_{\text{data}}(\mathbf{x}_s)}[\log D_s(\mathbf{x}_s)] + \\ & \mathbb{E}_{\mathbf{x}_t \sim p_{\text{data}}(\mathbf{x}_t)}[1 - \log D_s(G(\mathbf{x}_t))]\end{aligned} \tag{4.1}$$

$$\begin{aligned}\mathcal{L}_{\text{adv2}}(F, D_t, \mathbf{X}_s, \mathbf{X}_t) =& \mathbb{E}_{\mathbf{x}_t \sim p_{\text{data}}(\mathbf{x}_t)}[\log D_t(\mathbf{x}_t)] + \\ & \mathbb{E}_{\mathbf{x}_s \sim p_{\text{data}}(\mathbf{x}_s)}[1 - \log D_t(G(\mathbf{x}_s))]\end{aligned} \tag{4.2}$$

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.

$$\mathcal{L}_{\text{cyc1}}(G, F, \mathbf{X}_t) = \mathbb{E}_{\mathbf{x}_t \sim p_{\text{data}}(\mathbf{x}_t)}[||\mathbf{x}_t - F(G(\mathbf{x}_t))||_1] \tag{4.3}$$

$$\mathcal{L}_{\text{cyc2}}(F, G, \mathbf{X}_s) = \mathbb{E}_{\mathbf{x}_s \sim p_{\text{data}}(\mathbf{x}_s)}[||\mathbf{x}_s - G(F(\mathbf{x}_s))||_1] \tag{4.4}$$

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.

$$\mathcal{L}_{\text{id1}}(G, \mathbf{X}_t) = \mathbb{E}_{\mathbf{x}_t \sim p_{\text{data}}(\mathbf{x}_t)}[||G(\mathbf{x}_t) - \mathbf{x}_t||_1] \tag{4.5}$$

$$\mathcal{L}_{\text{id2}}(F, \mathbf{X}_s) = \mathbb{E}_{\mathbf{x}_s \sim p_{\text{data}}(\mathbf{x}_s)}[||F(\mathbf{x}_s) - \mathbf{x}_s||_1] \tag{4.6}$$

Total Loss

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

$$\begin{aligned} \mathcal{L}_{\text{total}} = & \mathcal{L}_{\text{adv1}}(G, D_s, \mathbf{X}_t, \mathbf{X}_s) + \\ & \alpha \mathcal{L}_{\text{cyc1}}(G, F, \mathbf{X}_t) + \\ & \beta \mathcal{L}_{\text{id1}}(G, \mathbf{X}_t) \\ & \mathcal{L}_{\text{adv2}}(F, D_t, \mathbf{X}_s, \mathbf{X}_t) + \\ & \gamma \mathcal{L}_{\text{cyc2}}(F, G, \mathbf{X}_s) + \\ & \delta \mathcal{L}_{\text{id2}}(F, \mathbf{X}_s) \end{aligned} \tag{4.7}$$

, where $\alpha, \beta, \gamma, \delta$ are the weight parameters:

$$\alpha = \frac{\mathcal{L}_{\text{adv1}}}{\mathcal{L}_{\text{cyc1}}}, \beta = \frac{\mathcal{L}_{\text{adv1}}}{\mathcal{L}_{\text{id1}}}, \gamma = \frac{\mathcal{L}_{\text{adv2}}}{\mathcal{L}_{\text{cyc2}}}, \delta = \frac{\mathcal{L}_{\text{adv2}}}{\mathcal{L}_{\text{id2}}} \tag{4.8}$$

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:

$$\min_{G,F} \max_{D_t, D_s} \mathcal{L}_{\text{total}}(G, F, D_t, D_s, \mathbf{X}_t, \mathbf{X}_s). \tag{4.9}$$

, 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.

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.

4.4.5 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 4.3, is trained solely with labelled source domain data ($\mathbf{X}_s$, $\mathbf{Y}_s$). The binary cross-entropy loss:

$$\mathcal{L}_{\text{class}}(C) = -\sum_s \left(y_s \log(C(\mathbf{x}_s)) + (1 - y_s) \log(1 - C(\mathbf{x}_s))\right) \tag{4.10}$$

, 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 4.6, this transformed data is then processed by the TempCNN crop mapper to generate mapping results:

$$\mathbf{Y}_t = C(G(\mathbf{X}_t)) \tag{4.11}$$

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.

4.5 Experiment Setup and Results

Table 4.6: 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.

(a)

	CropSTGAN			**CropTGAN**		
	OA	F1	Kappa	OA	F1	Kappa
S1	0.9602	0.9613	0.9205	0.9477	0.9503	0.8952
S2	0.9619	0.9630	0.9239	0.9346	0.9403	0.8681
S3	0.9641	0.9657	0.9281	0.9440	0.9477	0.8874
Avg	<u>0.9621</u>	<u>0.9633</u>	<u>0.9242</u>	0.9421	0.9461	0.8836

(b)

	STDAN			**TempCNN**		
	OA	F1	Kappa	OA	F1	Kappa
S1	0.9651	0.9670	0.9299	0.9688	0.9704	0.9376
S2	0.9550	0.9530	0.9122	-	-	-
S3	0.9549	0.9582	0.9093	-	-	-
Avg	0.9583	0.9594	0.9171	**0.9688**	**0.9704**	**0.9376**

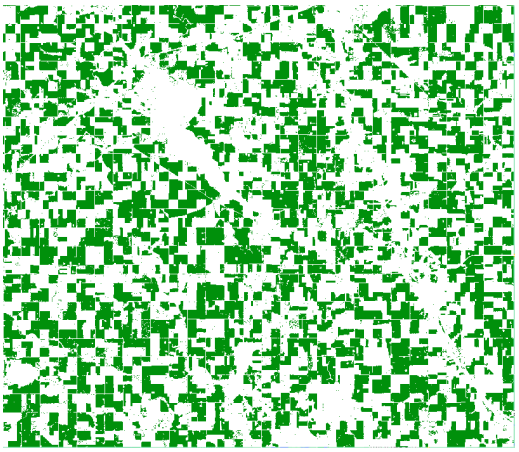
4.5.1 Experiment Setup

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

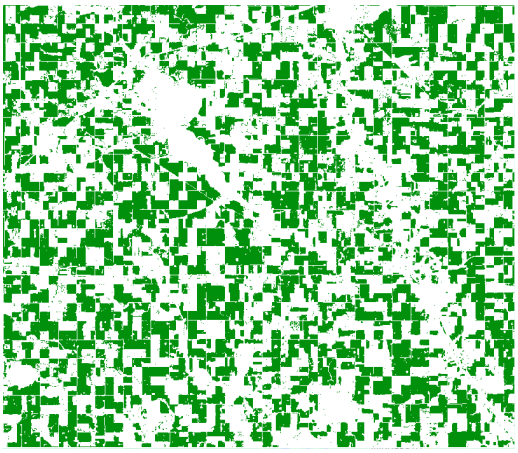
- The first set of experiments investigates the cross-year scenario, where Jackson County from 2020 to 2021 is considered as the target domain, and Jackson County in 2019 serves as the source domain. The discrepancies in data distribution between these domains are relatively slight.
- The second series 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.

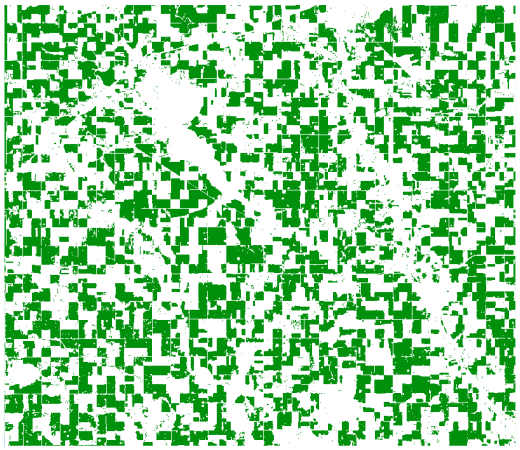
Furthermore, a series of experiments based on CropSTGAN were conducted for the cross-region analysis using MSI data collected at intervals of 50, 60, 70, 80, and 90 days starting from May 1st or the 150th day of the year, in order to determine the impact of different time-series input data lengths on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



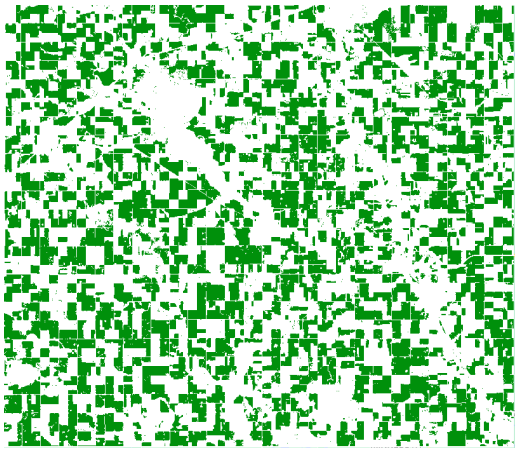
(a)



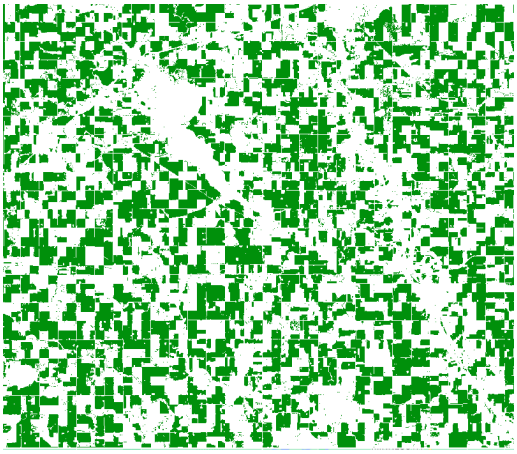
(b)



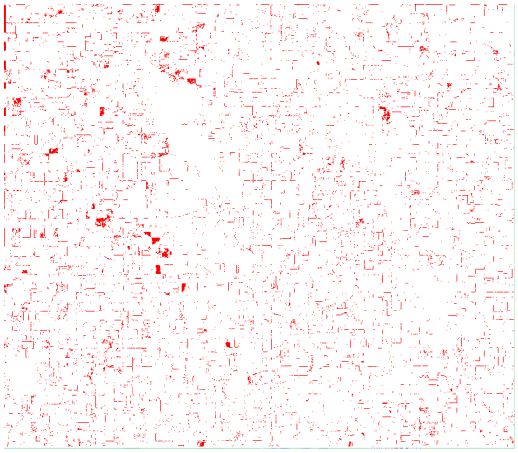
(c)



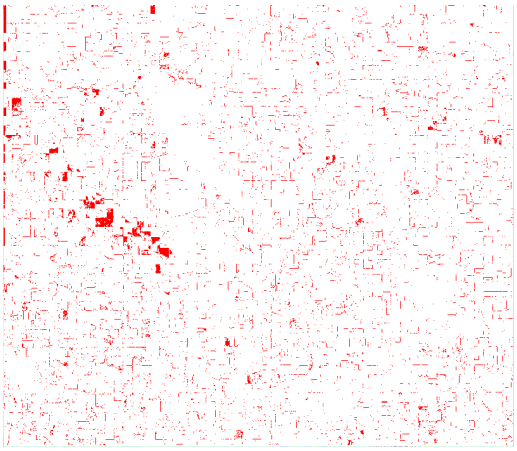
(d)



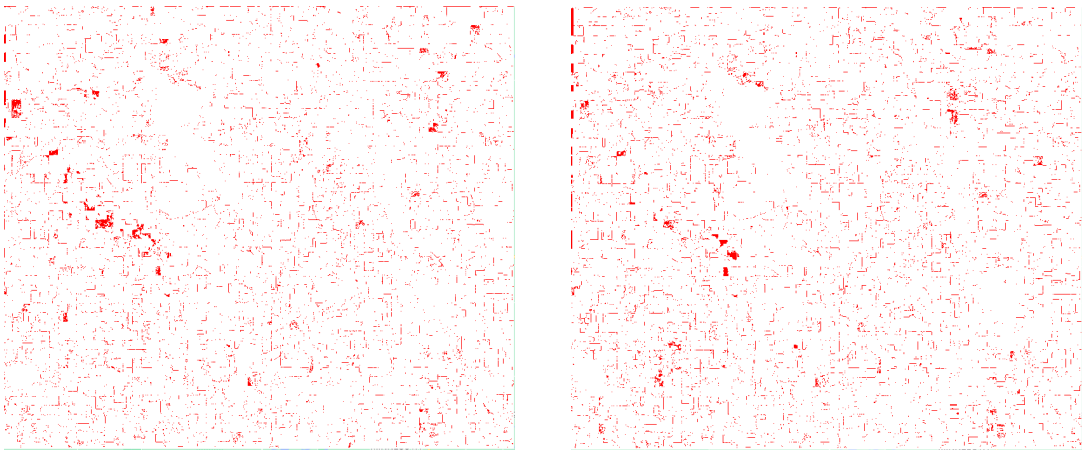
(e)



(f)



(g)



(h) (i)

Figure 4.8: 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, (d) for CropTGAN, (f) for STDAN, and (h) for TempCNN. In this visualization, green denotes corn, and white represents other types. The corresponding error images are illustrated in panels (c), (e), (g), and (i) for CropSTGAN, CropTGAN, STDAN, and TempCNN, respectively. Red highlights the misclassified pixels, and white represents correctly classified pixels.

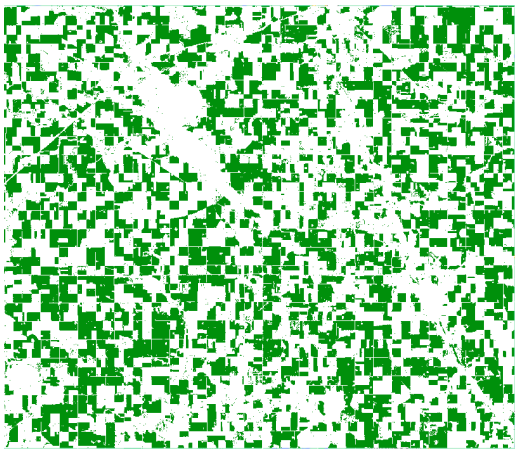
Table 4.7: Cross-Year Experiment Metrics Comparison: Targeting Jackson County 2021. The best metrics are indicated in bold, while the second-best metrics of the baseline methods are underlined.

(a)

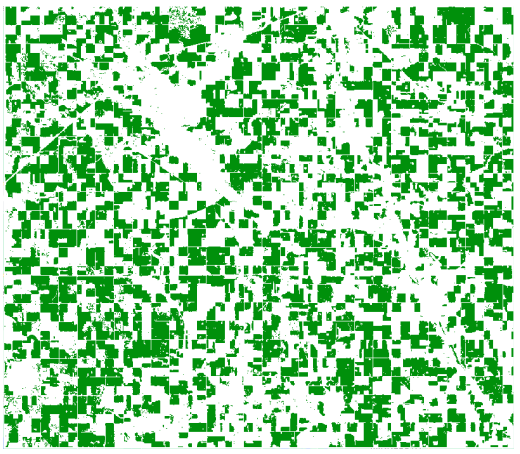
	CropSTGAN			**CropTGAN**		
	OA	F1	Kappa	OA	F1	Kappa
S1	0.9608	0.9616	0.9216	0.9581	0.9573	0.9163
S2	0.9701	0.9701	0.9401	0.9396	0.9396	0.8792
S3	0.9660	0.9663	0.9320	0.9523	0.9534	0.9047
Avg	**0.9656**	**0.9660**	**0.9312**	0.9500	0.9501	0.9001

(b)

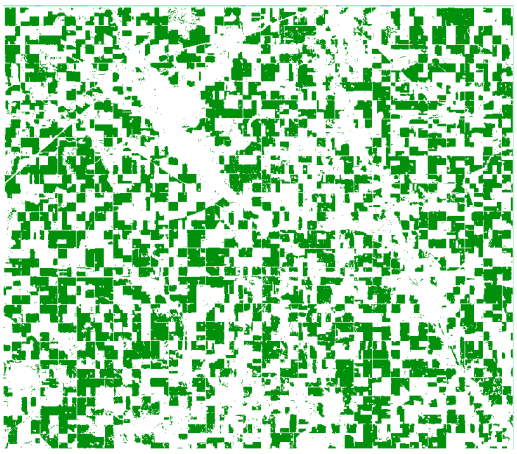
	STDAN			**TempCNN**		
	OA	F1	Kappa	OA	F1	Kappa
S1	0.9597	0.9604	0.9195	0.9338	0.9387	0.8669
S2	0.9673	0.9685	0.9345	-	-	-
S3	0.9677	0.9688	0.9354	-	-	-
Avg	<u>0.9649</u>	<u>0.9659</u>	<u>0.9298</u>	0.9338	0.9387	0.8669



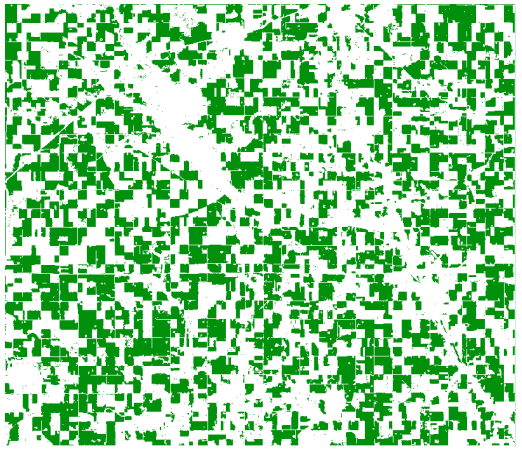
(a)



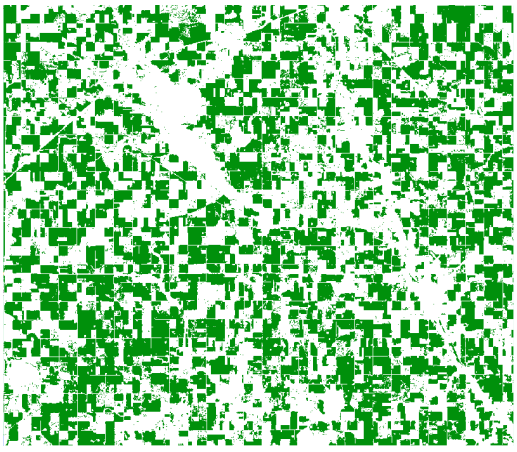
(b)



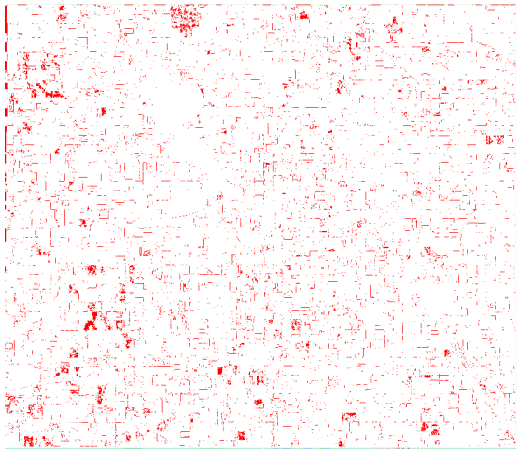
(c)



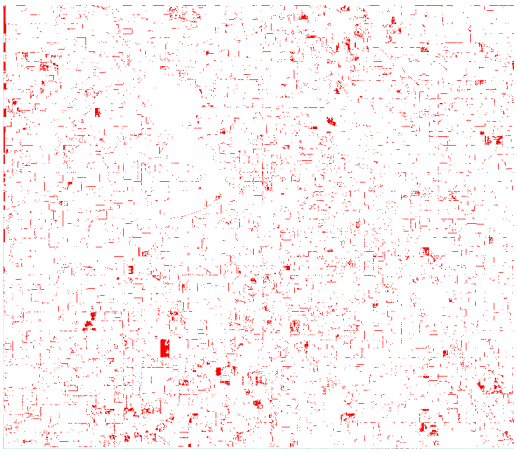
(d)



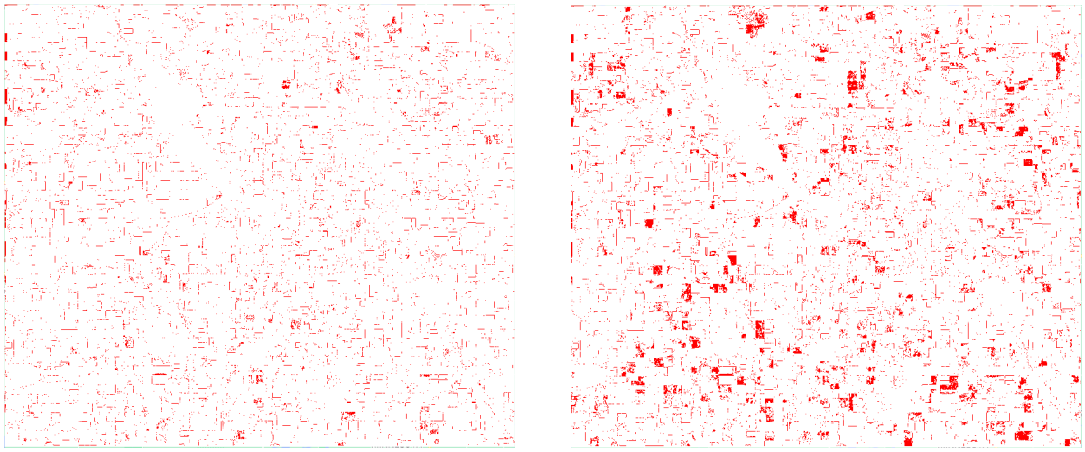
(e)



(f)



(g)



(h) (i)

Figure 4.9: Corn Crop Mapping Results Comparison with Jackson County 2021 as the Target Domain. (a) displays the ground truth. The crop mapping results are depicted in (b) for CropSTGAN, (d) for CropTGAN, (f) for STDAN, and (h) for TempCNN. In this visualization, green denotes corn, and white represents other types. The corresponding error images are illustrated in panels (c), (e), (g), and (i) for CropSTGAN, CropTGAN, STDAN, and TempCNN, respectively. Red highlights the misclassified pixels, and white represents correctly classified pixels.

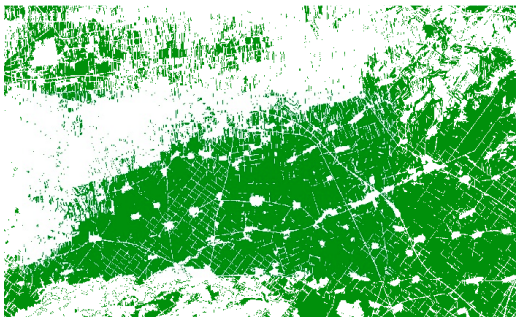
Table 4.8: Cross-Region Experiment Metrics Comparison: Targeting China 2019. The best metrics are indicated in bold, while the second-best metrics of the baseline methods are underlined.

(a)

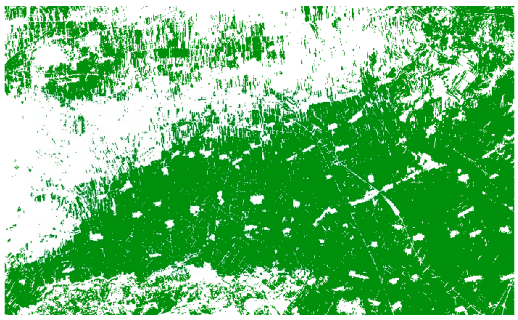
	CropSTGAN			**CropTGAN**		
	OA	F1	Kappa	OA	F1	Kappa
S1	0.8790	0.8901	0.7557	0.8539	0.8744	0.7019
S2	0.8935	0.9033	0.7847	0.8457	0.8675	0.6850
S3	0.8842	0.8968	0.7654	0.8663	0.8755	0.7311
Avg	**0.8856**	**0.8967**	**0.7686**	<u>0.8553</u>	<u>0.8725</u>	<u>0.7060</u>

(b)

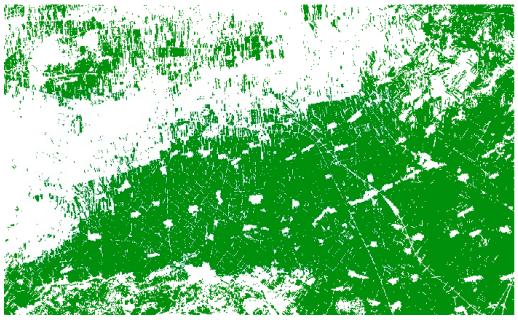
	STDAN			**TempCNN**		
	OA	F1	Kappa	OA	F1	Kappa
S1	0.8416	0.8613	0.6771	0.8448	0.8657	0.6826
S2	0.8353	0.8576	0.6630	-	-	-
S3	0.8339	0.8583	0.6590	-	-	-
Avg	0.8369	0.8591	0.6664	0.8448	0.8657	0.6826



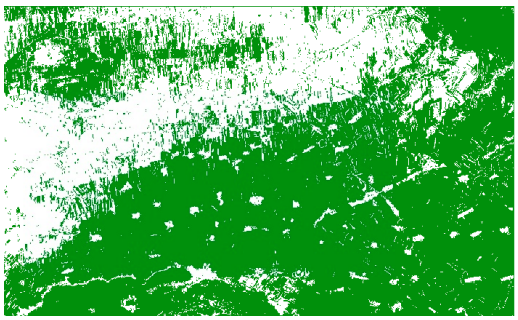
(a)



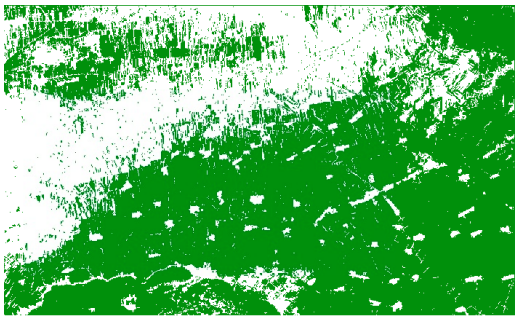
(b)



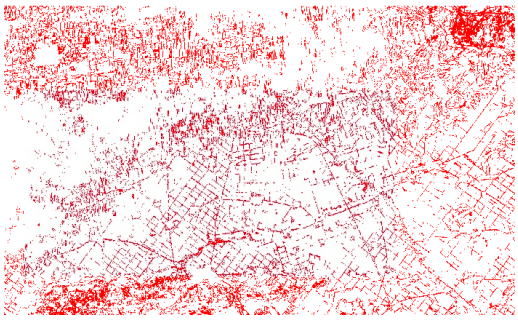
(c)



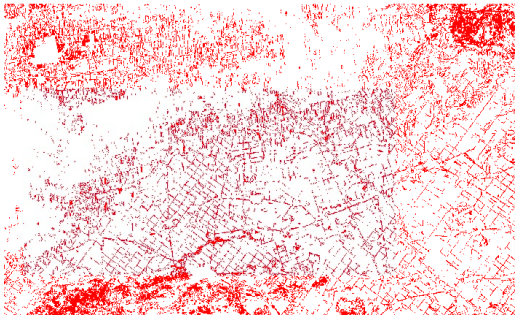
(d)



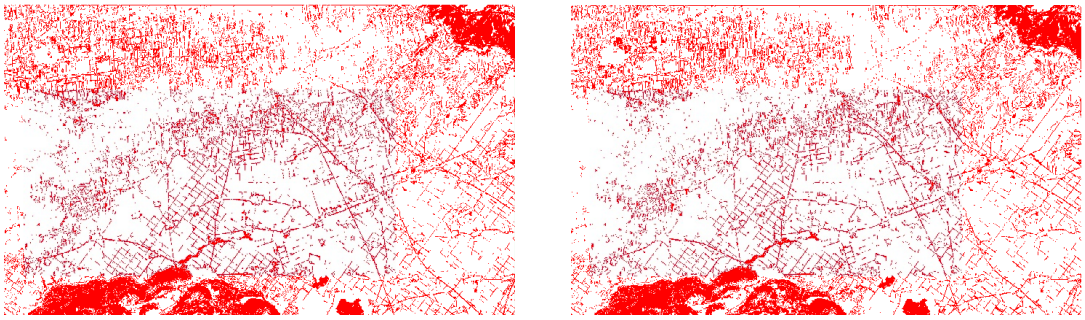
(e)



(f)



(g)



(h) (i)

Figure 4.10: Corn Crop Mapping Results Comparison with the Study Area of China 2019 as the Target Domain. (a) displays the ground truth. The crop mapping results are depicted in (b) for CropSTGAN, (d) for CropTGAN, (f) for STDAN, and (h) for TempCNN. In this visualization, green denotes corn, and white represents other types. The corresponding error images are illustrated in panels (c), (e), (g), and (i) for CropSTGAN, CropTGAN, STDAN, and TempCNN, respectively. Red highlights the misclassified pixels, and white represents correctly classified pixels.

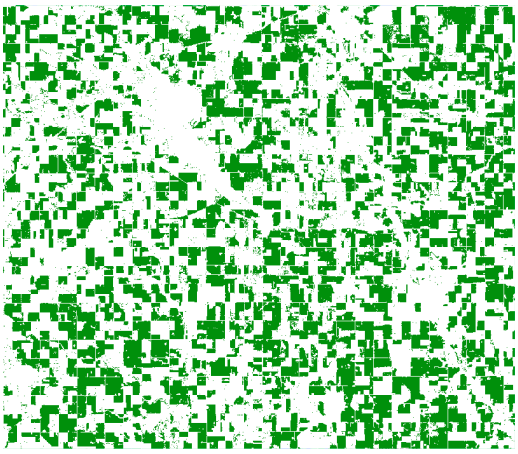
Table 4.9: Cross-Region Experiment Metrics Comparison: Targeting Jackson County 2019. The best metrics are indicated in bold, while the second-best metrics of the baseline methods are underlined.

(a)

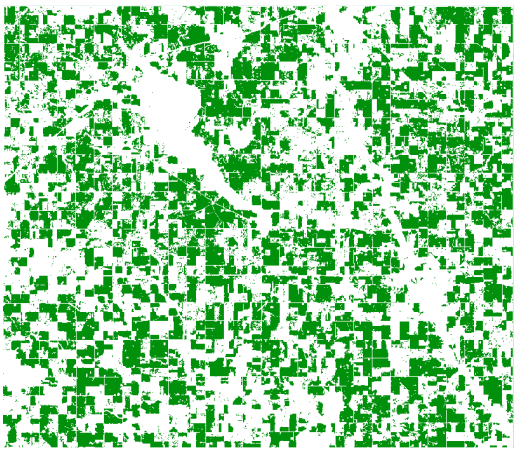
	CropSTGAN			**CropTGAN**		
	OA	F1	Kappa	OA	F1	Kappa
S1	0.8790	0.8901	0.7557	0.8409	0.8585	0.6783
S2	0.8743	0.8802	0.7485	0.8428	0.8594	0.6812
S3	0.8687	0.8799	0.7351	0.8361	0.8505	0.6690
Avg	**0.8740**	**0.8834**	**0.7464**	0.8399	0.8561	0.6762

(b)

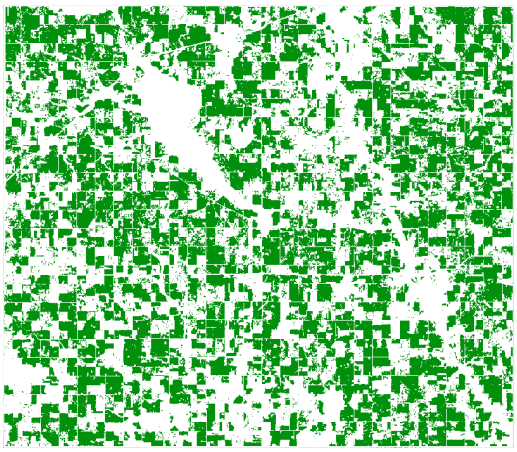
	STDAN			**TempCNN**		
	OA	F1	Kappa	OA	F1	Kappa
S1	0.8156	0.8364	0.6270	0.8465	0.8620	0.6900
S2	0.8383	0.8549	0.6735	-	-	-
S3	0.8431	0.8496	0.6856	-	-	-
Avg	0.8323	0.8470	0.6620	<u>0.8465</u>	<u>0.8620</u>	<u>0.6900</u>



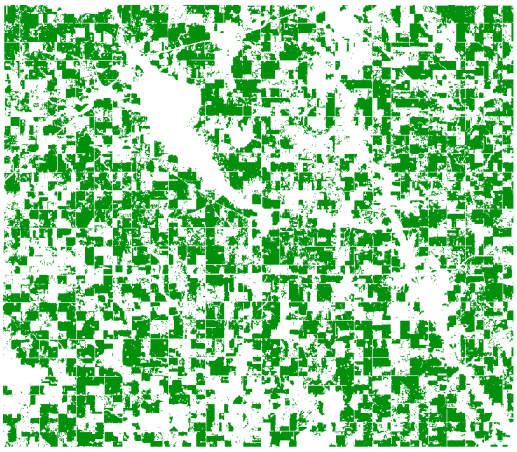
(a)



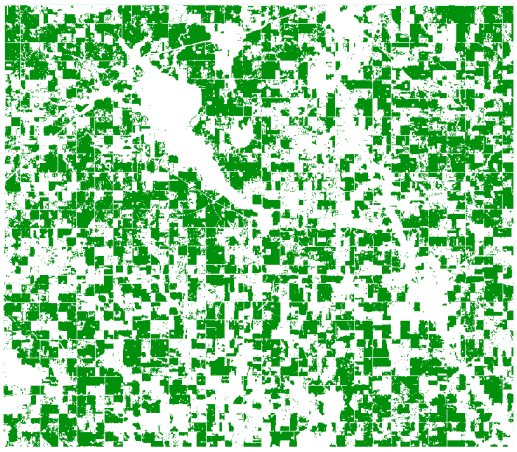
(b)



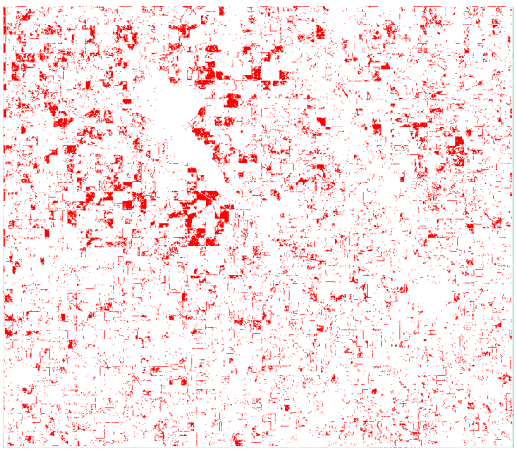
(c)



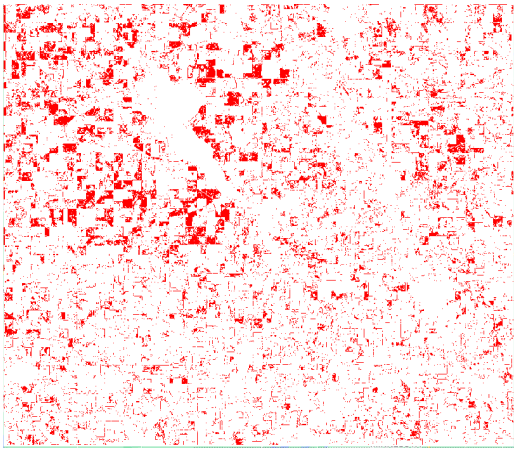
(d)



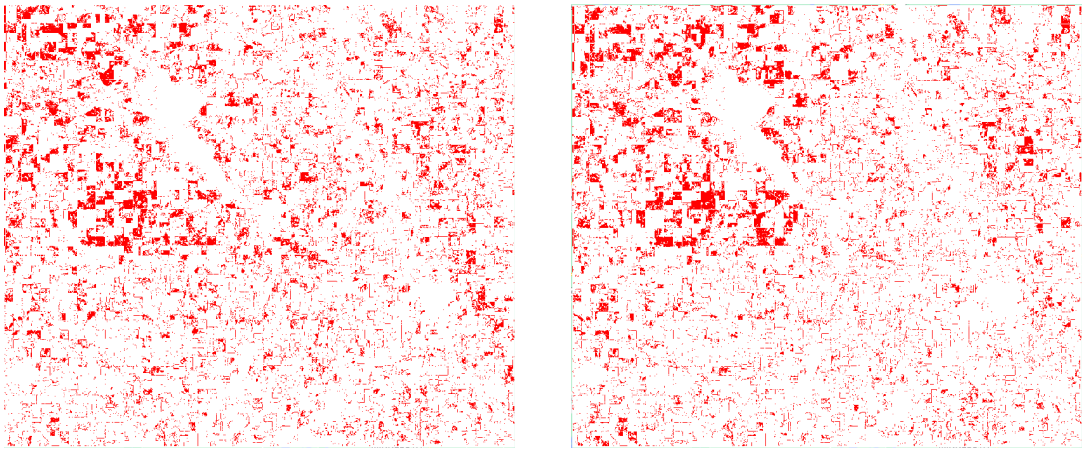
(e)



(f)



(g)



(h) (i)

Figure 4.11: Corn Crop Mapping Results Comparison with Jackson County 2019 as the Target Domain. (a) displays the ground truth. The crop mapping results are depicted in (b) for CropSTGAN, (d) for CropTGAN, (f) for STDAN, and (h) for TempCNN. In this visualization, green denotes corn, and white represents other types. The corresponding error images are illustrated in panels (c), (e), (g), and (i) for CropSTGAN, CropTGAN, STDAN, and TempCNN, respectively. Red highlights the misclassified pixels, and white represents correctly classified pixels.

Table 4.10: 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.

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

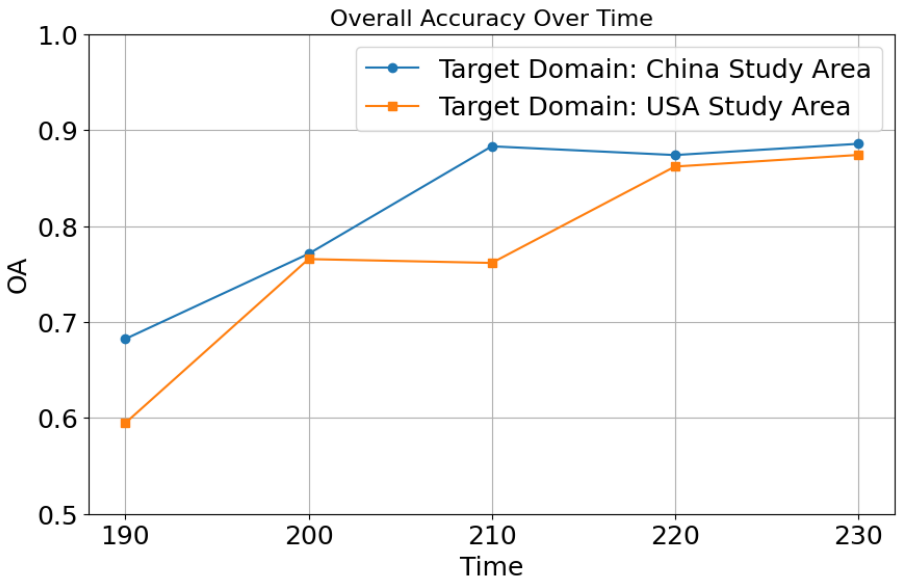


Figure 4.12: The Mapping Overall Accuracy with Different Time-Series Length Input using CropSTGAN.

structures is removed from the domain mapper generator, and one convolutional layer is removed from both the domain mapper discriminator and the crop mapper.

4.5.2 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.

4.5.3 Evaluation Metrics

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

Overall Accuracy (OA)

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

F1 Score

is a single metric that combines precision and recall to provide an overall measure of a model's accuracy in classification tasks. It balances the trade-off between correctly identifying positive instances and minimizing false positives and false negatives.

The Kappa coefficient

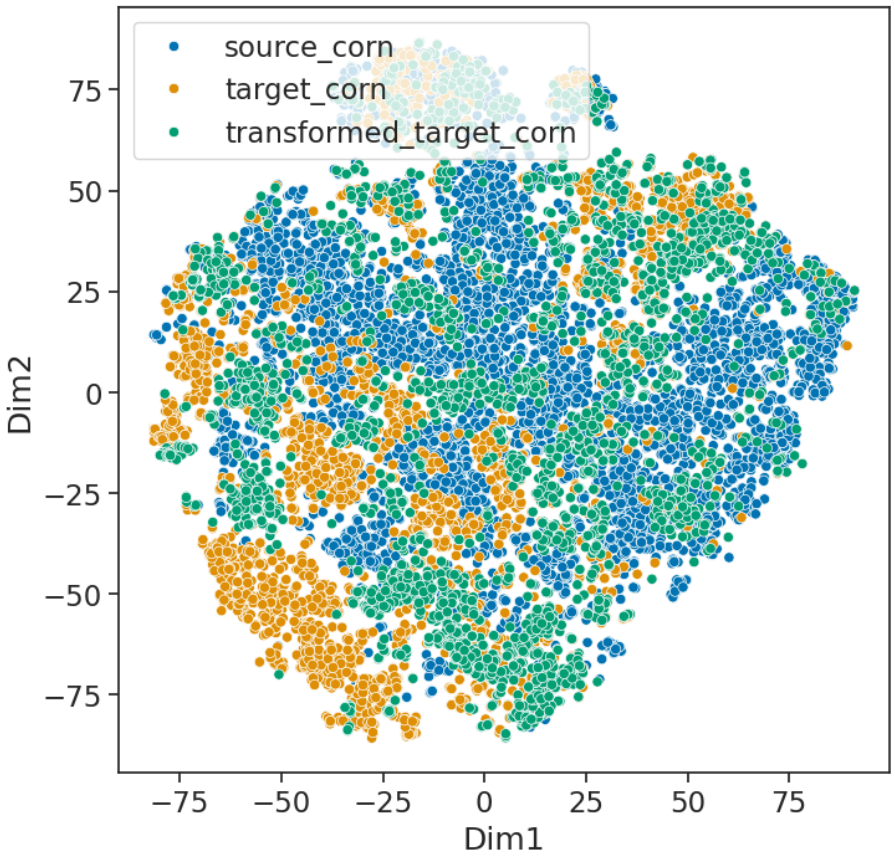
measures the agreement between the predicted and observed classifications, taking into account the agreement that would occur by chance alone.

4.5.4 Results

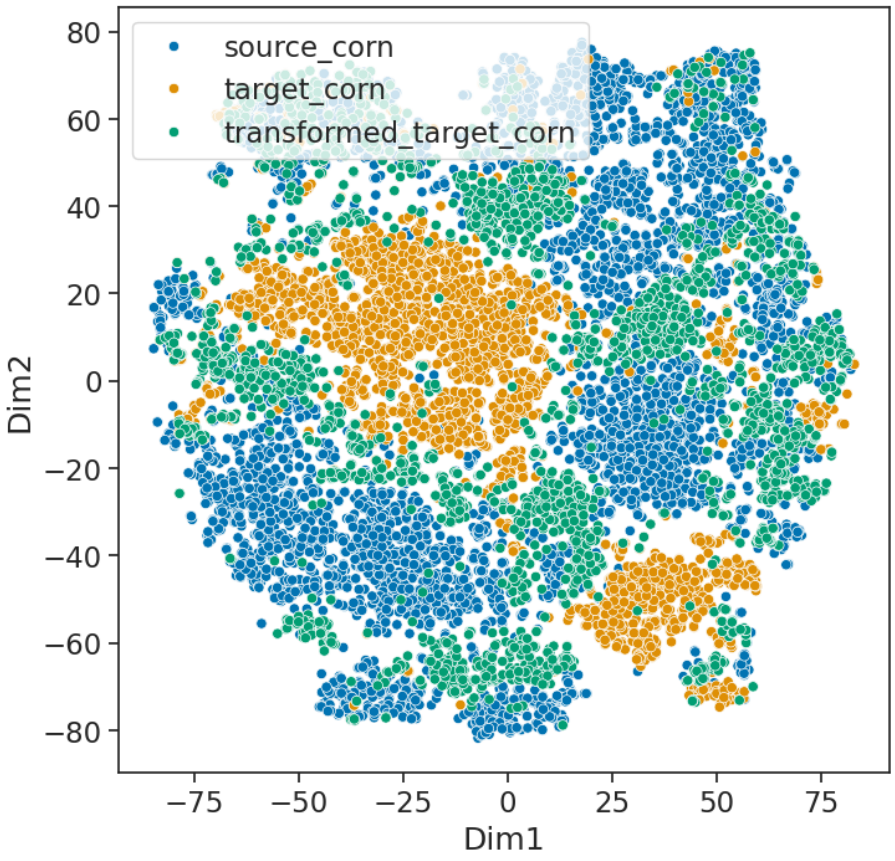
Tables 4.6 and 4.7 present the results for the first set of cross-year experiments. Tables 4.8 and 4.9 present the results for the second set of cross-region experiments. Across two sets of experiments, our CropSTGAN method demonstrates the highest average metrics, except for the cross-year experiment targeting Jackson County 2020.

In the first set of experiments targeting Jackson County 2020, the CropSTGAN achieved remarkable results, securing 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 4.8 and 4.9 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 4.10 and 4.11 present the results and error images from cross-region experiments.

Furthermore, Table 4.10 presents the cross-region experiment results of early crop mapping using different lengths of time-series MSI data. As shown in Figure 4.12, 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

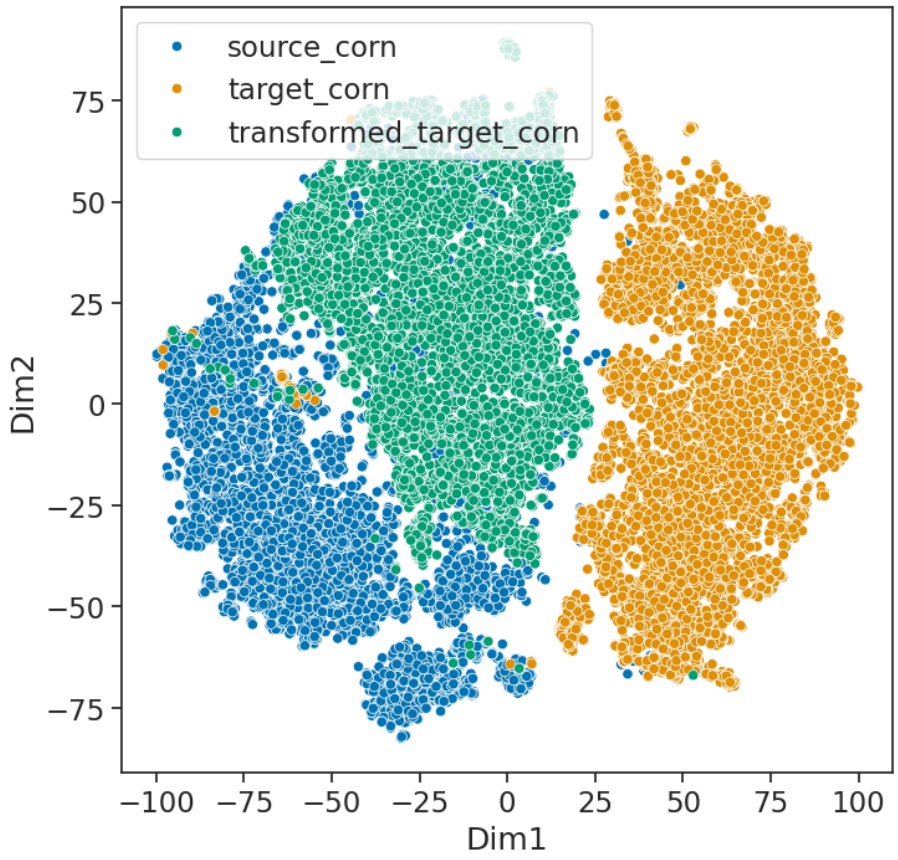


(a)

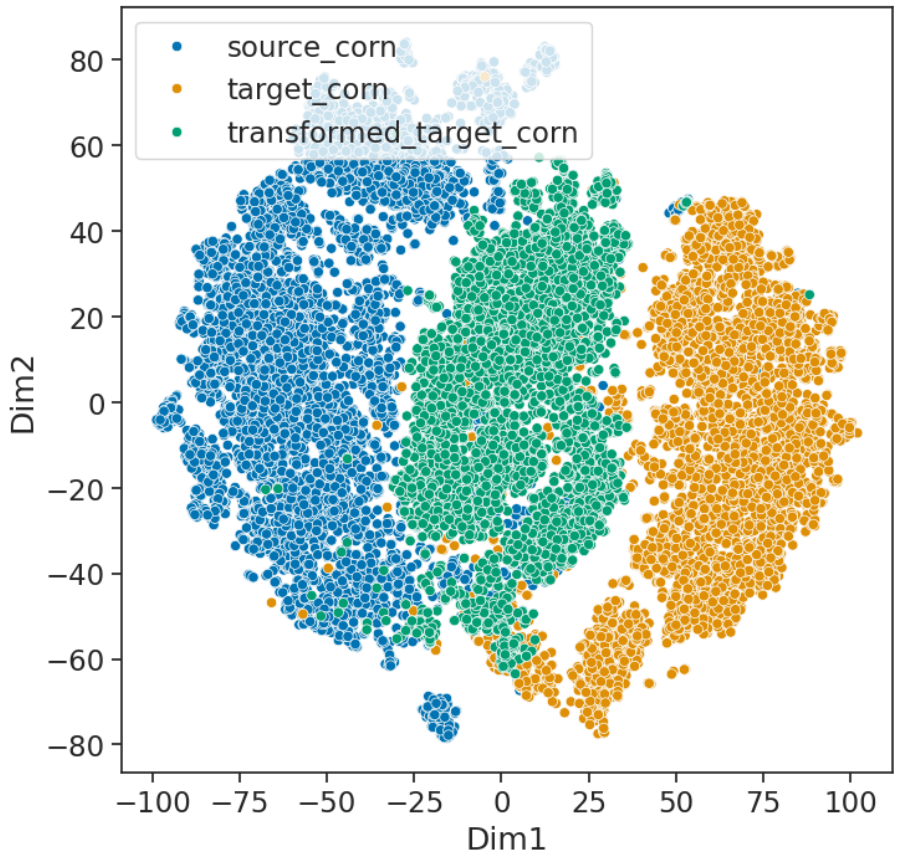


(b)

Figure 4.13: 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.



(a)



(b)

Figure 4.14: 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.

150th day of the year to the 230th day, which is the longest duration tested in these experiments.

4.5.5 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) [62] 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 4.13 and 4.14 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 results than the direct method, but not as good as the results in the cross-year experiments.

4.6 Discussion

4.6.1 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 4.13 and 4.14, 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.

4.6.2 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.

4.6.3 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.

4.7 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.

4.8 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.

5 Cross Domain Early Crop Mapping with Label Spaces Discrepancies using MultiCropGAN

Mapping target crops before the harvest season for regions lacking crop-specific ground truth is critical for global food security. Utilizing multispectral remote sensing and domain adaptation methods, prior studies strive to produce precise crop maps in these regions (target domain) with the help of the crop-specific labelled remote sensing data from the source regions (source domain). However, existing approaches assume identical label spaces across those domains, a challenge often unmet in reality, necessitating a more adaptable solution. This chapter introduces the Multiple Crop Mapping Generative Adversarial Neural Network (MultiCropGAN) model, comprising a generator, discriminator, and classifier. The generator transforms target domain data into the source domain, employing identity losses to retain the characteristics of the target data. The discriminator aims to distinguish them and shares the structure and weights with the classifier, which locates crops in the target domain using the generator's output. This model's novel capability lies in locating target crops within the target domain, overcoming differences in crop type label spaces between the target and source domains. In experiments, MultiCropGAN is benchmarked against various baseline methods. Notably, when facing differing label spaces, MultiCropGAN significantly outperforms other baseline methods. The Overall Accuracy is improved by about 10%.

This chapter is based on the work published in the following research paper:

- Wang, Yiqun, Hui Huang, and Radu State. "Cross Domain Early Crop Mapping with Label Spaces Discrepancies using MultiCropGAN." ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences 10 (2024): 241-248.

Contents

5.1 Overview

Mapping target crops in the early stages before the harvest season is crucial for a variety of agricultural applications such as agricultural planning, resource allocation, crop insurance, and risk management [22, 23]. To locate the target crops, the primary method involves analyzing time-series multispectral images (MSI), while MSI provides detailed spectral information, essential for understanding vegetation's spectral characteristics, influenced by its structural composition, leaf biochemistry, and phenological stages. To generate precise crop cultivation maps from these images, various supervised Deep Learning (DL) methods, including Convolutional Neural Networks (CNN) [31, 32], Temporal Convolutional Neural Network (TempCNN) [19], and Long Short-Term Memory (LSTM) [33, 34], have been explored. Notably, the findings indicate that DL approaches outperform conventional techniques like Support Vector Machine (SVM) [35], Decision Trees (DTs) [36, 37], and Random Forest (RF) [21, 17].

These methods leverage publicly available datasets such as the United States Department of Agriculture (USDA)'s Cropland Data Layer (CDL) [13] as the ground truth (GT).

Unfortunately, the collection of GT for crop types is expensive. In instances where GT data is absent, prevailing approaches adopt the "direct transfer strategy". This strategy involves training a classifier using labelled data from different regions (source domain) and then applying this trained model to regions lacking GT (target domain) [28, 29]. However, the trained model performs badly due to differences in soil composition, climate conditions, and crop progress, leading to discrepancies between the distributions of source and target data, commonly known as cross-domain or domain shift issue [30].

To address domain shift, Domain Adversarial Neural Networks (DANN) [63] and its variants, such as Self-Training with Domain Adversarial Network (STDAN) [14], Phenology Alignment Network (PAN) [15], and Deep Adaptation Crop Classification Network (DACCN) [16], are employed. These models aim to extract invariant features from both target and source domain data, subsequently using these invariant features for crop mapping classification. Alternatively, Generative Adversarial Neural Networks (GAN) [69], such as the Crop Generative Adversarial Network (CropGAN) [18], are utilized. CropGAN transforms time-series MSI data from the target domain to the source domain while preserving local structures. This transformation enables a pre-trained crop mapper classifier, using labelled data from the source domain, to accurately locate the target crop using the transformed target domain data.

However, these methodologies are predicated on the assumption that the crop-type label spaces between the target and source domains are similar. While effective in addressing missing labels and mitigating the adverse effects of domain shifts, these strategies are limited by the disparities in label spaces between the domains. When discrepancies in label spaces occur between these domains, the extractor or generator still endeavors to align the data distributions of both domains. Consequently, some data from one domain must correspond to the other domain data associated with the labels absent in the first domain, leading to misclassification by the crop classifier. To address this label space discrepancies issue, the Multiple Crop Mapping Generative Adversarial Neural Network (MultiCropGAN) is presented in this chapter. Our contributions can be summarized as follows:

- Introduce the MultiCropGAN, comprising a generator, a discriminator, and a classifier, as a solution to mitigate the domain shift issue with label space discrepancies encountered in early crop mapping tasks.
- Propose to incorporate identity losses into the generator's loss function to ensure that the generator refrains from making unnecessary alterations to the data, thereby preserving its essential characteristics.
- Conduct the experiments based on study areas encompassing the USA and Canada. A comparative analysis was undertaken, pitting MultiCropGAN against various baseline methods, including the CropGAN, STDAN, DACCN, TempCNN, and RF. MultiCropGAN demonstrates the highest classification metrics when handling divergent label spaces in the target and source domains.

5.2 Related Works

Within a confined area, specific crop types often exhibit minimal variability in the growth period, offering reliable priors for accurate sample inference. However, outside these labelled regions or domains, significant phenological disparities arise within the same crop type due to variations in environmental conditions. This presents a considerable challenge for cross-domain classification. Existing approaches address this challenge through two distinct perspectives:

From a feature perspective, previous works introduced DANN variant methods to map samples from diverse regions into a shared feature subspace, thereby reducing differences in deep features. For instance, [14] presents the STDAN, a novel unsupervised domain adaptation framework for crop type classification. Moreover, the DACCN [15] and the PAN [16], extended the loss function using the Maximum Mean Discrepancy (MMD) and the Multiple Kernel variant of Maximum Mean Discrepancy (MK-MMD), achieving improved accuracy compared to CNN and LSTM methods without domain adaptation. From a sample perspective, two distinct methods emerge. The first method involves fine-tuning pre-trained models by utilizing a few high-quality samples from the target domain. This process enables the adaptation of the original model to the new distribution. For instance, in [65], deep models were refined for nationwide land cover classification by incorporating pseudo-labels with high confidence. Similarly, in [66], new samples from the target domains were annotated to adjust RF classifiers through active learning. However, this approach often requires labelling additional samples, making it impractical for extensive area research. On the other hand, the second method, exemplified by the CropGAN [18], involves employing a GAN model to learn a mapping function. This function transforms time-series MSI sample data from the target domain to the source domain while preserving local structures. This transformation allows a pre-trained crop mapper classifier, utilizing source domain labelled data, to accurately process the transformed data, thereby enabling high-accuracy crop mapping without labelled data in the target domain.

However, the approaches discussed earlier, regardless of their perspective, face limitations arising from differences in label spaces between the target and source domains. As far as our knowledge extends, we are the pioneering contributors to tackling label space discrepancies from a sample perspective. This is achieved through the utilization of a GAN model combined with specially designed identity losses.

5.3 Problem Statement

Let $\mathbf{X}$ denote the time series remote sensing input data and $\mathbf{Y}$ denote the GT. Each sample $\mathbf{x}$ can be expressed as a temporal form $[\mathbf{x}_1, \mathbf{x}_2, ..., \mathbf{x}_t]$, where $\mathbf{x}_i$ represents input at time i. $\mathbf{x}_i$ can be further expanded as $[\mathbf{x}_{i1}, \mathbf{x}_{i2}, ..., \mathbf{x}_{ib}]$, containing multi-spectral bands information from band 1 to band b. Each GT label, denoted as $\mathbf{y}$, is represented as a one-hot vector comprising four elements, which correspond to the categories of corn, soybean, spring wheat, 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 label space discrepancies issue. The source and target domains correspond to the study areas detailed in Section 5.3.1, while the specific challenges are elaborated upon in Sections 5.3.2 and 5.3.3.

5.3.1 Study Areas

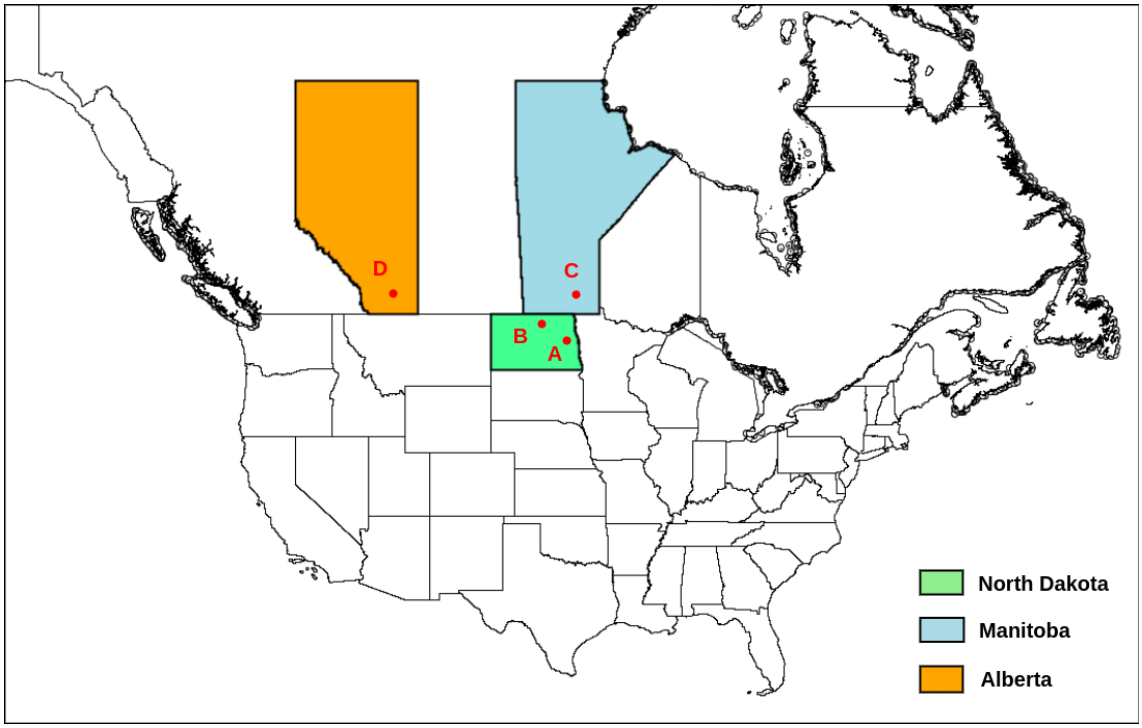
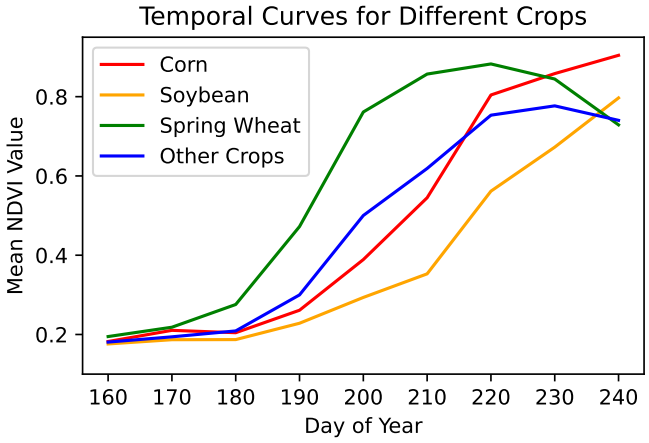
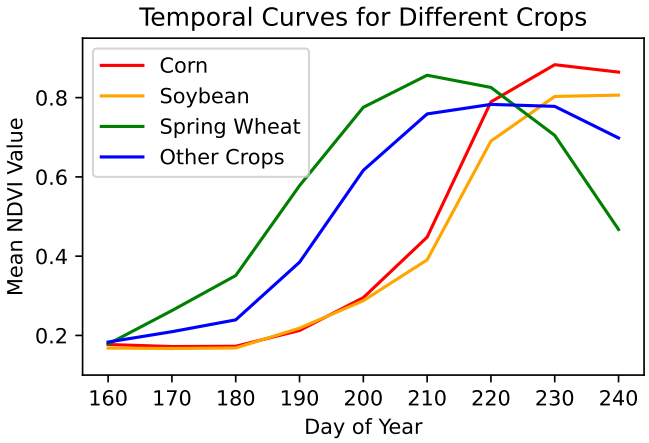


Figure 5.1: The Study Areas in the USA and Canada. The locations of our study areas are denoted by red dots.

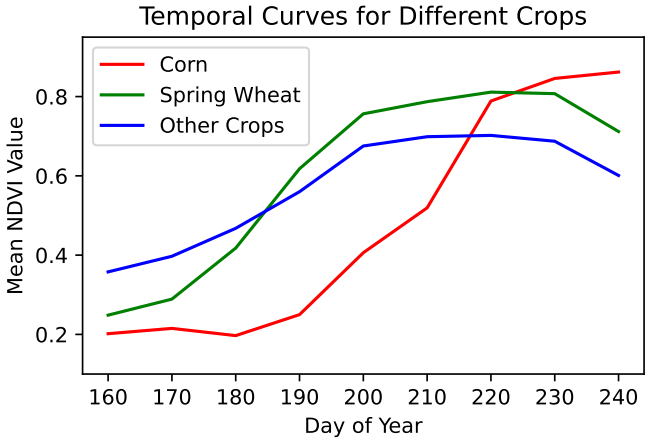
The study areas are located in two countries: the USA and Canada. As depicted in Figure 5.1, Study Areas A (Traill County) and B (Cavalier County) are located in North Dakota, the USA, whereas Study Areas C and D are in Manitoba and Alberta, Canada, respectively. The geographic coordinates for Study Area C range from longitudes −97.97 to −96.59 and latitudes 49.22 to 49.59. Study Area D is defined by longitudes −112.96 to −112.53 and latitudes 49.79 to 50.06.



(a)



(b)



(c)

Figure 5.2: The Temporal NDVI Curves of Different Crops during Their Growing Period in Study Areas. (a) Study Area A and B. (b) Study Area C. (c) Study Area D.

5.3.2 Cross Domain Issue

Table 5.1: The Environmental Conditions of the Study Areas. "S" denotes study areas, "T" annual average temperature, "P" average hourly precipitation, "E" average hourly evaporation rate, "R" surface net solar radiation, and "El" elevation.

S	**T**(K)	**P**(mm/h)	**E**(mm/h)	**R**(kJ/m^2)	**El** (m)
A	287.37	2.04	-1.01	4794.21	287.60
B	284.85	1.95	-0.93	4368.04	470.35
C	285.61	2.12	-0.95	4376.90	237.58
D	285.90	0.68	-0.64	5079.23	896.98

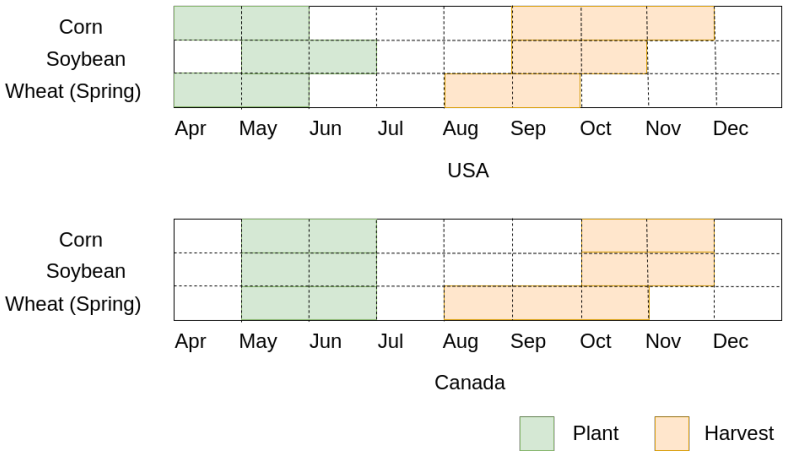


Figure 5.3: The Crop Calendar delineates the planting and harvesting schedules for target crops in the USA (above) and Canada (below).

The Normalized Difference Vegetation Index (NDVI) calculated as a ratio between the red (Red) and near-infrared (NIR) values by

$$\text{NDVI} = \frac{\text{NIR} - \text{Red}}{\text{NIR} + \text{Red}} \tag{5.1}$$

, is usually used to quantify vegetation greenness. Figure 5.2 displays the average NDVI value curves for the target crops and other crop types throughout their growth stages in different study areas. It is important to recognize that variations in NDVI curves for a specific crop across different regions can be attributed to different environmental conditions, as detailed in Table 5.1, and varying crop calendars, as illustrated in Figure 5.3. These factors contribute to distinct phenological characteristics in crops within these regions. Consequently, when a model trained on source domain data is directly applied to the target domain, it frequently underperforms, termed as the cross-domain issue.

5.3.3 Label Space Discrepancies Issue

Current cross-domain methods aim to align the data distribution between the source and target domains to mitigate the domain shift problem, assuming that the crop types between these domains are identical. However, in the context of real-world applications, these domains often display a variety of crop types. For instance, Table

Table 5.2: Crop Types in Study Area A, B, C, and D. "✓" denotes the presence of a crop, while "-" indicates its absence.

Crops	A	B	C	D
Corn	✓	✓	✓	✓
Spring wheat	✓	✓	✓	✓
Soybean	✓	✓	✓	-
Sunflower	✓	✓	✓	-
Barley	✓	✓	✓	✓
Canola	✓	-	✓	✓
Peas	✓	-	-	✓
Sugarbeets	-	✓	-	✓
Dry Beans	✓	✓	-	-
Pasture and Forages	-	-	✓	✓
Canary seed	-	-	✓	-
Oats	-	-	✓	-
Rye	-	-	✓	-
Flax seed	-	-	✓	-
Potatoes	-	-	-	✓
Lentils	-	-	-	✓

5.2 illustrates the diverse crop types present in the study areas. The absence of certain crop types in each domain leads to a significant difference between the data distributions, which the current cross-domain methods cannot resolve, known as the label space discrepancies issue.

5.4 Methodology

To address the cross-domain issue and label space discrepancies, the Multicrop-GAN model is proposed. It transforms the target domain data into the source domain while preserving the characteristics of the target domain data related to crop types absent in the source domain. Figure 5.4 illustrates our proposed MultiCropGAN model, consisting of a generator, a discriminator, and a classifier. The primary goal of the generator is to transform data from the target domain into the source domain. Meanwhile, the discriminator's objective is to differentiate between the transformed target data and the original source data. The classifier and the discriminator are structurally identical and share weights, with the exception of the output layer. Trained with source data and labels, the classifier's purpose is to categorize crop types in the target domain using the transformed target data produced by the generator.

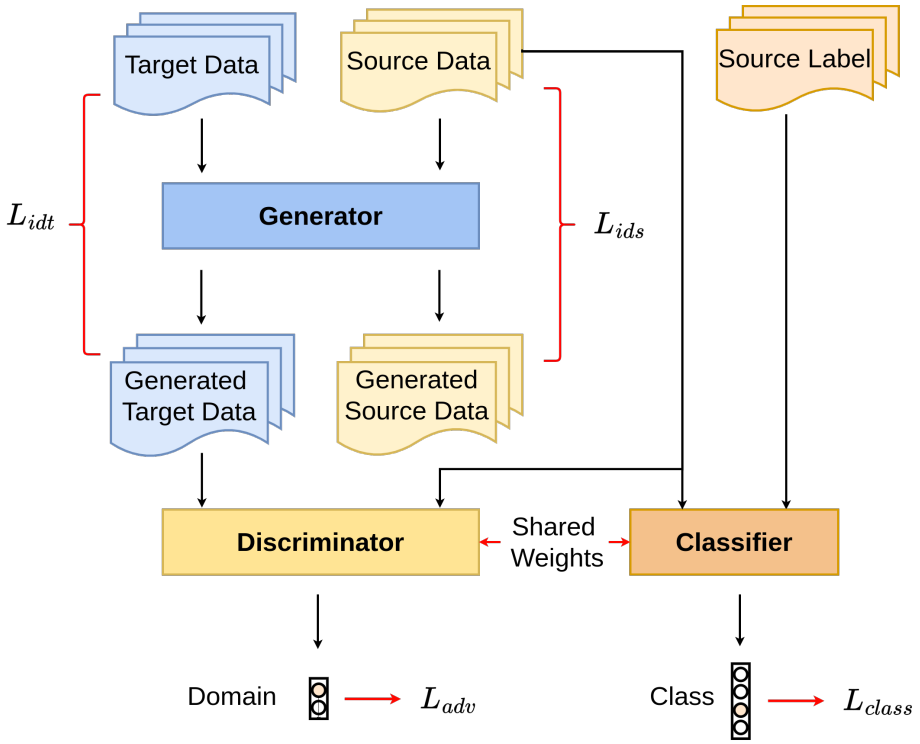


Figure 5.4: The MultiCropGAN Structure with Training Dataflow. It comprises three essential components: the generator, the discriminator, and the classifier.

5.4.1 Data Preprocessing

The preprocessing, shown in Figure 5.5, aims at providing complete time-series remote sensing data by filling gaps between MSI images due to cloud cover, atmospheric interference, or sensor limitations. There are four dataset used in the preprocessing:

- The Sentinel-2 MSI images, widely recognized and employed in numerous agricultural applications within the scientific community [31, 67], are utilized as remote sensing data. Sentinel-2 captures high-resolution MSI images up to 10 meters, with a 5-day revisit time, enabling frequent crop growth monitoring.
- The Dynamic World dataset [68] is utilized, providing detailed class probabilities

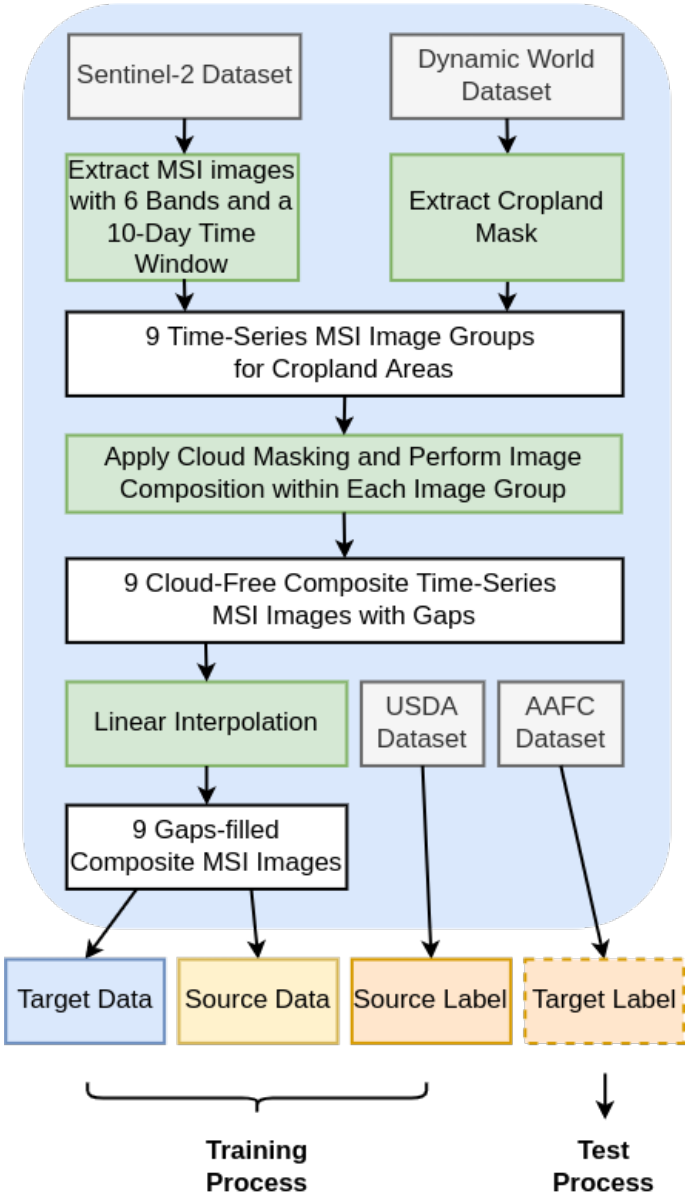


Figure 5.5: The Preprocessing.

and labels with a 10-meter resolution for nine distinct land categories, notably including cropland.

- The CDL [13], a crop-specific land cover raster map accessible for the entire contiguous U.S. land area at a 30-meter resolution, is provided by the USDA. It serves as the training GT for the USA.
- Similarly, the Canada Agriculture and Agri-Food Canada (AAFC) Annual Crop Inventory [70], with a 30-meter spatial resolution, is employed as the GT for Canada in the test process.

The procedure, shown in Figure, starts by acquiring MSI images with six bands, including B2 (Blue), B3 (Green), B4 (Red), B8 (NIR), B11 (Shortwave Infrared 1), and B12 (Shortwave Infrared 2), encompassing the entire designated study areas. These images originate from the Sentinel-2 Dataset and are captured at regular 10-day intervals as an image sequence. Our primary goal is to identify target crops with label space discrepancies and cross-domain problems during an early growth stage. Consequently, the time series remote sensing data should start after the planting period and end before the onset of crop harvest. As depicted in Figure 5.3, the earliest target crop harvesting season commences in August. As a result, nine image groups are compiled, spanning from May 1st to July 30th. Simultaneously, to eliminate non-agricultural lands in the MSI images, the process extracts the cropland mask during the crop growing season from the Dynamic World dataset and reprojects it to maintain a consistent 30-meter resolution. After excluding non-agricultural lands using the cropland mask extracted from the Dynamic World dataset, the MSI images in the nine image groups specifically focus on agricultural lands. Within each image group, Sentinel Hub's cloud detector [71] is utilized to implement cloud masks on the Sentinel-2 MSI images. This merging process retains the mean values of the cloud-free MSI images, resulting in a composite MSI image at a clear, 30-meter resolution. Consequently, a set of nine high-quality time-series MSI images is generated and characterized by their absence of clouds and gaps. These images encompass six bands, all with a spatial resolution of 30 meters.

5.4.2 Model Structure

In the MultiCropGAN framework, the generator operates as an autoencoder. The encoder is composed of three residual blocks, each featuring two one-dimensional convolutional (Conv1D) layers followed by a MaxPooling layer. The Conv1D layers are characterized by 32 filters with a kernel size of 3. Meanwhile, the MaxPooling layers have a pool size of 3 and a stride of 1. The encoder is followed by a flatten layer and a sequence of fully connected layers with dimensions 96, 64, 64, 64, and 96, respectively. The decoder, mirroring the encoder, employs Conv1DTranspose layers, which are specifically designed for transposing a Conv1D layer. This technique reverses the transformation performed by the corresponding Conv1D layers in the encoder, thus aiding in reconstructing the original input features from their encoded state. The discriminator and classifier adopt the same encoder structure as the generator. This is followed by two fully connected layers, featuring dimensions of 32 and 16, respectively. The discriminator's output layer is a fully connected layer with a dimension of 2, designed to distinguish between generated target data and original source data. Conversely, the classifier's output layer is a fully connected layer with a dimension of 4, tasked with identifying target crops (such as corn, soybean, and spring wheat) and other crop types.

5.4.3 Losses

5.4.3.1 Adversarial Loss

The adversarial loss function is defined by

$$\begin{aligned}\mathcal{L}_{\text{adv}}(G, D, \mathbf{X}_t, \mathbf{X}_s) =& \mathbb{E}_{\mathbf{x}_s \sim p_{\text{data}}(\mathbf{x}_s)}[\log D(\mathbf{x}_s)] + \\ & \mathbb{E}_{\mathbf{x}_t \sim p_{\text{data}}(\mathbf{x}_t)}[1 - \log D(G(\mathbf{x}_t))]\end{aligned} \tag{5.2}$$

, which aims to drive the generator to produce transformed data within the target domain that closely resembles real data from the source domain. The discriminator's objective is to distinguish the generated data as fake data, while the generator's goal is to craft realistic data to deceive the discriminator.

5.4.3.2 Identity Loss

The identity losses are defined by

$$\mathcal{L}_{\text{idt}}(G, \mathbf{X}_t) = \mathbb{E}_{\mathbf{x}_t \sim p_{\text{data}}(\mathbf{x}_t)}[||G(\mathbf{x}_t) - \mathbf{x}_t||_1] \tag{5.3}$$

$$\mathcal{L}_{\text{ids}}(G, \mathbf{X}_s) = \mathbb{E}_{\mathbf{x}_s \sim p_{\text{data}}(\mathbf{x}_s)}[||G(\mathbf{x}_s) - \mathbf{x}_s||_1] \tag{5.4}$$

, which encourage 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.

5.4.3.3 Class Loss

The class loss, denoted as the multi-class cross-entropy loss, is defined by the equation:

$$\mathcal{L}_{\text{class}}(C, \mathbf{X}_s, \mathbf{Y}_s) = -\sum \mathbf{y}_s \cdot \log(C(\mathbf{x}_s)) \tag{5.5}$$

This loss function encourages the classifier to extract features that encompass class-related information. These features are shared with the discriminator since the classifier and the discriminator have shared weights. The discriminator utilizes these features to discern whether the data originates from the original source domain or not.

5.4.3.4 Total Loss

The total loss is defined by

$$\begin{aligned}\mathcal{L}_{\text{total}} =& \alpha \mathcal{L}_{\text{adv}}(G, D, \mathbf{X}_t, \mathbf{X}_s) + \\ & \beta(\mathcal{L}_{\text{idt}}(G, \mathbf{X}_t) + \mathcal{L}_{\text{ids}}(G, \mathbf{X}_s)) + \\ & \sigma \mathcal{L}_{\text{class}}(C, \mathbf{X}_s, \mathbf{Y}_s)\end{aligned} \tag{5.6}$$

, where α, β, σ are the weight parameters. The total loss can be expressed as a minimax function:

$$\min_{G,C} \max_{D} \mathcal{L}_{\text{total}}(G, D, C, \mathbf{X}_t, \mathbf{X}_s, \mathbf{Y}_s) \tag{5.7}$$

, where the generator seeks to minimize the loss while the discriminator aims to maximize it.

5.5 Experiments

In this section, our proposed MultiCropGAN is compared with other state-of-the-art (SOTA) algorithms, including CropGAN, STDAN, DACCN, TempCNN, and RF, on two experiments of cross-domain time series early crop classification with discrepancies in label spaces. The classification encompasses four labels, comprising three types of target crops (corn, soybean, and spring wheat) and a category for other crop types.

5.5.1 Experimental Setup

Two experiments are set up in this chapter:

- The first experiment employs study areas A and B as the source domain, while study area C is designated as the target domain. Despite the alignment of target crop types with those in the source domain, variations exist in the types of other crops.
- In the second experiment, study areas A and B function as the source domain, and study area D serves as the target domain, specifically lacking soybean. The variations in the types of other crops are retained.

In the experiments, Our method and SOTA methods are categorized into two groups. MultiCropGAN, CropGAN, STDAN, and DACCN fall under the cross-domain methods category, while TempCNN and RF are classified as direct methods.

5.5.2 Training Setup

The experiment utilizes data sampled from 2019. From the source domain, $5,000$ data points are randomly selected for each of nine existing crop types, as detailed in Table 5.2, amounting to a total of $45,000$ data points. These are balanced and accompanied by source domain labels. In contrast, for each target domain, $150,000$ data points were randomly sampled, lacking label information and resulting in unbalanced data. For the training and evaluation of models, the source data was distributed as follows: 80% was dedicated to training, and the remaining 20% was set aside for evaluation, specifically for the application of early stopping criteria. Notably, all target data were utilized in the training process of the cross-domain methods and testing process of all methods. All methods are trained solely on the training sets to acquire well-trained models for these two experiments. Every compared method was repeatedly trained on each subset from scratch 5 times with the same training configuration. Our MultiCropGAN was trained with $\alpha = 1, \beta = 20, \sigma = 1$ in the loss function. The RF classifier was trained using specific parameters: tree_num = 50 and leaf_size = 15. The optimizer employed in all deep learning methods underwent a substitution, being replaced by the Adam optimizer, initialized with a learning rate of 0.0005 and configured with $\beta = (0.9, 0.998)$. The RF classifier underwent training using the sklearn library, while all deep models were constructed using Tensorflow. Training persisted until the completion of 500 epochs or upon convergence, as determined by an early stopping criterion set at 50 epochs.

5.5.3 Experimental Results

Two metrics, the Overall Accuracy (OA) and Weighted F1 Score, are utilized to assess the performance of the proposed MultiCropGAN model. Tables 5.3 and 5.4 present the OA and F1 score results of our method in comparison to the SOTA.

Table 5.3: Experiment Metrics for the First Experiment: Best metrics are indicated in bold, while the best metrics of the SOTA methods are underlined.

(a)

	MCGAN (Ours)		CropGAN		STDAN	
	OA (%)	F1 (%)	OA	F1	OA	F1
S1	80.90	80.91	<u>74.47</u>	<u>73.89</u>	70.52	69.66
S2	81.86	81.87	70.90	71.29	72.68	71.00
S3	81.64	81.54	72.81	72.02	68.98	67.51
S4	**82.96**	**82.93**	70.48	69.28	73.68	72.76
S5	81.08	81.04	73.39	73.87	64.88	61.77
Avg	**81.69**	**81.66**	<u>72.41</u>	<u>72.07</u>	70.15	68.54

(b)

	DACCN		TempCNN		RF	
	OA (%)	F1 (%)	OA	F1	OA	F1
S1	72.52	72.04	71.44	70.37	67.93	66.90
S2	68.06	67.05	71.06	70.19	68.23	67.11
S3	67.82	65.02	66.76	63.37	66.07	64.71
S4	71.68	71.03	71.80	70.39	66.95	65.59
S5	73.86	72.27	70.90	68.78	67.35	66.09
Avg	70.79	69.48	70.39	68.62	67.31	66.08

Across two experiments, our method demonstrates the highest average OA and F1 score.

In the first experiment, our approach achieved an average OA of 81.69% and an average F1 score of 81.66%. This marked a notable improvement, with a +9.28% increase in OA and a +9.59% boost in F1 score when compared to the CropGAN model's performance, which had an OA of 72.41% and an F1 score of 72.07%, securing the second position in the rankings. Among the cross-domain methods, DACCN claimed the third position, surpassing STDAN in performance. Within the category of direct methods, TempCNN exhibited slightly better results than STDAN but lagged behind DACCN. Notably, RF demonstrated the weakest performance among all the methods, with an OA of 67.31% and an F1 score of 66.08%.

In the second experiment, our approach achieved an OA of 79.69% and an F1 score of 80.54%. In contrast, the RF model secured the second position, attaining an OA of 77.01% and an F1 score of 77.09%. Noteworthy is the performance of the direct method TempCNN, which secured the third position with an OA of 73.53% and an F1 score of 74.32%, outperforming the other three cross-domain methods. Among these cross-domain methods, CropGAN demonstrated superior performance with an OA of 70.91% and an F1 score of 71.35% compared to STDAN, which achieved an OA of 68.75% and an F1 score of 69.09%. Furthermore, STDAN outperformed DACCN, which attained an OA of 67.63% and an F1 score of 67.94%.

The second experiment presents a more intricate scenario concerning discrepancies within the label space, in contrast to the first experiment. This complexity is characterized not only by a diversity of other crop types but also by the conspicuous absence of soybeans among the target crop categories. As depicted in Figure 5.6(a), it is evident that soybeans (depicted in yellow) are not present in the target domain,

Table 5.4: Experiment Metrics for the Second Experiment: Best metrics are indicated in bold, while the best metrics of the SOTA methods are underlined.

(a)

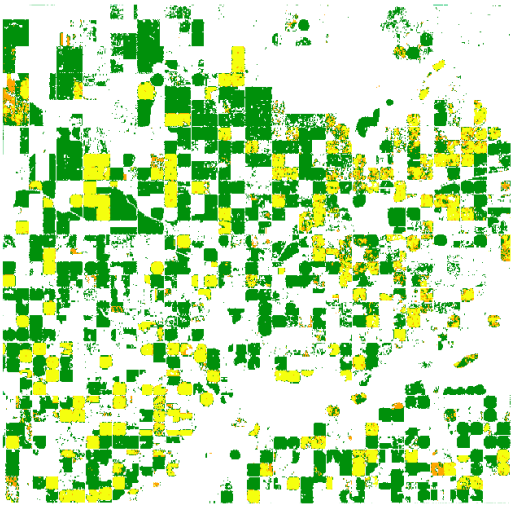
	MCGAN (Ours)		CropGAN		STDAN	
	OA (%)	F1 (%)	OA	F1	OA	F1
S1	79.96	80.27	70.00	69.25	64.19	63.06
S2	**80.89**	80.96	68.78	68.75	71.87	73.01
S3	79.53	80.55	70.94	72.97	69.40	70.76
S4	77.39	79.47	72.82	73.43	69.63	68.69
S5	80.68	**81.45**	72.03	72.37	68.65	69.92
Avg	**79.69**	**80.54**	70.91	71.35	68.75	69.09

(b)

	DACCN		TempCNN		RF	
	OA (%)	F1 (%)	OA	F1	OA	F1
S1	67.24	68.52	71.68	72.35	76.78	76.93
S2	70.01	70.96	73.88	74.41	76.93	77.00
S3	66.89	65.25	72.05	73.26	<u>77.27</u>	<u>77.25</u>
S4	66.94	67.89	74.76	75.52	76.99	77.13
S5	67.07	67.07	75.27	76.07	77.09	77.12
Avg	67.63	67.94	73.53	74.32	<u>77.01</u>	<u>77.09</u>

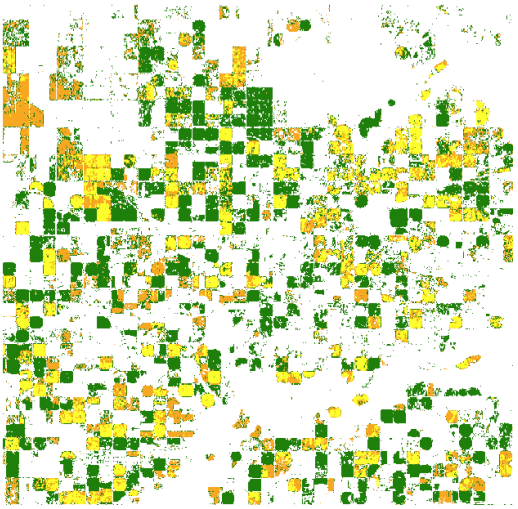
although they exist in the source domain. Figures 5.6 and 5.7 provide visual representations of the results obtained through both our methodology and SOTA approaches, specifically tailored to address the challenges of the second experiment. In the results, it becomes apparent that CropGAN and TempCNN misclassify parts of corn, spring wheat, or other crop types as soybean pixels. Moreover, STDAN, DACCN, and RF exhibit trends of misclassification of corn and spring wheat as other crops.

(a)

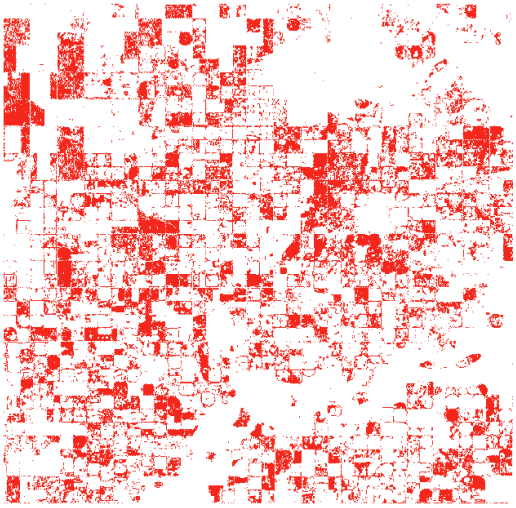


(b)

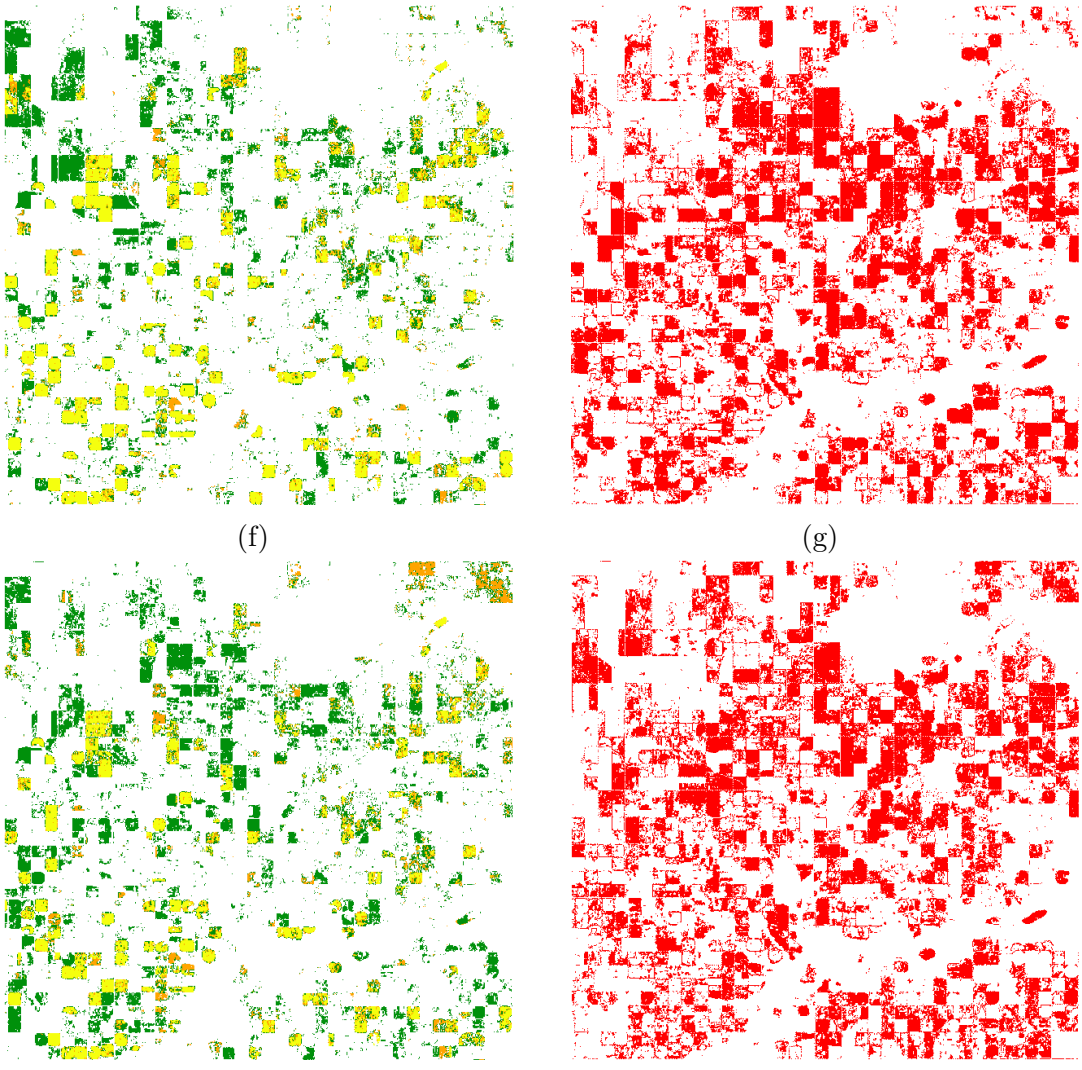
(c)



(d)



(e)



(f) (g)

(h) (i)

Figure 5.6: Visualization of the Second Experiment Results Employing Cross-Domain Deep Learning Methods. In this visualization, yellow denotes corn, orange signifies soybean, green indicates spring wheat, and white represents other crops. (a) displays the GT for crop types. The crop mapping results are depicted in (b) for MultiCropGAN, (d) for CropGAN, (f) for STDAN, and (h) for DACCN. The corresponding error images are illustrated in panels (c), (e), (g), and (i) for MultiCropGAN, CropGAN, STDAN, and DACCN, respectively. Red highlights the misclassified pixels.

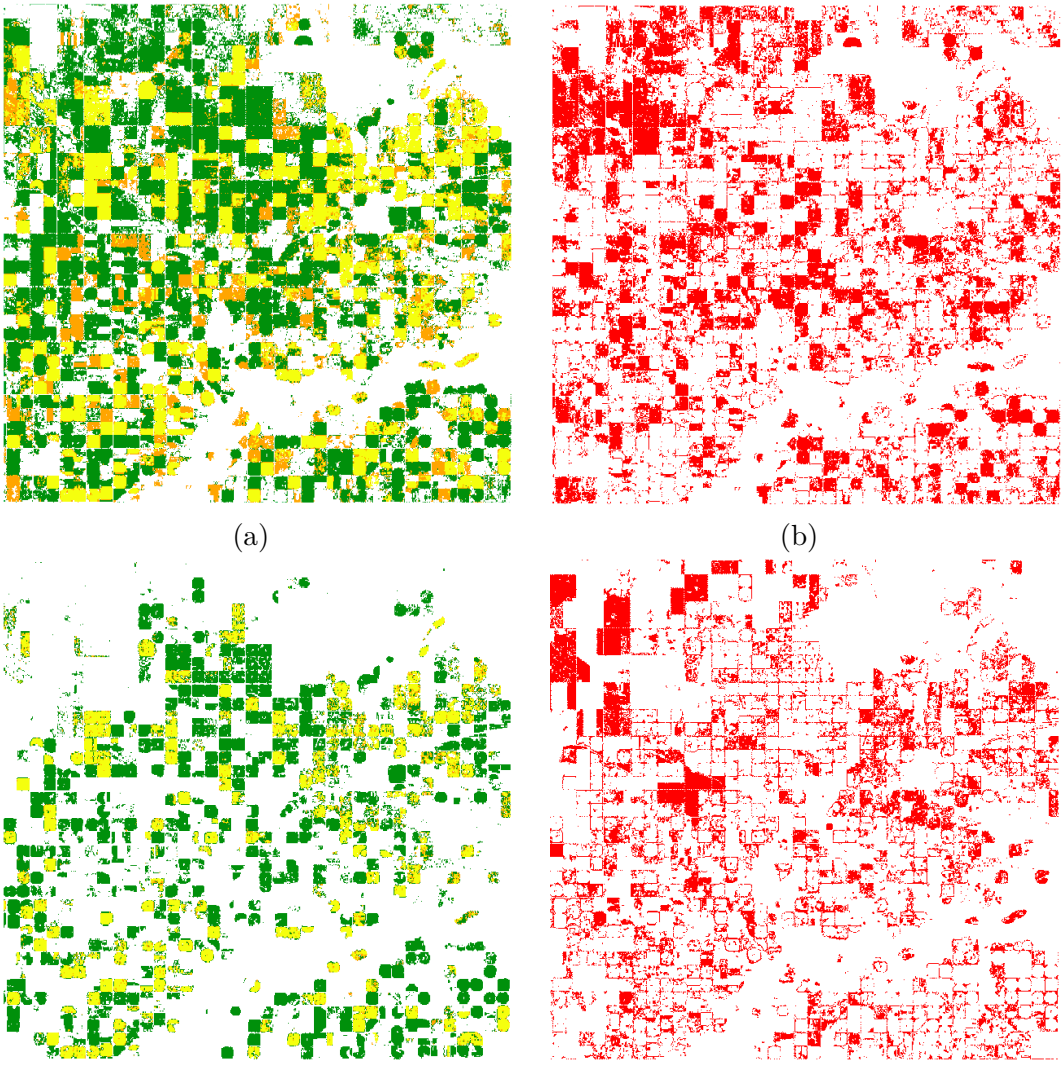


Figure 5.7: Visualization of the Second Experiment Results Employing the TempCNN and RF. The crop mapping results are depicted in (a) for TempCNN, and (c) for RF. The corresponding error images are illustrated in (b) for TempCNN, and (d) for RF.

5.6 Discussion

When there are moderate differences in label spaces across various cross-domain methods, such as those observed in the first experiment, cross-domain deep learning methods exhibit superior performance over TempCNN and RF. Notably, when DACCN incorporates the MMD loss to extract the invariant features more strictly, it demonstrates higher performance than STDAN. CropGAN outperforms both STDAN and DACCN, possibly due to its utilization of the identity loss calculated from the source domain data. Particularly noteworthy is the exceptional performance of MultiCropGAN, surpassing all the other methods, thereby establishing itself as the most effective method in this context.

However, in scenarios characterized by substantial disparities among target crop labels, as evident in the second experiment, TempCNN and RF demonstrated superior performance compared to several cross-domain methods (such as CropGAN, STDAN, and DACCN). This could be attributed to the attempts made by these cross-domain methods to align data distributions between the target and source domains, despite the inherent differences caused by the absence of soybeans in the target domain. Despite DACCN's more rigorous attempts to extract invariant features by employing MMD loss compared to STDAN, it exhibited less effective performance. Notably, CropGAN consistently demonstrated superior performance compared to both STDAN and DACCN. As expected, MultiCropGAN maintained its position as the top-performing approach, surpassing even TempCNN and RF. One of the key contributing factors to MultiCropGAN's success is its use of two identity losses, which ensure that the generator retains crucial data characteristics without unnecessary alterations.

5.7 Conclusion

This chapter introduces MultiCropGAN, an innovative generative adversarial neural network designed to tackle the issue of domain shift resulting from diverse regions and variations in label spaces during early crop mapping, leveraging multi-temporal multispectral input data. The MultiCropGAN model comprises three key components: a generator, a discriminator, and a classifier. Additionally, we introduce identity losses for both target and source domain data to ensure that the generator maintains essential data characteristics throughout the transformation process. The classifier is trained using source domain data and their respective labels, sharing model weights with the discriminator to enable the latter to leverage class-related features for distinguishing between generated target domain data and original source domain data.

Finally, our MultiCropGAN is evaluated against several SOTA methods in scenarios marked by variations in crop label spaces. Experiments conducted across the USA and Canada demonstrated that MultiCropGAN outperforms various SOTA methods, including CropGAN, STDAN, DACCN, TempCNN, and RF. The comparative analysis shows MultiCropGAN achieving the highest classification metrics, particularly effective in handling divergent label spaces in target and source domains. The results show that MultiCropGAN notably enhances classification outcomes for the target domain without utilizing any label information specific to the target domain.

6 Conclusions

In this dissertation, we presented three research focuses encompassing a suite of innovative approaches in remote sensing-based crop mapping, addressing various challenges through advanced methodologies and significantly enhancing the precision and reliability of early crop mapping.

The ECMDCM (Early Crop Mapping using Dynamic Clustering Method) introduces a novel approach to mapping target crops like soybean and corn earlier than the harvest period across the United States. Utilizing time-series NDVI and EVI data, the method employs a dynamic ecoregion clustering technique. The clustering process involves validating the optimal number of ecoregions using both the elbow and silhouette methods, followed by training an ecoregion clustering model using the Kmeans++ method. This model enables the generation of comprehensive ecoregion maps from 2013 to 2022, covering the entire cropland region within the Continental United States (CONUS). The method demonstrates a significantly higher mapping accuracy compared to traditional static clustering methods, illustrating its effectiveness in large-scale agricultural settings.

The CropSTGAN (Crop Spectral-temporal Generative Adversarial Neural Network) framework tackles the challenges posed by cross-domain variability due to inter-region and inter-year differences, which are significant in remote sensing-based crop mapping. The framework includes a pre-processor, a CropSTGAN domain mapper, and a TempCNN crop mapper, all designed to address substantial discrepancies in data distribution across domains. The domain mapper captures and transforms the temporal and spectral features from time-series multispectral imagery (MSI), aligning data from the target domain to closely resemble the source domain. This alignment facilitates the application of models across different agricultural contexts. Extensive testing across various regions and years in the USA and China shows that CropSTGAN outperforms the simpler yet analogous model named CropTGAN and existing methods such as TempCNN, and STDAN, confirming its superior performance and accuracy in managing cross-domain challenges in early crop mapping.

Lastly, the MultiCropGAN (Multiple Crop Mapping Generative Adversarial Neural Network) provides a robust solution to the domain shift issue with label space discrepancies in early crop mapping tasks. This framework is designed to address label inconsistencies across geographic domains and includes a generator, a discriminator, and a classifier. An innovative feature of this framework is the integration of identity losses into the generator's loss function to minimize unnecessary alterations to the data, thus preserving its essential characteristics and enhancing the authenticity of the geographical features. Comparative analysis in diverse agricultural regions of the USA and Canada against established methods such as CropGAN, STDAN, DACCN, TempCNN, and RF shows that MultiCropGAN exhibits superior classification metrics. It is particularly effective in scenarios involving divergent label

spaces between the target and source domains.

These methodologies contribute significantly to the advancement of agricultural remote sensing, highlighting the potential of specialized machine learning models to adapt to various geographic and temporal variations, ensuring more accurate and reliable crop mapping across different settings.

7 Future Works

In this chapter, we present potential future research directions that are in line with this dissertation.

Contents

Based on the methodologies and findings of our three main papers, there are several promising directions for future research that can further enhance the application and effectiveness of remote sensing-based crop mapping. Here, we explore potential advancements in two key areas: Early Warning Systems and Yield Estimation.

7.1 Early Warning System

Building on the dynamic clustering and generative adversarial network approaches developed in our papers, future work could focus on integrating these methods into an Early Warning System for agriculture. The primary goal would be to utilize our models to accurately locate and monitor target crops during the early growing period before harvest. Following the identification of these crops, we can integrate environmental time series data such as temperature, precipitation, and other relevant climatic factors to analyze trends and predict potential anomalies.

The proposed Early Warning System would leverage the spatial and temporal data processing capabilities of our models to forecast adverse conditions that might affect crop health and productivity. By predicting such anomalies early in the crop cycle, this system could provide crucial insights to farmers and agricultural stakeholders, enabling them to take preemptive measures to mitigate risks associated with extreme weather events, pest outbreaks, or other ecological threats. This approach not only enhances crop management practices but also supports sustainable agricultural strategies by reducing potential losses and improving resilience to environmental changes.

7.2 Yield Estimation based on the Early Cropping Results

Another promising area for future research is the application of our established methods in yield estimation. Utilizing the accurate and early identification capabilities of our models, such as the CropSTGAN and MultiCropGAN, future studies could focus on estimating crop yield based on the condition and coverage of crops identified during the early stages of growth. These estimations could then be used to predict market supply and potential price fluctuations, providing valuable information for economic planning and market stability.

The integration of yield estimation processes would involve analyzing the data captured by our models to assess the health and vigor of the crops. This analysis, coupled with historical yield data and current agricultural practices, could be used to model expected yields and anticipate market trends. Such predictive analytics could be invaluable for farmers in managing their crop sales strategies, for governments in planning food security measures, and for investors and businesses in navigating the agricultural market dynamics.

In summary, these future research directions aim to expand the utility of the sophisticated remote sensing technologies developed in our papers. By focusing on practical and impactful applications like Early Warning Systems and Yield Estimation, the advancements can directly contribute to improving agricultural productivity, sustainability, and economic efficiency. These efforts will continue to build on the strong foundation laid by the existing research, pushing the boundaries of what can be achieved with remote sensing and advanced data analytics in agriculture.

List of Papers

Papers included in this dissertation:

- **Wang, Yiqun**, Hui Huang, and Radu State. "Early Crop Mapping Using Dynamic Ecoregion Clustering: A USA-Wide Study." Remote Sensing 15.20 (2023): 4962.
- **Wang, Yiqun**, Hui Huang, and Radu State. "Cross Domain Early Crop Mapping using CropSTGAN." IEEE Access (2024).
- **Wang, Yiqun**, Hui Huang, and Radu State. "Cross Domain Early Crop Mapping with Label Spaces Discrepancies using MultiCropGAN." ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences 10 (2024): 241-248.

Bibliography

[1] L. Gevaux, *3D-hyperspectral imaging and optical analysis of skin for the human face*. PhD thesis, Université de Lyon, 2019.

[2] R. Sonobe, Y. Yamaya, H. Tani, X. Wang, N. Kobayashi, and K.-i. Mochizuki, "Mapping crop cover using multi-temporal landsat 8 oli imagery," *International Journal of Remote Sensing*, vol. 38, no. 15, pp. 4348–4361, 2017.

[3] F. Gao, M. C. Anderson, X. Zhang, Z. Yang, J. G. Alfieri, W. P. Kustas, R. Mueller, D. M. Johnson, and J. H. Prueger, "Toward mapping crop progress at field scales through fusion of landsat and modis imagery," *Remote Sensing of Environment*, vol. 188, pp. 9–25, 2017.

[4] S. Skakun, E. Vermote, J. C. Roger, and B. Franch, "Combined use of landsat-8 and sentinel-2a images for winter crop mapping and winter wheat yield assessment at regional scale," *AIMS geosciences*, vol. 3, no. GSFC-E-DAA-TN49944, 2017.

[5] M. Belgiu and O. Csillik, "Sentinel-2 cropland mapping using pixel-based and object-based time-weighted dynamic time warping analysis," *Remote sensing of environment*, vol. 204, pp. 509–523, 2018.

[6] B. D. Wardlow and S. L. Egbert, "Large-area crop mapping using time-series modis 250 m ndvi data: An assessment for the u.s. central great plains," *Remote Sensing of Environment*, vol. 112, no. 3, pp. 1096–1116, 2008.

[7] 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 Sensing of Environment*, vol. 251, p. 112048, 2020.

[8] P. HAO, H. Tang, Z. CHEN, Y. Le, and M. Wu, "High resolution crop intensity mapping using harmonized landsat-8 and sentinel-2 data," *Journal of Integrative Agriculture*, vol. 18, no. 12, pp. 2883–2897, 2019.

[9] M. Alami Machichi, L. El Mansouri, Y. Imani, O. Bourja, R. Hadria, O. Lahlou, S. Benmansour, Y. Zennayi, and F. Bourzeix, "Cerealnet: A hybrid deep learning architecture for cereal crop mapping using sentinel-2 time-series," in *Informatics*, vol. 9, p. 96, MDPI, 2022.

[10] Y. Cai, K. Guan, J. Peng, S. Wang, C. Seifert, B. Wardlow, and Z. Li, "A high-performance and in-season classification system of field-level crop types using time-series landsat data and a machine learning approach," *Remote sensing of environment*, vol. 210, pp. 35–47, 2018.

[11] S. Skakun, B. Franch, E. Vermote, J.-C. Roger, I. Becker-Reshef, C. Justice, and N. Kussul, "Early season large-area winter crop mapping using modis ndvi data, growing degree days information and a gaussian mixture model," *Remote Sensing of Environment*, vol. 195, pp. 244–258, 2017.

[12] R. Massey, T. T. Sankey, R. G. Congalton, K. Yadav, P. S. Thenkabail, M. Ozdogan, and A. J. S. Meador, "Modis phenology-derived, multi-year distribution of conterminous us crop types," *Remote sensing of Environment*, vol. 198, pp. 490–503, 2017.

[13] C. Boryan, Z. Yang, R. Mueller, and M. Craig, "Monitoring us agriculture: the us department of agriculture, national agricultural statistics service, cropland data layer program," *Geocarto International*, vol. 26, no. 5, pp. 341–358, 2011.

[14] G.-H. Kwak and N.-W. Park, "Unsupervised domain adaptation with adversarial self-training for crop classification using remote sensing images," *Remote Sensing*, vol. 14, no. 18, p. 4639, 2022.

[15] Z. Wang, H. Zhang, W. He, and L. Zhang, "Phenology alignment network: A novel framework for cross-regional time series crop classification," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 2940–2949, American Association for Artificial Intelligence, June 2021.

[16] 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 Journal of Photogrammetry and Remote Sensing*, vol. 199, pp. 102–117, 2023.

[17] Y. Wang, H. Huang, and R. State, "Early crop mapping using dynamic ecoregion clustering: A usa-wide study," *Remote Sensing*, vol. 15, no. 20, p. 4962, 2023.

[18] Y. Wang, H. Huang, and R. State, "Cross domain early crop mapping using cropstgan," *IEEE Access*, 2024.

[19] C. Pelletier, G. I. Webb, and F. Petitjean, "Temporal convolutional neural network for the classification of satellite image time series," *Remote Sensing*, vol. 11, no. 5, p. 523, 2019.

[20] Y. Wang, H. Huang, and R. State, "Cross domain early crop mapping with label spaces discrepancies using multicropgan," *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, vol. 10, pp. 241–248, 2024.

[21] D. C. Duro, S. E. Franklin, and M. G. Dubé, "A comparison of pixel-based and object-based image analysis with selected machine learning algorithms for the classification of agricultural landscapes using spot-5 hrg imagery," *Remote sensing of environment*, vol. 118, pp. 259–272, 2012.

[22] 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 Sensing*, vol. 7, no. 6, pp. 7959–7986, 2015.

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

[24] J. Xue, B. Su, *et al.*, "Significant remote sensing vegetation indices: A review of developments and applications," *Journal of sensors*, vol. 2017, 2017.

[25] 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 Sensing*, vol. 15, no. 8, p. 2014, 2023.

[26] 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, 2019.

[27] 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," *Scientific data*, vol. 8, no. 1, p. 41, 2021.

[28] P. Hao, L. Di, C. Zhang, and L. Guo, "Transfer learning for crop classification with cropland data layer data (cdl) as training samples," *Science of The Total Environment*, vol. 733, p. 138869, 2020.

[29] 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," *International Journal of Applied Earth Observation and Geoinformation*, vol. 102, p. 102451, 2021.

[30] 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 Sensing of Environment*, vol. 251, p. 112048, 2020.

[31] 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," *Computers and electronics in agriculture*, vol. 184, p. 106090, 2021.

[32] M. Hamidi, A. Safari, and S. Homayouni, "An auto-encoder based classifier for crop mapping from multitemporal multispectral imagery," *International Journal of Remote Sensing*, vol. 42, no. 3, pp. 986–1016, 2021.

[33] T. He, C. Xie, Q. Liu, S. Guan, and G. Liu, "Evaluation and comparison of random forest and a-lstm networks for large-scale winter wheat identification," *Remote Sensing*, vol. 11, no. 14, p. 1665, 2019.

[34] H. Crisóstomo de Castro Filho, O. Abílio de Carvalho Júnior, O. L. Ferreira de Carvalho, P. Pozzobon de Bem, R. dos Santos de Moura, A. Olino de Albuquerque, C. Rosa Silva, P. H. Guimaraes Ferreira, R. Fontes Guimarães, and R. A. Trancoso Gomes, "Rice crop detection using lstm, bi-lstm, and machine learning models from sentinel-1 time series," *Remote Sensing*, vol. 12, no. 16, p. 2655, 2020.

[35] A. Mathur and G. M. Foody, "Crop classification by support vector machine with intelligently selected training data for an operational application," *International Journal of Remote Sensing*, vol. 29, no. 8, pp. 2227–2240, 2008.

[36] K. Pittman, M. C. Hansen, I. Becker-Reshef, P. V. Potapov, and C. O. Justice, "Estimating global cropland extent with multi-year modis data," *Remote Sensing*, vol. 2, no. 7, pp. 1844–1863, 2010.

[37] A. Tariq, J. Yan, A. S. Gagnon, M. Riaz Khan, and F. Mumtaz, "Mapping of cropland, cropping patterns and crop types by combining optical remote sensing images with decision tree classifier and random forest," *Geo-Spatial Information Science*, vol. 26, no. 3, pp. 302–320, 2023.

[38] 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," *International Journal of Applied Earth Observation and Geoinformation*, vol. 34, pp. 103–112, 2015.

[39] 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 Sensing*, vol. 7, no. 5, pp. 5347–5369, 2015.

[40] R. Saini and S. K. Ghosh, "Crop classification on single date sentinel-2 imagery using random forest and suppor vector machine," *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, vol. 42, pp. 683–688, 2018.

[41] 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," *International journal of remote sensing*, vol. 40, no. 2, pp. 593–614, 2019.

[42] 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," *International Journal of Applied Earth Observation and Geoinformation*, vol. 98, p. 102313, 2021.

[43] T. Van Klompenburg, A. Kassahun, and C. Catal, "Crop yield prediction using machine learning: A systematic literature review," *Computers and Electronics in Agriculture*, vol. 177, p. 105709, 2020.

[44] J. Clevers, C. Büker, H. Van Leeuwen, and B. Bouman, "A framework for monitoring crop growth by combining directional and spectral remote sensing information," *Remote sensing of environment*, vol. 50, no. 2, pp. 161–170, 1994.

[45] L. Karthikeyan, I. Chawla, and A. K. Mishra, "A review of remote sensing applications in agriculture for food security: Crop growth and yield, irrigation, and crop losses," *Journal of Hydrology*, vol. 586, p. 124905, 2020.

[46] N. Y. Ramirez Cabral, L. Kumar, and F. Shabani, "Global alterations in areas of suitability for maize production from climate change and using a mechanistic species distribution model (climex)," *Scientific reports*, vol. 7, no. 1, pp. 1–13, 2017.

[47] X. Song, P. V. Potapov, A. Krylov, L. King, C. M. Di Bella, A. Hudson, A. Khan, B. Adusei, S. V. Stehman, and M. C. Hansen, "National-scale soybean mapping and area estimation in the united states using medium resolution satellite imagery and field survey," *Remote sensing of environment*, vol. 190, pp. 383–395, 2017.

[48] B. D. Wardlow and S. L. Egbert, "A comparison of modis 250-m evi and ndvi data for crop mapping: a case study for southwest kansas," *International Journal of Remote Sensing*, vol. 31, no. 3, pp. 805–830, 2010.

[49] G. Leng and M. Huang, "Crop yield response to climate change varies with crop spatial distribution pattern," *Scientific Reports*, vol. 7, no. 1, p. 1463, 2017.

[50] K. Choudhary, W. Shi, Y. Dong, and R. Paringer, "Random forest for rice yield mapping and prediction using sentinel-2 data with google earth engine," *Advances in Space Research*, vol. 70, no. 8, pp. 2443–2457, 2022.

[51] L. Zhong, P. Gong, and G. S. Biging, "Efficient corn and soybean mapping with temporal extendability: A multi-year experiment using landsat imagery," *Remote Sensing of Environment*, vol. 140, pp. 1–13, 2014.

[52] W. W. Hargrove and F. M. Hoffman, "Potential of multivariate quantitative methods for delineation and visualization of ecoregions," *Environmental management*, vol. 34, pp. S39–S60, 2004.

[53] F. M. Hoffman, W. W. Hargrove, R. T. Mills, S. Mahajan, D. J. Erickson, and R. J. Oglesby, "Multivariate spatio-temporal clustering (MSTC) as a data mining tool for environmental applications," *4th International Congress on Environmental Modelling and Software (Barcelona, Catalonia, Spain)*, 2008.

[54] L. Zhong, L. Yu, X. Li, L. Hu, and P. Gong, "Rapid corn and soybean mapping in us corn belt and neighboring areas," *Scientific reports*, vol. 6, no. 1, pp. 1–14, 2016.

[55] USDA, "District and County Boundary Maps by State." `https://www.nass.usda.gov/Charts{_}and{_}Maps\hfill\protect\penalty-\@M/Crops{_}County/boundary{_}maps/indexgif.php/`, 2023. [Online; Last Modified: 03/29/2023].

[56] L. Poggio, L. M. De Sousa, N. H. Batjes, G. Heuvelink, B. Kempen, E. Ribeiro, and D. Rossiter, "Soilgrids 2.0: producing soil information for the globe with quantified spatial uncertainty," *Soil*, vol. 7, no. 1, pp. 217–240, 2021.

[57] H. Hersbach, B. Bell, P. Berrisford, S. Hirahara, A. Horányi, J. Muñoz-Sabater, J. Nicolas, C. Peubey, R. Radu, D. Schepers, *et al.*, "The era5 global reanalysis," *Quarterly Journal of the Royal Meteorological Society*, vol. 146, no. 730, pp. 1999–2049, 2020.

[58] J. J. Danielson and D. B. Gesch, "Global multi-resolution terrain elevation data 2010 (gmted2010)," 2011.

[59] N. Gorelick, M. Hancher, M. Dixon, S. Ilyushchenko, D. Thau, and R. Moore, "Google earth engine: Planetary-scale geospatial analysis for everyone," *Remote sensing of Environment*, vol. 202, pp. 18–27, 2017.

[60] B. Bahmani, B. Moseley, A. Vattani, R. Kumar, and S. Vassilvitskii, "Scalable k-means++," *arXiv preprint arXiv:1203.6402*, 2012.

[61] F. Murtagh and P. Contreras, "Algorithms for hierarchical clustering: an overview," *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, vol. 2, no. 1, pp. 86–97, 2012.

[62] L. Van der Maaten and G. Hinton, "Visualizing data using t-sne.," *Journal of machine learning research*, vol. 9, no. 11, 2008.

[63] H. Ajakan, P. Germain, H. Larochelle, F. Laviolette, and M. Marchand, "Domain-adversarial neural networks," *arXiv preprint arXiv:1412.4446*, 2014.

[64] J. Zhu, T. Park, P. Isola, and A. A. Efros, "Unpaired image-to-image translation using cycle-consistent adversarial networks," in *Proceedings of the IEEE international conference on computer vision*, pp. 2223–2232, 2017.

[65] 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 Sensing of Environment*, vol. 237, p. 111322, 2020.

[66] 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 Journal of Photogrammetry and Remote Sensing*, vol. 171, pp. 76–100, 2021.

[67] 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 sensing of environment*, vol. 269, p. 112831, 2022.

[68] 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, *et al.*, "Dynamic world, near real-time global 10 m land use land cover mapping," *Scientific Data*, vol. 9, no. 1, p. 251, 2022.

[69] A. Creswell, T. White, V. Dumoulin, K. Arulkumaran, B. Sengupta, and A. A. Bharath, "Generative adversarial networks: An overview," *IEEE signal processing magazine*, vol. 35, no. 1, pp. 53–65, 2018.

[70] T. Fisette, P. Rollin, Z. Aly, L. Campbell, B. Daneshfar, P. Filyer, A. Smith, A. Davidson, J. Shang, and I. Jarvis, "Aafc annual crop inventory," in *2013 Second International Conference on Agro-Geoinformatics (Agro-Geoinformatics)*, pp. 270–274, IEEE, 2013.

[71] S. Skakun, J. Wevers, C. Brockmann, G. Doxani, M. Aleksandrov, M. Batič, D. Frantz, F. Gascon, L. Gómez-Chova, O. Hagolle, *et al.*, "Cloud mask intercomparison exercise (cmix): An evaluation of cloud masking algorithms for

landsat 8 and sentinel-2," *Remote Sensing of Environment*, vol. 274, p. 112990, 2022.