An Open and Large-Scale Dataset for Multi-Modal Climate Change-aware Crop Yield Predictions

Fudong Lin
University of Delaware
Newark, DE, USA

Kaleb Guillot
University of Louisiana at Lafeyette
Lafeyette, LA, USA

Summer Crawford
University of Louisiana at Lafeyette
Lafeyette, LA, USA

Yihe Zhang
University of Louisiana at Lafeyette
Lafeyette, LA, USA

Xu Yuan*
University of Delaware
Newark, DE, USA

Nian-Feng Tzeng
University of Louisiana at Lafeyette
Lafeyette, LA, USA

ABSTRACT

Precise crop yield predictions are of national importance for ensuring food security and sustainable agricultural practices. While AI-for-science approaches have exhibited promising achievements in solving many scientific problems such as drug discovery, precipitation nowcasting, *etc.*, the development of deep learning models for predicting crop yields is constantly hindered by the lack of an open and large-scale deep learning-ready dataset with multiple modalities to accommodate sufficient information. To remedy this, we introduce the CropNet dataset, the first terabyte-sized, publicly available, and multi-modal dataset specifically targeting climate change-aware crop yield predictions for the contiguous United States (U.S.) continent at the county level. Our CropNet dataset is composed of three modalities of data, *i.e.*, Sentinel-2 Imagery, WRF-HRRR Computed Dataset, and USDA Crop Dataset, for over 2200 U.S. counties spanning 6 years (2017-2022), expected to facilitate researchers in developing versatile deep learning models for timely and precisely predicting crop yields at the county-level, by accounting for the effects of both short-term growing season weather variations and long-term climate change on crop yields. Besides, we develop the CropNet package, offering three types of APIs, for facilitating researchers in downloading the CropNet data on the fly over the time and region of interest, and flexibly building their deep learning models for accurate crop yield predictions. Extensive experiments have been conducted on our CropNet dataset via employing various types of deep learning solutions, with the results validating the general applicability and the efficacy of the CropNet dataset in climate change-aware crop yield predictions. We have officially released our CropNet dataset on Hugging Face Datasets https://huggingface.co/datasets/CropNet/CropNet and our CropNet package on the Python Package Index (PyPI) https://pypi.org/project/cropnet. Code and tutorials are available at https://github.com/fudong03/CropNet.

*Corresponding author: Dr. Xu Yuan (xyuan@udel.edu)

KDD '24, August 25–29, 2024, Barcelona, Spain

ACM ISBN 979-8-4007-0490-1/24/08
https://doi.org/10.1145/3637528.3671536

CCS CONCEPTS

• **Computing methodologies** → **Artificial intelligence**; **Machine learning**.

KEYWORDS

Crop Dataset, Crop Yield Predictions, AI for Science

ACM Reference Format:
Fudong Lin, Kaleb Guillot, Summer Crawford, Yihe Zhang, Xu Yuan, and Nian-Feng Tzeng. 2024. An Open and Large-Scale Dataset for Multi-Modal Climate Change-aware Crop Yield Predictions. In *Proceedings of the 30th ACM SIGKDD Conference on Knowledge Discovery and Data Mining (KDD '24), August 25–29, 2024, Barcelona, Spain.* ACM, New York, NY, USA, 12 pages. https://doi.org/10.1145/3637528.3671536

1 INTRODUCTION

Precise crop yield prediction is essential for early agricultural planning [22], timely management policy adjustment [52], informed financial decision making [1], and national food security [44]. Recent advancements in deep neural networks (DNNs) have achieved impressive performance across various domains [6, 7, 11, 13, 16, 20, 26, 30, 33, 34, 38, 41, 43, 46, 50, 54, 58, 63, 66]. Building upon these advancements, plenty of studies have employed spatial-temporal DNNs [17, 27, 29, 31, 35, 39, 40, 42, 60, 61, 64] to predict crop yields with increased timeliness and precision [8, 12, 14, 22, 23, 32, 62]. However, they often applied their personally curated and limit-sized datasets, with somewhat mediocre prediction performance. There is an urgent need for new large-scale and deep learning-ready datasets tailored specifically for wide use in crop yield predictions.

Recently, some studies [2, 5, 9, 10, 14, 19, 28, 45, 51, 55] have developed open and large-scale satellite imagery (or meteorological parameter) datasets, flexible for being adopted to agricultural-related tasks, *e.g.*, crop type classification [51]. Unfortunately, two limitations impede us from applying them directly to crop yield predictions in general. First, they lack ground-truth crop yield information, making them unsuitable for crop yield predictions. Second, they provide only one modality of data (*i.e.*, either satellite images or meteorological parameters), while accurate crop yield predictions often need to track the crop growth and capture the meteorological weather variation effects on crop yields simultaneously, calling for multiple modalities of data. To date, the development of a large-scale dataset with multiple modalities, targeting specifically for county-level crop yield predictions remains open and challenging.

In this work, we aim to craft such a dataset, called CropNet, the first terabyte-sized and publicly available dataset with multiple

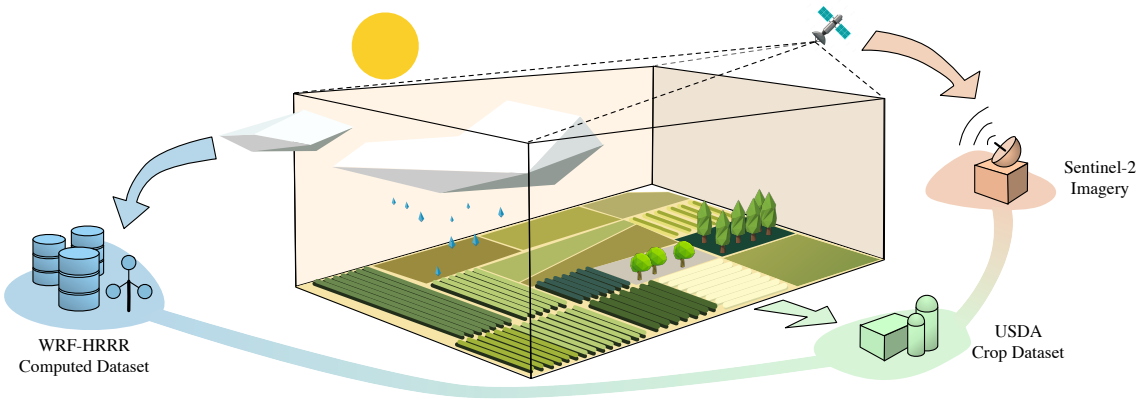


Figure 1: Our CropNet dataset is composed of three modalities of data, *i.e.*, Sentinel-2 Imagery, WRF-HRRR Computed Dataset, and USDA Crop Dataset, providing satellite images, meteorological parameters, and county-level crop yield information, respectively.

Table 1: Dataset comparison

Dataset	Size (GB)	Data Modality
SEVIR [55]	970	satellite imagery
DENETHOR [25]	254	satellite imagery
PASTIS [14]	29	satellite imagery
WorldStrat [9]	107	satellite imagery
RainNet [5]	360	satellite imagery
ENS-10 [2]	3072	meteorological parameters
Our CropNet dataset	2362	satellite imagery meteorological parameters crop information

modalities, designed specifically for county-level crop yield predictions across the United States (U.S.) continent. As shown in Figure 1, the CropNet dataset is composed of three modalities of data, *i.e.*, Sentinel-2 Imagery, WRF-HRRR Computed Dataset, and USDA Crop Dataset, covering a total of 2291 U.S. counties from 2017 to 2022. In particular, the Sentinel-2 Imagery, acquired from the Sentinel-2 mission [47], provides two categories of satellite images, *i.e.*, agriculture imagery (AG) and normalized difference vegetation index (NDVI), for precisely monitoring the crop growth on the ground. The WRF-HRRR Computed Dataset, obtained from the WRF-HRRR model [21], offers daily and monthly meteorological parameters, accounting respectively for the short-term weather variations and the long-term climate change. The USDA Crop Dataset, sourced from the USDA Quick Statistic website [53], contains annual crop yield information for four major crops, *i.e.*, corn, cotton, soybean, and winter wheat, grown on the contiguous U.S. continent, serving as the ground-truth label for crop yield prediction tasks. Table 1 summarizes the dataset comparison between our CropNet dataset and pertinent datasets.

Since the data in our CropNet dataset are obtained from different data sources, we propose a novel data alignment solution to make Sentinel-2 Imagery, WRF-HRRR data, and USDA crop yield data spatially and temporally aligned. Meanwhile, three modalities of data are stored in carefully designed file formats, for improving the accessibility, readability, and storage efficiency of our CropNet dataset. The key advantage of our CropNet dataset is to facilitate researchers in developing crop yield prediction models that are aware of climate change, by taking into account the effects of (1) the short-term weather variations, governed by daily parameters during the growing season, and (2) the long-term climate change, governed by monthly historical weather variations, on crop growth. Furthermore, we have developed the CropNet package, including three types of APIs, expected to assist researchers and practitioners in (1) dynamically downloading the CropNet data based on the specific time and region of interest and (2) flexibly building climate change-aware deep learning models for accurate crop yield predictions at the county level.

Our experimental results validate that the CropNet dataset can be easily adopted by the prominent deep learning models, such as Long Short-Term Memory (LSTM)-based, Convolutional Neural Network (CNN)-based, Graph Neural Network [24] (GNN)-based, and Vision Transformer [11] (ViT)-based models, for timely and precise crop yield predictions. Additionally, our CropNet dataset demonstrates its versatile applicability to boost the generalization capabilities of deep neural networks (DNNs), thanks to its abundant visual satellite imagery and numerical meteorological data.

2 DATA SOURCES

Our CropNet dataset is crafted from three different data sources, as listed below.

Sentinel-2 Mission. The Sentinel-2 mission [47], launched in 2015, serves as an essential earth observation endeavor. With its 13 spectral bands and high revisit frequency of 5 days, the Sentinel-2 mission provides wide-swath, high-resolution, multi-spectral satellite images for a wide range of applications, such as climate change, agricultural monitoring, *etc.*

WRF-HRRR Model [21]. The High-Resolution Rapid Refresh (HRRR) is a Weather Research & Forecasting Model (WRF)-based forecast modeling system, which hourly forecasts weather parameters for the whole United States continent with a spatial resolution of 3km. We take the HRRR assimilated results archived in the University of Utah for use, which provides several crop growth-related parameters, *e.g.*, temperature, precipitation, wind speed, relative humidity, radiation, *etc.*, beginning with July 2016.

USDA. The United States Department of Agriculture (USDA) [53] provides annual crop information for major crops grown in the U.S., including corn, cotton, soybeans, wheat, *etc.*, at the county level. The statistical data include the planted areas, the harvested areas, the production, and the yield for each type of crop, dating back to 1850 at the earliest.

3 OUR CROPNET DATESET

3.1 Motivation

The large-scale data with multiple modalities comprising satellite images, numerical meteorological weather data, and crop yield statistic data, are essential for tracking crop growth and correlating the weather variation's effects on crop yields, to be used for timely and precisely predicting crop yields at the county level. To date, such an open and large-scale dataset intended for county-level crop yield prediction is still absent. In this benchmark article, we plan to design and publish such an open and large-scale dataset, called CropNet, with multiple modalities, consisting of visual satellite images, numerical meteorological parameters, and crop yield statistic data, across the U.S. continent. Notably, not all U.S. counties are suitable for crop planting, so our dataset only includes the data corresponding to 2291 U.S. counties over 3143 counties in total (see Figure 2 for its geographic distribution). Such a multi-modal dataset is valuable for researchers and practitioners to design and test various deep learning models for crop yield predictions, by taking into account the effects of both short-term growing season weather variations and long-term climate change on crop yields.

3.2 Overview of Our CropNet Dataset

Our CropNet dataset is composed of three modalities of data, *i.e.*, Sentinel-2 Imagery, WRF-HRRR Computed Dataset, and USDA Crop Dataset, spanning from 2017 to 2022 (*i.e.*, 6 years) across 2291 U.S. counties. Figure 2 shows the geographic distribution of our dataset. Since crop planting is highly geography-dependent, Figure 2 also provides the number of counties corresponding to each crop type in the USDA Crop Dataset (see the rightmost bar chart). Notably, four of the most popular crops, *i.e.*, corn, cotton, soybeans, and winter wheat, are included in our CropNet dataset, with satellite imagery and the meteorological data covering all 2291 counties. Table 2 overviews our CropNet dataset. Its total size is 2362.6 GB, with 2326.7 GB of visual data for Sentinel-2 Imagery, 35.5 GB of numerical data for WRF-HRRR Computed Dataset, and 2.3 MB of numerical data for USDA Crop Dataset. Specifically, Sentinel-2 Imagery contains two types of satellite images (*i.e.*, AG and NDVI), both with a spatial resolution of around 40 meters (covering an area of 9x9 km with 224x224 pixels) as well as a revisit frequency of 14 days. Figures 3a (or 3b) and 3c (or 3d) respectively depict examples of AG and NDVI images in the summer (or winter). The WRF-HRRR Computed Dataset provides daily (or monthly) meteorological parameters gridded at the spatial resolution of 9 km in a one-day (or one-month) interval. Figures 4a and 4b visualize the temperature in the WRF-HRRR Computed Dataset for the summer and the winter, respectively. The USDA Dataset offers crop information for four types of crops each on the county-level basis, with a temporal resolution of one year. Figure 5 shows the example for the USDA Crop Dataset, depicting 2022 soybeans yields across the U.S. continent.

3.3 Data Collection and Preparation

Sentinel-2 Imagery. We utilize the Sentinel Hub Processing API [48] to acquire satellite images from the Sentinel-2 mission at a processing level of Sentinel-2 L1C, with a maximum allowable cloud coverage of 20%, three spectral bands (*i.e.*, B02, B08, and B11) for AG images and two bands (*i.e.*, B04 and B08) for NDVI images. Satellite images are obtained at the revisit frequency of 14 days instead of the original highest revisit frequency of 5 days. The reason is that the 5-day revisit frequency under our cloud coverage setting results in a large number of duplicate satellite images, according to our empirical study (refer to Appendix A.1 for details). As precisely tracking the crop growth on the ground requires high-spatial-resolution satellite images, we partition a county into multiple grids at the resolution of 9x9 km, with each grid corresponding to one satellite image. Figures 6a and 6b illustrates an example of county partitioning (refer to Appendix A.2 for more details). The downloaded satellite images for one U.S. state (including all counties therein) spanning one season are stored in one Hierarchical Data Format (HDF5) file. Three reasons motivate us to employ the HDF5 file format. First, it can significantly save the hard disk space. That is, the collected satellite images with a total of 4562.2 GB shrank to 2326.7 GB (*i.e.*, 0.51x smaller space occupancy) in the HDF5 file. This can facilitate researchers and practitioners for lower hard disk space requirements and faster data retrieval. Second, it allows for storing data in the form of multidimensional arrays, making satellite images easy to access. The HDF5 file for Sentinel-2 Imagery is organized in the form of (F, T, G, H, W, C), where F represents the FIPS code (*i.e.*, the unique number for each U.S. county) used for retrieving one county's data, T indicates the number of temporal data in a 14-day interval with respect to one season, G represents the number of high-resolution grids for a county, and (H, W, C) are the width, height, and channel numbers for the satellite image. Third, it can store descriptive information for the satellite image, such as its revisit day, the latitude and longitude information it represents, among others.

WRF-HRRR Computed Dataset. The WRF-HRRR Computed Dataset is sourced from the WRF-HRRR model [21], which produces GRID files on the hourly basis, containing meteorological parameters over the contiguous U.S. continent at a spatial resolution of 3x3 km. To lift the domain knowledge required for using the WRF-HRRR data, our CropNet dataset includes 9 carefully chosen and crop growth-relevant meteorological parameters, with 6 parameters obtained directly from the WRF-HRRR model, *i.e.*, averaged temperature, precipitation, relative humidity, wind gust, wind speed, downward shortwave radiation flux, and other 3 parameters computed by ourselves, *i.e.*, maximal temperature, minimal temperature, vapor pressure deficit (VPD). Table 3 presents details of meteorological parameters in the WRF-HRRR Computed Dataset. Notably, VPD describes the difference between the amount of moisture in the air and the maximum amount of moisture the air can hold at a specific temperature, which is an important concept in understanding the environmental conditions that affect plant growth and transpiration. Given two meteorological parameters, *i.e.*, the temperature measured in Kelvin T_K and the relative humidity RH,

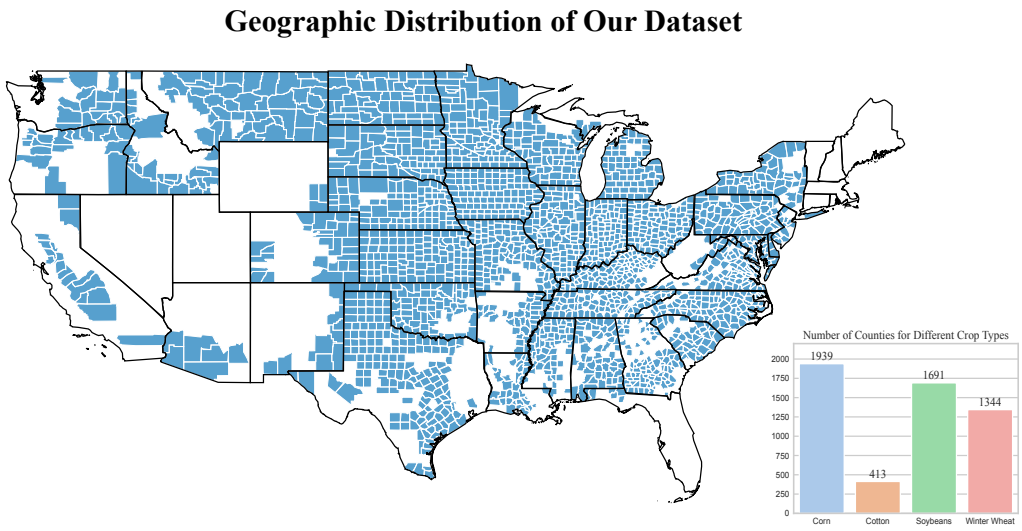
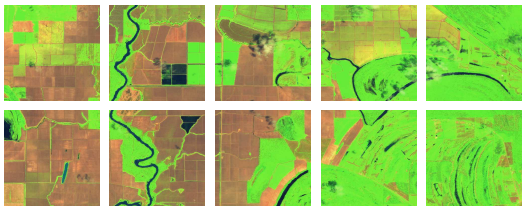


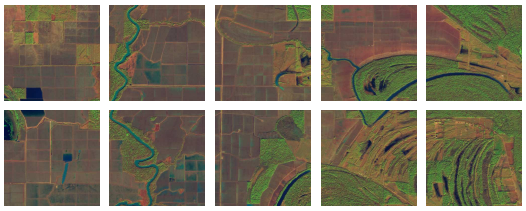
Figure 2: Geographic distribution of our CropNet dataset across 2291 U.S. counties. The rightmost bar chart provides the number of counties corresponding to each of the four crop types included in the USDA Crop Dataset.

Table 2: Overview of our CropNet dataset

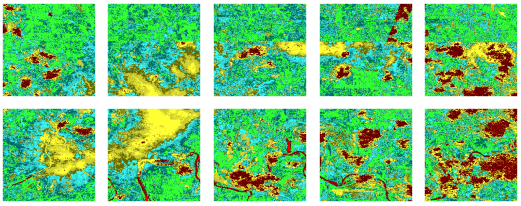
Data Modality	Size	Spatial Resolution	Temporal Resolution	Content
Sentinel-2 Imagery	2326.7 GB	40 m	14 days	satellite images with 224x224 pixels
WRF-HRRR Computed Dataset	35.5 GB	9x9 km	1 day or 1 month	weather parameters
USDA Crop Dataset	2.3 MB	county-level	1 year	crop information



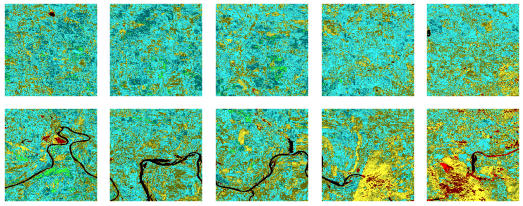
(a) AG images in the summer



(b) AG images in the winter



(c) NDVI images in the summer



(d) NDVI images in the winter

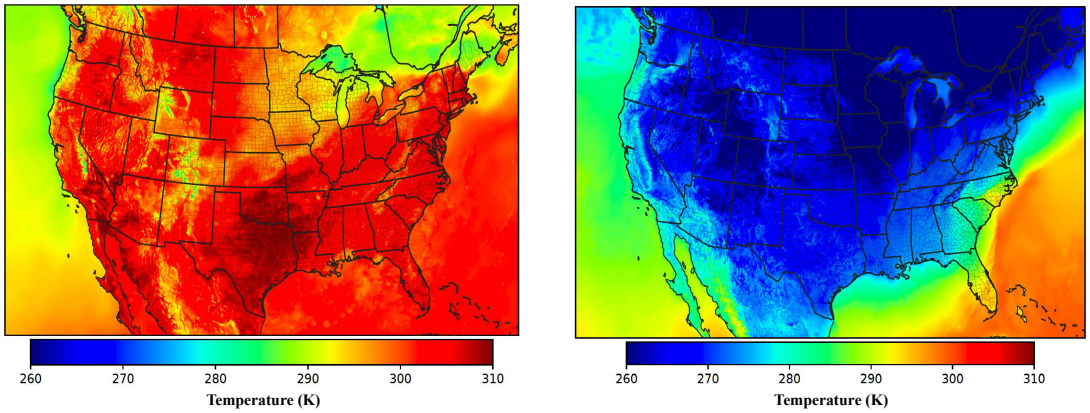
Figure 3: Examples of agriculture imagery (AG, see 3a and 3b) and normalized difference vegetation index (NDVI, see 3c and 3d) in Sentinel-2 Imagery.

VPD is calculated by the following equations:

$$\begin{aligned} T_C &= T_K - 273.15, \\ VP_{\text{sat}} &= \frac{610.7 \times 10^{(7.5\times T_C)/(237.3+T_C)}}{1000}, \\ VP_{\text{air}} &= VP_{\text{sat}} \times \frac{RH}{100}, \\ VPD &= VP_{\text{sat}} - VP_{\text{air}}. \end{aligned} \tag{1}$$

Two challenges impede us from efficiently and effectively extracting meteorological parameters from GRID files. First, the resolution in the WRF-HRRR Computed Dataset should align with the one in the Sentinel-2 Imagery, *i.e.*, 9x9 km[1]. A novel solution is proposed to address this issue. We first follow the Sentinel-2 Imagery by

[1]Note that acquiring satellite images at a spatial resolution of 3x3km is infeasible in practice due to its tremendous space size requirement (*i.e.*, over 20 TB).



(a) Temperature in the summer (b) Temperature in the winter

Figure 4: Examples of the temperature parameters in the WRF-HRRR Computed Dataset.

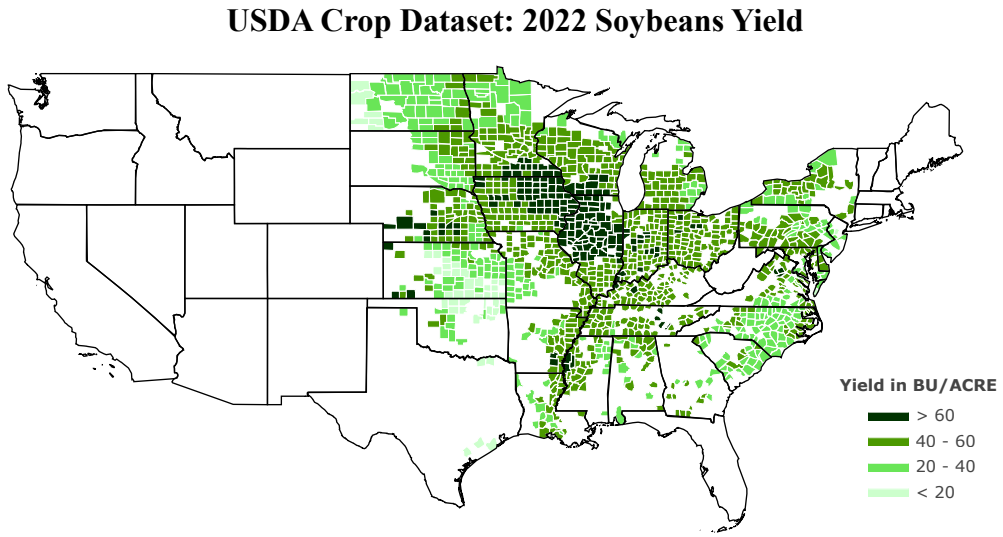
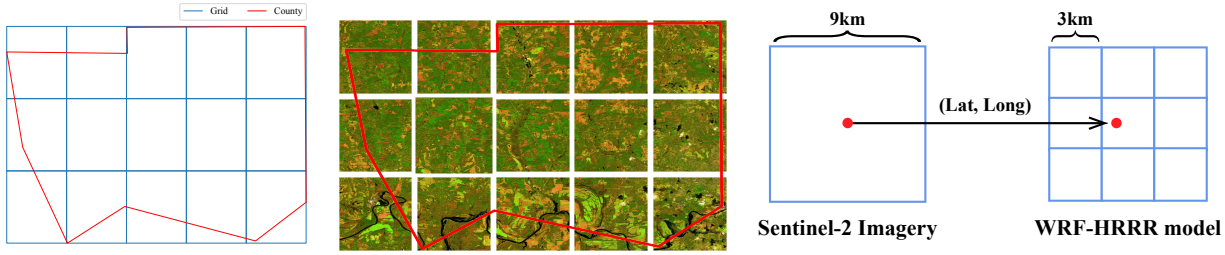


Figure 5: Illustration of USDA Crop Dataset for 2022 soybeans yields.



(a) Boundaries (b) Satellite images (c) Spatial resolution alignment

Figure 6: Illustration of county partitioning (*i.e.*, 6a and 6b) and spatial resolution alignment (*i.e.*, 6c). (a) Boundaries for one county (*i.e.*, the red line) and the corresponding high-resolution grids (*i.e.*, the blue line). (b) Satellite images in the Sentinel-2 Imagery for representing the county. (c) One 3x3km and its surrounding eight grids in the WRF-HRRR model are used for aligning with one 9x9km grid in the Sentinel-2 Imagery.

Table 3: Details of WRF-HRRR Computed Dataset

Source	Parameters	Description
WRF-HRRR model	Averaged Temperature	2 metre averaged temperature during a day/month. Unit: K
	Precipitation	Total precipitation. Unit: kg/m^2
	Relative Humidity	2 metre relative humidity. Unit: %
	Wind Gust	Wind gust on the ground. Unit: m/s
	Wind Speed	Wind speed on the ground. Unit: m/s
	Downward Shortwave Radiation Flux	The total amount of shortwave radiation that reaches the Earth's surface. Unit: W/m^2
Computed by us	Maximal Temperature	2 metre maximal temperature during a day/month. Unit: K
	Minimal Temperature	2 metre minimal temperature during a day/month. Unit: K
	Vapor Pressure Deficit (VPD)	The amount of drying power the air has upon the plant. Unit: kPa

partitioning one county into multiple grids at the spatial resolution of 9x9 km. Then, we utilize the latitude and longitude of the centric point in the 9x9km grid to find the nearest 3x3km grid in the WRF-HRRR model. Next, meteorological parameters in the 3x3 km grid and its surrounding 8 grids can be used for representing a region gridded at 9x9 km, as shown in Figure 6c. In this way, our dataset allows researchers to capture the immediate effects of atmospheric weather variations occurring directly above the crop-growing area on crop yields. Second, extracting meteorological parameters from GRID files is extremely time-consuming as searching the nearest grids requires to match geo-grids across the continental United States. To handle this challenge, we develop a global cache solution by pre-storing the nearest grid information corresponding to a pair of latitude and longitude for each location, reducing the required extraction time from 60 days to 42 days (*i.e.*, 1.42x faster than the one without global caching).

The daily meteorological parameters are computed out of the hourly data extracted from the GRID file, while the monthly weather parameters are derived from our daily data to significantly reduce the frequency of accessing the GRID file. Finally, daily and monthly meteorological parameters are stored in the Comma Separated Values (CSV) file, making them readable by researchers and accessible for deep learning models. The CSV file also includes additional valuable information such as the FIPS code of a county and the latitude and longitude of each grid. This provides easy and convenient access to relevant data for researchers.

USDA Crop Dataset. The data in the USDA Crop Dataset is retrieved from the USDA Quick Statistic website [53] via our newly developed web crawler solution. For each crop type, the USDA website provides its crop information at the county level in a one-year interval, with a unique key for identifying the data for one crop type per year, *e.g.*, "85BEE64A-E605-3509-B60C-5836F6FBB5F6" for the corn data in 2022. Our web crawler first retrieves the unique key by specifying the crop type and the year we need. Then, it utilizes the unique key to obtain the corresponding crop data in one year. Finally, the downloaded crop data is stored in the CSV file. Notably, other useful descriptive information, *e.g.*, FIPS code, state name, county name, *etc.*, are also contained in the CSV file for facilitating readability and accessibility.

However, the crop statistic data from the USDA Quick Statistic website is not deep learning-friendly. For example, it uses two columns, *i.e.*, "Data Item" and "Value", to keep all valuable crop information. That is, if the description of the "Data Item" column refers to the corn yield, then the numerical data in the "Value" column represents the corn yield. Otherwise, the data in "Value" may signify other information, *e.g.*, the corn production, the soybeans yield, *etc.* New data pre-processing techniques are developed to unify the data format, making the production and yield information stored in two independent columns for facilitating Python libraries (*e.g.*, pandas) to access them.

Our CropNet dataset specifically targets county-level crop yield predictions across the contiguous U.S. continent. We utilize the FIPS code to rapidly fetch the data of each county, including a list of HDF5 files for Sentinel-2 Imagery, two lists of CVS files respectively for daily and monthly meteorological parameters, and one CVS file for the USDA Crop Dataset, with configurations stored in the JSON file for increasing accessibility.

4 EXPERIMENTS AND RESULTS

Three scenarios of climate change-aware crop yield predictions, *i.e.*, **Crop Yield Predictions**, **One-Year Ahead Predictions**, and **Self-Supervised Pre-training**, are considered to exhibit the general applicability of our CropNet dataset to various types of deep learning solutions.

4.1 Experimental Settings

Approaches. The LSTM-based, CNN-based, GNN-based, and ViT-based models are represented respectively by **ConvLSTM** [49], **CNN-RNN** [23], **GNN-RNN** [12], and **MMST-ViT** [32] in our experiments, targeting crop yield predictions. Meanwhile, two self-supervised learning (SSL) techniques, *i.e.*, **MAE** [15], and **MM-SSL** in the MMST-ViT, serving respectively as unimodal and multimodal SSL techniques, are applied under the self-supervised pre-training scenario. The aforementioned methods are modified slightly to make them fit the CropNet data in our experiments.

Metrics. Three performance metrics, *i.e.*, **Root Mean Square Error (RMSE)**, **R-squared (R^2)**, and **Pearson Correlation Coefficient (Corr)**, are adopted to evaluate the efficacy of the CropNet dataset for crop yield predictions. Note that a lower RMSE value and a higher R^2 (or Corr) value represent better prediction performance.

Details of utilizing our CropNet data for conducting experiments are deferred to Appendix B for conserving space.

Table 4: Overall performance for 2022 crop yield predictions, where the yield of cotton is measured in pounds per acre (LB/AC) and those of the rest are measured in bushels per acre (BU/AC)

Method	Corn			Cotton			Soybeans			Winter Wheat		
	RMSE (↓)	R^2 (↑)	Corr (↑)	RMSE (↓)	R^2 (↑)	Corr (↑)	RMSE (↓)	R^2 (↑)	Corr (↑)	RMSE (↓)	R^2 (↑)	Corr (↑)
ConvLSTM	19.2	0.795	0.892	56.7	0.834	0.913	5.3	0.801	0.895	6.0	0.798	0.893
CNN-RNN	14.3	0.867	0.923	54.5	0.826	0.899	4.1	0.853	0.915	5.6	0.823	0.906
GNN-RNN	14.1	0.871	0.917	55.1	0.813	0.881	4.1	0.868	0.929	5.3	0.845	0.912
MMST-ViT	13.2	0.890	0.943	50.9	0.848	0.921	3.9	0.879	0.937	4.8	0.864	0.929

4.2 Performance Evaluation for 2022 Crop Yield Predictions

We conduct experiments on the CropNet dataset for 2022 crop yield predictions by using satellite images and daily weather conditions during growing seasons, as well as monthly meteorological conditions from 2017 to 2021, running under the ConvLSTM, CNN-RNN, GNN-RNN, and MMST-ViT models. Table 4 presents each crop's overall performance results (*i.e.*, RMSE, R^2, and Corr) in aggregation. We have two observations. First, all models achieve superb prediction performance with our CropNet data. For example, ConvLSTM, CNN-RNN, GNN-RNN, and MMST-ViT achieve small RMSE values of 5.3, 4.1, 4.1, and 3.9, respectively, for soybeans yield predictions (see the 8th column). These results validate that our CropNet dataset is well-suited for LSTM-based, CNN-based, and GNN-based, and ViT-based models, demonstrating its general applicability. Second, MMST-ViT achieves the best performance results under all scenarios, with lowest RMSE values of 13.2, 50.9, 3.9, and 4.8, as well as highest R^2 (or Corr) values of 0.890 (or 0.943), 0.848 (or 0.921), 0.879 (or 0.937), and 0.864 (or 0.929), respectively for predicting corn, cotton, soybeans, and winter wheat yields. This is due to MMST-ViT's novel attention mechanisms [18, 36, 37, 54, 57], which perform the cross-attention between satellite images and meteorological parameters, able to capture the effects of both growing season weather variations and climate change on crop growth. This experiment exhibits that our CropNet dataset can provide crop yield predictions timely and precisely, essential for making informed economic decisions, optimizing agricultural resource allocation, *etc.*

4.3 Performance of One-Year Ahead Predictions

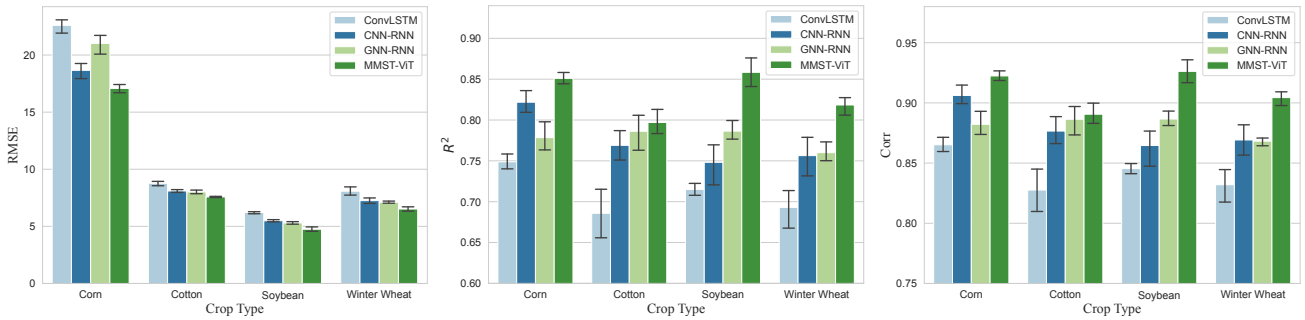
Crop yield predictions well in advance of the planting season are also critical for farmers to make early crop planting and management plans. Here, we apply the CropNet dataset one year before the planting season for predicting the next year's crop yields. Figure 7 shows our experimental results for 2022 crop yield predictions by using our CropNet data during the 2021 growing season. We observe that all models can still maintain decent prediction performance. For instance, ConvLSTM, CNN-RNN, GNN-RNN, and MMST-ViT achieve the averaged RMSE values of 6.2, of 5.4, of 5.3, and of 4.7, respectively, for soybeans predictions. Meanwhile, MMST-ViT consistently achieves excellent Corr values, averaging at 0.922 for corn, 0.890 for cotton, 0.926 for soybeans, and 0.904 for winter wheat predictions, only slightly inferior to the performance results for the regular 2022 crop yield predictions (see the last row in Table 4). This can be attributed to MMST-ViT's ability to capture the indirect influence of 2021's weather conditions on crop growth in the subsequent year through the utilization of long-term weather parameters, which further underscores how our CropNet dataset enhances climate change-aware crop yield predictions.

4.4 Improving the Generalization Capabilities of DNNs

Self-supervised learning (SSL) techniques [3, 4, 15, 56, 59, 65, 67] have significantly advanced the generalization capabilities of deep neural networks (DNNs), especially in vision transformers (ViTs). Our CropNet dataset with a total size of over 2 TB of data can benefit both deep-learning and agricultural communities by providing large-scale visual satellite imagery and numerical meteorological data for pre-training DNNs. To exhibit the applications of our CropNet dataset to self-supervised pre-training, we adopt the MMST-ViT for crop yield predictions by considering three scenarios, *i.e.*, MMST-ViT without the SSL technique (denoted as "w/o SSL"), MMST-ViT with the SSL technique in MAE (denoted as "MAE"), and MMST-ViT with the multi-modal SSL technique proposed in [32] (denoted as "MM-SSL"). Figure 8 illustrates the performance results for four crop types under three performance metrics of interest (*i.e.*, RMSE, R^2, and Corr). We discover that without the SSL technique (*i.e.*, the gray line), the MMST-ViT model exhibits limitations in generalization capabilities, resulting in suboptimal crop yield prediction performance across all tested scenarios. Besides, pre-training MMST-ViT with the SSL technique in MAE (*i.e.*, the blue line) improves its performance results (compared to the "w/o SSL"), with decreased RMSE values by 3.8, 9.6, 1.3, and 1.7 for corn, cotton, soybeans, and winter wheat predictions, respectively. This statistical evidence confirms that our CropNet dataset can improve the generalization capabilities in vision models. Furthermore, MMST-ViT with the multi-modal SSL technique (*i.e.*, the green line) achieves the best performance results under all scenarios. In comparison to the "w/o SSL" scenario, it decreases RMSE values by 6.4, 18.3, 2.6, and 3.6, respectively, for predicting corn, cotton, soybeans, and winter wheat. The effectiveness of the multi-modal SSL technique may stem from its ability to integrate visual satellite imagery with numerical meteorological data found in the CropNet dataset. This integration enhances the generalization capabilities of the MMST-ViT model by improving its ability to effectively discern the influence of weather conditions on crop growth patterns during the pre-training phase.

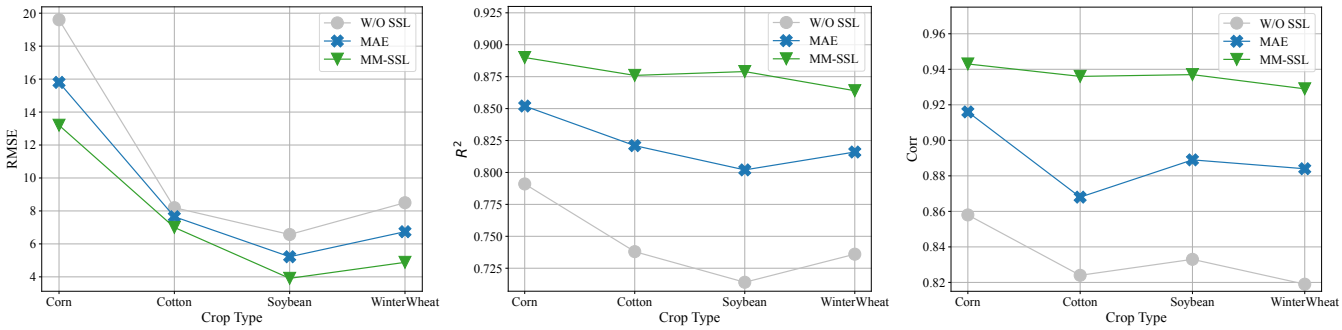
4.5 Significance of Each Modality of Our CropNet Dataset

To show the necessity and significance of each modality data in our CropNet dataset, we examine five scenarios. First, we drop the temporal satellite images (denoted as "w/o temporal images") by randomly selecting only one day's imagery data. Second, we discard the high-resolution satellite image (denoted as "w/o high-resolution



(a) RMSE (b) R^2 (c) Corr

Figure 7: The performance of one-year ahead crop yield predictions, with the cotton yield measured by LB/AC and other crop yields measured by BU/AC. In Figure 7a, we present the square root of the RMSE values for the cotton yield to enhance visualization.



(a) RMSE (b) R^2 (c) Corr

Figure 8: Illustration of how our CropNet dataset benefits self-supervised learning techniques. Notably, Figure 8a depicts the square root of RMSE values for the cotton yield to improve visualization.

Table 5: Ablation studies for different modalities of the CropNet dataset, with five scenarios considered and the last row presenting the results by using all modalities

Modality	Scenario	Corn			Soybeans		
		RMSE (↓)	R^2 (↑)	Corr (↑)	RMSE (↓)	R^2 (↑)	Corr (↑)
Sentinel-2 Imagery	w/o temporal images	22.1	0.758	0.870	5.72	0.773	0.879
	w/o high-resolution images	27.9	0.656	0.810	7.80	0.631	0.794
WRF-HRRR Computed Dataset	w/o WRF-HRRR data	20.6	0.758	0.871	5.78	0.764	0.874
	w/o short-term data	18.6	0.796	0.892	5.04	0.816	0.903
	w/o long-term data	15.3	0.854	0.924	4.72	0.825	0.908
All	–	13.2	0.890	0.943	3.91	0.879	0.937

images") by using only one satellite image to capture the whole county's agricultural information. Third, we ignore the effects of weather variations on crop yields by dropping all meteorological data, denoted as "w/o WRF-HRRR data". Similarly, "w/o short-term data" and "w/o long-term data" represent masking out the daily and monthly meteorological parameters, respectively. We also include prediction results by using all modalities of the CropNet (denoted as "All") for performance comparison. Note that the USDA Crop Dataset provides the label for crop yield predictions; hence, no ablation study requires.

Table 5 presents the experimental results under the MMST-ViT model [32]. We have four observations. First, discarding the temporal satellite images (*i.e.*,"w/o temporal images") degrades performance significantly, raising the RMSE value by 8.9 (or 1.81) and lowering the Corr value by 0.073 (or 0.058) for corn (or soybeans) yield predictions. This is due to that a sequence of satellite images spanning the whole growing season are essential for tracking crop growth. Second, "w/o high-resolution images" achieves the worst prediction performance, with a largest RMSE vaue of 27.9 (or 7.8) and a lowest Corr value of 0.810 (or 0.794) for corn (or soybeans) yield predictions. The reason is that high-resolution satellite images

are critical for precise agricultural tracking. Third, dropping meteorological parameters (*i.e.*, w/o WRF-HRRR data) makes MMST-ViT fail to capture meteorological effects on crop yields, leading to the increase of RMSE value by 7.4 (or 1.87) and the decease of Corr value by 0.072 (or 0.063) for predicting corn (or soybeans) yields. Fourth, discarding either daily weather parameters (*i.e.*,"w/o short-term data") or monthly meteorological parameters (*i.e.*,"w/o long-term data") lowers crop yield prediction performance. The reason is that the former is necessary for capturing growing season weather variations, while the latter is essential for monitoring long-term climate change effects. Hence, we conclude that each modality in our CropNet dataset is important and necessary for accurate crop yield predictions, especially for those crops which are sensitive to growing season weather variations and climate change.

5 THE CROPNET PACKAGE

In addition to our CropNet dataset, we also release the CropNet package, including three types of APIs, at the Python Package Index (PyPI), which is designed to facilitate researchers in developing DNNs for multi-modal climate change-aware crop yield predictions, with its details presented as follows.

```
# Download the 2023 Sentinel-2 Imagery
downloader.download_Sentinel2(fips_codes=["01003"], years
    =["2023"])

# Download the 2023 WRF-HRRR Computed data
downloader.download_HRRR(fips_codes=["01003"], years=["
    2023"])

# Download the 2023 USDA Soybean data
downloader.download_USDA("Soybean", fips_codes=["01003"],
     years=["2023"])
```

Figure 9: Example of our DataDownloader API.

```
# Retrieve the Sentinel-2 Imagery data for two counties
retriever.retrieve_Sentinel2(fips_codes=["01003","01005"
    ],years=["2022"])

# Retrieve the WRF-HRRR Computed data for two counties
retriever.retrieve_HRRR(fips_codes=["01003", "01005"],
    years=["2022"])

# Retrieve the USDA data for two counties
retriever.retrieve_USDA(fips_codes=["01003", "01005"],
    years=["2022"])
```

Figure 10: Example of our DataRetriever API.

DataDownloader. This API allows researchers to download the CropNet data over the time/region of interest on the fly. For example, given the time and region (*e.g.*, the FIPS code for one U.S. county) of interest, Figure 9 presents how to utilize the DataDownloader API to download the up-to-date CropNet data.

DataRetriever. This API enables researchers to conveniently obtain the CropNet data stored in the local machine (*e.g.*, after you have downloaded our curated CropNet dataset) over the time/region of interest, with the requested data presented in a user-friendly format. For instance, Figure 10 shows how to employ the DataRetriever API to obtain the CropNet data for two U.S. counties.

```
from torch.utils.data import DataLoader

# The base directory for the CropNet dataset
base_dir = "/mnt/data/CropNet"
# The JSON configuration file
config_file = "data/soybeans_train.json"

# The PyTorch dataloaders for each modality of data
sentinel2_loader = DataLoader(Sentinel2Imagery(base_dir,
    config_file))
hrrr_loader = DataLoader(HRRRComputedDataset(base_dir,
    config_file))
usda_loader = DataLoader(USDACropDataset(base_dir,
    config_file))
```

Figure 11: The PyTorch example of our DataLoader API.

DataLoader. This API is designed to assist researchers in their development of DNNs for crop yield predictions. It allows researchers to flexibly and seamlessly merge multiple modalities of CropNet data, and then expose them through a DataLoader object after performing necessary data preprocessing techniques. A PyTorch example of using our DataLoader API for training (or testing) DNNs is shown in Figure 11.

6 CONCLUSION

This work presented our crafted CropNet dataset, an open, large-scale, and multi-modal dataset targeting specifically at county-level crop yield predictions across the contiguous United States continent. Our CropNet dataset is composed of three modalities of data, *i.e.*, Sentinel-2 Imagery, WRF-HRRR Computed Dataset, and USDA Crop Dataset, containing high-resolution satellite images, daily and monthly meteorological conditions, and crop yield information, aligned in both the spatial and the temporal domains. Such a dataset is ready for wide use in deep learning, agriculture, and meteorology areas, for developing new solutions and models for crop yield predictions, with the consideration of both the effects of growing season weather variations and climate change on crop growth. Extensive experimental results validate the general applicability of our CropNet dataset to various types of deep learning models for both the timely and one-year ahead crop yield predictions. Besides, the applications of our CropNet dataset to self-supervised pre-training scenarios demonstrate the dataset's versatile utility in improving the generalization capabilities of deep neural networks (DNNs). In addition to our crafted dataset, we have also developed the CropNet package, which allows researchers and practitioners to (1) construct the CropNet data on the fly over the time/region of interest and (2) flexibly build their deep learning models for climate change-aware crop yield predictions. Although our initial goal of crafting the CropNet dataset and developing the CropNet package is for precise crop yield prediction, we believe its future applicability is broad and deserved further exploration. It can benefit the deep learning, agriculture, and meteorology communities, in the pursuit of more interesting, critical, and pertinent applications.

ACKNOWLEDGMENTS

This work was supported in part by NSF under Grants 2019511, 2348452, and 2315613. Any opinions and findings expressed in the paper are those of the authors and do not necessarily reflect the view of funding agencies.

REFERENCES

[1] Javad Ansarifar, Lizhi Wang, and Sotirios V Archontoulis. 2021. An interaction regression model for crop yield prediction. *Scientific reports* (2021).

[2] Saleh Ashkboos, Langwen Huang, Nikoli Dryden, Tal Ben-Nun, Peter Dueben, Lukas Gianinazzi, Luca Kummer, and Torsten Hoefler. 2022. ENS-10: A Dataset For Post-Processing Ensemble Weather Forecasts. In *NeurIPS*.

[3] Hangbo Bao, Li Dong, Songhao Piao, and Furu Wei. 2022. BEiT: BERT Pre-Training of Image Transformers. In *ICLR*.

[4] Ting Chen, Simon Kornblith, Mohammad Norouzi, and Geoffrey E. Hinton. 2020. A Simple Framework for Contrastive Learning of Visual Representations. In *ICML*.

[5] Xuanhong Chen, Kairui Feng, Naiyuan Liu, Bingbing Ni, Yifan Lu, Zhengyan Tong, and Ziang Liu. 2022. RainNet: A Large-Scale Imagery Dataset and Benchmark for Spatial Precipitation Downscaling. In *NeurIPS*.

[6] Ziheng Chen, Fabrizio Silvestri, Gabriele Tolomei, Jia Wang, He Zhu, and Hongshik Ahn. 2022. Explain the explainer: Interpreting model-agnostic counterfactual explanations of a deep reinforcement learning agent. *IEEE Transactions on Artificial Intelligence* (2022).

[7] Ziheng Chen, Fabrizio Silvestri, Jia Wang, He Zhu, Hongshik Ahn, and Gabriele Tolomei. 2022. Relax: Reinforcement learning agent explainer for arbitrary predictive models. In *CIKM*. 252–261.

[8] Minghan Cheng, Xiyun Jiao, Lei Shi, Josep Penuelas, Lalit Kumar, Chenwei Nie, Tianao Wu, Kaihua Liu, Wenbin Wu, and Xiuliang Jin. 2022. High-resolution crop yield and water productivity dataset generated using random forest and remote sensing. *Scientific Data* (2022).

[9] Julien Cornebise, Ivan Orsolic, and Freddie Kalaitzis. 2022. Open High-Resolution Satellite Imagery: The WorldStrat Dataset – With Application to Super-Resolution. In *Neural Information Processing Systems Datasets and Benchmarks Track*.

[10] Adrian Cottam, Xiaofeng Li, Xiaobo Ma, and Yao-Jan Wu. 2024. Large-Scale Freeway Traffic Flow Estimation Using Crowdsourced Data: A Case Study in Arizona. *Journal of Transportation Engineering, Part A: Systems* 150, 7 (2024).

[11] Alexey Dosovitskiy, Lucas Beyer, Alexander Kolesnikov, Dirk Weissenborn, Xiaohua Zhai, Thomas Unterthiner, Mostafa Dehghani, Matthias Minderer, Georg Heigold, Sylvain Gelly, Jakob Uszkoreit, and Neil Houlsby. 2021. An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale. In *ICLR*.

[12] Joshua Fan, Junwen Bai, Zhiyun Li, Ariel Ortiz-Bobea, and Carla P. Gomes. 2022. A GNN-RNN Approach for Harnessing Geospatial and Temporal Information: Application to Crop Yield Prediction. In *AAAI*. 11873–11881.

[13] Jinglun Feng, Liang Yang, Ejup Hoxha, Biao Jiang, and Jizhong Xiao. 2023. Robotic inspection of underground utilities for construction survey using a ground penetrating radar. *Journal of Computing in Civil Engineering* 37, 1 (2023).

[14] Vivien Sainte Fare Garnot and Loïc Landrieu. 2021. Panoptic Segmentation of Satellite Image Time Series with Convolutional Temporal Attention Networks. In *ICCV*.

[15] Kaiming He, Xinlei Chen, Saining Xie, Yanghao Li, Piotr Dollár, and Ross B. Girshick. 2022. Masked Autoencoders Are Scalable Vision Learners. In *CVPR*.

[16] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. 2016. Deep Residual Learning for Image Recognition. In *CVPR*.

[17] Wenchong He, Zhe Jiang, Marcus Kriby, Yiqun Xie, Xiaowei Jia, Da Yan, and Yang Zhou. 2022. Quantifying and Reducing Registration Uncertainty of Spatial Vector Labels on Earth Imagery. In *KDD*, Aidong Zhang and Huzefa Rangwala (Eds.). 554–564.

[18] Wenchong He, Zhe Jiang, Tingsong Xiao, Zelin Xu, Shigang Chen, Ronald Fick, Miles Medina, and Christine Angelini. 2023. A Hierarchical Spatial Transformer for Massive Point Samples in Continuous Space. In *NeurIPS*.

[19] Wenchong He, Zhe Jiang, Chengming Zhang, and Arpan Man Sainju. 2020. CurvaNet: Geometric deep learning based on directional curvature for 3D shape analysis. In *KDD*. 2214–2224.

[20] Yi He, Fudong Lin, Nian-Feng Tzeng, et al. 2021. Interpretable minority synthesis for imbalanced classification. In *IJCAI*.

[21] HRRR. 2023. *The High-Resolution Rapid Refresh (HRRR).* https://home.chpc.utah.edu/~u0553130/Brian_Blaylock/cgi-bin/hrrr_download.cgi

[22] Saeed Khaki, Hieu Pham, and Lizhi Wang. 2021. Simultaneous corn and soybean yield prediction from remote sensing data using deep transfer learning. *Scientific Reports* (2021).

[23] Saeed Khaki, Lizhi Wang, and Sotirios V Archontoulis. 2020. A cnn-rnn framework for crop yield prediction. *Frontiers in Plant Science* (2020).

[24] Thomas N. Kipf and Max Welling. 2017. Semi-Supervised Classification with Graph Convolutional Networks. In *ICLR*.

[25] Lukas Kondmann, Aysim Toker, Marc Rußwurm, Andrés Camero, Devis Peressuti, Grega Milcinski, Pierre-Philippe Mathieu, Nicolas Longépé, Timothy Davis, Giovanni Marchisio, Laura Leal-Taixé, and Xiaoxiang Zhu. 2021. DENETHOR: The DynamicEarthNET dataset for Harmonized, inter-Operable, analysis-Ready, daily crop monitoring from space. In *NuerIPS Datasets and Benchmarks*.

[26] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E. Hinton. 2012. ImageNet Classification with Deep Convolutional Neural Networks. In *NeurIPS*.

[27] Shaojie Li, Yuhong Mo, and Zhenglin Li. 2022. Automated pneumonia detection in chest x-ray images using deep learning model. *Innovations in Applied Engineering and Technology* (2022), 1–6.

[28] Tianyi Li, Joshua Klavins, Te Xu, Niaz Mahmud Zafri, and Raphael Stern. 2023. Understanding driver-pedestrian interactions to predict driver yielding: naturalistic open-source dataset collected in Minnesota. *arXiv preprint arXiv:2312.15113* (2023).

[29] Tianyi Li, Mingfeng Shang, Shian Wang, Matthew Filippelli, and Raphael Stern. 2022. Detecting stealthy cyberattacks on automated vehicles via generative adversarial networks. In *International Conference on Intelligent Transportation Systems (ITSC)*. 3632–3637.

[30] Zhenglin Li, Yangchen Huang, Mengran Zhu, Jingyu Zhang, JingHao Chang, and Houze Liu. 2024. Feature manipulation for ddpm based change detection. *arXiv preprint arXiv:2403.15943* (2024).

[31] Zhenglin Li, Hanyi Yu, Jinxin Xu, Jihang Liu, and Yuhong Mo. 2023. Stock market analysis and prediction using LSTM: A case study on technology stocks. *Innovations in Applied Engineering and Technology* (2023), 1–6.

[32] Fudong Lin, Summer Crawford, Kaleb Guillot, Yihe Zhang, Yan Chen, Xu Yuan, et al. 2023. MMST-ViT: Climate Change-aware Crop Yield Prediction via Multi-Modal Spatial-Temporal Vision Transformer. In *ICCV*.

[33] Fudong Lin, Xu Yuan, Lu Peng, and Nian-Feng Tzeng. 2022. Cascade variational auto-encoder for hierarchical disentanglement. In *CIKM*. 1248–1257.

[34] Fudong Lin, Xu Yuan, Yihe Zhang, Purushottam Sigdel, Li Chen, Lu Peng, and Nian-Feng Tzeng. 2023. Comprehensive transformer-based model architecture for real-world storm prediction. In *ECML-PKDD*. 54–71.

[35] Shun Liu, Kexin Wu, Chufeng Jiang, Bin Huang, and Danqing Ma. 2023. Financial time-series forecasting: Towards synergizing performance and interpretability within a hybrid machine learning approach. *arXiv preprint arXiv:2401.00534* (2023).

[36] Weimin Lyu, Xinyu Dong, Rachel Wong, Songzhu Zheng, Kayley Abell-Hart, Fushen Wang, and Chao Chen. 2022. A Multimodal Transformer: Fusing Clinical Notes with Structured EHR Data for Interpretable In-Hospital Mortality Prediction. In *American Medical Informatics Association Annual Symposium (AMIA)*.

[37] Weimin Lyu, Songzhu Zheng, Tengfei Ma, and Chao Chen. 2022. A Study of the Attention Abnormality in Trojaned BERTs. In *NAACL*. 4727–4741.

[38] Haixu Ma, Donglin Zeng, and Yufeng Liu. 2022. Learning Individualized Treatment Rules with Many Treatments: A Supervised Clustering Approach Using Adaptive Fusion. In *NeurIPS*.

[39] Xiaobo Ma, Abolfazl Karimpour, and Yao-Jan Wu. 2023. Eliminating the impacts of traffic volume variation on before and after studies: a causal inference approach. *Journal of Intelligent Transportation Systems* (2023), 1–15.

[40] Xiaobo Ma, Abolfazl Karimpour, and Yao-Jan Wu. 2024. Data-driven transfer learning framework for estimating on-ramp and off-ramp traffic flows. *Journal of Intelligent Transportation Systems* (2024), 1–14.

[41] Yuhong Mo, Hao Qin, Yushan Dong, Ziyi Zhu, and Zhenglin Li. 2024. Large Language Model (LLM) AI Text Generation Detection based on Transformer Deep Learning Algorithm. *International Journal of Engineering and Management Research* 14, 2 (2024), 154–159.

[42] Zhaobin Mo, Yongjie Fu, and Xuan Di. 2024. PI-NeuGODE: Physics-Informed Graph Neural Ordinary Differential Equations for Spatiotemporal Trajectory Prediction. In *International Conference on Autonomous Agents and Multiagent Systems*. 1418–1426.

[43] Zhaobin Mo, Wangzhi Li, Yongjie Fu, Kangrui Ruan, and Xuan Di. 2022. CVLight: Decentralized learning for adaptive traffic signal control with connected vehicles. *Transportation research part C: emerging technologies* 141 (2022), 103728.

[44] Spyridon Mourtzinis, Paul D Esker, James E Specht, and Shawn P Conley. 2021. Advancing agricultural research using machine learning algorithms. *Scientific reports* (2021).

[45] Kangrui Ruan, Xin He, Jiyang Wang, Xiaozhou Zhou, Helian Feng, and Ali Kebarighotbi. 2024. S2E: Towards an End-to-End Entity Resolution Solution from Acoustic Signal. In *International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. 10441–10445.

[46] Kangrui Ruan, Junzhe Zhang, Xuan Di, and Elias Bareinboim. 2023. Causal Imitation Learning via Inverse Reinforcement Learning. In *ICLR*.

[47] Sentinel-2. 2023. *The Copernicus SENTINEL-2 Mission.* ttps://sentinel.esa.int/web/sentinel/missions/sentinel-2

[48] Sentinel-Hub. 2023. *Sentinel Hub Process API.* https://docs.sentinel-hub.com/api/latest/api/process/

[49] Xingjian Shi, Zhourong Chen, Hao Wang, Dit-Yan Yeung, Wai-Kin Wong, and Wang-chun Woo. 2015. Convolutional LSTM Network: A Machine Learning Approach for Precipitation Nowcasting. In *NeurIPS*.

[50] Han Song, Cong Liu, and Huafeng Dai. 2024. Bundledslam: An accurate visual slam system using multiple cameras. In *Advanced Information Technology, Electronic and Automation Control Conference (IAEAC)*, Vol. 7. 106–111.

[51] Gabriel Tseng, Ivan Zvonkov, Catherine Lilian Nakalembe, and Hannah Kerner. 2021. CropHarvest: A global dataset for crop-type classification. In *Neural Information Processing Systems Datasets and Benchmarks Track*.

[52] Matteo Turchetta, Luca Corinzia, Scott Sussex, Amanda Burton, Juan Herrera, Ioannis Athanasiadis, Joachim M Buhmann, and Andreas Krause. 2022. Learning

long-term crop management strategies with CyclesGym. In *NeurIPS*.
[53] USDA. 2023. *The United States Department of Agriculture (USDA).* https://quickstats.nass.usda.gov
[54] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Lukasz Kaiser, and Illia Polosukhin. 2017. Attention is All you Need. In *NeurIPS*.
[55] Mark S. Veillette, Siddharth Samsi, and Christopher J. Mattioli. 2020. SEVIR : A Storm Event Imagery Dataset for Deep Learning Applications in Radar and Satellite Meteorology. In *NeurIPS*.
[56] Yihe Wang, Yu Han, Haishuai Wang, and Xiang Zhang. 2023. Contrast Everything: A Hierarchical Contrastive Framework for Medical Time-Series. In *NeurIPS*.
[57] Yihe Wang, Nan Huang, Taida Li, Yujun Yan, and Xiang Zhang. 2024. Medformer: A Multi-Granularity Patching Transformer for Medical Time-Series Classification. *arXiv preprint arXiv:2405.19363* (2024).
[58] Zhenyi Wang, Li Shen, Tiehang Duan, Donglin Zhan, Le Fang, and Mingchen Gao. 2022. Learning to learn and remember super long multi-domain task sequence. In *CVPR*. 7982–7992.
[59] Zhenyi Wang, Li Shen, Donglin Zhan, Qiuling Suo, Yanjun Zhu, Tiehang Duan, and Mingchen Gao. 2023. Metamix: Towards corruption-robust continual learning with temporally self-adaptive data transformation. In *CVPR*. 24521–24531.
[60] Zepu Wang, Peng Sun, Yulin Hu, and Azzedine Boukerche. 2022. A novel mixed method of machine learning based models in vehicular traffic flow prediction. In *International Conference on Modeling Analysis and Simulation of Wireless and Mobile Systems*. 95–101.
[61] Zepu Wang, Dingyi Zhuang, Yankai Li, Jinhua Zhao, Peng Sun, Shenhao Wang, and Yulin Hu. 2023. ST-GIN: An uncertainty quantification approach in traffic data imputation with spatio-temporal graph attention and bidirectional recurrent united neural networks. In *International Conference on Intelligent Transportation Systems (ITSC)*. 1454–1459.
[62] Xiaocui Wu, Xiangming Xiao, Jean Steiner, Zhengwei Yang, Yuanwei Qin, and Jie Wang. 2021. Spatiotemporal changes of winter wheat planted and harvested areas, photosynthesis and grain production in the contiguous United States from 2008–2018. *Remote Sensing* (2021).
[63] Donglin Zhan, Yusheng Dai, Yiwei Dong, Jinghai He, Zhenyi Wang, and James Anderson. 2024. Meta-adaptive stock movement prediction with two-stage representation learning. In *SIAM International Conference on Data Mining (SDM)*. 508–516.
[64] Donglin Zhan, Shiyu Yi, Dongli Xu, Xiao Yu, Denglin Jiang, Siqi Yu, Haoting Zhang, Wenfang Shangguan, and Weihua Zhang. 2019. Adaptive Transfer Learning of Multi-View Time Series Classification. *arXiv preprint arXiv:1910.07632* (2019).
[65] Dan Zhang, Fangfang Zhou, Yuwen Jiang, and Zhengming Fu. 2023. MM-BSN: Self-Supervised Image Denoising for Real-World with Multi-Mask based on Blind-Spot Network. In *CVPR Workshop*. 4189–4198.
[66] Zijian Zhang, Yujie Sun, Zepu Wang, Yuqi Nie, Xiaobo Ma, Peng Sun, and Ruolin Li. 2024. Large Language Models for Mobility in Transportation Systems: A Survey on Forecasting Tasks. *arXiv preprint arXiv:2405.02357* (2024).
[67] Fangfang Zhou, Zhengming Fu, and Dan Zhang. 2023. High Dynamic Range Imaging with Context-aware Transformer. In *IJCNN*. 1–8.

OUTLINE

This document provided supplementary materials to support our main paper. Section A provides details of data collection. Section B presents additional experimental settings.

A DETAILS OF DATA COLLECTION

A.1 Significance of Our Cloud Coverage Setting and Revisit Frequency for Sentinel-2 Imagery

This section supplements the main paper by demonstrating the necessity and importance of our cloud coverage setting (*i.e.*, $\leq$ 20%) and revisit frequency (*i.e.*, 14 days) for Sentinel-2 Imagery. Figures 12 and 13 present examples of Sentinel-2 Imagery under the original revisit frequency of 5 days with and without our cloud coverage setting, respectively. Figure 14 illustrates satellites images under our revisit frequency of 14 days and our cloud coverage setting (*i.e.*, $\leq$ 20%).

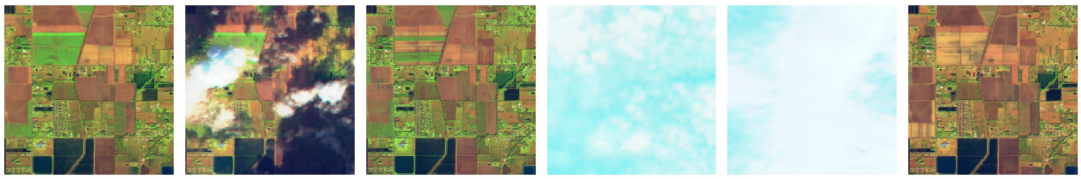
From Figure 12, we observed that the cloud coverage may significantly impair the quality of Sentinel-2 Imagery (see Figures 12b, 12d, and 12e). Worse still, the extreme cases of cloud coverage (refer to Figures 12d and 12e) degrade satellite images into noisy representations. This demonstrates the significance of our cloud coverage setting for discarding low-quality satellite images. Unfortunately, under the original sentinel-2 revisit frequency of 5 days, our cloud coverage setting would result in a large proportion of duplicate satellite images, *e.g.*, 50% (*i.e.*, 3 out of 6 satellite images) as depicted in Figure 13. This is because if the cloud coverage in our requested revisit day exceeds 20%, Processing API [48] will download the most recent available satellite images, whose cloud coverage satisfies our condition (*i.e.*, $\leq$ 20%). In sharp contrast, extending the revisit frequency from 5 days to 14 days markedly decreases the occurrence of duplicate satellite images. For example, there are no duplicate satellite images observed in Figure 14. Hence, our revisit frequency of 14 days for Sentinel-2 Imagery is necessary as it can significantly improve storage and training efficiency.

A.2 County Partitioning

In our main paper, we have introduced partitioning one county into multiple high-spatial-resolution grids for precise agricultural tracking. Here, we provide the details for such a partition. A naive way to achieve this is to expand a county's geographic boundary to a rectangle area by using its maximal and minimal latitude and longitude, and then evenly divide such a rectangle area into multiple grids. Unfortunately, such a partition solution may result in a large number of grids outside the county polygon for some large counties (see Figure 15a). To handle this matter, we develop a novel solution by dropping the grids outside the county's boundary (see Figure 15b). Compared to the naive solution, our solution enjoys two advantages. First, it can significantly reduce the disk space storage size. Take Coconino County in Arizona for example, by employing our solution, its total number of grids degrades from 1023 to 729, which is 0.71x less than that from the naive solution. Second, our solution can evade the negative effect incurred by regions outside the county's boundary on crop yield predictions.

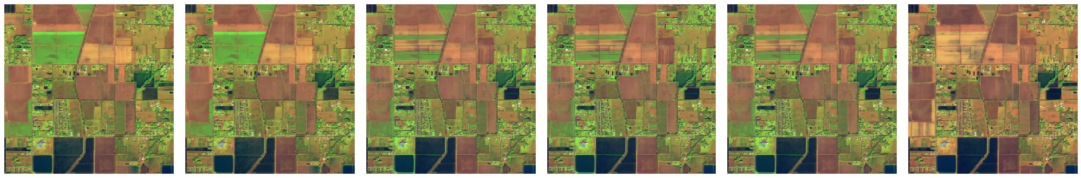
B SUPPORTING EXPERIMENTAL SETTINGS

CropNet Data. Due to the limited computational resources, we are unable to conduct experiments across the entire United States. Consequently, we extract the data with respect to five U.S. states, *i.e.*, Illinois (IL), Iowa (IA), Louisiana (LA), Mississippi (MS), and New York (NY), to exhibit the applicability of our crafted CropNet dataset for county-level crop yield predictions. Specifically, two of these states (*i.e.*, IA and IL) serve as representatives of the Midwest region, two others (*i.e.*, LA and MS) represent the Southeastern region, and the fifth state (*i.e.*, NY) represents the Northeastern area. Four of the most popular crops are studied in this work, *i.e.*, corn, cotton, soybeans, and winter wheat. For each crop, we take the aligned Sentinel-2 Imagery and the daily data in the WRF-HRRR Computed Dataset during growing seasons in our CropNet dataset, respectively for precise agricultural tracking and for capturing the impact of growing season weather variations on crop growth. Meanwhile, the monthly meteorological parameters from the previous 5 years are utilized for monitoring and quantifying the influence of climate change on crop yields.



(a) 2022-12-01 Cloud: 0% (b) 2022-12-06 Cloud: 35.8% (c) 2022-12-11 Cloud: 0% (d) 2022-12-16 Cloud: 97.2% (e) 2022-12-21 Cloud: 100% (f) 2022-12-26 Cloud: 2.7%

Figure 12: Examples of Sentinel-2 Imagery under the original revisit frequency of 5 days without our cloud coverage setting, with the revisit date and the cloud coverage listed below each image.



(a) 2022-12-01 Original (b) 2022-12-06 Duplicate (c) 2022-12-11 Original (d) 2022-12-16 Duplicate (e) 2022-12-21 Duplicate (f) 2022-12-26 Original

Figure 13: Examples of Sentinel-2 Imagery under the original revisit frequency of 5 days and our cloud coverage setting. The revisit date is listed below each image. "Duplicate" (or "Original") indicates whether the satellite image is duplicate (or not) under our cloud coverage setting.

(a) 2022-12-01 Original (b) 2022-12-15 Original (c) 2023-01-01 Original (d) 2023-01-15 Original (e) 2023-02-01 Original (f) 2023-02-15 Original

Figure 14: Examples of Sentinel-2 Imagery under our revisit frequency of 14 days and our cloud coverage setting, with the revisit date listed below each image. We would like to highlight that there are no duplicate satellite images observed.

(a) Naive solution (b) Our solution

Figure 15: Difference between the naive solution and our solution. (a) The naive solution leads to a significant number of grids falling outside the county's polygon. (b) By using our solution, the boundaries of grids (*i.e.*, the blue line) align perfectly with the county's boundary (*i.e.*, the red line).