A Conditional Generative Adversarial Networks-Augmented Case-Based Reasoning Framework for Crop Yield Predictions with Time-Series Remote Sensing Data

Yanbing Bai[1], Dong Chen[2] and Ziyue Zhang[1,*]

[1]Center for Applied Statistics, School of Statistics, Renmin University of China, Beijing BJ 100872, China
[2]China State Information Center, Beijing BJ 100038, China

Abstract

The lack of remote sensing images during the crop harvest period has posed a major challenge in estimating crop yields. This study overcomes this challenge by using Conditional Generative Adversarial Networks-augmented Case-Based Reasoning Framework (CGANa-CBR) to generate remote sensing images of farmland during the harvest period. Specifically, the study used the CGANa-CBR model to generate remote sensing images of farmland and then used these generated images to supplement the real remote sensing images that lacked harvest period data, enabling data augmentation. Subsequently, a convolutional neural network (CNN) model was trained to improve the accuracy of average yield prediction. The results showed that the CNN model incorporating both real data and CGANa-CBR-generated data achieved better predictive performance, with an average reduction of 6.3% in RMSE compared to the baseline CNN model trained only on real remote sensing data. The study also found that training duration and the amount of data used significantly impacted model performance, suggesting the need for further research in this area.

Keywords

remote sensing imagery, Conditional Generative Adversarial Networks (CGAN), Case-Based Reasoning Framework, predictions of average yields on farmland

1. Introduction

Agricultural monitoring holds immense importance for developing countries as it can help prevent famines and support humanitarian efforts. However, crop yield estimation has been one of the major challenges, particularly predicting crop yields before harvest. Existing methods rely on survey data and other variables related to crop growth, such as weather and soil characteristics, to estimate crop yields. These methods have been successful in the developed countries, where data is abundant, of relatively high quality, and weather parameters and land cover types are comprehensively surveyed and publicly available, greatly facilitating yield prediction tasks. However, in developing countries, there is often a lack of information on weather, soil characteristics, and precise land cover data, despite the fact that these countries are most in need of reliable yield predictions.Existing research has shown significant discrepancies between future crop yield and food demand forecasts, making it difficult for decision-makers to formulate appropriate food policies. There is still a great demand for more accurate crop yield estimates [1].

Multi-spectral satellite images have relatively high spatial and temporal resolution, containing abundant information related to vegetation growth and agricultural outcomes. These satellite images cover the entire globe, with openly accessible data, eliminating issues of data accessibility. Consequently, yield prediction based on satellite images is a cheap and efficient method. Nevertheless, due to the high-dimensional and unstructured nature of the data, extracting useful features can be challenging. In the past, manual feature extraction methods were commonly used for yield estimation, but these methods tended to extract coarse features. With the advancement of deep learning, methods based on

ICCBR AI Track'24: Special Track on AI for Socio-Ecological Welfare at ICCBR2024, July 1, 2024, Mérida, Mexico
*Corresponding author.
ybbai@ruc.edu.cn (Y. Bai); chendong@sic.gov.cn (D. Chen); 2022103700@ruc.edu.cn (Z. Zhang)

convolutional neural networks (CNNs) and recurrent neural networks (RNNs) have shown efficient performance in this area [2, 3]. While researchers have achieved good results in predicting crop yields using remote sensing data, the performance of models often falls short when the time span of remote sensing data is short or lacks yield data during harvest periods.

While researchers have achieved good results in predicting crop yields using remote sensing data, the performance of models often falls short when the time span of remote sensing data is short or lacks yield data during harvest periods. This may be due to inadequate collection of plant growth information from remote sensing image sequences. Additionally, in previous studies predicting crop yields using remote sensing images, models were typically based on simple CNN or RNN structures. Furthermore, there have been relatively few studies on predicting crop yields before harvest.

To address the above-mentioned issues, this study employs a Conditional Generative Adversarial Networks-augmented Case-Based Reasoning Framework (CGANa-CBR) model, which generates realistic images by learning the relationship between real and generated images. The model is utilized to generate image data of future farmland, which is then used in conjunction with convolutional neural networks for crop yield prediction. Compared to previous methods that used simple CNN or RNN structures for crop yield prediction, the approach employed in this study is expected to better capture the complex variations in farmland environments and plant growth trends.

The main contributions of this paper are as follows:

- This research combined the CBR framework with the CGAN network, achieving better predictive performance in the task of completing unharvested remote sensing satellite images.
- Inspired by the work of Suman Ravuri et al. [4] on precipitation rate prediction tasks, we apply the CGANa-CBR model to generate future remote sensing images based on past remote sensing images. In practical application, it was found that this model performed exceptionally well in specific prediction tasks.

The rest of the paper is structured as follows. Section 2 provides an overview of background information and related work. In Section 3, a brief overview of the dataset and the preprocess of data. Section 4 outlines model architecture implemented in the research. The training outcomes and the evaluation of the model are presented in Section 5. Lastly, Section 6 serves as the conclusion of the paper.

2. Related Work

2.1. Crop Yield Estimation Model

Estimating crop yields presents a significant challenge in implementing effective monitoring, influenced by a multitude of factors such as weather and soil conditions, crop types, fertilizer usage, and seed variety [5, 6]. The effects of these elements differ across various stages of crop development. Actual data on soil types, solar radiation, precipitation, and temperature fluctuations, along with various assumptions, are incorporated into models for seed development and plant growth. However, most mechanistic models are tailored to specific crops [7]. Recent research has shown promising results with the use of semi-empirical crop models [8]. Despite their effectiveness, these models tend to be costly in terms of both time and financial resources, making them impractical for widespread application and strategic agricultural planning.

In recent years, an increasing number of researchers have turned to machine learning to supplement or replace these methods [9]. It is worth noting that while related to yield prediction, plant count estimation is distinct; accurately predicting corn field yields requires an understanding of both plant counts and the yield per plant. In recent years, Khaki et al. [10] used DNNs to predict the performance of corn hybrids across the United States from a dataset that included detailed hybrid information (including genetic markers) and environmental data such as weather. Barbosa et al. [11] relied on machinery data, such as planting, spraying, and harvesting information, without any imaging data, to predict field yields. With the rapid advancement of machine learning technologies in recent years, remote sensing techniques have seen significant progress.The convenience, wide coverage, and rich

information content of remote sensing imagery have made it a powerful tool for improving the accuracy and stability of yield estimation.

2.2. Application of Machine Learning in Crop Yield Prediction.

Machine learning, as a mature field within computer science, has demonstrated broad prospects across various research areas. Given the enormous and ever-growing volume of agriculture-related data, machine learning techniques can significantly aid in analyzing these large datasets. Elisa Kamir [12]. utilized climate records and satellite image time series as inputs and, upon comparing the effectiveness of several models, found that Support Vector Regression demonstrated the highest learning efficiency in estimating wheat yield. Nath [13] predicted the yield of wheat in India using the Box-Jenkins AutoRegressive Integrated Moving Average model ARIMA(1,1,0). Hansanee Fernando [14] created random forest regression models to predict canola yield using remotely sensed flowering information from PlanetScope satellite imagery combined with derived soil and topography parameters. Parthsarthi Pandya [15] developed cotton and groundnut crop yield prediction models using multiple linear regression , artificial neural networks with Multilayer Perceptron, and random forest.

Deep learning, a significant branch of machine learning, has achieved successful applications in various domains, including computer vision, medicine, and conversational agents [16, 17]. Recently, its application has extended to the agricultural sector, primarily focusing on crop classification [18], prediction, and yield estimation [19]. Jie Sun [20] developed a deep convolutional neural network-long short-term memory (CNN-LSTM) model for predicting soybean yields , trained on crop growth variables and environmental factors including weather, MODIS Land Surface Temperature (LST), and MODIS Surface Reflectance (SR) data. Shruti Kulkarni [21] developed a data-driven model that leverages historical soil and rainfall data to analyze and predict crop yield across different seasons in various districts.

2.3. Generative Adversarial Networks

Generative Adversarial Networks [22] (GANs), as the state-of-the-art generative models for image synthesis, have also been gradually applied in the agricultural sector. Feng Yang [23] developed a corn variety yield prediction model based on a Generative Adversarial Network and Graph Attention Network to fill in missing trait attributes and achieve corn variety yield prediction. Similarly, Ahmed Ali Gomaa [24] employed both convolutional neural network-based and generative adversarial network-based classification models for the early identification of plant infection, prior to the onset of severe disease symptoms, at different life stages of a tomato plant infected with Tomato Mosaic Virus. However,in the vast array of studies on crop yield prediction, the use of GANs models is rare. Given that GANs can generate realistic data, they are perfectly suited to fill the gaps in the data required for crop yield prediction. Therefore, this study employs a CGANa-CBR model, which learns the relationship between real and generated images to produce lifelike images. This model will generate future farmland image data and further use convolutional neural networks for crop yield prediction.

3. Data Source and Preprocessing

3.1. Data Source

This research leverages a diverse dataset that includes soybean yield data from counties across the United States, boundary vector data for states and counties, and remote sensing imagery. The yield data, provided by the U.S. Department of Agriculture, details soybean production in bushels per acre, alongside essential identifiers such as state, county, year, and yield. Boundary vectors, derived from the 2010 U.S. Census, aid in delineating areas for exporting MODIS (Moderate Resolution Imaging Spectroradiometer) remote sensing data via Google Earth Engine.

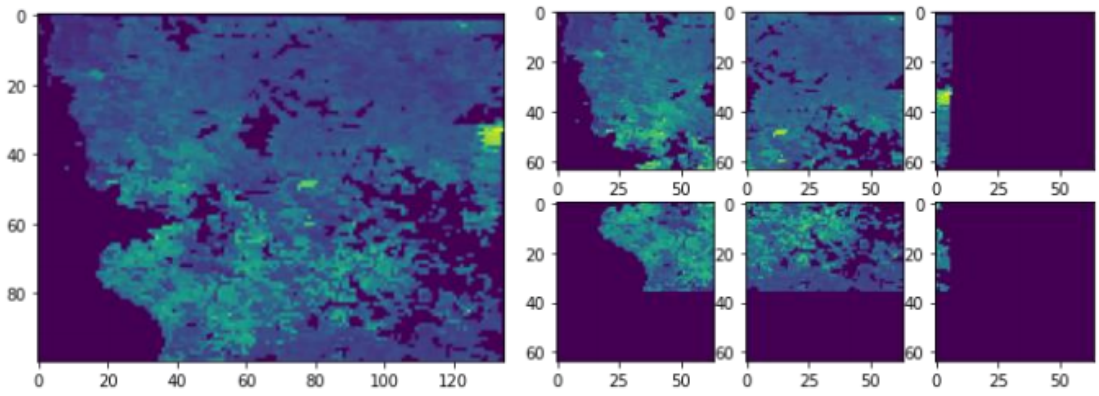

Figure 1: Left: Original image; Right: The original image cropped into 64x64 patches with zero-padding around patches smaller than 64.

The utilized remote sensing dataset encompasses 9 bands, offering estimates of surface spectral reflectance across 7 bands, with the remaining 2 bands delivering land surface and cloud temperature insights. Data collection occurs every 8 days, beginning on the first day of each year, yielding 46 samples annually.

Focusing on 11 states which collectively contribute over 75% to the U.S. soybean output, the study spans from 2003 to 2015. For remote sensing data, county boundaries facilitate area delineation, while the Global Land Cover dataset, accessed through Google Earth Engine, ensures the exclusion of non-crop pixels.

3.2. Data Preprocessing

To ensure consistent image shapes as input to the model for predicting agricultural remote sensing images, the image is first cropped spatially in units of 64 x 64 during training. For parts of the image dimensions that are not divisible by 64, the remaining portions are padded with zeros (Figure 1). A total of 2113 x 13 = 27469 sets of image sequences will be used for training, with each image sequence containing remote sensing images of all 9 bands. Each image sequence is divided into shapes of 46 x 9 x 64 x 64.

Furthermore, the problem addressed in this study is to conduct more accurate yield prediction before the harvesting season, with the remote sensing images during the harvesting period being the desired output of the model. For soybean planting, the planting period is approximately from the 49th day to the 305th day of each year, and the agricultural data around this period do not contain crop growth information. Therefore, each set of remote sensing images is temporally cropped in this study, removing a total of 14 samples before the soybean planting period and after the harvesting period to ensure that the selected data only includes useful information. Based on this, the study divides the remaining 32 samples from the planting period to the harvesting period into time intervals of $[t, t+6]$, where $t \in 7, 8, ..., 22$, generating 12 sets of data points. Each set of data consists of the first 5 sets of remote sensing images of 9 bands as inputs to the model, using the subsequent 1 set of images as the expected output. During the prediction task, the model can accomplish predictions at different time steps by moving the window, which involves re-entering the predicted results into the model for the next step of prediction. Since the time steps of input and output images are adjustable parameters in the model, explaining the model structure using the output of 1 set of images is not convenient. Therefore, the description of the model structure is based on inputting 20 sets of images and outputting the subsequent 12 sets of images, which is also a future direction of this study.

Table 1
Training results of the two models

Year	RMSE1 baseline	RMSE2 baseline	RMSE average baseline
2009	5.344475	5.696371	5.520423
2010	7.738574	6.264664	7.001619
2011	11.68791	7.464125	9.576017
2012	8.707651	7.067769	7.88771
2013	6.608574	7.77421	7.191392
2014	6.183367	6.533232	6.3583
2015	6.198505	6.704858	6.451681
average	7.495579	6.786461	**7.14102**

4. Methodology Formulation

Problem Definition: Let $I = \{x_1, x_2, ..., x_n\}$ be the set of agricultural remote sensing images for a given area, including time series images which may contain missing values for certain periods. Our goal is to predict soybean yield using remote sensing satellite images that are missing during the harvest period.

4.1. Case Retrieval and Case Reuse

You et al. [2] discretized the remote sensing satellite images into a fixed number of bins, creating a histogram for each band to achieve dimensionality reduction. Subsequently, two main deep learning models, Long Short-Term Memory (LSTM) and Convolutional Neural Network (CNN), were employed to process and analyze this histogram data. Finally, by integrating spatial and temporal information in crop yield prediction deep learning models using Deep Gaussian Process, it was determined that the model combining CNN with Deep Gaussian Process was optimal.

As this research aims to enhance accuracy in forecasting crop yields using remote sensing satellite images that are missing during the harvest period, we firstly use the CNN+GP model trained on data from the first 20 non-harvest sampling instances to predict crop yield. Table 1 shows the results.

CGANa-CBR model: Recognizing the crucial role of satellite images in predicting yield during the harvest period, this study employ the CGANa-CBR model for crop yield prediction. Specifically, the study leverages these images with missing values to enhance prediction accuracy. Figure 2 shows the structure of the CGANa-CBR model. The CGAN model consists primarily of two components: the generator and the discriminator. Sections 4.2.1 to 4.2.2 provide a detailed explanation of the model structure.

4.2. Case Revise

4.2.1. Generator

The generator (G) consists of three main components: a conditional representation module,a latent conditioning module and a sampler.

The conditional representation module(Figure 3) is a feed-forward convolutional neural network that generates the conditional representation based on the previous 20 observations of agricultural remote sensing images in 9 bands. Firstly, by downsampling each of the 9 bands of the image by a factor of 2, the four subsampled results are stacked along the band dimension, transforming each $64 * 64 * 9$ remote sensing observation into a $32 * 32 * 36$ input. Then, each remote sensing observation is processed individually to ensure uniform treatment of each image as they are essentially the same data. Next, four downsampling residual blocks are used, which lower the image resolution and double the number

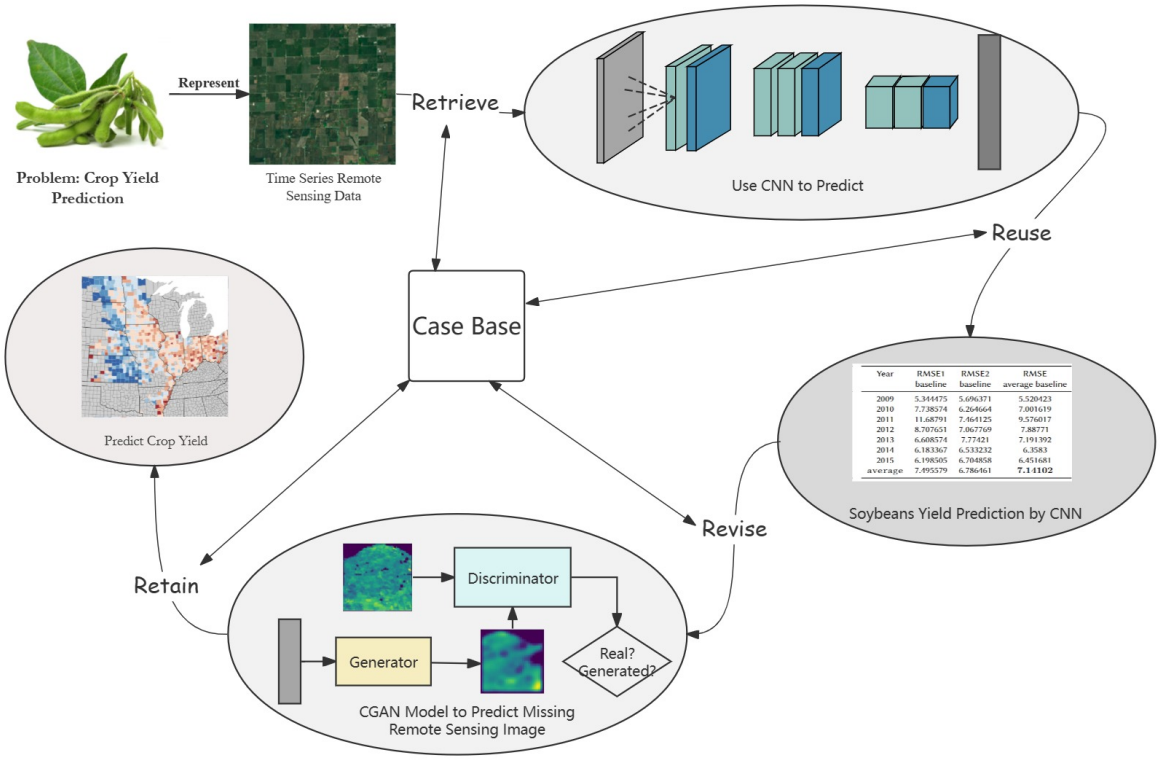

Figure 2: The technical roadmap of this article

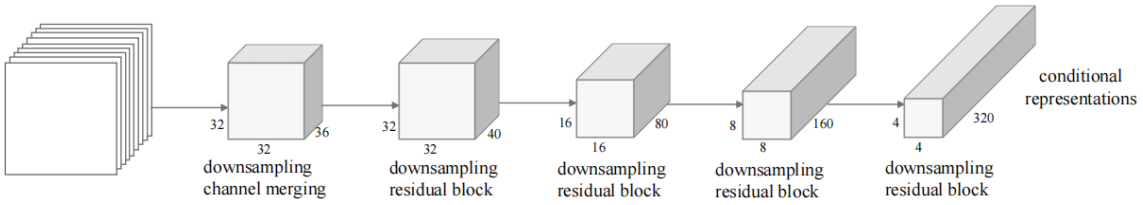

Figure 3: Processing of 9 bands inputs in conditional representation module.

of image channels. The outputs of each residual block are concatenated along the channel dimension, and for each output, a $3 * 3$ spectral normalization convolution is applied to reduce the number of channels by half, resulting in conditional representations of size $32 * 32 * 40, 16 * 16 * 80, 8 * 8 * 160$, and $4 * 4 * 320$ produced through rectified linear units.

For the latent condition module, the random input vector of size $9 * 4 * 4$is independently sampled from a normal distribution $N(0, 1)$. The last two dimensions of the random vector represent the height and width, which are $1/16$ of the height and width of the $64 * 64 * 9$ radar observation values. This random vector will then be passed through a $3 * 3$ convolutional layer, 3 modified residual blocks (L-Blocks), a spatial attention module, and another L-Block. The L-Block used here is a special type of residual block designed to increase the number of channels in the corresponding input.

The sampler(Figure 4) is responsible for generating 12 future predictions of remote sensing images from this conditional representation and latent conditions generated by random noise. The sampler consists of a stack of four ConvGRU units, which is a neural network architecture that combines convolutional neural networks (CNN) and gated recurrent units (GRU), integrating the functions of both convolutional and recurrent layers to effectively retain spatiotemporal information and perform efficient computations. In this architecture, the input at each time step undergoes convolutional operations with sliding convolution kernels and is fused with previous state information to generate new state

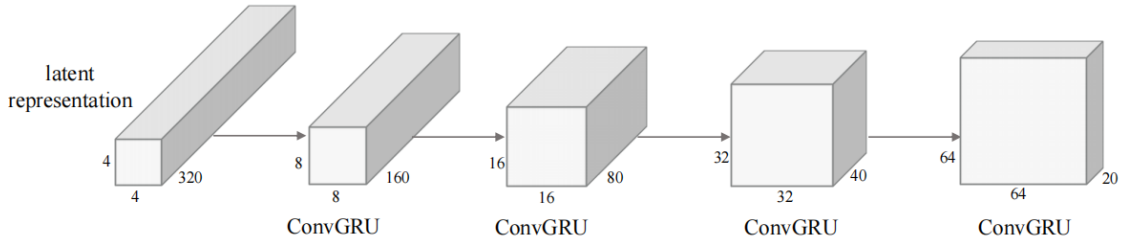

Figure 4: Processing in sampler module

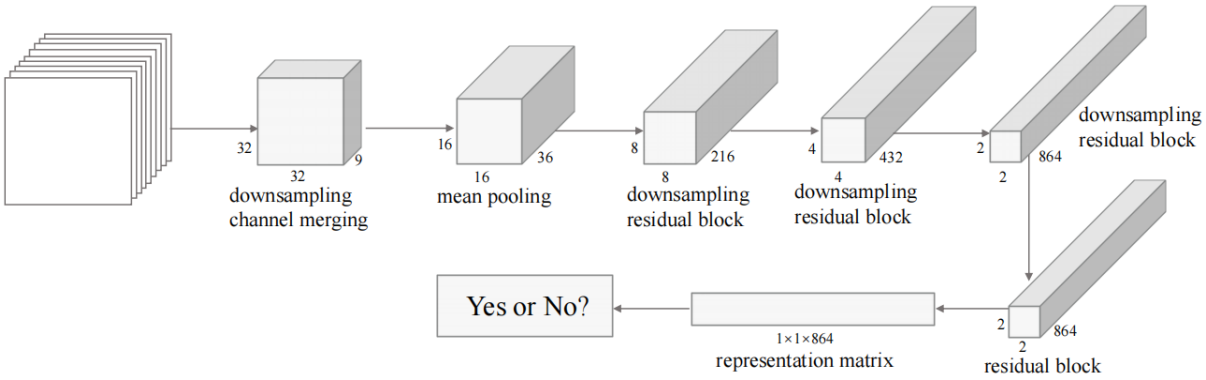

Figure 5: Processing in spatial discriminator

representations. The known conditional representation of the image serves as the initial state for each recurrent module. Starting from the initial state, 12 copies of a latent representation of size $4 * 4 * 320$ generated by the latent conditioning module (one for each forecast period) are fed as inputs to the lowest-resolution ConvGRU block.

The output of the latent condition module is repeated 12 times and used as the input to the first ConvGRU unit. Then, the output of each ConvGRU unit is upsampled as the input to the next ConvGRU. Within each ConvGRU unit, an independent spectral normalization convolutional layer and two residual blocks process all 12 latent representations, where the second residual block uses nearest-neighbor interpolation to double the input image resolution and halve the number of channels. After the final ConvGRU, the intermediate feature vector has a size of $64 * 64 * 20$.This feature vector is batch-normalized, applied with a ReLU activation function, and passed through a $1 * 1$spectrally normalized convolution to produce an output of size $32 * 32 * 36$.Similar to super-resolution, the 12 outputs can be converted to 12 predictions of size $64 * 64 * 9$by applying a depth-to-space upsampling operation on these 12 outputs.

4.2.2. Discriminator

The spatial and temporal discriminators used for training generative adversarial networks are similar to [25], as they simultaneously operate on the generated image sequences (generator part) and the combination of generated images and real images (discriminator part). The spatial discriminator(Figure 5) randomly selects 6 out of 20 consecutive images, first downsampling them to $32 * 32 * 9$ using a $2 * 2$ mean pooling, then transforming them into 16*16*36 inputs by stacking the downsampled images into the channel layer (from plane to depth). This is followed by three residual blocks, each reducing the resolution by half while doubling the number of channels. The outputs of the three modules are $8 * 8 * 216, 4 * 4 * 432, and 2 * 2 * 864$, respectively. After being processed by another residual block (maintaining spatial resolution and channel number), the result of merging along the height and width dimensions is the representation matrix of a single image. Finally, all 6 representation matrices are summed up, processed through spectral normalization linear layer, and applied ReLU for binary

classification output to distinguish between the image sequence of generated results and real data. The input of the temporal discriminator is a sequence composed of four remote sensing images that are either sequentially connected predicted images (generator part) or a combination of predictions and targets (discriminator part). Rectangles of height and width $32 * 32$ are randomly cropped from the sequence, and then each image in the sequence is transformed into a $16 * 16 * 36$ image through spatial-to-depth downsampling. This output is processed by two 3D modules, similar to the processing of the first two residual modules in the spatial discriminator, but with a $3 * 3 * 3$ spectral-normalized convolution process, where the first 3D block does not apply ReLU before the first $3 * 3 * 3$ convolution. The resulting representation matrix of length 3, $8 * 8 * 216$, is handled by two residual blocks with the same structure as the spatial discriminator. The remaining steps are the same as those after the last residual block of the spatial discriminator.

4.3. Case Retain

For a set of agricultural remote sensing images with a shape of $46 * 9 * width * height$ representing a full-year observation sequence, the images from each band in each observation are first converted into histograms, with the minimum boundary set to 1, the maximum boundary set to 5000, and the number of bins set to 32, resulting in a three-dimensional histogram with a shape of $46 * 9 * 32$. Subsequently, the time cropping method used for data processing in the aforementioned CGAN model is applied to this three-dimensional histogram. The first two dimensions of this three-dimensional matrix are then swapped to obtain a three-dimensional histogram with a shape of $9 * 32 * 32$ as the input data for the CNN model. For remote sensing image sequences with observation durations less than $6 + 32 = 38$ times and for remote sensing images where all pixels are 0, the corresponding histogram information will be padded with 0s.
In this study, a CNN architecture is utilized to establish a nonlinear mapping model for crop yield estimation, which consists of several convolutional modules and fully connected layers. The three-dimensional histogram with a shape of $9 * 32 * 32$ is input into the model and passes through six convolutional modules, each consisting of a convolutional layer, a batch normalization layer, and a dropout layer. Additionally, a ReLU activation function is applied to the data at the end of each convolutional module to enhance the model's nonlinear fitting capability. After passing through the six convolutional modules, the model generates an intermediate feature matrix of size $512 * 8 * 8$, which is then flattened into a vector of length 8192. Finally, this vector is fed into two fully connected layers to output high-precision estimates of average crop yield.

5. Results

The experiments in this study are divided into two parts. The first part involves evaluating the effectiveness of the CGANa-CBR model. In the second part, images generated by the generative model are used to train a CNN model, and the prediction results are evaluated.

5.1. Training and Performance Evaluation of the CGANa-CBR Model

In this study, the CGANa-CBR model was employed to generate remote sensing data for agricultural fields three months in advance, in order to optimize the prediction performance of the CNN model during the harvest season when data was missing. The study utilized a method where five consecutive remote sensing images were input and a single prediction output was made for model training and prediction. This approach allowed for a more flexible prediction time step by using a sliding window method and reduced GPU memory consumption during model training, thereby increasing the number of model parameters and achieving more accurate prediction results.

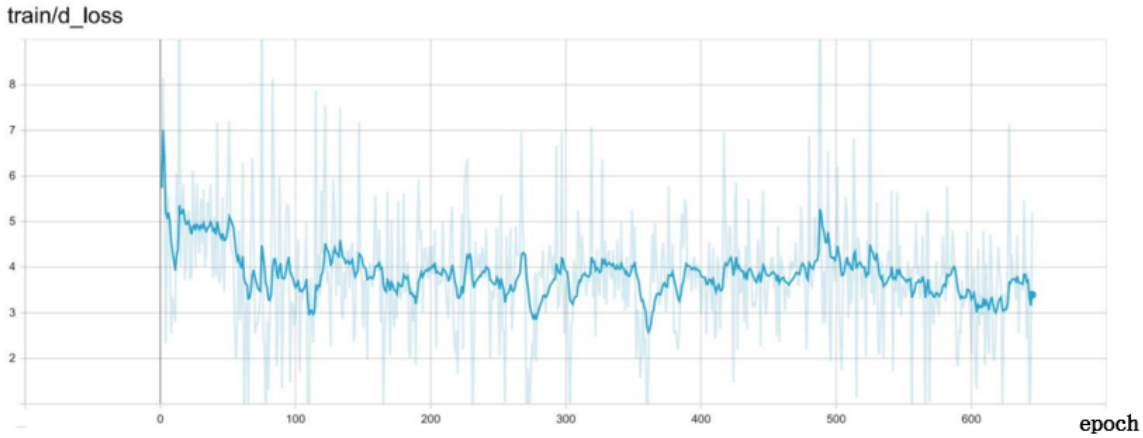

Figure 6: Discriminator Loss Curve

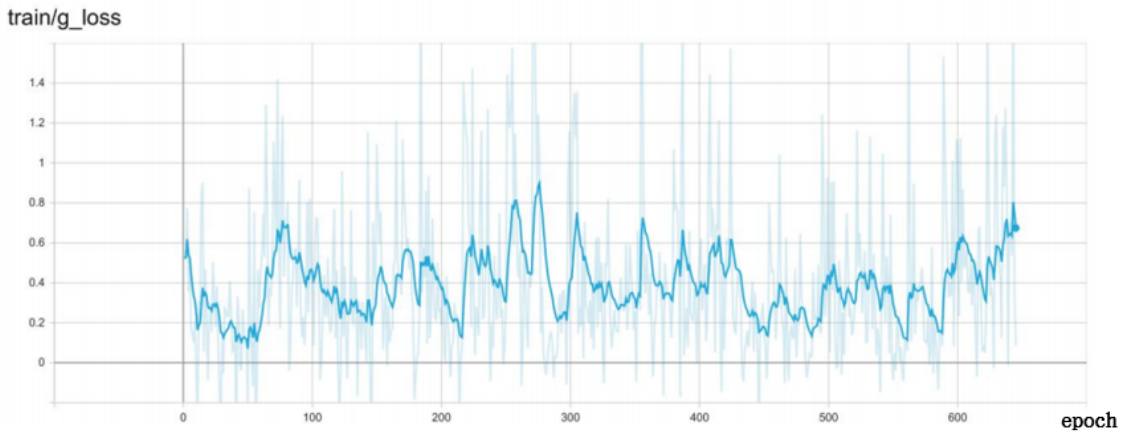

Figure 7: Generator Loss Curve

In the model training process, this study selected:

$$min_G max_D \sum_{i=1}^{n} D(\boldsymbol{x_i}, \boldsymbol{y_i}) - D(\boldsymbol{x_i}, \boldsymbol{z_i}|\boldsymbol{y_i}) + \lambda L_{reg}(\boldsymbol{G}). \tag{1}$$

as the optimization target of the model, where x_i and y_i represent real data pairs, z_i represents the noise vector, λ represents the weight of the regularization term, and $L_{reg}(\boldsymbol{G})$ represents the regularization term loss function of the generator. The optimization objective was to minimize the loss function of the generator, enabling it to generate high-quality fake data while adversarially challenging the discriminator's discriminative ability. The model had approximately 735 million parameters, and this study trained on a single A40 GPU with 0.25% of the dataset for 12 hours, totaling 5 epochs. We recorded the loss curve during the training process for evaluating the effectiveness of the Generative Adversarial Network. Figure 6 and Figure 7 are the loss curves during the model training process:
It can be observed that the discriminator's loss curve experiences a rapid decrease at the beginning of the model training, followed by fluctuations around a relatively low level, while the generator's loss curve consistently fluctuates at a low level. This indicates that the generative model and the discriminative model achieved good adversarial effects during the training process.
Figure 8 is the display of the prediction results and real images of future images using the model described above:

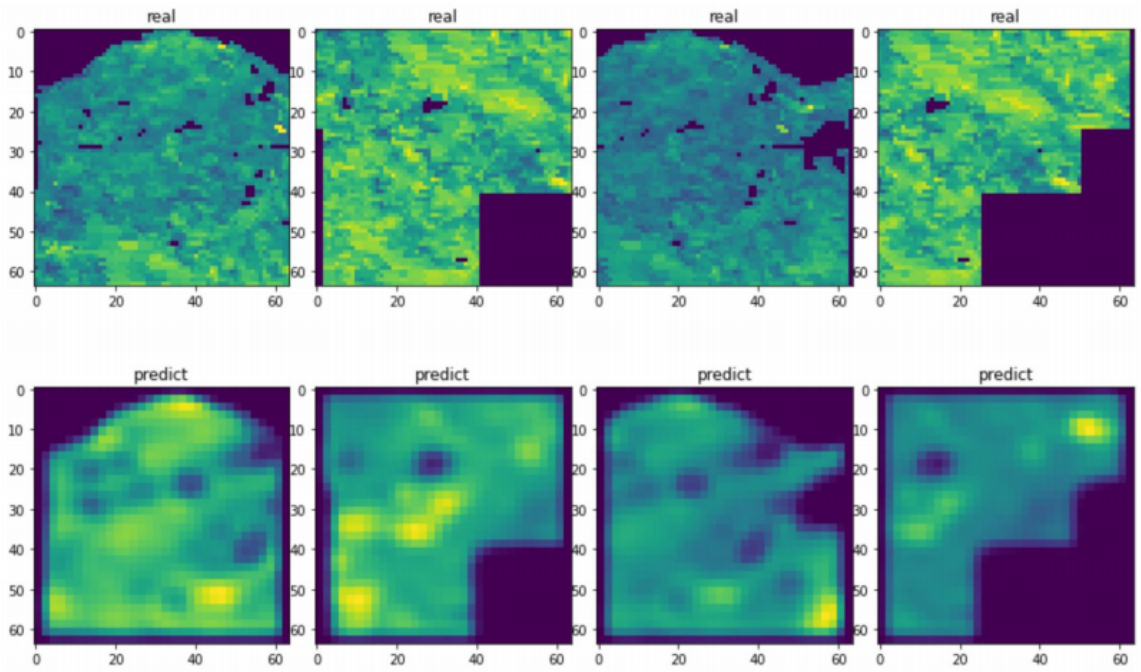

Figure 8: Real Image (Above) and Predicted Image (Below)

It can be noticed that the generator has learned the shape features of the images, maintaining the original terrain boundaries during prediction without generating unnecessary noise in the masked non-agricultural areas. Additionally, the generator's predictions of the distribution of spectral reflectance levels are reasonably accurate. For example, in the first set of comparison images on the left, the generator made good inferences about the high reflectance in the upper and lower right areas of the farmland. However, the predicted images exhibit a significant issue of being overly blurry, which may be attributed to the excessively large model parameters (735M) and the short training time (12 hours, 0.25% dataset, 2 epochs).

Furthermore, as the CNN model for yield estimation is trained on input data in the form of three-dimensional histograms derived from remote sensing images, the images generated by the generator need to be further transformed into histograms for application in CNN model training. Figure 9 is a display of histograms derived from integrated annual remote sensing images consisting of real and generated data.

The vertical axis in figure 9 represents the sequential sampling time of remote sensing images, while the horizontal axis indicates the boundaries of the histograms. In this experiment, the first 20 remote sensing images were used as input data for the generative model, hence the histogram data for the first 20 sampling times in the figure are identical. The figure illustrates the differences between the histograms of real and predicted images. It can be observed that the issue of blurriness in the images generated by the model is somewhat controlled after conversion into histograms, as similar pixel values will fall into the same interval.

Furthermore, it can be seen that the results generated by the model are relatively stable in the first 3-5 prediction steps, but the model seems to lose control over the image data in subsequent prediction steps. This phenomenon arises from the prediction method used in this experiment, where five inputs are used to predict one output. By the sixth step, the model's input data will entirely come from the previous model outputs, which may not sufficiently reflect the past image features. As a result, the model's predicted noise accumulates, leading to a decrease in prediction accuracy over time steps.

Moreover, this could also be attributed to the fact that past information cannot entirely predict the future. Past images can only provide some prediction for the near future when the time gap is too long, the prediction accuracy tends to decrease.

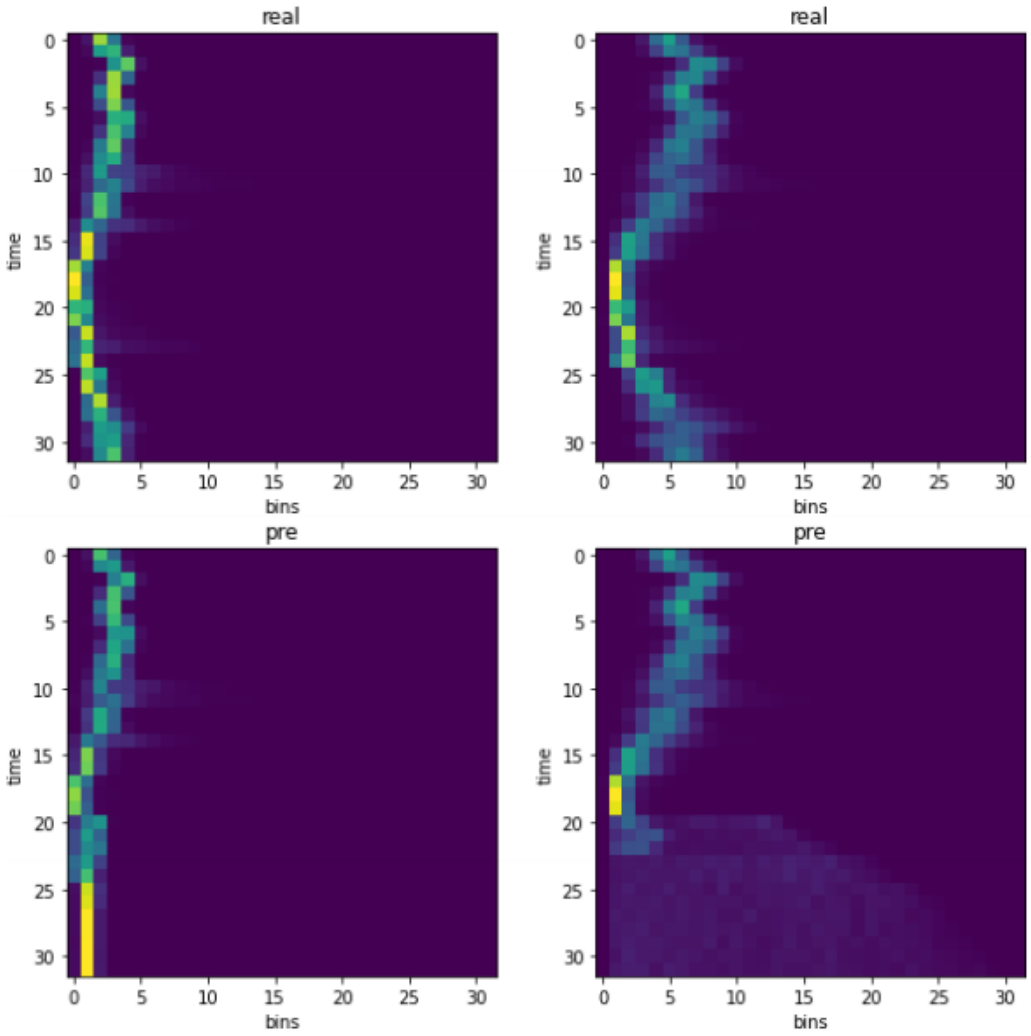

Figure 9: Real Image (Above) and Predicted Image (Below)

5.2. Evaluation of the Average Yield Prediction Model

We utilized agricultural remote sensing data from the first 20 sampling instances and data generated from the next 12 sampling instances using Conditional Generative Adversarial Networks (CGAN) to construct a CNN+GP model, which was then utilized for predicting average crop yield. We conducted two training and evaluation cycles using data from the years 2009 to 2015. The results of this model were compared with those obtained by the CNN+GP model trained using only the first 20 non-harvest sampling instances. Table 2 are the results of the models.

From the table, it is evident that in both independent training iterations, the CNN+GP models utilizing CGAN-generated data exhibited superior predictive performance compared to the baseline CNN model across most datasets. Specifically, the CNN models trained on CGAN-generated data showed an average reduction of 6.3% in RMSE values, indicating that fine-tuning CNN models with generated remote sensing imagery can effectively enhance the accuracy of average yield predictions.

6. Conclusion

This study employed a method of utilizing CGANa-CBR-generated remote sensing data and CNN models for predicting average crop yield. Experimental results indicate that the CNN models trained on CGANa-CBR-generated data exhibited superior predictive performance across most datasets, with

Table 2
Training results of the two models

Year	RMSE1 baseline	RMSE2 baseline	RMSE average baseline	RMSE1 CGAN+CNN	RMSE2 CGAN+CNN	RMSE average CGAN+CNN
2009	5.344475	5.696371	5.520423	5.786561	5.43374	5.61015
2010	7.738574	6.264664	7.001619	6.471627	5.766839	6.119233
2011	11.68791	7.464125	9.576017	9.109857	7.060263	8.08506
2012	8.707651	7.067769	7.88771	7.015368	7.042023	7.028696
2013	6.608574	7.77421	7.191392	6.852591	6.82692	6.839756
2014	6.183367	6.533232	6.3583	6.278881	5.976492	6.127687
2015	6.198505	6.704858	6.451681	7.649933	6.465012	7.057473
average	7.495579	6.786461	**7.14102**	7.023546	6.367327	**6.695436**

average reductions of 10% and 12% in RMSE and ME values, respectively. This suggests that fine-tuning CNN models with generated remote sensing imagery can enhance the accuracy of average yield predictions.

The method proposed in this study can provide valuable insights for agricultural production, aiming to enhance agricultural productivity and food security. Additionally, the CGANa-CBR approach introduced in this study offers a new perspective for enhancing and applying remote sensing data. Future research could further explore the application of GAN-based methods for enhancing and predicting other agricultural data.

However, this study has some limitations that need to be addressed. Firstly, the dataset used in this study only consists of data from a single crop, neglecting the impact of different crops and regions. Therefore, targeted adjustments may be necessary for practical applications. Additionally, due to limited computational resources, this research only utilized a small-scale remote sensing dataset of farmland, which may impose certain restrictions on the model's generalization ability. Moreover, it is important to acknowledge that insufficient training time for the model may affect its performance, and further data validation is required in real-world applications to confirm the accuracy and practicality of the model. In the future, advancements can be made by further exploring additional network architectures and training strategies to enhance the predictive performance and generalization ability of the model. Additionally, experiments should be conducted on larger-scale datasets to validate the findings.

Acknowledgments

This work was jointly supported by National Natural Science Foundation of China (NSFC) under grant 62206301; Public Health & Disease Control and Prevention, Fund for Building World-Class Universities (Disciplines) of Renmin University of China. Project No. 2024PDPC; the Major Project of the MOE (China) National Key Research Bases for Humanities and Social Sciences (22JJD910003); and the Wine Group research grant Project No. 09202188. This research was supported by Public Computing Cloud, Renmin University of China. We sincerely thank Mr. Jingzhou Xu of Renmin University of China for providing data processing and experiment support.

References

[1] O. Satir, S. Berberoglu, Crop yield prediction under soil salinity using satellite derived vegetation indices, Field crops research 192 (2016) 134–143.

[2] J. You, X. Li, M. Low, D. Lobell, S. Ermon, Deep gaussian process for crop yield prediction based on remote sensing data, in: Proceedings of the AAAI conference on artificial intelligence, volume 31, 2017.

[3] L. Baghdasaryan, R. Melikbekyan, A. Dolmajain, J. Hobbs, Deep density estimation based on

multi-spectral remote sensing data for in-field crop yield forecasting, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2022, pp. 2014–2023.

[4] S. Ravuri, K. Lenc, M. Willson, D. Kangin, R. Lam, P. Mirowski, M. Fitzsimons, M. Athanassiadou, S. Kashem, S. Madge, et al., Skilful precipitation nowcasting using deep generative models of radar, Nature 597 (2021) 672–677. URL: http://creativecommons.org/licenses/by/4.0/, available under a CC BY 4.0 License. Code adapted with changes.

[5] X. Xu, P. Gao, X. Zhu, W. Guo, J. Ding, C. Li, M. Zhu, X. Wu, Design of an integrated climatic assessment indicator (icai) for wheat production: A case study in jiangsu province, china, Ecological indicators 101 (2019) 943–953.

[6] T. Van Klompenburg, A. Kassahun, C. Catal, Crop yield prediction using machine learning: A systematic literature review, Computers and Electronics in Agriculture 177 (2020) 105709.

[7] B. Basso, D. Cammarano, E. Carfagna, et al., Review of crop yield forecasting methods and early warning systems, in: Proceedings of the first meeting of the scientific advisory committee of the global strategy to improve agricultural and rural statistics, FAO Headquarters, Rome, Italy, volume 241, 2013.

[8] Y. Chen, R. J. Donohue, T. R. McVicar, F. Waldner, G. Mata, N. Ota, A. Houshmandfar, K. Dayal, R. A. Lawes, Nationwide crop yield estimation based on photosynthesis and meteorological stress indices, Agricultural and Forest Meteorology 284 (2020) 107872.

[9] T. Van Klompenburg, A. Kassahun, C. Catal, Crop yield prediction using machine learning: A systematic literature review, Computers and Electronics in Agriculture 177 (2020) 105709.

[10] S. Khaki, L. Wang, Crop yield prediction using deep neural networks, Frontiers in plant science 10 (2019) 452963.

[11] A. Barbosa, R. Trevisan, N. Hovakimyan, N. F. Martin, Modeling yield response to crop management using convolutional neural networks, Computers and Electronics in Agriculture 170 (2020) 105197.

[12] E. Kamir, F. Waldner, Z. Hochman, Estimating wheat yields in australia using climate records, satellite image time series and machine learning methods, ISPRS Journal of Photogrammetry and Remote Sensing 160 (2020) 124–135.

[13] B. Nath, D. Dhakre, D. Bhattacharya, Forecasting wheat production in india: An arima modelling approach, Journal of Pharmacognosy and Phytochemistry 8 (2019) 2158–2165.

[14] H. Fernando, T. Ha, K. A. Nketia, A. Attanayake, S. Shirtliffe, Machine learning approach for satellite-based subfield canola yield prediction using floral phenology metrics and soil parameters, Precision Agriculture (2024) 1–18.

[15] P. Pandya, N. K. Gontia, Early crop yield prediction for agricultural drought monitoring using drought indices, remote sensing, and machine learning techniques, Journal of Water and Climate Change 14 (2023) 4729–4746.

[16] R. Wu, S. Yan, Y. Shan, Q. Dang, G. Sun, Deep image: Scaling up image recognition, arXiv preprint arXiv:1501.02876 7 (2015) 4.

[17] R. Miotto, L. Li, J. T. Dudley, Deep learning to predict patient future diseases from the electronic health records, in: Advances in Information Retrieval: 38th European Conference on IR Research, ECIR 2016, Padua, Italy, March 20–23, 2016. Proceedings 38, Springer, 2016, pp. 768–774.

[18] Y. Cai, K. Guan, J. Peng, S. Wang, C. Seifert, B. Wardlow, 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 210 (2018) 35–47.

[19] P. Nevavuori, N. Narra, T. Lipping, Crop yield prediction with deep convolutional neural networks, Computers and electronics in agriculture 163 (2019) 104859.

[20] J. Sun, L. Di, Z. Sun, Y. Shen, Z. Lai, County-level soybean yield prediction using deep cnn-lstm model, Sensors 19 (2019) 4363.

[21] S. Kulkarni, S. N. Mandal, G. S. Sharma, M. R. Mundada, et al., Predictive analysis to improve crop yield using a neural network model, in: 2018 international conference on advances in computing, communications and informatics (ICACCI), IEEE, 2018, pp. 74–79.

[22] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, Y. Bengio, Generative adversarial networks, Communications of the ACM 63 (2020) 139–144.

[23] F. Yang, D. Zhang, Y. Zhang, Y. Zhang, Y. Han, Q. Zhang, Q. Zhang, C. Zhang, Z. Liu, K. Wang, Prediction of corn variety yield with attribute-missing data via graph neural network, Computers and Electronics in Agriculture 211 (2023) 108046.

[24] A. A. Gomaa, Y. M. Abd El-Latif, Early prediction of plant diseases using cnn and gans, International Journal of Advanced Computer Science and Applications 12 (2021).

[25] P. Luc, A. Clark, S. Dieleman, D. de Las Casas, Y. Doron, A. Cassirer, K. Simonyan, Transformation-based adversarial video prediction on large-scale data, ArXiv abs/2003.04035 (2020).