

# Exploiting Hierarchical Features for Crop Yield Prediction based on 3D Convolutional Neural Networks and Multi-kernel Gaussian Process

Mengjia Qiao, Xiaohui He, Xijie Cheng, Panle Li, Haotian Luo, Zhihui Tian, and Hengliang Guo

**Abstract**—Accurate and timely prediction of crop yield based on remote sensing data is important for food security. However, crop growth is a complex process, which makes it quite difficult to achieve better performance. To address this problem, a novel 3D convolutional neural multi-kernel network (3DMKGP) is proposed to capture hierarchical features for predicting crop yield. First, a full 3D convolutional neural network (3D CNN) is constructed to maximally explore deep spatial-spectral features from multispectral images. Then, a multi-kernel learning (MKL) approach is proposed for fusion of intra-image deep spatial-spectral features and inter-sample spatial consistency features. Specifically, we assign a group of nonlinear kernels for each feature in the MKL framework, which provides a robust way to fit features extracted from different domains. Finally, the probability distribution of prediction results is obtained by a kernel based-method. We evaluate the performance of the proposed method on China wheat yield prediction and offer detailed and systematic analyses of the performance of the proposed method. In addition, our method is compared with several competing methods. Experimental results demonstrate that the proposed method has certain advantages and can provide better prediction performance than the competitive methods.

**Index Terms**—crop yield, 3D convolutional neural network, spatial consistency, multi-kernel learning.

## I. INTRODUCTION

CROP yield prediction is an essential and active part of improving food security. A reliable and timely estimation of crop yield before harvest is of great significance and has been a topic of interest for decades. Traditionally, researchers focused on predicting crop yields through crop simulation models (CSMs) [1], [2]. Two main dilemmas arise in adopting CSMs in crop yield prediction. First, a substantial number of indicators are needed for model calibration that restrict these models from broad applications. Second, CSMs are commonly designed for specific regions based on the situations of current interest, which makes it difficult to accurately estimate yields in other regions. To overcome these drawbacks, later studies cast yield prediction problems from the perspective of machine learning methods, such as support vector machines (SVM) [3], [4], random forests (RF) [5], [6], artificial neural networks

Manuscript submitted Sep 25, 2020.

This work was supported by the National Key R&D Program of China under Grant 2018YFB0505000 (Corresponding author: Xiaohui He).

Xiaohui He, Haotian Luo, Zhihui Tian and Hengliang Guo is with the School of Geoscience and Technology, Zhengzhou University, 450001, China (e-mail: hexh@zzu.edu.cn).

Mengjia Qiao, Xijie Cheng, and Panle Li are with School of Information Engineering, Zhengzhou University, Zhengzhou 450001, China.

(ANN) [7]–[9], etc. These algorithms can build empirical predictive models using a fixed pattern of features extracted from remote sensing images (RSIs) without abundant parameters. However, the performance is still limited because hand-crafted features mainly leverage human ingenuity and prior knowledge [10], which are not robust for all cases. Furthermore, such algorithms are computationally intensive and become intractable on large-scale datasets. All such difficulties make crop yield prediction challenging and lead to unsatisfactory performance.

With the development of deep learning methods, an important trend in crop yield prediction is the use of deep models. Compared with traditional approaches, these models can automatically extract informative features from the raw image, which enable generalization for coping with different situations [11], [12]. Ma et al. [13] introduced a stacked autoencoder (SAE) to extract the deep spectral features for crop yield prediction using multispectral images. However, this method flattens images into one-dimensional spectral vectors that suffer from spatial information loss. To address this issue, [14]–[17] proposed the use of convolutional neural networks (CNNs) for crop yield prediction. Compared with SAEs, CNNs allow using spatial patches as input that provide a natural way to incorporate spatial information and enhance the performance [18]. Moreover, Qi et al. [19] developed a two-branch CNN to separately extract spectral and spatial features. These methods have achieved better results than traditional approaches and have proven that the deep spatial and spectral features excavated by CNNs can create a generalized representation for crop yield prediction.

However, in the real world, crop growth is a highly complex trait determined by many factors such as soil properties, precipitation, etc., which are not fully captured by RSI. In such cases, merely excavating intra-image deep spatial and spectral features based on CNNs may be insufficient. It has been observed that these crop-related factors have a strong consistency trend in the spatial dimension [20], [21]. In this scenario, the use of this spatial consistency between different data points could be of help to incorporate properties of other missing factors into the task of prediction. As demonstrated by Anselin [22], spatial correlation was prevalent in corn yield response models and should be critically considered in the analysis of yield monitor data. Peralta et al [23] pointed that without accounting for spatial consistency, prediction models may lead to inaccurate estimations. Following these rationales, the spatial consistency property has been intensively incorporated in rice [24], grain [20] and soybean [21] prediction

and achieved improved performance. Hence, we assume that the integrated use of deep spatial-spectral features and spatial consistency can help as an ensemble of multiple crop-related characters for predicting crop yield, which would effectively improve the results. Nevertheless, since these two features are typically obtained from different sources, determining how to simultaneously integrate these two heterogeneous levels of information into the prediction model is a substantial challenge. To overcome this problem, we propose a multi-kernel learning approach associated with a kernel-based Gaussian process (GP) for integrating heterogeneous features as well as a feature selection strategy for inferring the contribution of each feature.

In this paper, we aim at fully exploiting and utilizing the discriminative power of hierarchical features for crop yield prediction. Based on this intention, we propose a novel architecture (3DMKGP) for crop yield prediction, which is a combination of 3D CNN and Multi-Kernel Gaussian Process. Although 3D CNN has been used for crop yield prediction in earlier works [25], [26], to the best of our knowledge, we are first to apply it for excavating the joint spatial-spectral features from MSIs. Note that gaussian process has been explored in [14], where a linear GP is utilized for final prediction and the spatial consistency is only taken as the residual of the prediction. In our work, a novel multi-kernel learning strategy is not only employed for seamlessly generalizing deep features and spatial consistency within a kernel function, but is also used for measuring the uncertainty of the predictions. The main contributions of this paper are listed as follows:

1 A 3D CNN is first applied for excavating spatial-spectral features in the crop yield prediction assessment. By applying 3D convolutions, the robust spatial-spectral features can be simultaneously extracted.

2 A multi-kernel GP with a new 'spatial-spectral-spatio' composite Gaussian kernel is concatenated on the top of the 3D CNN. The new kernel is derived from deep features and location data and can flexibly encode deep feature characteristics and spatial consistency.

3 County-level wheat yield in China is predicted to show the effectiveness of the proposed method. Experimental results show that our method achieves a promising modeling capability and is capable of predicting crop yield with higher quality.

The rest of the paper is organized as follows. Section II presents an introduction of crop yield prediction and multi-kernel learning. Section III describes the proposed 3DMKGP in detail. Section IV then provides the dataset information and experimental results. Finally, Section V presents the conclusions.

## II. RELATED WORKS

### A. Traditional-based Methods for Crop Yield Prediction

Crop yield prediction has been a vital problem for decades, and many methods have been proposed to forecast crop yields. The most commonly used approach for prediction yield is crop yield models. Generally, we can classify all previous crop yield models into two categories. The models in the first category forecast crop yields by simulating multiple

physiological characteristics of crops, such as meteorological data, soil properties, biogeochemical fluxes, relevant socio-economic indicators, etc. The most widely used models including AquaCrop [27], the CERES-Maize model [28] and the APSIM model [29]. However, these kinds of models require a massive number indicators as input, which are hard to obtain, especially in developing counties. Furthermore, these models exhibit weak spatial generalization, which limits these models from wider application. The second category is the machine learning-based models. These kinds of models are based on hand-crafted features, such as the normalized difference vegetation index (NDVI) and the enhanced vegetation index (EVI). A considerable body of work in the crop yield prediction community falls into this category. For example, in [30], NDVI and EVI from MODIS were used to establish an empirical approach to forecast maize yield in the United States. Moreover, [31] also demonstrated good correlation between sunflower yield and the calculated NDVI from pixel-based multispectral images based on an SVM. Similar studies evaluated different methods for various crop types, e.g., regression trees for wheat [32] and corn [33], multilinear regression (MLR) for barely [34] and rice prediction [35], and so on. After that, in order to incorporate more complex modeling problems, Bose [9] later introduced spiking neural networks for crop yield estimation from NDVI image time series, which have been shown to outperform traditional machine learning classifiers. However, these traditional machine learning-based methods are still limited by hand-crafted features, since they are mostly mathematical combinations of a few fixed bands and are not robust for all cases [36].

### B. Deep Learning-based Methods for Crop Yield Prediction

Recently, deep learning methods have shown great potential in the field of urban planning [37], [38], land cover classification [39], [40], and gradually developed in crop yield prediction. Compared to traditional approaches, the deep learning methods can automatically extract robust features from the original image using a hierarchical representation architecture. Peerlinck et al. [41] introduced a stacked autoencoder (SAE) to extract high-level features for crop yield prediction based on multispectral images. In this work, the images were flattened into a one-dimensional spectral vector that could not exploit the spatial information. To overcome this limitation, You [14] first introduced a convolutional neural network (CNN) and a long short-term memory (LSTM) network to automatically discover crop-related features from raw images for crop yield prediction in America, and the results showed that both the CNN and LSTM achieve better results than the traditional approach. Compared to the SAE, the CNN can preserve local spatial information from spatial patches through two-dimensional convolution filters. Based on this intention, many researches have applied CNNs to the task of excavating spatial and spectral features in crop yield prediction [15], [42]. Moreover, Wang and Kaneko [16], [17] provided an extension to new regions with a transfer learning approach, which further exhibited the superiority of deep learning methods in extracting spatial and spectral features. More recently, Qi et

al. [19] presented a two-branch CNN architecture for rich grain yield estimation so that spectral and spatial features can be separately extracted. The above studies provided evidence regarding the advantages of deep spatial and spectral features over hand-crafted features. However, these methods do not take full advantage of spatial-spectral features due to the limitation of 2D networks.

### C. 3D Convolutional Neural Networks

The 3D CNN was first proposed for extracting features from spatial and temporal dimensions in human action recognition [43]. Then, the development of the 3D CNN provided a more effective way to excavate spatial-spectral features. More recently, the 3D CNN was developed for hyperspectral image (HSI) classification [44], [45]. Instead of extracting spatial and spectral features separately in 2D networks, the 3D CNN can capture the joint spatial-spectral features, which is more suitable for three-dimensional cube data, and has achieved excellent results. For instance, Chen et al. [46] proposed a 3D CNN-based feature extraction model to extract effective spatial-spectral features of HSI. In this way, the 3D hypercube with joint spatial-spectral information can be obtained simultaneously. Yang et al. [47] proposed a recurrent 3D CNN method, where the 3D convolution operator is utilized to exploit both spatial context and spectral correlation through gradually shrinking the patch, and greatly improved performance in HSI classification was achieved. To increase the classification accuracy of HSIs, Sun et al. [48] designed a spectral-spatial attention network through spectral and spatial models with 3D convolution to extract the joint spectral-spatial features. Then, the spectral-spatial attention model was embedded between

the above two models to suppress the effects of interfering pixels and capture attention areas in an HSI cube. In [49], a multiscale recurrent neural network (RNN) is presented with spectral-spatial features, where the 3D CNN is used to extract the local spectral-spatial features and the RNN is used to capture the spatial dependency. The results have proven that the model can capture the deep spectral-spatial features simultaneously, and it outperforms the SAE and 2D CNN models. The above studies have proven that the 3D CNN has great superiority in exploiting spatial-spectral features. However, the datasets used in crop yield prediction are different from HSI, and the process of crop growth is also typical. Therefore, determining how to construct a proper 3D CNN for predicting yield is a vital problem.

### D. Multi-kernel Learning

Kernel-based methods, such as GPs or SVMs can model nonlinear data distributions by an implicit mapping function input to the kernel space through kernel embedding [50], [51]. However, in most general case, a single kernel is not able to fully model the highly complex nonlinear relationship. Hence, a more flexible multi-kernel learning approach is proposed to address multimodality features by combining multiple kernels. In the multi-kernel framework, each kernel stands for a certain kind of feature. The fusion of features through multi-kernels can and has achieved state-of-the-art performance in many fields. For example, Cao et.al [52] developed a multi-kernel-based feature fusion and selection method with a kernel logistic regression model for modeling multiple visual characteristics. To incorporate high-dimensional feature space, [53]

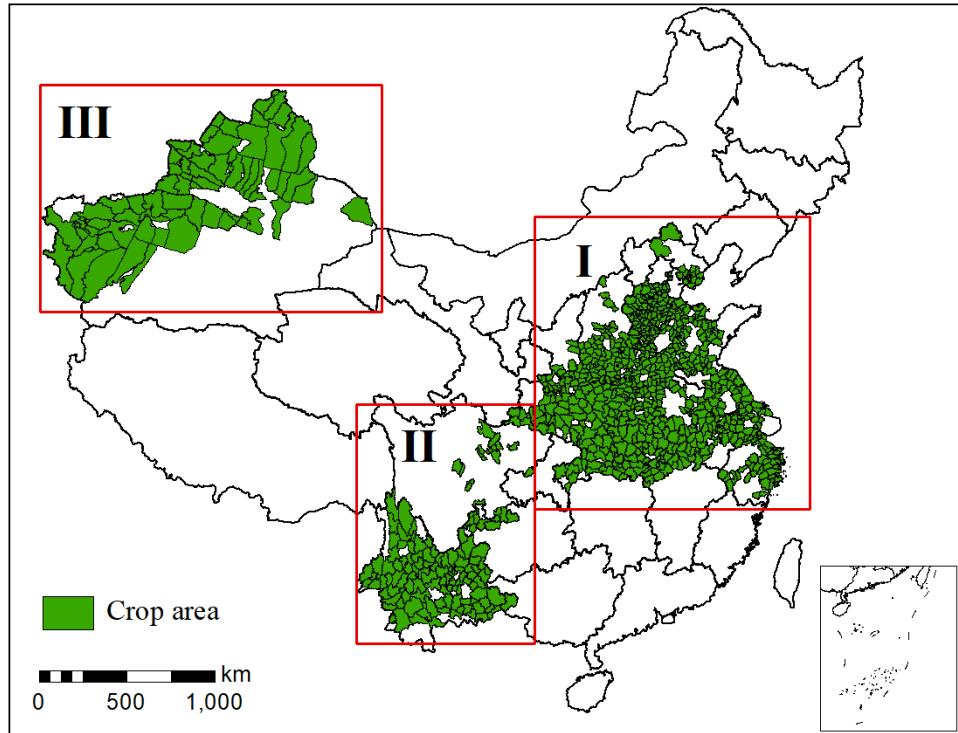


Fig. 1. Winter wheat study area in China.

proposed a kernel alignment-based model based on the automatic optimization of the linear combination of kernels, where the weights of each kernel are automatically merged after training the model. Yeh [54] presented the MKL framework for fusion of features from different domains. Their method is able to select the class-specific weights for different types of features via the one-vs-rest learning strategy. They then verify the feasibility of their method in feature fusion under MKL. More recently, studies have developed MKL for combining multisource features obtained from different sensors. For example, a new MKL model was proposed to integrate heterogeneous features from LiDAR and MSI achieved great performance in urban area classification [55]. Later, MKL was also applied for fusion of features from different modalities and yielded improved performance in damage detection. The results revealed the integration of multiple features by MKL led to an additional improvement of approximately 3% [56].

In this work, we aim to explore and integrate hierarchical features for crop yield prediction. Therefore, we first apply a 3D CNN for excavating deep spatial-spectral features from raw images within a single region. We then incorporate a multi-kernel GP to simultaneously capture multiple features through a composite kernel. The experimental results indicate that our method is superior to competing methods with a significant improvement in the wheat yield prediction in China.

### III. METHODS

#### A. Problem Setting

In this section, we start by formalizing the crop yield prediction problem. The purpose of crop yield prediction is to predict crop yield before harvest. More specifically, given a set of remote sensing images ( $I^0, I^1, \dots, I^t$ ) from time 0 to  $t$ , the model is required to make a reliable prediction yield at time  $t + 1$ . For this paper, we are interested in the average crop yield within several counties. Let  $X \in \mathcal{R}^{B \times HW}$  denote the

$H \times W$  multispectral image with  $B$  bands within each county as the training input, and  $Y$  is the crop yield label; Given any new image  $X^*$  from the testing set, we aim to predict the probability distribution  $P(Y^*|X^*, Y, X)$  of the corresponding crop yield  $Y^*$ .

#### B. Study Area

In this paper, we focus on validation our method on winter wheat prediction in China. The main planting areas at county level of wheat are presented in Fig.1. We further divided the planting areas into three typical types, including the Northern winter wheat part (Zone I), the Southwest winter wheat part (Zone II) and the Xinjiang winter wheat part (Zone III).

#### C. An Overview of the Proposed Approach

Crop yield prediction has been a vital problem and many studies have been proposed for seeking better performance. In this study, we focus on exploring and integrating hierarchical features for crop yield prediction.

First, a 3D CNN is introduced for learning spatial-spectral features from remote sensing image. In the 3D CNN, several 3D convolution kernels are applied for deep feature extraction from multispectral images. Second, the prediction process is carried out using a multi-kernel GP based on the deep spatial-spectral features and spatial location data. Since the multi-kernel GP is capable of systematically aggregating different representations, it can simultaneously capture the deep feature similarities as well as the spatial consistency.

Then, we design a new 3DMKGP framework by integrating the 3D CNN and multi-kernel GP, as shown in Fig.2. Two parts are included in our 3DMKGP: the first part is the feature space, including deep spatial-spectral feature extraction from the 3D CNN and location data (Loc data) capture. In addition, the multi-kernel GP is connected on the top of the 3D CNN as the kernel space. Finally, the probability distributions of prediction

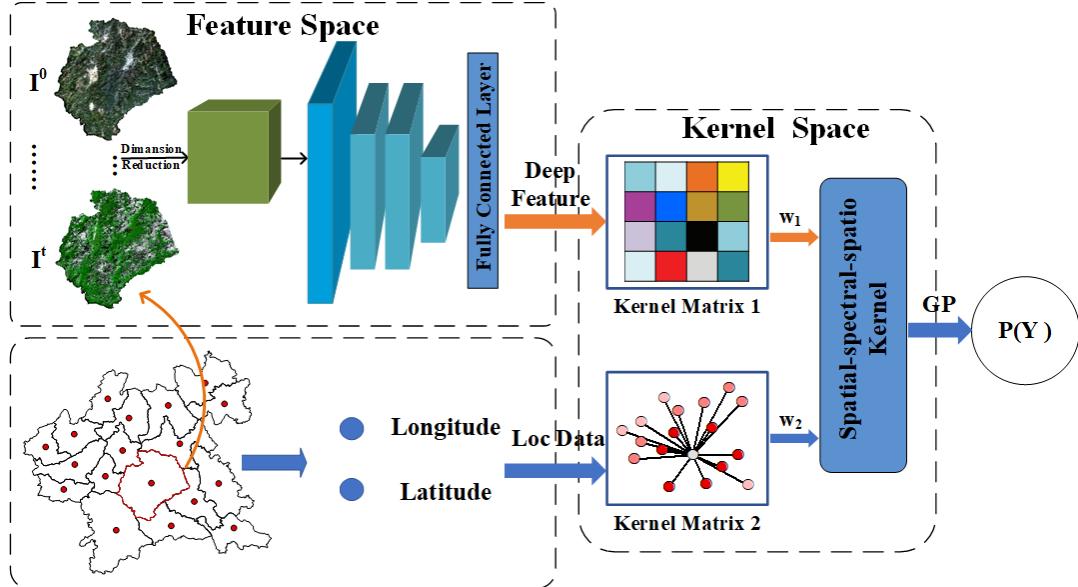


Fig. 2. Architecture of proposed 3DMKGP.

results are established after the kernel space. The proposed 3DMKGP can fully excavate spatial-spectral features from remote sensing images and make reliable prediction jointly from deep features and location data.

#### D. 3D Convolutional Neural Networks

The remote sensing images used for crop yield prediction in this study are represented by a 3D cube, which contains a two-dimensional spatial context and one-dimensional spectral information. The traditional 2D CNN uses the 2D convolution kernels during the process of feature learning and moves in two directions ( $x, y$ ) to calculate low-dimensional features from the image data. The output shape is also a two-dimensional matrix, which neglects the spectral information along the third dimension. The 3D CNN applies three-dimensional convolution kernels to the dataset, and the convolution kernels move in three directions ( $x, y, z$ ) to calculate the spatial and spectral feature representations. Most importantly, the output shape after 3D convolution is a three-dimensional volume space such as a cube or cuboid, which can jointly learn the spatial and spectral features via different kernels. The difference between 2D CNNs and 3D CNNs can be seen from Fig.3 . Hence, in order to fully explore both spatial and spectral discrimination simultaneously, the 3D CNN is adopted in the proposed method, including 3D convolution, 3D pooling, and 3D batch normalization.

1) *3D Convolution*: As shown in Fig.3, for an input  $I \in \mathbb{R}^3$ , the value at position  $(x, y, z)$  on the  $j$ th feature map in the  $i$ th convolution layer is represented as:

$$v_{ij}^{xyz} = f(b_{ij} + \sum_m \sum_{p=0}^{P_i-1} \sum_{q=0}^{Q_i-1} \sum_{r=0}^{R_i-1} w_{ijm}^{pqr} v_{(i-1)m}^{(x+p)(y+q)(z+r)}) \quad (1)$$

where  $f$  is an activation function and  $m$  stands for the number of feature maps in the  $(i-1)$ th layer (previous layer), and  $w_{ijm}^{pqr}$

is the weight at the position  $(p, q, r)$  in the 3D convolution layer that connects to the feature maps of adjacent layers.  $P_i, Q_i, R_i$  are the height, width and depth of the kernel, respectively, and  $b_{ij}$  denotes the bias.

2) *3D Pooling*: Max pooling is used to reduce the number of training parameters of the 3D CNN in this paper. Different from 2D pooling, 3D max pooling can perform downsampling by dividing three-dimensional input into cuboidal pooling regions and computing the maximum of each region. The feature after pooling at position  $(x, y, z)$  is defined as:

$$O^{x,y,z} = \max_t F_t^{x,y,z} \quad (2)$$

where  $F$  stands for the features extracted from 3D convolution kernels before the pooling layer.

3) *3D Batch Normalization*: Assume that  $X$  is a mini-batch input; the output of a normalized 3D batch can be represented as :

$$y = \frac{\gamma x}{\sqrt{Var(x) + \epsilon}} + (\beta - \frac{\gamma E(x)}{\sqrt{Var(x) + \epsilon}}) \quad (3)$$

where  $E(x)$  and  $Var(x)$ , respectively, represent the mean and standard deviation that are calculated over the mini-batch,  $\gamma$  and  $\beta$  stand for the trainable parameters.

#### E. Multi-kernel Gaussian Process

The kernel space employs a kernel-based method, the Gaussian process, by integrating the deep features extracted from the 3D CNN and the location of the study regions. In this section, we first present a brief introduction about GP, and then the multi-kernel GP is adopted to integrate the deep features and spatial consistency.

1) *Gaussian Process*: A Gaussian process is a collection of random variables, any finite number of which have a joint Gaussian distribution [57]. Moreover, GP is a nonparametric

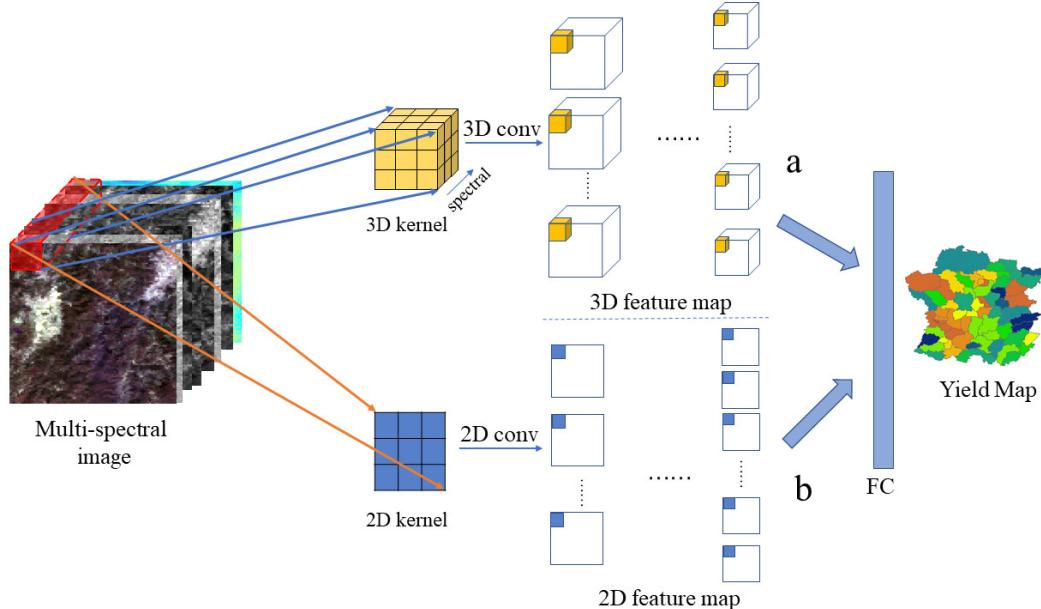


Fig. 3. Difference between 3D convolution(a) and 2D convolution(b) on crop yield prediction task.

probability model that is completely specified by the mean and covariance functions, and the distribution can be denoted as:

$$f(x) \sim gp(m(x), k(x, x')) \quad (4)$$

where  $x$  and  $x'$  stand for any random variables, and the mean function  $m(x)$  represents the expectation  $E[f(x)]$  and the kernel function  $k(x, x')$  defines the covariance  $cov(f(x), f(x'))$ .

Usually, for notational simplicity, we will take the mean function to be zero, and furthermore, we add noise  $\epsilon \sim gp(0, \sigma^2)$  so that the function distribution is closer to the real data.

For the regression problem, we consider:

$$y = f(x) + \epsilon \quad (5)$$

Moreover, we can infer the distribution of the observations  $y$ :

$$y \sim N(0, K(x, x') + \sigma_n^2 I_n) \quad (6)$$

where  $I_n$  is an  $n$ -dimensional identity matrix.

Hence, given the input latent variable set  $X = \{x_1, x_2, \dots, x_N\}$ , and we define the crop yield as  $Y = \{y_1, y_2, \dots, y_N\}$  which is a multivariate Gaussian process indexed by  $x_i$ ; then, we have:

$$\begin{aligned} P(Y|X, \theta) &= \prod_{i=1}^N p(y_i|x_i, \theta) \\ &= \frac{1}{(2\pi)^{\frac{DN}{2}} |K_Y|^{\frac{D}{2}}} \exp\left(-\frac{1}{2} \text{tr}(K_Y^{-1} Y Y^T)\right) \end{aligned} \quad (7)$$

where  $K_Y$  is an  $n \times n$ -order symmetric positive covariance matrix, and the element in matrix  $(K_Y)_{ij} = k_Y(x_i, x_j)$  is used to measure the similarity between  $x_i$  and  $x_j$ .  $I_n$  is an  $n$ -dimensional identity matrix.

2) *Spatial-spectral-spatio Kernel*: Based on Eq. 7, we can infer that the performance of prediction is dependent on the selection of the kernel function. Hence, in this section, a compound kernel for GP under the multi-kernel framework is constructed for capturing hierarchical features. The kernel employed consists of two separate components: the deep feature kernel and spatial feature kernel. The weight of each kernel is learned during an optimization phase.

### Deep spatial-spectral feature kernel

In this part, we intend to model the distance of the deep features extracted from the 3D CNN. For this purpose, a radial basis function kernel (RBF) is used to build the deep feature kernel based on the spatial-spectral features extracted near the top of the 3D CNN. Let the  $X = x_1, x_2 \dots x_i$  be the features generated after the 3D CNN. Then, the feature kernel can be expressed as:

$$K_{dfeature} = \sigma_f^2 \exp\left[-\frac{\|x - x'\|_2^2}{2l_f^2}\right] \quad (8)$$

where  $x$  denotes the normalized deep feature of the training dataset, and  $x'$  is the feature of the test data.  $\sigma_f^2$ ,  $\sigma_s^2$ ,  $l_f$ ,  $l_s$ ,  $l_y$  are the hyper-parameters of this kernel, which are defined as  $\theta_1 = \{\theta_l | l = 1 \dots M\}$

### Spatial consistency feature kernel

For the sake of capturing spatial consistency, we also construct the spatial kernel based on the RBF kernel. The kernel can be expressed as:

$$K_{spatial} = \sigma_s^2 \exp\left[-\frac{\|\Delta_{loc}\|_2^2}{2l_s^2}\right] \quad (9)$$

In the spatial kernel, the similarity is measured among the instances of the spatial location of the data points, where  $\Delta_{loc} = |loc - loc'|$  represents the distance metric between training and testing data. In this study, we used the central longitude and latitude of each county as the measurement of spatial distance. The hyper-parameters in this kernel can be defined as:  $\theta_2 = \{\theta_l | l = 1 \dots K\}$

### Spatial-spectral-spatio kernel

To simultaneously aggregate the information within deep features and the spatial consistency, we construct a new 'spatial-spectral-spatio' kernel by combining the deep feature kernel and spatial feature kernel using an MKL approach. In this new kernel, we construct two similarity matrices for each data source through Euclidean distance. The new kernel can be expressed by:

$$K = \sum w_s k_s(x_i, x_j) \quad (10)$$

where  $W_S = \{w_1, w_2\}$  are the weights of the subkernels. We define the weights positive and  $w_1 + w_2 = 1$ . The optimized weights  $w_1$  and  $w_2$  are capable of reflecting the contributions of each kind of feature.  $K_S = \{K_{dfeature}, K_{spatial}\}$  stands for the deep feature kernel and spatial kernel, respectively. As the parameters may be different for different kernels, we define the parameter vector as  $\theta_m = \{\theta_1, \theta_2, w_1, w_2\}$  for the new kernel. The detailed parameter settings will be discussed in Section IV .

3) *Prediction*: Finally, based on the new 'spatial-spectral-spatio' kernel, let  $Y$  denote the training yield dataset, and  $y'$  is the yield that needs to be predicted. The joint distribution of observations  $Y$  and predictions  $y'$  can be expressed as:

$$\begin{bmatrix} Y \\ y' \end{bmatrix} \sim \left(0, \begin{bmatrix} K(X, X) + \sigma_n^2 I_n & K(X, x') \\ K(x', X) & k(x', x') \end{bmatrix}\right) \quad (11)$$

$K(X, x') = K(x', X)^T$  represent the  $n \times 1$ -order covariance matrix between prediction variables  $x'$  and training set  $X$ .

Based on the joint distribution 11, we can calculate the probability distributions of prediction yield  $y'$ :

$$P(y' | x', X, Y) \sim N(E[y'], cov(y')) \quad (12)$$

where

$$E[y'] = K(x', X)[K(X, X) + \sigma_n^2 I_n]^{-1} y \quad (13)$$

And

$$\begin{aligned} cov(y') &= k(x', x') - K(x', X) \\ &\quad \times [K(X, X) + \sigma_n^2 I_n]^{-1} K(X, x') \end{aligned} \quad (14)$$

Then, we adopt  $E[y']$  as the final prediction result. The  $cov(y')$  is used as the measurement of uncertainty.

#### IV. EXPERIMENTS

In this section, we examine the superiority of the proposed 3DMKGP for crop yield prediction. Toward this end, several experiments have been conducted.

##### A. Data Description and Processing

1) *Remote Sensing Series*: In this paper, we focus on the winter wheat yield prediction in China. The remote sensing data used in this paper are acquired from the MODIS satellite, including two imagery products. The Surface Reflectance (MOD09A1) dataset [58] contains seven bands that provide rich spectral information. The Land Surface Temperature dataset (MYD11A2) [59] contains two bands that provide the necessary temperature information. Both of them provide eight-day composites. Based on the wheat growth period, only the image taken from the 280th day of the year to the 150th day of the next year are selected. In addition, our study only focuses on the crop area, so the MODIS Annual Land Cover data [60] are used to remove the noncrop pixels. All datasets are acquired from Google Earth Engine [61]. Finally, the remote sensing dataset ( $I_1^{(h \times w \times b)}, \dots, I_t^{(h \times w \times b)}$ ) is used for crop yield prediction and includes four dimensions. For such a high-dimensional training labeled data, it will be easy to overfit by training an end-to-end deep network directly based on the original data. To avoid this problem, we employ the same dimension reduction method proposed in [14]. After this, the original four-dimensional data are transformed into a three-dimensional histogram ( $I^{bins*time*bands}$ ).

2) *Crop Yield Data*: County-level wheat yields from 2001 to 2018 were taken from the Agricultural Statistic Yearbook [62] and Resource Discipline Innovation Platform [63]. All yields are reported in the units of metric tonnes per kilometer. Furthermore, because the MODIS cropland mask does not distinguish wheat from other crops, we ignored the regions that contributed the bottom 5% of total production in China. Hence, our model is only trained on regions with significant wheat crop cover and noisy crop yield values are filtered from regions. Finally, a total of 14287 counties are selected as the final study area. And then, we select the crop data of 2001–2014 as the training dataset. The crop data from 2015 to 2018 are used as the test dataset. Additionally, for the multi-kernel GP and traditional algorithms, we did not take the entire time series data as our test and training dataset because of the calculation limitations. Only the years from 2006–2014 are used as the training data, and the years of 2015, 2016, 2017 and 2018 as the test data. The number of the training and testing data samples displayed in Table I

##### B. Experiment Setup

1) *Implementation Details*: The proposed 3DMKGP and other deep learning methods are designed, trained and tested based on the TensorFlow framework [64]. The parameter settings of the proposed 3D CNN are listed in Table II. We use Adam as the optimizer and employ dropout after the dense layer to avoid overfitting. The batch normalization (BN) technique is also utilized at the end of all convolutional layers

TABLE I  
THE NUMBER OF TRAINING AND TESTING SAMPLES OF CROP YIELD DATA

	Training	Testing			
		2015	2016	2017	2018
3D CNN	11352	816	765	691	663
MKGP & Traditional	6377	816	765	691	663

to accelerate the convergence and improve the prediction accuracy. Our 3D CNN is trained with an initial learning rate of 0.003, and weight decay is  $1e^{-5}$ . A minibatch is set to 20. For hardware system configuration, all the following experiments are completed on a 64-b Intel Core CPU i7-6900K @ 3.20 GHz with an NVIDIA GeForce 1080 GPU. The RAM memory is 62 GB. Only one GPU is used, under CUDA version 9.0.176.

TABLE II  
THE CONFIGURATIONS OF 3D CNN STRUCTURE.

Name	Kernel	Stride	Output Size
Input	-	-	$32 \times 32 \times 9$
Cov1-1	$3 \times 3 \times 3$	1	$32 \times 32 \times 9 \times 64$
Pool1	$2 \times 2 \times 2$	2	$16 \times 16 \times 5 \times 64$
Cov2-1	$3 \times 3 \times 3$	1	$16 \times 16 \times 5 \times 128$
Pool2	$2 \times 2 \times 2$	2	$8 \times 8 \times 3 \times 128$
Cov3-1	$3 \times 3 \times 3$	1	$8 \times 8 \times 3 \times 256$
Cov3-2	$3 \times 3 \times 3$	1	$8 \times 8 \times 3 \times 256$
Pool3	$2 \times 2 \times 2$	2	$4 \times 4 \times 2 \times 256$
FC3	-	-	1024

2) *General Information*: To validate the effectiveness of the proposed 3DMKGP, it is compared with the most widely used crop yield prediction methods. They are summarized as follows:

(1) SVM [3], [65]: which is an effective approach for crop yield prediction. Here, an SVM is constructed based on the RBF kernel with the complexity parameter. The degree of the kernel function and the penalty factor C in the SVM are determined by cross-validation with the degree varying in the range from 1 to 7, and the C parameter is chosen from the set  $10^{-4}, 10^0, \dots, 10^6$ .

(2) RF [6], [66]: which is a supervised ensemble learning algorithm that acts based on decision trees. A grid search for model performance optimization is carried out with the five-fold cross-validation technique based on the  $R^2$  metric. In this paper, we search the number of generated trees from 50 to 500, and we allow a maximum tree depth of 7.

(3) Decision Tree (DT) [67]: which is another effective algorithm used for regression with a several leaf nodes. Here, we set the max depth of the trees at 10.

(4) 2D CNN [15]: a widely used deep learning method that has proven to achieve great success in crop yield prediction. The iterations are set up to 50000. The initial learning rate is 0.003.

TABLE III  
ACCURACY COMPARISON OF CROP YIELD PREDICTION AT COUNTY-LEVEL

	2015			2016			2017			2018			Avg.		
	RMSE	R <sup>2</sup>	MAPE												
DT-NDVI	1.27	0.44	23.99	1.34	0.38	26.12	1.24	0.46	25.58	1.17	0.44	22.99	1.25	0.43	24.67
RF-NDVI	1.16	0.53	22.31	1.17	0.53	23.16	1.19	0.50	23.64	1.21	0.43	23.33	1.20	0.43	24.22
SVM-NDVI	1.12	0.57	21.46	1.12	0.57	22.83	1.23	0.47	25.18	1.07	0.63	18.06	1.13	0.56	21.88
DT-Hist	1.20	0.51	24.96	1.13	0.56	23.20	1.13	0.55	23.63	1.09	0.51	20.81	1.14	0.53	23.15
RF-Hist	1.10	0.60	22.15	1.12	0.59	20.79	1.05	0.63	20.68	1.03	0.59	18.55	1.08	0.60	20.54
SVM-Hist	1.01	0.66	20.38	1.07	0.63	20.89	1.02	0.66	18.99	0.97	0.63	18.62	1.02	0.64	19.72
LSTM	0.93	0.70	17.70	1.12	0.57	23.14	0.97	0.67	19.06	0.93	0.64	17.34	0.98	0.64	19.31
2D CNN	0.96	0.68	19.51	0.89	0.72	17.53	0.96	0.67	19.70	0.95	0.63	19.49	0.94	0.67	19.05
3DMKGP	<b>0.76</b>	<b>0.79</b>	<b>14.68</b>	<b>0.71</b>	<b>0.82</b>	<b>13.88</b>	<b>0.73</b>	<b>0.81</b>	<b>14.87</b>	<b>0.73</b>	<b>0.78</b>	<b>13.98</b>	<b>0.73</b>	<b>0.80</b>	<b>14.35</b>

(5) LSTM [6], [68]: another widely used deep-learning method that has shown promising performance in classification and regression tasks.

3) *Evaluation Matrix*: In this paper, in order to evaluate the performance of different crop yield prediction methods, we consider the most common metrics for evaluating regression results, namely, RMSE, R<sup>2</sup> and MAPE, between the ground truth yield and the estimated yield as the evaluation criteria, which can defined as follows:

$$RMSE = \sqrt{\frac{1}{m} \sum_{i=1}^m (p_i - y_i)^2} \quad (15)$$

$$R^2 = 1 - \sum_{i=1}^m (p_i - y_i)^2 / \sum_{i=1}^m (p_i - \bar{y})^2 \quad (16)$$

$$MAPE = \frac{100\%}{m} \sum_{i=1}^m \left| \frac{p_i - y_i}{p_i} \right| \quad (17)$$

where m is the number of predicting data points, y<sub>i</sub> and p<sub>i</sub> stand for the label and predicting yield, respectively, and  $\bar{y}$  denotes the average value of the predicted results.

### C. Comparisons with Competing Methods

In this section, we compare the proposed method with three traditional models (SVM, RF and RT) and two deep learning models (2D CNN and LSTM) to prove the necessity and advancement of the proposed 3DMKGP in crop yield prediction. In this experiment, the input of deep learning methods is a three-dimensional histogram with the size of 32×32×9. The performance of traditional models is evaluated on two datasets, including the most widely used hand-crafted feature NDVI (DT-NDVI, RF-NDVI, SVM-NDVI) and the histogram utilized in deep learning methods (DT-Hist, RF-Hist, SVM- Hist). The NDVI features are extracted from the original remote sensing image of the region of interest during the experimental time series, and the same data processing and dimension reduction method is employed. Therefore, the NDVI images are transformed into a 32\*32 NDVI histogram and then flatten into a 1024- dimensional vector as the first input. And the histogram is also transformed into a flatten vector with the size of 32\*32\*9=9216 as the final input. All the hyper-parameters in traditional machine learning algorithms

are determined by fivefold cross-validation, and for the hyper-parameters  $\theta_m$  of the new 'spatial-spectral-spatio' kernel described in Section III, we first optimized the  $\theta_1$  and  $\theta_2$  within each kernel individually. For the  $\sigma_f$  and  $\sigma_s$ , the algorithm searches from 1 to 19.  $l_f$  and  $l_s$  are searching from  $10^{-2}$  to  $10^3$ . Finally, the weights of each kernel are also optimized between 0 – 1 at 0.1 intervals. The RMSE of all competitors and the proposed method on the crop yield prediction task can be found in Table III.

We first compare the proposed method with traditional methods for a baseline comparison. For the traditional models, the performance of DT-NDVI, RF-NDVI and SVM-NDVI is inferior to DT-Hist, RF- Hist, SVM- Hist. This is because the histogram can provide more spectral features compared with the NDVI. However, the proposed 3DMKGP effectively surpasses all traditional models and obtains the best results in all of the testing years. Regarding the traditional methods, the SVM-hist obtained a better performance among them. Compared to the SVM-hist, the decreases in RMSE achieved by 3DMKGP are 0.25, 0.41, 0.29 and 0.24. The improvements in R2 are 0.13, 0.19, 0.15 and 0.15 for four testing years. These findings proved that our methods have better advantages over traditional models in crop yield prediction. To further verify the superiority of 3DMKGP, we complement various deep learning methods for crop yield prediction. The deep learning methods perform better than the traditional methods. However, our method still surpasses these deep learning methods in all testing years, with the highest R<sup>2</sup> of 0.80 and lowest RMSE and MAPE values of 0.73 and 14.35, respectively. Compared with the LSTM, the 3DMKGP reduced the RMSE by 25.5% and MAPE by 25.7% on average. Compared with the 2D CNN, the average R<sup>2</sup> of 3DMKGP increased by 19.4% and the RMSE and MAPE decreased by 22.3% and 24.6%, respectively.

To further present the effectiveness of the proposed method, we divided the testing data points into three main parts according to the Fig.1, and visualize the prediction errors of these three wheat producing areas. Corresponding to Table III, similar conclusions can be drawn from the prediction error maps presented in Figs. 4 – 6 from which it is obvious that the maps provided by the proposed model achieve the lowest error in most counties. Specifically, the regions that are seriously overestimated and underestimated by competing methods are apparently corrected in our 3DMKGP. These results reveal the

TABLE IV  
COMPARISON ANALYSIS OF CRUCIAL COMPONENT IN 3DMKGP

	2015			2016			2017			2018			Avg.		
	RMSE	R <sup>2</sup>	MAPE												
LSTMGP	0.95	0.69	18.32	1.00	0.66	19.77	0.94	0.62	18.48	0.88	0.68	16.61	0.94	0.66	18.32
2DGP	0.98	0.66	20.00	0.88	0.73	17.74	0.96	0.68	19.96	0.90	0.67	18.48	0.93	0.68	19.04
3DGP	0.77	0.79	20.61	0.75	0.80	15.54	0.80	0.77	18.02	0.76	0.76	15.21	0.77	0.78	17.34
LSTM-MKGP	0.94	0.69	17.83	0.96	0.68	18.88	0.91	0.71	17.99	0.84	0.70	15.86	0.91	0.69	17.64
2DMKGP	0.86	0.74	16.41	0.81	0.77	15.40	0.83	0.76	15.82	0.83	0.72	15.78	0.83	0.74	15.85
3DSGP	<b>0.72</b>	<b>0.82</b>	<b>14.41</b>	0.74	0.81	13.36	0.77	0.79	17.33	0.76	0.76	13.20	0.75	0.79	14.57
3DMKGP	0.76	0.79	14.68	<b>0.71</b>	<b>0.82</b>	<b>13.88</b>	<b>0.73</b>	<b>0.81</b>	<b>14.87</b>	<b>0.73</b>	<b>0.78</b>	<b>13.98</b>	<b>0.73</b>	<b>0.80</b>	<b>14.35</b>

fact that our 3DMKGP model can construct a more powerful crop-related representation. The first reason is that the 3D CNN in the 3DMKGP can make full use of both spatial and spectral information from multispectral images. Second, the new 'spatial-spectral-spatio' kernel makes it possible for the 3DMKGP to simultaneously capture deep features and spatial correlation and automatically adjust the weight of each feature, which enables 3DMKGP with better suitability for crop yield prediction.

#### D. Analysis of the Proposed Method: Effect of Different Components

In this section, based on above explanation, extensive experiments are conducted to verify the effectiveness of the crucial components in 3DMKGP. Here, we analyze the superiority of the 3D CNN in extracting spatial-spectral features, and then we compare the effect of the multi-kernel GP strategy. These experiments demonstrate different contributions of components and provide more insights of our proposed method.

1) 3D CNN Component: We first evaluated the effectiveness of the 3D CNN in our framework. For this purpose, we replace the 3D CNN component in the proposed framework with different 2D networks, including 2D CNN and LSTM (CNN-MKGP, LSTM-MKGP). The experimental results shown in Table IV clearly demonstrate that 3DMKGP outperforms CNN-MKGP and LSTM-MKGP in all testing years, and achieves a lower RMSE of 0.1 and 0.18, respectively. CNN-MKGP has slightly improved compared with LSTM-MKGP; however, it also leads to unsatisfactory performance because they do not fully exploit the spatial-spectral features. Compared to the CNN-MKGP, the 3DMKGP obtains 8% gains in R<sup>2</sup>. Due to the clear improvement of the 3DMKGP, it can be confirmed that the 3D convolutions make it possible to fully use spatial-spectral information.

2) Multi-kernel GP Component: In this experiment, we investigate the effectiveness of the designed multi-kernel GP (MKGP). For this purpose, we first constructed a single kernel GP where the kernel is structured with the deep features

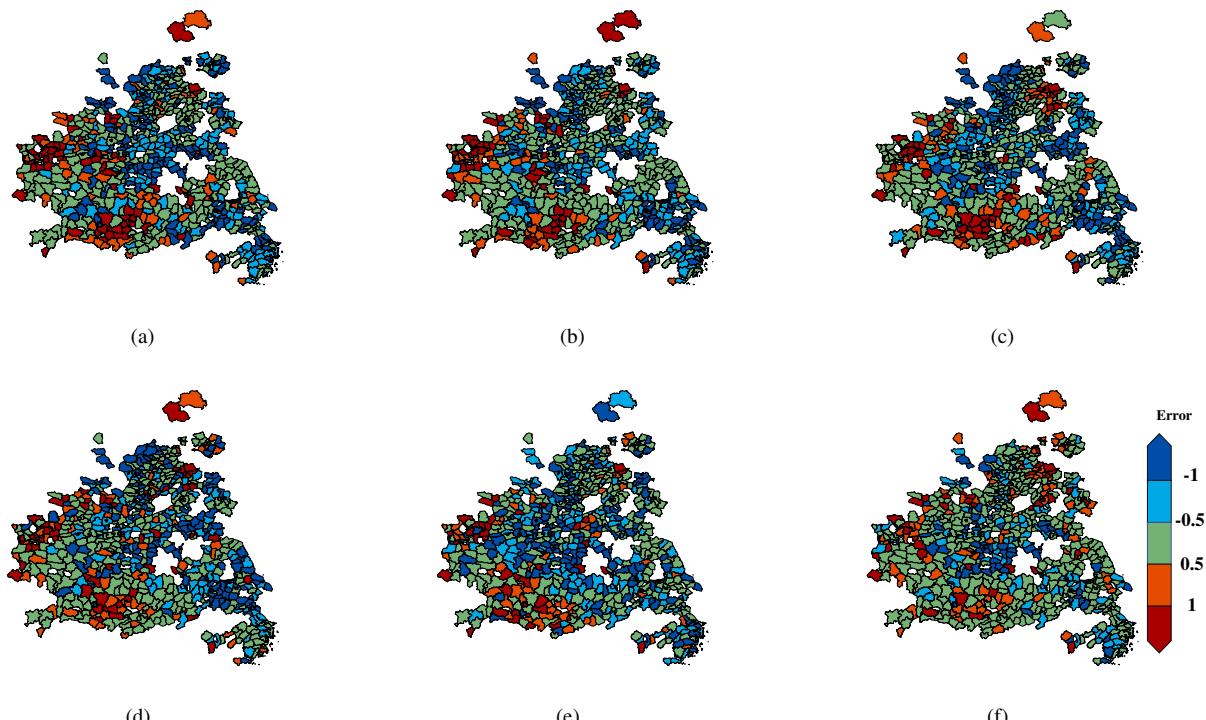


Fig. 4. County-level error maps of Northern winter wheat region. (1)DT. (2)RF. (3)SVM. (4)LSTM. (5)2D CNN. (6) 3DMKGP.

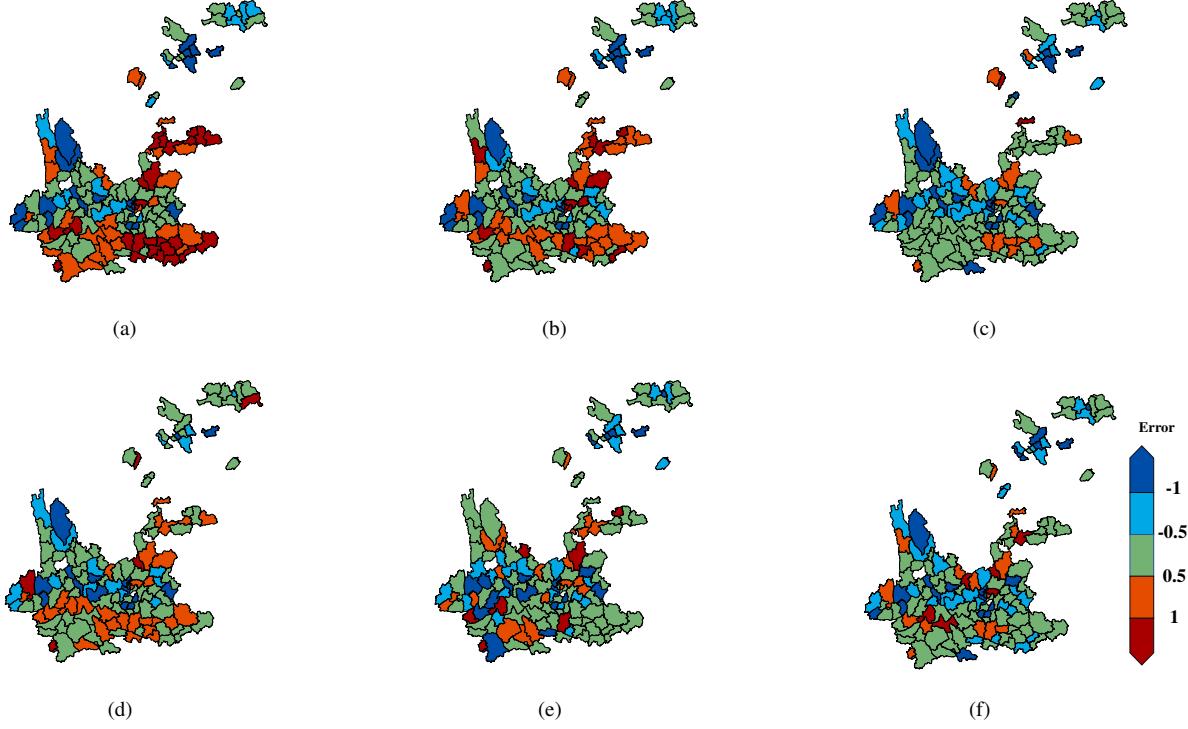


Fig. 5. County-level error maps of Southwest winter wheat region. (1)DT. (2)RF. (3)SVM. (4)LSTM. (5)2D CNN. (6) 3DMKGP.

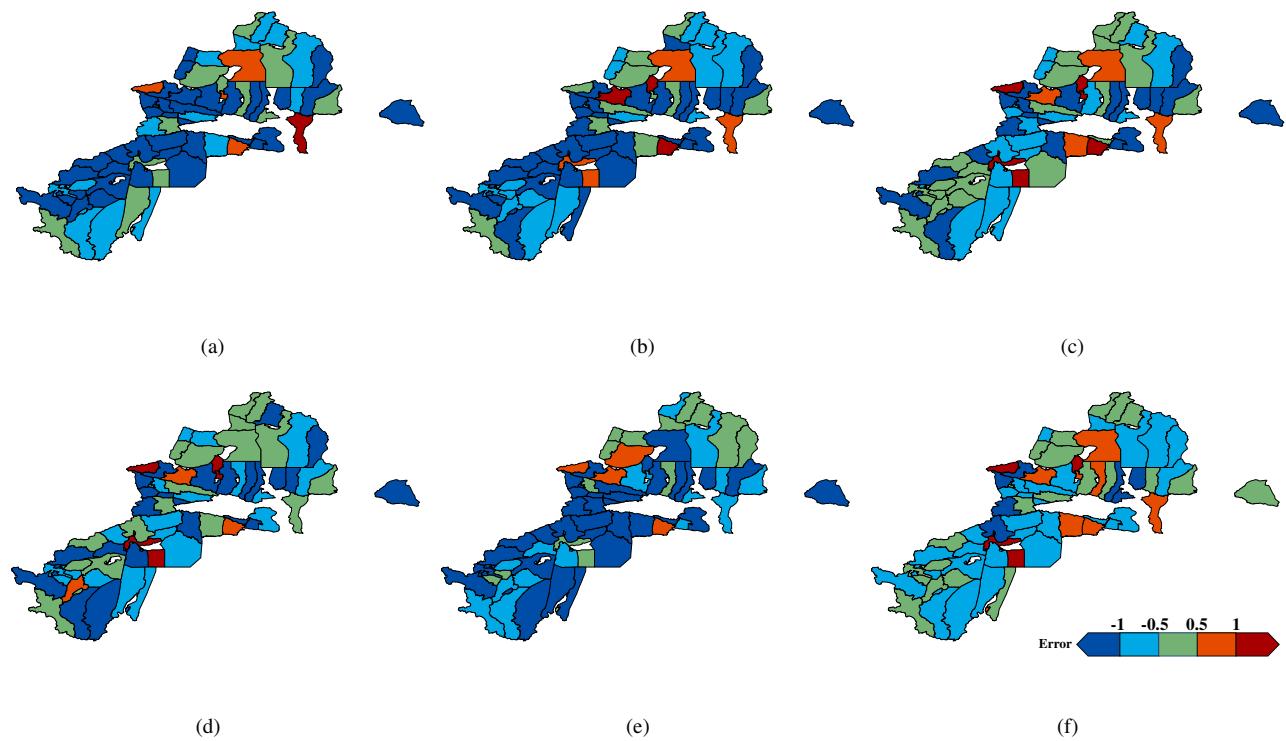


Fig. 6. County-level error maps of Xinjiang winter wheat region. (1)DT. (2)RF. (3)SVM. (4)LSTM. (5)2D CNN. (6) 3DMKGP. The error bar represents the difference between prediction and real yield.

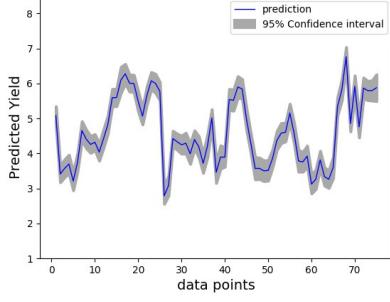


Fig. 7. The uncertainty estimation of prediction results. The shades of gray represent 95% confidence.

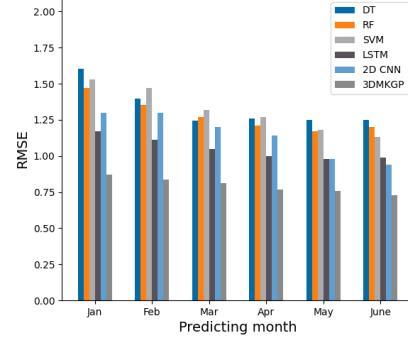


Fig. 8. Experiments of different models with different growth period at county level prediction.

extracted from LSTM, 2DCNN and 3DCNN (LSTMGP, 2DGP and 3DGP). And then, we compared them with our multi-kernel GP that is also implemented on three different deep learning networks (LSTMGP, 2DMKGP and 3DMKGP). Moreover, since the gaussian process is also utilized for incorporating the spatial dependency in [14]. We further implement the experiment to compare our multi-kernel GP with the GP in [1]. Especially, we have changed the deep network of [14] with our proposed 3D CNN (3DSGP) to avoid the influence of network structure. The comparison results are shown in Table IV.

It is noticed that the results obtained by MKGP by integrating all features are always superior in all testing years. In greater detail, the RMSE provided by MKGP with the 3D CNN, 2D CNN and LSTM are 0.04, 0.10 and 0.03 lower than that of the DGP in RMSE, respectively. MKGP improves on the performance of the DGP and obtains 0.02, 0.06 and 0.03 gains in R2. Hence, it can be concluded that by modeling the long-term dependencies in the spatial dimension, it can provide a higher performance for crop yield prediction. These results demonstrate that our MKGP better exploits the hierarchical features and is more effective than a single feature for the crop yield prediction. Besides, comparing with the 3DSGP, our method obtains a better result in three of the four testing years. This further improve the effectiveness of the proposed multi-kernel GP.

3) Uncertainty Estimation: In real word applications, one of the key points of the prediction problem is offering the uncertainty of the results. The 3DMKGP can learn a probability distribution of the predictions which provides a natural

way for uncertainty estimation. In this part, the 95% prediction intervals that are calculated from the 3DMKGP based on Eq.14 have been used to estimate the random error in the crop yield prediction. Fig.7 illustrates the predictions and corresponding uncertainty obtained from 3DMKGP.

#### E. Crop Yield Prediction Before Harvest

The ultimate goal of crop yield prediction is predicting the crop yield before harvest. To this end, we train and test our model based on different time windows. The starting of windows is triggered by the sowing period (October), and the ending point is varying from February to June(harvest) of the following year. Fig. shows the averaged RMSE in different time windows. It can be observed that with the ending point closer to the sowing month, the performance of all models gradually improved. All the other models except 3DMKGP did not perform well in the early months since the information in the early months is not obvious. Moreover, our proposed method 3DMKGP outperforms other methods consistently at all time windows, demonstrating the robustness of the proposed method. Besides, it is noticed that the RMSE remains steady after May in the proposed method, suggesting our method can predict the crop yield two months before harvest.

#### F. Parameter and Feature Analysis

1) Experiments with Different Kernel Depths : One of the contributions in this study is that the 3D CNN is applied to explicitly spatial-spectral features. Different from the 2D

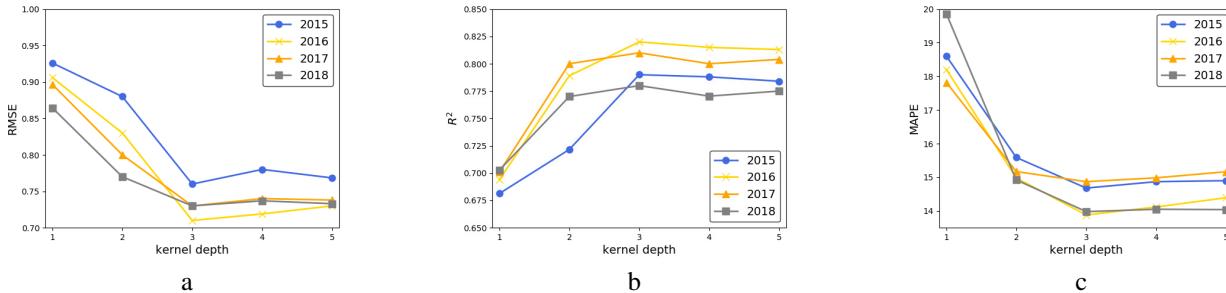


Fig. 9. Evaluation performance of different kernel depth in terms of RMSE (a),  $R^2$  (b) and MAPE (c).

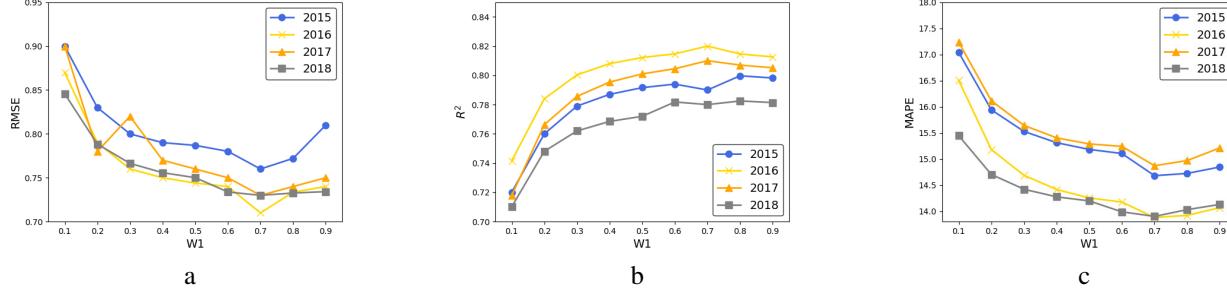


Fig. 10. Evaluation performance of different feature weights in terms of RMSE (a),  $R^2$  (b) and MAPE (c),  $w_1$  stand for the weight of spatial-spectral feature kernel.

CNN, the third dimension of 3D convolution layers can explore the spectral correlation through the three-dimensional kernel. To explore the influence of the kernel size, we vary the kernel depth while keeping all other common settings fixed. Here, we experiment with different kernel depths that are set from 1 to 5. Note that depth=1 has the same architecture as the 2D CNN. The experimental results of the 3D CNN with different kernel spectral depths are shown in Fig.9. It is observed that depth=3 achieves the best results among all the kernels. As expected, depth=1 has the worst results since it is equal to a 2D CNN that cannot fully excavate spatial-spectral features.

**2) Experiments with Different Weights of Hierarchical Features:** The new 'spatial-spectral-spatio' kernel is the combination of two kernels, and the optimized weights of each kernel will directly give a ranked importance to each kind of feature. To show that each feature has contributed to the overall result, we highlight the contribution of the deep spatial-spectral feature and the spatial consistency feature by adjusting the weight of the deep feature kernel  $w_1$  in Eq.10. From Fig.10, we can see that the optimized weights with  $w_1 = 0.7$  and  $w_2 = 0.3$  achieve the best performance, which indicates that deep features play a dominant role in predicting crop yield. With the value of  $w_1$  decreasing, the RMSE tends to continuously increase. Additionally, the spatial consistency feature also plays an important role in predicting crop yield with a weight of 0.3. The analysis of  $w_i$  can help illustrate each feature's contribution and provide more insight for feature selection in crop yield prediction.

## V. CONCLUSION

In this article, an effective model is proposed that involves hierarchical features for crop yield prediction. We first utilize the superiority of the 3D CNN that enables extracting spatial-spectral features from raw RSIs. To further account for the features beyond the RSIs, we employ an MKL framework that fuses these deep features with spatial consistency features between data points. Moreover, the prediction process is implemented in the kernel-based GP method.

County-level winter wheat in China is predicted in this paper. Extensive experimental results demonstrate that our method not only outperforms other traditional and deep learning methods but also extracts more discriminative feature representations, which demonstrates its potential application

in different prediction problems. In terms of further research, we can fuse more crop-related features with the multi-kernel approach. Furthermore, our method can be generalized for yield prediction of other crop types.

## ACKNOWLEDGMENT

The authors would like to thank the 13th Five-year Informatization Plan of Chinese Academy of Sciences(Grant No.XXH13505-07) for making the crop yield data available.

## REFERENCES

- [1] K. Dzotsi, B. Basso, and J. Jones, "Development, uncertainty and sensitivity analysis of the simple salus crop model in dssat," *Ecological Modelling*, vol. 260, pp. 62–76, 2013.
- [2] R. Srinivasan, X. Zhang, and J. Arnold, "Swat ungauged: hydrological budget and crop yield predictions in the upper mississippi river basin," *Transactions of the ASABE*, vol. 53, no. 5, pp. 1533–1546, 2010.
- [3] H. Aghighi, M. Azadkabt, D. Ashourloo, H. S. Shahabi, and S. Radim, "Machine learning regression techniques for the silage maize yield prediction using time-series images of landsat 8 oli," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 11, no. 12, pp. 4563–4577, 2018.
- [4] N. Gandhi, L. J. Armstrong, O. Petkar, and A. K. Tripathy, "Rice crop yield prediction in india using support vector machines," in *2016 13th International Joint Conference on Computer Science and Software Engineering (JCSSE)*. IEEE, 2016, pp. 1–5.
- [5] U. Saeed, J. Dempewolf, I. Becker-Reshef, A. Khan, A. Ahmad, and S. A. Wajid, "Forecasting wheat yield from weather data and modis ndvi using random forests for punjab province, pakistan," *International journal of remote sensing*, vol. 38, no. 17, pp. 4831–4854, 2017.
- [6] J. Cao, Z. Zhang, F. Tao, L. Zhang, Y. Luo, J. Zhang, J. Han, and J. Xie, "Integrating multi-source data for rice yield prediction across china using machine learning and deep learning approaches," *Agricultural and Forest Meteorology*, vol. 297, p. 108275, 2021.
- [7] M. Kaul, R. L. Hill, and C. Walther, "Artificial neural networks for corn and soybean yield prediction," *Agricultural Systems*, vol. 85, no. 1, pp. 1–18, 2005.
- [8] B. Ji, Y. Sun, S. Yang, and J. Wan, "Artificial neural networks for rice yield prediction in mountainous regions," *The Journal of Agricultural Science*, vol. 145, no. 3, p. 249, 2007.
- [9] P. Bose, N. K. Kasabov, L. Bruzzone, and R. N. Hartono, "Spiking neural networks for crop yield estimation based on spatiotemporal analysis of image time series," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 54, no. 11, pp. 6563–6573, 2016.
- [10] G. Cheng and J. Han, "A survey on object detection in optical remote sensing images," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 117, pp. 11–28, 2016.
- [11] O. A. Penatti, K. Nogueira, and J. A. Dos Santos, "Do deep features generalize from everyday objects to remote sensing and aerial scenes domains?" in *Proceedings of the IEEE conference on computer vision and pattern recognition workshops*, 2015, pp. 44–51.

- [12] C. Chen, K. Li, S. G. Teo, G. Chen, X. Zou, X. Yang, R. C. Vijay, J. Feng, and Z. Zeng, "Exploiting spatio-temporal correlations with multiple 3d convolutional neural networks for citywide vehicle flow prediction," in *2018 IEEE international conference on data mining (ICDM)*. IEEE, 2018, pp. 893–898.
- [13] J.-W. Ma, C.-H. Nguyen, K. Lee, and J. Heo, "Regional-scale rice-yield estimation using stacked auto-encoder with climatic and modis data: a case study of south korea," *International Journal of Remote Sensing*, vol. 40, no. 1, pp. 51–71, 2019.
- [14] J. You, X. Li, M. Low, D. Lobell, and S. Ermon, "Deep gaussian process for crop yield prediction based on remote sensing data," in *Thirty-First AAAI conference on artificial intelligence*, 2017.
- [15] P. Neuvuori, N. Narra, and T. Lipping, "Crop yield prediction with deep convolutional neural networks," *Computers and electronics in agriculture*, vol. 163, p. 104859, 2019.
- [16] A. Kaneko, T. Kennedy, L. Mei, C. Sintek, M. Burke, S. Ermon, and D. Lobell, "Deep learning for crop yield prediction in africa."
- [17] A. X. Wang, C. Tran, N. Desai, D. Lobell, and S. Ermon, "Deep transfer learning for crop yield prediction with remote sensing data," in *Proceedings of the 1st ACM SIGCAS Conference on Computing and Sustainable Societies*, 2018, pp. 1–5.
- [18] J. Gu, Z. Wang, J. Kuen, L. Ma, A. Shahroud, B. Shuai, T. Liu, X. Wang, G. Wang, J. Cai *et al.*, "Recent advances in convolutional neural networks," *Pattern Recognition*, vol. 77, pp. 354–377, 2018.
- [19] Q. Yang, L. Shi, J. Han, Y. Zha, and P. Zhu, "Deep convolutional neural networks for rice grain yield estimation at the ripening stage using uav-based remotely sensed images," *Field Crops Research*, vol. 235, pp. 142–153, 2019.
- [20] A. Haghatalab, J. Crain, S. Mondal, J. Rutkoski, R. P. Singh, and J. Poland, "Application of geographically weighted regression to improve grain yield prediction from unmanned aerial system imagery," *Crop Science*, vol. 57, no. 5, pp. 2478–2489, 2017.
- [21] M. Maimaitijiang, V. Sagan, P. Sidike, S. Hartling, F. Esposito, and F. B. Fritsch, "Soybean yield prediction from uav using multimodal data fusion and deep learning," *Remote Sensing of Environment*, vol. 237, p. 111599, 2020.
- [22] L. Anselin, R. Bongiovanni, and J. Lowenberg-DeBoer, "A spatial econometric approach to the economics of site-specific nitrogen management in corn production," *American Journal of Agricultural Economics*, vol. 86, no. 3, pp. 675–687, 2004.
- [23] N. R. Peralta, Y. Assefa, J. Du, C. J. Barden, and I. A. Ciampitti, "Mid-season high-resolution satellite imagery for forecasting site-specific corn yield," *Remote Sensing*, vol. 8, no. 10, p. 848, 2016.
- [24] Y.-S. Shiu and Y.-C. Chuang, "Yield estimation of paddy rice based on satellite imagery: comparison of global and local regression models," *Remote Sensing*, vol. 11, no. 2, p. 111, 2019.
- [25] H. Russello, "Convolutional neural networks for crop yield prediction using satellite images," *IBM Center for Advanced Studies*, 2018.
- [26] P. Neuvuori, N. Narra, P. Linna, and T. Lipping, "Crop yield prediction using multitemporal uav data and spatio-temporal deep learning models," *Remote Sensing*, vol. 12, no. 23, p. 4000, 2020.
- [27] M. Abedinpour, A. Sarangi, T. Rajput, M. Singh, H. Pathak, and T. Ahmad, "Performance evaluation of aquacrop model for maize crop in a semi-arid environment," *Agricultural Water Management*, vol. 110, pp. 55–66, 2012.
- [28] C. M. T. Soler, P. C. Sentelhas, and G. Hoogenboom, "Application of the csm-ceres-maize model for planting date evaluation and yield forecasting for maize grown off-season in a subtropical environment," *European Journal of Agronomy*, vol. 27, no. 2-4, pp. 165–177, 2007.
- [29] B. A. Keating, P. S. Carberry, G. L. Hammer, M. E. Probert, M. J. Robertson, D. Holzworth, N. I. Huth, J. N. Hargreaves, H. Meinke, Z. Hochman *et al.*, "An overview of apsim, a model designed for farming systems simulation," *European journal of agronomy*, vol. 18, no. 3-4, pp. 267–288, 2003.
- [30] D. K. Bolton and M. A. Friedl, "Forecasting crop yield using remotely sensed vegetation indices and crop phenology metrics," *Agricultural and Forest Meteorology*, vol. 173, pp. 74–84, 2013.
- [31] F. A. Vega, F. C. Ramirez, M. P. Saiz, and F. O. Rosúa, "Multi-temporal imaging using an unmanned aerial vehicle for monitoring a sunflower crop," *Biosystems Engineering*, vol. 132, pp. 19–27, 2015.
- [32] G. Ruß and R. Kruse, "Feature selection for wheat yield prediction," in *Research and Development in Intelligent Systems XXVI*. Springer, 2010, pp. 465–478.
- [33] J. Frausto-Solis, A. Gonzalez-Sanchez, and M. Larre, "A new method for optimal cropping pattern," in *Mexican International Conference on Artificial Intelligence*. Springer, 2009, pp. 566–577.
- [34] M. Zaefizadeh, A. Jalili, M. Khayatnezhad, R. Gholamin, and T. Mokhtari, "Comparison of multiple linear regressions (mlr) and artificial neural network (ann) in predicting the yield using its components in the hulless barley," *Advances in Environmental Biology*, pp. 109–114, 2011.
- [35] X. Zhou, H. Zheng, X. Xu, J. He, X. Ge, X. Yao, T. Cheng, Y. Zhu, W. Cao, and Y. Tian, "Predicting grain yield in rice using multi-temporal vegetation indices from uav-based multispectral and digital imagery," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 130, pp. 246–255, 2017.
- [36] J. Xue and B. Su, "Significant remote sensing vegetation indices: A review of developments and applications," *Journal of Sensors*, vol. 2017, 2017.
- [37] W. Zhou, D. Ming, X. Lv, K. Zhou, H. Bao, and Z. Hong, "So-cnn based urban functional zone fine division with vhr remote sensing image," *Remote Sensing of Environment*, vol. 236, p. 111458, 2020.
- [38] P. Li, X. He, X. Cheng, X. Gao, R. Li, M. Qiao, D. Li, F. Qiu, and Z. Li, "Object extraction from very high-resolution images using a convolutional neural network based on a noisy large-scale dataset," *IEEE Access*, vol. 7, pp. 122 784–122 795, 2019.
- [39] N. Kussul, M. Lavreniuk, S. Skakun, and A. Shelestov, "Deep learning classification of land cover and crop types using remote sensing data," *IEEE Geoscience and Remote Sensing Letters*, vol. 14, no. 5, pp. 778–782, 2017.
- [40] C. Zhang, I. Sargent, X. Pan, H. Li, A. Gardiner, J. Hare, and P. M. Atkinson, "Joint deep learning for land cover and land use classification," *Remote sensing of environment*, vol. 221, pp. 173–187, 2019.
- [41] A. Peerlinck, J. Sheppard, and B. Maxwell, "Using deep learning in yield and protein prediction of winter wheat based on fertilization prescriptions in precision agriculture," in *International Conference on Precision Agriculture (ICPA)*, 2018.
- [42] S. Khaki and L. Wang, "Crop yield prediction using deep neural networks," *Frontiers in plant science*, vol. 10, p. 621, 2019.
- [43] S. Ji, W. Xu, M. Yang, and K. Yu, "3d convolutional neural networks for human action recognition," *IEEE transactions on pattern analysis and machine intelligence*, vol. 35, no. 1, pp. 221–231, 2012.
- [44] S. Mei, J. Ji, Y. Geng, Z. Zhang, X. Li, and Q. Du, "Unsupervised spatial-spectral feature learning by 3d convolutional autoencoder for hyperspectral classification," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 57, no. 9, pp. 6808–6820, 2019.
- [45] F. I. Alam, J. Zhou, A. W.-C. Liew, X. Jia, J. Chanussot, and Y. Gao, "Conditional random field and deep feature learning for hyperspectral image classification," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 57, no. 3, pp. 1612–1628, 2018.
- [46] Y. Chen, H. Jiang, C. Li, X. Jia, and P. Ghamisi, "Deep feature extraction and classification of hyperspectral images based on convolutional neural networks," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 54, no. 10, pp. 6232–6251, 2016.
- [47] X. Yang, Y. Ye, X. Li, R. Y. Lau, X. Zhang, and X. Huang, "Hyperspectral image classification with deep learning models," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 56, no. 9, pp. 5408–5423, 2018.
- [48] X. Mei, E. Pan, Y. Ma, X. Dai, J. Huang, F. Fan, Q. Du, H. Zheng, and J. Ma, "Spectral-spatial attention networks for hyperspectral image classification," *Remote Sensing*, vol. 11, no. 8, p. 963, 2019.
- [49] C. Shi and C.-M. Pun, "Multi-scale hierarchical recurrent neural networks for hyperspectral image classification," *Neurocomputing*, vol. 294, pp. 82–93, 2018.
- [50] N. D. Lawrence, "Gaussian process latent variable models for visualisation of high dimensional data," in *Advances in neural information processing systems*, 2004, pp. 329–336.
- [51] B. Gu, V. S. Sheng, Z. Wang, D. Ho, S. Osman, and S. Li, "Incremental learning for  $\nu$ -support vector regression," *Neural networks*, vol. 67, pp. 140–150, 2015.
- [52] P. Cao, X. Liu, J. Yang, D. Zhao, W. Li, M. Huang, and O. Zaiane, "A multi-kernel based framework for heterogeneous feature selection and over-sampling for computer-aided detection of pulmonary nodules," *Pattern Recognition*, vol. 64, pp. 327–346, 2017.
- [53] D. Tuia, G. Camps-Valls, G. Matasci, and M. Kanevski, "Learning relevant image features with multiple-kernel classification," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 48, no. 10, pp. 3780–3791, 2010.
- [54] Y.-R. Yeh, T.-C. Lin, Y.-Y. Chung, and Y.-C. F. Wang, "A novel multiple kernel learning framework for heterogeneous feature fusion and variable selection," *IEEE Transactions on multimedia*, vol. 14, no. 3, pp. 563–574, 2012.
- [55] Y. Gu, Q. Wang, X. Jia, and J. A. Benediktsson, "A novel mkl model of integrating lidar data and msi for urban area classification," *IEEE*

- transactions on geoscience and remote sensing*, vol. 53, no. 10, pp. 5312–5326, 2015.
- [56] A. Vetrivel, M. Gerke, N. Kerle, F. Nex, and G. Vosselman, “Disaster damage detection through synergistic use of deep learning and 3d point cloud features derived from very high resolution oblique aerial images, and multiple-kernel-learning,” *ISPRS journal of photogrammetry and remote sensing*, vol. 140, pp. 45–59, 2018.
  - [57] C. E. Rasmussen, “Gaussian processes in machine learning,” in *Summer School on Machine Learning*. Springer, 2003, pp. 63–71.
  - [58] E. Vermote, “Mod09a1 modis/terra surface reflectance 8-day l3 global 500m sin grid v006,” *NASA EOSDIS Land Processes DAAC*, vol. 10, 2015.
  - [59] Z. Wan, S. Hook, and G. Hulley, “Myd11a2 modis/aqua land surface temperature and the emissivity 8-day l3 global 1km sin grid. nasa lp daac,” 2015.
  - [60] M. Friedl and D. Sulla-Menashe, “Mcd12q1 modis/terra+ aqua land cover type yearly l3 global 500m sin grid v006 [data set],” *NASA EOSDIS Land Processes DAAC*, vol. 10, 2015.
  - [61] N. Gorelick, M. Hancher, M. Dixon, S. Ilyushchenko, D. Thau, and R. Moore, “Google earth engine: Planetary-scale geospatial analysis for everyone,” *Remote sensing of Environment*, vol. 202, pp. 18–27, 2017.
  - [62] N. B. of Statistics, “China rural statistical yearbook,” [EB/OL], <http://www.stats.gov.cn/> Accessed April 4, 2010.
  - [63] C. A. of Science, “Resource discipline innovation platform,” [EB/OL], <http://www.data.ac.cn/server/database.html/> Accessed April 4, 2010.
  - [64] M. Abadi, P. Barham, J. Chen, Z. Chen, A. Davis, J. Dean, M. Devin, S. Ghemawat, G. Irving, M. Isard *et al.*, “Tensorflow: A system for large-scale machine learning,” in *12th {USENIX} symposium on operating systems design and implementation ({OSDI} 16)*, 2016, pp. 265–283.
  - [65] Y. Cai, K. Guan, D. Lobell, A. B. Potgieter, S. Wang, J. Peng, T. Xu, S. Asseng, Y. Zhang, L. You *et al.*, “Integrating satellite and climate data to predict wheat yield in australia using machine learning approaches,” *Agricultural and forest meteorology*, vol. 274, pp. 144–159, 2019.
  - [66] D. Gómez, P. Salvador, J. Sanz, and J. L. Casanova, “Potato yield prediction using machine learning techniques and sentinel 2 data,” *Remote Sensing*, vol. 11, no. 15, p. 1745, 2019.
  - [67] J. Han, Z. Zhang, J. Cao, Y. Luo, L. Zhang, Z. Li, and J. Zhang, “Prediction of winter wheat yield based on multi-source data and machine learning in china,” *Remote Sensing*, vol. 12, no. 2, p. 236, 2020.
  - [68] R. A. Schwalbert, T. Amado, G. Corassa, L. P. Pott, P. V. Prasad, and I. A. Ciampitti, “Satellite-based soybean yield forecast: Integrating machine learning and weather data for improving crop yield prediction in southern brazil,” *Agricultural and Forest Meteorology*, vol. 284, p. 10786, 2020.