

## Research Article

# Depth and Width Changeable Network-Based Deep Kernel Learning-Based Hyperspectral Sensor Data Analysis

Jing Liu <sup>1</sup>, Tingting Wang <sup>2</sup>, and Yulong Qiao <sup>1</sup>

<sup>1</sup>College of Information and Communication Engineering, Harbin Engineering University, Harbin 150001, China

<sup>2</sup>School of Electronics and Information Engineering, Harbin Institute of Technology, Harbin, Heilongjiang 150001, China

Correspondence should be addressed to Yulong Qiao; [qiaoyulonghrbeu@126.com](mailto:qiaoyulonghrbeu@126.com)

Received 25 September 2020; Revised 28 November 2020; Accepted 20 January 2021; Published 22 February 2021

Academic Editor: Chi-Hua Chen

Copyright © 2021 Jing Liu et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Sensor data analysis is used in many application areas, for example, Artificial Intelligence of Things (AIoT), with the rapid developing of the deep neural network learning that promotes its application area. In this work, we propose the Depth and Width Changeable Deep Kernel Learning-based hyperspectral sensing data analysis algorithm. Compared with the traditional kernel learning-based hyperspectral data classification, the proposed method has its advantages on the hyperspectral data classification. With the deep kernel learning, the feature is mapped through many times mapping and has the more discriminative ability. So, the deep kernel learning has the better performance compared with the multiple kernels learning. And it has the ability to adjust the network architecture for hyperspectral data space, with the optimization equation of the span bound. The experiments are implemented to testified the feasibility and performance of the algorithms on the hyperspectral data analysis, with the classification accuracy of hyperspectral data. The comprehensive analysis of the experiments shows that the proposed algorithm is feasible to hyperspectral sensor data analysis and its promising classification method in many areas data analysis.

## 1. Introduction

With the developing of the deep learning theory, the researchers have proved the deep architecture is feasible and effective in several applications. Deep kernel learning is widely used in image analysis, and the results show it is the effective framework [1, 2], including image annotation [3], classification [4], segmentation [5], and anomaly object detection [6]. The previous works present the deep multiple kernels learning [7] and the novel objective function [8], and in this deep kernel learning framework, the adaptive back propagation is used to update coefficients and weights of the network to improve the ability of representation [9–12].

In the deep kernel framework, the accuracy and speed of visual recognition are effectively improved by using random gradient descent, which provides a research framework for the follow-up deep kernel learning. The researcher proposed a multilayer kernel machine based on the deep confidence network in 2012, which describes the structure from the support vector machine to deep multilayer more completely

[12]. From the perspective of popular learning, Brahma et al. analyzed the principle of improving the performance of kernel learning under the deep neural network from the perspective of popular learning and gave a relatively perfect theoretical framework of deep kernel learning [13]. For the classification problem, Wong et al. proposed a regularized depth Fisher kernel mapping method in 2011. The learning sample space was explicitly mapped to the feature space, and the deep neural network was used to improve the separability of features according to the Fisher criteria [14]. Based on the traditional KPCA, The research results show that FKNN and DLF effectively improve the image classification performance [15]. To solve the clustering problem, the researcher proposed a  $k$ -means algorithm based on RBF kernel based on the nonlinear kernel and multilayer architecture [16], the researchers used the kernel  $k$ -means method to detect the change of the satellite image [17], and realized the pixel representation of a pseudo training set of the nonlinear clustering partition by deep learning areas. For the recursive problem, the researcher proposed a local deep

kernel learning algorithm to accelerate the prediction performance of nonlinear support vector machines and realize the function of optimizing the spatial tree structure [18]. In view of the problem of the deep model selection, Strobl et al. proposed an idea based on multicore deep learning and effectively improved the performance of deep kernel learning through learning multilayer and multiple kernel functions [8]. For the improvement of the deep kernel learning algorithm with different application backgrounds, Li et al. proposed a data recognition algorithm combining CNN and multicore learning for emotional recognition, which effectively improved the recognition performance [19]. On tumor classification of ultrasound images, a research framework based on multicore deep learning was proposed [20]. On the problem of discourse level analysis, Poria proposed a data recognition algorithm based on the multicore model and deep CNN. For the problem of protein-protein interaction in biological information, Yang proposed a classification algorithm based on deep kernel learning [21]. Deep kernel learning has to be applied to many areas, for example, image recognition [33] and pattern prediction [34]. The structure variable deep kernel learning architecture is very feasible, because the width and depth are adjusted. The features are extracted with the different kernel networks, and the performance are different with the different parameters. These parameters are adjusted through the data training. The number of training data is different in the different applications.

In our recent work [35], the novel multiple kernel learning-based deep neural network algorithm work was proposed for the data classification. The work presented the algorithm framework, and it has the advantage on the nonlinear data classification. For the hyperspectral data classification, the algorithm [35] has its ability for the data analysis, but it must be optimized for the hyperspectral data analysis. The hyperspectral data analysis is different from the general data classification in [35]. We have the analysis on the kernel learning with the advantage on the hyperspectral data analysis in the previous work [36], but the algorithm in [36] is the traditional kernel learning framework, with less performance than the deep learning. So, it is effective to improve the performance of kernel learning on the hyperspectral learning. The similar results are testified on the SAR image classification in the recent work [37], only many kernel functions are combined with some combination weights, and the features are mapped with only one time of projection compared with the traditional kernel, but the structure in [37] is unchangeable, and the performance is limited.

So, motivated by the work [35–37], we propose the novel structured changeable deep kernel for hyperspectral sensing data analysis. Different from the work [35], the parameter of structure is optimized with the optimization algorithm, so compared with the algorithm in [35], the algorithm has the ability of adjusting the architecture for the more performance on the hyperspectral data analysis compared with the algorithm in [36, 37]. In the framework, the architecture parameters are solved with the optimization equation using the span bound. The performance on the hyperspectral sensor data is testified and evaluated for hyperspectral applications.

## 2. Proposed Variable Structure-Based Deep Learning Network

*2.1. Framework.* As an excellent method of nonlinear feature extraction and classification, the kernel learning method can be extended to the deep learning network architecture and enhance the ability of multilayer feature extraction. Although the deep learning network and kernel learning algorithm have been combined and the algorithm is presented in the previous work [37], it cannot achieve the adaptive adjustment of the network structure. So, the network is not adaptive to heterogeneous data.

For the hyperspectral data analysis, the hyperspectral data  $D = \{(x_1, y_1), (x_2, y_2), \dots, (x_m, y_m)\}$  where the hyperspectral vector  $x_i \in R^n$  with the labels vector  $y_i \in \{-1, 1\}$ ,  $i = 1, 2, \dots, m$ . With the kernel function  $k(x_i, x_j)$ , the function is calculated with  $k(x_i, x_j) = \phi(x_i)\phi(x_j)$  with the nonlinear mapping features of the input hyperspectral data. The hyperspectral data matrix,  $K$ , is calculated with the following equation. And the classification criterion function is shown as

$$f(x) = \text{sign} \left( \sum_{i=1}^m \alpha_i y_i k(x_i, x) + b \right), \quad (1)$$

where  $\alpha_i$  is the coefficient vector and  $b$  is the bias vector for the classification criterion  $f(x)$ . For the hyperspectral data classification application, the optimization equation is listed as follows:

$$\begin{aligned} \min_{(\alpha, b, \xi)} \frac{1}{2} \|\omega\|^2 + C \sum_{i=1}^m \xi_i, s.t. y_i (\alpha_i k(x_i, x) + b) \\ \geq 1 - \xi_i, \xi_i \geq 0, C > 0, \quad i = 1, 2, \dots, m, \end{aligned} \quad (2)$$

where  $\omega = \sum_{i=1}^m \alpha_i y_i x_i$ , and the parameters  $\xi_i$  and  $C$  are chosen with the crossvalidation method; then, the algorithm can be defined as

$$K(x_i, x) = \sum_{i=1}^M \theta_i k_i(x_i, x), s.t. \theta_i \geq 0, \sum_{i=1}^M \theta_i = 1, \quad (3)$$

where  $M$  is the number of the different kernel functions, and  $\theta_i$  is the parameter of combining the kernel functions. Then, under the multiple kernel functions, the classification decision criterion is transformed as

$$f(x) = \text{sign} \left( \sum_{k=1}^M \theta_k \sum_{i=1}^m \alpha_i y_i k(x_i, x) + b \right). \quad (4)$$

And with this learning model, the combined kernel of each layer in the DWS-MKL algorithm is a linear combination of multiple kernel functions. The general framework is shown in Figure 1.  $M$  is the total number of basis kernel functions, and  $\theta_i$  is the combination coefficient of the combined kernel function. In the previous work [35], we have present the MKL framework.

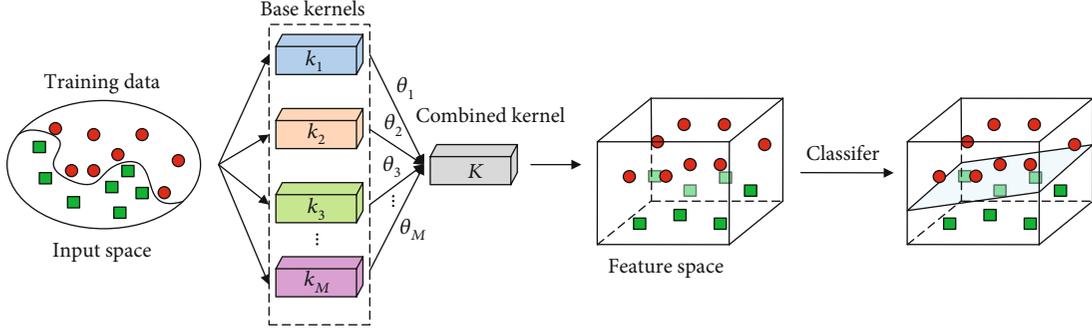


FIGURE 1: The framework of multiple kernel learning.

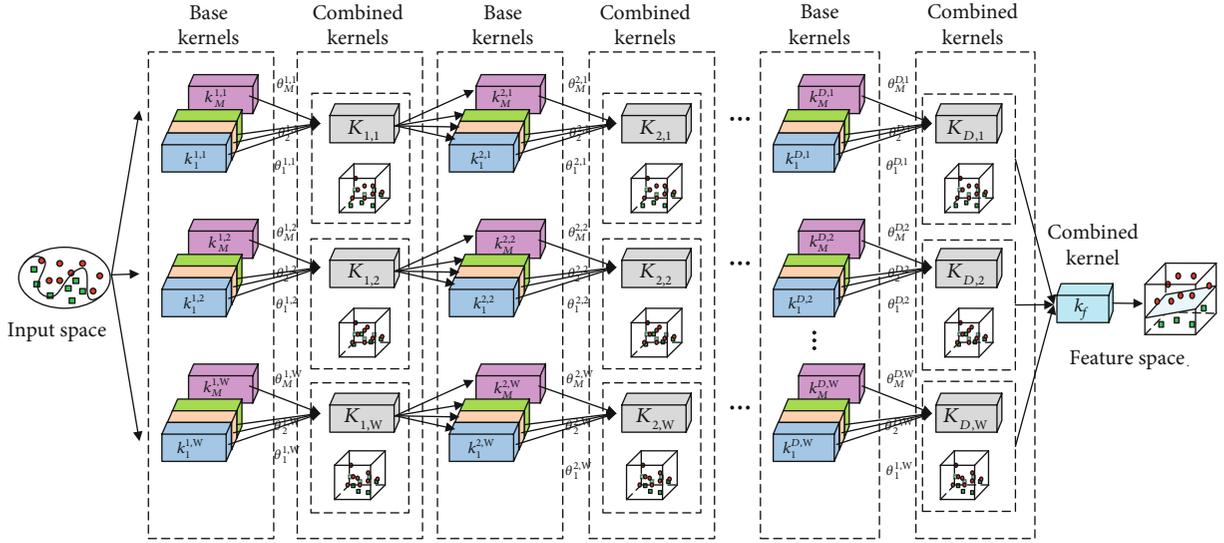


FIGURE 2: The framework of depth-width-scaling multiple kernel learning.

The overall architecture of the DWS-MKL algorithm's variable depth and width combined structure is shown in Figure 2. Since the layers of the DWS-MKL algorithm are directly cascaded, the output of the combined kernel function of the previous layer is the input of the base kernel function of the next layer, and the channels are independent of each other. Therefore, it is easy to adjust the depth and width combination of the algorithm. For the same data set, the classification accuracy of the DWS-MKL algorithm under different depth and width combined architectures is different. For data sets of different scales, if the depth and width combination structure are fixed, the classification accuracy rate under all data sets cannot be maximized. Therefore, it is necessary and reasonable to determine the complexity of the model according to the complexity of the problem to be solved. Adjust the structure of the algorithm flexibly according to the data set, which is the main feature of the DWS-MKL algorithm.  $K_{D,m}$  is the first combined core of the last layer except  $K_f$  for the deep and wide combined structure of the DWS-MKL algorithm.  $\theta_m^{d,w}$  is the combination coefficient corresponding to the base kernel function.

The DWS-MKL algorithm uses SVM as a classifier to solve basic data classification problems. The input of the SVM classifier is the output feature of the combined kernel

$K_f$ . The specific architecture of the classifier is shown in Figure 3. The decision function of the SVM is

$$f(x) = \text{sign} \left( \sum_{i=1}^n \alpha_i y_i (x_i \cdot x) + b \right), \quad (5)$$

Where  $\alpha_i$  is the dual coefficient and  $b$  is the bias of the decision function  $f(x)$ . The optimization problem of SVM is

$$\begin{aligned} \min_{(\alpha, b, \xi)} & \frac{1}{2} \|\omega\|^2 + C \sum_{i=1}^m \xi_i s.t. y_i (\alpha_i k(x_i, x) + b) \\ & \geq 1 - \xi_i, \xi_i \geq 0, C > 0, \quad i = 1, 2, \dots, m, \end{aligned} \quad (6)$$

where  $\omega = \sum_{i=1}^m \alpha_i y_i x_i$ ,  $\xi_i$  is the slack variable, and  $C$  is the regularization coefficient. The basic structure of the DWS-MKL algorithm is a multiple combined kernel, and each combined kernel is a weighted linear combination of multiple basic kernel functions. Therefore, the decision function of the algorithm can be written as

$$f(x) = \text{sign} \left( \sum_{k=1}^M \theta_k \sum_{i=1}^m \alpha_i y_i k(x_i, x) + b \right), \quad (7)$$

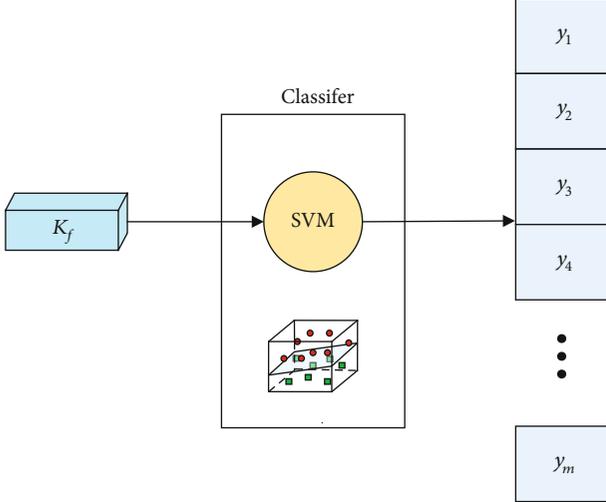


FIGURE 3: Classification network.

TABLE 1: Formulation of 3-order kernel functions.

Kernel function	Formulation
Linear kernel	$k^{(3)}(x, y) = k^{(2)}(x, y) = k(x, y) = x \cdot y$
Polynomial kernel	$k^{(2)}(x, y) = (\alpha(k(x, y) + \beta))^p$ $k^{(3)}(x, y) = (\alpha(k^{(2)}(x, y) + \beta))^p$
Gaussian kernel	$k^{(2)}(x, y) = e^{-(1/\sigma^2)(1-k(x,y))}$ $k^{(3)}(x, y) = e^{-(1/\sigma^2)(1-k^{(2)}(x,y))}$

where  $\theta_k$  is the combination coefficient of the deep and width combination architecture of the DWS-MKL algorithm. Without loss of generality, the decision function of the DWS-MKL algorithm is uniformly written as

$$f(x) = \text{sign} \left( \sum_{i=1}^m \alpha_i y_i k(x_i, x; \theta_i) + b \right), \quad (8)$$

where combination parameters  $\theta$ , dual coefficients  $\alpha$ , and offsets  $b$  are learned through the DWS-MKL algorithm.

**2.2. Algorithm.** Extending the combined kernel cascade to  $L$ -layers, the combined kernel function of the first layer of a deep multiple kernel learning algorithm can be expressed as

$$K^{(L)}(x, y) = \phi^{(L)} \left( \phi^{(L-1)} \left( \dots \phi^{(1)}(x) \right) \right) \cdot \phi^{(L)} \left( \phi^{(L-1)} \left( \dots \phi^{(1)}(y) \right) \right), \quad (9)$$

where  $x$  and  $y$  are the input vectors of the algorithm.  $\phi^{(L)}$  is the nonlinear mapping kernel function. In addition to cascading multilayer combined kernel, the DWS-MKL algorithm extends multiple deep structures to multiple channels, and the channels are independent of each other. Finally, the feature output of each channel is input to a com-

```

Input:  $I = \{(x_i, y_i) \mid i = 1, 2, \dots, m\}, \theta^0, 1/W, D, W, lr, C.$ 
Output: Parameter matrix.
1 Initialize  $k$ , compute  $K$ ;
2 for  $i \leftarrow 1$  to  $N$  do
3   compute the optimal vector with  $K_f$ ;
4   for  $w \leftarrow 1$  to  $W$  do
5     for  $d \leftarrow 1$  to  $D$  do
6       compute  $\tilde{K}_{sv}, B, \partial S_p^2 / \partial \theta; \partial T_{span} / \partial \theta$ 
7       update  $\theta^{i-1} \leftarrow \theta^i + lr \cdot \partial T_{span} / \partial \theta$ ;
8     end
9   end
10 end

```

ALGORITHM 1: Model = DWS-MKL ( $I, \theta^0, D, W, lr, C$ ).

bined kernel  $K_f$  according to the sum and average rule. The combined kernel formula is as follows:

$$K_f = \frac{1}{W} \sum_{m=1}^W K_{D,m}, \quad (10)$$

where  $K_f$  is the combined kernel of the last layer except for the deep and wide combined structure of the DWS-MKL algorithm. The total number of combined kernel in each layer is determined by the number of independent channels of the algorithm. The combined kernel  $K_{D,W}$  of the DWS-MKL algorithm in each layer, and channel is defined as

$$K_{d,w} = \sum_{m=1}^M \theta_m^{d,w} k_m^{d,w}, \quad d = 1, 2, \dots, D, w = 1, 2, \dots, W \quad (11)$$

$$s.t. \sum_{m=1}^M \theta_m^{d,w} = 1, \theta_m^{d,w} > 0 \quad m = 1, 2, \dots, M,$$

where  $k_m^{d,w}$  is the base kernel function of the layer channel, and  $\theta_m^{d,w}$  is the combination coefficient corresponding to the base kernel function.

There are many choices for the basis kernel function of the combined kernel, such as the linear kernel function, polynomial kernel function, RBF kernel function, Laplace kernel function, and sigmoid kernel function. The basic kernel function of the DWS-MKL algorithm can be combined according to specific practical applications. In the example test of the DWS-MKL algorithm of the present invention, three kernel functions are selected as the basis kernel functions of each combined kernel, namely, linear kernel, polynomial kernel, and RBF kernel. The specific formula is shown in Table 1.

**2.3. Procedure.** The depth and width of the algorithm are limited to  $D \in [1, 2, 3], W \in [1, 2, 3]$ . After that, we input the training data for iterative training and use a crossvalidation algorithm to determine the penalty coefficient  $C$  and the optimal depth  $D$  and width  $W$ . After training, we obtain the SVM classification model with classification capability and the combination coefficient of the variable depth and width

TABLE 2: Performance evaluation on the deep kernel learning.

Targets	Training/test samples	2-DNN [37]	3-DNN [37]	Proposed
2S1&BRDM_2	1000/1000	89.0%	92.9%	93.8%
BRDM_2&ZSU234	1000/1000	89.7%	90.4%	91.5%
2S1&BRDM_2	400/400	91.7%	92%	93%
D7&T62	400/400	100%	100%	100%
T62&2S1	400/400	85.0%	85.5%	86.5%

TABLE 3: Detection results on deep kernel learning.

Targets	Training/test samples	2-DNN [37]	3-DNN [37]	Proposed
2S1	1000/1000	80.5%	83.2%	84.5%
BRDM_2	1000/1000	85.6%	85.2%	86.4%
ZSU234	1000/1000	85.0%	89.2%	90.5%
D7	1000/1000	65.3%	67.0%	68.0%
ZIL131	1000/1000	68.5%	72.0%	73.5%

architecture of the algorithm. We represent the training set data as  $I = \{(x_i, y_i) \mid i = 1, 2, \dots, m\}$ .  $lr$  is the learning rate.

### 3. Experiments

*3.1. Algorithm Feasibility Evaluation.* Firstly, we evaluate the feasibility of the deep kernel learning, on the MSTAR data.

In the experiment, the same SAR image target as the deep belief network experiment is used as the verification object, and ten kinds of chariot targets under different angles are selected as the data set. Firstly, the original images in MSTAR data set are preprocessed and cut to the same size. Then, 1000 images are selected as the training set and 1000 images as the test set. In the experiments, in order to test the influence of the number of samples in the data set on the results of feature extraction, a comparative experiment was carried out with 400 training sets and 400 test sets as the data objects under the condition that other conditions remain unchanged.

Different data sets are constructed to verify the effect of feature extraction in the two modes, and the target classification results after feature extraction of depth kernel mapping are compared with other common methods. Firstly, the classification task is used to verify the effect of feature extraction. The ordinary SVM only supports two classification methods. Therefore, this part carries out pairwise combination for different categories of chariot targets, then constructs data sets according to the combination structure, and verifies the effect of deep kernel mapping structure feature extraction in classification under two different data set sizes. The results are shown in Table 2.

Through five groups of two classification tasks, the feature extraction results of depth kernel mapping are verified for different objects and different size data sets. The performance of the same group of network parameters in the classification task of different targets is slightly different, but the overall accuracy rate has been improved with the depth of the structure, and the increase of the amount of data contained in

TABLE 4: Accuracy performance on the different learning methods.

Method	RBF-SVM	DBN	CNN	3-kernel deep [37]	Proposed
Accuracy	83.5%	88.2%	92.3%	95.9%	96.5%

the data set will improve the classification accuracy, but the calculation time will also increase accordingly. Therefore, in the follow-up research, we need to focus on the actual application of the indicators of operation speed and algorithm performance. The network structure is optimized and adjusted. After verifying the classification results, some classification methods are implemented to evaluate the algorithms, and some results are shown in Table 3.

The detection accuracy of five kinds of targets is different, which shows that the differences between each category of targets and other categories are different, and the performance of features extracted by different targets is also different. However, with the increase of structure depth, the detection performance is also improved. It shows that the depth structure can improve the ability of kernel mapping in remote sensing image feature extraction, but for all kinds of specific targets and detection, it is also necessary to optimize the structural parameters to improve its performance. Under the same computing resources and data objects, we compare the classification performance of the common target feature extraction methods in the target classification task. The support vector machine with RBF kernel is selected in the single kernel mapping, the four-layer network structure with good effect is selected for the deep confidence network, and the common AlexNet model is selected for the convolution network. The classification results of each method are shown in Table 4.

By comparing the classification results of feature extraction based on depth kernel mapping and other classification methods, it can be found that the feature extraction effect of deep multicore mapping is the best. Compared with the common single kernel mapping algorithm, the accuracy has been greatly improved, which shows that the depth structure can improve the feature extraction performance of the kernel mapping algorithm. In this paper, we achieve the better performance with the depth structure of the foundation, the existing loss function, and the number and category of the kernel function. So, it is effective for the mapping structure with the parameter optimization algorithm.

Secondly, we have the evaluations on the UCI dataset, to test the performance with different depths and depth widths. 7 public data sets from the UCI benchmark are chosen for the binary and multiple classification tasks. For the comparisons, we have implemented other algorithms, including EasyMKL [22], SimpleMKL [23], SM1MKL [24], and L2MKL [25, 26]. These algorithms with the optimal parameter in the references and 5-fold cross validation method are used to choose the parameter for our algorithm. Some results are shown in Table 5.

From a statistical point of view, there are 8 frameworks that are combined by 3-width in our method regardless of the depth. This work proves that a wider width can improve

TABLE 5: The classification performance by comparing with existing MKL algorithm.

Dataset	EasyMKL <sup>1</sup> [22]	SimpleMKL <sup>2</sup> [23]	SM1MKL <sup>3</sup> [24]	L <sub>2</sub> MKL <sup>4</sup> [25, 26]	Our method (D,W)
Australian	84.63 (3)	84.30 (4)	85.79 (1)	84.05 (5)	85.50 (3,3) (2)
Breast	69.06 (5)	69.78 (4)	70.50 (1)	70.50 (1)	70.50 (1,1) (1)
Cleve	81.08 (4)	80.40 (5)	81.75 (2)	81.75 (2)	84.45 (3,3) (1)
Cloud	69.78 (4)	75.43 (2)	73.29 (3)	61.20 (5)	99.80 (1,3) (1)
Diabetes	74.21 (5)	74.73 (4)	76.56 (2)	76.56 (2)	77.34 (3,3) (1)
Pima	78.38 (1)	78.12 (3)	77.34 (4)	77.34 (4)	78.38 (3,3) (1)
Sonar	86.66 (1)	76.69 (3)	72.38 (5)	74.28 (4)	84.76 (1,3) (2)
Average rank	3.27	2.14	2.57	3.29	1.28

<sup>1</sup>Available at <https://github.com/okbalefthanded/EasyMKL>. <sup>2</sup>Available at <https://github.com/maxis1718/SimpleMKL>. <sup>3</sup>Available at <https://sites.google.com/site/xinxingxu666/>. <sup>4</sup>The implementation is obtained in footnote 4.

TABLE 6: Performance evaluation on different kernel learnings.

Datasets Methods	Indian pines dataset		Pavia University dataset	
	OA (%)	KC (%)	OA (%)	KC (%)
RBF-SKL [27]	73.23	63.73	75.43	68.76
POLY-SKL [27]	75.77	66.28	78.08	72.07
Mahal-RBF-SKL [31]	76.92	67.79	76.26	69.87
Mahal-Poly-SKL [31]	77.64	68.62	79.07	73.24
SK-CV-RBF-SKL [30]	67.03	64.23	75.71	69.14
SK-POLY-SKL [30]	69.37	66.96	77.62	71.37
EasyMKL [22]	68.13	65.53	76.82	72.14
SimpleMKL [23]	69.22	66.78	73.56	66.76
SM1MKL [24]	77.34	69.62	79.98	74.34
L2MKL [25, 26]	77.37	74.85	79.67	75.87
NMF-MKL [32]	67.48	64.81	71.57	64.42
KNMF-MKL [32]	68.22	65.63	72.80	65.81
MKL1 [28]	74.23	69.63	78.69	72.11
MKL2 [29]	76.07	72.84	79.24	72.92
QMKL	79.28	75.85	80.96	75.34
Proposed deep	81.53	78.34	83.14	78.76

the performance of the framework when combined with the appropriate depth. In summary, DWS-MKL outperforms the MKL algorithm and improves the classification performance of the algorithm.

**3.2. Performance on the Hyperspectral Data Classification.** In this paper, we have the experiments on the Indian Pines dataset, Pavia University data, and these datasets are the same to our previous work [36] for the comparisons. The Kappa coefficient (KC) and Overall accuracy (OA) are used to evaluate the algorithms. 10 basis kernels are used to learning, and [0.01, 2] is the scale parameter  $\sigma$ . On the classifier settings, a standard multiclass SVM is used to classification, and the optimal parameters are chosen with the crossvalidation method. The original spectral features are reduced with the PCA algorithm. From the comparisons, the 15 methods also are implemented on two hyperspectral databases, including RBF-SKL [27], POLY-SKL [27], Mahal-RBF-SKL [31], Mahal-Poly-SKL [31], SK-CV -RBF-SKL [30], SK-POLY-

SKL [30], EasyMKL [22], SimpleMKL [23], SM1MKL [24], L2MKL [25, 26], NMF-MKL [32], KNMF-MKL [32], Euclidean-MKL1 [28], Euclidean-MKL2 [29], and QMKL [36]. The detail algorithm description can refer to work [36].

As shown in Table 6, compared with other kernel learning methods, the proposed method achieves the highest recognition rate. The results show that the deep structure can effectively improve the discrimination ability of data. In the comparison of the algorithms, including single kernel learning, multicore learning, and variable structure kernel learning, the variable structure deep kernel learning proposed in this paper achieves the best results, which shows that the algorithm is feasible. The database used in the experiment is actually collected database, which has good practicability. It can be shown that the method proposed in this paper can be used in practical applications.

**3.3. Discussion.** From the above experimental results, we can see that the framework of deep kernel learning is effective and can effectively solve the problem of classification. The results on hyperspectral data sets show that the proposed variable depth kernel learning architecture is also suitable for practical data classification problems and can improve the accuracy of classification. In addition, this method has the characteristics of small amount of calculation and small scale of network parameters, which is especially suitable for application scenarios with limited computing resources. In the paper, three kinds of data have their own characteristics, so the UCI data, radar data, and hyperspectral data have the representative meaning for the practical application. So, in the practical applications, the algorithms should have the corresponding performances to the results. Moreover, the proposed algorithm can be implemented in the limited computing ability of platform, for example, DSP and FPGA. So, it can be used for the many computing practical application areas.

## 4. Conclusion

In this paper, we propose the novel deep kernel learning for the hyperspectral data classification, and the solve method with the gradient projection is used to optimize the error of the classifier. The Depth and Width Changeable Deep Kernel Learning-based hyperspectral sensing data analysis

algorithm is proposed. Some experiments show that the proposed method has its advantages on the hyperspectral data classification. The experiments are implemented to testified the feasibility and performance of the algorithms on the hyperspectral data analysis. In the future research, the computing efficiency should be improved, because it can be used in the video recognition and image processing with small computing ability.

### Data Availability

We have not used specific data from other sources for the simulations of the results. The two popular hyperspectral datasets in this paper, Indian Pines dataset and Pavia University data, are free download with the website: [http://www.ehu.es/ccwintco/index.php?title=Hyperspectral\\_Remote\\_Sensing\\_Scenes](http://www.ehu.es/ccwintco/index.php?title=Hyperspectral_Remote_Sensing_Scenes). The proposed algorithm is implemented in python.

### Conflicts of Interest

The authors declare that they have no conflicts of interest.

### Acknowledgments

We would like to thank Dr. Li Li, Prof. Junbao Li, and Prof. Yanfeng Gu to provide the kernel-based learning programs of their papers for comparison in the experiments. This work is supported by the National Science Foundation of China under Grant No. 61871142.

### References

- [1] J. Ker, L. Wang, J. Rao, and T. Lim, "Deep learning applications in medical image analysis," *IEEE Access*, vol. 6, pp. 9375–9389, 2017.
- [2] G. Wu, M. Kim, Q. Wang, B. C. Munsell, and D. Shen, "Scalable high-performance image registration framework by unsupervised deep feature representations learning," *IEEE Transactions on Biomedical Engineering*, vol. 63, no. 7, pp. 1505–1516, 2016.
- [3] M. Jiu and H. Sahbi, "Nonlinear deep kernel learning for image annotation," *IEEE Transactions on Image Processing*, vol. 26, no. 4, pp. 1820–1832, 2017.
- [4] C. Wang, J. Shi, Q. Zhang, and S. Ying, "Histopathological image classification with bilinear convolutional neural networks," in *2017 39th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pp. 4050–4053, South Korea, 2017.
- [5] W. Li, X. Qian, and J. Ji, "Noise-tolerant deep learning for histopathological image segmentation," in *2017 IEEE International Conference on Image Processing (ICIP)*, pp. 3075–3079, Beijing, China, 2017.
- [6] N. Kumar, A. V. Rajwade, S. Chandran, and S. P. Awate, "Kernel generalized gaussian and robust statistical learning for abnormality detection in medical images," in *2017 IEEE International Conference on Image Processing (ICIP)*, pp. 4157–4161, Beijing, China, 2017.
- [7] J. Zhuang, I. W. Tsang, and S. C. H. Hoi, "Two-layer multiple kernel learning," in *Proceedings of the Fourteenth International Conference on Artificial Intelligence and Statistics*, pp. 909–917, Ft. Lauderdale, FL, USA, April 2011.
- [8] E. V. Strobl and S. Visweswaran, "Deep multiple kernel learning," in *2013 12th International Conference on Machine Learning and Applications*, pp. 414–417, Miami, FL, USA, 2013.
- [9] I. Rebai, Y. BenAyed, and W. Mahdi, "Deep multilayer multiple kernel learning," *Neural Computing and Applications*, vol. 27, no. 8, pp. 2305–2314, 2016.
- [10] I. Rebai, Y. BenAyed, and W. Mahdi, "Deep architecture using multi-kernel learning and multi-classifier methods," in *2015 IEEE/ACS 12th International Conference of Computer Systems and Applications (AICCSA)*, pp. 1–6, Marrakech, Morocco, 2015.
- [11] I. Rebai, Y. BenAyed, and W. Mahdi, "Deep kernel-SVM network," in *2016 International Joint Conference on Neural Networks (IJCNN)*, pp. 1955–1960, Vancouver, BC, Canada, 2016.
- [12] L. Le, J. Hao, Y. Xie, and J. Priestley, "Deep kernel: learning kernel function from data using deep neural network," in *Proceedings of the 3rd IEEE/ACM international conference on big data computing, applications and technologies*, pp. 1–7, New York, NY, USA, 2016.
- [13] P. P. Brahma, D. Wu, and Y. She, "Why deep learning works: a manifold disentanglement perspective," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 27, no. 10, pp. 1997–2008, 2016.
- [14] W. K. Wong and M. Sun, "Deep learning regularized fisher mappings," *IEEE Transactions on Neural Networks*, vol. 22, no. 10, pp. 1668–1675, 2011.
- [15] S. Wang, Y. Jiang, F. L. Chung, and P. Qian, "Feedforward kernel neural networks, generalized least learning machine, and its deep learning with application to image classification," *Applied Soft Computing*, vol. 37, pp. 125–141, 2015.
- [16] K. Ni and R. Prenger, "Learning features in deep architectures with unsupervised kernel k-means," in *2013 IEEE Global Conference on Signal and Information Processing*, pp. 981–984, Austin, TX, USA, 2013.
- [17] V. Harikrishnan and A. Paulose, "A deep learning kernel K-means method for change detection in satellite," *International Journal of Science and Research*, vol. 3, no. 12, pp. 1220–1226, 2014.
- [18] C. Jose, P. Goyal, P. Aggrwal, and M. Varma, "Local deep kernel learning for efficient non-linear SVM prediction," in *Proceedings of the 30th International Conference on Machine Learning*, vol. 28, pp. 486–494, Atlanta, Georgia, USA, June 2013.
- [19] W. Li, F. Abtahi, and Z. Zhu, "A deep feature based multi-kernel learning approach for video emotion recognition," in *Proceedings of the 2015 ACM on International Conference on Multimodal Interaction*, pp. 483–490, New York, NY, USA, 2015.
- [20] X. Liu, J. Shi, and Q. Zhang, "Tumor classification by deep polynomial network and multiple kernel learning on small ultrasound image dataset," in *Proceedings of 6th International Workshop on Machine Learning in Medical Imaging*, vol. 9352, pp. 313–320, Munich, Germany, October 2015.
- [21] S. Poria, E. Cambria, and A. Gelbukh, "Deep convolutional neural network textual features and multiple kernel learning for utterance-level multimodal sentiment analysis," in *Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing*, pp. 17–21, Lisbon, Portugal, 2015.

- [22] F. Aiolli and M. Donini, "EasyMKL: a scalable multiple kernel learning algorithm," *Neurocomputing*, vol. 169, pp. 215–224, 2015.
- [23] A. Rakotomamonjy, F. Bach, S. Canu, and Y. Grandvalet, "SimpleMKL," *Journal of Machine Learning Research*, vol. 9, pp. 2491–2521, 2008.
- [24] X. Xu, I. W. Tsang, and D. Xu, "Soft margin multiple kernel learning," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 24, no. 5, pp. 749–761, 2013.
- [25] C. Cortes, M. Mohri, and A. Rostamizadeh, "L2 regularization for learning kernels," in *Proceedings of the Twenty-fifth Conference on Uncertainty in Artificial intelligence*, pp. 109–116, Montreal Quebec, Canada, June 2009.
- [26] M. Kloft, U. Brefeld, S. Sonnenburg, and A. Zien, "lp-norm multiple kernel learning," *Journal of Machine Learning Research*, vol. 12, pp. 953–997, 2011.
- [27] D. Wang, D. S. Yeung, and E. C. C. Tsang, "Weighted mahalanobis distance kernels for support vector machines," *IEEE Transactions on Neural Networks*, vol. 18, no. 5, pp. 1453–1462, 2007.
- [28] M. Gonen and E. Alpaydin, "Multiple kernel learning algorithms," *Journal of Machine Learning Research*, vol. 12, pp. 2211–2268, 2011.
- [29] Y. Gu, C. Wang, D. You, Y. Zhang, S. Wang, and Y. Zhang, "Representative multiple kernel learning for classification in hyperspectral imagery," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 50, no. 7, pp. 2852–2865, 2012.
- [30] L. Li, C. Sun, L. Lin, J. Li, and S. Jiang, "A dual-layer supervised Mahalanobis kernel for the classification of hyperspectral images," *Neurocomputing*, vol. 214, pp. 430–444, 2016.
- [31] L. Li, C. Sun, L. Lin, J. Li, and S. Jiang, "A Mahalanobis metric learning-based polynomial kernel for classification of hyperspectral images," *Neural Computing and Applications*, vol. 29, no. 4, pp. 1103–1113, 2018.
- [32] Y. Gu, Q. Wang, H. Wang, D. You, and Y. Zhang, "Multiple kernel learning via low-rank nonnegative matrix factorization for classification of hyperspectral imagery," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 8, no. 6, pp. 2739–2751, 2015.
- [33] J. E. Liu and F. P. An, "Image classification algorithm based on deep learning-kernel function," *Scientific Programming*, vol. 2020, Article ID 7607612, 14 pages, 2020.
- [34] T. Wang, L. Zhang, and W. Hu, "Bridging deep and multiple kernel learning: a review," *Information Fusion*, vol. 67, pp. 3–13, 2021.
- [35] T. Wang, H. Su, and J. Li, "DWS-MKL: depth-width-scaling multiple kernel learning for data classification," *Neurocomputing*, vol. 411, pp. 455–467, 2020.
- [36] J. Liu and Y. Qiao, "Quasiconformal mapping kernel machine learning-based intelligent hyperspectral data classification for internet information retrieval," *Wireless Communications and Mobile Computing*, vol. 2020, Article ID 8873366, 14 pages, 2020.
- [37] X. Chen, X. Peng, R. Duan, and J. Li, "Deep kernel learning method for SAR image target recognition," *Review of Scientific Instruments*, vol. 88, no. 10, p. 104706, 2017.