SEMISUPERVISED CLASSIFICATION OF HYPERSPECTRAL IMAGES USING DISCRETE NONLOCAL VARIATION POTTS MODEL

LINYAO GE, BAOXIANG HUANG, WEIBO WEI AND ZHENKUAN PAN*

College of Computer Science and Technology
Qingdao University
Qingdao, 266071, China

(Communicated by Guanglu Zhou)

Abstract. The classification of Hyperspectral Image (HSI) plays an important role in various fields. To achieve more precise multi-target classification in a short time, a method for combining discrete nonlocal theory with traditional variable fraction Potts models is presented in this paper. The nonlocal operator makes better use of the information in a certain region centered on that pixel. Meanwhile, adding the constraint in the model can ensure that every pixel in HSI has only one class. The proposed model has the characteristics of non-convex, nonlinear, and non-smooth so that it is difficult to achieve global optimization results. By introducing a series of auxiliary variables and using the alternating direction method of multipliers, the proposed classification model is transformed into a series of convex subproblems. Finally, we conducted comparison experiments with support vector machine (SVM), K-nearest neighbor (KNN), and convolutional neural network (CNN) on five different dimensional HSI data sets. The numerical results further illustrate that the proposed method is stable and efficient and our algorithm can get more accurate predictions in a shorter time, especially when classifying data sets with more spectral layers.

1. Introduction. Hyperspectral image (HSI) is an especially important image type. Every pixel in the image has a unique spectral structure, which can be used to identify ground objects that cannot be distinguished by the visible spectrum. Meanwhile, the classification task is a particularly significant part of HSI research. People can understand changes of targets and take steps in agriculture, environment [36], and region [8]. But the significant characteristic of the HSI is that the number of the bands is usually more than hundreds, which brings challenges to the classification task. Thus, precise classification is an important aspect of analyzing the features of the ground objects.

For the HSI classification task, semi-supervised methods are more frequently used, which uses both labeled and unlabeled data to fit the model [5]. It can reduce the amounts of human annotations and improve accuracy. Among the many supervised classification methods, semi-supervised K-nearest neighbor (KNN) is
popularly used because it has a simple design and does not require training. Besides, KNN has been successfully used for hyperspectral image classification [18, 2]. It can consider the spectral features of each point as a set of high-dimensional data. However, the KNN algorithm calculates the Euclidean distance from the points to the decision boundary. Accordingly, KNN is a computationally intensive method and plays a substantial role in the HSI classification. Another popular classification method is the semi-supervised support vector machine (SVM) [3], which is not only effective in traditional image classification but also has superior performance in HSI [14, 25]. Kernel trick is the main point of innovation. SVM employs kernel to map the input space to high-dimensional feature space. Therefore, we need to choose kernel functions according to the feature number, such as Radial Basis kernel functions [33] and polynomial kernel functions [39, 6]. Furthermore, a composite kernel structure that combines spatial and spectral information is proposed in [7]. Due to the SVM classifiers need to map the input to a high dimensional space, as the number of HSI bands increases, the time it takes to export the predictions increases as well.

The aforementioned methods belong to machine learning, but another important branch of machine learning is deep learning. Deep learning algorithms have astounding performance in the area of image recognition, especially convolutional neural networks (CNN) [15]. The designing idea of CNNs is to process the data with multiple arrays such as a color image [22]. Whereas, the bands of hyperspectral images can be seen as a form of multiple arrays. Consequently, the convolutional neural networks are also utilized in the assignment of classifying hyperspectral images. 1D-CNNs employed one-dimensional convolution kernels that treat spectral information as a high-dimensional vector [17]. However, we need to consider neighborhood information and the spatial structure of HSI. 3D-CNN methods combined the spatial structure and region feature by dividing the image into some patches [16, 24]. Besides, the 3D-CNNs adopted a three-dimensional kernel, which augments the performance of the network. Besides, methods were proposed to extract both spectral and spatial information simultaneously using three-dimensional wavelets [31, 20], especially Gabor filters [34]. In the process of the CNN method recognition, it needs to set epochs to train the convolution model until the loss of prediction converges. As the scale of the data increase, the model requires more training iterations. Therefore, principal component analysis (PCA) is widely used [10]. Dimension reduction of the image before classification can reduce training rounds and improve performance. In addition to principal component analysis, there are other dimensionality reduction methods such as discrete wavelet transform [21] and independent component analysis [32].

Unlike machine learning and deep learning, variational methods process images by partial differentiation. On the other hand, the variational methods are model-driven algorithms. Merriman et al. [27] presented Merriman-Bence-Osher (MBO) method, which incorporated the diffuse interface model proposed by Bertozzi and Flenner [1] for high-order data classification. This method is not only applied to multi-classification problems [11, 19] but also has outstanding results in hyperspectral image classification [26]. To make better use of information about the region where the pixel is centered, Gilboa and Osher proposed a framework of a nonlocal image processing system [12]. Furthermore, Zhu et al. [38] proposed an unsupervised nonlocal total variation method.
According to the problem of the above methods, we proposed a new semi-supervised discrete nonlocal variation Potts (NLVP) model to classify the hyperspectral image. Firstly, in our proposed model, we adopt the nonlocal operator to compute the similarity of patches in the search window. Using the nonlocal operator, the classification of one point can utilize a broader range of region information. Therefore, the proposed method can have a higher classification accuracy of hyperspectral images. Secondly, the proposed method in this paper classifies unknown test sets with the training set and initializes $\phi_0$ with the training set, thus, it is semi-supervised. Semi-supervised can make sure output one-to-one correspondence to ground feature class so that it can help analysts distinguish the category of each part of the output and perform corrected analysis. Thirdly, the Potts model [23, 30] is developed based on the classical model Mumford-Shah [28] of variational image segmentation and is a convex relaxation model with a binary label function. Compared with the classic Potts model, the proposed model only contains nonlocal variation terms $\|\nabla w \phi\|_{L_1}$ and label constraints $\sum_{i=1}^{k} \phi(x)_i = 1$. The model satisfying the label constraint ensures that the final classification results are free of omissions and double-division. Finally, we employ an alternating direction method of multipliers (ADMM) [29] algorithm to transform the model into an alternating-optimized structure for solving variables in this model.

This paper is organized as follows. The discrete nonlocal variation Potts model and ADMM algorithm are presented in Section II. Section III presents the numerical results. Section IV presents the conclusions.

2. Discrete nonlocal Potts model and ADMM algorithm. The nonlocal method is an extension for image processing techniques based on the variational method and the partial differential equations (PDEs), it addresses the staircase effect to some extent. Compared with the traditional total variational (TV) method [4], the nonlocal framework has better results in image processing, including the RGB image, texture image, and so on. Using Gaussian kernels and Euclidean Metric to calculate similarities between points can be highly productive in image segmentation tasks. In the nonlocal operation, the nonlocal derivative is

$$\frac{\partial \phi(x)}{\partial x} = \frac{\phi(y) - \phi(x)}{d(x, y)} \quad x, y \in \Omega$$

$\Omega$ is the domain of the pixels. $\phi = (\phi_1, \phi_2, \ldots, \phi_n)$ is the label function, $n$ is the number of the categories. The $d(x, y)$ is the distance between the pixels $x$ and $y$. The $d(x, y)$ can is used to measure the similarity between the two pixels. Many distance formulas can be chosen, for example, Euclidean distance and Cosine distance.

By utilizing the label function, the classification problem can be formulated as the following minimization problem.

$$E(\phi) = \gamma \sum_{i=1}^{n} \|\nabla w \phi_i\|_{L_1}$$

To avoid duplication and omission of classes, we can add constraints to (2) and transformed the penalty function into

$$E(\phi) = \gamma \sum_{i=1}^{n} \|\nabla w \phi_i\|_{L_1} + \frac{\mu_0}{2} \left(\sum_{i=1}^{n} \phi_i - 1\right)^2$$

(3)
\( \nabla_w \phi \) is the nonlocal gradient, it can be defined as the collection of all partial derivatives.

\[
\nabla_w \phi(x, y) = \frac{\partial \phi(x)}{\partial x} = \sqrt{w(x, y)}(\phi(y) - \phi(x))
\]

(4)

The nonlocal divergence \( \text{div}_w \phi \) is defined as

\[
\text{div}_w \phi(x) = \int_{\Omega} \sqrt{w(x, y)}\phi(x, y) - \sqrt{w(y, x)}\phi(y, x)dy
\]

(5)

\( w(x, y) \) is the weight matrix. Whether calculating nonlocal gradient or nonlocal divergence, the weight matrix is an important factor.

2.1. Weight matrix. The weight matrix is used to describe the degree of similarity between pixels on the graph. Thus, we need to calculate the distance to construct a weight matrix. In nonlocal methods, we use a patch to compute the generalized distance. As shown in Fig. 1, \( x, y \) are the center points of the patches and the size of the patch is less than the search window. Usually, the patch size is chosen as \( 3 \times 3 \) for the classification task and the radius of the search window is twice as big as the corresponding patch.

![Figure 1. Patch and Search window](image)

The distance between patches is as following.

\[
d^2(x, y) = \int_{\Omega} G_\sigma \ast (f(x + t) - f(y + t))^2dt
\]

(6)

The \( G_\sigma \) is the Gaussian of standard deviation \( \sigma \). \( t \) is the radius of the patch, such as Fig. 1 \( t = 2 \). Meanwhile, the weight between \( x \) and \( y \) can be obtain by the following formula.

\[
w(x, y) = e^{-\frac{\int_{\Omega} G_\sigma \ast (f(x + t) - f(y + t))^2dt}{\pi\sigma^2}}
\]

(7)

\( h \) is to make sure that the projection of the value of the weight between 0 and 1. The weight matrix in our approach is of critical importance, and we will discuss how to reach the numerical approximation of formula (2) using the ADMM algorithm with the nonlocal gradient and nonlocal divergence.
2.2. Application of ADMM to discrete nonlocal Potts model. The alternating direction method of multipliers (ADMM) algorithm is an important algorithm to solve the convex optimization problem with a separable structure [13]. It applies mostly to machine learning and is an extension of the ALM (Augmented Lagrangian method) algorithm. The ADMM algorithm for the discrete nonlocal Potts model is derived from the variable splitting, Lagrange multiplier, and penalty function methods. We employ a vector \( \vec{w} = [\vec{w}_1, \vec{w}_2, \ldots, \vec{w}_n]^T \) instead of the \( \nabla w \phi \) in (3). Therefore, the (3) can be transformed into the following alternate optimization form,

\[
(w_i^{k+1}, \phi_i^{k+1}) = \arg \min_{\phi, \vec{w}} E(\phi, \vec{w}) = \gamma \sum_{i=1}^{n} \sum_{x \in V} |\vec{w}_i^k(x)| + \sum_{i=1}^{n} \sum_{x \in V} |\vec{w}_i^k - \nabla w \phi_i^k|^2
+ \sum_{x \in V} \lambda_i^k \sum_{j=1}^{n} (\phi_j - 1) + \frac{\mu_0}{2} \sum_{x \in V} \sum_{j=1}^{n} (\phi_j - 1)^2
\]  

(8)

where \( \lambda \) is the Lagrange multiplier, \( \mu \) is the penalty parameter, and \( \lambda_i^k, \lambda_0^k \) are the \( \lambda_i^k, \lambda_0^k \) at step \( k \) or initial value at step \( 0 \), respectively. First of all, we need to initialize the \( \vec{w}_i^0, \phi_0^0 \), \( \lambda_0^0 \) at step \( k = 0 \). Then, we can solve for the value at step \( k + 1 \) by fixing the value at step \( k \) until convergence. This process reflects the primary thinking of the alternating direction method of multipliers. Taking advantage of the ADMM algorithm, therefore, we can divide the iterative process from step \( k \) to step \( k + 1 \) in optimization problem (8) into the following four suboptimization problems,

\[
\phi_i^{k+1} = \arg \min_{\phi_i} \{ E(\phi_i^k, w_i^k) \}
\]  

(9)

\[
w_i^{k+1} = \arg \min_{\vec{w}_i} \{ E(\phi_i^{k+1}, w_i^k) \}
\]  

(10)

\[
\lambda_i^{k+1} = \lambda_i^k + \mu (w_i^{k+1} - \nabla w \phi_i^{k+1})
\]  

(11)

\[
\lambda_0^{k+1} = \lambda_0^k + \mu_0 \left( \sum_{j=1}^{n} \phi_j^{k+1} - 1 \right)
\]  

(12)

Finally, we use the Euler-Lagrange equation of (8) to get the \( \phi_i^{k+1} \). The Euler-Lagrange can be described as follows,

\[
-\mu \nabla w \cdot (\nabla w \phi_i - w_i^k - \frac{\lambda_i^k}{\mu}) + \mu_0 \sum_{j=1}^{n} (\phi_j - 1) + \frac{\lambda_i^k}{\mu_0} = 0
\]  

(13)

2.2.1. Computing the primal variable \( \phi_i \). According to the nonlocal gradient and the nonlocal divergence, \( \phi_i^{k+1} \) can be computed. We take the gradient and the
divergence to (8), the transform equations as follows,

\[ F_i = -\mu \nabla w \cdot (-w^k_i - \frac{\lambda^k_i}{\mu}) \]  

\[ S = \mu_0 \left( \sum_{j=1}^{n} \phi_j - 1 + \frac{\lambda^0_i}{\mu_0} \right) \]  

\[ \phi^{k+1}_i = \mu \sum_{i=1}^{n} \sum_{x \in V} \phi_i(y)w(x, y) + F_i + S \]  

sometimes \( \mu \sum_{i=1}^{n} \sum_{x \in V} w(x, y) = 0 \). Thus, we should add an extremely small values like \( 10^{-6} \).

2.2.2. The minimization with respect to \( \vec{w}_i \). The soft threshold formula proposed in [9] is typically utilized to evaluate the \( \vec{w}_i \). This algorithm is widely used find the minimum value of a variable and is one of the most classical algorithms. \( \phi_i(x)^{k+1} \) is corrected by (16) and we calculate the corrected value of \( \phi_i(x)^{k+1} \) and \( \lambda^k \). The Euler-Lagrange equation of (10) is given.

\[ \gamma \frac{\vec{w}_i}{|\vec{w}_i|} + \mu (\vec{w}_i - \nabla w \phi^{k+1}_i + \frac{\lambda^k_i}{\mu}) = 0 \]  

Moreover, \( \vec{w}_i \neq 0 \). The solution to this equation can be expressed analytically by the generalized soft threshold formula.

\[ \vec{w}_i^{k+1} = \max \left( \left| \frac{\nabla w \phi^{k+1}_i}{\mu} - \frac{\lambda^k_i}{\mu} \right|, 0 \right) \]  

2.2.3. Update of Lagrange multipliers. After calculating \( \vec{w}_i^{k+1} \), the Lagrange multipliers \( \lambda^k \) and \( \lambda_0 \) need to be updated. It can actually be calculated by the following formula.

\[ \lambda^{k+1}_i = \lambda^k_i + \mu (\vec{w}_i^{k+1} - \nabla w \phi^{k+1}_i) \]  

\[ \lambda^{k+1}_0 = \lambda_0^k + \mu_0 \left( \sum_{i=1}^{n} \phi^{k+1}_i - 1 \right) \]  

After the update of Lagrange multipliers, we need to threshold the \( \phi^{k+1}_i \) to \( \phi_{ihard} \).

2.2.4. Output \( \phi_i \) is projected to \( \phi_{ihard} \). \( \phi_i \) is a label function that describes a class of every point, thus, the label function is either 1 or 0. To avoid being greater than 1 and less than 0, we use \( \phi^{k+1}_i = \max(\min(\phi^{k+1}_i, 1), 0) \) fix the value between 0 and 1. Then, we adopt the following threshold function binarize \( \phi_i \).

\[ \phi^{k+1}_i = \begin{cases} 
1 & \text{if } \phi^{k+1}_i > 0.5 \\
0 & \text{if } \phi^{k+1}_i < 0.5 
\end{cases} \]  

The proposed NLVP method is demonstrated in pseudocode format in Algorithm 1. After explaining the proposed algorithm, we demonstrate the validity of the proposed method by conducting experiments on five datasets.
Algorithm 1 Discrete nonlocal Potts model

Initialization: $\gamma$, $\mu$, $\mu_0$

$\phi^0_i = \text{Training data for class } i$

$w^0, \lambda^0, \lambda^0_0 \leftarrow 0$

Output: $\sum_{i=1}^{n} \phi_i$

1: while energy not converged do
2:   $\phi^{k+1} \leftarrow \phi^{k}, w^{k}, \lambda^k, \lambda^k_0$
3:   $w^{k+1} \leftarrow \phi^{k+1}, \lambda^k, \lambda^k_0$
4:   $\lambda^{k+1} \leftarrow \phi^{k+1}, w^{k+1}, \lambda^k$
5:   $\lambda^{k+1}_0 \leftarrow \phi^{k+1}, \lambda^k$
6:   Thresholding $\phi_i$
7: end while

3. Numerical results. We compared with other semi-supervised classification methods include support vector machines (SVM), K-Nearest Neighbor (KNN), and Convolutional Neural Network (CNN). All experiments were run on the windows machine with Intel Core i5, 2.80GHz with 16GB of DDR3 RAM. The GPU is a GeForce GTX 1060. In each dataset experiment, all the algorithms used the same training set, which accounted for 10 percent of the dataset. For the NLVP method, we initial $\phi_0$ using training set.

SVM is realized through the “Libsvm” toolbox on MATLAB, and the kernel function is Polynomial kernel. In the KNN method, setting K to 8 has high accuracy in the experiment.

The CNN method in the experiment adopts the 3D-CNN with a double convolution pooling structure in [24]. The double convolution structure includes the 3D convolution, Batch Normalization, and ReLU activation functions. After the double convolution structure, the maximum pooling is set for the downsampling, and the output results are input into the full connection. Use Dropout after full connection to avoid overfitting, and finally use Softmax classifier for classification.

The convolution kernel is set as $3 \times 3 \times 3$, and the stride of the first convolution is set as 5 to reduce the dimension of the image. The maxpooling size is set to $2 \times 2 \times 2$, and the input pixel block is $11 \times 11 \times$ spectral dimension. The learning rate is 0.003. CNN is carried out on GPU, and the ultimate accuracy and running time are the average of ten times.

SVM, KNN, NLVP, and CNN carried out experiments on MATLAB R2020a.

3.1. Hyperspectral image datasets.

3.1.1. XiongAn New Area (Matiwan village) data set. The spectral range of aerial hyperspectral remote sensing image is 400-1000nm, with 250 bands. The image size is $3750 \times 1580$ pixels, and the spatial resolution is 0.5m. Through field research on the category of ground objects, the image marked 19 categories of ground objects, mainly cash crops such as rice, peach tree, and lotus leaf [35].

3.1.2. Fanglu Tea Farm data set. This data set was obtained by aerial flight at the tea tree planting base in fanglu village, Changzhou city, Jiangsu province, China. The hyperspectral data were obtained by the airborne Pushbroom Hyperspectral Imager, which contained $348 \times 512$ pixels with a spatial resolution of 2.25 m and 80 spectral bands. The data set includes ten typical classes consists of masson pine,
Table 1. Comparison of numerical results in the Fanglu and Indian data sets

| Algorithm | Fanglu | Indian |
|-----------|--------|--------|
|           | Run-Time(s) | Accuracy(%) | Run-Time(s) | Accuracy(%) |
| SVM       | 5.625 | 93.626 | 2.318s | 79.002 |
| KNN       | 121.741 | 93.015 | 11.001s | 66.387 |
| CNN       | 11.002(GPU) | 99.196 | 4.756s(GPU) | 84.590 |
| NLVP      | **16.809** | **99.961** | **37.449** | **99.618** |

Table 2. Comparison of numerical results in the Salinas and Pavia University data sets

| Algorithm | Salinas | Pavia University |
|-----------|---------|------------------|
|           | Run-Time(s) | Accuracy(%) | Run-Time(s) | Accuracy(%) |
| SVM       | 17.938 | 93.003 | 6.275 | 93.098 |
| KNN       | 283.259 | 89.227 | 90.298 | 86.076 |
| CNN       | 25.804(GPU) | 98.029 | 12.984(GPU) | 99.578 |
| NLVP      | **31.175** | **98.889** | **20.835** | **99.783** |

bamboo forest, tea tree, reeds, rice, sweet potato, coriander, weeds, pond, buildings, roads [37].

3.1.3. Indian Pines data set. The scene of this dataset was collected in northwest Indian by AVIRIS. After removing bands containing the region of water absorption, the image contained 145 × 145 pixels with 200 bands. Two-thirds of this data set were agriculture, one-thirds were forest or other natural plants, and the ground truth was marked 16 classes.

3.1.4. Salinas scene data set. This scene was covered comprises 512 lines by 217 samples and collected by the 224-band AVIRIS sensor over Salinas Valley. It includes vegetables, bare soils, and vineyard fields. The data set ground truth contains 16 classes.

3.1.5. Pavia University scene data set. This data set was acquired by the ROSIS sensor and the number of spectral is 103 for Pavia University. It includes water, trees, and asphalt. The image ground truths differentiate 9 classes.

3.2. Fanglu Tea Farm data set. The Fanglu data set was collected in a tea garden in Jiangsu province. Most of the land in the data set were crops, and there were only 80 bands in the data set. With fewer bands, we can see from Table 1 that each method has a high accuracy rate. In the accuracy rate, the NLVP method is close to the result of the CNN method and higher than SVM and KNN approaches. Thus, the method proposed in this paper works well on datasets at lower spectral bands.

---

1Available at: [http://www.ehu.eus/ccwintco/index.php?title=Hyperspectral_Remote_Sensing_Scenes](http://www.ehu.eus/ccwintco/index.php?title=Hyperspectral_Remote_Sensing_Scenes).
3.3. **Indian Pines data set.** The classification results are shown in Fig. 3. In the graph of the results of the SVM and KNN methods, we can visually see that there are many misclassification points in each category. But in general, the results of these visualizations are confusing. While the method proposed in this paper is smoother in each classification region and better enables analysts to perform ground feature analysis.

The size of the Indian dataset is $145 \times 145$ and the number of bands is up to 200. However, the overall amount of data is small so that the computation time is relatively less. From Table 1, the method proposed in this paper requires more time to output results than SVM and KNN.

Although we had to sacrifice some time to obtain the results, the classification accuracy achieved more than 99% and is significantly higher than other algorithms.

3.4. **Salinas scene data set and Pavia University data set.** The Salinas scene dataset is relatively large, which consists of $512 \times 217$ and collects 224 bands. From Table 2, when above several data sets managed to know that all methods consume more time on this dataset. During the experiment, obtaining a classification map
(a) Ground Truth  (b) SVM  (c) KNN  (d) CNN  (e) NLVP

Figure 3. Classification results of Indian data set. The picture of the left is the ground truth, and the remaining four images are the classification result of the SVM, KNN, CNN, and NLVP.

Table 3. The accuracy of all algorithms in each category of the Salinas Scene data set

| Label | Samples | SVM(%) | KNN(%) | CNN(%) | NLVP(%) |
|-------|---------|--------|--------|--------|---------|
| 1     | 2009    | 97.412 | 94.873 | 99.104 | 99.801  |
| 2     | 3726    | 98.631 | 98.148 | 99.973 | 99.866  |
| 3     | 1976    | 98.330 | 97.217 | 99.241 | 99.798  |
| 4     | 1394    | 98.350 | 98.494 | 100    | 99.785  |
| 5     | 2678    | 97.498 | 95.183 | 99.664 | 98.394  |
| 6     | 3959    | 98.232 | 98.055 | 99.949 | 99.571  |
| 7     | 3579    | 98.044 | 97.346 | 99.721 | 99.749  |
| 8     | 11271   | 89.726 | 84.624 | 95.333 | 97.746  |
| 9     | 6203    | 98.436 | 97.469 | 99.919 | 99.532  |
| 10    | 3278    | 94.356 | 87.340 | 99.847 | 96.827  |
| 11    | 1068    | 96.536 | 88.390 | 98.315 | 97.659  |
| 12    | 1927    | 97.146 | 96.886 | 100    | 100     |
| 13    | 916     | 97.489 | 96.397 | 100    | 98.690  |
| 14    | 1070    | 95.234 | 86.729 | 98.972 | 93.738  |
| 15    | 7268    | 64.364 | 54.100 | 96.863 | 98.005  |
| 16    | 1807    | 97.620 | 95.849 | 99.779 | 97.731  |
Figure 4. Classification results of Salinas Scene data set. The picture of the left is the ground truth, and the remaining four images are the classification result of the SVM, KNN, CNN, and NLVP.

using KNN consumed 200 sec. Nevertheless, the proposed method has a running time of around 31 sec merely 6 sec more than the Indian dataset. Therefore, when the scale of the image is bigness, the proposed method in this paper also has comparatively stable performance.

Table 4 statistics the accuracy of the four methods for each category in the Salinas dataset. We can identify that SVM and KNN have lower classification precision in some categories. For example, both the SVM and KNN methods are less than 70% accurate for label 15, while both CNN and the method presented in this paper achieve more than 90% accuracy.

The Pavia University Data Set is different from the above several data sets. For example, the “Trees” class is distributed on both sides of the “Asphalt” class, which is more scattered. Therefore, the NLVP method requires less time during iteration than other data sets. Although NLVP takes three times longer to run than SVM, the precision is improved by 99%.

For this dataset, we also performed a statistical analysis of the classification precision of each Label. Furthermore, it is demonstrated that the method proposed in this paper has greater accuracy in the classification of each Label. For instance, the
second and ninth labels were both given perfect classification with 100%. Meanwhile, the classification of other tags also got better results, achieving 99%.

![Classification results of Pavia University data set.](image)

**Figure 5.** Classification results of Pavia University data set. The picture of the left is the ground truth, and the remaining four images are the classification result of the SVM, KNN, CNN, and NLVP.

3.5. **XiongAn data set.** The size of this data set is $3750 \times 1580$ pixels and has 250 bands. Therefore, a portion of the source dataset is selected for experimentation. The selected experimental data size is $459 \times 591$ pixels, with 250 bands.

In the experiments, the SVM and KNN methods have very long run times without dimensionality reduction. Table 4 shows that the SVM method runs for more than 1000 seconds on this dataset and KNN is much longer than SVM. To shorten the runtime a bit, PCA is used to dimensionality reduction of the data in the SVM and KNN methods. The running time of the proposed method in this paper is similar to that of CNN on the GPU. As a result, the proposed method does not substantially
Table 4. The accuracy of all algorithms in each category of the Pavia University Scene data set

| Label | Samples | SVM(%) | KNN(%) | CNN(%) | NLVP(%) |
|-------|---------|--------|--------|--------|---------|
| 1     | 6631    | 96.139 | 87.830 | 97.994 | 99.774  |
| 2     | 18649   | 96.960 | 98.665 | 99.882 | 100     |
| 3     | 2099    | 74.845 | 64.078 | 94.140 | 99.809  |
| 4     | 3064    | 93.603 | 77.742 | 99.641 | 98.172  |
| 5     | 1345    | 99.628 | 99.257 | 100    | 99.777  |
| 6     | 5029    | 85.942 | 47.345 | 99.920 | 99.980  |
| 7     | 1330    | 80.000 | 82.556 | 91.429 | 99.925  |
| 8     | 3682    | 88.403 | 84.465 | 98.588 | 99.891  |
| 9     | 947     | 100    | 99.789 | 99.578 | 100     |

Figure 6. Classification results of XiongAn data set. The picture of the left is the ground truth, and the remaining four images are the classification result of the SVM, KNN, CNN, and NLVP.

Table 5. Comparison of numerical results on the XiongAn data set

| Algorithm | Run-Time(s) | Accuracy(%) |
|-----------|-------------|-------------|
| SVM       | 1571.159    | 86.361      |
| PCA-SVM   | 75.926      | 80.188      |
| PCA-KNN   | 515.325     | 80.114      |
| CNN       | 135.598(GPU)| 93.037      |
| NLVP      | 135.789     | 99.507      |
increase the running time as the number of spectral bands increases and the image size expands.

For data downscaling, SVM and KNN have a lower accuracy of 80 percent. The CNN and NLVP methods both achieve 90 percent, and the NLVP method achieves 99%, which is higher than CNN. It can be seen that the method proposed in this paper still has a more stable operational efficiency and classification accuracy on a larger scale of hyperspectral images.

4. Conclusion remarks. The present research aimed to improve the precision of the classification for hyperspectral images using a semi-supervised method. A highly accurate and smooth graph of classification results is significant for relevant analysts to analyze the detail of agriculture and the environment, etc. The discrete nonlocal variation Potts model proposed in this paper contributes to helping the analysts obtain a better classification figure in a short time peculiarly larger-scale data. Furthermore, the model was tested on the five HSI data sets and compared with SVM, KNN, and CNN.

We summarize the following points based on numerical experiments. Firstly, the discrete nonlocal variation Potts model is simple and easy to understand since it only includes nonlocal variation and penalties. Moreover, this semi-supervised method is a model-driven algorithm so that it does not require a training process. Secondly, the method proposed in this paper performs better on widely used data sets such as the Indian data set and the Pavia University data set. Meanwhile, the output image of the classification is smoother and can clearly distinguish the classified regions. Finally, as the size of the HSI increase, the time consumed by the proposed method is within reasonable limits and does not increase substantially.

A limitation of this study is that the numerical experiment adapts to the five HSI data sets. Every HSI has its own unique data characteristics so that the experiment should be conducted on a larger number of hyperspectral images. In numerical experiments, the method can be found inadequate in running efficiency. Thus, further research should focus on improving the performance of the model, and propose a model more suitable for the hyperspectral image classification task.

5. Acknowledgments. This work was supported in part by the National Natural Science Foundation of China under Grant 61602269, in part by the Natural Science Foundation of Shandong Province, China, under Grant ZR2019LZH002, in part by the China Postdoctoral Science Foundation under Grant 2015M571993.

REFERENCES

[1] A. L. Bertozzi and A. Flenner, Diffuse interface models on graphs for classification of high dimensional data, Multiscale Model. Simul., 10 (2012), 1090–1118.
[2] C. Bo, H. Lu and D. Wang, Spectral-spatial K-Nearest Neighbor approach for hyperspectral image classification, Multimedia Tools Appl., 77 (2018), 10419–10436.
[3] B. E. Boser, I. M. Guyon and V. N. Vapnik, A training algorithm for optimal margin classifier, Proceedings of the Fifth Annual Workshop on Computational Learning Theory, 5 (1992), 144–152.
[4] A. Buades, B. Coll and J. M. Morel, A review of image denoising algorithms, with a new one, Multiscale Model. Simul., 4 (2015), 490–530.
[5] Y. Cai, X. F. Zhu and Z. Sun, Semi-supervised and ensemble learning: A review, Comput. Sci., 44 (2017), 7–13.
[6] G. Camps-Valls and L. Bruzzone, Kernel-based methods for hyperspectral image classification, IEEE Transactions on Geoscience and Remote Sensing, 43 (2005), 1351–1362.
[7] G. Camps-Valls, L. Gomez-Chova and J. Munoz-Mari, et al., Composite kernels for hyperspectral image classification, *IEEE Geoscience and Remote Sensing Letters*, 3 (2006), 93–97.

[8] C.-I. Chang, *Hyperspectral Imaging: Techniques for Spectral Detection and Classification*, Springer, 2003.

[9] Z. Dou, B. Zhang and X. Yu, A new alternating minimization algorithm for image segmentation, 6th International Conference on Wireless, Mobile and Multi-Media (ICWMMN), 2015.

[10] M. D. Farrell and R. M. Mersereau, On the impact of PCA dimension reduction for hyperspectral detection of difficult targets, *Geoscience and Remote Sensing Letters*, 2 (2005), 192–195.

[11] C. Garcia-Cardona, E. Merkurjev and A. L. Bertozzi, et al., Multiclass data segmentation using diffuse interface methods on graphs, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 36 (2014), 1600–1613.

[12] G. Gilboa and S. Osher, Nonlocal Operators with Applications to Image Processing, *Multiscale Model. Simul.* 7 (2008), 1005–1028.

[13] T. Goldstein, B. O’Donoghue, S. Setzer and R. Baraniuk, Fast alternating direction optimization methods, *SIAM J. Imaging Sci.*, 7 (2014), 1588–1623.

[14] J. A. Gualtieri and R. F. Cromp, Support vector machines for hyperspectral remote sensing classification, *Proc. SPIE*, 3584 (1999).

[15] X. Hao, G. Zhang and S. Ma, Deep learning, *International J. Semantic Computing*, 10 (2016), 417–439.

[16] M. He, B. Li and H. Chen, Deep multi-scale 3D deep convolutional neural network for hyperspectral image classification, *2017 IEEE International Conference on Image Processing (ICIP)*, (2017), 3904–3908.

[17] W. Hu, Y. Huang, L. Wei, F. Zhang and H. Li, Deep convolutional neural networks for hyperspectral image classification, *J. Sensors*, 2015 (2015), 1–12.

[18] K. Huang, S. Li, X. Kang and L. Fang, Spectral-spatial hyperspectral image classification based on KNN, *Sensing and Imaging*, 17 (2016).

[19] G. Huo, S. X. Yang, Q. Li and Y. Zhou, A robust and fast method for sidescan sonar image segmentation using nonlocal despeckling and active contour model, *IEEE Transactions on Cybernetics*, 47 (2017), 855–872.

[20] S. Jia, L. Shen and Q. Li, Gabor feature-based collaborative representation for hyperspectral imagery classification, *IEEE Transactions on Geoscience and Remote Sensing*, 53 (2015), 1118–1129.

[21] S. Kaewpijit, J. Le Moigne and T. El-Ghazawi, Automatic reduction of hyperspectral imagery using wavelet spectral analysis, *IEEE Transactions on Geoscience and Remote Sensing*, 41 (2003), 863–871.

[22] Y. LeCun, Y. Bengio and G. Hinton, Deep learning, *Nature*, 521 (2015), 436–444.

[23] F. Li, M. K. Ng, T. Y. Zeng and C. Shen, A multiphase image segmentation method based on fuzzy region competition, *SIAM J. Imaging Sci.*, 3 (2010), 277–299.

[24] G. Li, C. Zhang, F. Gao and X. Zhang, Doubleconvpool-structured 3D-CNN for hyperspectral remote sensing image classification, *J. Image and Graphics*, 24 (2019), 639–654.

[25] F. Melgani and L. Bruzzone, Classification of hyperspectral remote sensing images with support vector machines, *IEEE Transactions on Geoscience and Remote Sensing*, 42 (2004), 1778–1790.

[26] E. Merkurjev, J. Sunu and A. L. Bertozzi, Graph MBO method for multiclass segmentation of hyperspectral stand-off detection video, *2014 IEEE International Conference on Image Processing (ICIP)*, (2014), 689–693.

[27] B. Merriman, J. K. Bence and S. J. Osher, Motion of multiple junctions: A level set approach, *J. Comput. Phys.*, 112 (1994), 334–363.

[28] D. Mumford and J. Shah, Optimal approximation by piecewise smooth function and associated variational problems *Comm. Pure Appl. Math.*, 42 (1989), 577–685.

[29] M. Myllykoski R. Glowinski, T. Karhunen and T. Rossi, A new augmented Lagrangian approach for $L^1$-mean curvature image denoising, *SIAM J. Imaging Sci.*, 8 (2015), 95–125.

[30] R. B. Potts, Some generalized order-disorder transformations, in *Mathematical Proceedings of the Cambridge Philosophical Society*, 48, Cambridge Philosophical Society, 1952, 106–109.

[31] S. Jia, L. Shen and Q. Li, Gabor feature-based collaborative representation for hyperspectral imagery classification, *IEEE Transactions on Geoscience and Remote Sensing*, 49 (2011), 5039–5046.
[32] J. Wang and C.-I. Chang, Independent component analysis-based dimensionality reduction with applications in hyperspectral image analysis, *IEEE Transactions on Geoscience and Remote Sensing*, **44** (2006), 1586–1600.

[33] P. Wang and X. Y. Zhu, Model selection of SVM with RBF kernel and its application, *Comput Engrg. Appl.*, (2003), 72–73.

[34] Y. Wang and L. Wang, Local Gabor convolutional neural network for hyperspectral image classification, *Comput. Sci.*, **47** (2020), 151–156.

[35] C. Yi, L. F. Zhang and X. Zhang, et al., Aerial hyperspectral remote sensing classification dataset of Xiongan New Area (Matiwan Village), *J. Remote Sensing*, (2019).

[36] L. Zhang, H. Sun, Z. Rao and H. Ji, Hyperspectral imaging technology combined with deep forest model to identify frost-damaged rice seeds, *Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy*, **229** (2020).

[37] X. Zhang, B. Zhang, L. Zhang and Y. Sun, Hyperspectral remote sensing dataset for tea farm, *Global Change Data Repository*, (2017).

[38] W. Zhu, V. Chayes and A. Tiard, et al., Unsupervised classification in hyperspectral imagery with nonlocal total variation and primal-dual hybrid gradient algorithm, *IEEE Transactions on Geoscience and Remote Sensing*, **55** (2017), 2786–2798.

[39] S. Zhuo, X. S. Guo and J. Wan, et al., Fast classification algorithm for polynomial kernel support vector machines, *Comput. Engrg.*, **33** (2007).

Received October 2020; 1st revision January 2021; 2nd revision February 2021.

E-mail address: linyaoge@foxmail.com
E-mail address: hbx3726@163.com
E-mail address: njustwwb@163.com
E-mail address: zkpan@126.com