HYPER-VOXEL BASED DEEP LEARNING FOR HYPERSPECTRAL IMAGE CLASSIFICATION

Atif Mughees and Linmi Tao*

Department of Computer Science,Tsinghua University, Beijing, China

ABSTRACT

Classification of distinct classes in hyperspectral images (HSI) is one of the most pervasive problem in remote sensing field. Deep learning has recently proved its efficiency in HSI classification. However, incorporating spatial/contextual features along with spectral information in deep network is still a challenging task. In this paper, for an effective spectral-spatial feature extraction, an improved deep network, spatial updated hyper-voxel stacked auto-encoder (HVSAE) approach is proposed which exploits spatial context within spectrally similar contiguous pixels for effective HSI classification. The proposed approach involves two key steps-firstly, we compute adaptive boundary adjustment based segmentation whose size and shape can be adapted according to the spatial structures and which consists of spatially contiguous pixels with similar spectral features, followed by an object-level classification using stacked auto-encoder (SAE) based decision fusion approach that merges spatial-segmented outcome and spectral information into a SAE framework for robust spectral-spatial HSI classification. In addition, instead of directly using a large number of spectral bands, band preference and correlation based band selection approach is used to select the most informative bands without compromising the original content in HSI. Use of local spatial structural regularity and spectral similarity information from adaptive boundary adjustment based process, and fusion of spatial context and spectral features into SAE has significant effect on the accuracy of the final HSI classification. Experimental results on real divergent hyperspectral imagery with different contexts and resolutions validates the classification accuracy of the proposed method over several existing techniques.

Index Terms— image classification, hyperpsectral image, stacked auto-encoder, segmentation.

1. INTRODUCTION

The advancements in remote sensing technology has enabled the sensors to acquire hyperspectral images (HSI) with detailed spectral and spatial information of the scene. The incredible amount of spectral and spatial information if successfully exploited, can yield higher classification accuracies and more detailed class taxonomies [1]. Moreover, due to advances in hyperspectral technology, the rich spatial resolution of recently operated sensors makes it possible to analyze the small spatial structures in the images [2].

HSI comprises of hundreds of spectral bands. The increase in spectral and spatial information poses new challenges in classification. In general, dimensionality reduction techniques, such as Principle Component Analysis (PCA) and Independent Component Analysis (ICA) are applied for feature extraction and to reduce the dimensions for HSI classification, which results in loss of detailed information inevitably and hence effect the classification results. Moreover, the local structural information may also be lost during feature extraction process using these traditional feature extraction methods.

In general, integrating spatial information along with spectral channels in HSI classification process is of paramount importance [3, 4] as with the advancement in imaging technology, hyperspectral sensors can deliver excellent spatial resolution. Recently deep learning based machine learning architecture is employed for HSI classification which has shown promising results [5, 6]. Deep learning based algorithms with more layers attempts to extract more abstract and invariant features. In recent years, some deep models have been proposed for HSI processing [6, 7, 2]. Deep models require sufficient amount of training data to train different layers, which is a major limitation of HSI's. Moreover, these methods are unable to incorporate the spatial information into deep models.

In this paper, we propose a hyper-voxel based auto-encoder model that efficiently exploits the spatial contextual features of the HSI where pixel by pixel information is replaced by the local contextual voxel information. Weighted hyper-voxel segmentation has twofold effect, first, it retains the spectral correlations in spectral dimensions and match the actual structure in spatial dimensions, secondly, it implicitly selects the best informative bands and removes noisy and redundant bands without compromising the original contents. Proposed method takes full advantage of the available detailed spectral information in the presence of limited training samples. Each hyper-segmented region [8] in HSI can be considered as a local spatial region whose size and shape can be adaptively adjusted for different spatial structures. The hyper-voxel based stacked auto-encoder (HVSAE) first employs an effi-

*Corresponding author. E-mail: maoz14@mails.tsinghua.edu.cn

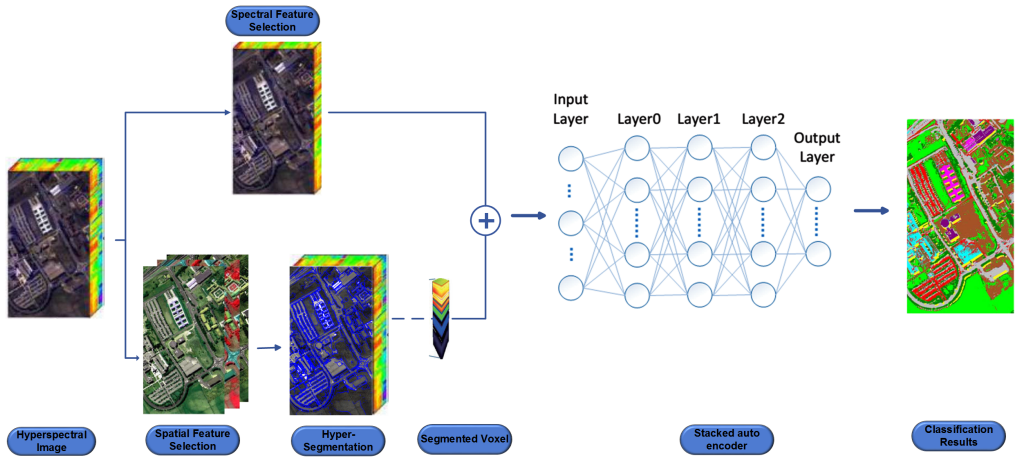


Fig. 1: Framework of the HVSAE based classification.

cient hyper-voxel based segmentation approach [8] to divide the HSI into many spatially connected regions. Then, pixels in each hyper-segmented structures are assumed to have very similar spectral characteristics, and their correlations are exploited via a stacked auto-encoder. Proposed method exploit multi-layer stacked auto-encoder (SAE) to learn shallow and deep features of hyperspectral data. The rest of the paper is formulated as follows. The proposed methodology is presented in section 2. Experimental analysis and data sets are explained in section 3 followed by conclusion in section 4.

2. HV BASED SAE FOR HSI CLASSIFICATION

Feature extraction process should take into account two major facts 1) There is a high probability that data with similar spectral signatures share the same class 2) There is also a high probability that neighboring data with correspondence in spectral signatures should also share the same class. By keeping these facts in mind, we have improved the SAE classification process to extract discriminative spectral-spatial features. Auto-encoder considers every data sample individually without any correlation with neighboring samples and only pays attention on minimizing the error between the input and the reconstructed output. In order to control the feature learning structure to follow the above mentioned facts, we improve SAE to keep the correlation between sampled data by providing SAE contextual/spatial information along with the selected spectral features. That results into an improved classification performance.

Hyper-segmentation is an adaptive boundary adjustment based model [8] which results into a spatial region whose shape and size can be flexibly attuned according to different spatial structures present in the HSI. Proposed HVSAE algorithm extends this model for HSI classification and adopts the SAE to effectively exploit spectral–spatial information within and among each segmented region. In general, the proposed HVSAE algorithm mainly involves three parts: 1) Feature selection 2) creation of spatially adaptive segments in HSI and 3) exploration of spectral–spatial information of these segmented regions via SAE.

2.1. Spatial Feature Extraction by Hyper-segmentation

The hyper voxel is extracted by exploiting efficient hyper-segmentation approach [8] that extract spatial voxels through the energy function:

$$A(p, q_i) = \sqrt{|x_p - g_i|^2 + \lambda \tilde{n}_i(p)|Grad(p)|} \qquad (1)$$

Tri-factor weight model is applied to extract the structural information which is evaluated based on the movement of edges till the actual structural boundary is encountered [8]. To reduce the computational cost, before the segmentation, an efficient feature selection approach [9] is applied on the original HSI to obtain best features that should contain the most important variational information for the whole image, it is used as the base image for the hyper-segmentation. The segmentation process is illustrated in Fig. 1.

2.2. Extraction of Spectral-Spatial Information of Spatial Segments via SAE

Auto-encoder (AE) is a basic deep learning architecture model based on a symmetrical neural network in which input signal is reconstructed at the output layer by going through an intermediate layer. It can be used to learn the deep and abstract features of the data in an unsupervised manner [10]. In general, the input layer of the auto-encoder maps the input x to a hidden representation through an encoder function and the hidden layer can be considered as a new feature representation. The hidden layer is then utilized to reconstruct an estimate of the input through a decoder function as shown in Fig. 2. Non linear sigmoid function is the most commonly

Algorithm 1: spectral-spatial Segmentation based SAE classification

Input : Hyperspectral image $\mathcal{I}$, having pixels p with intensity vectors x_p

Output: HSI classification

To improve accuracy, noisy/redundant bands are removed based on [9].;

To get the spatial contextual feature, adaptive boundary adjustment based hyper-segmentation [8] is applied on to the selected bands and R structural regions are obtained.

for *each spatial dominated region* R **do**

  Calculate the number of pixels M in each region R;

  Flatten the array into a feature vector of size M.

end

Scale M into unit interval. ;

Normalize the whole initial image onto unit interval.;

for *each pixel* **do**

  Add spectrum of each pixel on tail of each pixel's feature vector i.e, rows in M,

end

Train Stacked auto-encoder with M.

used function for encoder and decoder. Encoder and decoder functions can be mathematically represented as

$$y = f(w_y x + b_y), z = f(w_z y + b_z) \tag{2}$$

where $\mathbf{y}$ is obtained from input x by weights w_y and bias b_y. To train the AE, the error between x and z must be minimized by determining the optimized values of parameters, i.e.

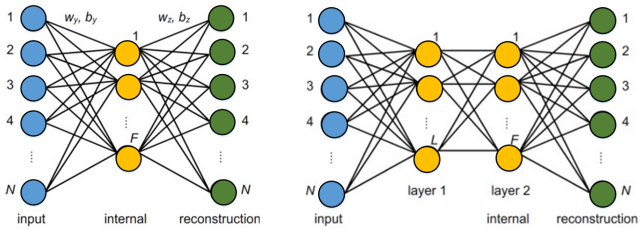
$$\underset{w_y, w_z . b_y, b_z}{\arg\min} \; [error(x, z)] \tag{3}$$

SAE is an expansion of this process where AEs are stacked together. SAE has emerged as one of the powerful abstract feature extractor model [7]. The cross entropy based cost function [7] in accordance with the sigmoid activation function for the SAE is given as

$$c = -\frac{1}{m} \sum_{i=1}^{m} \sum_{k=1}^{d} [x_{ik} \log(z_{ik}) + (1 - x_{ik}) \log(1 - z_{ik})] \tag{4}$$

where d and m represent the vector size and mini-batch size respectively. $x_{ik}(z_{ik})$ denotes the kth element of the ith input. Mini-batch stochastic gradient descent is used to optimize this equation [7]. Let (X, Y) represents the training set, where x^n is the corresponding label for y^n and N is the number of samples. The complete process can be briefed as follows:

1. Each training sample x^n, is encoded by the encoder $z^n = f(x^n)$, where $f(.)$ is the encoding function.



(a) Basic Auto-encoder (b) Stacked auto-encoder

Fig. 2: General SAE model. It learns a hidden feature from input.

2. Decode z^n to the reconstruction by $h^n = g(z^n)$, where $g(.)$ is the decoding function.

3. Optimize the encoding and decoding functions parameters by reducing the error between the reconstructions and the inputs over the whole training set.

In order to supervise the learned spectral-spatial features from SAE hidden layers into different classes, we add multinomial logistic regression (MLR) as an output layer.

3. EXPERIMENTAL RESULTS AND PERFORMANCE COMPARISONS

In order to validate the effectiveness, proposed method was evaluated on two real hyperspectral datasets, Pavia University and Indian Pine captured with different sensors.

3.1. Data Set Description

Indian Pine data set was acquired by AVIRIS sensor in 1992. It consists of 220 spectral channels each with the dimension of 145×145 and contain 16 classes. Pavia University dataset was captured by ROSIS sensor over the area of Pavia University. It consists of 610×340 pixels with 103 spectral bands after removing the water absorption bands. It comprises of 9 different classes. We randomly selected 10% samples for training and remaining 90% for testing in both the datasets as shown in Table 1 and Table 2.

3.2. Parameter Setting

We conducted our experiment on windows 7 system, on 4.0 GHz processor with NVIDIA GeForce GTX 970. The code was implemented in Theano. For both the datasets, we used 180 units in the first hidden layer and 100 units in the second hidden layer, as experimentally shown in [7] . According to the research's observation, number of hidden units is more imperative than the number of hidden layers. The proposed method HVSAE is compared with well known existing methods such as stacked auto-encoder with

Table 1: Classification accuracy of each class for the Indian Pine dataset and comparison with existing approaches.

Class	Training	Test	LORSAL-MLL	SAE-LR	HVSAE
1	5	49	**100**	93.33	94.11
2	143	1291	87.46	84.66	**89.42**
3	83	751	81.23	84.39	**88.70**
4	23	211	88.41	73.08	**89.05**
5	50	447	**97.49**	93.47	95.54
6	75	672	**97.34**	93.41	96.31
7	3	23	**100**	**100**	**100**
8	49	440	**97.98**	95.11	96.89
9	2	18	**100**	**100**	**100**
10	97	871	83.64	85.78	**90.73**
11	247	2221	86.31	83.46	**89.11**
12	61	553	**91.79**	81.62	87.27
13	21	191	**100**	98.52	98.87
14	129	1165	**97.04**	91.77	94.45
15	38	342	86.18	81.79	**90.95**
16	10	85	**100**	98.88	99.05
	Overall Accuracy		87.18	86.85	**90.08**
	Average Accuracy		90.83	89.95	**93.09**
	Kappa Coefficient		0.8536	0.8495	**0.8875**

logistic regression(SAE-LR) [7] and augmented lagrangian-multilevel logistic (LORSAL-MLL) [11].

3.3. Spectral-Spatial Classification Results

For Indian Pine dataset we used first six components of PCA and a window size of 7×7 for SAE-LR. Classification results of Indian Pine dataset and its comparison is shown in Table 1 and Fig. 3. Mixed pixel is the major challenge in this dataset due to its low spatial resolution and small size. Results approve that spectral-spatial classification using contextual feature extraction has significant effect on the classification accuracy because spatial features help prevent the salt and paper noise.

For Pavia University dataset, we used first six components of PCA and a window size of 9×9 for SAE-LR. Classification results and its comparison is shown in Table 2 and Fig. 3.

Overall, experimental results demonstrates the significant improvement in HSI classification by incorporating spatial information and spectral feature selection. The algorithm has performed significantly well on the low spatial resolution dataset.

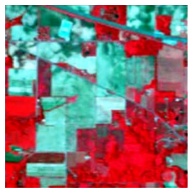
4. CONCLUSION

This paper proposes a new HVSAE approach to exploit spatial contextual features via hyper-voxel segmentation for efficient HSI classification. The dimension of each hyper-voxel can be flexibly modified according to the HSI spatial structures, which results into an improved exploitation of spatial information. HVSAE then utilizes the multi-layer SAE to efficiently exploit the HV based spatial features and selected spectral features within and among segmented hyper-voxels. Experimental performance shows that proposed technique produces improved results than other present techniques in HSI classification, particularly in HSI with minor spatial structures. Moreover, HVSAE as one of the efficient feature extractor, performs well in diverse images, specially for complex urban scenes.

One of our future research track is to develop a more efficient approach for selecting the number of hyper-voxels based on diverse spatial contextual information. Furthermore, HVSAE approach can effectively be applied to other HSI applications such as surveillance, noise detection, and object identification and classification.

Table 2: Classification accuracy of each class for the Pavia University dataset and comparison with existing approaches.

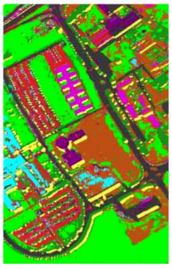
Class	Training	Test	LORSAL-MLL	SAE-LR	HVSAE
1	597	6034	**100**	96.88	98.78
2	1681	16971	98.50	98.30	**99.01**
3	189	1910	**96.27**	91.09	94.50
4	276	2788	91.03	99.15	**99.40**
5	121	1224	98.45	99.85	**99.89**
6	453	4576	97.24	**96.44**	97.30
7	120	1210	**97.94**	94.12	95.96
8	331	3351	**99.94**	93.27	95.14
9	85	862	**100**	**100**	**100**
	Overall Accuracy		98.95	97.12	**98.98**
	Average Accuracy		97.28	96.56	**97.46**
	Kappa Coefficient		0.9819	0.9615	**0.9875**



(a) Indian Pine

(b) Classification Result

(c) Pavia University



(b) Classification Result

Fig. 3: classification Results of Pavia and Indian Pine

5. REFERENCES

[1] Suju Rajan, Joydeep Ghosh, and Melba M Crawford, "An active learning approach to hyperspectral data classification," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 46, no. 4, pp. 1231–1242, 2008.

[2] Yushi Chen, Hanlu Jiang, Chunyang Li, Xiuping Jia, and Pedram 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.

[3] Antonio Plaza, Javier Plaza, and Gabriel Martin, "Incorporation of spatial constraints into spectral mixture analysis of remotely sensed hyperspectral data," in *2009 IEEE International Workshop on Machine Learning for Signal Processing*. IEEE, 2009, pp. 1–6.

[4] Yuntao Qian and Minchao Ye, "Hyperspectral imagery restoration using nonlocal spectral-spatial structured sparse representation with noise estimation," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 6, no. 2, pp. 499–515, 2013.

[5] Konstantinos Makantasis, Konstantinos Karantzalos, Anastasios Doulamis, and Nikolaos Doulamis, "Deep supervised learning for hyperspectral data classification through convolutional neural networks," in *Geoscience and Remote Sensing Symposium (IGARSS), 2015 IEEE International*. IEEE, 2015, pp. 4959–4962.

[6] Luis Gómez-Chova, Devis Tuia, Gabriele Moser, and Gustau Camps-Valls, "Multimodal classification of remote sensing images: a review and future directions," *Proceedings of the IEEE*, vol. 103, no. 9, pp. 1560–1584, 2015.

[7] Yushi Chen, Zhouhan Lin, Xing Zhao, Gang Wang, and Yanfeng Gu, "Deep learning-based classification of hyperspectral data," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 7, no. 6, pp. 2094–2107, 2014.

[8] Atif Mughees, Xiaoqi Chen, and Linmi Tao, "Unsupervised hyperspectral image segmentation: Merging spectral and spatial information in boundary adjustment," in *Society of Instrument and Control Engineers of Japan (SICE), 2016 55th Annual Conference of the*. IEEE, 2016, pp. 1466–1471.

[9] Atif Mughees, Xiaoqi Chen, Rucheng Du, and Linmi Tao, "Ab3c: adaptive boundary-based band-categorization of hyperspectral images," *Journal of Applied Remote Sensing*, vol. 10, no. 4, pp. 046009–046009, 2016.

[10] Geoffrey E Hinton and Ruslan R Salakhutdinov, "Reducing the dimensionality of data with neural networks," *Science*, vol. 313, no. 5786, pp. 504–507, 2006.

[11] Jun Li, José M Bioucas-Dias, and Antonio Plaza, "Hyperspectral image segmentation using a new bayesian approach with active learning," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 49, no. 10, pp. 3947–3960, 2011.