Deep neural networks-based relevant latent representation learning for hyperspectral image classification

Akrem Sellami, Salvatore Tabbone

Deep neural networks-based relevant latent representation learning for hyperspectral image classification

Akrem Sellami[a], Salvatore Tabbone[a]

[a] *Universit de Lorraine, LORIA, CNRS, UMR 7503, Nancy Grand Est, Vandoeuvre-ls-Nancy, F54506, France*

Abstract

The classification of hyperspectral image is a challenging task due to the high dimensional space, with large number of spectral bands, and low number of labeled training samples. To overcome these challenges, we propose a novel methodology for hyperspectral image classification based on multi-view deep neural networks which fuses both spectral and spatial features by using only a small number of labeled samples. Firstly, we process the initial hyperspectral image in order to extract a set of spectral and spatial features. Each spectral vector is the spectral signature of each pixel of the image. The spatial features are extracted using a simple deep autoencoder, which seeks to reduce the high dimensionality of data taking into account the neighborhood region for each pixel. Secondly, we propose a multi-view deep autoencoder model which allows fusing the spectral and spatial features extracted from the hyperspectral image into a joint latent representation space. Finally, a semi-supervised graph convolutional network is trained based on thee fused latent representation space to perform the hyperspectral image classification. The main advantage of the proposed approach is to allow the automatic extraction of relevant information while preserving the spatial and spectral features of data, and improve the classification of hyperspectral images even when the number of labeled samples

URL: akrem.sellami@inria.fr (Akrem Sellami), salvatore.tabbone@univ-lorraine.fr (Salvatore Tabbone)

is low. Experiments are conducted on three real hyperspectral images respectively Indian Pines, Salinas, and Pavia University datasets. Results show that the proposed approach is competitive in classification performances compared to state-of-the-art.
Keywords: Deep learning, representation learning, hyperspectral image classification, feature extraction

1. Introduction

Hyperspectral imagery (HSI) can contain a large number of spectral bands, which provide a rich information on Earth's surface in both spectral and spatial domains. Therefore, HSI can measure radiance values of different ground objects, and is widely used in several fields such as defence, mineralogy, or agriculture [1, 2]. Each pixel $\mathbf{x}_i$ in an HSI $\mathbf{X}$ is a 1-D vector with hundreds of spectral values corresponding to various spectral bands. Due to the high dimensionality of HSI, especially, the large number of pixels and spectral bands, HSI classification is proving to be very challenging [3, 4, 5]. Moreover, when there is few labeled training samples, HSI misclassification does often occur. The large number of spectral bands and the low number of training samples lead to the problem of the curse of dimensionality, which can significantly harm the performance in terms of classification accuracy [6, 7, 8].

To address these challenges, dimensionality reduction (DR) is applied as a preprocessing phase before the spectral-spatial classification. It aims to reduce the number of spectral bands, and obtaining a better classification accuracy preserving the discrimination capability of the spectro-spatial features. DR approaches can be decomposed into two main categories that are feature extraction and band selection. Feature extraction aims to project the whole HSI into a very low dimensional subspace, whereas feature selection selects a subset of relevant spectral bands, i.e., by discarding irrelevant and redundant ones.

Feature extraction can be categorized into linear and non-linear techniques. Linear feature extraction approaches can include principal component analy-

sis (PCA) [9], independent component analysis (ICA) [10], or linear discriminant analysis (LDA) [5, 11]. Non-linear feature extraction methods seek to obtain non-linear feature spaces like laplacian eigenmaps (LE) [12], kernel PCA (KPCA) [13], or locality preserving projection (LPP) [14].

Band selection methods aim to select a subset of relevant spectral bands across the initial HSI by using a specific criteria, including, entropy, variance, and distance between labeled classes and spectral bands, etc. Usually, band selection techniques can be categorized into three groups: supervised, unsupervised, and semi-supervised methods [15, 16]. Supervised band selection is based on a searching algorithm associated with an optimization criteria, including class separability measures and information theoretic [17, 18]. It aims to find informative bands using class labels as a priori information. However, in the HSI field, there is often very few a priori information on desired ground objects. Unsupervised band selection methods aim to find discriminative and distinctive spectral bands, with no a priori knowledge or training labeled samples [19, 20]. Semi-supervised band selection methods aim to find a subset of discriminative and informative spectral bands using unlabeled and labeled training samples. Most techniques are based on graph clustering [21] or manifold learning [22].

To summarize, DR is an important step to overcome all issues related to the high dimensionality of HSI. Moreover, in recent years, spatial information has been growing more and more important for spectral-spatial classification of HSI. In this context, we propose a novel methodology allowing to reduce the dimensionality of data by preserving the spectral and spatial features in order to improve the classification of HSI based on multi-view deep representation learning with only a small number of labeled samples.

The remainder of the paper is organized as follows. In Section II, we present some works related to the spectral-spatial classification of HSI. In Section III, we detail the proposed methodology called Multi-View Deep Neural Networks (MV-DNNet) which includes the multi-view deep autoencoder (MVDAE) that allows fusing spectral- and spatial-features, and the semi-supervised graph convolutional network (SSGCN) model. In Section IV, we first describe the HSI

dataset; then we detail our experimental protocol. Quantitative evaluations demonstrate the good performances of the MV-DNNet for classification tasks and we conclude in Section V.

2. Related Works

Recently, various deep learning-based have been developed for the spectral-spatial classification of HSI [23, 24, 25, 26]. Liang et al. [27] proposed a deep multi-scale feature fusion model for spectral-spatial classification of HSI. Zhong et al.[28] developed an end-to-end spectro-spatial residual network (SSRN), where the model learns discriminative features from spatial contexts and abundant spectral signatures. An 3-D CNN framework is designed to preserve spectro-spatial deep features, which are discriminative features. In [29], a deep architecture based on deep belief network (DBN) has been proposed to combine the spectral-spatial feature extraction and classification together improving the classification accuracy. Their framework is based on PCA, hierarchical learning-based feature extraction, and logistic regression (LR). Furthermore, a method based on fully convolutional neural network (F-CNN) for HSI classification has been proposed in [30]. Specifically, some works have been proposed to extract jointly spectral and spatial features from HSI to perform the classification [31]. A common approach consists to extract for each pixel a neighboring region and flatten it into a $1 \times D$ vector. Then, the spatial vector and the spectrum vector are concatenated and fed into deep learning models [32, 29, 33]. In [34], Zhou et al. proposed a compact and discriminative stacked autoencoder model for HSI classification HSI, which can learn discriminative low-dimensional feature mappings and train an effective classifier progressively. Cheng et al. [35] proposed a method to extract hierarchical deep spatial feature for HSI classification by exploring the power of off-the-shelf CNN models. In [36], Cheng et al. proposed a method to learn discriminative CNNs to boost the performance of remote sensing image scene classification. Wan et al. proposed a multiscale dynamic GCN which employs dynamic graphs to encode the intrinsic similar-

ities among regions to improve the HSI classification [37]. Moreover, in [38] a CNN-enhanced GCN has been proposed to consider pixel- and superpixel-level features for HSI classification . Other methods seeks to extract spectral-spatial features by averaging all spectral vectors within a spatial region. Then, the averaged spectral vector is feed into a deep neural network [39, 40]. Moreover, instead of directly extracting the spatial feature within a neighboring region, some filtering methods, including, Gabor filtering [41] or attribute profiles [42], were proposed to process the original HSI data seeking to extract more relevant spatial features. In [23], the authors proposed a fused 3-D CNN for spectral-spatial classification of HSI, which seeks to fuse multiple 3-D CNN applied on a set of groups of similar spectral bands. However, the combination of multiple supervised 3-D CNN is very expensive in time and computations.

Based on previous works, we can notice that deep learning models have shown their high performance in enhancing HSI classification by extracting effective spectral and spatial features. However, there are several issues, especially for CNN, requiring a large number of labeled samples for training and classification. However, it is generally very difficult to obtain enough training labeled samples for HSIs. In addition, most of spectral-spatial feature extraction methods aim to concatenate or average the spectrum vector with the neighboring region. However, some features are not useful for classification and may be noisy. In this regard, we propose an unsupervised multi-view deep autoencoder (MVDAE) model to fuse both spatial and spectral features into joint latent representation in order to improve the classification of HSI. The aim of the proposed MVDAE is to extract only useful features by discarding the noise and finding a shared latent representation, which can be effective for the classification. Moreover, we propose to develop a semi-supervised graph convolutional neural network (SSGCN) in order to consider local vertex features and graph topology in the convolutional layers preserving the spectral-spatial features in the classification of the HSI and using a limited set of labeled training samples.

3. Proposed methodology

This section details the proposed methodology MV-DNNet, which is composed of three phases as shown in Figure 1. The first phase consists in extracting the spectral and spatial features based on a simple deep autoencoder (AE), which seeks to automatically extract relevant features while preserving the spatial property of HSI. In the second phase, we develop a multi-view deep autoencoder (MVDAE) to combine both views, i.e., spectral and spatial features. Then, we construct the graph for multi-view latent representation. It seeks to take into account the spatial features by considering distances between neighboring pixels. Afterwards, we propose a semi-supervised graph convolutional network (SSGCN) which integrates graph topology and local vertex features in the convolutional layers, in order to improve the HSI classification by preserving the spectro-spatial features. The main advantage of the proposed methodology is to allow the automatic extraction of relevant spectral and spatial features, and improve the HSI classification by using a few number of labeled samples.

3.1. Spectral and spatial feature construction

In this section, we propose two parallel modules to extract and build a set of spectral and spatial features. The aim is to automatically fuse these extracted features with a multi-view representation learning model in order to improve the classification of HSIs. The obtained combined latent representation contains spectral and spatial features that can be useful for classification.

3.1.1. Spectral feature $\mathbf{X}^{spe}$

In spectral feature extraction, we consider the raw data, i.e., all pure spectral features of the HSI $\mathbf{X}$. Usually, a spectral signature is represented as a one dimensional spectrum vector $(1 \times D)$ for each pixel, where D is the number of spectral bands. Hence, in order to exploit rich spectral information and leverage limited prior knowledge, we take into consideration the responses of all spectral bands as input. We obtain then a spectral matrix $\mathbf{X}^{spe} \in \mathbb{R}^{N \times D}$, where N is the number of pixels, and D is the number of spectral bands. Therefore, each

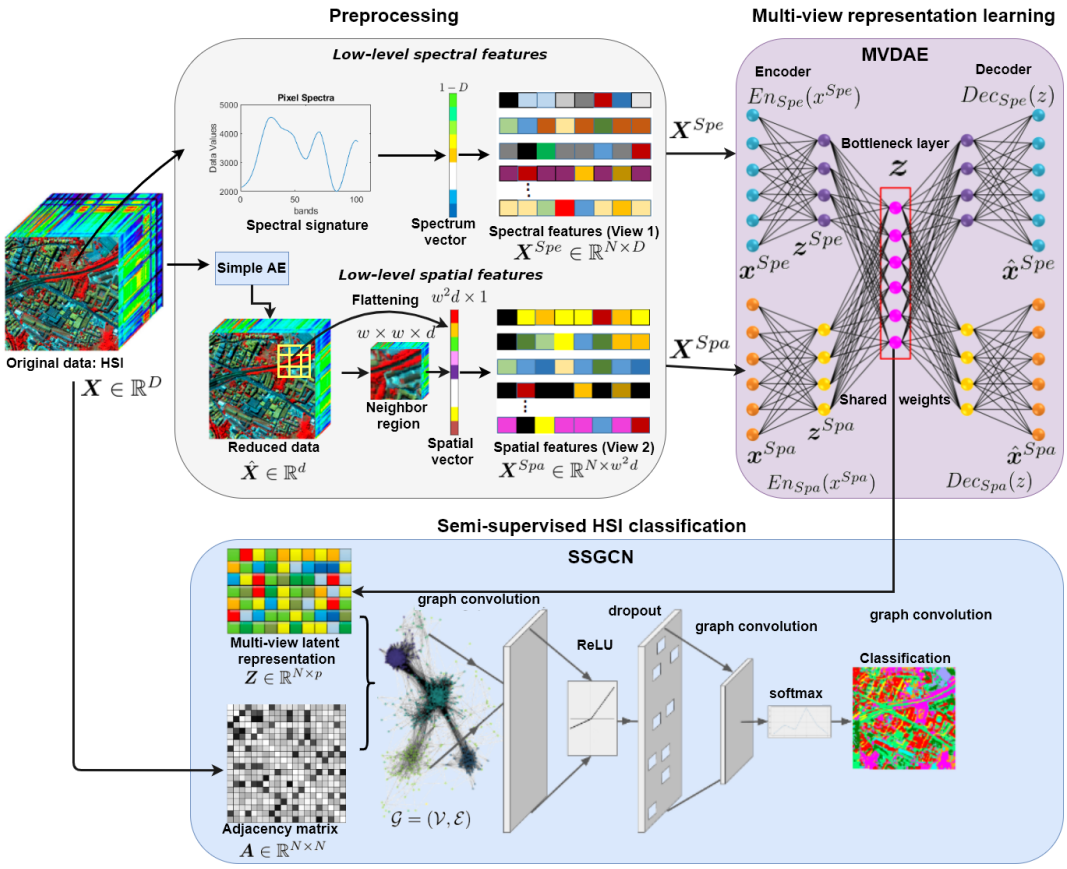

Figure 1: Flowchart of the proposed methodology MV-DNNet

row at location i of the spectral matrix $\mathbf{X}^{spe}$ is the spectral signature of the pixel p_i.

3.1.2. Spatial feature $\mathbf{X}^{spa}$

The aim here is to extract spatial information around each pixels neighborhood and takes all neighbor regions into consideration. Therefore, firstly a simple deep autoencoder (AE) is conducted to reduce the high dimensionality of HSI data from D to d ($d << D$), by preserving its spatial structure. We use AE model along the spectral dimension of $\mathbf{X}$ and only retain several features according to the reconstruction error. Formally, an AE takes as input the original HSI $\mathbf{X} \in \mathbb{R}^{N \times D}$. It includes an encoder noted $E_\theta(\mathbf{X})$ and a decoder noted $D_\phi(\mathbf{z})$ ($\mathbf{z}$ is the bottleneck layer). The encoder E_θ non-linearly projects $\mathbf{X} \in \mathbb{R}^{N \times D}$ into a new latent representation space $\mathbf{z} = E_\theta(\mathbf{X})$ ($\mathbf{z} \in \mathbb{R}^{N \times d}$) from which the decoder D_ϕ seeks to recover $\mathbf{X}$, i.e. $D_\phi(E_\theta(\mathbf{X})) \approx \mathbf{X}$. The AE aims to minimize the reconstruction error between the input $\mathbf{X}$ and its output (reconstructed input) $\hat{\mathbf{X}}$ using a Mean Squared Error (MSE) criterion (see Fig. 2):

$$\mathcal{L}(\theta, \phi; \mathbf{x}) = \mathbb{E}\left[(\mathbf{x} - E_\theta(D_\phi(\mathbf{x})))^2\right] \tag{1}$$

where $E(.)$ and $D(.)$ are parameterized by θ and ϕ, respectively. The parameters (θ, ϕ) are learned together to reconstruct data $\hat{\mathbf{x}}$ same as the initial input $\mathbf{x}$.

Secondly, a neighbor region is extracted around the pixels in the reduced data, i.e., latent representation matrix $\mathbf{Z} \in \mathbb{R}^{N \times d}$, which has only d features ($d << D$) in spectral dimension. For each pixel, we extract a $s \times s$ neighbor pixels. Given that d is the number of extracted features with AE model and a pixel can be considered as a box with a size of $s \times s \times d$. Finally, we flatten the box into $1 \times D$ vector with a size of $s^2 d \times 1$ elements. All 1-D vectors are then concatenated into the spatial features matrix $\mathbf{X}^{spa}$. Therefore, $\mathbf{X}^{spa} \in \mathbb{R}^{N \times s^2 d}$ contains the spatial features of each pixel taking into account their neighbor regions. Figure 3 reports the main architecture of the spatial feature extraction procedure.

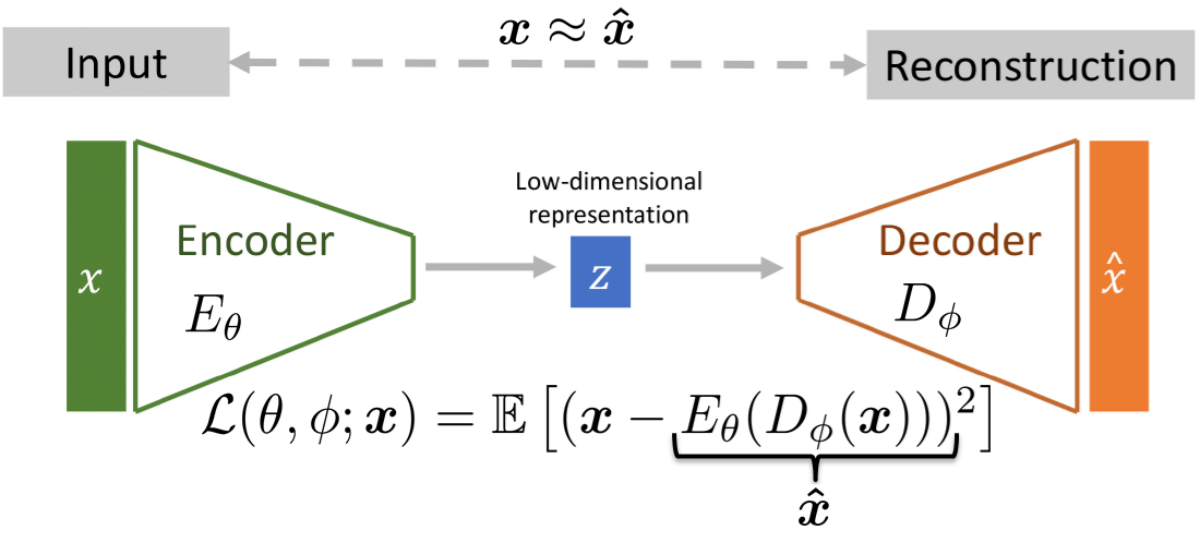

Figure 2: The architecture of the deep autoencoder model (AE)

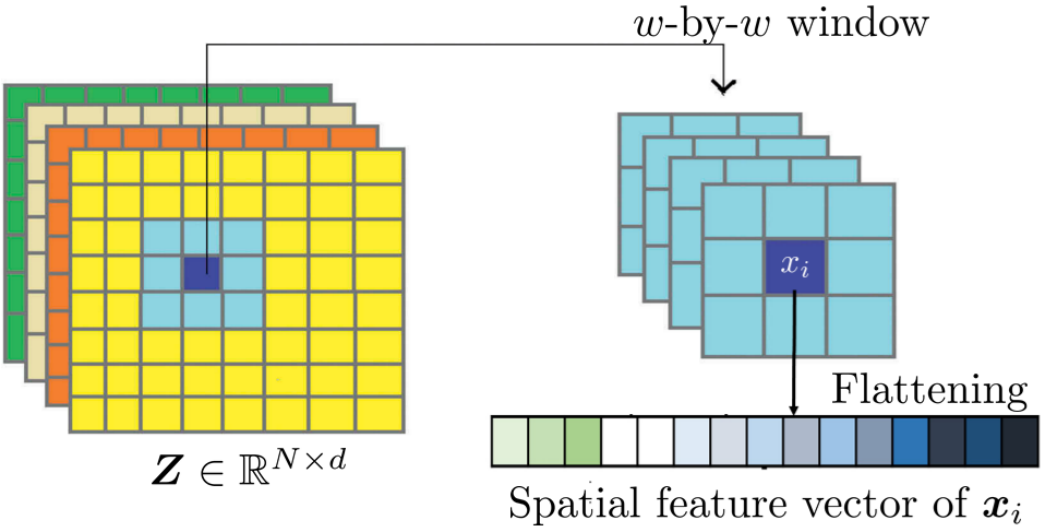

Figure 3: Spatial feature extraction from $\mathbf{Z}$

3.2. Representation learning with multi-view deep autoencoder (MVDAE)

In the classification context, a spectrum of each pixel can contain important and useful information for discriminating different kinds of ground classes. Moreover, with the spatial information, in a neighbor region the statistics of the pixels decreases the intra-class variance which can lead to improve classification performances. Furthermore, we designed a multi-view deep autoencoder (MVDAE) that seeks at extracting high level features from the combination of multiple input views, i.e., spectral and spatial features, from which these input views can be reconstructed. It relies on our case on the assumption that spectral and spatial features are indeed complementary. Our goal is to extract a relevant multi-view latent representation to learn an accurate classification model. The MVDAE includes one encoder per view noted E_{spe} and E_{spa} for spectral and spatial features inputs. Each encoder is a multi layer neural network model that non linearly transforms the input view into a new representation latent space. We note $\mathbf{z}_{spe}$ and $\mathbf{z}_{spa}$ the corresponding latent representations extracted by the two encoders, $\mathbf{z}_{spe} = E_{spe}(\mathbf{x}_{spe})$ and $\mathbf{z}_{spa} = E_{spa}(\mathbf{x}_{spa})$.

A latent representation, $\mathbf{z}$, is extracted from the encoding of the two views $\mathbf{X}^{spe}$ and $\mathbf{X}^{spa}$, then this latent multi-view latent representation $\mathbf{z}$ is input to two decoders D_{spe} and D_{spa} that each aims at reconstructing both views, $\hat{\mathbf{x}}_{spe} = D_{spe}(\mathbf{z})$ and $\hat{\mathbf{x}}_{spa} = D_{spa}(\mathbf{z})$. The decoders are nonlinear neural networks including one to three hidden layers. The learning criterion of the MVDAE is the sum of the reconstruction error criterion of both views, we used the MSE:

$$\mathcal{L}(\mathbf{x}_{spe}, \mathbf{x}_{spa}; \theta) = \sqrt{\frac{1}{D}||\mathbf{x}_{spe} - \hat{\mathbf{x}}_{spe}||^2 + \frac{1}{w^2 d}||\mathbf{x}_{spa} - \hat{\mathbf{x}}_{spa}||^2} \tag{2}$$

The aim of the MDAVE is to find a shared representation from the two encoding data $\mathbf{z}_{spe}$ and $\mathbf{z}_{spa}$ using a specific merging layer actually implemented as a dense layer $\mathbf{z} = E(W_{spe} \times \mathbf{z}_{spe} + W_{spa} \times \mathbf{z}_{spa})$. We instead use another possibility which relies on sharing weights and define the multi-view representation as : $\mathbf{z} = E(W \times \mathbf{z}_{spe} + W \times \mathbf{z}_s)$. This seeks to find a common space, i.e., shared representation between two views: spectral and spatial features matrices. This

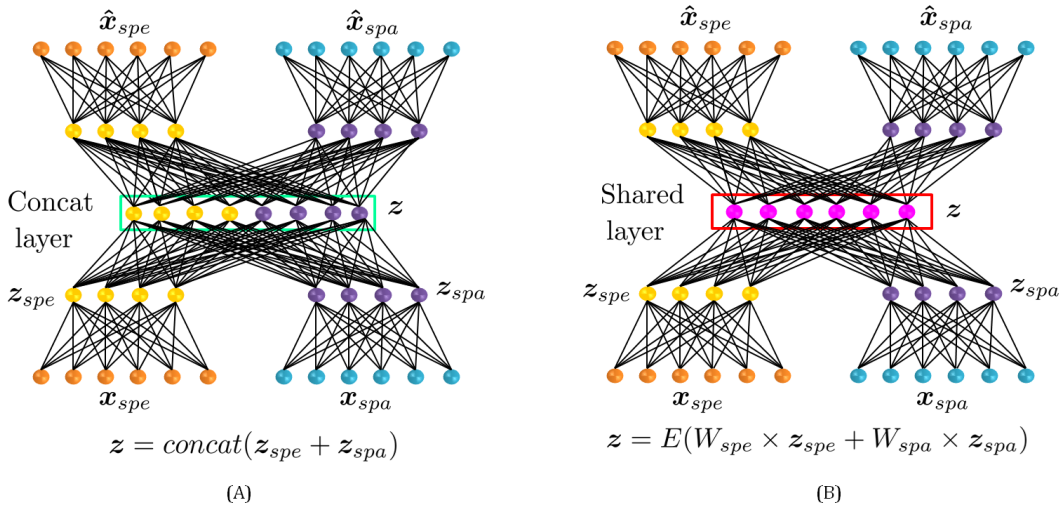

Figure 4: Different architectures can be considered of MVDAE. (A) a multi-view deep autoencoder model trained on the concatenation of both latent representations $\mathbf{x}_{spe}$ and $\mathbf{x}_{spa}$ ($\mathbf{z} = concat(\mathbf{x}_{spe} + \mathbf{x}_{spa})$ ($\mathbf{z}$ is the bottleneck layer). (B) a multi-view deep autoencoder model trained on the shared representations: $\mathbf{z} = E(W_{spe} \times \mathbf{z}_{spe} + W_{spa} \times \mathbf{z}_{spa})$.

is a subtle difference in this scheme that introduces constraints on how the latent representation is defined (see Fig. 4).

3.3. Spectro-Spatial Graph Construction

After applying an AE upon HSI $\mathbf{X}$, we obtain a reduced 3D cube $z = [x_1, ..., x_d] \in \mathbb{R}^{N \times d}$. We build then an undirected graph $\mathcal{G} = (\mathcal{V}, \mathcal{E}, \mathcal{W})$, where $\mathcal{V} = \{v_1, ..., v_N\}$ is a set of nodes corresponding to the pixels, $\mathcal{E}$ is the set of edges, and $\mathcal{W} \in \mathbb{R}^{N \times N}$ is the weighted adjacency matrix of $\mathcal{G}$, where $w_{i,j}$ is a weight attributed to the edge $e_{i,j} = (v_i, v_j) \in \mathcal{E}$. Each vertex v_i in the graph $\mathcal{G}$ has a feature vector of size d (i.e., the number of extracted features with AE, i.e., the multi-view latent representation matrix $\mathbf{Z}$). Then, we define an edge $e_{i,j}$ between two nodes (pixels) v_i and v_j based on a similarity criterion, which is computed by taking into account the spatial information (pixels neighborhoods) as well as the spectral information (intensity values). Formally, each vertex v_i is connected to v_j if x_j belongs to the neighborhood of x_i in some p-by-p window.

The weight $w_{i,j}$ for the edge $e_{i,j}$ is computed using the following formula:

$$w_{i,j} = \begin{cases} \exp^{-\phi(x_i,x_j)} \times \exp \frac{-dist(x_i,x_j)}{t} & \text{if } dist(x_i,x_j) < p \\ 0 & \text{otherwise} \end{cases} \tag{3}$$

where the parameter p is the neighborhood size (window), $t \in \mathbb{R}$ is a heat kernel which is a parameter that is used to compute the weight matrix W, $\phi(\cdot)$ is the spectral angle mapper which is the spectral distance given by the angle between the feature vectors of pixels x_i and x_j using Min/Max normalization, and $dist(\cdot)$ is the spatial distance. $\phi(x_i, x_j)$ is computed as follows:

$$\phi(x_i, x_j) = cos^{-1}\left(\frac{< x_i, x_j >}{\parallel x_i \parallel \cdot \parallel x_j \parallel}\right) \tag{4}$$

The spatial distance $dist(x_i, x_j)$ is calculated as follows:

$$dist(x_i, x_j) = \sqrt{((r_i - r_j)^2 + (l_i - l_j)^2)} \tag{5}$$

where (r_i, l_i) and (r_i, l_i) are the coordinates of x_i and x_j respectively. Algorithm 1 reports the different steps for the spectro-spatial graph construction.

3.4. Semi-Supervised Classification with Graph Convolutional Network (SSGCN)

In this section, we present our proposed semi-supervised graph convolutional network (SSGCN). The main goal of this model is to perform the spectral-spatial classification by considering the constructed graph $\mathcal{G}$. For semi-supervised learning, let $TD_l = \{\mathbf{z}_i, y_i\}_{i=1}^{L}$ be a set of labeled training dataset of size L, where $\mathbf{z}_i$ indicates a feature vector of the i^{th} labeled pixel, and y_i is its corresponding label. Moreover, let $TD_{nl} = \{\mathbf{z}_i\}_{i=L+1}^{L+NL}$ be a set of unlabeled training samples of size NL ($L + NL = N$). The aim of semi-supervised learning is to predict the labels of unlabeled training samples TD_{nl}, using a non linear function $f(\mathbf{Z}, W)$ such as ReLu [43].

3.4.1. Convolutional layers

Convolution on graphs can be computed by multiplying each graph signal $\bar{\mathbf{z}}$ by a filter g_θ parametrized by the Fourier coefficient $\theta \in \mathbb{R}^N$ [44]. Usually, the

Algorithm 1 Spectro-Spatial Graph Construction

Input: $\mathbf{X} \in \mathbb{R}^{N \times D}$, $\mathbf{Z} \in \mathbb{R}^{N \times d}$

p (window size): integer; t (heat kernel): float

Output: $G = (V, E, W)$ Spectro-Spatial Graph of X and $Diag$: diagonal degree matrix

Initialization:$V \leftarrow [\,]$, $E \leftarrow [\,]$, $W \leftarrow [\,]$, $Diag \leftarrow [\,]$

// Compute the list of vertices V

for $i = 1 : N$ **do** // N: Number of pixels

  for $j = 1 : d$ **do** // d Number of extracted features

    $V[i][j] \leftarrow Z[i][j]$

  end for

end for

// Find the list of edges E for each voxel v_i and compute their weights

for $i = 1 : N$ **do**

  for $j = 2 : N - 1$ **do**

    if $dist(x_i, x_j) < p$ **then**

      $E \leftarrow E \cup \{(x_i, x_j)\}$

      $W[i][j] \leftarrow \exp^{-\phi(x_i, x_j)} \times \exp \frac{-dist(x_i, x_j)}{t}$

    else

      $W[i][j] \leftarrow 0$

    end if

  end for

end for

// Compute $Diag$ based on the weighted adjacency matrix W

for $i = 1 : N$ **do**

  $Diag[i] \leftarrow 0$

  for $j = 1 : N$ **do**

    $Diag[i] \leftarrow Diag[i] + W[i][j]$

  end for

end for

return $V, E, W, Diag$

graph *Fourier transform* for a signal z is defined as:

$$\mathscr{F}(z) = U^T z = \hat{z} \in R^n \tag{6}$$

where $\mathscr{F}^{-1}(z) = U\hat{z}$, $\mathbf{U}$ is the matrix of eigenvectors of the normalized graph Laplacian $\mathbf{L}_n = \mathbf{I}_N - Diag^{\frac{-1}{2}}\mathbf{W}Diag^{\frac{-1}{2}} = \mathbf{U}\Sigma\mathbf{U}^T$, $Diag$ is the diagonal degree matrix of $\mathcal{G}$, $\mathbf{I}_N$ is the identity matrix, and Σ is the diagonal matrix of eigenvalues. The graph convolution of the latent representation z with a filter $g \in R^n$ is calculated using:

$$x *_{\mathcal{G}} g = \mathscr{F}^{-1}(\mathscr{F}(x) \odot \mathscr{F}(g)) = U(U^T(x) \odot U^T(g)) \tag{7}$$

where $\odot$ denotes the element wise product. According to [45], we can efficiently compute an approximated convolution of G as follows:

$$g_\theta \star \bar{\mathbf{z}} = \theta\, B\, \bar{\mathbf{z}} \tag{8}$$

where $B = \mathbf{I}_N + Diag^{\frac{-1}{2}}\mathbf{W}Diag^{\frac{-1}{2}} + (Diag^{\frac{-1}{2}}WDiag^{\frac{-1}{2}})^2$.

3.4.2. Training SSGCN

For the semi-supervised learning, the optimal neural network weights

$$\mathbf{W}^{(0)}, \mathbf{W}^{(1)}, \ldots, \mathbf{W}^{(K)}$$

can be trained using the labeled set of training samples $TD_L = (i, y_i)_{i=1}^{L}$, by minimizing the standard cross-entropy loss function:

$$\mathcal{L}oss = -\sum_{i=1}^{L} y_i \ln \mathbf{M}_i \tag{9}$$

where $\mathbf{M}_i$ is the label output of node i in the final layer.

3.4.3. Semi-supervised classification

The proposed SSGCN aims to predict the labels of unlabeled pixels $\mathbf{z}_i \in TD_{nl}$ which will go through various propagation layers. Formally, given a input multi-view latent feature matrix $\mathbf{Z}$ and a weighted adjacency matrix $\mathbf{W}$, our

SSGCN applies a layer-wise propagation rule using the Rectified Linear Unit ($ReLu$) as a non linear activation function and $softmax()$ as a classifier:

$$\begin{aligned} \mathbf{Z}^{(1)} &= ReLu(B\,\mathbf{Z}^{(0)}\,\mathbf{W}^{(0)}) \\ &\vdots \\ \mathbf{Z}^{(K-1)} &= ReLu(B\,\mathbf{Z}^{(K-2)}\,\mathbf{W}^{(K-2)}) \\ \mathbf{Z}^{(K)} &= softmax(B\,\mathbf{Z}^{(K-1)}\,\mathbf{W}^{(K-1)}) \end{aligned} \tag{10}$$

where $\mathbf{Z}^{(0)} = \mathbf{Z}$, $\{\mathbf{Z}^{(1)}, \mathbf{Z}^{(2)}, \ldots, \mathbf{Z}^{(K-1)}\}$ are the feature map outputs of the different layers and $\mathbf{Z}^{(K)} = M$ is the label output of the final layer, i.e., $\mathbf{M}_i$ is the label of vertex v_i.

4. Experimental Results

4.1. HSI Description

To evaluate the effectiveness and the performance of the proposed methodology, we perform our experiments on three real HSIs [1]:

- The Indian Pines HSI collected by the Airborne Visible/ Infrared Imaging Spectrometer (AVIRIS) sensor, which represents the north-western Indiana. It consists of 145×145 pixels with a spatial resolution of 20 m per pixel and 220 spectral bands in the wavelength range from 0.4 to 2.5 μm. The ground truth contains 16 classes. Fig. 5 reports the false color image and its ground truth.

- The Salinas image collected by the AVIRIS sensor over Salinas, California, which consists of 512×217 pixels with a spatial resolution of 3.7 m per pixel and 224 spectral bands. The ground truth contains 16 classes (see Fig. 6).

[1] `http://www.ehu.eus/ccwintco/index.php?title=Hyperspectral_Remote_Sensing_Scenes`

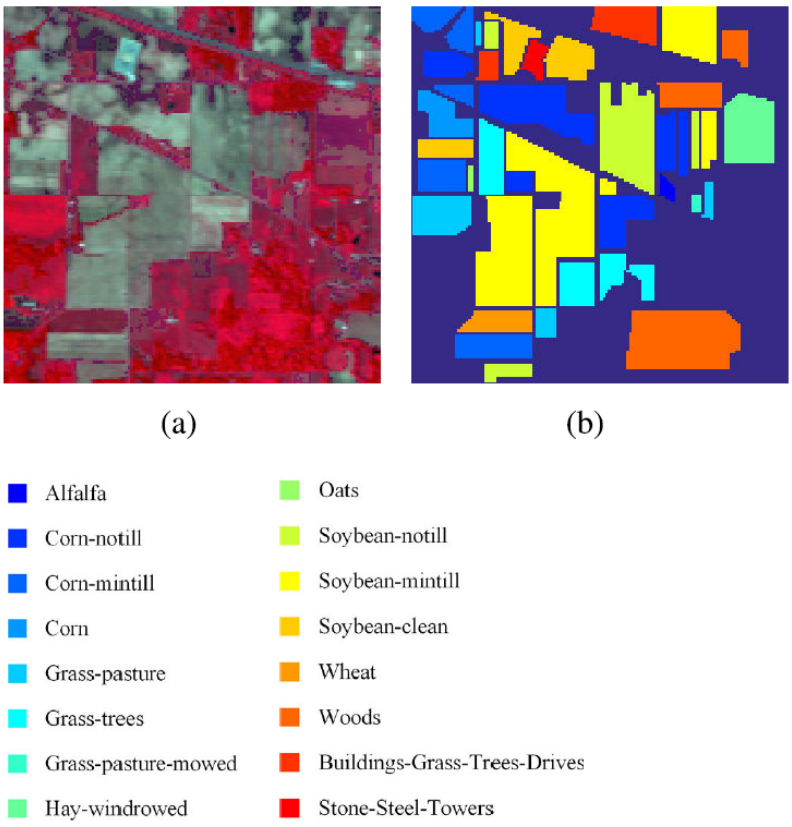

Figure 5: (a) False color image of the Indian Pines Dataset. (b) Ground-truth classification map of Indian Pines Dataset.

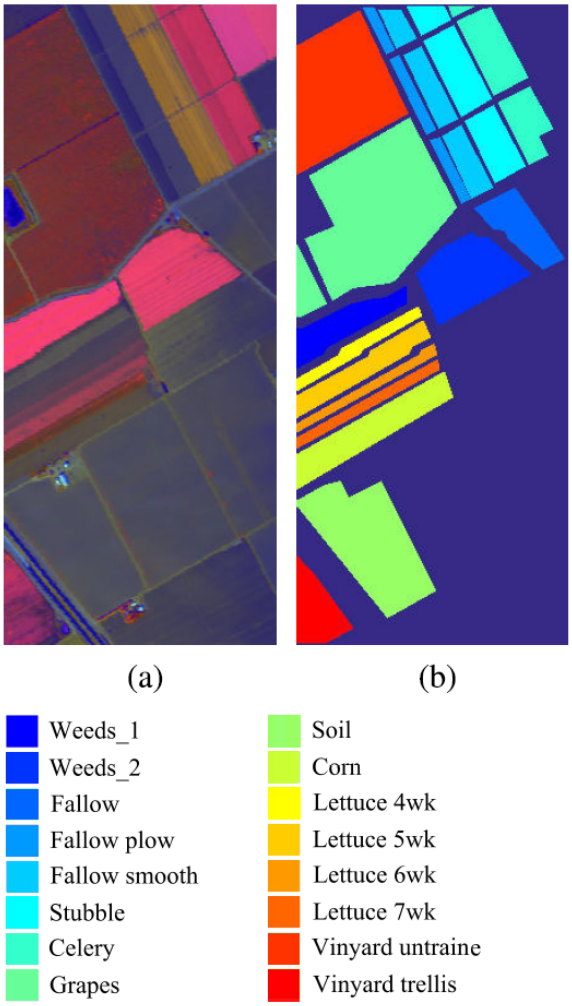

Figure 6: (a) False color image of the Salinas Dataset. (b) Ground-truth classification map of Salinas Dataset.

- The last HSI is the Pavia University data collected by the Reflective Optics System Imaging Spectrometer (ROSIS-03) sensor, which consists of 610 × 610 pixels, and 115 spectral bands in the range from 0.43 to 0.86 μm, with a spatial resolution of 1.3 m. The ground truth contains 9 classes (see Fig. 7).

4.2. Performance Evaluation Metrics and Parameters Setting

In order to train our MV-DNNet model, we randomly choose 10% of the samples per class as training samples and the rest as testing samples (see table 1). After several tests, we choose stochastic gradient descent (SGD) optimizer for the training. The learning rate lr is fixed to 10^{-3} as training parameter, the training epoch to 200, and the batch size to 300, with 10000 iterations, a weight decay of 5.10^{-4}, and a momentum of 0.9. We variate several encoding dimensions for each HSI from 2 to 100 ($d \in [2, .., 100]$). We tested different pairs of activation functions (hidden layers, output layer) of the MVDAE model: (*linear, linear*), (*relu, linear*), and (*relu, sigmoid*). It is implemented using the *keras* toolkit. We fixed the window size $s = 3$ of the AE in the first step of the spatial feature construction. The diagonal degree matrix $Diag$ is fixed also with a window size $p = 3$ to take into account the 8-neighbors pixels. We repeated our experiments 10 times with random training samples to get stable classification accuracy. We adopted some metrics of performance to assess the classification rate: overall accuracy (OA), average accuracy (AA), and kappa coefficient (k). The OA is the number of corrected classified pixels divided by the total number of testing pixels, whereas AA is the mean value of classification accuracy of all classes. The k index is a statistical measurement of consistency between the classification maps and the ground truth.

4.3. Analysis of the reconstruction error

In this section, we present the reconstruction errors obtained by different representation learning models, including, PCA, ICA, AE(*linear,linear*), AE(*relu,linear*), and AE(*relu,sigmoid*). We opted for the MSE loss function

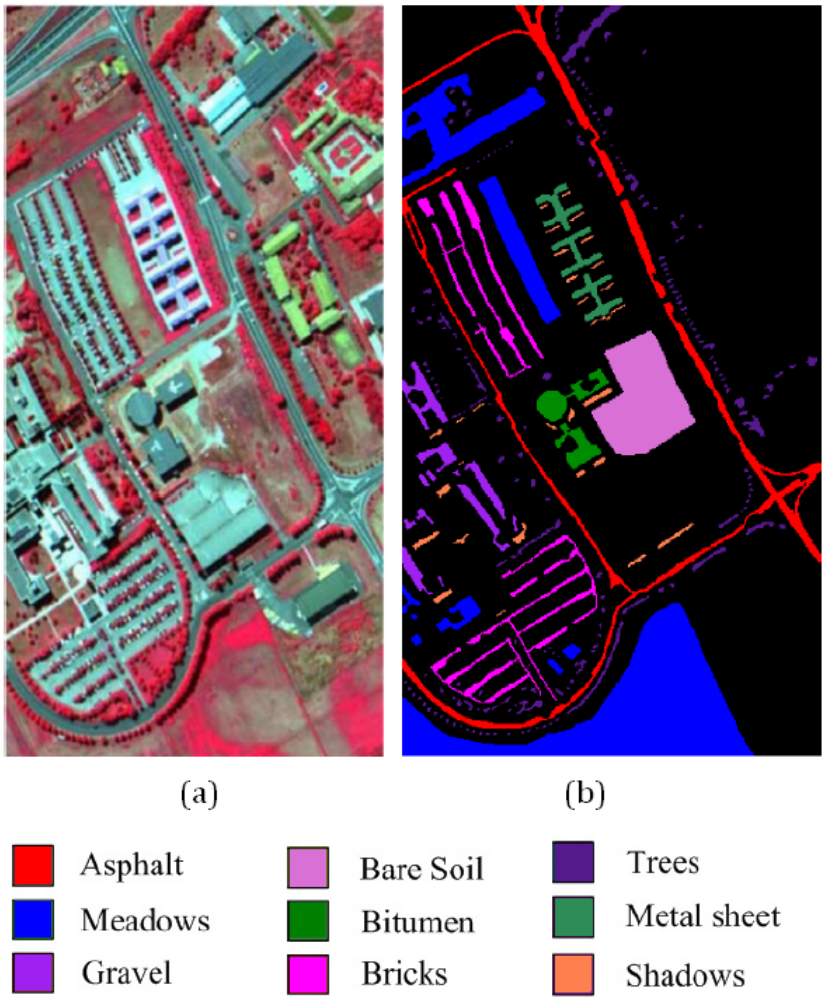

Figure 7: (a) False color image of the Pavia University Dataset. (b) Ground-truth classification map of Pavia University Dataset.

Table 1: Training and testing sets for Indian Pines, Salinas, and Pavia datasets.

Indian Pines			Salinas			Pavia University		
Class	Samples		Class	Samples		Class	Samples	
	Train	Test		Train	Test		Train	Test
Alfalfa (C1)	5	49	Broc-W-1 (C1)	200	1809	Asphalt (C1)	132	6499
Build-G (C2)	38	342	Broc-W-2 (C2)	372	3354	Bare-S (C2)	100	4929
Corn (C3)	23	211	Fallow (C3)	197	1779	Bitumen (C3)	26	1304
Corn-M (C4)	83	751	Fallow-P (C4)	139	1255	Gravel (C4)	41	2058
Corn-N (C5)	143	1291	Fallow-S (C5)	267	2411	Meadows (C5)	373	18321
Grass-W (C6)	5	21	Stubble (C6)	395	3564	Painted-M (C6)	26	1319
Grass-P (C7)	50	447	Celery (C7)	357	3222	Self-B (C7)	73	3609
Grass-T (C8)	75	672	Grapes-U (C8)	1127	10144	Shadows (C8)	19	928
Hay-W (C9)	49	440	Soil-V-D (C9)	620	5583	Tree (C9)	61	3003
Oats (C10)	2	18	Corn-W (C10)	327	2951			
Soyb-C (C11)	62	552	Let-4wk (C11)	106	962			
Soyb-M (C12)	247	2221	Let-5wk (C12)	192	1735			
Soyb-N (C13)	97	871	Let-6wk (C13)	91	825			
Stone-S (C14)	10	85	Let-7wk (C14)	107	963			
Wheat (C15)	21	191	Viney-U (C15)	726	6542			
Woods (C16)	130	1164	Viney-T (C16)	180	1627			

to compute the reconstruction error between the initial HSI $\mathbf{X}$ and the reconstructed input $\hat{\mathbf{X}}$ and we reported then the average MSE versus to the encoding dimension for Indian Pines, Salinas, and Pavia University (see Fig. 8). Thus, we can interpret that the AE(*relu,linear*) is the appropriate one for representation learning for three HSIs data with an MSE value equal to 0.069 and an encoding dimension set to 20 for Indian Pines, 0.064 for an encoding dimension of 30 for Salinas, and 0.058 for a dimension defined to 20 for Pavia University. For the other methods, the best average MSE, i.e. $MSE < 0.1$ is obtained when the size of the latent representation is greater than ≈ 80 features. However, in our case we need a lower encoding dimension due the curse of dimensionality and the overfitting problem. Moreover, Fig. 9 reports the reconstruction error and standard error values of the best model AE(*relu,linear*) from 1 to 4 hidden dense layers, where the encoding dimension is 20. We can notice from the obtained results that the best obtained MSE is equal to $0.069(\pm 0.006)$ for

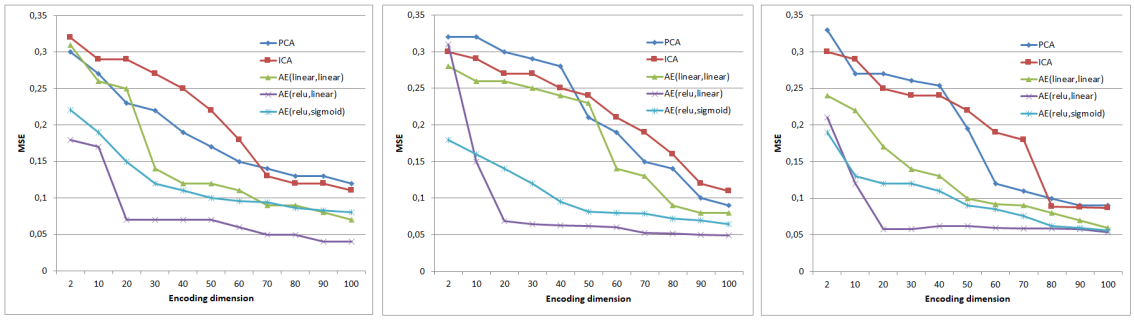

Figure 8: MSE of reconstruction versus encoding dimension. (A) Indian Pines, (B) Salinas, and (C) Pavia University

Indian Pines, where the number of hidden layers is 3. For the Salinas, the best MSE is equal to 0.067($\pm$0.009), for again a number of hidden layers equals to 3, i.e., the same architecture of AE. Also, for the Pavia University the best MSE is equal to 0.058($\pm$0.007) when the number of hidden layers is 3. Therefore, following these experiments we set the number of layers to 3 and we fix the best configuration (size and number of dense layers) of the architecture as follows: $\mathbf{Z}^{(1)} = 170$, $\mathbf{Z}^{(2)} = 130$, and $\mathbf{Z}^{(3)} = 20$.

In order to evaluate the effectiveness of the proposed MVDAE model on the classification task, we performed a comparative study using the concatenation of inputs, i.e., the concatenation of spectral and spatial features denoted by $\mathbf{x}_{spe} + \mathbf{x}_{spa}$, and the concatenation of latent representations $\mathbf{z}_{spe} + \mathbf{z}_{spa}$. The aim here is to demonstrate the potential of the shared representation obtained by our model MV-DNNet. Tables 2, 3, and 4 report the obtained best average MSE and OA on Indian Pines, Salinas, and Pavia University, respectively. Based on the obtained results, we can notice that the fusion of latent representation $\mathbf{z}_{spe}$ and $\mathbf{z}_{spa}$ have shown their added value in the classification task. Moreover, we can observe that the proposed MVDAE model based on the fusion of latent representation can achieve a better classification than the concatenation of spectral and spatial features, i.e., $\mathbf{x}_{spe} + \mathbf{x}_{spa}$.

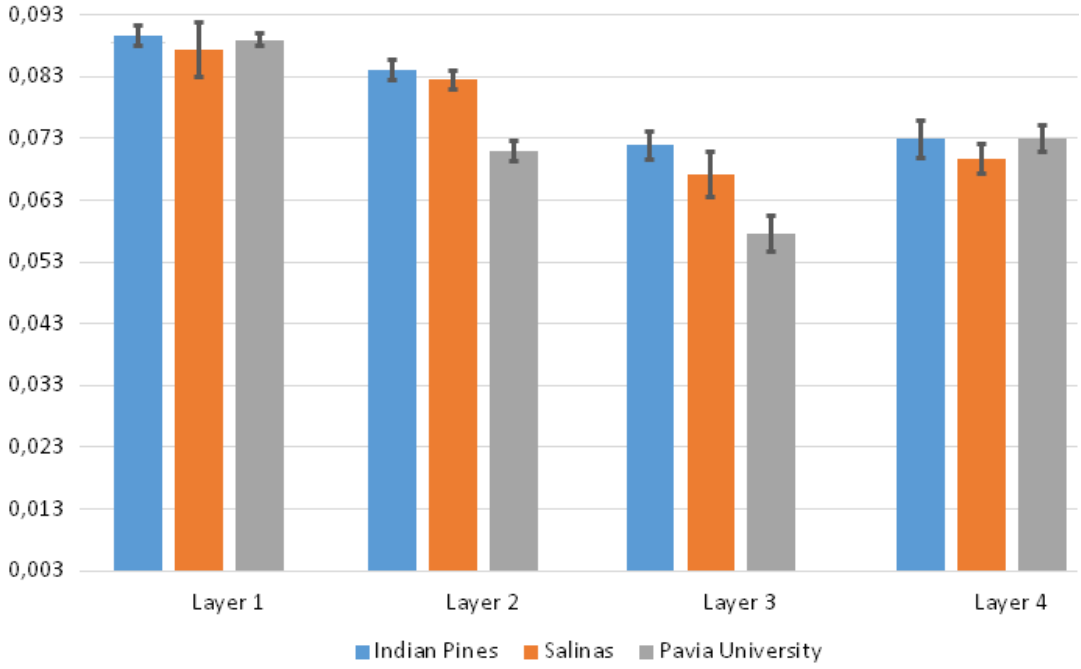

Figure 9: MSE of reconstruction error versus encoding dimension using four different layers ($d = 20$)

Table 2: Best average MSE and OA (± standard error) using spectral and spatial features of Indian Pines data based on PCA, ICA, and different AE/MVDAE models (3 layers)

	Concatenated inputs ($\mathbf{x}_{spe} + \mathbf{x}_{spa}$)		Fused latent rep. ($\mathbf{z}_{spe} + \mathbf{z}_{spa}$)	
	Avg MSE	Avg OA	Avg MSE	Avg OA
PCA	0.121 (± 0.023)	94.23 (± 0.047)	N/A	N/A
ICA	0.118 (± 0.019)	95.44 (± 0.092)	N/A	N/A
AE/MVDAE (*lin, lin*)	0.093 (± 0.011)	96.23 (± 0.014)	0.085 (± 0.038)	96.52 (± 0.023)
AE/MVDAE (*relu, lin*)	**0.071** (± 0.020)	**96.94** (± 0.013)	**0.068** (± 0.095)	**97.68** (± 0.042)
AE/MVDAE (*relu, sig*)	0.094 (± 0.063)	96.14 (± 0.028)	0.090 (± 0.036)	96.19 (± 0.012)

Table 3: Best average MSE and OA (± standard error) using spectral and spatial features of Salinas based on PCA, ICA, and different AE/MVDAE models (3 layers)

	Concatenated inputs ($\mathbf{x}_{spe} + \mathbf{x}_{spa}$)		Fused latent rep. ($\mathbf{z}_{spe} + \mathbf{z}_{spa}$)	
	Avg MSE	Avg OA	Avg MSE	Avg OA
PCA	0.092 (± 0.039)	95.84 (± 0.013)	N/A	N/A
ICA	0.101 (± 0.026)	95.62 (± 0.069)	N/A	N/A
AE/MVDAE (*lin, lin*)	0.083 (± 0.014)	96.59 (± 0.091)	0.071 (± 0.043)	96.84 (± 0.081)
AE/MVDAE (*relu, lin*)	**0.073** (± 0.046)	**96.72** (± 0.022)	**0.063** (± 0.082)	**98.24** (± 0.036)
AE/MVDAE (*relu, sig*)	0.087 (± 0.054)	96.33 (± 0.082)	0.074 (± 0.024)	96.50 (± 0.044)

Table 4: Best average MSE and OA (± standard error) using spectral and spatial features of Pavia University based on PCA, ICA, and different AE/MVDAE models (3 layers)

	Concatenated inputs ($\mathbf{x}_{spe} + \mathbf{x}_{spa}$)		Fused latent rep. ($\mathbf{z}_{spe} + \mathbf{z}_{spa}$)	
	Avg MSE	Avg OA	Avg MSE	Avg OA
PCA	0.089 (± 0.092)	96.81 (± 0.077)	N/A	N/A
ICA	0.094 (± 0.074)	96.34 (± 0.021)	N/A	N/A
AE/MVDAE (*lin, lin*)	0.075 (± 0.013)	96.84 (± 0.034)	0.071 (± 0.043)	96.84 (± 0.081)
AE/MVDAE (*relu, lin*)	**0.070** (± 0.029)	**97.12** (± 0.047)	**0.051** (± 0.063)	**99.16** (± 0.016)
AE/MVDAE (*relu, sig*)	0.079 (± 0.031)	96.51 (± 0.011)	0.073 (± 0.031)	96.72 (± 0.072)

Table 5: Classification performances using DSVM, SAE, CNN, DCNN,R-VCA, SKCR, GF, 3DCNN, F-3DCNN, and MV-DNNet (Ours): Indian Pines HSI ($d = 20$)

Class	Models									
	DSVM	SAE	CNN	DCNN	R-VCA	SKCR	GF	3DCNN	F-3DCNN	Ours
C1	88.22	93.92	93.41	94.21	95.02	93.81	94.01	94.46	96.32	**96.42**
C2	92.31	91.24	**96.80**	95.41	94.91	95.63	95.69	92.19	96.26	96.65
C3	95.12	93.10	95.26	96.98	97.36	97.58	93.48	**98.05**	96.97	98.01
C4	95.96	89.51	93.82	97.56	96.99	97.06	92.69	97.09	96.99	**98.12**
C5	97.03	93.18	96.22	95.11	94.78	95.06	96.21	92.16	**97.42**	97.12
C6	83.03	88.21	94.02	96.51	98.89	96.28	96.32	95.10	95.18	**97.08**
C7	89.64	80.21	95.04	95.63	96.14	97.15	98.01	94.06	97.02	**98.61**
C8	92.03	94.31	98.01	97.10	95.31	96.14	97.22	**98.91**	96.89	98.10
C9	96.12	92.10	**97.02**	94.12	93.36	92.78	92.21	97.03	96.56	96.91
C10	91.22	87.03	98.04	91.85	94.21	95.03	94.02	94.02	97.07	**98.07**
C11	85.13	88.21	95.32	93.45	94.21	94.09	96.72	95.21	97.31	**97.52**
C12	91.12	91.04	86.14	89.48	92.14	95.47	96.12	95.03	96.84	**97.08**
C13	86.12	82.23	94.06	94.25	93.84	95.23	95.06	96.21	96.21	**96.89**
C14	95.01	86.07	97.13	93.85	94.96	96.98	96.21	96.12	97.26	**97.94**
C15	95.02	93.06	95.23	94.61	95.95	94.86	95.04	94.18	96.61	**97.08**
C16	92.38	94.26	96.10	96.85	95.91	96.21	97.02	95.19	96.95	**98.51**
OA	92.24	91.85	94.99	94.86	94.26	96.23	96.52	96.54	96.98	**97.68**
AA	92.13	91.78	94.81	96.72	94.08	96.11	96.39	96.47	96.85	**97.55**
k	92.25	91.77	94.85	94.78	94.18	96.21	96.52	96.40	96.89	**97.62**
Time (s)	230	241	322	361	298	267	239	278	274	211

Table 6: Classification performances using DSVM, SAE, CNN, DCNN,R-VCA, SKCR, GF, 3DCNN, F-3DCNN, and MV-DNNet (Ours): Salinas HSI ($d = 30$)

Class	Models									
	DSVM	SAE	CNN	DCNN	R-VCA	SKCR	GF	3DCNN	F-3DCNN	Ours
C1	95.22	97.12	98.04	96.95	95.75	96.23	97.07	95.26	97.01	**97.84**
C2	94.93	95.84	96.28	94.21	93.87	95.41	96.18	95.92	**96.68**	96.71
C3	93.16	95.89	97.87	96.81	97.25	96.45	97.13	95.86	96.88	**97.67**
C4	95.14	95.12	96.24	95.24	96.11	96.36	96.87	**96.99**	95.98	96.69
C5	96.42	96.12	96.85	97.01	96.81	96.74	97.62	96.74	96.75	**98.26**
C6	96.24	95.29	97.17	96.91	96.85	97.11	97.09	96.12	97.88	**98.92**
C7	96.14	95.85	97.42	97.21	97.53	98.01	97.82	96.40	97.51	**98.72**
C8	97.02	96.62	97.17	96.12	96.34	97.04	**97.80**	96.87	96.94	97.55
C9	96.26	96.84	96.92	97.11	99.94	97.25	97.84	96.74	97.26	**98.62**
C10	97.21	94.11	97.21	95.23	96.16	96.98	97.12	96.21	96.23	**98.43**
C11	94.62	95.81	97.14	96.74	96.84	97.14	97.47	97.22	97.10	**98.32**
C12	93.14	94.06	96.21	95.99	96.47	96.81	97.12	97.08	96.28	**97.89**
C13	94.98	96.16	96.82	96.95	97.03	96.96	97.08	97.16	97.21	**98.45**
C14	96.24	93.81	97.12	97.23	96.83	97.01	96.92	96.13	97.16	**97.51**
C15	96.08	94.92	97.74	97.15	98.12	98.04	98.83	97.84	97.28	**98.68**
C16	96.47	96.11	97.94	97.62	96.95	97.74	98.32	97.61	97.62	**98.22**
OA (%)	95.12	96.24	97.12	96.56	96.74	97.12	97.18	96.36	97.65	**98.24**
AA (%)	95.02	96.19	97.03	96.31	96.61	96.98	97.11	96.22	97.52	**98.17**
$k \times 100$	95.07	96.22	97.15	96.30	96.47	97.07	97.16	96.28	97.49	**98.20**
Time (s)	320	289	295	344	332	297	230	265	289	259

Table 7: Classification performances using DSVM, SAE, CNN, DCNN,R-VCA, SKCR, GF, 3DCNN, F-3DCNN, and MV-DNNet (Ours) : Pavia University HSI (d=20)

Class	Models									
	DSVM	SAE	CNN	DCNN	R-VCA	SKCR	GF	3DCNN	F-3DCNN	Ours
C1	95.70	97.89	96.49	96.52	96.87	95.89	96.16	96.93	97.24	**98.21**
C2	95.24	96.24	97.21	96.49	96.84	98.45	96.33	95.35	98.07	**99.22**
C3	96.25	97.41	97.82	96.91	97.24	97.62	97.24	96.39	98.21	**98.87**
C4	96.72	96.35	96.84	96.07	95.89	96.84	96.79	96.93	97.18	**99.22**
C5	96.91	97.24	96.86	96.66	96.15	96.87	96.06	96.02	98.00	**98.62**
C6	96.82	96.89	97.63	96.97	97.12	96.84	98.10	95.52	97.93	**99.01**
C7	96.24	96.16	97.79	96.81	96.54	97.03	97.21	95.75	97.71	**98.24**
C8	95.95	96.89	96.79	96.88	96.99	97.14	97.26	96.59	98.09	**98.56**
C9	96.84	97.26	96.83	96.96	96.91	97.03	97.06	96.96	98.90	**99.63**
OA (%)	95.98	96.94	97.12	96.74	96.81	97.08	96.92	96.85	97.83	**99.16**
AA (%)	95.78	96.84	97.05	96.58	96.79	97.01	96.81	96.64	97.71	**99.04**
$k \times 100$	95.81	96.87	97.11	96.63	96.80	97.06	96.87	96.82	97.77	**98.07**
Time (s)	298	195	149	221	197	211	194	201	175	189

4.4. Classification performances using the proposed MV-DNNet with different deep learning-based models

In this section, we compare the obtained classification results using the proposed approach MV-DNNet with other deep learning-based methods, including, deep suppport vector machines (DSVM) [46], stacked autoencoder (SAE) [32], CNN [40], discriminative convolutional neural network (DCNN) [47], rolling guidance filter and vertex component analysis network (R-VCA) [48], structural-kernel collaborative representation (SKCR)[49], gabor filtering (GF) [41], 3D convolutional neural network (3DCNN) [50], and fused 3D CNN (F-3DCNN) [23]. DSVM seeks to use several kernels in the deep SVM model (exponential radial basis function, gaussian radial basis function, neural, and polynomial) to improve the HSI classification. SAE uses the AE model and PCA technique to preserve the spectral and spatial information in the classification task. The CNN model aims to perform the HSI classification by considering both spatial context and spectral features. DCNN uses triplet loss to improve the HSI classification. R-VCA aims to incorporate the spatial information and spectral characteristics

in the classification task using the rolling guidance filter and vertex component analysis network. SKCR seeks to preserve the spatial neighborhood of the pixels in a superpixel belonging to the same class. Furthermore, GF is used to preserve the spatial information in order to improve the performance of the classification. 3D CNN uses a 3D convolution operation to take into account simultaneously the spectral and spatial features. Finally, F-3DCNN is proposed to fuse several 3D CNN in order to enhance the classification rates.

Table 5 reports the obtained classification accuracies OAs for the Indian Pines HSI. Based on these results, we can notice that the proposed model MV-DNNet gives better classification performance, compared to other deep learning-based methods. In fact, the obtained OA is 97.68%, AA is 97.55%, and k is 97.62%. However, for few classes our results are slightly less. For instance, the 3DCNN and F-3DCNN methods give the best classification rates for three classes 'Corn', 'Corn-N' and 'Grass-T', with an OA of 98.05%, 97.42% and 98.91%, respectively and the CNN model for the 'Building-G' and 'Hay-W', with an OA of 96.80% and 97.02%, respectively. For the remaining 11 classes of Indian Pines, the proposed method MV-DNNet gives better classification performance. Fig. 10 reports a visual classification maps for Indian Pines HSI with the corresponding classification rates OA for different models. As shown in this figure, the SAE and DSVM models present noisy classification results because they only exploit the concatenated spectral and spatial information into a single vector without select useful features. Moreover, the CNN, DCNN, R-VCA, SKCR, and GF methods can provide smoother classification performances. Furthermore, due to the limited number of training labeled samples, the CNN, 3DCNN, and F-3DCNN also present noisy classification results. In contrast, the proposed MV-DNNet model not only delivers better classification performances but also achieves accurate classification on the edges area.

For the Salinas HSI, the obtained classification rate OA with MV-DNNet is 98.24%, AA is 98.17%, and k is 98.29% (see Table 6). Also, we can notice that the MV-DNNet method gives better classification rates for 13 out of 16 classes, with a number of features $d = 30$. Most classification rates are greater

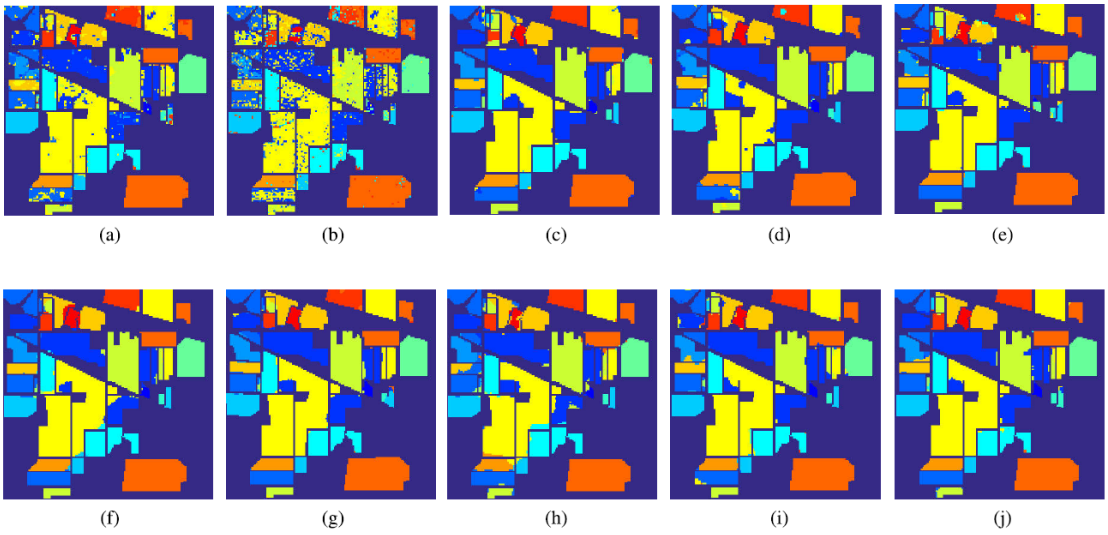

Figure 10: Classification maps for the Indian Pines HSI obtained by (a) DSVM, (b) SAE, (c) CNN, (d) DCNN, (e) R-VCA, (f) SKCR, (g) GF, (h) 3DCNN, (i) F-3DCNN, and (j) MV-DNNet.

than 97%. Again, for very few classes our model is slightly less. Indeed, the GF method gives the best classification rates for the class 'Grapes-U', with an OA of 97.80%. Also, the 3DCNN model is better for the class 'Fallow-R-P', with an OA of $96,99\%$, and the F-3DCNN is better for the class 'Brocoli-G-W-2', with an OA of 96.68%. Fig. 11 shows the thematic maps of different classification methods for Salinas Dataset. Overall, we obtained a higher performance in classification with the proposed MV-DNNet model. This prove the effectiveness of the incorporation of the deep multi-view representation in the classification task. However, in the rest of classification methods, there are still some mistakes in different thematic maps.

Table 7 reports the obtained classification results for the Pavia University HSI. From this table, we can see that the OA with MV-DNNet is 99.16%, AA is 99.04%, and k is 99.07%. The MV-DNNet model gives better classification rates for all the 9 classes of Pavia University dataset. Also, most classification rates OAs are greater than 98%. In Fig. 12, we compare also the obtained classifica-

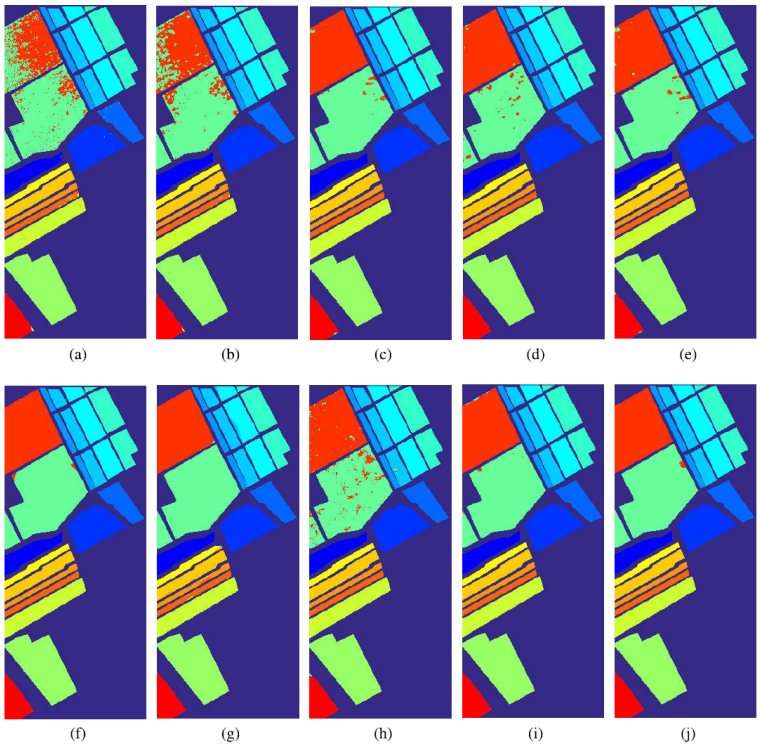

Figure 11: Classification maps for the Salinas HSI obtained by (a) DSVM, (b) SAE, (c) CNN, (d) DCNN, (e) R-VCA, (f) SKCR, (g) GF, (h) 3DCNN, (i) F-3DCNN, and (j) MV-DNNet.

tion maps with MV-DNNet based on multi-view latent representation, i.e., fused both spectral and spatial features with those only relying on spectral and spatial features, and average or concatenated spectral and spatial features. We can notice that the MV-DNNet can extract effective features to perform the classification. In this case, multi-view deep representation learning as a spectralspatial feature extraction model also provides a better classification performance than the standard spectral and spatial classification methods.

According to the obtained classification results, we can state that the proposed method is more effective than many other state-of-the-art deep learning-based methods for spectral-spatial classification of HSI. Furthermore, our model can extract the relevant spectral and spatial features of HSI with limited labeled samples by preserving useful information simultaneously, and it also provides a good spectral-spatial classification by exploiting the SSGCN model.

Overall, we can explain the high performance of our MV-DNNet model compared to other classical methods of spectral-spatial classification of HSI, that the multi-view latent representation obtained simultaneously with the deep AE model, i.e. the shared representation is more effective as a relevant features for classification than a simple concatenation or average of the features. We have shown then the added value of deep multi-view representation learning in the classification of HSI compared to other existing models that only use spatial features, like the GF or spectral features like the SAE model. In terms of computation time, the proposed model MV-DNNet is quite fast compared to some deep learning-based methods, since the training of the multi-view latent representation requires much less time.

5. Conclusion

In this paper, we proposed a novel approach for spectral-spatial classification of HSI, called MV-DNNet, which is based on multi-view deep autoencoder (MVDAE) and semi-supervised graph convolutional network (SSGCN). The advantage of such an approach is that it works with very small number of labeled

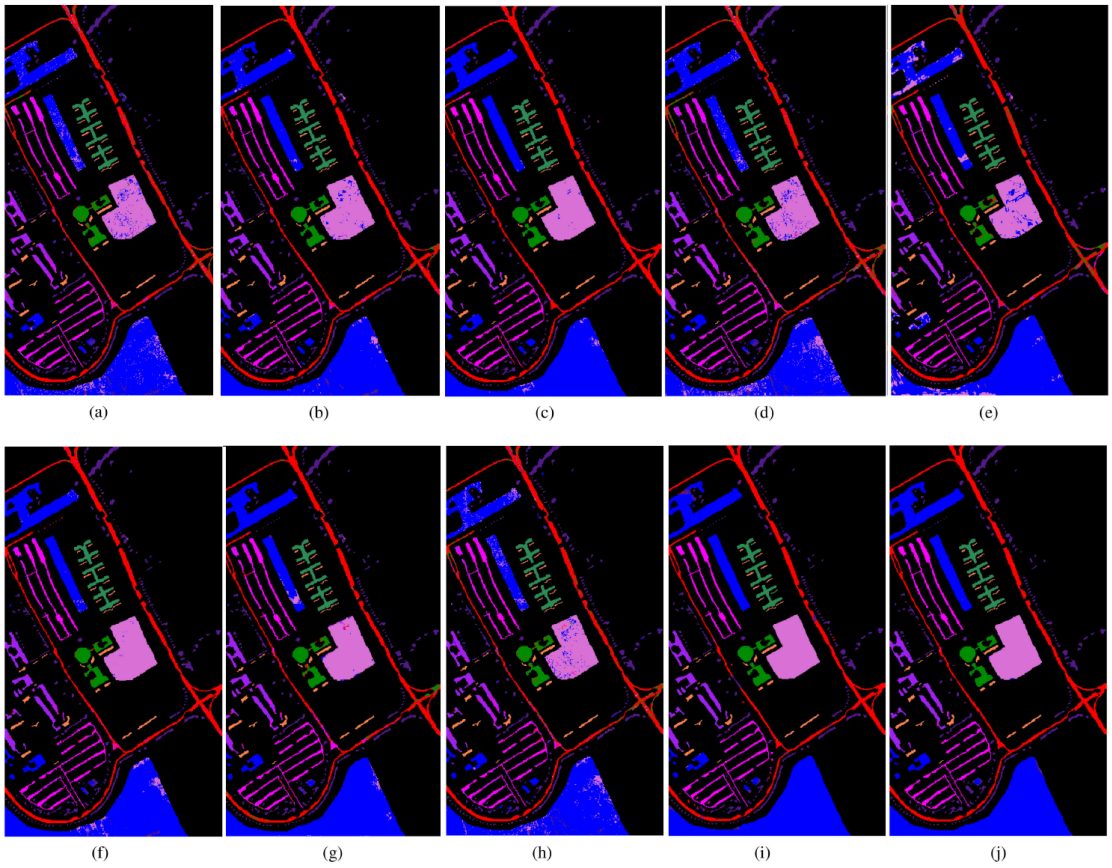

Figure 12: Classification maps for the Pavia University HSI obtained by (a) DSVM, (b) SAE, (c) CNN, (d) DCNN, (e) R-VCA, (f) SKCR, (g) GF, (h) 3DCNN, (i) F-3DCNN, and (j) MV-DNNet.

samples. Furthermore, the MVDAE model can extract relevant features, while preserving the useful spatial-spectral information for classification. Moreover, a SSGCN has been developed in order to preserve the spectro-spatial features of the multi-view latent representation. Finally, this approach has been used for the classification of HSI using few samples. Experimental results have shown that the proposed approach is more effective compared to other DL-based classification methods, including GCN, and CNN-based methods. As future work, our approach can be applied on other real hyperspectral data and extended to other field of applications where few labeled training samples are available.

References

[1] T. Qiao, Z. Yang, J. Ren, P. Yuen, H. Zhao, G. Sun, S. Marshall, J. A. Benediktsson, Joint bilateral filtering and spectral similarity-based sparse representation: a generic framework for effective feature extraction and data classification in hyperspectral imaging, Pattern Recognition 77 (2018) 316–328.

[2] J. Liang, J. Zhou, L. Tong, X. Bai, B. Wang, Material based salient object detection from hyperspectral images, Pattern Recognition 76 (2018) 476–490.

[3] Q. Zhao, S. Jia, Y. Li, Hyperspectral remote sensing image classification based on tighter random projection with minimal intra-class variance algorithm, Pattern Recognition 111 (2021) 107635.

[4] W. Li, F. Feng, H. Li, Q. Du, Discriminant analysis-based dimension reduction for hyperspectral image classification: A survey of the most recent advances and an experimental comparison of different techniques, IEEE Geoscience and Remote Sensing Magazine 6 (1) (2018) 15–34.

[5] H. Wu, S. Prasad, Semi-supervised dimensionality reduction of hyperspectral imagery using pseudo-labels, Pattern Recognition 74 (2018) 212–224.

[6] Z. Wang, N. M. Nasrabadi, T. S. Huang, Spatial–spectral classification of hyperspectral images using discriminative dictionary designed by learning vector quantization, IEEE Transactions on Geoscience and Remote Sensing 52 (8) (2014) 4808–4822.

[7] C. Deng, X. Liu, C. Li, D. Tao, Active multi-kernel domain adaptation for hyperspectral image classification, Pattern Recognition 77 (2018) 306–315.

[8] Y. Shao, N. Sang, C. Gao, L. Ma, Spatial and class structure regularized sparse representation graph for semi-supervised hyperspectral image classification, Pattern Recognition 81 (2018) 81–94.

[9] W. Sun, Q. Du, Graph-regularized fast and robust principal component analysis for hyperspectral band selection, IEEE Transactions on Geoscience and Remote Sensing 56 (6) (2018) 3185–3195.

[10] R. J. Johnson, J. P. Williams, K. W. Bauer, Autogad: An improved ica-based hyperspectral anomaly detection algorithm, IEEE Transactions on Geoscience and Remote Sensing 51 (6) (2013) 3492–3503.

[11] H. Huang, Z. Li, H. He, Y. Duan, S. Yang, Self-adaptive manifold discriminant analysis for feature extraction from hyperspectral imagery, Pattern Recognition 107 (2020) 107487.

[12] L. Cheng, L. Ma, W. Cai, L. Tong, M. Li, P. Du, Integration of hyperspectral imagery and sparse sonar data for shallow water bathymetry mapping, IEEE Transactions on Geoscience and Remote Sensing 53 (6) (2015) 3235–3249.

[13] A. Romero, C. Gatta, G. Camps-Valls, Unsupervised deep feature extraction for remote sensing image classification, IEEE Transactions on Geoscience and Remote Sensing 54 (3) (2016) 1349–1362.

[14] A. Sellami, I. R. Farah, High-level hyperspectral image classification based on spectro-spatial dimensionality reduction, Spatial Statistics 16 (2016) 103–117.

[15] M. G. Asl, M. R. Mobasheri, B. Mojaradi, Unsupervised feature selection using geometrical measures in prototype space for hyperspectral imagery., IEEE Trans. Geoscience and Remote Sensing 52 (7) (2014) 3774–3787.

[16] A. Sellami, M. Farah, I. R. Farah, B. Solaiman, Hyperspectral imagery semantic interpretation based on adaptive constrained band selection and knowledge extraction techniques, IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing 11 (4) (2018) 1337–1347.

[17] H. Yang, Q. Du, H. Su, Y. Sheng, An efficient method for supervised hyperspectral band selection, IEEE Geoscience and Remote Sensing Letters 8 (1) (2011) 138–142.

[18] S. Patra, P. Modi, L. Bruzzone, Hyperspectral band selection based on rough set, IEEE Transactions on Geoscience and Remote Sensing 53 (10) (2015) 5495–5503.

[19] Q. Wang, F. Zhang, X. Li, Optimal clustering framework for hyperspectral band selection, IEEE Transactions on Geoscience and Remote Sensing 56 (10) (2018) 1–13.

[20] M. Zhang, J. Ma, M. Gong, Unsupervised hyperspectral band selection by fuzzy clustering with particle swarm optimization, IEEE Geoscience and Remote Sensing Letters 14 (5) (2017) 773–777.

[21] Q. Liu, Y. Sun, R. Hang, H. Song, Spatial–spectral locality-constrained low-rank representation with semi-supervised hypergraph learning for hyperspectral image classification, IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing 10 (9) (2017) 4171–4182.

[22] Q. Wang, J. Lin, Y. Yuan, Salient band selection for hyperspectral image classification via manifold ranking, IEEE transactions on neural networks and learning systems 27 (6) (2016) 1279–1289.

[23] A. Sellami, A. B. Abbes, V. Barra, I. R. Farah, Fused 3-d spectral-spatial deep neural networks and spectral clustering for hyperspectral image classification, Pattern Recognition Letters 138 (2020) 594–600.

[24] C. Shi, C.-M. Pun, Superpixel-based 3d deep neural networks for hyperspectral image classification, Pattern Recognition 74 (2018) 600–616.

[25] Y. Li, W. Xie, H. Li, Hyperspectral image reconstruction by deep convolutional neural network for classification, Pattern Recognition 63 (2017) 371–383.

[26] W. Zhao, S. Du, Spectral–spatial feature extraction for hyperspectral image classification: A dimension reduction and deep learning approach, IEEE Transactions on Geoscience and Remote Sensing 54 (8) (2016) 4544–4554.

[27] M. Liang, L. Jiao, S. Yang, F. Liu, B. Hou, H. Chen, Deep multiscale spectral-spatial feature fusion for hyperspectral images classification, IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing 11 (8) (2018) 2911–2924.

[28] Z. Zhong, J. Li, Z. Luo, M. Chapman, Spectral–spatial residual network for hyperspectral image classification: A 3-d deep learning framework, IEEE Transactions on Geoscience and Remote Sensing 56 (2) (2018) 847–858.

[29] Y. Chen, X. Zhao, X. Jia, Spectral–spatial classification of hyperspectral data based on deep belief network, IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing 8 (6) (2015) 2381–2392.

[30] J. Li, X. Zhao, Y. Li, Q. Du, B. Xi, J. Hu, Classification of hyperspectral imagery using a new fully convolutional neural network, IEEE Geoscience and Remote Sensing Letters 15 (2) (2018) 292–296.

[31] S. Li, W. Song, L. Fang, Y. Chen, P. Ghamisi, J. A. Benediktsson, Deep learning for hyperspectral image classification: An overview, IEEE Transactions on Geoscience and Remote Sensing 57 (9) (2019) 6690–6709.

[32] Y. Chen, Z. Lin, X. Zhao, G. Wang, Y. Gu, Deep learning-based classification of hyperspectral data, IEEE Journal of Selected topics in applied earth observations and remote sensing 7 (6) (2014) 2094–2107.

[33] X. Ma, H. Wang, J. Geng, Spectral–spatial classification of hyperspectral image based on deep auto-encoder, IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing 9 (9) (2016) 4073–4085.

[34] P. Zhou, J. Han, G. Cheng, B. Zhang, Learning compact and discriminative stacked autoencoder for hyperspectral image classification, IEEE Transactions on Geoscience and Remote Sensing 57 (7) (2019) 4823–4833.

[35] G. Cheng, Z. Li, J. Han, X. Yao, L. Guo, Exploring hierarchical convolutional features for hyperspectral image classification, IEEE Transactions on Geoscience and Remote Sensing 56 (11) (2018) 6712–6722.

[36] G. Cheng, C. Yang, X. Yao, L. Guo, J. Han, When deep learning meets metric learning: Remote sensing image scene classification via learning discriminative cnns, IEEE transactions on geoscience and remote sensing 56 (5) (2018) 2811–2821.

[37] S. Wan, C. Gong, P. Zhong, B. Du, L. Zhang, J. Yang, Multiscale dynamic graph convolutional network for hyperspectral image classification, IEEE Transactions on Geoscience and Remote Sensing 58 (5) (2019) 3162–3177.

[38] Q. Liu, L. Xiao, J. Yang, Z. Wei, Cnn-enhanced graph convolutional network with pixel-and superpixel-level feature fusion for hyperspectral image classification, IEEE Transactions on Geoscience and Remote Sensing.

[39] X. Sun, F. Zhou, J. Dong, F. Gao, Q. Mu, X. Wang, Encoding spectral and spatial context information for hyperspectral image classification, IEEE Geoscience and Remote Sensing Letters 14 (12) (2017) 2250–2254.

[40] S. Mei, J. Ji, J. Hou, X. Li, Q. Du, Learning sensor-specific spatial-spectral features of hyperspectral images via convolutional neural networks, IEEE Transactions on Geoscience and Remote Sensing 55 (8) (2017) 4520–4533.

[41] X. Kang, C. Li, S. Li, H. Lin, Classification of hyperspectral images by gabor filtering based deep network, IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing 11 (4) (2017) 1166–1178.

[42] E. Aptoula, M. C. Ozdemir, B. Yanikoglu, Deep learning with attribute profiles for hyperspectral image classification, IEEE Geoscience and Remote Sensing Letters 13 (12) (2016) 1970–1974.

[43] T. N. Kipf, M. Welling, Semi-supervised classification with graph convolutional networks, arXiv preprint arXiv:1609.02907.

[44] A. Qin, Z. Shang, J. Tian, Y. Wang, T. Zhang, Y. Y. Tang, Spectral–spatial graph convolutional networks for semisupervised hyperspectral image classification, IEEE Geoscience and Remote Sensing Letters 16 (2) (2018) 241–245.

[45] M. Defferrard, X. Bresson, P. Vandergheynst, Convolutional neural networks on graphs with fast localized spectral filtering, in: Advances in Neural Information Processing Systems, 2016, pp. 3844–3852.

[46] O. Okwuashi, C. E. Ndehedehe, Deep support vector machine for hyperspectral image classification, Pattern Recognition 103 (2020) 107298.

[47] K.-K. Huang, C.-X. Ren, H. Liu, Z.-R. Lai, Y.-F. Yu, D.-Q. Dai, Hyperspectral image classification via discriminative convolutional neural network with an improved triplet loss, Pattern Recognition 112 (2021) 107744.

[48] B. Pan, Z. Shi, X. Xu, R-vcanet: A new deep-learning-based hyperspectral image classification method, IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing 10 (5) (2017) 1975–1986.

[49] B. Tu, C. Zhou, X. Liao, G. Zhang, Y. Peng, Spectral-spatial hyperspectral classification via structural-kernel collaborative representation, IEEE Geoscience and Remote Sensing Letters (2020) 861–865.

[50] L. Li, H. Ge, J. Gao, A spectral-spatial kernel-based method for hyperspectral imagery classification, Advances in Space Research 59 (4) (2017) 954–967.

AUTHOR BIOGRAPHY

Akrem Sellami :

Akrem Sellami received the Ph.D. degree in Signal, Image, Vision (SIV) from the University of Bretagne Loire, IMT Atlantique, Brest, France, in 2017. He is currently a Postdoctoral researcher at Lorraine Research Laboratory in Computer Science and its Applications (LORIA), University of Lorraine and INRIA/CNRS, Grand Est Nancy, France. He was a postdoc at Qarma (Machine Learning) team, LIS Lab, Aix-Marseille University. Since Oct. 2017, he was ATER at Paris Descartes University. His main research include graph deep learning, multi-view representation learning, dimensionality reduction, hyperspectral image classification, and image processing.

Salvatore Tabbone :

Salvatore Tabbone is professor in computer science at Universit de Lorraine (France) and the director of the Institute of Digital science, Management and Cognition (IDMC). He received the Ph.D. degree in Computer Science from Institut Polytechnique de Lorraine, France, in 1994 and the Habilitation degree in 2005. From 2007 to 2016, he was the Leader of QGAR team at LORIA Laboratory. Since 2010, he is the president of the scientific french assocation GRCE and he has been the leader of several national and international projects funded by French and European institutions. He is author/co-author of more than 100 articles in refereed journal and conferences. He serves as PC member for several international and national conferences. His area is on pattern recognition methods to compute useful features for image and document indexing. His main contributions are in the area of algorithms and methods for image and shape recognition.

Declaration of interests

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.