Morphological Convolutional Neural Networks for Hyperspectral Image Classification

Swalpa Kumar Roy, *Student Member, IEEE,* Ranjan Mondal, Mercedes E. Paoletti, *Senior Member, IEEE,* Juan M. Haut, *Senior Member, IEEE,* and Antonio Plaza, *Fellow, IEEE*

Abstract—**Convolutional neural networks (CNNs) have become quite popular for solving many different tasks in remote sensing data processing. The convolution is a linear operation which extracts features from the input data. However, nonlinear operations are able to better characterize the internal relationships and hidden patterns within complex remote sensing data, such as hyperspectral images (HSIs). Morphological operations are powerful nonlinear transformations for feature extraction that preserve the essential characteristics of the image, such as borders, shape and structural information. In this paper, a new end-to-end morphological deep learning framework (called `MorphConvHyperNet`) is introduced. The proposed approach efficiently models nonlinear information during the training process of HSI classification. Specifically our method includes spectral and spatial morphological blocks to extract relevant features from the HSI input data. These morphological blocks consist of two basic 2D morphological operators (erosion and dilation) in the respective layers, followed by a weighted combination of the feature maps. Both layers can successfully encode the nonlinear information related to shape and size, playing an important role in classification performance. Our experimental results, obtained on five widely used HSIs, reveal that our newly proposed `MorphConvHyperNet` offers comparable (and even superior) performance when compared to traditional 2D and 3D CNNs for HSI classification. The source code is publicly available at https://github.com/mhaut/MorphConvHyperNet.**

Index Terms—**Deep learning (DL), morphological transformations, latent feature space transfer, convolutional neural networks (CNNs), hyperspectral images (HSIs), classification.**

I. INTRODUCTION

HYperspectral images (HSIs) contain rich spectral and spatial information comprised by hundreds of highly-correlated and near-contiguous spectral bands, which are simultaneously captured over an observation area (along different wavelengths within the electromagnetic spectrum). The large and rich information contained HSI data cubes has been successfully exploited in many different applications, such as environmental management, surveillance, precision agriculture and crop analysis, among others.

S.K.Roy is with the Computer Science and Engineering at Jalpaiguri Government Engineering College, 735102, India (email: swalpa@cse.jgec.ac.in).

R. Mondal is with the Electronics and Communication Sciences Unit at Indian Statistical Institute, Kolkata 700108, India (email: ranjan15_r@isical.ac.in).

J.M. Haut is with the Department of Communication and Control Systems, Higher School of Computer Engineering, National Distance Education University, 28015 Madrid, Spain (e-mail: juanmariohaut@unex.es).

M. E. Paoletti and A. Plaza are with the Hyperspectral Computing Laboratory, Department of Technology of Computers and Communications, Escuela Politécnica, University of Extremadura, 10003 Cáceres, Spain (email: mpaoletti@unex.es, aplaza@unex.es).

Classification is an important technique for HSI data exploitation. The goal of classification is to assign a land-cover class to each spectral pixel by analyzing both the spectral and spatial information contained in the image. However, HSI classification poses two main challenges: the large spatial variability of the pixel-based spectral signatures, and the lack of available labeled samples. These challenges aggravate the curse of dimensionality problem, which hinders the training of any supervised algorithm and prevents the achievement of desirable performance levels, i.e., the obtained classification accuracies may not always be satisfactory.

To tackle the above problems, deep learning (DL) has received a lot of attention in the field of HSI classification [1], [2]. Initially, stacked autoencoders (SAEs) [3] and deep belief networks (DBNs) [4] were introduced as accurate unsupervised methods to extract layer-wise trained deep features. However, their standard fully-connected architecture imposes a feature flattening process before the classification, leading to the loss of spatial-contextual information. On the contrary, convolutional neural networks (CNNs) are able to automatically extract spectral-spatial features from the raw input data through a series of linear transformations (combined with nonlinear activations) to facilitate the recognition of patterns. In fact, the stack of convolutions layers is inspired by the natural visual cortex, where the spatial dimensions of the convolution kernel define the receptive field, identifying the presence of certain features and refining the feature extraction procedure along the entire stack. In this sense, the convolution kernel can be easily adopted to conduct HSI data analysis [5].

For instance, Bera and Shrivastava [6] explored the performance of the CNN model considering different optimizers. Similarly, Paoletti *et al.* [7] analyzed the impact of the input spatial size on model accuracy. Ramamurthy *et al.* [8] conducted image denoising and dimensionality reduction by combining autoencoders and CNNs. Also, Zhao *et al.* explored 2D CNNs [9] to extract spatial information from reduced HSI data using principal component analysis (PCA) for classification, but failed to exploit the entire range of spectral information contained in the HSI. To overcome this issue, Konstantinos *et al.* introduced 2D CNNs to extract spectral-spatial information separately, combining these two sources of information to improve classification performance [10]. This is due to the convolution operation (the basic building block of the CNN), which simply computes a linear combination of the input, followed by an activation function to introduce non-linearity in the feature learning process. Zhong *et al.* [11] introduced efficient spectral and spatial residual blocks to

extract discriminative features for HSI classification. Roy *et al.* [12] introduced a new hybrid deep model which combines 3D and 2D convolutions to improve spatial–spectral feature representation. Furthermore, Roy *et al.* [13] combined convolution kernels with generative adversarial minority oversampling to enhance the model performance by addressing the imbalanced data challenge imposed by HSI classification. Wang *et al.* [14] proposed an end-to-end cubic CNN, which applies convolutions in different directions of the feature volume to fully exploit spatial and spatial–spectral features. Driven by the goal of extracting and exploiting the best possible features, Alipour *et al.* [15] and Roy *et al.* [16] explored new architectural designs to make the convolutional kernel more flexible.

However, in order to fully exploit the spatial-contextual information contained in the HSI, the shapes and contours of the border regions should be well-preserved when extracting features (by keeping their geometry unchanged). Despite the success of previous works focused on capturing high-level features, complex spatial features and relationships can be missed in this context due to the kernel and subsequent pooling operations. In this context, morphological operators have been widely used to address the aforementioned issues and better capture the spatial-contextual information [17], [18]. In the following, a number of related works combining morphological operations and deep architectures are presented.

A. Related works

A constant critique of deep networks is their feature extraction process, which is completely opaque. Indeed, kernels self-adjust through the forward-backward procedure, without any control over what features they are extracting. In this context, several efforts have been conducted to open the black box and provide an interpretation of the extracted features, some of them inspired by morphological operators. In these operators, the structuring element (SE) plays an important role and helps to preserve the semantic meaning of structures according to the size and shape of the SE. For instance, Shen *et al.* attempted to learn both the SE and the morphological operations. They have also defined the residual morphological neural network with the help of subtraction of dilation and erosion operation [19]. Furthermore, it has been proved that nonlinear functions can better capture the intrinsic structure of abstract features [20]. In this context, the intrinsic linear combination operations within the CNN model can be replaced by nonlinear morphological operations to reduce the number of activation functions, while maintaining (or even increasing) the performance of the model. In addition, Mellouli *et al.* have defined a soft version of dilation and erosion using Counter-Harmonic Mean (CHM), validating the method in digits recognition, where the proposed CHM-based layer achieved higher performance than conventional models [21]. Also, Nogueira *et al.* [22] conducted an extensive study on the combination of deep models and multiple morphological operations such as opening, closing, top-hat operations, which have been combined with CNN to perform classification task on aerial images.

Mathematical morphology (MM) is well-known for its capacity to analyze and recover specific structures within

images using combinations of nonlinear filtering operations, such as dilation ($\oplus$) and erosion ($\ominus$) [23]. MM operations have been successfully applied in many areas of computer vision, such as feature extraction (FE), semantic image segmentation, denoising and edge detection, among others [24]. Traditionally, standard methods for HSI classification consist of two stages: FE and classification, using for instance support vector machines (SVM) [25]. Commonly used FE techniques based on MM are morphological attribute profiles (MAPs) [26], morphological profiles (MPs) [27], derivatives of morphological profiles (DMPs) [28], and extended morphological profiles (EMPs) [29]. These FE approaches have been widely used in the HSI research domain, normally by reducing the HSI to a few representative components using, for instance, PCA. Franchi *et al.* [30] introduced a morphological pooling layer similar to convolutional max-pooling. The resulting network is used for image de-noising and edge detection. The great success of MM features has inspired us to design a completely new approach for HSI classification which combines the dilation and erosion operations with the conventional CNN in a layered fashion design, without increasing model complexity (in terms of trainable weight parameters) [30]. In fact, two new blocks have been designed to improve the FE procedure conducted by the deep CNN model:

- On the one hand, a newly designed `SpectralMorph` block implements a dual-path module, where the first path applies the erosion operation over the data and the second path performs the dilation operation. The obtained features are processed along the channel dimension by two lightweight 1×1 convolution layers, and then combined to obtain the final `SpectralMorph` feature maps.
- On the other hand, the `SpatialMorph` block also implements a dual-path module. However, the obtained (eroded and dilated) features are processed along the spatial dimension by two 3×3 convolution layers, and then combined to obtain the final feature maps.

To the best of our knowledge, this is the first time in the literature that morphological operators such as dilation and erosion are integrated into the conventional CNN architecture for extracting structural information and classifying HSIs in an end-to-end fashion.

The remainder of the paper is structured as follows. Section II provides the architectural details of the proposed model, detailing the pre-processing step and the considered morphological operations. Section III discusses the experimental results obtained with several HSI classifiers and the proposed method, using five HSI datasets. Finally, Section IV concludes paper with some remarks and hints at plausible future research lines.

II. PROPOSED CLASSIFICATION FRAMEWORK

In the following, we provide the details of our new deep morphological CNN model for remote sensing HSI data classification. Our model is composed of two trainable morphological blocks called `SpectralMorph` and `SpatialMorph`. They respectively implement a spectral and a spatial conv2D operation followed by 2D erosion and dilation. The overall
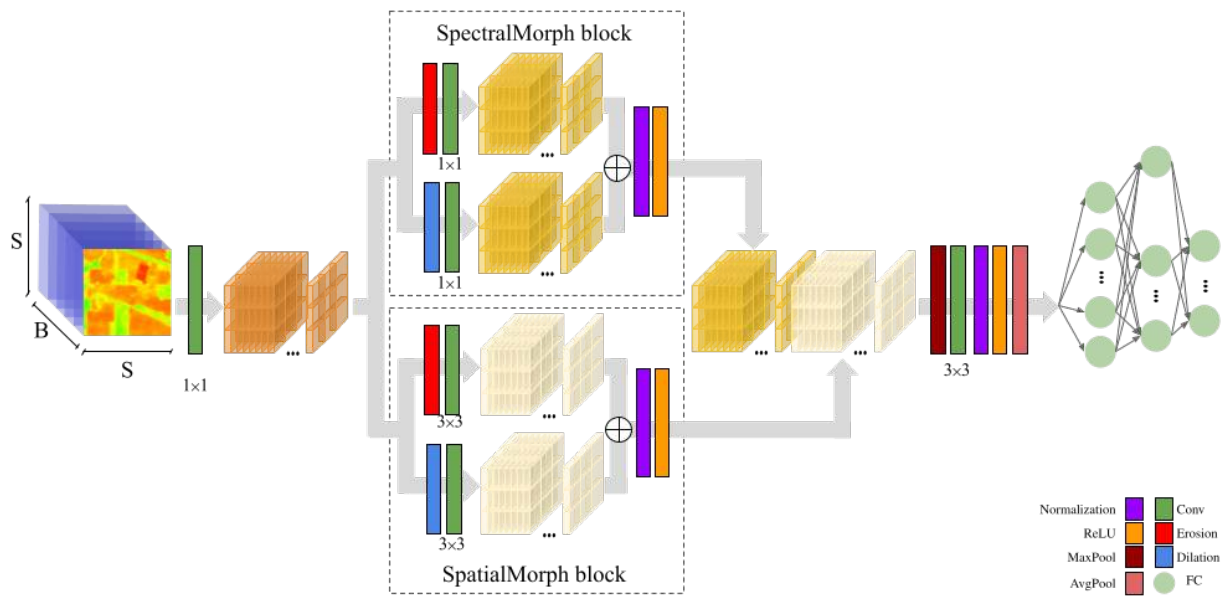
Fig. 1. Graphical overview of the proposed mophological convolutional network (`MorphConvHyperNet`) for spectral-spatial HSI classification.

architecture of our model is shown in Fig. 1. The flow of the proposed framework involves 3D HSI patch extraction, morphological operations and automatic selection of morphological features using 2D erosion and dilation. In the following, we describe these stages in detail.

A. Pre-processing of HSI data

A spectral-spatial HSI can be represented by a 3D tensor with dimensions W (width), H (height) and B (channels), defined as the data cube $\mathbf{X} \in \mathbb{R}^{H \times W \times B}$. The range of spectral features should be standardized to prevent features with higher variances (or wider ranges) from dominating the deep model optimization metrics. In this regard, the data are standardized by removing the mean and scaling to unit variance as a pre-processing step, forcing all features to contribute equally to the model performance. For this purpose, the Z-score method is implemented following Eq. (1):

$$\hat{\mathbf{x}}_{i,j} = \frac{\mathbf{x}_{i,j} - \mu}{\sigma}, \tag{1}$$

where $\mathbf{x}_{i,j} \in \mathbb{R}^D = [x_{i,j,1}, \ldots, x_{i,j,D}]$ is a pixel from $\mathbf{X}$, $\forall i \in [1, H]$, $j \in [1, W]$ and $\hat{\mathbf{x}}_{i,j} \in \mathbb{R}^D$ defines its standardized counterpart, with zero mean and unit standard deviation. μ and σ are the mean and the variance, respectively.

Hereinafter and with the aim of simplifying the nomenclature, when we refer to as $\mathbf{X}$, we mean the reduced data cube that has been standardized. Then, to capture both spectral and spatial information, $\mathbf{X}$ is cropped into overlapping 3D input patches of size $\mathbf{x}_{i,j} \in \mathbb{R}^{S \times S \times B}$, $\forall i \in [1, H]$, $j \in [1, W]$, with a stride of 1. Finally, the obtained patches $\mathbf{x}_{i,j}$ are sent to the neural model to be processed, so that every position (i, j) needs to be associated with one of L land-cover classes defined in advance.
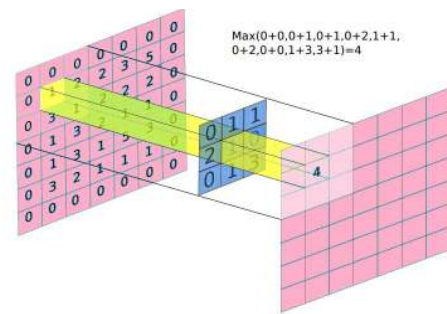

Fig. 2. Graphical visualization of the dilation operation. An input image patch of size (7×7×1) is dilated with a structuring element (SE) of size (3×3×1) and produces same output size if padded.
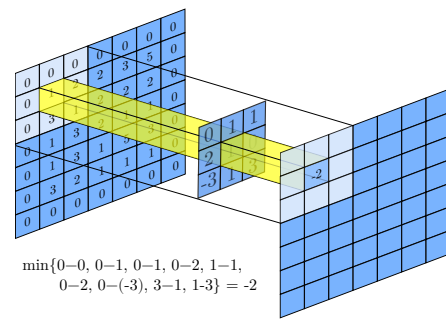

Fig. 3. Graphical visualization of the erosion operation. An input image patch of size (7×7×1) is eroded with a structuring element (SE) of size (3×3×1), removing the irrelevant data.

B. Morphological operations

Morphological operations are very powerful in terms of capturing the shape and size of objects in the image. In this work, a deep network based on two elementary morphological operations is proposed. In particular, morphological dilation and erosion operations are considered. Let $\mathbf{I} \in \mathbb{R}^{M \times N \times C}$ be

an intermediate feature map extracted from the HSI data, with spatial size $M \times N$ and C channels. The dilation ($\oplus$) and erosion ($\ominus$) operations over the feature map centered at spatial location (i, j) can be defined as follows:

$$(\mathbf{I} \oplus \mathbf{S}^d)(i,j) = \max_{(\hat{i},\hat{j},\hat{k}) \in U} (\mathbf{I}_{i+\hat{i},j+\hat{j},\hat{k}} + S^d_{i,\hat{j},\hat{k}}), \qquad (2)$$

$$(\mathbf{I} \ominus \mathbf{S}^e)(i,j) = \min_{(\hat{i},\hat{j},\hat{k}) \in U} (\mathbf{I}_{i+\hat{i},j+\hat{j},\hat{k}} - \mathbf{S}^e_{i,\hat{j},\hat{k}}), \qquad (3)$$

where $U = \{ (\hat{i}, \hat{j}, \hat{k}) \mid \hat{i} \in \{1, 2, 3, .., M\}; \hat{j} \in \{1, 2, 3, .., N\}; \hat{k} \in \{1, 2, 3, .., C\}\}$, and $\mathbf{S}^d$ and $\mathbf{S}^e$ are the SEs for the dilation and erosion operations, respectively.

Fig. 2 depicts the dilation operation with an SE of size 3×3. It may be noted that the dilation and erosion operations are nonlinear and piece-wise differentiable. A gray scale image can be viewed as a surface over the image plane. The dilation will increase the surface of the particular feature in the feature space (according to the size and shape of the SE). Similarly, the erosion will suppress the feature in the surface. Fig. 3 provides a graphical representation of the erosion operation For HSIs, morphological erosion and dilation can be applied in band-by-band fashion.

C. SpectralMorph and SpatialMorph Blocks

Dilation and erosion are shape-sensitive operations. This property is quite helpful to extract discriminative spatial-contextual information during the training stage. In this context, we designed our network using nonlinear MM filters. As mentioned in [20], a single-layer dilation or erosion (followed by a linear combination) can be used for complex classification tasks. Dilation and erosion operations on morphological feature maps generate dilated and eroded feature maps. To combine the resulting feature maps we can take a linear combination of these maps as follows:

$$\mathbf{I}^2_{(i,j)} = b + \sum_{k=1}^{C} w_k \mathbf{I}^2_{(i,j,k)}, \qquad (4)$$

where the feature maps of $\mathbf{I}_1$ are combined linearly in order to generate $\mathbf{I}_2$. The linear combination can be viewed as a 1×1 convolution. To generate additional features, we may apply multiple dilation/erosion operations (and generate multiple linear combinations of dilation and erosion).

In this work, shape features have been incorporated by introducing trainable (and therefore, learnable) MM operations, i.e. dilation and erosion into the conventional CNN model. Consequently, we have defined two separate morphological blocks:

- First, the SpectralMorph is built as a spectral morphological block, which comprises two parallel MM operations (dilation or erosion) followed by a linear combination of dilated and eroded feature maps. We further add (in element-wise fashion) the resulting feature maps. Fig. 1 provides the graphical representation of the SpectralMorph block.
- Similarly, 3×3 convolutions can be used instead of linear combinations of feature maps. This helps to extract spatial

features from dilated and eroded feature maps. In this case, we call the resulting block SpatialMorph (see Fig. 1).

Before training, for all the considered SEs, the weights of the linear combination and the convolution weights are initialized randomly. In particular, network weights have been set through He's operator [31], the well-known variance-scaling initializer which enhances the network performance when ReLU activation functions are implemented. Moreover, biases have been set to 0. For simplicity, in this work we design a network that focuses on simple nonlinear MM operations (erosion and dilation) and explore their performance in the context of HSI classification. However, more sophisticated MM operations such as opening, closing, reconstruction-based operations, etc., can also be included in future developments. In the following subsection we describe in details the adopted network configuration.

D. Proposed Morphological CNN

Our morphological CNN comprises several convolution and morphological layers. In each layer, we have taken multiple convolutions and dilation/erosion operations to generate multiple feature maps. Fig. 1 illustrates the overall architecture of the proposed network. A layer-wise detailed summary of the proposed model is provided in Table I.

As it can be seen, for the design of the proposed network we have employed two symmetric morphological blocks designated as SpectralMorph and SpatialMorph. These blocks are intended to extract MM features from the feature maps. It should be noted that the dilation and erosion operations may also work as redundant layers. For example, a dilation operation using a SE of size 3×3 with a center element set to 0 and all other elements set to $-\inf$ propagates the input to the next layer without changing the value of the input. As a result, multiple dilation or erosion operations help to generate multiple morphological feature maps.

In our experiments, we have considered $B/4$ dilations and $B/4$ erosions in each block. It should also be noted that dilation and erosion are based on $\min / \max$ operations, so they may produce many zero gradient values while conducting the back-propagation step. To boost the gradient, we use a convolution layer to specifically enhance the desired output of each operation. Then normalization and activation functions are applied to each block. The obtained feature maps are then concatenated and processed by a spatial down-sampler, i.e. a MaxPool layer, followed by a stack of convolution, normalization and activation layers. Finally, a global average pooling is applied to reshape the data into vectors suitable for the processing of fully connected (FC) layers, which are used for classification purposes.

The SEs and the convolutional kernels are all 3×3 pixels in size, and all of them are initialized randomly before training. The network is trained in end-to-end fashion using the back-propagation algorithm. In the following section we quantitatively and qualitatively verify the performance of our network using real HSI data. We also provide an ablation study to validate if morphological layers help extracting features that further contribute to the final classification performance.

TABLE I
ARCHITECTURAL DETAILS OF OUR MORPHCONVHYPERNET MODEL

Layer ID	Kernel/Neurons	BatchNorm	Act. function	Stride
CNN	$B/4 \times 1 \times 1 \times B$	No		1
S-MORPH	$B/4$	Yes	ReLU	1
SP-MORPH	$B/4$	Yes	ReLU	1
MAX pool	2×2			1
CNN	$128/2 \times 3 \times 3 \times B/2$	Yes	ReLU	1
AVG pool	8×8			8
FC1	$n_{classes}$	No	Softmax	

Fig. 4 visualizes different feature maps extracted from several input samples at particular stages of the network architecture. As we can see, although the feature maps obtained by the initial convolution layer are quite smooth, the SpectralMorph and SpatialMorph blocks are able to extract valuable information through their erosion and dilation paths, which is combined to obtain the final output. As a result, the block outputs contain rich information, improving the data representation by means of highly discriminating information.
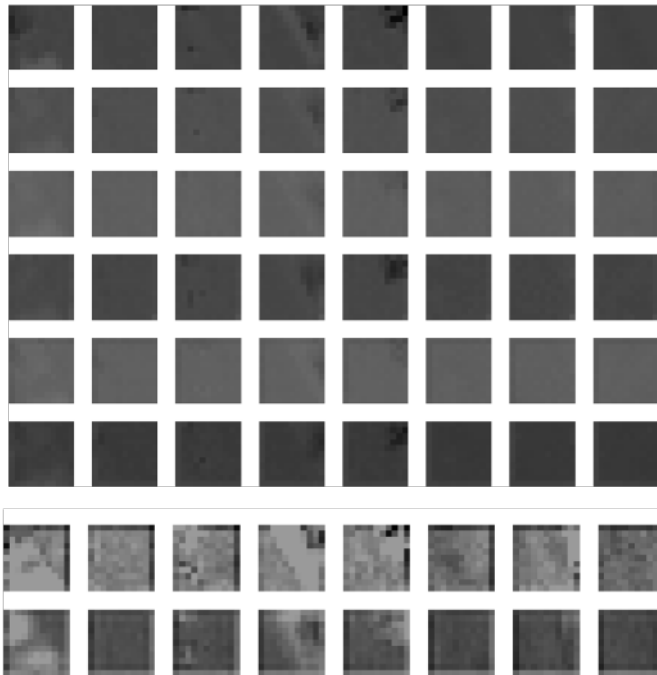

Fig. 4. Graphical visualization of the obtained features. Each column indicates a different sample, whilst each row provides the obtained feature obtained after being processing through the different filters and blocks of the network. In this sense: row 1 provides the original sample; rows 2 provide the extracted feature maps after the first convolution layer; rows 3 and 4 provide the extracted feature maps after spectral erosion and dilation; rows 5 and 6 provide the extracted feature maps after spatial erosion and dilation, and finally rows 7 and 8 provide the feature maps obtained by the SpectralMorph block and the SpatialMorph block, respectively.

III. EXPERIMENTAL RESULTS

A. HSI datasets

In order to evaluate the performance of the proposed MorphConvHyperNet, five different HSI scenes[1] have been
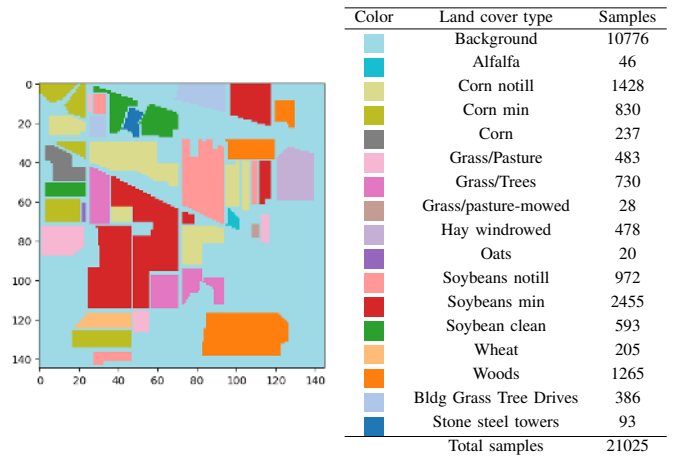
[1] Available online: http://dase.grss-ieee.org/

Color	Land cover type	Samples
	Background	10776
	Alfalfa	46
	Corn notill	1428
	Corn min	830
	Corn	237
	Grass/Pasture	483
	Grass/Trees	730
	Grass/pasture-mowed	28
	Hay windrowed	478
	Oats	20
	Soybeans notill	972
	Soybeans min	2455
	Soybean clean	593
	Wheat	205
	Woods	1265
	Bldg Grass Tree Drives	386
	Stone steel towers	93
	Total samples	21025

Fig. 5. The ground truth, type associated with the land-cover classes, and number of available samples in the Indian Pines (IP) dataset.

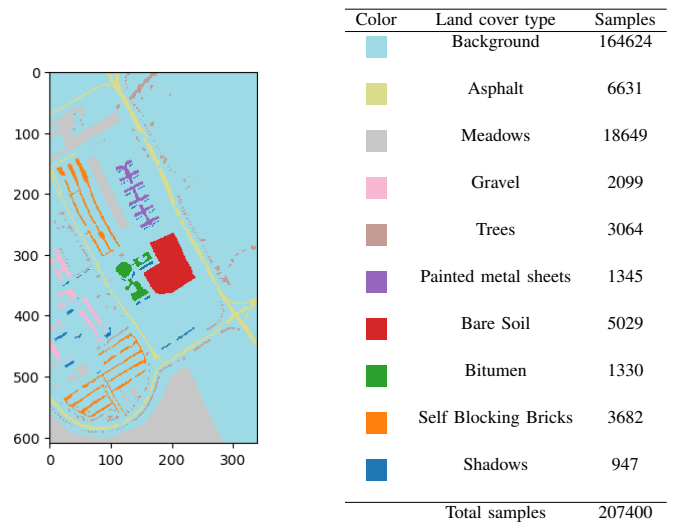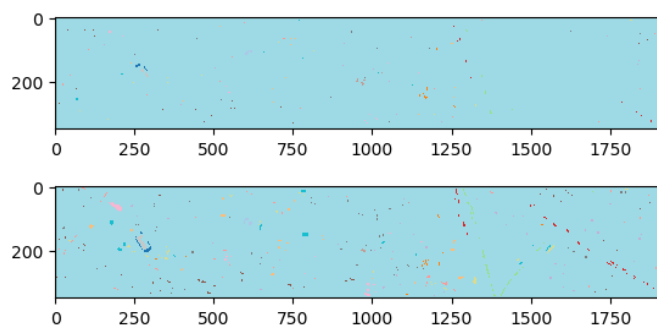| Color	Land cover type	Samples
	Background	164624
	Asphalt	6631
	Meadows	18649
	Gravel	2099
	Trees	3064
	Painted metal sheets	1345
	Bare Soil	5029
	Bitumen	1330
	Self Blocking Bricks	3682
	Shadows	947
	Total samples	207400

Fig. 6. The ground truth, type associated with the land-cover classes, and the number of available samples in the University of Pavia (UP) dataset.

considered: Indian Pines (IP), University of Pavia (UP), University of Houston (UH), Salinas Valley (SV) and Bostwana (BW) scenes. Figs. 5, 6, and 7 respectively show a detailed summary of the IP, UP and UH scenes, including their corresponding ground truth, the type associated with the land-cover classes, and the number of available labeled samples per class. In the following, we describe the datasets considered in this paper:

- The **Indian Pines** (IP) dataset was gathered by the Airborne Visible/Infrared Imaging Spectrometer (AVIRIS) [32] over the Indian Pines test site in North-western Indiana. It contains 224 spectral bands within a wavelength range of 400 to 2500 nm. The 24 null and corrupted bands have been removed. The spatial size of the image is 145×145 pixels, and it comprises of 16 mutually exclusive vegetation classes. The spatial resolution is 20 meters per pixel (mpp).

- The **University of Pavia** (UP) dataset was acquired by the Reflective Optics System Imaging Spectrometer (ROSIS) sensor during a flight campaign over the university campus at Pavia, Northern Italy [33]. It consists of 610×340 pixels

(a) Disjoint Train   (b) Disjoint Test

Fig. 8. Spatially disjoint training and test samples for the IP dataset (DIP dataset).

Color	Land cover type	Samples train	Samples test
	Background	662013	652648
	Grass-healthy	198	1053
	Grass-stressed	190	1064
	Grass-synthetic	192	505
	Tree	188	1056
	Soil	186	1056
	Water	182	143
	Residential	196	1072
	Commercial	191	1053
	Road	193	1059
	Highway	191	1036
	Railway	181	1054
	Parking-lot1	192	1041
	Parking-lot2	184	285
	Tennis-court	181	247
	Running-track	187	473
	Total samples	664845	664845

Fig. 7. The ground truth, type associated with the land-cover classes, and the number of available samples in the University of Houston (UH) dataset, where the image at the top is the training set and the image at the bottom is the testing set.

with 103 spectral bands in the wavelength range from 430 to 860 nm and 2.5 mpp. It comprises 9 urban land-cover classes.

- The IEEE Geoscience and Remote Sensing Society published the **University of Houston** (UH) dataset –collected by the Compact Airborne Spectrographic Imager (CASI)– in 2013 [34], as part of its Data Fusion Contest. It is composed of 340×1905 pixels with 144 spectral bands. The spatial resolution of this dataset is 2.5 mpp with a wavelength ranging from 0.38to 1.05 μm. Finally, the ground truth comprises 15 different land-cover classes

- The **Salinas Valley** (SV) dataset was acquired using AVIRIS sensor over an agricultural area on Salinas Valley, California. It contains 512×217 pixels with 224 spectral bands. For classification purpose, 20 absorption and noise bands were removed (108th - 112th, 154th - 167th and 224th). The spatial resolution is 3.7 mpp, and the ground-truth considers 16 different land-cover classes.

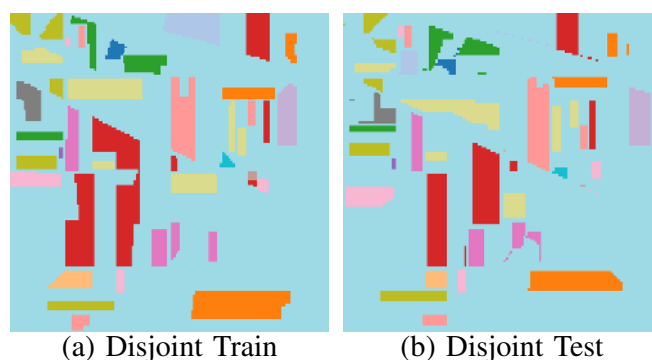- The **Botswana** (BW) dataset was collected using the Hy-

perion instrument aboard the NASA EO-1 satellite, which captured the scene on the Okavango Delta, Botswana. The spatial resolution of this dataset is 30 mpp, with 1497×256 pixels. It contains 145 spectral bands with a wavelength range between 400-2500 nm. Before the classification task, 97 uncalibrated and water-corrupted bands were removed. The ground-truth contains 14 land-cover classes.

In addition, several experiments have been conducted using the Disjoint Indian Pines (DIP) scene and the Disjoint University of Pavia (DUP) image. This addresses an important drawback associated with 2D/3D kernel-based models and the spatially overlapping data. Indeed, CNNs are based on neighbourhood windows, which must be extracted as cropped windows from the HSI scene. In a clear contrast to other remote sensing applications, this represents a serious limitation within HSI processing, as the same HSI scene is used for extracting both training and test samples. As a result, when extracting neighbourhood windows for the training samples, these windows may overlap the test information, including it in the training stage. This may unfairly benefit the model, which will provide overrated classification results. To overcome this drawback, recent HSI classification work encourages the use of both disjoint and random sampling strategies, providing a trade-off between these different approaches. Inspired by these works, DIP, DUP and DUH have been taken into account from the IEEE GRSS Data and Algorithm Standard Evaluation (DASE) website. These disjoint datasets have mutually exclusive training and test samples (as the ground-truth for the UH dataset), i.e., there is no spatial overlapping between the training and test data. These spatial disjoint. Figs. 8 and 9 show the disjoint splits given for the IP and UP datasets, respectively. Moreover, Table II details the number of pixels per class. It is noteworthy that disjoint train-test sets can be even more challenging compared to randomly selected training and test samples, as they do not ensure class balance. In fact DIP and DUP have highly class imbalanced issues, which is very interesting to test the robustness of the proposed model. The disjoint training and test splits for the UH dataset are given in Fig. 7.

Finally, to better illustrate the performance of the proposed model on random train-test splits, the BW and SV datasets are used only when labeled samples are available. Here, sample patches are extracted from the raw HSI data and random
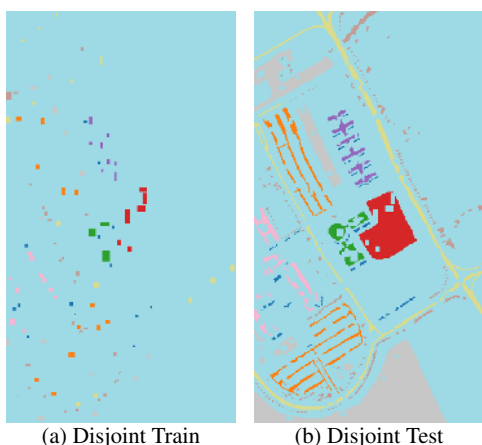
(a) Disjoint Train   (b) Disjoint Test

Fig. 9. Spatially disjoint training and test samples for the UP dataset (DUP dataset).

TABLE II
NUMBER OF TRAINING SAMPLES (TRS) AND TEST SAMPLES (TES)
OBTAINED FROM THE DISJOINT TRAIN-TEST IP AND UP DATASETS

Class	Disjoint IP		Disjoint UP	
	TRS	TES	TRS	TES
1	29	25	548	6304
2	762	675	540	18146
3	435	404	392	1815
4	146	99	524	2912
5	232	274	265	1113
6	394	354	532	4572
7	16	2	375	981
8	235	250	514	3364
9	10	10	231	795
10	470	503		
11	1424	1065		
12	328	282		
13	132	80		
14	728	545		
15	291	99		
16	57	44		

sampling is conducted to select the training samples, while the remaining labeled samples are used for testing.

B. Experimental settings

In order to validate the effectiveness of the proposed `MorphConvHyperNet`, a detailed comparison have been conducted between several state-of-the art models, which include classical machine learning and representative deep learning methods. These methods are available on [5][2]. In particular, the methods considered include: multinomial logistic regression (MLR) [35], SVM with radial basis function [25], gated recurrent unit (GRU) [36], long short term memory (LSTM) [37], CNN-1D [38], CNN-2D [5] and CNN-3D [39], [40]. Moreover, to feed spatial-based models, a spatial neighborhood of size 11×11 has been considered to create patches for all the HSI datasets. The network parameters are randomly initialized and trained with the Adam optimizer [41], using a learning rate of 0.001 while minimizing the widely used cross-entropy loss.

The classification performance of the proposed model has been evaluated considering four widely used quantitative

[2]https://github.com/mhaut/hyperspectral_deeplearning_review

metrics: per-class accuracy, overall accuracy (OA), average accuracy (AA), and kappa coefficient (κ) [42], respectively. In this sense, the ratio of correctly classified samples among the total test samples is determined by the OA, while the mean of class-wise accuracy is determined by the AA. Finally, the κ represents a strong mutual agreement between the generated classification maps of one network model and the provided ground truth. All the considered models have been run 5 times, using 200 epochs per iteration, and we collect and report the average results.

The hardware environment used for experiments is composed by an Intel i9-9940X processor with 128GB of DDR4 RAM, NVidia Titan RTX with 24GB of DDR4 RAM. The source code of our framework was implemented by using the Keras library with TensorFlow as the backend.

C. Classification Results over Disjoint Datasets

To illustrate the generalization ability of our newly proposed method in comparison with other traditional approaches, the DIP (Fig. 8), DUP (Fig. 9) and DUH (Fig. 7) datasets have been considered. As pointed out before, these datasets prevent spatial overlapping between training and test, and introduce certain challenges, as the different land cover classes are unbalanced in terms of number of samples (such as DIP and DUP).

In this regard, DIP is first considered. It contains class-specific imbalanced training samples. As shown in Fig. 5, class-wise data variation is particularly observed in two classes: "*Oats*" and "*Soybean-mintill*". These classes respectively contain 20 and 2455 samples. According to the experimental settings discussed in Section III-B, the quantitative results in terms of OA, AA, κ and per-class accuracy for the DIP dataset (using all the considered classification models) are reported in Table III, where we display in bold typeface the highest achieved results across all the compared classification methods. On the one hand, Table III reveals that the proposed `MorphConvHyperNet` model achieves superior performance in terms of OA, and κ. On the other hand, the highest AA is achieved by the SVM. It can also be seen that the CNN2D can provide better classification performance than CNN1D and CNN3D, achieving better class-specific accuracy for a few of the classes. Overall, the proposed `MorphConvHyperNet` framework consistently outperforms the traditional CNN1D, CNN2D and CNN3D by a large margin (particularly in comparison with CNN1D and CNN3D). The other classification models (i.e., MLP, LSTM, and GRU) achieve similar accuracy, whilst RNN and MLR provide the lowest OA values.

In a similar way, Table IV reports the results obtained for the disjoint UH dataset. As pointed before, the total number of class-wise training and test samples for the UH dataset are shown in Fig. 7. The results presented in Table IV reveal that the proposed network exhibits constant performance gains in all the considered measurements i.e., OA, AA, and κ with respect to CNN1D, CNN2D and all the other considered methods. It can also be observed that the OA achieved by the CNN1D is significantly better than that achieved by the

TABLE III
CLASSIFICATION RESULTS OBTAINED BY MLR, SVM, RNN, LSTM, GRU, CNN1D, CNN2D, CNN3D, AND MorphConvHyperNet ON THE DISJOINT TRAIN-TEST DATASET FOR THE IP SCENE (DIP).

Class	MLR	SVM	MLP	RNN	LSTM	GRU	CNN1D	CNN2D	CNN3D	MorphConvHyperNet
1	80.0±0.0	88.0±0.0	73.6±7.42	58.4±4.8	89.6±1.96	77.6±9.33	80.8±12.75	73.64±14.77	48.18±22.84	92.27±3.55
2	81.48±0.0	80.0±0.0	81.45±1.07	75.5±1.48	82.22±1.26	81.1±2.77	79.38±4.16	83.12±6.12	85.12±7.88	84.05±8.03
3	54.11±0.12	69.55±0.0	64.55±2.85	63.37±1.93	64.16±5.44	70.35±1.36	74.26±6.12	81.98±3.89	77.22±13.04	79.34±3.45
4	38.38±0.0	48.48±0.0	47.07±10.41	29.49±5.4	55.35±11.15	53.33±11.15	31.92±11.55	45.39±6.36	50.11±10.04	52.14±6.24
5	91.97±0.0	87.23±0.0	86.94±1.07	87.59±1.5	89.27±0.68	88.4±0.85	90.73±1.07	89.11±5.55	80.28±6.52	91.66±1.69
6	94.63±0.0	96.33±0.0	95.93±0.97	95.31±0.83	96.39±0.87	96.38±1.14	96.39±0.9	95.02±5.68	89.81±4.03	95.74±2.28
7	0.0±0.0	50.0±0.0	10.0±20.0	0.0±0.0	0.0±0.0	0.0±0.0	0.0±0.0	0.0±0.0	0.0±0.0	0.0±0.0
8	100.0±0.0	100.0±0.0	99.84±0.2	99.52±0.3	99.2±0.91	99.12±0.64	99.84±0.32	99.96±0.13	95.96±6.78	100.0±0.0
9	0.0±0.0	50.0±0.0	80.0±15.49	56.0±10.2	76.0±8.0	66.0±4.9	50.0±8.94	26.66±15.87	77.78±21.66	44.44±19.88
10	66.76±0.08	76.54±0.0	75.35±5.02	71.13±5.93	81.51±2.76	78.53±4.64	81.83±4.69	77.44±8.99	77.9±6.2	80.77±3.77
11	84.13±0.0	87.7±0.0	83.19±1.51	78.86±1.45	80.4±2.43	82.29±1.82	80.39±3.65	89.4±5.47	82.73±3.81	88.54±5.03
12	66.31±0.0	77.3±0.0	78.58±2.95	71.91±5.07	76.31±1.39	83.19±1.16	84.75±7.5	87.72±3.06	82.64±14.49	88.46±4.26
13	95.0±0.0	97.5±0.0	98.0±0.61	97.0±1.7	97.25±0.94	97.75±0.94	97.75±0.5	95.28±4.74	89.72±6.89	87.64±3.43
14	90.64±0.0	91.38±0.0	92.92±1.48	90.28±1.09	94.13±1.18	92.88±1.79	93.32±2.34	98.94±0.55	98.31±1.41	98.82±1.01
15	89.9±0.0	80.81±0.0	87.88±3.78	75.56±6.43	90.71±2.34	93.54±1.64	89.9±4.78	82.02±14.83	55.17±27.57	69.44±15.86
16	97.73±0.0	97.73±0.0	87.27±4.45	88.64±4.31	94.09±2.32	95.45±2.49	96.82±2.32	82.0±6.69	82.5±12.5	84.0±4.21
OA	80.33±0.02	84.12±0.0	82.95±0.23	79.07±0.33	83.55±0.39	84.2±0.21	84.0±0.28	87.25±1.03	83.6±1.41	87.45±1.01
AA	70.69±0.01	79.91±0.0	77.66±1.98	71.16±0.87	79.16±0.75	78.49±0.36	76.76±0.75	75.48±2.12	73.34±3.46	77.33±1.56
k(x100)	77.47±0.02	81.87±0.0	80.56±0.26	76.12±0.4	81.27±0.44	82.01±0.26	81.81±0.35	85.48±1.15	81.36±1.62	85.75±1.14

TABLE IV
CLASSIFICATION RESULTS OBTAINED BY MLR, SVM, RNN, LSTM, GRU, CNN1D, CNN2D, CNN3D, AND MorphConvHyperNet ON THE DISJOINT TRAIN-TEST DATASET FOR THE UH SCENE.

Class	MLR	SVM	MLP	RNN	LSTM	GRU	CNN1D	CNN2D	CNN3D	MorphConvHyperNet
1	82.24±0.06	82.34±0.0	81.23±0.28	82.22±0.28	82.76±0.36	82.58±0.35	82.28±0.98	82.25±0.65	82.1±0.39	82.43±0.33
2	82.5±0.07	83.36±0.0	82.29±0.55	82.87±0.33	80.19±1.36	81.64±0.71	91.78±6.46	84.15±0.28	84.14±0.45	84.42±0.19
3	99.8±0.0	99.8±0.0	99.72±0.1	99.72±0.2	99.68±0.16	99.88±0.1	99.92±0.16	90.31±4.41	77.85±4.8	97.21±1.23
4	98.3±0.0	98.96±0.0	87.58±1.28	93.5±2.02	91.23±1.29	93.22±2.84	94.36±3.12	87.24±3.21	89.24±1.1	92.37±0.33
5	97.44±0.0	98.77±0.0	97.35±0.49	97.76±0.29	97.65±0.31	97.37±0.16	98.77±0.13	99.51±0.48	98.97±0.59	99.77±0.53
6	94.41±0.0	97.9±0.0	94.55±0.28	95.1±0.0	97.06±1.79	98.32±1.63	95.8±1.88	96.43±2.14	98.91±1.44	99.46±1.15
7	73.37±0.07	77.43±0.0	75.24±2.27	81.4±0.43	78.88±1.0	77.03±2.18	82.78±2.23	86.44±2.18	85.48±1.98	88.07±1.78
8	63.82±0.0	60.3±0.0	57.0±6.97	40.06±1.07	40.11±1.92	53.62±2.97	75.5±6.71	70.03±3.96	62.06±3.01	73.09±3.5
9	70.23±0.04	76.77±0.0	75.58±2.86	76.54±2.96	81.55±4.12	79.06±1.61	81.44±2.0	79.53±6.38	80.81±4.32	84.09±2.73
10	55.6±0.0	61.29±0.0	48.78±2.27	47.44±1.44	47.37±2.29	49.54±2.61	68.71±14.55	60.22±4.2	54.75±4.63	62.86±3.08
11	74.21±0.04	80.55±0.0	76.25±0.46	76.24±0.81	76.38±1.09	80.82±0.71	85.24±2.83	82.93±7.68	66.78±3.34	89.15±6.86
12	70.41±0.0	79.92±0.0	75.31±3.75	76.33±3.09	79.98±3.32	84.15±3.13	89.93±4.29	92.87±3.31	93.83±1.92	93.02±3.32
13	67.72±0.0	70.88±0.0	73.19±2.15	69.12±1.61	71.37±3.54	72.63±3.68	74.88±5.14	86.21±2.65	82.34±2.49	89.61±1.34
14	98.79±0.0	100.0±0.0	99.84±0.32	100.0±0.0	99.11±0.47	99.92±0.16	99.68±0.16	98.92±1.8	96.31±3.67	99.19±1.3
15	95.56±0.0	96.41±0.0	97.8±0.51	97.59±0.47	98.14±0.31	98.22±0.59	98.48±0.24	77.63±2.91	75.85±2.69	97.04±4.47
OA	78.97±0.01	81.86±0.0	78.22±0.36	77.95±0.68	78.16±0.28	80.21±0.27	86.42±1.64	83.27±0.8	80.24±0.55	86.51±0.71
AA	81.63±0.01	84.31±0.0	81.45±0.37	81.06±0.55	81.43±0.32	83.2±0.27	87.97±1.38	84.98±0.74	81.96±0.75	88.78±0.68
k(x100)	77.3±0.01	80.43±0.0	76.55±0.39	76.23±0.71	76.52±0.3	78.66±0.29	85.27±1.77	81.89±0.86	78.62±0.59	85.4±0.76

TABLE V
CLASSIFICATION RESULTS OBTAINED BY MLR, SVM, RNN, LSTM, GRU, CNN1D, CNN2D, CNN3D, AND MorphConvHyperNet ON THE DISJOINT TRAIN-TEST DATASET FOR THE UP SCENE (DUP).

Class	MLR	SVM	MLP	RNN	LSTM	GRU	CNN1D	CNN2D	CNN3D	MorphConvHyperNet
1	77.68±0.0	82.23±0.0	84.53±1.89	83.08±3.3	82.63±2.39	77.25±6.92	87.18±2.11	93.4±1.89	85.66±4.0	94.52±1.9
2	58.79±0.01	65.81±0.0	75.13±2.4	67.9±2.92	78.74±1.99	80.1±5.12	89.64±2.53	96.84±1.93	95.88±1.71	97.12±1.08
3	67.21±0.02	66.72±0.0	68.37±5.17	65.17±7.99	60.73±11.0	54.79±14.82	71.1±5.98	65.48±13.94	68.11±6.47	85.08±4.53
4	74.27±0.05	97.77±0.0	93.5±2.32	90.72±2.56	97.1±1.22	92.05±2.31	95.32±1.49	95.55±2.14	97.02±0.83	97.0±1.03
5	98.88±0.04	99.37±0.0	99.37±0.08	99.23±0.09	99.28±0.08	99.51±0.12	99.48±0.26	98.03±0.92	98.9±0.56	99.25±0.22
6	93.53±0.02	91.62±0.0	89.94±4.14	85.07±3.14	65.94±5.92	74.86±11.38	88.28±2.33	80.52±9.39	68.85±11.29	93.92±3.88
7	85.08±0.05	87.36±0.0	87.2±3.05	82.94±3.79	84.95±4.02	90.17±3.9	86.77±3.38	89.29±9.48	73.09±9.53	84.98±10.74
8	87.58±0.01	90.46±0.0	90.37±1.24	85.85±4.97	88.89±7.83	90.42±4.39	90.43±3.34	94.5±5.44	95.21±1.69	96.62±2.21
9	99.22±0.05	93.71±0.0	98.44±1.17	94.52±4.79	98.29±1.47	93.51±7.93	97.33±3.31	95.8±0.76	93.54±1.76	97.05±0.46
OA	72.23±0.0	77.8±0.0	82.05±0.88	77.07±0.95	80.38±0.52	80.7±0.56	89.09±0.97	92.55±1.02	89.43±1.37	95.51±0.66
AA	82.47±0.01	86.12±0.0	87.43±1.03	83.83±0.72	84.06±0.74	83.63±2.03	89.5±1.03	89.94±1.37	86.25±1.98	93.95±0.96
k(x100)	65.44±0.0	72.06±0.0	76.89±1.07	70.84±1.04	74.32±0.68	74.76±1.02	85.5±1.22	89.9±1.42	85.61±1.94	93.95±0.88

CNN2D and CNN3D. Focusing on recurrent models, GRU outperforms the classification results obtained by the standard RNN and LSTM. MLP and MLR seem to perform very similarly, achieving the lowest accuracy results.

Finally, to determine the generalization power of the proposed MorphConvHyperNet model in highly imbalanced scenarios, the DUP dataset (Fig. 9) has been considered too. As we can observe, Fig. 6 provides the class-wise number of samples. Among all the classes, the "*Shadows*" is the minority class which contains the minimum number of samples (947), while the "*Meadows*" class contains the maximum number of samples (18649). From Table V, it can be observed that the proposed MorphConvHyperNet framework achieves OA, AA, and κ values that consistently outperform those obtained by the other traditional classification models. In particular, the proposed model exhibits better scores than the convoutional-based models, i.e. CNN1D, CNN2D and CNN3D. This is due to the presence of similar textures over most spectral
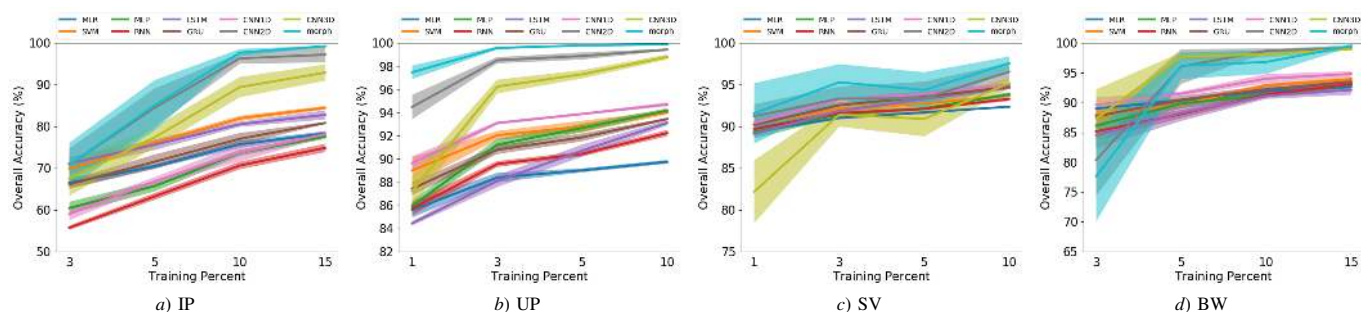
Fig. 10. Overall accuracy (OA) obtained by different classification methods with different training percentages over (a) IP dataset, (b) UP dataset, (c) SV dataset, and (d) BW dataset.

bands on three classes, namely: "Asphalt," "Self-blocking bricks," and "Shadows," where the morphological layers help to better capture the shape information and distinguish these classes, whilst its simplicity avoids overfitting problems. As compared to the CNN1D and CNN3D, the CNN2D provides significantly better OA due to the poor generalization ability exhibited by the CNN1D and the overfitting and complexity problems raised by the CNN3D. The CNN2D also improves the performance obtained by the other traditional classification models. Focusing on recurrent models, the LSTM and the GRU exhibit similar OA performances, whereas the RNN achieves the worst results. Comparing MLR, SVM and MLP, the MLP and SVM perform very similarly.

D. Performance on random sampling with different Training Percentages

To evaluate the generalization ability of the proposed `MorphConvHyperNet` network, it is important to analyze the performance improvements obtained using randomly selected and varying training sets. Fig. 10(a)-(d) respectively show the overall accuracy (OA) –obtained by different classification methods, using different training percentages– for the IP, UP, SV, and BW datasets. Specifically, we randomly select 3%, 5%, 10% and 15% of the available labeled samples from the IP and BW datasets for training, whilst the 1%, 3%, 5% and 10% of the available labeled samples have been randomly selected from the UP, and SV datasets to train the models (the remaining samples are used for testing).

It can be observed from Fig. 10(a) that the proposed `MorphConvHyperNet` model utterly outperforms all the standard machine learning methods, i.e., SVM, MLR and MLP. Also, recurrent networks (RNN, GRU and LSTM) reaches lower OA results in comparison with the proposed network. In this sense, the proposed `MorphConvHyperNet` takes advantage of the spatial information provided by the input neighbourhood windows in a natural way, which can significantly reduce the uncertainty introduced by spectrally complex images, such as IP. These results are corroborated by the CNN1D, the OA of which is considerably lower than that obtained by spatial and spectral-spatial models (i.e. CNN2D and CNN3D) and the proposal. In fact, purely spectral models (SVM, MLR, MLP, RNN, GRU, LSTM and CNN1D) have the worst accuracy, and are also the most affected by the lack of training samples in IP scene. On the contrary, CNN2D,

CNN3D and `MorphConvHyperNet` achieve the best OAs, in particular the proposed network far exceeds the results obtained by the CNN3D, performing more accurately when few training samples are available. This may be due to the high complexity introduced by the kernels of the CNN3D model, which introduce a large number of parameters that must be carefully trained and adjusted to extract the most discriminative features in order to improve classification. However, these kernels consume a great amount of training samples, therefore the model quickly tends to stagnate and overfit its parameters. On the other hand, the CNN2D is far less complex than the CNN3D, performing similarly to the proposed model when there are few training samples. Nevertheless, its OA results drop in comparison to the proposed model, in particular in the last two cases. This is due to the behaviour of the kernels themselves, which work as black boxes. Indeed, there is no control of the features that the CNN2D is obtaining, so redundant and irrelevant information may be being extracted by the model, which in the end produces worse classification results. In sharp contrast, the proposed model extracts better features, producing abstract and discriminative data representations that greatly help the model to improve classification results.

In a similar way, Fig. 10(b) provides the OA results obtained by the considered models on UP scene using training set sizes comprising 1%, 3%, 5% of and 10% of randomly selected samples from the available labelled samples. Once more, the spatial and spectral-spatial networks greatly outperform the classification results obtained by the purely spectral models. Moreover, Fig. 10(b) demonstrates the significant OA performance gains achieved by the proposed `MorphConvHyperNet` network compared with CNN2D and CNN3D. Whilst the complexity of kernels produces a fast degradation of the CNN3D model, the CNN2D does not extract the full potential of the data. In this sense, the proposed model is much simpler and takes advantage of the rich spectral-spatial information extracted from the morphological operations, achieving better OA results even when there are few training samples available.

Fig. 10(c) illustrates the obtained results in terms of OA with 1%, 3%, 5% and 10% of randomly selected training samples from the available labeled ones in the SV dataset. The image of SV is characterised by its regular patches of different crops and the spectral complexity of several lettuce crops, which differ in the stage of ripening. In this regard, spectral models show

very similar results, where the MLR is the worst of all, whilst the GRU is the only one able to outperform the 94% of OA. Focusing on the spatial and spectral-spatial models, CNN3D is really affected by spectral mixing, in particular when the training data is small to cover the full variability of the samples. Thus, its kernels are incapable of being adjusted to extract the most representative information. Although CNN2D is less affected than CNN3D by spectral similarity, it does not achieve the best results either. Finally, the proposed model achieves the best classification result, outperforming all the compared methods when the number of training samples is small (i.e., with 1% and 3%), although its standard deviation is higher than spectral models.

Finally, Fig. 10(d) provides the obtained results using 3%, 5%, 10% and 15% of randomly selected samples from the BW scene to train the classification models considered. Similarly to IP scene, this dataset is characterized by its low spatial resolution (even lower than IP, as BW has 30 mpp and IP 20 mpp) and its high spectral mixing, which makes it a very challenging image for HSI classification algorithms, which have to devote a higher effort to properly exploit the information contained on low spatial resolution HSI satellite images such as this one. As a result, it is not uncommon to note that many related papers in the literature have not reported accuracies in the BW datasets precisely because of their difficulty. In this regard, the OA of spectral models remain between 80% and 94% of accuracy, where the CNN1D stands out as the best classifier. On the contrary, spatial and spectral-spatial models are highly affected by both the low spatial resolution and the high spectral mixing. It is notable that in this scene, the proposed model suffers a great degradation and does not outperform either CNN2D or CNN3D until it has at least 15% of training data, whilst with 5% of training data it performs quite similar to CNN2D. In this sense, erosion and dilation operations are affected by the spectro-spatial characteristics of the data. One way to overcome this drawback is to try different window sizes or different numbers of morphological operations, in order to better focus the operations and match them to the type of data, although the aim of this paper has been from the beginning to provide a general architecture that works accurately with a large number of different scenes. For instance with images of low spatial resolution and high spectral mixture such as IP, scenes with better spectrally-separated classes and rich spatial information like UP, or with samples with a lot of spectral similarity as SV. Indeed, focusing on BW dataset, the results of the proposed network are utterly better than those obtained by the spectral models from the 5% of training samples with a significant margin, performing similarly to the CNN2D and the CNN3D at the end of the experiment.

Overall, the plots in Fig. 10 reveal that the features extracted by the proposed MorphConvHyperNet exhibit better generalization ability, leading to superior classification performance for training sets with different sizes and different spectral-spatial characteristics. It should also be mentioned that the proposed model achieves excellent performance with the considered HSI datasets. Morphological operations are highly non-linear. For example, dilation and erosion operations are composed of simple maximum and minimum operations.

This removes a significant amount of complexity from the model. In particular, our work implements spatial and spectral dilation/erosion operations, which selectively ignore noise and redundancies in the feature maps. In fact, this improves feature extraction and representation. Whilst CNN takes spatial linear combinations, including redundancies and noises, the proposed network applies its morphological operations to remove irrelevant information. As a result, when the network is trained with few examples, the standard CNN wastes too many resources on modelling irrelevant data and is not able to find the noise pattern, while the morphological operations completely ignore the noise. Therefore, noise- and redundancy-free feature maps always play an important role in achieving high accuracy, even on a small number of training samples. The fact that CNN1D under performs when compared to other strategies is expected, as it only relies on spectral information, whilst the other tested methods include also spatial-contextual information.

E. Ablation study

The proposed MorphConvHyperNet network employs two commonly used morphological operations, i.e., dilation and erosion, within its underlying architecture. To further evaluate the effectiveness of the morphological layers in the proposed framework, an ablation study has been conducted in order to evaluate the accuracy of the proposed MorphConvHyperNet and its baseline counterpart, the CNN2D. In this context, the CNN2D model has been designed upon the same network architecture of MorphConvHyperNet but without including the morphological blocks. Whilst conducting the experiments, we kept the experimental settings unchanged (as discussed in Section III-B).

The **CNN2D** and **MorphConvHyperNet** columns of Tables III, IV and V provide the results of the ablation studies in terms of OA, AA, and κ over the DIP, UH, and DUP datasets. Focusing on the obtained results reported from DIP scene in Table III (characterized by its low spatial resolution, its high spectral mixing and great class-imbalance), the CNN2D and MorphConvHyperNet models achieves similar OA and Kappa values (slightly higher in the case of the proposed model and with a lower standard deviation), nevertheless the AA is almost two percentage points better in the proposed model, indicating that the classification by class is more accurate than in the CNN2D model. On the other hand, focusing on Table IV which reports the obtained results on the class-balanced UH scene, the proposed MorphConvHyperNet far outperforms the classification results obtained by the baseline, exhibiting a lower standard deviation, thus its classification performance is much more stable than CNN2D. Finally, on Table V, we can evaluate the behaviour of proposed MorphConvHyperNet and CNN2D in classifying the class-imbalanced DUP scene. Once more, the proposed model greatly outperforms the OA, AA and Kappa values achieved by the CNN2D model, exhibiting higher stability and generalization ability with significantly lower standard deviation. This is because the proposed model extracts more robust and effective features through the spectral and spatial morphology
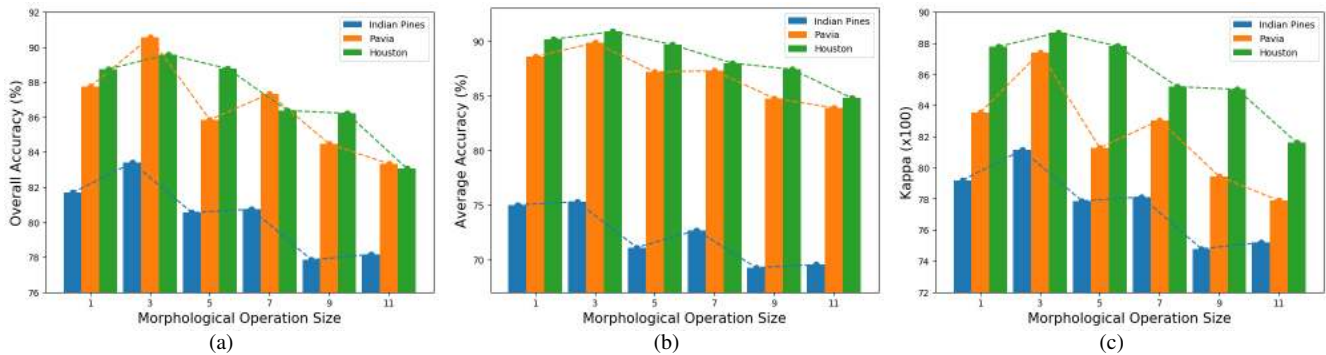
Fig. 11. (a) Overall accuracy (OA), (b) Average accuracy (AA), and (c) κ ($\times 100$) obtained by the proposed network with various sizes of the dilation and erosion structuring elements using the DIP (blue bars), DUP (orange bars), and UH (green bars) datasets.
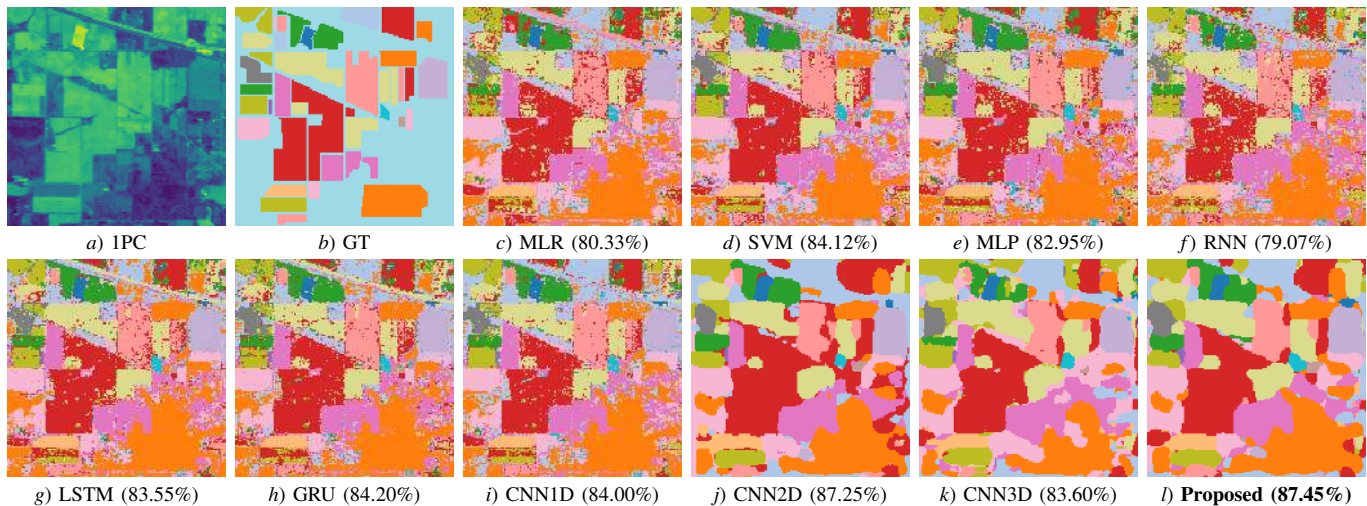

Fig. 12. (a) False color representation of the first PC obtained from the IP scene. (b) Ground truth and classification maps obtained for the DIP dataset by: (c) MLR, (d) SVM, (e) MLP, (f) RNN, (g) LSTM, (h) GRU, (i) CNN1D, (j) CNN2D, (k) CNN3D, and (l) `MorphConvHyperNet` models.

blocks, which are constructed using a combination of dilation and erosion layers. Such MM layers extract and learn more informative feature maps, which emphasize the role of spatial-contextual information during the training stage of the network as compared to the baseline CNN2D.

F. Effect of Using Different Structuring Element Sizes for the Dilation and Erosion Operations

In order to evaluate the effect of using different sizes of the SEs considered for the dilation and erosion operations in the proposed `MorphConvHyperNet` network, several experiments have been conducted over DIP, DUP and UH datasets considering SEs of size 1, 3, 5, 7, 9 and 11 in the `SpectralMorph` and `SpatialMorph` blocks of the proposed network, while keeping the other experimental settings unchanged (as discussed in Section III-B).

The achieved performances are shown in Figs. 11(a)-(c) in terms of OA, AA and κ, respectively for the DIP, DUP and UH scenes (differentiated with the colours blue, orange and green). In general, the accuracy of the model varies with different SE sizes, reaching the highest peak in terms of OA, AA and κ when the size of the SE is 3 and decreasing for the remaining sizes, in particular when using large structuring elements. This is due to the presence of varying shape information that can

be better captured through an adequate size of the SE used to implement the morphological operation. It is also clear from Figs. 11(a)-(c) that the proposed model generally achieves the best OA, AA and κ scores with SEs of size 3, while SEs of size of 11 generally provide the lowest scores for the aforementioned metrics.

G. Visual Analysis of the Obtained Classification Maps

In order to provide a qualitative visual comparison between the classification maps provided by the proposed `MorphConvHyperNet` and the other compared methods MLR, SVM, MLP, RNN, LSTM, GRU, CNN1D, CNN2D and CNN3D), Figs. 12, 13, and 14 display the obtained classification maps for the IP, UH, and UP datasets, respectively.

Focusing on DIP scene, Figs. 12 shows that spectral-based models MLR, SVM, MLP, RNN, GRU, LSTM, and CNN1D contain "salt and pepper" noise due to the miss-classification of many land-cover pixels surrounded by spectrally mixed neighboring pixels. Indeed, although spectral models ultimately differentiate the different areas visually, at their edges there is an important problem of "salt and pepper" noise, which produces many pixels of different classes totally isolated in a completely random mix at the edges between regions. This also happens in the inner-areas of the different land-cover

| *a)* 1PC | *b)* GT | *c)* MLR (72.23%) | *d)* SVM (77.80%) | *e)* MLP (82.05%) | *f)* RNN (77.07%) |

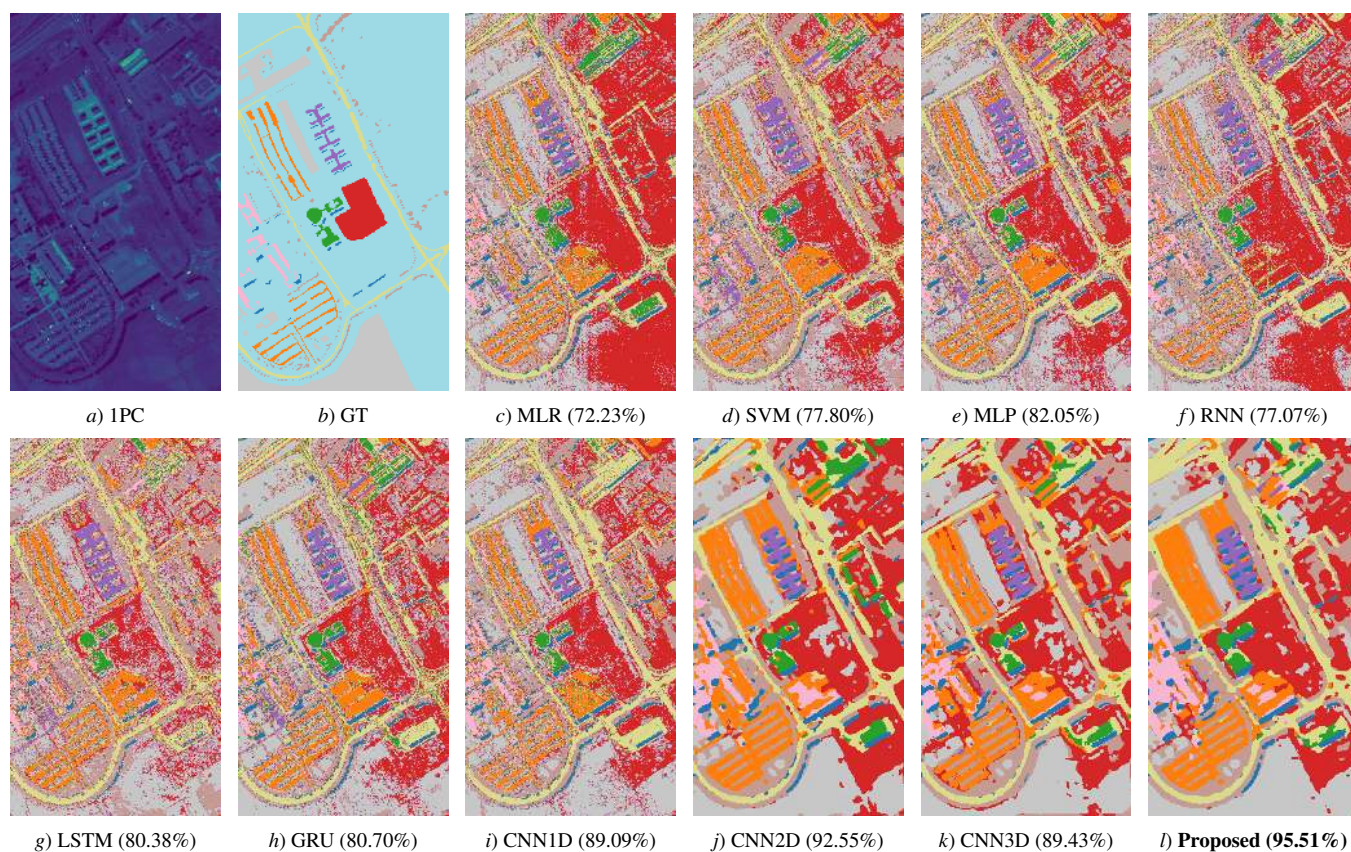| *g)* LSTM (80.38%) | *h)* GRU (80.70%) | *i)* CNN1D (89.09%) | *j)* CNN2D (92.55%) | *k)* CNN3D (89.43%) | *l)* **Proposed (95.51%)** |

Fig. 13. (a) False color representation of the first PC obtained from the UP scene. (b) Ground truth and classification maps obtained for the DUP dataset by: (c) MLR, (d) SVM, (e) MLP, (f) RNN, (g) LSTM, (h) GRU, (i) CNN1D, (j) CNN2D, (k) CNN3D, and (l) `MorphConvHyperNet` models.
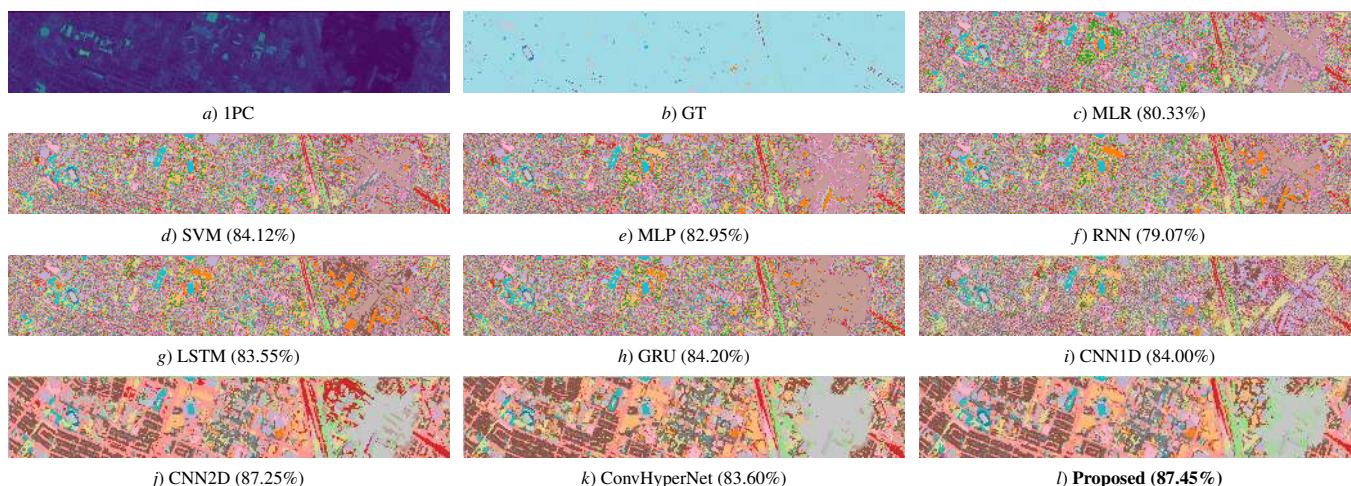

| *a)* 1PC | *b)* GT | *c)* MLR (80.33%) |

| *d)* SVM (84.12%) | *e)* MLP (82.95%) | *f)* RNN (79.07%) |

| *g)* LSTM (83.55%) | *h)* GRU (84.20%) | *i)* CNN1D (84.00%) |

| *j)* CNN2D (87.25%) | *k)* ConvHyperNet (83.60%) | *l)* **Proposed (87.45%)** |

Fig. 14. (a) False color representation of the first PC obtained from the UH scene. (b) Ground truth and classification maps obtained for the DUH dataset by: (c) MLR, (d) SVM, (e) MLP, (f) RNN, (g) LSTM, (h) GRU, (i) CNN1D, (j) CNN2D, (k) CNN3D, and (l) `MorphConvHyperNet`.

regions. In particular, the bottom left area with the "Corn min" ground cover is highly affected by the spectral mixture, being miss-classified mostly as "Soybeans min". On the contrary, spatial and spectral-spatial models overcome this drawbacks, including spatial information to mitigate the effects of spectral variability and the uncertainty introduced by it during the classification. In general, spatial and spectral-spatial models attempt to consistently delimit the different regions in such a way that they create sharply defined borders between one region and another. As a result, the frontiers are better defined,

with few samples miss-classified within the areas of different land cover classes. Of course, there are miss-classified border areas, as CNN2D shows. In particular, in contrast with CNN2D and CNN3D, the proposed `MorphConvHyperNet` can accurately identify those regions covered by "Alfalfa", "Grass/pasture-mowed" and their surrounding areas, without introducing 'Soybeans min" areas like CNN2D and following more the trend of spectral models.

This behaviour is repeated within DUP and UH scenes, as we can observe in Figs. 13, and 14. On the one hand,

MLR, SVM, MLP, RNN, GRU, LSTM, and CNN1D generally contain an important amount of "salt and pepper" noise, although visually, the different regions are easily discernible. On the other hand, spatial and spectral-spatial models produce the characteristic classification maps, with less noise artifacts and more solid regions in the sense that they attempt to remove different land-cover pixels from inner regions. It is also worth noting that the classification maps generated using CNN2D and CNN3D contain noise artefacts in some classes, whereas the classification maps of the proposed `MorphConvHyperNet` are more accurate, smoother and with better delineation of borders.

IV. CONCLUSION

This paper introduces `MorphConvHyperNet`, a new HSI classification framework based on morphological CNNs. Our network replaces the traditional linear convolution layer with basic nonlinear morphological operations that are able to extract better spectral and spatial-contextual information from the raw remote sensing data, using a less complex structure. The morphological convolution layer consists of two widely used (and easily learnable) morphological filters: dilation and erosion. This layer extracts highly discriminative features from the original HSI data using two spectral-spatial morphological blocks, i.e., `SpectralMorph` and `SpatialMorph`. Our experiments, conducted using five widely used HSI datasets, indicate that the proposed `MorphConvHyperNet` outperforms the baseline architecture without the morphological layers (ConvHyperNet) and all the other the compared methods. The effect of using dilation and erosion operations with different SE sizes are also thoroughly reported and investigated in terms of OA, AA and κ metrics. In the future, we will use more sophisticated morphological operations for the design of the convolution layer, including opening and closing by reconstruction and directional morphological operations (which may be particularly useful in urban environments).

ACKNOWLEDGMENT

REFERENCES

[1] S. Li, W. Song, L. Fang, Y. Chen, P. Ghamisi, and J. A. Benediktsson, "Deep learning for hyperspectral image classification: An overview," *IEEE Transactions on Geoscience and Remote Sensing*, 2019.

[2] M. Paoletti, J. Haut, J. Plaza, and A. Plaza, "Deep learning classifiers for hyperspectral imaging: A review," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 158, pp. 279 – 317, 2019.

[3] L. Zhang, L. Zhang, and B. Du, "Deep learning for remote sensing data: A technical tutorial on the state of the art," *IEEE Geoscience and Remote Sensing Magazine*, vol. 4, no. 2, pp. 22–40, 2016.

[4] Y. Chen, X. Zhao, and 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*, vol. 8, no. 6, pp. 2381–2392, 2015.

[5] M. Paoletti, J. Haut, J. Plaza, and A. Plaza, "Deep learning classifiers for hyperspectral imaging: A review," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 158, pp. 279–317, 2019.

[6] S. Bera and V. K. Shrivastava, "Analysis of various optimizers on deep convolutional neural network model in the application of hyperspectral remote sensing image classification," *International Journal of Remote Sensing*, vol. 41, no. 7, pp. 2664–2683, 2020.

[7] M. E. Paoletti, J. M. Haut, J. Plaza, and A. Plaza, "A new deep convolutional neural network for fast hyperspectral image classification," *ISPRS journal of photogrammetry and remote sensing*, vol. 145, pp. 120–147, 2018.

[8] M. Ramamurthy, Y. H. Robinson, S. Vimal, and A. Suresh, "Auto encoder based dimensionality reduction and classification using convolutional neural networks for hyperspectral images," *Microprocessors and Microsystems*, vol. 79, p. 103280, 2020.

[9] W. Zhao and S. Du, "Spectral–spatial feature extraction for hyperspectral image classification: A dimension reduction and deep learning approach," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 54, no. 8, pp. 4544–4554, 2016.

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

[11] Z. Zhong, J. Li, Z. Luo, and M. Chapman, "Spectral–spatial residual network for hyperspectral image classification: A 3-d deep learning framework," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 56, no. 2, pp. 847–858, 2017.

[12] S. K. Roy, G. Krishna, S. R. Dubey, and B. B. Chaudhuri, "Hybridsn: Exploring 3-d–2-d cnn feature hierarchy for hyperspectral image classification," *IEEE Geoscience and Remote Sensing Letters*, vol. 17, no. 2, pp. 277–281, 2019.

[13] S. K. Roy, J. M. Haut, M. E. Paoletti, S. R. Dubey, and A. Plaza, "Generative adversarial minority oversampling for spectral-spatial hyperspectral image classification," *IEEE Transactions on Geoscience and Remote Sensing*, pp. 1–15, 2021.

[14] J. Wang, X. Song, L. Sun, W. Huang, and J. Wang, "A novel cubic convolutional neural network for hyperspectral image classification," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 13, pp. 4133–4148, 2020.

[15] T. Alipour-Fard, M. Paoletti, J. M. Haut, H. Arefi, J. Plaza, and A. Plaza, "Multibranch selective kernel networks for hyperspectral image classification," *IEEE Geoscience and Remote Sensing Letters*, 2020.

[16] S. K. Roy, S. Manna, T. Song, and L. Bruzzone, "Attention-based adaptive spectral-spatial kernel resnet for hyperspectral image classification," *IEEE Transactions on Geoscience and Remote Sensing*, 2020.

[17] R. Van Den Boomgaard and A. Smeulders, "The morphological structure of images: The differential equations of morphological scale-space," *IEEE transactions on pattern analysis and machine intelligence*, vol. 16, no. 11, pp. 1101–1113, 1994.

[18] S. K. Roy, B. Chanda, B. B. Chaudhuri, D. K. Ghosh, and S. R. Dubey, "Local morphological pattern: A scale space shape descriptor for texture classification," *Digital Signal Processing*, vol. 82, pp. 152–165, 2018.

[19] Y. Shen, X. Zhong, and F. Y. Shih, "Deep morphological neural networks," *arXiv preprint arXiv:1909.01532*, 2019.

[20] R. Mondal, S. Santra, and B. Chanda, "Dense morphological network: an universal function approximator," *arXiv:1901.00109*, 2019.

[21] D. Mellouli, T. M. Hamdani, J. J. Sanchez-Medina, M. B. Ayed, and A. M. Alimi, "Morphological convolutional neural network architecture for digit recognition," *IEEE transactions on neural networks and learning systems*, vol. 30, no. 9, pp. 2876–2885, 2019.

[22] K. Nogueira, J. Chanussot, M. D. Mura, W. R. Schwartz, and J. A. d. Santos, "An introduction to deep morphological networks," *arXiv preprint arXiv:1906.01751*, 2019.

[23] M. C. Tobar, C. Platero, P. M. González, and G. Asensio, "Mathematical morphology in the hsi colour space," in *Iberian Conference on Pattern Recognition and Image Analysis*. Springer, 2007, pp. 467–474.

[24] P. Soille, *Morphological image analysis: principles and applications*. Springer Science & Business Media, 2013.

[25] F. Melgani and L. Bruzzone, "Classification of hyperspectral remote sensing images with support vector machines," *IEEE Transactions on geoscience and remote sensing*, vol. 42, no. 8, pp. 1778–1790, 2004.

[26] M. Dalla Mura, J. A. Benediktsson, B. Waske, and L. Bruzzone, "Morphological attribute profiles for the analysis of very high resolution

images," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 48, no. 10, pp. 3747–3762, 2010.

[27] M. Dalla Mura, A. Villa, J. A. Benediktsson, J. Chanussot, and L. Bruzzone, "Classification of hyperspectral images by using extended morphological attribute profiles and independent component analysis," *IEEE Geoscience and Remote Sensing Letters*, vol. 8, no. 3, pp. 542–546, 2010.

[28] H. G. Akçay and S. Aksoy, "Automatic detection of geospatial objects using multiple hierarchical segmentations," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 46, no. 7, pp. 2097–2111, 2008.

[29] J. A. Benediktsson, J. A. Palmason, and J. R. Sveinsson, "Classification of hyperspectral data from urban areas based on extended morphological profiles," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 43, no. 3, pp. 480–491, 2005.

[30] G. Franchi, A. Fehri, and A. Yao, "Deep morphological networks," *Pattern Recognition*, vol. 102, p. 107246, 2020.

[31] K. He, X. Zhang, S. Ren, and J. Sun, "Delving deep into rectifiers: Surpassing human-level performance on imagenet classification," in *Proceedings of the IEEE international conference on computer vision*, 2015, pp. 1026–1034.

[32] R. O. Green, M. L. Eastwood, C. M. Sarture, T. G. Chrien, M. Aronsson, B. J. Chippendale, J. A. Faust, B. E. Pavri, C. J. Chovit, M. Solis *et al.*, "Imaging spectroscopy and the airborne visible/infrared imaging spectrometer (aviris)," *Remote sensing of environment*, vol. 65, no. 3, pp. 227–248, 1998.

[33] X. Huang and L. Zhang, "A comparative study of spatial approaches for urban mapping using hyperspectral rosis images over pavia city, northern italy," *International Journal of Remote Sensing*, vol. 30, no. 12, pp. 3205–3221, 2009.

[34] X. Xu, J. Li, and A. Plaza, "Fusion of hyperspectral and lidar data using morphological component analysis," in *2016 IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*. IEEE, 2016, pp. 3575–3578.

[35] J. Haut, M. Paoletti, A. Paz-Gallardo, J. Plaza, A. Plaza, and J. Vigo-Aguiar, "Cloud implementation of logistic regression for hyperspectral image classification," in *Proc. 17th Int. Conf. Comput. Math. Methods Sci. Eng.(CMMSE)*, vol. 3. Cádiz, Spain: Costa Ballena (Rota), 2017, pp. 1063–2321.

[36] K. Cho, B. Van Merriënboer, D. Bahdanau, and Y. Bengio, "On the properties of neural machine translation: Encoder-decoder approaches," *arXiv preprint arXiv:1409.1259*, 2014.

[37] M. E. Paoletti, J. M. Haut, J. Plaza, and A. Plaza, "Scalable recurrent neural network for hyperspectral image classification," *The Journal of Supercomputing*, vol. 76, no. 11, pp. 8866–8882, 2020.

[38] W. Hu, Y. Huang, L. Wei, F. Zhang, and H. Li, "Deep convolutional neural networks for hyperspectral image classification," *Journal of Sensors*, vol. 2015, 2015.

[39] A. Ben Hamida, A. Benoit, P. Lambert, and C. Ben Amar, "3-d deep learning approach for remote sensing image classification," *IEEE Transactions on geoscience and remote sensing*, vol. 56, no. 8, pp. 4420–4434, 2018.

[40] S. Mei, J. Ji, Y. Geng, Z. Zhang, X. Li, and Q. Du, "Unsupervised spatial–spectral feature learning by 3d convolutional autoencoder for hyperspectral classification," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 57, no. 9, pp. 6808–6820, 2019.

[41] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," *arXiv preprint arXiv:1412.6980*, 2014.

[42] G. M. Foody, "Status of land cover classification accuracy assessment," *Remote sensing of environment*, vol. 80, no. 1, pp. 185–201, 2002.