IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received 15 February 2024, accepted 6 March 2024, date of publication 11 March 2024, date of current version 18 March 2024.

Digital Object Identifier 10.1109/ACCESS.2024.3376258

APPLIED RESEARCH

An End-to-End Framework for the Classification of Hyperspectral Images in the Wood Domain

ROBERTO CONFALONIERI[1], (Member, IEEE), PHYU PHYU HTUN[2], BOYUAN SUN[3], AND TAMMAM TILLO[4]

[1]Department of Mathematics, University of Padua, 35131 Padua, Italy
[2]University of Computer Studies at Yangon, Yangon 11052, Myanmar
[3]Nantong Academy of Intelligent Sensing, Nantong 226019, China
[4]Indraprastha Institute of Information Technology, New Delhi, Delhi 110020, India

Corresponding author: Roberto Confalonieri (roberto.confalonieri@unipd.it)

This work was supported by the Hyperspectral Images for Inspection Applications (H2I) Project through the European Regional Development Fund [Europäischer Fonds für Regionale Entwicklung (EFRE)]-Fondo Europeo di Sviluppo Regionale (FESR) Program (2014–2020) under Grant CUP: I56C19000100009.

ABSTRACT Hyperspectral images consist of a multitude of spectral bands for each pixel. Spectral bands provide information about wavelengths that may cover a larger spectrum of what the human eye may see. In the hyperspectral domain, the classification of hyperspectral images is usually addressed by taking into account only the spectral information. However, in the wood domain, spatial information is also relevant. To bridge this gap, this paper proposes a CNN-based end-to-end framework for the classification of hyperspectral images in the wood domain. The proposed framework consists of a spatial and spectral classifier that are integrated to make the final prediction. Each classifier is built by adapting a general image classifier, which is suitable for the classification of three-band images, to handle hyperspectral images. The framework is trained and validated on a real dataset, provided by a company working in the wood domain to detect wood fungi. The results obtained have shown that the proposed framework is a lightweight and effective approach for the recognition of wood fungi categories. The framework outperforms a benchmark classifier by 17% and can generate a classification map of hyperspectral images of wood boards of any size with an accuracy of 96%.

INDEX TERMS Application of hyperspectral images, detection of wood fungi.

I. INTRODUCTION

Hyperspectral imaging is an evolving field used in a variety of applications, such as astronomy, molecular biology, physics, medicine, surveillance and mineralogy [1]. Unlike RGB images, which are associated with three wavelength bands, hyperspectral images encompass a more extensive range of bands, and they can capture a broader spectrum than what the human eye can perceive.

In the field of hyperspectral image (HSI) analysis, pixel-level classification has been applied for a variety of applications, including image segmentation and object recognition. For HSI classification, two primary elements must be considered: two dimensions representing the image's spatial characteristic, and the third representing its spectral features. In contrast to common RGB images, spectral bands provide valuable spectral information. However, the increased complexity associated with these bands requires advanced classification solutions that differ from conventional Deep Learning techniques.

In the literature, approaches for HSI classification fall under two categories: i) studies that follow a two-steps approach in which, first, features are manually extracted, and, then, a classification model is developed based on these extracted features; ii) studies that adopt Deep Learning techniques to combine feature extraction and classification in a single framework. In the first category, various extensions of kernel-based methods like SVM have been investigated [2], [3], [4] on the basis of feature extraction algorithms like PCA [5]. While these methods can yield competitive results,

The associate editor coordinating the review of this manuscript and approving it for publication was Gustavo Callico.

their performance is significantly influenced by the expertise of expert users in feature extraction. In the second category, approaches based on Deep Belief Networks [6], Stacked Autoencoders [7] and CNNs [8], [9], [10], [11] can be used to extract hierarchical features automatically. Most of these approaches mainly focus on exploiting the spectral information, since it is typically richer and more informative than the spatial information in satellite or air-borne images.

In the wood domain, on the other hand, spatial information is also available. Thus, making use of both spatial and spectral information is worth investigating. To take full advantage of the characteristics of HSIs, we propose a CNN-based end-to-end framework for hyperspectral images classification (Fig. 1) that combines spatial and spectral classification. The framework is trained and validated on a real dataset shared by a company working in the wood domain. The classification task consists of recognizing four different categories of fungi by which some wood boards can be affected. The results obtained prove that the end-to-end framework offers various advantages: (i) it is lightweight in the sense that training the framework does not require a large dataset, or special GPU and memory requirements; (ii) it combines a spatial and a spectral classifier, which can be trained following independent strategies, improving the performance of both. Furthermore, once the spatial and spectral branches have been trained, the framework does not need any further training; (iii) it obtains multiclass classification accuracy close to 96%; (iv) it outperforms a benchmark classifier by 17% of accuracy.

II. RELATED WORK

There is an extensive use of deep-learning techniques in hyperspectral image classification in the remote image sensing domain [9], [12], [13], [14], [15], [16], [17], [18]. Nonetheless, all these approaches work under the assumption that they have only one hyperspectral image. Thus, training and testing are performed on this single image only. Our approach is intrinsically different, since the objective of our work is to train a model on a set of images of wood boards and then use the trained model to classify images which might be different from the images which were used during the training process. Thus, the testing data in our approach does not originate from the same hyperspectral image as might happen in the remote image sensing domain mentioned above.

The rest of this section is going to consider works related to the proposed work from two perspectives: 1) works related to wood classification, and 2) works related to classification of hyperspectral images using spatial and spectral information.

The work described in [19] uses a neural networks to classify the input RGB image of wood into two categories, namely sapwood and non-sapwood. In this work, the input image is divided into blocks. Then, three histograms of the red, green, and blue channels of each block are evaluated. The input data to the network is related to the generated histograms. Different approaches were investigated to generate the input data. The approach that provided the highest sapwood detection accuracy (which yielded an accuracy of 77.8%) is the one that uses information related to the histograms of the block to be classified and the four non-adjacent neighboring blocks as input data. The outcome of this work might be considered to indicate the importance of considering the spatial information of neighboring blocks while classifying the current block. A direct comparison of our proposed work with this work might not be regarded as appropriate since the proposed work uses hyperspectral images, whereas this work uses RGB images.

The work presented in [20] is about the recognition of wood species by using hyperspectral images of the wood. In this work, Principle Component Analysis (PCA) was used to reduce the dimensionality of the data. Then, features related to these reduced dimensionality data were generated. These generated features were used as input to train a neural network. This recognition task might be regarded as a relatively easier task with respect to the task of recognizing different parts that constitute a board of wood (if the assumption that differences within the species are smaller than the differences between different species holds). In this work, the reported recognition rate of five species is 96.5%.

Another work which aims at recognizing the species of the wood is described in [21]. This work investigates the effectiveness of three variants of neural networks, namely Artificial Neural Network (ANN), Deep Neural Network (DNN), and Convolutional Neural Network (CNN), in classifying different species of softwood lumber using Near-Infrared Spectroscopy (NIR). The performance of each network variant was evaluated based on the precision of the classification task. The results obtained revealed that the CNN-based model outperformed the other two models and attained a validation accuracy of 99.3%, 99.9%, and 100% for raw spectra, standard normal variate (SNV) spectra, and Savitzky-Golay second derivative spectra, respectively. Additionally, the CNN-based model was found to be stable during the training process, indicating its robustness and reliability. The study highlights the potential of CNNs as a valuable tool for the accurate and reliable classification of various types of materials in a variety of industrial applications.

In a similar vein, a recent study [22] proposed a novel approach based on 'cognitive spectroscopy' to classify different hardwood species using hyperspectral imaging (HSI) images obtained from a Near-Infrared HSI camera. This framework involved feature extraction from the complex spectroscopic data, specifically 120 hyperspectral samples representing 38 different hardwood species, followed by the principal components (PC1-PC6) image extraction and the application of a Deep Neural Network (DNN) for classification. The authors reported an overall accuracy of 90.5% that exceeded the accuracy of 56% achieved in conventional

visible images. These results demonstrate the effectiveness of cognitive spectroscopy in accurately identifying different species of hardwood, even in complex and challenging scenarios.

In another study [20], the authors used hyperspectral data of five typical wood species located at Northeast Forestry University, Northeast of China. This wood species sample was obtained in the range of 400-2500 nm. They used PCA to reduce the dimension of the original hyperspectral image data, and the principal component image sequence was obtained. The parameters of the wood texture feature are extracted from the image sequence of the principal component. Each three species has 100 samples, including 500 experimental samples. The recognition rate of the Probabilistic Neural Network (PNN) classifier is 96.5%. PNN is a kind of radial basis function network. Their experiment shows that their proposed method is suitable for solving the problem of recognizing wood species.

The work presented in [6] uses spectral information and spatial information to classify pixels of the hyperspectral image of remote sensing. In this work, it seems that the training and testing data are related to the same image, where, in particular, two images were used, namely the Indian Pines image and the Pavia image. Similarly, the work presented in [23] seems to use spectral information and spatial information to classify the pixels of a hyperspectral image. In this work, it seems that training data and testing data are related to the same image.

The work presented in [8] compares three classification approaches for hyperspectral images. These three approaches use convolutional neural networks. In this work, 50% of the pixels in an image were used to train the model and the remaining pixels were used to test the model. The Salinas Valley and Indian Pines images were used to assess these three approaches. The reported results on both images showed that using information related to neighboring pixels with respect to the pixel considered for classification provides the highest classification accuracy. It should be noted that in at least two reported cases, the area from which the neighboring pixels were taken is the 21×21 pixels around the pixel being considered. Considering that 50% of pixels are used to make the training data and considering the size of the neighboring areas, this seems to suggest that there is a relatively high probability of having the testing data among the neighboring pixels which were considered for the training process.

The work in [24] proposes a framework that combines Locality Preserving Projections (LPP), Deep Convolutional Neural Network (DCNN), and logistic regression for effective hyperspectral image classification. LPP is used to process hyperspectral image data and reduce its dimensionality. Subsequently, a DCNN is constructed using autoencoders to extract deep features from the data. These deep features are then fed into logistic regression for the final classification. The work in [25] proposes a spatial-spectral

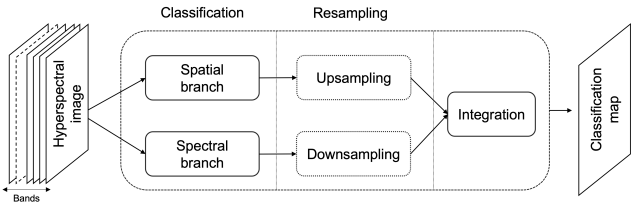

FIGURE 1. Conceptual framework for HSI classification.

classification framework that mitigates the loss of valuable bands (for example, due to noise from water absorption and corrections) through interpolation. The framework uses PCA and LPP to extract hybrid features that contain local and global spatial information, making classification more efficient.

Readers interested in approaches to hyperspectral image classification in the remote image-sensing domain may refer to the surveys in [26] and [27].

III. PROBLEM STATEMENT

The ability to recognize wood defects automatically and at an early stage can have significant economic implications for the wood industry and is worth investigating. Grayscale or RGB images are inadequate for the satisfactory detection of fungi in wood, making the acquisition of HSIs over a chosen electromagnetic range necessary.

There exist different types of fungi that affect wood [28]. Some types of fungi are easy to spot and are known as fruiting bodies (e.g., deadwood conks, mushrooms). Others are more harmful. Harmful fungi are divided into two macro-categories, depending on the damage they cause: wood-destroying fungi and wood-staining fungi. The former changes the chemical properties of the wood and weakens it, while the latter causes discoloration of the wood. In this paper, three categories of fungi will be taken into account and distinguished (classified) from *clear wood*: *soft rot*, belonging to the category of wood-destroying fungi, *brown stain* and *blue stain*, belonging to the category of wood-staining fungi.

In the context of wood detection, several studies have applied two-stage techniques to classify wood fungi from HSIs: a feature extraction stage, followed by a classification stage based on the features extracted from the previous step (e.g., [20], [29], [30]). Nonetheless, to the best of our knowledge, a unified framework, combining spatial and spectral information, has not been proposed yet, nor has a CNN architecture been utilized for the classification of HSIs of wood. Finally, there has never been a method for recognizing multiple species of fungi from HSIs.

IV. END-TO-END FRAMEWORK

Fig. 1 shows a conceptual framework for the classification of HSI. The framework takes as input a HSI with b bands and returns a classification map. The framework consists of the following components.

Classification: The spatial and spectral information of a HSI are classified by a spatial and spectral classifier. These classifiers can be developed through structural adaptation and fine-tuning of existing pre-trained CNNs architectures (e.g., Cifar10Net, VGG16, etc.). Classifiers can operate at different resolution levels. For example, spatial classification can be based on blocks (e.g., blocks of size $h \times w \times b$), whereas spectral classification can be based on pixels.

Resampling: Since the spatial and spectral classifiers can operate at different resolution levels, a resampling step is needed to align their resolutions. For instance, in the case where the spatial classifier is block-based and the spectral classifier is pixel-based, this step consists of two options: either the spatial classification is upsampled from blocks to pixels, or the spectral classification is downsampled from pixels to blocks.

Integration The two classification branches are integrated to calculate the final prediction for the classification task. This step can be implemented by some forms of information fusion (weighted average, max operator, etc.) or by layers in the neural network architecture.

In the context of this paper, the conceptual framework described above is instantiated as follows. We focus on the block-based classification, the resolution alignment is based on the downsampling of the spectral branch, and the integration is performed at network level. To merge the spatial and spectral branches, the weights and biases of the last convolutional layer from each branch are combined into a new convolutional layer. An additional convolutional layer is introduced to calculate the average scores obtained from the two branches (for more information on the integration of the two branches, see Section IV-F). The input hyperspectral images contain 337 bands and the pre-trained CNN classifier taken as a basis is Cifar10Net.[1] Cifar10Net takes as input RGB images of size $32 \times 32 \times 3$ and performs multi-class classification (10 classes). Its architecture consists of 3 convolution layers (5×5, symmetric padding of 2, stride equal to 1), each followed by ReLU and max-pooling units. The network ends with two Fully Connected (FC) layers that generate 64 and 10 feature maps, respectively, the last of which is fed into a softmax unit.

A. DATA PREPARATION

The raw data consists of 88 HSIs of eucalyptus boards of size $897 \times 512 \times 384$ that comes from a VisNIR (visible and near-infrared spectroscopy) hyperspectral camera with nominal sensitivity in the range 400-1000nm. The data was made available by a company working in the wood industry.

[1]Cifar10Net was chosen based on the input size required by the input layer. The size of 32×32 was particularly suitable for extracting an adequate number of cuboids with spatial information from the hyperspectral images. Although it would have been feasible to extract cuboids of larger sizes for clear wood data, it would not have been possible for the other categories representing wood fungi, such as soft rot, brown stain, and blue stain, due to the small portion of the wood affected by the fungi themselves.

TABLE 1. Average results of the best configuration of learning rates for the spatial branch (10^{-4}, 10^{-3}, 10^{-3} for Phase 1, Phase 2 and Phase 3, respectively).

	Phase1 (%)	Phase2 (%)	Phase3 (%)	GPU Time (min)	CPU Time (min)
avg.	43.27	71.94	90.25	130.40	323.70
std.	9.26	2.08	1.37	1.52	1.32

TABLE 2. Confusion matrix of the spatial branch performing block-based classification.

	clear w.	soft r.	brown s.	blue s.
clear w.	19	0	0	1
soft r.	0	19	1	0
brown s.	2	1	17	0
blue s.	0	0	2	18
accuracy	91.25%			

Domain experts labeled the spatial region in which one of the four classes considered is present. An exploratory analysis of the spectral bands revealed that, in all 88 hyperspectral images, bands 1 to 47 consistently had a mean and standard deviation of 0 since they fell outside the sensitivity range of the VisNIR hyperspectral camera. Consequently, these 47 bands were excluded, and the spectral bands considered for the hyperspectral images were limited to 337 bands in the range $[48, 384]$. The training data consists of cuboids with a spatial size of 32×32 pixels extracted from the hyperspectral data. 80 and 20 samples were extracted for each class to create the training and testing set, respectively. Each cuboid contains pure information about the classes considered.

B. EXPERIMENT SETUP

A grid search was carried out to identify optimal hyperparameters to train the spatial and spectral branches. The search involved learning rates in the range $[10^{-3}, 10^{-4}]$ and the number of training epochs in the range $[500, 1000, 1500]$. Regarding the mini-batch size, considering that the two branches are trained on data with different sizes (cuboids vs. pixels), the explored values were $[10, 25, 50]$ for the spatial branch and $[100, 200, 300]$ for the spectral branch.

All training experiments were repeated five times. To ensure that training was not dependent on a specific portion of the dataset, in each run, the training and testing data (80-20) were selected from different portions of the dataset, thereby implementing a 5-fold cross-validation method.

The simulations carried out for the training strategies of the spatial and spectral branches were performed on two different machines: i) a laptop with CPU Intel Core i5 dual-core at 1.6 GHz and memory of 8Gb at 1600MHz; ii) a machine with CPU Intel i7-9700k 8core at 3.6GHz, installed memory of 16Gb at 3600MHz and GPU RTX 2080Ti with 12GB memory. The computation times for the training of the

spatial and spectral branches are reported in Tables 1 and 3, respectively.

C. SPATIAL BRANCH

1) ARCHITECTURE

The architecture of the hyperspectral spatial classifier is defined by manipulating the input unit and the outputs unit of the pre-trained Cifar10Net classifier as follows. First, the input unit is restructured so that it can process 337 bands instead of three RGB channels. For this purpose, the filter of the first convolution layer is modified from $5 \times 5 \times 3$ to $5 \times 5 \times 337$ to handle input images of size $32 \times 32 \times 337$. Then, since the number of classes for the given domain is 4 instead of 10 (the number of output categories of the Cifar10 dataset), the last fully connected layer of the general image classifier, composed of 10 output units, is replaced with a fully connected layer with 4 output units. This structural adaptation shapes the spatial classifier that can handle the HSIs of wood and classify them as one of the four wood categories.

2) TRAINING

The training strategies described below share the following training hyperparameters: the number of training epochs is 1500, the mini-batch size is 50, stochastic gradient descent with momentum (SGDM) equal to 0.9. The spatial classifier is trained with all possible combinations of fixed learning rates (namely, 10^{-3} and 10^{-4}) for all three phases and 5 independent simulations are repeated to validate the results. The classifier is trained following a training strategy consisting of three phases. In Phase 1, the output unit of the HSI classifier is tuned. In Phase 2, only the input unit is trained, while in Phase 3 both input and output units are tuned. It is worth noting that both input and output layers in this strategy are trained twice to fine-tune their parameters and fully exploit the tuning process. Experiments were also carried out by training the network with a subset of the above-mentioned phases, e.g., by training the network in one phase or in two phases. However, the results in terms of testing accuracy show that training the classifier with all the three phases yields the highest average accuracy (Table 1) and the overall highest accuracy (Table 2). The CPU training times achieved on a common laptop are not significantly high, showing that the developed classifier can be trained without the need of a large dataset, nor of special GPU or memory requirements.

3) CLASSIFICATION

Finally, to produce the classification map of an HSI of arbitrary size, the classifier needs to be transformed into a Fully Convolutional Network (FCN), that is, a network with 1×1 convolutions that accomplish the function of FC layers and remove the input size constraint proper of CNN architectures that end with one or more FC layers. The spatial classifier produces a block-based prediction by classifying each spatial 32×32 region of the input HSI.

TABLE 3. Average results of the best configuration of learning rates for the spectral branch (10^{-3}, 10^{-3}, 10^{-4} for Phase 1, Phase 2 and Phase 3, respectively).

	Phase1 (%)	Phase2 (%)	Phase3 (%)	GPU Time (min)	CPU Time (min)
avg.	77.35	77.82	78.71	44.84	210.20
std.	0.78	0.77	0.77	0.45	1.07

TABLE 4. Confusion matrix of the spectral branch performing pixel-level classification.

	clear w.	soft r.	brown s.	blue s.
clear w.	19484	0	16	980
soft r.	32	17961	2063	424
brown s.	2876	5374	9467	2763
blue s.	1778	558	157	17987
accuracy	79.22%			

D. SPECTRAL BRANCH

1) ARCHITECTURE

The spectral classifier is based on the Cifar10Net architecture, manipulated to process the spectral component of the hyperspectral image rather than the spatial component. To adapt Cifar10Net to focus on processing spectral information, the kernel size of the convolution layers is changed from 5×5 to 5×1, and the output of the last fully connected layer is changed from 10 to 4. For the max-pooling layers, the original 3×3 kernel is replaced by a 3×1 kernel. Training data samples are first extracted in vector format (i.e., $1 \times 1 \times 337$) and then reshaped into $337 \times 1 \times 1$. The reshaping process allows the spectral classifier to process the spectral components exclusively.

2) TRAINING

The spectral classifier is trained on selected training samples by phases. To reduce the negative effects caused by noisy data, training data is divided into different groups based on the classification scores achieved by a simple classifier, such as a neural network-based classifier, or classical methods based on SVM or random forest. Based on this, three groups of data whose associated probability score is $90-100\%$, $80-90\%$, and $70-80\%$ were created. These groups were used to train the classifier in 3 phases. In Phase 1, the network is trained with data that is correctly predicted with probability greater than 90%. In Phase 2, the last two fully connected layers are set to be trainable. Then, the network is trained with data that is correctly predicted with a probability between 80% and 90%. In Phase 3, only the last fully-connected layer is set to be trainable and the network is fine-tuned with data that has prediction probability between 70% and 80%. To find the optimal performance, two fixed learning rates are used for each stage, namely 10^{-3} and 10^{-4}. The training options are kept the same except for the learning rates: the number of

training epochs is 100, mini-batch size is 128, and stochastic gradient descent with momentum algorithm is used, where momentum is 0.9. The optimal average test accuracy for each phase and the computation time are shown in Table 3. The performance of the spectral classifier is evaluated on the test set of extracted pixel vectors: we report the confusion matrix and the corresponding accuracy in Table 4. The CPU training times achieved on a common laptop are not significantly high.

3) CLASSIFICATION

The spectral branch operates at pixel-level resolution. The classifier receives as input a hyperspectral image of size $n \times m \times b$, and produces a prediction for each pixel, producing an output of size $n \times m \times n\_classes$, with the number of classes equal to 4 in our case. The spectral branch applies 3D convolutions and 3D pooling layers, i.e., it applies sliding cuboidal convolution filters to the 3D input that moves along the input horizontally, vertically, and along the depth dimension. The use of 3D layers is necessary to process the spectral information without considering the spatial one.

E. DOWNSAMPLING

The spatial and spectral branches operate at different resolution levels namely, block-based and pixel-based, respectively. Before integrating these two branches, a resampling step is needed to make their resolutions compatible. In this paper, we focus on downsampling from pixels to blocks in the spectral branch.[2]

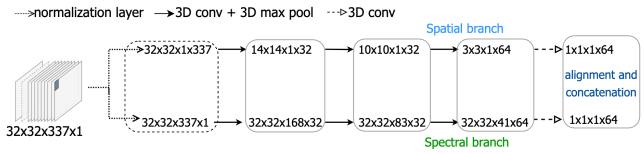

FIGURE 2. Architecture and volume transformation of the end-to-end framework, after the downsampling and integration steps.

TABLE 5. Confusion matrix of the downsampled spectral classifier performing block-based classification.

	clear w.	soft r.	brown s.	blue s.
clear w.	20	0	0	0
soft r.	0	18	2	0
brown s.	3	5	9	3
blue s.	1	0	0	19
accuracy	82.50%			

The downsampling was implemented as follows. After the fourth convolution layer (Fig. 2), a total of 32 features are extracted for each of the 32 × 32 pixels of the HSI. This is obtained by applying 64 3D convolution filters of size 1 × 1 × 39 on the input dimension of size 32 × 32 × 41 × 64. This is where the downsampling technique is applied, with the aim of obtaining 64 features for the entire 32 × 32 spatial block. To this end, the fourth convolution layer is modified to include 64 filters of size 32 × 32 × 39, which applied to the input features of size 32 × 32 × 41 × 64 produces as output 64 features for the entire original block (Fig. 2, bottom). The performance of the spectral classifier after the downsampling step is evaluated on the test set. The confusion matrix and the corresponding accuracy are reported in Table 5.

[2]Upsampling from blocks to pixels could be also achieved, for instance, by adopting transposed 3D convolutions in the spatial branch.

TABLE 6. Confusion matrix of the end-to-end framework performing block-based classification.

	clear w.	soft r.	brown s.	blue s.
clear w.	20	0	0	0
soft r.	0	20	0	0
brown s.	2	1	17	0
blue s.	0	0	0	20
accuracy	96.25%			

After downsampling, there are 64 extracted features after the fourth convolution layers in both branches. However, the volume transformations at each layer of the spatial branch are 3-dimensional, while those of the spectral branch are 4-dimensional. To compensate for this difference, the spatial branch convolutional filters are converted to 3D convolutions: each convolution layer composed by n filters of size $f \times f \times c$ is converted into a 3D convolution layer composed by n filters of size $f \times f \times 1 \times c$, with the weights and biases reshaped accordingly. It was tested that this conversion does not affect the performance of the classifier, and it finally allows to align the dimensions of the features extracted after the first four convolutional layers (Fig. 2). This alignment supports the integration of the two branches, as it will be described in the following section.

F. INTEGRATION

The integration of the two branches into a single end-to-end framework requires two final steps: input adjustment and extracted features concatenation.

1) INPUT ADJUSTMENT

The spatial branch expects an input volume of size 32 × 32 × 1 × 337, while the spectral one expects an input volume of size 32 × 32 × 337 × 1. Furthermore, the spatial branch applies zero-centred normalization for each of the bands (1 × 337 mean values), while the spectral branch applies zero-centred normalization by subtracting a single value. These differences needed to be resolved to align the two branches. To this end, an intermediate layer called the *normalization layer* was created. In the spatial branch, the *normalization layer* performs both reshaping and normalization functions. It is a 3D convolution layer made up of 337 filters of size 1 × 1 × 337 × 1, such that for each filter i, with $1 \leq i \leq 337$, the i^{th} value of the filter is equal to 1, and the others are equal to 0. Since the 3D convolution layer applies an element-wise multiplication, it copies the value

FIGURE 3. Example of classification map produced by the end-to-end framework on a sample HSI wood board of size 897 × 512 × 337.

of the corresponding element from the third to the fourth dimension, converting the volume from $32 \times 32 \times 337 \times 1$ to $32 \times 32 \times 1 \times 337$. At the same time, the 337 mean values of the input layer are used as bias values to emulate the normalization. In the spectral branch, the *normalization layer* is composed of a convolution layer that emulates the zero-centred normalization. The size of the filters is $1 \times 1 \times 1 \times 1$ (to preserve the dimension of the input), the weights are made by 1s (to preserve the values), and the bias is obtained by negating the mean value. This solution let the framework be flexible w.r.t. input sizes of the spatial and spectral branch.

2) EXTRACTED FEATURES CONCATENATION

To combine the predictions of both branches into a single final prediction, the aligned features that are produced after the fourth convolutional layer are concatenated along the fourth dimension, to produce a total of 128 extracted features. Moreover, the weights and biases of the fourth 3D convolution layer for each branch are re-used and concatenated into a new 3D convolution layer. Finally, a last 3D convolution layer is added. This layer performs the average of the scores coming from the two branches, producing 4 scores that are fed into a softmax unit to perform classification.

V. EXPERIMENTS AND RESULTS

The final end-to-end framework architecture allows us to process HSIs, to extract features from both spectral and spatial information, and combine them to produce a final prediction of one of the four categories under consideration. The performance of the complete framework was evaluated on the test set: results are reported in Table 6. The results clearly show the benefit of combining spectral and spatial information to produce HSI classification. Specifically, the combined framework yields an improvement of around 5% and 17% over the spatial and spectral branches, respectively. Furthermore, a detailed analysis of the misclassified samples revealed that 3 samples were misclassified by both independent branches, out of a total of 7 and 14 errors for the spatial and spectral classifier, respectively. The combined

TABLE 7. Precision and Recall of PLS-DA vs End-to-end framework.

	PLS-DA		Framework	
	Prec.	**Recall**	**Prec.**	**Recall**
clear wood	0.8	1	**0.91**	**1**
soft rot	0.74	0.85	**0.95**	**1**
brown stain	0.73	0.4	**1**	**0.85**
blue stain	0.86	0.9	**1**	**1**

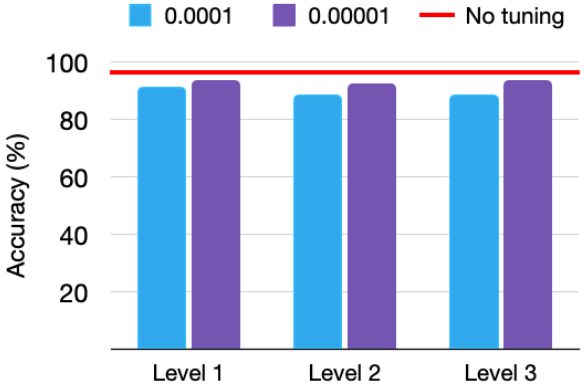

FIGURE 4. Accuracy of the combined framework after being trained following three levels of tuning.

classifier reported exactly the 3 errors that were shared by both branches. Consequently, the combination and interaction of the spectral and spatial information allowed 15 of 15 errors to be corrected in cases where one of the two branches had correctly classified the input HSI. Another advantage of the framework is that it is independent of the spatial and spectral branch. The two branches can be trained using different strategies.

Finally, the architecture of the end-to-end classifier is able to produce a classification map of HSIs of arbitrary size. The complete end-to-end framework was tested on the original HSI of wood boards (Fig. 3).

A. COMPARISON WITH BENCHMARK CLASSIFIER

We compared the obtained results with a benchmark classifier used by the company that provides the data. The classifier in question is Partial Least Square Discriminant Analysis (PLS-DA), a supervised classification algorithm. The two classifiers are trained and tested on the same samples. The PLS-DA classifier obtains a testing accuracy of 78.75%. Consequently, the end-to-end framework developed outperforms PLS-DA by approximately 17%. It also improves precision and recall for each category (see Table 7), highlighting the benefit of integrating spatial and spectral information in a single framework.

B. TUNING THE FRAMEWORK

Further experiments were conducted on the combined framework. Specifically, the combined framework was trained again after the downsampling and integration steps, with the aim of tuning some layers after the combination of the branches. The framework was trained with two different

learning rates (10^{-4} and 10^{-5}) and with three levels of tuning: (i) freezing the spatial and spectral branches, tuning the layers after the concatenation layer (level 1); (ii) tuning the layers after concatenation together with the last layer of each branch (level 2); (iii) tuning the layers after concatenation together with the last two layers of each branch (level 3). The results in term of testing accuracy revealed that none of the tuning techniques reported an improvement over the approach without further training. This seems to suggest that retraining the end-to-end classifier is not needed once the two branches have been trained separately and integrated.

VI. CONCLUSION AND FUTURE WORKS

This paper proposed a CNN-based end-to-end framework for hyperspectral image classification, investigating a case study in the detection of wood fungi. The proposed framework consists of a spatial and spectral classifier that are combined to produce a final classification. Each classifier is built on the basis of a pre-trained RGB general image classifier, without the need of a large dataset, nor of special GPU or memory requirements. The framework is trained and validated on a real dataset provided by a company working in the wood domain, with the aim of recognizing four different categories: *clear wood*, *soft rot*, *brown stain* and *blue stain*. The proposed classifier outperforms a benchmark classifier by 17% and it produces a classification map of HSI wood boards of any size with an accuracy of 96%. The framework is available at https://github.com/rconfalonieri/hsi-framework.

As future work, we plan to apply the proposed framework to additional hyperspectral image datasets in different domains and to an extended version of the dataset described in this paper.

ACKNOWLEDGMENT

The present work was carried out when Roberto Confalonieri and Boyuan Sun were affiliated with the Free University of Bozen-Bolzano, Italy. The authors would like to thank D. Cremonini for the work conducted in his Master thesis at the Free University of Bozen-Bolzano, and Matteo Caffini, Simone Faccini, and Marco Boschetti, for the wood dataset.

REFERENCES

[1] H. Grahn and P. Geladi, *Techniques and Applications of Hyperspectral Image Analysis*. Hoboken, NJ, USA: Wiley, 2007.

[2] F. Melgani and L. Bruzzone, "Classification of hyperspectral remote sensing images with support vector machines," *IEEE Trans. Geosci. Remote Sens.*, vol. 42, no. 8, pp. 1778–1790, Aug. 2004.

[3] G. Camps-Valls and L. Bruzzone, "Kernel-based methods for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 43, no. 6, pp. 1351–1362, Jun. 2005.

[4] K. Kavitha, S. Arivazhagan, and I. K. Sangeetha, "Hyperspectral image classification using support vector machine in ridgelet domain," *Nat. Acad. Sci. Lett.*, vol. 38, no. 6, pp. 475–478, Dec. 2015.

[5] C. Chen, W. Li, H. Su, and K. Liu, "Spectral–spatial classification of hyperspectral image based on kernel extreme learning machine," *Remote Sens.*, vol. 6, no. 6, pp. 5795–5814, Jun. 2014.

[6] Y. Chen, X. Zhao, and X. Jia, "Spectral–spatial classification of hyperspectral data based on deep belief network," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 8, no. 6, pp. 2381–2392, Jun. 2015.

[7] Y. Chen, Z. Lin, X. Zhao, G. Wang, and Y. Gu, "Deep learning-based classification of hyperspectral data," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 7, no. 6, pp. 2094–2107, Jun. 2014.

[8] T.-H. Hsieh and J.-F. Kiang, "Comparison of CNN algorithms on hyperspectral image classification in agricultural lands," *Sensors*, vol. 20, no. 6, p. 1734, Mar. 2020.

[9] W. Hu, Y. Huang, L. Wei, F. Zhang, and H. Li, "Deep convolutional neural networks for hyperspectral image classification," *J. Sensors*, vol. 2015, pp. 1–12, Aug. 2015.

[10] Q. Gao, S. Lim, and X. Jia, "Hyperspectral image classification using convolutional neural networks and multiple feature learning," *Remote Sens.*, vol. 10, no. 2, p. 299, Feb. 2018.

[11] J. Steinbrener, K. Posch, and R. Leitner, "Hyperspectral fruit and vegetable classification using convolutional neural networks," *Comput. Electron. Agricult.*, vol. 162, pp. 364–372, Jul. 2019.

[12] K. Makantasis, K. Karantzalos, A. Doulamis, and N. Doulamis, "Deep supervised learning for hyperspectral data classification through convolutional neural networks," in *Proc. IEEE Int. Geosci. Remote Sens. Symp. (IGARSS)*, Jul. 2015, pp. 4959–4962.

[13] W. Zhao and S. Du, "Spectral–spatial feature extraction for hyperspectral image classification: A dimension reduction and deep learning approach," *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 8, pp. 4544–4554, Aug. 2016.

[14] Y. Li, H. Zhang, and Q. Shen, "Spectral–spatial classification of hyperspectral imagery with 3D convolutional neural network," *Remote Sens.*, vol. 9, no. 1, p. 67, Jan. 2017.

[15] W. Zhao, Z. Guo, J. Yue, X. Zhang, and L. Luo, "On combining multiscale deep learning features for the classification of hyperspectral remote sensing imagery," *Int. J. Remote Sens.*, vol. 36, no. 13, pp. 3368–3379, Jul. 2015.

[16] A. Ben Hamida, A. Benoit, P. Lambert, and C. Ben Amar, "3-D deep learning approach for remote sensing image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 56, no. 8, pp. 4420–4434, Aug. 2018.

[17] Y. Luo, J. Zou, C. Yao, X. Zhao, T. Li, and G. Bai, "HSI-CNN: A novel convolution neural network for hyperspectral image," in *Proc. Int. Conf. Audio, Lang. Image Process. (ICALIP)*, Jul. 2018, pp. 464–469.

[18] S. Yu, S. Jia, and C. Xu, "Convolutional neural networks for hyperspectral image classification," *Neurocomputing*, vol. 219, pp. 88–98, Jan. 2017.

[19] A. Ziadi, F. Ntawiniga, and X. Maldague, "Neural networks for color image segmentation: Application to sapwood assessment," in *Proc. Can. Conf. Electr. Comput. Eng.*, 2007, pp. 417–420.

[20] H. Wang, H. Wang, W. Yu, and H. Li, "Research on wood species recognition method based on hyperspectral image texture features," in *Proc. 4th Int. Conf. Mech., Control Comput. Eng. (ICMCCE)*, Oct. 2019, pp. 413–4133.

[21] S.-Y. Yang, O. Kwon, Y. Park, H. Chung, H. Kim, S.-Y. Park, I.-G. Choi, and H. Yeo, "Application of neural networks for classifying softwood species using near infrared spectroscopy," *J. Near Infr. Spectrosc.*, vol. 28, nos. 5–6, pp. 298–307, Oct. 2020.

[22] H. Kanayama, T. Ma, S. Tsuchikawa, and T. Inagaki, "Cognitive spectroscopy for wood species identification: Near infrared hyperspectral imaging combined with convolutional neural networks," *Analyst*, vol. 144, no. 21, pp. 6438–6446, 2019.

[23] 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 Trans. Geosci. Remote Sens.*, vol. 57, no. 9, pp. 6808–6820, Sep. 2019.

[24] S. Singh and S. S. Kasana, "Efficient classification of the hyperspectral images using deep learning," *Multimedia Tools Appl.*, vol. 77, no. 20, pp. 27061–27074, Oct. 2018.

[25] S. Singh and S. S. Kasana, "A pre-processing framework for spectral classification of hyperspectral images," *Multimedia Tools Appl.*, vol. 80, no. 1, pp. 243–261, Jan. 2021.

[26] S. Jia, S. Jiang, Z. Lin, N. Li, M. Xu, and S. Yu, "A survey: Deep learning for hyperspectral image classification with few labeled samples," *Neurocomputing*, vol. 448, pp. 179–204, Aug. 2021.

[27] N. Audebert, B. Le Saux, and S. Lefevre, "Deep learning for classification of hyperspectral data: A comparative review," *IEEE Geosci. Remote Sens. Mag.*, vol. 7, no. 2, pp. 159–173, Jun. 2019.

[28] G. Thomasson, J. Capizzi, F. Dost, J. Morrell, and D. Miller, "Wood preservation and wood products treatment—Training manual," Oregon State Univ., Corvallis, OR, USA, Tech. Rep. EM 8403, 2006.

[29] P. Zhao and C.-K. Wang, "Hardwood species classification with hyperspectral microscopic images," *J. Spectrosc.*, vol. 2019, pp. 1–14, Jun. 2019.
[30] I. Burud, L. R. Gobakken, A. Fløe, K. Kvaal, and T. K. Thiis, "Hyperspectral imaging of blue stain fungi on coated and uncoated wooden surfaces," *Int. Biodeterioration Biodegradation*, vol. 88, pp. 37–43, Mar. 2014.

ROBERTO CONFALONIERI (Member, IEEE) received the Ph.D. degree (cum laude) in artificial intelligence from the Politechnic University of Catalonia—Barcelona Tech, Spain, in 2011. He is currently an Associate Professor of computer science with the Department of Mathematics "Tullio Levi-Civita," University of Padua, Italy. His research interests include AI, particularly trustworthy and explainable AI, knowledge representation and applied ontologies and deep learning, particularly computer vision and hyperspectral images, and computational creativity, particularly concept invention, concept evaluation, and concept refinement. He is a Senior Editor of *Cognitive Science Research* journal (Elsevier) and an Associate Editor of *Neurosymbolic Artificial Intelligence* journal (IOS Press).

PHYU PHYU HTUN received the master's degree in computer science from the University of Computer Studies at Mandalay (UCSM), Myanmar, in 2007. She is currently pursuing the Ph.D. degree with the University of Computer Studies at Yangon (UCSY), Myanmar. Since 2019, she has been working in the hyperspectral image classification area.

BOYUAN SUN received the B.Eng. degree in electronic and communication engineering from Xi'an Jiaotong–Liverpool University, Suzhou, China, in 2010, the M.Sc. degree in electrical engineering from the University of Bristol, in 2010, and the M.Phil. degree in electrical engineering from the University of Liverpool, in 2019. Currently, he is a Hyperspectral Data Engineer with Nantong Academy of Intelligent Sensing, Nantong, China. From 2020 to 2022, he was a Research Assistant with the Faculty of Computer Science, Free University of Bolzano-Bozen. His research interests include computer vision, machine learning, and artificial intelligence.

TAMMAM TILLO received the Diploma degree in electrical engineering from Damascus University, Damascus, Syria, in 1994, and the Ph.D. Diploma degree in electronics and communication engineering from Politecnico di Torino, Italy, in 2005. In 1996, he completed the military service in Syria. From 2005 to 2008, he was with the Image Processing Laboratory, Politecnico di Torino. In 2008, he joined Xi'an Jiaotong–Liverpool University, China. In 2017, he joined the Free University of Bozen-Bolzano, Italy. In 2021, he joined the Indraprastha Institute of Information Technology Delhi, Delhi, India.

Open Access funding provided by 'Università degli Studi di Padova' within the CRUI CARE Agreement