# A multiscale deep learning approach for high-resolution hyperspectral image classification.

Kazem Safari, Saurabh Prasad, *Senior Member, IEEE*, Demetrio Labate, *Member, IEEE*,

***Abstract*—Hyperspectral imagery (HSI) has emerged as a highly successful sensing modality for a variety of applications ranging from urban mapping to environmental monitoring and precision agriculture. Despite the efforts by the scientific community, developing reliable algorithms of HSI classification remains a challenging problem especially for high-resolution HSI data where there is often larger intraclass variability combined with scarcity of ground truth data and class imbalance. In recent years, deep neural networks have emerged as a promising strategy for problems of HSI classification where they have shown a remarkable potential for learning joint spectral-spatial features efficiently via backpropagation. In this paper, we propose a deep learning strategy for HSI classification that combines different convolutional neural networks especially designed to efficiently learn joint spatial-spectral features over multiple scales. Our method achieves an overall classification accuracy of 66.73% on the 2018 IEEE GRSS hyperspectral dataset – a high-resolution dataset that includes 20 urban land-cover and land-use classes.**

***Index Terms*—Hyperspectral data, convolutional neural networks, deep learning, multiscale analysis.**

## I. Introduction

Hyperspectral imagery (HSI) has gained wide recognition due to its success in a variety of applications including remote sensing for ground cover analysis, urban mapping and environmental monitoring. As technological advances have increased the spatial and spectral resolution available for data acquisition, the problem of achieving accurate classification is becoming more challenging [1]. As a result, there is a critical need to develop improved image analysis algorithms tailored to high resolution HSI. As a part of the effort to foster innovation in classification algorithms, the IEEE Geoscience and Remote Sensing Society (GRSS) has made available a new high-resolution hyperspectral dataset containing 20 classes that include urban land-cover and land-use classes [2] over the University of Houston campus and surrounding areas.

Traditional approaches to HSI classification, such as machine learning and kernel based methods [3], [4] typically operate following a workflow that includes feature extraction followed by design and optimization of a classifier acting in the resulting feature space. A major challenge for such algorithms is high dimensionality combined with small number of ground truth data. As the number of spectral bands can be of the order of a hundred, the volume of the feature space increases very rapidly with the number of bands and a huge amount of data would be required to model this space. Therefore effective methods for feature reduction are critical for accurate classification [5]. Taking advantage of spatial information has been also widely exploited to control the problem of high dimensionality. Neighboring pixels in HSI data are highly correlated since land structures are typically bigger than the pixel size and presence of a material at one pixel affects the likelihood of the same material being present in a neighboring pixel. For this reason, spatial-spectral methods designed to jointly estimate groups of pixels whose properties are constrained with one another have also become very successful [6], [7].

Following their success in several classification tasks, deep convolutional neural networks (CNNs) have been applied for HSI classification problems and achieved state-of-the-art performance [8]. Compared with conventional machine learning methods where features are manually engineered, CNNs automatically learn hierarchical features from raw input data through a sequence of convolutional and pooling layers, followed by a fully connected layer. In contrast to linear transforms and kernel methods that generates features by convolution with fixed filters, in a CNN the convolutional filters are learned from training data via backpropagation. However, to achieve a high classification accuracy, CNNs typically require a large number of training samples – a requirement that is not practical in many HSI applications. Despite the limited number of available training samples, several methods have been proposed to adapt deep neural networks to HSI classification. Chen et al. [9] introduced a deep learning framework in combination with principal component analysis to integrate spatial and spectral features. Makantasis et al. [10] proposed a combination of a CNN to conduct the task of high-level features construction followed by a Multi-Layer Perceptron (MLP) for the classification task. More recently, a number of papers have proposed CNNs with 3D convolutional kernel aimed at learning discriminative spatial-spectral features with higher efficiency [1], [11], [12], including methods based on residual networks [13], [14], CapsNets [15] and recurrent CNNs [16]. Inspired by DenseNet architecture [17], Wang et al. [18] proposed a Fast Dense Spectral-Spatial Convolution Network (FDSSCN) that learns spectral and spatial features separately but more efficiently than a conventional 3D CNN thanks to appropriate strategies of dimensionality reduction. However, the performance of these methods is sensitive to the dataset and direct application of existing strategies to more challenging high-resolution data like the one we consider here typically do not yield classification results comparable to

SP acknowledges support of NASA New Investigator (Career) project NNX14AI47G and DL acknowledges NSF-DMS projects 1720487 and 1720452. Authors are grateful to the Hewlett Packard Enterprise Data Science Institute at the University of Houston for its support.

S. Prasad is with the Hyperspectral Image Analysis Laboratory in the Department of Electrical and Computer Engineering, University of Houston. (e-mail: saurabh.prasad@ieee.org).

K. Safari (e-mail: mkazem.safari@gmail.com) and D. Labate (e-mail: dlabate@math.uh.edu) are with the Department of Mathematics at the University of Houston.

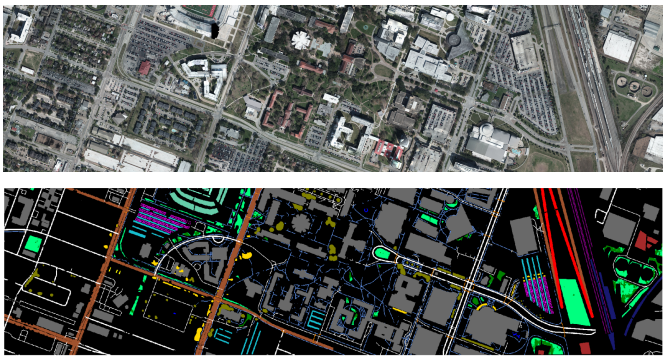

Fig. 1. RGB (above) and ground truth (below) images of training region in the 2018 IEEE GRSS Data Fusion Contest.

| | Class | Training pixels | Test pixels |
|---|---|---|---|
| | Unlabeled | 927,928 | - |
| 1 | Healthy grass | 9,799 | 20,000 |
| 2 | Stressed grass | 32,502 | 20,000 |
| 3 | Artificial turf | 684 | 20,000 |
| 4 | Evergreen trees | 13,595 | 20,000 |
| 5 | Deciduous trees | 5,021 | 20,000 |
| 6 | Bare earth | 4,516 | 20,000 |
| 7 | Water | 266 | 1,628 |
| 8 | Residential bld.s | 39,772 | 20,000 |
| 9 | Nonresidential blds | 223,752 | 20,000 |
| 10 | Roads | 45,866 | 20,000 |
| 11 | Sidewalks | 34,029 | 20,000 |
| 12 | Crosswalks | 1,518 | 5,345 |
| 13 | Major thoroughfares | 46,348 | 20,000 |
| 14 | Highways | 9,865 | 20,000 |
| 15 | Railways | 6,937 | 11,232 |
| 16 | Paved parking lots | 11,500 | 20,000 |
| 17 | Unpaved parking lots | 146 | 3,524 |
| 18 | Cars | 6,547 | 20,000 |
| 19 | Trains | 5,369 | 20,000 |
| 20 | Stadium seats | 6,824 | 20,000 |

TABLE I
LIST OF CLASSES IN THE 2018 IEEE GRSS HYPERSPECTRAL DATASET AND NUMBER OF TRAINING AND TESTING PIXELS PER CLASS (THE LAST COLUMN IS DERIVED FROM THE TEST CONFUSION MATRIX).

those reported for smaller datasets such as Indian Pines and UH2013. This is due to more classes, finer spatial resolution and lower number of spectral bands than those found in other standard dataset as well as class imbalance, intraclass variability and separation of training and test regions.

In this work, we developed a modified deep learning strategy based on a number of observations. We found that channel-wise data normalization reduces spectral discrimination for our data (note the relatively small number of spectral bands) hence we did not use channel-wise normalization. We also observed that batch normalization, while prevalent in neural network applications, has a negative impact on this dataset as it reduces generalization (high training classification accuracy with poor test accuracy). Since, in this dataset, training and test data regions are well separated – unlike smaller datasets commonly used in HSI literature – overfitting is a more critical issue. Hence we did not use batch normalization layers in our network architectures. Removing batch normalization lead us to make modification in the selection of nonlinearities, as we discuss below. Finally, we found that some objects of interest occur over multiple spatial scales and are not efficiently captured using a single patch size. Hence, we considered input patch sizes of different dimensions in our networks. We show below that our strategy based on these observations is effective to achieve improved classification performance as compared to conventional and state-of-the-art HSI classification methods.

The rest of the paper is organized as follows. In Sec. II, we present the proposed classification approach. In Sec. III, we describe the application of our method to the 2018 IEEE GRSS hyperspectral dataset and demonstrate its efficacy. We provide concluding remarks in Sec. IV.

## II. PROPOSED APPROACH

We introduce a deep learning strategy for HSI classification aimed at addressing challenges found in high-resolution hyperspectral images such as the 2018 IEEE GRSS hyperspectral dataset that was released as part of the 2018 IEEE GRSS Data Fusion Contest [2], [19]. Such data (cf. Fig. 1) was acquired by the National Center for Airborne Laser Mapping (NCALM) at the University of Houston covering an urban area of over 4 km$^2$ and 48 spectral bands (380-1050 nm spectral range) at 1 meter/pixel resolution. The 20 urban land-cover/land-use classes in this set are more detailed than those found in prior studies [20] and include various kinds of vegetation, soil and urban classes as listed in Table I. This classification problem is particularly challenging due to the spectral and spatial variability of the various material classes in the scene, the subtle differences between some classes, class imbalance and separation of training and test regions. Our strategy aims at addressing these challenges via a robust spectral-spatial feature extraction over multiple spatial scales.

### A. Deep learning strategy for HSI classification

We designed three network architectures to process the 2018 IEEE GRSS hyperspectral dataset.

The first one is a modified 3D CNN consisting of 3 convolutional layers of sizes 16, 32 and 64 with (3x3x3) filters, followed by flattening, dropout, linear layer, and removal of batch normalization layers. We found heuristically that removing batch normalization improves discrimination of certain classes (e.g., "Railways") in the test set and that, after doing so, LeakyReLU performs better than other types of nonlinearities.

Our second architecture, shown in Fig. 2, is a modified Fast Dense Spectral Spatial Convolution (FDSSC) network that we adapted from [18]. It is designed to handle first the spectral and then the spatial information. Unlike the original design, our version uses 2D convolutions in the spatial block to improve computational cost, does not use any batch normalization layers for reasons stated above and uses PReLU nonlinearities.

Our third architecture is a Multi-Layer Perceptron (MLP) with 2 linear layers and dropout, each with 1024 and 48 units, followed by a linear layer. Again, we include no batch normalization layers. We found heuristically that our MLP performs better without any nonlinearity.

We refer below to these 3 modified networks as *Conv3d*$^*$, *FDSSC*$^*$ and *MLP*$^*$, respectively, with $()^*$ indicating our modifications in contrast to conventional architectures that use batch normalization layers and ReLU.

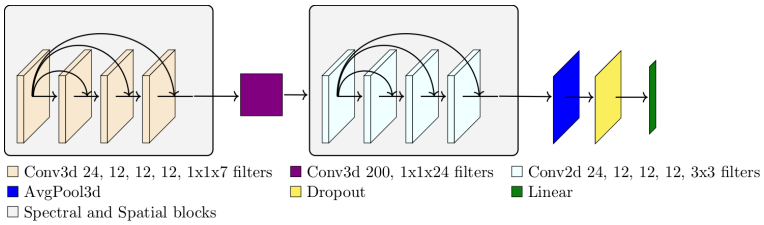


Fig. 2. Modified Fast Dense Spectral Spatial Convolution (FDSSC*) network. The first block handles spectral information and the second block handles spatial information. Each convolutional layer is followed by a PReLU.

Additionally, we found that the accuracy of a class assignment is sensitive to the choice of the patch size used as input of the network. Specifically, classes including healthy grass, artificial turf, bare earth, water, railways and highways are accurately classified using input spatial patch size 1x1 or 3x3 whereas classes including buildings, trains and roads benefit from using a 7x7 spatial patch size. We attribute this behavior to the higher relevance of spatial characteristics in the second class group. In addition, some structures occur over multiple spatial scales and a fix patch size may be insufficient for accurate detection. Therefore, we used (3x3x48) and (7x7x48) input patch sizes for our FDSSC$^*$ network model. For our Conv3d$^*$, we used input patch size (3x3x48). By construction, MLP$^*$ has input size (1x1x48). Hence, in total, to allow for multiple patch sizes, we developed 4 classification models based on our architectures: MLP$^*$, Conv3d$^*$, FDSSC3$^*$ and FDSSC7$^*$ (numbers 3 and 7 in last two models indicate that input patch sizes are (3x3x48) and (7x7x48)).

## III. Results

We report below the hyperspectral classification results on the 2018 IEEE GRSS hyperspectral dataset obtained with our deep neural network approach. Table II lists the classification performance of our proposed network architectures and our fusion model, whose classification map is shown in Fig. 3. We describe below how we carried our numerical experiments including the preprocessing, training and postprocessing.

### A. Experimental setup

Our dataset is a $1202 \times 4172 \times 48$ hyperspectal cube and the training set is a $601 \times 2384 \times 48$ hyperspectral subcube, with the test set being its complement, cf. Fig. 3. The output of our classifier is a prediction map of size $1202 \times 4172$ consisting of integer values between 1 to 20, corresponding to the 20 classes of interest. For this dataset, the nonpublic test ground truth is a raster at 0.5-m GSD superimposable to airborne image. Therefore any prediction map must be upsampled by a factor of 2, resulting in a map of size $2404 \times 83440$ before being uploaded to the IEEE GRSS server at `http://dase.grss-ieee.org` to generate test results consisting of overall and average accuracy, kappa, class accuracies and confusion matrix. According to rules established by the committee handling this dataset, accuracy parameters are computed with respect to undisclosed test samples.

### B. Preprocessing

To generate a model for each deep learning network, we divided training pixels into a training set and a validation set. We randomly selected 100 samples per class for validation and left the remaining set for training. We did not normalize the dataset and selected a batch size of 8096. To alleviate the issue of class imbalance, we used the Pytorch's function `WeightedRandomSampler` that is designed to sample uniformly from each class on the fly in each batch during training [22]. Due to consistent failure in predicting the class "Unpaved parking lots" and because of its small test size (about $1\%$), we removed this class from training.

### C. Training

For training, we used the Adam optimizer with a learning rate of $0.001$ and the cross-entropy loss function. All our models were trained for 20 epochs while monitoring the validation overall accuracy as our metric. However, due to empirical observations and overfitting concerns, the training was stopped early if the validation accuracy surpassed thresholds of $63\%$, $70\%$ and $72\%$ for various configurations.

### D. Postprocessing (Fusion model)

To combine the advantages of our 4 (trained) single network models, we designed a postprocessing step to build a new classifier (called *Fusion* in Table II) by computing a weighted average of their prediction probabilities. Specifically, we denote our network models as $N_i$, $i = 1, \ldots, 4$ and the corresponding computed class accuracies as $A_i = (a_{i,1}, \ldots, a_{i,20})$, where $a_{i,j}$ is the test accuracy of the network model $i$ for the class $j$, $j = 1, \ldots, 20$. For any test pixel $x$, each model $N_i$ generates a probability vector $P_i(x) = (p_{i,1}(x), \ldots p_{i.20}(x)$ where $p_{i,j}(x)$ is the probability that $x$ belongs to class $j$. Corresponding to each $P_i$, we define a class weight $W_i = (w_{i,1}, \ldots, w_{i,20})$ of the form $w_{i,j} = \frac{a_{i,j}}{\sum_{k=1}^{4} a_{k,j}}$. The fusion model's prediction probability vector at $x$ is $\bar{P}(x) = (\bar{p}_1(x), \ldots, \bar{p}_{20}(x)) = \sum_{i=1}^{4} W_i \odot P_i(x)$, where $\odot$ denotes the element-wise product. Finally, the fusion model's predicted class label for $x$ is $C(x) = \text{argmax}_{j=1,\ldots,20}(\bar{p}_j(x))$.

### E. Discussion

Table II reports the test classification accuracy on the 2018 IEEE GRSS hyperspectral dataset using our modified networks Conv3d$^*$, FDSSC$^*$, MLP$^*$ and our Fusion postprocessed

| | Baseline comparison | | | | | | Proposed methods | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Classes | JPlay | SSRN | pResNet | Conv1d | Conv2d | Conv3d | MLP* | Conv3d* | FDSSC3* | FDSSC7* | Fusion |
| 1 | 96 | 94.72±5.76 | 95.94±5.1 | 95.68±3.59 | 97.7±2.54 | 98.02±1.11 | 83.25±2.95 | 98.6±0.25 | 98.27±1.04 | 97.94±0.54 | 97.89 |
| 2 | 73.97 | 75.14±11.16 | 73.43±12.22 | 84.49±4.72 | 74.29±8.49 | 82.28±4.75 | 52.6±6.22 | 73.2±4.19 | 78.43±4.81 | 73.2±3.59 | 78.03 |
| 3 | 72.1 | 45.27±24.91 | 29.87±20.24 | 21.84±4.77 | 34.28±25.38 | 51.25±15.21 | 99.85±0.08 | 90.01±6.43 | 88.04±5.36 | 83.5±25.41 | 98.62 |
| 4 | 86.87 | 93.29±3.05 | 92.85±3.76 | 76.83±2.62 | 89.77±4.1 | 90.3±2.33 | 87.38±2.28 | 93.66±1.16 | 93.52±2.01 | 93.25±1.58 | 95.2 |
| 5 | 70.8 | 61.11±8.53 | 51.93±12.23 | 68.89±5.15 | 63.08±11.33 | 69.72±4.63 | 74.8±3.53 | 71.25±5.99 | 68.35±10.71 | 61.72±7.67 | 80.45 |
| 6 | 38.8 | 44.17±24.53 | 49.13±30.45 | 71.58±8.66 | 53.08±19.83 | 72.18±5.22 | 93.92±2.4 | 94.47±1.47 | 90.64±8.06 | 97.63±0.81 | 97.02 |
| 7 | 97.05 | 34.74±24.79 | 71.89±18.78 | 7.35±2.76 | 54.37±31.24 | 40.96±19.5 | 95.14±7.28 | 85.08±18.55 | 93.22±5.48 | 95.01±2.62 | 96.07 |
| 8 | 37.3 | 48.22±7.12 | 35.42±8.86 | 27.62±2.12 | 24.04±4.44 | 29.74±3.32 | 38.07±6.07 | 47.51±6.75 | 40.9±3.97 | 38.08±4.63 | 45.93 |
| 9 | 44.74 | 62.09±5.84 | 57.63±10.92 | 43.84±4.82 | 59.5±6.55 | 60.1±4.72 | 46.64±4.52 | 50.12±3.07 | 48.51±4.16 | 50.96±4.56 | 53.23 |
| 10 | 26.76 | 55.3±10.37 | 57.04±12.3 | 34.02±12.3 | 54.1±5.42 | 45.85±12.69 | 20.37±12.56 | 39.75±9.97 | 37.5±14.12 | 37.89±10.86 | 39.32 |
| 11 | 31.41 | 54.12±4.59 | 50.06±5.47 | 31.27±6.24 | 46.09±5.53 | 43.37±10.69 | 36.47±7.17 | 40.7±7.59 | 44.6±6.7 | 39.28±3.3 | 33.77 |
| 12 | 32.33 | 27.59±17.75 | 18.24±11.28 | 30.71±11.97 | 18.05±12.35 | 23.2±9.01 | 28.61±15.33 | 40.11±5.66 | 32.84±4.48 | 28.19±8.03 | 33.3 |
| 13 | 30.63 | 38.4±6.38 | 35.29±11.9 | 16.84±8.98 | 30.15±8.4 | 36.67±17.17 | 22.7±15.3 | 24.91±6.24 | 29.72±11.96 | 22.22±7.52 | 44.94 |
| 14 | 50.81 | 20.97±12.16 | 24.31±24.79 | 45.38±10.28 | 17.66±12.03 | 24.98±9.34 | 57.09±9.09 | 51.89±4.62 | 65.1±2.9 | 65.64±4.29 | 61.54 |
| 15 | 87 | 6.39±0.94 | 3.94±2.47 | 5.5±0.69 | 3.25±1.96 | 5.03±1.05 | 42.58±21.62 | 87.85±11.88 | 83.71±12.72 | 67.46±26.07 | 95.19 |
| 16 | 27.65 | 31.34±10.5 | 35.11±12.7 | 39.84±8.21 | 19.54±8.64 | 51.35±14.07 | 19.29±5.87 | 29.32±5.22 | 23.45±7.57 | 29.36±3.07 | 30.45 |
| 17 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 18 | 34.66 | 34.95±4.01 | 37.12±14.89 | 36.02±5.82 | 42.0±5.72 | 45.92±4.07 | 39.57±4.76 | 46.82±0.77 | 38.86±3.43 | 38.7±3.73 | 43.92 |
| 19 | 72.24 | 58.82±12.46 | 55.66±19.39 | 66.19±5.2 | 76.5±8.41 | 52.49±8.89 | 70.14±3.87 | 77.48±5.36 | 82.46±4.07 | 72.79±14.23 | 82.39 |
| 20 | 82.38 | 54.74±10.16 | 69.72±27.04 | 46.07±4.21 | 74.85±15.96 | 49.79±6.09 | 61.1±9.77 | 50.13±13.53 | 65.04±21.01 | 64.16±16.6 | 87.32 |
| OA | 55.16 | 51.88±2.59 | 50.54±1.8 | 47.89±2.22 | 50.78±3.39 | 53.63±1.65 | 55.16±0.99 | 61.27±0.76 | 61.84±2.87 | 59.69±3.12 | **66.73** |
| AA | 54.67 | 47.07±3.18 | 47.23±1.52 | 42.5±1.75 | 46.62±2.32 | 48.66±1.79 | 53.48±0.89 | 59.64±1.14 | 60.16±2.83 | 57.85±3.08 | 64.73 |
| Kappa | 0.53 | 0.49±0.03 | 0.48±0.02 | 0.45±0.02 | 0.48±0.04 | 0.51±0.02 | 0.53±0.01 | 0.59±0.01 | 0.6±0.03 | 0.57±0.03 | 0.65 |

TABLE II

TEST CLASSIFICATION ACCURACIES (CLASSES 1-20 IN TABLE I). BASELINE COMPARISON INCLUDES CONVENTIONAL 1D, 2D, 3D CNNS (CONV1D, CONV2D, CONV3D, RESPECTIVELY) AND STATE-OF-THE-ART ALGORITHMS J-PLAY [21], SSRN [14] AND PRESNET [13]. PROPOSED METHODS ARE THE MODIFIED MLP (MLP*), 3D-CNN (CONV3D*), FDSSC WITH 3X3 OR 7X7 INPUT PATCHES (FDSSC3* AND FDSSC7*) AND OUR FUSION POSTROCESSED MODEL. TEST ACCURACY RESULTS ARE AVERAGES OVER 10 RUNS WITH THE STANDARD DEVIATION.

classification model. For baseline comparison, we consider conventional 1D, 2D and 3D CNNs (with batch normalization layers and ReLUs), denoted as *Conv1d, Conv2d* and *Conv3d*, respectively; we also consider the state-of-the-art hyperspectral classification algorithms *J-Play* [21] – based on an improved subspace learning technique –, the Spectral-Spatial Residual Network (SSRN) [14] – employing 3D convolutional layers and spectral and spatial residual blocks – and the Pyramidal Residual Networks (pResNet) [13] – employing 2D convolutional layers and pyramidal bottleneck residual blocks.

Comparing networks Conv3d and Conv3d* in the table shows that our modification improves the overall accuracy by $8\%$ and the average accuracy by $12\%$ with major improvements for "Railways", "Artificial turf", "Bare Earth", "Trains" and "Highways". We attribute this improvement mostly to the enhanced spectral discrimination due to the removal of batch normalization layers. For the same reason, a relatively simple network such as MLP* (with no batch normalization layers) outperforms the conventional CNNs. Note though that Conv3d* performs better than MLP* overall, with most improvements for "Railways", "Roads", grasses and buildings. We explain this improvement to the better efficiency of Conv3d* in capturing spatial information. Also, our FDSSC* architecture outperforms Conv3d* for "Trains" and "Stadium Seats" due to its ability to integrate spatial and spectral information, even though the overall performances is comparable to Conv3d*. By combining the advantages of the different network models, our postprocessed Fusion model is able to improve the overall classification accuracy by almost 5% with respect to our individual network models.

Among the methods we used for baseline, the J-Play algorithm performs better than conventional CNNs but not as well as our Conv3d* and FDSSC*. The algorithms SSRN and pResNet perform comparably to Conv2d and Conv3d but worse than our Conv3d* and FDSSC*. We attribute their inferior performance to the presence of batch normalization layers that limits the ability of the model to generalize effectively. We remark the separation of train and test regions, higher spatial resolution (1 meter/pixel) and lower number of spectral bands (48 bands) available in this dataset as compared to the standard datasets on which J-Play, SSRN and pResNet were previously tested, namely Indian Pines, University of Pavia and the University of Houston 2013. Note that we ran SSRN and pResNet in their standard configuration (7x7 input patch size). To train J-Play, we experimentally found an optimal setting to be the heat kernel (out of 3 possible kernels) and selection of 200 random samples in each class for training.

## IV. CONCLUSIONS

We presented a new deep learning approach for HSI classification designed to address challenges found in high-resolution data such as the 2018 IEEE GRSS hyperpsectral dataset, namely, more classes, finer spatial resolution, lower number of spectral bands, class imbalance, intraclass variability and separation of training and test regions. To address limitations of existing methods for HSI classification, our proposed approach applies a deep learning strategy that integrates spatial and spectral, and employs different input patch sizes to efficiently capture structure occurring over multiple scales. A major novelty of our approach is to remove dataset normalization per channel and batch normalization layers. As discussed above, this modification is critical to improve classification accuracy by increasing class separability and reducing overfitting. Removing batch normalization also increases the sensitivity to weight initialization but this was addressed by modifying the nonlinearities of the network. A further investigation of this effect is beyond the scope of this letter and will be addressed in a separate work.

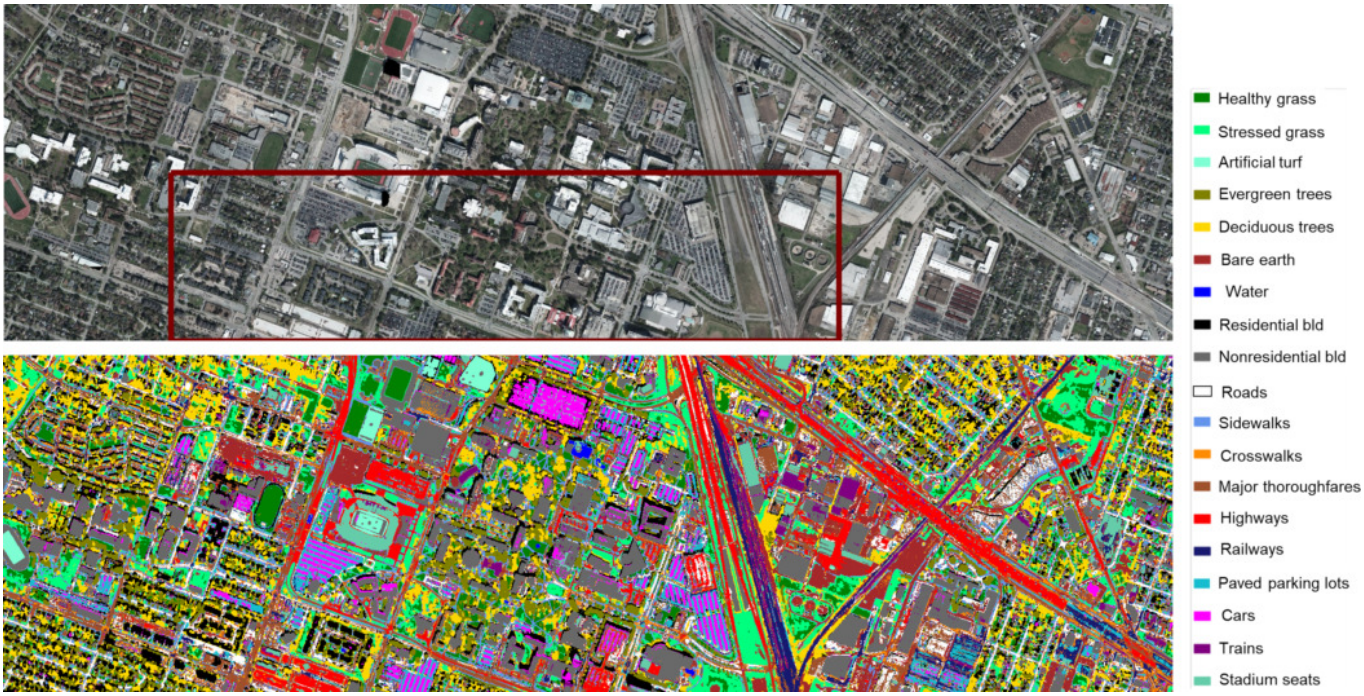


Fig. 3. RGB image of the 2018 IEEE GRSS dataset (above) and corresponding (upsampled) Fusion classification map (below) obtained from our deep convolutional network approach. It achieves a 66.73% overall test accuracy. The training set is contained within the red rectangle whereas the test set is in the complement region (the two regions are disjoint).

## REFERENCES


[1] 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, Aug 2016.

[2] B. Le Saux, N. Yokoya, R. Hansch, and S. Prasad, "2018 ieee grss data fusion contest: Multimodal land use classification [technical committees]," *IEEE Geoscience and Remote Sensing Magazine*, vol. 6, no. 1, pp. 52–54, March 2018.

[3] U. B. Gewali, S. T. Monteiro, and E. Saber, "Machine learning based hyperspectral image analysis: a survey," *arXiv:1802.08701*, 2018.

[4] G. Camps-Valls and L. Bruzzone, "Kernel-based methods for hyperspectral image classification," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 43, no. 6, pp. 1351–1362, 2005.

[5] Y. Zhou, J. Peng, and C. L. P. Chen, "Dimension reduction using spatial and spectral regularized local discriminant embedding for hyperspectral image classification," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 53, no. 2, pp. 1082–1095, Feb 2015.

[6] M. Fauvel, Y. Tarabalka, J. A. Benediktsson, J. Chanussot, and J. C. Tilton, "Advances in spectral-spatial classification of hyperspectral images," *Proceedings of the IEEE*, vol. 101, no. 3, pp. 652–675, 2012.

[7] L. He, J. Li, C. Liu, and S. Li, "Recent advances on spectral-spatial hyperspectral image classification: An overview and new guidelines," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 56, no. 3, pp. 1579–1597, 2017.

[8] 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.

[9] Y. Chen, Z. Lin, X. Zhao, G. Wang, and Y. Gu, "Deep learning-based classification of hyperspectral data," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 7, no. 6, pp. 2094–2107, June 2014.

[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)*, July 2015, pp. 4959–4962.

[11] M. Paoletti, J. 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.

[12] Y. Li, H. Zhang, and Q. Shen, "Spectral-spatial classification of hyperspectral imagery with 3d convolutional neural network," *Remote Sensing*, vol. 9, no. 1, 2017.

[13] M. E. Paoletti, J. M. Haut, R. Fernandez-Beltran, J. Plaza, A. J. Plaza, and F. Pla, "Deep pyramidal residual networks for spectral-spatial hyperspectral image classification," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 57, no. 2, pp. 740–754, Feb 2019.

[14] 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, Feb 2018.

[15] M. E. Paoletti, J. M. Haut, R. Fernandez-Beltran, J. Plaza, A. Plaza, J. Li, and F. Pla, "Capsule networks for hyperspectral image classification," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 57, no. 4, pp. 2145–2160, April 2019.

[16] 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, Feb 2018.

[17] G. Huang, Z. Liu, L. Van Der Maaten, and K. Q. Weinberger, "Densely connected convolutional networks," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017, pp. 4700–4708.

[18] W. Wang, S. Dou, Z. Jiang, and L. Sun, "A fast dense spectral-spatial convolution network framework for hyperspectral images classification," *Remote Sensing*, vol. 10, no. 7, 2018.

[19] IEEE-GRSS, www.grss-ieee.org/community/technical-committees/data-fusion/2018-ieee-grss-data-fusion-contest/.

[20] C. Debes, A. Merentitis, R. Heremans, J. Hahn, N. Frangiadakis, T. van Kasteren, W. Liao, R. Bellens, A. Pižurica, S. Gautama, W. Philips, S. Prasad, Q. Du, and F. Pacifici, "Hyperspectral and lidar data fusion: Outcome of the 2013 grss data fusion contest," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 7, no. 6, pp. 2405–2418, June 2014.

[21] D. Hong, N. Yokoya, J. Xu, and X. Zhu, "Joint and progressive learning from high-dimensional data for multi-label classification," in *The European Conference on Computer Vision*, September 2018.

[22] pytorch.org, "Weightedrandomsampler in pytorch," 2019, pytorch.org/docs/stable/data.html#torch.utils.data.WeightedRandomSampler.