IGARSS 2020 - 2020 IEEE International Geoscience and Remote Sensing Symposium | 978-1-7281-6374-1/20/$31.00 ©2020 IEEE | DOI: 10.1109/IGARSS39084.2020.9324302

TRAINING CAPSNETS VIA ACTIVE LEARNING FOR HYPERSPECTRAL IMAGE CLASSIFICATION

[1]*Mercedes E. Paoletti, Student Member, IEEE*, [1]*Juan M. Haut, Member, IEEE*,
[1]*Javier Plaza, Senior Member, IEEE*, [1]*Antonio Plaza, Fellow, IEEE*
[1]Hyperspectral Computing Laboratory, Department of Technology of Computers and Communications,
Escuela Politécnica, University of Extremadura, E-10003 Cáceres, Spain. (e-mail: mpaoletti@unex.es)

ABSTRACT

Hyperspectral imaging (HSI) gathers hundreds of images along the electromagnetic spectrum for the same area on the surface of the Earth, collecting a rich amount of spatial and spectral information. Deep learning classifiers have achieved significantly high precision results when analyzing HSI data. In particular, capsule networks (CapsNets) can provide robust classification results, overcoming the limitations of traditional convolutional neural networks (CNNs) by enriching the feature presentation capability and applying dynamic routing mechanisms. As a result, CapsNets are now widely regarded as the state-of-the-art within deep learning field. However, as it is the case for CNNs, the performance of CapsNets strongly depends on the quantity and quality of the available training samples, which in HSI tends to be scarce and noisy. Moreover, obtaining labeled data is expensive and time-consuming, and the high dimensionality of HSI data makes it difficult to accurately design classifiers based on limited training samples. This is mainly due to the strong intra-class variability present in the HSI data. Active learning (AL) can alleviate the aforementioned problems by selecting a small set of highly-representative labeled samples from a pool of unlabeled data, in iterative fashion. This paper presents a new AL-based approach for HSI data classification that integrates the spectral and the spatial information contained in the HSI data and enhances the performance of CapsNets when very limited training samples are available. **Code:** https://github.com/mhaut/AL-CapsNet-HSI.

1. INTRODUCTION

Remotely sensed hyperspectral imaging (HSI) –also called imaging spectroscopy– is an active topic in Earth observation field. It collects a large amount of spectral-spatial information from an observed surface by gathering hundreds of images (at different wavelength channels) along the electromagnetic spectrum [1]. As a result, 3-dimensional data cubes are obtained, where each pixel reflects the behaviour of terrestrial materials in the presence of electromagnetic radiation by measuring their degree of reflection, emission and absorption into a unique spectral signature for each material optically detected by the spectrometer. This rich spectral information is very useful in pattern recognition tasks, such as modeling and mapping of natural resources, allowing for a very accurate characterization of the materials located on the imaged area. In particular, HSI data have been widely used for land-cover classification tasks, exploiting a wide variety of classifiers – ranging from unsupervised to supervised methods (through semi-supervised techniques)– considering different types of information. i.e. spectral, spatial and spectral-spatial.

In fact, classification can be seen as an optimization problem, where a mapping function $f(\cdot, \theta)$ with parameters θ receives as input each pixel $\mathbf{x}_i \in \mathbb{R}^B$ of the HSI scene $\mathbf{X} \equiv \{\mathbf{x}_1, \cdots, \mathbf{x}_N\} \in \mathbb{R}^{N \times B}$ (where B is the number of spectral bands and N the number of pixels), and obtains a land-cover classification label $y_i = \{1, \cdots, K\}$ as output by adjusting θ in order to minimize the loss between the predicted and the desired outputs. Focusing on supervised classifiers, they are designed with a two-stage procedure consisting of training and test/inference stages. The first stage carries out the adjustment of model parameters, by feeding the classifier with labeled samples that compose the training set $\mathcal{D}_{train} = \{\mathbf{x}_i, y_i\}_{i=1}^{N_l}$. Thus, $\mathcal{D}_{train}$ should contain enough representative information about the HSI scene in order to perform an accurate inference for the remaining unlabeled data contained within the test set $\mathcal{D}_{test} = \{\mathbf{x}_i\}_i^{N_t}$

In recent years, many supervised classification algorithms have been developed by the HSI community, with artificial neural networks (ANNs) being quite popular because of their ability to discover hidden patterns and non-linear relationships, without prior information about the data distribution. Particularly, deep learning (DL)-based models exhibit great generalization power [2], extracting abstract features through a hierarchical stack of operational layers. On this wise, the full classifier $f(\cdot, \theta)$ is divided into several concatenated sub-mapping functions that usually apply affine linear transformations and point-wise nonlinear functions over the data in order to obtain a final representation that is finally processed by a classification function (usually, a *softmax*). In particular, convolutional neural networks (CNNs) stand out due to their feature extraction ability, which is enhanced by their architecture

Work supported by Spanish Ministry (FPU15/02090), FEDER/Junta de Extremadura (GR18060), and European Union (734541-EXPOSURE).

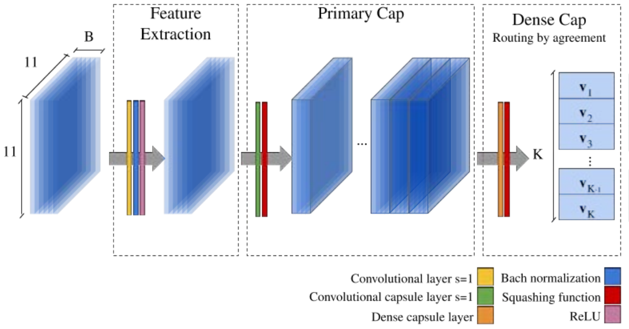


Fig. 1. Topology of CapsNet with final vector $\mathbf{z}$.

based on locally-connected kernels, allowing a natural and detailed analysis of both the spectral information (contained in each HSI pixel) and the spatial-contextual information (provided by its neighbors) [3]. Inspired by the CNN, the capsule network (CapsNet) [4] was developed to overcome some limitations observed in CNNs, such as the massive drop in performance when spatial transformations are applied over the data, the loss of the representativeness of spatial information during pooling operations, or the susceptibility to adversarial samples. Besides, the CapsNet enriches the feature presentation ability by implementing capsules (which provide a vectorized output or *pose*), while improving the back-propagation stage of CNNs through a second training algorithm (called *dynamic routing*). This model has been shown to be highly reliable in HSI classification tasks [5].

However, the performance of these classifiers strongly depends on both the quality and quantity of the available training samples. On the one hand, HSI data are often too complex to be processed because of their great spectral dimensionality, which can introduce a significant amount of class-variability due to perturbations and noise during the acquisition process. This forces classifiers to consume more training data to learn a large amount of spectral features, and to be more robust to subtle variations in the test data. On the other hand, labeled HSI data are actually quite scarce and difficult to obtain, so very limited labeled data are often available for HSI classification tasks. These restrictions generally lead to overfitting of DL-based models, which converge too early on the available training data without improving their inference results. In this context, active learning (AL)-based approaches have been employed to enhance the classification of HSI data by selecting wisely the most descriptive unlabeled samples from the original dataset, providing more information to the neural model with fewer samples and reducing the cost of labeled data acquisition [6].

In this paper, we present a new AL-inspired model for HSI data classification that combines spectral and spatial information, with the aim of enhancing the performance of CapsNets performance when very limited labeled data are available. To this end, two different acquisition functions have been implemented and applied over the network: the breaking ties (BT) criterion [7] and simple random acquisition. Our experimental results with real HSI data indicate that the proposed AL-based approach can enhance the performance of CapsNets even with very limited training data.

2. METHODOLOGY

The CapsNet is composed by three operational blocks [5]: i) feature extractor (FE) layer, ii) primary capsule (PC) layer, and iii) dense capsule (DC) layer. As we can observe in Fig. 1, the FE-layer receives as input patches $\mathbf{x} \in \mathbb{R}^{11\times 11\times B}$ extracted from the original HSI $\mathbf{X}$ and centered over each pixel from $\mathcal{D}_{train}$, applying a spectral-spatial convolutional layer to extract a feature representation of the input data. This volume is processed by the PC layer, where capsule i obtains the *pose* vector $\mathbf{u}_i$. Each $\mathbf{u}_i$ is routed to capsules j in the DC layer by using the routing-by-agreement algorithm, obtaining a vote $\hat{\mathbf{u}}_{j|i} = \mathbf{W}_{ij}\mathbf{u}_i$ describing how much the capsule i affects the capsule j, and applying a coupling coefficient $c_{i,j}$, which indicates the strength of the connection between both capsules as follows:

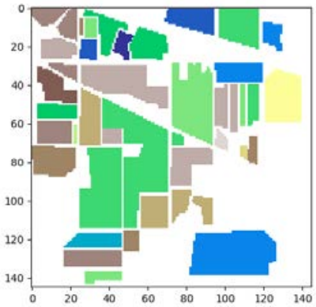
$$\mathbf{v}_j = \frac{||\mathbf{s}_j||^2}{1+||\mathbf{s}_j||^2}\frac{\mathbf{s}_j}{||\mathbf{s}_j||}, \text{ with } \mathbf{s}_j = \sum_i c_{ij}\hat{\mathbf{u}}_{j|i} \qquad (1)$$

In the end, one activity vector $\mathbf{v}_j$ per class is obtained, encoding additional details about the features –such as orientation, *pose* or size– and obtaining the probability of belonging to that class (through its length). Finally, these vectors are reshaped to be the input of a multi-layer perceptron (MLP) that obtains the final classification.

Layer	*Size*	*Stride*	*BatchNorm*	*Act. Funct.*
FE	$64 \times 3 \times 3 \times B$	1	Yes	ReLU
PC	$8 \times 64 \times 3 \times 3 \times 64$	1	No	Squashing
DC	$K \times 16$	-	No	Squashing
FC1	328	-	No	Sigmoid
FC2	129	-	No	Sigmoid
FC3	$B \times D \times D$	-	No	Linear

Table 1. Summary of the parameters in each layer of the considered network. The dynamic routing is set to 3. Three fully-connected (FC) layers perform the final classification.

With the aim of applying AL, the original HSI dataset $\mathbf{X}$ is divided into four subsets: $\mathcal{D}_{train}$, initially composed by two samples per class. From the remaining samples, 10% are included in $\mathcal{D}_{val}$. $\mathcal{D}_{pool}$ and $\mathcal{D}_{test}$ are respectively composed by 50% of the remaining samples per class. The AL procedure is iterated 80 times, conducting three stages: i) model training over $\mathcal{D}_{train}$ with 100 epochs, validating in each epoch with $\mathcal{D}_{val}$ and employing Adam optimizer with learning rate $1e-3$, ii) evaluation of $\mathcal{D}_{pool}$, and iii) ranking and selection of 10 unlabeled samples from $\mathcal{D}_{pool}$ to be included in $\mathcal{D}_{train}$. For this purpose, the proposed method obtains (for each sample in $\mathcal{D}_{pool}$) the corresponding activity vectors, and



Color	Land cover type	Samples
	Background	10659
■	Alfalfa	54
■	Corn-notill	1434
■	Corn-min	834
■	Corn	234
■	Grass/Pasture	497
■	Grass/Trees	747
■	Grass/pasture-mowed	26
■	Hay-windrowed	489
■	Oats	20
■	Soybeans-notill	968
■	Soybeans-min	2468
■	Soybean-clean	614
■	Wheat	212
■	Woods	1294
■	Bldg-Grass-Tree-Drives	380
■	Stone-steel towers	95
	Total samples	21025

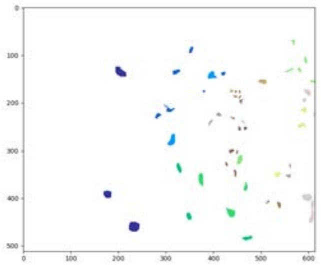
Fig. 2. Available labeled samples in the IP scene.

constructs $\mathbf{z} \in \mathbb{R}^K$ with their lengths, which can be easily understood as the probability vector of the CapsNet for each input sample. Then, the BT criterion is applied over $\mathbf{z}$ in order to decide which unlabeled samples $\mathbf{x} \in \mathcal{D}_{pool}$ will be queried by an external *oracle* and included into the $\mathcal{D}_{train}$ set in the next iteration. This acquisition function intends to select those samples that provide the most discriminative information to the classifier, being focused on the boundary region between two classes with the aim of obtaining more diversity in the composition of $\mathcal{D}_{train}$. The ten best ranked samples are then selected, labeled, and included into the training set. The final inference stage is conducted over the $\mathcal{D}_{test}$ set.

3. EXPERIMENTAL RESULTS

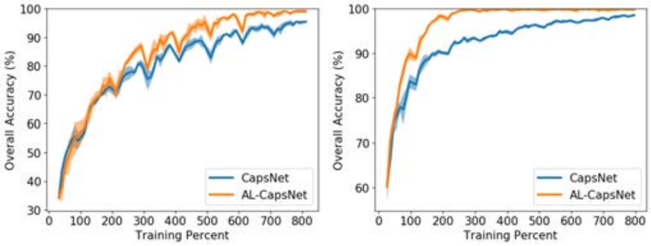
In order to discuss the performance of the proposed method, an experimental comparison has been conducted between an AL-based CapsNet model trained with the BT criterion, and simple random selection as acquisition function. Two widely used HSI scenes gathered by the Airborne Visible Infra-Red Imaging Spectrometer (AVIRIS) have been considered: Indian Pines (IP) and Kenedy Space Center (KSC) images. The first one (see Fig. 2) contains 145×145 samples of crops and forest, with 16 land-cover classes and 200 spectral bands in the range 0.2-2.4 microns, with spatial resolution of 20 meters per pixel. The second one (see Fig. 3) is comprised by 512×614 pixels with 176 bands and 13 ground-truth classes. To assess the classification performance, the overall (OA) and average (AA) accuracy, as well as the kappa coefficient, have been measured considering 5 Monte Carlo runs.

Fig. 4 shows the OA evolution as the training set grows with the new samples selected from the pool of unlabeled samples $\mathbf{D}_{pool}$. In both scenes, the BT criterion (AL-CapsNet) is able to reach very good performance. For instance, in the KSC experiment, only 30 iterations of the AL algorithm (i.e., $30 \cdot 10 + (2 \cdot K) = 326$ labeled samples) are neede to obtain the maximum (100%) OA. This represents only 15.98% of the total number of samples labeled for the KSC dataset. The proposed method also obtains excellent results with the IP scene, reaching close-to-optimal performance after 80 AL iterations (832 samples, only 12.45% of the available ground-truth). This means that the proposed AL-based sampling criterion can accurately identify the most discriminative samples from the available labeled ones.



Color	Land cover type	Samples
	Background	309157
■	Scrub	761
■	Willow-swamp	243
■	CP-hammock	256
■	Slash-pine	252
■	Oak/Broadleaf	161
■	Hardwood	229
■	Swap	105
■	Graminoid-marsh	431
■	Spartina-marsh	520
■	Cattail-marsh	404
■	Salt-marsh	419
■	Mud-flats	503
■	Water	927
	Total samples	314368

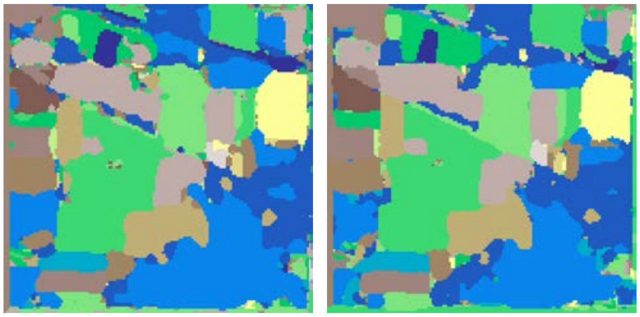
Fig. 3. Available labeled samples in the KSC scene.



(a) IP scene

(b) KSC scene

Fig. 4. OA evolution with the number of labeled samples.



CapsNet (94.43%)

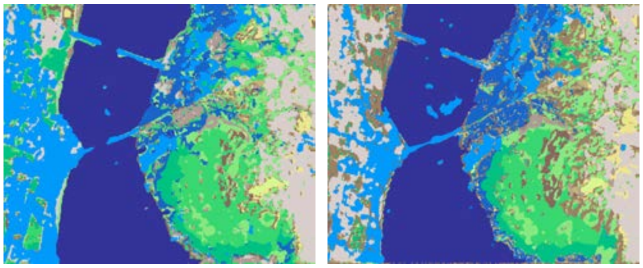
AL-CapsNet (**98.85%**)

Fig. 5. Classification maps for IP scene (80 AL iterations).

Table 2 explains in detail the classification results obtained after 80 AL iterations (IP scene) and 30 AL iterations (KSC scene). The table shows that the application of an

Class	**IP scene**		**KSC scene**	
	CapsNet	**AL-CapsNet**	**CapsNet**	**AL-CapsNet**
0	86.25±12.44	**97.50**±2.50	99.88±0.23	**100.0**±0.00
1	94.00±3.65	**97.82**±1.40	69.73±4.22	**98.72**±0.73
2	93.95±2.93	**98.05**±1.19	81.93±6.04	**100.0**±0.00
3	86.79±1.76	**99.06**±1.63	56.25±2.65	**98.39**±1.04
4	94.82±1.23	**99.66**±0.38	74.93±3.61	**98.59**±2.82
5	98.78±0.00	**99.85**±0.26	85.29±5.23	**99.22**±1.14
6	95.83±7.22	**100.0**±0.00	99.57±0.85	**100.0**±0.00
7	**100.0**±0.00	99.53±0.57	96.37±2.64	**100.0**±0.00
8	87.50±0.00	**100.0**±0.00	99.91±0.17	**100.0**±0.00
9	92.14±1.79	**99.54**±0.36	99.34±0.64	**100.0**±0.00
10	95.24±2.05	**98.26**±1.80	99.89±0.21	**100.0**±0.00
11	81.30±10.0	**98.59**±0.86	96.11±0.33	**99.91**±0.18
12	100.0±0.00	100.0±0.00	100.0±0.00	100.0±0.00
13	99.34±0.34	**99.74**±0.36	-	-
14	87.72±1.43	**99.71**±0.29	-	-
15	**100.0**±0.00	98.17±2.02	-	-
OA	94.43±0.94	**98.85**±0.38	93.43±0.41	**99.78**±0.14
AA	93.35±1.60	**99.09**±0.21	89.17±0.50	**99.60**±0.28
K(x100)	93.65±1.07	**98.69**±0.43	92.68±0.46	**99.75**±0.16

Table 2. Accuracy results for the IP (80 AL iterations, 832 samples) and KSC (30 AL iterations, 326 samples) scenes.



CapsNet (93.43%)

AL-CapsNet (**99.78%**)

Fig. 6. Classification maps for KSC scene (30 AL iterations).

AL-based criterion significantly increases the OA results by the CapsNet. In addition, the AL-based criterion exhibits a more robust behaviour in terms of standard deviation. Figs. 5 and 6 respectively provide the classification maps obtained in the aforementioned cases. As we can observe, the classification maps are more homogeneous spatially. Moreover, if we compare the two classification maps in Fig. 5 (IP scene), we can see that our proposed AL-based method provides a better delineation of the borders of the classes. Although the limited labeled data available for the KSC scene makes more difficult the interpretation of the classification maps in Fig. 6, it can be also observed that the location and distribution of the considered land-cover classes is more balanced. For instance, the classification map obtained by the proposed AL-CapsNet contains large areas of scrubs and slash-pine at the leftmost part of the scene, which is more realistic than the other displayed map according to ground knowledge.

Finally, Table 3 provides a comparison of the proposed AL-CapsNet with other AL-based classifiers using the KSC (30 iterations). The table shows that the proposed model is able to reach the best OA with the fewest amount of labeled samples, when compared with the probabilistic AL-based CNN and multinomial logistic regresion (MLR) in [6].

Algorithm	**Overall Accuracy**						
	70%	75%	80%	85%	90%	95%	99%
AL-MLR [6]	**26**	**36**	66	116	216	616	–
AL-CNN [6]	56	86	96	126	166	206	276
AL-CapsNet	36	46	66	**76**	**126**	**156**	**236**

Table 3. Number of samples that different AL-based models need to reach a given % of accuracy (KSC scene). Best model in bold and second-best model in blue.

4. CONCLUSIONS

This paper presents a new AL-based CapsNet model for HSI classification using spectral and spatial features. The proposed approach successfully captures the uncertainty of the data, offering robustness to overfitting with small training sets and improving the generalization ability by including intelligently selected unlabeled training samples (avoiding the curse of dimensionality). As future work, we will model further the epistemic uncertainty of the network.

5. REFERENCES

[1] A.F.H. Goetz, G. Vane, J.E. Solomon, and B.N. Rock, “Imaging Spectrometry for Earth Remote Sensing,” *Science*, vol. 228, no. 4704, pp. 1147–1153, 1985.

[2] M.E. Paoletti, J.M. 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] 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.

[4] S. Sabour, N. Frosst, and G.E. Hinton, “Dynamic routing between capsules,” in *Advances in neural information processing systems*, 2017, pp. 3856–3866.

[5] 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, 2018.

[6] J.M. Haut, M.E. Paoletti, J. Plaza, J. Li, and A. Plaza, “Active learning with convolutional neural networks for hyperspectral image classification using a new bayesian approach,” *IEEE Transactions on Geoscience and Remote Sensing*, vol. 56, no. 11, pp. 6440–6461, 2018.

[7] J. Li, J.M. Bioucas-Dias, and A. Plaza, “Hyperspectral image segmentation using a new bayesian approach with active learning,” *IEEE Transactions on Geoscience and Remote Sensing*, vol. 49, no. 10, pp. 3947–3960, 2011.