Hyperspectral Image Classification Using Deep Matrix Capsules

Anirudh Ravikumar
Information Science & Engineering
Ramaiah Institute of Technology
Bengaluru, India
anirudhr40@gmail.com

Rohit P N
Information Science & Engineering
Ramaiah Institute of Technology
Bengaluru, India
rohitnagraj.99@gmail.com

Mydhili K Nair
Professor & Head, B.Tech
School of Computer Science & Engg, RV University
Bengaluru, India
mydhilinair@rvu.edu.in

Vimal Bhatia
Professor, Electrical Engineering
IIT Indore
Indore, India
vbhatia@iiti.ac.in

2022 International Conference on Data Science, Agents & Artificial Intelligence (ICDSAAI) | 979-8-3503-3384-8/22/$31.00 ©2022 IEEE | DOI: 10.1109/ICDSAAI55433.2022.10028853

***Abstract*—Hyperspectral image (HSI) classification is used in multiple domains like precision agriculture, mineral exploration, remote sensing, and others. Conventionally, convolutional neural networks (CNNs) were used in HSI classification, however they have limitations in exploiting spectral-spatial relationships, which is a key factor in understanding HSI. Even though deeper CNN architectures and use of 3-D-CNNs mitigate the above problem to a certain extent, they have increased computational complexity, which inhibits their use in resource-limited devices like IoT and edge computing devices. In this paper, we propose a novel method based on the concept of matrix capsules with Expectation-Maximization (EM) routing algorithm which is specifically designed to accommodate the nuances in the HSI data to efficiently tackle the aforementioned problems. The capsule units enable effective identification of spectral signatures and part-whole relationships in the data while EM routing ensures viewpoint-invariance. Three representative HSI data sets are used to verify the effectiveness of the proposed method. The empirical results demonstrate that the proposed method is better than the current state-of-the-art methods in terms of accuracy while having 25 times fewer model parameters and requiring over 65 times less storage space. The source code can be found at https://github.com/DeepMatrixCapsules/DeepMatrixCapsules.**

Keywords—Capsule network, convolutional neural network (CNN), classification, deep learning, expectation-maximization (EM) routing, hyperspectral image (HSI) classification, matrix capsule, remote sensing, spectral–spatial, viewpoint-invariance.

I. INTRODUCTION

Hyperspectral images (HSI) are images captured across hundreds of frequencies that convey complex spectral-spatial relationships which can uncover pivotal information for the given region of interest. It has numerous applications in fields of remote sensing [1]–[3], precision agriculture [4]–[6], mineral exploration [7] and many more [8], [9]. One of the primary components employed in the above fields is the classification of HSI images, sometimes with limited computing resources. However, the presence of hundreds, if not thousands, of bands containing coherent spectral-spatial information makes it an extremely challenging task to accurately classify HSI images as a consequence of the curse of dimensionality [10], [11]. In addition to the above challenges, the supply of labelled training samples is limited due to the high cost of labelling the massive amount of data.

Several approaches have been used for the classification of HSI among which, SVMs were used as the primary classifiers for the methods that employed manual feature detection techniques [12]–[16]. However, the traditional approaches are affected by the Hughes phenomenon [17] as the spectral resolution increases which causes the accuracy to first increase with the number of spectral bands before it decreases drastically. To overcome the deficiencies caused by Hughes phenomenon, a number of new methods to decrease the dimensionality were introduced [18]–[21]. Other popular hand designed feature extraction approaches are fusing correlation coefficient, multiscale superpixels and guided filter and so on [22]–[27]. Most of these techniques are computationally inexpensive; however, they lack the performance required for specific applications.

Recently, convolutional neural networks (CNNs) have become the preferred approach due to the performance gain over the conventional machine learning-based approaches [28], [29]. It has shown promising results in a variety of applications in image processing, such as classification of images, depth estimation, face anti-spoofing, image denoising etc. [30]–[40]. Further, HSI classification has, also made a huge progress with the help of CNNs [41]–[45]. 3-D-CNNs were employed by [47] to extract features for HSI classification. An attention-based mechanism was developed by [48] for the classification of HSI images that use attention maps to optimize the extracted feature maps. Roy *et al.* [49] introduced HybridSN which employs 3-D-CNN targeted for the spectral bands and a 2-D-CNN aiming to extract spatial information in the HSI images. SpectralNET is put forward by [50] which employs a wavelet CNN architecture for HSI classification.

For HSI classification tasks, CNNs have proven to be a useful tool. However, CNN-based techniques face a few problems. The first being, as the structure deepens, gradients gradually vanish in the forward propagation phase, resulting in a slow convergence rate. Second, CNNs employ pooling, which invariably lowers the quantity of spatial information owing to downsampling. Third, the CNN models are resource heavy, both in terms of compute and storage owing to their enormous number of parameters. Finally, CNNs are unable to identify object pose information since the convolutional filters are incapable of representing activities linked to transformation of features. Various data

augmentation approaches have been used by [47] aiming to increase the robustness of CNNs to spatial transformations and to train on small data sets. However, these approaches fail to fundamentally learn and understand object poses and relationships. Due to the aforementioned reasons, CNNs are incapable of modelling relative object relationships effectively and efficiently.

To overcome the limitations of conventional CNN-based methods, [51] introduced the idea of a capsule. A capsule is a collection of neurons that describe an entity's pose and probability of existence. As a result, the capsule includes more information about the entity's properties than a scalar neuron in a CNN. Sabour *et al.* [52] proposed dynamic routing between capsules to ensure the effective use of information stored in the capsule neurons. This network called the capsule network (CapsNet) encodes the likelihood of the presence of an object and the pose of the object using the length and orientation of an activity vector, respectively. Considering the above advantages of capsules networks, [53] utilized it for the classification of HSI.

Furthermore, Hinton *et al.* [54] proposed matrix capsules with EM routing which overcome certain deficiencies of the dynamic routing algorithm by [52] as follows:

1. To get the agreement between the two pose vectors, dynamic routing [52] employs cosine of the angle between them. Though this worked well when the difference between the pose vectors was significant, it failed to capture the intricate differences between a good agreement and a great agreement due to the cosine function saturating at 1. To overcome this, matrix capsules with EM routing [54] uses the negative log variance of a Gaussian cluster to measure the agreement.
2. In dynamic routing [52], a pose is represented by a vector of length n which results in a transformation matrix with n^2 elements. On the other hand, matrix capsules with EM routing [54] represents a pose using a matrix with only n elements.

With the aforementioned advantages, an innovative method is introduced in this work based on the notion of capsule networks using EM routing [54] which is specifically designed to accommodate the nuances in the HSI data enabling the architecture to effectively and efficiently capture spectral signatures and part-whole relationships while maintaining viewpoint-invariance. This results in interpretation of the HSI characteristics at a higher abstraction level compared to the other methods. The following section describes the proposed method.

II. Proposed Method

We propose a novel method developed on the concept of matrix capsules using EM routing with the objective of capturing the highest degrees of spectral-spatial non-linearity present in the HSI data effectively while maintaining the size and complexity of the model.

A. Data Processing

An HSI image is represented in the form of a data cube $\mathbf{X}$ which spans over 3 dimensions, $\mathbf{X} \in \mathbb{R}^{H\times W\times C}$ where $\mathbf{X}$ is the original HSI image, where H, W, C represent the height, width and number of spectral bands respectively, i.e. a group of H×W pixels, each representing a target in $\mathbb{R}^C$. Every HSI pixel in $\mathbf{X}$ maps to the land-cover category in the form of a scalar $y = i \in \mathbb{R}$, $1 \leqslant i \leqslant D$ the total number of land cover categories is denoted by D. Generally, the HSI contain a large number of spectral dimensions leading to spectral sparsity. To tackle this, we apply principal component analysis (PCA), a dimensionality reduction technique on $\mathbf{X}$ along the spectral dimension. This results in a data cube $\mathbf{X} \in \mathbb{R}^{H\times W\times B}$, where $B \leqslant C$ while spatial information remains unaltered. Further, we create S×S overlapping 3-D patches by dividing the data cube $\mathbf{X} \in \mathbb{R}^{H\times W\times B}$ into $\mathbf{P} \in \mathbb{R}^{S\times S\times B}$, each centered around (α, β) in the spatial dimensions and spanning all B spectral bands. These patches have the same label as the central pixel $x_{\alpha,\beta}$ and are passed to the model as the input data. These image cubes provide better receptive field and context to the model.

B. Capsule Network

The Capsule network is made up of multiple matrix capsules. The set of matrix capsules in each layer L is depicted by Ω_L. A matrix capsule i in the layer L captures the activation probability a_i, along with a 4×4 pose matrix $\mathbf{M}_i$. The pose matrix depicts an object's translation and rotation, which is equivalent to a change in the object's viewpoint. A viewpoint invariant transformation matrix $\mathbf{W}_{ij}$ exists between a capsule i in Ω_L and capsule j in the Ω_{L+1} which is learned during the training process through backpropagation and it represents the part-whole relationships in the image. A capsule i in Ω_L votes for the pose matrix of the capsules j in the Ω_{L+1}. The viewpoint invariant transformation matrix $\mathbf{W}_{ij}$ is multiplied by the pose matrix of a capsule i denoted by $\mathbf{M}_i$ to determine the vectorized vote $\mathbf{v}_{ij}$ ($\mathbf{v}_{ij} = \mathbf{M}_i \times \mathbf{W}_{ij}$). A weight is assigned to the votes $\mathbf{v}_{ij}$ by multiplying it with an assignment coefficient R_{ij}. The Expectation-Maximization (EM) algorithm is then used to update the coefficients R_{ij} iteratively for each data point. This is calculated so that the output of each capsule in layer L is routed to a capsule in the next layer $L + 1$ that receives a cluster of similar votes.

1) EM Routing: The EM algorithm, as described in procedure 1, is used to finalise the pose parameters and capsule activations in Ω_{L+1}. The EM algorithm uses two alternating steps called the E-step and the M-step with the goal of trying to fit a mixture of Gaussian distributions to the data. The assignment probability R_{ij} of each capsule i in Ω_L to a capsule j in Ω_{L+1} is determined by the E-step. Further, the Gaussian distributions' parameters are re-calculated in the M-step based on R_{ij} . This iteration is repeated until convergence, to get activation a_j of each capsule j in Ω_{L+1}.

The inputs a and V in Procedure 1 are the activations and the capsules' votes in the layer L respectively. In M-step, we calculate the assignment probabilities R_{ij} , mean of capsule j's Gaussian model μ^h_j and variance $(\sigma^h_j)^2$ based on the activation a_i of capsules from Ω_L and the votes V foreach dimension h. M-step also re-calculates the $cost^h$ using the costs β_u and β_a where β_u is the cost generated when a capsule is not activated and β_a is the cost incurred when a capsule is activated to encode its mean (μ) and variance (σ^2). Subsequently, the activation a_j is calculated for capsule in layer $L + 1$. In E-step, the new μ, σ and a_j are used to

recalculate the assignment probability R_{ij} . As the vote gets closer to the μ of the updated Gaussian distribution, there is a corresponding increase to the value of the assignment probability.

prediction and the margin. The margin m begins at 0.2 and linearly increases to 0.5 with an increase of 0.05 at each epoch. Dead capsules are avoided in the early training phase by using a low margin at the beginning. Spread loss is same

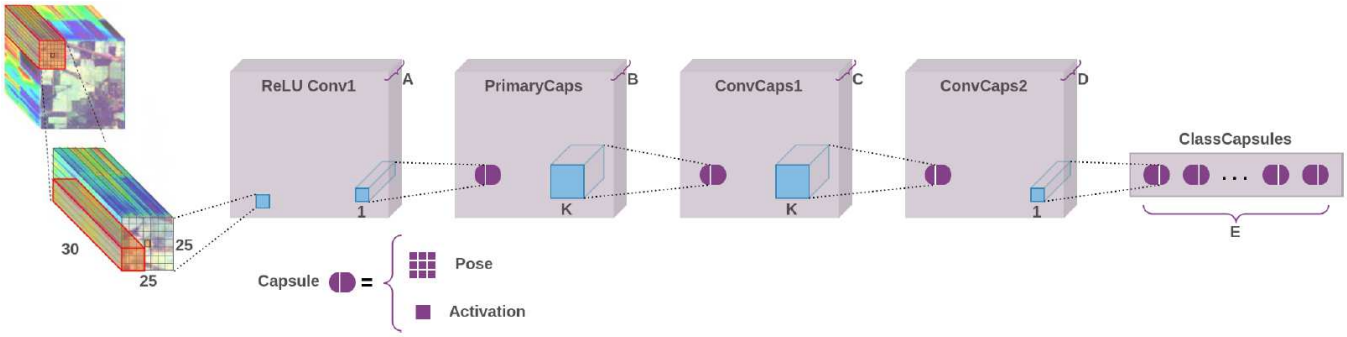


Fig. 1. Proposed Deep Matrix Capsule architecture for HSI Classification.

Procedure 1 EM Routing Algorithm

procedure EM ROUTING$(\mathbf{a}, \mathbf{V})$
  $\forall i \in \Omega_L, j \in \Omega_{L+1}: R_{ij} \leftarrow 1/|\Omega_{L+1}|$
  for t iterations **do**
    $\forall j \in \Omega_{L+1}:$ M-STEP$(\mathbf{a}, \mathbf{R}, \mathbf{V}, j)$
    $\forall i \in \Omega_L:$ E-STEP$(\mu, \sigma, \mathbf{a}, \mathbf{V}, i)$
  end for
  return $\mathbf{a}, \mathbf{M}$

procedure M-STEP$(\mathbf{a}, \mathbf{R}, \mathbf{V}, j)$
  $\forall i \in \Omega_L: R_{ij} \leftarrow R_{ij} * a_i$
  $\forall h: \mu_j^h \leftarrow \frac{\sum_i R_{ij} V_{ij}^h}{\sum_i R_{ij}}$
  $\forall h: (\sigma_j^h)^2 \leftarrow \frac{\sum_i R_{ij} (V_{ij}^h - \mu_j^h)^2}{\sum_i R_{ij}}$
  $cost^h \leftarrow \left(\beta_u + \log(\sigma_j^h)\right) \sum_i R_{ij}$
  $a_j \leftarrow logistic\left(\lambda(\beta_a - \sum_h cost^h)\right)$

procedure E-STEP$(\mu, \sigma, \mathbf{a}, \mathbf{V}, i)$
  $\forall j \in \Omega_{L+1}: p_j \leftarrow \frac{1}{\sqrt{\prod_h^H 2\pi(\sigma_j^h)^2}} e^{-\left(\sum_h^H \frac{(V_{ij}^h - \mu_j^h)^2}{2(\sigma_j^h)^2}\right)}$
  $\forall j \in \Omega_{L+1}: R_{ij} \leftarrow \frac{a_j p_j}{\sum_{k \in \Omega_{L+1}} a_k p_k}$

TABLE I
ARCHITECTURE OF DEEP MATRIX CAPSULESWITH WINDOW SIZE 25×25. THE LAST LAYER IS BASED ON THE INDIAN PINES DATA SET

INPUT CONVOLUTION LAYER					
Layer ID	Input Shape	Kernel Size	Stride	Padding	Output Shape
conv1	25×25×30	5×5	2	2	13×13×64
PRIMARY CAPSULE LAYER (PrimaryCaps)					
Layer ID	Input Shape	Kernel Size	Stride	Padding	Output Shape
pose	13×13×64	1×1	1	0	13×13×32×4×4
a	13×13×64	1×1	1	0	13×13×32
CONVOLUTIONAL CAPSULE LAYER 1 (ConvCaps-1)					
Layer ID	Input Shape	Kernel Size	Stride	Padding	Output Shape
pose	13×13×32×4×4	3×3	2	0	6×6×16×4×4
a	13×13×32	3×3	2	0	6×6×16
CONVOLUTIONAL CAPSULE LAYER 2 (ConvCaps-2)					
Layer ID	Input Shape	Kernel Size	Stride	Padding	Output Shape
pose	6×6×16×4×4	3×3	1	0	4×4×16×4×4
a	6×6×16	3×3	1	0	4×4×16
CLASS CAPSULE LAYER (ClassCaps)					
Layer ID	Input Shape	Kernel Size	Stride	Padding	Output Shape
pose	4×4×16×4×4	1×1	1	0	4×4×16
a	4×4×16	1×1	1	0	16

2) Loss Function: We use spread loss to train the transformation matrix **W** and to calculate the costs β_u and β_a during backpropagation. Spread loss is utilized as it dampens the effect of the initialization of weights of the model. Further, the impact of hyper-parameters of the model is also reduced on the training process with the use of spread loss, easing the process of hyper-parameter tuning as a result. The spread loss for class i is calculated as:

$$L_i = (max(0, m - (a_t - a_i)))^2 \quad (1)$$

where the margin, the activation of target class t and the activation of class i are denoted by m, a_t and a_i respectively. The total cost is depicted by:

$$L = \sum_{i \neq t} L_i \quad (2)$$

If the difference between the prediction of the true class and that of the other classes is less than the margin m, the model is penalized by the squared difference between the as the squared hinge loss with the margin equal to 1.

3) Architecture: Fig. 1 depicts the proposed architecture, which is divided into four main components.

a) The first convolutional layer (ReLU Conv1) processes the input data cube with a 5×5 kernel with 30 input channels and A = 64 output channels. It has a stride of 2, paired with ReLU as the activation function to extract features of interest from the 3-D data cube.

b) These features are then passed into a primary capsule layer (PrimaryCaps). The primary capsule layer transforms each of the A = 64 channels into B = 32 primary capsules each comprising a scalar activation value and a 4×4 pose matrix. The 4×4 pose matrix of each of the B = 32 primary capsules is a learned linear transformation of the output of all the previous layer ReLU activations centred at that location. The activations of the primary capsules are calculated by applying the sigmoid function to the weighted sums of the same set of previous layer ReLU activations.

c) Further, the output of the PrimaryCaps is passed to ConvCaps1 with $C = 16$ convolutional capsules using a 3×3 kernel ($K = 3$) with stride 2. A convolutional capsule layer is similar to a conventional convolutional layer, except that it computes the output of the capsule via EM routing. The output from ConvCaps1 is then fed into the next convolutional capsule layer (ConvCaps2) with $D = 16$ convolutional capsules and stride 1.

d) Lastly, ConvCaps2's outputs are directed to $E = 16$ class capsules using a 1×1 kernel and it outputs one capsule per class. The predicted class is obtained by applying the softmax function on the output activations from the class capsules.

The architecture of the proposed method is depicted in Table I with window size 25×25 which is utilized in the experiments for the Indian Pines data set.

III. Experiments

A. Data Set Description

During the course of the experiments, three HSI data sets[1] were utilized, namely Indian Pines, University of Pavia, and Salinas Scene.

AVIRIS sensor was used to accumulate the data from various agricultural areas in Northwestern Indiana, USA to constitute the Indian Pines data set. The Indian Pines consists of 200 useful spectral bands with an image resolution of 145×145 pixels. 16 mutually exclusive vegetation classes are identified as the available ground truth.

University of Pavia data set consists of 610×340 pixels, having 103 spectral bands captured over University Of Pavia, Northern Italy by the ROSIS sensor. 9 mutually exclusion urban land-cover classes are identified as the ground truth.

Salinas Scene data set consists of 512×217 pixels, having 200 spectral bands captured over California, USA using AVIRIS sensor. The ground truth consists of 16 mutually exclusive classes.

B. Experiment Settings

All the experiments were carried out on a cloud-based virtual machine allocated with an NVIDIA Tesla P100 GPU. The experimental framework consists of PyTorch v1.6.0 with cuda 10.1. The classification results were used to optimize the learning rate to find the optimal value of 3×10^{-3}.

C. Classification Results

In this work, we validate the performance of the classification of HSI using the following performance metrics :

1) Overall Accuracy (OA)
2) Average Accuracy (AA)
3) Kappa Coefficient (Kappa), which is a statistic that is used to measure inter-rater reliability between the classification map and ground truth map.

The training, validation and testing data sets are created by splitting the data into 50%, 10%, and 40% buckets respectively.

The OA, AA and Kappa metrics are compared for the Deep Matrix Capsules with 2-D-CNN, 3-D-CNN, DBDA, Capsule Network, HybridSN and SpectralNET over the three representative data sets in Table II. From Table II, it can be deduced that Deep Matrix Capsules perform better than all the other methods on the representative HSI data sets. The classification performance of Deep Matrix Capsules demonstrate the effectiveness of capsules in identifying spectral signatures and part-whole relationships in the data while EM routing ensures viewpoint-invariance. To arrive at this conclusion, we used the publicly available code [2] of the methods compared.

Fig. 2-4 show the classification maps for 2-D-CNN, 3-DCNN, DBDA, Capsule Network, HybridSN, SpectralNET and Deep Matrix Capsules for the Indian Pines, Salinas Scene and University of Pavia data sets. The classification maps clearly depict that Capsule Network, HybridSN, SpectralNET and Deep Matrix Capsules are far superior compared to the other techniques. On closer inspection, the Deep Matrix Capsules' classification maps has clearer separation of regions compared to the maps of Capsule Network, HybridSN and SpectralNET.

The loss and accuracy curves for 50 epochs are portrayed in Fig. 5, which shows that the validation loss gradually decreases till the 50th epoch and the validation accuracy steadily increases till 45 epochs. The training time for 50 epochs on the Indian Pines data set was approximately 47 minutes with the aforementioned experimental settings.

Empirical results of the classification performance for Deep Matrix Capsules across various spatial window sizes is shown in Table III. The classification performance from Table III shows that 25×25, 25×25 and 19×19 window sizes are optimal for Indian Pines, Salinas Scene and University of Pavia data sets respectively. Table IV compares the number of parameters and the model size over the three representative data sets for HybridSN, SpectralNET and Deep Matrix Capsules. It can be observed that the Deep Matrix Capsules have a significant reduction in the number of parameters, by over 25 times, and a reduction of over 65 times in the size of the model stored on the disk compared to the current state of-the-art methods. This is achieved while improving on the classification performance, leading to not only highly accurate inferencing, but also cheaper and more efficient storage.

IV. Conclusion

A novel method, namely Deep Matrix Capsules, was developed on the idea of matrix capsules with EM routing, with the objective of classifying remotely sensed HSI data effectively and efficiently. Deep Matrix Capsules consists of matrix capsule units that allow effective extraction of abstract spectral-spatial features in the remotely sensed data. The effectiveness can be attributed to the ability of capsules to capture complex spectral signatures and part-whole relationships while EM routing ensures viewpoint-invariance. Consequently, the information gain resulting from the highly effective feature extraction by the Deep Matrix Capsules, leads to a computationally simpler

[1] http://www.ehu.eus/ccwintco/index.php/Hyperspectral_Remote_Sensing_Scenes

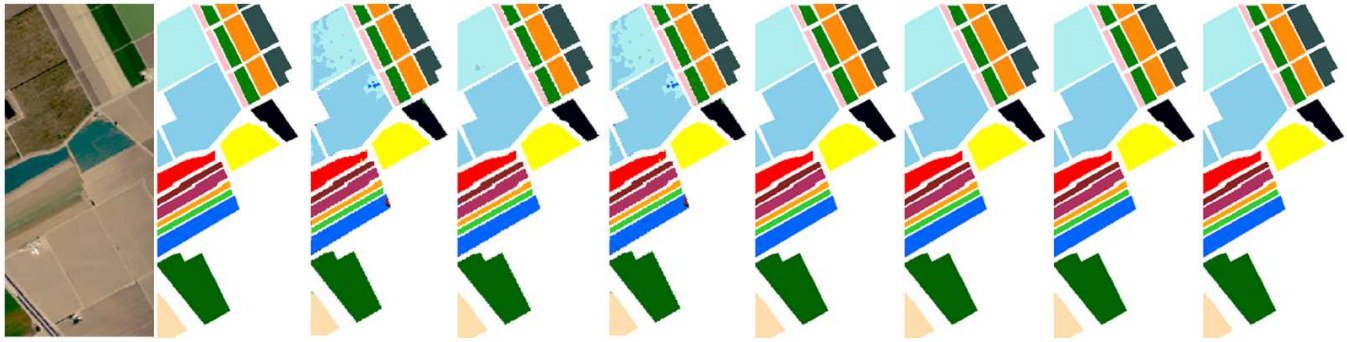
[2] https://github.com/eecn/Hyperspectral-Classification, https://github.com/lironui/Double-Branch-Dual-Attention-Mechanism-Network, https://github.com/gokriznastic/HybridSN, https://github.com/tanmay-ty/SpectralNET

TABLE II
CLASSIFICATION RESULTS (IN PERCENTAGE) ON THE INDIAN PINES, SALINAS SCENE AND UNIVERSITY OF PAVIA DATA SETS

Methods	Indian Pines Dataset			Salinas Scene Dataset			University of Pavia		
	OA	Kappa	AA	OA	Kappa	AA	OA	Kappa	AA
2-D-CNN	89.48 ± 0.20	87.96 ± 0.50	86.14 ± 0.80	97.38 ± 0.07	97.08 ± 0.10	98.84 ± 0.10	97.86 ± 0.20	97.16 ± 0.50	96.55 ± 0.10
3-D-CNN	91.10 ± 0.40	89.98 ± 0.50	91.58 ± 0.20	93.96 ± 0.20	93.32 ± 0.50	97.01 ± 0.60	96.53 ± 0.10	95.51 ± 0.20	97.57 ± 1.30
DBDA	95.38 ± 0.32	94.74 ± 0.40	96.47 ± 0.24	97.51 ± 0.21	97.23 ± 0.35	98.00 ± 0.40	96.00 ± 0.52	94.67 ± 0.42	96.45 ± 0.15
Capsule Network	99.45 ± 0.13	99.37 ± 0.14	99.34 ± 0.40	99.81 ± 0.03	99.79 ± 0.03	99.92 ± 0.01	99.95 ± 0.02	99.93 ± 0.03	99.90 ± 0.05
HybridSN	99.75 ± 0.10	99.71 ± 0.10	99.63 ± 0.20	100 ± 0.00	100 ± 0.00	100 ± 0.00	99.98 ± 0.00	99.98 ± 0.00	99.97 ± 0.00
SpectralNET	99.86 ± 0.20	99.84 ± 0.20	99.98 ± 0.10	100 ± 0.00	100 ± 0.00	100 ± 0.00	99.99 ± 0.10	99.98 ± 0.10	99.98 ± 0.10
Deep Matrix Capsules	99.93 ± 0.07	99.89 ± 0.10	99.98 ± 0.02	100 ± 0.00	100 ± 0.00	100 ± 0.00	99.99 ± 0.01	99.98 ± 0.02	99.99 ± 0.01

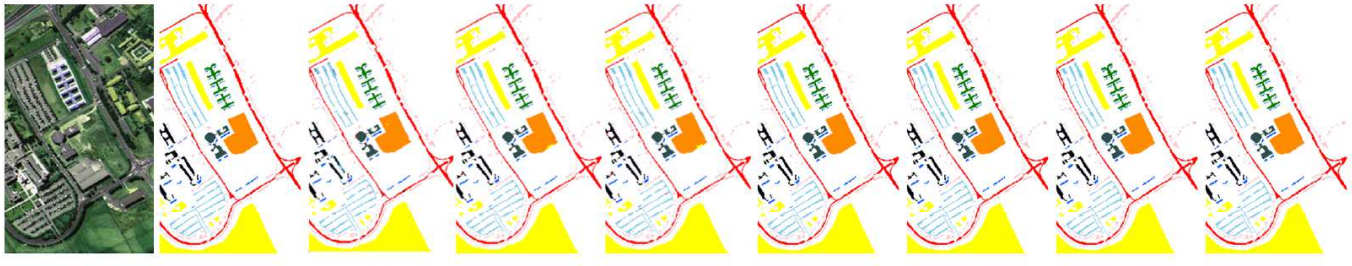
(a) (b) (c) (d) (e) (f) (g) (h) (i)

Fig. 2. Classification maps for the Indian Pines data set. (a) RGB composition. (b) Ground-truth classification map. (c)–(i) Classification maps corresponding to Table II [2-D-CNN, 3-D-CNN, DBDA, Capsule Network, HybidSN, SpectralNET and Deep Matrix Capsules respectively].



(a) (b) (c) (d) (e) (f) (g) (h) (i)

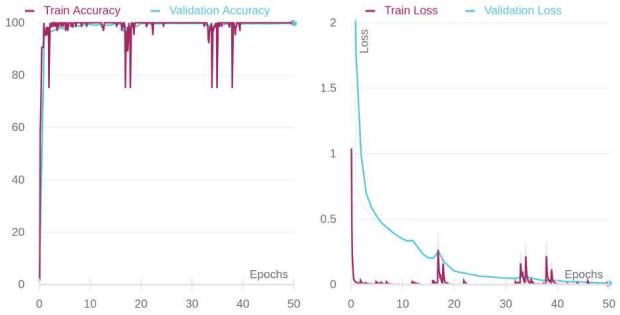
Fig. 3. Classification maps for the Salinas Scene data set. (a) RGB composition. (b) Ground-truth classification map. (c)–(i) Classification maps corresponding to Table II [2-D-CNN, 3-D-CNN, DBDA, Capsule Network, HybidSN, SpectralNET and Deep Matrix Capsules respectively].



(a) (b) (c) (d) (e) (f) (g) (h) (i)

Fig. 4. Classification maps for the University Of Pavia data set. (a) RGB composition. (b) Ground-truth classification map. (c)–(i) Classification maps corresponding to Table II [2-D-CNN, 3-D-CNN, DBDA, Capsule Network, HybidSN, SpectralNET and Deep Matrix Capsules respectively].

architecture, ensuring fast and accurate model convergence. This was empirically validated by using three benchmark HSI data sets namely, Indian Pines, Salinas Scene and University of Pavia. The experimental results show that



(a) (b)

Fig. 5. (a) Accuracy versus epochs and (b) loss versus epochs over the Indian Pines data set.

Deep Matrix Capsules achieve better results on all the benchmark data sets compared to the current state-of-the-art methods while having 25 times fewer parameters and requiring over 65 times less storage space. Accordingly, we draw a conclusion that Deep Matrix Capsules enable highly accurate inferencing while being more efficient in terms of compute and storage.

V. ACKNOWLEDGMENT

We thank IIT Indore and Grant Agency of Excellence, University of Hradec Kralove, Faculty of Informatics and Management, Czech Republic for all the support. We would also like to thank all the support and facilities provided by the Department of Information Science and Engineering, Ramaiah Institute of Technology, Bengaluru, India.

REFERENCES

[1] D. Landgrebe, "Hyperspectral image data analysis," in *IEEE Signal Processing Magazine*, vol. 19, no. 1, pp. 17–28, Jan. 2002.

[2] J. M. Bioucas-Dias *et al.*, "Hyperspectral remote sensing data analysis and future challenges," in *IEEE Geoscience and Remote Sensing Magazine*, vol. 1, no. 2, p. 6–36, Jun. 2013.

[3] G. Camps-Valls *et al.*, "Advances in hyperspectral image classification: Earth monitoring with statistical learning methods," in *IEEE Signal Processing Magazine*, vol. 31, no. 1, pg. 45–54, Jan. 2014.

[4] D. Cabellero *et al.*, "Chapter 3.3 - Hyperspectral imaging in crop fields: precision agriculture," in *Data Handling in Science and Technology*, vol. 32, pg. 453-473, 2019.

[5] X. Zhang *et al.*, "Crop classification based on feature band set construction and object-oriented approach using hyperspectral images," in *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 9, no. 9, pg. 4117–4128, Sep. 2016.

[6] K. Manjunath *et al.*, "Identification of indices for accurate estimation of anthocyanin and carotenoids in different species of flowers using hyperspectral data," in *Remote Sensing Letters*, vol. 7, no. 10, pg. 1004–1013, 2016.

[7] E. Bedini, "The use of hyperspectral remote sensing for mineral exploration: a review," *in Journal of Hyperspectral Remote Sensing*, vol. 7, no. 4, 2017.

[8] M. Fauvel *et al.*, "Advances in spectral-spatial classification of hyperspectral images," in *Proceedings of the IEEE*, vol. 101, no. 3, pg. 652-675, Mar. 2013.

[9] M.A. Calin *et al.*, "Hyperspectral Imaging in the Medical Field: Present and Future," in *Applied Spectroscopy Reviews*, pg. 435-447, 2014.

[10] Bellman, R., "Dynamic programming," in *Science*, 153(3731), pg. 34–37, 1966.

[11] P. Ghamisi *et al.* "Advanced spectral classifiers for hyperspectral images: A review," in *IEEE Geoscience and Remote Sensing Magazine*, Mar. 2017.

TABLE III
IMPACT OF THE SPATIAL WINDOW SIZE ON THE OVERALL ACCURACY (OA) OF DEEP MATRIX CAPSULES

Window Size	Indian Pines (%)	University Of Pavia (%)	Salinas Scene (%)
19 x 19	99.13	99.99	99.74
21 x 21	99.84	99.70	99.69
25 x 25	99.93	99.64	100.00
31 x 31	99.61	99.54	99.42

SALINAS SCENE AND UNIVERSITY OF PAVIA DATA SETS FOR HYBRIDSN, SPECTRALNET AND DEEP MATRIX CAPSULES

Model Name	Data set	Parameters	Model Size (kB)
HybridSN	Indian Pines	5,122,176	62,976
	Salinas Scene	5,122,176	62,976
	University of Pavia	4,662,322	57,344
SpectralNET	Indian Pines	6,805,072	54,736
	Salinas Scene	6,805,072	54,736
	University of Pavia	6,194,133	49,840
Deep Matrix Capsules	Indian Pines	198,336	800
	Salinas Scene	198,336	800
	University of Pavia	180,530	729

[12] L. Fang *et al.*, "A New Spatial–Spectral Feature Extraction Method for Hyperspectral Images Using Local Covariance Matrix Representation," in *IEEE Transactions on Geoscience and Remote Sensing*, vol. 56, pg. 3534-3546, Jun. 2018.

[13] G. Camps-Valls *et al.*, "Composite kernels for hyperspectral image classification," in *IEEE Geoscience and Remote Sensing Letters*, vol. 3, pg. 93-97, Jan. 2006.

[14] Osuna E. *et al.*, "Support Vector Machines: Training and Applications," in *A.I. Memo No. 1602, Artificial Intelligence Laboratory*, MIT, 1997.

[15] Zhang, Y. *et al.*, "Support Vector Machine Classification Algorithm and Its Application, Information Computing and Applications," in *ICICA*, 2012.

[16] X. Zhang *et al.*, "Crop classification based on feature band set construction and object-oriented approach using hyperspectral images," in *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, Sep. 2016.

[17] D. Landgrebe, "Hyperspectral image data analysis," in *IEEE Signal Processing Magazine*, vol. 19, pg. 17-28, Jan. 2002.

[18] X. Wei *et al.*, "Reconstructible Nonlinear Dimensionality Reduction via Joint Dictionary Learning," in *IEEE Transactions on Neural Networks and Learning Systems*, vol. 30, pg. 175-189, Jan. 2019.

[19] X. Wei, H. Shen and M. Kleinsteuber, "Trace Quotient Meets Sparsity: A Method for Learning Low Dimensional Image Representations," 2016 *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016, pp. 5268-5277.

[20] Q. Wang, Z. Meng and X. Li, "Locality Adaptive Discriminant Analysis for Spectral–Spatial Classification of Hyperspectral Images," in *IEEE Geoscience and Remote Sensing Letters*, vol. 14, no. 11, pp. 2077-2081, Nov. 2017.

[21] Q. Wang *et al.*, "Salient Band Selection for Hyperspectral Image Classification via Manifold Ranking," in *IEEE Transactions on Neural Networks and Learning Systems*, vol. 27, pg. 1279-1289, June 2016.

[22] Q. Gao, S. Lim and X. Jia, "Hyperspectral Image Classification Using Joint Sparse Model and Discontinuity Preserving Relaxation," in *IEEE Geoscience and Remote Sensing Letters*, vol. 15, no. 1, pp. 78-82, Jan. 2018.

[23] P. Gao, J. Wang, H. Zhang and Z. Li, "Boltzmann Entropy-Based Unsupervised Band Selection for Hyperspectral Image Classification," in *IEEE Geoscience and Remote Sensing Letters*, vol. 16, no. 3, pp. 462-466, March 2019.

[24] P. Hu *et al.*, "Band Selection of Hyperspectral Images Using Multiobjective Optimization-Based Sparse Self-Representation," in *IEEE Geoscience and Remote Sensing Letters*, vol. 16, March 2019.

[25] B. Tu *et al.*, "Hyperspectral Image Classification via Fusing Correlation Coefficient and Joint Sparse Representation," in *IEEE*

Geoscience and Remote Sensing Letters, vol. 15, pg. 340-344, March 2018.

[26] T. Dundar and T. Ince, "Sparse Representation-Based Hyperspectral Image Classification Using Multiscale Superpixels and Guided Filter," in *IEEE Geoscience and Remote Sensing Letters*, vol. 16, Feb. 2019.

[27] G. Camps-Valls *et al.*, "Advances in hyperspectral image classification: Earth monitoring with statistical learning methods," in *IEEE Signal Processing Magazine*, Jan. 2014.

[28] A. Krizhevsky *et al.*, "ImageNet classification with deep convolutional neural networks," in *Advances in Neural Information Processing Systems*, pg. 1097–1105, 2012.

[29] S. Albawi *et al.*, "Understanding of a convolutional neural network," in *International Conference on Engineering and Technology (ICET)*, 2017.

[30] Y. Li, "Research and Application of Deep Learning in Image Recognition," in *IEEE 2nd International Conference on Power, Electronics and Computer Applications (ICPECA)*, pg. 994-999, Mar. 2022.

[31] S. K. Roy *et al.*, "LiSHT: Non-parametric linearly scaled hyperbolic tangent activation function for neural networks," 2019, *arXiv:1901.05894*, unpublished.

[32] S. Ren, K. He, R. Girshick, and J. Sun, "Faster R-CNN: Towards real-time object detection with region proposal networks," *Advances in Neural Information Processing Systems*, 2015, pp. 91-99.

[33] K. He *et al.* "Mask R-CNN," 2017 *IEEE International Conference on Computer Vision (ICCV)*, pg. 2980-2988, 2017.

[34] X. Peng *et al.*, "SIDE: Center-Based Stereo 3D Detector With Structure-Aware Instance Depth Estimation," in *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)*, pg. 119-128, Jan. 2022.

[35] C. Nagpal and S. R. Dubey, "A performance evaluation of convolutional neural networks for face anti spoofing," in *Proceedings IEEE Internation Joint Conference Neural Networks (IJCNN)*, pg. 1-8, Mar. 2019.

[36] Sakshi Indolia *et al.*, "Conceptual Understanding of Convolutional Neural Network- A Deep Learning Approach," in *International Conference on Computational Intelligence and Data Science*, 2018.

[37] P. Kollapudi *et al.*, "A Novel Faster RCNN with ODN-Based Rain Removal Technique," in *Mathematical Problems in Engineering*, vol. 2022, Article ID 4546135, 11 pages, May 2022.

[38] Jair Cervantes, "A comprehensive survey on support vector machine classification: Applications, challenges and trends," in N*eurocomputing*, Volume 408, 2020.

[39] Y. Bazi and F. Melgani, "Gaussian process approach to remote sensing image classification," in *IEEE Transactions on Geoscience and Remote Sensing*, Jan. 2010.

[40] W. Li *et al.* , "Hyperspectral image classification using deep pixel-pair features," in *IEEE Transactions on Geoscience and Remote Sensing*, Feb. 2017.

[41] Y. Yu *et al.*, "An Unsupervised Convolutional Feature Fusion Network for Deep Representation of Remote Sensing Images," in *IEEE Geoscience and Remote Sensing Letters*, vol. 15, pg. 23-27, Jan. 2018.

[42] W. Li *et al.*, "Data Augmentation for Hyperspectral Image Classification With Deep CNN," in *IEEE Geoscience and Remote Sensing Letters*, vol. 16, no. 4, pp. 593-597, April 2019.

[43] W. Song *et al.*, "Hyperspectral Image Classification With Deep Feature Fusion Network," in *IEEE Transactions on Geoscience and Remote Sensing*, vol. 56, no. 6, pp. 3173-3184, June 2018.

[44] G. Cheng *et al.*, "Exploring Hierarchical Convolutional Features for Hyperspectral Image Classification," in *IEEE Transactions on Geoscience and Remote Sensing*, vol. 56, pg. 6712-6722, Nov. 2018.

[45] J. Yang *et al.*, "Hyperspectral image classification using two-channel deep convolutional neural network," in *Proceedings IEEE International Geoscience and Remote Sensing Symposium*, Jul. 2016.

[46] K. He *et al.*, "Deep residual learning for image recognition", in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2016.

[47] Y. Chen *et al.*, "Deep Feature Extraction and Classification of Hyperspectral Images Based on Convolutional Neural Networks," in *IEEE Transactions on Geoscience and Remote Sensing*, vol. 54, Oct. 2016.

[48] Li, R. *et al.*, "Classification of Hyperspectral Image Based on Double-Branch Dual-Attention Mechanism Network," in *Remote Sensing*, vol. 12, pg. 582, 2020.

[49] S. K. Roy *et al.*, "HybridSN: Exploring 3-D–2-D CNN Feature Hierarchy for Hyperspectral Image Classification," in *IEEE Geoscience and Remote Sensing Letters*, vol. 17, pg. 277-281, Feb. 2020.

[50] T. Chakraborty and U. Trehan., "SpectralNET: Exploring Spatial-Spectral WaveletCNN for Hyperspectral Image Classification," 2021, *arXiv:2104.00341*., unpublished.

[51] G. E. Hinton *et al.*, "Transforming Auto-Encoders" in *Proceedings of International Conference of Artificial Neural Networks*, pg. 4451, 2011.

[52] S. Sabour *et al.*, "Dynamic routing between capsules," in Proceedings of Advances in *Neural Information Processing Systems 30 (NIPS 2017)*, pg. 3859-3869, 2017.

[53] H. Zhang *et al.*, "1D-Convolutional Capsule Network for Hyperspectral Image Classification," 2019, *arXiv:1903.09834*, unpublished.

[54] G. E. Hinton *et al.*, "Matrix capsules with EM routing," in *International Conference on Learning Representations (ICLR)*, 2018.

[55] Baumgardner *et al.*, "220 Band AVIRIS Hyperspectral Image Data Set: June 12, 1992 Indian Pine Test Site 3," in *Purdue University Research Repository*, 2015.

[56] J. Hui, "Understanding Matrix capsules with EM Routing (Based on Hinton's Capsule Networks)", 2017, https://jhui.github.io/2017/11/14/Matrix-Capsules-with-EM-routing-Capsule-Network/.

[57] Q. Wang *et al.*, "Locality adaptive discriminant analysis for spectralspatial classification of hyperspectral images," in *IEEE Geoscience and Remote Sensing Letters*, Nov. 2017.

[58] T. Hahn *et al.*, "Self-Routing Capsule Networks," in *Part of Advances in Neural Information Processing Systems 32, NeurIPS*, 2019.

[59] V. Mazzia *et al.*, "Efficient-CapsNet: capsule network with self-attention routing," in *Sci Rep 11*, *14634*, 2021.

[60] L. Yang, M. Hasani, "Matrix Capsules with EM Routing," 2018, https://github.com/yl-1993/Matrix-Capsules-EM-PyTorch.