International Journal of
INTELLIGENT SYSTEMS AND APPLICATIONS IN ENGINEERING
ISSN:2147-6799 www.ijisae.org Original Research Paper

Hyperspectral Image Classification Using Dimensionality Reduction Deep Networks

[1]D. Kavitha, [2]S. Shobana, [3]S. Rajkumar, [4]Ganta Raghotham Reddy, [5]Dr. Ramesh Babu P., [6]Papiya Mukherjee

Submitted: 02/10/2023 **Revised**: 20/11/2023 **Accepted**: 02/12/2023

Abstract: In this research, we apply a convolutional neural network (CNN) to three publicly available hyperspectral datasets to determine which of these four models is the most effective when it comes to reducing the number of dimensions. The findings demonstrate that the models have a higher rate of classification accuracy on the smaller datasets when compared to the other techniques. It would appear from the observations that employing SuperPCA results in an overall improvement in the classifier level of effectiveness.

Keywords: *Convolutional neural network, image classification, hyperspectral imaging, dimensionality reduction*

1. Introduction

The quantity of information can be conceptualized along two dimensions, the first of which is the number of variables, and the second of which is the number of bits that are used to characterize each variable. This is one method to think about the volume of information. These two dimensions have the potential to acquire extremely high values, which can lead to complications during the process of analyzing and understanding the data [1].

Therefore, approaches to data processing that enhance an individual comprehension of the value contained within the data are an absolute requirement. The processing of high-dimensional data not only calls for a significant increase in the amount of time and memory that is accessible, but it also has the potential to reduce the efficiency of the algorithms that are used. This is because high-dimensional data are more difficult to organize and represent [2].

The reduction in the total number of parameters was one of the first suggestions that was put forward as a potential solution to this problem that was brought up. The objective here is to zero in on the elements of the design that are of the utmost importance and then either isolating them or doing away with them. By only including data that contributes to an increase in the precision and dependability of the dataset, the dimensionality of the problem can be made more manageable from the perspective of the signal space of the instances. This can be accomplished by excluding any data that does not contribute to either of those goals [3].

This simplification has two primary goals: the first is to get rid of unnecessary particulars, and the second is to make the data more accessible for additional analysis by using a broader variety of techniques. Dimensionality reduction has been shown to be capable of successfully projecting high-dimensional data to a lower-dimensional location while preserving most of the information that is intrinsic to the data as well as the information that differentiates the data from one another. Because of this, the data can be processed in a fashion that is both more easily manageable and more effective [4].

The spectral-spatial approaches to HSI classification have been found to be effective, most approaches continue to rely predominantly on shallow descriptors that are produced by hand. However, most custom-built features are intended to carry out a specific task and call for specialized knowledge during the process of parameter setup. The usefulness of these features is limited in environments that are more complicated. There is a probability that the capability to depict hand-crafted features is insufficient to differentiate between small differences between classes and large differences within the same classes. This is because hand-crafted features tend to be more subtle. It is generally accepted that determining more discriminative characteristics is one of the most essential steps in the HSI classification process [5].

Deep learning methods, on the other hand, employ several

[1]*Associate Professor, Department of Information Technology, PVP Siddhartha Institute of Technology, Vijayawada, Andhra Pradesh, India. Email ID: kavitha.d@pvpsiddhartha.ac.in*
[2]*Assistant Professor, R.M.K Engineering College, R.S.M Nagar, Kavaraipettai, Tamil Nadu, India. Email ID: ssh.it@rmkec.ac.in*
[3]*Associate Professor, School of Computer Science and Engineering, Vellore Institute of Technology, Chennai, Tamil Nadu, India. Email ID: rajkumar.srinivasan@vit.ac.in*
[4]*Professor, Department of ECE, Kakatiya Institute of technology and science, Warangal, Telangana, India. Email ID: grrece9@gmail.com*
[5]*Associate Professor, Department of Computer Science, College of Engineering and Technology, Wollega University, Nekemte, Oromia Region, Ethiopia. Email ID: drprb2009@gmail.com*
[6]*Assistant Professor, Koneru Lakshmaiah Education Foundation, Department of CSE, Green Fields, Vaddeswaram, Guntur, Andhra Pradesh, India. Email ID: paps.friend@kluniversity.in*

hierarchical layers to extract useful features from raw data, in contrast to the more laborious and manual approaches. The structure of these levels is arranged in a manner like that of a tree. Layers that appear before them can retrieve simple characteristics such as information about textures and edges. In addition, the deeper layers of the model allow for the representation of characteristics with a greater degree of complexity [5].

Deep learning is an approach to machine learning that emphasizes the elimination of human intervention in the learning process. When it comes to extracting features for use in the HSI classification process, deep learning is widely recognized as a technique that is both effective and efficient.

2. Related Works

The technological improvements that are made to support hyperspectral imaging will lead to an increase in both the degree of precision and the breadth of information that can be extracted from these images. There will be a sizeable improvement made to one of the aspects, specifically the spectrum resolution, while the other aspect, specifically the spatial resolution, will see a much more significant advancement made to it [6].

However, as hyperspectral imagers continue to advance, the amount of data that is stored in hyperspectral images will continue to expand, as will the number of applications that make use of hyperspectral images. The amount of information that is stored in hyperspectral images is already quite considerable [7].

Attempting to analyze and manipulate hyperspectral images has been met with a wide variety of obstacles at every single step ever since the field first got started. On the other hand, the first spectroscopic technique that it developed was hindered by subpar hyperspectral sensors and a dearth of information [8].

Nevertheless, it was successful in developing the technique. Because hyperspectral imaging utilizes a significantly greater number of imaging bands than multispectral imaging does, the images that are produced by hyperspectral imaging have a significantly better spectral resolution than those that are produced by multispectral imaging [9].

The hyperspectral image classification method faces several obstacles, the vast majority of which are problems that have not been resolved yet. Both the high dimensionality of the hyperspectral data and the similarities between the spectra and the mixed pixels contribute to these difficulties [10].

3. Proposed Dimensionality Reduction with classification

It is essential to devise methods of interpretation that are richer in nuances so that they can be used in subsequent processes. Dimensionality reduction algorithms take large, multidimensional datasets and streamline their presentation in a format that is simpler to work with. This is accomplished by selecting and ranking the aspects of the data that are most significant. Dimensionality reduction algorithms are used in many different fields.

A wide range of applications can be developed for large image collections, and each of these applications requires pinpoint accuracy. Because of this, it is of the utmost importance that images be retrieved from large databases in a timely fashion.

Dimensionality reduction that is both accurate and quick is required to achieve the goals of simplifying data, getting an understanding of its local and global patterns, developing theories, and making decisions that are driven as much as possible by data. Regardless of whether the original data structure was low- or high-dimensional, dimensionality reduction can be successfully accomplished with several different approaches, each of which is predicated on a particular presumption pertaining to the starting point of the data. In the sections that follow, we will explain the methods of dimensionality reduction that were used in our research.

3.1. SuperPCA

Several bands with a high resolution of between 5 and 10 nm cover each pixel in an HSI cube photograph. These bands encompass both the visible and infrared spectrums and have a resolution of between 5 and 10 nm. One way to think about the proximity of these noise clumps to one another is as if they were situated side by side.

$$X \in R^{M\times N\times L} \tag{1}$$

where,

M, N and L - rows, columns, and wavelengths.

$X=[x_1, x_2,\ldots,x_P]\in R^{L\times P}(P=MN)$

The research transforms the cube from having three dimensions to having only two of them. One pixel vector is represented in each column of the new matrix that has been created. This vector represents the energy spectrum of the materials that are located within the geographic coverage zone of a single pixel. These materials can be found in the coverage zone of the pixel.

Denote $x_i \in R^L(1\leq i\leq P)$ the i^{th} pixel vector of the observed HSI cube $X\in R^{L\times P}$,

$$x_i = [x_{i1}, x_{i2},\cdots,x_{iL}]^T \tag{2}$$

Utilizing the low-dimensional representation computation that is made available by principal component analysis enables one to optimize data variance in regions in which it has been dimensionally reduced. This is possible because PCA offers a low number of dimensions. It outlines a linear

mapping from the original L-dimensional space $X \in R^{L \times P}$ to a low-d-dimensional space denoted by $Y = [y_1, y_2, \cdots, y_P] \in R^{d \times P}$, $d < L$.

Principal component analysis, also known as PCA, is a type of data preprocessing that is utilized in a considerable number of HSI-based systems. This is since PCA is not only efficient but also resistant to noise. A high-speed interconnect (HSI) is made up of many sections that are extremely similar to one another. It is much more probable for pixels that are located inside of a particular area to belong to the same class than it is for pixels that are located outside of that area to belong to the same class.

This transformation takes a linear approach to the data. The computations performed by the algorithm have resulted in this transformation. For the sake of clarity, we will refer to the transformation matrix as W ($y_i = W^T x_i$), because this will prevent any confusion from occurring. It is essential to find a solution for the objective function that is provided by the equation to acquire the linear transformation matrix.

$$W^* = \underset{W^T W = I}{\arg\max} Tr\left(W^T Cov\left(X\right) W\right) \tag{3}$$

where

Cov(X) - covariance matrix, and

$Tr(X)$ - trace

nxn - square matrix.

3.2. 3D CNN

We introduce a stacked block convolutional neural network (SB-3D-CNN) that considers both spectral and geographic data. This network makes use of 3D convolution, and the research presented here details how that functionality is implemented. The strategy that is being proposed can be broken down into three discrete steps at this point.

Principal component analysis (PCA) and neighborhood extraction are utilized in the first stage of the process, which is known as dimensional reduction. These methods are utilized to cut down on the total number of spectral and geographic dimensions and get rid of any superfluous data. The integrity of the most important spectral parameters can be preserved by utilizing this technique, which makes this possibility feasible.

The subsequent step involves carrying out the process of screening and separating out individual features. At this point in the process, we are utilizing a five-layer framework that is constructed with a concentration layer as one of its components. The overall number of layers in each unit is three-dimensional due to the presence of one- and two-stride convolutional layers in each of its individual units.

To accomplish downsampling in the pooling layer, you must first achieve downsampling in the convolutional layers by setting the cadence to 2 at regular intervals. This will make it possible to get more precise outcomes. Because the design that is being suggested does not make use of a sharing layer, the data that is contained in the HSI tensor will not be protected against access by unauthorized parties.

Incorporating padding into your design is one way to mitigate the impact of sharp edges while also ensuring that your output is of the same scale as your input. Padding can also be used to ensure that your output is of the same scale as your input. You can accomplish this by making sure that the scale of your output is the same as the scale of your intake. After each convolutional layer, a batch normalization procedure is executed, and this could help speed up the convergence process even further.

A layer that has complete connections between all its nodes is utilized to classify the spectral-spatial characteristics. When it comes to the classification stage of the procedure, we make use of a technique called softmax activation. When there are n distinct types of data, the completely connected layer will produce an output consisting of conditional probabilities because of its processing. The following formula can be used to calculate the probability that a given input number, n, corresponds to a specific category.

$$P\left(Y = m \middle| X', k, r\right) = f\left(kX' + r\right) = \frac{e^{k_1 X' + r_1}}{\sum_j e^{k_j X' + r_j}} \tag{5}$$

where

k - weight value and

r - bias term.

The 3D-architecture is illustrated in Fig 1.

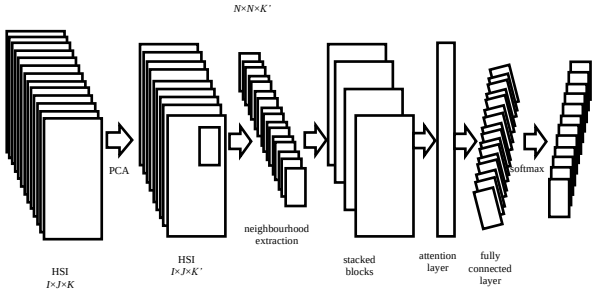

Fig 1: 3D-CNN

3.3. Data Sets

The training set is made up of three hundred instances that were picked at random from each of the three different

database class objects. These instances make up the training set. However, when the databases from Indian Pines and Kennedy Space Center were combined, there were fewer than two hundred unique records in the combined database. For instance, the presence of class items 1, 7, and 16 in the Indian Pines dataset caused the distribution of the population to be skewed in an abnormal orientation. This was since these items were included in the dataset. In the training collection, approximately 80% of the examples were taken from the same pool of fewer than two hundred unique items each. After finishing all the stages that came before it, the practice then made use of the information that was still available. In addition, to verify the accuracy of the model, a random subset of seventy-five% of the data that was employed in the training process was chosen.

3.3.1. Indian Pines

At Indian Pines, which is found in Northwestern Indiana, an investigation into the application of remote sensing was carried out. The AVIRIS (Airborne Visible Infrared Imaging Spectrometer) instrument was mounted on an aircraft, which allowed for the accumulation of data to be carried out with its assistance. The file has a horizontal resolution of 20 meters per pixel and a vertical resolution of 145 pixels. The range of wavelengths that are included in the spectrum covering starts at 0.4 millimeters and goes all the way down to 2.5 microns which is shown in figure 2.

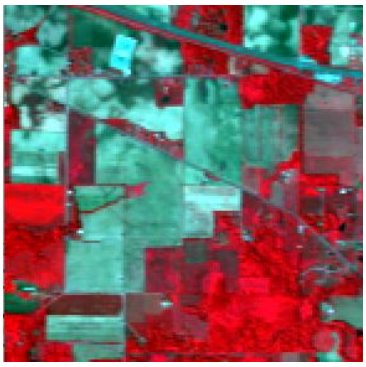

Fig 2. Indian Pines dataset

3.3.2. Pavia University

Pavia University, which is situated in the region of northern Italy, was the source of the information that was used for the data collection.. Since it has a resolution of 1.3 meters per pixel and measurements that are 384 pixels wide by 616 pixels high, this dataset is one of those that offers the highest degree of granularity out of all those that are available. The process known as holography was used to produce it, and it features 115 distinct colors that cover a wide range of frequencies, from 0.43 microns to 0.86 microns. Holography was also used to create it. Before beginning the experiment, we took the initial set of 103 bands and subtracted 12 noise bands from it. This allowed us to conclude the experiment with just 103 bands instead of the original set of 103 bands. In the land-cover dataset that is administered by Pavia University, there are more than 42,000 individual instances that have been included which is shown in figure 3.

Fig 3. Pavia University dataset

3.3.3. Kennedy Space Center (KSC)

The AVIRIS system aboard the Kennedy Space Center in Florida is able to collect data about the center even as the airplane is flying above it. The pixel size of the photograph that was obtained through remote sensing is comparable to 18 meters in the real world. This photograph is included in one of the data sets. After eliminating the bands from the experiment that had a high water absorption rate and a low signal-to-noise ratio, the experiment continued with the use of the 176 bands that were left over after the previous bands were eliminated. There are a total of 5213 examples included in the dataset that was compiled by the KSC, in addition to 13 distinct classifications of land vegetation.

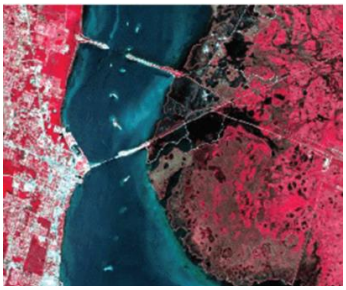

Fig 4. Kennedy Space Center dataset

4. Results

Recently, many novel classification models have been developed, and we make use of a few of those to evaluate how well the recommended classification model performs in comparison to other models. When they were used in conjunction with one another for this investigation, it was possible to draw reliable similarities between the results of the various aspects of the investigation. To ensure the reliability and consistency of the results, each of the three datasets contains experiments that were performed ten times to guarantee accurate results.

4.1.1. Experimental Results

Each experiment was carried out a total of ten times to ensure that the outcomes were reliable and accurate. The findings of the investigation provided, about the variable that was under consideration, both an average and a standard variation. Every experiment produces its own one-of-a-kind collection of ten random samples that are then entered into the random number generator. This is done to eliminate any possibility that the sequence in which the training and testing samples were collected could have an impact on the findings.

The results of the classification that was performed on the Indian Pines dataset, the KSC dataset, and the Pavia University dataset with the assistance of the dimensionality reduction method known as SuperPCA are presented in Tables 1 - 3, in that order. Each table contains its own collection of findings. When compared to more traditional algorithms such as ICA, PCA, and Haar wavelet, the superPCA algorithm can reduce the dimensionality of the HSI while simultaneously increasing its precision in all aspects. This is possible because of the algorithm ability to perform dimensionality reduction. Because the accuracy of categorization varies depending on the number of training samples, we can deduce that the accuracy of the classifier improves as the number of training samples increases.

Table 1. Classification results of Indian Pines University

Class	CNN (%)	AlexNet (%)	Proposed CNN (%)
Alfalfa	82.08	84.73	71.5
Corn-no till	81.67	84.41	85.33
Corn-min till	82.26	83.61	84.88
Corn	75.38	80.03	84.81
Grass/pasture	81.76	84.37	80.98
Grass/trees	78.96	83.53	82.06
Grass/pasture-mowed	82.55	84.75	83.74
Hay-windrowed	78.14	80.88	85.18
Soybeans-no till	81.16	82.92	80.81
Soybeans-min till	82.71	84.59	83.44
Soybeans-clean till	80.36	83.42	85.35
Wheat	79.32	81.84	81.98
Woods	80.68	83.64	82.35
Buildings-grass-trees	77.18	80.72	84.11
Stone-steel towers	81.42	83.11	82.91
Oats	82.71	84.45	83.33

Table 2: Classification results of KSC University

Class	CNN (%)	AlexNet (%)	Proposed CNN (%)
Scrub	86.32	83.78	86.59
Willow swamp	86.64	79.27	85.6
Cabbage palm	79.49	79.99	84.58
Cabbage oak	75.85	85.14	81.92
Slash pine	73.52	79.21	74.25
Oak hammock	81.97	83.56	85.58
Hardwood swamp	65.56	64.5	80.55
Graminoid marsh	83.52	80.82	89.99
Spartina marsh	86.02	83.62	86.31
Cattail marsh	86.5	83.61	84.73
Salt marsh	84.4	86.07	87.12
Mud flats	84.98	83.21	89.52
Water	88.31	84.79	86.75

Table 3: Classification results of Pavia University

Class	CNN (%)	AlexNet (%)	Proposed CNN (%)
Asphalt	83.59	83.29	85.59
Meadows	83.92	83.51	85.81
Gravel	82.9	82.51	84.93
Trees	83.74	82.7	85.27
Metal Sheets	82.54	81.18	85.35
Bare Soil	83.69	83.1	85.83
Bitumen	81.56	82.61	84.21
Bricks	83.16	83.11	85.08
Shadows	80.94	82.46	85.21

According to the findings, the general accuracy of the model increases as the number of training samples increases, but it decreases at a rate that is variable depending on the batch size and the learning rate. This might be interpreted as a sign that the precision as a whole is getting better. The findings also indicate that the SSRN is more reliable than the various other methodologies that were taken into consideration.

3.5 Summary

In this investigation, we apply SuperPCA dimensionality

reduction methods to three distinct datasets so that we can determine which of these four models is the most effective when it comes to reducing the number of dimensions. In addition, we evaluate the performance of 3DCNN on each of these datasets to determine how well it performs on each of these datasets and to see how well it performs overall. According to the findings of the study, superPCA performed noticeably better in terms of categorization precision than the other two models when applied to either the Alexnet or CNN model. In addition, we evaluate the performance of 3DCNN on each of these datasets to determine how well it performs on each of these datasets and to see how well it performs overall. According to the findings of the study, superPCA performed noticeably better in terms of categorization precision than the other two models when applied to either the Alexnet or CNN model. The super PCA model is obviously superior to the other reduction models because it can generate the greatest decrease in dimensionality in the HSI images. This is the main reason why it is the model that is currently being used. This is the primary factor that contributes to the super PCA model undeniable advantage. The findings demonstrate that the models have a higher rate of classification accuracy on the smaller datasets when compared to the other techniques. This is demonstrated by the comparison. It would appear from the observations that employing SuperPCA results in an overall improvement in the classifier level of effectiveness. This method of dimensionality reduction might benefit from making use of the linear transformation matrix that is provided by SuperPCA. This is one possible way in which the method could be improved. SuperPCA has kindly supplied us with this matrix.

References

[1] Ding, Y., Zhang, Z., Zhao, X., Hong, D., Cai, W., Yu, C., ... & Cai, W. (2022). Multi-feature fusion: graph neural network and CNN combining for hyperspectral image classification. *Neurocomputing*, *501*, 246-257.

[2] Yu, C., Han, R., Song, M., Liu, C., & Chang, C. I. (2020). A simplified 2D-3D CNN architecture for hyperspectral image classification based on spatial–spectral fusion. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, *13*, 2485-2501.

[3] Cao, X., Yao, J., Xu, Z., & Meng, D. (2020). Hyperspectral image classification with convolutional neural network and active learning. *IEEE Transactions on Geoscience and Remote Sensing*, *58*(7), 4604-4616.

[4] Dong, Y., Liu, Q., Du, B., & Zhang, L. (2022). Weighted feature fusion of convolutional neural network and graph attention network for hyperspectral image classification. *IEEE Transactions on Image Processing*, *31*, 1559-1572.

[5] Ge, Z., Cao, G., Li, X., & Fu, P. (2020). Hyperspectral image classification method based on 2D–3D CNN and multibranch feature fusion. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, *13*, 5776-5788.

[6] Bhatti, U. A., Yu, Z., Chanussot, J., Zeeshan, Z., Yuan, L., Luo, W., ... & Mehmood, A. (2021). Local similarity-based spatial–spectral fusion hyperspectral image classification with deep CNN and Gabor filtering. *IEEE Transactions on Geoscience and Remote Sensing*, *60*, 1-15.

[7] Vaddi, R., & Manoharan, P. (2020). Hyperspectral image classification using CNN with spectral and spatial features integration. *Infrared Physics & Technology*, *107*, 103296.

[8] Chang, Y. L., Tan, T. H., Lee, W. H., Chang, L., Chen, Y. N., Fan, K. C., & Alkhaleefah, M. (2022). Consolidated convolutional neural network for hyperspectral image classification. *Remote Sensing*, *14*(7), 1571.

[9] Yu, C., Han, R., Song, M., Liu, C., & Chang, C. I. (2021). Feedback attention-based dense CNN for hyperspectral image classification. *IEEE Transactions on Geoscience and Remote Sensing*, *60*, 1-16.

[10] He, X., Chen, Y., & Lin, Z. (2021). Spatial-spectral transformer for hyperspectral image classification. *Remote Sensing*, *13*(3), 498.

[11] 16) Janani, S., Dilip, R., Talukdar, S.B., Talukdar, V.B., Mishra, K.N., Dhabliya, D. IoT and machine learning in smart city healthcare systems (2023) Handbook of Research on Data-Driven Mathematical Modeling in Smart Cities, pp. 262-279.

[12] Juneja, V., Singh, S., Jain, V., Pandey, K.K., Dhabliya, D., Gupta, A., Pandey, D. Optimization-based data science for an IoT service applicable in smart cities (2023) Handbook of Research on Data-Driven Mathematical Modeling in Smart Cities, pp. 300-321.

[13] Pandey, J.K., Veeraiah, V., Talukdar, S.B., Talukdar, V., Rathod, V.M., Dhabliya, D. Smart city approaches using machine learning and the IoT (2023) Handbook of Research on Data-Driven Mathematical Modeling in Smart Cities, pp. 345-362.