ISSN: 2350-0328

International Journal of Advanced Research in Science, Engineering and Technology

Vol. 8, Issue 4 , April 2021

A Survey on Dimensionality Reduction of Hyperspectral Images Classification using Deep Learning

Khusbboo Singh, Dr. Avinash Sharma

PG Scholar CSE, MIT Bhopal , Associate prof. & Head CSE, MIT Bhopal

ABSTRACT: Hyperspectral image data is a progression of spectral bands collected over visible and infrared regions of the electromagnetic spectrum. These datasets hold relevant useful information and also accommodate noise and redundancy, leading to sparseness. Correlation between the bands is inversely proportional to sparseness and this is referred as Curse of Dimensionality. It is imperative to pre-process hyper spectral data to efficiently extract meaningful information. Dimensionality reduction techniques such as Principal component analysis (PCA), Non-negative matrix factorization (NMF), Independent component analysis (ICA) and Singular value decomposition (SVD) were used in this research. This paper explores the dependency of the standard PCA, NMF, ICA, and SVD algorithms on the selected number of dimensions (L). Unsupervised clustering algorithms K-Means and Fuzzy C-Means (FCM) were utilized to identify the influence of L on the dimensionality reduction techniques through clustering accuracy. As L value increases, each algorithm yields different accuracy and optimum L value is difficult to determine. Comments were made on the above dimensionality reduction techniques while gauging the number of dimensions (L) value. Hyperspectral imaging (HSI) is one of the progressive remote sensing techniques. HSI captures data in large number of continuous spectral bands with the spectral range from visible light to (near) infrared, so it is capable of detecting and identifying the minute differences of objects and their changes in temperature and moisture. But its high dimensional nature makes its analysis complex. Various methods have been developed to reduce the dimension of hyperspectral image by feature extraction. This paper highlights the advantages and drawbacks of number of classical dimension reduction algorithms in machine learning communities for HSI classification.

This Research proposal addresses the issues of dimension reduction algorithms in Deep Learning(DL) for Hyperspectral Imaging (HSI) classification, to reduce the size of training dataset and for feature extraction ICA(Independent Component Analysis) are adopted. The proposed algorithm evaluated uses real HSI data set. It shows that ICA gives the most optimistic presentation it shrinks off the feature occupying a small portion of all pixels distinguished from the noisy bands based on non Gaussian assumption of independent sources. In turn, finding the independent components to address the challenge. A new approach DL based method is adopted, that has greater attention in the research field of HSI. DL based method is evaluated by a sequence prediction architecture that includes a recurrent neural network the LSTM architecture. It includes CNN layers for feature extraction of input datasets that have better accuracy with minimum computational cost.

KEYWORDS : Dimensionality Reduction Techniques, Hyperspectral Images, Unsupervised clustering, Classification, PCA, ICA, NMF, SVD, feature extraction, classification, Deep Learning (DL), Minimum Noise Fraction(MNF). LSTM architecture & CNN.

I. INTRODUCTION

Hyperspectral imaging(HSI) has its significance and versatility through its wide array of applications in recent years. HSI mainly involved remote sensing applications, with the emer- gence of ground-based HSI, applications for electronic imaging for food inspection, image-guided surgery, estimation of annual crop-yield and vegetation, cancer detection, forestry fire control, and military uses are made possible[1]. Hyperspectral imaging is a process of dividing the elec- tromagnetic spectrum into hundreds to thousands of contiguous spectral bands resulting in rich data and higher dimensionality.

ISSN: 2350-0328

International Journal of Advanced Research in Science, Engineering and Technology

Vol. 8, Issue 4 , April 2021

Due to large number of spectral bands in HSI the raw data produced is not suitable for use. The abundance of data becomes problematic as data redundancy is prevalent, along with the increase in sparsity and noise, rendering it nearly impossible to extract any meaningful infor- mation. In order to utilize the hyperspectral data, we must first pre-process the hyperspectral data by reducing the number of dimensions to a reasonable size. The data becomes appropriate for evaluation as meaningful information may be derived. While reducing the data from a high dimension to lower dimensions, objects and features must be preserved to maintain integrity ofthe data.

HSI is a foremost research area in Remote Sensing. Hyperspectral remote sensing, which provides very high spectral resolution image data with hundreds of contiguous and narrow spectral bands, has been widely used for discriminating the various land–cover types in HSIs. [1]. It is also referred as imaging spectroscopy, which partition the spectral regions into abundant bands and generates visible images from them. Hyperspectral data usually consists of the spectral bands depicting the ultraviolet (200-400 nm), visible (400-700 nm), near infrared (700-1000 nm), and short-wave infrared (1000- 4000 nm).HSI are generally preferred over typical images for applications like environmental monitoring, forestry and crop analysis, thin films, remote sensing, security and defense, medical diagnose, mineral exploration, food analysis and surveillance. [2].

Recently, there is a noticeable increase in the number of Hyperspectral sensors on-board various satellites/airborne platforms. The data can be collected from different hyperspectral sensors like Airborne Visible Infrared Imaging Spectrometer (AVIRIS), Hyperspectral Digital Imagery Collection Experiment (HYDICE), Hyperspectral Imager (HySI), HYMAP, Compact Airborne Spectrographic Imager (CASI), Digital Airborne Imaging Spectrometer (DAIS), Reflective Optics System Imaging Spectrometer (ROSIS), Airborne Imaging Spectrometer for Applications (AISA) and Hyperion.[3]

Human eye is able to distinguish between objects based on the respective spectral responses those responses are limited part of electromagnetic spectrum[1]. However, multispectral imaging sensors acquire in Fig.1,the electromagnetic spectrum bands. Its observed that the multispectral images uses only three bands among spectral range with huge band interludes, extends around visible region upto infrared regions of the spectrum.

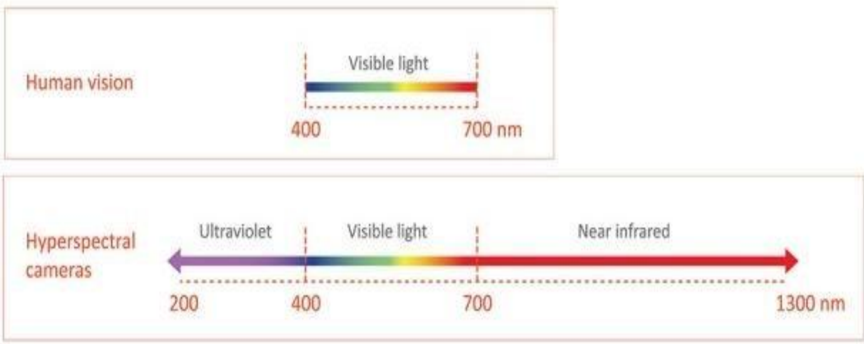
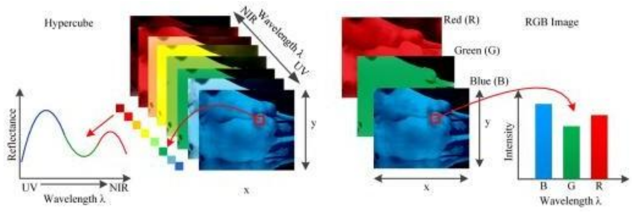


Fig.1: EM Spectral Bands

Images in simplest form whose bands are not necessarily contagious as seen in Fig.1 the electromagnetic spectrum bands. Its observed that the multispectral images uses only three bands among spectral range with huge band interludes, extends around visible region upto infrared regions of the spectrum. However due to this less number of spectral bands discrimination between various materials is the limiting factor [2]. With the involvement of hyper spectral sensing it is made possible to acquire several hundred bands with extremely intense frequency spectral resolution that can execute to discriminate subtle differences and ground covers[4] a clear illustration can be viewed in Fig.2.

ISSN: 2350-0328
International Journal of Advanced Research in Science, Engineering and Technology
Vol. 8, Issue 4 , April 2021

Fig.2: Multispectral Vs Hyperspectral Imaging



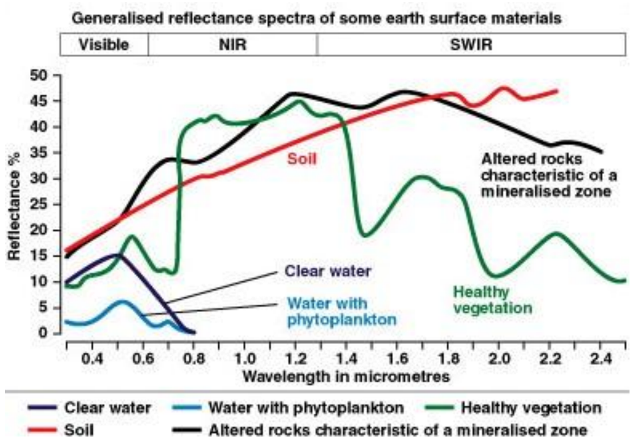


Fig.3: Spectral signatures of earth materials in different spectral regions

As hyperspectral imaging sensors acquires images in a very close spectral bands that results in a high dimensional data feature sets in spectral and spatial resolutions that contains redundant information[5].The increased spectral resolution of HSI allows in subtle discrimination of land services and different material by exploiting a property that each material has their respective spectral signatures, as seen in Fig..3, which was not possible previously due to low spectral resolution of multispectral imaging sensors[3].

II.NEED FOR DIMENSION REDUCTION

Dimensionality reduction is a branch of mathematics that deals with the complexities of huge data sets and enterprise to reduce the dimensionality of the data while capturing its important features. As the complexity of sensors have increased the ability to store enormous amounts of data, reduction algorithms are becoming more essential today. Hyperspectral sensors captures roughly a hundred times the amount of information compared to typical optical sensor. The number of bands as well as the correlation between the bands in HSI data is plentiful and strong. Copious of the data brings difficulties to data storage as well as for data processing. This somewhat limits the applying of the HSI data in some degrees. Also traditional methods which have been designed for multi-spectral image data cannot be easily applied to HSI. So, dimensional reduction in HSI without losing significant information about objects of interest is much essential.

The curse of dimensionality coined by Bellman in 1961 refers to the concept of the expo- nential increase in data in direct relation to the number of dimensions [2]. As the number of dimensions increase, the space between available data points increase making it become sparse. Curse of dimensionality becomes obvious as important features and objects become sparse. Thus, making it

ISSN: 2350-0328
International Journal of Advanced Research in Science, Engineering and Technology
Vol. 8, Issue 4 , April 2021

difficult to process the data efficiently with machine learning techniques. To solve this issue, dimensionality reduction techniques are necessary as a critical pre-processing step to prepare the data for further processing and analysis.

III. DIMENSIONALITY REDUCTION TECHNIQUES

A. Principal Component Analysis (PCA)

As mentioned above, hyperspectral data sets contain highly correlated spectral bands and often having redundant information. Based on this fact, most researchers across the globe [3, 4] have been using Principal Component Analysis (PCA) to extract the band dependency or correlation through statistical properties. [5] Practical implementation of this algorithm was based on [6].

$$X_i = [x_1, x_2, x_3, ..., x_N]^T \quad (1)$$

$$M = 1/M \sum_{i=\text{ to } N}^{i} [x_1, x_2, x_3, ..., x_N]^T \quad (2)$$

In Eq 1, X_i represents a vector of pixels for all the dimensions at a specific pixel location i. Similarly, x_i represents each pixel in the respective i^{th} dimension and the total number of dimensions are N . Total number of pixel vectors depend on the size of each band $M = m * n$ where m is the number of rows and n is the number of columns. The PCA depends on the eigenvalue decomposition of the covariance matrix and Eq (4) represents the decomposition.

$$C_x = ADA^T \quad (3)$$

where $D = diag(\lambda_1, \lambda_2, \lambda_3, ..., \lambda_N)$(4)

is the diagonal matrix of eigenvalues λ_i, i = 1, 2, 3, ...N of the covariance matrix C_x. A is the orthonormal matrix with eigenvectors $a_k(k = 1, 2, \quad , N)$ of C_x. Each PCA pixel vector after the transformation is defined by:

$$y_I = A^T x_I(I = 1, 2, 3, ..., M), \quad (5)$$

Each of these transformed pixel vectors contain compressed information of the entire data set and only a first few bands contribute to useful information. However, the number of vectors to be chosen for each data set varies with size and dimensions of the data set.

B. Independent Component Analysis (ICA)

Similar to PCA, there are many other methods which were formulated to avoid negative values in the transformed pixels due to the orthogonality constraint[7]. The primary assump- tion for ICA is to have a statistically independent input data at any given time with at most one source is a Gaussian source. The goal of ICA is to create maximally independent compo- nents for an given data[8, 9] and this method is based on the blind source separation[9]. ICA representation of the images is based on [10],

$$x = As, \quad (6)$$

where $x_i = \sum_{j=1}^{N} a_{ij} s_j$ is a pixel in an image, A is a basis function matrix and s is image source.

x_i is represented as a linear combination of each image source s_j and weight coefficient a_{ij}. N is the number of image sources in the data set. All pixels of each image source as linearly transformed with a matrix of filters W , so that the resulting vector:

$$u = Wx, \quad (7)$$

recovers the underlying causes, s, in a different order and re-scaled. It shows partial hindrance to process data with noise and becomes computationally expensive with hyper spectral data.

ISSN: 2350-0328

International Journal of Advanced Research in Science, Engineering and Technology

Vol. 8, Issue 4 , April 2021

C. Non Negative Matrix factorization (NMF)

The canonical methods mentioned above are examples of low rank approximations. Later, Paatero and Tapper [14] contributed a novel framework named Positive Matrix Factorization, which was later supported by Lee and Seung[15] as Non Negative Matrix Factorization. NMF incorporates non-negativity constraint on the data and factorizes the matrices. This yields parts-based representation and interpretability of the factorized matrices. The NMF that was implemented in this paper was based on Lee and Seungs multiplicative updating rules with alternating least squares algorithm.

Given a data matrix $X \epsilon R_{\geq 0}^{M \times N}$, M is number of rows and also equivalent to the dimension of a random vector x_j, $j = 1, 2, 3, \ldots, N$. N is number of columns for matrix X and also equivalent to the number of random vector observations . NMF decomposes the matrix X into non-negative $(M \times L)$ basis matrix U and $(L \times N)$ coefficient matrix V. Where:

$$U = [u_1, u_2, u_3, \ldots\ldots, u_L] \epsilon R_{\geq 0}^{M \times L} \quad (8)$$

$$V = [v_1, v_2, v_3, \ldots\ldots, v_N] \epsilon R_{\geq 0}^{L \times N} \quad (9)$$

The multiplicative updating rules for every iteration of UV *calculation* are:

$$\bar{V}_{ij} = V_{ij} \frac{(U^T X)_{ij}}{(U^T U V)} \quad (10)$$

$$\bar{U}_{ij} = U_{ij} \frac{(X V^T)_{ij}}{(U V V^T)} \quad (11)$$

and the objective function of the method to minimize the error approximation is:

$$F = min_{(U,V)} ||X - UV||_{Frob}, \quad (12)$$

The initial condition for a stable and accurate algorithm is to have

$$L << min(M, N).$$

NMF was also used for hyperspectral image data sets in comparison with other prevailing dimensionality reduction techniques. However, this method also has it down- sides, the initialization of this method creates a non-unique solution and the multiplicative update rules converges to a local minimum of the objective function.

D. Truncated Singular Value Decomposition (SVD)

The original Singular value decomposition (SVD) is a matrix decomposition method used for reducing a matrix into 3 separate matrices[17]. Equation(3.15) is the SVD equation:

$$A = USV^T \quad (13)$$

SVD can be used as a dimensionality reduction method in which it is referred to as truncated SVD. A reduced rank approximation to A can be found by setting the first k largest singular value $(0 < k < n)$ along with the corresponding left singular vector and right singular vector [20]. Then a matrix U of size $m \times k$ with reduced number of dimensions k can be derived. Equation (3.X) shows the truncated, reduced SVD, as follow:

$$A_{m\times n} = U_{m\times k} S_{k\times k} V^T_{k\times n} \tag{14}$$

E. Clustering

Clustering is an unsupervised machine learning method that partitions data points into groups(clusters) with similar features. There are two types of clustering, soft clustering and hard clustering. In hard clustering, any one data point belongs to a single cluster with no overlapping between clusters[15]. Soft clustering differs as each data point can be associated to more than one cluster at a time, hence overlapping between clusters is allowed[16]. Each data point has a degree of membership u_{ij} representing the probabilities of belonging to each cluster, with the summation of the probabilities(u_{ij}) equating to 1. This soft clustering method may also be referred to as fuzzy clustering. For this article, K-means (hard clustering algorithm) and Fuzzy C-Means (soft clustering algorithm) was used for classification of the hyperspectraldataset.

IV. CONCLUSIONS

All the dimensionality reductions were efficient in reducing the size of data. The efficiency decreased with increase in *L* dimensions. Out of all, NMF showed consistency with both clus- tering algorithms with increase in *L* value. With this available data, correct number of L dimensions is still difficult to determine. Missing features from reducing the data can lead to misclassification or misinterpretation. Therefore, properly selecting the correct spectral bands is crucial and it is important for dimensionality reduction techniques to preserve the quality of the data and the structural integrity. The Dimensionality Reduction of Hyperspectral Imaging using Deep Learning is an ICA-stationed feature extraction method that extracts independent components later detects targets for various data sets of hyper spectral imaging sensors.

REFERENCES

- "Band Selection for Dimension Reduction in Hyper Spectral Image Using Integrated Information Gain and Principal Components Analysis Technique "on International Journal of Machine Learning and Computing, Vol. 2, No. 3, June 2012

  Addison-Wesley Publishing Company. pp.148-56.
- Anthony J. Bell, Terrence J. Sejnowski, The independent components of natural scenes are edge filters, Vision Research, Volume 37, Issue 23, 1997, Pages 3327-3338, ISSN 0042-6989.
- Asgeir Bjorgan,Lise Lyngsnes Randeberg ,"Real-Time Noise Removal for Line-Scanning Hyperspectral Devices Using a Minimum Noise Fraction-Based Approach". Sensors 2015, 15, 3362-3378. OPEN ACCESS sensors ISSN 1424-8220 www.mdpi.com/journal/sensors.
- Bezdek, J.C. Partition Recognition with Fuzzy Objective Function Algorithms; Plenum Press: NewYork, NY, USA, 1981.
  - Bioucas-Dias, J. M., Plaza, A., Camps-Valls, G., Scheunders, P., Nasrabadi, N., & Chanussot, J. (2013). Hyperspectral remote sensing data analysis and future challenges. IEEE Geoscience and remote sensing magazine, 1(2), 6-36.
- C. Lee and D. A. Landgrebe, ―Analyzing high-dimensional multispectral data,‖ IEEE Trans. Geosci. Remote Sens., vol. 31, no. 4, pp. 792–800, Jul. 1993
- D. Lee and H. Seung, Learning the Parts of Objects by Non-Negative Matrix Factor- ization, Nature, vol. 401, no. 6755, pp. 788-791, 1999.
- Deepa.P and K.Thilagavathi, „Data reduction Techniques of hyperspectral images: A comparative study", in IEEE Explore. DOI: 10. 1109 / ICSCN. 2015. 7219866. (Published in International Conference on Signal Processing, Communication and Networking (ICSCN-2015).
  - Gonzalez, R., and R. Woods. 1993. Digital image process- in. Reading, Massachusetts,
- Hyperspectral Remote Sensing Scenes. (n.d.). Retrieved from http://www.ehu.eus/ ccwintco/index.php/Hyperspectral Remote Sensing Scenes
- Hyperspectral Remote Sensing-SPIE APRS Symposium 2016. Presymposium Tutorial.
- Hyvrinen, A., & Oja, E. (2000). Independent component analysis: algorithms and ap- plications. Neural networks, 13(4-5), 411-430.
- Jolliffe.I "Principal Component Analysis"(2ed., Springer,2002) (518s)_MVSa.pdf
- K.Thilagavathi, R.Nagendran,"Remote Sensing Methods for Forestry Applications – A Survey", International Journal for Science and Advance Research in Technology, Volume 3, Issue 5 in May 2017. ISSN [Online]: 2395-1052.
- Kitti Koonsanit, Chuleerat Jaruskulchai, and Apisit Eiumnoh,
- Klema, V., & Laub, A. (1980). The singular value decomposition: Its computation and some applications. IEEE Transactions on Automatic Control, 25(2), 164-176. doi:10.1109/tac.1980.1102314
- L. Zhang, Y. Zhong, B. Huang, J. Gong, and P. Li, "Dimensionality reduction based on clonal selection for hyperspectral imagery," IEEE Trans. Geosci. Remote Sens., vol. 45, no. 12, pp. 4172–4186, Dec. 2007.

ISSN: 2350-0328

International Journal of Advanced Research in Science, Engineering and Technology

Vol. 8, Issue 4 , April 2021

- Lee, S. Y. (2005). Blind source separation and independent component analysis: A review. Neural Information Processing-Letters and Reviews, 6(1), 1-57.
- Lori Mann Bruce, Cliff H. Koger, and Jiang Li. "Dimensionality Reduction of Hyperspectral Data Using Discrete Wavelet Transform Feature Extraction". IEEE Transactions on Geoscience and Remote Sensing, Vol. 40, NO. 10, October 2002.
- M. J. Khan, H. S. Khan, A. Yousaf, K. Khurshid and A. Abbas, "Modern Trends in Hyperspectral Image Analysis: A Review," in IEEE Access, vol. 6, pp. 14118-14129, 2018. doi:10.1109/ACCESS.2018.2812999
- Maya R. Gupta and Nathaniel P. Jacobson, "Wavelet Principal Component Analysis and Its Application to Hyperspectral Images", IEEE, 2006.
- P. Bajorski. "Statistical inference in pca for hyperspectral images". IEEE J. Sel. Topics Signal Process., 5(3):438–445, June 2011.
- P. Paatero and U. Tapper, Positive Matrix Factorization: A nonnegative Factor Model with Optimal Utilization of Error Estimates of Data Values, Environmetrics, vol. 5, no.2, pp.111-126, 1994.
- Peg Shippert," Introduction to Hyperspectral Image Analysis, Research Systems, Inc.

  preprint arXiv:1404.1100.
- Q. Shi, L. Zhang, and B. Du, ―Semisupervised discriminative locally enhanced alignment for hyperspectral image classification,‖ IEEE Trans. Geosci. Remote Sens., vol. 51, no. 9, pp. 4800–4815, Sep. 2013.
- Randall B Smith. (2012), "Introduction to Hyperspectral Imaging",International Journal of Microimages.
- Richards, J. A., & Richards, J. A. (1999). Remote sensing digital image analysis (Vol. 3). Berlin et al.: Springer.
- Rodarmel, C., & Shan, J. (2002). Principal component analysis for hyperspectral image classification. Surveying and Land Information Science, 62(2), 115-122.
- S. Chen and D. Zhang, ―Semisupervised dimensionality reduction with pairwise constraints for hyperspectral image classification,‖ IEEE Geosci. Remote Sens. Lett., vol. 8, no. 2, pp. 369–373, Mar. 2011.
- Shlens, J. (2014). A tutorial on principal component analysis. arXiv
  - Varshney, P. K., & Arora, M. K. (2004). Advanced image processing techniques for remotely sensed hyperspectral data. Springer Science & Business Media.
- Verleysen, M., & Franois, D. (2005, June). The curse of dimensionality in data mining and time series prediction. In International Work-Conference on Artificial Neural Networks (pp.758-770). Springer, Berlin, Heidelberg.
- Villa, J. A. Benediktsson, J. Chanussot, and C. Jutten."Hyperspectral image classification with independent component discriminant analysis". IEEE Trans. Geosci. RemoteSens., 49(12):4865–4876, Dec. 2011.
- W. Li, S. Prasad, J. E. Fowled, and L. Mann Bruce, "Locality-preserving dimensionality reduction and classification for hyperspectral image analysis," IEEE Trans. Geosci. Remote Sens., vol. 50, no. 4, pp.1185– 1198, Apr.2012.

W. Li, S. Prasad, J. E. Fowler, and L. M. Bruce, ―Locality-preserving dimensionality reduction and classification for hyperspectral