A novel approach for Hyper Spectral Images using Transfer Learning

1Rohit Bharti, 2Dipen Saini, 3Rahul Malik
1,2,3SCSE, Lovely Professional University, Phagwara, Punjab
1rohit.bharti8211@gmail.com, 2er.dipensaini@gmail.com, 3rahul.23360@lpu.co.in

Abstract: The spectral analysis and spatial analysis of high dimensional images are very important and in this paper we tried to cover some aspects that how this problem can be handled and proposed a way through which we can overcome the problem of the time constraint and using some deep learning novel approach like transfer learning for getting the best results while performing the actual computations and the results which we obtained. The dataset used is EuroSAT in which by using a VGG network, the accuracy is achieved 95 per cent and the validation accuracy achieved is 92 per cent. Also, the Kappa score which we got for this observation is 0.95. The tool used for the implementation purpose is TensorFlow with GPU which is also discussed in the paper.

Keywords: spectral, spatial, deep learning, transfer learning, EuroSAT, Landscape

1. Introduction:

Nowadays with the advanced technologies, there has been a tremendous advancement in the hyperspectral images which consists of several image pixels and have hundreds of spectral channel, range from ultraviolet and goes up to infrared having several channels captured from remote sensors. In our image, there are several small spatial structures which can be sensed with the high-resolution spatial sensors. The structure of the pixels can be analyzed properly with high definition spectral resolution sensors.[1].

When we analyze the images of hyperspectral then we want our images should be classified properly. The applications of hyperspectral images are the development of urbanization, development in forest areas, mapping of the land-cover and also managing the risk.

While working with hyperspectral images the same scene is repeated while working with them and when we work with hyperspectral image ten bands or close to this number are available to us. When we try to increase the dimensionality of the images which we have acquired then various problems of theoretical and practical start to arise and then we have to work with the limited training set and due to the increase in features the problem of accuracy arises [2]. And these problems are considered as dimensionality curse.

The hyperspectral images are considered as a list by conventional spectral classifiers [3]. As an example, lately, SVM classifiers received great attention as their capability of handling the high dimensional data [4] with very less training set data. Their efficiency is seen when we see their results of the accuracy in applications which are a lot [5], [6]. If we consider images classification on spatial dependencies, then SVM is opposite for it due to which distinct pixels become an issue and for remote sensing images pixel neighbourhood are not able to provide the important information for which are considered the most and it becomes a disadvantage for SVM [7]. Seeing these problems, both spatial, as well as spatial techniques for classification, received a lot of attention and a way for combining these two is going on.

The classification of salt and pepper using spatial information and then spatial domain to extract the relevant information is important for a given pixel. That is the reason their combination works better. We can use morphological filters [1] and also the usage of morphological levelling [8], image segmentation [9], and the Markov random fields [10].
Markov random fields come from the family of the probabilistic model and they are a 2-d stochastic process of the discrete pixel lattices of MRFs [11]. For combining spatial information with contextual information for classification framework is considered a powerful tool [12]. Lately, as this study enhances [13], a new model called hidden MRF is introduced which is considered as a special case for the previous HMM and whose base is MRF, and not Markov Chains in HMM and therefore the stochastic underlying process for HMM is MRF. They are not restricted to one-dimensional images and even they can go beyond it and extract from Two dimensional and three-dimensional images and thus called the special case of the Hidden Markov Model.

When we combine the imaging technology with the spectroscopy technology, then the hyperspectral technology working on remote sensing can easily differentiate the continuous data and we can get a valuable tool for the Earth surface monitoring [14]. The hyperspectral image when exploited successfully could bring high accuracies in the classification and then it can also provide a better and detailed taxonomy of different classes also [15].

The conventional classification methods for HSI used the method related to the information which is based on spectral for the classification algorithms k-nearest neighbour, logistic regression [16]. HSI deals with high dimensional data having few labelled data, so a curse of the dimensionality is the problem of HSI [17]. Methods for tackling this problem like PCA, wavelet analysis and ICA are used [18].

The rest of the paper is divided into three sections as in section 2, the methodology is discussed, in section 3, the Result analysis is discussed, in section 4, the conclusion and future work is discussed.

2. Literature Review:

M. Fauvel et al. [1] addressed various aspects of hyperspectral images including both spatial as well as spectral aspects of classification. Various approaches were studies and finally it was seen that for primary classification steps, information of spatial and spectral classification is needed. For more adaption and features, the spatial features extraction is done at the object level. The algorithms like EMP–KPCA and the MSSC–MSF performed far better in accuracy. SVM is considered all over as due to its dimensionality reduction capabilities and it is also seen as far more flexible as well.

G. F. Hughes et al. [2] proposed a study in which the pattern of the classifiers especially Bayes classifier is calculated. Various parameters were taken. A probability model of two based on two classes is used and also in them no Gaussian or statistical approach is used. Graphs based on accuracies including optimal as well as maximum accuracies were plotted. Metric functions for calculating metric loss function and also squared error loss functions were considered.

P. F. Hsieh et al. [3] published a thorough study about the classification of high dimensionality of data when the training samples are limited to us. The motivation behind this arisen due to the problem of pattern recognition when we do not have enough samples during the training phase, like this many times these types of problems arise. These type of problems also become critical when we have to classify a sample which is a remote sensing data. They proposed a binary tree model which is able to extract the information of features and also act as a classifier with the limited training samples.

V. N. Vapnik [4] published about various statistical approaches and provided a way how we can use different classifiers and also chose SVM if the dimensionality is proving a different approach. Consistencies and Entropy of real values is noted and different approaches are applied on them a approach is shown how we can implement different classifiers by visualizing the results statistically.

L. Gómez-Chova et al. [5] provide a review of how kernel methods are playing an important role in the area of remote sensing images. Potentially high dimensionality data which is acquired through
satellite sensors is very low when provided as a sample for the training samples, so various SVM classifiers like v-SVM, SVDD are used. For regression, support vector regressor, Gaussian vector regression is also used for classification purpose. Kernel methods are used for conversion of any linear feature into a non-linear feature and after all this we can use linear algebra on it and also perform operations on it. Different machine learning methods can be used in near future for increasing the accuracy when we acquire remote data from satellite sensors.

P. Mantero et al. [6] proposed a approach in which classification is not done on a complete training dataset but despite a sample is taken from an unknown dataset and then using Bayesian decision rule. Probability density functions are found out using SVM and Recursive procedures are used. Both synthetic and real datasets are used for doing the experiments. Very high accuracies are achieved using this proposed model which is able to detect samples from known and unknown samples.

G. Moser et al. [7] proposed a framework in which SVM is integrated with Markov random filed models so that a rigorous and novel framework can be created. As far SVM has acquired a great attention due to its capability of handling the high dimensionality data, which in this case are the images obtained through remote sensing satellites. A spatial contextual classification using this integration is proposed here. Powell algorithm accompanying Ho-Kashyap is used so that MRF model of SVM can be optimized.

M. Pesaresi and J. A. Benediktsson [8] proposed a new method which is dependent upon the morphological characteristics. The images which are too thin or enveloped acquired by remote sensing satellites are segmented properly using Multi segmentation. The radiometric contrast is kept low so that the method performs well which is proposed. Ambiguity and some other features like cross border are kept in consideration. It covers even small areas which are significant for doing a proper segmentation.

P. Ghamisi et al. [9] introduced a new method of multilevel thresholding. The motivation behind is that the multispectral images consist of many data channels, so due to this the data becomes of high dimensionality, and it is very difficult to find the exact accuracies of these remote sensing hyperspectral images. So this method is introduced for extracting features more efficiently. The method consists of fractional-order Darwinian particle swarm optimization, it is capable of finding out many solutions which co-exist at the same time by using swarm intelligence. Then this new approach is then combined with SVM and the results are way more better when simply SVM is used.

A. A. Farag et al. [10] proposed a approach for doing a maximum a posteriori in the remote sensing images, the data acquired by them so that accuracies can be increased. With MAP, class conditional probability is also used, and then the posterior probability is calculated. Now along with conditional class probability SVM is used so that high dimensional data can be tackled more accurately. For measuring the powerful performance of the proposed algorithm, synthetic and actual datasets are used and the experimental results show that the proposed method has greater accuracy.

H. Derin and P. A. Kelly [11] deals with different markov processes in one and two dimensions which is called Discrete type random markov process. The previous and new results are compared and then it is seen that there are two fundamental properties of markov as first when it is defined in terms of conditional probability and then second when they are defined in strict sense linear minimum mean squared error in a more wide sense. Models are assumed and then classification is done on these models and also compared to one another.

G. Moser et al. [12] reviewed Markov random fields and also the different frameworks which is used for classification of the remote sensing high definition images. Various lands which are covered i.e., land cover mapping of high definition images are focused here. For monitoring the environment and also making sure that land exploitation is understood these frameworks are studied efficiently and also future reference is given for increasing the classification of these land coverings even more efficiently.
Y. Zhang et al. [13] proposed a novel approach for analyzing the brain images segmentation by using hidden markov models. It is a stochastic process, and also it is difficult to observe state sequence directly. HMRF is new and far better than FM model, and also EM algorithms are used with HMRF so that robust segmentation and accuracy is increased by incorporating bias field Guillemaud and Brady framework for enhancing the capability in three dimensional approach and also making the system automatic.

A. A. Gowen et al. [14] discussed various aspects of hyper spectral images as how they can be classified and how feature engineering can be done on these high definition and high dimensionality images which are remote sensing images acquired through satellite and are air borne. Various images of agricultural land, waste land, land coverings, and many other images are discussed and then further classifiers are implemented on those and then accuracy can be calculated.

S. Rajan et al. [15] proposed a system in which active learning plays an important role. There are semisupervised and fullysupervised models beforehand but this proposed system has shown great results. The motivation behind this approach is that the data we acquire from the remote sensing satellites is not enough i.e., it is somehow unlabelled and we have to create a complete labelled data so that we can make it work under semisupervised algorithms and models, which becomes so hectic and takes a lot of time. So, using this approach we can also work with the unknown data by using limited principle lables and then apply every algorithm which we want and the results acquired are better and the time is also utilized efficiently.

G. M. Foody et al. [16] compared different classifiers for multi class image classification to find out the relative accuracy when compared with different classifiers which are derived from discriminant analysis and some other algorithms like decision tree algorithms. The whole classification when done on these algorithms depends on the correlation with the training samples. And in this case they are positively correlated. When the p value of the training set is less than 0.05 percent, then SVM performed better approaching the accuracy of 93.7 percent, which is grater than the other algorithms with which it is compared.

P. J. Hsieh et al. [17] proposed a study for overcoming the problem of a small sample size which causes Hughes phenomenon. Now for classifying a image it is important that we extract features by using algorithms like Linear Discriminant Analysis and also like non parametric weighted feature extraction. Everytime we have to apply this preprocessing step so that we can combine and reduce some features which are of higher order and then classes separation for between the class and then one more as with class seperation are used. Indian pine dataset is used for the experimental purpose and the accuracy is calculated which is showing the greater accuracy of this proposed model.

Liu Ying et al. [18] proposed a new algorithm for extraction of the features of of a Hyper spectral image. The algorithm is called Selective Principal Component Analysis which is based upon genetic algorithms based upon subgroups and it is specific about the features which are to be chosen. The bands which have to be chosen for the feature extraction are used for the classification purpose here. The dataset of AVIRIS is used in the proposed algorithm and a great accuracy is achieved. When it is compared with other algorithms then this proposed algorithm is able to increase the efficiency and accuracy by 1.77 percent.

In the table 1, the summary of the literature review in which the year, and the findinds of the authors and the datasets which they used can be seen as:
Table 1: summary of the literature review

| Sr.No. | Author                        | Year | Findings                                                                                                                                 |
|-------|-------------------------------|------|------------------------------------------------------------------------------------------------------------------------------------------|
| 1     | G. Cheng et al. [20]          | 2020 | In this review paper how different deep learning approaches can be used for the Remote sensing images and the work already done by others is shown. |
| 2     | K. Li et. al [19]             | 2019 | A large scale DIOR datset was proposed for the optical Remote sensing images benchmark study in their review work.                          |
| 3     | P. Ghamisi et al. [9]         | 2014 | A new method called fractional-order Darwinian particle swarm optimizatton were used with SVM for handling the multi-channel data.          |
| 4     | A. A. Gowen et al. [14]       | 2014 | The emergence and its use in different fields like microbiology and High resolution images, was discussed.                                  |
| 5     | P. J. Hsieh et al. [17]       | 2014 | Using preprocessing for the hyperspectral images by using Linear Discriminant Analysis and nonparametric weighted feature extraction and the classifier used was SVM. |
| 6     | M. Fauvel et al. [1]          | 2013 | SVMs are used for classification and a minimum spanning forest algorithm is used with a multiple classifier.                             |
| 7     | G. Moser and S. B. Serpico [7]| 2013 | With SVM, the Ho–Kashyap and Powell algorithms were combined for increasing the results on the multi-spectral images and accessing the training size more efficiently. |
| 8     | G. Moser et al. [12]          | 2013 | Markov Models were discussed as how they can prove to increase the results on the Very High Resolution images.                          |
| 9     | L. Gómez-Chova et al. [5]     | 2011 | Kernel methods were used for handling the data, kernel SVMs were used and the results were dramatically improved. Kernel SVMs proved to be more efficient in handling the high dimensional data. |
| 10    | S. Rajan et al. [15]          | 2008 | A semisupervised learning method was proposed for making the classifiers learn more from the less data available.                   |
| 11    | Liu Ying et al. [18]          | 2006 | A new algorithm for doing the Feature extraction and selection as selective principal component analysis based on genetic algorithm with subgroups. The dataset used was AVIRIS and promising results were found. |
| 12    | A. A. Farag et al. [10]       | 2005 | By using the class conditional probability with Markov Random Field and then applying SVM, increases the segmentation accuracy upto 10 percent of the remote sensing images. |
3. Methodology:

3.1 Outline of the Methodology:

The method which we proposed consists of the pre-trained neural network VGG16 and the weights which we used were of ImageNet. ImageNet consists of 14 million images consisting of natural images of different classes like Plants, Animals, Geological information etc. The methodology which we proposed is shown in figure 1 as:
3.2 Model Architecture:

The VGG stands for Visual Geometry Group. The VGG16 Network is used for using the trained weights of the ImageNet. The VGG16 consists of 13 layers of the convolution and 3 layers are fully connected layers. The convolution layer has pooling layers, there are five of them. The convolution layers start with 64 channels with 3 x 3 filter size and end with 512 channels with 3 x 3 filters.

While using the network we changed the head of the network, in which we replaced the network’s last layer with 2048 x 2048, 2048 x 1024, 1024 x 10. Where the last 10 units were used as there are ten classes which are needed to be classified. We have used the Tensorflow 2.0 for the implementation purpose with 8 GB RAM and GTX 1050 GEFORCE GTX GPU.

The VGG16 comprised of two models as VGG16 and VGG19, VGG19 comprises of the more layers but still the VGG16 performs better on many datasets and why it is so, one reason is we implemented on it and got better results and the other one is it is designed in such a way that the feature extraction is far better than the other deep learning networks which are most common in these days.

3.3 Dataset Description:

The Dataset which we used is called EuroSAT. The Sentinel-2A satellite has taken the pictures of the different regions of Europe and then classified them in separate classes. All the images are of 64 * 64 consisting of different electromagnetic channels. The more thorough description of the dataset can be found at Helber et al. 2019 [19] where they described the data thoroughly. The distribution of the data can also be seen in figure 2 and in figure 3 the visualization of the data, on which this work is implemented can be seen. As the data comprises of electromagnetic channels so this is considered one of the benchmark dataset on which the hyperspectral classification algorithms can be tested so that in further times we can used these tested algorithms on some more different datasets and more better and accurate results can be achieved.
3.4 Data Augmentation:

The data augmentation plays a very important role when we are using a computer to get to understand it. We humans can easily detect images if it little misplaces in the images, it is not that of a big deal to us, but for a computer, it is. We apply data augmentation also when we do not want to add more data or we do not have enough data. It also prevents underfitting and overfitting of the data during the training phase. The techniques which we apply for doing the augmentation are height shift range, width shift range, brightness, horizontal flip, vertical flip etc. we decide these techniques according to the data.
which we are using. We have applied rotation range of 90, width shift range of 20 per cent, height shift range of 10 per cent, shear range of 10 per cent, the zoom range of 20 per cent, validation split of 20 per cent, and we have also used horizontal and vertical flip.

3.5 Description of the training phase:

We split the dataset into training and testing in the ratio of 75:25, means 75 per cent for the training and 25 per cent for the testing. Out of a total of 21600 images, we kept 5400 images for the testing purpose. We have trained for 40 epochs on 400 images with a batch size of 500. We used an Adam optimizer with a learning rate of 0.0001. The total parameters were 14, 735, 178, from which the trainable parameters were 20, 490 and non-trainable parameters were 14, 714, 688.

4. Results Analysis:

We have summarized our findings in table 2 as the highest accuracy, training and validation, the loss which we had in predicting the classes. The different Augmentation parameters which we used for increasing the performance. The already trained weights prove beneficial in this approach of hyperspectral images as well. It is better to start with the weights which are pre-trained than the weights which we have to start training from the beginning and random. The accuracy and computational timings which we achieved are quite promising. This technique can also be applied to other satellite images and results can be compared as well.

| Dataset   | Augmentation parameters | CNN hyper-parameters |
|-----------|-------------------------|----------------------|
|           | Rotation | Width shift | Height shift | Learning rate | Accuracy | val-Accuracy |
| EuroSAT   | 90       | 0.2         | 0.2          | 0.1           | 0.0001    | 0.95        | 0.92        |

The accuracy and loss plots can be seen in figure 4 and the conclusion can be drawn that the accuracy and loss are converging nicely as can be visualized with graphs as:

![Figure 4: train and loss plots](image-url)
5. Conclusion and Future Scope:

We applied the transfer learning and achieved great results in which the accuracy was 0.95 and the validation accuracy was 0.92. This is a new approach we can be used on other datasets also like UC Merced, RSSCN7, SAT4 etc and we are sure the results will be promising. We can also use different pre-trained neural networks like ResNet, GoogleNet, Xception etc, and achieve great results. The most important step is that we start with trained weights and we do not need to get started with Random weights.

References:

[1] M. Fauvel, Y. Tarabalka, J. A. Benediktsson, J. Chanussot, and J. C. Tilton, “Advances in spectral-spatial classification of hyperspectral images,” Proc. IEEE, vol. 101, no. 3, pp. 652–675, 2013.

[2] G. F. Hughes, “On the Mean Accuracy of Statistical Pattern Recognizers,” IEEE Trans. Inf. Theory, vol. 14, no. 1, pp. 55–63, 1968.

[3] P. F. Hsieh and D. Landgrebe, “Classification of high dimensional data,” ECE Tech. Reports, p. 52, 1998.

[4] V. N. Vapnik, “Statistical Learning Theory,” 1998.

[5] G. Gómez-Chova, J. Muñoz-Marí, V. Laparra, J. Malo-López, and G. Camps-Valls, A Review of Kernel Methods in Remote Sensing Data Analysis. 2011.

[6] P. Mantero, G. Moser, and S. B. Serpico, “Partially supervised classification of remote sensing images using SVM-based probability density estimation,” 2003 IEEE Work. Adv. Tech. Anal. Remote. Sensed Data, vol. 00, no. C, pp. 327–336, 2004.

[7] G. Moser and S. B. Serpico, “Combining Support Vector Machines and Markov.pdf,” vol. 51, no. 5, pp. 2734–2752, 2013.

[8] M. Pesaresi and J. A. Benediktsson, “A new approach for the morphological segmentation of high-resolution satellite imagery,” IEEE Trans. Geosci. Remote Sens., vol. 39, no. 2, pp. 309–320, 2001.

[9] P. Ghamisi, M. S. Couceiro, F. M. L. Martins, and J. A. Benediktsson, “Multilevel image segmentation based on fractional-order Darwinian particle swarm optimization,” IEEE Trans. Geosci. Remote Sens., vol. 52, no. 5, pp. 2382–2394, 2014.

[10] A. A. Farag, R. M. Mohamed, and A. El-Baz, “A unified framework for MAP estimation in remote sensing image segmentation,” IEEE Trans. Geosci. Remote Sens., vol. 43, no. 7, pp. 1617–1634, 2005.

[11] H. Derin and P. A. Kelly, “Discrete-Index Markov-Type Random Processes,” Proc. IEEE, vol. 77, no. 10, pp. 1485–1510, 1989.

[12] G. Moser, S. B. Serpico, and J. A. Benediktsson, “Land-cover mapping by Markov modeling of spatial-contextual information in very-high-resolution remote sensing images,” Proc. IEEE, vol. 101, no. 3, pp. 631–651, 2013.

[13] Y. Zhang, M. Brady, and S. Smith, “Segmentation of brain MR images through a hidden Markov random field model and the expectation-maximization algorithm,” IEEE Trans. Med. Imaging, vol. 20, no. 1, pp. 45–57, 2001.
[14] A. A. Gowen, E. Gaston, and J. Burger, “Hyperspectral imaging,” *Food Eng. Ser.*, pp. 199–216, 2014.

[15] S. Rajan, J. Ghosh, and M. M. Crawford, “An active learning approach to hyperspectral data classification,” *IEEE Trans. Geosci. Remote Sens.*, vol. 46, no. 4, pp. 1231–1242, 2008.

[16] G. M. Foody and A. Mathur, “A relative evaluation of multiclass image classification by support vector machines,” *IEEE Trans. Geosci. Remote Sens.*, vol. 42, no. 6, pp. 1335–1343, 2004.

[17] P. J. Hsieh, C. H. Li, K. C. Chen, and B. C. Kuo, “A kernel-based feature extraction method for hyperspectral image classification,” *Int. Geosci. Remote Sens. Symp.*, pp. 513–516, 2014.

[18] Liu Ying, Gu Yanfeng, and Zhang Ye, “Hyperspectral Feature Extraction using Selective PCA based on Genetic Algorithm with Subgroups,” pp. 652–656, 2006.

[19] P. Helber, B. Bischke, A. Dengel, and D. Borth, “Eurosat: A novel dataset and deep learning benchmark for land use and land cover classification,” *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.*, vol. 12, no. 7, pp. 2217–2226, Jul. 2019.

[20] G. Cheng, X. Xie, J. Han, L. Guo, and G. S. Xia, “Remote Sensing Image Scene Classification Meets Deep Learning: Challenges, Methods, Benchmarks, and Opportunities,” *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.*, vol. 13, pp. 3735–3756, 2020.