Hyperspectral Image Classification with CNN

Ahmet Gülmez

Introduction to Deep Learning - Final Report

Student ID:3687544
1st March, 2021

1 Introduction

Image classification or object recognition has used image datasets with a single color or primary colors like the RGB spectrum. In contrast, hyperspectral imaging (HSI) is a sensing technique that contains images of a large number of closely spaced wavelengths. Every pixel in a hyperspectral image consists of a discrete spectrum which is, in general, more than 100 spectral bands. Therefore it is three-dimensional data that needs to be processed differently from an ordinary image processing or classification. Below, an example of hyperspectral data is shown.

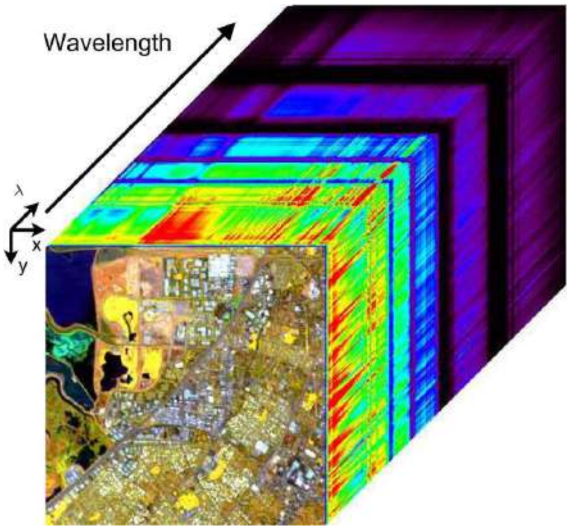


Fig. 1. Example of a hyperspectral cube data [1]

2 Related Works

HSI has arisen from remote sensing; NASA has conducted different researches and became a pioneer in this regard [2]. Recently, deep learning has been used in HSI because of its advantages in image processing for this kind of data. Besides this, HSI is an emerging method in remote sensing. It has been used in various areas such as archaeology, autonomous vehicles, vegetation, and water resource control [3], food quality and safety control [4], forensic medicine [5], biomedicine [6] [7]. This project uses two datasets from named Indian Pines, and Pavia University are explained later. The model used in the project is a 3D CNN model, a state-of-the-art model for video recognition algorithms. The reason for that will be explained later. However, the project is based on the paper [8] which presents a new approach. The project can be seen as an extension and fine-tuning to this approach.

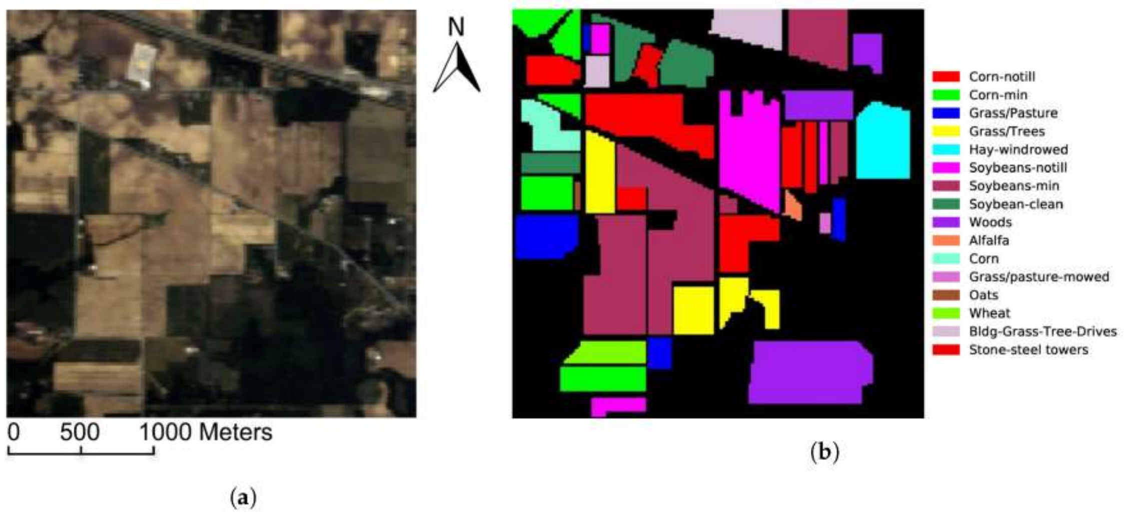


Fig. 2. a) Indian Pines scene b) ground reference data with labels [9]

3 Dataset

HSI datasets mostly consist of a scene with a significant number of spectral bands. Two different datasets will be used for the project: the Indian Pines dataset and the Pavia University dataset. These are the generic datasets that are used to evaluate the models in hyperspectral image classification. The scene in Indian Pines, which can be seen as a satellite scene, was collected by an AVIRIS sensor. The scene that is a test site image from North-western India consists of 145*145 pixels with 224 spectral bands. The dataset has a labeled ground reference data, which is shown in the figure 2.

The Pavia University dataset consists of 610*340 pixels with 103 spectral bands. It has nine classes of interests that show different types of materials, shown in the figure 8. The samples of both datasets are the pixels that have a vector size of the spectral band.

Besides this, the data is split into train, validation, and test sets. This is done with the help of sci-kit learn library. The split was done in a stratified fashion. This means that the proportion of the classes in train, validation, and test set is preserved. This is important because the proportion of the data for each label varies in both datasets. Therefore, the tranining could not be reliable without a stratified split. Besides, the splitting is done randomly.

Eighty percent of data is used for training. Firstly the whole data split into training and test sets with a 0.9 ratio. Then the remaining training set is split again into training and validation set with a 0.875 ratio. The random train and test split of data are shown in the following figures:

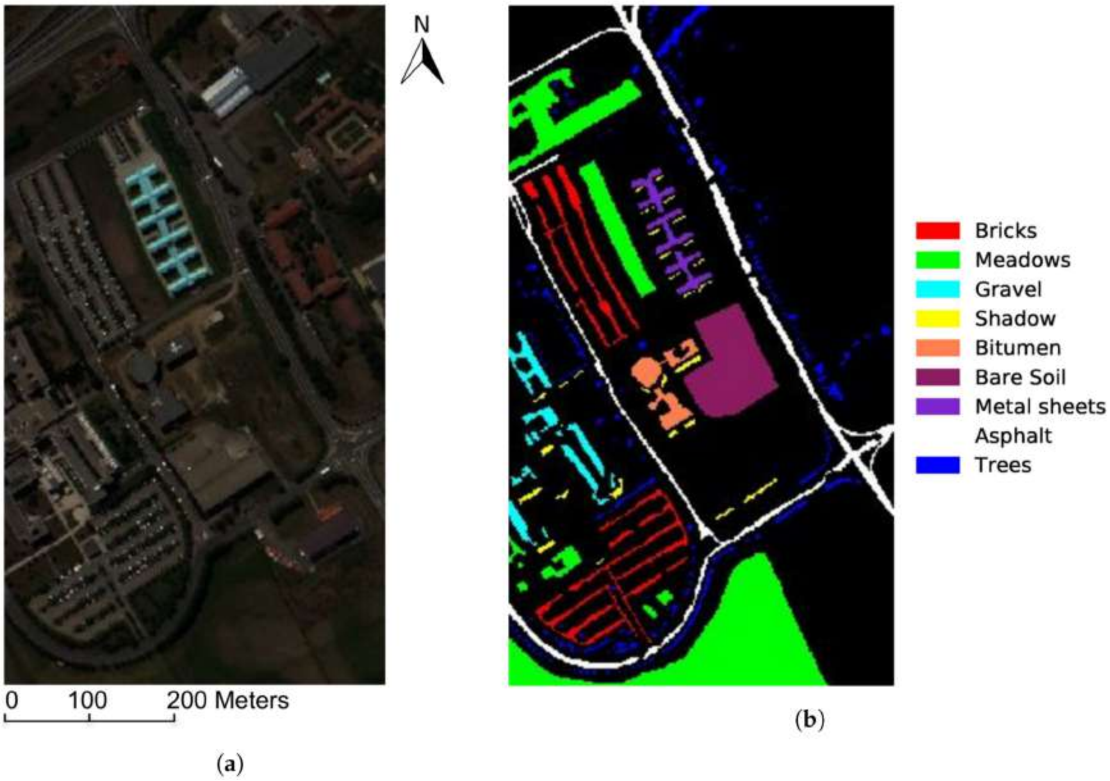
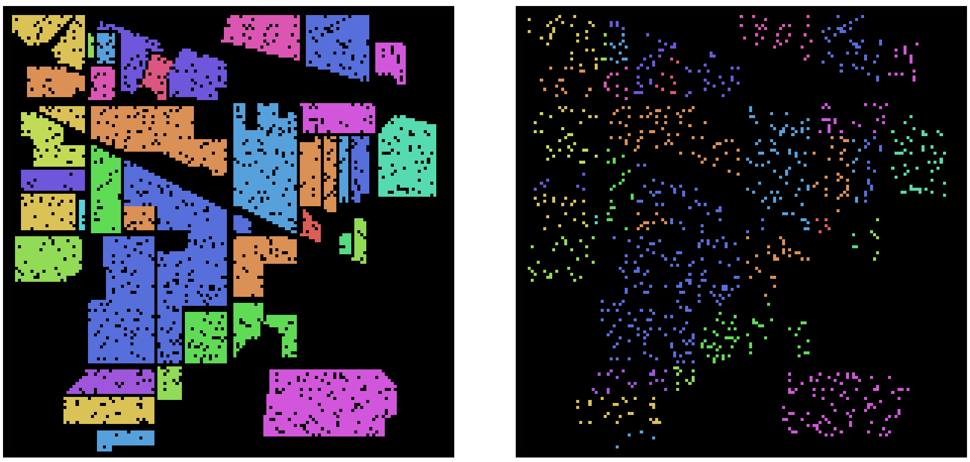


Fig. 3. a) Pavia University scene b) ground reference data with labels [9]



(a) Train ground truth **(b)** Test ground truth

Fig. 4. Training and test set after splitting on ground truth

4 Final Model Structure

Based on the article, it can be said that 3D CNN has certain advantages. Generally, 2D CNN is used for image classification. However, spectral information in hyperspectral data is essentially key to classification. Before that, a band selection was made using the spectral band compression algorithms such as PCA in various articles. And then, 2D CNN was used for the resulting data. However, as seen in this article, simultaneous spatial and spectral properties' evaluation yielded better results. One of the report aims is to get a better result by changing the hyperparameters and model structure. At the same time, to be more prone to hyperspectral data analysis for my future personal research.

In the first phase, the paper's model structure has been implemented, and the results are denoted in this regard. Then the model has changed empirically by the depth of the CNN and the hyperparameters such as kernel size, sample size, etc. The model structure in the paper was implemented with two 3D-CNN's.

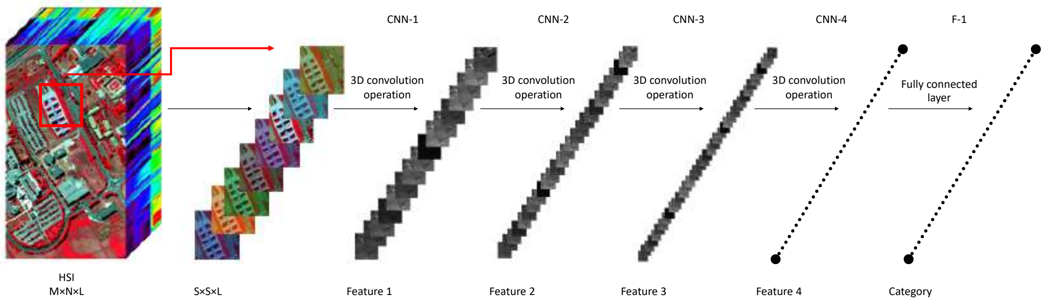


Fig. 5. Model Structure

The project's model structure has four sequential 3D-CNN's. At the end it includes a fully connected layer in order to accomplish the classification. The activation function used is a Rectified Linear Unit (ReLu) which is generally used in this type of classification algorithms in hyperspectral images. To obtain the model structure several experiments are concluded with a hierarchical model searching procedure. This experiments will be explained and shown in the section 5. As shown in the figure the whole dataset is seperated into samples with SxSxL size. These samples are given into network in batches. The sample size that is chosen in this project is a 9x9xL patch.

With the help of the validation set, it was seen that the network was underfitted in this number of iterations. Therefore the strategy was to increase the depth of the network. Instead of using only 2 CNN layers, it was added one or two-layer and observed the results.

As shown in the figure 5, the whole dataset is separated into samples with SxSxL size. These samples are given into the network in batches. A batch size of 64 is used in training. This batch size is not problematic because it is seen that the network does not have a lack of generalization. Therefore it is gained a faster training with larger batches. Empirically it is found out that the optimal sample size is 9x9xL. The reason for that might be that the spatial features are better obtained with a larger sample size instead of 7x7 or 5x5.

By choosing the number of CNN layers, it is again an empirical strategy was pursued. The strategy was to use more layers as possible until it is overfitted early. By comparing 3 and 4 CNN layers training, it is found out that the test accuracies are almost identical, but the convergence speed was different. It is observed that four-layer CNN was converged faster in comparison to a three-layer structure. Besides, it is seen that more than four layers are prone overfitting and also the noise in the dataset. Additionally, it is used an Adam optimizer with a learning rate of 0.0001.

The hyperparameters in the CNN's layers are shown in the following table:

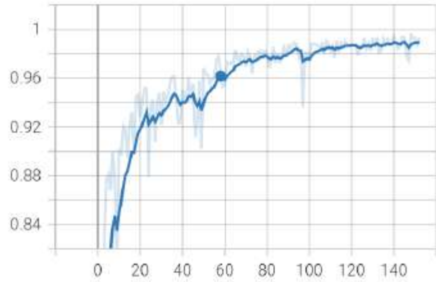
Layer	**Kernel Size**	**Kernel Number**	**Padding**	**Stride**
CNN-1	9x3x3	1	4	1
CNN-2	7x3x3	2	3	1
CNN-3	5x3x3	4	2	1
CNN-4	3x3x3	8	1	1

Model structure with its hyperparameters

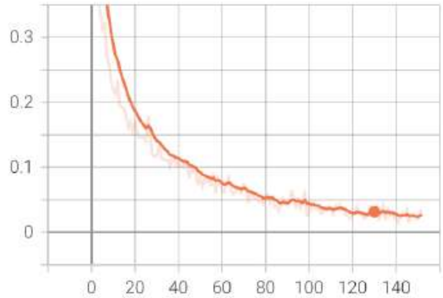
These hyperparameters were the best among others that are experimented with different layers and sets. Some of these experiments are shown in the following by comparing the test accuracies.

5 Evaluation and Discussion

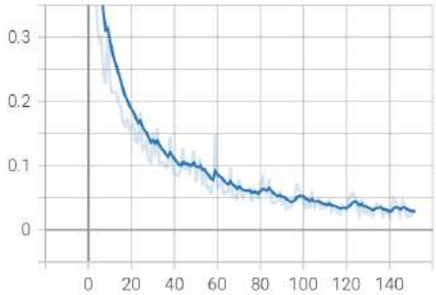
In the previous section, the model was evaluated according to its model structure and hyperparameters. In this section, I will conclude the qualitative and quantitative analysis. First of all, as stated in the proposal, the accuracy of the test was evaluated according to the successful pixel classification. This means it is calculated by the ratio of the number of successful classified pixels and the total number of pixels. Therefore it is also understandable that this problem set is not a segmentation but a classification of each pixel. Before this evaluation, the accuracy in the validation set was calculated to determine the required hyperparameters. Accordingly, the optimal hyperparameters were set. Finally, the final result was obtained with a test set that was completely independent of this. You can see these train and validation precision and losses in the figure below.



(a) Validation Accuracy

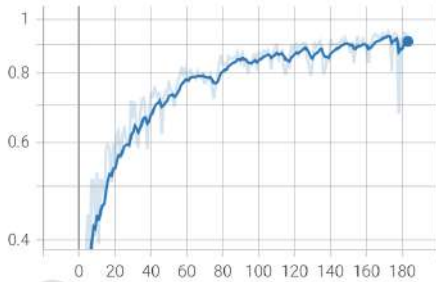


(b) Train Loss

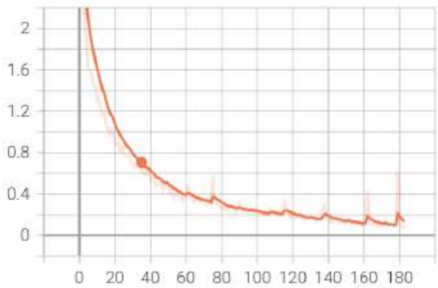


(c) Validation Loss

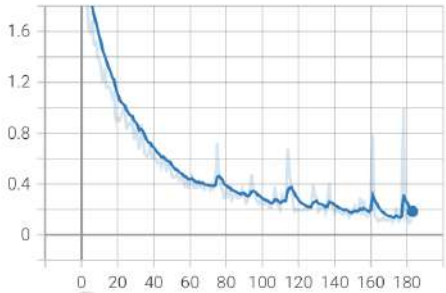
Fig. 6. Train and validation set results from Pavia University dataset



(a) Validation Accuracy



(b) Train Loss



(c) Validation Loss

Fig. 7. Train and validation set results from Indian Pines dataset

As it can be observed training with the Pavia leads to a validation and a test accuracy to 99 percent. Therefore the measurement is cut-off earlier than 200 epoch. The Indian dataset will converge to almost 94 accuracy in 200 epoch. The prediction of the pixels can be seen in the following figure which can be compared to ground truth:

Fig. 8. Prediction from the Indian Pines Dataset

Additionally the experiments that are mentioned in the section 4 will be demonstrated as follows:

	Sample Size	Kernel-1	Kernel-2	Kernel-3	Kernel-4	Accuracy in %
1.	7	(7,3,3)	(3,3,3)	-	-	81,89
2.	7	(3,3,3)	(3,3,3)	(3,3,3)	-	78,15
3.	9	(3,3,3)	(3,3,3)	(3,3,3)	-	89,97
4.	9	(3,3,3)	(3,3,3)	(3,3,3)	(3,3,3)	87,75
5.	9	(9,3,3)	(7,3,3)	(5,3,3)	(3,3,3)	93,95

Accuracy of networks with different kernel sizes in Indian Pines

	Sample Size	Kernel-1	Kernel-2	Kernel-3	Kernel-4	Accuracy in %
1.	7	(7,3,3)	(3,3,3)	-	-	98,66
2.	7	(3,3,3)	(3,3,3)	(3,3,3)	-	97,26
3.	9	(3,3,3)	(3,3,3)	(3,3,3)	-	99,48
4.	9	(3,3,3)	(3,3,3)	(3,3,3)	(3,3,3)	98,46
5.	9	(9,3,3)	(7,3,3)	(5,3,3)	(3,3,3)	99,43

Accuracy of networks with different kernel sizes in Pavia University

In the tables the kernel sizes are demonstrated according to their layers. These results are from some of the most informative experiments among all in order to obtain the optimal hyperparameters. As seen with deeper networks, it is better to use wider kernels in CNN. Therefore the optimal network is obtained empirically with these experiments. The results of 3 and 5 experiments seem to be equally successful. The reason to use a deeper network, as in 5, is because it converges quickly to the optimum accuracy in earlier epochs.

Additionally, the results from the paper [8] in the Indian Pines dataset seem to be pretty higher than my network. One of the possible reasons is the fewer iterations in my training. For both datasets, the number of training iterations is about 10000 - 25000 iterations which is much lower than the paper's network. This is due to the hardware and time constraints. Another reason is that the Indian Pines dataset has a big radiation noise by measurements. A noise reduction might be implemented in the paper's network, which is not implemented in my network. The other reason is that the Indian Pines dataset has a big variation in the each label's number of data. Training the network with a random split might lead to a different rate there. Another reason might be the lack of augmentation. In my experiments, I used only flip augmentation, which did not lead to better performance. The paper might have other data augmentation tricks in order to obtain better accuracy.

6 Conclusion

Hyperspectral imaging processing has various opportunities in the medical and food industry. Deep learning is a promising solution to this type of classification. The approach concluded in this report is a successful approach in both test accuracy and time consumption. The intuition that is gained from this project was very important for my further researches and implementations in hyperspectal imaging. As future work, I will be more dealt with the spectral feature of these datasets, radiational noise and its effects.

7 References

1. Christophe, E., Mailhes, C. & Duhamel, P. Hyperspectral image compression: Adapting SPIHT and EZW to anisotropic 3-D wavelet coding. *IEEE Transactions on Image Processing* **17,** 2334–2346 (2008).
2. Goetz, A. F. Three decades of hyperspectral remote sensing of the Earth: A personal view. *Remote Sensing of Environment* **113,** S5–S16 (2009).
3. Govender, M., Chetty, K. & Bulcock, H. A review of hyperspectral remote sensing and its application in vegetation and water resource studies. *Water SA* **33,** 145–151 (2007).
4. Gowen, A. A. & Frias, J. M. Hyperspectral imaging-an emerging process analytical tool for food quality and safety Trends in Food Science & Technology food quality and safety control biting into a busy monolith . Information technology implementation research : a technological diffusion approach , the gyro integrator , at first glance , is unstable . (2007).
5. Malkof, D. B. & Oliver, W. R. Hyperspectral imaging applied to forensic medicine. **3920,** 108–116 (2000).
6. Carrasco, O., Gomez, R. B., Chainani, A. & Roper, W. E. Hyperspectral imaging applied to medical diagnoses and food safety. *Geo-Spatial and Temporal Image and Data Exploitation III* **5097,** 215 (2003).
7. Lu, G. & Fei, B. Medical hyperspectral imaging: a review. *Journal of Biomedical Optics* **19,** 010901 (2014).
8. Li, Y., Zhang, H. & Shen, Q. Spectral-spatial classification of hyperspectral imagery with 3D convolutional neural network. *Remote Sensing* **9** (2017).
9. Carranza-García, M., García-Gutiérrez, J. & Riquelme, J. C. A Framework for Evaluating Land Use and Land Cover Classification Using Convolutional Neural Networks. *Remote Sensing* **11** (2019).