A REPORT

ON

AUTOMATIC LAND COVER CLASSIFICATION OF MULTI-TEMPORAL SATELLITE IMAGES

BY

Guntaas Singh 2018A7PS0269P

Nisarg Vora 2018A7PS0254P

AT

Indian Institute of Remote Sensing, Dehradun

A Practice School - I station of

Birla Institute of Technology and Science, Pilani

June, 2020

A REPORT
ON

AUTOMATIC LAND COVER CLASSIFICATION OF MULTI-TEMPORAL SATELLITE IMAGES

BY

ID Number	Name	Branch
2018A7PS0269P	Guntaas Singh	B.E. (Hons.) Computer Science
2018A7PS0254P	Nisarg Vora	B.E. (Hons.) Computer Science

Prepared in the partial fulfillment of the

Practice School - I course

AT

Indian Institute of Remote Sensing, Dehradun

A Practice School - I station of

Birla Institute of Technology and Science, Pilani

June, 2020

Acknowledgments

We would sincerely like to thank the Director of Indian Institute of Remote Sensing, Dr. Prakash Chauhan, for giving us an opportunity to work in this organization and gain a significant amount of exposure to corporate work culture, ethics, and etiquette to be followed while working in a professional environment.

We would like to extend our most sincere gratitude to our project in-charge, Dr. Hari Shanker Srivastava, Head of the Programme Planning and Evaluation Group (PPEG) at IIRS, for providing us with the opportunity to work with him on this project, and his guidance and mentorship during the same.

We wish to extend our gratitude to the faculty in charge of the PS-I program at IIRS, Dr. Rekha A., Assistant Professor at BITS Pilani - Bangalore Center, for her guidance and advice during the PS-I program, and her helpfulness and responsiveness while addressing all the concerns we raised during the same.

In addition, we would also like to thank the members of the Practice School Division, who have worked very hard for operating the PS-I programme remotely to ensure that we have a seamless learning experience.

BIRLA INSTITUTE OF SCIENCE AND TECHNOLOGY
PILANI (RAJASTHAN)
Practice School Division

Station: Indian Institute of Remote Sensing
Centre: Dehradun
Duration: From 18th May, 2020 to 27th June, 2020
Date of start: 18th May, 2020
Date of submission: 4th June, 2020
Title of project: Automatic Land Cover Classification of Multi-temporal Satellite Images

ID Number	**Name**	**Branch**
2018A7PS0269P	Guntaas Singh	B.E. (Hons.) Computer Science
2018A7PS0254P	Nisarg Vora	B.E. (Hons.) Computer Science

Name of guide: Dr. Hari Shanker Srivastava
Designation: Scientist/Engineer - SG. Group Head, Programme Planning and Evaluation Group (PPEG).
Name of PS faculty: Dr. Rekha A.

Abstract

Blah

BIRLA INSTITUTE OF SCIENCE AND TECHNOLOGY
PILANI (RAJASTHAN)
Practice School Division
Response Option Sheet

Station: Indian Institute of Remote Sensing
Centre: Dehradun

ID Number	Name	Branch
2018A7PS0269P	Guntaas Singh	B.E. (Hons.) Computer Science
2018A7PS0254P	Nisarg Vora	B.E. (Hons.) Computer Science

Title of project: Automatic Land Cover Classification of Multi-temporal Satellite Images

Code No.	Response Option	Course No.(s) and Name
1	A new course can be designed out of this project.	
2	The project can help modification of the course content of some of the existing Courses	
3	The project can be used directly in some of the existing Compulsory Discipline Courses (CDC)/ Discipline Courses Other than Compulsory (DCOC)/ Emerging Area (EA), etc. Courses	
4	The project can be used in preparatory courses like Analysis and Application Oriented Courses (AAOC)/ Engineering Science (ES)/ Technical Art (TA) and Core Courses.	
5	This project cannot come under any of the above mentioned options as it relates to the professional work of the host organization.	

Signature
Date:

Contents

1. Introduction

1.1 About IIRS

Formerly known as Indian Photo-interpretation Institute (IPI), the Institute was founded on 21st April 1966 under the aegis of Survey of India (SOI). It was established with the collaboration of the Government of The Netherlands on the pattern of Faculty of Geo-Information Science and Earth Observation (ITC) of the University of Twente, The Netherlands. The original idea of setting the Institute came from India's first Prime Minister Pandit Jawahar Lal Nehru during his visit to The Netherlands in 1957. Since its establishment in 1966, IIRS is a key player for training and capacity building in geospatial technology and its applications through training, education and research in Southeast Asia. The training, education and capacity building programmes of the Institute are designed to meet the requirements of Professionals at working levels, fresh graduates, researchers, academia, and decision makers. IIRS is also one of the most sought after Institute for conducting specially designed courses for the officers from Central and State Government Ministries and stakeholder departments for the effective utilization of Earth Observation (EO) data. Keeping pace with the technological advances, the Institute has enhanced its capability with time, to fulfill the increased responsibility and demand from Indian and international community. Today, it has programmes for all levels of users, i.e. mid-career professionals, researchers, academia, fresh graduates and policy makers. The sustained efforts by its dedicated faculty and the management have made the institute remain in the forefront throughout its journey of about four and a half decades from a photo-interpretation institute to an institute of an international stature in the field of remote sensing and geo-information science.[3][4]

1.2 Remote Sensing

Remote sensing is the acquisition of information about an object or phenomenon without making physical contact with the object and thus in contrast to on-site observation, especially the Earth. Remote sensing is used in numerous fields, including geography, land surveying and most Earth science disciplines (for example, hydrology, ecology, meteorology, oceanography, glaciology, geology); it also has military, intelligence, commercial, economic, planning, and humanitarian applications. It may be split into "active" remote sensing (when a signal is emitted by a satellite or aircraft to the object and its reflection detected by the sensor) and "passive" remote sensing (when the reflection of sunlight is detected by the sensor). Passive sensors gather radiation that is emitted or reflected by the object or surrounding areas. Reflected sunlight is the most common source of radiation measured by passive sensors. Examples of passive remote sensors include film photography, infrared, charge-coupled devices, and radiometers. Active collection, on the other hand, emits energy in order to scan objects and areas whereupon a sensor then detects and measures the radiation that is reflected or backscattered from the target. RADAR and LiDAR are examples of active remote sensing where the time delay between emission and return is measured, establishing the location, speed and direction of an object. [12]

2. Methodology

2.1 Image Classification

Image classification is a standard task in computer vision. In general, the image classification problem involves assigning one label out of a given fixed set of discrete labels to the input image on the basis of its visual content. While this is a trivial task for humans, robust image classification is a big challenge for a machine. To the computer, the image is just a grid of numbers which entirely change in unreliable ways with variations in viewpoint, illumination, occlusion, etc. As a result, there is no obvious algorithm which solves this problem. However, a data driven approach of providing the machine with many examples of each class and use of machine learning techniques has shown to be useful.[7]

There are different ways in which these techniques can be applied for classification of satellite imagery.

Pixel Based Approach

In typical satellite images, pixel sizes are generally similar in size to the objects of interest. Most of the methods for image analysis using remote sensing data work on a per-pixel basis. However, with advances in remote sensing technology, the spatial resolution has become finer than the typical objects of interest, leading to an increase in within-class variability.[5]

Object Based Approach

The term "objects" represents meaningful semantic entities or scene components that are distinguishable in an image.[5] This approach involves the partition of the image into meaningful geographical objects that share relatively homogeneous spectral, color, etc.

Semantic Approach

This aims to label each scene image with a specific semantic class. Here, a scene image usually refers to a local image patch manually extracted from large scale remote sensing images that contain explicit semantic classes.[5]

2.2 Deep Learning and Neural Networks

Application of traditional machine learning techniques requires handcrafted features, developing which demands a considerable amount of engineering skill and domain expertise. This, however, is not true for neural networks, which automatically learn these features from data using a general-purpose learning procedure.[5, 7] Despite having been around for decades, neural networks have garnered much attention only in the last few years on account of the availability of increased computational power and large amounts of data.

A standard neural network consists of many simple, connected processors called neurons, each producing a sequence of real-valued activations. Input neurons get activated through sensors perceiving the environment, other neurons get activated through weighted connections from previously active neurons. [13] Each neuron can be seen as a single unit applying a non-linear activation function (such as sigmoid, tanh, ReLU) to a linear combination of the input activations to the neuron.[10]

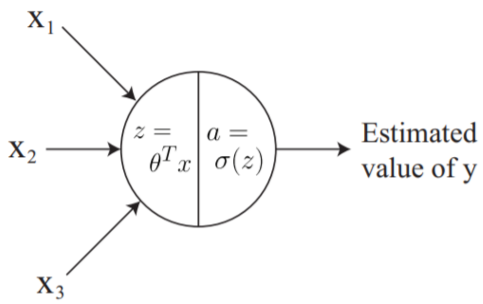

Figure 2.1: A single neuron.
[10]

These single neurons can be stacked so that one neuron passes its output as input into the next neuron. The resulting network of neurons can, hence, consist of several layers of neurons, each with their own learnable weights and biases. Used in conjunction with an appropriate loss function and optimization algorithm, such a network can be used to learn any complex function, if sufficient data is available for training. Forward propagation through the network yields its prediction for a given input. This prediction is compared with the actual class label, and the loss is computed. Backward Propagation is used to compute the gradients of the loss function with respect to the parameters, which are then used by the optimization algorithm to adjust the parameters and minimize the loss over a number of iterations.

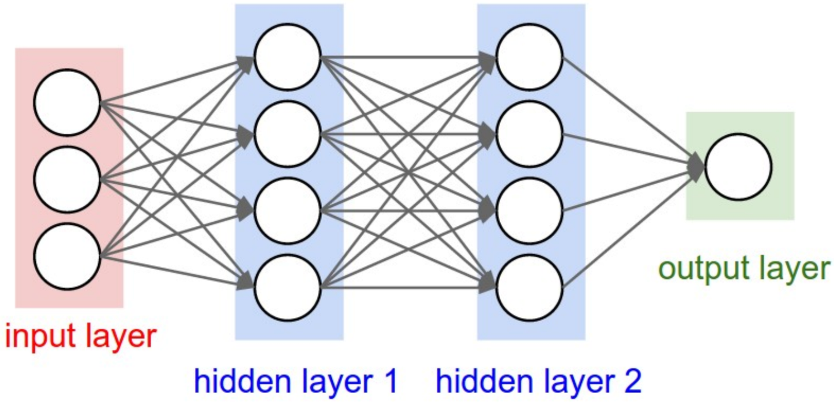

Figure 2.2: A two layer neural network with fully connected layers.
[7]

2.3 Convolutional Neural Networks

Regular neural networks do not scale well to full images. If the input to the neural network is a 200x200 RGB image, the number of weights for each neuron will be 200*200*3 = 120,000 weights. For large networks, the total number of learnable parameters become very large and lead the model to potentially overfit the training data, unless the training set is adequately large.

A convolutional neural network (CNN) is a sequence of layers. Each layers transforms an input volume (images are represented as a three dimensional matrix) of activations to another with some differentiable function which may or may not have parameters.[7, 11] These layers are of three main types:

Convolutional Layer

This is the core building block for convolutional networks. It is based on the convolution operation on images.

Each convolutional layer of a CNN consists of N kernels or filters of a certain volume of neurons sized $f \times f \times d$, with f being the spatial dimension and d being the number of feature channels of the kernel, which is same as the number of channels in the image at its input (D_i). Every one of these filters is convolved with a corresponding volume of the input image, and is slid through the entire image of size $H_i \times W_i \times D_i$. Convolution refers to the summation of the element-wise dot product of the neurons in each filter with the corresponding values in the input, for each position in which the filter is aligned with the image. Based on this notion, a convolution with a single filter at each layer results in a two dimensional output of a certain size. [7, 9, 11]

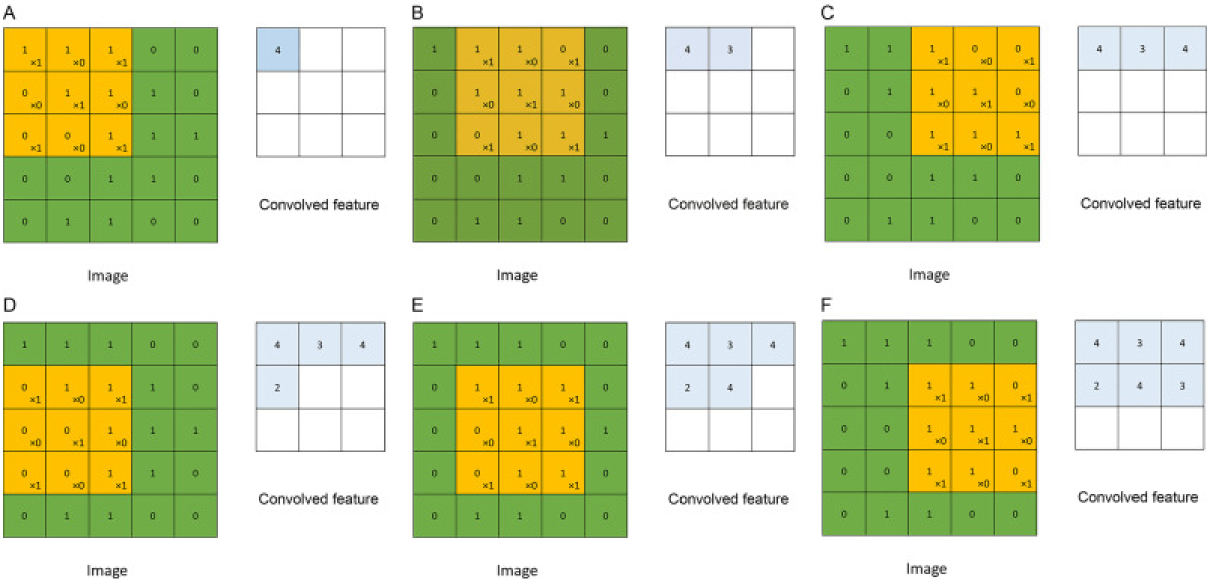

Figure 2.3: The convolution operation performed using a 3x3 filter on a 5x5x1 image. [1]

The intervals with which the filter moves in each spatial dimension is decided by the **stride** s. In order to prevent undue shrinkage of the volume along the spatial dimension, the image can be padded with pixels along the outer edges. The width of this **padding**, in pixels, is given by another hyper-parameter, p.[11, 9] The convolution operation is repeated for each of the N filters, and the resulting N activation maps are stacked together across the third dimension giving an output volume of dimensions:

$$H_o = \frac{H_i - f + 2p}{s} + 1$$
$$W_o = \frac{W_i - f + 2p}{s} + 1$$
$$D_o = N$$

In a convolutional layer, each neuron is connected to only a local region of the input volume, called the receptive field of that neuron. The extent of connectivity is limited to the filter size along the spatial dimensions, but is full along the depth axis.[7] It should also be noted that all activations belonging to a particular channel in the output volume correspond to a single filter applied on the input volume, and hence depend on the same shared parameters. Local connectivity and parameter sharing not only help reduce the number of learnable parameters, but also make the CNN good at capturing **translation invariance**. This makes them an ideal choice for the image classification problem.

Pooling Layer

It is common to periodically insert a Pooling layer in-between successive convolutional layers in a CNN. Its function is to progressively reduce the spatial size of the representation to reduce the amount of parameters and computation in the network, and hence to also control overfitting.[7] The most common form of pooling layer in CNN architectures employs filters of size 2×2 with a stride of 2, taking a max over 4 cells of the input image. It is worth noting that while this halves the width and height of the image, the depth remains unaffected as the max operation is applied independently to each channel of the image. This down-sampling process effectively discards 75% of the activations.

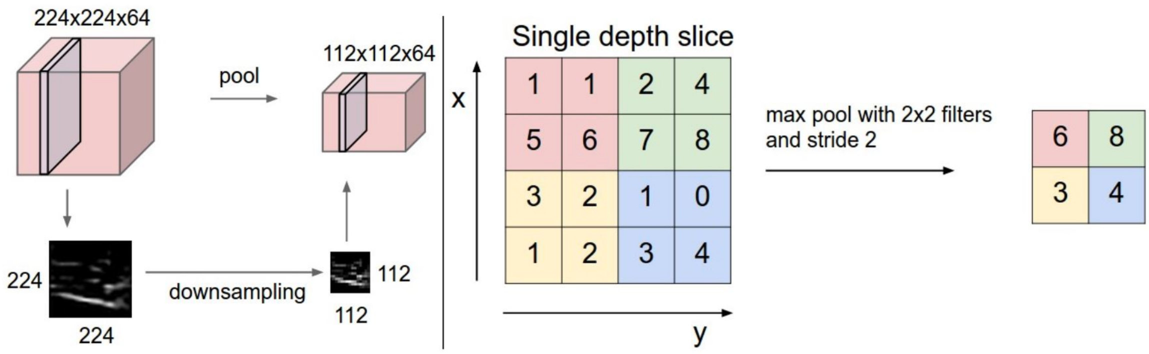

Figure 2.4: (1) A typical max pooling layer. (2) The max pooling operation. [7]

Fully Connected Layer

Once higher level features are detected from the preceding convolution and pooling layers, a fully connected layer is usually attached at the end of the network. This layer is fully connected to all activations in the previous layer, as in regular neural networks, allowing all the features learned by the network to be taken into account by the output layer.

In practice, fully connected layers have an equivalent representation as a convolutional layer having N filters with dimensions equal to those of the input image. The output of this layer will thus be a volume of dimensions $1 \times 1 \times N$. This simple change allows the same CNN to be applied on images with arbitrary spatial dimensions and classify them in a single pass of forward propagation, instead of iterating on different crops of adequate size. This is the basic intuition behind **Fully Convolutional Networks**.[8]

All of these different types of layers can be stacked together in various ways to form a CNN.

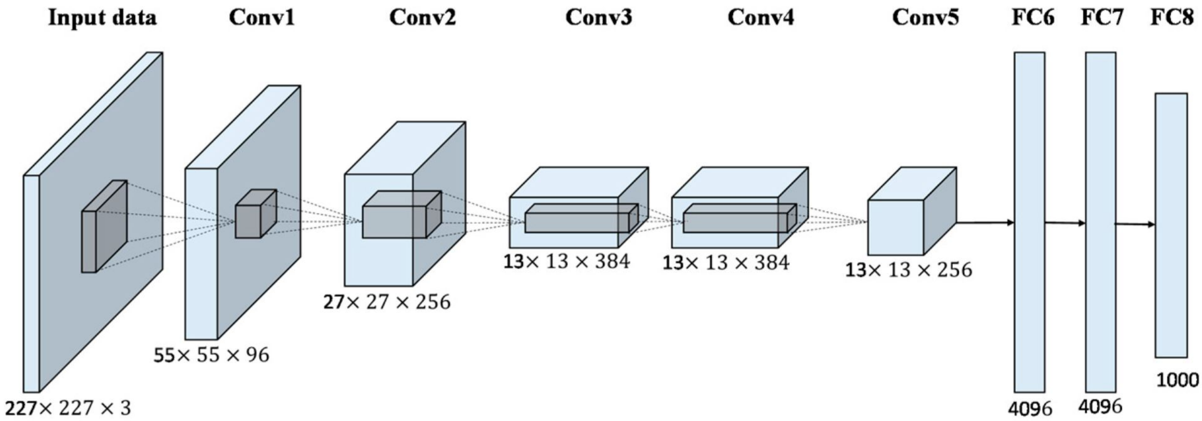

Figure 2.5: AlexNet - the first work that popularized the use of CNNs and GPUs to accelerate deep learning.

[6, 2]

2.4 Semantic Segmentation

References

[1] Ehsan Fathi and Babak Maleki Shoja. "Deep neural networks for natural language processing". In: *Handbook of statistics*. Vol. 38. Elsevier, 2018, pp. 229–316. (Visited on 06/02/2020).

[2] Xiaobing Han et al. "Pre-trained alexnet architecture with pyramid pooling and supervision for high spatial resolution remote sensing image scene classification". In: *Remote Sensing* 9.8 (2017), p. 848.

[3] *History—Indian Institute of Remote Sensing*. URL: `https://www.iirs.gov.in/historyandobjectives` (visited on 06/01/2020).

[4] *Indian Institute of Remote Sensing*. URL: `https://www.iirs.gov.in/institute-profile` (visited on 06/01/2020).

[5] Manideep Kolla, Aniket Mandle, and Apoorva Kumar. *Eye in the sky*. GitHub Page. 2018. URL: `https://github.com/manideep2510/eye-in-the-sky` (visited on 06/01/2020).

[6] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton. "Imagenet classification with deep convolutional neural networks". In: *Advances in neural information processing systems*. 2012, pp. 1097–1105.

[7] Fei-Fei Li, Justin Johnson, and Serena Yeung. *cs231n - Lecture Slides*. 2017. URL: `http://cs231n.stanford.edu/slides/2017/` (visited on 06/01/2020).

[8] Jonathan Long, Evan Shelhamer, and Trevor Darrell. "Fully convolutional networks for semantic segmentation". In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2015, pp. 3431–3440.

[9] Shivaprakash Muruganandham. *Semantic segmentation of satellite images using deep learning*. 2016.

[10] Andrew Ng and Kian Katanforoosh. *cs229 - Lecture Notes*. 2018. URL: `http://cs229.stanford.edu/notes/` (visited on 06/01/2020).

[11] Andrew Ng, Kian Katanforoosh, and Younes Bensouda Mourri. *Convolutional Neural Networks*. 2017. (Visited on 06/02/2020).

[12] *Remote Sensing - Wikipedia*. URL: `https://en.wikipedia.org/wiki/Remote_sensing` (visited on 06/01/2020).

[13] Jürgen Schmidhuber. "Deep learning in neural networks: An overview". In: *Neural networks* 61 (2015), pp. 85–117.