Multi-Classification of Satellite Images

Mr. Vibhor Sharma[1] Shivan Nawal[2]
[1]Assistant Professor
[1,2]Department of Information Technology
[1,2]Maharaja Agrasen Institute of Technology, Sector-22, Rohini, New Delhi-110086, India, India

Abstract— Every minute, the world loses an area of forest the size of 48 football fields. And deforestation in the Amazon Basin accounts for the largest share, contributing to reduced biodiversity, habitat loss, climate change, and other devastating effects. But better data about the location of deforestation and human encroachment on forests can help governments and local stakeholders respond more quickly and effectively. Convolution layers are used to extract the features from input training samples. Each convolution layer has a set of filters that helps in feature extraction. In general, as the depth of CNN model increases, complexity of features learnt by convolution layers increases. For example, first convolution layer captures simple features while the last convolution layer captures complex features of training samples. Another name for feed forward networks like these are Shift Invariant or Space Invariant Artificial Neural Networks (SIANN), as their architecture is mainly based on their weight sharing attribute and transient architecture. With upcoming Deep Learning techniques every year, the field of computer vision research has been advancing at a fast pace resulting in building robust models that have been created benchmarks on a world scale.

Key words: Space Invariant, Artificial, Convolution, Neural Network

I. Introduction

The world has developed a lot from its primordial history. Technology has advanced by leaps and bounds and has alleviated the way of life of everyone on the planet. But the problems that are present today are also a side effect of this advancement. With deforestation, mining and other such activities for extracting precious minerals from the earth, we have resorted to destroying our environment and have decided to ignore the consequences of the same. But this technology can also be put to judicious use. Initiatives have been taken to use computers, especially, algorithms capable of actually predicting and classifying the environment present in captured images. Planet, designer and builder of the world's largest constellation of Earth-imaging satellites, will soon be collecting daily imagery of the entire land surface of the earth at 3-5 meter resolution. While considerable research has been devoted to tracking changes in forests, it typically depends on coarse-resolution imagery from Landsat (30 meter pixels) or MODIS (250 meter pixels). This limits its effectiveness in areas where small-scale deforestation or forest degradation dominate.

In this paper, we try to develop a model based on already developed computer vision research that will help us in classifying satellite based images and paint a more accurate picture of the environment contained within an image. The accuracy for this problem will be selected as a metric that is widely used to measure the actual precision of the model against the classification of such images.

II. Understanding Satellite Data

Fig. 1: Map location for the Amazon

The data for this project is based off of the Amazon basin and the areas adjoining to it. The rainforest in itself is quite flourishing in flora and fauna and has been chosen as the specific location to serve as a primary model for our images taken from the satellite. These images are available in JPeg and GeoTiff formats. To assemble this data set we set out with an initial specification of the phenomena we wished to find and include in the final data set. From that initial specification we created a "wish list" of scenes where we included a ballpark number of scenes required to get a sufficient number of chips to demonstrate the phenomena. This initial set of scenes was painstakingly collected by the Berlin team using Planet Explorer. All told, this initial set of scenes numbered approximately 1600 and covered a land area of thirty million hectares. This initial set of scenes was then processed using a custom product processor to create the jpg and 4-band tif chips. Any chip that did not have a full and complete four band product was omitted. This initial set of over 150,000 chips was then divided into two sets, a "hard" and an "easy" set. The easy set contained scenes that the Berlin team identified as having easier-to-identify labels like primary rainforest, agriculture, habitation, roads, water, and cloud conditions. The harder set of data was derived from scenes where the Berlin team had selected for shifting cultivation, slash and burn agriculture, blow down, mining, and other phenomenon. The chips for this competition were derived from Planet's full-frame analytic scene products using our 4-band satellites in sun-synchronous orbit (SSO) and International Space Station (ISS) orbit. The set of chips for this competition use the GeoTiff format and each contain four bands of data: red, green, blue, and near infrared. The specific spectral response of the satellites can be found in the Planet documentation. Each of these channels is in 16-bit digital number format, and meets the specification of the Planet four band analytic ortho scene product.

III. Classification

Classification is the problem of identifying to which of a set of categories a new observation belongs, on the basis of a training set of data containing observations (or instances) whose category membership is known in machine learning and statistics. For example, putting email into "spam" or "non-spam" classes or assigning a diagnosis to a given patient as described by observed characteristics of the patient (gender, blood pressure, presence or absence of certain symptoms, etc.). Classification is an example of pattern recognition. In the terminology of machine learning, classification is considered an example of supervised learning, i.e. learning where a training set of correctly identified observations is available. The corresponding unsupervised procedure is known as clustering, and involves grouping data into categories based on some measure of similar characteristic or distance. An algorithm that implements classification, especially in a concrete implementation, is known as a classifier. The term "classifier" sometimes also refers to the mathematical function.

In supervised training, both the inputs and the outputs are provided. The network then processes the inputs and compares its resulting outputs against the desired outputs. Errors are then propagated back through the system, causing the system to adjust the weights which control the network. This process occurs over and over as the weights are continually tweaked. The set of data which enables the training is called the "training set." During the training of a network the same set of data is processed many times as the connection weights are ever refined. A supervised learning algorithm analyses the training data and produces an inferred function, which can be used for mapping new examples. An optimal scenario will allow for the algorithm to correctly determine the class labels for unseen instances. The class labels for this task were chosen in collaboration with Planet's Impact team and represent a reasonable subset of phenomena of interest in the Amazon basin. The labels can broadly be broken into three groups: atmospheric conditions, common land cover/land use phenomena, and rare land cover/land use phenomena.

IV. Labels

A. *Cloud Cover Labels*

I have chosen to include a cloud cover label for each chip. These labels closely mirror what one would see in a local weather forecast: clear, partly cloudy, cloudy, and haze. For our purposes haze is defined as any chip where atmospheric clouds are visible but they are not so opaque as to obscure the ground. Cloudy images have 90% of the ship obscured by opaque cloud cover

B. *Primary Rainforest*

The overwhelming majority of the data set is labeled as "primary", which is shorthand for primary rainforest, or what is known colloquially as virgin forest. Generally speaking, the "primary" label was used for any area that exhibited dense tree cover.

C. *Water (River and Lakes)*

Rivers, reservoirs, and oxbow lakes are important features of the Amazon basin, and we used the water tag as a catch-all term for these features.

D. *Habitation*

This includes anything from dense urban centers to rural villages along the banks of rivers.

E. *Roads*

For our data, all types of roads are labeled with a single "road" label. Some rivers look very similar to smaller logging roads, and consequently there may be some noise in this label.

F. *Slash and Burn*

Slash-and-burn agriculture can be considered to be a subset of the shifting cultivation label and is used for areas that demonstrate recent burn events. This is to say that the shifting cultivation patches appear to have dark brown or black areas consistent with recent burning.

G. *Blooming*

Blooming is a natural phenomenon found in the Amazon where particular species of flowering trees bloom, fruit, and flower at the same time to maximize the chances of cross pollination. These trees are quite large and these events can be seen from space.

H. *Blow Down*

Blow down, also called windthrow, is a naturally occurring phenomenon in the Amazon. Briefly, blow down events occur during microbursts where cold dry air from the Andes settles on top of warm moist air in the rainforest.

I. *Agriculture*

Commercial agriculture, while an important industry, is also a major driver of deforestation in the Amazon. For the purposes of this dataset, agriculture is considered to be any land cleared of trees that is being used for agriculture or range land.

J. *Selective Logging*

The selective logging label is used to cover the practice of selectively removing high value tree species from the rainforest (such as teak and mahogany).

V. Resnet

The authors of ResNet observed, no matter how deep a network is, it should not be any worse than the shallower network. Hence, it might be useful to explicitly force the network to learn an identity mapping, by learning the residual of input and output of some layers (or subnetworks). Suppose the input of the subnetwork is x, and the true output is H(x). The residual is the difference between them: $F(x)=H(x)-x$. As we are interested in finding the true, underlying output of the subnetwork, we then rearrange that equation into $H(x)=F(x)+x$. Where traditional neural nets will learn H(x) directly, ResNet instead models the layers to learn the residual of input and output of subnetworks. This will give the network an option to just skip subnetworks by making $F(x)=0$, so that $H(x)=x$. In other words, the output of a particular subnetwork is just the output of the last subnetwork.

VI. TRAINING

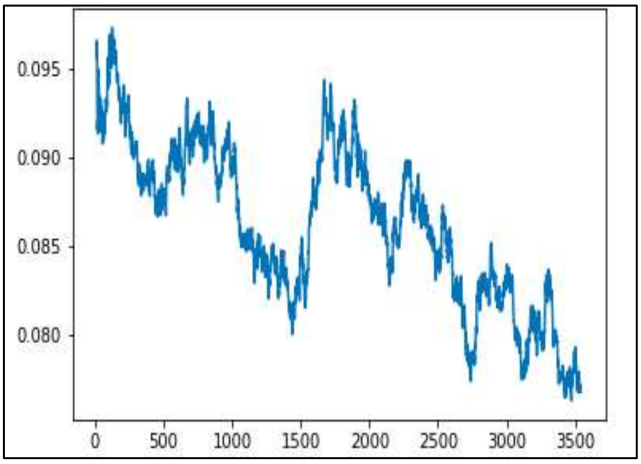

Fig. 2: Training Loss

In the training phase, the correct class for each record is known (this is termed supervised training), and the output nodes can therefore be assigned "correct" values -- "1" for the node corresponding to the correct class, and "0" for the others. It is thus possible to compare the network's calculated values for the output nodes to these "correct" values, and calculate an error term for each node (the "Delta" rule). These error terms are then used to adjust the weights in the hidden layers so that, hopefully, the next time around, hopefully the values will be closer to the "correct" values.

VII. FINAL ACCURACY

```
f2 Score: 0.9285756738073999
f2 Score: 0.9325735931488134
f2 Score: 0.9345646226806884
f2 Score: 0.9331467241762751
f2 Score: 0.9349772800026489
```

Fig. 3: Accuracy on the F2 Metric

In this project, we studied about Neural Networks in theory and applied these concepts mathematically to create an algorithm that could actually differentiate between images and was able to predict the various labels associated with the satellite images with an accuracy close to 94%

VIII. CONCLUSION

The results and observations show that on training the Resnet model on the satellite image data, we were able to gain an accuracy of around 94% on the F2 metric which is the best suited metric for classification problems. Studying Convolutional Neural Networks helped in the understanding of the practical aspect of computer vision and was useful in carrying out the practical aspects of the same. The amount of overfitting used can be increased using regularization techniques like Dropout.

ACKNOWLEDGMENT

I would like to thank Mr. Vibhor Sharma for his immense help and support, useful discussions and valuable recommendations.

REFERENCES

[1] https://en.wikipedia.org/wiki/Convolutional_neural_network
[2] https://wiseodd.github.io/techblog/2016/10/13/residual-net/
[3] http://cv-tricks.com/cnn/understand-resnet-alexnet-vgg-inception/
[4] https://www.kaggle.com/c/planet-understanding-the-amazon-from-space
[5] https://www.kaggle.com/c/planet-understanding-the-amazon-from-space/data
[6] https://arxiv.org/pdf/1512.03385.pdf
[7] https://en.wikipedia.org/wiki/Supervised_learning

