AMSat: An Holistic System to Classify Temporal Satellite Imagery

Aman Agarwal, Aditya Mishra, Priyanka Sharma

Institute of Technology, Nirma University

{15bce006, 15bce003, priyanka.sharma}@nirmauni.ac.in

Introduction

Detecting changes on the earth's surface is vital to predict and avoid catastrophic events, monitor the progression of buildings, country's growth rate, vegetation cover, or climatic effects. These changes can be noticed from different kinds of low and high resolution or multi-spectral and multi-temporal satellite images. There are different kinds of change detection techniques to observe changes in the images like principal component analysis, spectral change vector analysis, post-classification method, kernel method, etc. Deep learning based change detection provides better accuracy because these methods are directly based on pixel comparison and do not require hand-engineered features. The change detection accuracy on satellite images may depend on various factors like the image resolution, type of object to be monitored, or atmospheric noises. Through this work, we present AMSat, a deep learning based approach to classify temporal satellite images in an end-to-end fashion through an inflated 3D network architecture. We also show how a small dataset can be used to train a network which generalizes well on the unseen data without any post-processing, besides, giving a good accuracy.

Methods

Data Preparation

We collected around 450 images belonging to 80 different locations around the world, differing by seasons and cloud cover and categorized them into five classes viz. airfield, commercial (construction), residential (construction), road, and unknown. The images were collected over a span of 4 years showing regions in different phases of construction. To increase the data size synthetic data was generated using various random data augmentation methods. Fig. 1 shows the input to the network with random data augmentations applied.

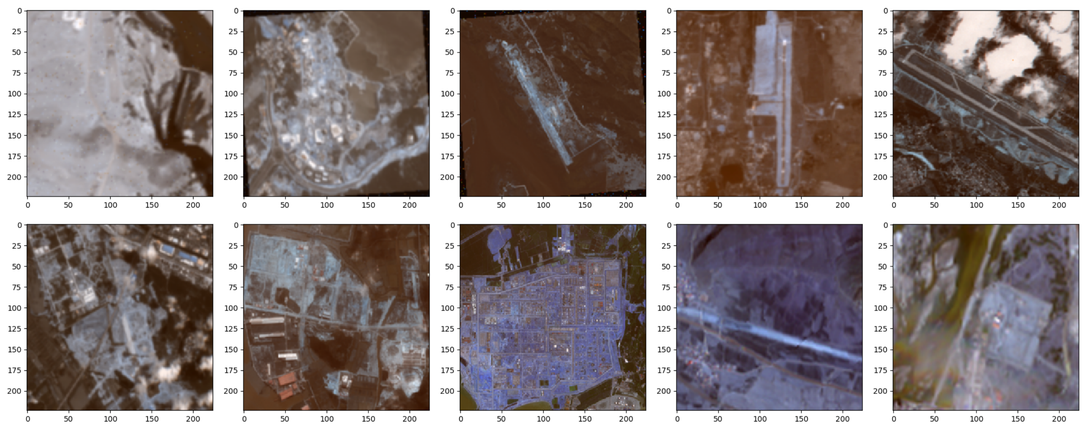

Figure 1: Sample of input images just before feeding into the model. The images were normalized by subtracting the mean of the ImageNet data.

Model

A pre-trained I3D network was selected as the backbone which was originally trained for the task of video classification. It is an extension of the 2D Inceptionv3 network to 3D, leveraging the understanding of images from the ImageNet dataset to classify video frames. A schematic representation of the I3D network is shown in Fig 2.

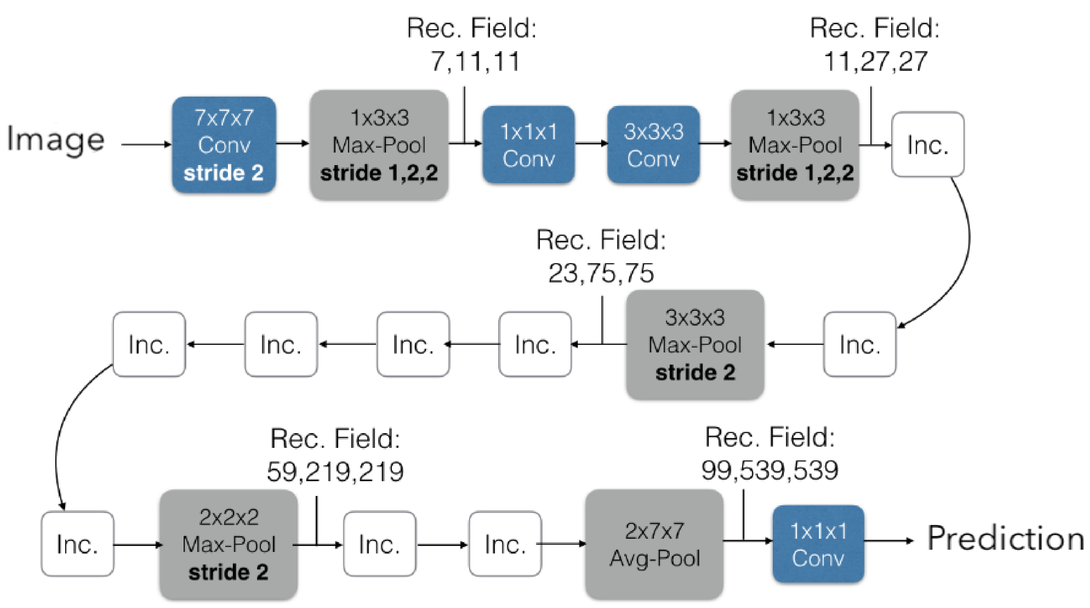

Figure 2: Inflated inception (I3D) model architecture.

Training

The input to the network was fed by combining four image frames taken from the same location in chronological order to make the network incorporate temporal features. AWS EC2 instance was used for training purposes which used Nvidia V100 GPU as the accelerator. A batch size of 8 was used with categorical cross-entropy as the loss function and Adam optimizer, with the learning rate of $1e-5$.

Deployment

The deployment was done on the same AWS EC2 instance on the cloud, with the workflow as shown in Fig. 3. The average time taken (per run) by the whole deployment process over 100 target sites was nearly 60-90 minutes.

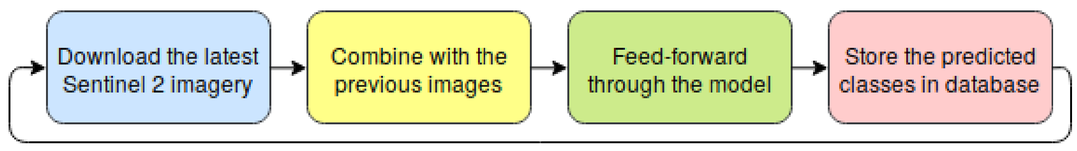

Figure 3: Workflow of the model deployment process.

Results

After training for around 1500 epochs, the model reached saturation, giving the training accuracy of 100% (small dataset) while the average validation accuracy of 85%. Table 1 shows a detailed analysis on the validation set while Fig. 4 shows some of the classification results.

Table 1: Class-wise results on the validation set.

Class	Samples	Accuracy
Airfield	23	83%
Commercial	34	94%
Residential	21	91%
Road	34	86%
Unknown	18	72%
Total	**130**	**85%**

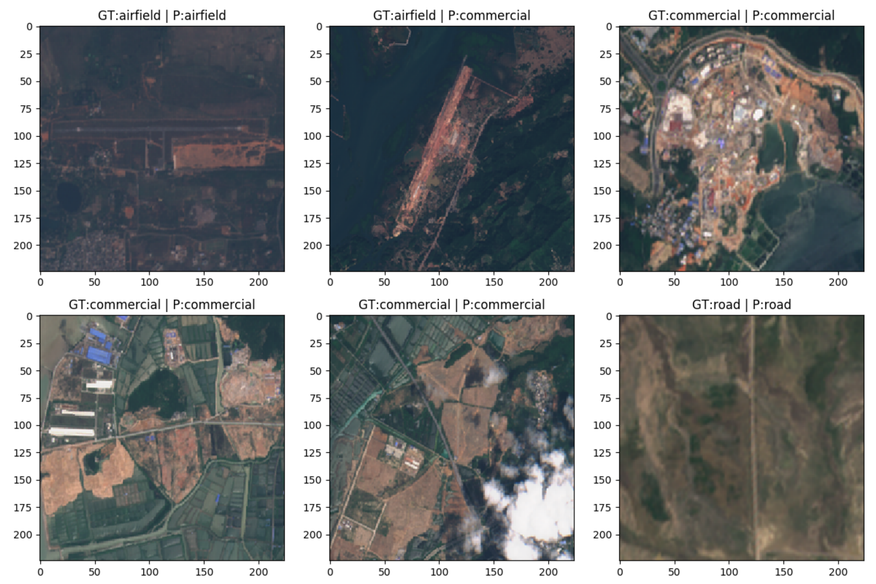

Figure 4: Results on the validation set

Conclusion & Future work

Through this work, we showcase the full pipeline of building a machine learning model starting from the data gathering phase to deploying the model for inference. We also showed how a very small dataset can leverage the power of transfer learning and data augmentation methods and be used for training a deep neural network. AMSat achieved an overall accuracy of 85% on the validation set consisting of 130 images from various categories. Such systems can be used by organizations, builders, or farmers to monitor their assets like construction sites, lands, crop fields, or even the GDP of a country.

Acknowledgement

We would like to thank the Institute of Technology, Nirma University for giving us this opportunity to showcase our work.

We would also like to thank our colleagues, friends, and family members who supported and encouraged us throughout the course of this work.