Land Cover Classification using Satellite Data

Kevin Patel
kevin.patel@iitgn.ac.in
IIT Gandhinagar

Soham Pachpande
pachpande.soham@iitgn.ac.in
IIT Gandhinagar

Smeet Vora
smeet.vora@iitgn.ac.in
IIT Gandhinagar

ABSTRACT

Land is a key resource, especially for a country like India which has a strong agriculture sector. As India urbanises, there has been an increase in non-agricultural land usage and in turn misuse of land which leads to irreversible environmental damage. The knowledge about land usage is becoming increasingly important to design proper policies and infrastructure to tackle the resource crunch of arable land and water.

The recent developments in space technology and the availability of satellite imagery presents an opportunity to identify and maintain real-time records of land usage. We propose a deep learning model to identity land usage and land cover type of a patch of pixels with high accuracy. The end goal is to have a complete land classification picture of India and to study environmental state and its transition over time span of data available which would be an aid to researchers and policy makers.

CCS CONCEPTS

• **Computing methodologies → Machine learning**.

KEYWORDS

neural networks, transfer learning, pixel-bassed tagging

ACM Reference Format:

Kevin Patel, Soham Pachpande, and Smeet Vora. 2019. Land Cover Classification using Satellite Data. In *Proceedings of Machine Learning (IIT Gandhinagar '19).* ACM, New York, NY, USA, 5 pages. https://doi.org/10.1145/nnnnnnn.nnnnnnn

1 INTRODUCTION

The knowledge about land cover has become increasingly important for a nation as it plans to overcome problems like uncontrolled development, depletion of resources, loss of prime agricultural lands, destruction of wetlands and loss of wildlife habitat.

Traditionally in India, the statistics of land usage are compiled from the village records which are manually surveyed by employees. This approach is very inefficient in terms of maintaining the relevance of the data and scanning large geographical areas. We propose a technique which uses satellite imagery from Sentinel-2, the Earth observation mission from the EU Copernicus Programme. We use Eurosat dataset [2] to train our machine learning models and test our work over satellite images of various cities.

2 RELATED WORK

Since last few years, there have been significant technological advances which have led to an increase in the availability of Very Fine Spatial Resolution (VSFR) imagery. Land cover classification is not being widely used for the purpose of extracting meaningful information from these images. However, land cover and land use classification is a challenging task as land usage cannot be directly interpreted from the tone, texture, color or shape of an image feature. This is why LC classification is still an open problem as there is no single internationally accepted LC classification system.

Since the last decade, a lots of effort has been made in developing a Land cover classification system from satellite imagery. There are two approaches for classifying land cover namely pixel-based method and object based method. Pixel based classification system uses only the individual pixels for classifying them into a land cover category without taking into consideration its neighboring pixels. This method often leads to image speckle or a few misclassified pixels in between the pixels of a different land cover category which results overall inaccuracies when applied to high resolution images.

Another method commonly used for land cover classification is object based classification. In this method, image segmentation is used to aggregate individual pixels into a set of homogeneous image objects which are highly likely to be in the same class. Now these objects are classified into land cover classes. Major challenge in object based classification method is selection of segmentation scales to obtain objects that correspond to a specific class. This method often results in under segmentation and over segmentation. This means once an object is defined it can only have a single class, so if there exists a different class inside that object it wont be classified in a different category.

A lot of new techniques have emerged in the last few years for image classification tasks. Techniques such as Scale-Invariant Feature Transform (SIFT), Historgram of Oriented Gradients (HOG) are invariant to geometric and photometric transformations. Using these techniques, we can get good accuracies, especially in tasks like object recognition. But given, the high dimensionality of the feature space, we need to able to obtain a more compact but expressive representation of the image.

Bag of Visual Words (BOVW) is a common tool used to represent an image into a set of features. Features consist of keypoints and descriptors. Keypoints are the standout points in an image that do not change when the image is rotated, shrinked or expanded. Descriptors are the descriptions of these keypoints. We use these keypoints and descriptors to construct dictionaries and represent an image as frequency histogram of features that are in the image. Eventually these histograms are fed to a classifier. This paper [9] has successfully applied SIFT-BOVM approach for land use classification task.

Dense SIFT can also be used for feature extraction, the major difference is that in dense SIFT we get a descriptor at every location, while in normal SIFT we get descriptors at locations determined by Lowe's algorithm. In [8] histogram of dense SIFT features are

IIT Gandhinagar '19, April 23–24, 2019, Gujarat, India

ACM ISBN 978-x-xxxx-xxxx-x/YY/MM…$15.00
https://doi.org/10.1145/nnnnnnn.nnnnnnn

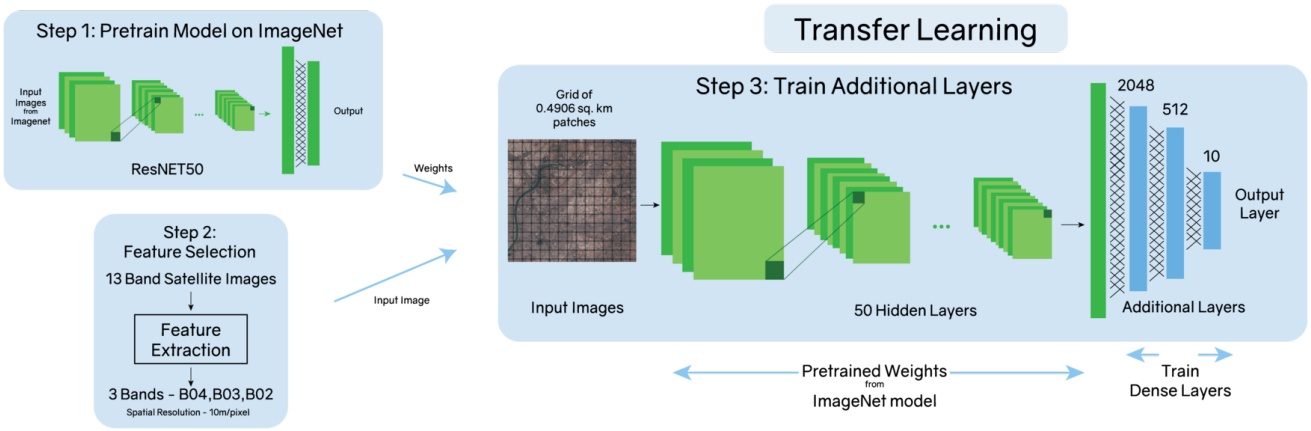

Figure 1: Our Transfer Learning Model Implementation

matched through a fast approximation of the Earth movers distance. This paper [3] proposes another way to improve histogram based feature extraction by using histograms regarded as Dirichlet distributed probability mass functions and then transform using a Fisher kernel to enhance their discriminative power. Another variant is the so called max-SIFT, which is obtained from the maximum of a SIFT descriptor and its flipped copy.

Recent papers, have been more focused on using CNN for prediction. Fully convolution networks (FCN) have been used in [4] which can be used to apply convolutions and pooling operations to the entire image rather than to a patch, this leads to set of signals at low spatial resolution. Deconvolution networks [5] use an encoder-decoder strategy. In this approach encoder part is similar to a CNN, but the decoder part does the upsampling of the low resolution signals to high resolution images. The structure of decoder typically resembles the mirror image of the encoder.

Land cover classification has also been approached by using more complex architectures. The work by Castelluccio et all in Land use classification in Remote Sensing Images using CNN[1] uses CaffeNet and GoogLeNet. They considered and compared three approaches in this paper (1) training both the networks from scratch (2) fine tuning the network using pre-trained weights and (3) taking the output of the penultimate layer as feature vector for classification.

In the work on Classification of land cover and land use based on CNN [7], different variants of SegNet and LiteNet are used for classification. Segnet is a deep convolution encoder-decoder architecture for multiclass pixelwise segmentation. This paper uses these architectures for pixel based classification of land cover and land use.

3 DATA AND DATASET

EuroSAT dataset was used to train models in this study. The images were captured by sentinel 2 satellite. The dataset consists of 27000 images of size 64x64 pixels. It consists of 10 different classes with 2000 to 3000 images per class. The resolution is 10 meter per pixel. It covers 13 spectral bands including RGB which was majorly used in this study. There are different land cover and land use classes in the dataset. Agricultural land use classes include annual crop, permanent crop and pasture. Artificial surfaces include highway, residential buidings and industrial buildings. Water bodies include river, sea and lake. Green cover classes include foreset and herbaceous vegetation.

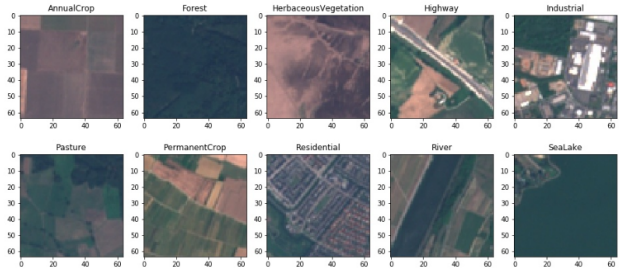

Figure 2: Sample Images from all 10 classes

4 APPROACH

4.1 Baseline Models

We start our baseline approach by implementing Random Forrest and Naive Bayes. We first break down a large satellite image into 64×64 sized tiles. We use a pixel based approach to classify land cover and land usage. In this approach, we will take the 64×64 tiles and classify them into 10 classes.

We flatten 64×64 pixel images and pass bit values corresponding to spectral bands to the models as input. We performed Nested Cross Validation to find optimal parameters. The Sentinel-2 Satellite program records Multi-spectral data with 13 bands in the visible, near infrared, and short wave infrared part of the spectrum. We trained and tested models on the RGB bands and found accuracy of 62%.

4.2 Applying Dimensionality reduction

After flattenning the 64×64 images, we are left with 786 features. We try to reduce the dimensionality of this data to remove the

unimportant and uncorrelated features by using PCA. PCA was used to reduce the number of features from 786 to a set of different lower values, this reduced the complexity of data and now it is easier to apply a variety of Machine Learning algorithms on this data. We applied various machine learning classification algorithms but there was no improvement in the accuracy.

4.3 Neural Networks

Convolutional Neural Networks (CNNs) are a type of Neural Networks [13] which give impressive results on image classification challenges [12], [21], [23].
In CNNs, each layer is comprised of various sublayers of neurons which operate in parallel with the previous layer, to extract a number of features at once. It consists of filters which are learned and which elaborates the three input color bands. A CNN architecture typically comprises of these layers -

- **Convolution layers :** These layers compute the convolution of the input image by sliding a small kernel over the image. Layers at the start learn low level features in an image while deeper layers learn a larger portion of the image by combining the low level ones computed by the starting layers. Each layer has a few parameters like number of filters and stride between windows.
- **Pooling layers :** These layers reduce the size of image, so as to reduce the number of parameters to be learned. Relevant hyper parameters are the support of the pooling window and the stride between the windows.
- **Normalization layers :** These layers provide generalisation. These are typically used with sigmoid neurons.
- **Fully-connected layers :** These are layers are generally added towards the end of the neural network. They basically summarize all the information conveyed by the lower level layers.

The caveat is that large labelled datasets are required to train the models. When training CNNs on the available remote-sensing datasets, we run into the major problem of limited training data. The labelled satellite image datasets available are far too small for accurately training a large CNN.
To address this problem, we pre-train our base CNN models on a similar data and problem and re-purpose it by fine tuning additional layers to the task of interest. Particularly, optical remote sensing images from satellite in the Red, Green and Blue bands are similar to general purpose optical images, for example the images found in ImageNet data set(Cite). Additionally, the task of fine tuning additional layers is faster and computationally less expensive than training entire network from scratch which may take weeks to complete. We use this approach and train all our models on ImageNet dataset to get a set of weights. The fine tuning can be done in two methods:

(1) Build a classifier on top of the classification of base CNN model.
(2) Take the output of penultimate layer in the base model and add additional dense layers at the end depending on the task at hand. Fix the pre-trained weights and use the training images to fine tune the weights of the additional layers at the end and some layers of the base model.

While the first method is easier to implement, it would not adapt well to the spatial and temporal variations of remote sensing image data.The second solution is of more interest, as it exploits the full potential of CNNs and giving a deeper adaptation to the data of interest. In this paper, we implement the second solution where after removing the last layer from pre trained model of ResNet50, two more trainable dense layers were added of sizes 2048 and 512. Then one final classification layer of size 10 was added. One needs to decide which layers to freeze in the base model and which layers be allowed to learn new weights. Considering the characteristics of our data and amount of training data available we keep the additional layers trainable and freeze all the early layers. The reasoning is that the base layers would capture low level features which would better fit our problem.
For our base model, we use VGG16, VGG19 and ResNet50 [cite] architectures. We train these models on ImageNet dataset which contains general purpose images [mention details of this dataset]. We delete the last softmax output layer and add 3 trainable Dense layers including the output layer. We only choose Red, Green and Blue bands of the multi spectral image available and scale 64×64 pixel RGB tiles to 224×224 pixel and use these for training the models. The reasoning is explained in the Experimental Setup section. After tuning hyper parameters, we found ResNet50 to perform better than the rest of the models as discussed in the Results section.

5 EXPERIMENTAL SETUP

We get our test images on different regions of India and the world from Sentinel Hub made by Synergise and European Space Agency. While we have data about 13 bands with differenct frequencies, we only choose Red, Green and Blue bands of the multi spectral image available. This is done because the ImageNet dataset is in Red, Green and Blue bands and the rest of 10 bands from Sentinel sensors will no correlate with the weights obtained from pre-training phase. This is due to difference in respective charecteristics. We also scale 64×64 pixel RGB images to 224×224, the input size of ImageNet images according to the pre-trained model and pass them as input. The scaling is done as the spatial resolution is low when compared to general purpose images in ImageNet which leads to loss of information during forward propagation. This lead to an improvements in the result. The figure below shows the working of our model on the satellite image tile taken of New York City.

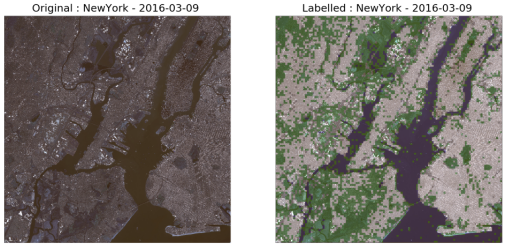

Figure 3: New York City

6 RESULTS

We use 3 baseline models along dimensionality reduction to extract important features from the images. We achieve reasonable results, giving us a good baseline to judge the Neural Network approach.

Model	Accuracy	F1 Score
Decision Tree	36.41	0.37
Naive Bayes	59.17	0.605
Random Forrest	32.29	0.36

Accuracy of Baseline Models

We perform dimensionality reduction by using Support Vector Decomposition(SVD). SVD gives us the most important eigenvectors of the input matrices which are images in this case. This helps us remove unnecessary noise and information. This technique has been successfully used for face recognition [6]. In practise we observed a minor increase in our accuracies and F1-scores.

Model	Accuracy	F1 Score
Decision Tree	40.42	0.405
Naive Bayes	58.75	0.62
Random Forrest	43.95	0.46

Accuracy with Dimensionality Reduction

We train our transfer learning models on 70% of train data and test the models on the Test set. We attain highest accuracy of 82.33% using ResNet50 model after fine tuning additional dense layers. The following table describes our results.

Model	Accuracy
VGG 16	82.33
VGG 19	81.9
ResNet50	78.2

Accuracy of Transfer Learning Models

The following confusion matrices showcase the ability of our models to correctly classify satellite images. Uncertainty on some classes like Permanent Crop, Pasture and Forrest is seen. These classes with higher confusion can be merged with other similar classes, for example, Permanent Crop and Forest can be merged into 1 class named Green Cover.

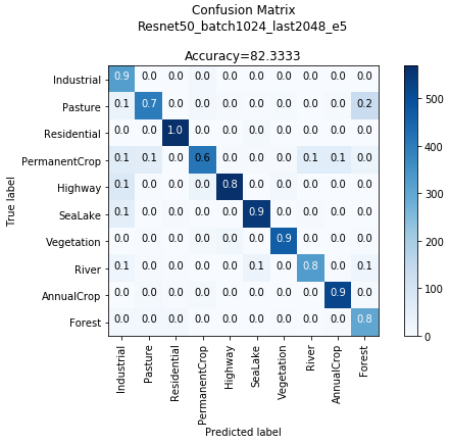

Figure 4: Confusion matrix for ResNet-50 Model

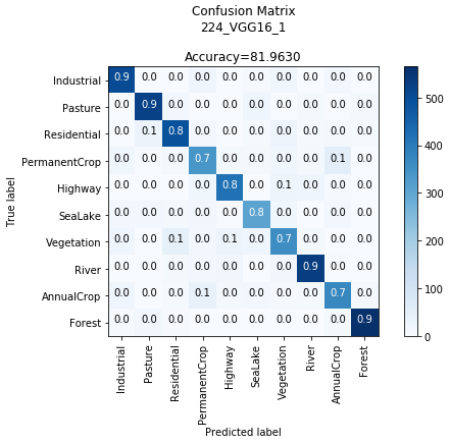

Figure 5: Confusion matrix for VGG-16 Model

7 APPLICATION

We create a platform which visualizes land cover and land usage patterns over a period of time for several regions in India as well as the world. We used satellite images from Sentinel-2 for all the test cases in the platform. We also provide an option to upload a satellite image from Sentinel-2 or similar and the platform will return a image with land cover and classification information in form of overlaid tiles.

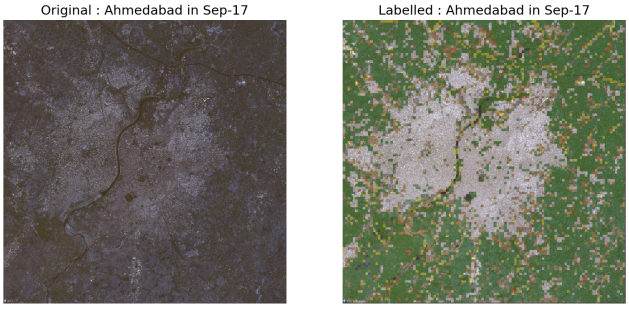

Figure 6: Ahmedabad region in Monsoon

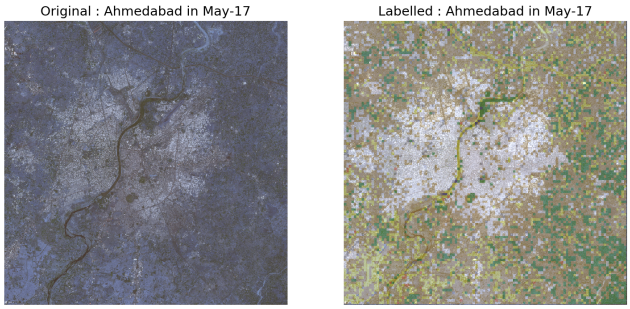

Figure 7: Ahmedabad region in Summer

For the application, we merge the classes Annual Crop, Permanent Crop, Pasture, Vegetation and Forrest into 1 class named Green cover. We merge the Sea-Lake and River classes to form Water-body class and we merge the Residential and Industrial classes to form Urban class. We merge the classes as these don't add any significant information over a period of time and introduce noise as a negative consequence. We observe some interesting patterns. The classification of satellite images from Ahmedabad city which lies in the semi-arid dry region of Western India are very interesting. They clearly portray a distinct pattern in the hot months around May and the monsoon period around September.

8 CONCLUSION

In the course of this paper, we have described the satellite data available for classification, the various baseline algorithms, deep neural network and transfer learning approaches and the results. We also propose an accessible platform to make the work done during this project accessible and of utility to the people.

REFERENCES

[1] Marco Castelluccio, Giovanni Poggi, Carlo Sansone, and Luisa Verdoliva. 2015. Land use classification in remote sensing images by convolutional neural networks. *arXiv preprint arXiv:1508.00092* (2015).

[2] Patrick Helber, Benjamin Bischke, Andreas Dengel, and Damian Borth. 2018. Introducing EuroSAT: A Novel Dataset and Deep Learning Benchmark for Land Use and Land Cover Classification. In *IGARSS 2018-2018 IEEE International Geoscience and Remote Sensing Symposium*. IEEE, 204–207.

[3] Takumi Kobayashi. 2014. Dirichlet-based histogram feature transform for image classification. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 3278–3285.

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

[5] Hyeonwoo Noh, Seunghoon Hong, and Bohyung Han. 2015. Learning deconvolution network for semantic segmentation. In *Proceedings of the IEEE international conference on computer vision*. 1520–1528.

[6] Matthew A Turk and Alex P Pentland. 1991. Face recognition using eigenfaces. In *Proceedings. 1991 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*. IEEE, 586–591.

[7] Yi Yang and Shawn Newsam. 2010. Bag-of-visual-words and spatial extensions for land-use classification. *Proceedings of the 18th SIGSPATIAL International Conference on Advances in Geographic Information Systems - GIS âĂŹ10* (2010). https://doi.org/10.1145/1869790.1869829

[8] Yasen Zhang, Xian Sun, Hongqi Wang, and Kun Fu. 2013. High-resolution remote-sensing image classification via an approximate earth mover's distance-based bag-of-features model. *IEEE Geoscience and Remote Sensing Letters* 10, 5 (2013), 1055–1059.

[9] Jinyi Zou, Wei Li, Chen Chen, and Qian Du. 2016. Scene classification using local and global features with collaborative representation fusion. *Information Sciences* 348 (2016), 209–226.