2020 International Conference on Computational Performance Evaluation (ComPE),
North-Eastern Hill University, Shillong, Meghalaya, India. July 2–4, 2020

A Lightweight CNN Architecture for Land Classification on Satellite Images

Gourab Patowary
Dept. of ECE
Gauhati University
Guwahati, India
gourabmit@gmail.com

Meenakshi Agarwalla
AI Division
Manvi R&D
Guwahati, India
meenakshi267@gmail.com

Sumit Agarwal
Dept. of EEE
IIT Guwahati
Guwahati, India
a.sumit@iitg.ac.in

Manash Pratim Sarma
Dept. of ECE
Gauhati University
Guwahati, India
mpsarma@gauhati.ac.in

***Abstract*—Land cover classification using satellite images is an important tool in the study of terrestrial resources. Satellite based information is presently available as huge sets of high resolution images from a large number of satellites like Sentinel, Landsat-8, etc. Land cover classification from these images is a difficult task because of very large sized data and high variation types. Deep Neural Networks can play a vital role in this regard and can perform classification on these large sized data. Related works in this field have used lighter models and included a large number of handcrafted parameters which requires domain knowledge on the subject. It is realised that most models are too shallow for such a complicated image. In this paper, a deeper Convolutional Neural Network (CNN) model without any satellite image specific parameters is proposed. On SAT4 and SAT6 images, our 13-layered network has achieved better accuracy upto 99.84% and 99.47% which is state-of-the-art. It is still called lightweight model because most models in Artificial Intelligence(AI)-CNN are much deeper and larger than ours.**

***Index Terms*—SimpleNet, Satellite Imagery, Deep Neural Networks, SAT4, SAT6, Remote Sensing, Scene Classification, CNN**

I. Introduction

Classification of large satellite imagery is a challenging task for understanding and portraying land cover information. Land cover is the physical land which includes trees, crop fields, barren lands, rivers, forests, etc. Information about land cover is an input for classifying, planning, monitoring and devising ways to use earth resources potentially in greater interest of the human race. This classification is important for various geospatial application like agriculture, environmental and urban management. Accurate and up-to-date information about land cover goes a long way in helping various government and other agencies to update their plans on regular basis. Traditional methods of gathering land cover information are field surveys that are time consuming as well as include much physical labour. Also, data collected gets stale as it is not feasible to recollect information in short intervals of time. Satellite remote sensing images are a viable source of gathering effective land cover information due to their large view and repetitive coverage area. Raw images taken by satellites cannot be processed directly as they contain multiple bands with large data size. Because of the involvement of such huge data and higher variability in land cover classes, it is not easy to determine the land cover types. The land cover types most useful are:

1) Forestlands / Trees
2) Grasslands
3) Barrenlands
4) Water Bodies

There are two approaches to classify land cover from these images which include:

- Supervised Learning [1], [2]
- Unsupervised Learning [2]

Supervised learning involves training by using data tagged with the correct answer, known as labeled data. This helps in predicting the unforeseen data. Various techniques in Satellite image processing can be classified as supervised learning. The disadvantage with these techniques is that they cannot be scaled to the large amount of data easily. Getting labeled data in supervised learning for satellite data is difficult, hence unsupervised learning algorithms are gaining popularity. Unsupervised learning deals with unlabeled data and the model is left on its own to discover useful features from the input. Deep learning and Convolutional Neural Networks (CNNs) are a class of Unsupervised learning techniques which have shown promising results in land cover classification. Deep neural network has the capability to classify unlabeled data in hierarchical fashion. Recent advances show that deeper features learn from simple level to higher level patterns in an orderly manner. In recent years, applying deep learning to satellite images has become an ongoing topic of research.

This work presents a CNN model which is a modification of the Simplenet [3] model for performing land cover classification on SAT4 and SAT6 [4] datasets. Special features of our proposed model are:

- Light weight model with less number of parameters.
- Efficient in terms of time saving.
- Requires less storage space.
- Reaches a good accuracy in less number of epochs.

It has been able to achieve an accuracy of 99.84% on SAT4 and 99.47% on SAT6 with savings in time as compared to related work.

This paper is divided into six sections. Section I introduces the topic. Section II reviews related work and Section III provides details about the datasets. Section IV highlights the architecture of our proposed neural network. Section V provides the comparison and results and Section VI concludes the paper.he results along with comparison and lastly, the conclusion is included in Section VI

II. Related Work

The spatial resolutions of earlier satellites were so low that most objects of interest were of pixel size. Hence, during the early period of remote-sensing, per pixel analysis was the norm. With the availability of high resolution images, objects of interest such as buildings, roads, etc. cover several pixels. A classic paper [5] hence questions the use of pixel level classification in the realm of high resolution images and from there, the journey of advanced scene classification begins. Some authors call it Object Based Image Analysis (OBIA) [6]. Due to their usefulness, these object based techniques dominated the remote sensing field for the last two decades. But they show their limitations when it comes to classifying the grasslands from forest lands, residential buildings from other constructions and so on. Traditional techniques used in classifying objects in Remote Sensing images are based on Supervised learning. Popular among them are Support Vector Machines (SVM) and Random Forests (RF). But Unsupervised Learning helps to learn features from large amount of data and proves to be a boon in the study of satellite images. Among the various learning models, Deep Belief Network (DBN) [4] was the first one to produce results better than the traditional models. They are also credited with creation of the SAT4 and SAT6 datasets which is used now almost universally. They used a combination of supervised and unsupervised learning algorithms. Various statistical and satellite specific features of the input image are extracted and normalized before feeding them to the DBN model. Specifically, the image was passed through a Restricted Boltzman Machine and a Contrastive Divergence Algorithm. It was seen that with addition of more neural network layers, the accuracy decreased. Even the CNN used by them is no more than 6 layers. The accuracy obtained on SAT4 reached 87% compared to 82% of DBN. To further improve the accuracy, it was proposed to use the traditional features used previously in satelite image processing. Examples are the statistical mean, various central moments, entropy and such other 50 descriptors. They were ranked using Distribution Separability Criterion and the best 22 were selected. They were normalized to be in the range of [0,1] and fed to the DBN. The accuracy on SAT4 improved to 98% which was state-of-the-art. Later Ma [7] used a form of CNN derived from GoogleNet's inception module. It was found to have a large set of hyper-parameters to choose from. A genetic algorithm was found to be useful in getting the best hyper-parameters. Also they used data augmentation to increase the size of the dataset. The model with optimized hyper-parameters reached accuracy of 98.4%. In this design the absence of Satellite image related features is to be noted.

One of the most widely cited papers in this field [8] introduces a new dataset called NWPU-RESISC45 with 45 land cover classes. A number of standard CNN models were applied and the resulting accuracies were in the range of 75-85 %.

Zhong et. al [9] proposes SatCNN, a CNN model consisting of convolutional layers followed by fully connected layers on SAT datasets. The experiment is performed on GPU machine to accelerate the process and obtains an accuracy of 99.65 % on SAT4 and 99.54 % on SAT6 within 40 minutes.

Gong et. al. [10] uses a modified version of Siamese Network [11], called D-DSML which helps in achieving good accuracy by decreasing the redundancy between the information gained through the batch training of high resolution images. It also takes into consideration all the possible combinations of pairs of images which are neglected in Siamese Networks.

Liu [12] uses the features obtained by traditional methods [4]. Instead of passing the features to a DBN , they are concatenated in the FCN layer of a CNN. It gives greatly improved accuracy of 99.90 % on SAT4 and 99.84 % on SAT6. In [13], a CNN model is proposed to perform land cover classification in Indonesia using images from Sentinel-2. It reports an accuracy of 98 % considering various parameters like hue, texture, shape.

But our proposed model uses 13 convolutional layers which helps in identifying more and more useful information without using any satellite specific features, thereby gearing up the training accuracy upto 99.84 % for SAT4 and 99.47 % on SAT6 datasets.

III. Datasets

Our experimental data includes SAT4 and SAT6 datasets. SAT4 has 500,000 images with four types of landscapes. SAT6 has 6 classes of landscapes with 405,000 images. The different landscapes include trees, water bodies, agricultural fields, rural areas, urban areas, etc. These images of 1 m resolution were taken from a US database. Each image is sized to 28×28 pixels covering an area of around 28×28 m. The size of images are selected to be of 28 m as it represents the size of a typical land-holding or an urban building. These images have four channels namely, Red, Green, Blue and Near InfraRed (NIR).

IV. Our Proposed Model

Our proposed architecture is almost similar to original SimpleNet architecture which consists of 13 layers with each convolution layer having 3×3 kernel except the 11^{th} and 12^{th} layers which have 1×1 kernel. The 1×1 kernel is not used for early layers as it avoids local information in the input although it increases the non-linearity of the model. The 2×2 kernels are used for pooling operation. Our model differs from SimpleNet with number of kernels used in each layer as illustrated in Fig. **??**. The kernels of the first hidden layer compute huge feature information from the input. So the feature map of the first hidden layer is richer in information compared to the input satellite image which consists of R,G,B and NIR bands with pixel values ranging from (0, 0, 0, 0) to (255, 255, 255, 255). As the given layer feeds directly from the

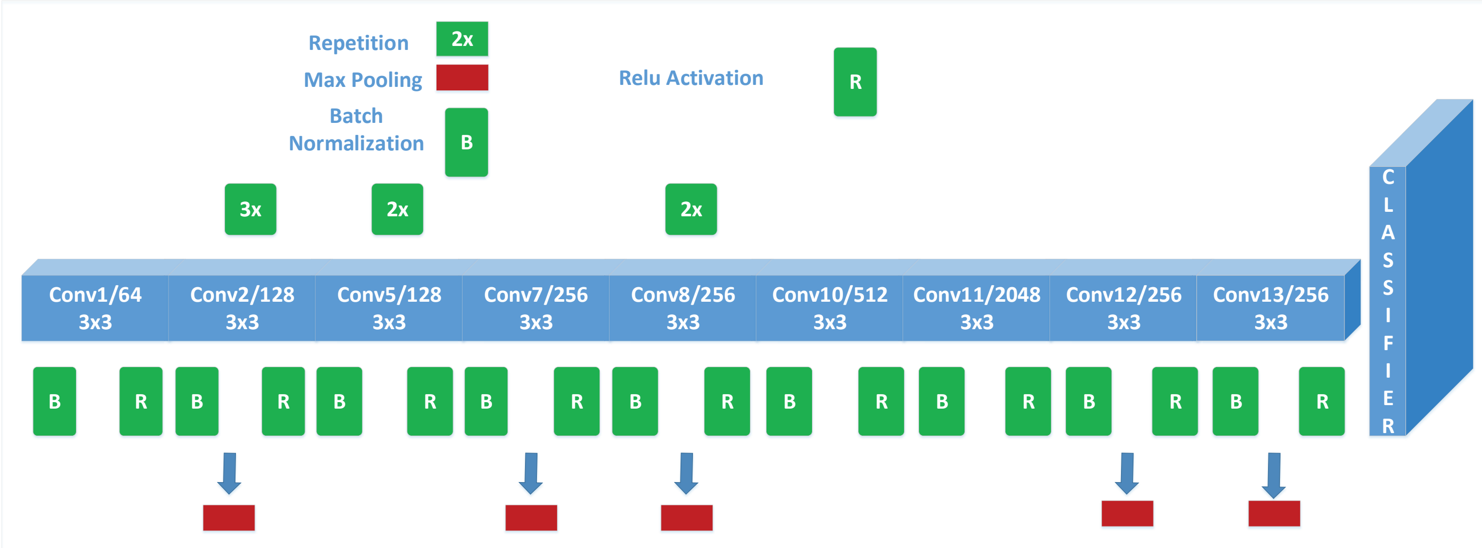

Fig. 1: Modified Simplenet architecure

previous layer to form new kernels by combination of features of the previous layer, the richness of information increases per layer. To perfectly extract the increasingly richer and richer information from the previous hidden layers, the number of kernels used in each layer has been increased. Due to the use of activation function in multiple layers, the loss function converges to zero which makes the network impossible to train. To overcome this problem, Batch Normalisation has been used. The experimental dataset has compressed images and it may consist of lots of noise. Hence, the model learns noisy information along with the detail which impacts the performance of the model negatively and this is called over-fitting. The Dropout technique has been used to deal with this over-fitting problem, with probability 0.2 such that upper layers can be more informative. The results show that our model outperforms other models in less number of epochs.

V. Results

The experiment has been carried out on a NVIDIA based GPU machine for GPU acceleration with python 3.6 and tensorflow framework. Testing and Training accuracy and loss of the model is calculated to determine the performance of the proposed architecture and the results have been compared with the related works.

Fig. 2 shows the variation of training and testing (classification) accuracy with each epoch for SAT4 Dataset. It can be inferred that after training for 15 numbers of epochs with batch size of 50, our model achieves an accuracy of 99.84 % which is better than the previous results. Fig. 3 shows that the loss gradually decreases as the number of epochs increases but small rise in loss function at some points may be due to noise in the dataset.

Fig. 4 shows the variation in testing and training accuracy for SAT6 dataset. Our model is trained for 30 epochs with

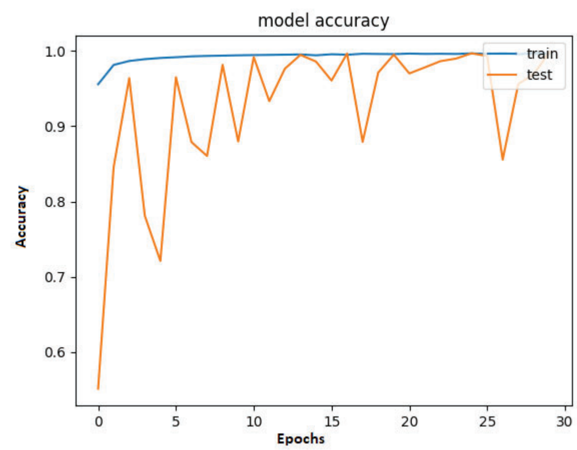

Fig. 2: Accuracy on SAT 4 dataset

batch size of 50 and the highest accuracy achieved is 99.47%. From Fig. 5, it can be observed that with increase in number of epochs, loss approaches to zero.

Table I shows the comparison of training accuracy of our proposed SimpleNet model with related works.

From the results, it can be concluded that our proposed model has been able to achieve an accuracy of 99.84% and 99.47% on SAT 4 and SAT 6 datasets respectively in just 30 epochs.

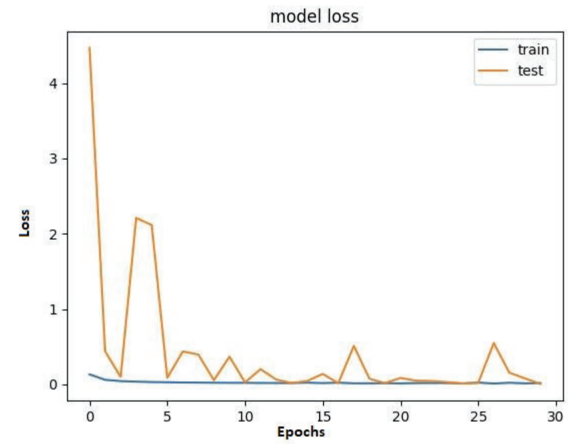

Fig. 3: Loss on SAT 4 dataset

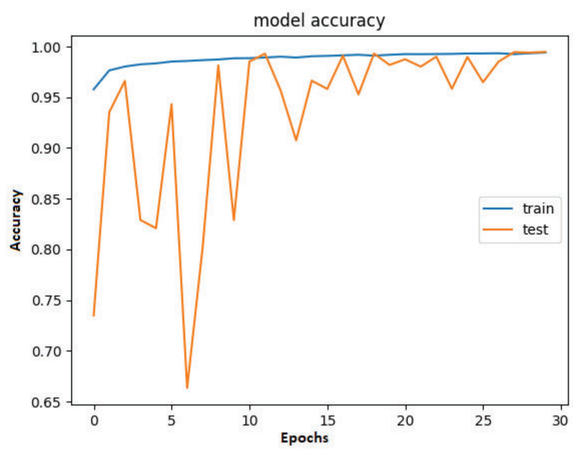

Fig. 4: Accuracy on SAT 6 dataset

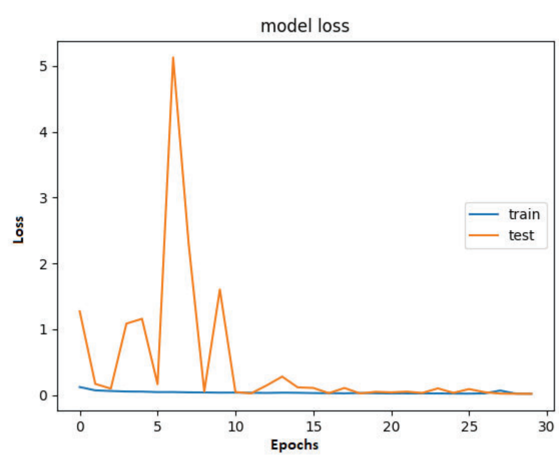

Fig. 5: Loss on SAT 6 dataset

TABLE I: Comparison of Accuracy achieved with number of Epochs on SAT4 and SAT6 Datasets

Model	**SAT 4** (Accuracy %)	**SAT 6** (Accuracy %)	**Epochs**	**Satellite image-specific Features**
DBN [4]	81.78	76.47	300	Yes
SDAE [4]	78.98	78.43	300	Yes
CNN [4]	86.13	79.10	300	Yes
DeepSat [4]	97.95	93.92	–	Yes
DCNN [7]	98.41	96.04	–	Yes
SatCNN [9]	99.69	99.61	–	No
D-DSML [10]	99.51	99.58	–	Yes
Siamese Network Modified [12]	99.76	99.71	–	Yes
Modified SimpleNet (Our proposed model)	99.84	99.47	30	No

VI. Conclusion

This paper proposes a CNN architecture for extracting scene information from Satellite images. Most other models proposed in the literature use a light CNN model and augment it with features specific to satellite images. Such domain knowledge is not generally available with the AI community. The model is designed on lines of VGG but has fewer number of parameters. Also, the techniques of Batch Normalization and Dropout has helped our model to outperform all the other architectures with an accuracy of 99.84% and 99.47% on SAT4 and SAT6 satellite image datasets respectively. The amount of time required for training and testing is just 30 epochs which is very less as compared to others. Another advantage is that the images need not be pre-processed and hence, it can be useful to batch-process large number of images. It is planned to use this model to process the data of entire state and ascertain its land resources.

References

[1] Krizhevsky, Alex, I. Sutskever, and G. E. Hinton, "ImageNet classification with Deep Convolutional Neural Networks," *Advances in Neural Information Processing Systems*, pp. 1097–1105, 2012.

[2] M. J. Canty, *Image Analysis, Classification and Change Detection in Remote Sensing with Algorithms for ENVI/IDL and Python*. CRC Press, 3 ed., 2014.

[3] S. H. Hasanpour, M. Rouhani, M. Fayyaz, and M. Sabokrou, "Lets keep it simple, Using Simple Architectures to outperform deeper and more complex Architectures," *Computer Vision and Pattern Recognition*, vol. arXiv:1608.06037, February 2018.

[4] S. Basu, S. Ganguly, S. Mukhopadhyay, R. DiBiano, M. Karki, and R. Nemani, "Deepsat-A Learning framework for Satellite Imagery," *arXiv preprint*, vol. arXiv:1509.03602, September 2015.

[5] T. Blaschke and J. Strobl, "What's wrong with pixels? Some recent developments interfacing remote sensing and GIS," *GeoBIT/GIS*, vol. 6, no. 1, pp. 12–17, 2001.

[6] T. Blaschke, "Object based Image Analysis for Remote Sensing," *ISPRS J. Photogramm. Remote Sens*, vol. 65, no. 1, pp. 2–16, 2010.

[7] C. L. Zhong Ma, Zhuping Wang and X. Liu, "Satellite Imagery Classification based on Deep Convolution Network," *International Journal of Computer and Information Engineering*, vol. 10, no. 6, pp. 1156–1159, 2016.

[8] G. Cheng and J. Han, "Remote Sensing Image scene Classification: Benchmark and State-of-the-art," *Proceedings of the IEEE*, vol. DOI:10.1109/JPROC.2017.2675998, 2017.

[9] Y. Zhong, F. Fei, Y. Liu, B. Zhao, H. Jiao, and L. Zhang, "SatCNN: Satellite Image Dataset Classification using Agile Convolutional Neural Networks," *Remote Sensing Letters*, vol. 8, no. 2, p. 136–145, 2017.
[10] Z. Gong, P. Zhong, Y. Yu, and W. Hu, "Diversity-Promoting Deep Structural Metric Learning for Remote Sensing Scene Classification," *IEEE Transactions on Geoscience and Remote Sensing*, January 2018.
[11] J. Wang, "Learning fine-grained image similarity with Deep Ranking," *IEEE Conf. on Computer Visual Pattern Recognition*, June 2014.
[12] Q. Liu, S. Basu, S. Ganguly, S. Mukhopadhyay, R. DiBiano, M. Karki, and R. Nemani, "DeepSat V2: Feature Augmented Convolutional Neural Nets for Satellite Image Classification," *Remote Sensing Letters*, Nov 2019.
[13] Y. Heryadi and E. Miranda, "Land Cover Classification Based on Sentinel-2 Satellite Imagery Using Convolutional Neural Network Model: A case study in Semarang Area, Indonesia," *Asian Conference on Intelligent information and Database Systems (ACIIDS)*, pp. 191–206, 2019.