Land Classification using Convolutional Neural Networks

Anees Fatima Khan, Bhavya P, R. Ravinder Reddy

Abstract**: *Identifying the physical aspect of the earth's surface (Land cover) and also how we exploit the land (Land use) is a challenging problem in environment monitoring and much of other subdomains. One of the most efficient ways to do this is through Remote Sensing (analyzing satellite images). For such classification using satellite images, there exist many algorithms and methods, but they have several problems associated with them, such as improper feature extraction, poor efficiency, etc. Problems associated with established land-use classification methods can be solved by using various optimization techniques with the Convolutional neural networks(CNN). The structure of the Convolutional neural network model is modified to improve the classification performance, and the overfitting phenomenon that may occur during training is avoided by optimizing the training algorithm. This work mainly focuses on classifying land types such as forest lands, bare lands, residential buildings, Rivers, Highways, cultivated lands, etc. The outcome of this work can be further processed for monitoring in various domains.***

Keywords: Convolution Neural Networks(CNN), Deep Learning, Land Classification

I. INTRODUCTION

Land-cover and land-use change (LUCC), that is associated with imbalances in ecosystems, biodiversity, and global climate changes, show the consequences of human activities and climatic changes on the ecological atmosphere of the Earth's surface. The international community gives more significance to LUCC because of the key content of worldwide environmental change research. Since the 1990s, various research institutions such as Food and Agriculture Organization(FAO), International Geosphere-Biosphere Project(IGBP), International Institute for Applied Systems Analysis(IIASA), etc. have launched several LUCC related projects [15]. Remote sensing is a potential tool for keeping track of the Earth's surface and also a fundamental element for works that use classification and identification technologies to look over the land-use status [16].

While performing field surveys is more comprehensive and authoritative, it's an upscale task and mostly takes an extended time to update. It is made easy with remote sensing. Land use or land cover change detection is very important for a greater understanding of topography dynamics.

Land use and land cover changes could also be an omnipresent and accelerating process, primarily driven by natural circumstances and anthropogenic undertakings, that successively reflects changes that may affect natural ecosystems. Remote sensing images assist us in discovering solutions to the increasing number of environmental challenges that we face today. It helps not only in getting a bird's eye view of what's around but also in discovering parts of the world that are hardly seen. The potential of categorizing land use and land cover allows humans to more competently utilize natural resources and thereby decreasing waste and deprivation. In spite of its potential to be incredibly useful, satellite data is very huge and confusing, and making sense out of it requires complex analysis.

Understanding land use/land cover has become more and more vital because the country plans to overcome the issues of hazard, unrestrained development, loss of superior agricultural lands, worsening environmental standards, demolition of important wetlands, and so forth. If standards and living conditions are to be upgraded or maintained at modern-day ranges, land use records are needed in the analysis of environmental activities.

Classification of hyperspectral and multispectral data has become crucial in detecting land-use change. Upgrading classification techniques using extensively available satellite data has largely hindered in recent years though there have been many approaches for land classification. The problems related to these customary land classification algorithms are a speedy increase in the dimensionality of satellite data, poor running efficiency, and inefficient feature extraction, etc. To surpass these, various strategies are being developed. Information ashore land use or land cover and opportunities for their appropriate use is essential for the selection, planning, and application of land use schemes to meet the growing needs of primary human needs and welfare. This knowledge also helps in tracking the dynamics of land use on account of the growing needs of the increasing population.

II. LITERATURE REVIEW

Land cover and Land use are two separate terms that we often use interchangeably. Land cover refers back to the bodily traits of the earth's floor, captured within the distribution of flora, soil, water, and different physical characteristics of the land, together with the ones created completely via human activities. Utilization of land by people and their habitat refers to land-use, usually with the emphasis at the utilitarian position of land for economic utilizations eg. construction of buildings, agricultural lands, etc.

Revised Manuscript Received on June 22, 2020.

Anees Fatima Khan, Student, Department of Computer Science and Engineering, Chaitanya Bharathi Institute of Technology, Hyderabad, Telangana, India.

Bhavya P, Student, Department of Computer Science and Engineering, Chaitanya Bharathi Institute of Technology, Hyderabad, Telangana, India.

R. Ravinder Reddy, Associate Professor, Department of Computer Science and Engineering, Chaitanya Bharathi Institute of Technology, Hyderabad, Telangana, India.

Land use or land cover mapping has been a beneficial and specific way to improve the selection of regions designed to agricultural, urban, and business regions of a community with the discovery of remote sensing and Geographical Information System(GIS) approaches.

Applications of satellite data made it possible to study the adjustments in land cover at low cost, in less time, and with better accuracy in integration with GIS that acts as an effective platform for records evaluation or analysis, replacement, and retrieval. Several types of land-use classification algorithms or standards exist, which include several classes and are reasonable for the complex features of land-use/and-cover types, and these traits pose complexities for precise classification. Determining the classification method, followed by selecting the precise classifier is essential in the classification of remote sensing images. Classification methods consist of supervised/unsupervised algorithms, the authentic use of feature extraction from spectral information and, hard or soft classification. Due to differences in the basic unit of classification, classification strategies can be divided into pixel-based and object-oriented classifications[6]. Regarding classifier selection, the conventional technique used is the statistical procedure for low-level feature extraction, along with distance[7], K-nearest neighbor[8], maximum likelihood[9], and logistic regression[10] classifiers. High-resolution remote sensing (HRRS) images are applied in land-use classification with the fast evolution of aerospace, computer, and sensor technologies[13]. The range of objects inside a given class expands as does the similarity of objects in various classes because of spectral confusion in these high-resolution remote sensing images. These properties decrease the effectiveness of conventional classification techniques based on low-level features/attributes. Therefore, the method built on modeling the mid-level features has evolved based on the low-level feature technique[14]. The bag-of-visual-words (BoVW) is one of the three types of mid-level feature extraction methods that describe image features. The performance of BoVW-based procedures relies on the extraction of features in practical applications. The other two mid-level feature extraction techniques are latent Dirichlet allocation(LDA) and machine learning model. The machine learning models unaided perform data analysis and feature extraction and remove the pattern of the extracted features in line with pre-established rules. Thus, these models are able to obtain enhanced classification results when put in with the complex images. Sparse coding, support vector machines, neural networks, and deep learning are commonly used machine learning methods[11]. The deep learning networks are made of various non-linear layers, which constitute a state-of-the-art method of intelligent pattern recognition and are critical new conduct in the province of remote sensing image processing.

Research on applying Convolution neural networks to remote sensing images has surfaced over recent years. The Hinton team successfully topped the ImageNet image classification contest and minimized the error rate(top-5) of 1000 images from 26.2 to 15.3%[12]. Hu et al was the first to use a Convolution neural network model to classify HRRS images[5]. Chen et al integrated a CNN classification technique that utilizes spatial information, pixel spectral information, and explored the significance of spatial information in the classification of High-resolution remote sensing(HRRS) images[17].

III. METHODOLOGY

In this work, the implementation is carried out as steps mentioned below which are also depicted in Fig. 1.

A. Data Collection and Data Preparation
B. Simple CNN model
C. Improvised CNN
D. Using Transfer Learning

These steps are explained in the following subsections.

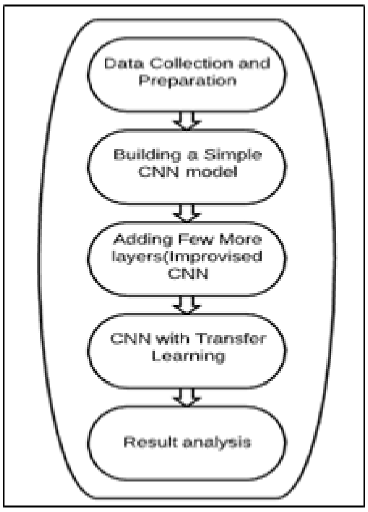

Fig. 1.Process flow of the proposed solution

A. Data Collection and Data Preparation

For this work, we have used open-source EuroSAT Sentinel-2 satellite images from the German Research Center for Artificial Intelligence, which can be downloaded locally [link]. The dataset which we have considered has about 27,000 labeled images. And these images are of 10 different land use classes, which are Highway, Forest, Herbaceous Vegetation, Annual Crop, Industrial, etc. Each multispectral image consists of 13 different color bands that represent different wavelengths of light or color and different resolutions. These different light bands help distinguish parts of the landscape that reflect certain sorts of light, especially ways.

Preprocessing of satellite images involves radiometric calibration and geometric correction. The motive of preprocessing is to convey the information contained in the activity of image synthesis further to furnish it in the neighborhood to the actual image. The elimination of image contortion caused by radiation errors is voided by radiation correction. Geometric correction ensures that the total error of the corrected position is no more than one pixel. And also images are translated from 'tif' form to 'jpeg' form as required using rasterio. And it is very important to split the dataset into train, tests, and valid folders.

B. Simple CNN Model

The first convolutional neural network explored in this work was a simple CNN with a three-layer architecture: one convolutional input layer, one flatten layer in the middle and one dense output layer as shown in Fig. 2.

The simplicity of the first model is a relatively fast test that the data preparation was sufficient and that the neural network was set up with the right parameters. However, this model has a very low accuracy of around 30% as shown in Table-III. Since satellite data is pretty complex this model will require more layers for it to learn the distinct features of each category of the dataset. So we have improvised the CNN model by adding a few more layers in the next step.

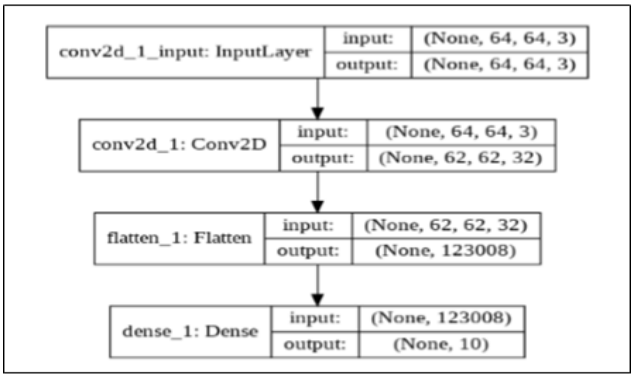

Fig. 2.Working of a simple CNN model used

C. Improvised CNN

In this network, we have added a few more layers as shown in Table-I to learn the distinct features. Each image will be fed into the CNN model. The first layer, with a 5x5 kernel size, is a Convolutional 2D layer. The results of convolution are feature maps with size 28x28. After that, the Max Pooling layer is performed as subsampling with kernel 2x2 (non-overlapping). The results of the max-pooling layer are feature maps with size 14x14 (because of non-overlapping so stride or movement of the kernel is 2). The convolution layer and Max Pooling layer are then performed. Each output in the convolutional layer is often activated by using ReLU. The dropout technique is incorporated here as a regularization technique. Using this CNN model the accuracy is successfully improved to 90%. But to improve the accuracy furthermore, transfer learning with CNN is used in the next step.

Table-I: Summary of Improvised CNN model

Layer (type)	Output Shape	Param #
conv2d_2 (Conv2D)	(None, 62, 62, 32)	896
max_pooling2d_2 (MaxPooling2	(None, 31, 31, 32)	0
conv2d_3 (Conv2D)	(None, 29, 29, 32)	9248
max_pooling2d_3 (MaxPooling2	(None, 14, 14, 32)	0
dropout_1 (Dropout)	(None, 14, 14, 32)	0
flatten_1 (Flatten)	(None, 6272)	0
dense (Dense)	(None, 128)	802944
dropout_2 (Dropout)	(None, 128)	0
dense_1 (Dense)	(None, 10)	1290

D. CNN with Transfer Learning(VGG16 Model)

CNN with transfer learning is used to improve the performance on the dataset. Transfer learning let's one adapt a large pre-trained model to their work. In this case, the ImageNet VGG16 model was used. The VGG16 model is a CNN that was trained on 15 million images in over 22 thousand categories. While the first several layers of the model will be kept in their pre-trained state, we will train the last set of layers to learn the specific features of this dataset, which only has a few thousand images and ten categories. The VGG16 model architecture is an input layer followed by a pattern of two or three convolutional layers and a max-pooling layer that repeats five times as shown in Table-II. It changed the model to match the input size of the data, as well as the output size of ten categories freezing the first 12 layers so that only the final 10 would be trained. Various techniques such as dropout, data augmentation, etc. have been used to improve the efficiency of the model. Using this model, an accuracy of 95% (as shown in Table-III) is achieved successfully.

Table-II: The architecture of the VGG16 model.

Layer (type)	Output Shape	Param #
block1_conv1 (Conv2D)	(None, 64, 64, 64)	1792
block1_conv2 (Conv2D)	(None, 64, 64, 64)	36928
block1_pool (MaxPooling2D)	(None, 32, 32, 64)	0
block2_conv1 (Conv2D)	(None, 32, 32, 128)	73856
block2_conv2 (Conv2D)	(None, 32, 32, 128)	147584
block2_pool (MaxPooling2D)	(None, 16, 16, 128)	0
block3_conv1 (Conv2D)	(None, 16, 16, 256)	295168
block3_conv2 (Conv2D)	(None, 16, 16, 256)	590080
block3_conv3 (Conv2D)	(None, 16, 16, 256)	590080
block3_pool (MaxPooling2D)	(None, 8, 8, 256)	0
block4_conv1 (Conv2D)	(None, 8, 8, 512)	1180160
block4_conv2 (Conv2D)	(None, 8, 8, 512)	2359808
block4_conv3 (Conv2D)	(None, 8, 8, 512)	2359808
block4_pool (MaxPooling2D)	(None, 4, 4, 512)	0
block5_conv1 (Conv2D)	(None, 4, 4, 512)	2359808
block5_conv2 (Conv2D)	(None, 4, 4, 512)	2359808
block5_conv3 (Conv2D)	(None, 4, 4, 512)	2359808

IV. RESULTS AND DISCUSSIONS

After the data preparation, the first convolutional neural network(CNN) explored was a simple CNN with a three-layer architecture. The model was compiled using Adam optimizer, which is a stochastic optimization function that's computationally efficient. The accuracy of the model during training stayed constant at 30% for almost all training epochs. The confusion matrix of the model is evaluated and it turned out that this model is way too simple. This model must not have learned much, because it guessed every test image as the same category. No wonder the accuracy was around 30%, which is what one would expect if they were randomly guessing for a dataset with 10 categories.

Table-III: Result analysis of CNN models mentioned

Model	Accuracy	No of epochs	Time Taken
Simple 3 layered CNN	30%	10	< 1 hour
Improvised CNN	90%	100	< 9 hours
CNN with Transfer Learning	95%	10	more than a day

Since the simple model was not very good, the next step was to use an improvised CNN in which we personalized the architecture to improve the performance of the CNN model by adding a few more layers. Using this CNN model, accuracy of 90% is achieved as shown in Table-III.

To improve the accuracy of the classification, the next step was to use transfer learning to improve the performance of this model on the dataset. In this case, the ImageNet VGG16 model was used. Data Augmentation and Dropout regularization are incorporated to boost the effectiveness of the model. Data augmentation increases the size of the dataset by creating multiple versions of the same image that are slightly transformed. Even though this model took very long to run for ten epochs, it out-performed the simple CNN by an impressive amount. This model achieved 95% accuracy, guessing most categories in the test set correctly. We then experimented with the Batch Normalization technique in this model. Batch normalization improves the speed and performance of a neural network by normalizing the data to a standard scale in the model's learning process. With ten epochs of training, the tuned model achieved a 56% accuracy. Batch normalization is known to speed up model training, allowing for larger learning rates and fewer epochs. With Batch Normalization, the model took less time to compile than before. Despite the difference in running time, the tuned model's accuracy is continuing to improve up until the last epoch, and would likely continue improving if given more epochs to train. For the purpose of this work, all model parameters are kept the same, but it might have tuned the model much faster, and reap the benefits of batch normalization if the learning rate is increased to something larger than 0.0001.

V. CONCLUSION AND FUTURE WORK

In this work, the first simple model is explored just as a relatively fast test and as it has very few layers, it has failed to learn the features. To better the efficiency of the CNN model or to learn the features more effectively, few more layers are added to simple CNN which is improvised CNN. It successfully improved the classification accuracy, but to further improve the accuracy, CNN with Transfer learning is explored. Some optimization techniques such as data augmentation, dropout regularization, etc. have been used.

This model which is CNN with Transfer Learning achieved the highest accuracy. Since this work addresses an environmental question and most environmental entities are in the public, nonprofit, or university sectors, resources and realism is a very important consideration. Even for small startups that have limited funding and need to prove their worth before the next funding round, taking advantage of limited resources is a big factor in deciding which models to use in production.

Changing the learning rate while using Batch Normalization might have helped to get the highest accuracy in very little time. So it would be best to explore this optimization technique. Another aspect of this work to explore would be the base model used in transfer learning. In this work, the ImageNet VGG16 model is used, but there are many others to choose from. Another popular model for transfer learning is MobileNet, which was built to work well on web browsers and on mobile devices. MobileNet models take up much less memory and have fewer parameters then ImageNet models, making them more flexible for use on computers, tablets, and smartphones. In exchange for this flexibility, MobileNet is not as accurate as of its larger counterparts but still performs rather well. Using a MobileNet model as the base for this experiment would be useful if the model were to be used on a smartphone for real-time land use classification using the phone's camera, for example.

REFERENCES

1. Baoxuan Jin., Peng Ye., Xueying Zhang., Weiwei Song., Shihua Li(2019).. Object-Oriented Method Combined with Deep Convolutional Neural Networks for Land-Use-Type Classification of Remote Sensing Images.
2. Zhang, Ce. (2018). Deep Learning for Land Cover and Land Use Classification. 10.17635/lancaster/thesis/428.
3. Jin, Y. T., Yang, X. F., Gao, T., Guo, H. M., & Liu, S. M. (2018). The typical object extraction method based on object-oriented and deep learning. Remote Sensing for Land and Resources, 30(1), 22–29.
4. Al Khawlani, M. M., Elmogy, M., & Elbakry, H. M. (2015). Content-based image retrieval using local features descriptors and bag-of-visual words. International Journal of Advanced Computer Science & Applications, 6(9), 212–219.
5. Hu, F., Xia, G. S., Hu, J. W., & Zhang, L. P. (2015). Transferring deep convolutional neural networks for the scene classification of high-resolution remote sensing imagery. Remote Sensing, 7(11), 14680–14707.
6. Zheng, Y., Wu, F. D., & Liu, Y. F. (2010). A feature analysis approach for object-oriented classification. Geography and Geo-Information Science, 26(2), 19–22.
7. Tzeng, Y. C. (2006). Remote sensing images classification/data fusion using distance weighted multiple classifiers systems. Journal of Chinese Institute of Engineers, 31(4), 639–647.
8. Meng, Q., Cieszewski, C. J., Madden, M., & Borders, B. E. (2007). K nearest neighbor method for forest inventory using remote sensing data. GI Science & Remote Sensing, 44(2), 149–165.
9. Bruzzone, L., & Prieto, D. F. (2001). Unsupervised retraining of a maximum likelihood classifier for the analysis of multitemporal remote sensing images. IEEE Transactions on Geoscience and Remote Sensing, 39(2), 456–460.
10. Lee, S. (2005). Application of logistic regression model and its validation for landslide susceptibility mapping using GIS and remote sensing data. International Journal of Remote Sensing, 26(7), 1477–1491.
11. Zhang, L., Zhang, L., & Du, B. (2016). Deep learning for remote sensing data: A technical tutorial on the state of the art. IEEE Geoscience and Remote Sensing Magazine, 4(2), 22–40.

12. Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2017). ImageNet classification with deep convolutional neural networks. Communications of the ACM, 60(6), 84–90.
13. Hu, F., Xia, G. S., Hu, J. W., & Zhang, L. P. (2015). Transferring deep convolutional neural networks for the scene classification of high-resolution remote sensing imagery. Remote Sensing, 7(11), 14680–14707.
14. Bosch, A., Muñoz, X., & Martí, R. (2007). Which is the best way to organize/classify images by content? Image and Vision Computing, 25(6), 778–791.
15. Sands, R. D., & Leimbach, M. (2003). Modeling agriculture and land use in an integrated assessment framework. Climatic Change, 56(1–2), 185–210.
16. Song, X., Duan, Z., & Jiang, X. (2012). Comparison of artificial neural networks and support vector machine classifiers for land cover classification in Northern China using a SPOT-5 HRG image. International Journal of Remote Sensing, 33(10), 3301–3320.
17. Chen, Y. S., Jiang, H. L., Li, C. Y., Jia, X. P., & Chamisi, P. (2016). Deep feature extraction and classification of hyperspectral images based on convolutional neural networks. IEEE Transactions on Geoscience and Remote Sensing, 54(10), 6232–6251.

AUTHORS PROFILE

Anees Fatima Khan is a final year student studying Computer Science and Engineering in Chaitanya Bharathi Institute of Technology. Her interests include Data Analytics, Machine Learning and VR Software.

Bhavya P is a final year student studying Computer Science and Engineering in Chaitanya Bharathi Institute of Technology. Her interests include Machine Learning and Deep Learning

R. Ravinder Reddy is working as Associate Professor in Chaitanya Bharathi Institute of Technology with qualifications in Ph. D. (CSE), M. Tech(CSE). His interests include Computer Science and Deep Learning