

## Review on Deep Learning in Remote Sensing Image Classification

Preeti Rajkumar Komati<sup>1</sup> Rajkumar Damodar Komati<sup>2</sup>

### **Abstract:**

*Remote sensing image scene classification has a wide scope of applications and hence has been receiving remarkable attention. Nowadays, Deep learning (DL) algorithms have seen a massive rise in popularity for remote-sensing image analysis over the past few years. This paper first provides a comprehensive review of the recent progress in remote sensing applications. The existing remote-sensing classification methods are categorized into four main categories according to the features they use: manually feature-based methods, unsupervised feature learning methods, supervised feature learning methods, and object based methods[15]. This article then focuses on evaluating the available and public remote-sensing datasets. Finally, a conclusion regarding the current state of art methods and directions for future research are presented.*

**Keywords** – Remote sensing, deep learning DL models

### **I. INTRODUCTION**

The analysis and classification of remote-sensing images is very important in many practical applications, such as natural hazards and geospatial object detection, precision agriculture, urban planning, vegetation mapping, and military monitoring [1]. The research analysis for remote-sensing images analysis still remains low. The main objective of this paper is to present a literature review on the recent deep learning based techniques for remote sensing and the available training and testing datasets.

The remote-sensing methods are developed for improving the performance aspects of image, such as preprocessing, segmentation, and classification. Neural networks, the basis of deep learning (DL) algorithms, have been used in the remote sensing for many years. However, prior to the development of DL, the focus from neural networks was shifted to support vector machine (SVM) and ensemble classifiers, e.g. random forest (RF), for image classification. SVM can handle high dimensionality data and perform well with limited training samples, among other things [2], while RF gained popularity and ease of use (e.g. relatively insensitive to classification parameters) and high accuracy [3]. In more recent years, however, the advent of DL has led to renewed interest in neural networks.

In this paper presents review on four categories of remote sensing classification methods: manually feature based method, unsupervised feature learning methods, supervised feature learning methods, and object based methods. In supervised methods, the DL models for CNN (convolutional neural network) and RNN (recurrent neural network) models are explained. In unsupervised methods, the DL models for AE (Auto encoders), DBN (Deep Belief Network), and GAN (Generative adversarial network) are explained.

This paper also presents a literature review on the publicly available remote sensing images datasets.

### **II. IMAGE CLASSIFICATION METHODS FOR REMOTE SENSING**

The existing remote sensing image classification methods could be summarized into four main categories according to the features they use: 1) manually feature based methods, 2) unsupervised classification methods, 3) supervised learning methods and 4) object-based methods.

*1. Manually Feature Based Methods:* The early works for scene classification are mainly based on manually handcrafted features. These methods mainly focus on using a considerable amount of engineering skills and domain expertise to design various human engineering features, such as color, texture, shape, spatial and spectral information, or their combination that are the primary characteristic of a scene image and hence carry useful information used for scene classification [5]. Some of the most common and essential features that are used for scene classification are: 1) Color histograms - Texture

descriptors, 2) GIST: describe orientations of a scene, 3) SIFT: describe sub-regions of a scene, 4) HOG: describe gradient of objects. [5].

*2 Unsupervised Classification Methods:* The limitations of manually feature based methods could be overcome by self-learning features from images, called unsupervised learning method. The alternative to manually feature based method is unsupervised feature learning from unlabeled input data. In this method first the image pixels are grouped into clusters based on their properties. By learning features from images instead of relying on manually designed features, we can obtain more discriminative feature that is better suited for the classification problem [7]. Such clustering algorithms are: principal component analysis (PCA), k-means clustering, sparse coding, and so on. The commonly used DL models for unsupervised image classification are AE (Auto encoders), DBN (Deep Belief Network), and GAN (Generative adversarial network).

*2.1 AE (Auto encoders):* Auto-Encoders consists of three layers viz inputs layer, output layer and hidden layer. In this method, mapping of the input layer to the hidden layer takes place first and produces the output layer. This process is done by encoder. After this, mapping of hidden layer to output layer is done by decoder. A decoder produces the output layer having the same number of nodes as the input layer [8]. Normally, autoencoders (AEs) are designed to learn a compressed and distributed data set representation. The number of hidden units in one hidden layer is smaller compared with the input or the output, and this is the most important feature of AE. Therefore, an AE can accomplish the purpose of data compression and dimensionality reduction through one hidden layer. Hence, AEs are mostly used for the processing of feature hierarchy [4].

Stacked AE, as the name implies, a stacked AE (SAE) is a neural network consisting of multiple layers of AEs, where the outputs of each layer are wired to the inputs of the following layer. It is formed by stacking AE layers. In the field of remote sensing, such multilayer AEs are usually used for feature representation and have produced good effects particularly in spectral-spatial feature learning [8].

*2.2 Restricted Boltzmann machine (RBM) and Deep Belief Network (DBN):* A Restricted Boltzmann Machine is a unique type of Markov random field which contains one layer of stochastic hidden unit and one layer of stochastic visible or observable unit [9]. In this method all visible units are joined to all hidden units, and visible-visible or hidden-hidden connections are not available [10].

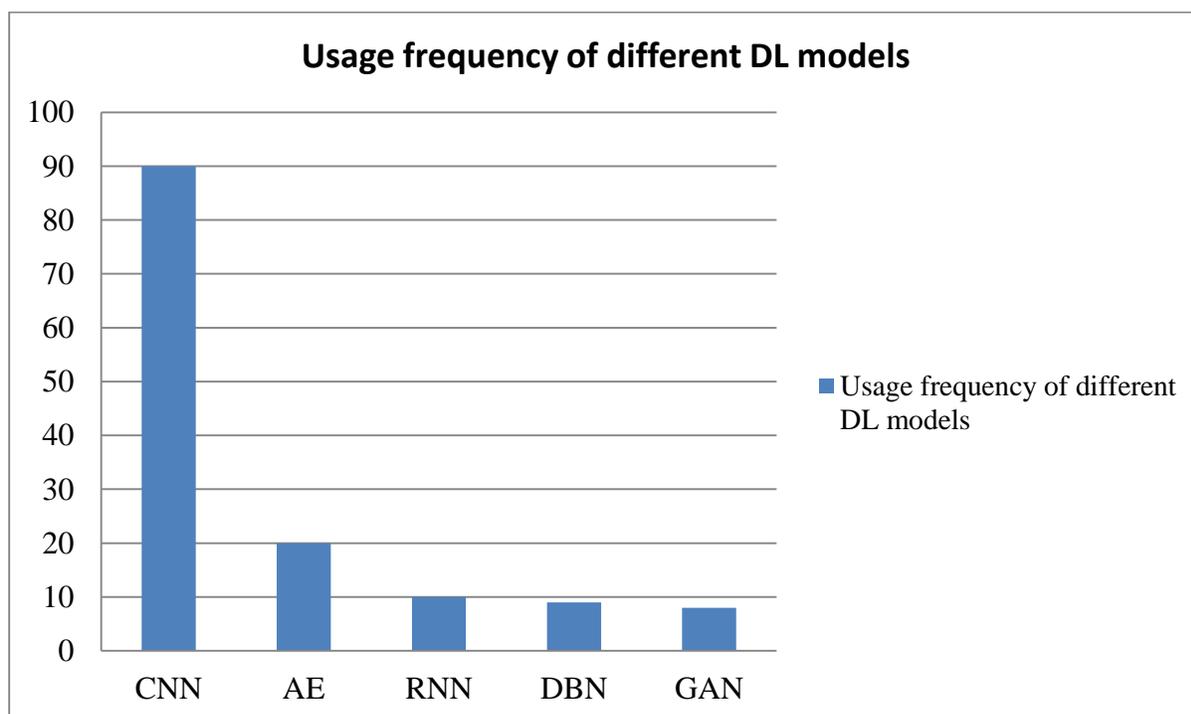
DBNs each consist of multiple layers of RBMs, except that the individual layers in DBNs are trained using the RBM model. A linear classifier is added to the top layer of DBN to implement a supervised optimization. This usually yields good results [8].

*2.3 Generative adversarial networks (GANs):* The GAN contains of two networks contesting with each other: a generative network and discriminative network. The generative network maps the particular data of interest, while the discriminative network discriminates between the real data and the generated data produced by the generative network. The target of training the generative network is to “fool” the discriminative network by producing examples that appear realistic that have the true data distribution [8]. The discriminative network is generally a standard convolutional network to produce probabilities. Both networks try to optimize a different and opposing loss function in a zero-sum game. In the last three years, GANs have been successfully applied in many computer vision and image processing applications [12].

*3. Supervised Classification Methods:* Supervised learning methods use labeled data to extract features. The commonly used DL models for supervised classification are Convolutional Neural Networks (CNNs) and Recurrent Neural Network (RNNs).

**3.1 Convolutional Neural Networks (CNNs) [4]:** CNN, one of the most extensively used DL models, was originally designed to process data in the form of multiple arrays. The convolutional neural network (CNN) is well-suited for processing multiband remote-sensing image data in which pixels are arranged regularly. The CNN consists of three layers : convolution layers, pooling layers, and fully connected layers. At each layer, the input image is convolved with a set of K kernels  $W = \{W_1, W_2, \dots, W_k\}$  and added biases  $\gamma = \{b_1, \dots, b_k\}$ , each generating a new feature map  $X_k$ . These features are subjected to elementwise nonlinear transform  $\sigma(\cdot)$ , and the same process is repeated for every convolutional layer. Compared with traditional MLPs(Multi-Layer Perceptrons), in CNNs the values of pixels within a neighborhood of a certain size are aggregated using a permutation invariant function, typically the max or mean operation. At the end of the convolutional stream of the network, fully connected layers (i.e., regular neural-network layers) are usually added, where weights are no longer shared. Some popular CNN architectures are ALEXNET , VGG NETWORKS , RESNET and also a recent development for GoogleNet called Inception-v4.

**3.2 Recurrent Neural Network:** The RNN model was traditionally used for a discrete sequence analysis. In an RNN, the length of input and output data vary. Therefore, RNN are used for speech and language processing. It is difficult to learn and store information for very long period. Thus to address this problem, an explicit memory is used to augment the networks. Therefore, several specialized memory units are developed, like the long short-term Memory cell and gated recurrent unit. With the development of the architecture and ways of training, RNNs have been successfully and extensively applied in predicting the next character in the text, or the next word in a sequence and have been extended to other more complex tasks of remote-sensing images.



The CNN model has been the most commonly used for remote-sensing image analysis, followed by the AE model[4]. The RNN, DBN, and GAN models, are much less commonly used. The higher popularity of CNN is likely because it has unique characteristics that make it highly suitable for processing multiband remote-sensing image data in which pixels are arranged regularly.

*4 Object-Based Methods* : The object-based image classification groups pixels into representative shapes and sizes. Each group is assigned to a semantic object. This process relies on multi-resolution segmentation. Multi-resolution segmentation produces homogenous objects by grouping pixels and generate objects with different scales in an image simultaneously. These objects are more meaningful because they represent features in the image [13]. When you have high spatial resolution, object-based image classification is superior to traditional pixel-based classification [14].

### III. PROPOSED REMOTE SENSING IMAGE DATASETS

In the past years, several high resolution remote-sensing image datasets have been introduced by different groups to enable research for scene classification and to evaluate different methods in this field [15]. Some publicly available image datasets are given in Table 1. The table below shows the number of scene classes, images per class, size of images, and spatial resolution.

Data set	Scene classes	Image per class	Spatial Resolution	Image size
AID	30	200-400	High	600x600
Patter Net	38	800	Upto 0.8	256x256
RSI-CB-256	35	Varies	0.3-3	256x256
SAT_4 & SAT_6		Patches	Low	28x28
UC-Merged Land Use	21	100	0.3	256x256
WHU-RS19	19	~50	Upto 0.5	600x600
SIRI-WHU	12	200	2	200x200
RSSCN7	7	400	-	400x400
RSC11	11	~100	0.2	512x512
Brazilian Coffee	2	1438	Low	64x64
NWPU-RESISC45	45	700	~30-0.2	256x256

The most images in these datasets are imported from Google Earth Engine and cover the areas of: agricultural, airplane, baseball diamond, beach, buildings, chaparral, dense residential, forest, freeway, golf course, harbor, intersection, medium density residential, mobile home park, overpass, parking lot, river, runway, sparse residential, storage tanks, and so on [15].

### CONCLUSION

For remote sensing image processing, DL models including AEs and CNNs have been considerably successful recently in remote sensing and DL models such as RNNs and GANs are expected to be introduced into this field for further developments. Thus Deep learning methods can offer better feature representations for the related remote-sensing images. There is a scope for researchers dedicated to learning better features in image classification tasks by utilizing appropriate deep learning methods.

### REFERENCES:

- [1] Zhang, L., Xia, G., Wu, T., Lin, L., Tai, X.: Deep learning for remote sensing image understanding. *J. Sens.* 2016,1 –2 (2016)
- [2] Mountrakis, G., Im, J., Ogole, C.: Support vector machines in remote sensing: a review. *ISPRS J. Photogramm. Remote Sens.*66 (3), 247-259. (2011)
- [3] Belgiu, M., Dragut, L.,: Random forest in remote sensing: a review of applications and future directions. *ISPRS J. Photogramm. Remote Sens.*114, 24-31.(2016)
- [4] Lei Ma, Yu Liu, Xueliang Zhang, YuanxinYe, Gaofei Yin, Brian Alan Johnson: Deep learning in remote sensing applications: A meta-analysis and review. *ISPRS Journal of Photogrammetry and Remote Sensing.*152,166-177.(2019)

- [5] Cheng, G., Han, J., Lu, X.: Remote sensing image scene classification: benchmark and state of the art. *Proc. IEEE* 105(10), 1–17 (2017)
- [6] Mekhalfi, M., Melgani, F., Bazi, Y., Alajlan, N.: Land-use classification with compressive sensing multifeature fusion. *IEEE Geosci. Remote Sens.* 12(10), 2155–2159 (2015)
- [7] Cheriyyadat, A.: Unsupervised feature learning for aerial scene classification. *IEEE Trans. Geosci. Remote Sens.* 52(1), 439–451 (2014)
- [8] M. P. Vaishnave, K. Suganya Devi , P. Srinivasan: A Study on Deep Learning Models for Satellite Imagery. *International Journal of Applied Engineering Research* ISSN 0973-4562 Volume 14, Number 4 (2019) pp. 881-887
- [9] Powers, D.M.W (2011) Evaluation: From Precision, Recall And F-Measure To Roc, Informedness, Markedness & Correlation. *Journal Of Machine Learning Technologies*:2(1),37-63.
- [10] Q. Zou, L. Ni, T. Zhang, and Q. Wang (2015) Deep learning based feature selection for remote sensing scene classification. *IEEE Geosci. Remote Sens.* 12( 11), 2321-2325.
- [11] Goodfellow, I., Abadie, J., Mirza, M., Xu, B., Farley, D., Ozair, S., Courville, A., Bengio, Y., 2014. Generative adversarial nets, arXiv:1406.2661v1.
- [12] Isola, P., Zhu, J.Y., Zhou, T., Efros, A.A., 2017. Image-to-image translation with conditional adversarial networks. arXiv preprint arXiv:1611.07004.
- [13] Blaschke, T.: Object based image analysis for remote sensing. *ISPRS J. Photogramm. Remote Sens.* 65(1), 2–16 (2010)
- [14] GIS Geography. <http://gisgeography.com/image-classification-techniques-remote-sensing/>. Accessed Feb 16 2018
- [15] Mayar Shafaey, Mohammed A.M. Salem, H. M. Ebied, M. N. Al-Berry, and M. F. Tolba: Deep Learning for Satellite Image Classification. Springer Nature Switzerland AG (2019)