EAI Endorsed Transactions
on Cognitive Communications
Research Article EAI.EU

Comparative Analysis of Scene Classification Methods for Remotely Sensed Images using Various Convolutional Neural Network

P. Deepan[1*] and L.R. Sudha[2]

[1]Research Scholar, Department of Computer Science and Engineering, Annamalai University, Chidambaram, Tamil nadu.
[2]Associate Professor, Department of Computer Science and Engineering, Annamalai University, Chidambaram, Tamil nadu.

Abstract

Remote sensing image (RSI) scene classification has received growing attention from the research community in recent days. Over the past few decades, with the rapid development of deep learning models particularly Convolutional Neural Network (CNN), the performance of RSI scene classifications has been drastically improved due to the hierarchical feature representation learning through traditional CNN. But, we found that these models suffer from characterizing complex patterns in remote sensing imagery because of small inter class variations and large intra class variations. In order to tackle these problems, we have finetuned and proposed three different CNN models namely, Dilated CNN (D-CNN), RSI Scene Classification model (RSISC-16 Net) and fused the features of CNN and RSISC-16Net to improve the performance of RSI scene classification. The aim of proposed CNN models is to incorporate more relevant information by increasing the receptive field of convolutional layer. In addition, we have performed feature fusion of two CNN models and finetuned by varying hyper parameters such as activation function, dropout probability and batch size to reduce over fitting problem and to improve the performance of our proposed work. For evaluating the proposed approach, we have collected 7,000 remote sensing images from NWPU 45-class dataset and the experiments are carried out using different CNN models and results. The obtained accuracy is 89.85%, 94.7% and 96.5% respectively.

Keywords: Remote sensing images, scene classification, dilated convolutional, convolutional layer and feature fusion.

Received on 11 January 2021, accepted on 05 February 2021, published on 11 February 2021

doi: 10.4108/eai.11-2-2021.168714

*Corresponding author. Email:deepanp87@gmail.com

1. Introduction

With the rapid development of earth observation technology, image scene classification plays a significant role in the field of RSI. It is used for variety of applications ranges from agriculture monitoring, environmental monitoring, land use/ land cover planning, scene classification, urban planning, surveillance, geo-graphic mapping, disaster control, object detection and etc., [1-2]. Several techniques have been developed for image scene classification during the last decades. These techniques are broadly categorized into two types based on the features they use, namely low-level features learning based and high level or deep features learning based methods. Earlier, image scene classification was based on the low level features or handcraft feature learning method [4]. This method was mainly used for designing the handcraft or human engineering features, such as color [3], shape, texture, spatial and spectral information. The histogram of gradients (HoG), color histogram (CH), gray level co-occurrence matrix (GLCM), local binary pattern (LBP) and scale in-variant feature transform (SIFT) are some of the familiar handcraft feature extraction methods that are used for image scene classification [5-6]. These low level features are producing better results, but they require domain expertise and consume more time for the limited data. In addition, handcrafted features require an artificial dilation for extracting the features.

To overcome the limitations of handcraft features, learning the features automatically from images is considered as best way. In recent years [7], deep learning method has great success in the field of image scene classification. It is composed of multiple layers that can learn more powerful feature extraction of data with multiple levels of abstraction. In addition, the deep layers of representations have great potential to characterise robust features with complex patterns and semantics, such as land use, land cover, functional sites and etc. Currently, so many deep learning models are available such as Convolutional Neural Network, Recurrent Neural Network (RNN) with Long Term Short Memory (LSTM), Auto Encoder (AE), Deep Belief Network (DBN) and Generative Adversarial Network (GAN).

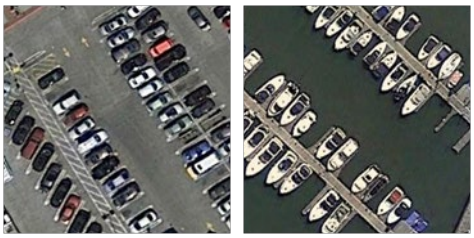

Figure 1. Sample image labelled with (a) parking lot and (b) harbor

The main reasons for the popularity of deep learning are the highly improved parallel processing capability of hardware especially the general-purpose graphical processing units (GPUs), the substantially increased size of data available for training, and the recent improvements in machine learning algorithms. These advancements enable deep learning methods to effectively utilize complex data, compositional nonlinear functions, learn distributed and hierarchical features automatically by utilizing both labeled and unlabeled data effectively. Figure 1a and 1b, are the images from NWPU 45-class dataset in which images have similar visual perceptions; they are correctly classified as car in parking lot and ship in harbor using deep learning models. So, successful deep learning application requires a very large amount of data to train the model as well as GPU to process the data rapidly. Especially, the CNN models are familiar and widely used for image classification and have achieved better results.

The remainder of the paper is structured as follows: Section "Related works" contains the literature survey of CNN classification for remote sensing images; Section "Proposed work" presents the newly developed different CNN models such as dilated convolutional neural network, RESISC-16 model and feature fusion of CNN and VGG-16; Section "Experimental result and analysis" discusses how the performance is improved from traditional CNN to new proposed convolutional neural network models; and in Section "Conclusion" we reiterate the focus of the paper and summarize the work presented.

2. Related Works

The first CNN model was developed by LeCun et al.[8-9] which is similar to the traditional neural network and also it is the foundation for modern CNN. The structure of the CNN model is inspired by the neurons in animal and human brains. In recent days, researchers have developed many models which are related to image classification problems. For example, Xuning Liu et al.[10] developed Siamese networks for Remote sensing scene classification. The results showed that the Siamese CNN model performance is efficient and better than the VGG-16 (Visual Geometry Group) results. The research in [11] proposed CNN model for road recognition system from remote sensing images. The research by Wong et al.[12] presented a smart object detection system for blind people. This method captured object scene by webcam and then extracted the features by using convolutional layer. After that, audio detector was used to analyse the detected object for the blind people. Chih-Yuan Koh et al.[13] proposed a bird sound classification model, in which features of ResNet model and Inception model are combined. Yu Weng et al.[14] introduced an effective framework for solving different image scene classification based on convolutional neural architecture search (CNAS).

Souleyman Chaib et al.[15] developed a feature fusion model for high resolution remote sensing image scene classification, where VGG-16 and Inception model features are combined. In [16] a deep learning fusion framework was introduced for improving the classification accuracy of remote sensing images. This method used feature fusion of three state-of-the-art models namely traditional CNN, VGG-16 and ResInception and obtained higher accuracy than the individual models. In [17] a deep CNN model was proposed for classification and detection of plant leaf diseases. Yunya Dong et al.[18] introduced a combined deep learning model for High Resolution-RSI scene classification. This model combines CNN features representation with LSTM model for improving the accuracy of scene classification. Gong Cheng et al.[19] proposed a discriminative CNN model to improve the performance of RSI scene classification, in which within class diversity and between class similarity problems are addressed. Abdul Qayyum et al.[20] proposed an efficient method for scene classification of aerial images by CNN based sparse coding learning techniques.

Wei Zhang et al.[21] introduced a capsule network for RSI-scene classification. This model first extracted the features based on CNN and then the extracted features are fed into capsule network to obtain better classification accuracy. Peng Ding et al.[22] performed object detection model for RSI images by faster regional CNN approach. This model reduces the detection time (test-time) and memory requirements. In addition, the proposed model can detect small objects in RSI more efficiently. Antonio-Javier Gallego et al. performed automatic ship classification by combining CNN and k-Nearest Neighbor method (k-NN) to improve the performance [23]. Maher Ibrahim et al. [24], classified Very High Resolution(VHR) aerial photographs to classify the several land cover classes namely, building,

barren land, dense area, grassland, road, shadow and water body of Selangor. Malaysia. Grant J. Scott et al.[25] presented a fusion algorithm in which multiple deep CNN models such as CaffeNet, GoogLeNet, and ResNet50 are used and features were extracted for land cover classification of HRI. In [26] a dilated CNN model was proposed for scene classification of remote sensing images. All the above mentioned models are not efficient as they require more computational time to train and validate the data. Taking the above disadvantages into consideration, we have proposed three convolutional neural networks namely dilated convolutional model, RSISC-16 Net and fused the features of CNN and RSISC-16 for scene classification of remote sensing images.

3. Proposed Work

In this section, we have proposed three different convolutional neural network models such as Dilated Convolutional Neural Network, fine tuning the hyper parameter of RSISC-16 Deep CNN model and Feature fusion of CNN and RSISC-16 by replacing the traditional CNN for improving the classification accuracy of RSI scene classification. To deal with more complex situation and to achieve better performance of network, we have incorporated more relevant information by increasing the receptive field of convolutional layer in traditional model and also increased the number of convolutional layers. In order to handle these problems, which are present in the previous study, we have introduced different kinds of CNN instead of traditional CNN. In general, the traditional CNN consists of four major steps namely, convolutional layer, sub-sampling layer, activation function and fully connected layer.

3.1. Dilated Convolutional Neural Network

The Dilated CNN model consists of N number of dilated convolution layers followed by N number of pooling layers and two fully connected layers. The major problem in deep learning techniques is overfitting while training these structures. Data augmentation, optimizer, dense, dropout and drop connect are some of the techniques developed to avoid overfitting problems.

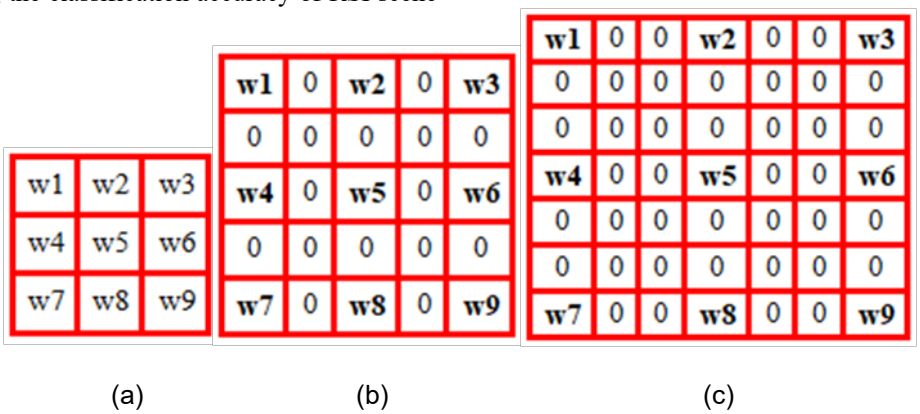

Figure 2. (a) Traditional convolutional with kernel size 3×3 (b) dilated convolution with dilation rate 2 and kernel size is 5×5 (c) dilated convolution with dilation rate 3 and kernel size is 7×7.

Figure 3 shows the traditional and dilated convolution kernel over an image of size10 × 10, where (a) is a traditional 3×3 convolution kernel, a zero is inserted between each point in the matrix in (a) and transformed into (b) is called dilation rate 2, similarly, (c) is a dilation rate 3 kernel. As shown in Figure 2, the kernel's receptive field are 3×3, 5×5 and 7×7 respectively. The receptive field size is increased by adding the zero between the matrices; however, the number of parameters in all the dilated convolution kernels is same. Therefore, using such a dilated convolutional kernel to process images, we can get more information from the convolution kernel without increasing the computation. In dilated convolution, a small kernel size w×w is extended to w + (w-1) (dr-1) with dilate rate dr. In traditional convolutional kernel with size of 3×3, and its receptive field is 3×3. While performing dilated convolutional kernel with size of 3×3, it's receptive field is 5×5 when dilation rate dr = 2, and 7×7 when dr = 3. The receptive field is generally defined as [k + (k − 1) (i −1)] × [k + (k − 1) (i − 1)] when dr = i.

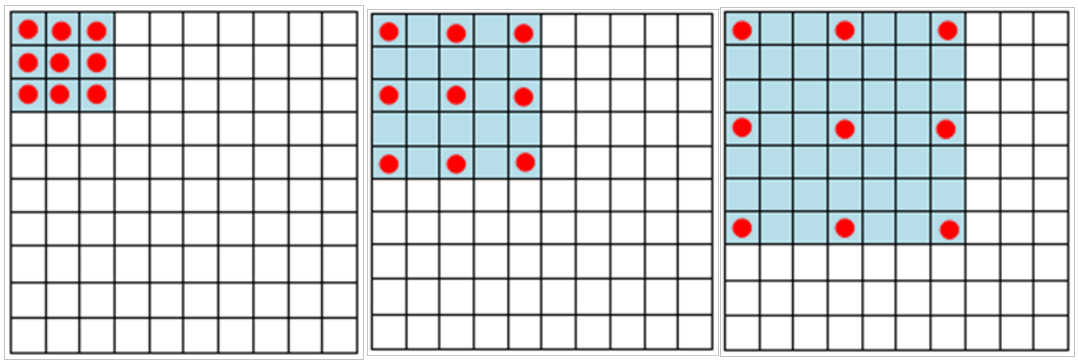

Figure 3. (a) Conceptual illustration of traditional and dilated convolution; (a) traditional convolution (b) dilated convolution with dilation rate 2; (c) dilated convolution with dilation rate 3.

Dilated Convolutional Layer

The convolution layer is the most important layer in the CNN which is the origin of the "convolutional neural network". The aim of convolution layer is to learn feature representations of the inputs. The convolution layer is a three dimensional matrix with size of h×w×c with corresponding weight for each point, where h represents height of the inputs, w represents width of the inputs and c represents the depth of the channel. A kernel of convolution is a neuron, and the size of the convolutional kernel is called as neuron's receptive field. Like neural networks, convolutional networks use convolution operation rather than matrix multiplication process. The general form of convolution is defined as:

$$s(i,j) = \sum_{k=1}^{n_{in}} (X_k \times W_k)(i,j) + b \qquad (1)$$

where n_{in} represents the input matrices of the tensor. X_k is k^{th} input matrix. W_k is the k^{th} sub-convolution kernel matrix of the convolution kernel. s(i,j) represents the output values for matrix of corresponding elements to the kernel w. For example, 10×10 two-dimensional matrix as a input with padding size of 1(11×11 input size) and the size of convolution kernel is 5×5 matrix, the size of stride is set to 1, the output of corresponding convolution size is 7×7 and convolution process is shown in Figure 4.

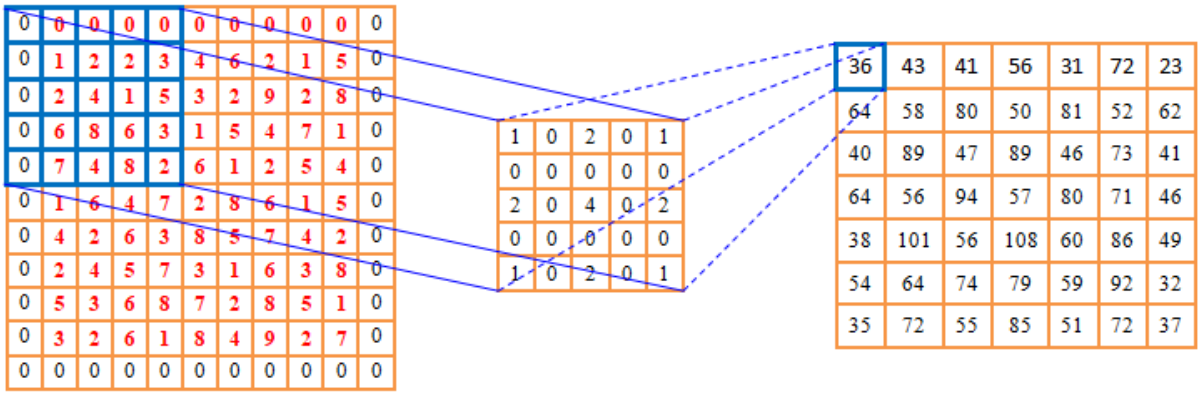

Figure 4. Pictorial representation of dilated convolutional layer

Pooling Layer

The sub sampling or pooling layer is used to reduce the feature resolution. This layer reduces the number of connection between the convolutional layers, so it will lower the computational time also. There are three types of pooling: max pooling, min pooling and average pooling. In each case, the input image is divided into non-overlapping two dimensional spaces. For example, if the input size is 4×4 and sub sampling size is 2×2, a 4×4 image is divided into four non-overlapping of matrices 2×2. For max pooling, the maximum value of the four values is selected. In the case of min pooling, the minimum value of the four values is selected. The figure 5 shows the operation of max pooling and average pooling process.

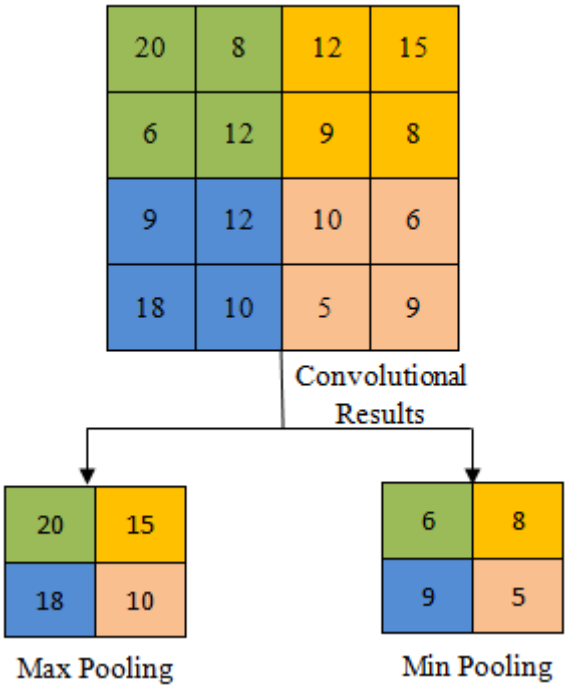

Figure 5. Pictorial representation of max-pooling and min-pooling process

The CNN model ends with fully connected layer and softmax function. In these layers, sum of all the weight of previous layer features is calculated and the specific output is determined. Finally, fully connected layers reduce the dimension into 2048 and classify the ten class object using softmax function. The Activation Function improves the Deep CNN performance. In general, there many activation functions are available such as tanh, sigmoid, ReLU and etc., for solving the non linear problems. The Rectified linear unit (ReLU) is one of the standard and popular activation functions in the last few years.

$$b_{i,j,k} = \max(a_{i,j,k}, 0) \quad (2)$$

where, $a_{i,j,k}$ is the input of the activation function at location (i, j) on the k^{th} channel. In this layer we remove every negative value from the filtered images and replace it with zeros. The Figure 6 elaborates the process of activation function.

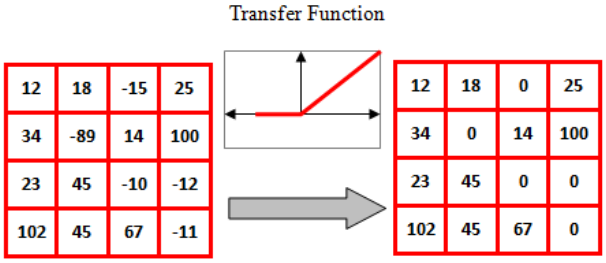

Figure 6. Pictorial representation of activation function

Fully Connected Layer

After the several convolutional and pooling layer processes, the two-dimensional data is converted into one-dimensional vector. The one dimensional data will be the input for fully connected layers. There may be one or more hidden layers which perform high level reasoning. Each neuron uses the data from the previous layers and multiplies them with the connection weights and then adds a bias value. The working principle of flatten layer is shown in Figure 7. The output of final fully connected layers is fed into the classifier ie. softmax function. The softmax function is used to classify the object. The general form of softmax is defined as in Eq. (3)

$$\text{class}_j = \frac{\exp(sf_j)}{\sum_q(sf_q)} \quad (3)$$

where, exp (sf_j) is the probabilities of each target class where as sf_q is possible of all the target classes.

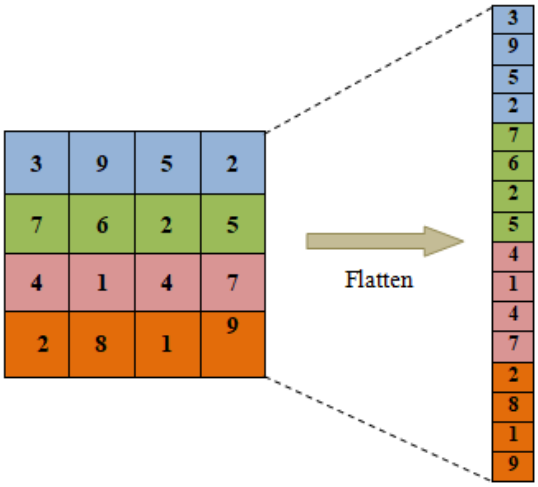

Figure 7. Pictorial representation of flattening process

3.2. Fine tuning the hyper parameter of RSISC-16 Model

This section describes the RSISC-16 deep CNN model for scene classification of remote sensing images using deep convolutional neural networks. Figure 8. shows the architecture diagram of RSISC-16 deep CNN model. It takes input image at the low level and processes them through a sequence of computational unit and obtains the necessary values for classification in the higher layer. This model consists of 13 convolutional layers with 3×3 filter sizes, 5 sub sampling/ pooling layer with size of 2×2, two fully connected layers with activation function and softmax function.

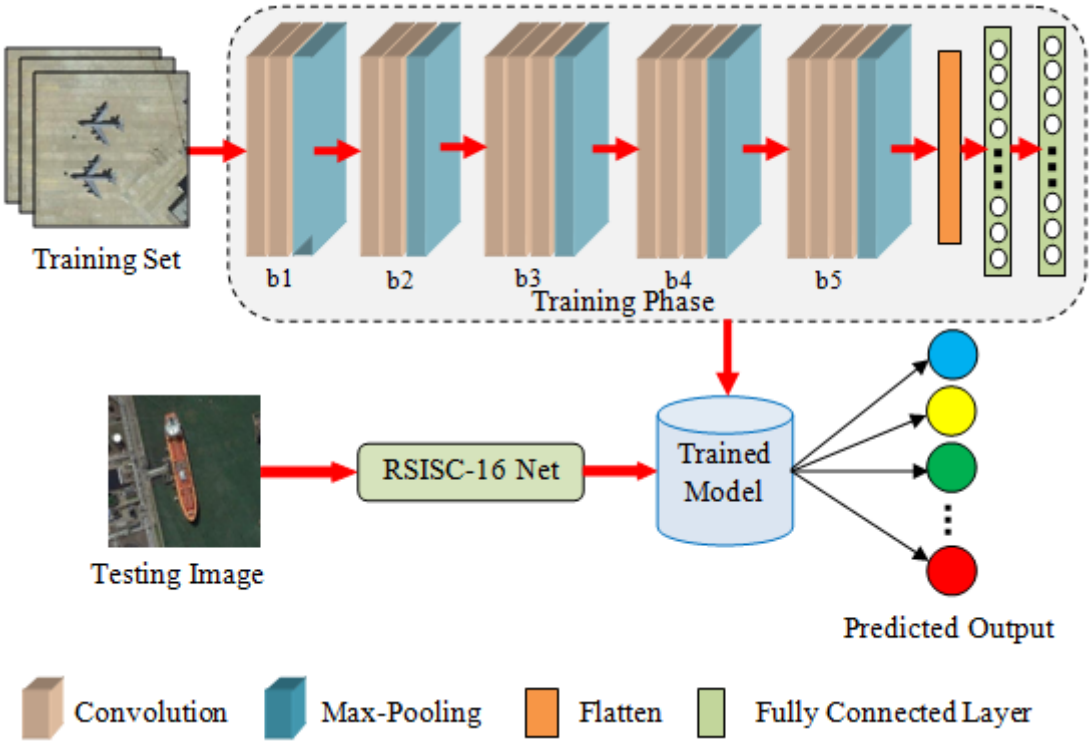

Figure 8. Architecture of RSISC-16 Model

The convolutional layers extract features from the input images. The 13 convolutional layers are distributed in five blocks. The first two blocks contain two convolutional layers in each block. Similarly, the remaining three blocks consist of 3 convolutional layers in each block. The first block convolutional layer extracts low level features such as lines and edges. Higher level layer extracts high level features. Every convolutional filter has a kernel size of 3×3 filters with stride 1.The filter size of convolutional layer is gradually increased from 64 to 512.

The overfitting is an unneglectable problem in RSISC-16 deep CNN model that can be reduced by regularization. In this paper, we use the effective regularization technique of dropout is used. The dropout was introduced by Hinton et al. and it has been proved that it is effective in reducing over fitting problem. The dropout techniques are used in the fully connected layer and we can specify the different level of dropout parameter like 0.2, 0.3 and 0.5. The RSISC-16 deep CNN model was trained with back propagation algorithm and Root Mean Square property (RMS prop). The RMS prop is used to reduce the loss function of RESISC-16 deep CNN model.

3.3. Feature Fusion of CNN and RSISC-16 Model

In this work, we focus on the ensemble model for improving the performance and increasing the predication rate of remote sensing image scene classification. We have introduced two standard Convolutional Neural Networks such as CNN and RSISC-16 as our base models. For each base model, we have fine tuned the hyper-parameters and trained the two models independently with 10 class dataset. Then we have calculated the average of all the features after fully connected and applied softmax function to classify the data. The architecture diagram of the proposed method is shown in Figure 9 and the details of the components in the architecture diagram are discussed in the following sections.

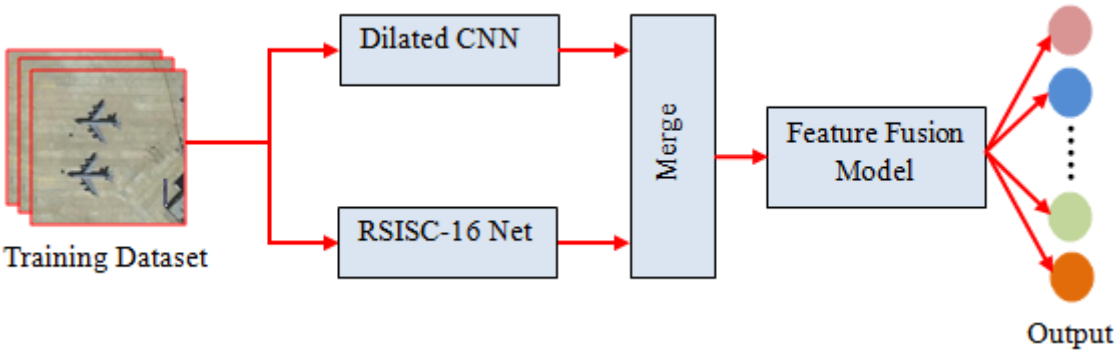

Figure 9. Architecture of average feature fusion model

Convolutional Neural Network

The CNN model consists of four hidden layers (three convolutional- sub sampling layer and one fully connected layer), one input layer and one output layer. The input layer contains 224×224×3 neurons, indicating the RGB values for a 224×224×3 image. The convolutional- sub sampling layer use the size 3×3 stride lengths and followed by 2×2 regions. The fully connected layer contains 256 neurons with Rectified Linear Unit (ReLU). Finally, the output layers used soft-max function to produce the class of objects in RSIs.

RSISC-16 Model

This model is one of the most powerful deep convolutional neural networks which have been proposed by Simonyan et al. It consists of 13 convolutional layers with 3×3 filter size, 5 sub sampling/ max pooling layer with size of 2×2, two fully connected layers with activation function and soft-max function. Each block is made by consecutive 3×3 convolutions and followed by a max pooling layer. To avoid the problem of over-fitting, we need to eliminate the redundant features by adding the Dropout.

Average Feature Fusion Classifiers

Average feature fusion technique is one of the familiar ensemble approaches for classifying remote sensing images. It takes average of output score divided by probability of all base CNN model and reports it as predicted score divided by probability. Due to the high capacity of Deep CNN, the average feature fusion model improves the performance substantively. Taking average of multiple models will reduce the variance. The number of convolutional layer, pooling layer in each model is showed in Table 1.

Table 1. Trainable and Non- Trainable Parameters of CNN and RSISC-16 Model

S. No.	Model	No. of Convolutional Layer	No. of Pooling Layer	Total No. of Parameters
1.	CNN	3	3	14,62,863
2.	RSISC-16	13	5	1,47,16,227
3.	**Feature Fusion Model**	**18**	**8**	**1,61,79,090**

4. Experimental Analysis

In this section, we have discussed and analyzed various remote sensing image scene classification methods using deep learning techniques based on Dilated CNN, RSISC-16 model, feature fusion of CNN and RSISC-16. First, we introduce the benchmark datasets for RSI scene classification, then analyzed the performance of traditional CNN model, and finally presented the experimental results for three proposed models with same dataset and parameters.

4.1. Dataset Description

For experimental evaluation, we have used publicly available large-scale benchmark dataset for remote sensing image scene classification. The dataset is North Western Polytechnical University (NWPU) 45-class dataset [4] which contains 45 classes and totally 31,500 images. Each class consists of 700 images with resolution of 256×256 pixels. As far as we know, the NWPU45 class is the most challenging dataset in very high resolution (VHR) image scene classification tasks because it has larger scale scene categories and image number than other datasets.

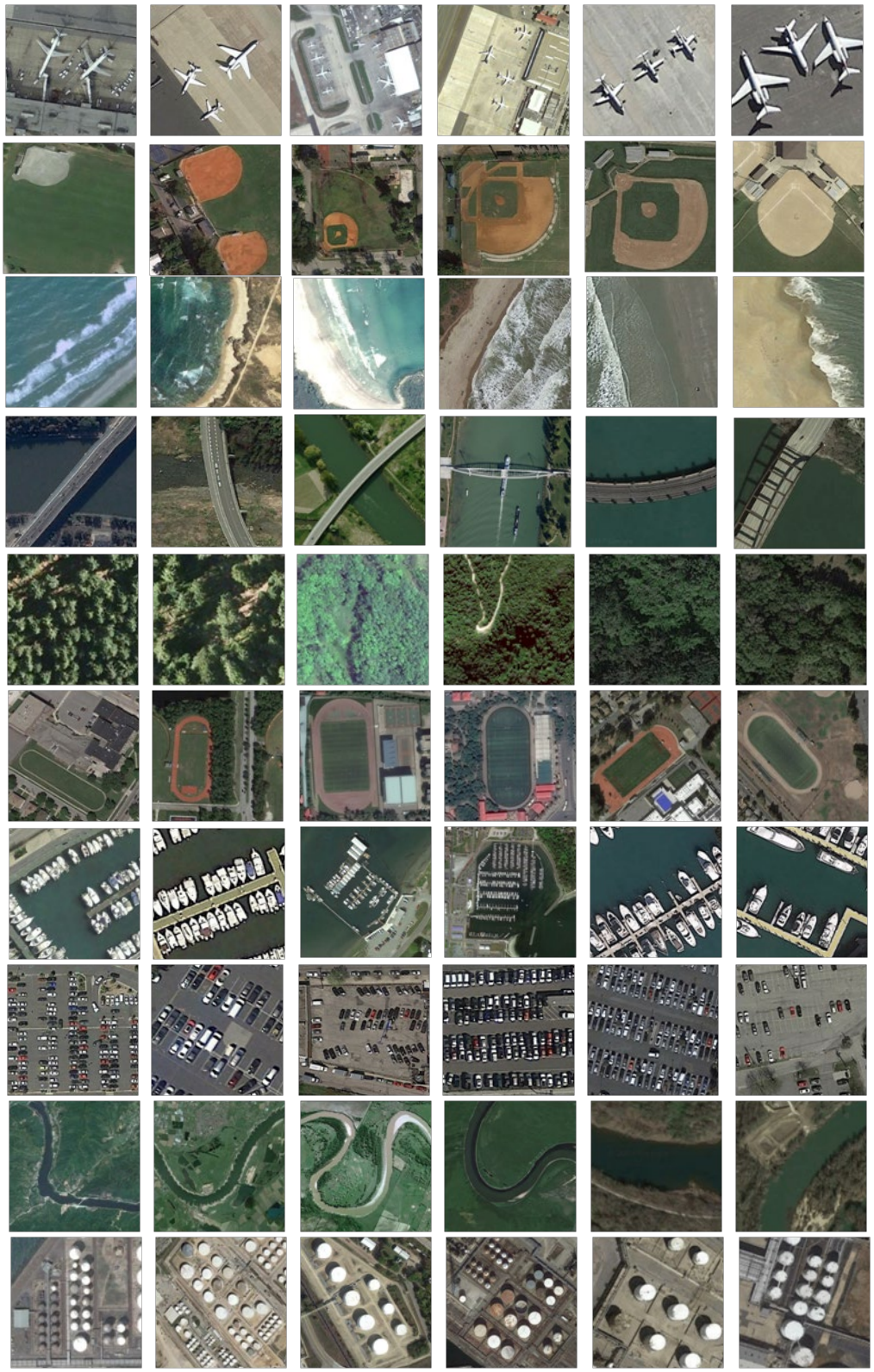

Figure 10. The sample images from NWPU 45-class dataset of our proposed models

In addition, each image category in NWPU-45 class dataset have rich variations, like illumination changes, resolution, shooting angle, background, etc., which also increases the difficulty of scene classification. The spatial resolution of images ranges from 0.2 to 30m. The dataset was collected from more than 100 countries and extracted by Google Earth. For our proposed work, we have chosen ten classes namely airplane, baseball diamond, beach, bridge, forest, ground track, harbor, parking lot, river and storage tank for remote sensing image scene classification. A sample image from benchmark dataset is shown in Figure 10.

4.2. Performance Metrics

We have evaluated the performance of a proposed model by using various performance metrics such as Accuracy, Precision, Recall, F1-measure and Mean Square Error (MSE). The Accuracy can be calculated by the number of properly classified data in a dataset divided by the total number of samples, as shown in the equation (4).

$$\text{Accuracy} = \frac{t}{n} \tag{4}$$

where, t is a number of properly classified samples and n is a total number of samples in a dataset. The precision can be measured by number of properly classified data in a datasets divided by total number of all samples in a class. Precision value of the class c, P_c can be shown in equation (5) where, t_c is a total number of properly classified samples in class c and n_c is a total number of samples in the class c.

$$P_c = \frac{t_c}{n_c} \tag{5}$$

The recall can be measured by number of properly classified data which are divided by the number of all relevant samples in the corresponding class. Recall value of the class c, R_c can be shown in equation (6) where, t_c is a total number of properly classified samples in class c and k_c is number of samples classified as relevant to class c.

$$R_c = \frac{t_c}{k_c} \tag{6}$$

The F1-measure (harmonic mean) is used to show the balance between the precision and recall measures. F1-score value can be calculated using equation (7):

$$F_1 = \frac{2*(P_c*R_c)}{P_c+R_c} \tag{7}$$

The MSE value returns prediction error rate between the original input image and predicted image. It is calculated by the equation below:

$$\text{MSE} = \frac{\sum_{i=1}^{H}\sum_{j=1}^{W}(O(i,j)-\tilde{O}(i,j))^2}{HxW} \tag{8}$$

Where, O and $\tilde{O}$ are represented as observed original input image and predicted image respectively and value of H and W represented as Height and Width of the images.

	P	N
Y	True Positive	False Positive
N	False Negative	True Negative

Figure 11. Concepts of Confusion Matrix

It is essential to find the confusion matrix while calculating the performance measures. Confusion matrix is a technique used to summarise results and used for validating classification methods. There are two common classes, which are usually dealt with confusion matrix namely positive class and negative class. These two common classes can be further divided into four categories namely, True Positive (TP), False Positive (FP), True Negative (TN) and False Negative (FN). TP is an outcome, where the model that has correct classification of positive example. FN is an outcome, where the model that has incorrect classification of positive examples. FP is an outcome, where the model that has incorrect classification of positive examples. TN is an outcome, where the model that has correct classification of negative examples.

4.3. Experimental Results of different Scene Classification Methods

The proposed model is trained and tested with NWPU 45-class dataset using tensor flow in Core i7 CPU 2.6GHz, 1 TB of Hard Disk and 16 GB of RAM. The proposed models are developed in Python and Anaconda IDE tools with various python libraries like keras, numpy and open CV. The experimental results of our three proposed models are compared with traditional CNN model with same parameter and configurations. The experimental setting of traditional CNN and Dilated CNN model consists of three convolutional layer and max pooling. For avoiding the problem of over fitting concepts, we have used dropout and Adam optimizers. Figure 12, shows the performance metrics of NWPU 45- class dataset (traditional CNN and dilated CNN model) for 15 epochs on both training and validation data. In first experimental set, Dilated CNN model has the highest performance accuracy because the receptive fields which are increased more than traditional CNN model.

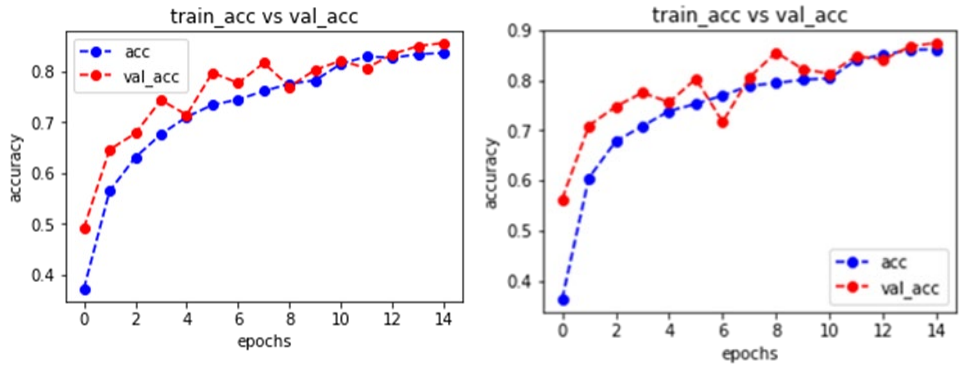

Figure 12. Performance analysis of traditional CNN vs. dilated CNN model

Similarly, The RSISC-16 model was trained by varying three hyper parameters namely, activation function, dropout probability and batch size. In this analysis, we found that the activation function "ReLU" has higher results when compared with other two functions namely, ELU and tanh. By using ReLU activation functions, the RSISC-16 Deep CNN model has achieved 94.7% accuracy as a result. The experiment by varying dropout probability gives better performance for 0.3. Also, the efficiency of RSISC-16 deep CNN model was compared with different batch size such as 4, 8, 12 and 16 and found that the batch size 4 gave better results. In order to improve the accuracy of remote sensing image scene classification, we have proposed feature fusion of CNN and fine tuned RSISC-16 model. Figure 13, shows the performance metrics of NWPU 45- class dataset of RSISC-16 model and feature fusion of CNN and RSISC-16 model for 15 epochs on both training and validation data.

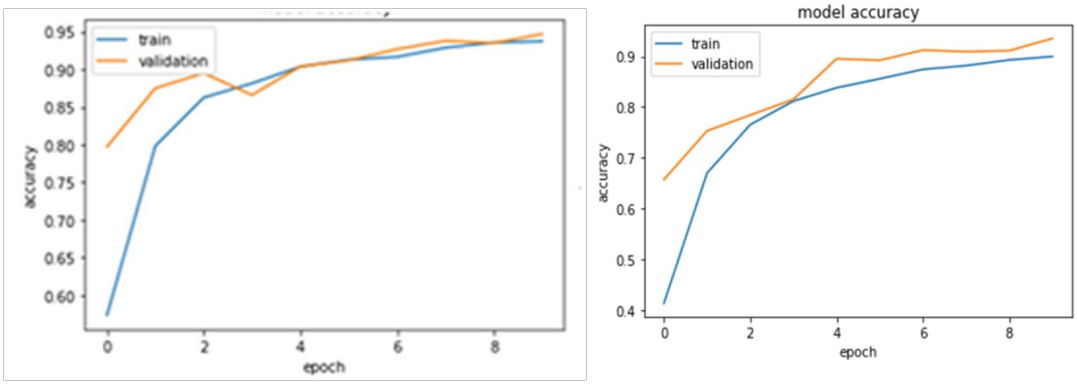

Figure 13. Performance analysis of fine tuned RSISC-16 vs. Feature fusion model

Table 1. Trainable and Non- Trainable Parameters of CNN and RSISC-16 Model

S. No.	Model	Accuracy	Precision	Recall	F1-Score	MSE
1.	Traditional CNN	85.85	86.21	85.86	85.73	3.44
2.	Dilated CNN	89.85	89.49	89.43	89.37	2.13
3.	RSISC-16 Model	94.7	95	95	95	0.86
4.	Feature Fusion of CNN and RSISC-16 Model	96.5	95.5	96	96.1	0.23

Based on experimental results, it is clear that, our three proposed CNN models have higher accuracy than traditional CNN model. The dilated CNN model has 4% higher accuracy than the traditional CNN model. Similarly, RSISC-16 model has 9% higher than the traditional model and feature fusion of CNN and RSISC-16 model has 11% higher accuracy than traditional model. Table 2 and Figure 14, shows the performance analysis of our three proposed CNN models with the traditional CNN model.

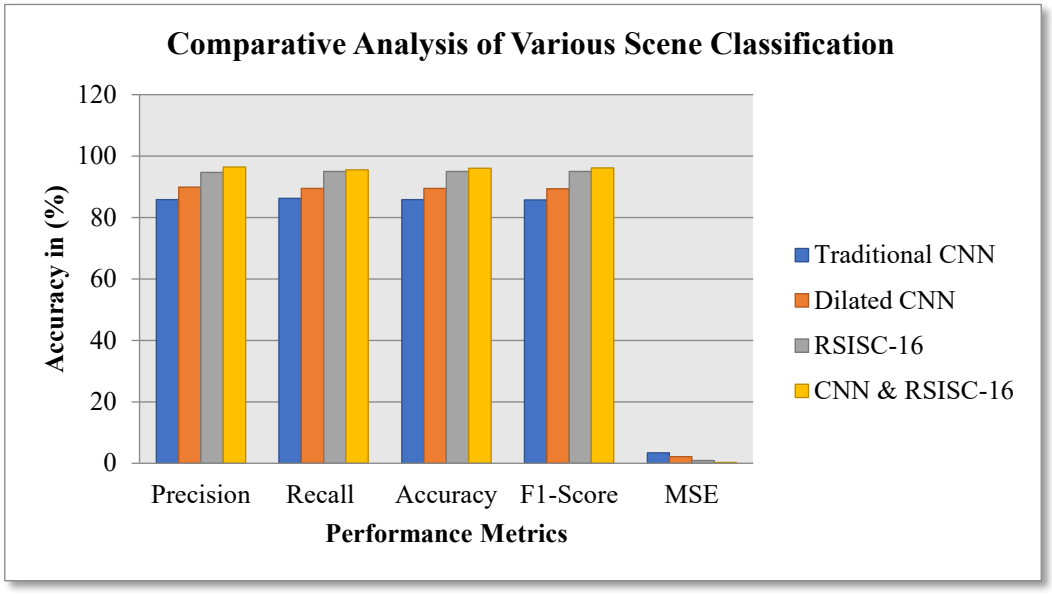

Figure 14. Comparative Analysis of various Proposed CNN Scene Classification methods

5. Conclusion

Scene Classification in remote sensing images is a challenging problem because objects of same category have often a diverse appearance. So, we have proposed three different CNN model for remote sensing image scene classification by replacing the kernel of traditional CNN; dilated convolutional model, fine tuned the hyper parameters of RSISC-16 model and feature fused D-CNN and RSISC-16Net model for accurate scene classification. In order to demonstrate the efficiency of proposed models, experiments are conducted on a 10 class dataset which are selected from NWPU RESISC-45 and achieved the accuracy as 89.85%, 94.7% and 96.5% respectively. We have observed that all the three proposed CNN models have given better accuracy than traditional CNN model. In future, we have planned to incorporate our proposed convolutional neural network model for remote sensing object detection and also the same will be implemented in GPU environment for reducing the computational time.

References

[1] Lei M., Yu L., Xueliang Z., Yuanxin Y., Gaofei Y and Brian Alan J. Deep learning in remote sensing applications: A meta-analysis and review, ISPRS Journal of Photogrammetry and Remote Sensing: 2018, pp. 166–177.

[2] Deepan P. and Sudha L.R. Object Classification of Remote Sensing Image Using Deep Convolutional Neural Network, The Cognitive Approach in Cloud Computing and Internet of Things Technologies for Surveillance Tracking Systems, 2020, pp. 107-120.

[3] Yu H., Yang W., Xia G.S. and Liu G. A color-texture-structure descriptor for high resolution satellite image classification, Jo. of Remote Sensing, 2016, pp. 259-269.

[4] Cheng G., Han J. and Lu X. Remote Sensing Image Scene Classification: Benchmark and State of the Art, Proceedings of the IEEE, 2017, pp. 1-19.

[5] Maxwell A., Warner T.A. and Fang F. Implementation of machine-learning classification in remote sensing: an applied review, International Journal of Remote Sensing, 2018, pp. 2784-2817.

[6] Khalid S., Khalil T. and Nasreen S. A Survey of Feature Selection and Feature Extraction Techniques in Machine Learning International conference on Science and Information, 2014, pp. 372–378.

[7] Zhang L. and Du B. Deep learning for remote sensing data: A technical tutorial on the state of the art," IEEE Geosci. Remote Sens. Mag., Vol.(4), 2016, pp. 22–40.

[8] Lecun Y., Bottou L., Bengio Y. and Haffner P. Gradient-based learning applied to document recognition, Proceedings of the IEEE, vol. (86), 2015, pp. 2278–2324.

[9] O'Shea K. and Nash R. An Introduction to Convolutional Neural Networks, International Journal of Computer Vision and Pattern Recognition, 2015, pp. 2-11.

[10] Liu X., Zhou Y., Zhao J., Yao R., Liu B. and Zheng Y. Siamese Convolutional Neural Networks for Remote Sensing Scene Classification, IEEE Geoscience and Remote Sensing Letters, 2019, pp. 1-5.

[11] Deepan P., Abinaya S., Haritha G. and Iswarya V. Road Recognition from Remote Sensing Imagery using Machine Learning, International Research Journal of Engineering and Technology, 2018, pp. 3677-3683.

[12] Wong Y.C., Lai J.A., Ranjit S.S., Syafeeza A.R. and Hamid N. A. Convolutional Neural Network for Object Detection System for Blind People, Journal of Telecommunication, Electronic and Computer Engineering, 2019, pp. 1-6.

[13] Koh C., Chang J., Tai C., Huang D., Hsieh H. and Liu Y. Bird Sound Classification using Convolutional Neural Networks, International Journal of computer vision, 2019, pp. 1-10.

[14] Wen Y., Zhou T., Liu L. and Xia C. Automatic Convolutional Neural Architecture Search for Image Classification under Different Scenes, IEEE Transaction on Innovation and Application in Edge Computing, 2019, pp. 38495- 38506.
[15] Chaib S., Liu H., Gu Y. and Yao H. Deep Feature Fusion for VHR Remote Sensing Scene Classification, IEEE Transactions on Geoscience and Remote Sensing, Vol.2(10) , 2017, pp. 1-10.
[16] Deepan P. and Sudha L.R. Fusion of Deep Learning Models for Improving Classification Accuracy of Remote Sensing Images, Journal Of Mechanics of Continua and Mathematical Sciences, 2019, Vol. 3(12), pp. 189-201.
[17] Akila M. and Deepan P. Detection and Classification of Plant Leaf Diseases by using Deep Learning Algorithm, International Journal of Engineering Research & Technology (IJERT), 2018, pp. 1-6.
[18] Dong Y. and Zhang Q. A. Combined Deep Learning Model for the Scene Classification of High-Resolution Remote Sensing Image, IEEE Geoscience and Remote Sensing Letters, 2019, pp. 1-5.
[19] Cheng G., Yang C., Yao X., Guo L. and Han J. When Deep Learning Meets Metric Learning: Remote Sensing Image Scene Classification via Learning Discriminative CNNs, IEEE Transactions on Geoscience and Remote Sensing, 2018, pp. 1-11.
[20] Qayyum A., Malik A., Saad N.M., Iqbal M., Abdullah M.F., Rasheed W., Abdullah T. and Jafaar M. Scene classification for aerial images based on CNN using sparse coding technique, International Journal of Remote Sensing, 2017, pp. 1-24.
[21] Zhang W., Tang P. and Zhao L. Remote Sensing Image Scene Classification Using CNN-CapsNet, Remote Sens., 2019, pp. 1-22.
[22] Ding P., Zhang Y., Deng W., Jia P. and Kuijper A. A light and faster regional convolutional neural network for object detection in optical remote sensing images, ISPRS Journal of Photo. and Remote Sensing, 2018, pp.208–218.
[23] Gallego A., Pertusa A. and Gil P. Automatic Ship Classification from Optical Aerial Images with Convolutional Neural Networks, MDPI Journal of remote sensing, 2018,pp. 1-20.
[24] Sameen M., Pradhan B. and Aziz O. Classification of Very High Resolution Aerial Photos Using Spectral-Spatial Convolutional Neural Networks, Journal of Sensors, Vol.1(10), 2020, pp. 1-13.
[25] Scott G.J., Marcum R.A., Davis C.H., and Nivin T.W. Fusion of Deep Convolutional Neural Networks for Land Cover Classification of High-Resolution Imagery, IEEE Geoscience and Remote Sensing Letters, Vol.2(5), 2017, pp. 1-5.
[26] Deepan P. and Sudha L.R. Remote sensing image scene classification using dilated convolutional neural networks, International Journal of Emerging Trends in Engineering Research, Vol. 8, No.7, 2020, pp.3622-3630.