ISSN 2347 - 3983
Volume 8. No. 7, July 2020
International Journal of Emerging Trends in Engineering Research
Available Online at http://www.warse.org/IJETER/static/pdf/file/ijeter120872020.pdf
https://doi.org/10.30534/ijeter/2020/120872020

Remote Sensing Image Scene Classification using Dilated Convolutional Neural Networks

P. Deepan[1], L.R. Sudha[2]
[1]Research Scholar, Department of Computer Science and Engineering, Annamalai University, Tamilnadu, India, deepanp87@gmail.com
[2]Associate Professor, Department of Computer Science and Engineering, Annamalai University, Tamilnadu, India, sudhaselvin@ymail.com

ABSTRACT

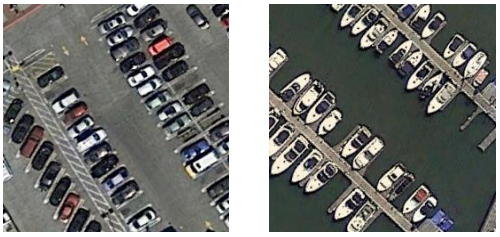
Remote sensing image (RSI) scene classification has received growing attention from the research community in recent days. Over the past few decades, with the rapid development of deep learning models particularly convolutional neural networks (CNN), the performance of RSI scene classification have been drastically improved due to the hierarchical feature representation learning through CNN. But, we found that these models suffer for characterizing complex patterns in remote sensing imagery because of small inter class variations and large intra class variations. In order to solve these problems, we have proposed a Dilated Convolutional Neural Network (D-CNN) to improve the performance of RSI scene classification. The aim of dilated convolution filter is to incorporate more relevant information by increasing the receptive field of convolutional layer. In addition to traditional CNN model, it increases CNN efficiency and reduce computational time. For evaluating the proposed approach, we have collected three publicly available benchmark datasets namely, NWPU 45-class, PatternNet and Aerial Image Dataset (AID). Finally, experimental results are demonstrated for our proposed model using above dataset and achieved 89.85%, 92.35% and 97.18% respectively, which is outperformed traditional CNN model.

Key words: Convolutional neural networks, deep learning, dilated convolutional, remote sensing images, dilation rate and scene classification.

1. INTRODUCTION

With the rapid development of earth observation technology, image scene classification plays a significant role in the field of RSI. It's applications ranges from agriculture monitoring, environmental monitoring, land use/ land cover planning, scene classification, urban planning, surveillance, geo-graphic mapping, disaster control, object detection, 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 feature learning based and high level or deep feature learning based method. 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), scale in-variant feature transform (SIFT) are some of the familiar handcraft feature extraction methods 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 limitation of handcraft features, automatically learning the features from images are 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 etc. Currently, there are so many deep learning models are available such as Convolutional Neural Network, Recurrent Neural Network (RNN) with Long Term Short Memory (LSTM), Auto Encoder , Deep Belief Network and Generative Adversarial Network.



(a) (b)

Figure 1: Sample image labelled with parking lot and harbor

The main reason 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 advances in machine learning algorithms. These advances enable deep learning methods to effectively utilize complex, compositional nonlinear functions, to automatically learn distributed and hierarchical

features, by effectively utilizing both labeled and unlabeled data. Figure 1a and 1b, are images from NWPU 45-class dataset though these 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 rapidly process the data[30]. Especially, the CNN models are familiar and widely used for image classification and have achieved better results. The rest 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 dilated convolutional model; Section "Experimental result and analysis" discusses how the performance is improved from traditional CNN to dilated convolutional model; 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 related to image classification problems. For example, Xuning Liu et al.[10] developed Siamese networks for Remote sensing scene classification. The results showed that 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 capture object scene by webcam and then extract 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 extract 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 Sameen et al.[24], classify 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 model such as CaffeNet, GoogLeNet, and ResNet50 features were extracted for land cover classification of HRI. 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 a dilated convolutional model for scene classification of remote sensing images.

3. DILATED CONVOLUTIONAL NEURAL NETWORKS

In this section, we have proposed a dilated convolutional neural network by replacing the traditional CNN for improving the classification accuracy of RSI scene classification. The new dilated convolution filter expands the receptive field without increasing parameters. So, we can improve the performance of this model and also reduce the computational time. With a deep convolutional network, the traditional CNN model with small convolution kernels needs to learn more relevant information, which has a more computational complexity. To deal with more complex situation and achieve better performance of network, by increasing the depth of CNN in traditional model. In order to handle these problems, we have introduced dilated convolution kernel instead of traditional convention kernel. Dilated convolution kernel is based on the idea of expanding or increasing kernel receptive field without increasing the number of parameters and by adding zero weight values in the filters.

The Dilated CNN model consists of N numbers of dilated convolution layers followed by N numbers of pooling layers and two fully connected layers. The architecture of proposed Dilated CNN model is shown in Figure 2 and layer1, layer2 and layer3 represent three levels of dilated convolutional layers with corresponding 'ReLu' activation function and max pooling are used for feature reduction. 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.

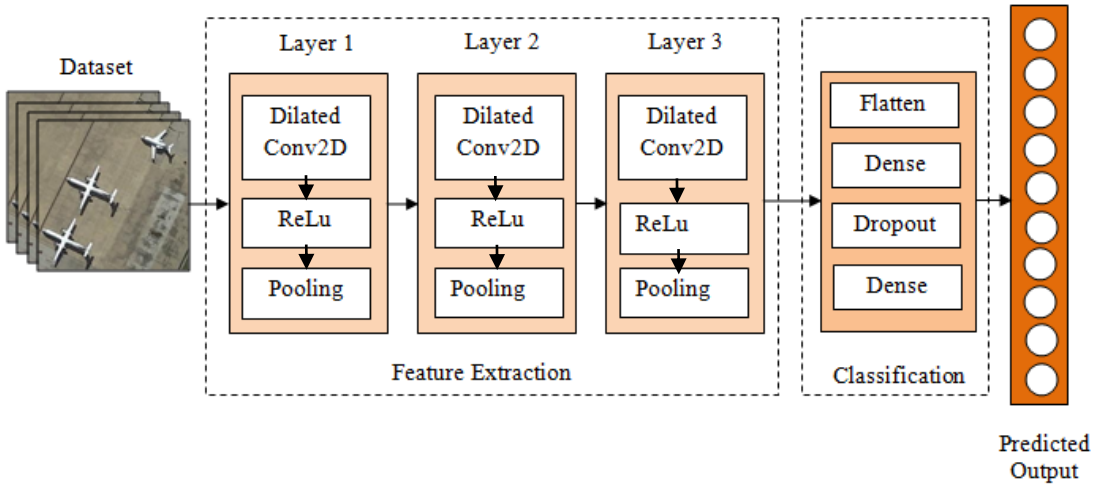
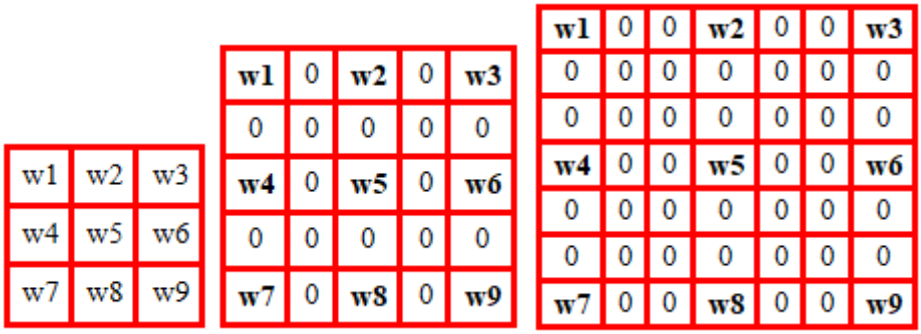


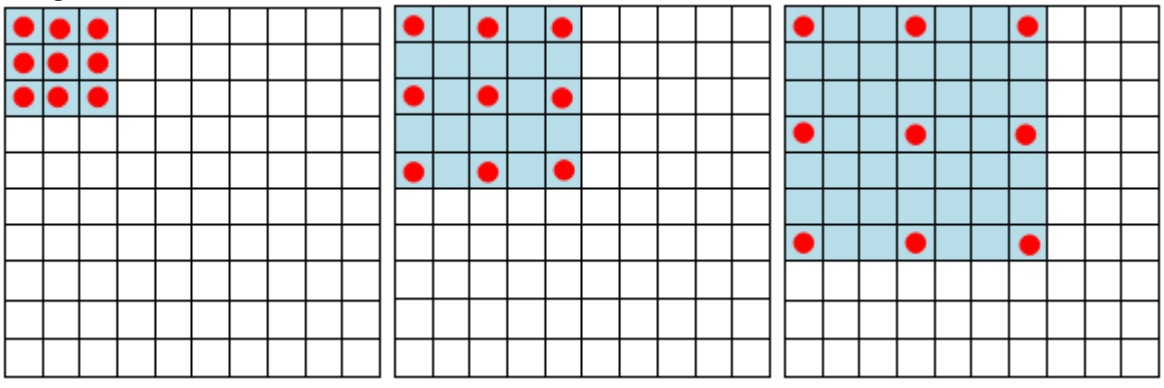
Figure 2: Architecture of proposed dilated convolutional neural network



(a) (b) (c)

Figure 3: (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 4 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 3, the kernel's receptive field is 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 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, the receptive field is 3×3. While performing dilated convolutional kernel with size of 3×3, its receptive field is 5×5 when dilation rate dr = 2, and 7×7 when dr = 3. The receptive field is generally defined as $[k + (k - 1) (i - 1)] \times [k + (k - 1) (i - 1)]$ when dr = i.



(a) (b) (c)

Figure 4: Conceptual illustration of traditional and dilated convolution; (a) traditional convolution (b) dilated convolution with dilation rate 2; (c) dilated convolution with dilation rate 3.

3.1 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 network uses 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 5.

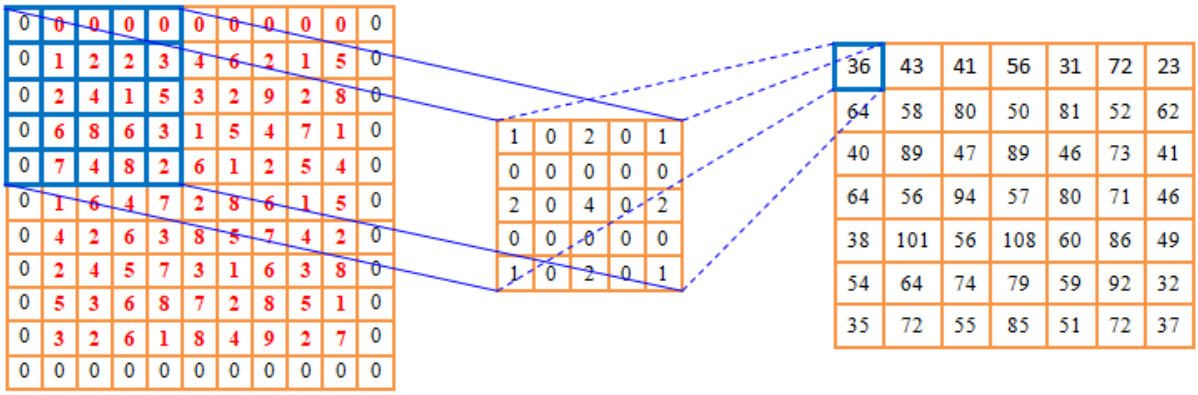


Figure 5: Pictorial representation of dilated convolution process

3.2 Activation Function

The Activation Function is mainly used to improve the performance of CNN model. There are many activation functions available such as ReLU, ELU, tanh, sigmoid and maxout. In this paper, standard and familiar Rectified Linear Unit (ReLU) activation function has been used. The ReLU activation function is defined as:

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

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 replace every negative value from the filtered images with zeros. The Figure 6. elaborates the process of activation function.

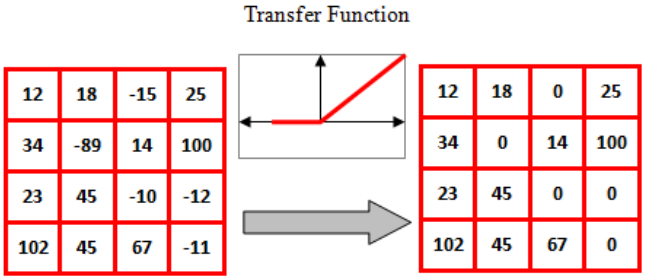


Figure 6: Pictorial representation of activation function

3.3 Pooling or Sub-sampling layer

The pooling process is used to reduce the dimensionality of feature maps that have passed through convolutional layer and activation functions. These processes reduce the number of connection between convolutional layers, so it will reduce the computational time also. The well known pooling types are max pooling, min pooling and average pooling. In each case, the input image is divided into non-overlapping two dimensional spaces. The input size is 4×4 and sub sampling size is 2×2. A 4×4 image is divided into six non-overlapping of matrices 2×2. The Figure 4 shows the operation of sub sampling process. For max pooling operations, the maximum value of the four values is selected. In the case of min pooling, the minimum value of the four values is selected and similarly average pooling takes average value of input. The main advantage of pooling process is to provide reduction of input image. Figure 7, shows the operation of max pooling and average pooling process.

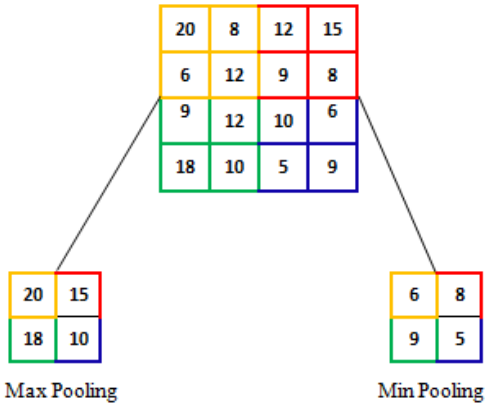


Figure 7: Pictorial representation of Pooling processes

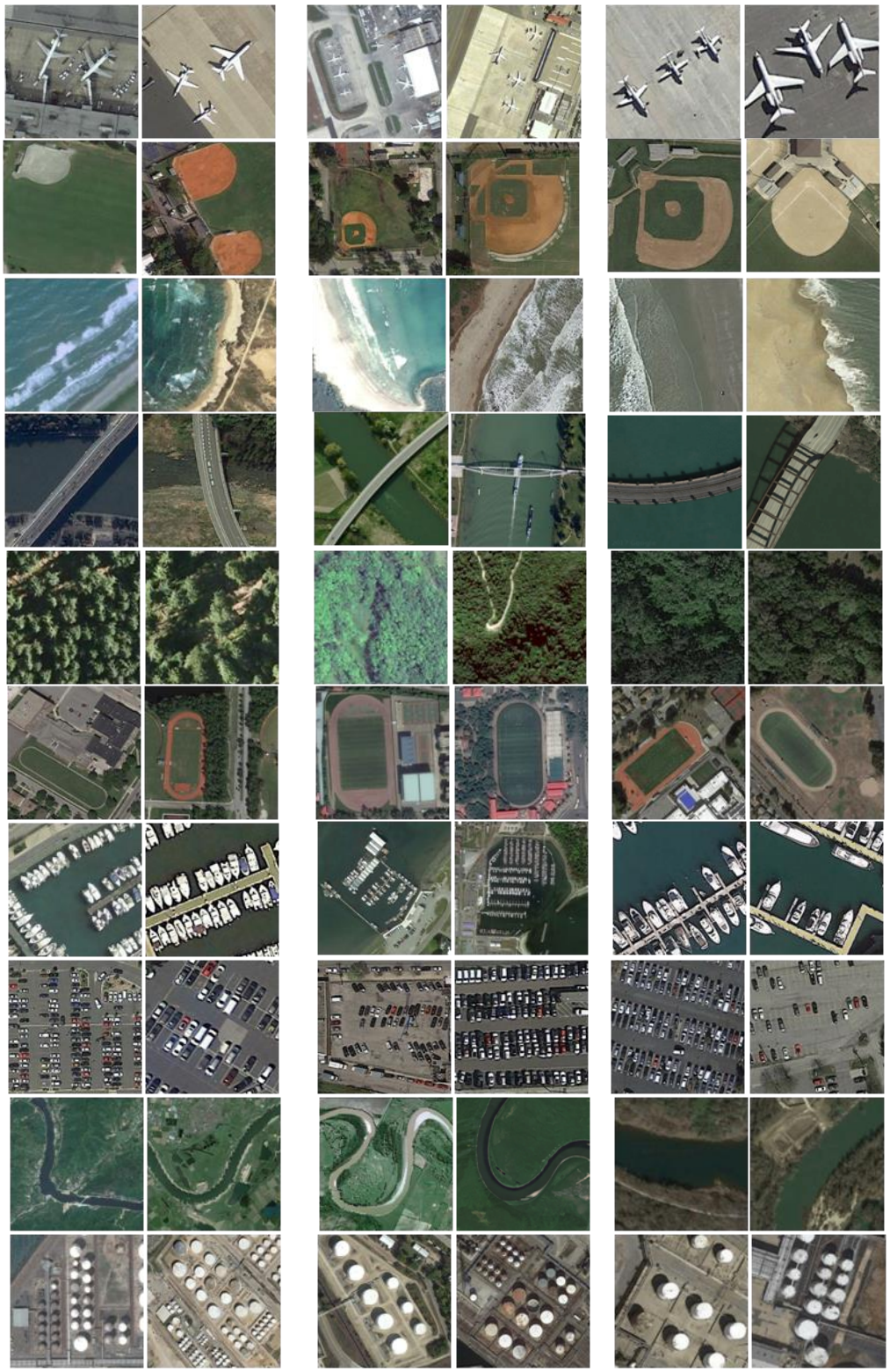
3.4 Fully Connected Layer

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 to multiplies the connection weights and add a bias value. The output of final fully

connected layers 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.



(a) (b) (c)

Figure 8: The sample images from three benchmark datasets. (a) NWPU 45-class dataset (b) AID Dataset (c) PatternNet

4. EXPERIMENTAL RESULTS AND PERFORMANCE METRICS

In this section, we focus on performance and effectiveness of RSI scene classification based on Dilated CNN model and traditional CNN model under the same parameters and conditions. 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 proposed D-CNN with different dilation rates. The proposed dilated convolutional network model is developed on Python and Anaconda IDE tools.

4.1 Dataset Descriptions

For experimental evaluation, we have used three publicly available benchmark datasets for remote sensing image scene classification. The first dataset is Aerial Image Dataset[27] which contains 30 classes and totally 10 000 images. Each class ranges from 220-420 images with resolution of 600 × 600 pixels. The spatial resolution of images ranges from 0.5 to 8m. The second 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 256 × 256 pixels. 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. The final dataset is PatternNet[28] which contains 38 classes and totally 30 400 images. Each class consists of 800 images with resolution 256 × 256 pixels. The spatial resolution of images ranges from 0.062 to 4.69m. For our proposed work, we have chosen common classes namely airplane, baseball diamond, beach, bridge, forest, ground track, harbor, parking lot, river and storage tank for remote sensing image scene classification from the three benchmark datasets. Sample images from three benchmark datasets are shown in Figure 8.

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 datas 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):

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

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

Figure 9: Concepts of confusion matrix

The working principle of confusion matrix processes is shown in figure 9. 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 deal with confusion matrix namely positive class and negative class. These two common classes can be further divided into four categories. True Positive is an outcome, where the model that has correct classification of positive example. False Negative is an outcome, where the model that has incorrect classification of positive examples. False Positive is an outcome, where the model that has incorrect classification of positive examples. True Negative is an outcome, where the model that has correct classification of negative examples.

4.3 Experimental Results of Traditional CNN and Dilated CNN

The proposed model was trained and tested with three benchmark datasets using tensor flow in Core i7 CPU 2.6GHz, 1 TB of Hard Disk and 16 GB of RAM. The proposed dilated convolutional models, experimental results are compared with traditional CNN model with same parameter and configurations. Like traditional CNN model, the dilated convolutional with dilation rate of 2 is as it same as parameter and configuration but, the receptive field is increased as 5×5. So, in order to increases of receptive field is incorporated with more relevant information and increase the performance of proposed model as well as reduce the computational time. Similarly, the dilated convolutional with dilation rate of 3 is as it same as parameter and configuration in traditional CNN model but, the receptive field is increased as 7×7.

Table 1: Performance metrics of NWPU 45-class dataset

S. No.	Dataset	Model	Accuracy	Precision	Recall	F1-Score
1	NWPU 45-class	Traditional CNN	85.85	86.21	85.86	85.73
2		Dilated CNN-1	87.71	88.35	87.71	87.42
3		Dilated CNN-2	89.85	89.49	89.43	89.37

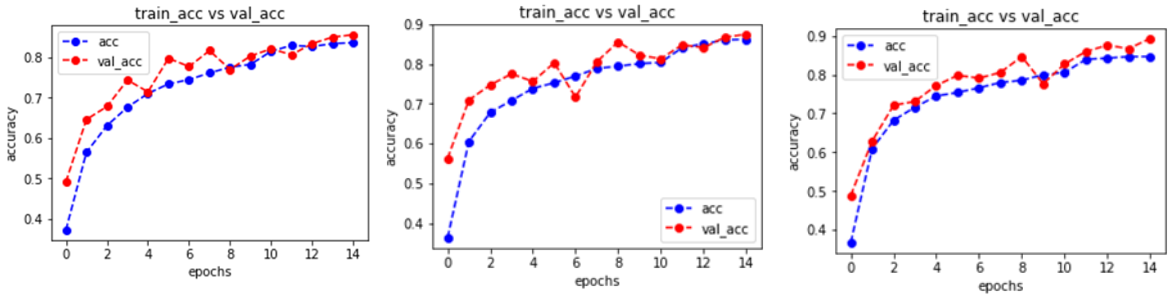


Figure 10: Classification accuracy for NWPU 45-class dataset

To evaluate the effectiveness of traditional CNN and Dilated CNN for RSI scene classification, we have conducted experiments on three datasets. These three datasets contain different spatial and spectral information. We compare the performance of traditional CNN model, Dilated CNN-1 and Dilated CNN-2 with the state-of-the-art results in these three datasets and performance metrics are summarized in Table 1, 2 and 3 respectively. The experimental setting of traditional CNN and Dilated CNN model consists of three convolutional layer and max pooling. For avoiding the problem of overfitting concepts, we have used dropout and Adam optimizers. Figure 10. shows the performance metrics of NWPU 45- class dataset (traditional CNN, dilated CNN-1 and Dilated CNN-2 model) for 15 epochs on both training and validation data. In first experimental set, Dilated CNN-2 model has highest performance accuracy because the receptive fields are increased than other two models. Figure 11 shows performance metrics of AID dataset.

Table 2: Performance metrics of Aerial Image Dataset

S. No.	Dataset	Model	Accuracy	Precision	Recall	F1-Score
1	AID Dataset	Traditional CNN	90.88	91.39	90.88	90.69
2		Dilated CNN-1	91.76	92.27	91.76	91.61
3		Dilated CNN-2	92.35	92.66	92.35	92.28

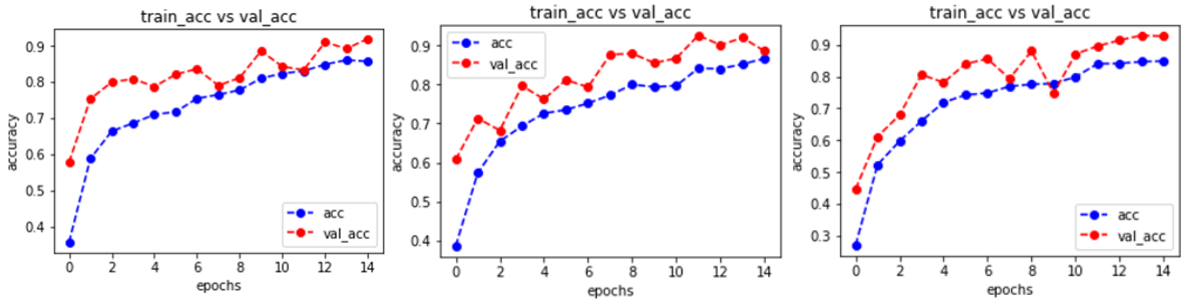


Figure 11: Classification accuracy of AID Dataset

In third experiment set, we have trained pattern net dataset for traditional CNN, dilated CNN-1 and dilated CNN-2 model. Figure 12 shows performance metrics of PatternNet dataset.

Table 3: Performance metrics of PatternNet Dataset

S. No.	Dataset	Model	Accuracy	Precision	Recall	F1-Score
1	PaternNet Dataset	Traditional CNN	94.85	95.44	94.86	94.89
2		Dilated CNN-1	96.85	97.17	96.86	96.89
3		Dilated CNN-2	97.18	97.41	97	96.98

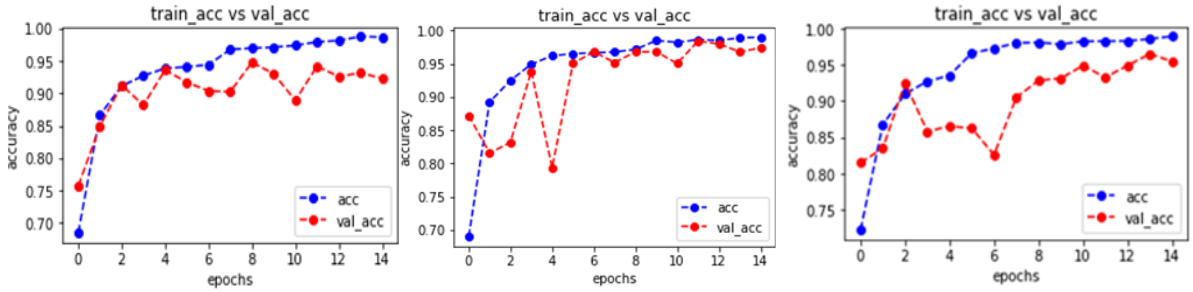


Figure 12: Classification accuracy of PaternNet Dataset

4.4 Performance Comparison of Proposed Model with Traditional CNN

Based on experimental results show that, the dilated-1 CNN model and dilated-2 CNN model have higher accuracy than traditional CNN model and also computation time is less compared to traditional model. The dilated CNN-1model has 2% higher than the traditional CNN model. Similarly, dilated CNN-2 model has 3% higher than the traditional model. The performance comparison of proposed model is shown in figure 13.

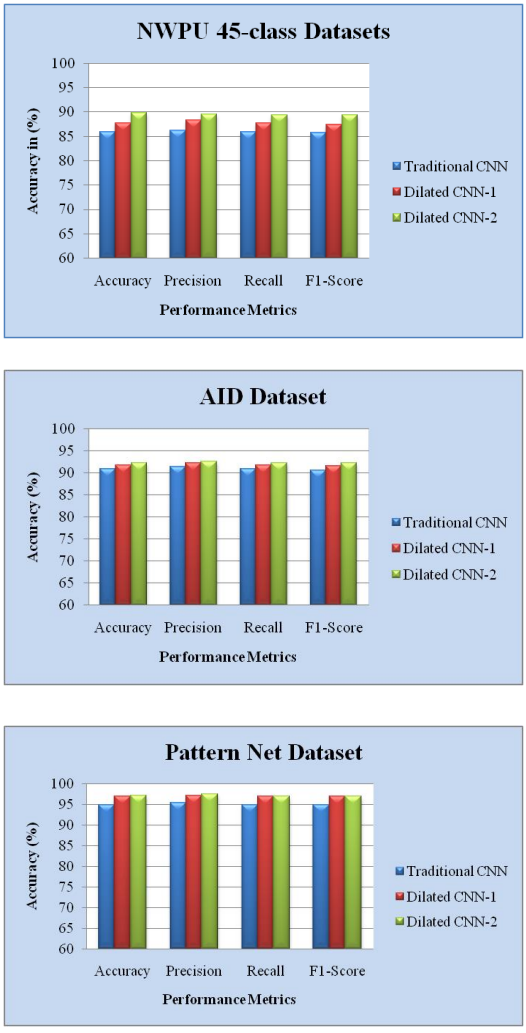


Figure 13: Performance analysis of proposed model with traditional model

5. CONCLUSION

In this paper, we have proposed a Dilated Convolutioanl Neural Network model for remote sensing image scene classification by replacing the kernel of traditional CNN with dilated convolutional kernel. We have trained dilated convolutional models for remote sensing image scene classification for three benchmarks datasets namely NWPU 45-class, PatternNet and Aerial Image Dataset with dilation rate of 2 & 3 and achieved better performance than traditional CNN. In future, we have planned to incorporate our proposed dilated convolutional neural network model in remote sensing object detection system and by implementing 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: 166–177, 2019.
2. Deepan, P., and Sudha, L.R., **Object Detection in Remote Sensing Aerial Images: A Review**, International Journal of Scientific Research in Computer Science Applications and Management Studies: 1-8, 2018.
3. Yu, H., Yang, W., Xia, G.S., and Liu, G., **A color-texture-structure descriptor for high resolution satellite image classification**, Journal of Remote Sensing: 259-269, 2016. https://doi.org/10.3390/rs8030259
4. Cheng, G., Han, J., and Lu, X., **Remote Sensing Image Scene Classification: Benchmark and State of the Art**, Proceedings of the IEEE: 1-19, 2017.
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: 2784-2817, 2018.
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: 372–378, 2014. https://doi.org/10.1109/SAI.2014.6918213
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, 22–40, 2016.

8. Lecun, Y., Bottou, L., Bengio, Y., and Haffner, P., **Gradient-based learning applied to document recognition**, Proceedings of the IEEE, vol. 86, 2278–2324, 2015.
9. O'Shea, K., and Nash, R., **An Introduction to Convolutional Neural Networks**, International Journal of Computer Vision and Pattern Recognition: 2-11, 2015.
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: 1-5, 2019. https://doi.org/10.1109/LGRS.2019.2894399
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: 3677-3683, 2018.
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 Engineerin:1-6, 2019.
13. Koh, C., Chang, J., Tai, C., and Huang, D., Hsieh, H., and Liu, Y., **Bird Sound Classification using Convolutional Neural Networks,** International Journal of computer vision: 1-10, 2019.
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: 38495- 38506, 2019.
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) : 1-10, 2017.
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, Vol. 3(12): 189-201, 2019. http://doi.org/10.26782/jmcms.2019.10.00015.
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): 1-6, 2018.
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:1-5, 2019. https://doi.org/10.1109/LGRS.2019.2902675
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: 1-11, 2018.
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: 1-24, 2017.
21. Zhang, W., Tang, P., and Zhao, L., **Remote Sensing Image Scene Classification Using CNN-CapsNet**, Remote Sens.: 1-22, 2019.
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 Photogrammetry and Remote Sensing: 208–218, 2018.
23. Gallego, A., Pertusa, A., and Gil, P., **Automatic Ship Classification from Optical Aerial Images with Convolutional Neural Networks**, MDPI Journal of remote sensing: 1-20, 2018. https://doi.org/10.3390/rs10040511
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): 1-13, 2018.
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): 1-5, 2017.
26. 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: 107-120, 2020. https://doi.org/10.1016/B978-0-12-816385-6.00008-8.
27. Gui-Song, X., Jingwen, H., Fan, H., Baoguang, S., Xiang, B., Yanfei, Z., and Liangpei, Z., **AID: A Benchmark Data Set for Performance Evaluation of Aerial Scene Classification**, IEEE Transactions on Geoscience and Remote Sensing: pp.1-17, 2017.
28. Weixun, Z., Shawn, N., Congmin, L., and Zhenfeng, S., **PatternNet: A benchmark dataset for performance evaluation of remote sensing image retrieval**, ISPRS Journal of Photo and Remote Sen., pp. 1-13, 2018.
29. Simonyan, K., and Zisserman, A., **Very deep convolutional networks for large-scale image recognition**, in: Proceedings of the International Conference on Learning Representations (ICLR): 1-8, 2015.
30. Prasad , M.V.D., Syed Inthiyaz and Teja kiran kumar, M., **Human activity recognition using Deep learning,** International Journal of Emerging Trends in Engineering Research, Volume 7, No. 11, 2019. https://doi.org/10.30534/ijeter/2019/227112019
31. Alvin S.A., Cherry D. C., and Mon Arjay F., **A YOLOv3 Inference Approach for Student Attendance Face Recognition System,** International Journal of Emerging Trends in Engineering Research, Volume 8, No. 2, 2020. https://doi.org/10.30534/ijeter/2020/24822020