ISSN (Online) 2394-2320

International Journal of Engineering Research in Computer Science and Engineering (IJERCSE)

Volume 11 Issue 3 March 2024

Study of Deep Learning Pre-Trained Models: VGG-19, Inception-V3, Densenet-169 for Remote Sensing Image Classification Utilizing Transfer Learning

[1] Nisha Gupta, [2] Ajay Mittal, [3] Satvir Singh

[1] MRSPTU, Bathinda
[2] Aryabhatta Group of Institutes, Barnala
[3] IKGPTU, Jalandhar
Corresponding Author Email: [1] nishabs2019gmail.com, [2] mittalajay@gmail.com, [3] drsatvir.in@gmail.com

Abstract*— Image Classification accuracy results are strongly based upon feature extraction methods adopted. Feature extraction is the dimensionality reduction process that efficiently represents only the meaningful parts of the im- age as a comparative lower dimensional feature vector. Traditional methods could not produce optimum results in case of remote sensing images due to much more complexity of remote sensing images as compare to normal images. Classification becomes tedious for remote sensing images as level of abstraction converges from pixel to objects. Thus traditional methods encoding color, texture and shape features proved to be inefficient for classify- ing complex remote sensing images. Scene categorization requires high level of features that are comparatively more representational than local and global features based on color, shape and texture. These traditional feature extraction approaches were drifted towards convolutional neural networks as these networks efficiently extracted abstract features. Pre-trained models employing transfer learning could produce superior results as compared to earlier traditional machine learning as well as neural networks based methods used for feature extraction and classification. Pre-trained models utilizes transfer learning in which learned features in form of weights for one general task is used for extracting features of a particularly specific task. Thus, transfer learning speeds up the training phase. This letter is to enrich the accuracy of image classification employing transfer learning with study of deep learning models models VGG-19, Inception-v3 and DenseNet-169.

Index Terms*— Image Classification, Feature extraction, Pre-trained models, Transfer Learning.

I. INTRODUCTION

Remote sensing image classification is the efficient execution of image categorization of high spatial resolution images for large remote sensing archives [1]. High performance of image categorization is directly based upon efficient image feature extraction. Before widespread adoption of deep learning in field of remote sensing image classification the feature extraction stage relied on man- ually designed low level features mainly focusing on basic features color, shape and texture [2]. However, these tra- ditional handcrafted methods due to its lower perfor- mance were eventually replaced by convolutional neu- ral networks[3] which efficiently extracted abstract fea- tures. In the domain of remote sensing prominent re- sults in image classification could be achieved through convolution neural networks combined with transfer learning [4]. Categorization of scene images into a distinct set of meaningful groups based on content or features of the images is significant in domain of remote sensing [5]. Remote sensing has wide range of applications in vari- ous fields like identifying natural hazards, recognizing spatial objects, retrieving geographical images, creating vegetation maps, monitoring the environment and contributing to urban planning and many more.

II. RELATED LITERATURE SURVEY

A neural network trained on large dataset gains knowl- edge from this data and this acquired knowledge termed as weights of the network [6]. Only the learned features in the form of weights can be extracted then transferred to any other neural network instead of training that neu- ral network from the initial stage. Instead of building a model from starting point pre-trained models are already trained on large dataset are used as a feature extractor by removing output layer and using entire network as a fixed feature extractor by freezing weights of initial lay- ers whereas retraining only higher layers for new problem specific dataset [7]. Correct weights for network are iden- tified for the network by multiple forward and backward iterations. The weights and architecture acquired by pre- trained models previously trained on huge datasets are used directly and apply the learned weights on our tar- get problem known as transfer learning. Fine tuning is the most important phase of transfer learning as the ex- perimental dataset is small and selected images are differs from various images in source domain. Model is fine-tuned by freezing the initial layers of pre-trained model [8]. Fine tuning of pre-trained models involves convolution operators learn broad characteristics in first layer and then move to features more unique to the dataset. Model is trained in final layer. The early and central layers of neu- ral networks are used for transfer learning. It utilizes the labelled data from task it was initially trained on and only the last layers are retrained. Various researchers have used pre-trained models for image classification. Huang, Gao and Liu, Zhuang and Van Der Maaten, Lau- rens and Weinberger, Kilian [9] introduced the DenseNet (dense convolutional network) which links each and ev- ery layer to the other layers in a feed-forward fashion. DenseNets solves the vanishing-gradient problem. Author has achieved accuracy of 97.44% on AID, 99.50% on UC- Merced and 95.89% on Optimal and 94.98% on NWPU- RESIS45 dataset. Zhang, Jianming and Lu, Chaoquan and Li, Xudong and Kim, Hye-Jin and Wang, Jin [10].The authors have proposed DenseNet pre-trained model for re- mote sensing image classification. Experiments are done on AID (Aerial Image Dataset) dataset, UCM dataset, NWPU-RESISC45 dataset, and Optimal-31 dataset. The author has achieved the accuracy of 98.67% (50% train- ing ratio), 99.50% (80% training ratio) on UCM dataset and 95.37% (20% training ratio), 97.19% (50% train- ing ratio) on AID dataset and 95.41% (80% training ratio) on optimal-31 and 92.90 (10% training ratio), 94.95 (20% training ratio) on NWPU-RESISC45 dataset. Thirumaladevi, S and Swamy, K Veera and Sailaja, M [11] Image classification is performed by using transfer learning using pre-trained AlexNet and variants of Visual Geometry Group (VGG) networks VGG-16 and VGG-19. Experiments are performed using UCM and SIRI-WHU dataset. Author has achieved accuracy of 93.57% on UCM and 91.34 on SIRI-WHU using AlexNet. Further accuracy achieved is 94.08% on UCM and 92.78% on SIRI-WHU using VGG-16 and 95% accuracy achieved on UCM and 93.4% accuracy achieved on SIRI-WHU using VGG-19 pre-trained model. Tan, Pooi Shiang and Lim, Kian Ming and Tan, Cheah Heng and Lee, Chin Poo[12] has proposed DenseNet-121 deep learning pre-trained using transfer learning to extract significant features of the image. Com- putation resources are greatly reduced using DenseNet. Three benchmark datasets: soundscapes1, soundscapes2 and urbansound8k are used for experimentation. The pro- posed pre-trained model DenseNet-121 with multilayer perceptron outperforms existing works on soundscapes1, soundscapes2 and urbansound8k datasets with the F1- scores of 80.7%, 87.3%, and 69.6%, respectively.

III. PRE-TRAINED MODEL VGG

VGG is expansion for Visual Geometry Group. It is so termed as it was created by VGG group by K. Simonyan and A. Zisserman at University of Oxford in 2012. VGG model could score 92.7% top-5 test accuracy in ImageNet. It was the great improvement over AlexNet by VGG. Variant of the VGG architecture are VGG-11, VGG-16 and VGG-19. VGG-16 has 16 weight layers in convolution layer. VGG-19 has 19 layers additional three convolution layers than VGG-16. These additional three layers are ca- pable to extract more abstract features of remote sensing images[13].

A. Layered Structure Of VGG-19

VGG is a family of Convolution Neural Networks.VGG-19 constitutes total 19 layers First 16 convolution layers are used for feature extraction. Next 3 fully connected layers are used for classification[14].

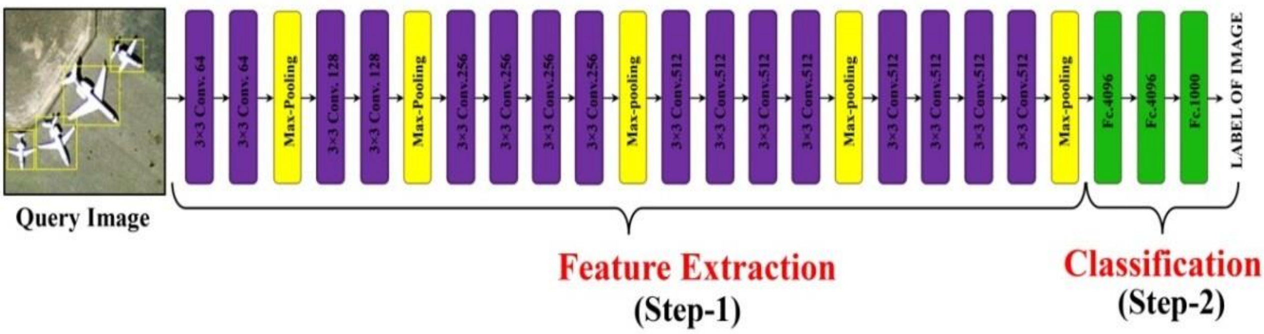

Fig. 1. VGG-19 Feature Extraction and Classification Layers

Input image is passed through weight layers of VGG-19 after pre-processing. Training image samples go through a pile of convolution layers. The 19 layers of VGG include:VGG-19 uses multiple 3×3 filters in every convolutional layer. First 16 convolutional layers are used for feature extraction and next 3 layers work for classification [15]. Fully Connected layers involve 4096 channels. Each channel follows fully connected layer with 1000 channels for predicting 1000 labels.At the last, fully connected layer has softmax layer for classification. VGG- 19 is trained on the large scale ImageNet database. Im- ageNet is the collection of millions of images of 1000 categories[16]. This whole dimensionality reduction pro- cess is presented in fig2.

ISSN (Online) 2394-2320

International Journal of Engineering Research in Computer Science and Engineering (IJERCSE)

Volume 11 Issue 3 March 2024

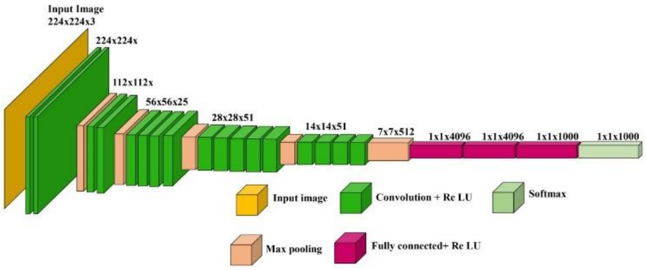

Fig. 2. VGG-19 Model Main Architecture

IV. PRE-TRAINED MODEL INCEPTION-V3

As the CNN's grow deeper by adding multiple convolution layers, the major concerning issue is the problem of over fitting of CNN deep neural networks. To resolve this research problem Inception or GoogleNet, pre-trained model came into existence. Inception-V3 was developed by the team of Google. It was introduced as GoogleNet- 2014.Inception-v3 model proved to better as compared to previous CNN on ImageNet LargeScale Visual Recogni-tion challenge (ILSVRC) benchmark having lesser error rate than its predecessors. Inception-v3 is a CNN, deep learning pre-trained model used for image classification.

A. Inception-v3 Blocks

It is third version of Inception network architecture introduced in 2015 by Google. It uses a new Inception module known as Factorization based Inception module. Factorization technique reduces number of parameters in neural network [17]. Inception-v3 also uses batch normalization layer that is placed after each convolution layer to improve performance. Inception -v3 is the successor of Inception-v1(naive form) model [18]. It is having 1×1 convolutional layer before each convolution layer. Replacing 5×5 convolution by 1×1 inculcates reduced dimensions and faster computations in the network as shown in fig3.

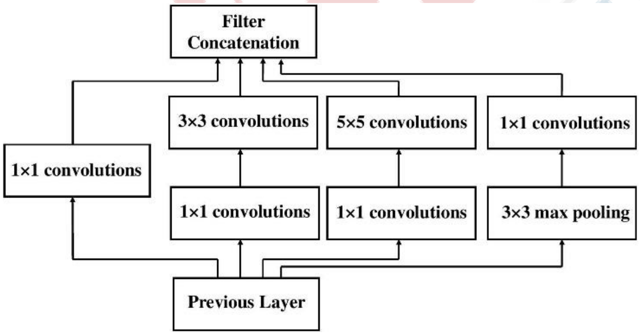

Fig. 3. Inception-v3

Inception-v3 model has higher efficiency as cheaper in computations as it uses an auxiliary classifier. After employing these four improvements the Inception-v3 was introduced as the advanced version of Inception-v1 and Inception-v2.

Inception-v3 is made up of both symmetric and asymmetric blocks. It includes convolutions, average pooling, max pooling, concatenations, dropouts and fully con- nected layers. Batch normalization is applied for acti- vating inputs to make the network stable. Loss is com- puted using Softmax at the last layer. After performing

a. Factorization into Smaller Convolutions
b. Spatial Factorization into Asymmetric Convolutions
c. Utility of Auxiliary Classifiers
d. Efficient Grid Size Reduction

All the above discussed improvements, the final Inception V3 model is represented in Inception V3 model has 21 layers higher. The efficiency of Inception-v3 is highly im- proved than its older versions. The output size of each module is the input size of the next module as shown in fig4

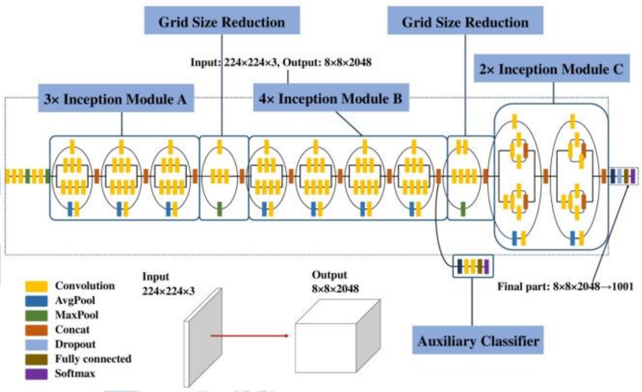

Fig. 4. Inception-v3

V. PRE-TRAINED MODEL DENSENET

In DenseNet architecture each layer is in direct connection with every other layer thus named as Densely Connected Network. For L number of layers, DenseNet is having L(L+1)/2 direct and shorter connections between layers close to input and output proves to be much efficient to train and achieves much more accuracy [19]. DenseNet is a feed forward network in which each layer is directly con-nected to the front layers [73].Input of ithlayer can be the output of (i-1)th layer and (i-2)th and (i-n)th layer. Direct connection of each layer in the network uses Batch nor- malization (BN) to normalize the input of each layer that reduces the absolute difference between data.Connection between all the layers ensures the maximum information flow. Additional inputs from all preceding layers passes on its features maps all other subsequent layers [20]. The framework of DenseNet is represented in fig5.

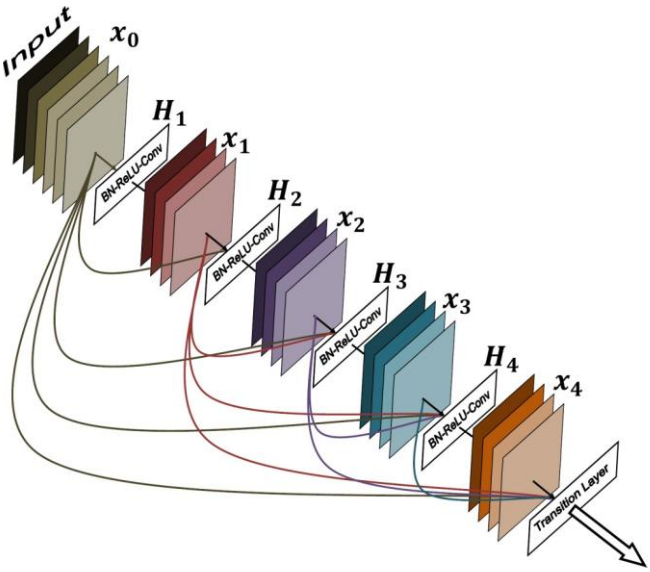

Fig. 5. DenseNet

VI. CONCLUSION

Pre-trained models saves time and computational resources and offers best features learned from large datasets. These pre-trained models are already trained on ImageNet database having more than millions of images and can classify into 1000 object categories. Pre-trained neural network employing transfer learning is much faster than training a neural network from scratch. Pre-trained models differ with respect to its prevalent characteristics that contribute while selecting a neural network to ap ply to a particular domain. The main characteristics of pre-trained neural networks are accuracy, speed and size. Selecting a particular neural network for a particular ap- plication is a tradeoff between these main characteristics.

REFERENCES

[1] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton. Imagenet classification with deep convolutional neural networks. Advances in neural information processing systems, 25, 2012.

[2] Mengluan Huang, Le Dong, Weisheng Dong, and Guangming Shi. Supervised contrastive learning-based deep hash retrieval for remote sensing image. In IGARSS 2022-2022 IEEE International Geoscience and Remote Sensing Symposium, pages 1512–1515. IEEE, 2022.

[3] Weixun Zhou, Shawn Newsam, Congmin Li, and Zhenfeng Shao. Learning low dimensional convolutional neural networks for high-resolution remote sensing image retrieval. Remote Sensing, 9(5):489, 2017.

[4] Yuze Wang, Rong Xiao, Ji Qi, and Chao Tao. Cross-sensor remote-sensing images scene understanding based on transfer learning between heterogeneous networks. IEEE Geoscience and Remote Sensing Letters, 19:1–5, 2021.

[5] Zhiqi Zhang, Wen Lu, Xiaoxiao Feng, Jinshan Cao, and Guangqi Xie. A discriminative feature learning approach with distinguishable distance metrics for remote sensing image classification and retrieval. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 16:889–901, 2022.

[6] Yishu Liu, Liwang Ding, Conghui Chen, and Yingbin Liu. Similarity-based unsupervised deep transfer learning for remote sensing image retrieval. IEEE Transactions on Geoscience and Remote Sensing, 58(11):7872–7889, 2020.

[7] Liangli Zhen, Peng Hu, Xi Peng, Rick Siow Mong Goh, and Joey Tianyi Zhou. Deep multimodal transfer learning for cross-modal retrieval. IEEE Transactions on Neural Networks and Learning Systems, 33(2):798–810, 2020.

[8] Abebaw Alem and Shailender Kumar. Deep learning models performance evaluations for remote sensed image classification. IEEE Access, 10:111784–111793, 2022.

[9] Gao Huang, Zhuang Liu, Laurens Van Der Maaten, and Kilian Q Weinberger. Densely connected convolutional networks. In Proceedings of the IEEE conference on computer vision and pattern recognition, pages 4700–4708, 2017.

[10] Jianming Zhang, Chaoquan Lu, Xudong Li, Hye-Jin Kim, and Jin Wang. A full convolutional network based on densenet for remote sensing scene classification. Mathematical Biosciences and Engineering, 16(5):3345–3367, 2019.

[11] S Thirumaladevi, K Veera Swamy, and M Sailaja. Remote sensing image scene classification by transfer learning to augment the accuracy. Measurement: Sensors, 25:100645, 2023.

[12] Pooi Shiang Tan, Kian Ming Lim, Cheah Heng Tan, and Chin Poo Lee. Pre-trained densenet-121 with multilayer perceptron for acoustic event classification. IAENG International Journal of Computer Science, 50(1), 2023.

[13] Chiranjibi Sitaula and Mohammad Belayet Hossain. Attention-based vgg-16 model for covid-19 chest x-ray image classi- fication. Applied Intelligence, 51:2850–2863, 2021.

[14] A Victor Ikechukwu, S Murali, R Deepu, and RC Shivamurthy. Resnet-50 vs vgg-19 vs training from scratch: A comparative analysis of the segmentation and classification of pneumonia from chest x-ray images. Global Transitions Proceedings, 2(2):375–381, 2021.

[15] Srigiri Krishnapriya and Yepuganti Karuna. Pre-trained deep learning models for brain mri image classification. Frontiers in Human Neuroscience, 17:1150120, 2023.

[16] Dimitrios Marmanis, Mihai Datcu, Thomas Esch, and Uwe Stilla. Deep learning earth observation classification using imagenet pretrained networks. IEEE Geoscience and Remote Sensing Letters, 13(1):105–109, 2015.

[17] Cheng Wang, Delei Chen, Lin Hao, Xuebo Liu, Yu Zeng, Jianwei Chen, and Guokai Zhang. Pulmonary image classification based on inception-v3 transfer learning model. IEEE Access, 7:146533–146541, 2019.

[18] Debanjan Pathak and USN Raju. Content-based image retrieval using feature-fusion of groupnormalized-inception-darknet-53 features and handcraft features. Optik, 246:167754, 2021.

[19] Guoqing Li, Meng Zhang, Jiaojie Li, Feng Lv, and Guodong Tong. Efficient densely connected convolutional neural networks. Pattern Recognition, 109:107610, 2021.

[20] Farsana Salim, Faisal Saeed, Shadi Basurra, Sultan Noman Qasem, and Tawfik Al-Hadhrami. Densenet-201 and xception pre-trained deep learning models for fruit recognition. Electronics, 12(14):3132, 2023.