International Journal For Technological Research In Engineering
Volume 6, Issue 7, March-2019 ISSN (Online): 2347 - 4718

SCENE IMAGE CATEGORIZATION USING DEEP LEARNING CONVOLUTIONAL NEURAL NETWORK

A. Sudalaimathi[1], Dr. V.P.M.B. Aarthi[2], M. Ramesh[3]
[1]PG Student, [2,3]Assistant Professor
[1,2]Department of ECE, VPMM Engineering College for Women, Krishnankoil, India
[3]Department of ECE, Kamaraj College of Engineering & Tech, Virudhunagar, India

Abstract: In Image processing classification results in one of the predefined classes. The machine learning algorithms use these extracted features to classify the images. It has become very difficult task to classify the images into interpretative classes. At present, the focus is on deep learning techniques for feature extraction and classification of images. Convolutional Neural Networks (CNN) in the recent past due to its accuracy in image retrieval and categorizing of images. Convolutional Neural Networks is a process which does a function on an image and gives another image. The experimental observation shows that the performance of fine tuned models using pretrained configuration on the self-collected dataset is far better than directly using trained models. In this paper, AlexNet under the Deep Learning Algorithm for the Outdoor scene image classification. The technique holds good because of its simplicity and efficient usage of trained data set. Further it provides us with a much accurate results and proper labeling with reduced computation time. It can be noted that as it does not require any kind of manual annotations it increase the scalability by our technique.
Index Terms: Classification; Deep Learning; AlexNet; Convolutional Neural Network;

I. INTRODUCTION

The machine learning algorithm that is capable of localizing and classifying the objects from image and video frames is an interesting topic in artificial intelligence (AI), especially in computer vision. The long-time research in AI focuses on how to make a computer to work like a human. The power of AI and machine learning is automating the substantial number of computing tasks. The focus is to develop an algorithm that can teach the computer how to recognize and track an object like a human or better than a human. The demand of the deep neural network (DNN) based approach for real-time object detection is increasing rapidly. Deep Learning is an artificial intelligence function that mimics the neural networks of the human brain. Deep learning technique focus on the application of multi-label and multi-class image classification [1]. A sample from the training data is randomly selected and provide to the inputs of the network, which computes the outputs on a layer-by-layer basis until the output layer is attained [4]. The softmax classifier categorizes an input image into one of the classes using in trained dataset [2]. AlexNet is faster training effect in large data conditions by using two GPUs [3]. Object detection on the single class dataset is comparatively less challenging than in multiclass. This project focus with a more challenging and demanding task of object detection by collecting a multi-class object dataset from outdoor premises, train existing object detection frameworks and test the performance of trained detectors on the self-collected dataset implemented using deep learning CNN for outdoor scene image classification.

II. SYSTEM MODEL

a) ALEXNET

AlexNet revolutionized the state-of-the-art in object recognition. The architecture of AlexNet is shown in figure 1.1 comprises of 8 layers among which 5 are convolutional layers and 3 are fully connected layers. The size of input RGB images is 227 x 227 x 3 with 154587 pixels.

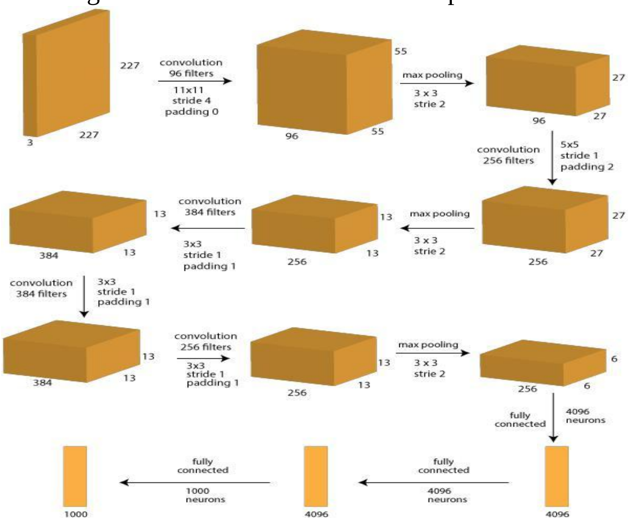

Figure 1.1 AlexNet Architecture

Convolution layer acts as a Feature Extractor that extract features of the inputs such as edges, corners and endpoints. The spatial size of the image is reduced by Pooling layer. It is otherwise named as a down sampling layer. Non linearity or Rectified Linear Unit layer is the activation functions for both feed forward and back propagation. Fully connected layer have full connections to all activations in the previous layer acts as classifier. A Softmax Layer is quite simply, that generates a softmax distribution over all of the outputs. This is commonly used in multi-class classification to generate probabilities of the instance belonging to each class. Two Graphics Processing Unit (GPU) can directly read and write both memories.

b) Accuracy Metrics

In machine learning, precision is the fraction of retrieved items over all items that are present. In object detection case, precision is the sum of the correctly detected object divided by the total population of the object that is detected by the detector. Precision takes the entire detected object into account.

$$Precision = \frac{(relevant\ objects) \cap (retrieved\ objects) \in [0,1]}{retrieved\ \ objects}$$

To understand the terms used in performance measurement, it is wise to consider the binary classification problem. The output of the classifier is positive or negative based on whether it is classified correctly or not. The detection of the single object can be considered as classification task. The clear and concise way to understand the idea behind the binary classification accuracy measure is using the confusion matrix.

Figure 1.2 Training and Categorization of Images
(a). Input sample images
(b). Categorized images based on transferring of weights from pre-trained CNN

The proposed method transfers the learning from pre-trained neural net (AlexNet) and replaces the last three layers with new fully connected layer by training them with new set of images. The figure 1.2 shows that the total number of images considered for training is 16 the timing report of 70% and 30% for training and testing respectively.

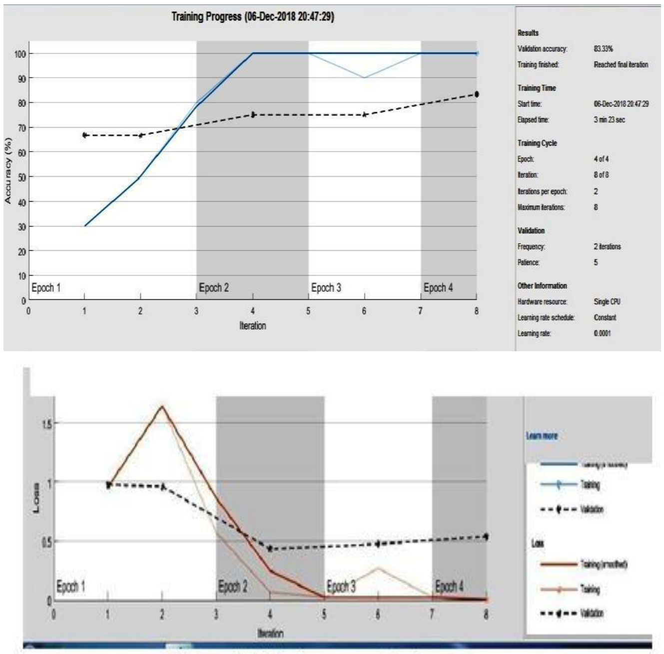

Figure 1.3 *Performance Analysis*

a) *Accuracy vs Iteration graph* ***b)*** *Loss vs Iteration graph*

The figure 1.3 (a) shows the plot between Accuracy and Iteration followed by plot between Loss and Iteration in figure 1.3 (b), continuous lines they represent the training period and dotted lines represent the validation accuracy and loss. The resulting accuracy measurement shows 83.3%.

c) CAFFE MODEL

Convolution Architecture For Feature Extraction (CAFFE) is a Deep Learning Framework developed by the Berkeley Vision and Learning Center (BVLC). CAFFE model is a fast and well tested code with seamless switch between CPU and GPU. For the given input image this model will identify the scene type of the image among the existing 205 classes. There are 4 steps in training a CNN using CAFFE models including data preparation, model definition, solver definition and model training. As AlexNet CAFFE can accept an input image of specified size it can cover wider image categories up to 1000 object classes. It consists of 25 layers. Each successive layer in AlexNet will increase the accuracy of prediction and widens the object categorization ability. These layers are conv 1, conv 2, conv 3, conv 4, conv 5, pool 1, pool 2, relu 1, relu 2, fc 8 etc,. The last layer classifies the output and depicts the top 5 probabilities accurately.

III. RESULTS & DISCUSSION

In multi-label classification the Neural Network models have performed best in all evaluation metrics among all the models in spite of having more or less same features given to all the models extracted from Convolutional Neural Networks which gives a good representation of an image into a vector. The proposed AlexNet & CAFFE model under the Deep Learning Algorithm for the Outdoor Scene Categorization is shown in figure 1.4 the river image considered as the input is figure 1.4 (a). This model performs the good evaluation metrics resulting in the maximum probability of classification as river when compared to hot spring, rain forest, creek and valley. Similarly figure 1.4 (b, c, d) shows that the proposed model is applied for Ocean, ice berg, mountain and mountain snowy images and the accurate classification with proper labeling is attained at a reduced computation time.

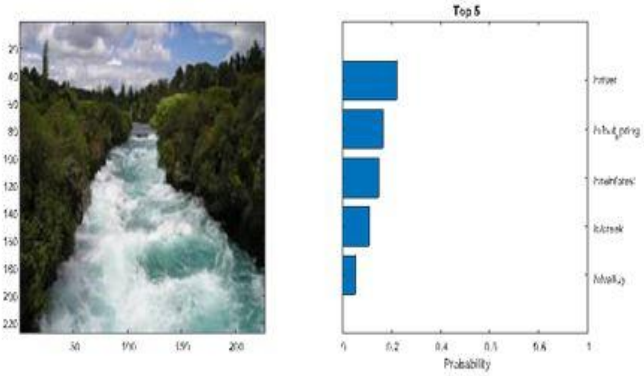

Figure 1.4 (a)

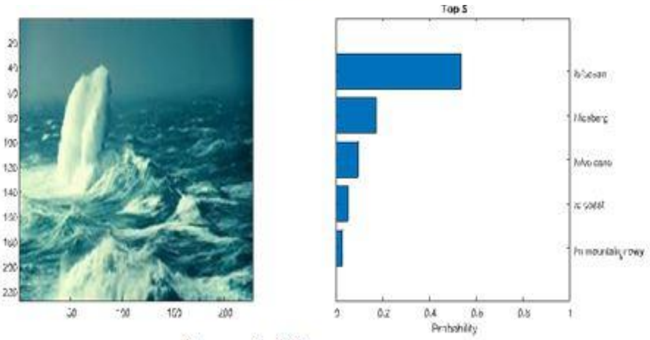

Figure 1.4 (b)

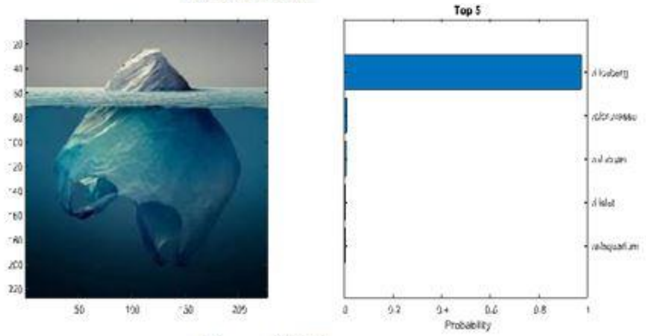

Figure 1.4 (c)

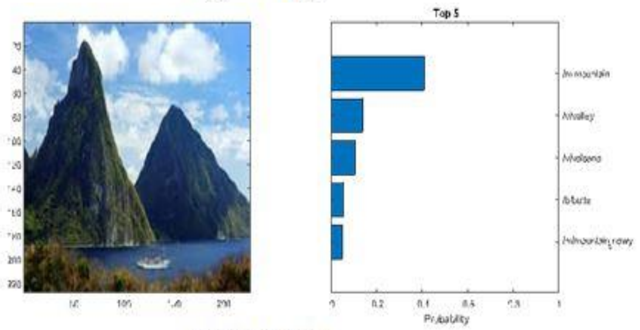

Figure 1.4 (d)

***Figure 1.4 a)** River image **b)** Ocean image*
***c)** Iceberg image **d)** Mountain Image*

The advantage of our approach is that it is simple and it makes the best use of the trained image dataset. It can be noted that as it does not require any kind of manual annotations the proposed technique becomes more scalable. In future the technique can be extended for Indoor scene image classification.

REFERENCES

[1] Aparna R. Rout, Prof. Dr. Sahebrao B. Bagal, “Natural Scene Classification Using Deep Learning” IEEE 10.1109 / ICCUBEA 2017, 8463727, Aug. 2017.
[2] H. M. Bui, M. Lech, E. Cheng, K. Neville and I. S. Burnett, "Object Recognition Using Deep Convolutional Features Transformed by a Recursive Network Structure," in IEEE Access, vol. 4, pp. 10059-10066, 2016.
[3] J. Shao, C. Qu and J. Li, "A performance analysis of convolutional neural network models in SAR target recognition," 2017 SAR in Big Data Era: Models, Methods and Applications (BIGSARDATA), Beijing, 2017, pp. 1-6.
[4] M. A. Mohandes, T. O. Halawani, S. Rehman, and A. A. Hussain, “Support vector machines for wind speed prediction,” Renew. Energy, vol. 29, no. 6, pp. 939–947, 2004.
[5] S. Lekha, Member, IEEE, and M. Suchetha, Member, IEEE “A Novel 1-D Convolution Neural Network With SVM Architecture for Real-Time Detection Applications” Published in: IEEE Sensors Journal (Volume: 18, Issue: 2, Jan.15, 15 2018) Page(s): 724 – 731
[6] Maria Grazia De Giorgi, Antonio Ficarella and ArimeLay-Ekuakille,“CavitationRegime Detection by LS-SVM and ANN With Wavelet decomposition Based on Pressure Sensor Signals” IEEE Sensors Journal (Volume: 15, Issue: 10, Oct. 2015) Page(s): 5701 - 5708.
[7] L. Zhang, D. Zhang, X. Yin, and Y. Liu, “A novel semi-supervised learning approach in artificial olfaction for E-nose application,” IEEE Sensors J., vol. 16, no. 12, pp. 4919–4931, Jun. 2016.
[8] Y. Bar, N. Levy, and L. Wolf, “Classification of artistic styles using binarized features derived from a deep neural network,” in Proc. Eur. Conf. Comput. Vis. Workshops, 2014, pp. 71–84.

