Published by :
http://www.ijert.org

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 9 Issue 07, July-2020

Classification of Natural Scene using Convolution Neural Network

Supriya R Iyer
M Tech Student, Signal Processing,
ECE Dept. GECBH
Thiruvananthapuram, Kerala, India

Rishidas S
Professor, HOD, ECE Dept. GECBH
Thiruvananthapuram,
Kerala,India

***Abstract*— Classification is a process in which objects are recognized, differentiated and understood. Image classification is classifying image into one of the predefined classes. In conventional way, people use different computer vision techniques to extract features from images and different machine learning algorithms use these extracted features to classify the images. Nowadays, accuracy and performance of the model depends mainly on trained dataset and algorithm used. Neural networks are found to be extremely effective in classification of our data. A Convolution Neural Network concept is used. Natural scenes has objects we would ideally want computer to recognize automatically. Object is recognized on the basis of shape and textual characteristics of regions of interest. MATLAB tool is used to classify the images into their classes. There are different classes of natural scenes to be identified into their respective categories. Here we are mainly concerned with 12 different categories. Dataset containing several thousands of images of natural scene is used to train the model. Neural network model used is Alex-Net model. The histogram analysis of images are carried out at each classification of image. The accuracy of the image as well as the loss occurred in the model is also found out. The performance of the model is calculated with the help of confusion matrix which represents true value corresponding to each class.**

Keywords- Supervised learning, Machine learning technique, Convolution Neural Network, Alex-Net Model, Image classification Histogram analysis

I. INTRODUCTION

Classification in machine learning and statistics is a supervised learning approach in which the computer program learns from the data given to it and make new observations or classifications. Classification is a process of categorizing a given set of data into classes. It can be performed on both structured and unstructured data. The process starts with predicting the class of given data points. The classes are often referred to as target, label or categories. In Image Classification, we classify an image into one of the predefined classes or multiple classes at the same time. In Multi Label Image classification, an image can have multiple classes present among the set of classes where as in simple Image classification an image contains only one class among the set of classes.

Generally in supervised learning, an object is represented by a feature vector and it is represented with a class label. Let us assume X as the feature space and Y the set of class labels. Now, the task is to learn a function which is f: X → Y from a given data set. The above method is existing and it is successful, but there are many problems associated with real world where this work does not prove to be correct. A real world problem may be related with a number of instances and labels simultaneously. One of the main difficulties in applying the supervised learning is that we require large amount of training images (trained dataset). It becomes very difficult to label these amount of images as it is expensive and also requires more time. To overcome such situation, there are two different methods for solving such type of problems. The first one is called as the problem transformation methods and other one is algorithm adaptation method. Deep learning model for image classification has recently attracted the attentions. Several algorithms prove its efficiency in image classification. Image classification is one of the most widely studied subject in the field of Machine Learning which has developed many algorithms for it. Convolution Neural Network is one such technique. This work focuses on the application of CNN algorithms for multi-class Image Classification.

II. PROPOSED SYSTEM

CNN is a technique to learn complex relationship or high level features from data. CNN consists of more than one hidden layers. For multi-label image classification we have proposed a Alex-Net model and its performance is found out. For this the datasets used were images of Landscapes such as Mountains, Sunset, Desert, Water, Trees and combination of these landscapes thereby creating a class of 12 categories. These were downloaded from Google and resized accordingly. Total of around 5000 images were taken

In this paper an Alex-Net neural network model has been developed that infers the images and classifies it into the respective classes according to the dataset. This paper presents a new approach to enhance the performance of image classification.

Alex-Net model is one of the accurate and most reliable methods compared to the other methods. This paper presents a new technique to intensify the performance of image classification. Figure 1 shows the block diagram of the proposed method.

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 9 Issue 07, July-2020

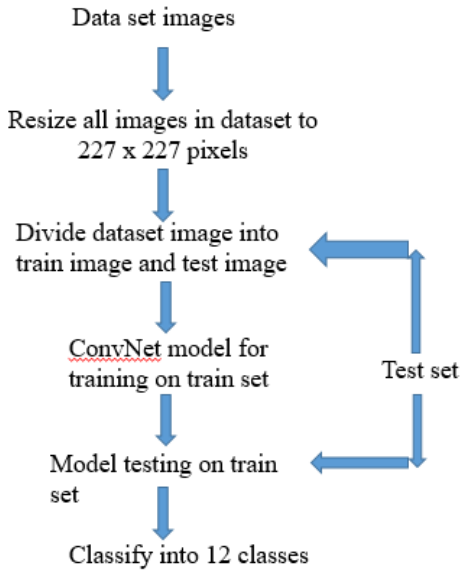

Fig. 1. Block diagram of proposed system

A. *Graphical User Interface Image Classification*

Graphical user interfaces (**GUIs**), provide point-and-click control of your software applications, eliminating the need for others to learn a language or type commands in order to run the application. This is applicable for use within MATLAB and also as standalone desktop or web apps. This is used to create an app user interface by writing the code itself. For added control over design and development, here MATLAB functions are used to define the layout and behavior. In this approach, a figure is acting to serve as the container for user interface and add components to it programmatically. In this work, GUI technique for multi label image classification is used. The combination of ConvNet with the same architecture with GU interface for multi label image classification is implemented. The classified images are stored as data values in matrix form in MATLAB. To represent the image, GUI converts the matrix form into images.

B. *Formation of dataset*

For Multi Class image Classification, dataset consisting of different classes such as Trees, Sunset, sea, desert, mountains, etc. Total of around 5000 images were collected. The tabular form of different products with their number of images taken is shown. Here each product represents a class. The figure 2 shows the sample of images which are used in the natural scene dataset. Before the dataset images are passed through the model foe classification every image is resized to the size acceptable by the model.

C. *Preprocessing*

First resize all images into 227 x 227 because CNN (here Alex-net so 227 x 227 dimension) requires fixed size image as input. And split dataset into 80% and 20% where test set has some examples from each class and training set has images. Resized all images to 227 x 227 pixels and created two sets one is train set and other is test set. The labels for each image will be given different class score to represent it as different classes for easy identification.

Table 1. Dataset of natural scene

Label set	Images
Desert	822
Mountain	689
Sea	601
Sunset	645
Desert Mountain	110
Desert Sunset	276
Desert Sea	139
Trees	733
Desert trees	302
Mountain trees	105
Desert sunset mountain	122
Desert sunset trees	116

Fig. 2. Some sample images for natural scene dataset

III. ARCHITECTURE OVERVIEW

Convolutional Neural Networks are very similar to ordinary Neural Networks. They are made up of neurons that have learnable weights and biases. Each neuron receives some inputs, performs a dot product and optionally follows it with a non-linearity. The whole network takes input as the raw image pixels on one end to class scores at the other. And they still have a loss function (e.g. here it's Softmax) on the last (fully-connected) layer.

ConvNet architectures make the explicit assumption that the inputs are images, which allows us to encode certain properties into the architecture. These then make the forward function more efficient to implement and vastly reduce the amount of parameters in the network.

Neural Networks receive an input (a single vector), and transform it through a series of *hidden layers*. Each hidden layer is made up of a set of neurons, where each neuron is fully connected to all neurons in the previous layer, and where neurons in a single layer function completely independently and do not share any connections. The last fully-connected layer is called the "output layer" and in classification settings it represents the class scores.

For feature extraction we used Convolution neural network (CNN) and for classification we used affine network which is nothing but fully connected neural network (NN). The CNN and NN works together and hence called as ConvNet. The ConvNet is a combination of several Convolutional layers followed by pooling layer and whole followed by affine layer. After data collection the main task is to extract good features from images that is to make 3

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 9 Issue 07, July-2020

dimension tensor into a one dimension tensor for this we used a series of five Convolution Neural Network (CNN) for feature extraction and Affine Network (fully connected network) for the classification purpose. ConvNet architectures make the direct assumption that the inputs are images, which allows to encode certain properties into the architecture. This then make the forward function more efficient to implement and vastly reduce the amount of parameters in the network. The fig. 3 is a ConvNet which stacks the neurons in the form of width, height and depth as visualized in one of the layers. Each layer of a ConvNet converts the 3D input volume to a 3D output volume of neuron activations.

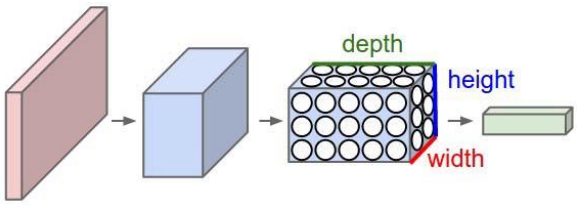

Fig. 3. A ConvNet Model

As mentioned earlier in the above section, a general ConvNet is a series of layers and each layer of a ConvNet transforms one volume of activations to another through a differentiable function. Three different types of layers are used to build ConvNet architectures. These layers are: Convolution Layer, Pooling Layer and Fully Connected Layer. These layers are combined to form a full ConvNet architecture.

- INPUT [227x227x3] holds the unprocessed pixel values of the image, in this situation an image has width of 227, height of 227, and with three colour channels R, G, B with a depth of 3.
- The CONV layer calculates the output of the neurons which are connected to local regions in the input, each of which computes a dot product between their weights and a small region which are connected to the input volume. This may result in a volume such as [227x227x32] if we make use of 32 filters. Similarly after the second layer convolution resultant volume is obtained.
- RELU layer is used to apply an element wise activation function. The width, height, result in volume
- POOL layer performs a down sampling operation. Mostly the max pooling process is done. This includes selecting maximum from the pooling region of interest. This is fed to next layer as input.
- FC (i.e. fully-connected) layer computes the class scores. Each neuron in this layer is connected to all the other neurons in the previous layers. The result of the convolution and max pooling process is taken by this layer and use them to classify the image with a label.

IV. PROCESS FLOW

The model of a Convolution Neural Network written in MATLAB is Alex-Net model architecture. The model is used to train a large network, uses feature map technique and hence classified with respect to their class score. The following steps are involved for each image to be classified.

A. Data loading and Pre-processing

During the data loading and preprocessing phase, the image dataset is split in the ratio 80:20. The height and width dimensions of each image are changed to uniform size. A batch size for processing at a time is also defined during the initial process of this phase in order to increase the processing speed. If required the intensity values are changed to the requirement ranging from 0 to 255. Then the images from the training image dataset is loaded into the training generator. Similarly images from the validation image dataset is loaded into the validation generator. The classification mode is set and class indices are assigned. The model and the necessary functions are loaded from the MATLAB add- on functions.

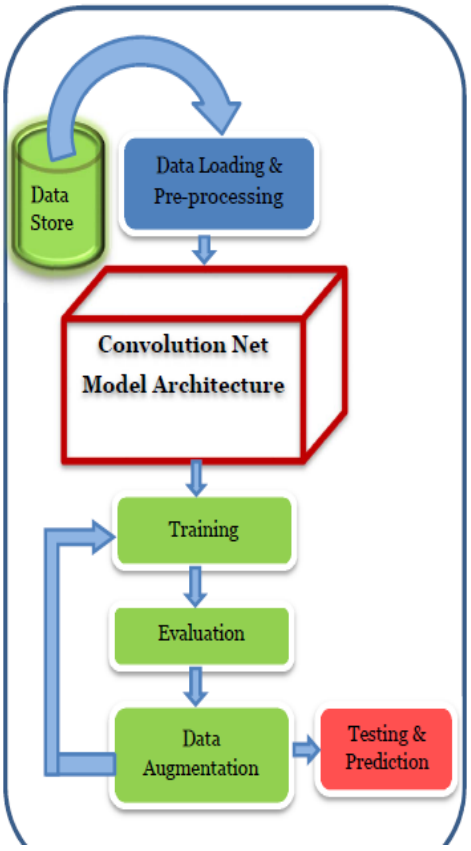

Fig. 4. Process flow

B. Convolution Network Model Architecture

The various parameters such as the number of convolution layers, activation functions and other important parameters required for the CNN architecture say Alex-Net are set. And in this way the model of this architecture is set. More the number of convolution layer more the training accuracy will be present in order to extract more features. The model architecture is shown in the figure 4.

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 9 Issue 07, July-2020

The convolution layer thereby has a problem that it can be conceived very easily and so the output of one layer gets transferred to another layer as input. In order to avoid such issues of curse of dimensionality the convolution layers are interlaced with the pooling layer. The main purpose of the pooling layer is to reduce the spatial size by reducing the amount of parameters and computation in the networks. Also to control the over fitting pooling layer plays an important role. Mostly the max pooling is done.

A non-saturating activation function ReLU-Rectified Linear Unit function given by

$$f(x) = \max(0,x)$$

is generally used. This is done in order to increase the non-linearity property as well as to improve the overall network. In the final step of this phase, after several convolutional and max pooling layers a high level reasoning in the neural network is done via fully connected layers. This fully connected layers have connection with all the Relu activation function. These are connected by the neurons. Thus the output of fully connected layer is given to softmax layer for different scores (classes).

C. Training

The training image set is then trained on the model by setting the number of stages of training (Epochs). In this phase the inputs (training set images) are completely passed through the network. Also a backward phase takes place where the weights are updated.

D. Evaluation

In this phase, the model is run by the validation dataset which was generated in the beginning. This is mainly done to compute the accuracy of the model. The accuracy of training and validation data set for each of the epoch if required can be plotted. This will indicate progress of the validation dataset.

E. Data Augmentation

To improve the performance of the network with the existing samples we hereby artificially enhance the training dataset by applying some random transformations such as rotation, cropping, rescaling, etc. of images. These transformations on training and validation dataset increases the volume of the dataset. After this augmentation is done, the model is allowed to recompile and re-evaluate on the newly generated training dataset. Hence, the accuracy of the model is generally improved after this phase.

F. Testing and Prediction

From different classes test images are randomly chosen and pre-processed as it was done for training and validation dataset which is mainly the image dimension and intensity normalization. The test images are then subjected into the model for classification. The model thereby predicts the image into their perspective classes and labeled correctly.

V. ALEX-NET ARCHITECTURE

Alex-Net model is the most representative model of CNN. Because of its superior performance, less training parameters and strong robustness, the model suits well for classification of images. Alex-Net is a deep CNN model with multiple hidden layers, including an input layer, convolutional layers, pooling layers, fully connected layers and an output layer with different output class labels (here 12 classes).

Deep CNN reduces the dimensions of image by increasing the number of hidden layers (convolution layer and max pooling layers) and therefore extracts the sparse image features in low dimensional space.

Alex-Net model directly uses Relu non-linearity in the data structures to make the initialized method more consistent with the theory and also begin to train the network directly from the starting point to enhance the training speed.

A. Input Layer

The images are reduced to 227 x 227 dimensions and selected as input to training network. The red, green and blue are the main colors of images.

B. Convolution Layer (C1)

First convolution layer is used for feature extraction and so in this way we obtain 96 feature map with size 55 x 55 dimension. This is obtained by using the convolutional kernel of size 11 x 11 dimension.

The size of the convolution kernel is to extract the effective local features in the range of convolution kernel with representation ability. Therefore the proper setting of the convolutional kernel is very important to extract the effective image features and improve the performance of the CNN.

Convolution layer filters the 227 x 227 image with 96 convolution kernels of size 11 x 11 with stride of 4 pixels for sampling frequency, namely the convolution kernel is spread over every unit of size 11 x 11. Finally, we get 96 feature maps with size of (227/4-1)(227/4-1)= 55 x 55.

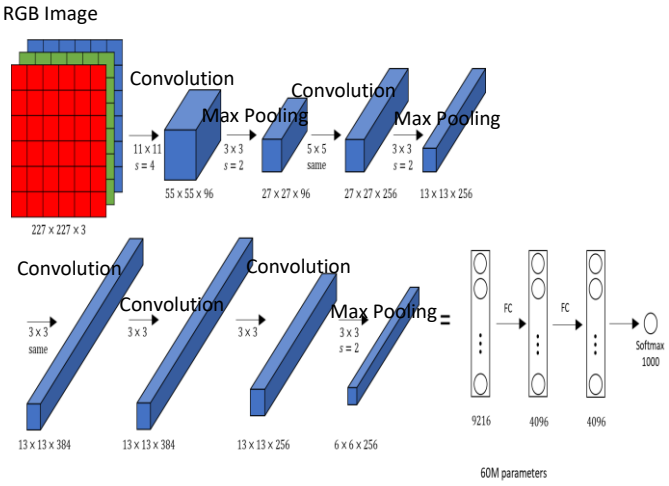

Fig 5. Alex- Net Architecture

C. Max- Pooling Layer

The pooling process is to select the maximum in each of pooling regions as the value of the area after pooling.

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 9 Issue 07, July-2020

In this layer, we choose a max pooling layer over a 3 x 3 region in order to control the speed of dimensionality reduction, because the decline in dimension is decreasing exponentially. If the speed is falling faster, the images are tougher and many details are lost. So such a situation should be avoided.

D. Convolution Layer (C2)

The 2nd layer extracts features similar to 1st convolution layer. C2 takes input from the output of 1st convolution layer and filters with 256 feature maps with size of 27 x 27. Each feature maps in pooling layer as input for convoluting again. Reason behind this is sparsely connected mechanism keeps the number of connections in a reasonable range. The asymmetry of network enables the different combinations to extract various features.

E. Remaining convolution layers and pooling layers

With continuous increase of depth of convolution, the extracted features are more abstract showing more expressive power. The classification process with accuracy increases when depths are more added and so performance level of CNN increases.

F. Fully Connected (FC) Layer

The objective of fully connected layer is to take the results of the convolution and pooling process and use them to classify the image into a label. Then they pass forward to the output layer in which every neuron represents a classification label. The output of fully connected layer is fed to softmax which produce a distribution over 12 classes. Output is 12 classes with its labels.

VI. RESULTS

The results of Natural Scene classification using CNN are as follows:

A. Classification

To check the classification done by the model, different sets of trained images were passed through the model. For classification, the trained images which are 227 x 227 in dimension are loaded in the respective code in jpg format. The resized images with kb size is classified along with label. When the code is run, then GUI displays a window showing 12 different switches/pushbuttons along with confusion matrix switch and a display portion of accuracy and loss. The right hand side of the window displays the classified image with its label. This happens when a pushbutton corresponding to its code is evoked. For example, a Desert Mountain pushbutton is clicked, then the code corresponding to it is called and then whichever image is uploaded it will display it with its correct label. For example: if desert mountain switch is clicked then the image uploaded corresponding to its code gets classified as desert mountain. This is how an image is classified to its correct class. Like these 12 classes can be classified. The figure shows the classification of Desert Mountain.

Fig 6. Classification of Desert Mountain

B. Histogram Analysis

Histogram prediction helps in analysis of an image in depth. Histogram analysis is a graphical representation of the tonal distribution in an image. It plots the number of pixels for each tonal value. To understand the entire tonal distribution of a specific image in glance histogram plays a vital importance. The horizontal axis of the graph represents the tonal variations, while the vertical axis represents the total number of pixels in that particular tone.

To the left side of the horizontal axis represents the pixels of the dark areas, whereas the middle part represents the mid tonal values and the right side represents the light areas. The vertical axis represents the size of the area that is the total number of pixels captured in each one of these zones. Thus, the histogram for a very dark image will have most of its data points on the left side and center of the graph. Whereas, the histogram with a very bright image will have most of its data points on the right side and center of the graph. Each image will show its histogram graph corresponding to classified image. Histogram analysis of image is shown in the figure.

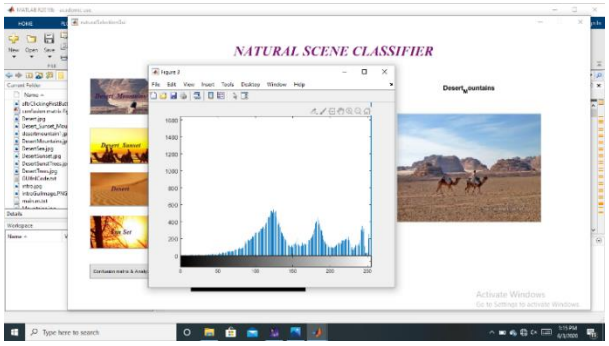

Fig 7. Histogram Analysis

C. Accuracy of the Model

The accuracy of the model is quite high and thus the model performs well with good classification. Since Relu function is used instead of Tanh to add non linearity so this accelerates the speed by 6 times at the same accuracy. Here the accuracy of the Alex-Net model is obtained 89.6%. This is comparatively high. The whole trained images are tested and accuracy found to be good. The accuracy is thus displayed.

D. Loss of the Model

The loss obtained in classification of Alex-Net model is 10%. Here, hamming loss is obtained. Hamming loss is the

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 9 Issue 07, July-2020

fraction of the wrong labels to the total number of labels. So in this model its quite less.

E. Confusion Matrix

A confusion matrix is a table that is often used to describe the performance of a classification model on a set of test data for which the true values are known. It therefore allows the visualization of the performance of an algorithm. A confusion matrix sometimes is known as error matrix also. That is if one class is commonly mislabeled as the other, so this matrix table helps in finding the error. It allows easy identification of confusion between the classes. Most performance measures are computed from the confusion matrix. The number of correct and incorrect predictions are summarized with count values and hence broken down by each class.

This is the main thing done in confusion matrix. Hence, confusion matrix shows the ways in which the classification model is confused when it makes predictions. It helps us to find the types of errors that are being made.

VII. CONCLUSION

In this paper, we proposed an Alex-Net model under CNN algorithm for the Natural Scene image classification. From different class test accuracy it is evident that automatic classification using CNN technique is simple and more accurate in comparison to other feature extraction techniques.

Multi class classification technique and building a model using different parameters namely activation function, number of convolution layers, number of training cycles could be used to classify images. In multi label the Neural Network models have thereby performed best as its accuracy performance is far better among all the models inspite of having more or less same features given to all models extracted from CNN (Alex-Net model) which gives a good representation of an image.

Our approach involved the following crucial tasks:

1. Creating Alex-Net model.
2. Classification of 12 different Natural Scene classes.
3. Accuracy of the model being represented with maximum testing of images.
4. Histogram of each image with its intensity of pixel values.
5. Loss occurred in the model during the classification process.
6. Confusion matrix in order to test the performance of the model.

The advantage of our approach is that it is simple and it makes the best use of the image dataset. It also provides much accurate results and proper labelling at a lesser time.

It has proved that the many intermediate layers used in Alex-net of CNN provide comparable and more information for image classification. Information mostly is the result of accuracy of the test images.

The graph of histogram of each images specify the intensity level of the pixels with more accuracy. The accuracy and loss of the model is found out independently which proves Alex-Net efficiency to be high with increased accuracy and decreased data loss. The confusion matrix also reflected the performance efficiency of the model with most of the classes predicting true classes.

Thus it can be noted that the Classification of Natural Scene using Alex-Net model becomes more scalable.

REFERENCES

[1] Sameer Singh, Markos Markou & John Haddon, Natural Object Classification Using Artificial Neural Networks
[2] Scene Classification – A general description
[3] Zhi-Hua Zhou, Min-Ling Zhang, Multi-Instance Multi-Label Learning with Application to Scene Classification
[4] Zhixin Li, Yu Shen, Nan Huang, Liang Xiao, 2017 IEEE, Supervised Classification of Hyperspectral Images Via Heterogeneous Deep Neural Networks
[5] R. Raja, S.Md.Mansoor Roomi, D.Dharmalakshmi, S.Rohini, 2013 IEEE, Classification of Indoor/Outdoor Scene
[6] Liang Ye, Zhiguo Cao, and Yang Xiao, DeepCloud: Ground-Based Cloud Image Categorization Using Deep Convolutional Features
[7] Tutorial GUI MATLAB
[8] Alexis David P. Pascual, Lei Shu, Justin Szoke-Sieswerda, Kenneth McIsaac, Gordon Osinski , IEEE 2019, Towards Natural Scene Rock Image Classification with Convolutional Neural Networks