International Journal of Advances in Engineering and Management (IJAEM)
Volume 3, Issue 3 Mar. 2021, pp: 1319-1326 www.ijaem.net ISSN: 2395-5252

An Indoor-Outdoor Scene Classification Analysis Based on VGG-16 CNN Model

Nardeep Singh, Mr.Abhinash Singla
M.tech (CSE) Student Bhai Gurdas Institute Of Engineering and Technology,Sangrur,PB
Assistant Professor Bhai Gurdas Institute Of Engineering and Technology,Sangrur,PB

Submitted: 30-03-2021 Revised: 06-04-2021 Accepted: 09-04-2021

ABSTRACT-Computer vision has developed a different stage that required robots from laboratory environments to explore the outside world. Also with progress in this area, robots face difficulties in understanding their environment. Scene classification is a primary stage for scene understanding. In several ways Scene classification can be used, such as tracking camera, autonomous driving, residential robot & database imaging. Surveillance cameras are now installed all over the place. Scene classification of indoor-outdoor approaches has a poor accuracy problem. This research aims to enhance the accuracy by using the Convolution Neural Network Model in VGG-16. Indoor/Outdoor scene classification. This paper proposes a new approach to VGG-16 to classify images into their classes. The algorithm results are tested using SUN397- indoor-outdoor dataset and the experimental data reveal that the methodology proposed is superior to the existing technology for the scene classification of the indoor-outdoor. In this paper, We propose Very Deep Convolutional Networks for Large-Scale Image Recognition. In ImageNet, which is a dataset with more than 14 million images from 1000 classes, 92.7% achieved top-5 test accuracy. It improves on AlexNet by replacing large filters (11 & 5 in 1st& 2nd convolutional layers), with multiple 3×3 kernel size filters one by one. We attain Training loss = 10% and Training Accuracy = 96% in our proposed work.

Index Term—Classification, Indoor-outdoor Classification, Deep Learning, Neural network model VGG-16.

I. INTRODUCTION

One of the holy grail problems of computer vision is the scene classification (SC) problem. Provided an arbitrary image, we want to explain what kind of semantic scene it shows. Very little work has now been performed in this field probably because the issue is very complex and perhaps because scene description terminology has not been agreed upon. Low-level image processing, which rarely aims to overcome the difference to semantic scene classification, is the main computer vision study. the indoor and outdoor scene recognition methods are vividly implemented in handheld assistance to help visually challenged people in different environments of unknown public places like the library, temple, an airport terminal, cafeteria, etc. In the robotics field, scene classification algorithms help robots in recognizing the type of environment in which they are working. Also, many pictures are being clicked by photographers at different places across the world every day [1]. Classification of image scenes is an important component of certain aspects. Indoor & Outdoor classification is an important component of scene processing since it is a basis for multiple approaches to semantic scene evaluation. [2].

Numerous new technologies have been implemented to tackle this issue, but each method depends on an image database which reduces the trust in any method's performance. The classification of the scene is a challenging task because the high-level entities that are considered representative of one kind of scene may be a different kind. Related objects as plants can, for example, occur in either class in the indoor/outdoor classification. Different techniques have been proposed for automatic classification with different performance levels which focus mostly on low color space as well as texture features. Study on the scene classification has grown considerably in the last decades. Rapid technological advancement in CBIR and growing demand for online storage spaces, improved organization & retrieval of the image database. CBIR is the use of image retrieval & scanning for digital images in broad databases via computer vision techniques. Content-based means the search is based on image content and not on metadata including keywords, tags, or image-related descriptions. Colors, shapes, textures, or any other data that can be resulting from an image can be related to in this way. The Scene classification has (SC) been known as the core field of CBIR study.

Using Bayesian binary classifier to classify images by using color and texture feature in the hierarchical category. Classification indoor/outdoor scene using FCM clustering where accuracy level is insufficient[3] future modern classification method depended on image edge straightness analysis as indoor images include man-made objects with straight edges while tree-like outdoor, a mountain without straight edges [4].

One of the difficult tasks in computer vision is the SC problem. Any scene classification challenge specified any arbitrary image is that computer must accurately associate it with a given category like indoor scene. For several years, the issue of SC has been examined from numerous angles in literature. Different techniques were suggested & good performance in specific image datasets was obtained. Even so, a clear scene classification technique cannot be accepted and will solve the scene classification problem perfectly. About all classification approaches to the Indoor-Outdoor scene can be summarized. There are usually 2 stages called training & classification phases. Extract image features are the first step for both training and classification phases. Researchers have developed different types of features to demonstrate a contrast between the indoor & outdoor images. Feature extraction is commonly considered to be critical for the classification of the indoor-outdoor scene [5].

CNN is a type of directed acyclic graph. Such kinds of networks will have the ability to learn immensely high non-linear functions. A neuron is a primary unit in a CNN. Every layer in CNN is comprised of numerous neurons. The convolutional features have a more distinguish portrayal of scene images than those of features extracted by image processing methods. The convolutional features are learning-based features that contain rich semantic information, which is more potent and better applicable for scene classification. We should note that the low-level features that contain descriptive details cannot be ignored [6].

The paper is planned as follows. In Section 2; we extant the related work. Section 3 provides an overview of the Proposed methodology defines a methodology for scene radiation Section 4 experiment result of the proposed SSDCNW approach. Section 5 draws some conclusions and section 6 discusses our future work.

II. LITERATURE SURVEY

The contribution of various research papers is explored in this section that exhilarated our understanding of the problem definition and helpful in determining the challenges, gaps, and issues available in the field of scene classification in indoor and outdoor images. In recent times, there are many improvements and developing are made in image recognition which is used in scene classification that is to differentiate the different classes using neural networks. Scene recognition and classification or scene categorization have been broadly carried out in various environments. Following are the different state of an art methods in the domain of scene radiance. Several enhancements have been suggested in recent years to learn rich features from color pictures. One tactic of this type is coevolution Neural Networks (CNN's) with dense net teaching proposals use image region proposals and another way is to discuss contextual Knowledge that is meant to distinguish the various images according to their groups and properties between different image segments.

L. Ru et al. (2020)A CorrFusion module was suggested that fuses components highly correlated into two-term embeddings. They first investigate the depth of the two-time inputs with DNN. Then extracted parameters are projected onto a smaller space to isolate correlated components & determine the correlation of the instance stage. cross-temporal is carried out based on the simulated CorrFusion module correlation. With SoftMax layers, the final scene classification findings are produced. They also added a new method in the optimization problem to more accurately and stably measure the temporal correlation. The methodological considerations of back-propagation gradients are also provided for the proposed module. Also, they introduced and performed detailed studies with a much wider-scale scene changes data set with more sentences. The study results have shown that our suggested CorrFusion module can increase the classification and identification results of multi-temporal scenes remarkably. [7].

D. Alajajiet al. (2020) Preliminary findings report using UC Merced as well as optimal31 data sets from both RS scene. Remote sensing (RS) is now a major development topic for scene classification. A typical approach is to mark a sufficient number of RS scenes with expert opinions as appropriate, and then to learn how to correctly identify other new scenes. A broad dataset was needed for training in the best performing DL models. Therefore, an intelligent ML algorithm needs to be developed which can learn to identify RS datasets from just a handful of samples, with new unseen classes. This is called ML with few shots. In this work, they establish a DL method for classifying RS scenes. The suggested procedure is built on DNN models and CNN pre-trained for image processing together with the Squeeze Net. [8].

International Journal of Advances in Engineering and Management (IJAEM)
Volume 3, Issue 3 Mar. 2021, pp: 1319-1326 www.ijaem.net ISSN: 2395-5252

A.Rafique et al. (2020)A model for understanding & distinguish a scene using depth information to let computer view real-time scenes as human beings. The proposed recognition method is a new segmentation system, that usages multi-object statistical segmentation to learn & segregate objects in the scene. unique features are then stripped for further recognition using linear SVM from these separated objects. Lastly, features & weight of scene recognition are given for multilayer perceptron. Our system has significantly built on state-of-the-art systems. Sport & security system proposed effectively in autonomous vision systems like robotic vision, GPS location Finder, etc. The intelligence capacity of computers is the day by day by technological advances. Investigators are devoted to humanly equipping devices. The machines can currently sense and process sensor information. Though, the desire to think and understand real scenes remains enormous. Understanding the scene is now a day for study [9].

P. G. Pawar & V. Devendran (2019) Humans are extremely proficient to understand high-level systems of visually perceiving natural scenes. Scene understanding lately is a challenge and a major computer vision issue. Images are visual but visual data may have different characteristics similar to shape, edges, texture & color. Identifying objects in the image and where they are all located is the main aim behindhand object detection. The understanding of a scene includes meaningful information to extract semantic connections & patterns at multiple levels. interaction of separate objects for human beings is most innate and natural. Compared with object detection, a scene containing describes the object targets as well as the target distribution in a scene. To perceive, analysis, and analyze visual scenes access to different areas of science the scene understanding has a major effect on computer vision. In this text, we discuss the definition of scene comprise the extraction of roles, classification and scene identification, and related datasets. Finally, certain challenges in the detection of objects remain unresolved and may be used for further analysis [10].

L. Zhang et al. (2019) Propose an indoor-outdoor classification ensemble based on cellular data from an LTE commercial network based on a traditional municipal field. The variables are derived by KPIs and radio propagation information of network core performance indicators. The DT grows & breaks into the Gini index of sampled characteristics, depending on these key variables. All the decision trees are then arranged to create an ensemble scheme and thus boost the capacity to distinguish. The self-validation outcomes indicate that the ensemble model achieves highly detailed classification in both indoor and outdoor environments (with an out-of-bag error of less than 1 percent). Also, prominent variables are chosen depending on the variable reputation of initial training. In comparison to other classical ML methods reconfigured model depended on a smaller number of variables as well as less weak learners often achieves the highest accuracy & relatively short computation period. Through the development and introduction of mass-mobile devices for the next 5G era, wireless Big Data has attracted considerable interest. For personalized services in a smart world, the context information about these devices is essential. The ongoing change in scenarios, however, challenges the network operator [11].

O. Ye et al. [2018] Proposes a mine video SC system, enhanced CNNs. algorithm increases the accuracy of feature extraction for complex background videos by the growing depth of the initial CNN network. This network communication structure contains 10 neuron layers. The first 7 layers are wrapped, and all layers are related, and features are extracted. The SoftMax loss function is used to define the whole connection layer.Experimental findings indicate that the approach introduced in this paper effectively solves the issue of the difficult background video scene classification. Classification of the scene is a significant computer vision research content that has been widely used in many areas, including retrieval of video and images, machine vision & robotics, monster medical, and video tracking. Issues of classification with a difficult background for video scenes are how to certify the accuracy of video processing & classification [12].

S. Paul et al. (2016) Aims to select a subset of revealing & diverse images that are used to efficiently learn the classification model. Depending upon the budget for manual labeling, the scale of the subset would be calculated. While effectively studying algorithms can solve the problem of identifying insightful images, several labeled images are required for the initial model construction that our approach does not need, since the best samples are recognized at one shot. They integrate ideas of a strong and poor teacher to support students efficiently learn model with a small manual labeling budget. The efficacy of our algorithm is shown by two difficult scene classification sets. The enormous time needed for building several labeled images to train a classifier has driven physicists to create the most detailed training images so that they are labeled inappropriate ways to achieve a significant accuracy in classification [13].

R. Raja et al. (2015) The exponential rise in digital image storage, Retrieval has become an impending problem by the exponential growth of the storage of digital images. Such a large data set requires a long time to obtain images, other than to pick images relevant to the query. The search time is even longer, despite success in introducing useful features. The time of search could be reduced in such a scenario by categorizing the datasets indoors or outdoors. The paper aims to divide the picture into an internal or external scene. The effective illumination & rotation of arbitrary low-level characteristics like the color of the HSV model & texture (GBWHGOPCA) in combination with the Sparse Representative Classifier (SRC) have been used to enable automated SC at the concept level. As these image features show characteristic disparity amid images comprising indoor or outdoor scenes, better performance as respects grade quality is achieved by the proposed technique. This paper is evaluated on the data set of IITM-SCID2 (scene classification image database) & 15 scene categories with a data set of 3442 images composed by authors from the web [14].

III. PROPOSED METHODOLOGY

In this section, we analyzed the different experimentation for indoor and outdoor classes. We have 6 indoor classes and 6 outdoor classes. A well known pretrained network 'VGG-16' has been choosen for our proposed transfer learning model.

3.1 Problem Statement

It is observed that the Previously proposed model is highly efficient while classifying outdoor classes and getting a little bit confused while classifying indoor scene categories. Airport terminal and Gymnasium are the two major classes, where the model is less efficient in classifying accurately. The major reason could be that the activations of these classes did not fetch the desired features for training. Recovering images or pictures from this large compilation of databases is a highly time-consuming and complex task. The deep learning approach in many computer visions tasks has recently achieved good performance. But the classification of the indoor-outdoor scene is also difficult to explain. Thus, it also needs to be multidisciplinary scholars, in particular academics like neurobiology and machine learning, to focus on indoor-outdoor problems for scene classification.

3.2 Methodology

The work proposes a VGG16 deep CNN model for indoor-outdoor scene classification from images of various public scene environments. A well-known pretrained network, 'VGG16' has been chosen for our proposed learning model. In this first, we have to perform preprocessing. Preprocessing is the overall term for all the transformation of the data, including centering, normalization, rotation, shifting, shear, etc., before being transformed into the model.

3.2.1 Image Data Generator

The validation dataset and the evaluation dataset can also be defined by a data generator. A separate instance of ImageDataGenerator is also used that has the same configuration for pixel scaling (not covered) as that utilized for Image Generator's training dataset. The reason is that the data augmentation is only employed to artificially expand the training data set to boost model efficiency on an augmented dataset. To increase our training data, we employed Keras ImageDataGenerator. It offers different transformations to increase image data such as scale, rotating, shear, brightness, zoom, channel shift, width and height changes, and horizontal and vertical shifts. We applied geometric transforms such as Scaling, zoom, Horizontal flip, Image size, Batch size, Images, Classes, Color channel, Test data image, and Validation image in the proposed method. In our scaling 1. /255, image zoom in 20%, a horizontal flip is True, our image size of 227*227, batch size 64, and we used images 4132, classes 2, Color channel 3(RGB), Test data image 518 and Validation image is 515.

3.2.2 Data augmentation (DA)

The data augmentation plays an important role in improving the model's efficiency in the proposed process.DA of the image is a method that may be utilized to artificially increase the size of the training dataset by producing updated image versions in a dataset. Training DLNN models in more data will cause skilled models & increase techniques that can produce image variations that enhance the capability of fit models to simplify what they have learned into new images. Kera's DNN learning library gives the capability to fit image data increase models through the ImageDataGenerator class.

3.2.3 Neural network model VGG16

Deep learning has strong performance in image classificationand several deep learning models like AlexNet, VGGNet, and InceptionNet have been used in recent years. In this work, we have used VGG-16 in CNN for this purpose. CNN is a form of ANN that utilizes several perceptrons that evaluate image inputs also have

International Journal of Advances in Engineering and Management (IJAEM)
Volume 3, Issue 3 Mar. 2021, pp: 1319-1326 www.ijaem.net ISSN: 2395-5252

learnable weights & bases to many parts of images up to separate each other. One benefit of used CNN is that it uses local spatial coherence of the input images to reduce their weight when different parameters are exchanged. Research in architectural design has accelerated the efficient use of CNNs in image recognition tasks. CNN architectures have a basic and efficient design principle. Their architecture, called VGG, was modular in layers pattern.

In the paper, VGG16 is a CNN model presented by A. Zisserman and K. Simonyan from Oxford University's "Very Deep Convolutional Networks of Large-Scale Image Recognition". In ImageNet, a dataset of over 14 million images belonging to 1000 classes, this model achieves 92.7% of the highest test accuracy. This was one of the ILSVRC-2014's popular models. It enhanced on AlexNet by replacing broad kernel-sized filters (5, & 11, respectively, on 1st& 2nd layers of convolution) by multiple 3×3 kernel-sized filters one by one.

VGG 16 is a CNN 16-layer, pre-trained 100-class model. The Sun397 image network dataset is trained in this VGG 16 model. We utilized this model as a feature extraction tool & extracted from each image 4132 features & stored them in hdf5 file format. To resize all the images to mentioned dimensions VGG 16 models need images of 224 x 224 dimensions. We have such a good result. VGG-16 has an excellent ability to extract the image to get a strong image classification effect.

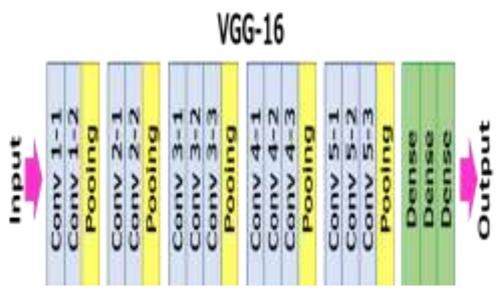

Figure 3.1: VGG16 Model

3.3 Loss Function

CategoricalCrossEntropy: For categorical classification, cross-entropy loss contributed by training data point i (x_i, y_i), is simply the "negative log-likelihood (NLL)":

$$L_i = -\log(p_{y_i})$$

since the ground truth probability is one for the correct label y_i and zero for every other label.

3.4 Adam Optimizer (Adaptive Moment Estimation)

It is an algorithm for gradient descent optimization. When dealing with big problems with a lot of data or parameters, this approach is very effective. It needs less storage and is effective. Intuitively, the "gradient descent" algorithm, as well as the "RMSP" algorithm, are combined.

3.4.1 Mathematical Aspect of Adam Optimizer

Taking the formulas used in the above two methods, we get:

$$m_t = \beta_1 m_{t-1} + (1-\beta_1)\left[\frac{\delta L}{\delta \omega_t}\right] v_t$$
$$= \beta_1 v_{t-1} + (1-\beta_1)\left[\frac{\delta L}{\delta \omega_t}\right] \quad \ldots (1)$$

m_t and v_t are estimates of 1st moment (mean) & 2nd moment (uncentered variance) of gradients correspondingly, henceforth method name. As m_t & v_t are set as vectors of 0's, researchers of Adam detect that they are biased towards 0, especially throughout initial time steps, & especially when decay rates are small (such that β_1 & β_2 are nearly 1).

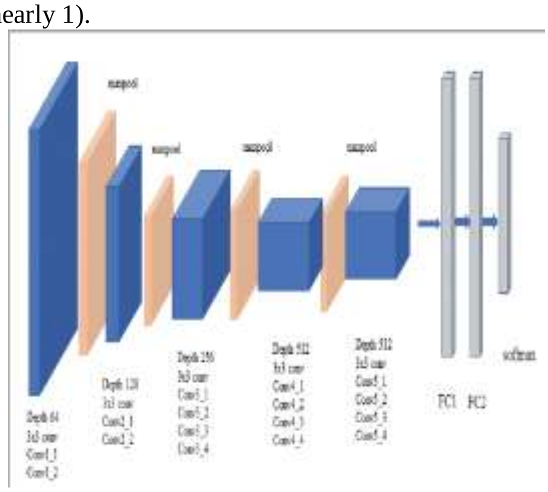

Figure 3.2: Architecture of VGG-16-CNN model

We modified VGG16 with different layers like the Batch normalization layer after every convolution layer and the dropout layer after every Dense layer.

Algorithm: Proposed Algorithm

Step 1: Collect the input images from SUN397 Dataset.
Step 2: Preprocess the image
Step 3: Training with CNN model
Step 4: Modify the VGG 16 with different layers.
Step 5: Test the images.
Step 6: Predicted Results.

International Journal of Advances in Engineering and Management (IJAEM)
Volume 3, Issue 3 Mar. 2021, pp: 1319-1326 www.ijaem.net ISSN: 2395-5252

3.5 Proposed Model

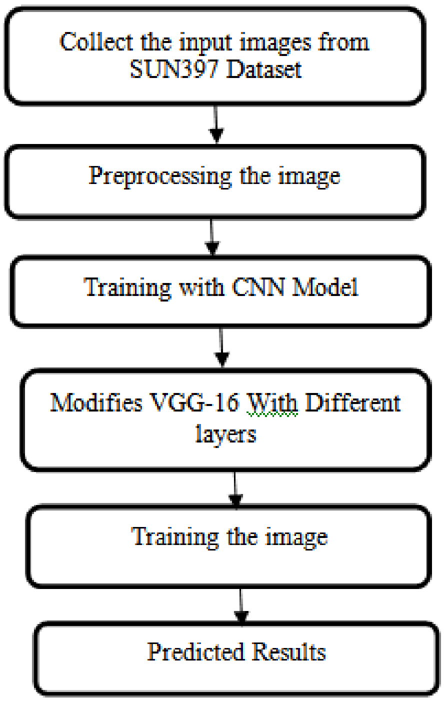

Figure 3.3: Block chart of Proposed work

The above figure shows the flowchart of pseudocode in which we will use the additional approach to decrease the loss and increase the accuracy percentage. The CNN architecture transfers the output of one layer to the next layer as an input. Layers of convolution are Made with weights and biases used to filter a representation of the input. A collection of filtered images is the output of a convolution sheet. This, this, Activations of the layer are called performance. These activations are a 3-D series, where a channel is also considered the third dimension. Dense Net is a recent discovery of visual object detection in NNs. With some basic variations, DenseNet is very different from ResNet. ResNet uses an additive approach (+) to fuse the prior layer (identification) with the future layer, whereas DenseNet concatenates (.). Dense connections are a kind of layer in DNN that uses a linear operation, where each input is associated by weight with each output.

IV. EXPERIMENT AND DISCUSSION

This work has been implemented using Python programming to test the proposed approach. After training the networks separately for indoor and outdoor classes, it is observed that the accuracy of indoor classes is better than that of outdoor classes from the chosen dataset. The dataset used for the purpose is publicly available SUN397dataset.

4.1 Dataset Description

SUN397 is a wider scene benchmark of 397 categories like indoor, man-made & natural categories (at the least before places). The SUN397dataset is used for public use. This dataset is very demanding not only for a large number of categories however also because of the smaller number of trained data and a much wider variability of object and layout properties (50 images per category). It is generally accepted as the scene classification reference benchmark. Our experiments consider seven scales which are 227x227 by scale images.

Table 4.1: Model Summary

Model: "sequential"

Layer (type)	Output Shape	Param #
conv2d (Conv2D)	(None, 224, 224, 64)	1792
conv2d_1 (Conv2D)	(None, 224, 224, 64)	36928
max_pooling2d (MaxPooling2D)	(None, 112, 112, 64)	0
Batch_normalization Batch No	(None, 112, 112, 64)	256
conv2d_2 (Conv2D)	(None, 112, 112, 128)	73856
conv2d_3 (Conv2D)	(None, 112, 112, 128)	147584
max_pooling2d_1 (MaxPooling2	(None, 56, 56, 128)	0
batch_normalization_1 (Batch	(None, 56, 56, 128)	512
conv2d_4 (Conv2D)	(None, 56, 56, 256)	295168
conv2d_5 (Conv2D)	(None, 56, 56, 256)	590080
conv2d_6 (Conv2D)	(None, 56, 56, 256)	590080
max_pooling2d_2 (MaxPooling2	(None, 28, 28, 256)	0
batch_normalization_2 (Batch	(None, 28, 28, 256)	1024
conv2d_7 (Conv2D)	(None, 28, 28, 512)	1180160
conv2d_8 (Conv2D)	(None, 28, 28, 512)	2359808
conv2d_9 (Conv2D)	(None, 28, 28, 512)	2359808
max_pooling2d_3 (MaxPooling2	(None, 14, 14, 512)	0
batch_normalization_3 (Batch	(None, 14, 14, 512)	2048
conv2d_10 (Conv2D)	(None, 14, 14, 512)	2359808
conv2d_11 (Conv2D)	(None, 14, 14, 512)	2359808
conv2d_12 (Conv2D)	(None, 14, 14, 512)	2359808
max_pooling2d_4 (MaxPooling2	(None, 7, 7, 512)	0
batch_normalization_4 (Batch	(None, 7, 7, 512)	2048
flatten (Flatten)	(None, 25088)	0
dense (Dense)	(None, 4096)	102764544
dropout (Dropout)	(None, 4096)	0
dense_1 (Dense)	(None, 4096)	16781312
dropout_1 (Dropout)	(None, 4096)	0
dense_2 (Dense)	(None, 2)	8194

Total params: 134,274,626
Trainable params: 134,271,682
Non-trainable params: 2,944

Table 4.2: Parameters information

Parameters	value

International Journal of Advances in Engineering and Management (IJAEM)
Volume 3, Issue 3 Mar. 2021, pp: 1319-1326 www.ijaem.net ISSN: 2395-5252

Dataset	sun397
Scaling	1./255
zoom	20%
Horizontal flip	True
Image size	227*227
Batch size	64
Images	4132
Classes	2
Color channel	3 (RGB)
Test data image	518
Validation image	515
Neural network model	VGG16
Epoch	100

4.2 Results Analysis

This subsection represents the outcomes study obtained by the proposed model.

Table 4.3: Comparison of accuracy, Loss, Val_loss and Val_AccuracyforBase and Proposed Model

Model	**Loss**	**Accuracy**	**Val_loss**	**Val_Accuracy**
Base model	0.2623	0.9113	1.7929	0.5615
Proposed model	0.1029	0.9637	1.2143	0.8913

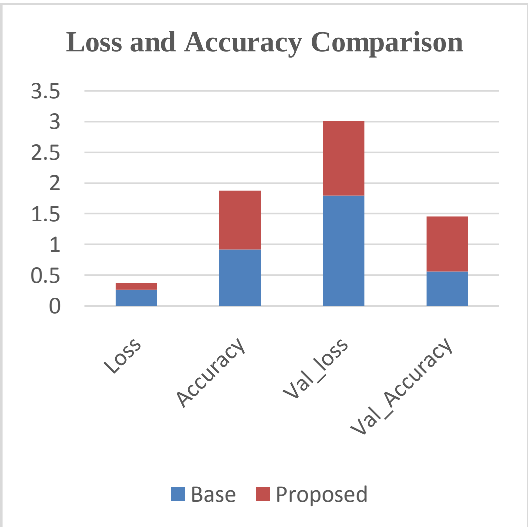

Figure 4.1: Graph comparison of accuracy, Loss, Val loss & Val AccuracyforBase and Proposed Model

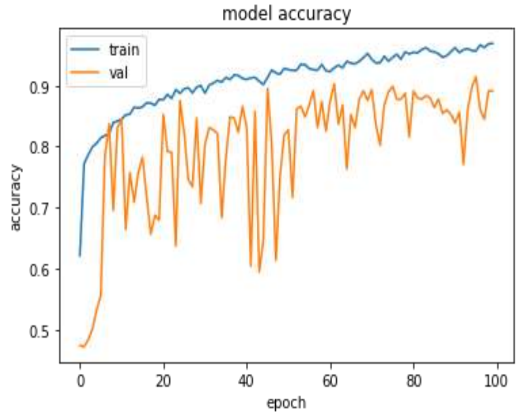

Figure 4.2: Line Graph for Model Training Accuracy

Figure 4.2 represents a line graph for model training accuracy. This process continues up to 100 epochs. It shows training and validation accuracy. Initially, it starts training accuracy from 62%, which is gradually increased up to 96% accuracy at 100 epochs. It also shows validation accuracy. Initially, it starts validation accuracy at approximately 23%, which has a variable increment inaccuracy. it constantly increases approximately 89% accuracy.

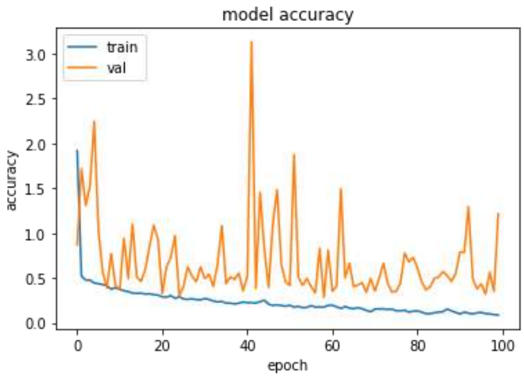

Figure 4.3: Line Graph for Model Loss Accuracy

Figure 4.3 represents a line graph for model loss accuracy. This process continues up to 100 epochs. It shows training and validation accuracy. Initially, it starts loss accuracy from 62%, which is gradually increased up to 96% accuracy at 100 epochs. It also shows validation accuracy. Initially, it starts validation accuracy at approximately 23%, which has a variable increment inaccuracy. it constantly increases approximately 89% accuracy.

V. CONCLUSION

Scene classification is one of those applications in which the computer plays an important role in scene analysis. Scene classification is a method in which a view of a computer vision a scene and the machine then attempts to categories the scene based on machine learning. The classified scene can be an indoor scene as with bakery, airport, garage, bedroom, etc., or an outdoor area including such mountain and roadside, beaches, desserts, etc. Classification of indoors scenes is more difficult than the classification of the outside scenes because of variability. Various methods for classification indoor scenes have been invented over recent years and face different challenges and accuracy is the biggest challenge. In this paper, we suggest Very Deep Convolutional Networks for Large-Scale Image Recognition." In ImageNet, which is a dataset with more than 14 million images from 1000 classes, 92,7% achieves top-5 test results. It improves on AlexNet by replacing broad filters (11 & 5 in 1st& 2nd convolutional layer, respectively) with multiple 3×3 kernel-sized filters one after another. In our VGG-16 model we attain Training loss = 10% and Training Accuracy = 96% in our proposed work.

VI. FUTURE SCOPE

The deep learning method has recently worked well in several computer vision tasks. But how to classify the outdoor-indoor scene perfectly is still rarely explained. Therefore, multidisciplinary researchers do need to focus on the Indoor-Outdoor scene classification problem, in particular from researchers, like neuro-biology & machine learning, for more analysis. We will create an Indoor-Outdoor dataset based on current existing databases in future work. We would also test with a deep learning model and associate them with previous studies.

REFERENCES

[1] A.Payne and S. Singh, “Indoor vs. outdoor scene classification in digital photographs,” Pattern Recognition, 2005
[2] D. Deng and J. Zhang, “Combining multiple precisionboosted classifiers for Indoor-outdoor scene classification,” ICITA 2005,Dunedin, New Zealand, July 4-7, 2005, pp. 720-725.
[3] L. van G. Fei Guo, Lukas Bossard, “Indoor-Outdoor Image Classification,” 2011.
[4] L. Tao, Y. H. Kim, and Y. T. Kim, “An efficient neural network-based indoor-outdoor scene classification algorithm,” 2010, DOI: 10.1109/ICCE.2010.5418764.
[5] Julia Vogel,Bernt Schiele, “Semantic Modeling of Natural Scenes for Content-Based Image Retrieval,” Int. J. Comput. Vision, 72:133-157, July 2006, DOI: 10.1007/s11263- 006-8614-1.
[6] H. Noh, S. Hong, B. Han, Learning deconvolution network for semantic segmentation, in Int. Conf. on Computer Vision (ICCV), 2015.
[7] L. Ru, B. Du and C. Wu, "Multi-Temporal Scene Classification and Scene Change Detection With Correlation Based Fusion," in IEEE Transactions on Image Processing, vol. 30, pp. 1382-1394, 2021, doi: 10.1109/TIP.2020.3039328.
[8] D. Alajaji, H. S. Alhichri, N. Ammour & N. Alajlan, "Few-Shot Learning For Remote Sensing Scene Classification," 2020 Mediterranean and Middle-East Geoscience & Remote Sensing Symposium (M2GARSS), Tunis, Tunisia, 2020, pp. 81-84, doi: 10.1109/M2GARSS47143.2020.9105154.
[9] A. A. Rafique, A. Jalal, & K. Kim, “Statistical Multi-Objects Segmentation for Indoor/Outdoor Scene Detection and Classification via Depth Images,” in 2020 17th International Bhurban Conference on Applied Sciences & Technology (IBCAST), Jan. 2020, pp. 271–276, DOI: 10.1109/IBCAST47879.2020.9044576.
[10] P. G. Pawar and V. Devendran, “Scene Understanding: A Survey to See the World at a Single Glance,” in 2019 2nd International Conference on Intelligent Communication and Computational Techniques (ICCT), 2019, pp. 182–186, DOI: 10.1109/ICCT46177.2019.8969051.
[11] L. Zhang, Q. Ni, M. Zhai, J. Moreno, and C. Briso, “An Ensemble Learning Scheme for Indoor-Outdoor Classification Based on KPIs of LTE Network,” IEEE Access, 2019, DOI: 10.1109/ACCESS.2019.2914451.
[12] O. Ye, Y. Li, G. Li, Z. Li, T. Gao, and T. Ma, “Video scene classification with complex background algorithm based on improved CNNs,” 2018, DOI: 10.1109/ICSPCC.2018.8567752
[13] S. Paul, J. H. Bappy, and A. K. Roy-Chowdhury, “Efficient selection of informative and diverse training samples with applications in scene classification,” 2016, DOI: 10.1109/ICIP.2016.7532406.
[14] R. Raja, S. M. M. Roomi, and D. Dharmalakshmi, “Robust indoor/outdoor scene classification,” in 2015 Eighth International Conference on Advances in Pattern Recognition (ICAPR), Jan. 2015, pp. 1–5, DOI: 10.1109/ICAPR.2015.7050698.

International Journal of Advances in Engineering and Management
ISSN: 2395-5252

IJAEM

Volume: 03 Issue: 03 DOI: 10.35629/5252

www.ijaem.net
Email id: ijaem.paper@gmail.com