World Journal of Research and Review (WJRR)
ISSN: 2455-3956, Volume-13, Issue-1, July 2021 Pages 27-35

A Novel Deep Learning Model for Indoor-Outdoor Scene Classification Using VGG-16 Deep CNN

Deepika Bhardwaj, Vinod Todwal

***Abstract*— Machines have begun to rule human beings as machines are performed nearly all the task what people are capable of doing in today's world. The description of the scene is one word that gains significance in that machines imitate a human being's behavior. Scene classification can either be conducted on indoor or outdoor scenarios by means of different aspect feature extraction techniques. Indoor/Outdoor scenes' classification is found to be more demanding in these two categories. 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 scenario classification. This paper proposes a new approach to VGG-16 to classify images into their classes. The algorithm results are tested using the SUN397- indoor-outdoor dataset. The experimental data reveals that the methodology proposed is superior to the existing technology for indoor-outdoor scene classification. From experimental results, we create that model shown the accuracy of 93.66 percent for indoor classes & 98.91 percent for outdoor classes. Effective tests show the validity of the proposed method.**

***Index Terms*— Scene Understanding, Scene Classification, Indoor – Outdoor Classification, Deep Learning, CNN, VGG-16.**

I. INTRODUCTION

understanding detects a scene first, identifies it, and then understands it. In addition, edges need to detect visual features such as the edges to learn, adapt, weigh alternative solutions & establish new analytical & interpretation techniques, which include a renewed computer vision feature. Understand a scene system needs to be able to adjust current environment variations, foresee, and adapt to predict & communicate with people and other systems. Understanding the scene is a challenge still open where all images must also be evaluated properly defined [1].

The understanding of a scene [2] seeks to explain the image material, no. of objects present, locations & semantic relationships among various objects [3]. Scene identification means scene interpretation and categorization process, and people may often at a glance identify natural scenes containing objects such as people, buildings, cars, roads, and shopping malls. There is a common ability for people to comprehend natural scenes in a short time. Identification of scene is a scientific idea for identifying categories of the scene.

Scene classification & recognition algorithms research into understanding verbal context scene [4] [5]. Recognition of scene ensures that giving data & procedure of image captioning in solitary word by types of methods and techniques.

Computer vision has attained a new level, allowing robots from a laboratory's confines to discover the outside world. Even as robots progress in this field, they face difficulties understanding their environment. The classification of the scene is the primary phase toward understanding the scene. Scene recognition can be used in certain applications, such as a surveillance camera, autonomous drive, domestic robot, and database image retrieval. Surveillance cameras are now everywhere mounted. The need for a camera to be installed in public places and at home, as the crime rate increases day after day. The demand of operators would rise to control all operations on the camera. However, a human error can be made, and even information can not be monitored. So scene classification can help certain activities solve these problems [6].

The scene classification aims to classify scene image 1 to one of the pre-described categories of the scene (like kitchen, bakery, & beach), depends upon the image's ambient content, objects, as well as their layout. Understanding the visual scene involves reasoning about our everyday lives' diverse and complicated environment. Recognizing visual categories, including objects, actions, and events is undoubtedly a prerequisite for a visual system [7].

A major purpose of Image Classification is to distinguish various objects in an image. Various methods are used to identify various objects [8]. It divides images into one between several different classes. Image classification is classified into two classes as a classification of the indoor and outdoor images. As indoor/outdoor classification, laptop vision is a / and additionally difficult task in 20-year image retrieval [9] [10]. System mastering & deep gaining knowledge of techniques were utilized to get incredible results of scene classification [11]. Under extraordinary lighting conditions for

an indoor or outdoor form that includes increased shading processing, video processing & sample reputation, multiple procedures have been proposed [2].

The classification of indoor-outdoor scenes is a major issue in the scene classification, and the findings in the classification of indoor-outdoor scenes help to generalize the classification of the scenes [12][13][14][15]. The classification of indooroutdoor scenes also draws substantial interest from scientists interested in content-based image retrieval [16][17]. In addition to assuming that pictures are normally taken indoors and outdoors in various illumination

Deepika Bhardwaj, , IT Department, RCEW , Jaipur
Vinod Todwal Asst. Prof. CS Department, RCEW Jaipur

environments, more imaging applications, including image processing orientation detection [18], map depth creation [19], and color constancy improvement [20] and robot applications, can also be decided [21].

DL was recently the fastest-growing development in big data analysis as well as, for its outstanding performance relative to conventional learning algorithms, has been extensively and effectively implemented in numerous fields, for instance, natural language processing, image recognition, and speech enhancement. A different DL design of CNNs has successfully obtained effective facts in computer vision, which is attributable to the deep structure that enables the capturing & generalization of filtering processes by image-domain convolutions leading to very abstract & efficient features [22]. DCNN has attained new advances in the characterization of the scene in particular. This CNN is able to learn a variety of different attributes from the images. In the world of computer vision, CNN also has huge promise. It is also important that CNN will play the main role in the future creation of scene classification [23].

II. LITERATURE REVIEW

Jing Sun et al. (2016) Proposes a novel scene classification algorithm, which is dependent upon a classical Alex-Net model and SVM (Support Vector Machine), and which can learn deep characteristics of images. They use scene image classification Lib-SVM training model and compare it with the regression model classification method; Finally, we conducted experiments in this work on two common datasets. The test results have shown that the image features can be efficiently extracted by DCNN (Deep Convolutional Neural Network). Meanwhile, the qualified scene model is even more generalized and achieves the latest classification accuracy [24].

Yashwanth. A et al. (2019) In this article, a new methodology was suggested for the transfer learning method focused on Alexnet to identify images in their classes. Twelve classes were chosen from the publicly available SUN397 dataset, of which six were indoor, & the other six were outdoor. Model is independently trained indoor-outdoor also outcomes have compared. Based on experimental findings, they found that the model showed 92% accuracy for indoor classes & 98 percent accuracy for outdoor classes [25].

G. Memiş and M. Sert (2019) Standard deviation (sigma) features have been added to improve accuracy and reduce computational complexity. To assess its effectiveness, they have made comparisons on their proposed dataset with selected machine learning algorithms. The findings from their data set show that the multimodal sigma-based features provide 81.60 percent of the best classification accuracy for the proposed DNN system. Moreover, the classification accuracy of 79.04 percent was based on its proposed DNN system without sigmabased characteristics [26].

I. Saffar et al. (2019) empirically test the successful use of modern real-time radio-data for semi-supervised learning methods, with partial ground truth information collected from a wide variety of users (indoor & outdoor) in traditional and varied locations. Analyzes of such schemes compared with the latest supervised approaches like SVM and Deep Learning are also presented [27].

O. Sen and H. Yalim Keles (2019) propose two new deep learning frameworks to solve the problem of classification of scenes by aerial images. The results of the two models built using a part of the ResNET50 pre-trained model are examined. The dataset contained 45 categories in model assessments, and 31500 samples were one of the biggest open-access datasets, i.e., NWPU-RESIS45 data collection. The accuracy of 95.7 percent that the developed models achieve is comparable with state-of-the-art techniques [28].

M. Ye et al. (2019) Using the deep convolutional neural network from ASC (Acoustic Scene
Classification) transmission learning. To this end, the Residual Neural Network (Resnet) has a strong and common deeplearning architecture. For TUT Urban Acoustic Scenes 2018, Transfer learning is utilized for the improvement of a pretrained ResNet model. In addition, the focal loss improves overall performance. A data increase strategy based on mix-up is applied to reduce the chances of overfitting. In terms of classwise accuracy about DCASE (Detection and Classification of Acoustic Scenes & Events) 2018 baseline system in TUT Urban Acoustic Scenes 2018 data set, their best system improved by more than 10 percent [29].

Z. Chen et al. (2020) A new classification method built on the CNN-based scene proposed. A spatially sensitive block designed CNN-based scene classification system is effective in extracting abundant spatial features; however, it can also modify functional responses to optimize the role of informative features in results for classification. In comparison, the HRRS imagery-based learning technique is used to achieve the initial model for fine-tuning model parameters, which dramatically minimizes training time. The proposed approach was illustrated using 2 HRRS data sets, and experimental findings demonstrated the superiority of the proposed method [30].

K. Abdullah et al. (2020) In this paper, they suggest a 3G network user classification algorithm based on machine learning indoor/outdoor (IO) in cellular systems. They consider several scenarios. The experimental findings demonstrate that an improving algorithm with an accuracy of 88.9 percent is the best machine learning algorithm for IO classification [31].

III. RESEARCH METHODOLOGY

A. PROBLEM STATEMENT

In this work, we analyze scene classification's basic problem about scene classification for indoor-outdoor. They have used AlexNet DCNN for the scene classification of indoor-outdoor in previous work. This model's depth is much smaller, and thus it is hard to learn from the collection of images. More time is required to produce better results. AlexNet stacks fewer layers and maximum size filters. To deal with such an issue, we have used VGG16 deep CNN model scene understanding and classification that intends to understand the activations from images of various public scene environments with the help of CNN.

B. PROPOSED METHODOLOGY

The work proposes a VGG16 deep CNN model for indooroutdoor scene classification from images of various

public scene environments. A well-known pre-trained 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. The preprocessing objective is to enhance image data that eliminates unwanted distortions or optimizes some image features needed for more processing. In contrast, geometric image transformation is classed among pre-processing methods here, given that similar technologies are utilized for the preprocessing process.

1. Data augmentation (DA)

The data augmentation plays an important role in improving the model's efficiency in the proposed process. The data augmentation concerns the modification of image data and a sequence of operations so that the changed image remains in the same class. Data augmentation helps generalize the data input, reflecting a better test accuracy. There are three key strategies for augmenting training data: expanding data set, in-place or fly augmentation, incorporating data set, and on-site augmentation.

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 the dataset. Training deep learning neural network models in more data will cause skilled models & increase techniques that can produce image variations that enhance the ability of fit models to generalize what they have learned into new images. Kera's deep neural network learning library provides the capability to fit image data increase models through the ImageDataGenerator class.

- ***ImageDataGenerator***

A data generator can also define the validation dataset and the evaluation dataset. A separate instance of ImageDataGenerator is also used that has the same configuration for pixel scaling (not covered) as that utilized for ImageGenerator'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 the unaugmented dataset.

We will concentrate on five major forms of image DA strategies; in particular:

- Image shifts through *height_shift_range* & *width_shift_range* arguments.
- The image flips through *vertical_flip* & *horizontal_flip* arguments.
- Image rotations through *rotation_range* argument • Image brightness through *brightness_range* argument.
- Image zoom through *zoom_range* argument.

In order 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.

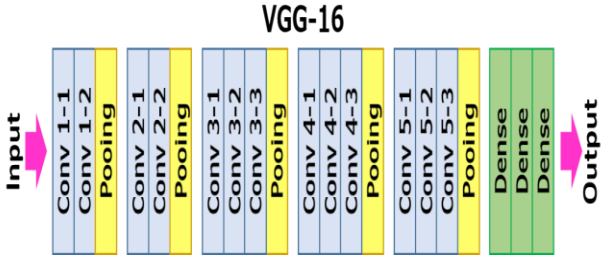
- **2. Convolutional Neural Network (CNN)**

CNN is a form of ANN that utilizes several perceptrons that evaluate image inputs and have learnable weights & bases to many parts of images that can separate each other. One benefit of using CNN is that it uses local spatial coherence of the input images to reduce their weight when different parameters are exchanged. This is an efficient memory and complexity method. Research in architectural design has accelerated the efficient use of CNNs in image recognition tasks. Simonyan et al. suggested that CNN architectures have a basic and efficient design principle. Their architecture, called VGG, was modular in layers pattern.

- **3. Neural network model VGG16**

Deep learning has strong image classification performance, and a number of 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. In the paper, VGG16 is a CNN model presented by A. Zisserman and K.
Simonyan from the 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 improves on AlexNet by replacing broad kernel-sized filters (5 & 11, respectively, on 1st & 2nd layers of convolution) with 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 [11] 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. In order 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 in order to get a strong image classification effect.



- Figure 1**:** VGG16 Model

- **Loss Function**

CategoricalCrossEntropy: By calculating the categorical Crossentropy Loss Function, an example loss is

calculated by:

where $\hat{y_i} = \hat{\imath} - th$ scalar value in model output y_i = corresponding target value, output size is no. of scalar values in model output.

$$Loss_{output} = -\sum_{i=1}^{size} y_i * log\ \hat{y_i}$$

This loss is a strong estimate of the distinguishing distributions of two distinct probabilities. In this case, y is likely to occur with event i & the sum of all y_i is 1, which means that an event will occur exactly. A minus sign means that the loss becomes smaller as distributions get closer together.

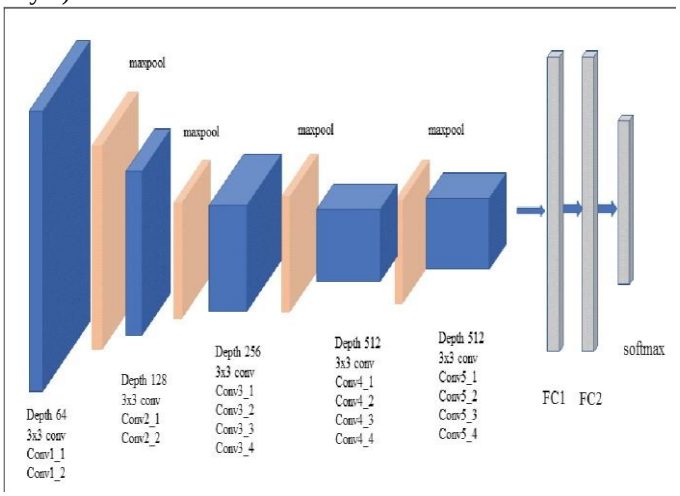
- **Adam Optimizer**

The optimizer is an image file smaller service, product, or library. In general, an image optimizer reduces an image file size by compressing and resizing it, preferably without losing image quality. We used Adam optimizer in this research. Adaptive Moment Estimation (Adam) is an adaptive learning rate calculation tool per parameter. Adam also retains an exponential decay average of previous square gradients v_t Such as Adadelta and the RMSprop, close to momentum, and an exponentially decaying average of previous gradients m_t. Although momentum can be considered a slope-running ball, Adam acts as a heavy friction ball, preferring flat minima on the error surface. We calculate as follows the decaying averages of m_t and v_t from past and past square gradients:

$mt = \beta 1 mt{-}1 + (1 - \beta 1)t$

$vt = \beta 1 vt{-}1 + (1 - \beta 1)t2$

$m_t$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 initialized 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 low (such that β_1& β_2 are nearly 1).



- Figure 2: Proposed model Overview

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

Table 1: Model Summary

conv2d_1 (Conv2D)	(None, 112, 112, 128)		73856
max_pooling2d (MaxPooling2D)	(None, 112, 112, 64)		0
batch_normalization (BatchNo	(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, 6)		24582

Total params: 134,291,014
Trainable params: 134,288,070 Non-trainable params: 2,944

IV. SIMULATION RESULTS

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.

A. DATASET DESCRIPTION

The SUN397dataset is used for public use. The dataset includes 108,753 images 397 categories that have been used in the scene understanding (SUN) benchmark. There are at least 100 images per category, but the number of images varies according to category. SUN397 is a wider scene benchmark of 397 categories like indoor, artificial & natural categories (at the least before places). This dataset is very demanding for many categories 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 2: PARAMETERS INFORMATION

Parameters	value
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

B. Results Analysis

This subsection represents the analysis of the results obtained by the proposed model.

Table 3: Comparison of accuracy, Loss, Val_loss and Val_Accuracy for Base and Proposed indoor/outdoor Model

Model	Loss	Accuracy	Val_loss	Val_Accuracy
Base Indoor	0.2623	0.9113	1.7929	0.5615
Base Outdoor	0.2097	0.9308	2.6702	0.5119
Proposed Indoor	0.2112	0.9366	1.4925	0.7469
Proposed Outdoor	0.0373	0.9891	3.8100	0.5019

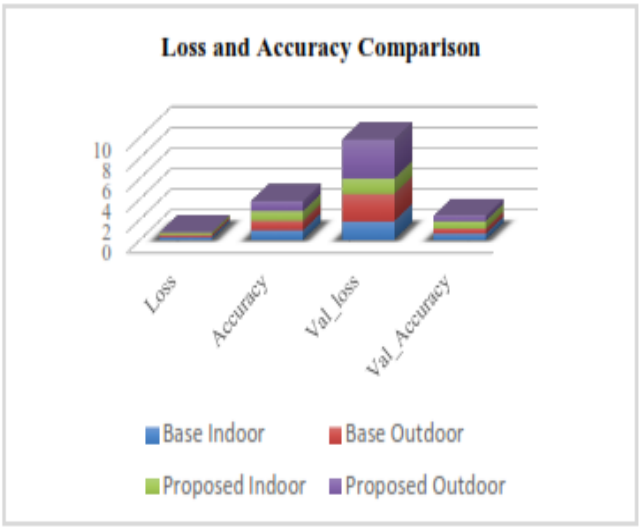


Figure 3: Graph comparison of accuracy, Loss, Val_loss & Val_Accuracy for Base and Proposed indoor/outdoor Model

C. Model Training Accuracy proposed results of Indoor and outdoor

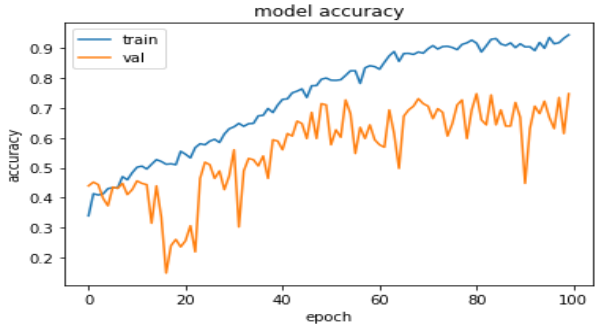


Figure 4: Line Graph for Indoor model accuracy

Figure 4 represents a line graph for Indoor model accuracy. This process continues up to 100 epochs. It shows training and validation accuracy. Initially, it starts training accuracy from 35%, which is gradually increased to 91% accuracy at 100 epochs. It also shows validation accuracy. Initially, it starts validation accuracy from 43%, which has a variable increment in accuracy, but it goes down 10% at 18 epochs. Then, it constantly increases approximately 51% accuracy.

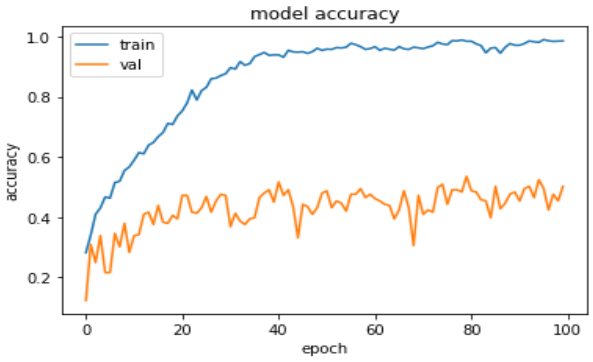


FIGURE 5: LINE GRAPH FOR OUTDOOR MODEL ACCURACY

Figure 5 represents a line graph for outdoor model accuracy. This process continues up to 100 epochs. It shows training and validation accuracy. Initially, it starts training accuracy from 33%, which gradually increases 82% at 40 epochs. Then it constantly increases up to 98.9 % accuracy. It also shows validation accuracy. Initially, it starts validation accuracy from 10%, which has a variable increment in accuracy. It constantly increases approximately 50% accuracy.

C. Performance Measurements

The following performance measurement is utilized to determine the proposed model's reliability.

1. **Accuracy:** Accuracy is the degree of closeness to the true value.

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FN + FP} = \frac{TP + TN}{P + N}$$

2. **Precision:** Precision is the degree to which an instrument or process will repeat the same value.

$$\text{Precision} = \frac{TP}{TP + FP}$$

3. **Recall:** It is a fraction of related documents that are fruitfully retrieved.

$$\text{Recall} = \frac{TP}{TP + FP}$$

4. **F1-score:** F-score is the harmonic mean of precision & recall.

$$F_\beta = \frac{(1 + \beta^2)(\text{Precision} * \text{Recall})}{\beta^2 * (\text{Precision} + \text{Recall})}$$

Here, True Positive (TP), True Negative (TN), False Positive (FP), False Negative (FN).

The following figure & tables show the confusion Metrix without normalizing indoor and outdoor scene labels.

Table 4: Comparison of confusion Metrix, without normalization of indoor scene classification labels of Precision, Recall, F1Score, and Support

	Precision	**Recall**	**F1-Score**	**Support**
airport_terminal	0.44	0.49	0.47	110
church	0.17	0.08	0.11	25
classroom	0.17	0.09	0.11	23
florist_shop	0.00	0.00	0.00	16
gymnasium	0.13	0.21	0.16	38
library	0.13	0.28	0.12	37
macro avg	0.17	0.16	0.16	249
weighted avg	0.27	0.28	0.27	249

Table 5: Comparison of confusion Metrix, without normalization of outdoor scene classification labels of Precision, Recall, F1Score, and Support

	Precision	**Recall**	**F1-Score**	**Support**
Amusement park	0.56	0.57	0.57	75
Botanical Garden	0.54	0.81	0.65	27
Crosswalk	0.20	0.05	0.08	20
Gas Station	0.45	0.38	0.41	34
Market	0.63	0.67	0.65	85
Temple	0.36	0.35	0.36	37

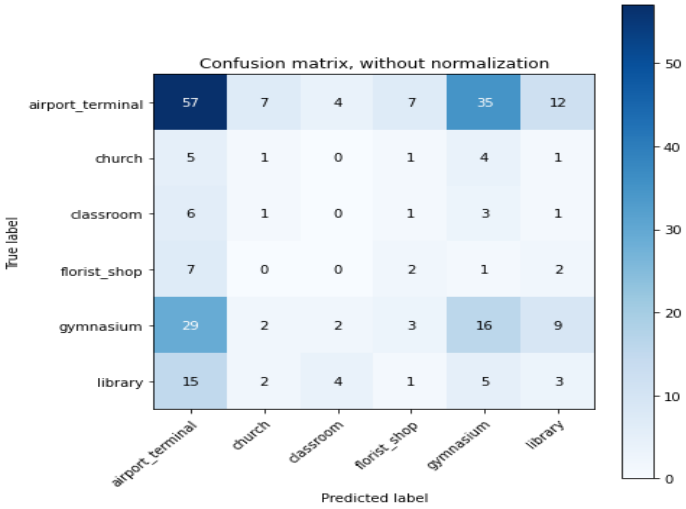


Figure 6: Line Graph for Indoor Confusion matrix without normalization

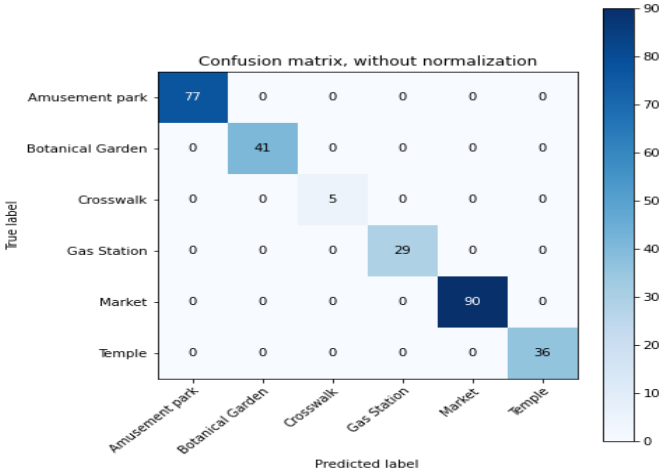


Figure 7: Line Graph for outdoor Confusion matrix without normalization

The diagonal of the matrix represents the correctly classified figure into their respective labels. The actual categories are represented in rows, and columns indicate the predicted label.

V. CONCLUSION

macro avg	0.46	0.47	0.45	248
weighted avg	0.51	0.54	0.52	248

In this article, the two-class scene classification model is clear and efficient, based on an indoor-outdoor classification proposed. Indoor/outdoor scenes have a significant feature used as a guideline to build that model: they have spatially repeating properties. Although the concept of indoor or outdoor scenes has been studied over the years, to our understanding, all the study was dedicated to a VGG-16 where both indoor and outdoor scenes are categorized. In order to explain images and important issues in computer graphics for geometry comprehension, scene understanding is still an extremely difficult issue in the computer view. The primary objective is to teach machines how to interpret scene classification. We also introduced a methodology that classifies all indoor/outdoor scenes and provides an indoor/outdoor scene label, unlike previous works. We then performed a comprehensive performance assessment on the SUN397 dataset for validating the model. We found from the test results that 93.66% accuracy of indoor classes & 98.91% accuracy of outdoor classes demonstrated the model's accuracy.

REFERENCES

[1] Aarthi, S., & Chitrakala, S. (2017). Scene understanding — A survey. 2017 International Conference on Computer, Communication and Signal Processing (ICCCSP). doi:10.1109/icccsp.2017.7944094

[2] Pawar, P. G., & Devendran, V. (2019). Scene Understanding: A Survey to See the World at a Single Glance. 2019 2nd International Conference on Intelligent Communication and Computational Techniques (ICCT). doi:10.1109/icct46177.2019.8969051

[3] Rucai Zhou, Kuojian Lu, Yi Long. "A Survey on Social Image Understanding," International Conference on Behavioral, Economic, Socio-cultural Computing (BESC), pp. 1-5,2018.

[4] Devi Parikh" Human-Machine Crfs For Identifying Bottlenecks In Scene Understanding" IEEE Transactions On Pattern Analysis And Machine Intelligence, Vol. 38, No. 1, January 2016

[5] C. Lawrence Zitnick, Ramakrishna Vedantam, And Devi Parikh. "Adopting Abstract Images For Semantic Scene Understanding," IEEE Transactions On Pattern Analysis And Machine Intelligence, Vol. 38, No. 4, April 2016

[6] Patel, T. A., Dabhi, V. K., & Prajapati, H. B. (2020). Survey on Scene Classification techniques. 2020 6th International Conference on Advanced Computing and Communication Systems (ICACCS). doi:10.1109/icaccs48705.2020.9074460

[7] Delu Zeng, Minyu Liao, Mohammad Tavakolian, Yulan Guo Bolei Zhou, Dewen Hu, Matti Pietikainen, and Li Liu," Deep Learning for Scene Classification: A Survey," arXiv:2101.10531v2 [cs.CV] 20 Feb 2021

[8] Mitisha Narottambhai Patel, Purvi Tandel. "A Survey on Feature Extraction Techniques for Shape-based Object Recognition," International Journal of Computer Applications, Volume. 137, No.6, March 2016

[9] Jeff Donahue, Lisa Anne Hendricks, Sergio Guadarram, Marcus Rohrbach. "Long-term Recurrent Convolutional Networks for Recognition and Description." IEEE Transactions on Pattern Analysis and Machine Intelligence, Volume: 39, Issue: 4, pp. 26252634, 2016.

[10] Jianhui Chen, Wenqiang Dong, Minchen Li. "Image Caption Generator Based On Deep Neural Networks." Springer, pp. 1-28, 2016.

[11] Sang Jun Lee, Sag Woo Kim. "Recognition of Slab Identification Numbers using a Deep Convolutional Neural Network," 15th IEEE International Conference on Machine Learning and Applications, pp. 1- 4, 2016.

[12] Matthew R Boutell, Jiebo Luo, Xipeng Shen, and Christopher M Brown. Learning multi-label scene classification. Pattern recognition, 37(9):1757–1771, 2004.

[13] Li-Jia Li, Hao Su, Li Fei-Fei, and Eric P Xing. Object bank: A highlevel image representation for scene classification & semantic feature sparsification. In Advances in neural information processing systems, pages 1378–1386, 2010.

[14] Ariadna Quattoni and Antonio Torralba. Recognizing indoor scenes. In Computer Vision and Pattern Recognition, 2009. CVPR 2009. IEEE Conference on pages 413–420. IEEE, 2009.

[15] Jianxin Wu and Jim M Rehg. Centrist: A visual descriptor for scene categorization. IEEE transactions on pattern analysis and machine intelligence, 33(8):1489–1501, 2011.

[16] Aditya Vailaya, M´ario AT Figueiredo, Anil K Jain, and Hong-Jiang Zhang. Image classification for content-based indexing. IEEE transactions on image processing, 10(1):117–130, 2001.

[17] Ritendra Datta, Dhiraj Joshi, Jia Li, and James Z Wang. Image retrieval: Ideas, influences, and trends of the new age. ACM Computing Surveys (Csur), 40(2):5, 2008.

[18] Lei Zhang, Mingjing Li, and Hong-Jiang Zhang. Boosting image orientation detection with indoor vs. outdoor classification. In

Applications of Computer Vision, 2002. (WACV 2002). Proceedings. Sixth IEEE Workshop on pages 95–99. IEEE, 2002.

[19] Sebastiano Battiato, Salvatore Curti, Marco La Cascia, Marcello Tortora, and Emiliano Scordato. Depth map generation by image classification. In Electronic Imaging 2004, pages 95–104. International Society for Optics and Photonics, 2004.

[20] Simone Bianco, Gianluigi Ciocca, Claudio Cusano, and Raimondo Schettini. Improving color constancy using indoor–outdoor image classification. IEEE Transactions on image processing, 17(12):2381–2392, 2008.

[21] Jack Collier and Alejandro Ramirez-Serrano. Environment classification for indoor/outdoor robotic mapping. In Computer and Robot Vision, 2009. CRV'09. Canadian Conference on, pages 276–283. IEEE, 2009.

[22] Li, Y., Zhang, H., Xue, X., Jiang, Y., & Shen, Q. (2018). Deep learning for remote sensing image classification: A survey. WIREs Data Mining and Knowledge Discovery, 8(6), [e1264]. https://doi.org/10.1002/widm.1264

[23] Zhehang Tong," A Review of Indoor-Outdoor Scene Classification," 2nd International Conference on Control, Automation, and Artificial Intelligence (CAAI 2017), Advances in Intelligent Systems Research, volume 134

[24] Jing Sun, Xibiao Cai, Fuming Sun and J. Zhang, "Scene image classification method based on Alex-Net model," 2016 3rd International Conference on Informative and Cybernetics for Computational Social Systems (ICCSS), Jinzhou, China, 2016, pp. 363-367, doi: 10.1109/ICCSS.2016.7586482.

[25] Yashwanth. A et al.," A novel approach for indoor-outdoor scene classification using transfer learning," International Journal of Advance Research, Ideas and Innovations in Technology, Volume 5, Issue 2,2019

[26] G. Memiş and M. Sert, "Detection of Basic Human Physical Activities With Indoor–Outdoor Information Using Sigma-Based Features and Deep Learning," in IEEE Sensors Journal, vol. 19, no. 17, pp. 7565-7574, 1 Sept.1, 2019, doi: 10.1109/JSEN.2019.2916393.

[27] I. Saffar, M. L. A. Morel, K. D. Singh, and C. Viho, "SemiSupervised Deep Learning-Based Methods for Indoor Outdoor Detection," ICC 2019 - 2019 IEEE International Conference on Communications (ICC), Shanghai, China, 2019, pp. 1-7, doi: 10.1109/ICC.2019.8761297.

[28] O. Sen and H. Yalim Keles, "Scene Recognition with Deep Learning Methods Using Aerial Images," 2019 27th Signal Processing and Communications Applications Conference (SIU), Sivas, Turkey, 2019, pp. 1-4, doi: 10.1109/SIU.2019.8806616.

[29] M. Ye, H. Zhong, X. Song, S. Huang, and G. Cheng, "Acoustic Scene Classification Using Deep Convolutional Neural Network via Transfer Learning," 2019 International Conference on Asian Language Processing (IALP), Shanghai, China, 2019, pp. 19-22, doi: 10.1109/IALP48816.2019.9037692.

[30] Z. Chen, Y. Wang, W. Han, R. Feng, and J. Chen, "An Improved Pretraining Strategy-Based Scene Classification With Deep Learning," in IEEE Geoscience and Remote Sensing Letters, vol. 17, no. 5, pp. 844-848, May 2020, doi: 10.1109/LGRS.2019.2934341.

[31] K. Abdullah et al., "A Machine Learning-Based Technique for the Classification of Indoor/Outdoor Cellular Network Clients," 2020 IEEE 17th Annual Consumer Communications & Networking Conference (CCNC), Las Vegas, NV, USA, 2020, pp. 1-2, doi: 10.1109/CCNC46108.2020.9045473