iJRASET
International Journal For Research in Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 11 **Issue:** X **Month of publication:** October 2023

DOI: https://doi.org/10.22214/ijraset.2023.56213

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 11 Issue X Oct 2023- Available at www.ijraset.com

Monument Tracker: Deep Learning Approach for Indian Heritage

Jacob Kuriakose[1], Dr. Pinki Nayak[2], Lokesh Meena[3], Jyoti Parashar[4]
Dr. Akhilesh Das Gupta Institute of Technology & Management

Abstract: ***Monuments are physical structures built to commemorate a person or event. Their importance to the region necessitates their documentation and upkeep. Due to the many variations in how various monuments are built, monument recognition is a challenging task in the field of picture classification. The various angles of the building are critical in identifying the monuments in photographs. As more international landmarks and monuments are covered, there is a greater need to connect a structure's physical presence to its digital presence. As a result, the monument's automated recognition is enabled. Monuments represent the culturally rich legacy of people of all ethnicities, castes, and faiths. It reflects tremendous achievements in art and architecture, and it also serves as the backbone of the surrounding region's socioeconomic progress through tourism. As an important historical and cultural heritage asset, the monument must be digitally recognized and archived. The monument photographs should be identified and described to aid in the preservation of people's cultures from various locations. The goal of this project is to present a method for classifying different monuments based on the characteristics of the monument photographs. Machine Learning and Deep Learning are advancing, speeding up advances in image recognition and allowing computer vision to reach new heights. The results with Baseline Model had an overall accuracy of 73.2%. After using Transfer Learning, we achieved an overall accuracy of 94.5% with VGG16 Architecture, Inception with an accuracy of 91.2% and Resnet50 with an accuracy of 85.5%.***

Keywords: Machine Learning, Deep Learning, computer vision, Transfer Learning, VGG16

I. INTRODUCTION

A monument is a tangible building or artwork made in honor of a person, an occasion, or a goal. With the development of machine learning and deep learning, image recognition is increasing, and computer vision is reaching new heights. Consequently, the monument's automatic identification becomes relevant. The effectiveness of the Deep Learning architectures for automating monument prediction(A. Saini, 2017) is examined and further enhanced. Transfer learning(Tammina, 2019) will be considered for image classification, and a few of the architectures like VGG16, ResNet50, and inceptionv3. After computation, the best model with higher accuracy will be chosen. Deep learning(Schmidhuber) is a subset of machine learning that uses artificial neural networks to model and solve complex problems. It has been used to achieve state-of-the-art performance in a variety of applications, including image recognition and computer vision. In the context of monument prediction, deep learning can be used to automatically identify and classify different types of monuments in images. One way to do this is by using a pre-trained model, such as VGG16, ResNet50(He , et al., 2015), or InceptionV3, and fine-tuning it on a dataset of images. This process is known as transfer learning, and it can significantly improve the performance of the model. Transfer learning can be used to train a model on a smaller dataset of images, by using the knowledge learned by the pre-trained model on a large dataset of images. This allows the model to learn from the patterns and features that are common among different types of monuments and improve its accuracy. After the training process, the model can then be tested on a separate dataset to evaluate its performance. The model with the highest accuracy will be chosen as the best model for monument prediction. This can be used to automatically identify and classify monuments(Ninawe, et al., 2020) in images, which can be useful in a variety of applications, such as heritage preservation and tourism.

II. LITERATURE SURVEY

In the realm of monument recognition, the pivotal role of deep learning techniques is underscored by a series of seminal studies. Saini and Thakur (2017)(A. Saini, 2017) showcase the effectiveness of Convolutional Neural Networks (CNNs) in their 'Image-based Indian Monument Recognition Using Convolutional Neural Networks.' The ability of CNNs to extract intricate features is highlighted as a critical factor in achieving accurate recognition. Similarly, He, Zhang, Ren, and Sun (2015)(He , et al., 2015) introduce the Residual Network (ResNet) architecture in 'Deep Residual Learning for Image Recognition,' a breakthrough that enhances recognition accuracy through efficient training.

Ninawe et al. (2020)(Ninawe, et al., 2020) employ TensorFlow, a powerful deep learning framework, to recognize cathedrals and Indian Mughal monuments, exemplifying the practical application of deep learning for heritage site identification. Simonyan and Zisserman (2014) introduce very deep convolutional networks, laying the foundation for multi-layered deep neural networks with enduring impact. Tammina (2019)(Tammina, 2019) demonstrates the utility of transfer learning with the VGG-16 architecture, showcasing its significance in leveraging pre-existing knowledge for accurate image classification. Additionally, Lyashenko (2022)(Lyashenko, 2022) underscores the vital role of data augmentation techniques in 'Data Augmentation in Python: Everything You Need to Know.'

This article highlights how data augmentation contributes to increased model robustness and generalization, which are paramount in the development of effective monument recognition models. Collectively, these studies and articles emphasize that the techniques and algorithms underpinning deep learning, along with data augmentation strategies, are pivotal in advancing monument recognition, offering accurate, efficient, and practical solutions.

III. METHODOLOGY

CNNs (Simonyan, et al., 2014)employ a type of multilayer perceptron that is supposed t need minimum pre-processing. An input and output layer, as well as several hidden layers, comprise a convolutional neural network. Convolutional layers, activation functions, pooling layers, fully connected layers, and normalizing layers are common components of a CNN's hidden layers.

1) *Convolutionlayer:* With each neuron computing a dot product between its weights and a small input volume region it is connected to, the convolution layer will calculate the output of neurons that are connected to local input regions.
2) *Poolinglayers:* By combining the outputs of neuron clusters at one layer into a single neuron at the following layer, pooling layers reduce the dimensions of the data. Small clusters are combined through local pooling, typically 2 x 2.
3) *Fullyconnectedlayers:* Through fully connected layers, every neuron in one layer is linked to every neuron in the following layer. It functions in a similar manner to a typical multi-layer perceptron neural network. To identify the photos, the flattened matrix is passed through a fully linked layer.Following the first convoluted layer, is a dropout layer that avoids overfitting. The feature maps are subjected to maximum pooling. This is followed by another convolution layer that generates 32 feature maps, a dropout layer, and a pooling layer. A single vector of pixel values is sent into the Artificial Neural Network, which is created by flattening the pooling layer. There are two fully connected layers, one with 1024 nodes and the other with 512 nodes, with a dropout rate of 20.

A baseline CNN (Convolutional Neural Network) model with 73.7% accuracy refers to a model that has been trained on a dataset using a CNN architecture and has achieved an accuracy of 73.7% on a test set. This is the starting point or "baseline" performance for the model, and any future improvements or changes to the model will be compared to this baseline. The baseline CNN(Indoliaa, et al., 2018) model can be a simple architecture, such as a shallow CNN, or a more complex architecture, such as a deep CNN with multiple layers and feature maps. The accuracy of 73.7% can be considered as a benchmark or a reference point to compare with other models or architectures. It's important to note that the accuracy of 73.7% can be considered as a relative performance, it could be considered as a good performance or a bad performance depending on the complexity of the problem, the quality of the dataset and the performance of other models.

a) When building a deep learning model, training data must be given to the learning algorithm, which is what deep learning algorithms do. The model artefact produced by the training process is referred to as a "deep learning model".
b) The target or target attribute, sometimes referred to as the correct response, must be present in the training data. The learning algorithm finds patterns in the training data that map the input data attributes to the target (the answer that you want to predict), and it outputs a deep learning model that captures the spatters.
c) You can use the deep learning model to get predictions on new data for which you donot know the target. For instance, suppose you wish to teach a deep learning model todetermine whether or notan email is spam.

A. *Transfer Learning*

In the realm of machine learning, transfer learning is a research topic. It saves the information obtained while addressing one problem and applies it to another, but similar, problem. For example, knowledge obtained when learning to distinguish cats may be applied to recognising cheetahs. Transfer learning is a deep learning approach that involves training a neural network model on a problem comparable to the one being solved. Transfer learning offers the advantage of reducing learning model training time and can result in decreased generalization error.

In machine learning, information or data obtained while resolving one problem is labelled, saved, and then used to another problem that is unrelated but yet appropriate.For instance, a machine learning algorithm that learns to recognise cars may subsequently transmit that knowledge to another machine learning model that is being created to learn to recognise other kinds of vehicles, like trucks.

B. VGG16 Architecture

VGG16(Tammina, 2019) is a 16-layer deep neural network, as its name suggests. With 138 million parameters in all, the VGG16 network is therefore quite large by today's standards. The fundamental draw of the VGGNet16 architecture is, however, its simplicity. The most crucial characteristics of convolutional neural networks are included in the VGG16 design. Consisting of small convolution filters, VGG networks. VGG16 comprises 13 convolutional layers and three fully linked layers.

Below is a description of the VGG architecture:

1) VGGNet receives an image input of 224224 pixels. By removing a 224224 square from the centre of each image submitted for the ImageNet competition, the model's developers were able to maintain a constant image input size.

2) Convolutional layers—the VGG convolutional filters have a 33 receptive field, which is the smallest conceivable. A 11-convolution filter is also used by VGG to linearly transform the input.

3) The next innovation by AlexNet for reducing training times is the Rectified Linear Unit Activation Function (ReLU) component.ReLU is a linear function that produces a matching output for positive inputs and zero for negative inputs. To maintain the spatial resolution after convolution, VGG has a predetermined convolution stride of 1 pixel.

4) Hidden layers—All of the hidden layers of the VGG network employ ReLU rather than Local Response Normalisation, like in AlexNet. With no improvement in overall accuracy, the latter method lengthens training sessions and uses more memory.

5) Pooling layers: After numerous convolutional layers, a pooling layer is added to help minimise the dimensionality and number of parameters in the feature maps that each convolution step produces. Given the quick increase in the number of possible filters from 64 to 128, 256, and finally 512 in the last layers, pooling is essential.

6) Layers that are totally connected—VGGNethas three layers that are entirely connected.Each of the first two layers contains 4096 channels, while the third layer contains 1000 channels—one for each class.

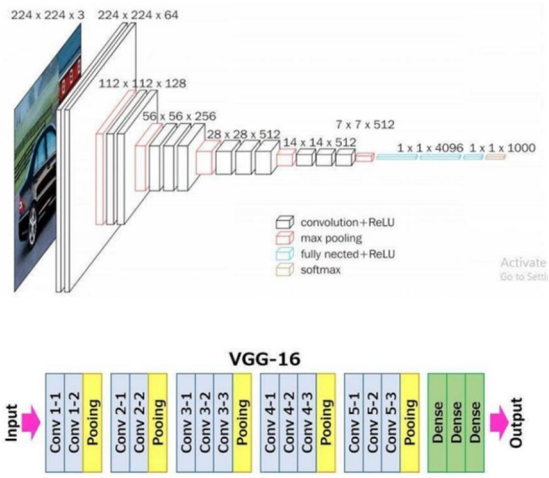


Figure 1: VGG16 Architecture

C. Data Analysis

Data is critical to the effectiveness of a deep learning model in much research. In the case of monument recognition, the availability and quality of data can have a significant impact on the model's performance. Kaggle, a platform for machine learning and data science competitions, is one common source of data. However, the data given on Kaggle may not be sufficient for training a deep learning model in some circumstances. Using web scraping to get extra data is one way to address this issue.

To enhance our deep learning model's performance, we utilized data from Kaggle and performed web scraping, as illustrated in Figure 2. This approach significantly augmented the dataset used for model training, leading to improved results. Table 1 offers a comprehensive overview of the gathered dataset.

Table 1: Training and Validation Data

S.No	Attribute Name	Training Image Count	Validation Image Count
1	Amber Fort	587	33
2	Brihadeeswarar Temple	654	43
3	Fatehpur Sikri	589	35
4	Golden Temple	799	84
5	Hawa Mahal	760	73
6	Humayun Tomb	749	61
7	Lotus Bahai Temple	674	41
8	MeenakshiAmman Temple	685	51
9	Mysore Palace	655	40
10	Purana Qila	645	36
11	Qutab Minar	708	54
12	Safdarjung's Tomb	656	38
13	Sanchi Stupa	648	34
14	Taj Mahal	756	64
15	Virupaksha Temple	740	59

Figure 2: Dataset Image Samples

D. Result and Discussion

Metrics are used to monitor and measure the performance of a model (During Training and Validation), and don't need to be differentiable. The loss is decreasing, and the accuracy is increasing in the desired way, The model is henceforth converging and achieving the goal with best accuracy possible on the available dataset.

```
Epoch 1/10
322/322 [==============================] - 8189s 25s/step - loss: 1.4127 - accuracy: 0.5735 - val_loss: 0.6902 - val_accuracy: 0.8378
Epoch 2/10
322/322 [==============================] - 145s 449ms/step - loss: 0.7543 - accuracy: 0.7685 - val_loss: 0.5210 - val_accuracy: 0.8539
Epoch 3/10
322/322 [==============================] - 144s 446ms/step - loss: 0.5930 - accuracy: 0.8157 - val_loss: 0.4534 - val_accuracy: 0.8606
Epoch 4/10
322/322 [==============================] - 144s 445ms/step - loss: 0.5021 - accuracy: 0.8443 - val_loss: 0.3847 - val_accuracy: 0.8901
Epoch 5/10
322/322 [==============================] - 144s 446ms/step - loss: 0.4341 - accuracy: 0.8637 - val_loss: 0.3713 - val_accuracy: 0.8887
Epoch 6/10
322/322 [==============================] - 143s 443ms/step - loss: 0.3819 - accuracy: 0.8811 - val_loss: 0.3431 - val_accuracy: 0.8995
Epoch 7/10
322/322 [==============================] - 143s 446ms/step - loss: 0.3448 - accuracy: 0.8926 - val_loss: 0.3153 - val_accuracy: 0.9142
Epoch 8/10
322/322 [==============================] - 143s 443ms/step - loss: 0.3148 - accuracy: 0.9002 - val_loss: 0.3344 - val_accuracy: 0.8887
Epoch 9/10
322/322 [==============================] - 142s 440ms/step - loss: 0.2932 - accuracy: 0.9105 - val_loss: 0.2977 - val_accuracy: 0.9129
Epoch 10/10
322/322 [==============================] - 142s 443ms/step - loss: 0.2629 - accuracy: 0.9176 - val_loss: 0.3133 - val_accuracy: 0.9048
```

Figure 3:Training Metric

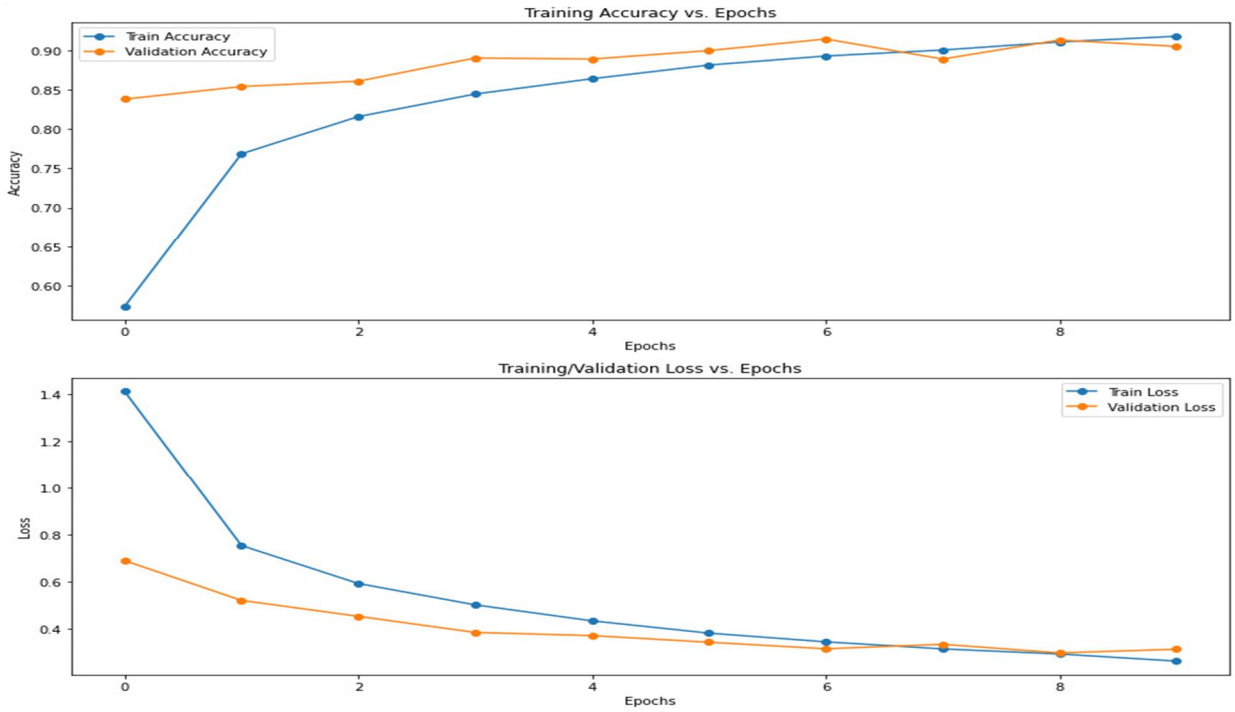


Figure 4:VGG16 Performance

A classification report Figure 5 is a machine learning performance evaluation metric. It serves as evidence of the trained classification model's accuracy, recall, F1 Score, and support.

1) precision = Only true Positive no false positive then the precision is 1.0 (False positive are worse than false negative).
2) recall = having no false negative then recall is 1.0 (False negative are worse than false positive).
3) F1 score = combination of precision and recall. The perfect model has 1.0.
4) support = the no. of sample each metric has calculated on
5) Accuracy = accuracy of a models
6) macro-avg = avg precision, recall, F1 Score between classes. macro-avg doesn't take classes imbalance into account class imbalance means the split of the values (if SUPPORT is 50 50 approx. the no class imbalance)

7) Weighted-avg = weighted-avg precision, recall, F1 Score between classes. weighted means each metric is calculated with respect to how many samples are there in each class (will give high value when one class have more samples that other)

	precision	recall	f1-score	support
Amber Fort	1.00	0.73	0.85	15
Brihadeeswarar Temple	0.77	0.91	0.83	11
Fatehpur Sikri	1.00	0.92	0.96	13
Golden Temple	1.00	1.00	1.00	28
Hawa Mahal	1.00	0.97	0.98	33
Humayun Tomb	1.00	1.00	1.00	18
Lotus Bahai Temple	1.00	1.00	1.00	18
Meenakshi Amman Temple	1.00	1.00	1.00	14
Mysore Palace	0.82	1.00	0.90	14
Purana Qila	0.65	0.92	0.76	12
Qutab Minar, Delhi	1.00	0.95	0.98	21
Safdarjung's Tomb	0.92	0.85	0.88	13
Sanchi Stupa	0.94	0.94	0.94	16
Taj Mahal	1.00	1.00	1.00	14
Virupaksha Temple	1.00	0.82	0.90	11
accuracy			0.94	251
macro avg	0.94	0.93	0.93	251
weighted avg	0.95	0.94	0.95	251

Figure 5: Classification Report

Resnet50: 85.5%, Inception: 91.2%, VGG16: 94.5%. VGG16 has the highest accuracy among all other Transfer Learning Models(Nwankpa, et al., 2018)Figure6 illustrates the comparative analysis of models.

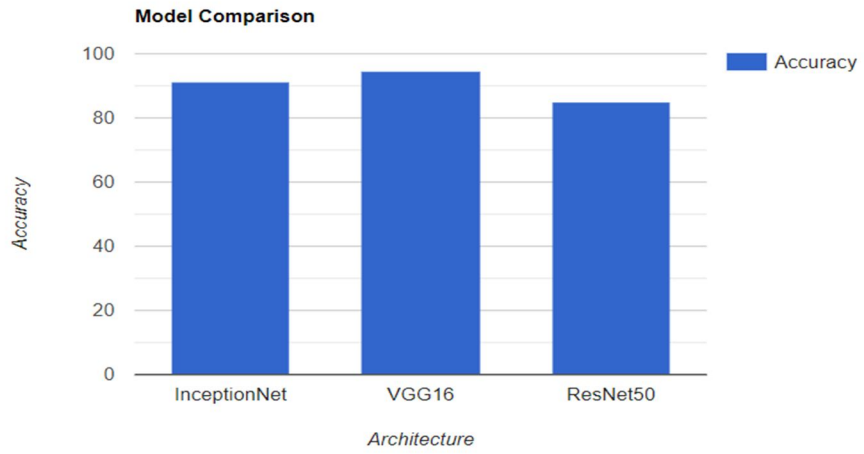


Figure 6: Model Comparison

E. Confusion Matrix

An error matrix, often referred to as a confusion matrix Figure 6(Bhandari, 2020),is a table that is frequently used to illustrate how well a classification model (or "classifier") performs on a set of test data where the true values have been established. It makes it possible to see how well an algorithm performs. It enables quick diagnosis of class labelling confusion, such as when one class is frequently mislabeled as the other. The confusion matrix's essential feature is that it counts both correct and incorrect guesses, breaking them down by class rather than just counting mistakes.

The correctness of our output from our training data was evaluated using a "Confusion matrix" analysis.

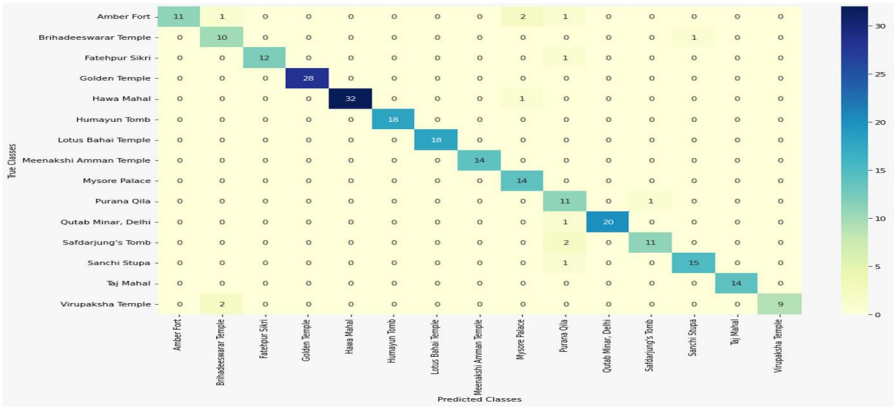


Figure7: Confusion Matrix

If we take an example of "Amber Fort" from Figure 7, we can observe that:
True Positives (TP) = 11
False Positives (FP) = 0
False Negatives (FN) = 4
True Negatives (TN) = 236
The above values are calculated as depicted in Figure 8.

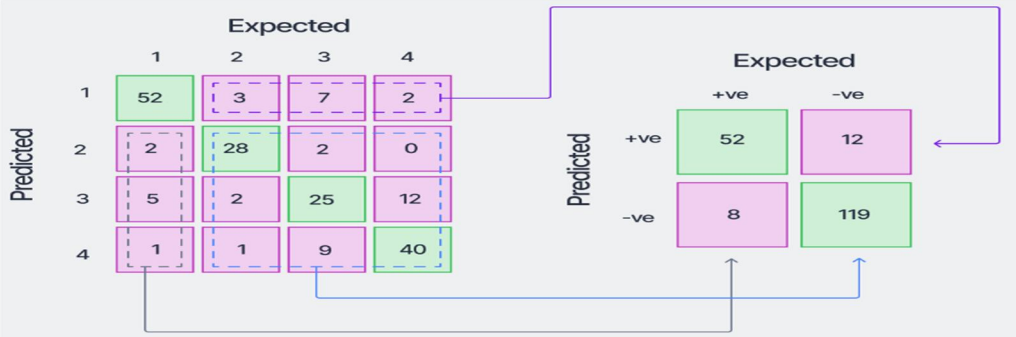


Figure 8: Calculating values of Confusion Matrix

When dealing with large amounts of data, such as in the -omics field, deep learning models are regarded cutting-edge for classification and grouping. Nonetheless, we are still a long way from having DL models for -omics data that can be utilized in precision medicine because the proposed approaches have yet to be verified in clinical practice. The success of DL is dependent on developing an architecture that fits the research topic and is capable of handling the data. Various DL approaches have been created throughout the years, making the selection of the best suited method a difficult road. For example, LSTM networks are a sophisticated form of RNN, capsNets attempt to address CNN restrictions like as data perspectives, and GANs by treating the unsupervised issue as supervised, promising results for automatically training a generative model may be obtained.

IV. CONCLUSION AND FUTURE SCOPE

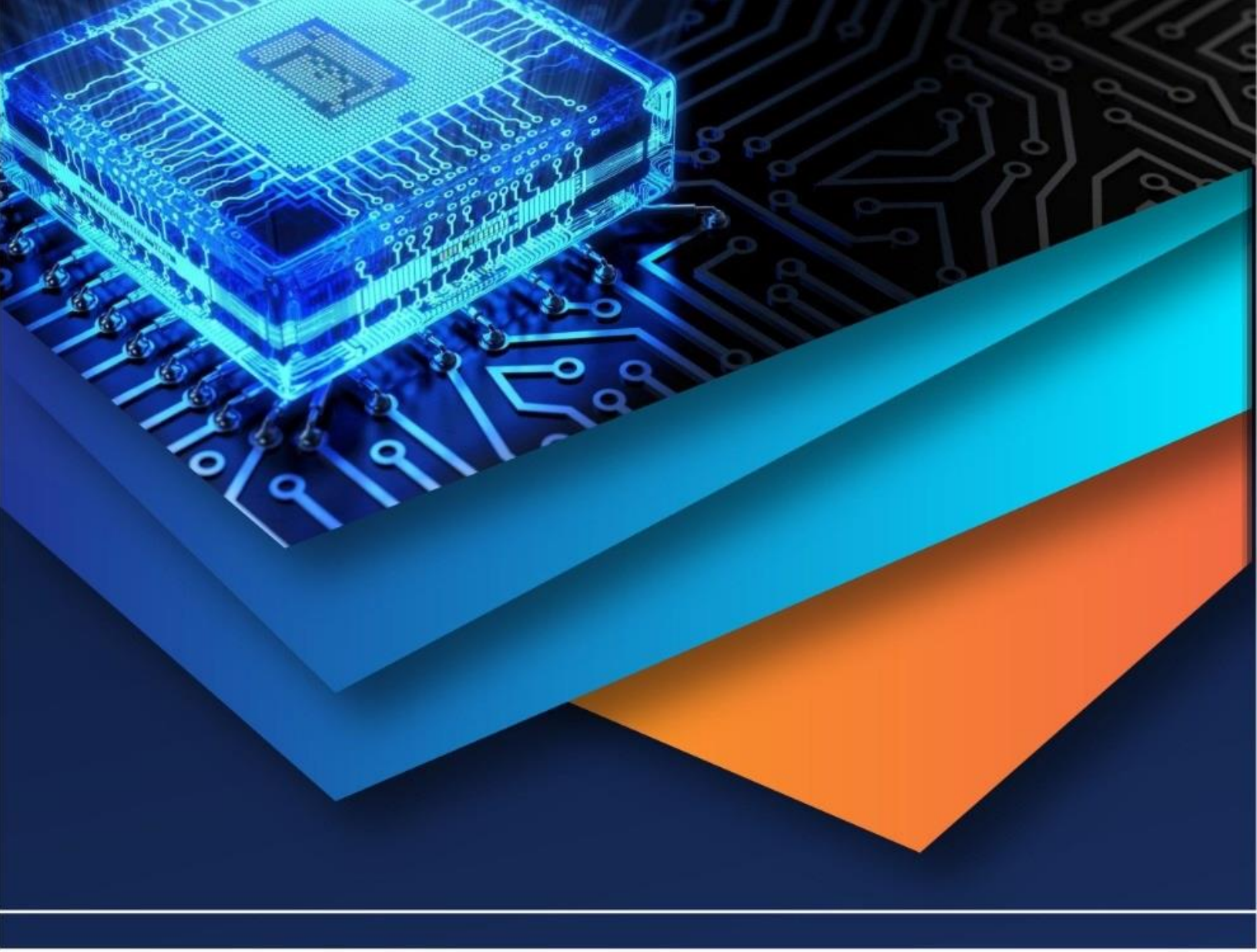
In conclusion, the experiment of using transfer learning for monument recognition has yielded impressive results. The baseline model, which was trained on a limited dataset, achieved a test accuracy of 73.7%. However, by using pre-trained models such as VGG16, ResNet, and Inception(Szegedy, et al., 2014) as a starting point, the test accuracy was significantly improved. The VGG16 model achieved a test accuracy of 94%, ResNet model achieved 85%, and the Inception model achieved 91.2%.
These results demonstrate the effectiveness of transfer learning for improving the performance of image recognition models. The high accuracy achieved by the pre-trained models suggests that they have learned useful features from a large dataset that are transferable to the monument recognition task. This approach can be highly beneficial for future work in monument recognition and other related tasks that may have limited data available and can help to improve the performance and accuracy of the model significantly.

There are several areas where deep learning techniques for monument recognition can be further developed and improved in the future:

1) Real-time recognition: Developing models that can recognize monuments in real-time, such as using mobile devices, can make the process more accessible and convenient for users.
2) Multi-modal recognition: Combining visual recognition with other modalities such as audio, text, and GPS data can improve the accuracy and robustness of monument recognition.
3) Multi-lingual recognition: Developing models that can recognize monuments in different languages can make the process more accessible to a wider range of users.
4) Handling variability: Monuments can vary greatly in terms of style, architecture, and condition, developing models that can handle this variability can improve the accuracy of recognition.
5) Augmented Reality: Developing models that can integrate with augmented reality technology can provide interactive and engaging experiences for tourists and educational.
6) Age and Damage detection: Developing models that can detect the age and damage of the monuments can help in their preservation and restoration. Developing Generative models such as GANs and VAEs, can be used to generate new images of monuments and to restore the images of damaged monument
7) Unsupervised and semi-supervised learning: Developing models that can learn from limited labelled data or unsupervised data can be useful in situations where labelled data is scarce or difficult to obtain.

REFERENCES

[1] A. Saini, T. Gupta, R. Kumar, A. K. Gupta, M. Panwar, and A. Mittal. 2017 . Image based Indian monument recognition using convoluted neural networks. [Online] 2017. https://www.researchgate.net/publication/328957743_Image_based_Indian_monument_recognition_using_convoluted_neural_networks.
[2] Bhandari, Aniruddha. 2020. Everything you Should Know about Confusion Matrix for Machine Learning. Analytics Vidhya. [Online] 4 17, 2020. https://www.analyticsvidhya.com/blog/2020/04/confusion-matrix-machine-learning/.
[3] Donahue, Jeff , et al. 2013. A Deep Convolutional Activation Feature for Generic Visual Recognition. [Online] October 6, 2013. https://arxiv.org/abs/1310.1531.
[4] He , Kaiming , et al. 2015. Deep Residual Learning for Image Recognition. [Online] December 10, 2015. https://www.cvfoundation.org/openaccess/content_cvpr_2016/papers/He_Deep_Residual_Learning_CVPR_2016_paper.pdf.
[5] Indoliaa, Sakshi , et al. 2018. Conceptual Understanding of Convolutional Neural Network- A Deep Learning Approach.[Online]2018.https://reader.elsevier.com/reader/sd/pii/S1877050918308019?token=632C4F02B083C6A263302456F7E31038B918FFA7A95D658BE613A675F8D7CB05AA1BD7B69F24E5F2B2067F4C734A51D4&originRegion=eu-west-1&originCreation=20230117170557.
[6] Lyashenko, Vladimir . 2022. Data Augmentation in Python: Everything You Need to Know . [Online] MLOPS blog, December 13, 2022. https://neptune.ai/blog/data-augmentation-in-python.
[7] Mhatre1, Mitali S , et al. 2015. A Review paper on Artificial Neural Network: A Prediction Technique. [Online] December 12, 2015. https://www.ijser.org/researchpaper/A-Review-paper-on-Artificial-Neural-Network--A-Prediction-Technique.pdf.
[8] Mikołajczyk, , Agnieszka and Grochowski , Michał . Data augmentation for improving deep learning in imageclassificationproblem.[Online]https://www.researchgate.net/publication/325920702_Data_augmentation_for_improving_deep_learning_in_image_classification_problem.
[9] Ninawe, Aniket, et al. 2020. Cathedral and Indian Mughal Monument Recognition Using Tensorflow. [Online] Springer Link, August 15, 2020. https://link.springer.com/chapter/10.1007/978-3-030-51992-6_16.
[10] Nwankpa, Chigozie Enyinna , et al. 2018. Activation Functions: Comparison of Trends in Practice and Research for Deep Learning. [Online] November 8, 2018. https://arxiv.org/pdf/1811.03378.pdf#:~:text=Activation%20functions%20are%20functions%20used,can%20be%20fired%20or%20not.
[11] Schmidhuber, Jürgen . Deep learning in neural networks: An overview. [Online] https://www.sciencedirect.com/science/article/abs/pii/S0893608014002135?via%3Dihub.
[12] Simonyan, Karen and Zisserman, Andrew . 2014. Very Deep Convolutional Networks for Large-Scale Image Recognition. [Online] Cornell University, September 4, 2014. https://arxiv.org/abs/1409.1556v6.
[13] Szegedy, Christian , et al. 2014. Going deeper with convolutions. [Online] September 17, 2014. https://arxiv.org/pdf/1409.4842.pdf.
[14] Tammina, Srikanth . 2019. Transfer learning using VGG-16 with Deep Convolutional Neural Network for ClassifyingImages.[Online]October10,2019.https://www.researchgate.net/publication/337105858_Transfer_learning_using_VGG-16_with_Deep_Convolutional_Neural_Network_for_Classifying_Images.
[15] Thite, Avinash . 2022. Introduction to VGG16 | What is VGG16? [Online] Great Learning, November 18, 2022. https://www.mygreatlearning.com/blog/introduction-to-vgg16/.



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY