Journal of Data Acquisition and Processing
ISSN: 1004-9037
https://sjcjycl.cn/
DOI: 10.5281/zenodo.98549421

HERITAGE IDENTIFICATION OF MONUMENTS USING DEEP LEARNING TECHNIQUES

Dr S Murugesan[1, a)], Dr N Ramshankar[1, b)], Hiba Mariam H [2,c)], Kalapoorani P [3,d)], Kalpika K [4,e)]

Author Affiliations

[1]Associate Professor, Department of Computer Science and Engineering, R.M.D Engineering College, Kavaraipettai, Tamil Nadu, India

[2,3,4] Student, Department of Computer Science and Engineering, R.M.D Engineering College, Kavaraipettai, Tamil Nadu, India

Author Emails

[a)] Corresponding Author: bsmurugesan@gmail.com

[b)]nrs.cse@rmd.ac.in

[c)]ucs19204@rmd.ac.in

[d)]ucs19213@rmd.ac.in

[e)]ucs19214@rmd.ac.in

Abstract. The goal of this work is to present a web application that uses machine learning methods to aid travelers in recognising Indian monuments. The process of digitizing this cultural heritage preservation involves collecting large datasets of images, cleaning the data, and training a model based on Indian Monuments using Pytorch. The model was trained using MobileNet V1 architecture for monument prediction and MobileNet V2 for satellite identification. By deploying this model into a Python Flask Web application, users can capture an image of a monument and receive information about it, such as nearby tourist spots, hotels, user ratings, price range, reviews, a description of the monument, official website, and a link to a 360-degree view. The results showed that MobileNet V1 achieved an accuracy of 97% for monument identification, and MobileNet V2 achieved 93.3% for satellite identification. Applying data augmentation to the latter model resulted in an accuracy of 95%. This paper offers a solution to the lack of proper entry and tourism facilities for less familiar monuments and the shortage of trained guides. It also provides an interactive learning process that can stimulate users' interest in cultural heritage.

Keywords—Indian Monuments, MobileNet , Satellite view, Web Application

I. INTRODUCTION

The preservation of cultural heritage is crucial in the age of digitization, and digital technologies are playing an increasingly important role in this effort. Machine learning techniques have been found to be effective in addressing preservation challenges, but they face obstacles such as dataset quality and size. Transfer learning architectures, which involve training deep convolutional neural networks on large image datasets and then applying them to smaller data, can help overcome dataset size problems. Machine learning approaches can also enhance heritage objects using image reconstruction methods. Convolutional neural networks, a type of deep learning algorithm, can improve monument recognition by identifying and distinguishing different aspects of objects in images. This technology can be used to create monument recognition applications that provide interactive learning experiences for tourists and locals alike. These applications can be used to obtain information about monuments, even in the absence of guides, and make the learning process more engaging and stimulating.

II. LITERATURE SURVEY

Recent advancements in deep learning, particularly in convolutional neural networks, which perform astonishingly well with computer vision, have opened up several possibilities and find use in a variety of industries. The monument is one application of this technical advancement, which is especially useful for those who travel the world and see new places. An efficient and practical way to research and learn about the nearby cultural tourist attractions with only a few clicks.

[1] A paper titled "Monument Recognition using Deep Neural Networks" was published in 2017 by Siddhant Gada, Prof. Purva Raut, Viraj Mehta, Chahat Jain, and Karan Kanchan.

In their effort, they use the Inception v3 deep learning architecture model to categorize famous Monuments of India. This architectural design consists of numerous distinct layers with the peculiar property of continuously analyzing the dataset and providing a better model. The Rectified Linear Unit (ReLU), in this example, which offers smooth performance and quicker training times, is an essential part of the neural network's effective train. Additionally, they introduce the concept of pooling and more specifically, the max pooling, a well-known machine learning strategy that consolidates the existence of features in an input image [8], strengthening a model. Using a retrained model with 400–600 pictures for each landmark, they were able to achieve a cross-entropy close to 0.067, a training accuracy of 99.4%, and a comparable testing accuracy estimated from 96–99% after 4000 iterations.

[2] V.Palma, “A Monument Recognition Mobile App”

Attempts to create a smartphone app that can recognise and classify monuments. The application's fundamental idea depends on deep learning techniques, particularly convolutional neural networks. The programme can identify architectural objects by using convolutional neural network algorithms. To further explain, the aforementioned mobile application has two main sections: the first allows you to upload the desired document and potentially related data, and the second uses machine learning software to provide details on monuments that are stored in an online database. He creates two different apps, one for Android OS and the other for Apple's iOS. This model can identify the necessary monument with an accuracy of 95% or better by giving 50-100 photos for each of the 46 different monuments and utilizing data augmentation techniques to enhance the dataset.

[3] Chakkrit Termritthikun, Raisarn Muneesawang, and Surachet Kanprachar at the "NU-LiteNet: Mobile Landmark Recognition using Convolutional Neural Networks' 'in 2018.

They describes a new CNN model, called NU-LiteNet, which is designed to improve accuracy and reduce the size of the model for mobile applications. The authors conducted experiments on two datasets, one with images of Singapore landmarks and the other with images of Paris landmarks, and compared the results with those of other models, including AlexNet, GoogLeNet, and SqueezeNet. They found that NU-LiteNet achieved promising accuracy scores than the other models, while also being smaller in size. The authors also developed an Android application using NU-LiteNet for mobile landmark recognition.

[4] Ajay Kumar Mallicka, Vikash Yadav, Aniket Ninawe, Hifzan Ahmad, Sah, Cornel Barna and Dinesh Kumar, , "Cathedral and Indian Mughal Monument Recognition using Tensorflow"

In doing so, they tried to resolve the issue of image classification by identifying if a monument or architectural construction belonged to the Indian Mughal or the Cathedral. To successfully enable the algorithm to learn potential changes between monuments of the same class, they conduct the research using 1000 monuments of both cathedral and Indian Mughal architecture. By utilizing convolutional neural networks to extract general information and get train weights, they are able to improve their performance by more than 7% and reach 80% accuracy.

[5] Aradhya Saini, Rajat Kumar, Tanu Gupta, Akshay Kumar Gupta, Ankush Mittal, Monika Panwar with the "Image based Indian Monument Recognition using Convolutional Neural Networks" published back in 2017

They are finding that recognizing
monuments is a difficult process since you have to take into account the vast variances and various orientations of the structures. Their endeavor is based primarily on the features that might be acquired from the monuments. They extract information to identify the monuments using manual and convolutional neural network techniques. The Histogram of Oriented Gradients (HOG), Local Binary Patterns (LBP), and (GIST) features methodologies make up the main idea behind the classification of monuments based on hand-crafted features. For the experiment, 50 photos were used for each of the 100 unique monuments in India.Famous classification algorithms including Support Vector Machine (SVM), k-Nearest Neighbor (KNN), and Random Forest were used in conjunction with a variety of different methodologies. Convolutional neural models perform better than hand-crafted examples, with accuracy estimated at 92.7%, according to the results. They also offer the idea of Graph-Based Visual Saliency (GBVS), a tool that can identify an image's salience and produce superior outcomes.

III. PROPOSED SYSTEM

The proposed system for monument classification uses two methods for classification: one using 2D images with MobileNet v1 and the other using satellite images with MobileNet v2. For the first approach, images of monuments are captured from various angles and distances, and MobileNet v1 is used for classification. The model is trained using a large dataset of monument images, and once trained, it can classify new images with an accuracy of approximately 97%. Additional features include recommendations for nearby tourist spots, hotels, user reviews and images, and links to official websites and link to 360-degree views of the monuments. The second approach uses satellite images to classify monuments based on factors such as location, size, and shape. MobileNet v2 is used for this method, allowing for classification in remote locations or areas that are difficult to access using traditional methods.

IV. SYSTEM ARCHITECTURE

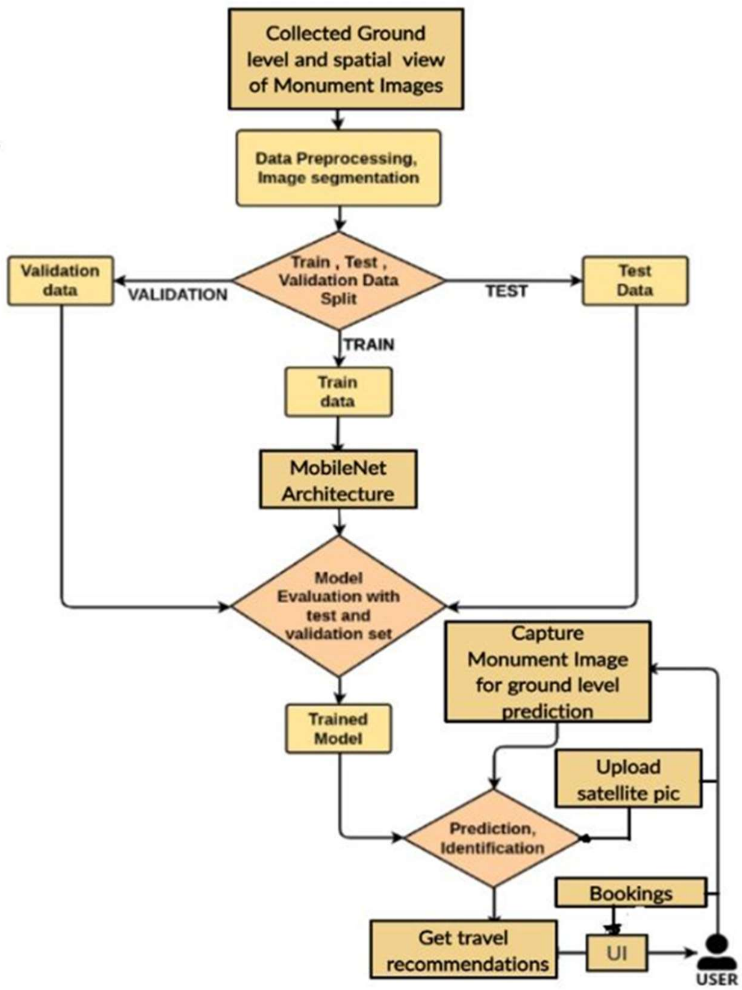

FIGURE 1 The system architecture diagram

V. THE METHODOLOGIES AND MODULES

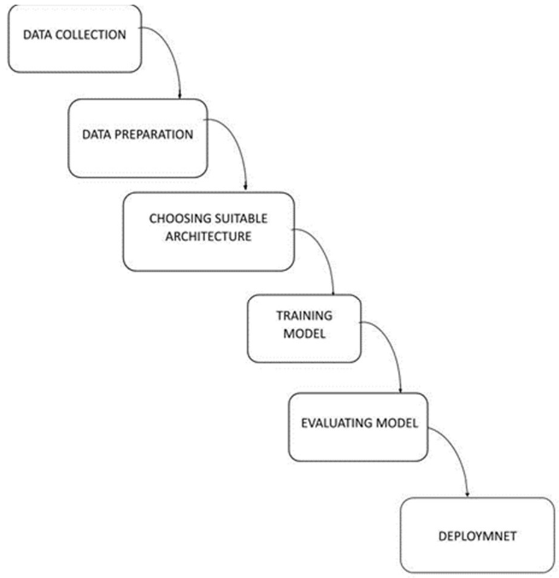

FIGURE 2 Methodologies and Modules involved

5.1 Data Collection

5.1.1 Street Level View of Monuments

We gathered pictures of the ten different Indian Historic Monuments for the research. We collect the dataset by manually collecting images of the monuments from the web because there was no existing dataset pertinent to our objectives. We make an effort to retain variety in our photographs because lighting and various monument angles are two aspects that directly affect a monument's coloration and possibly its shape. This approach ensures a higher likelihood of a reliable and accurate prediction, even in a variety of situations. https://kaggle.com/datasets/3d55dd9d1eb17 99118b66f1f84a9f432ef094d7a5bdfcaeec22ebf84f 779dfe4.

Serial Number	Monument Name	No of Images
1.	Hawa Mahal	269
2.	The Gateway of India	221
3.	Brihadeeshwarar Temple	123
4.	Taj Mahal	452
5.	Kaanch Mahal	125
6.	The Lotus Temple	112
7.	Se Cathedral	167
8.	The Sanchi Stupa	137
9.	The Mysore Palace	131
10.	Iron Pillar	229

TABLE 1. Count of images for each monument in the dataset

5.1.2 Satellite View of Monuments

Five renowned monuments of India were captured using Google Earth Pro's satellite imagery for the purpose of data collection. The aerial images were captured from a height of 200 meters, providing a clear and detailed view of each structure. The monuments include the

Lotus Temple, the Brihadeeshwara temple, the Taj Mahal, the Golden Temple, and Global Vipassana Pagoda.
https://kaggle.com/datasets/728ca2442933
d561da052f2aeea4835c08fa54e2de83c864f985d52 94b29dfa6

FIGURE 3 (a), (b),(c),(d) Sample images of collected street-level view dataset

5.2 Data Preparation

Pre-processing is a crucial stage in machine learning. In this subsection, we make sure that the data collection is done with a standard level of quality while also taking the required steps to get the data ready for the training section. This step involves Checking the quality, renaming the images, resizing the images (street view- (160,160),satellite view-(224,244) for street view of monuments and , and annotating the images. The dataset is split into two as train dataset and test dataset with a ratio 80:20. On dividing we obtained 1466 images in train and 367 images in test

5.2.1 Choosing Suitable Architecture

MobileNets is a series of very competent models that performs admirably, especially when utilized for mobile and other embedded devices. This specific class of models is built on an efficient architecture that makes use of depth-wise divisible convolutions to build a compact convolutional neural network that provides better performance and low latency.

5.2.2 MobileNet V1

MobileNets replace the conventional convolution layer with two new ones: a depthwise convolution and a point-wise convolution filter. This modifies the standard convolutional neural network design previously examined.

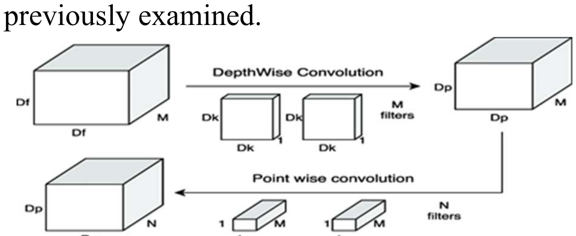

FIGURE 4 The working of Depthwise convolution

5.2.3 MobileNet V2

The MobileNetV1 architecture saw a few important upgrades that significantly improved the model's accuracy. The use of the ReLU6 activation function in place of ReLU and the addition of inverted residual blocks and linear bottlenecks were the key modifications made to the architecture. Even using satellite photographs, we were able to capture additional aspects of monuments with the help of the architecture's squeezing and stimulation.

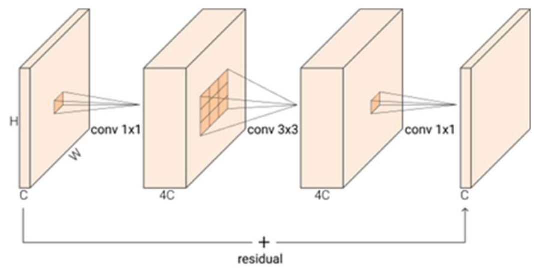

FIGURE 5 MobileNetV2 inverted bottleneck structure

5.3 Training the Model

In the training process, the model generates probabilistic outputs for each image, the probabilistic output represents the confidence of the model that the image belongs to a particular class. The performance of the model is measured using a loss function, which calculates the difference between the predicted and ground truth labels for each image. The optimizer helps to reduce the loss between the predicted and actual labels by updating the weights. The weights are then adjusted in the direction of the negative gradient after computing the gradients of the loss function with respect to the model parameters. This process is repeated for each batch of images in the training set, until the model's performance on the training set reaches a satisfactory level. By effectively training the model, on an average we got 99% accuracy for ground-level view of Indian monuments and 98% in spatial-view of the Indian monuments.

5.4 Evaluating the Model

Evaluating the model and metrics is a crucial step in the machine learning workflow. It helps you understand how well the model performs on the test dataset and determine whether it's suitable for the intended purpose.

Accuracy: The proportion of accurate predictions to total predictions is known as accuracy. It gauges how frequently the model is accurate in its forecasts.

Precision: The ratio of true positives to the total of true positives and false positives is known as precision. It counts the proportion of projected positive values that really are positive.

Recall: The proportion of true positives to the total of true positives and false negatives is known as recall. It counts the proportion of positive values that were really anticipated to be positive.

F1 Score: The harmonic mean of recall and precision is the F1 Score. It provides an equilibrium between recall and precision.

5.5 Deployment

A user-friendly web application is designed to provide almost accurate detection of the monuments. User opens the website and is displayed with two options

Identify the monuments

The user uploads the image of the monument. The model identifies the name of the monument with accuracy above 90% and displays relevant details and historic info about the monument.

Identify The Monument with Satellite Images

For this the user will have the option to upload a current top view image of the monument captured by satellite along with noise. The model identifies the monument with reasonable accuracy.

VI. RESULTS

On conducting a monument classification task on a dataset of images containing different types of heritage structures. We trained a deep learning model to classify the monuments into the correct type. Overall the accuracy is 97% for street-level monument classification and 95% for spatial view of monuments.

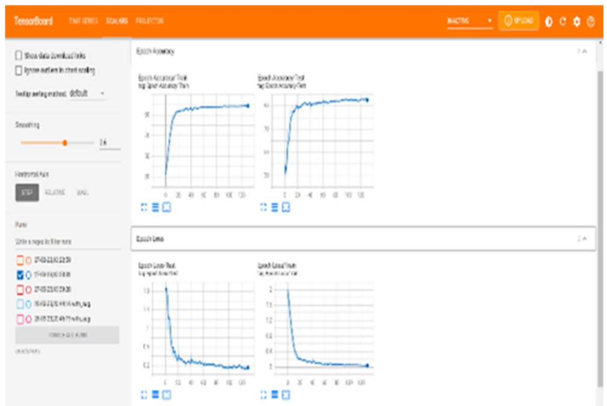

FIGURE 6 Visualization of metrics in tensorboard

VII CONCLUSION

Deep learning for heritage identification of monuments has emerged as a promising approach for heritage conservation and management. Our work has demonstrated the potential of this technology to accurately identify and classify different types of heritage structures. By training the model on a large and diverse dataset, we were able to capture the unique features and characteristics of each monument type. Although there are challenges such as capturing the full complexity of each monument and environmental factors, the use of deep learning for heritage identification is a promising approach that can revolutionize the field of heritage conservation and management. This technology can improve the accuracy and efficiency of heritage identification, which is critical for their preservation. Further research and development in this field could lead to more accurate and efficient methods of heritage identification, with broader applications in cultural heritage preservation

VIII REFERENCES

1. Gada, S., Mehta, V., Kanchan, K., Jain, C., & Raut, P. (2017). Monument Recognition using Deep Neural Networks. International Journal of Computer Applications, 167(1), 25-29.

2. Palma, V. (2019). A Monument Recognition Mobile App. In Proceedings of the 2019 IEEE Global Engineering Education Conference (EDUCON) (pp. 1671-1674). IEEE.

3. Termritthikun, C., Kanprachar, S., & Muneesawang, R. (2018, December). NU-LiteNet: Mobile Landmark Recognition using Convolutional Neural Networks. In 2018 IEEE 11th Conference on Service-Oriented Computing and Applications (SOCA) (pp. 212-215). IEEE.

4. Mallicka, A. K., Ninawe, A., Yadav, V., Ahmad, H., Kumar, D., Sah, H., & Barna, C. (2020). Cathedral and Indian Mughal Monument Recognition using Tensorflow. In 2020 12th International Conference on Computational Intelligence and Communication Networks (CICN) (pp. 47-51). IEEE.

5. Saini, A., Gupta, T., Kumar, R., Gupta, A. K., Panwar, M., & Mittal, A. (2017, December). Image based Indian Monument Recognition using Convolutional Neural Networks. In 2017 Second International Conference on Inventive Communication and Computational Technologies (ICICCT) (pp. 1422-1426). IEEE