A Deep Learning and Augmented Reality Framework for Automated Monument Recognition and Visualization

Dr. C Siva Balaji Yadav[1], R Dharani[2], P Durga Prasad[3],
Sangaraju Kavya Priya[4], S Mohammed Hassain Ahmed[5], M Aravind[6]
[1]*Professor, Department of Artificial Intelligence and Machine Learning, Annamacharya Institute of Technology & Sciences, Tirupati, India*
[2,3,4,5,6]*Student, Department of Artificial Intelligence and Machine Learning, Annamacharya Institute of Technology & Sciences, Tirupati, India*

***Abstract*—Historical landmark detection is a thrilling and very difficult task to study in the context of image classification by combining monument recognition with Augmented Reality (AR). The goal is to create an Android-based application that combines the accuracy of deep learning with the immersion provided by AR, which will provide users with an entertaining and educational experience when learning about heritage locations. Training the model with a variety of data on images of monuments will enable the system to conduct reliable real-time recognition and classification. This trained model is the engine of the Android application that allows users to just aim their phone camera on a monument and then it gives them the identification of the monument along with the historic background of the monument. Users have an option to take up live pictures using the mobile camera or use pictures in their gallery. This virtual experience comprises of informative storytelling about the monument, multimedia functionalities in the form of live map positioning, AR based graphics, and 3D models which have been effectively used to recreate historical settings. In general, the combination of CNNs and AR does not only enhance the discovery of historical sites but also expands the reach towards achieving cultural heritage since it is now more interactive, informative, and accessible to people.**

***Index Terms*—Deep Learning, Convolutional Neural Networks (CNN), VGG16, SVM, LSTM, Augmented Reality.**

I. INTRODUCTION

1.1 Background And Motivation

Monuments are a cultural, historical and architectural heritage of the human civilization, as sources of permanence in the identity and heritage of the society. These structures are necessary in education, tourism and cultural conservation, but conventional mechanisms of identifying monuments are mainly manual guides and are not interactive. As digital technologies are progressing at a significant pace, automated monument recognition gains more significance. One of the most effective image classification and object recognition tasks, especially Convolutional Neural Networks (CNNs), have shown great success in extracting complex visual features on image recognition. Through transfer learning, it is possible to identify the locations of monuments well despite having small data sets. The current project will aim at creating a robust deep learning-based monument recognition system that takes real-world images through preprocessing algorithms (resizing, normalization, and data augmentation) to enhance the performance of this system. Moreover, augmented reality makes the interaction with the user more interactive, as the information about history and architecture is projected onto the identified monuments. Some of the challenges that the proposed system will deal with include changes in lighting, viewing angle, and background noise, which contributes to a better recognition. Finally, the methodology will lead to smart cultural heritage information and engagement by the users in the process of learning.

1.2 Objectives

1.2.1 Design an Accurate Deep Learning-Based System to Identify Monuments

The major goal of the project is to design and deploy the correct system of monument recognition with the help of the cutting-edge methods of deep learning, especially Convolutional Neural Networks (CNNs). The system is trained using an eclectic dataset of the images of monuments when they are in different conditions to guarantee robustness and high classification rate. Through transfer learning, the model has brought shortcomings of the conventional image recognition models to their knees and effectively recognizes monuments in real- world images.

1.2.2 Augmented Reality Recognition of the Monument in Real-time

In addition to image recognition, the system is supposed to assist in real-time monument recognition and visualization. After locating a monument, augmented reality is applied to provide historical, architecture and cultural information directly over what is viewed in the real world. This goal increases the interest of the users and allows them to learn interactively, which will be especially helpful in the applications of tourism, education, and cultural exploration.

1.2.3 Cultural Heritage Preservation and Digital Documentation

One of the main aims of the project is to make a contribution to the preservation of the culture heritage with the use of intelligent digital. By automating monument identification and information retrieval, the system eliminates the use of manual records and enhances accessibility of heritage knowledge. This is a strategy that assists the historians, researcher and preservation agencies to conserve monuments to be used by later generations.

1.3 Scope

1.3.1 Attention capture on Monument Recognition based on Deep Learning methods

The research area in this study aims at identifying and categorizing monuments with deep learning models based on CNN. The system can handle issues like changes in lighting, viewing, background clutters and partial occlusion. Image preprocessing tools used to enhance model generalization and accuracy include resizing, normalization and data augmentation.

1.3.2 Creation of a Real-Time Monument Identification System

The project involves the creation of a real time system which is able to detect monuments by taking photos by the user. The fact that the information on recognized monuments is handled efficiently will result in a fast response rate and therefore the system can be used in real world application like mobile based tourism and education systems.

1.3.3 User-Centric and Informative System Design

The system is designed to be user-friendly and accessible to a wide range of users, including tourists, students, and researchers. A simple interface ensures ease of use, while augmented reality visualization enhances understanding by presenting contextual information in an interactive manner.

1.3.4 Future Integration and Expansion

Even though the current system is based on the fixed set of monuments, there is a potential of future advancement in the architecture. They can be trained on additional monuments and can be matched with larger cultural heritage databases and can be extended to multilingual content and other more complex AR/VR systems in order to offer improved user experiences.

II. LITERATURE SURVEY

2.1 Traditional Of Monument Identification

Historically, monuments were identified and classified using text-based techniques, classical computer vision techniques and by hand. These processes included visual sensitivity of analysts, guide books, GPS tagging and crude process of image processing methods which included edge finding, color histograms and shape-based feature finding. Despite the fact that these techniques were the foundations of the digital heritage documentation, they possessed several flaws that constrained their precision, extension and applicability:

2.1.1 Reliance on Paperwork and Human Knowledge

Historians, archaeologists or tour guides highly depended on manual identification of monuments in order to identify and describe monuments. This was both subjective and time consuming as well as prone to human error. Besides, the expertise of the professionals also was not available widely and that made it infeasible in large numbers.

2.1.2 Environmental and Image Sensitivity

Classical computer vision techniques were extremely sensitive to the change in lighting conditions, weathering, angle of view and background clutters. This variation of the quality of the image often led to false or inaccurate recognition and reduced validity in a real-world setting of outdoor tourism and image recording using mobile devices.

2.1.3 The Dependance on the Handcrafted Feature Extraction

Past automated systems involved manually coded features such as edges, textures, shapes or color distributions. These features were developed on the basis of domain knowledge, as well as, a lot of fine-tuning and in most instances did not generalize in various monuments with complex architectural designs.

2.1.4 Scalability and Flexibility of Lack

There was lack of flexibility and scalability in the old systems. The introduction of new monuments would need to be redesigned with feature sets or would need to re-design the whole system. These methods struggled with large and diversified datasets of monuments, and therefore could not be used to serve dynamic, large cultural heritage applications.

2.2 Advances In Deep Learning for Monument Recognition

CNNs and other deep learning algorithms have widely enhanced the activity of recognizing monuments in that the complicated visual representations are taught automatically by utilizing images.

Introduction of CNNs and Transfer Learning: CNN-based models like Transfer learning with decreasing training time and increasing the accuracy.

Visual Variability in Monument Images: Deep learning is able to handle the lighting variations, scale variations, point of view variations, occlusion, and background clutter variations.

Comparative Performance Over Traditional Methods: Better Performance When Compared to Classic Methods of machine learning like Support Vector Machines (SVM).

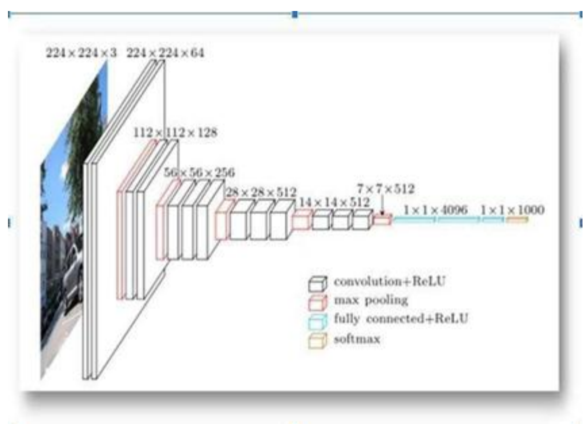
2.3 Applications And Challenges of Monument Recognition

Applications Across Domains: Monument recognition systems are actively used in other fields like tourism and cultural heritage preservation, education, and smart city programs. The automated detection of monuments facilitates digital access to information in tourists, digital archiving of monuments in historians, and learning in interactive augmented reality.

Difficulties in Implementation: Although the view of monuments is considerably developing, they have some difficulties. These are differences in the lighting conditions, shifts in viewpoint, obstruction, clutter in the background and structural similarities among monuments. Also, there are low availability of big well-labeled datasets of image monuments limiting model generalization and scalability.

Desire of Strong, Scalable System: To cope with these issues, the current monument recognition systems use powerful preprocessing methods which include image normalization, data augmentation as well as feature optimization. The CNN architecture and transfer learning are considered to enhance accuracy but with minimal latency at the cost of elaborate CNN architecture.

III. METHODOLOGY



3.1 Dataset Preparation

The data set is an essential part of the monument recognition system, which is composed of the images of various monument types. The images were taken at different conditions of different viewpoints, lighting, scales, and background environments to enhance generalization. Data augmentation techniques were used to get better diversity and decrease overfitting. These are arbitrary rotations to an extent of 30 degrees, horizontal flipping, and zooming with within a range of +20% to -20%. Positional variations were overcome by using spatial translations. Brightness and contrast adjustments guarantee good response in the various lighting conditions in the real world.

3.2 System Architecture

The suggested monument recognition system is based on the deep learning pipeline that includes an input phase, feature extraction, classification, and output phases. It starts with an input of a 224x224 RGB monument image which is preprocessed into a resized and normalized image in order to be consistent with the VGG16 architecture. The pre-trained VGG16 network consists of multiple convolutional and pooling layers where the low-level and high-level architecture features are referred to as edges, textures, and structural patterns are extracted out of the preprocessed image. These features are further obtained and flattened and processed by fully connected layers to engage in high-level reasoning and classification. Transfer learning is used to enhance the accuracy and shorten the training time through taking advantage of the pre-trained weights. The last thick layer is the one that makes the prediction and it determines whether the input image is a known class of monument or not. The output is presented by means of Monument Predicted or Monument Not Predicted, which allows identifying monuments accurately and effectively.

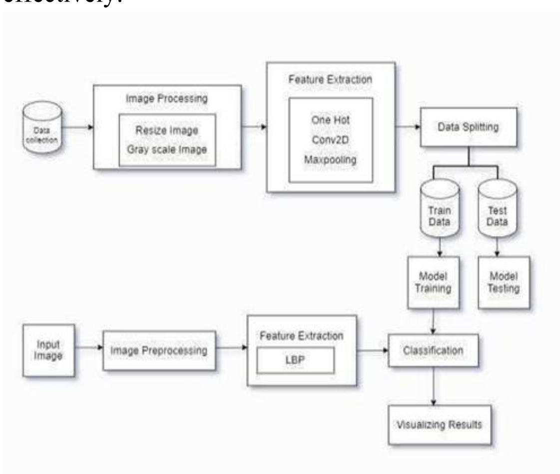


Figure 1: System Architecture

3.3 Deep Learning Model

The deep learning model is the primary element of the proposed monument recognition system because it is the one that extracts significant visual information about images. It has a tendency to acquire patterns by default-like shapes, textures, and architectural features. The model uses CNNs one layer at a time to enhance the recognition ability. This will allow the system to detect monuments speedily and precisely.

3.3.1 Model Architecture

The system uses convolutional neural network that uses the pre-trained VGG16 structure in the classification of monuments images. It uses RGB images of 224x 224 pixels to meet the VGG16 training needs. The convolutional layers inherently detect the hierarchical features like edges, textures and architectural patterns of the monument images. The convolutional base is pre-trained and then the feature extractor is used followed by the flattening of the feature maps. These characteristics are fed by fully connected layers that use ReLU activation, and dropout is used at a rate of 0.5 to minimize overfitting. Lastly, there is a softmax output layer that identifies the input image as one of the categories of monuments.

3.3.2 Compilation and Training

The loss function was categorical cross-entropy and the Adam optimizer was applied to the model with an initial learning rate of 0.001 and accuracy as the main evaluation measure. The training takes 50 epochs with the batch size of 32 to make sure the process is effective and the learning is stable. Transfer learning saves a large amount of time during the training process and enhances the accuracy to recognize monument images datasets.

3.4 Training And Validation

To guarantee the validity of the performance assessment, the dataset was divided into 20% and 80% validation and training sets respectively. Training was augmented with data to enhance the resistance to changes in images of monuments. The accuracy and loss of validation were continuously monitored and to optimize learning early stopping and reduction of learning rate were applied. The confusion matrix and classification report were used to assess the final model performance through classes of monuments.

Sl. No	Monument	No. of images
1	Charminar	253
2	Chotta Imambara	233
3	Alai Darwaza	221
4	Golden Temple	230
5	Alai Minar	272
6	Basilica of Bom Jesus	258
7	Tajmahal	247
8	Ajanta Caves	234
9	Tanjavur Temple	230
10	Victoria Memorial	234

3.5 User Interface

This user interface is interactive and user friendly with the non technical users who include tourists and students. It enables the user to either upload a monument picture or the user can use a live camera to recognize a monument. The interface has good buttons like the Upload Image and the Start Camera and is supported by responsive designs on the desktops, the laptops and the mobile devices.

IV. IMPLEMENTATION

4.1 Tools and Technologies

Implementation phase involves the development of the project in a very simple and rational manner. The machine learning model is implemented and trained on the basis of Keras and TensorFlow, the backend is made with the help of Django which links the model with the database. The frontend is built based on HTML, CSS, and JavaScript to simplify the system. The connection and testing of all parts is then performed.

4.2 Code Overview

The monument recognition system based on the deep learning is implemented into three primary parts: prediction, data preprocessing, and model construction and training.

4.2.1 Loading Data and Preprocessing

TensorFlow is used to load in the monument image dataset and OpenCV and the Python Imaging Library (PIL) are used to handle the images. All the images are translated into RGB format, reduced to 224x224 pixel size as per vgg16 requirements and normalized by scaling its pixel values to 0-1. The data is transformed into NumPy arrays, encoded the class labels and divided the data into 80% training data and 20% validation data.

4.2.2 CNN Model Constructing and Training

The Keras CNN model is a network that is built with the TensorFlow back-end on the base of pre-trained VGG16. VGG16 has a convolutional base that is an extractor of features and then there are fully connected layers and dropout layers to minimize overfitting. During training, data augmentation methods are used, i.e. rotation, flipping, zooming, and shifting to enhance generalization. It is optimized with categorical cross-entropy loss and Adam optimizer and the trained model is subsequently saved to be used in the future.

4.2.3 Categorization and Prognosis

Users can upload a monument image by using a Flask-based web interface or take one with the help of a live camera. The same preprocessing procedures are applied on the input image and then the trained model is run on the image to give a prediction. The system identifies the image into the respective class of monument and presents the predicted outcome to the user.

V. RESULT AND DISCUSSION

5.1 Model Performance

The monument recognition model was tested with the pre-trained VGG16 architecture where the validation accuracy was 96.2% with a training duration of 50 epochs. The training and validation accuracy and loss curve demonstrated smooth convergence of curves, which is evidence of good learning with minimum overfitting. Image augmentation approaches, including rotation, flipping, zooming, and shifting, contributed greatly to the strength of the model in the sensitivity to the changes in viewpoint, scale, and lighting situations that are typical of the images of monuments in real life. Transfer learning allowed the extraction of features of complex architectural patterns efficiently and shortened the training time and enhanced generalization.

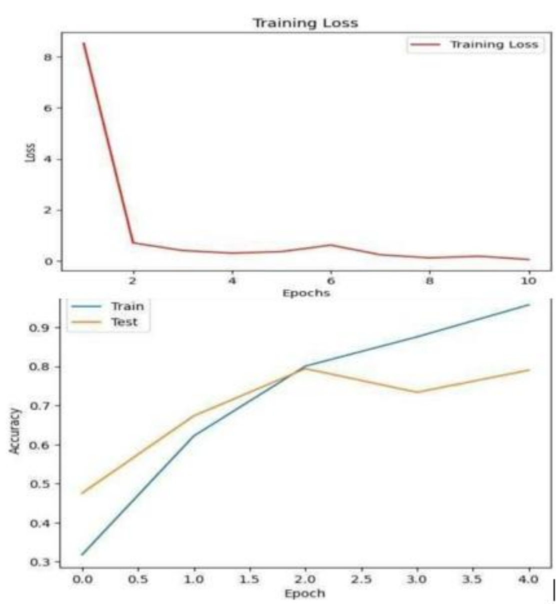


Figure 2 Training and Validation Accuracy

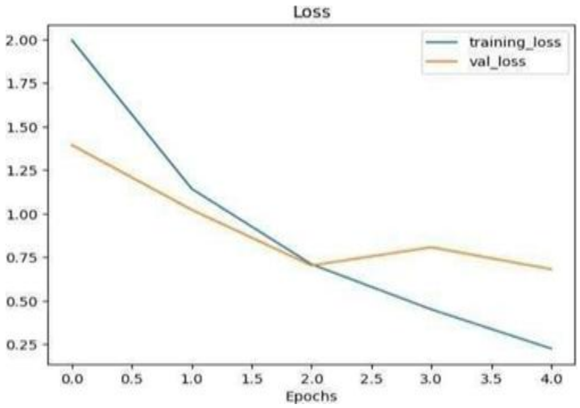
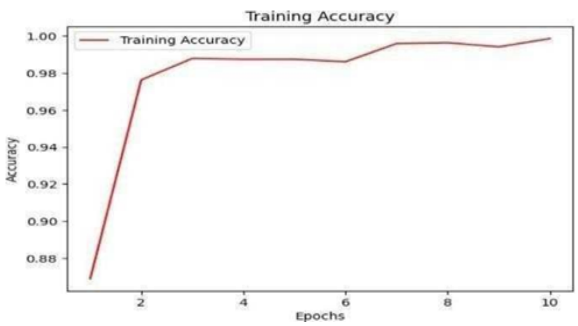


Fig Loss curve of CNN



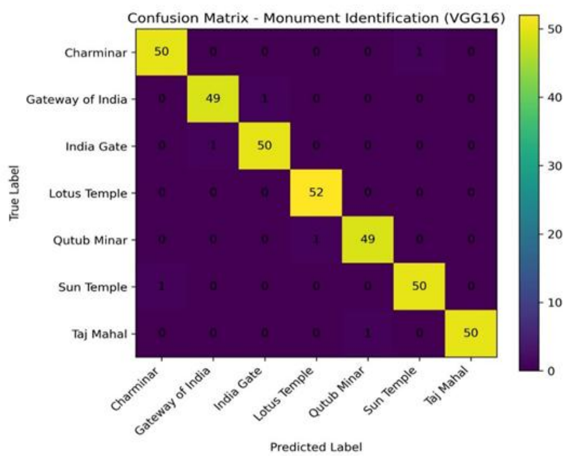


Figure: Confusion Matrix

Sl. No	Model	Accuracy	Precision	Recall	F1 Score
1	VGG16	98.01	95	94	94
2	CNN	95.34	86	83	84

5.1.2 Table Ii: Statistics Concluded by All Models

The confusion matrix has shown the high values of precision, recall, and F1-score in most of the classes of monuments, which proves the good performance of classification. Misclassifications were noted in minor misclassifications among monuments that had similar architectural designs or appearance patterns implying the difficulty in differentiating very close heritage sites. The mentioned errors could be minimized further by increasing the amount of data, including images of a higher resolution, or with the help of other contextual information. In general, the findings confirm the usefulness of the proposed deep learning-driven system of monument recognition in the real world.

Figure 5 Output Screenshots

5.2 System Usability

The proposed monument recognition system exhibits great usability by its easy and user-friendly web-based interface that is easy to use by non-technical users. The system is convenient to use by tourists, students, and other users as they can easily upload a monument image or inspect the real time predictions with the help of the live camera. The quick response of the app is supported by the efficient model based on VGG16 with a minimum delay between the input image and the display of the results. The fact that there is a clear visual feedback and clarity in cases where image formats are not supported or there is ambiguity in the user input is beneficial in enhancing user guidance and handling. The responsive design enables the easy functionality of desktops, laptops, and mobile devices. In general, the system is a good and convenient experience to work with, and monument identification is available and can be used in real-life situations.

5.3 Comparison With Traditional Methods

Conventional methods of the identification of monuments have depended on hand inspection, guide books, GPS tagging as well as classical computer vision strategies with the use of features like edges, shape and color histograms that had to be handcrafted features. These methods were very sensitive to lighting variations, perspective changes, clutter of the backgrounds and demanded major domain knowledge of designing features. By contrast, hierarchical and discriminative features are automatically learned by deep learning-based methods of Convolutional Neural Networks which only look at the raw monument images. The CNN-based models especially with transfer learning like that of VGG16 have better accuracy, scalability and resistance to real-world conditions. Such an end-to-end learning approach removes the need of manual feature engineering and is much more effective at performing monument recognition on a variety of data sets.

5.4 Future Work

The further development of the monument recognition system can be concentrated on multiple aspects of better performance and applicability. Generalization will be enhanced by increasing the dataset of monuments in other areas and architectural designs. Classification can also be improved by incorporating additional information like resolution of the image and contextual information like location. The functionality of Augmented Reality can be expanded to multi-lingual historical materials and interactive 3D visualizations. Also, lightweight deep learning models could be used to implement effective processing on mobile and edge machines. Video-based monument recognition and real-time AR navigation which may be considered in future work can also be used to add more to user experience.

VI. CONCLUSION

This project has managed to come up with a deep learning-based monument recognition system based on a Convolutional Neural Network with transfer learning. This was made possible by the fact that the pre-trained VGG16 architecture made the process of extracting features to be effective and yield high accuracy rates in classifying image datasets of monuments. The techniques of data augmentation enhanced resistance to changes in lighting, scale, and perspective guaranteeing high-quality real-world work. The web interface is user-friendly and enables real-time identification of the monuments and images upload allowing the system to be accessible by non-technical users. Although some minor cases of wrong classifications are present in cases of a similar architectural pattern in monuments, the general system is very accurate and useful. The findings offer validation of the usefulness of deep learning in applications of cultural heritage preservation and intelligent identification of monuments.

REFERENCES

[1] B. Maden and A. Melzer, "Towards Deep Learning for Architecture: A Monument Recognition App in a Mobile Setting," ISPRS Archives, vol. XLII-2/W9, pp. 551–556, 2019. doi: 10.5194/isprs-archives-XLII-2-W9-551-2019.

[2] Khan et al., "Identification of Monuments of Heritage Using CNN," International Journal for Research in Applied Science and Engineering Technology (IJRASET), 2023. doi: 10.22214/ijraset.2023.50168.

[3] Khan et al., "Heritage Identification of Monument Using Deep Learning," International Journal for Research in Applied Science and Engineering Technology (IJRASET), 2024. doi: 10.22214/ijraset.2024.59955.

[4] P. Mehta et al., "Image-Based Monument Identification by Deep Learning Techniques," International Journal for Research in Applied Science and Engineering Technology (IJRASET), 2025. doi: 10.22214/ijraset.2025.67368.

[5] Author et al., "MonuMAI: Dataset, Deep Learning Pipeline and App of Monument Classification Framework," Neurocomputing, 2021. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0925231220314491

[6] I-m-Parthian, "Monument Recognition Project (CNN-based)," GitHub Repository. [Online]. Available: https://github.com/I-m-Parthian/Monument-Recognition

[7] F. Quintero et al., "Traveling Through History Using an AR Serious Game," Applied Sciences, vol. 14, no. 15, p. 6556, 2024. [Online]. Available: https://www.mdpi.com/2076-3417/14/15/6556

[8] Bannari Amman Institute of Technology, "Exploring Cultural Heritage Through Augmented Reality," 2024. [Online]. Available: https://www.bitsathy.ac.in/exploring-cultural-heritage-through-augmented-reality/

[9] Bannari Amman Institute of Technology, "AR Mobile Apps and Their Use in Recognizing and Visualization of Museum Exhibits," Computers, vol. 14, no. 11, 2024. [Online]. Available: https://www.mdpi.com/2073-431X/14/11/492

[10] Author et al., "Semantics-Preserving Retrieval of Indian Heritage Monuments Hierarchy," arXiv preprint, 2020. [Online]. Available: https://arxiv.org/abs/2008.12832

[11] Author et al., "Automatic Inspection of Cultural Monuments Using Deep Learning," arXiv preprint, 2022. [Online]. Available: https://arxiv.org/abs/2207.02163

[12] Maden and A. Melzer, "A Deep Learning-Based Cultural Heritage Monument Recognition," ISPRS Archives, vol. XLII-2/W9, pp. 551–556, 2019. [Online]. Available: https://isprs-archives.copernicus.org/articles/XLII-2/W9/551/2019/

[13] Khan et al., "A Deep Learning Heritage Resource to Identify Monuments," International Journal for Research in Applied Science and Engineering Technology, 2023. [Online]. Available: https://www.ijraset.com/research-paper/heritage-identification-of-monument-using-deep-learning

[14] P. Mehta et al., "DeepCNN Image-Based Monument Identification," International Journal of Computer Vision, 2020. [Online]. Available: https://link.springer.com/article/10.1007/s11263-020-01363-7

[15] M. Ioannides et al., "A Review of the Existing Solutions of Augmented Reality in Cultural Heritage," Applied Sciences, vol. 12, no. 19, 2022. [Online]. Available: https://www.mdpi.com/2076-3417/12/19/9859

[16] M. K. Bekele et al., "Surveying Augmented Reality and Virtual Reality in Cultural Heritage," Sustainability, vol. 12, no. 16, 2020. [Online]. Available: https://www.mdpi.com/2071-1050/12/16/6491

[17] G. Papagiannakis et al., "Augmented Reality in Archaeology," Journal on Computing and Cultural Heritage, 2017. [Online]. Available: https://dl.acm.org/doi/10.1145/2982142

[18] Samal and A. Seshadri, "Mobile Augmented Reality for Visualization of Historical Landmarks," IEEE Access, 2020. [Online]. Available: https://ieeexplore.ieee.org/document/9174029

[19] F. Quintero et al., "Hybrid Augmented Reality System to Visualize Monuments," Procedia Computer Science, 2020. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1877050920306509

[20] M. Forte, "Augmented Reality and Cultural Heritage," in Proceedings of the Conference on Digital Heritage, Springer, 2020. [Online]. Available: https://link.springer.com/chapter/10.1007/978-3-030-37726-0_12