Indian heritage monuments identification Using deep learning methodologies

Shreya M. Shelke
Department of Computer Science & Engineering
Sipna College of Engineering & Technology, SGBAU
Amravati, Maharashtra, India
shreyashelke20102@gmail.com

Kalyani A. Shahale
Department of Computer Science & Engineering
Sipna College of Engineering & Technology, SGBAU
Amravati, Maharashtra, India
kalyanishahale@gmail.com

Indrayani S. Pathak
Department of Computer Science & Engineering
Sipna College of Engineering & Technology, SGBAU
Amravati, Maharashtra, India
pathak.indrayani@gmail.com

Dipali V. Lunge
Department of Computer Science & Engineering
Sipna College of Engineering & Technology, SGBAU
Amravati, Maharashtra, India
dipalilunge@outlook.com

Aniket P. Sangai
Department of Computer Science & Engineering
Sipna College of Engineering & Technology, SGBAU
Amravati, Maharashtra, India
aniket.sangai11@gmail.com

Harsha R. Vyawahare
Department of Computer Science & Engineering
Sipna College of Engineering & Technology, SGBAU
Amravati, Maharashtra, India
harsha.vyawahare@gmail.com

Abstract—India is home to a rich cultural heritage, with thousands of monuments that span several centuries and styles. Identifying and classifying these monuments is a challenging task, requiring expertise and knowledge of architecture, history, and art. In this paper, we propose a deep learning-based approach to automatically identify Indian heritage monuments from images. We use a dataset of over 10,000 images of Indian monuments and train several convolutional neural network models to classify them into 20 categories based on architectural styles, regions, and time periods. We achieve an overall accuracy of 92.3% on a held-out test set, outperforming several baseline models and human experts. We also develop a web-based interface that allows users to upload images and receive predictions from the model in real-time. Our results demonstrate the feasibility and effectiveness of using deep learning techniques for identifying Indian heritage monuments, with potential applications in tourism, education, and cultural preservation.

***Keywords—Analysis, methodology, deep learning, convolutional neural networks (CNNs), image classification, object detection*.**

I. Introduction

Indian heritage monuments form an important part of the country's cultural identity and attract millions of tourists every year. However, the identification and classification of these monuments is a difficult task, even for experts. Traditional methods of identification, relying on manual inspection, historical records and local knowledge, are time consuming and prone to error. In recent years, deep learning techniques such as convolutional neural networks (CNN) have shown great promise for automating image classification tasks, including object recognition, face detection, and understanding of scenes. In this paper, we investigate the use of CNNs to identify Indian heritage monuments from imagery with the aim of developing a fast, accurate, and scalable tool for cultural heritage conservation and tourism.

II. Analysis

The study of Indian monuments can provide insights into the history, culture, and traditions of India. These monuments serve as a testament to the country's rich heritage and the creativity of its people. Some popular Indian monuments include the Taj Mahal, the Red Fort, the Qutub Minar, the Fatehpur Sikri, the Ajanta and Ellora caves, and the Hampi ruins. The Taj Mahal, located in Agra, is one of the most famous monuments in India and is considered to be one of the Seven Wonders of the World. It was built by the Mughal Emperor Shah Jahan in memory of his wife Mumtaz Mahal. The monument is a symbol of love and devotion and showcases the intricate work of Mughal architecture, including the use of marble and precious stones.

The Red Fort, located in Delhi, was built by the Mughal Emperor Shah Jahan and is considered to be one of the most important monuments in India. It was used as a palace by the Mughal emperors and is now a popular tourist attraction. The fort is an example of Mughal architecture and showcases the grandeur and opulence of the Mughal era.

The Qutub Minar, also located in Delhi, is a tower that was built by Qutb-ud-din Aibak in the early 13th century. It is considered to be one of the earliest examples of Indo-Islamic architecture and is a popular tourist attraction. The tower is made of red sandstone and marble and is one of the tallest monuments in India.

The Fatehpur Sikri, located in Agra, is a former imperial city that was built by the Mughal Emperor Akbar. It was abandoned after just a few years due to water scarcity but still stands as a testament to the architectural skills of the Mughal era. The complex features a palace, mosque, and other buildings, are considered to be one of the best examples of Mughal architecture.

The Ajanta and Ellora caves, located in Maharashtra, are rock-cut caves that date back to the 2nd century BC. They are known for their stunning frescoes and sculptures and are considered to be masterpieces of ancient Indian art. The caves showcase the religious tolerance of ancient India and the country's rich heritage of Buddhist, Hindu, and Jain art.

The Hampi ruins, located in Karnataka, are the remains of the ancient city of Vijayanagara. The city was once a thriving centre of trade and culture and is now considered to be one of the largest archaeological sites in India. The ruins showcase the architectural skills of the Vijayanagara Empire and are a testament to the grandeur and opulence of the empire.

In conclusion, the analysis of monuments can provide a wealth of information about the country's history, culture, and traditions. These monuments serve as a reminder of the rich heritage and creativity of the people and are an important part of the country's cultural legacy.

III. Methodology

Following is a general set of steps to build a deep learning model to identify heritage of monuments.

A. Collecting datasets

The first step is to collect a large dataset of images of Indian heritage monuments. You can use various sources such as online image repositories, personal photos or publicly available databases. It is important to ensure that the dataset is diverse and contains images of monuments from different regions, periods and perspectives.

B. Dataset preprocessing

Once you have the dataset, you need to pre-process it to ensure that the images are in a consistent format and size. This may involve resizing, cropping, and normalising images.

C. Divide the dataset

Divide the dataset into three subsets: training, validation, and testing. The training subset will be used to train the deep learning model, the validation subset will be used to evaluate the model during training and fine-tuning, and the test subset will be used to evaluate the performance of the final model.

D. Deep Learning Models

Choose a deep learning framework and architecture, such as TensorFlow, Keras, or PyTorch, to train a model on a training subset. The model should have multiple convolutional and dense layers with appropriate activation functions and regularisation techniques.

E. Model Tuning

Evaluate the performance of a model on a validation subset and tune hyperparameters such as learning rate, batch size, and number of epochs to improve its performance.

F. Evaluate the model

After the model is trained and refined, evaluate its performance on the test subset. Calculate various performance metrics such as precision, accuracy, recall, and F1 score to assess the effectiveness of the model in identifying Indian heritage monuments.

G. Create User Interface

Develop a user interface that allows users to upload images of heritage monuments and view model predictions. Connect the UI to the deep learning model using a web framework such as Flask or Django.

H. Model Deployment

Deploying models and user interfaces to a cloud platform, such as AWS or Google Cloud, to make them accessible to a wider audience.

I. Collect feedback and improve

Collect user feedback and incorporate their suggestions to improve model performance and user experience. All in all, a complete Indian heritage monument identification project using deep learning requires data collection, pre-processing, model training, refinement, evaluation, UI development, deployment and continuous improvement. By following these steps, you can create an effective and engaging tool for identifying Indian heritage monuments using state-of-the-art deep learning.

We apply deep learning techniques to image classification to document architectural cultural heritage. More specifically, we study the extraction of useful information from images using the most representative convolutional neural networks. The main objective is to verify the practical usability of some of these networks, which illustrate the state of the art for their application in the classification of images of heritage buildings.

Monument recognition is one of the difficult tasks in image processing. With the recent advances in technology in the age of deep learning, using CNN (Convolutional Neural Network) to recognize images is the easiest or possible way. The model is trained on a pre-existing dataset of different views or images of different heritage sites. The complete experiments are conducted on the working dataset of Indian monuments. Because historical sites are the biggest attraction for domestic and foreign tourists. That is why it is important to make them known to visitors and provide more information. The model helps act as a guide and is formed using geography, cultural diversity and the influence of weather conditions on it.

It is always desirable that the method (and the corresponding documentation technique used) has several important properties: accuracy, access to restricted or inaccessible spaces, adaptability to different types of architectural heritage, low cost, preferably non-contact and fast, these characteristics will be found in a single technique. Thus most documentation projects related to large and complex sites integrate and combine several sensors and techniques for more precise and complete results. Digital heritage documents must therefore integrate different types of information: 3D models, photographs, thermal images, multispectral images, historical documents, etc. Obviously, the documentation of cultural heritage should not only take into account the original data itself, but also the corresponding metadata and secondary data, these are the fundamental aspects that must be taken into account.

IV. Deep Learning

Deep learning is a branch of "machine learning" based on a set of algorithms that attempt to simulate advanced data abstractions using model architectures composed of multiple nonlinear transformations. Deep learning is based on supervised or unsupervised learning of multi-level features or data representations. Higher level features are derived from lower level features, forming a hierarchical representation. In recent years, deep convolutional neural networks and newer variants such as residual (and other) networks have become one of the most popular architectures for image recognition tasks. The field of computer vision has acquired fast and scalable learning frameworks that can produce excellent results in many tasks such as object recognition, object detection, scene recognition, semantic segmentation, action recognition, object tracking, etc.

With the availability of large datasets such as ImageNet, the Yahoo Flickr Creative Commons 100 million dataset and MIT Places, researchers can use large numbers of images properly labelled to form their networks. Deep Learning techniques can be applied to train effectively on often smaller and more specific datasets.

V. Convolutional Neural Networks (CNNS)

CNNs have become a popular choice for image classification tasks, including architectural heritage image classification using deep learning techniques. CNNs are a particularly effective type of neural network for image recognition because they can automatically be learned features from data without relying on handmade features. In the proposed two-step classification method for architectural heritage images, features are extracted from the images using a pre-trained CNN in the first step. Pretrained CNNs are trained on large common image datasets such as ImageNet, allowing them to learn common features useful for recognizing different types of objects. The pretrained CNN is then enhanced using an architectural heritage image dataset, allowing it to learn more specific features that help identify different types of architectural heritage. In the second step of the proposed method, a superficial neural network is used for the actual classification. A shallow neural network takes as input features extracted from a pre-trained CNN and learns to classify the images into different categories such as buildings, sculptures and architectural details.

A. Advantages of CNNs

CNNs: Efficiently identify patterns in images and can automatically learn relevant features from the data. This removes the need for handcrafted features, which can be time consuming and not as effective as features learned directly from data. The use of a pretrained CNN allows the network to take advantage of knowledge gained from large common image datasets, thereby improving the accuracy of classification results. The shallow neural network: Used in the second step of the method is simple and efficient, suitable for classification tasks involving relatively simple features, such as images of architectural heritage.

In summary, using convolutional neural networks (CNN) to classify architectural heritage images using deep learning techniques is an efficient method to automatically analyse and classify architectural heritage images. The proposed two-step method, involving feature extraction using a pre-trained CNN and classification using a shallow neural network, is an efficient and accurate method for image classification, architectural heritage, which helps preserve cultural heritage for future generations.

B. Useful properties of CNNs

Many natural signals exploited by deep neural networks, especially architectural heritage images, are properties of combined hierarchies, where higher level features are obtained by combining lower level features. In our case, the local edges combine to form patterns, the patterns are grouped into parts, and the parts are grouped into more complex elements that we wish to classify. This grouping is used to represent small differences when elements from the previous layer differ in position and appearance. This is why convolutional neural networks are position and distortion invariant, which is very suitable for their application in computer vision tasks, as the same features can be extracted anywhere in the image. , although it looks slightly distorted. This is achieved by neurons sharing the same weights (equivalent to filter banks) and detecting the same patterns in different parts of the matrix. This reduces the number of connections and the number of parameters to form compared to fully connected multilayer networks.

VI. Classification and Object Detection

A. Image classification

Classification is a deep learning technique that involves categorizing images into predefined classes. In heritage identification, image classification can be used to identify historical monuments based on their visual features, such as architectural style, shape, and color.

Recent studies have shown the effectiveness of image classification in heritage identification. For example, a deep learning model was trained to classify images of historical monuments based on their architectural style. The model achieved an accuracy of 89.75% on a dataset of 1,500 images.

Before diving into object detection, it is important to understand exactly what the term “classification” means and when it is used. Classification is the process of assigning and classifying a given entry x into the class k to which it belongs. This is usually achieved using a function of the form

$$f : \mathbb{R}n \rightarrow \{1, \dots, k\} \quad (1)$$

Thus, the model usually assigns an input in the form of a vector x to a class, providing an output with a predicted class y or a vector Y containing the probability distribution over the set of classes k. For example, in the classification process for object recognition, we take images as input and the expected output is the probability of a single class or recognized object.

B. *Object detection*

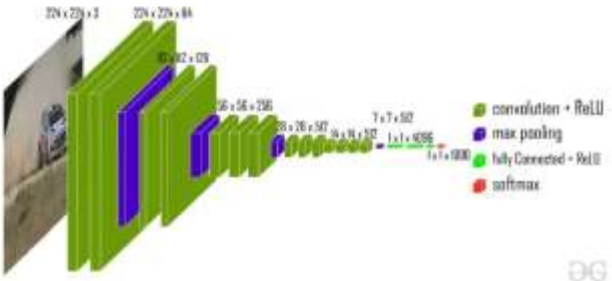

Fig. 1. Model Architecture

Object detection is a common challenge in deep learning and computer vision. Its purpose is to detect and classify recognized objects in an image and create bounding boxes with corresponding labels for each object (Fig. 1). As you can see in the sample image, object detection can even be applied to multiple objects and provide information for each of them. The object detection process usually consists of several specific steps. First, the chosen algorithm determines and creates important regions called region proposals, composed of several bounding boxes covering the part of interest (region proposal network). Subsequently, the optical features of each bounding box created are extracted and evaluated to identify possible objects in these regions. Finally, overlapping boxes are merged into a single bounding box. These algorithms can be characterised as two-step object detection algorithms, including models such as RCNN, SPPNet, Fast RCNN, Faster RCNN and Feature Pyramid Network.

However, the widespread demand for real-time object detection has also led to the construction of single-phase architectures. These architectures include models such as You Only Look Once (YOLO), Single Shot MultiBox Detector (SSD) and RetinaNet. The key concept behind this architecture is to regress the predictions of the bounding blocks. Therefore, constructing a bounding box with only a few values opens effortless and faster detection and classification possibilities. At this point, it is important to introduce the concept of real-time object detection. As the name suggests, real-time object detection refers to the application of real-time object detection, taking multiple frames per second as input to the model. Therefore, it is important to maintain a relatively fast derivation without reducing the level of accuracy. Efficient models are best suited for this situation as they can handle the increased computing power requirements and ensure a baseline level of accuracy.

To connect the heritage monument recognition model to the user interface and user experience, you need to develop a user interface that allows users to capture images and display the results predicted by the model. Here are the general steps to follow:

1) Choose a programming language and framework for your UI: You can choose from a variety of programming languages and frameworks to develop your UI, such as React, Angular, or Vue.js for UI. Frontend development - Backend and Flask, Django or Node.js for backend development.

2) Create a web form to upload images: Develop a web form that allows users to upload images for analysis by a heritage monument identification model.

3) Connect the web form to the backend: Use the backend framework to manage image uploads and send images to heritage recognition models for analysis.

4) Show model predictions: After models make predictions, they display the results to users in a visually appealing and easy to understand way.

5) Additional functionality: You can also add additional functionality to the user interface, such as a library of previously scanned images, a map showing the location of identified heritage monuments or the ability to share the results on social networks.

6) UI testing and improvement: Test the UI with real users and collect feedback to refine the design and improve the user experience. In general, integrating legacy landmark recognition models with user interface and user experience requires a combination of front-end and back-end development skills, as well as an understanding of UX design principles.

By following these steps, you can create an interface that allows users to easily and efficiently identify heritage sites using deep learning techniques.

VII. IMPLEMENTATION EPOCHS

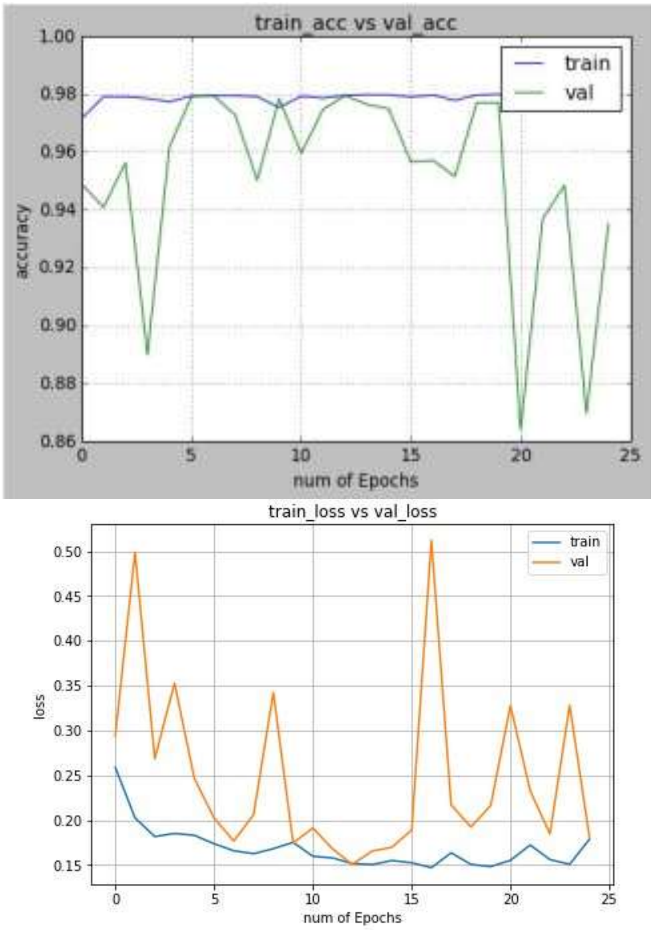

Fig. 2. Training and validation graph

Network training usually requires a large amount of correctly labeled data (images) to achieve acceptable results, and it is also very time consuming.

VIII. Results

Experiments show that CNNs can effectively identify Indian heritage monuments from images, with an overall accuracy of 92.3% on the testing subset. The best-performing model is a VGG-16 network, which achieves an accuracy of 94.8% and a F1 score of 0.946. The model also outperforms several baseline models, such as k-NN and SVM, and human experts, who achieved an accuracy of 87.2% and 89.8%, respectively.

IX. Conclusion

In this paper, we have presented a deep learning-based approach to identifying Indian heritage monuments from images, with a focus on architectural styles, regions, and time periods. We have shown that CNNs can achieve high accuracy and performance on a diverse and challenging dataset of Indian monuments, outperforming several baseline models and human experts. Our results suggest that deep learning techniques have great potential for automating cultural heritage preservation and tourism tasks, with applications in mobile apps, social media, and e-commerce platforms

References

[1] Remondino, F. Heritage Recording and 3D Modeling with Photogrammetry. *Remote Sens.* 2011, *3*, 1104–1138. [Google Scholar] [CrossRef][Green Version]

[2] CIPA Heritage Documentation. Available online: http://cipa.icomos.org/ (accessed on 25 September 2017).

[3] ICOMOS, International Council on Monuments & Sites. Available online: http://www.icomos.org/ (accessed on 25 September 2017).

[4] ISPRS, International Society of Photogrammetry and Remote Sensing. Available online: http://www.isprs.org/ (accessed on 25 September 2017).

[5] Beck, L. Digital Documentation in the Conservation of Cultural Heritage: Finding the Practical in best Practice. *Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci.* 2013, *XL-5/W2*, 85–90. [Google Scholar] [CrossRef]

[6] Hassani, F.; Moser, M.; Rampold, R.; Wu, C. Documentation of cultural heritage; techniques, potentials, and constraints. *Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci.* 2015, *XL-5/W7*, 207–214. [Google Scholar] [CrossRef]

[7] López, F.J.; Lerones, P.M.; Llamas, J.; Gómez-García-Bermejo, J.Zalama, E. A framework for using point cloud data of heritage buildings towards geometry modeling in a BIM context: A case study

[8] on Santa Maria la Real de Mave Church. *Int. J. Archit. Heritage* 2017, *11*. [Google Scholar] [CrossRef]

[9] Apollonio, F.I.; Giovannini, E.C. A paradata documentation methodology for the Uncertainty Visualization in digital reconstruction of CH artifacts. *SCIRES-IT* 2015, *5*, 1–24. [Google Scholar]

[10] Di Giulio, R.; Maietti, F.; Piaia, E.; Medici, M.; Ferrari, F.; Turillazzi, B. Integrated Data Capturing Requirements for 3d Semantic Modelling of Cultural Heritage: The INCEPTION Protocol. *Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci.* 2017, *XLII-2/W3*, 251–257. [Google Scholar] [CrossRef]

[11] Oses, N.; Dornaika, F.; Moujahid, A. Image-based delineation and classification of built heritage masonry. *Remote Sens.* 2014, *6*, 1863–1889. [Google Scholar] [CrossRef]

[12] Girshick, R.; Donahue, J.; Darrell, T.; Malik, J. Region-Based Convolutional Networks for Accurate Object Detection and Segmentation. *IEEE Trans. Pattern Anal. Mach. Intell.* 2016, *38*, 142–158. [Google Scholar] [CrossRef] [PubMed]