

Mumbai Lens: A Chatbot for Monument Recognition and Historical Exploration

Shubhada Labde
Computer Engineering,
KJ Somaiya Institute of Technology,
Sion
Mumbai, India
shubhada.l@somaiya.edu

Nisha Vanjari
Computer Engineering,
KJ Somaiya Institute of Technology,
Sion
Mumbai, India
nvanjari@somaiya.edu

Abstract—Artificial Intelligence has brought a great revolution in all fields including digital tourism and cultural education. With increasing interest in tourism and education, there is a high demand for tools that can help in connecting individuals with cultural knowledge. The intersection of Tourism and Artificial Intelligence has offered us new opportunities to digitally preserve and promote tourism. This research aims to explore how well image driven AI systems can improve cultural awareness by using passive visual inputs to transform into active learning experiences. This study aims to investigate the integration of rule-based chatbots with image recognition techniques to give interactive heritage exploration. The proposed system is a rule-based AI chatbot designed to improve historical exploration and cultural awareness for the city Mumbai. The system is built using a custom-curated dataset of over 700 images and 32 historic places across Mumbai. The system helps in significantly bridging the gap between cultural heritage and technology by proposing an interactive way to explore the rich heritage of the city.

Keywords—Object Detection, Transfer Learning, Chatbot, Smart Tourism

I. INTRODUCTION

In today's world, people are increasingly relying on AI-powered integrated systems to get instant responses for their queries in minimal time, especially for domains such as tourism and travel. Tourism and Travel industry has increased exponentially in recent years. Travelling is a hobby which is enjoyed by everyone and often seen as a means of taking a break from their busy and stressful life. A study shows that 87% of users would use a tourism chatbot if it saves them their money and time. As time flies, cities change and urban areas become more complex, tourists and even local residents sometimes find it difficult to find and learn about the historical monuments around them. Traditional methods such as using guidebooks require prior knowledge and manual effort, which can diminish the engagement with the heritage landscape. This has created a great need for a user-friendly utility which can offer instant information and context, by simply analyzing the image of the monument.

AI-powered chatbot systems are an effective way to tackle the situation. Such systems employ deep learning and machine learning algorithms in conjunction with Natural Language Processing(NLP) to offer human-like interaction in real-time. When these systems are integrated with computer vision, they provide a more dynamic and

immersive experience. The paper is aimed at developing an AI-based chatbot which uses deep learning techniques to detect and give information about any monument. This approach enhances cultural awareness and also contributes to the development of smart tourism technologies and digital heritage solutions.

II. LITERATURE SURVEY

AI is playing an important role in improving the user experience across tourism and heritage sectors. Recent studies show that combining image recognition and conversational systems can provide great interactive and informative experiences to users. The Deep Learning technologies particularly Convolutional Neural Networks (CNNs) have produced substantial results in terms of classification and identification of monuments on the basis of image datasets. In one study, a Long Term Short Memory(LSTM) neural network combined with an open source framework called Rasa was used to develop an Arabic chatbot. This allowed them to work efficiently even with a small dataset [1].

Conversational AI chatbots are shown to deliver more personalized and effective user experience with real-time information in Natural Language. Another study trained 3 different algorithms: Bag of Words, BERT, and RoBERTa, out of which BERT gave the highest accuracy of 90% and 90.1% F1-score. The developed model was used in a chatbot mobile application on android based on Flutter and Dart [2]. Through a comparative study, Deep Neural Networks with LSTM chatbot model, can be 20 percent in managing the user queries and provide any suitable recommendation to the final users once it is integrated with different Open APIs in recommendation and net ratings online[3].

AI-powered chatbots can be used in assisting users with queries regarding historical context, location and even itinerary suggestions. In a study, researchers developed an AI-based chatbot system which used a pretrained model over 3 datasets: DST dataset, Neo4J graph DB and MySQL DB servers[4]. Technologies used to develop these kinds of exceptional AI chatbots include classic query management, logs, and fuzzy search tools [5]. A work presented an Artificial Intelligence-powered chatbot information distribution in a rather diverse Indian city in terms of its culture and the lingo it does not speak [6].

Another model developed using YOLOv8(You Only Look Once) resulted in an impressive Average Mean

Precision of 91.5% primarily focused on Tamil Nadu monuments and a variety of diverse historical settlements located around the Indian subcontinent [7]. In a study of promoting tourism in places like Thailand, Ranong and Chumphon, researchers developed an AI-powered chatbot using XLM-RoBERTa (XLM-R) large language model (LLM) which achieved an F1-score of 99.56% for intent entity classification and 99.96% for named entity classification [9]. A study proposed the use of a hybrid CNN-RF model which successfully attained an accuracy of 80.89% and a highest F1-score of 85.81% [8].

A detailed survey showed the importance of monument recognition and classification and how it helps in improving smart tourism[10]. The survey reviewed both traditional techniques such as BoW and SIFT and modern techniques like CNN, DCNN and ensemble techniques[10]. The survey highlights the challenges in Indian Context such as inconsistent lighting and architectural overlaps[10].

A research article proposed by V. Kukreja et al.[11] demonstrated a model based on Mask R-CNN(MRCNN) for hemispherical dome recognition in Indian Heritage Architecture. The model was trained on a dataset of 600 dome images, successfully achieving a mean average precision of 95.6% at 50% Intersection of Union[11]. The model shows precise identification of dome structures with pixel level accuracy[11]

In [12], Sasithradevi et al. gave a novel deep learning architecture for classifying heritage monuments of Kolkata. The model enhances feature extraction by incorporating Dense Channel Attention and Parallel-Spatial Channel Attention modules, especially from intricate architectural designs[12]. Using a custom dataset of 50 images for each of 13 kolkata monuments, the model achieved an accuracy of 89% and precision of 87.77%[12].

III. PROPOSED METHODOLOGY

The presented system is a prototype, created as the part of the academic research and feasibility study in the field of AI and computer vision as well as heritage informatics. The prototype assists in the illustration of the application of the AI powered image recognition in smart tourism and digital heritage. The architecture of the designed system combines deep learning and the powerful backend database and dynamic chatbot interface and a robust backend database. It is structured in a way to support and help users with a smooth and interactive experience. The system's architecture is modular, scalable and designed for seamless integration of its components and modules. The key components of the proposed solution is given below:

A.Dataset

The image classification model of the proposed system was developed on a self-curated dataset of 32 historic places in Mumbai. The dataset contains more than 700 high resolution labelled images to train and test our model. This includes widely known places like Gateway of India and Chatrapati Shivaji Maharaj Terminus as well less popular places like Bandra Fort, Horniman Circle and David Sassoon Library.



Fig. 1. Dataset Samples

In the attempt to enhance the model performance and generalization capability, it was ensured that the images collected were of high quality, different lighting conditions, weather scenarios and angles. To ensure that the model generalizes well, an extensive preprocessing pipeline was applied. This included image resizing, augmentation techniques such as rotation, flipping, zooming, contrast adjustment, and noise removal.

A.System Design

The proposed system uses a MobileNetV2 Model pre-trained on Imagenet database, using Transfer Learning to detect the monuments in the supplied image. This method efficiently decreases training time and gives better performance even upon a small dataset. MobileNetV2 is a lightweight model and helps decrease computational costs.

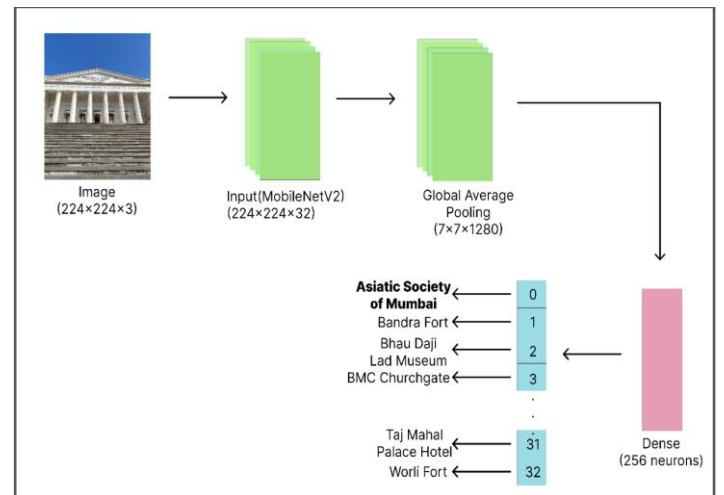


Fig. 2. Model working flowchart

The proposed architecture shown in Figure 2 shows the layers used in the model to identify monuments from images. Layers used in the model:

The proposed architecture shown in Figure 2 shows the layers used in the model to identify monuments from images. Layers used in the model:

1. Input MobilenetV2 layer: a light weighted CNN model which has been trained on Imagenet, and processes the image of dimension of 224x224 size to generate low level, and mid level spatial features like edges and shapes of objects using multiple Convolutional Layers.
2. Global Average Pooling Layer: assists in decreasing the spatial size of feature maps and maintains features and provides minimal computational cost. It averages each of the 7x7 grids resulting in a 1D vector output corresponding to the grid.

3. Dense Layer: Maps the 1280 features to a higher-dimensional space and passes to the next layer.
4. Dense Output Layer: The last fully linked many-to-one layer is converted to probabilities of class using the softmax activation operation.

The proposed system can use the current GPS coordinates of the user to give more accurate predictions, and hence decreases its false positives. To further improve the system capabilities, the proposed solution uses an interactive rule-based chatbot interface with information about the historic places stored in a NoSQL database. The complete system flow diagram for the process of detecting the monument name from the supplied image is illustrated in Figure 3.

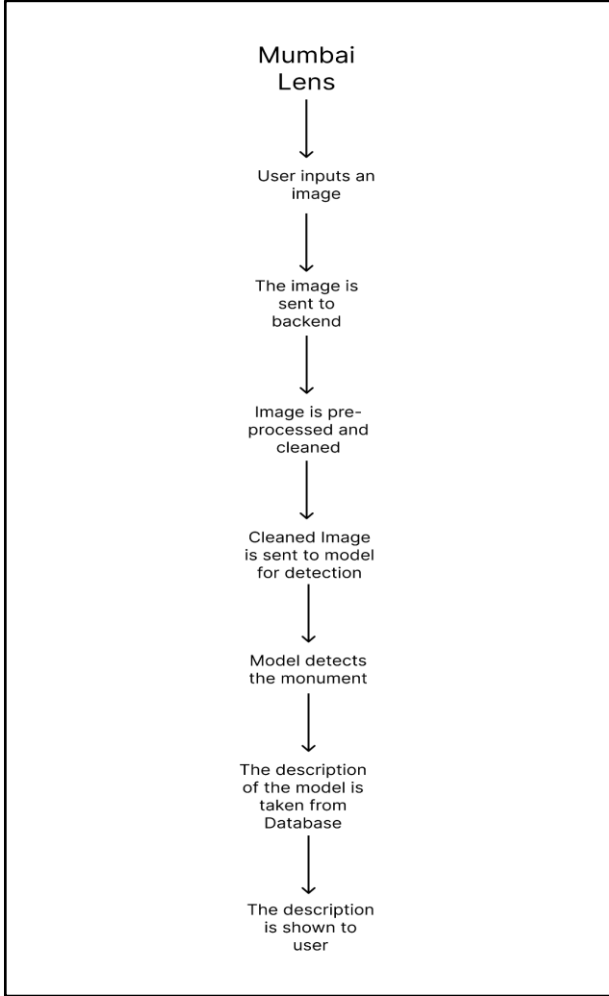


Fig. 3. System Flow Diagram of Mumbai Lens

IV. RESULTS AND ANALYSIS

According to the performance assessment of the model, it was revealed that the model has been able to get an accuracy of 94.04% with the dataset during the testing. This showed the model's strong generalization and classification ability. The results signified how the model had maintained a balanced ability to minimize false positives and false negatives across all monument categories. It was seen that the places with distinct architectural features like Gateway of India were well-identified and had higher classification accuracies while the monuments with less architectural features like the David Sassoon Library had slightly lesser classification accuracies. The model contains 32 output

classes which were represented by the names or the labels of those historic places. We have used MobileNetV2 pre-trained model with Transfer Learning to train a model, integrated with a rule based chatbot to give users an immersive and interactive experience.

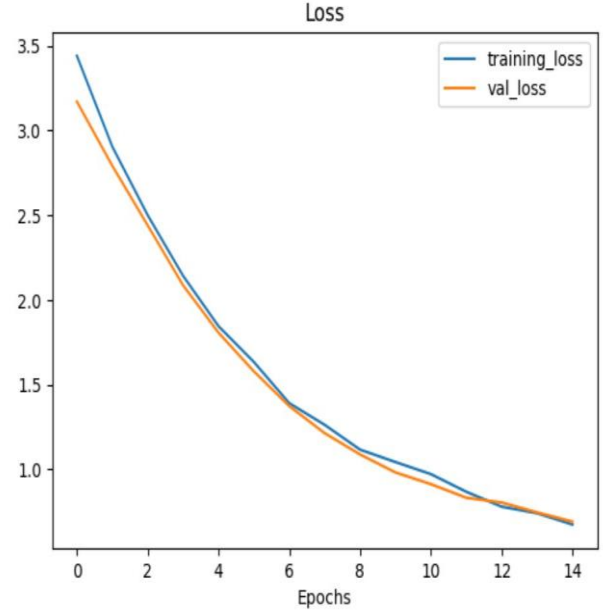


Fig. 4. Training and validation loss curves

The Training and Validation Curves of the utilized model is shown in Figure 4. The two curves show a constant decreasing trend where values of losses are reducing, in equal ratios, to be larger than 3.0 and smaller than 1.0. The model employs a generalized and stable learning procedure that can be confirmed by the close relation of training and validation loss curves. The lack of dispersion of the lines indicates the existence of very little probability of Model Overfitting.

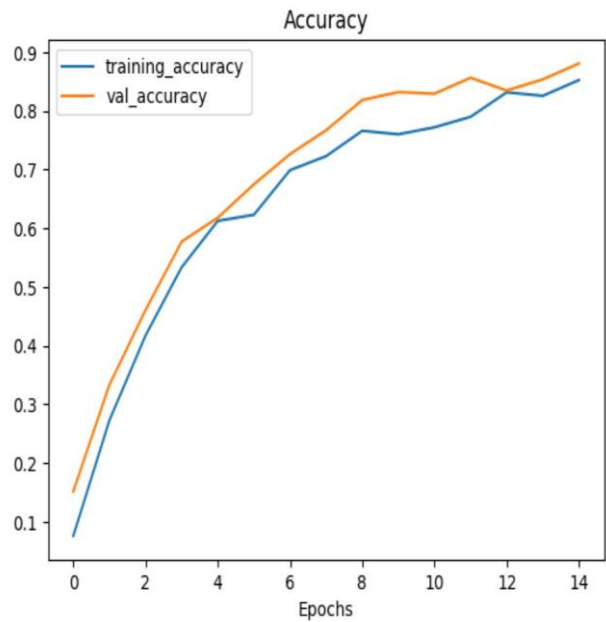


Fig. 5. Training and validation accuracy curves

The curves of Training and Validation Accuracy have been presented in Figure 5. The accuracy is seen to improve sharply over the initial epochs and finally plateaus as the

model converges. The graph shows how the validation accuracy surpasses the training accuracy curve, after approximately the 6th epoch while maintaining the lead till the end. This shows that the model is highly generalized and has well-tuned parameters.

The chatbot enhances system accessibility for general users, including tourists and students. By coupling monument recognition with interactive exploration, the interface transforms the application from a passive information tool into a highly engaging digital heritage chatbot.

V. CONCLUSION

The proposed system gives an ideal solution for removing the gap between urban heritage and real-time user interaction, through AI-powered monument detection and recognition. By integrating deep learning techniques with a conversational chatbot interface, it helps the user identify and learn about the history of any historical place in real time. This approach helps in enhancing user experience and contributes to cultural awareness in a better and interactive way, particularly tourists who are not aware of the city's architectural remarks.

The system being proposed was developed with respect to scalability and adaptability, helping it demonstrate strong potential for better application in smart tourism initiatives. Its architecture allows for future enhancements such as multilingual support, recommendations to similar or nearby historic places and even real-time crowd density updates at monument sites. The proposed system represents a great example of how technology can be used to support and promote cultural learning and foster deeper engagement towards modern urban environments.

REFERENCES

- [1] R. Alotaibi, A. Ali, H. Alharthi, and R. Almeahmdi, "AI chatbot for tourist recommendations: a case study in the city of Jeddah, Saudi Arabia," *International Journal of Interactive Mobile Technologies (IJIM)*, vol. 14, no. 19, p. 18, Nov. 2020, doi: 10.3991/ijim.v14i19.17201.
- [2] H. T. Nguyen, T. T. Tran, P. T. Nham, N. U. B. Nguyen, and A. D. Le, "AI chatbot for tourist recommendations: a case study in Vietnam," *Applied Computer Systems*, vol. 28, no. 2, pp. 232–244, Dec. 2023, doi: 10.2478/acss-2023-0023.
- [3] G. S. Saluja, N. Maheswari, T. S. P. Kumar, and M. Sivagami, "AI based intelligent travel chatbot for content oriented user queries," *International Journal of Applied Science and Engineering*, vol. 18, no. 2, pp. 1–8, Jan. 2021, doi: 10.6703/ijase.202106_18(2).006..
- [4] H. -C. Kang, K. -B. Kang, D. -H. Kim, M. -C. Jwa, T. -S. Ko and J. -W. Jwa, "Smart tourism chatbot system using Multi-domain Tourism Information DST," 2023 Fourteenth International Conference on Ubiquitous and Future Networks (ICUFN), Paris, France, 2023, pp. 608–612, doi: 10.1109/ICUFN57995.2023.10200288. keywords: {Transfer learning;Natural languages;Chatbots;Web servers;Information systems;AI-based Chatbot;Smart Tourism;Dialogue State Tracking;Pre-trained Language Model;Tourism Information Knowledgebase},
- [5] L. Benaddi, C. Ouaddi, A. Souha, A. Jakimi and B. Ouchao, "Chatbots in Tourism Sector: Classification, Evolution, and Functionalities," 2024 IEEE 12th International Symposium on Signal, Image, Video and Communications (ISIVC), Marrakech, Morocco, 2024, pp. 1–6, doi: 10.1109/ISIVC61350.2024.10577889. keywords: {Technological innovation;Databases;Tourism industry;Chatbots;Market research;Artificial intelligence;chatbot;tourism;classification of chatbots;functionalities;smart-tourism},
- [6] S. Anand, A. M. Abhishek Sai and M. Karthikeya, "Chatbot Enabled Smart Tourism Service for Indian Cities: An AI Approach," 2023 11th International Conference on Internet of Everything, Microwave Engineering, Communication and Networks (IEMECON), Jaipur, India, 2023, pp. 1–7, doi: 10.1109/IEMECON56962.2023.10092286. keywords: {Urban areas;Information services;Transportation;Switches;User interfaces;Chatbots;User experience;Chatbots;Smart City;Information Service;Neural Network;Tourism},
- [7] U. Surya, E. Subashree and G. Shobana, "Artificial Intelligence Techniques For Monuments Identification," 2024 International Conference on Innovations and Challenges in Emerging Technologies (ICICET), Nagpur, India, 2024, pp. 1–6, doi: 10.1109/ICICET59348.2024.10616302. keywords: {YOLO;Technological innovation;Accuracy;Urban planning;Focusing;Real-time systems;Object recognition;You Only Look Once (YOLOv8);Artificial Intelligence (AI);mean Average Precision (mAP);voice component;antiquarian research},
- [8] D. Kumar, M. Kumar, S. K. Ray and V. Kukreja, "Monument Preservation: Utilizing a Hybrid CNN-Random Forest Framework," 2024 International Conference on Automation and Computation (AUTOCOM), Dehradun, India, 2024, pp. 7–11, doi: 10.1109/AUTOCOM60220.2024.10486115. keywords: {Pollution;Costs;Computational modeling;Surface contamination;Forestry;Feature extraction;Generative adversarial networks;Monument preservation;Hybrid Model;CNN;Random Forest;Machine Learning},
- [9] N. Laosen, K. Laosen and P. Sriprasert, "Development of AI Chatbot for Tourism Promotion: A Case Study in Ranong and Chumphon, Thailand," 2024 21st International Joint Conference on Computer Science and Software Engineering (JCSSE), Phuket, Thailand, 2024, pp. 224–231, doi: 10.1109/JCSSE61278.2024.10613647. keywords: {Adaptation models;Databases;Training data;Prototypes;Machine learning;Chatbots;Data models;chatbot;intent classification;named entity recognition;tourism promotion;graph database},
- [10] A. J. Paul, S. Ghose, K. Aggarwal, N. Nethaji, S. Pal and A. Dutta Purkayastha, "Machine Learning Advances aiding Recognition and Classification of Indian Monuments and Landmarks," 2021 IEEE 8th Uttar Pradesh Section International Conference on Electrical, Electronics and Computer Engineering (UPCON), Dehradun, India, 2021, pp. 1–8, doi: 10.1109/UPCON52273.2021.9667619. keywords: {Industries;Geography;Computer vision;Economic indicators;Conferences;Computational modeling;Buildings;Artificial Intelligence;Computer Vision;Indian Heritage;Landmark Recognition;Monument Classification;Monument Detection},
- [11] V. Kukreja and D. Kumar, "An Application of MRCNN Model for Hemispherical Domes Recognition in India Heritage Monuments," 2023 International Conference on Sustaining Heritage: Innovative and Digital Approaches (ICSH), Sakhr, Bahrain, 2023, pp. 51–55, doi: 10.1109/ICSH57060.2023.10482824. keywords: {Instance segmentation;Computational modeling;Semantic segmentation;Object detection;Computer architecture;Cultural differences;Convolutional neural networks;Cultural Heritage;Public Spaces;Urban Development;MRCNN;Sustainable;Computer Vision},
- [12] Sasithradevi, A., nathan, S., Chanthini, B. *et al.* MonuNet: a high performance deep learning network for Kolkata heritage

image classification. *Herit Sci* 12, 242 (2024). <https://doi->

[org.library.somaiya.edu/10.1186/s40494-024-01340-z](https://doi.org/library.somaiya.edu/10.1186/s40494-024-01340-z)