JETIR.ORG ISSN: 2349-5162 | ESTD Year : 2014 | Monthly Issue
JOURNAL OF EMERGING TECHNOLOGIES AND INNOVATIVE RESEARCH (JETIR)
An International Scholarly Open Access, Peer-reviewed, Refereed Journal

INDISIGHT: A CNN BASED IMAGE RECOGNITION OF INDIAN MONUMENTS

Kiran Chandrashekar Datawad[1], Hemanth Mahesh[2] , K Preetham[3], J R Chandan[4]

Students of CSE Department, K. S. School of Engineering and Management, Bengaluru, Karnataka, India
Mrs. Belji T[5]
Associate professor, Department of Computer Science and Engineering, K. S. School of Engineering and Management, Bengaluru, Karnataka, India

ABSTRACT

In a world rich with cultural heritage and historical significance, preserving and celebrating the architectural wonders of India is of paramount importance. Our project, "IndiSight" introduces a groundbreaking solution for the seamless identification and appreciation of Indian monuments. This system utilizes convolutional neural networks (CNNs) to analyze images and provide detailed information about the monuments. With this system, we can explore India's rich cultural heritage and gain a deeper understanding of its history.

Keywords- Artificial Intelligence, Machine Learning, Deep Learning, Convolutional Neural Networks, Image Classification, Architectural Heritage, CNN(Convolutional neural networks), Monument Recognition.

I. INTRODUCTION

India boasts a rich cultural heritage adorned with a myriad of architectural marvels that span centuries of history. From the iconic Taj Mahal to the majestic temples of Khajuraho, the country's monuments stand as testaments to its diverse traditions, craftsmanship, and historical significance. With the advent of modern technologies, particularly advancements in machine learning and computer vision, there lies an exciting opportunity to delve deeper into the realm of Indian monument recognition.

Machine learning algorithms, particularly deep learning methods, have revolutionized the way we perceive and analyze visual data. By leveraging vast amounts of image data and computational power, these algorithms can discern intricate patterns and features within images, enabling them to recognize and categorize objects with remarkable accuracy. In the context of Indian monuments, this presents a fascinating avenue for automating the process of identification and classification.

The state-of-the-art Deep Convolutional Neural Networks (DCNN) is used for extracting representations. The model is trained on representations of different Indian monuments, obtained from cropped images, which exhibit geographic and cultural diversity. Urban elements within a city form interconnected narratives, often concealing inaccessible information. Sites, whether ancient, modern, or archaeological, encapsulate a tapestry of historical events across time. This wealth of information constitutes a virtual network, interlinking various data points. Storing this information in a database demands queries that pivot on diverse parameters. For instance, encountering an ancint statue might spark interest in sculptural technique, historical context, archaeological significance, or the creator's biography. Simultaneously, one might seek information about analogous specimens across different geographical regions.

Traditional solutions like audio tours, information panels, or QR codes fall short in addressing such diverse queries. Conversely, prevalent music recognition systems engage with environmental features to access data and services. Similarly, by connecting architectural objects, artworks, and spatial characteristics with digital content, mobile devices could serve as gateways enabling complex queries initiated by object recognition.

Cultural heritage management faces the challenge of resource optimization. Existing infrastructure in cultural sites demands tailored maintenance, posing challenges for both large sites unable to provide adequate services and smaller sites constrained by surveillance and maintenance costs. Mobile computing technologies present an avenue to surmount limitations, augmenting the visitor experience at cultural sites while enhancing accessibility to digital heritage collections.

II. RELATED WORKS

Unprecedented computational resources and burgeoning datasets have propelled the evolution of deep learning (DL) techniques. These algorithms, a subset of machine learning (ML), empower machines to comprehend complex concepts by hierarchically interpreting input objects. Convolutional Neural Networks (CNNs), a prominent DL model for image processing, find application in various fields like face recognition, handwriting recognition, and medical image analysis.

II.1 Image based recognition of the monuments in Prizren

The automatic monument recognition approach is still a new area of interest. The literatureand the field is only been found in a very small number of papers. There are some of the working groups an authors who have proposed and applied different methods for the landmark recognition. Some of the work and used techniques by this authors are also represented in this thesis.This chapter is divided in different parts to express the theoretical work of this thesis so that the readers could have a clarified understanding about it. Firstly, it will talk about the captured images and the digital image processing. And the other part will deal with the machine learning algorithms used for image classification, recognition and object detection.

The idea of this intelligent tour guide is that the tourists who visit Prizren, by only photographing the monument will be able learn about the history, the time of construction and the purpose of that particular monument.

II.2 Heritage identification of Monuments Using Deep learning

This project aims to introduce a web-based application leveraging machine learning techniques to assist travelers in identifying Indian monuments. Preserving this cultural heritage in a digital format involves gathering extensive image datasets, refining the data, and employing PyTorch to train models focused on Indian monuments. Specifically, the model utilizes the MobileNet V1 architecture for monument recognition and MobileNet V2 for satellite identification. Implementation within a Python Flask Web application enables users to capture monument images and access related information, including nearby attractions, hotels, user reviews, price ranges, monument descriptions, official websites, and links to 360- degree views.

It provides an equilibrium between recall and precision. A user-friendly web application is designed to provide almost accurate detection of the monuments.

II.3 Image based Indian Monument Recognition using Convoluted Neural Networks

This paper proposes an approach for classification of various monuments based on the features of monument images. Deepconvoluted neural networks is used for extractingrepresentations. This paper considered hundred monumentsin the dataset for classification with 90% accuracy. They usedGBVS (Graph Based Visual Saliency) method for monument recognition to find out saliency in monument images so that better image recognition algorithm work. The images are previously processed according to the GBVS method, which is in order to either keep the SURF or SIFT features , which are corresponding to the present actual monuments while the background “noise” which has been minimized. The monuments or the landmarks accurately identified using the classification.

The model is trained on representations of different Indian monuments, obtained from cropped images, which exhibit geographic and cultural diversity.

III. PROPOSED SYSTEM

The proposed system gives a method for recognition of various Indian Monuments. Some future enhancements for the proposed system can be to:

1) Increase the dataset and improve the scope of the model.
2) Make the system more user friendly by creating an application of the same.
3) Improve the information about the retrieval of the monuments by providing dynamic information from internet.
4) Help the tourists by providing directions to the monuments and live tracking the tourists using Global Position Service (GPS).
5) Improving the dynamicity of the images by live capturing and instant monument detection. The proposed system can be evolved to meet several other operations which are not included in this project. Expanding the system will result in more efficient and hassle free operations.

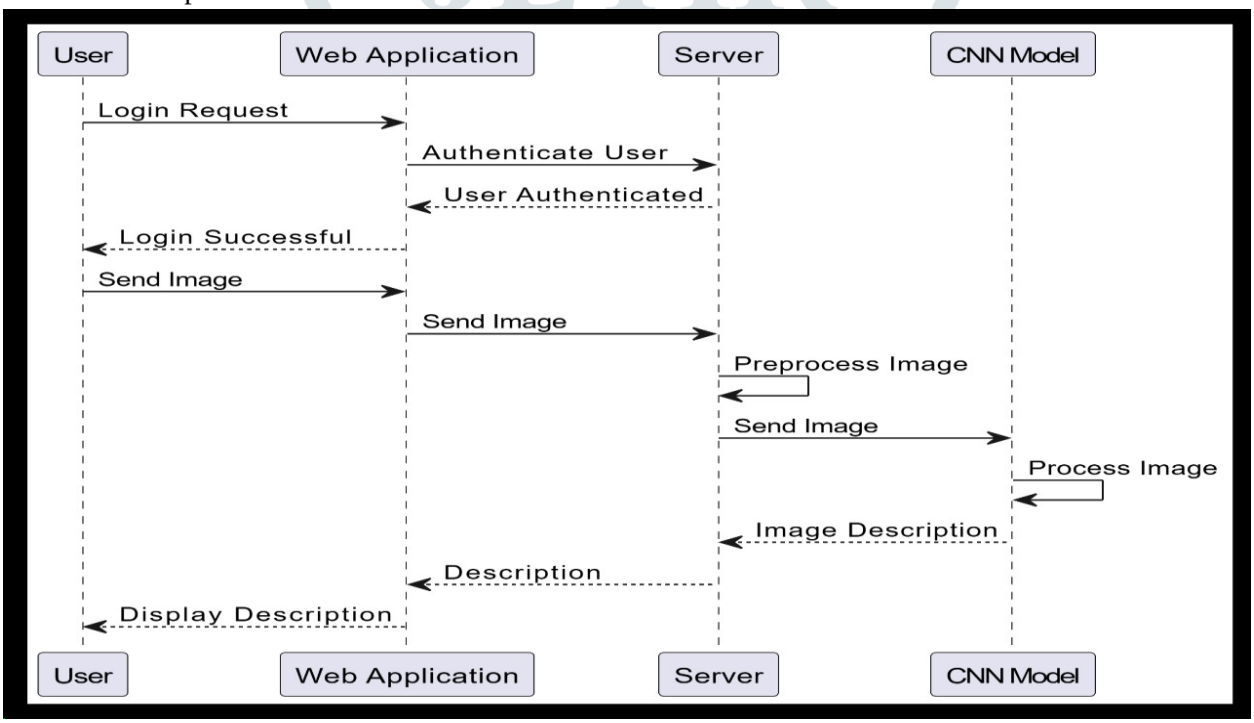

Fig 1: Sequence Diagram

1) User: A user is a casual name given to an individual who interacts with a website, online service, app or platform in any way.
2) Login: the process by which an individual gains access to a computer system or program by identifying and authenticating themselves.
3) Authentication: the process by which an individual gains access to a computer system or program by identifying and authenticating themselves.
4) Upload a Picture: An image description is a written caption that describes the essential information in an image.
5) Augmentation: augmentation means simply outsourcing workers for projects aligned with the organization's current staffing needs and objectives.

6) Database: A database is an organized collection of structured information, or data, typically stored electronically in a computer system.

IV. DATASET

The present application was developed for recognizing Indian monuments focusing on a diverse range of landmarks scattered across the country. Unlike other datasets for object recognition, capturing images of Indian monuments poses some unique challenges. First, these monuments are often fixed in their positions, resulting in pictures of the same monuments being quite similar for a model to distinguish from the surroundings. Second, the dataset includes monuments that are spatially and historically close together, leading to frequent occurrences of similar backgrounds and architectural features.

Starting the project from the scratch and having to document not only famous buildings, we took pictures specifically for the project. Our dataset labels the images of monuments of the India, with pictures spanning a complete overview of the architectural characteristics. The viewpoints have been chosen with particular care: the monuments were framed from different positions and we included the most common places for the visitors, details and panoramic views. We also chose different object lighting and camera exposure conditions to make the DL model deal with more difficult circumstances. We both used pictures taken from mobile phones and professional cameras.

Monuments not only show recurring architectural structures and decorations. In fact, they feature heterogeneous dimensions and conservation status: some are quite incomplete and poorly preserved, some are part of more recent buildings, some are large and composed of many different parts.

V. METHODOLOGY

1. DATA PREPARATION: The preparation of the dataset is one of the fundamental steps in ML. Good results require a careful treatment of data. We started by inspecting the available images for each class, then we applied data augmentation, formatting and cleaning.

Methods:

1. Data selection: we collected some samples of pictures for each monument. Since pictures may present different classes (more than one monument), we manually verified each picture and highlighted (*e.g.* by cropping the picture) the features to make the network "learn" properly the corresponding monument.

2. *Data augmentation:* The collected picture dataset does not contain enough images to train the network. Some platforms can artificially generate more data during the training, but we chose to extend the dataset in advance. In fact, the relatively small number of images available for each class makes it difficult to handle duplicate data if we cannot access the random image generation algorithms.

3. *Data cleaning:* each operation applied during the data augmentation is random. Since several duplicate images are expected, we adopted a classical technique to cleaned up the data, by creating a hash table for the images. For this, we used a JSON1 format database.

2. NETWORK TRAINING : The convolutional network works well for the classification task and is widely used for computer vision problems. The idea is to extract the features of the input image by using a sufficient number of convolutional layers and use a fully connected network at the final layer to perform the classification (*deep convolutional*

neural network). The mainpurpose of training is to tune the kernel parameters, togetherwith the weights and the biases for the fully connected layer. Moreover, we need to be able to use the inference model into amobile device.

Methods:

1. The model: after several trials, we chose to use a class of CNNs known as the MobileNets (Howard et al., 2017). It was developed by Google researchers to be an efficient candidate for mobile DL models. The model fits our needs because it is fast and light and it shows better accuracy when tested on the ImageNet database. The main feature of this architecture is the use of depthwise separable convolutions, composed by depthwise 3 3 and pointwise 1 X 1 convolution layers. This significantly reduces the number of parameters to be trained in the network. The MobileNet-224 architecture can be seen as follows:

In Table 1, the architecture of the mobileNet with the input image shape 224 224 3 is shown. Each convolution layer is followed by the batch normalization (Ioffe and Szegedy, 2015) and the activation function ReLU. The width multipliers of the MobileNet model are $\alpha =$ 0.25, 0.5, 0.75, 1.0, to reduce the size of the model and the width of the depthwise separable convolution.

2. Training: the MobileNet model is easy to train for a right choice of parameters. To this purpose we used some known techniques for debugging neural networks such as:

- Initialization of the weight
- Regularization of the tuning parameter
- Choice of the back-propagation function
- Fine tuning
- Dropout

3. Test : we tested the accuracy of our model for a set of 100 images not included in the training dataset.

3. BUILDING THE MODEL: The final step is to use the inference model on a website or a web app.

methods:

1. *Converting the inference model for website*: here we used TensorFlow lite4, an open source DL framework for website included in TensorFlow (Abadi et al., 2016).

VI. IMPLEMENTATION AND RESULTS

Fig 1 : Home Page

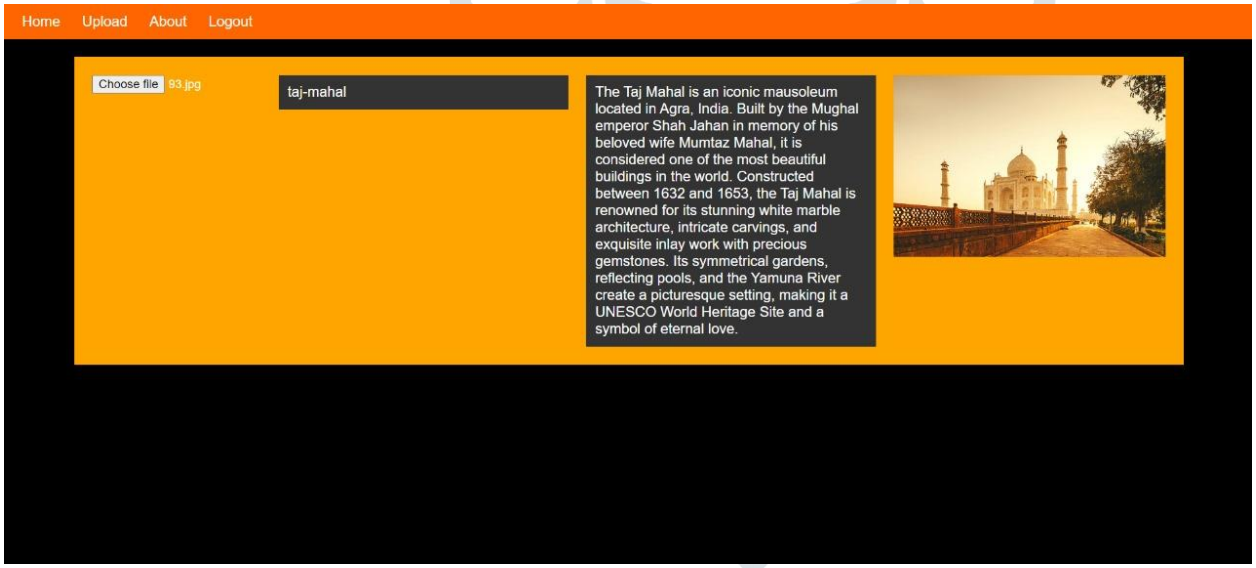

Fig 2 : Upload Page

VII. CONCLUSIONS AND FUTURE WORKS

As we highlighted in the introduction, the AI is changing the way we interact with things, in particular with the city. Search engines for information are still based mainly on texts, partially on images and sounds, and operate on general and unorganized datasets. Libraries, archives, academic studies and researches, constitute a framework with a logical structure that is not generally accessible to non-experts. In the presence of an archaeological or architectural art object, however, the users search for cultured or academic information that is general not easily available.

The Indian architectural and archaeological heritage represents an immense mine of information, studies, research or simply stories that deserve to be accessed in a simple way starting from real objects. The creation of information sets indexed by real objects seems to be one of the possible natural evolution of a paradigm that has remained stable in the last thousands of years

This approach has the additional advantage of bringing people away from virtual environments by putting closer to the real monuments that become a set of entry points for a labyrinth of knowledge on the history and culture.

The traditional archaeological guides, even when they are reorganized according to new technologies, remain prescriptive and rather intrusive objects. On the contrary, the model proposed in this paper is a sort of cultured automaton, a wise friend who, consulted in the right place, can lead the user on a cultural path that can radically change her the level of knowledge. From this point of view the tool described in these pages can become a useful educational tool.

The possibility of changing our awareness and knowledge of ancient and modern architecture can be realized by a smart and pragmatic use of artificial intelligence technologies. At the same time, the mixing of the disciplinary fields is also the starting point of a cultural transformation that no longer contemplates any difference between the humanistic, scientific and artistic disciplines.

VIII. REFERENCES

[1] https://www.tensorflow.org/tutorials.
[2] https://keras.io/
[3] https://pytorch.org/tutorials/
[4] https://developer.mozilla.org/en-US/
[5]https://www.coursera.org/specializations/deep-learning.
[6]https://www.coursera.org/learn/convolutional-neural-networks.