Indian Monument Recognition using CNN Algorithm

1st Neha Himesh
Department of Computer Science
Dayananda Sagar College Of Engineering
Bangalore, India

2nd Shriya R
Department of Computer Science
Dayananda Sagar College Of Engineering
Bangalore, India

3rd Gowthami PN
Department of Computer Science
Dayananda Sagar College Of Engineering
Bangalore, India

4th Benish Roshan
Department of Computer Science
Dayananda Sagar College Of Engineering
Bangalore, India

Abstract—Monument recognition is a challenging problem in the domain of image classification due to huge variations in the architecture of different monuments. Different orientations of the structure play an important role in the recognition of the monuments in their images. This paper proposes an approach for classification of various monuments using Convolutional Neural Network (CNN) Architecture using a GPU unit. The model is trained on representations of different Indian monuments, obtained from cropped images, which exhibit geographic and cultural diversity. Experiments have been carried out on the manually acquired dataset that is composed of 660 images of 10 different monuments where each monument has images from different angular views. The model was able to achieve accuracy above 90 percent for all the different monuments.
Index Terms—monument recognition;Convolutional Neural Networks; image classification

I. Introduction

A monument implies a structure that has been constructed in order to commemorate a person or an event. The term 'monument' is often applied to the buildings or structures that have been considered as examples of an important architectural and/or cultural heritage. The people belonging to the various cultures, castes, creeds and religions take pride in their culturally rich heritage bestowed upon them in the form of monuments. Monuments are also the tourist destinations in any country. They even are representations of great achievements present in art and architecture.It is therefore important to preserve them to help the present and the future generations understand and respect people who lived in different eras with different habits and traditions. Preservation of these monuments can happen by recognition of these different monuments and spreading the information about it to the other fellow citizens and tourists. A Convolutional Neural Network (ConvNet/CNN) is a Deep Learning algorithm which can take in an input image, assign importance (learnable weights and biases) to various aspects/objects in the image and be able to differentiate one from the other. The preprocessing required in a ConvNet is much lower as compared to other classification algorithms. While in primitive methods filters are hand-engineered, with enough training, ConvNets have the ability to learn these filters/characteristics. The architecture of a ConvNet is analogous to that of the connectivity pattern of Neurons in the Human Brain and was inspired by the organization of the Visual Cortex.

II. Architecture

A convolutional neural network consists of an input and an output layer, as well as multiple hidden layers. The hidden layers of a CNN typically consist of convolutional layers,activation function, pooling layers, fully connected layers and normalization layers.

A. Convolution layer

Convolution layer will compute the output of neurons that are connected to local regions input, each computing a dot product between their weights and a small region they are connected to in the input volume.

B. Pooling layer

Pooling layers reduce the dimensions of the data by combining the outputs of neuron clusters at one layer into a single neuron in the next layer. Local pooling combines small clusters,typically 2 x 2.

C. Fully Connected layer

Fully connected layers connect every neuron in one layer to every neuron in another layer. It is in principle the same as the traditional multi-layer perceptron neural network (MLP). The flattened matrix goes through a fully connected layer to classify the images.
The architecture of the CNN that we used is shown in the figure 1. Each images are resized to 170x170, is the first covoluted with 3x3 feature detectors to give 16 feature maps.

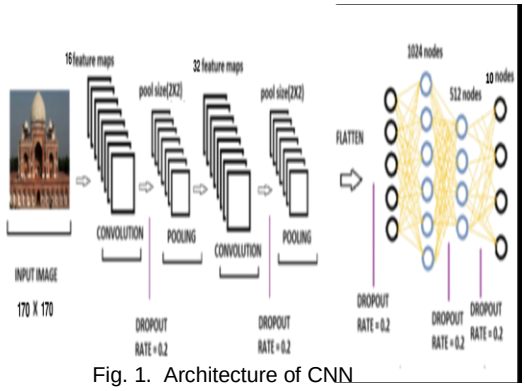

Fig. 1. Architecture of CNN

The first convoluted layer is followed by a dropout layer where it prevents in overfitting. The feature maps undergo max pooling .This is followed by another convolution layer outputting 32 feature maps, dropout layer and pooling layer follows. A single vector of pixel values is inputted to the Artificial Neural Network ,which is obtained from the pooled layer using flattening technique. Two fully-connected layers are used, one with 1024 nodes and other with 512 nodes along with dropout rate of 20

III. Dataset Description

A database comprising of 10 folders with each folder having around 66 images per monument is created, where each image is of size 170x170 pixels. These images are chosen from the online available dataset- https://goo.gl/ijKXY1 Mostly, the famous Indian monuments were considered for dataset. The naming of each folder is done according to the name which corresponds to the monument. The membership of each mon-ument in its respective category was verified using Wikipedia. The dataset is manually pruned to remove duplicates and incorrectly retrieved images. The dataset includes images of interior and exterior part of these monuments from varying an-gles and illumination to ensure that extracted features pertain only to the architecture of the monument. Furthermore, using both interior and exterior parts of monuments, the classifier can identify texture and wall art patterns that are distinctive features. Also, both high and lowresolution images have been included to make the dataset robust.

IV. System Implementation

The user is allowed to choose an image from the set of images. The image chosen is fed into CNN architecture. The image is analysed in the CNN architecture and the corresponding feature vector is formed. This feature vector uniquely identifies each image. This feature vector is classified as a corresponding monument and the result is displayed on the user interface. The user is allowed to then type a query to know more about the monument. The corresponding query is processed and the required answer is displayed on the user interface.The system is trained for default set of questions and corresponding answers to it. When the user types a question

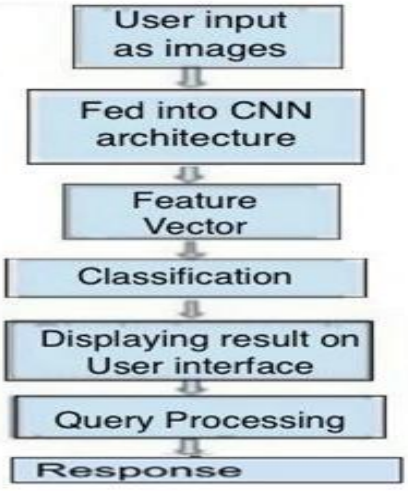

Fig. 2. Architecture of System Implementation

is corresponding answers from the set of trained answers is displayed. Thus, the monument is recognised and the user learns about the monument through its details. The architecture of system implementation is shown in figure 2.

V. Experimental Platform

The model was implemented on Intel(R) Core (TM) i3-7020U CPU with a 8GB of memory. The GPU consisted of 2GB of memory. The operating system was Microsoft windows 10.The software used is python IDE version 3.6.7.

VI. Experimental Analysis

Performance of the CNN network was analysed with dif-ferent settings of training epochs, activation function, number of convolution layers, dense layers, learning rate and change in the size of the images. The number of training epochs of the CNN network is one of the important aspects which affect the accuracy. A total of 20 training epochs were used in this model. A lesser number of epochs gave smaller accuracy and a larger number of epochs took longer time for execution with small increase in accuracy. After multiple experiments and trials, the number of epochs was chosen to be 20 as the higher accuracy did not give significant change in the accuracy. Learning rate was chosen to be 0.001. A total of 2 convolutional layers were used where the first convolu-tional layer consisted of 16 feature maps and the second convolutional layer consisted of 32 feature maps. The dropout rate was chosen to be 0.2 and the activation function was chosen to be relu because of the presence of more than two labels to be classified and the presence of non-negative values. The approach outperformed other existing approaches for monument recognition by a considerable margin. The model also performed well when used for classifying different styles of temple architectures.

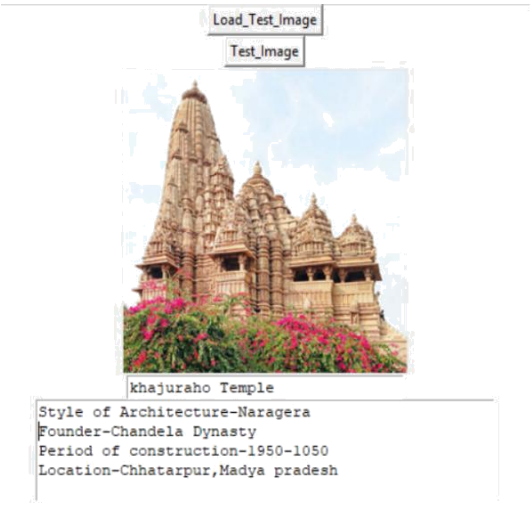

Fig. 3. Image of Khajuraho Temple

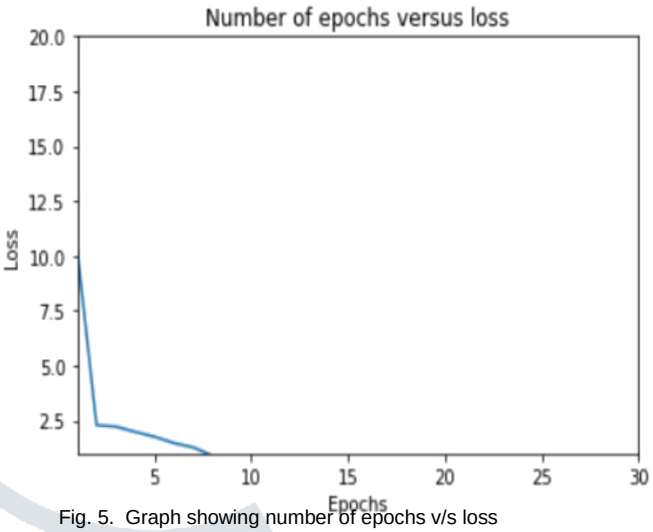

Fig. 5. Graph showing number of epochs v/s loss

Fig. 4. Image of Jama Masjid

```
epochs = 20
model.fit(X, y, epochs=epochs)
Epoch 12/20
660/660 [==============================] - 49s 74ms/step - loss: 0.1410 - acc: 0.9530
Epoch 13/20
660/660 [==============================] - 47s 71ms/step - loss: 0.1166 - acc: 0.9576
Epoch 14/20
660/660 [==============================] - 48s 72ms/step - loss: 0.1241 - acc: 0.9576
Epoch 15/20
660/660 [==============================] - 48s 73ms/step - loss: 0.0900 - acc: 0.9606
Epoch 16/20
660/660 [==============================] - 50s 76ms/step - loss: 0.1082 - acc: 0.9545
Epoch 17/20
660/660 [==============================] - 48s 73ms/step - loss: 0.0762 - acc: 0.9621
Epoch 18/20
660/660 [==============================] - 50s 76ms/step - loss: 0.0885 - acc: 0.9636
Epoch 19/20
660/660 [==============================] - 51s 77ms/step - loss: 0.0900 - acc: 0.9667
Epoch 20/20
660/660 [==============================] - 53s 80ms/step - loss: 0.0796 - acc: 0.9652
```

Fig. 6. The iteration rate and the accuracy percentage

VII. Output

The following are the few outputs obtained. The user when clicks on the button 'Load Test Image', a window opens from where the user is allowed to choose the required image for recognition. Once the image is chosen the button 'Test Image' is clicked. The image is displayed along with its label describing the name of the monument. Also, few attributes of the monument are displayed below the image.Thus the image is recognised. Figure 3 and 4 show the images of Khajuraho Temple and Jama masjid respectively which are being recognised by the model.

VIII. Results

Results The prediction percentage and the accuracy of the bounding boxes in the results depends on the following:
1.Batch size: It is the number of images that are trained per batch in one iteration of training. Batch sizes were not used in this experiment as the dataset was small. The batch size used here is 32. 2.Learning Rate: It is the training parameter that controls the size of weight and bias changes during learning. The learning rate used here was 0.001 3.Number of Iterations: It is the number of training iterations after which the network is optimally trained. The number of iterations used is 20. The following graph shows the loss rate with increase in number of epochs. As the number of epochs increase the loss rate during the training decreases. The ideal number of epochs is 20. The accuracy of the model is shown in the figure 5 with respect to number of epochs. As the number of epochs increase, the percentage of accuracy increases. The figure 6 shows a snap of the training displaying the number of epochs, accuracy and loss.

IX. Conclusion

The proposed system is based on Convolutional Neural Networks to detect and classify different types of monuments. The objective of the proposed system is achieved using the following modules. 1) Creating the customized dataset for the system consisting of images of all four types of classes. The four types of architecture is concentrated while creating the dataset. 2) The neural network is used consisting of totally 96 convolutional layers and a pool size of 2x2. 3) The system in implemented on a GPU having 2GB of memory. 4) The network was trained with different learning rates and for different number of iterations. Later, the weights with higher prediction percentage and lower error rate were used for testing. 5) The results showed the prediction of classes of the images. The network was able to retrieve features about it as well.

X. Future Work

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.

References

[1] P. Desai, J. Pujari, N. Ayachit, and V. K. Prasad. Classification of Archaeological monuments for different art forms with an application to cbir. In Advances in Computing, Communications and Informat-ics (ICACCI), 2013 International Conference on, pages 1108–1112. IEEE,2013.

[2] A. Ghosh, Y. Patel, M. Sukhwani, and C. Jawahar. Dynamic narratives for heritage tour. In European Conference on Computer Vision, pages 856–870. Springer, 2016.

[3] A. Goel, M. Juneja, and C. Jawahar. Are buildings only instances?: exploration in architectural style categories. In Proceedings of the Eighth Indian Conference on Computer Vision, Graphics and Image Processing, page 1. ACM, 2012.

[4] N. Hascoet and T. Zaharia. Building recognition with adaptive interest ¨ point selection. In Consumer Electronics (ICCE), 2017 IEEE Interna-tional Conference on, pages 29–32. IEEE, 2017.

[5] D. C. Hauagge and N. Snavely. Image matching using local symmetry features. In Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on, pages 206–213. IEEE, 2012.

[6] G. Kalliatakis and G. Triantafyllidis. Image based monument recognition using graph based visual saliency. ELCVIA, 12(2):88–97, 2013.

[7] Krizhevsky, A., Sutskever, I. and Hinton, G.E., 2012. Imagenet classifi-cation with deep convolutional neural networks. In Advances in neural information processing systems (pp. 1097-1105).

[8] Harel, J., Koch, C. and Perona, P., 2007. Graph-based visual saliency. In Advances in neural information processing systems(pp. 545-552).

[9] A. Goel, M. Juneja, and C. Jawahar. Are buildings only instances?: exploration in architectural style categories. In Proceedings of the Eighth Indian Conference on Computer Vision, Graphics and Image Processing, page 1. ACM, 2012]

[10] C. Doersch, S. Singh, A. Gupta, J. Sivic, and A. Efros. What makes paris look like paris? ACM Transactions on Graphics, 31(4), 2012.