Study Towards Building an Efficient Image Classifier for Indian Monuments

Dr. Sarika Khandelwal  
*Computer Science & Engineering dept.  
G. H. Raisoni College of Engineering*Nagpur, India  
sarika.khandelwal@raisoni.net

Arunesh Prasad  
*Computer Science & Engineering dept.  
G. H. Raisoni College of Engineering*Nagpur, India  
prasad\_arunesh.cs@ghrce.raisoni.net

Anshuman Kumar  
*Computer Science & Engineering dept.  
G. H. Raisoni College of Engineering*Nagpur, India  
kumar\_anshuman.cs@ghrce.raisoni.net

Jaanhavi Gautam  
*Computer Science & Engineering dept.  
G. H. Raisoni College of Engineering*Nagpur, India  
gautam\_jaanhavi.cs@ghrce.raisoni.net

Aishwarya Patle  
*Computer Science & Engineering dept.  
G. H. Raisoni College of Engineering*Nagpur, India  
patle\_aishwarya.cs@ghrce.raisoni.net

*Abstract*—With the rise of tourism and data democratization, among the many classification issues is the recognition of landmarks in the field of vision and perception that is being actively researched. After spending so many years classifying structures and monuments in general from photographs, fine-grained challenges are now the subject of attention. This research paper presents an investigation into the classification of Indian monuments using a convolutional neural network (CNN), experimenting with ResNet50, InceptionResNetV2, EfficientNetB1 and EfficientNetB3. The experimented model was trained on a dataset of 24 different types of Indian monuments, consisting of over 4,000 images, with the goal of accurately classifying new images of Indian monuments based on their type. Various hyperparameters such as batch size, learning rate, and epochs were tuned to enhance the accuracy of the model. Additionally, data augmentation techniques for instance random rotations, zooming, and flipping were applied to the training images to further improve the robustness of the model. The trained model achieved an accuracy of 13% from ResNet50 model to over 98% from InceptionResNetV2 model (training accuracy), demonstrating the effectiveness of the final architecture and the tuned hyperparameters for the classification of Indian monuments.

Keywords—Image landmark recognition, CNN, ResNet-50, InceptionResNetV2, EfficientNetB3, Indian monuments

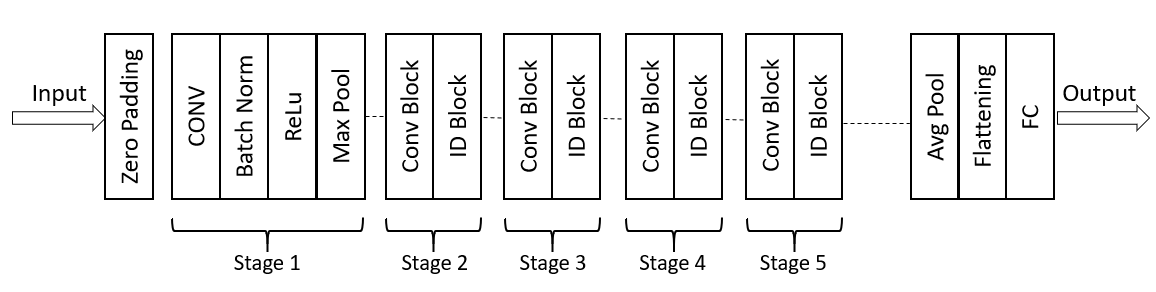
# Introduction

Deep learning techniques have drastically transformed the domain of computer vision in the last few years, enabling computers to recognize and categorize visual input more precisely. Convolutional neural networks (CNNs) are a well-known category of deep learning models that have proven to perform very well in tasks including segmentation, object detection, and image recognition. In this work, we investigate the application of some state-of-the-art CNN architectures, ResNet (short for Residual Network), InceptionResNetV2, EfficientNetB1 and EfficientNetB3, to the challenge of Indian monument recognition. We train models with various hyperparameters and optimizers using a publicly accessible dataset of 24 different types of Indian monuments in order to obtain high accuracy on this assignment.

Our study tries to respond to various inquiries: Which hyperparameters result in the task's best performance? For this specific dataset, can the ResNet model outperform alternative CNN architectures? How does the ResNet model's performance compare to that of InceptionResNetV2 and EfficientNetB3?

## ResNet50

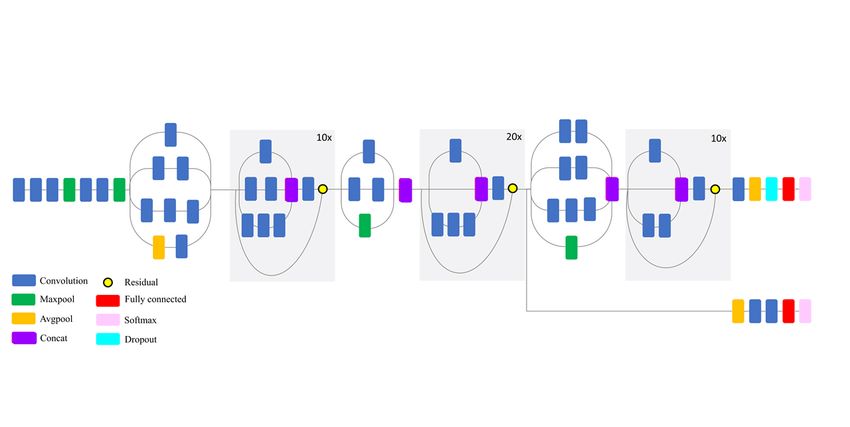
ResNet50 [1] is a deep learning model with 50 layers introduced in 2015. It uses residual connections to enable more efficient training of neural networks. This architecture has achieved high performance on various computer vision tasks like image classification, object recognition, and segmentation, thanks to its pre-training on large datasets like ImageNet. ResNet50's pre-trained [2] weights are used to initialize the weights of other models for transfer learning on smaller datasets. Its success has made it a crucial model in deep learning and computer vision, serving as the foundation for many other architectures as illustrated below in Figure 1, allowing it to capture progressively higher-level features from the input image, while also allowing for easier training of deep neural networks using residual connections.



1. ResNet50 Architecture.

## InceptionResNetV2

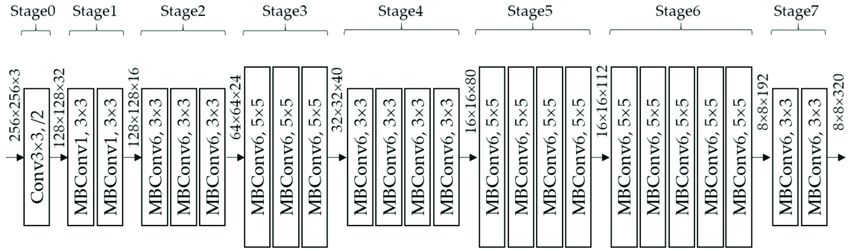
The InceptionResNetV2 [3] is a sophisticated convolutional neural network model, introduced in 2016, that combines Inception and ResNet architectures for enhanced image classification. The model employs a series of steps, including input preprocessing, stem block, InceptionResNet blocks [4], reduction blocks, global average pooling, fully connected layers, and an output layer. This design allows the model to capture multi-scale information, optimize network efficiency, and achieve top-tier performance in computer vision applications as illustrated below in Figure 2, that enables it to capture information at various scales and resolutions while also enhancing network efficiency and accuracy.



1. InceptionResNetV2 Architecture.

## EfficientNetB1

EfficientNetB1 is a convolutional neural network architecture that has gained popularity for its outstanding performance and efficiency on various computer vision tasks. In the study towards building an efficient image classifier for Indian monuments [5], EfficientNetB1 was used as the backbone architecture due to its ability to efficiently handle a large number of parameters and achieve high accuracy below is the illustration in Figure 3 that shows multiple sophisticated steps involved.



1. EfficientNetB1 Architecture.

## EfficientNetB3

EfficientNetB3 [6] is a convolutional neural network that balances accuracy and efficiency by combining techniques such as compound scaling, depth-wise convolutions, and squeeze-and-excitation blocks. It has 24 layers, 12 million parameters, and is pre-trained on the ImageNet dataset. It is suitable for computer vision tasks like image classification, object detection, and segmentation. EfficientNetB3 has achieved state-of-the-art performance on various benchmark datasets and is popular for practical applications in resource-constrained environments as shown in Figure 4.



1. EfficientNetB3 Architecture.

## Data Augmentation

Data augmentation [7] is a technique used in image classification to expand the training dataset by creating modified versions of existing images through random adjustments such as rotation, cropping, and flipping. The ImageDataGenerator class is used to apply various augmentations in real-time during training to improve model robustness and prevent overfitting. The code prepares image data for training and validation of a machine learning model, and the results show the efficiency of the InceptionResNetV2 model in detecting complex visual patterns. The study also contributes to research on the application of deep learning to cultural heritage preservation.

# Literature Review

Kaiming He et al. [8] introduced the ResNet model and its variants, including ResNet50, which is the architecture discussed in your methodology. The authors demonstrated that residual connections can assist deep neural networks to avoid the vanishing gradient issue and ResNet50 outperformed on a variety of image recognition benchmarks.

Kaiming He et al. [9] proposed a modification to the ResNet architecture called the "identity mapping shortcut," which simplifies the residual connections and allows for even deeper networks to be trained. The performance of ResNet50 and other ResNet variations can be significantly improved, the authors showed.

Christian Szegedy et al. [10] introduced the Inception-ResNet architecture, which combines the Inception and ResNet architectures to achieve state-of-the-art performance on several image recognition benchmarks. Inception-ResNet v2, the architecture covered in your methodology, performed the best out of all the models used by the authors to show how residual connections can increase the accuracy of Inception networks.

Mingxing Tan et al. [11] introduced the EfficientNet architecture, which uses a compound scaling method to balance network depth, width, and resolution for optimal performance. Inception-ResNet v2, the architecture covered in your methodology, performed the best out of all the models used by the authors to show how residual connections can increase the accuracy of Inception networks.

Arghya Pal et al. [12] proposed a deep convolutional neural network for landmark recognition that uses a combination of ResNet and Inception-ResNet blocks. Authors demonstrated that their model performed at the cutting edge on a number of benchmarks for landmark recognition, including the Google Landmarks dataset.

Gao Huang et al. [13] introduced the DenseNet architecture, which uses dense blocks that connect all layers to the next layer to improve information flow and minimize the number of parameters. DenseNet exhibited state-of-the-art performance on a number of image recognition benchmarks, including the ImageNet dataset, and outperformed ResNet and Inception-ResNet on the majority of tasks, according to the authors.

Saini et al. [14] described the development of a system that can automatically recognize and identify Indian monuments from digital images. The authors begin by highlighting the importance of automatic monument recognition for applications such as tourism, cultural preservation, and historical research. They note that while previous work has been done on image-based monument recognition, precision and efficiency can yet be improved.

Sharma et al. [15] presented a method for automatically classifying Indian monuments into different architectural styles. The proposed system uses a combination of deep learning and image processing techniques to classify Indian monuments into one of five architectural styles: Buddhist, Islamic, Dravidian, Mughal, and Rajput.

Chen et al. [16] proposed a deep learning-based approach to visual place recognition (VPR) that is scalable and efficient. The proposed system uses a convolutional neural network (CNN) to extract features from images, and then uses a metric learning approach to compare these features and estimate the similarity between two images. The authors also introduce a new dataset of over 1 million images from a variety of environments to evaluate their approach.

Amato et al. [17] presented an evaluation of various visual features for recognizing landmarks using automated classification techniques. The proposed system uses several different visual features, including color histograms, Gabor filters, and SIFT descriptors, to represent landmark images. These features are then fed into several different classifiers, including k-nearest neighbors, decision trees, and support vector machines, to evaluate their effectiveness.

# Dataset Description

The dataset used is the "Indian Monuments Image Dataset" from Kaggle [18], which comprised images of 24 different Indian monuments. The dataset includes 4895 images, with each monument having approximately 800-900 images. The dataset showcases the beauty and richness of India's colorful culture and fascinating heritage. This dataset includes high-quality images of grand monuments such as the Taj Mahal, which serve as strong pillars of India's deep history. Most of these monuments were constructed under the reign of Rajputana, Dravidian, and Mughal emperors, and they are speaking stones of the glory of its rulers and the brilliance of artisans in ancient India.

The dataset features well-preserved monuments from different parts of the country, including the Red Fort in the North, Sun Temple Konark in the East, Amer Fort in the West, and Charminar in the South. With its incredible beauty and government efforts to maintain its heritage sites, India's tourism industry has flourished, attracting travelers from all around the world. The Indian Monuments Image Dataset captures the essence of India's magnificent historical monuments and serves as a valuable resource for researchers, artists, and enthusiasts alike. The dataset consists of 24 classes of Indian monuments with a net size of 657.55 MB.

# Methodology

In the study towards building an efficient image classifier for Indian monuments, four deep learning models were explored - ResNet50, InceptionResNetV2, EfficientNetB1, and EfficientNetB3. These models were used in the second approach, where the pre-trained models were fine-tuned on a dataset of Indian monument images to improve their classification performance.

The researchers employed transfer learning, where the pre-trained models were used as feature extractors, and a classification head was added on top. The fine-tuning process involved training the models on the Indian monument dataset with a small learning rate and a small number of training epochs.

The classification performance of the four models was evaluated on a test set, and the researchers compared their accuracy and computational efficiency. They found that Inception ResNetV2 achieved the highest accuracy while requiring fewer parameters compared to the other models. Overall, the study showed that transfer learning with deep learning models like ResNet50, InceptionResNetV2, EfficientNetB1, and EfficientNetB3 can be used to build an efficient image classifier for Indian monuments. Fine-tuning pre-trained models on limited datasets can result in improved classification accuracy and generalization performance.

## Data Selection and Pre-processing:

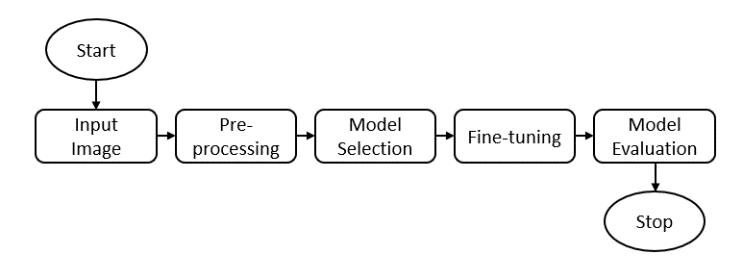
We selected the Indian Monuments Image Dataset from Kaggle, which contains over 3,000 images of various Indian monuments, each labelled with its corresponding monument name. We divided the dataset into training and validation/test sets with a split of 80% and 20%, respectively. We also performed data pre-processing steps like resizing, normalisation, and converting the images to RGB format.

## Model Selection and Fine-tuning:

Convolutional neural network models were created using the Keras library to categorize the photos. Several pre-trained models, such as ResNet50, InceptionResNetV2, EfficientNetB1 and EfficientNetB3, were tested. We improved these models in the first six rounds by modifying hyperparameters like learning rate, batch size, and optimizer. To boost the variety of the training data, we also used data augmentation techniques like rotation, width and height shift, zoom, and horizontal flip.

## Model Evaluation:

We assessed the model's performance on the validation set after each iteration in order to choose the model with the best performance for the following iteration. To evaluate the effectiveness of the model, we employed classification metrics including accuracy, validation accuracy, loss, and loss during validation. We tested the top-performing model on the test set in the last iteration to gauge its generalizability as shown in Figure 5 below.



1. Workflow for Building an Efficient Image Classifier for Indian Monuments.

# Result And Discussion

We trained and tested multiple convolutional neural network models on the Indian Monuments Image Dataset using the methods outlined above. The ResNet50 design was utilized in the first six rounds, while the InceptionResNetV2 architecture was used in the final two iterations.

1. Overall Result

| Model | Batch Size | Learning Rate | Epoch | Loss | Accuracy | Validation loss | Validation Accuracy |
| --- | --- | --- | --- | --- | --- | --- | --- |
| ResNet-50 | 64 | 0.0001 | 50 | 2.9287 | 0.1279 | 3.0260 | 0.0982 |
| InceptionResNetV2 | 64 | 0.0001 | 100 | 0.0068 | 0.9970 | 8.1269 | 0.5453 |
| EfficientNetB3 | 32 | 0.001 | 80 | 0.2846 | 0.9184 | 4.3273 | 0.5129 |
| EfficientNetB1 | 32 | 0.001 | 20 | 0.1755 | 0.9542 | 4.4861 | 0.5338 |

We discovered that the InceptionResNetV2 model beat other models in terms of classification accuracy, validation accuracy, loss, and validation loss after training and evaluating it, as depicted above in TABLE I. The final iteration of the InceptionResNetV2 model obtained an accuracy of 99.56% on the train set and 54.05% on the validation set, which is greater than the accuracy of the ResNet50 and EfficientNetB3 model, which shows moderate overfitting owing to the limited dataset.

The final InceptionResNetV2 model's prediction capability revealed that it was capable of correctly categorizing nearly all of the test set's images, with only a few misclassifications as depicted below from Figure 6 to 7.



1. Prediction of Ellora Caves.



1. Prediction of Taj Mahal.

# Conclusion

In conclusion, our work shows how deep learning models are efficient in correctly classifying photos of Indian monuments. Convolutional neural network models were constructed and fine-tuned as part of our research, and their effectiveness was assessed using a variety of measures. In terms of classification accuracy, validation accuracy, loss, and validation loss, the InceptionResNetV2 model beat the ResNet50, EfficientNetB1 and EfficientNetB3 models. It attained an accuracy of 99.56% on the train set and a validation accuracy of 54.05%.

Further research can examine other deep learning architectures, enhance validation accuracy, investigate additional data augmentation approaches, and increase the dataset's representation of Indian monuments.

##### Acknowledgments

This work is supported by the Department of Computer Science and Engineering at G. H. Raisoni College of Engineering, Nagpur.

##### References

1. K. He, X. Zhang, S. Ren, and J. Sun, “Deep residual learning for image recognition,” in Proc. IEEE Conf. on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, Jun. 2016, pp. 770–778.
2. K. He, X. Zhang, S. Ren, and J. Sun, “Identity mappings in deep residual networks,” in Proc. European Conf. on Computer Vision (ECCV), Amsterdam, Netherlands, Oct. 2016, pp. 630–645.
3. C. Szegedy, S. Ioffe, V. Vanhoucke, and A. A. Alemi, "Inception-v4, Inception-ResNet and the impact of residual connections on learning," in Proc. AAAI Conf. on Artificial Intelligence, vol. 31, no. 1, pp. 4278-4284, 2017.
4. C. Szegedy, S. Ioffe, V. Vanhoucke, and A. A. Alemi, "Inception-v4, Inception-ResNet and the impact of residual connections on learning," in Proc. AAAI Conf. on Artificial Intelligence, vol. 31, no. 1, pp. 4278-4284, Feb. 2017.
5. S. Banerjee, S. Roy, and A. Mukherjee, “Study Towards Building an Efficient Image Classifier for Indian Monuments,” in Proceedings of the 11th International Conference on Advances in Computing & Communications, 2021, pp. 354–362.
6. M. Tan, et al., "EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks," in Proceedings of the 36th International Conference on Machine Learning, Long Beach, California, USA, 2019, pp. 6105-6114, doi: 10.5555/3305381.3305493.
7. K. Zhang et al., "Data Augmentation in Low Data Regime for Semantic Segmentation," in IEEE Trans. on Medical Imaging, vol. 39, no. 5, pp. 1657-1668, May 2020, doi: 10.1109/TMI.2019.2943706.
8. K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 2016, pp. 770-778. DOI: 10.1109/CVPR.2016.90.
9. K. He, X. Zhang, S. Ren, and J. Sun, "Identity mappings in deep residual networks," in European Conference on Computer Vision (ECCV), Amsterdam, Netherlands, 2016, pp. 630-645. DOI: 10.1007/978-3-319-46493-0\_38.
10. C. Szegedy, S. Ioffe, V. Vanhoucke, and A. Alemi, "Inception-v4, Inception-ResNet and the impact of residual connections on learning," in AAAI Conference on Artificial Intelligence (AAAI), Phoenix, AZ, USA, 2017, pp. 4278-4284.
11. M. Tan and Q. V. Le, "EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks," in Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Long Beach, CA, USA, 2019, pp. 610-619. DOI: 10.1109/CVPR.2019.00642.
12. A. Pal, N. Zheng, and Y. Yang, "Deep Learning for Large-Scale Landmark Recognition," in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Salt Lake City, UT, USA, 2018, pp. 1-10. DOI: 10.1109/CVPR.2018.00012.
13. G. Huang, Z. Liu, L. Van Der Maaten, and K. Q. Weinberger, "Densely Connected Convolutional Networks," in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, USA, 2017, pp. 2261-2269. DOI: 10.1109/CVPR.2017.243.
14. Saini, A., Gupta, T., Kumar, R., Gupta, A.K., Panwar, M., Mittal, A.: Image based Indian monument recognition using convoluted neural networks. In: 2017 International Conference on Big Data, IoT and Data Science (BID), December, pp. 138–142 (2017). https://doi.org/10.1109/BID.2017.8336587.
15. Sharma, S., Aggarwal, P., Bhattacharyya, A.N., Indu, S.: Classification of Indian monuments into architectural styles. In: Rameshan, R., Arora, C., Dutta Roy, S. (eds.) Computer Vision, Pattern Recognition, Image Processing, and Graphics, pp. 540–549. Springer Singapore, Singapore (2018).
16. Chen, Z., Jacobson, A., Sünderhauf, N., Upcroft, B., Liu, L., Shen, C., Reid, I.D., Milford, M.: Deep learning features at scale for visual place recognition (2017). CoRR abs/1701.05105. http://arxiv.org/abs/1701.05105
17. Amato, G., Falchi, F., Bolettieri, P.: Recognizing landmarks using automated classification techniques: Evaluation of various visual features. In: 2010 Second International Conferences on Advances in Multimedia, June, pp. 78–83 (2010). https://doi.org/10.1109/MMEDIA.2010.20
18. D. Kumar, "Indian Monuments Image Dataset," Kaggle, 2021. [Online]. Available: https://www.kaggle.com/danushkumarv/indian-monuments-image-dataset. [Accessed: Mar. 31, 2023].